Ideals, Gröbner Bases, and PCPs

Prashanth Amireddy[*]   Amik Raj Behera[†]   Srikanth Srinivasan [‡]   Madhu Sudan[§]

Sophus Valentin Willumsgaard [¶]

November 6, 2025

Abstract

All known proofs of the PCP theorem rely on multiple "composition" steps, where PCPs over large alphabets are turned into PCPs over much smaller alphabets at a (relatively) small price in the soundness error of the PCP. Algebraic proofs, starting with the work of Arora, Lund, Motwani, Sudan, and Szegedy use at least 2 such composition steps, whereas the "Gap amplification" proof of Dinur uses $\Theta(\log n)$ such composition steps. In this work, we present the first PCP construction using just one composition step. The key ingredient, missing in previous work and finally supplied in this paper, is a basic PCP (of Proximity) of size $2^{n^{\varepsilon}}$, for any $\varepsilon > 0$, that makes $\mathcal{O}_{\varepsilon}(1)$ queries.

At the core of our new construction is a new class of alternatives to "sum-check" protocols. As used in past PCPs, these provide a method by which to verify that an m-variate degree d polynomial P evaluates to zero at every point of some set $S \subseteq \mathbb{F}_q^m$. Previous works had shown how to check this condition for sets of the form $S = H^m$ using $\mathcal{O}(m)$ queries with alphabet $\mathbb{F}_q^d$ assuming $d \geqslant |H|$. Our work improves this basic protocol in two ways: First we extend it to broader classes of sets S (ones closer to Hamming balls rather than cubes). Second, it reduces the number of queries from $\mathcal{O}(m)$ to an absolute constant for the settings of S we consider. Specifically when $S = (\{0,1\}_{\leqslant 1}^{m/c})^c$, where $T = \{0,1\}_{\leqslant b}^a \subseteq \mathbb{F}_q^a$ denotes the set of Boolean vectors of Hamming weight at most b in $\mathbb{F}_q^a$, we give such an alternate to the sum-check protocol with $\mathcal{O}(1)$ queries with alphabet $\mathbb{F}_q^{\mathcal{O}(c+d)}$, using proofs of size $q^{\mathcal{O}(m^2/c)}$. Our new protocols use insights from the powerful theory of Gröbner bases to extend previously known protocols to these new settings with surprising ease. In doing so, they highlight why these theories from algebra may be of further use in complexity theory.

Contents

1 Introduction

In this paper, we give a new general framework for constructing algebraic PCPs that leads to the first proof of the PCP theorem using only one "composition" step. Starting with the work of Arora and Safra [AS98], composition of PCPs has been a key ingredient in all previous PCP constructions. The original proof of the PCP theorem due to Arora, Lund, Motwani, Sudan, and Szegedy [ALMSS98] used two composition steps, while the novel alternate proof due to Dinur [Din07] uses $O(\log n)$ composition steps. Compositions improve various parameters of the construction at the cost of making the verifier less transparent. So it is a natural and long-sought goal to try to minimize the number of composition steps (to one, or even zero!). We achieve the weaker goal here.

The key to our construction is a new class of protocols replacing the "sum-check" protocol in PCP constructions. The sum-check protocol, due to Lund, Fortnow, Karloff, and Nisan [LFKN92] (also used in proofs of IP=PSPACE [Sha92] and MIP=NEXPTIME [BFL91]) has been a central ingredient in previous PCP constructions. In PCP constructions, the protocol is used to establish that an m-variate polynomial P over $\mathbb{F}_q$ given as an oracle from $\mathbb{F}_q^m \to \mathbb{F}_q$ is identically zero on the set $\{0,1\}^m \subseteq \mathbb{F}_q^m$ or more generally, on some set of the form H^m for $H \subseteq \mathbb{F}_q$. We refer to this latter task as "zero-on-variety testing". (The reason for the use of the term "variety" to describe the set H^m will become clearer later.) Other than the sum check protocol, the only other direct protocol for zero-on-variety testing is a protocol due to Ben-Sasson and Sudan [BS08], which also only works for varieties of the form H^m. In this work, we establish a new connection between the theory of Gröbner bases and the zero-on-variety test of [BS08] that allows us to get efficient zero-on-variety tests for a much broader class of varieties, including some varieties that are close to Hamming balls of constant radii. This latter setting which had eluded previous works and is key to our PCP construction.

Armed with this new class of protocols, we show how to significantly simplify the ALMSS PCP construction. We start by giving a new PCP construction that works relative to any variety $V \subseteq \mathbb{F}_q^m$ with performance depending on the "Gröbner basis complexity" of the variety, a notion we define. We then show how to specialize this PCP in two different ways by using two different varieties — the first giving the usual $\mathcal{O}(\log n)$ randomness and $\text{poly}(\log n)$ query PCP for NP, and the second giving an $\mathcal{O}(n^\varepsilon)$ randomness and $\mathcal{O}_\varepsilon(1)$ query PCP for NP, for any $\varepsilon > 0$. No natural PCP (built without composition) was known with the latter setting of parameters, and indeed, this has been the key bottleneck in reducing the number of composition steps in ALMSS. We stress that both our ingredient PCPs are instantiations of the same protocol — only the choice of the variety is different (and the analysis of the Gröbner basis complexity of the varieties is straightforward)! And furthermore, our PCPs are already "robust assignment testers" in the sense of Dinur and Reingold [DR06] (or equivalently, Robust PCPs of Proximity in the sense of Ben-Sasson, Goldreich, Harsha, Sudan and Vadhan [BGHSV06]) and thus immediately composable. Putting our two robust assignment testers together yields our final PCP. The resulting proof thus gives the following simplifications to the ALMSS protocol: It eliminates one composition step, it eliminates the need for the "Hadamard PCP" entirely, and it eliminates the need for the "parallelization/robustification" step in ALMSS [ALMSS98, Section 7].

In what follows, we describe our work in greater detail, starting with the basic notion of interest in this paper, the zero-on-variety testing problem, and the resulting PCPs.

1.1 Overview of Our Construction

PCPs of Proximity The central objects of interest in this paper are best described by the umbrella term "PCP of Proximity" (or equivalently "assignment tester"). Here a verifier $\mathcal{V}$ is given oracle access to some oracle $f : D \to \Sigma$ along with a proof oracle $\pi : S \to \Gamma$ where D, S, Σ and Γ are finite sets. A verifier for a property $\mathcal{F} \subseteq \{g : D \to \Sigma\}$ queries (f, π) and renders a verdict `Accept`/`Reject`, with the property that if $f \in \mathcal{F}$ then there exists a π such $\mathcal{V}^{f,\pi}$ always outputs `Accept`, while for $f \notin \mathcal{F}$ we have that for every proof π, $\mathcal{V}^{f,\pi}$ outputs `Reject` with probability[1] $\Omega(\delta_{\mathcal{F}}(f))$. The key parameters associated with the verifier are its randomness (usually $\mathcal{O}(\log|S|)$), its locality (or query complexity) ℓ which is the total number of queries to f and π, and the alphabet size $a = \max\{\log|\Sigma|, \log|\Gamma|\}$.

A good example of a PCP of Proximity is the low-degree test. Here $D = \mathbb{F}_q^m$ and $\Sigma = \mathbb{F}_q$. The property $\mathcal{F}_{m,d,q}$ is the set of evaluations of all m-variate polynomials over $\mathbb{F}_q$ of total degree at most d. When $d \leqslant q$ the best known low-degree tests achieve randomness of $\mathcal{O}(m \log q)$, locality $\ell = 2$, and alphabet size $d \log q$. (See Theorem 3.5.) The fact that the locality is a constant and d and m affect only the randomness and alphabet size is important in their use in PCPs.

Zero-on-variety testing The zero-on-variety testing problem is also a PCP of Proximity problem. It is described by parameters $\mathbb{F}_q, d, m$ and a variety $V \subseteq \mathbb{F}_q^m$. Here the verifier is given oracle access to a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ that is promised to be a degree d polynomial, and goal is to test for the property $\mathcal{F}_{V,d}$ that is the set of all degree d polynomials that are identically zero on V, or equivalently if the polynomial f lies in the ideal of polynomials $\mathbb{I}(V)$ vanishing on V. (In this description, we opt to describe this as a promise problem - though in PCP applications, the non-promised version of this problem is the one used. The two become essentially equivalent thanks to the existence of low-degree tests.)

Prior to this work, the only natural zero-on-variety tests considered the setting where $V = H^m$ for some $H \subseteq \mathbb{F}_q$. The protocol given by [BS08] uses the following identity. $f \in \mathbb{I}(H^m)$ if and only if there exist polynomials $f_1, \ldots, f_m$ of degree at most $d - |H|$ such that

$$f(X) = \sum_{i=1}^{m} f_i(X) Z_H(X_i), \tag{1}$$

where $Z_H(Y) = \prod_{\alpha \in H}(Y - \alpha)$ is the canonical univariate polynomial that vanishes on H. The identity above follows from Alon's Combinatorial Nullstellensatz [Alo99] and leads to a tester as follows: The zero-on-variety tester for H^m expects oracles for $f_1, \ldots, f_m$ as proof. It verifies using the low-degree test that each of these oracles has degree at most $d - |H|$ and then verifies Equation (1) for a random choice of $X = (a_1, \ldots, a_m)$. Modulo further details (involving local correction), this leads to an $\mathcal{O}(m)$ local tester with randomness $\mathcal{O}(m \log q)$ and alphabet size $\mathcal{O}(d \log q)$.

For our purposes, this choice of variety is insufficient. (Furthermore, the dependence of the locality on m is also problematic, but we'll address this later.) To remedy this, we use an alternate interpretation of the identity above. In this interpretation the identity holds because $\{Z_H(X_1), \ldots, Z_H(X_m)\}$ form a "Gröbner basis" of the ideal $\mathbb{I}(H^m)$ under a "graded monomial

---

[1] Here $\delta_{\mathcal{F}}(f) = \min_{g \in \mathcal{F}}\{\delta(f, g)\}$ and $\delta(f, g) = \Pr_{x \in D}[f(x) \neq g(x)]$ measure the distance of f from $\mathcal{F}$ is the normalized Hamming metric.

4

ordering". We won't define the exact notion of a Gröbner basis under different monomial orderings here — we don't need to. The notion that suffices for us is a notion we call a Gröbner generating set G of an ideal $\mathbb{I}$: G is a *Gröbner generating set* for $\mathbb{I}$ if for all polynomials $P \in \mathbb{I}$ there exist polynomials $h_g, g \in G$ such that $P = \sum_{g \in G} h_g \cdot g$ and $\deg(h_g \cdot g) \leqslant \deg(P)$ for every $g \in G$. And the above identity is the special case where $V = H^m$ with $\{Z_H(X_1), \ldots, Z_H(X_m)\}$ as the Gröbner generating set. The zero-on-variety test of [BS08] can now be extended to any variety that has a "small" Gröbner generating set. Indeed, this motivates our notion of the Gröbner complexity of a variety V (see Definition 4.2 for a formal definition) — which is the size of the smallest Gröbner generating set G of $\mathbb{I}(V)$. One crucial example for us is the following: The set of polynomials $\{X_i X_j | 1 \leqslant i < j \leqslant m\} \cup \{X_i^2 - X_i | i \in [m]\}$ form a Gröbner generating set for the variety $\{0,1\}_{\leqslant 1}^m$ consisting of Boolean points of Hamming weight at most 1 in $\mathbb{F}_q^m$. (Thus the variety $\{0,1\}_{\leqslant 1}^m$ has Gröbner complexity $\mathcal{O}(m^2)$.) We also use the following basic property of Gröbner generating sets: If $G_X \subseteq \mathbb{F}_q[X]$ is a Gröbner generating set for variety V_1 and $G_Y \subseteq \mathbb{F}_q[Y]$ is one for V_2 then $G_X \cup G_Y \subseteq \mathbb{F}_q[X,Y]$ is a Gröbner generating set for $V_1 \times V_2$. In particular this establishes that the Gröbner complexity of $(\{0,1\}_{\leqslant 1}^m)^c \subseteq \mathbb{F}_q^{mc}$ is $\mathcal{O}(cm^2)$. Applying the [BS08] protocol to Gröbner generating sets now gives us a $\mathcal{O}(k)$ query protocol for testing zero-on-V for variety V of Gröbner complexity k.

While this now gives many new varieties that have natural zero-on-variety tests, the locality of $\mathcal{O}(k)$ can be prohibitive. Our second contribution here is to give a new protocol to test this, that shifts the Gröbner complexity into the randomness of the protocol and achieves (a universal) constant locality. Specifically our verifier now expects an oracle for $\mathcal{M}_V(P)(X,Y) = \sum_{g \in G} Y_g h_g(X)$ where $Y = (Y_g | g \in G)$ is a new set of k variables. Specifically performing a low-degree test on $\mathcal{M}_V(P)$ along with a test that verifies $M_V(P)(X,0) \equiv 0$ ensures that $\mathcal{M}_V(P)$ is effectively giving access to all linear combinations of $h_g(X)$ with *constant* query complexity. Testing the identity $P(X) = \sum_{g \in G} h_g(X)g(X)$ at a random choice of X now requires only one query to P and one to $M_V(P)$! Modulo some standard use of self-correction, this gives us an $\mathcal{O}(1)$ locality protocol for zero-on-V testing with alphabet size $\mathcal{O}(d \log q)$ and randomness $\mathcal{O}((k+m) \log q)$ for any variety V of Gröbner complexity k. (See Section 5.)

PCPs from Zero-on-variety tests It is straightforward to build PCPs for NP-hard problems from low-degree tests and zero-on-variety tests. (Recall that a PCP verifier for say graph coloring is given as input a graph G and oracle access to a purported proof π with the feature that if G is 3-colorable that there exists a π such that V always accepts whereas if G is not 3-colorable then V rejects every proof π w.p. at least $1/2$. The parameters of interest to us are the same — the randomness, the locality, and the alphabet size of the proof.)

For example, the 3-coloring protocol, based on a similar proof from [BS08], goes as follows: Fix an odd prime power q.[2] For a variety $V \subseteq \mathbb{F}_q^m$ let its "extension degree" be the least integer d such that every function $f : V \to \mathbb{F}_q$ can be extended to a degree d polynomial in $\mathbb{F}_q[X_1, \ldots, X_m]$. Now, given a variety $V \subseteq \mathbb{F}_q^m$ of Gröbner complexity k and extension degree d we consider the 3-coloring problem on the vertex set V (same V). Note that the graph is given by an edge function $E : V \times V \to \{0,1\}$ which can be shown to be extendable to a degree $2d$ polynomial $\hat{E}$ from $\mathbb{F}_q^{2m} \to \mathbb{F}_q$. A proof that E is 3-colorable includes polynomials $\chi : \mathbb{F}_q^m \to \mathbb{F}_q$, $A : \mathbb{F}_q^m \to \mathbb{F}_q$ and

---

[2] We do this for simplicity here and allow us to assume $\{-1, 0, 1\} \subseteq \mathbb{F}_q$ can be used to represent 3 distinct colors. The protocol easily extends to other fields using some other set of 3 distinct elements of $\mathbb{F}_q$.

