

Low-Degree Method Fails to Predict Robust Subspace Recovery

He Jia
Aravindan Vijayaraghavan
Northwestern University

HE.JIA@NORTHWESTERN.EDU
 ARAVINDV@NORTHWESTERN.EDU

Editors: Steve Hanneke and Tor Lattimore

Abstract

The low-degree polynomial framework has been highly successful in predicting computational versus statistical gaps for high-dimensional problems in average-case analysis and machine learning. This success has led to the low-degree conjecture, which posits that this method captures the power and limitations of efficient algorithms for a wide class of high-dimensional statistical problems.

We identify a natural and basic hypothesis testing problem in \mathbb{R}^n which is polynomial time solvable, but for which the low-degree polynomial method fails to predict its computational tractability even up to degree $k = n^{\Omega(1)}$. Moreover, the low-degree moments match exactly up to degree $k = O(\sqrt{\log n / \log \log n})$.

Our problem is a special case of the well-studied robust subspace recovery problem. The lower bounds suggest that there is no polynomial time algorithm for this problem. In contrast, we give a simple and robust polynomial time algorithm that solves the problem (and noisy variants of it), leveraging anti-concentration properties of the distribution. Our results suggest that the low-degree method and low-degree moments fail to capture algorithms based on anti-concentration, challenging their universality as a predictor of computational barriers.

Keywords: Low-degree method, moment lower bounds, statistical vs computational gaps, robust subspace recovery, average-case complexity, anticoncentration

1. Introduction

Several problems in high-dimensional statistics and machine learning exhibit striking *statistical vs computational gaps* — these are gaps in the settings where it is possible to recover hidden structure or signal from polynomial samples, yet no known polynomial time algorithms succeed in doing so. Such statistical-computational gaps have been observed in a wide range of problems including planted clique, sparse PCA, community detection, low-rank matrix and tensor decomposition and estimation problems, mixtures of Gaussians, robust estimation and many others (Barak et al., 2019; Berthet and Rigollet, 2013; Hopkins and Steurer, 2017; Diakonikolas et al., 2017; Hopkins, 2018; Kunisky et al., 2019; Wein, 2025). A central goal in the theory of average-case analysis is to predict when such statistical-computational gaps arise, and to understand the limits of efficient algorithms for such statistical inference problems. Unlike the worst-case setting, where reductions and hardness conjectures (such as $P \neq NP$) organize our understanding, there has been less success in establishing strong computational hardness assuming a few central assumptions in the average-case world, barring some notable exceptions (Regev, 2009; Bruna et al., 2021; Brennan and Bresler, 2019; Brennan et al., 2018). The predominant approach has been to establish lower-bounds against specific families of algorithms, aiming to characterize the limits of broad algorithmic paradigms through lower bounds against statistical query algorithms, low-degree polynomials, convex relaxation hierarchies, local algorithms, and stable algorithms (Kearns, 1998; Feldman et al., 2017; Diakonikolas et al., 2023; Wein, 2025; Hopkins et al., 2017; Gamarnik, 2021).

The low-degree polynomial method has emerged as among the most popular approaches to understand statistical vs computational gaps. Consider a hypothesis testing problem, where the goal is to distinguish between a *planted* model P and a *null or reference* model Q . In this paper we consider a problem where we are given m i.i.d. samples $X_1, \dots, X_m \in \mathbb{R}^n$ where either $X_1, \dots, X_m \sim P$ or $X_1, \dots, X_m \sim Q$. The Low-Degree Advantage of degree k is given by

$$\text{LDA}_{\leq k}^{(m)}(P, Q) = \max_{f: \text{degree-}k \text{ polynomial}} \frac{|\mathbb{E}_{P^{\otimes m}}[f(X_1, \dots, X_m)] - \mathbb{E}_{Q^{\otimes m}}[f(X_1, \dots, X_m)]|}{\sqrt{\text{Var}_{Q^{\otimes m}}[f(X_1, \dots, X_m)]}} \quad (1)$$

Intuitively, this quantity measures how well degree- k polynomials of the samples can distinguish between the two distributions. The low-degree advantage also approximates the likelihood ratio between P and Q by restricting to low-degree statistics.

When the low-degree advantage is small e.g., $\text{LDA}_{\leq k}^{(m)}(P, Q) = O(1)$ or even better $o(1)$, it is considered an evidence of average-case hardness (Kunisky et al., 2019; Wein, 2025; Kunisky, 2025). In this regime, every degree- k polynomial fails to distinguish between P and Q , suggesting that no algorithm with comparable computational power can succeed. While the low-degree method does not capture Gaussian elimination and certain lattice-based algorithms (Diakonikolas and Kane, 2022; Zadik et al., 2022), these algorithms exploit rich algebraic structure and are brittle to a small amount of random noise. This is unlike low-degree polynomials which are inherently robust to random noise—a property that is particularly relevant in statistical estimation problems.

A rich line of work (Barak et al., 2019; Hopkins and Steurer, 2017; Hopkins et al., 2017) led to the low-degree conjecture (Hypothesis 2.1.5 and Conjecture 2.2.4) in the remarkable thesis of Hopkins (Hopkins, 2018) (see also Conjecture 1.6 in Ding et al. (2021), Conjecture 2.3 in Moitra and Wein (2023)), which in our setting can be phrased as follows:

Informal Conjecture 1 *For “natural” high-dimensional testing problems specified by m i.i.d. samples from P, Q , if $\text{LDA}_{\leq k}^{(m)}$ remains bounded as $m \rightarrow \infty$, then there is no (noise-tolerant) algorithm running in time $\exp(O(k/\text{polylog}(m)))$ that solves the distinguishing problem.*

The above conjecture is usually stated in a more general setting where the testing problem is specified by distributions P_N and Q_N with input size N ; our setting with m i.i.d. samples from a distribution (P or Q) over \mathbb{R}^n corresponds to an input size $N = mn$. This conjecture is informal since it does not specify a precise class of testing problems for which it is likely to hold. While the original conjecture was stated for a discrete domain (see Conjecture 2.2.4 in Hopkins (2018) for a precise version), the low-degree conjecture has been used for a much broader class of high-dimensional statistical problems over both discrete and continuous domains (see e.g., Ding et al., 2021; Moitra and Wein, 2023; Arpino and Venkataramanan, 2023; Kunisky, 2025).

Conversely it often holds in practice that if $\text{LDA}_{\leq k}^{(m)} \rightarrow \infty$ as $m, n \rightarrow \infty$, then there is typically a noise-tolerant distinguishing algorithm of running time roughly $(nm)^{O(k/\text{polylog}(mn))}$. The low-degree advantage seems to capture many existing algorithmic techniques including moment methods, spectral methods, local methods like approximate message passing (AMP), belief propagation, polynomial threshold functions and even certain Sum-of-Squares (SoS) relaxations (Kunisky et al., 2019; Kunisky, 2025; Wein, 2025; Hsieh et al., 2026). At the same time, we remark that the low-degree framework is viewed as a proxy for these paradigms rather than a formal characterization: it is not known to provably subsume all of them (see (Hsieh et al., 2026) for a careful account

of the rigorous implications of the low-degree heuristic). This low-degree conjecture has shown remarkable predictive power across a range of canonical problems in high-dimensional statistics including planted clique, sparse PCA, tensor PCA, clustering, spiked Wigner and Wishart models, tensor decomposition, community detection, Gaussian mixture models, certifying restricted isometry property and many others (see e.g., Hopkins et al., 2017; Ding et al., 2021; Moitra and Wein, 2023; Buhai et al., 2023; Arpino and Venkataramanan, 2023; Diakonikolas et al., 2023). See also the excellent surveys by Wein (2025); Kunisky (2025).

However, a recent exciting result of Buhai, Hsieh, Jain and Kothari (Buhai et al., 2025) provides a counterexample to the quasi-polynomial time variant of a concrete low-degree conjecture (Hopkins, 2018; Holmgren and Wein, 2020; Wein, 2025), by encoding a problem on list-decoding error-correcting codes into a statistical estimation task. Buhai et al. (2025) show that there is no low-degree advantage even for degree $k = n^{\Omega(1)}$, yet there is a quasi-polynomial-time algorithm for the problem. They also provide an example of a certain testing problem (with a single $n \times n$ matrix as input), that can be solved using eigenvalue computation but exhibits vanishing low-degree advantage for $k = \tilde{O}(n^{1/3})$. Despite the counterexample, the low-degree method remains one of the most compelling frameworks for understanding the limits of efficient algorithms in high-dimensional inference. This leads to the following compelling question:

Question: *Does the Low-Degree Advantage accurately predict the extent of natural algorithmic approaches for high-dimensional statistical problems? Are there algorithmic techniques that are not captured by the LDA method?*

1.1. Our Result

Our main contribution is a natural high-dimensional statistical problem where the Low Degree Advantage fails to predict computational tractability. Our problem is in fact a special case of the well-studied *robust subspace recovery* problem (Hardt and Moitra, 2013; Lerman and Maunu, 2018; Bhaskara et al., 2019; Bakshi and Kothari, 2021; Gao et al., 2026). In the general version of the problem, there is a (unknown) d -dimensional subspace $S \subset \mathbb{R}^n$, that we wish to recover. We are given samples $X_1, \dots, X_m \in \mathbb{R}^n$, where an α fraction of the samples are drawn from a distribution that is entirely supported on S , while the rest of the samples are drawn from a distribution fully supported on \mathbb{R}^n . The robust subspace recovery problem has an elegant polynomial time algorithm for sufficiently small d that leverages sufficient anti-concentration properties of the null distribution (Hardt and Moitra, 2013; Bhaskara et al., 2019). In contrast we identify natural distributions P (planted) and Q (null), which fools the low-degree method up to degree $k = n^{\Omega(1)}$, and other moment-based approaches. We also give an elementary polynomial time algorithm in this setting that achieves high noise tolerance.

We start by describing the hypothesis testing problem. We are given i.i.d. samples from a distribution \mathcal{D} . Our goal is to distinguish the following two hypotheses:

- **NULL:** \mathcal{D} is the rotationally invariant distribution Q_{rot} , given by a *scale mixture of spherical Gaussians*. Formally, a sample $X \in \mathbb{R}^n$ is drawn by first $\lambda \sim N(0, 1)$ and then $X \sim N(0, \lambda^2 I)$.
- **PLANTED:** \mathcal{D} is a distribution that places at least an $\alpha = 1/\text{poly}(n)$ probability mass on a subspace $S \subset \mathbb{R}^n$ of dimension at most $d = O(1)$.

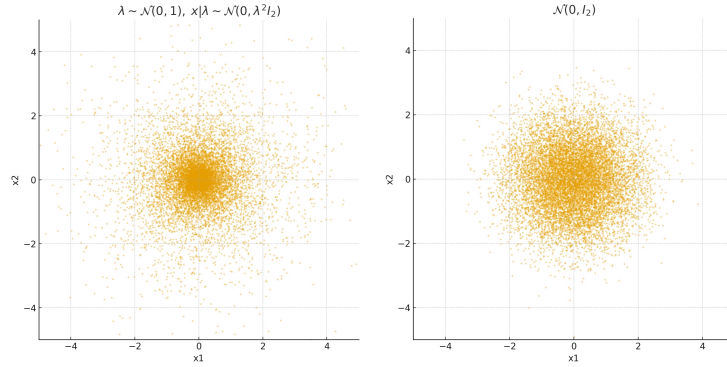


Figure 1: Comparison between a 2D scale mixture of Gaussians and a 2D standard Gaussian. The radial distribution (norm) under the scale mixture is more anti-concentrated.