$B : \mathbb{F}_q^{2m} \to \mathbb{F}_q$ satisfying (1) $A(X) = \chi(X) \cdot (\chi(X) - 1) \cdot (\chi(X) + 1)$, (2) A is zero-on-V, (3) $B(X, Y) = \hat{E}(X, Y) \prod_{i \in \{-2,-1,1,2\}}(\chi(X) - \chi(Y) - i)$ and (4) B is zero on $V \times V$. (Items (1) and (2) verify that χ is a 3-coloring of V with color set $\{-1, 0, 1\}$, while items (3) and (4) verify that χ is a valid coloring of the edges of E.) The V-verifier performs low-degree tests on all the four oracles and then tests identities (1) and (3) by picking a random value of the variables, and finally verifies items (2) and (4) using zero-on-V and zero-on-V^2 tests. By the aforementioned properties on Gröbner complexity and standard facts about extension degree we get that this PCP verifier achieves $\mathcal{O}(1)$ locality with randomness $\mathcal{O}((k + m) \log q)$ and alphabet $\mathcal{O}(d \log q)$ (matching those of the zero-on-V tests up to constant factors).

Instantiating the verifier above with $V = H^m$ where $|V| = n$, $|H| = \log n$, $m = \mathcal{O}(\log n / \log \log n)$ and $q = \mathcal{O}(\log^6 n)$ gives an $\mathcal{O}(1)$ locality PCP verifier for 3-coloring of n vertex graphs with randomness $\mathcal{O}(\log n)$ and alphabet size $\mathcal{O}(\text{poly} \log n)$. But a different instantiation *of the same PCP* with $m = n^\varepsilon$, $c = \frac{1}{\varepsilon}$, $q = \mathcal{O}_\varepsilon(1)$ and $V = (\{0, 1\}_{\leqslant 1}^m)^c$ gives an $\mathcal{O}(1)$ locality PCP verifier with randomness $O(n^{2\epsilon})$ and alphabet size $O_\epsilon(1)$! We note that even using $c = 1$ gives a completely new protocol matching the parameters of the Hadamard PCP in [ALMSS98, Section 5]. And using larger values of c gives us our new protocols. (See Section 6 for details.)

Furthermore, these PCPs are easily converted to "Robust PCPs of Proximity" (or "Robust assignment testers") in the sense of [DR06; BGHSV06] of constant robustness — since our PCPs have constant locality. This allows us to compose the PCPs above in a single composition step to get an $\mathcal{O}(1)$ locality PCP verifier with $\mathcal{O}(\log n)$ randomness and $\mathcal{O}(1)$ alphabet size — and thus the PCP theorem. (See Section 7 for details.)

2 Formal Statement of Our Results

We first introduce basic definitions needed to state our main result. Throughout this document, $\mathcal{V}^\Pi$ means that the algorithm/circuit/verifier has oracle access to the string Π, i.e., $\mathcal{V}$ can query $\Pi[i]$ for any $1 \leqslant i \leqslant |\Pi|$. We use the notation $\mathcal{V}^\Pi(x; R)$ to say that the algorithm $\mathcal{V}$ has oracle access to Π, has input x, and access to a random string R. In this notation, $\mathcal{V}^\Pi(x; R)$ is a deterministic algorithm and the randomness is in the choice of R.

2.1 PCPs

Definition 2.1 (Standard Verifier). *For functions $r, \ell, a : \mathbb{Z}^{\geqslant 0} \to \mathbb{Z}^{\geqslant 0}$, define a (r, ℓ, a)-standard verifier $\mathcal{V}$ as follows:*
Let $\Sigma = \{0, 1\}^{a(n)}$. On input $x \in \{0, 1\}^n$ of length n, a string[3] $R \in \{0, 1\}^{r(n)}$, and oracle access to a string $\Pi \in \Sigma^{\text{size}(n)}$ (i.e. Π is a string of length $\text{size}(n)$ on alphabet $\{0, 1\}^{a(n)}$), we have:

- *$\mathcal{V}^\Pi(x; R)$ outputs a subset $Q \subseteq [\text{size}(n)]$ of cardinality $\ell(n)$.*

- *$\mathcal{V}^\Pi(x; R)$ outputs a Boolean circuit $\mathcal{C}$ (depends on x and R) depending on $\ell(n) \cdot a(n)$ bits. The circuit C gets access to entries of Π as bits of length $a(n)$.*

- *$\mathcal{V}^\Pi(x; R)$ returns `Accept` if $C(\Pi|_Q) = 1$ and returns `Reject` if $C(\Pi|_Q) = 0$.*

---

[3] This string R is the random string fed into $\mathcal{V}$.

The maximum circuit size $|\mathcal{C}|$ over every possible choice of (x, R) will be referred to as the size of the standard verifier $\mathcal{V}$. The running time of the standard verifier $\mathcal{V}$ will be $\text{poly}(n \cdot 2^{r(n)})$.

Observe that in the above definition, $\mathcal{V}$ makes $\ell(n)$ queries to Π using the $r(n)$ coin tosses. In particular, $\mathcal{V}$ can only query a coordinate within range of $[0, \ell(n) \cdot 2^{r(n)} - 1]$. So from now on, we will always assume that the proof size $|\Pi|$ is $\mathcal{O}(\ell(n) \cdot 2^{r(n)})$.

Definition 2.2 (The class PCP). *For functions $r, \ell, a \in \mathbb{Z}^{\geq 0} \to \mathbb{Z}^{\geq 0}$, for $c, s \in (0, 1)$, define $\mathsf{PCP}_{c,s}[r, \ell, a]$ to be the class of languages L that have a standard (r, ℓ, a) verifier with completeness $\geq c$ and soundness $\leq s$, i.e.*

- **Completeness**: *For every $x \in L$, there exists a proof Π such that $\Pr_R[\mathcal{V}^\Pi(x; R) = \mathtt{Accept}] \geq c$.*

- **Soundness**: *For every $x \notin L$, for every Π, we have $\Pr_R[\mathcal{V}^\Pi(x; R) = \mathtt{Accept}] \leq s$.*

In this paper, we will usually focus on the language 3-COLOR of 3-colorable graphs.

Our main theorem (proved in Section 6) shows the following:

Theorem 2.3 (Main Theorem). *There exist constants c, ℓ such that the following holds for every q, m, d, k such that $q \geq cd^3$:*

Let $\mathbb{F}_q$ be a field of characteristic $\neq 2$ and let $V \subseteq \mathbb{F}_q^m$ have extension degree d and Gröbner complexity k. Then 3-COLOR on vertex set V is in $\mathsf{PCP}_{1,1/2}[c(k + m) \log q, \ \ell, \ cd \log q]$ with proofs of size $q^{c(k+m)}$.

Remark 2.4. *In Theorem 2.3, the assumption on characteristic $\neq 2$ is mostly for clarity in the proofs. We assign the vertices colors from the set $\{-1, 0, 1\}$, and these are three distinct colors only if the field is of characteristic $\neq 2$. For fields of characteristic 2 and with more than 3 elements, one could use colors $\{a, b, c\}$, where $a, b,$ and c are three distinct elements from the field. The proof is essentially the same.*

Lemma 2.5. *For every n, 3-COLOR on n-vertex graphs is in $\mathsf{PCP}_{1,1/2}\big[\mathcal{O}(\log n), \ \mathcal{O}(1), \ \mathcal{O}(\log^2 n/(\log \log n))\big]$ with proofs of size $n^{\mathcal{O}(1)}$.*

Proof. In Corollary 4.4 we show that if $V = H^m$ for some $H \subseteq \mathbb{F}_q$ then it has extension degree $(|H|-1) \cdot m$ and Gröbner complexity m. Taking $q \geq c \log^6 n$ a power of 3, $V = H^m$ for some subset $H \subseteq \mathbb{F}_q$ of size $\log n$ and $m = \log n / \log \log n$, we get the desired bounds. ∎

Lemma 2.6. *For every n, 3-COLOR on n-vertex graphs is in* $\mathsf{PCP}_{1,1/2}\big[\mathcal{O}(n^2),\ \mathcal{O}(1),\ \mathcal{O}(1)\big]$ *with proofs of size* $2^{\mathcal{O}(n^2)}$.

Proof. In Example 4.2.2 we show that if $V = \{0,1\}_{\leqslant 1}^m := \{(a_1, \ldots, a_m) \in \{0,1\}^m \,|\, \sum_{i=1}^m a_i \leqslant 1\}$ is the set of Boolean points in $\mathbb{F}_q^m$ of Hamming weight at most 1, then V has extension degree 1 and Gröbner complexity $\mathcal{O}(m^2)$. Picking q to be a large constant, and $V = \{0,1\}_{\leqslant 1}^m$ we get the desired bounds. ∎

Note that the above roughly matches the parameters of the Hadamard PCP of [ALMSS98] with a completely different proof!

Lemma 2.7. *For every $\varepsilon > 0$ and every n, 3-COLOR on n-vertex graphs is in* $\mathsf{PCP}_{1,1/2}\big[\mathcal{O}(n^\varepsilon),\ \mathcal{O}(1),\ \mathcal{O}\big(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\big)\big]$ *with proofs of size* $2^{\mathcal{O}(n^\varepsilon)}$.

Proof. In Corollary 4.5 we show that if $V = \big(\{0,1\}_{\leqslant 1}^m\big)^c$, then V has extension degree c and Gröbner complexity $\mathcal{O}(cm^2)$. Given $\varepsilon > 0$ picking $c = \mathcal{O}(\frac{1}{\varepsilon})$, $q = \mathrm{poly}(1/\varepsilon)$, $m = n^{\mathcal{O}(\varepsilon)}$, and $V = \big(\{0,1\}_{\leqslant 1}^m\big)^c$ we get the desired bounds. ∎

The above concludes the description of the atomic PCPs we construct. In Section 7 we show that these PCPs can be strengthened to "Robust assignment testers" (see Definition 7.1), and so can be composed together (see Lemma 7.3) to get the PCP theorem stated below (proved in Section 7).

Theorem 2.8 (PCP Theorem). *There exist universal constants ℓ, a, C such that for every n, 3-COLOR on n-vertex graphs is in* $\mathsf{PCP}_{1,1/2}[C\log n, \ell, a]$.

3 Preliminaries

For a field $\mathbb{F}_q$, we will use $\mathbb{F}_q[x_1, \ldots, x_m]$ to denote the multivariate polynomial ring in variables $x_1, \ldots, x_m$. For a degree parameter $d \in \mathbb{N}$, we will use $\mathcal{P}_d(\mathbb{F}_q^m) \subset \mathbb{F}_q[x_1, \ldots, x_m]$ to denote the subspace of degree $\leqslant d$ polynomials. For a polynomial $P \in \mathbb{F}_q[x_1, \ldots, x_m]$ and a set $V \subseteq \mathbb{F}_q^m$, we denote the restriction of P to V by $P|_V$. We will denote by $\mathbb{F}_q^\times$ the set of invertible elements of $\mathbb{F}_q$, i.e. $\mathbb{F}_q^\times = \mathbb{F}_q \setminus \{0\}$.

Theorem 3.1 (Polynomial Distance Lemma). *[Ore22; DL78; Sch80; Zip79]. Fix a field $\mathbb{F}_q$. For every degree parameter $d \in \mathbb{N}$ with $d \leqslant q$ and for every non-zero polynomial $P \in \mathbb{F}_q[x_1, \ldots, x_m]$, we have:*

$$\Pr_{\mathbf{a} \sim \mathbb{F}_q^m}[P(\mathbf{a}) = 0] \leqslant \frac{d}{q}.$$

An immediate and useful corollary of Theorem 3.1 is the following: If two degree $\leqslant d$ polynomials P and Q agree on strictly more than d/q-fraction of $\mathbb{F}_q^m$, then $P = Q$.

For any $c, m \in \mathbb{N}$ with $c \leqslant m$, we use $\{0,1\}_{\leqslant c}^m$ to denote the set of Boolean strings of Hamming weight $\leqslant c$. We say that two functions $f, g : S \to T$ are δ-close or δ-far if they differ on at most or at least a δ-fraction of their inputs, respectively.

Lines Table For every $m \in \mathbb{N}$, field $\mathbb{F}_q$, and points $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^m$, let $\ell_{\mathbf{a},\mathbf{b}} : \mathbb{F}_q \to \mathbb{F}_q^m$ (read as "line passing through $\mathbf{a}$ with slope $\mathbf{b}$") be defined as $\ell_{\mathbf{a},\mathbf{b}}(t) := \mathbf{a} + t\mathbf{b}$.

Definition 3.2 (Lines Table). *Fix a field $\mathbb{F}_q$. Let $d \in \mathbb{N}$ be the degree parameter and $m \in \mathbb{N}$ be the number of variables. For every degree $\leqslant d$ polynomial $f : \mathbb{F}_q^m \to \mathbb{F}_q$, we define the dth lines table for f $f_{\text{lines}}^{(d)} : \mathbb{F}_q^{2m} \longrightarrow (\mathbb{F}_q)^{d+1}$ as the function that maps an input $(\mathbf{a}, \mathbf{b}) \in \mathbb{F}_q^{2m}$ to $f(\ell_{\mathbf{a},\mathbf{b}}(t))$, where t is an indeterminate. We note that $f(\ell_{\mathbf{a},\mathbf{b}}(t))$ is indeed a univariate degree d polynomial in t and can be specified by the $d + 1$ coefficients of $t^0, t^1, \ldots, t^d$.*

3-colorability We state the 3-colorability language below. Note that the choice of 3-coloring as an NP-complete problem instead of one of many others is simply a matter of convenience.

Definition 3.3. *The decision problem* 3-COLOR *is the following problem:*
Given a graph $G = (V, E)$ with n vertices, decide whether there exists a proper coloring of G using 3 colors, i.e. for every edge $(u, v) \in E$, the vertices u and v are assigned different colors.

Lemma 3.4. *The decision problem* 3-COLOR *is* NP-*complete.*

3.1 Low-degree Testing

In this subsection, we discuss the standard point-vs-line test for low-degree testing from [ALMSS98]. We start by recalling the test and state its properties in Theorem 3.5. In the following discussion, we will switch between a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ and its evaluation vector $f \in (\mathbb{F}_q)^{q^m}$, as both are equivalent.