Our planted distribution has a significant mass on a d dimensional subspace S ; this setting with $d = O(1)$ corresponds to the easiest setting of the robust subspace recovery problem. In fact, the lower-bound result even works when the subspace dimension $d = 0$ i.e., $S = \{0\}$. The null distribution Q_{rot} is rotationally invariant: there is no low-dimensional subspace which has non-zero probability mass. Moreover, there is an elementary algorithm that distinguishes between the two distributions: take $O(d/\alpha)$ samples, and check to see if there are $d + 1$ points lying on a d -dimensional subspace!

Our first result shows that the moments of the two distributions NULL and PLANTED exactly match up to degree $k = O(\sqrt{\log n / \log \log n})$. This also implies a low-degree polynomial lower bound up to the same degree.

Theorem 2 (Moment Matching up to degree $k = \tilde{O}(\sqrt{\log n})$) *The two distributions over \mathbb{R}^n given by the NULL model and the family of planted distributions $\{\mathcal{P}_v : v \in \mathbb{S}^{n-1}\}$ parametrized by unit vectors $v \in \mathbb{S}^{n-1}$ as described above, satisfy the following properties:*

1. Q_{rot} is rotationally invariant, while \mathcal{P}_v is a mixture: with probability $\alpha = 1/\text{poly}(n)$, X is drawn from a distribution supported on $d = 1$ dimensional subspace along v ; with probability $1 - \alpha$, X is drawn from a full-dimensional distribution.
2. The first k moments of \mathcal{P}_v and Q_{rot} match for $k = O(\sqrt{\log n / \log \log n})$. Consequently, for any degree- k polynomial $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$|\mathbb{E}_{X \sim \mathcal{P}_v}[f(X)] - \mathbb{E}_{X \sim Q}[f(X)]| = 0.$$

3. There is an algorithm that runs in polynomial time (in n) and distinguishes between P and Q_{rot} with probability at least $1 - o(1)$. Moreover this algorithm is robust, both to adversarial perturbations of the points, and to rerandomizing a $p \in [0, 1)$ fraction of the samples from P to be drawn from Q_{rot} .

See Theorem 9 for the formal statement. The above lower bound shows that algorithms that rely on statistics of low-degree polynomials on the samples of degree $k = \tilde{O}(\sqrt{\log n})$ ¹ cannot distinguish

1. We use the notation $\tilde{O}(f(n))$ to denote $O(f(n)\text{poly} \log(f(n)))$, suppressing polylogarithmic factors of the argument.

between the two distributions. In fact $\text{LDA}_{\leq k}^{(m)} = 0$, and it is independent of the number of samples m . This rules out many polynomial time algorithms based on moment-based approaches — for statistical problems in n dimensions, algorithms based on degree k moments typically provide $n^{O(k)}$ guarantees both in terms of sample complexity and running time. However, certain special polynomials of degree $\Theta(\log n)$ can be evaluated in polynomial time; this particularly includes algorithms that are based on good approximations to the top eigenvalue of random matrices.

Our next result shows that the low-degree advantage is small even when the degree $k = n^{\Omega(1)}$. This suggests the absence of any polynomial time (or even quasi-polynomial time) algorithm, thereby making the polynomial time tractability of our problem interesting. In what follows, $c > 0$ can be chosen to be any constant e.g., $c = 1/2$.

Informal Theorem 3 (LDA Lower Bound up to degree $k = n^{\Omega(1)}$) *Let $C > 0$ be any constant. There exist rotational invariant distributions Q_{rot} and P over \mathbb{R}^n such that*

1. P has α probability mass at $X = 0$, while Q_{rot} places zero probability mass on every subspace of dimension at most $n - 1$.
2. For polynomials of degree at most k on m samples,

$$\text{LDA}_{\leq k}^{(m)}(P, Q) \leq \sqrt{(1 + C^2 \alpha^2 k)^m - 1}.$$

In particular, if $k = O(1/(\alpha^2 m))$, then $\text{LDA}_{\leq k}^{(m)}(P, Q_{\text{rot}}) = O(1)$.

3. *There is an algorithm that runs in polynomial time (running time $O(mn)$) and distinguishes between P and Q_{rot} with probability at least $1 - o(1)$. Moreover this algorithm is robust both to adversarial perturbations of the points, and to rerandomizing an ϵ fraction of the samples from P to be drawn from Q_{rot} .*

The formal statement is given in Theorem 12. The LDA upper bound in Theorem 3 (item 2) depends only on α , m and k : the advantage stays $O(1)$ as long as $k = O(1/(\alpha^2 m))$. Taking $\alpha = 1/\text{poly}(n)$ keeps the advantage bounded up to a polynomially large degree $k = n^{\Omega(1)}$. This holds for every dimension $n \geq 1$, since the proof reduces the n -dimensional advantage to a bound on the one-dimensional scale variable λ . When $m\alpha = \Omega(\log n)$, w.h.p. there are many points at $X = 0$ (or close to it, in the robust version) that can be checked easily in polynomial time. Hence, this special case of robust subspace recovery represents a natural statistical problem for which the predictions of the low-degree method fail.

Hopkins’ formal conjecture and why it is not a counterexample. We remark that the above theorem does *not* constitute a counterexample to the *formal* conjecture of Hopkins (2018). The original conjecture of Hopkins (2018) for discrete problems imposed additional restrictions of symmetry (permutation-invariance) and a product form for the null distribution Q_N . In our setting, the null distribution is comprised of m i.i.d. samples from Q_{rot} , which by itself is rotationally invariant: this is a natural symmetry condition for our problem. While Q_{rot} is not a product distribution, the null distribution for our input has a product form across the m samples (in fact, it is i.i.d.). Hence the joint distribution becomes closer to a product distribution when $m \gg n$.