Algorithm 1: Low-Degree Test $\mathcal{LDT}^{(\cdot)}$

    Input: Degree parameter d, string $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^m$, element $t \in \mathbb{F}_q^\times$, and oracle access to
        (f, f') where $f \in (\mathbb{F}_q)^{q^m}$ and $f' \in (\mathbb{F}_q^{d+1})^{q^{2m}}$

1 Query $f'[(\mathbf{a}, \mathbf{b})]$ and query $f[\mathbf{a} + t\mathbf{b}]$ // Two queries to (f, f')

2 **if** $f'[(\mathbf{a}, \mathbf{b})](t) \neq f[\mathbf{a} + t\mathbf{b}]$ // Running time is $\mathrm{poly}(m, d)$

3   **then**

4   return *Reject* **else**

5   return *Accept*

Theorem 3.5 (Low-degree Testing (see for instance [ALMSS98, Theorem A5])). *There exists absolute constants $0 < C, \delta_0$ such that for every $\delta < \delta_0$, for every $d, q \in \mathbb{N}$ with $q > Cd^3$, the following holds over $\mathbb{F}_q$:*

1. *If $f \in \mathcal{P}_d(\mathbb{F}_q^m)$, then $\mathcal{LDT}^{f, f_{\text{lines}}^{(d)}}(; \mathbf{a}, \mathbf{b}, t)$ returns $\texttt{Accept}$ with probability 1 over the random choice of $(\mathbf{a}, \mathbf{b}, t)$.*

2. *For every $f : \mathbb{F}_q^m \to \mathbb{F}_q$ and for every $f' : \mathbb{F}_q^{2m} \to \mathbb{F}_q$, we have:*

$$\Pr_{\mathbf{a}, \mathbf{b}, t}\left[\mathcal{LDT}_d^{f, f'}(; \mathbf{a}, \mathbf{b}, t) \text{ returns } \texttt{Reject}\right] \leqslant \delta \implies \delta(f, \mathcal{P}_d(\mathbb{F}_q^m)) \leqslant 4\delta.$$

Furthermore, $\mathcal{LDT}_d^{f, f'}$ makes 2 oracle queries, uses $\mathcal{O}(m \log q)$ bits of randomness, and runs in time $\mathrm{poly}(m, d)$.

Remark 3.6. *For low-degree testing, there has been a long line of work on achieving better parameters in terms of field size and soundness guarantee. We refer the interested reader to [HKSS24, Section 1] for a detailed overview of the results of low-degree testing and also for the state-of-the-art parameters (see [HKSS24, Theorem 1.2]). We use the low-degree testing from [ALMSS98] because the algorithm and analysis are done using the lines table.*

3.2   Local Correction of Low-Degree Polynomials

In this subsection, we discuss the local correction/self-correction algorithm for degree d polynomials over $\mathbb{F}_q^m$ from [ALMSS98]. We first describe the local corrector and then analyze it in Theorem 3.7. In the following discussion, we will switch between a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ and its evaluation vector $f \in (\mathbb{F}_q)^{q^m}$, as both are equivalent.

Algorithm 2: Local Corrector $\mathcal{LC}^{(\cdot)}$

Input: Degree parameter d, evaluation point $\mathbf{a} \in \mathbb{F}_q^m$, string $\mathbf{b} \in \mathbb{F}_q^m$, element $t \in \mathbb{F}_q^\times$, and oracle access to (f, f') where $f \in (\mathbb{F}_q)^{q^m}$ and $f' \in (\mathbb{F}_q^{d+1})^{q^{2m}}$

1 Query $f'[(\mathbf{a}, \mathbf{b})](t)$ and query $f[\mathbf{a} + t\mathbf{b}]$ // Two queries to (f, f')

2 **if** $f'[(\mathbf{a}, \mathbf{b})](t) \neq f[\ell_{\mathbf{a}, \mathbf{b}}(t)]$ // Running time is $\mathrm{poly}(m, d)$

3 **then**

4 $\quad$ **return** $\texttt{Reject}$

5 **return** $f'[(\mathbf{a}, \mathbf{b})](0)$

Theorem 3.7 (Local Correction (see e.g. [ALMSS98], Proposition 7.2.2.1)). *There exists an absolute constant $C > 0$ such that for every $d, q \in \mathbb{N}$ satisfying $q > Cd$, the following holds.*

1. *If f is a polynomial of degree d, then for every $\mathbf{a} \in \mathbb{F}_q^m$, $\mathcal{LC}^{f, f_{\text{lines}}^{(d)}}(\mathbf{a}; \mathbf{b}, t)$ is equal to $f(\mathbf{a})$ with*

probability 1 over the random choice of $(\mathbf{b}, t)$.

2. *Let* $f : \mathbb{F}_q^m \to \mathbb{F}_q$ *be any function with the condition that there exists a degree d polynomial P such that* $\delta(f, P) \leqslant \delta$. *Then for every* $f' : \mathbb{F}_q^{2m} \to \mathbb{F}_q^{d+1}$, *for every* $\mathbf{a} \in \mathbb{F}_q^m$, *we have:*
If $\mathcal{LC}_d^{(f,f')}(\mathbf{a})$ *does not return* `Reject`*, then* $\mathcal{LC}^{(f,f')}(\mathbf{a})$ *computes* $P(\mathbf{a})$ *exactly with high probability over the random choice of* $(\mathbf{b}, t)$, *i.e.*

$$\Pr_{\mathbf{b},t}[\mathcal{LC}_d^{(f,f')}(\mathbf{a}; \mathbf{b}, t) = P(\mathbf{a}) \quad OR \quad \mathcal{LC}_d^{(f,f')}(\mathbf{a}; \mathbf{b}, t) \text{ returns } \texttt{Reject}] \geqslant 1 - 2\sqrt{\delta} - \frac{d}{q-1}.$$

Furthermore, $\mathcal{LC}_d^{(f,f')}(\mathbf{a})$ *makes 2 oracle queries, uses* $\mathcal{O}(m \log q)$ *bits of randomness, and runs in time* $\mathrm{poly}(m, d)$.

4 Gröbner Generating Sets

In [Section 5](), we define a test to check if a polynomial vanishes on a subset $V \subseteq \mathbb{F}_q^m$. In this section, we introduce the relevant parameters of such subsets, which we use to describe the efficiency of such tests. We also show that these are well-behaved under Cartesian products. We first define the parameters.

Definition 4.1 (Extension degree)**.** *For a non-empty set* $V \subseteq \mathbb{F}_q^m$, *function* $f : V \to \mathbb{F}_q$ *and polynomial* $P \in \mathbb{F}_q[X_1, \ldots, X_m]$ *we say P extends f if for every $a \in V$, we have* $f(a) = P(a)$.
We define the extension degree *of V to be the smallest integer $d \in \mathbb{N}$ such that every function* $f : V \to \mathbb{F}_q$ *can be extended to a polynomial* $\widehat{f}$ *of total degree at most d.*

Definition 4.2 (Gröbner Complexity)**.** *For an ideal* $\mathbb{I} \subseteq \mathbb{F}_q[x_1, \ldots, x_m]$ *we say that a finite set* $\mathfrak{G} \subseteq \mathbb{I}$ *is a Gröbner generating set of* $\mathbb{I}$, *if every polynomial $P \in \mathbb{I}$ can be written as follows:*

$$P = \sum_{g \in \mathfrak{G}} h_g \cdot g, \quad \text{where } h_g \in \mathbb{F}_q[x_1, \ldots, x_m] \text{ and } \deg(h_g g) \leqslant \deg(P).$$

For a set $V \subseteq \mathbb{F}_q^m$, *let* $\mathbb{I}(V)$ *denote the ideal of polynomials that vanish on V. We define the* Gröbner complexity *of V to be the cardinality of the smallest Gröbner generating set of* $\mathbb{I}(V)$.

The naming convention comes from the fact that a Gröbner basis in a graded ordering is a Gröbner generating set, as we show in [Section A]().

Example 4.2.1. *Let H be a subset of* $\mathbb{F}_q$. *Then any polynomial in* $\mathbb{F}_q[x]$ *which vanishes on H is divisible by*

$$\prod_{h \in H}(x - h).$$

Furthermore, for any function $f : H \to \mathbb{F}_q$ we can find a degree $|H| - 1$ polynomial extending f. It follows H has Gröbner complexity 1, and extension degree $|H| - 1$.

11

Example 4.2.2. *Let $\{0,1\}_{\leqslant 1}^n \subseteq \mathbb{F}_q^n$ be the subset of $\{0,1\}^n$ consisting of points with Hamming weight at most one. Then*

$$\left(x_1^2 - x_1, \ldots, x_n^2 - x_n, x_1 x_2, \ldots, x_{n-1} x_n\right)$$

is a Gröbner generating set. To see this, note that any polynomial P can be written on the form

$$P(\mathbf{x}) = \sum_i^n h_i(\mathbf{x}) \cdot (x_i^2 - x_i) \; + \; \sum_{i,j}^n g_{i,j}(\mathbf{x}) \cdot (x_i x_j) \; + \; \ell(\mathbf{x})$$

where ℓ is a linear function, and $\deg(h_i), \deg(g_{i,j}) \leqslant \deg(P) - 2$. The first two summands vanish on $\{0,1\}_{\leqslant 1}^n$, so P vanishes on $\{0,1\}_{\leqslant 1}^n$ if ℓ also vanishes. However, if ℓ has a non-zero coefficient for any variable x_i, then ℓ takes different values on two points which only differ in the i-th coordinate. It follows that ℓ only vanishes when it is the zero polynomial, and so P vanishes on $\{0,1\}_{\leqslant 1}^n$ if and only if it can be written on the form

$$P(\mathbf{x}) = \sum_i^n h_i(\mathbf{x}) \cdot (x_i^2 - x_i) \; + \; \sum_{i,j}^n g_{i,j}(\mathbf{x}) \cdot (x_i x_j).$$

It follows that $\{0,1\}_{\leqslant 1}^n$ has Gröbner complexity at most $\frac{n(n+1)}{2}$ and extension degree 1.

The following lemma shows that we can upper-bound both the Gröbner complexity and the extension degree of Cartesian products.

Lemma 4.3 (Subadditivity of Gröbner complexity and extension degree). *Let $V_1 \subseteq \mathbb{F}_q^{m_1}$ and $V_2 \subseteq \mathbb{F}_q^{m_2}$, and consider their product $V_1 \times V_2 \subseteq \mathbb{F}_q^{m_1 + m_2}$. We then have:*

1. *if $\mathfrak{G}_1, \mathfrak{G}_2$ are Gröbner generating sets for $\mathbb{I}(V_1), \mathbb{I}(V_2)$ respectively, then $\mathfrak{G}_1 \cup \mathfrak{G}_2$ is a Gröbner generating set for $\mathbb{I}(V_1 \times V_2)$.*

2. *if V_1, V_2 have extension degrees d_1, d_2 respectively, then $V_1 \times V_2$ has extension degree at most $d_1 + d_2$.*

In particular, both the Gröbner complexity and extension degree are subadditive under Cartesian products.

Proof of Lemma 4.3. Let V be a subset of $\mathbb{F}_q^m$. We first note, that we can a find a monomial basis $S \subseteq \mathbb{F}_q[\mathbf{x}]$ for the space of functions $\mathbb{F}_q^V$, such that any polynomial is equivalent to a linear combination of monomials from S of same or lesser degree.

We argue as follows. Since V is finite, $\mathbb{F}_q^V$ is spanned by polynomials and so is also spanned by monomials. Then we can create S iteratively by degree, by first setting $S_0 = \{1\}$ as a basis of $\mathbb{F}_q^V \cap \mathcal{P}_{\leqslant 0}(\mathbb{F}_q^m)$, and S_{i+1} by extending S_i to a monomial basis of $\mathbb{F}_q^V \cap \mathcal{P}_{\leqslant i+1}(\mathbb{F}_q^m)$, and setting $S = \bigcup_i S_i$. Since any polynomial of degree i is contained in the span of S_i, it must then be equivalent to a sum of monomials of degree at most i, showing the desired property.

Now let S be such a basis. Then any polynomial P as a function from V to $\mathbb{F}_q$ is equivalent to a

linear sum

$$P \equiv \sum_{s \in S} c_s s$$

and so

$$P - \sum_{s \in S} c_s s \equiv 0$$

is a polynomial of degree at most $\deg(P)$ vanishing on V. It follows that a subset $\mathfrak{G} \subseteq \mathbb{F}_q[\mathbf{x}]$ is a Gröbner generating set of $\mathbb{I}(V)$ if and only if every polynomial P can be written in the form

$$P = \sum_{s \in S} c_s s + \sum_{g \in \mathfrak{G}} h_g g,$$

where each summand has degree at most $\deg(P)$.

Now, let V_1, V_2 be subsets with $S_1, \mathfrak{G}_1 \subseteq \mathbb{F}_q[\mathbf{x}]$ and $S_2, \mathfrak{G}_2 \subseteq \mathbb{F}_q[\mathbf{y}]$ as above, and set

$$S_{12} := \{s_1 s_2 \mid s_1 \in S_1, \ s_2 \in S_2\}.$$

Then S_{12} is a basis for functions $V_1 \times V_2 \to \mathbb{F}_q$ with the above mentioned property. We will show that $\mathfrak{G}_1 \cup \mathfrak{G}_2$ is a Gröbner generating set for $V_1 \times V_2$. Let $m_1(\mathbf{x})m_2(\mathbf{y})$ be a monomial in $\mathbb{F}_q[\mathbf{x}, \mathbf{y}]$, We can apply the above property to each monomial separately to get

$$m_1(\mathbf{x})m_2(\mathbf{y}) = \left(\sum_{s \in S_1} c_s s + \sum_{g \in \mathfrak{G}_1} h_g g \right) \left(\sum_{s' \in S_2} c_{s'} s' + \sum_{g' \in \mathfrak{G}_2} h_{g'} g' \right)$$

$$= \sum_{ss' \in S_{12}} c_s c_{s'} ss' + \sum_{\substack{g \in \mathfrak{G}_1 \\ \leqslant \deg(m_1)}} h_g g \left(\sum_{\substack{s' \in S_2 \\ \leqslant \deg(m_2)}} c_{s'} s' + \sum_{\substack{g' \in \mathfrak{G}_2 \\ \leqslant \deg(m_2)}} h_{g'} g' \right) + \sum_{\substack{g' \in \mathfrak{G}_2 \\ \leqslant \deg(m_2)}} h_{g'} g' \left(\sum_{\substack{s \in S_1 \\ \leqslant \deg(m_1)}} c_s s \right).$$

We see this gives a representation of the monomial as a linear combination of terms from S_{12} and $\mathfrak{G}_1 \cup \mathfrak{G}_2$ of degree at most $\deg(m_1) + \deg(m_2)$. Since we can write a polynomial as a sum of monomials, we get that $\mathfrak{G}_1 \cup \mathfrak{G}_2$ is a Gröbner generating set. This proves the first item of the lemma.

To show the second item of the lemma, note that any function $f : V_1 \times V_2 \to \mathbb{F}_q$ can be written as a finite sum

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} a_i(\mathbf{x}) \cdot b_i(\mathbf{y})$$

where a_i, b_i are functions from V_1, V_2 respectively to $\mathbb{F}_q$. Since all the functions a_i can be represented as degree d_1-polynomials and likewise for b_i, the above sum gives a representation of f as a

polynomial of degree at most $d_1 + d_2$. ∎

Corollary 4.4. *For any subset $H \subseteq \mathbb{F}_q$, $H^m \subseteq \mathbb{F}_q^m$ has Gröbner complexity at most m and extension degree at most $(|H| - 1) \cdot m$.*

Proof. Combine Lemma 4.3 and Example 4.2.1. ∎

Corollary 4.5. *The subset $\left(\{0,1\}_{\leqslant 1}^{n/c} \right)^c \in \mathbb{F}_q^n$ has Gröbner complexity at most $\frac{(n^2 + nc)}{2c}$ and extension degree at most c.*

Proof. Combine Lemma 4.3 and Example 4.2.2. ∎

The following two lemmas show that given a subset V of $\mathbb{F}_q^m$, we can efficiently compute its extension degree, find a polynomial extending a function $f : V \to \mathbb{F}_q$, and compute a minimal Gröbner basis.

Lemma 4.6 (Computing extension degree and low-degree extensions)**.** *For every $q, m \in \mathbb{N}$, there exists:*

1. *An algorithm, which takes a set of n points V in $\mathbb{F}_q^m$ as input and gives the extension degree d as output in time $\mathrm{poly}\left(n, \log q, \binom{m+d}{m} \right)$.*

2. *An algorithm, which takes a set of n points V in $\mathbb{F}_q^m$ and a function $f : V \to \mathbb{F}_q$ as input and gives a polynomial which extends f as output in time $\mathrm{poly}\left(n, \log q, \binom{m+d+1}{m} \right)$.*

Proof. We first show 1.

Computing extension degree   For every i, fix an ordering of the $\binom{m+i}{i}$ monomials of degree at most i and also an ordering of the points in V.

Then for every i, we can calculate the evaluation matrix E_i of dimension $n \times \binom{m+i}{m}$ where

$$(E_i)_{(jk)} = m_k(v_j),$$

v_j is the jth point of V and m_k the kth monomial. Then if the rank of E_i is equal to n, the monomials of degree at most i span the functions on V, and so the extension degree is i.

We can then find the extension degree by calculating the rank of E_i for every i, until E_i has rank n. Every E_i is a submatrix of E_{i+j} so we can reuse the computation for every i. It follows that the algorithm performs Gaussian elimination on a single $n \times \binom{m+d}{d}$ matrix.

Finding extending polynomial   We first find the extension degree using the previous step. Then since E_d has rank n, we can find a right inverse A such that $E_d \cdot A = I_n$. If we represent the function $f : V \to \mathbb{F}_q$ as a n-dimensional vector y, then $A \cdot y$ gives a vector in the monomial basis, which represents a polynomial extending y since

$$E_d \cdot (A \cdot y) = (E_d \cdot A) \cdot y = y.$$

It follows that the algorithm performs one calculation of a right inverse, and one matrix-vector multiplication. ∎

Lemma 4.7 (Computing a smallest Gröbner generating set). *For every $q, m \in \mathbb{N}$, there exists an algorithm which takes a set of n points V in $\mathbb{F}_q^m$ as input and gives a minimal (w.r.t. size) Gröbner generating set of $\mathbb{I}(V)$ as output in the monomial basis in time* $\mathrm{poly}\left(n, \log q, \binom{m+d+1}{m}\right)$, *where d is the extension degree of V.*

Proof. Given a set of polynomials S, define S_i to be the subset of S of polynomials of degree exactly i, and define $S_{\leqslant i}$ to be the subset of polynomials of degree $\leqslant i$.

We will then construct the subsets of the minimal Gröbner generating set $\mathfrak{G}_i$ inductively. We first define $\mathfrak{G}_0 = \varnothing$, and then set $\mathfrak{G}_i$ to be any basis of the quotient space $\mathbb{I}(V)_{\leqslant i}/L_i$, where

$$L_i = \left\{ \sum_j h_j \cdot g_j \ \middle| \ h_j \in \mathbb{F}_q[x_1, \ldots, x_m], g_j \in \mathbb{I}(V)_{\leqslant i-1}, \deg(g_j h_j) \leqslant i \right\}.$$

We repeat this step until $i = d + 1$, so in each step, we check whether or not the the monomials of degree at most i span all functions on $|V|$, to know when to stop, as expressed in the following algorithm:

Algorithm 3: Constructing Gröbner generating set.

Input: Subset V of $\mathbb{F}_q^m$ of size n
Output: A minimal generating set $\mathfrak{G}$ of $I(V)$

1 $\mathfrak{G}, \mathfrak{G}_0, \mathcal{A}_0, \mathcal{B}_0 \leftarrow \varnothing$.
2 **for** $i = 1, \ldots$ **do**
3   Compute the evaluation matrix $E_i \in \mathbb{F}_q^{n \times \binom{m+i}{i}}$ of all monomials of degree at most i on the points in V
4   Compute a basis $\mathcal{A}_i$ of $\ker(E_i) \subset \mathbb{F}_q^{\binom{m+i}{i}}$
5   Compute a basis $\mathcal{B}_i$ of $\mathcal{A}_{i-1} + \mathrm{span}\,(x_i a : \ a \in \mathcal{A}_{i-1}, \ 1 \leqslant i \leqslant m)$
6   Compute a minimal basis $\mathfrak{G}_i$ so that $\mathrm{span}\,(\mathfrak{G}_i) + \mathrm{span}\,(\mathcal{B}_i) = \mathrm{span}\,(\mathcal{A}_i)$
7      $\mathfrak{G} \leftarrow \mathfrak{G} \cup \mathfrak{G}_i$
8   **if** $\mathrm{rank}\, E_{i-1} = n$ **then**
9   **break**

10 **return** $\mathfrak{G}$

Correctness We will now show any set $\mathfrak{G}$ is a minimal Gröbner generating set if and only if $\mathfrak{G}_i$ is a basis for the quotient space $\mathbb{I}(V)_{\leqslant i}/L_i$. To see this, if every $\mathfrak{G}_i$ is a spanning set for $\mathbb{I}(V)_{\leqslant i}/L_i$, then any polynomial $P \in \mathbb{I}(V)_{\leqslant i}$ can be written as

$$P \equiv \sum_{g \in \mathfrak{G}_i} c_g \cdot g \mod L_i,$$

which is equivalent to

$$P = \sum_{g \in \mathfrak{G}_i} c_g \cdot g + \sum_{g \in \mathfrak{G}_{\leqslant i-1}} h_g \cdot g$$

where c_g are constants and $\deg(h_g \cdot g) \leqslant i$. This is the condition for $\mathfrak{G}$ being a Gröbner generating set, so $\mathfrak{G}$ is a Gröbner generating set if and only if each $\mathfrak{G}_i$ spans $\mathbb{I}(V)_{\leqslant i}/L_i$.

Furthermore, note that in the above characterization, the conditions on each degree i is independent of each other, so $\mathfrak{G}$ is minimal if and only if each $\mathfrak{G}_i$ is minimal, which is equivalent to each of them being a basis.

From the above, we see that a minimal Gröbner generating set does not contain any degree i polynomials if

$$\mathbb{I}(V)_{\leqslant i} = L_i.$$

We show this is true for any $i \geqslant d + 2$. Since any monomial m of degree $i - 1$, is equivalent to a polynomial P of degree $i - 2$, as $i - 2 \geqslant d$, it follows that $x_i m - x_i P$ is in L_i for any variable x_i. In particular, any polynomial of degree i is equal to a polynomial of degree at most $i - 1$ modulo L_i. Together with the fact that $\mathbb{I}(V)_{i-1} \subseteq L_i$, we have $\mathbb{I}(V)_i = L_i$.

Runtime   To analyze the runtime, in each step of the loop we perform Gaussian elimination on matrices of size at most $n \times \binom{m+d+1}{m}$, so each step takes at most time $\mathrm{poly}(n, \log q, \binom{m+d+1}{m})$, and so the total runtime must also be $\mathrm{poly}(n, \log q, \binom{m+d+1}{m})$. ∎

Claim 4.8. *The running times in Lemma 4.6 and Lemma 4.7 are both $q^{O(m)}$.*

Proof. This follows from the fact that $n = |V| \leqslant q^m$, the extension degree d is at most $q(m-1)$ (the extension degree of $\mathbb{F}_q^m$) and the following binomial estimate

$$\binom{m+d+1}{m} \leqslant \binom{mq+1}{m} \leqslant \left(\frac{emq+e}{m}\right)^m \leqslant e^m (q+1)^m.$$

∎

5   Zero-on-Variety Test

In this subsection, we discuss an efficient test to decide whether a given oracle vanishes on a subset of points/variety. Let $V \subset \mathbb{F}_q^m$ be a set with a Gröbner generating set $\mathfrak{G}(V)$ with extension degree d_V and complexity k (see Section 4 for formal definitions). Let $P : \mathbb{F}_q[x_1, \ldots, x_m] \to \mathbb{F}_q$ be a polynomial of degree d and say $d \leqslant d_V$. Informally, the main goal of this section is:

Design an efficient standard verifier to decide whether P is zero at all points of V.

Before we state our standard verifier, let us first discuss what could constitute as proof for the vanishing of $P|_V$. Suppose $P(\mathbf{x})$ is a degree d polynomial and $P|_V \equiv 0$. Using the definition of the

ideal $\mathbb{I}(V)$ and $\mathfrak{G}(V)$, we have:

$$P|_V \equiv 0 \iff P \in \mathbb{I}(V)$$
$$\iff \text{There exists polynomials } h_g \in \mathbb{F}_q[x_1, \ldots, x_m] \text{ for every } g \in \mathfrak{G}(V) \text{ such that}$$

$$P(\mathbf{x}) = \sum_{g \in \mathfrak{G}(V)} h_g(\mathbf{x}) \cdot g(\mathbf{x}), \quad \text{where for every } g, \ \deg(h_g(\mathbf{x}) \cdot g(\mathbf{x})) \leqslant d. \tag{2}$$

We will refer to the ordered tuple $(h_g : g \in \mathfrak{G}(V)) \in (\mathbb{F}_q[x_1, \ldots, x_m])^k$ in Equation (2) as a *vanishing certificate*[4] for the polynomial $P|_V$.

Definition 5.1 (Vanishing Certificate Polynomial). *Let P, $V \subseteq \mathbb{F}_q^m$ and $\mathfrak{G}(V)$ as defined above. A vanishing certificate polynomial $\mathcal{M}_V(P) : \mathbb{F}_q^{m+k} \to \mathbb{F}_q$ is a polynomial of degree $\leqslant d$ satisfying the following conditions:*

- *There exists polynomials $h_g \in \mathbb{F}_q[x_1, \ldots, x_m]$ for every $g \in \mathfrak{G}(V)$ such that*

$$\mathcal{M}_V(P)(\mathbf{x}, \mathbf{y}) = \sum_{g \in \mathfrak{G}(V)} h_g(\mathbf{x}) \cdot y_g.$$

- *If we substitute $g(\mathbf{x})$ for y_g for every $g \in \mathfrak{G}(V)$ in the polynomial $\mathcal{M}_V(P)(\mathbf{x}, \mathbf{y})$, it should be the polynomial $P(\mathbf{x})$, i.e.*

$$P(\mathbf{x}) = \mathcal{M}_V(P)(\mathbf{x}, (g(\mathbf{x}) : g \in \mathfrak{G}(V))),$$

where the above equality is equality as polynomials.

Whenever the subset $V \subseteq \mathbb{F}_q^m$ is clear from the context, we will use $\mathcal{M}_{P,\text{lines}}^{(d)}$ to refer to the dth lines table for $\mathcal{M}_V(P)$ (see Definition 3.2 for a formal definition of the lines table).
We record our discussion above using Definition 5.1 in the following observation. We use the same notation as in Definition 5.1.

Observation 5.2. *Let $P(\mathbf{x})$ be a degree d polynomial. Then $P|_V \equiv 0$ if and only if there exists a vanishing certificate polynomial $\mathcal{M}_V(P)$ of degree $\leqslant d$ (see Definition 5.1). We would like to emphasize that Definition 5.1 has a degree restriction on $\mathcal{M}_V(P)$ and this will be crucial for us, as we will see soon.*

Observation 5.2 says that if a verifier wants to test whether P vanishes on V, a valid proof Π is:

$$\Pi = \left(\mathcal{M}_V(P), \ \mathcal{M}_{P,\text{lines}}^{(d)} \right). \tag{3}$$

In other words, our verifier for the Zero-on-Variety test will accept the above Π with probability

---

[4] There could be multiple vanishing certificates for $P|_V$ satisfying Equation (2). We only use the fact that there always exists a vanishing certificate where *each* polynomial h_g has degree $\leqslant \deg(P)$. We are guaranteed of the existence of such a vanishing certificate due to $\mathfrak{G}(V)$.

1. For a degree d polynomial P where $P|_V \not\equiv 0$, no vanishing certificate polynomial exists, and we require our verifier to reject every "claimed" proof Π' with high probability.

We next describe a standard verifier $\mathcal{ZERO}$ (recall the definition of standard verifier in Definition 2.1) with oracle access to a function f and an arbitrary string Π to decide whether $f|_V \equiv 0$, in Algorithm 4. For convenience in writing, we define the following map:

$$\varphi : \mathbb{F}_q^m \;\to\; \mathbb{F}_q^k$$
$$(z_1, \ldots, z_m) \;\mapsto\; (g(\mathbf{z}) : g \in \mathfrak{G}(V))$$

Algorithm 4: Zero-on-Variety Test for V: $\mathcal{ZERO}^{(\cdot)}$

Input: Degree parameter d, subset $V \subseteq \mathbb{F}_q^m$, Gröbner generating set $\mathfrak{G}(V)$,
string $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^{m+k}$, $\boldsymbol{\alpha} \in \mathbb{F}_q^m$, element $t \in \mathbb{F}_q^{\times}$,
and oracle access to $(f, \mathcal{M}, \mathcal{M}')$ where $f \in (\mathbb{F}_q)^{q^m}$, $\mathcal{M} \in (\mathbb{F}_q)^{q^{m+k}}$, and
$\mathcal{M}' \in (\mathbb{F}_q^d)^{q^{2(m+k)}}$.