Algorithms and Noise Tolerance. The complementary algorithmic bound for $d = O(1)$ in the above Theorem 3 is achieved by the simple algorithms in Section 5. The algorithm samples $m = O(1/\alpha)$ points and checks whether there exists $d + 1$ samples which are close to being linearly dependent. Note that the planted distribution for our algorithmic result is more general than the specific planted distribution in the lower bound (Theorems 2 and 3). See Theorems 14 and 26 for formal statements.

Moreover, the algorithm is also noise tolerant: it is robust in several models of noise that are natural for the robust subspace recovery problem. First, a $p \in [0, 1]$ fraction of the samples drawn from the PLANTED distribution can be re-randomized to be drawn from the NULL distribution Q_{rot} , as in many random noise models for other problems. The work of (Holmgren and Wein, 2020) gave a counterexample based on error-correcting codes with this noise model, and proposed considering robustness to perturbations of each point in the real domain e.g., a small Gaussian perturbation to every entry when the NULL distribution is a standard spherical Gaussian $N(0, I)$.

For our NULL distribution Q_{rot} , there is no obvious choice for a natural average-case noise model as the points have varying length scales. We instead show robustness of the algorithm against two natural notions of adversarial perturbations \tilde{X} to each point X . The first kind are perturbations whose relative error is upper bounded in magnitude by a constant i.e., $\|\tilde{X} - X\|_2 \leq \epsilon \|X\|_2$ for $\epsilon = O_d(1)$. The second kind (additive error worst-case perturbations) bounds the magnitude of the $\|\tilde{X} - X\|_2 \leq \eta$ in an absolute sense (and is more meaningful when $S = \{0\}$). The ranges of ϵ, η allowed by the algorithm are within a constant factor of the best possible even for inefficient algorithms in this model. Please see Theorem 14 (relative error worst-case perturbations) and Theorem 26 (additive error worst-case perturbations) in Section 5 for formal statements.

Failure of low-degree polynomials in capturing anti-concentration. Our results together identify a natural candidate setting (robust subspace recovery) where the low-degree polynomial method fails. Perhaps more interestingly, it suggests that the low-degree method does not adequately capture algorithms that are based on *anti-concentration properties* of the distribution. This is unlike concentration properties and tail bound which are often captured by high-degree moments of the distribution. The algorithms for robust subspace recovery (Hardt and Moitra, 2013; Bhaskara et al., 2019) indeed rely on identifying a small but non-trivial α fraction of samples that are located in a very small region — a well-spread anti-concentrated distribution is unlikely to have any such high density region. In fact, the recent work of (Bakshi et al., 2024) studies the problem of certifying anti-concentration with a view towards obtaining guarantees for clustering problems for general distributions. They construct low-degree polynomials to certify that there is no small halfspace of small width α that contains $> C\alpha$ probability mass. However, the degree of the certificate grows as $\Omega(1/\alpha^2)$. For robust subspace recovery, we are in the regime where $\alpha = 1/\text{poly}(n)$. This seems to be challenging for such techniques including (Bakshi and Kothari, 2021; Bakshi et al., 2024).

New candidate instance for separations The main contribution of this work is a natural high-dimensional problem (an average-case setting of robust subspace recovery) which may serve as a candidate for proving separation results between different classes of algorithms. We believe that our candidate instance (and variants) could present a challenge to other algorithm classes including the Statistical Query (SQ) model, and the Sum-of-Squares (SoS). Concretely, we conjecture that the average-case robust subspace recovery problem studied in this work (e.g., in Theorem 2, or a generalization with an α -mass on a d -dimensional subspace, in the computationally tractable range) is also a hard instance for polynomial time algorithms that fit within the SQ framework.

Conjecture 4 (Conjectured Hardness for Statistical Query Algorithms) *Consider the hypothesis testing problem of distinguishing between the NULL distribution Q_{rot} and the family of PLANTED distributions with an α mass on a hidden $d = 1$ dimensional subspace as in Theorem 2. For $\alpha_k = 1/\text{poly}(n)$, any SQ algorithm that succeeds with probability $1 - o(1)$ requires either a super-polynomial $n^{\omega(1)}$ number of queries, or at least one query with tolerance $n^{-\omega(1)}$.*

This conjecture is motivated by two main observations. First, our low-degree lower bound (Theorem 2) shows that the first $k = O(\sqrt{\log n / \log \log n})$ moments of the single-sample distributions match exactly. This perfect moment-matching provides strong evidence of indistinguishability for any algorithm that relies on low-degree statistics. Connections to low-degree method lower bounds are also known in certain settings (Brennan et al., 2021). Second, our simple algorithms and the existing algorithms for robust subspace recovery (Hardt and Moitra, 2013; Bhaskara et al., 2019) for sub-constant α rely on identifying approximate linear dependence among a relatively small subset of the sampled points. This procedure is not believed to be captured by the SQ model, which is restricted to estimating aggregate statistical properties of the entire population. Moreover such certificates of anti-concentration seem quite challenging even for powerful algorithm classes like sum-of-squares, where current techniques seem to require degree bounds that are $\text{poly}(1/\alpha)$ (Bakshi et al., 2024).

We note that SQ hardness does not follow automatically from our low-degree lower bound via the SQ/low-degree equivalence of Brennan et al. (2021): their equivalence requires niceness conditions on the null (e.g., noise-robust, product-type structure) that fail here, since our planted distribution places an atom on a subspace and is singular with respect to Q_{rot} (and remains so after smoothing). We therefore view the instance as a genuine candidate for a separate SQ lower bound rather than a consequence of an existing reduction.

Comparison to Non-Gaussian Component Analysis (NGCA). Our results differ from typical hardness results in settings like NGCA (Blanchard et al., 2006; Diakonikolas et al., 2017, 2023). The objective in NGCA is typically to find a single hidden non-Gaussian direction in high dimensions. Lower bound constructions based on the NGCA problem and its variants have played a major role in proving evidence of hardness for several average-case problems (Diakonikolas et al., 2017; Bruna et al., 2021). For example, recent work (Diakonikolas et al., 2023) shows that a subspace recovery problem is hard for low-degree polynomials and Statistical Query algorithms when the non-Gaussian signal is confined to a subspace of co-dimension 1 (i.e., an $(n - 1)$ -dimensional subspace). However, in our constructions, we have an α fraction of mass on a 1-dimensional or 0-dimensional subspace; this makes an entire subspace of co-dimension $\geq n - 1$ to be non-Gaussian. In fact, this different flavor is expected – while the NGCA is believed to be computationally hard, our construction where the subspace is low-dimensional is algorithmically easy. This makes it significantly more difficult to design a planted distribution that matches all the mixed moments of the reference distribution. We hope that our new instance which feels significantly different from the NGCA instance will be helpful in understanding the comparative power of different algorithmic frameworks.

1.2. Technical Overview

Our main result is a separation, which we establish in two parts. First, we construct a hypothesis testing problem where a family of planted distributions P and a rotationally invariant null distribution Q_{rot} are indistinguishable by low-degree polynomials. We establish this indistinguishability

in two strong forms: exact moment matching up to degree $k = \tilde{O}(\sqrt{\log n})$, and a vanishing Low-Degree Advantage for degrees $k = n^{\Omega(1)}$. Both of these results leverage nice structural properties of the distribution Q_{rot} . Second, we provide a simple, efficient algorithm that robustly solves the same problem.

The Moment Matching Construction. The core of our first lower bound is the moment-matching claim. We first construct our null distribution Q as a scale mixture of Gaussians. This choice is crucial. A standard Gaussian is rotationally invariant, but its length ($\|X\|_2$) is highly concentrated. We require our null distribution Q to satisfy two properties: (1) *rotational invariance*, and (2) *anti-concentration of its length (scale)*.

Our chosen Q (a scale mixture with the scale $\lambda \sim \mathcal{N}(0, 1)$) satisfies both. Since both P and Q are designed to be rotationally invariant in the $(n - 1)$ -dimensional subspace irrelevant to the signal, two consequences follow: 1) The null distribution Q provides no information about the hidden subspace. 2) The moment-matching problem simplifies dramatically. We can analyze the moments along any arbitrary vector in the irrelevant subspace, reducing the n -dimensional problem to a 2D problem (one signal coordinate, one irrelevant coordinate).

We then focus on matching the 2D moment vectors of P and Q . Our strategy is to give a non-constructive proof to show the existence of P by effectively perturbing the moments of Q . The main technical challenge is to prove that this new, perturbed set of moments is valid, i.e., a distribution μ_2 defining them actually exists. In other words, we want to show that the moments of Q are sufficiently far from the boundary of the moment polytope.

We first establish a strong, two-sided anti-concentration bound, which relies on the Carbery-Wright inequality. Then a crucial step is to translate this two-sided bound into the one-sided bound required to lower-bound the Tukey depth. This is encapsulated in the following lemma.

Lemma 5 (Same as Lemma 19) *Suppose ν is a probability measure over \mathbb{R}^n that satisfies the following anti-concentration statements: for any multivariate polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$ of degree at most k ,*

$$\forall \delta > 0, \mathbb{P}_{x \sim \nu} \left[|p(x) - \mathbb{E}_\nu[p]| \leq \delta \cdot \sqrt{\text{Var}(p(x))} \right] \leq \eta_k(\delta).$$

Then for any $\delta > 0$ and any $\delta_1 > 0$ such that $\eta_k(\delta_1) \leq 1/2$, we have

$$\mathbb{P}_{x \sim \nu} \left[p(x) \geq \mathbb{E}_\nu[p(x)] + \delta \cdot \sqrt{\text{Var}(p(x))} \right] \geq \frac{\delta_1^2}{16} - \eta_k(\delta).$$

This positive Tukey depth guarantees that our perturbed moment vector is in the interior of the convex hull, which proves the existence of our moment-matching instance up to degree $k = O(\sqrt{\log n / \log \log n})$. These properties of the null distribution Q_{rot} also help in showing bounded low-degree advantage when for polynomial degree k .

Extending Low-Degree Method Hardness to Polynomial Degrees. While our moment-matching construction completely fools polynomials up to degree $k = \tilde{O}(\sqrt{\log n})$, standard heuristics suggest that low-degree lower bounds are strong evidence of hardness if they rule out polynomial of degree $k = \omega(\log(n))$. To robustly demonstrate that the low-degree framework fails to predict the tractability of this problem, we establish a bound on LDA for even higher degrees, i.e., $k = n^{\Omega(1)}$. We use the same null distribution Q_{rot} as in our moment-matching result. Its key structural properties, rotational invariance and strong anti-concentration of its length, remain crucial in proving a bound on the low-degree advantage.

We analyze a simplified planted distribution $P = (1 - \alpha)Q + \alpha\Delta_0$, which mixes the null distribution with a point mass at the origin 0. Using the anti-concentration of the length of Q , we first prove that the single-sample advantage $\text{LDA}_{\leq k}^{(1)}$ is $O(1)$. A key claim that allows us to prove the bound on the single-sample advantage is the following:

Claim 6 *Consider any degree- k polynomial f whose constant term is 0. Then we have*

$$\frac{|\mathbb{E}_Q[f(X)]|}{\sqrt{\text{Var}_Q(f(X))}} \leq O(\sqrt{k}),$$

where $Q = Q_{\text{rot}}$ is the rotational invariant scale mixture of Gaussians.