1 Run $\mathcal{LDT}_d^{\mathcal{M},\mathcal{M}'}(;\mathbf{a}, \mathbf{b}, t)$ (see Algorithm 1)   `// Two queries to` $(\mathcal{M}, \mathcal{M}')$

2 **if** $\mathcal{LDT}_d^{\mathcal{M},\mathcal{M}'}(;\mathbf{a}, \mathbf{b}, t)$ *returns* `Reject` **then**

3   **return** `Reject`

4 Run $\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \mathbf{0}); \mathbf{a}, t)$ (see Algorithm 2)   `// Two queries to` $(\mathcal{M}, \mathcal{M}')$

5 **if** $\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \mathbf{0}); \mathbf{a}, t) \neq 0$ **then**

6   **return** `Reject`

7 Run $\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t)$ `// Two queries to` $(\mathcal{M}, \mathcal{M}')$ `and time to evaluate` $\varphi(\boldsymbol{\alpha})$ `is`
    $\mathcal{O}(k \cdot q^{\mathcal{O}(m)})$

8 Query $f[\boldsymbol{\alpha}]$ `// One query to` f

9 **if** $\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) \neq f[\boldsymbol{\alpha}]$ **then**

10   **return** `Reject` **else**

11   **return** `Accept`

> **Lemma 5.3** (Zero-on-Variety Test). *There exists an absolute constant $C > 0$ such that for every $d, q \in \mathbb{N}$ satisfying $q > Cd^3$, for every subset $V \subset \mathbb{F}_q^m$ with extension degree $\leqslant d$ and Gröbner complexity k, the following holds. The standard verifier $\mathcal{ZERO}$ satisfies the following properties:*
>
> *Let $r = (\mathbf{a}, \mathbf{b}, t, \boldsymbol{\alpha})$. Then,*
>
> 1. ***Completeness:*** *For every degree d polynomial $f : \mathbb{F}_q^m \to \mathbb{F}_q$ with $f|_V \equiv 0$, there exists a proof Π over alphabet $\mathbb{F}_q^{d+1}$ and size $\mathcal{O}(q^{2(m+k)})$ such that the following holds:*
>
> $$\Pr_r[\mathcal{ZERO}_d^{(f,\Pi)}(;r) \text{ returns } \texttt{Accept}] = 1.$$
>
> 2. ***Soundness:*** *Let $f : \mathbb{F}_q^m \to \mathbb{F}_q$ be any function for which there exists a unique degree d polynomial $P(\mathbf{x})$ such that $\delta(f, P) = \delta < 0.01$ and $P|_V \not\equiv 0$. Then for every string Π, the following holds:*
>
> $$\Pr_r[\mathcal{ZERO}_d^{(f,\Pi)}(;r) \text{ returns } \texttt{Reject}] \geqslant 0.04.$$
>
> 3. ***Efficiency:*** *$\mathcal{ZERO}$ uses $\mathcal{O}((m+k)\log q)$ bits of randomness, makes 7 oracle queries to (f, Π), and runs in time $\mathcal{O}(k \cdot q^{\mathcal{O}(m)})$.*

Proof of Lemma 5.3. We first note that the efficiency immediately follows from the comments in Algorithm 4. We discuss completeness next.

Completeness Suppose f is a polynomial of degree at most d and $f|_V \equiv 0$. As observed in Observation 5.2, we know there exists a vanishing certificate polynomial $\mathcal{M}_V(f)$ of degree d, and let Π be as stated in Equation (3). From the first item of Theorem 3.5, we know that $\mathcal{LDT}^{\mathcal{M}_V(f), \mathcal{M}_{f,\text{lines}}^{(d)}}$ returns $\texttt{Accept}$ with probability 1. From the first item of Theorem 3.7, we know that $\mathcal{LC}^{\mathcal{M}_V(f), \mathcal{M}_{f,\text{lines}}^{(d)}}(\boldsymbol{\alpha}, \mathbf{0})$ is equal to 0 for every $\boldsymbol{\alpha} \in \mathbb{F}_q^m$ with probability 1. Similarly, we know that $\mathcal{LC}^{\mathcal{M}_V(f), \mathcal{M}_{f,\text{lines}}^{(d)}}(\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha}))$ is equal to $f(\boldsymbol{\alpha})$ for every $\boldsymbol{\alpha} \in \mathbb{F}_q^m$ with probability 1. It is not difficult to see that Algorithm 4 accepts Π with probability 1. This finishes the completeness part of Lemma 5.3.

Soundness Let f be any function for which there exists a degree d polynomial $P(\mathbf{x})$ such that $\delta(f, P) \leqslant \delta$ and $P|_V \not\equiv 0$. Consider the following events from Algorithm 4:

1. $\mathcal{E}_1$ denotes the event that $\mathcal{LDT}_d^{\mathcal{M}, \mathcal{M}'}(; \mathbf{a}, \mathbf{b}, t)$ returns $\texttt{Reject}$. It depends on the choice of $(\mathbf{a}, \mathbf{b}, t)$.

2. $\mathcal{E}_2$ denotes the event that $\mathcal{LC}_d^{\mathcal{M}, \mathcal{M}'}((\boldsymbol{\alpha}, \mathbf{0}); \mathbf{a}, t) \neq 0$. It depends on the choice of $(\boldsymbol{\alpha}, \mathbf{a}, t)$.

3. $\mathcal{E}_3$ denotes the event that $\mathcal{LC}_d^{\mathcal{M}, \mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) \neq f[\boldsymbol{\alpha}]$. It depends on the choice of $(\boldsymbol{\alpha}, \mathbf{a}, t)$.

In the proof below, to avoid cumbersome writing, we will avoid repeatedly mentioning the random bits that each event depends on.

If either of the events $\mathcal{E}_1$ or $\mathcal{E}_2$ happens with probability greater than 0.04, then we have the desired soundness. Assume that is not the case, i.e.,

$$\Pr_{\mathbf{a},\mathbf{b},t}[\mathcal{E}_1] \leqslant 0.04 \quad \text{and} \quad \Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{E}_2] \leqslant 0.04.$$

We now want to argue that $\mathcal{E}_3$ happens with probability at least 0.04.

Since $\mathcal{E}_1$ happens with probability at most 0.04, from Theorem 3.5, we know that there exists a degree d polynomial $\mathcal{R}(\mathbf{x},\mathbf{y})$ such that $\delta(\mathcal{M},\mathcal{R}) \leqslant 0.16$. We will show the following claim.

Claim 5.4. *Let $\mathcal{R}(\mathbf{x},\mathbf{y}) : \mathbb{F}_q^{m+k} \to \mathbb{F}_q$ be the degree d polynomial such that $\delta(\mathcal{M},\mathcal{R}) \leqslant 0.16$. Then,*

$$\mathcal{R}(\mathbf{x},\mathbf{0}) \equiv 0.$$

Proof of Claim 5.4. As mentioned above, we know that $\delta(\mathcal{M},\mathcal{R}) \leqslant 0.16$. Using Theorem 3.7, we get:

$$\Pr_{\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha},\mathbf{0});\mathbf{a},t) = \mathcal{R}(\boldsymbol{\alpha},\mathbf{0})] \geqslant 1 - 2\sqrt{0.16} - \frac{d}{q-1}, \qquad \text{for every } \boldsymbol{\alpha} \in \mathbb{F}_q^m$$

$$\Rightarrow \Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha},\mathbf{0});\mathbf{a},t) \neq \mathcal{R}(\boldsymbol{\alpha},\mathbf{0})] \leqslant 0.08 + \frac{d}{q-1}. \tag{4}$$

Since the event $\mathcal{E}_2$ happens with probability $\leqslant 0.04$, we have,

$$\Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha},\mathbf{0});\mathbf{a},t) \neq 0] \leqslant 0.04. \tag{5}$$

Using union bound on Equation (4) and Equation (5), we get,

$$\Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{R}(\boldsymbol{\alpha},\mathbf{0}) \neq 0] \leqslant 0.12 + \frac{d}{q-1}$$

$$\iff \quad \Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{R}(\boldsymbol{\alpha},\mathbf{0}) = 0] \geqslant 0.88 - \frac{d}{q-1}.$$

Since the event $(\mathcal{R}(\boldsymbol{\alpha},\mathbf{0}) = 0)$ does not depend on the random choice of $(\mathbf{a},t)$, we get,

$$\Pr_{\boldsymbol{\alpha}}[\mathcal{R}(\boldsymbol{\alpha},\mathbf{0}) = 0] \geqslant 0.88 - \frac{d}{q-1}.$$

We choose C in the statement of Lemma 5.3 large enough such that $0.88 - \frac{d}{q-1} > \frac{d}{q}$. The polynomial distance lemma (Theorem 3.1) then implies that $\mathcal{R}(\mathbf{x},\mathbf{0}) \equiv 0$. This finishes the proof of Claim 5.4. ∎

Claim 5.4 implies that $\mathcal{R}(\mathbf{x}, \mathbf{y})$ belongs to the ideal $\mathbb{I}(y_1, \ldots, y_k)$. This implies the existence of polynomials $R_1, \ldots, R_k \in \mathbb{F}_q[\mathbf{x}, \mathbf{y}]$ such that $\mathcal{R}(\mathbf{x}, \mathbf{y})$ can be expressed as follows:

$$\mathcal{R}(\mathbf{x}, \mathbf{y}) \;=\; \sum_{g \in \mathfrak{G}(H)} R_g(\mathbf{x}, \mathbf{y}) \cdot y_g.$$

Define the polynomial $R : \mathbb{F}_q^m \to \mathbb{F}_q$ as follow, $R(\mathbf{x}) := \mathcal{R}(\mathbf{x}, \varphi(\mathbf{x}))$. Observe that $P(\mathbf{x})$ and $R(\mathbf{x})$ are distinct polynomials, otherwise $P|_V \equiv 0$.

Since for every $g \in \mathfrak{G}(V)$, we know that $\deg(g) \leqslant d$ and we also have that $\deg(\mathcal{R}(\mathbf{x}, \mathbf{y})) \leqslant d$, we get $\deg(R(\mathbf{x})) \leqslant d^2$. Since $\delta(\mathcal{M}, \mathcal{R}) \leqslant 0.16$, we have

$$\Pr_{\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) = \mathcal{R}(\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha}))] \;\geqslant\; 1 - 2\sqrt{0.16} - \frac{d}{q-1}, \qquad \text{for every } \boldsymbol{\alpha} \in \mathbb{F}_q^m$$

$$\Rightarrow \Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) = \mathcal{R}(\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha}))] \;\geqslant\; 0.92 - \frac{d}{q-1}$$

$$\Rightarrow \Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) \neq R(\boldsymbol{\alpha})] \;\leqslant\; 0.08 + \frac{d}{q-1}. \tag{6}$$

Recall that $P \neq R$ and from the polynomial distance lemma (Theorem 3.1), we have:

$$\Pr_{\boldsymbol{\alpha}}[R(\boldsymbol{\alpha}) = P(\boldsymbol{\alpha})] \;\leqslant\; \frac{d^2}{q} \quad \implies \quad \Pr_{\boldsymbol{\alpha}}[R(\boldsymbol{\alpha}) = f[\boldsymbol{\alpha}]] \;\leqslant\; \delta + \frac{d^2}{q}. \tag{7}$$

Using Equation (6) and Equation (7) and applying union bound, we get,

$$\Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) = f[\boldsymbol{\alpha}]]$$

$$\leqslant \Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_d^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) \neq R(\boldsymbol{\alpha})] + \Pr_{\boldsymbol{\alpha}}[R(\boldsymbol{\alpha}) = f[\boldsymbol{\alpha}]] \;\leqslant\; 0.08 + \frac{2d^2}{q-1} + \delta.$$

By choosing C appropriately in the statement of Lemma 5.3, we can set $2d^2/(q-1) \leqslant 0.01$. Thus,

$$\Pr_{\boldsymbol{\alpha},\mathbf{a},t}[\mathcal{LC}_{d^2}^{\mathcal{M},\mathcal{M}'}((\boldsymbol{\alpha}, \varphi(\boldsymbol{\alpha})); \mathbf{a}, t) \neq f[\boldsymbol{\alpha}]] \;\geqslant\; 0.91 - \delta \;\geqslant\; 0.04,$$

where we are using $\delta \leqslant 0.01$. Thus $\mathcal{E}_3$ happens with probability $\geqslant 0.04$. Hence we have showed that either $\mathcal{E}_1$ or $\mathcal{E}_2$ happens with probability $\geqslant 0.04$, otherwise $\mathcal{E}_3$ happens with probability $\geqslant 0.04$. This finishes the soundness of Lemma 5.3 and also the proof of Lemma 5.3. $\blacksquare$

6 Proof of the Main Theorem (Theorem 2.3)

In this section, we give the proof of Theorem 2.3, which we recall below.

Theorem 2.3 (Main Theorem)**.** *There exist constants c, ℓ such that the following holds for every*

q, m, d, k such that $q \geqslant cd^3$:

Let $\mathbb{F}_q$ be a field of characteristic $\neq 2$ and let $V \subseteq \mathbb{F}_q^m$ have extension degree d and Gröbner complexity k. Then 3-COLOR on vertex set V is in $\mathsf{PCP}_{1,1/2}[c(k+m)\log q, \ \ell, \ cd\log q]$ with proofs of size $q^{c(k+m)}$.

Proof of Theorem 2.3. We will consider the NP-complete problem 3-COLOR for graphs (see Definition 3.3 for a formal definition of 3-COLOR). We will use $\mathcal{V}_{\mathsf{PCP}}$ to denote the standard verifier with parameters as stated in Theorem 2.3.

From Lemma 4.7 and Claim 4.8, we know that $\mathcal{V}_{\mathsf{PCP}}$ can compute the Gröbner generating set of complexity k in time $q^{\mathcal{O}(m)}$. Let $E : V \times V \to \{0,1\} \subset \mathbb{F}_q$ be the edge function for the input graph $G = (V, E)$, defined as $E(u,v) = 1$ if and only if $(u,v) \in E$. Let $\widehat{E} : \mathbb{F}_q^m \times \mathbb{F}_q^m \to \mathbb{F}_q$ denote an extension of E and from the second item of Lemma 4.3, we know that $\deg(\widehat{E}) \leqslant 2d$. By Lemma 4.6 and Claim 4.8, the standard verifier $\mathcal{V}_{\mathsf{PCP}}$ can compute both, the extension degree d and the extension $\widehat{E}$ in time $q^{\mathcal{O}(m)}$.

For simplicity, we will describe a standard verifier $\mathcal{V}$ for 3-COLOR and then $\mathcal{V}_{\mathsf{PCP}}$ will be repeating $\mathcal{V}$ for $\mathcal{O}(1)$ times. More particularly, the standard verifier $\mathcal{V}$ will have soundness γ for some absolute constant $\gamma \in (0,1)$, i.e. $\mathcal{V}$ rejects with probability at least γ. The standard verifier $\mathcal{V}_{\mathsf{PCP}}$ will repeat $\mathcal{V}$ for $\mathcal{O}(1/\gamma)$ times and return Reject if any one of the iterations return Reject. As $\mathcal{O}(1/\gamma) = \mathcal{O}(1)$, the number of random bits, queries, and running time of $\mathcal{V}_{\mathsf{PCP}}$ are a constant factor multiple of the number of random bits, queries, and running time of $\mathcal{V}$ respectively. So for rest of the proof, it will be sufficient to describe a standard verifier $\mathcal{V}$ which uses $c'(k+m)\log q$ random bits, makes ℓ' queries to proofs over alphabets of size $c' \cdot (d\log q)$, have soundness guarantee of γ, and has running time $q^{\mathcal{O}(m+k)}$, for some constants c', ℓ', and $\gamma \in (0,1)$. From the previous paragraph, we know that $\mathcal{V}_{\mathsf{PCP}}$ can compute the Gröbner generating set, extension degree d, and extension $\widehat{E}$, all in time $q^{\mathcal{O}(m)}$. So we will assume that our standard verifier $\mathcal{V}$ has access to all of them.