We combine this claim with a tensorization argument (in Lemma 13) to lift this to the multi-sample setting, showing that the multi-sample advantage remains bounded by a small constant. The bound in the above claim (for a single sample) is independent of the dimension n . We remark that this claim is not true for a spherical Gaussian $N(0, I)$ even though it is rotationally invariant; for example the degree-2 polynomial $f(X) = X^\top X$ is concentrated around $n \pm O(\sqrt{n})$. An important property that we exploit about Q_{rot} , we can write $X \sim Q_{\text{rot}}$ as a *product* of two independent random variables $X = \lambda g$, where $g \in \mathbb{R}^n$ is a random vector (drawn from a rotationally invariant distribution), and λ is a random variable capturing the scale; in our case, it is also drawn according to the normal distribution which has good anti-concentration properties. This allows us to establish the above claim (see Proof of Theorem 12). We remark that the low-degree lower bound construction also works for $n = 1$ dimension (where one can view $g \sim_u \{-1, +1\}$). The choice of a normal random variable for λ is not crucial; one can use other choices of the scale distribution λ that satisfy Claim 6 (potentially with a different dependence on k) to achieve other desirable properties.

The Simple Robust Algorithms. In contrast to the low-degree hardness, we show that this problem is computationally tractable by presenting two simple algorithms that sample subsets of points and check for approximate linear dependence. For d -dimensional hidden subspace, the first algorithm handles adversarial relative perturbations by normalizing samples and detecting if the $(d + 1)$ -th singular value of any subcollection of size $d + 1$ drops below a constant threshold. The second algorithm extends the algorithmic guarantees to the setting of worst-case additive perturbations. Both algorithms rely on the anti-concentration of Q_{rot} to ensure that random tuples are robustly linearly independent (soundness).

2. Preliminaries

2.1. Lifting Moment Matching to the Multi-Sample Setting

Moment Matching. We say two distributions P and Q over \mathbb{R}^n have matching moments up to degree k if for any multivariate monomials $M : \mathbb{R}^n \rightarrow \mathbb{R}$ of total degree at most k , we have

$$\mathbb{E}_{X \sim P}[M(X)] = \mathbb{E}_{X \sim Q}[M(X)].$$

The definition of LDA involves polynomials over multiple i.i.d. samples from a distribution. The following standard lemma shows that the single-sample moment matching property defined above directly implies that the expectation of any low-degree polynomial over the corresponding multi-sample distributions will also match. This allows us to focus our construction for the low-degree lower bound on a single-sample distribution.

Lemma 7 *Let P and Q be two distributions over \mathbb{R}^n whose moments match up to degree k . Let m be the number of samples, and let $P^{\otimes m}$ and $Q^{\otimes m}$ be the product distributions for m i.i.d. samples. Then for any polynomial $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ of total degree at most k , we have*

$$\mathbb{E}_{(X_1, \dots, X_m) \sim P^{\otimes m}} [f(X_1, \dots, X_m)] = \mathbb{E}_{(Y_1, \dots, Y_m) \sim Q^{\otimes m}} [f(Y_1, \dots, Y_m)].$$

3. Moment Matching Instance

In this section, our goal is to build a planted distribution whose first k moments agree with those of a rotationally invariant null distribution Q , while still hiding a 1-dimensional signal direction. Concretely, we use the scale parameter λ to control the overall spherical spread and a scalar z to carry the signal in one coordinate; the construction lets us tune a small mass α_k that lives on a 1-dimensional subspace while matching all mixed moments up to degree k . We also prove that the moments vectors of our planted and null distributions match exactly using the tools that are introduced in Section 3.1.

Definition 8 (Scale mixture of Gaussians) *An n -dimensional scale mixture of Gaussians is a distribution of X on \mathbb{R}^n where its scale parameter λ is governed by a distribution μ :*

$$\lambda \sim \mu; X \mid \lambda \sim \mathcal{N}(0, \lambda^2 I_n).$$

Theorem 9 *Let Q be a scale mixture of Gaussians with Gaussian scales $\mu = \mathcal{N}(0, 1)$ as in Definition 8. For any positive integer k , there exist $\alpha_k \in (0, 1)$ satisfying $\alpha_k \geq k^{-O(k^2)}$, and distributions $\mu_1(z), \mu_2(\lambda, z)$ that define a new distribution P , which has its first k moments matching those of Q . P is defined as follows for $X = (\bar{X}, X_n)$ over \mathbb{R}^n , where $\bar{X} \in \mathbb{R}^{n-1}, X_n \in \mathbb{R}$:*

- with probability α_k , $\bar{X} = 0, X_n \sim \mu_1$,
- with probability $1 - \alpha_k$, $(\lambda, z) \sim \mu_2; \bar{X} \mid \lambda \sim \mathcal{N}(0, \lambda^2 I_{n-1}), X_n = z$.

Our proof strategy is to first establish the moment-matching conditions, which, by rotational invariance, simplifies our task to matching the mixed moments of just two coordinates. We will show that the moments of our planted distribution P can be expressed as a convex combination of two components. The main technical challenge is to prove that the planted distribution, defined by distributions μ_1 and μ_2 , can be constructed to satisfy the resulting system of moment equations. To do this, we will reframe the problem as finding a target moment vector θ within the convex hull of the support of a related moment vector Φ . We then leverage the powerful tools related to anti-concentration and the geometry of moment spaces developed in Section 3.1 to prove that such a distribution μ_2 must exist.

3.1. Anticoncentration and Tukey Depth

As outlined in the proof of Theorem 9, our moment-matching construction hinges on proving the existence of a specific distribution μ_2 . To do this, we must show that a particular target vector of moments θ lies within the convex hull of the support of a random moment vector $\Phi(\lambda, z)$. This section develops the necessary anti-concentration tools to prove this claim.

We begin by defining a two-dimensional distribution ν over (λ, z) , which will serve as the basis for the random moment vector $\Phi(\lambda, z)$ used in Section 3. Consider the joint distribution ν over $\{(\lambda, z) : \lambda, z \in \mathbb{R}\}$ defined as follows:

1. $\lambda \in \mathbb{R}$ is drawn from a standard Gaussian $\mathcal{N}(0, 1)$
2. Conditioned on λ , $z \in \mathbb{R}$ is drawn from $\mathcal{N}(0, \lambda^2)$.

In other words, the joint distribution is given by

$$\nu(\lambda, z) = \frac{1}{\sqrt{2\pi}} e^{-\lambda^2/2} \cdot \frac{1}{|\lambda|\sqrt{2\pi}} e^{-z^2/(2\lambda^2)}. \quad (2)$$

The reference or null distribution over \mathbb{R}^n will be constructed based on the above two-dimensional measure ν . The parameter λ will determine the scale of the distribution, and z will determine the component along a special one-dimensional projection.

3.2. Anticoncentration Bounds

The following anti-concentration property of the distribution ν will be important for the lower bound construction. It ensures that no single value (or narrow band) has too much probability mass under any low-degree polynomial of our reference distribution. This property will later be used in Section B.1 to establish a positive Tukey depth for our moment vector, which is the key to proving its mean is in the interior of the convex hull of all valid moment vectors.

Lemma 10 *Suppose $k \in \mathbb{N}$. Consider the degree k polynomial*

$$p(\lambda, z) = \sum_{\ell_1, \ell_2 \text{ even}, 0 < \ell_1 + \ell_2 \leq k} \beta_{\ell_1, \ell_2} \lambda^{\ell_1} z^{\ell_2}$$

where β is the vector of coefficients $\beta = (\beta_{\ell_1, \ell_2} : \ell_1, \ell_2 \in \{0, 1, \dots, k\} \text{ s.t. } \ell_1 + \ell_2 \leq k)$. Then there exists a universal constant C such that

$$\forall \delta > 0, \sup_{t \in \mathbb{R}} \mathbb{P}_{(\lambda, z) \sim \nu} \left[|p_\beta(\lambda, z) - t| \leq \delta \cdot |\mathbb{E}[p_\beta]| \right] \leq C \cdot k \cdot \left(\sqrt{k} \delta \right)^{\frac{1}{2k}}. \quad (3)$$

We will use the following lemma that compares the mean and the variance of low-degree polynomials under a Gaussian to prove Lemma 10.

Lemma 11 *Let q be a polynomial in (λ, g) of degree at most k with zero constant term. Under $(\lambda, g) \sim N(0, I_2)$,*

$$|\mathbb{E}[q(\lambda, g)]| \leq C \sqrt{k \text{Var}(q(\lambda, g))}.$$

4. Extending Hardness to Higher Degrees via Bounded LDA

In this section, we prove the first part of Theorem 3. We extend our hardness result to polynomials of degree $k = \text{poly}(n)$ demonstrating that the LDA remains bounded in this regime. This effectively rules out any efficient algorithm that relies solely on low-degree polynomials. We will use the same null distribution Q as in Section 3, i.e., Q is a scale mixture of Gaussians with Gaussian scales $\mu = \mathcal{N}(0, 1)$. Define the planted distribution $P = (1 - \alpha)Q + \alpha 0$.

Theorem 12 (LDA Bound up to degree $k = n^{\Omega(1)}$) *For polynomials of degree at most k on m samples,*

$$LDA_{\leq k}^{(m)}(P, Q) \leq \sqrt{(1 + C^2 \alpha^2 k)^m - 1}.$$

In particular, if $k = O(1/(\alpha^2 m))$, then $LDA_{\leq k}^{(m)}(P, Q) = O(1)$.

The proof of the theorem has two steps. We first show a single-sample LDA bound and then lift it to m samples. The following lemma shows how low-degree advantage scales from one sample to m i.i.d. samples for any distribution P and Q .

Lemma 13 *Let P, Q be distributions on \mathbb{R}^n with finite moments up to degree $2k$. Suppose the single-sample low-degree advantage is $\text{LDA}_{\leq k}^{(1)}(P, Q) = \delta$. Then for any m ,*

$$\text{LDA}_{\leq k}^{(m)}(P, Q)^2 \leq (1 + \delta^2)^m - 1.$$

Proof [Proof of Theorem 12] We first bound the single-sample low-degree advantage. Let f be any polynomial of degree at most k on \mathbb{R}^n . Since adding a constant to f does not change LDA, we may assume the constant term of f is zero. Then

$$\mathbb{E}_P[f] - \mathbb{E}_Q[f] = -\alpha \mathbb{E}_Q[f].$$

Let $X \sim Q$. We can write $X = \lambda g$ with independent r.v. $g \sim \mathcal{N}(0, I_n)$ and $\lambda \sim \mathcal{N}(0, 1)$. Define $h(\lambda) = \mathbb{E}_g[f(\lambda g)]$. Then $\mathbb{E}_\lambda[h(\lambda)] = \mathbb{E}_Q[f(X)]$ and by the law of total variance,

$$\text{Var}_Q(f(X)) \geq \text{Var}_\lambda(h(\lambda)).$$

Moreover $h(0) = 0$ i.e., $h(\lambda)$ has no constant term. Hence we have reduced it to an analogous claim about the one dimensional scale distribution λ . We will now upper bound $|\mathbb{E}_\lambda[h(\lambda)]|/\sqrt{\text{Var}_\lambda[h(\lambda)]}$. Apply Lemma 11 to h by taking $q(\lambda, g) = h(\lambda)$,

$$|\mathbb{E}_Q[f(X)]| = \mathbb{E}_\lambda[h(\lambda)] \leq C\sqrt{k\text{Var}_\lambda(h(\lambda))} \leq C\sqrt{k\text{Var}_Q(f(X))}.$$