Oracles We now describe oracles that the standard verifier $\mathcal{V}$ expects in a proof Π. In particular, if $G \in$ 3-COLOR, then our standard verifier always returns Accept. Our oracles will be evaluation tables of polynomials and their corresponding lines table. In the following, we also mention the size of each oracle that appears in the proof.

1. Let $\chi : V \to \{-1,0,1\} \subset \mathbb{F}_q$ be a coloring assignment to every vertex in the input $G = (V,E)$. Here we use $\{-1,0,1\}$ to denote three distinct colors.

   Let $\widehat{\chi} : \mathbb{F}_q^m \to \mathbb{F}_q$ denote an extension of χ of degree d. Let $\widehat{\chi}_{\text{lines}}^{(d)}$ be the dth lines table for $\widehat{\chi}$. Size of $(\widehat{\chi}, \widehat{\chi}_{\text{lines}}^{(d)})$ is $2^{\mathcal{O}(m\log q)}$ over alphabet of size $\mathcal{O}(d\log q)$.

2. Define the polynomial $A : \mathbb{F}_q^m \to \mathbb{F}_q$ as follows:

$$A(\mathbf{x}) \; := \; \widehat{\chi}(\mathbf{x}) \cdot (\widehat{\chi}(\mathbf{x}) - 1) \cdot (\widehat{\chi}(\mathbf{x}) + 1).$$

We have $\deg(A) \leqslant 3 \cdot \deg(\widehat{\chi}) \leqslant 3d$. Let $A_{\text{lines}}^{(3d)}$ be the $(3d)^{th}$ lines table for A.
Size of $(A, A_{\text{lines}}^{(3d)})$ is $2^{\mathcal{O}(m \log q)}$ over alphabet of size $\mathcal{O}(d \log q)$.

Observation 6.1. *For a vertex $\mathbf{u} \in V$, $A(\mathbf{u}) = 0$ if and only if $\widehat{\chi}(\mathbf{u}) \in \{-1, 0, 1\}$. This implies that $A|_V \equiv 0$ if and only if for every vertex $\mathbf{u} \in V$, we have $\widehat{\chi}(\mathbf{u}) \in \{-1, 0, 1\}$.*

Let[5] $\mathcal{M}_A : \mathbb{F}_q^{m+k} \to \mathbb{F}_q$ denote a vanishing certificate polynomial for $A|_V$ (see Definition 5.1 for a formal definition). Let $d_1 := \deg(\mathcal{M}_A) \leqslant \deg(A) \leqslant 3d$. Let $\mathcal{M}_{A,\text{lines}}^{(3d)}$ be the $(3d)^{th}$ lines table for $\mathcal{M}_A$.
Size of $(\mathcal{M}_A, \mathcal{M}_{A,\text{lines}}^{(3d)})$ is $2^{\mathcal{O}(m \log q + k \log q)}$ over alphabet of size $\mathcal{O}(d \log q)$.

3. Define the polynomial $B : \mathbb{F}_q^m \times \mathbb{F}_q^m \to \mathbb{F}_q$ as follows:

$$B(\mathbf{x}, \mathbf{y}) \; := \; \widehat{E}(\mathbf{x}, \mathbf{y}) \cdot \prod_{a \in \{\pm 1, \pm 2\}} (\widehat{\chi}(\mathbf{x}) - \widehat{\chi}(\mathbf{y}) - a) \cdot$$

We have $\deg(B) \leqslant \deg(\widehat{E}) + 4 \deg(\widehat{\chi}) \leqslant 6d$. Let $B_{\text{lines}}^{(6d)}$ be the $(6d)^{th}$ lines table for B.
Size of $(B, B_{\text{lines}}^{(6d)})$ is $2^{\mathcal{O}(m \log q)}$ over alphabet of size $\mathcal{O}(d \log q)$.

Observation 6.2. *Suppose $\widehat{\chi}(\mathbf{u}) \in \{-1, 0, 1\}$ for every $\mathbf{u} \in V$. For any two vertices $\mathbf{u}$ and $\mathbf{v}$, $B(\mathbf{u}, \mathbf{v}) = 0$ if and only if either $(\mathbf{u}, \mathbf{v}) \notin E$ or $\widehat{\chi}(\mathbf{u}) \neq \widehat{\chi}(\mathbf{v})$.*

Let $\mathcal{M}_B$ denote a vanishing certificate polynomial for $B|_V$. Let $d_2 := \deg(\mathcal{M}_B) \leqslant \deg(B) \leqslant 6d$. Let $\mathcal{M}_{B,\text{lines}}^{(6d)}$ be the $(6d)^{th}$ lines table for $\mathcal{M}_B$.
Size of $(\mathcal{M}_B, \mathcal{M}_{B,\text{lines}}^{(6d)})$ is $2^{\mathcal{O}(m \log q + k \log q)}$ over alphabet of size $\mathcal{O}(d \log q)$.

The proof Π consists of the following oracles:

$$\Pi = \left(\widehat{\chi}, \; \widehat{\chi}_{\text{lines}}^{(d)}, A, \; A_{\text{lines}}^{(3d)}, \; \mathcal{M}_A, \; \mathcal{M}_{A,\text{lines}}^{(3d)}, \; B, \; B_{\text{lines}}^{(6d)}, \; \mathcal{M}_B, \; \mathcal{M}_{B,\text{lines}}^{(6d)} \right) \tag{8}$$

As we have mentioned, the size of each of the components in Π, we get that the size of the proof Π is $2^{\mathcal{O}(m \log q + k \log q)} = q^{\mathcal{O}(m+k)}$ over an alphabet of size $\mathcal{O}(d \log q)$.

Description of the standard verifier $\mathcal{V}$ We are now ready to describe the standard verifier $\mathcal{V}$ to test whether a graph G is 3-colorable or not. In the following description, we interpret that the proof Π consists of the oracles as stated in Equation (8), i.e., $\mathcal{V}$ will interpret the proof Π as a long string with sub-strings forming the structure in Equation (8). We will show that $\mathcal{V}$ is a

---

[5] Recall k is the Gröbner complexity of $\mathfrak{G}(V)$.

standard verifier which uses $\mathcal{O}(m + k)$ random bits, makes $\mathcal{O}(1)$ queries to proofs over alphabets of size $\mathcal{O}(d \log q)$, have soundness guarantee of γ, and has running time $q^{\mathcal{O}(m+k)}$, for some constant $\gamma \in (0, 1)$.

Algorithm 5: Test by the Verifier $\mathcal{V}$

Input: Degree parameter d, subset V, Gröbner generating set for V, polynomial $\widehat{E}$,
       strings $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^m$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{F}_q^{2m}$, $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathbb{F}_q^{m+k}$, $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{F}_q^{2(m+k)}$, $t \in \mathbb{F}_q^{\times}$,
       and oracle access to $\Pi = (\widetilde{\chi},\ \widetilde{\chi}',\ \widetilde{A},\ \widetilde{A}', \widetilde{\mathcal{M}}_A, \widetilde{\mathcal{M}}'_A,\ \widetilde{B},\ \widetilde{B}', \widetilde{\mathcal{M}}_B, \widetilde{\mathcal{M}}'_B)$

1   Query $\widetilde{\chi}[\mathbf{a}], \widetilde{\chi}[\mathbf{b}], \widetilde{A}[\mathbf{a}], \widetilde{B}[\mathbf{a}, \mathbf{b}]$ `// 4 queries to Π`

2   Run $\mathcal{LDT}_d^{\widetilde{\chi},\widetilde{\chi}'}(;\mathbf{a}, \mathbf{b}, t), \mathcal{LDT}_{3d}^{\widetilde{A},\widetilde{A}'}(;\mathbf{a}, \mathbf{b}, t)$, and $\mathcal{LDT}_{6d}^{\widetilde{B},\widetilde{B}'}(;\boldsymbol{\alpha}, \boldsymbol{\beta}, t)$ (see Algorithm 1)
     `// 6 queries to Π and runs in time poly(m, d)`

3   **if** *either of the above three $\mathcal{LDT}$ test returns* `Reject` **then**
4 | **return** `Reject`

5   **if** $\widetilde{A}[\mathbf{a}] \neq \widetilde{\chi}[\mathbf{a}] \cdot (\widetilde{\chi}[\mathbf{a}] - 1) \cdot (\widetilde{\chi}[\mathbf{a}] + 1)$ *OR*
     $\widetilde{B}[\mathbf{a}, \mathbf{b}] \neq \widehat{E}(\mathbf{a}, \mathbf{b}) \cdot \prod_{i \in \{\pm 1, \pm 2\}}(\widetilde{\chi}[\mathbf{a}] - \widetilde{\chi}[\mathbf{b}] - i)$ `// Runs in time poly(n, q^{O(m)})`
6   **then**
7 | **return** `Reject`

8   Run $\mathcal{ZERO}_{3d}^{\widetilde{A},\widetilde{\mathcal{M}}_A,\widetilde{\mathcal{M}}'_A}(;\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \mathbf{a}, t)$ and $\mathcal{ZERO}_{6d}^{\widetilde{B},\widetilde{\mathcal{M}}_B,\widetilde{\mathcal{M}}'_B}(;\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\alpha}, t)$ (see
     Algorithm 4)   `// 14 queries to Π and running time poly(n, q^{O(m+k)})`

9   **if** *either of the above two $\mathcal{ZERO}$ tests return* `Reject` **then**
10 | **return** `Reject`

11   **return** `Accept`

Efficiency   The random string used by $\mathcal{V}$ is the tuple $(\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, t)$. It is clear from here that these are $\mathcal{O}((m + k) \log q)$ random bits. From the comments in Algorithm 5, it is clear that $\mathcal{V}$ makes $\mathcal{O}(1)$ queries to the string Π, which is over an alphabet of size $\mathcal{O}(d \log q)$. From the comments in Algorithm 5, it is also clear that the running time of $\mathcal{V}$ is $q^{\mathcal{O}(m+k)}$.

Completeness   Let $G = (V, E) \in 3\text{-COLOR}$. This means there exists a coloring $\chi : V \to \{-1, 0, 1\}$ such that for every edge $(\mathbf{u}, \mathbf{v}) \in E$, we have $\chi(\mathbf{u}) \neq \chi(\mathbf{v})$. Let Π be as stated in Equation (8). From the first item of Theorem 3.5, we know that all three low-degree tests $\mathcal{LDT}$ return `Accept` with probability 1. From the definition of $\widehat{E}, \widehat{\chi}, A$, and B, we know that $\mathcal{V}$ never returns `Reject` in Line 6 of Algorithm 5. Using Observation 6.1 and Observation 6.2, we know that both $A|_V \equiv 0$ and $B|_{V \times V} \equiv 0$. From the completeness part of Lemma 5.3, we know that both $\mathcal{ZERO}^{A,\mathcal{M}_A,\mathcal{M}_{A,\text{lines}}^{(3d)}}$ and $\mathcal{ZERO}^{B,\mathcal{M}_B,\mathcal{M}_{B,\text{lines}}^{(6d)}}$ return `Accept` with probability 1. Hence $\mathcal{V}^{\Pi}(G)$ always return `Accept`

and thus has completeness 1.

Soundness Instead of the basic soundness claim, we prove a stronger claim that will also be useful in Section 7.

Claim 6.3 (Soundness). *For any constant $\varepsilon > 0$ there is a constant $\gamma > 0$ such that the following holds. Suppose there exists no proper 3-coloring $\psi : V \to \{-1, 0, 1\}$ of G such that $\delta(\widetilde{\chi}, \widehat{\psi}) \leqslant \varepsilon$, where $\widehat{\psi}$ is a degree d extension of ψ. Then $\mathcal{V}$ returns* `Reject` *with probability at least γ. In particular, if G is not 3-colorable, then $\mathcal{V}$ rejects with some constant probability.*

Proof of Claim 6.3. Consider the following events:

- $\mathcal{E}_1$ denotes the event that at least one of the three $\mathcal{LDT}$ test returns `Reject`. This event depends on the choice of $(\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, t)$.

- $\mathcal{E}_2$ denotes the event that

$$\widetilde{A}[\mathbf{a}] \neq \widetilde{\chi}[\mathbf{a}] \cdot (\widetilde{\chi}[\mathbf{a}] - 1) \cdot (\widetilde{\chi}[\mathbf{a}] + 1) \quad \text{OR} \quad \widetilde{B}[\mathbf{a}, \mathbf{b}] \neq \widehat{E}(\mathbf{a}, \mathbf{b}) \cdot \prod_{i \in \{\pm 1, \pm 2\}} (\widetilde{\chi}[\mathbf{a}] - \widetilde{\chi}[\mathbf{b}] - i).$$

  This event depends on the choice of $(\mathbf{a}, \mathbf{b})$.

- $\mathcal{E}_3$ denotes the event that $\mathcal{ZERO}_{3d}^{\widetilde{A}, \widetilde{\mathcal{M}}_A, \widetilde{\mathcal{M}}'_A}$ returns `Reject`. This event depends on the choice of $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \mathbf{a}, t)$.

- $\mathcal{E}_4$ denotes the event that $\mathcal{ZERO}_{6d}^{\widetilde{B}, \widetilde{\mathcal{M}}_B, \widetilde{\mathcal{M}}'_B}$ returns `Reject`. This event depends on the choice of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\alpha}, t)$.

Let $0 < \gamma < 0.01$ be an appropriate constant that we will choose later. If either of events $\mathcal{E}_1, \mathcal{E}_2$, or $\mathcal{E}_3$ happens with probability $> \gamma$, then we are done. Assume each of the events $\mathcal{E}_1, \mathcal{E}_2$, and $\mathcal{E}_3$ happens with probability $\leqslant \gamma$. We will show that $\mathcal{E}_4$ happens with probability $> \gamma$.

Since $\mathcal{E}_1$ happens with probability $\leqslant \gamma$, Theorem 3.5 implies:

- There exists degree d polynomial $P_{\widetilde{\chi}}(\mathbf{x})$ such that $\delta(P_{\widetilde{\chi}}, \widetilde{\chi}) \leqslant 4\gamma$.

- There exists degree $(3d)$ polynomial $P_{\widetilde{A}}(\mathbf{x})$ such that $\delta(P_{\widetilde{A}}, \widetilde{A}) \leqslant 4\gamma$.

- There exists degree $(6d)$ polynomial $P_{\widetilde{B}}(\mathbf{x})$ such that $\delta(P_{\widetilde{B}}, \widetilde{B}) \leqslant 4\gamma$.

We show the following claim on the relation between $P_{\widetilde{\chi}}, P_{\widetilde{A}}$, and $P_{\widetilde{B}}$.

Claim 6.4. *Let the polynomials $P_{\widetilde{\chi}}, P_{\widetilde{A}}$, and $P_{\widetilde{B}}$ be as mentioned above. Then for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^m$,*

$$P_{\widetilde{A}}(\mathbf{x}) \;=\; P_{\widetilde{\chi}}(\mathbf{x}) \cdot (P_{\widetilde{\chi}}(\mathbf{x}) - 1) \cdot (P_{\widetilde{\chi}}(\mathbf{x}) + 1). \tag{9}$$

and

$$P_{\widetilde{B}}(\mathbf{x}, \mathbf{y}) \;=\; \widehat{E}(\mathbf{x}, \mathbf{y}) \cdot \prod_{i \in \{\pm 1, \pm 2\}} (P_{\widetilde{\chi}}[\mathbf{x}] - P_{\widetilde{\chi}}[\mathbf{y}] - i). \tag{10}$$

25

Proof of Claim 6.4. The idea is to show that each pair of polynomials in either Equation (9) or Equation (10) agree on a large fraction of their respective domains. Since these are all low-degree polynomials, the polynomial distance lemma will imply that they are, in fact, equal.

Since the event $\mathcal{E}_2$ happens with probability $\leqslant \gamma$, we have the following two inequalities:

$$\Pr_{\mathbf{a} \sim \mathbb{F}_q^m} \left[\widetilde{A}[\mathbf{a}] \neq \widetilde{\chi}[\mathbf{a}] \cdot (\widetilde{\chi}[\mathbf{a}] - 1) \cdot (\widetilde{\chi}[\mathbf{a}] + 1) \right] \leqslant \gamma \tag{11}$$

and also

$$\Pr_{\mathbf{a},\mathbf{b} \sim \mathbb{F}_q^m} \left[\widetilde{B}[\mathbf{a},\mathbf{b}] \neq \widehat{E}(\mathbf{a},\mathbf{b}) \cdot \prod_{i \in \{\pm 1, \pm 2\}} (\widetilde{\chi}[\mathbf{a}] - \widetilde{\chi}[\mathbf{b}] - i) \right] \leqslant \gamma. \tag{12}$$

Using $\delta(P_{\widetilde{\chi}}, \widetilde{\chi}) \leqslant 4\gamma$, $\delta(P_{\widetilde{A}}, \widetilde{A}) \leqslant 4\gamma$, and Equation (11) together, via triangle inequality, we get:

$$\Pr_{\mathbf{a} \sim \mathbb{F}_q^m} \left[P_{\widetilde{A}}(\mathbf{a}) \neq P_{\widetilde{\chi}}(\mathbf{a}) \cdot (P_{\widetilde{\chi}}(\mathbf{a}) - 1) \cdot (P_{\widetilde{\chi}}(\mathbf{a}) + 1) \right] \leqslant 9\gamma$$

Since both the polynomials in the above inequality have degree $\leqslant (3d)$ and because $3d/q < 9\gamma$, the polynomial distance lemma (Theorem 3.1) implies Equation (9). An analogous argument shows Equation (10). This finishes the proof of Claim 6.4. ∎

We assumed that event $\mathcal{E}_3$ happens with probability $\leqslant \gamma$. Recall that $\gamma < 0.01 < 0.04$. From Lemma 5.3, we can infer that $P|_{\widetilde{A}}$ vanishes on V. In other words, using Equation (9), we know that $P_{\widetilde{\chi}}$ is an extension of a valid vertex coloring $\widetilde{\chi}$ (i.e. $\widetilde{\chi}$ assigns each vertex a color from the set $\{-1, 0, 1\}$). We will now show that event $\mathcal{E}_4$ happens with probability $> \gamma$.

Recall that $\widetilde{\chi}$ is ε-far from any degree d extension of a proper 3-coloring of G. In particular, for $\gamma < \varepsilon/4$, we know that $P_{\widetilde{\chi}}$ is not an extension of a proper 3-coloring. In particular, there exists vertices $\mathbf{u}, \mathbf{v}$ such that $(\mathbf{u}, \mathbf{v}) \in E$ and $P_{\widetilde{\chi}}(\mathbf{u}) = P_{\widetilde{\chi}}(\mathbf{v})$. In other words, $P_{\widetilde{B}}$ does not vanish on $V \times V$. From Lemma 5.3, we know that event $\mathcal{E}_4$ happens with probability $\geqslant (0.04 - 4\gamma) > \gamma$. Hence we have shown that either one of $\mathcal{E}_1, \mathcal{E}_2$, or $\mathcal{E}_3$ happens with probability $> \gamma$, otherwise $\mathcal{E}_4$ happens with probability $> \gamma$. This finishes the proof of Claim 6.3.

∎

Hence we have shown that the standard verifier $\mathcal{V}$ has completeness 1, soundness γ for some constant $\gamma \in (0, 1)$, uses $\mathcal{O}((m + k) \log q)$ random bits, makes $\mathcal{O}(1)$ queries to a proof of size $q^{\mathcal{O}(m+k)}$ over an alphabet of size $\mathcal{O}(d \log q)$, and runs in time $q^{\mathcal{O}(m+k)}$. As we discussed earlier, $\mathcal{V}_{\mathsf{PCP}}$ repeats $\mathcal{V}$ for $\mathcal{O}(1/\gamma) = \mathcal{O}(1)$ times to achieve soundness of $1/2$, and the other parameters remain the same upto $\mathcal{O}(1)$ factor. This finishes the proof of Theorem 2.3. ∎

7   The PCP Theorem with One Composition

In this section, we use the main theorem (Theorem 2.3) to give a proof of the PCP theorem with a single composition, composing two different instantiations of our basic PCP.

More precisely, we need an extension of our PCP to a *Robust Assignment-Tester* (or equivalently a *Robust PCP of Proximity* [BGHSV06]). The definition below is due to Dinur and Reingold [DR06]. We will assume throughout this section that there is a single growing parameter n and all other parameters $(R, q, \varepsilon, \ldots)$ are (possibly constant) functions of n.

Notation. Recall that two functions $f, g : S \to T$ are said to be δ-*close* if they differ on at most a δ-fraction of their inputs and δ-*far* if they differ on at least a δ-fraction of their inputs.

Definition 7.1 (Robust Assignment-Testers, combining Definitions 3.1 and 3.4 from [DR06]). *A Robust Assignment-Tester with parameters $(R, s, \ell, \delta, \varepsilon, w, \rho)$ is a reduction whose input is a Boolean circuit φ of size n over Boolean variables X. The reduction runs in time $\mathrm{poly}(n, R)$ and outputs a system of $R(n)$ Boolean circuits $\psi = \{\psi_1, ..., \psi_R\}$, each ψ_i of size at most $s(n)$ over Boolean variables X and auxiliary variables Y such that the following conditions hold.*

- *Each ψ_i depends on a set of ℓ variables from $X \cup Y$, which we denote $\mathrm{Vars}(\psi_i)$. The variables in Y take values in an alphabet Σ, and are accessible to ψ_i as a tuple of $w = \lceil \log_2(|\Sigma|) \rceil$ bits.*

- *For every Boolean assignment $\sigma : X \to \{0, 1\}$,*

  1. *Completeness: If σ satisfies the input circuit φ, then there exists an assignment $\tau : Y \to \Sigma$ such that $\sigma \cup \tau$ satisfies all of $\psi_1, \ldots, \psi_R$.*

  2. *Robust Soundness: If σ is δ-far from any satisfying assignment of the input circuit φ, then for any assignment $\tau : Y \to \Sigma$ and for any $(1-\varepsilon)$-fraction of the ψ_i's, the restricted assignment $(\sigma \cup \tau)|_{\mathrm{Vars}(\psi_i)}$ is ρ-far from any satisfying assignment of ψ_i.*

A (not necessarily robust) Assignment-Tester with parameters $(R, s, \ell, \delta, \varepsilon, w)$ is defined as above, except that we replace the Robust Soundness condition with the following weaker Soundness condition.

- *Soundness: If σ is δ-far from any satisfying assignment of φ, then for any assignment $\tau : Y \to \Sigma$ and for a $(1 - \varepsilon)$ fraction of the ψ_i's, the restricted assignment $(\sigma \cup \tau)|_{\psi_i}$ is not a satisfying assignment of ψ_i.*

[DR06] showed that the existence of one assignment-tester with specific parameters is enough to prove the PCP theorem. We state the observation formally below.

Observation 7.2 (Section 3.1 in [DR06]). *To prove the PCP theorem i.e.,*

$$\mathsf{NP} \subseteq \mathsf{PCP}_{1, 1-\Omega(1)}[\mathcal{O}(\log n), \mathcal{O}(1), \mathcal{O}(1)],$$

it suffices to show that there is an assignment-tester with parameters $R(n) = \mathrm{poly}(n)$, $s, \ell, w \in \mathcal{O}(1)$ and $\delta, (1 - \varepsilon) \in \Omega(1)$.

We will prove the existence of such an assignment-tester by composing two *robust* assignment-testers with suitable parameters. The process of composition starts with two robust assignment-testers $\mathcal{A}_1$ and $\mathcal{A}_2$ and produces a robust assignment-tester $\mathcal{A}_3$ with many parameters (e.g. s, ℓ, w) that are dictated by the parameters of $\mathcal{A}_2$. The robust assignment-tester $\mathcal{A}_3$ tests assignments to the Boolean circuit φ that is an input to $\mathcal{A}_1$.

The construction is simple: The tester $\mathcal{A}_3$ first runs $\mathcal{A}_1$ with input φ and then runs $\mathcal{A}_2$ on each

of the circuits produced by $\mathcal{A}_1$. The high-level idea behind it is that we test an assignment to the input circuit φ to $\mathcal{A}_1$ by testing that each of the 'local views' of the various circuits produced by $\mathcal{A}_1$ is close to a satisfying assignment, where the latter process is carried out by the circuits produced by reduction $\mathcal{A}_2$. This leads us to the following observation of Dinur and Reingold [DR06, Lemma 3.5], which we need only as a black box.

Lemma 7.3 (Composing robust assignment-testers: Lemma 3.5 in [DR06]). *Assume that there exist $\mathcal{A}_1$ and $\mathcal{A}_2$ are robust assignment-testers with parameters $(R_1, s_1, \ell_1, \delta_1, \varepsilon_1, w_1, \rho_1)$ and $(R_2, s_2, \ell_2, \delta_2, \varepsilon_2, w_2, \rho_2)$ respectively. If $\rho_1 \geqslant \delta_2$, then there also exists a robust assignment-tester $\mathcal{A}_3$ with parameters $(R_3, s_3, \ell_3, \delta_3, \varepsilon_3, w_3, \rho_3)$ where*

- $R_3(n) = R_1(n) \cdot R_2(s_1(n))$,

- $s_3(n) = s_2(s_1(n))$,

- $\ell_3(n) = \ell_2(s_1(n))$,

- $\delta_3(n) = \delta_1(n)$,

- $\varepsilon_3(n) = \varepsilon_1(n) + \varepsilon_2(s_1(n)) - \varepsilon_1(n) \cdot \varepsilon_2(s_1(n))$,

- $w_3(n) = w_2(n)$,[6]

- $\rho_3(n) = \rho_2(n)$.

Turning our PCPs into robust assignment-testers. To use the above lemma, we instantiate our main theorem with the two different choices of varieties used in Lemma 2.5 and Lemma 2.7. While these statements only yield PCPs (a weaker object than assignment-testers) for the 3-COLOR problem, we show that they easily yield assignment-testers for Boolean circuits using two basic ingredients: basic properties of standard reductions showing 3-COLOR is NP-hard and the local correctability of the polynomial witnesses in the PCP proof of Theorem 2.3 using Theorem 3.7.

The following is a basic property of standard reductions form Circuit-SAT to 3-COLOR. One can prove it e.g. by using the standard Tseitin transformation (reducing Circuit-SAT to 3-SAT) followed by the reduction from 3-SAT to 3-COLOR in [GJS76, Theorem 2.1].

Lemma 7.4. *There is a polynomial-time reduction from Circuit-SAT to 3-coloring that satisfies the following. On input a Boolean circuit φ, the graph $G = (V, E)$ produced by the reduction has the property that for any satisfying assignment $\sigma : X \to \{0, 1\}$ of φ, there is a proper 3-coloring $\chi : V \to \{-1, 0, 1\}$ such that $|V| = O(|\varphi|)$, χ restricts to σ on a fixed subset V_0 of V (here V_0 is in 1-1 correspondence with X) and $\chi(v_0) = -1$ for some fixed $v_0 \in V \backslash V_0$. Furthermore, any proper 3-coloring $\chi : V \to \{-1, 0, 1\}$ satisfying $\chi(v_0) = -1$, upon restriction to V_0, yields a satisfying assignment $\sigma' : X \to \{0, 1\}$ of the circuit φ.*

Given the above, we can easily modify the PCPs from the proofs of Theorem 2.3 to yield assignment-testers for CircuitSAT. We will show how to do this below.

Moreover, we will use an idea of Dinur and Reingold to make these assignment-testers *robust* using the simple process of encoding the input symbols by an explicit asymptotically good error-correcting

---

[6] This is not explicitly stated in [DR06] but it follows trivially from the proof.

code. More precisely, we use the following lemma.

Lemma 7.5 (Robustization: Lemma 3.6 in [DR06]). *There is an absolute constant c_1 such that if there is an assignment-tester $\mathcal{A}$ with parameters $(R, s, \ell, \delta, \varepsilon, w)$, then there is also a robust assignment-tester $\mathcal{A}'$ with parameters $(R' = R, s' = c_1 \cdot s, \ell' = \ell, \delta' = \delta, \varepsilon' = \varepsilon, w' = \text{poly}(w), \rho' = 1/(c_1 \cdot \ell))$.*

We are now ready to state the main lemma of this section, which says that the PCPs from the previous section can be turned into robust assignment-testers with suitable parameters.

Lemma 7.6. *There exist absolute constants c, c_1 such that the following holds. Assume that for every $n \geqslant 1$, there exist $q = q(n), m = m(n), d = d(n), k = k(n)$ such that $q \geqslant cd^3$ and a variety $V_n \subseteq \mathbb{F}_q^m$ of size $\omega(n)$ constructible in time $\text{poly}(n)$ with extension degree d and Gröbner complexity k. Then, for any $\delta = \delta(n) > 0$, there is a robust assignment-tester with parameters*

$$\left(q^{O(k+(m/\delta))}, \text{poly}((d \log q)/\delta), \ell = O(1/\delta), \delta, 1/4, q^{O(d)}, \frac{1}{c_1 \cdot \ell} \right)$$

Proof. The constants c and c_1 are chosen from Theorem 2.3 and Lemma 7.5 respectively. We can assume without loss of generality that $c \geqslant 100$.

We can now describe the reduction $\mathcal{A}$ behind the assignment-tester (we will make it robust below). The reduction first reduces the given instance φ of CircuitSAT to an instance $G = (V, E)$ of 3-COLOR using the polynomial-time reduction described in Lemma 7.4 (so $|V| = O(n)$). Fix $V_0 \subseteq V$ (which is 1-1 correspondence with the set of variables X of φ) and $v_0 \in V \backslash V_0$ as mentioned above.

The reduction $\mathcal{A}$ now constructs a variety V_n as in the hypothesis of the lemma. Note that $|V_n| = \omega(n) \geqslant |V|$. After adding some isolated vertices to G if required, we can identify V with V_n.

We now consider the algorithm Tester specified in Algorithm 6 with c_2 and c_3 being absolute constants that we will choose below. For any choice of these constants, the algorithm uses $r = O((k + (m/\delta)) \log q)$ many random bits, makes $O(1/\delta)$ queries to its input oracles σ and τ where the former is defined over the Boolean alphabet and the latter is defined over an alphabet of size $O(d \log q)$. The reduction $\mathcal{A}$ iterates over all sequences b of r random coin tosses used by the algorithm and for each it produces a Boolean circuit ψ_b that performs the checks specified in the algorithm.

Algorithm 6: Tester

Input: Degree parameter d and oracle access to (σ, τ) where $\sigma : X \to \{0, 1\}$,
$\tau : Y \to \Sigma$, where Σ is the alphabet of the PCP verifier from Theorem 2.3
and $|Y| = q^{O(k+m)}$ is the length of the proof.

1 Run the PCP verifier from Theorem 2.3 on τ independently c_2 times. If the PCP
   verifier rejects even once, then **return** `Reject`   `// Random bits` $r' = O(c_2(k + m) \log q)$`,`
   `number of queries` $\ell' = O(1)$`, alphabet size` $O(d \log q)$`.`

2 Run $\mathcal{LC}_d^{\hat{\chi}, \hat{\chi}'}(v_0)$ and if the algorithm rejects or returns anything other than -1, then
   return `Reject`   `// Random bits` $O(m \log q)$`, number of queries` $O(1)$`.`

3 Repeat c_3/δ times: Sample random $v \in V_0$ and run $\mathcal{LC}_d^{\hat{\chi}, \hat{\chi}'}(v)$. If the algorithm
   rejects or returns anything other than $\sigma(v)$, then **return** `Reject`   `// Random bits`
   $O((c_3/\delta) \cdot m \log q)$`, number of queries` $O(c_3/\delta)$`.`

4 **return** `Accept`

The number of circuits produced is $2^r = q^{O(k + (m/\delta))}$, each of which queries $\ell = O(1/\delta)$ locations in the string $\sigma \cup \tau$. The computations performed by ψ_b are all efficiently computable functions of the queried bits, and hence the circuit ψ_b has size that is polynomial in the number of bits queried, which is $O((d \log q)/\delta)$. The alphabet Σ is the same as that of the PCP verifier and hence has cardinality $q^{O(d)}$.

To show that this is an assignment tester, it suffices to argue the completeness and soundness criteria. Completeness is trivial from the definition of the PCP and the properties of the NP-completeness reduction argued above.

For soundness, let us assume that $\sigma : X \to \{0, 1\}$ is δ-far from any satisfying assignment to φ. We need to show that the probability that Algorithm 6 rejects is at least $3/4$, since this also means that at least $3/4$ of the circuits ψ_b reject.

By the soundness of the PCP verifier Claim 6.3 we know that it rejects with constant probability unless $\hat{\chi}$ is at distance at most $\eta = 0.01$ from a degree d polynomial $P : \mathbb{F}_q^m \to \mathbb{F}_q$ that is a low-degree extension of a proper 3-coloring of G. Since we repeat the tests of the PCP verifier c_2 many times, the acceptance probability in case this does not hold is $\exp(-\Omega(c_2)) \leqslant 1/4$ as long as c_2 is large enough.

So from now on, we assume that $\hat{\chi}$ is η-close to a degree d polynomial P which is a low-degree extension of a proper 3-coloring $\chi : V \to \{-1, 0, 1\}$. If $\chi(v_0) \neq -1$, then by Theorem 3.7, the next step rejects with probability at least

$$1 - 2\sqrt{\eta} - \frac{d}{q-1} \geqslant 1 - 2\sqrt{\eta} - \frac{1}{c} \geqslant \frac{3}{4}.$$

Thus, we can assume that $\chi(v_0) = -1$.

Finally, since χ is a proper 3-coloring of G and $\chi(v_0) = -1$, it follows that the restriction of χ to V_0

defines a satisfying assignment of φ. Since σ is δ-far from any satisfying assignment of φ, it follows that for each $v \in V_0$ chosen in the next step, the probability that σ and χ differ at v is at least δ. Further, by Theorem 3.7, the probability that $\mathcal{LC}_d^{\hat{\chi},\hat{\chi}'}(v)$ returns $\chi(v)$ or rejects is at least $3/4$ (as in the previous paragraph). Thus, the chance that this step does not reject is bounded by

$$\left(1 - \frac{3\delta}{4}\right)^{c_3/\delta} \leqslant 1/4$$

as long as c_3 is a large enough constant. This concludes the proof of the soundness of $\mathcal{A}$.

We have thus shown that $\mathcal{A}$ is an assignment-tester with parameters

$$(q^{O(k+(m/\delta))}, \operatorname{poly}((k + (m/\delta))\log q), \ell = O(1/\delta), \delta, 1/4, q^{O(d)}).$$

In order to make the assignment-tester robust, we simply apply the Robustization lemma Lemma 7.5. This concludes the proof of Lemma 7.6. ∎

Proving the PCP theorem. To conclude the proof of the PCP theorem, we apply the above lemma to the PCPs given by specific instantiations of V already seen above.

Lemma 7.7 (Outer robust assignment-tester). *For any $\delta_1 = \delta_1(n) > 0$, we have a robust assignment-tester with parameters*

$$(R_1 = n^{O(1/\delta_1)}, s_1 = \operatorname{poly}((\log n)/\delta_1), \ell_1 = O(1/\delta_1), \delta_1, \varepsilon_1 = 1/4, w_1 = \operatorname{poly}(n), \rho_1 = 1/(c_1 \cdot \ell_1)).$$

Proof. Set $q = \operatorname{poly}(\log n)$ a power of 3, $V = H^m$ for $H \subseteq \mathbb{F}_q$ of size $\log n$, and $m = O(\log n/\log\log n)$. The lemma follows from Lemma 7.6 and Corollary 4.4. ∎

Lemma 7.8 (Inner robust assignment-tester). *For any $c = c(n), \delta_2 = \delta_2(n) > 0$, we have a robust assignment-tester with parameters*

$$(R_2 = c^{O(cn^{2/c}/\delta_2)}, s_2 = \operatorname{poly}(c/\delta_2), \ell_2 = O(1/\delta_2), \delta_2, \varepsilon_2 = 1/4, w_2 = c^{O(c)}, \rho_2 = 1/(c_1 \cdot \ell_2)).$$

Proof. Set $V = (\{0,1\}_{\leqslant 1}^{m/c})^c \subseteq \mathbb{F}_q^m$ for a suitably large $q = \operatorname{poly}(c)$ and $m = O(n^{1/c})$. The lemma follows from Lemma 7.6 and Corollary 4.5. ∎

We can now prove the PCP theorem.

Proof of Theorem 2.8. We compose the robust assignment-testers from Lemma 7.7 and Lemma 7.8 above.

Choosing δ_1 to be any absolute constant in $(0,1)$, c to be a large enough absolute constant such that $R_2(s_1(n)) = \operatorname{poly}(n)$, and $\delta_2 \leqslant \rho_1$ another absolute constant, we obtain two robust assignment testers that can be composed using Lemma 7.3. This leads to a robust assignment-tester with parameters $R_3 = \operatorname{poly}(n)$, $s_3, \ell_3, w_3 \in O(1)$ and $\delta_3, 1 - \varepsilon_3 \in \Omega(1)$, which by Observation 7.2 implies the PCP theorem. ∎

References

[Alo99]   Noga Alon. "Combinatorial Nullstellensatz". In: *Combinatorics, Probability and Computing* 8.1–2 (1999), pp. 7–29. DOI: 10.1017/S0963548398003411 (cit. on p. 4).

[ALMSS98]   Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. "Proof Verification and the Hardness of Approximation Problems". In: *J. ACM* 45.3 (1998), pp. 501–555. DOI: 10.1145/278298.278306. URL: https://doi.org/10.1145/278298.278306 (cit. on pp. 3, 6, 8–10).

[AS98]   Sanjeev Arora and Shmuel Safra. "Probabilistic Checking of Proofs: A New Characterization of NP". In: *J. ACM* 45.1 (1998), pp. 70–122. DOI: 10.1145/273865.273901. URL: https://doi.org/10.1145/273865.273901 (cit. on p. 3).

[BFL91]   László Babai, Lance Fortnow, and Carsten Lund. "Non-Deterministic Exponential Time has Two-Prover Interactive Protocols". In: *Comput. Complex.* 1 (1991), pp. 3–40. DOI: 10.1007/BF01200056. URL: https://doi.org/10.1007/BF01200056 (cit. on p. 3).

[BGHSV06]   Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. "Robust PCPs of Proximity, Shorter PCPs, and Applications to Coding". In: *SIAM J. Comput.* 36.4 (2006), pp. 889–974. DOI: 10.1137/S0097539705446810. URL: https://doi.org/10.1137/S0097539705446810 (cit. on pp. 3, 6, 27).

[BS08]   Eli Ben-Sasson and Madhu Sudan. "Short PCPs with Polylog Query Complexity". In: *SIAM J. Comput.* 38.2 (2008), pp. 551–607. DOI: 10.1137/050646445. URL: https://doi.org/10.1137/050646445 (cit. on pp. 3–5).

[DL78]   Richard A. DeMillo and Richard J. Lipton. "A probabilistic remark on algebraic program testing". In: *Information Processing Letters* 7.4 (1978), pp. 193–195. DOI: 10.1016/0020-0190(78)90067-4 (cit. on p. 8).

[Din07]   Irit Dinur. "The PCP theorem by gap amplification". In: *J. ACM* 54.3 (2007), p. 12. DOI: 10.1145/1236457.1236459. URL: https://doi.org/10.1145/1236457.1236459 (cit. on p. 3).

[DR06]   Irit Dinur and Omer Reingold. "Assignment Testers: Towards a Combinatorial Proof of the PCP Theorem". In: *SIAM J. Comput.* 36.4 (2006), pp. 975–1024. DOI: 10.1137/S0097539705446962. URL: https://doi.org/10.1137/S0097539705446962 (cit. on pp. 3, 6, 27–29).

[GJS76]   M.R. Garey, D.S. Johnson, and L. Stockmeyer. "Some simplified NP-complete graph problems". In: *Theoretical Computer Science* 1.3 (1976), pp. 237–267. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(76)90059-1. URL: https://www.sciencedirect.com/science/article/pii/0304397576900591 (cit. on p. 28).

[HKSS24]   Prahladh Harsha, Mrinal Kumar, Ramprasad Saptharishi, and Madhu Sudan. "An Improved Line-Point Low-Degree Test". In: *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024, pp. 1883–1892. DOI: 10.1109/FOCS61266.2024.00113. URL: https://doi.org/10.1109/FOCS61266.2024.00113 (cit. on p. 10).

[LFKN92]   Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. "Algebraic Methods for Interactive Proof Systems". In: *J. ACM* 39.4 (1992), pp. 859–868. DOI: 10.1145/146585.146605. URL: https://doi.org/10.1145/146585.146605 (cit. on p. 3).

[Ore22]   Øystein Ore. "Über höhere kongruenzen". In: *Norsk Mat. Forenings Skrifter* 1.7 (1922), p. 15 (cit. on p. 8).

[Sch80]   Jacob T. Schwartz. "Fast Probabilistic Algorithms for Verification of Polynomial Identities". In: *J. ACM* 27.4 (1980), pp. 701–717. DOI: 10.1145/322217.322225 (cit. on p. 8).

[Sha92]   Adi Shamir. "IP = PSPACE". In: *J. ACM* 39.4 (1992), pp. 869–877. DOI: 10.1145/146585.146609. URL: https://doi.org/10.1145/146585.146609 (cit. on p. 3).

[Zip79]   Richard Zippel. "Probabilistic algorithms for sparse polynomials". In: *Symbolic and Algebraic Computation*. Springer Berlin Heidelberg, 1979, pp. 216–226. DOI: 10.1007/3-540-09519-5_73 (cit. on p. 8).

A   Relation between Gröbner Bases and Gröbner Generating Sets

In Section 4 we introduced Gröbner generating sets. We here show that Gröbner bases are Gröbner generating sets, and that Gröbner bases are also well-behaved under Cartesian products.

Definition A.1. *An ordering of monomials in $k[\mathbf{x}]$ is called admissible, if every monomials $M, N, L \in k[\mathbf{x}]$ satisfies*

1. *$M \leqslant N$ implies $ML \leqslant NL$.*

2. *$M \leqslant ML$.*

if an admissible ordering further satisfies

3. *$\deg(M) < \deg(N)$ implies $M < N$*

we call it a graded ordering. For a polynomial P, we define $\mathrm{LM}(P)$ as the monomial in P of maximal order.

Example A.1.1. *For the lexicographic ordering we have $M < N$ if there exist i such that the exponent of x_j in M is equal to the exponent of x_j in N for $j < i$ and the exponent of x_i in M is strictly smaller than the exponent of x_i in N.*

For the graded lexicographic, we have $M < N$ if $\deg(M) < \deg(N)$ or if $\deg(M) = \deg(N)$ and $M < N$ in the lexicographic ordering.

Definition A.2. *fix an admissible ordering of monomials in $k[\mathbf{x}]$. A generating set G of an ideal $I \subseteq k[\mathbf{x}]$ is a Gröbner basis with respect to that ordering, if the leading monomial of every polynomial in I is a multiple of a leading monomial of a polynomial in G.*

If the ordering is graded, we say that G is a graded Gröbner basis.

Lemma A.3. *Let $\mathfrak{G}$ be a Gröbner basis for a graded ordering. Then $\mathfrak{G}$ is a Gröbner generating set.*

Proof. If $\mathfrak{G}$ is a Gröbner basis for $\mathbb{I}$, then a polynomial f is in $\mathbb{I}$ if every complete lead reduction of f results in the zero polynomial. Every step of the reduction will be of the form

$$f' - \frac{\mathrm{LM}(f')}{\mathrm{LM}(g)} g$$

for some $g \in \mathfrak{G}$ and f' being an intermediate result in the reduction.

For a graded ordering we have by definition $\deg(\mathrm{LM}(f)) = \deg(f)$ for every f, so

$$\deg\left(\frac{\mathrm{LM}(f')}{\mathrm{LM}(g)} g\right) = \deg(f') \leqslant \deg(f)$$

$\blacksquare$

Remark A.4. *If $\mathfrak{G}$ is a Gröbner basis of smallest size, then is not necessarily a Gröbner generating set of smallest size. For example, the two polynomials spanning the ideal*

$$\left(x_1^2, x_1 x_2 - x_2^2\right)$$

form a Gröbner generating set, as they both are homogeneous of degree 2. However, we also have

$$x_2^3 = x_2 \cdot x_1^2 - (x_1 + x_2) \cdot \left(x_1 x_2 - x_2^2\right)$$

so $x_2^3 \in \left(x_1^2, x_1 x_2 - x_2^2\right)$, and so in the graded lexicographic ordering $x_1^2, x_1 x_2, x_2^3$ are all leading monomials. Since no monomials of degree 2 can divide two of these monomials, a Gröbner basis must contain at least 3 elements.