Thus,

$$\text{LDA}_{\leq k}^{(1)}(P, Q) \leq C\alpha\sqrt{k}.$$

Apply Lemma 13 with $\delta = C\alpha\sqrt{k}$,

$$\text{LDA}_{\leq k}^{(m)}(P, Q) \leq \sqrt{(1 + C^2\alpha^2k)^m - 1}.$$

■

The specific choices of g and λ being Gaussian were for convenience; in Appendix C.1 we explain precisely which structural property of the scale distribution μ (captured by the Christoffel function of μ at 0) governs the single-sample LDA bound, and how the argument generalizes to other scale distributions.

5. Simple Noise-Tolerant Algorithm

In this section we show that there are simple efficient algorithms that are not based on the low-degree polynomial method that can solve this problem efficiently. Consider the setting when the $\alpha = 1/\text{poly}(n)$ fraction lies on a $d \leq 1$ dimensional subspace. An elegant and clever algorithm for robust subspace recovery due to Hardt and Moitra (Hardt and Moitra, 2013) uses anti-concentration properties of $(1 - \alpha)$ points not on the subspace S : it gives efficient algorithms when $\alpha \geq \frac{\dim(S)}{n}$. Subsequent work by Bhaskara, Chen, Perreault and Vijayaraghavan (Bhaskara et al., 2019) gave

an improved algorithm in the smoothed analysis setting where the points not on the subspace are randomly perturbed (by a random vector of average length $\rho = 1/\text{poly}(n)$), and gave guarantees of recovery when $\alpha \gtrsim (d/n)^\ell$ for any constant $\ell > 0$ in $(nd)^{O(\ell)}$ time. In our setting where $d = O(1)$ and $\alpha = 1/n^{O(1)}$, we can pick an appropriate $\ell = O(1)$ to get algorithmic guarantees. However this algorithm is only moderately noise tolerant: every point can be adversarially perturbed (moved) by $1/n^{O(\ell)}$. We will instead give a simple algorithm specialized for the $d = O(1)$ setting that is elementary and also allows points to be moved by an amount $\epsilon = \Omega(1)$.

We consider the hypothesis testing version of the problem. We are given samples from a distribution \mathcal{D} over \mathbb{R}^n , and the goal is to distinguish between the following two hypotheses:

- **NULL:** $\mathcal{D} = Q_{\text{rot}}$ where Q_{rot} is the rotationally invariant distribution defined in Section 3.
- **PLANTED:** \mathcal{D} is any distribution with at least an α probability mass supported on a planted subspace S of dimension d .

Note that the PLANTED case here is more general than the planted distribution in the lower bound example. Our algorithms are noise tolerant: the m i.i.d. samples may be re-randomized (each redrawn from Q_{rot} independently with probability p) and then each point adversarially perturbed to $\tilde{x}_i = x_i + z_i$, bounded in relative error $\|z_i\| \leq \epsilon \|x_i\|$ or additive error $\|z_i\| \leq \eta$ (Definitions 22 and 25). Both algorithms check for approximate linear dependencies between $(d + 1)$ -tuples of samples; the first tolerates relative-error perturbations of $\epsilon = O_d(1)$.

Theorem 14 (Noise tolerant algorithm with relative error perturbations) *There exists a universal constants $c, C > 0$ such that the following holds when $m \geq C(d + 1)/(\alpha(1 - p))$ and any $0 \leq \epsilon \leq \frac{1}{8(d+1)}$. There is an algorithm (given in Figure 2) that runs in time $O(\binom{m}{d+1})$ that given as input noisy samples $\tilde{x}_1, \dots, \tilde{x}_m$ as described above with $\|\tilde{x}_i - x_i\| \leq \epsilon \|x_i\|$, can solve with high probability the testing problem.*

The following algorithmic result handles the setting where there is a worst-case perturbation of (additive) magnitude at most η as described below.

Theorem 15 (Noise tolerant algorithm with additive error perturbations) *There exists a universal constants $c, c' > 0, C \geq 1$ such that the following holds when $m \geq 2(d + 1) \log(4/\delta)/(\alpha(1 - p))$ and any $\delta \in (0, 1), 0 \leq \eta \leq \frac{\alpha(1-p)\sqrt{n}}{C(d+1)\log^2(4/\delta)^2}$. Then there is an algorithm (given in Figure 3 with $\tau = \frac{\alpha(1-p)\sqrt{n}}{c\log^2(4/\delta)}$) running in time $O(\binom{m}{d+1})$ that given as input $\tilde{x}_1, \dots, \tilde{x}_m$ generated as described above with $\|\tilde{x}_i - x_i\|_2 \leq \eta$ can solve the testing problem with probability at least $1 - \delta$.*

Acknowledgments

We thank Alex Wein, Miklos Racz and Sidhanth Mohanty for feedback on an early draft, and Sanchit Kalhan for helpful discussions related to the moment matching bounds. The authors were supported by the NSF-funded Institute for Data, Econometrics, Algorithms and Learning (IDEAL) through the grant NSF ECCS-2216970.

References

- Gabriel Arpino and Ramji Venkataramanan. Statistical-computational tradeoffs in mixed sparse linear regression. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 921–986. PMLR, 12–15 Jul 2023. URL <https://proceedings.mlr.press/v195/arpino23a.html>.
- Ainesh Bakshi and Pravesh K Kothari. List-decodable subspace recovery: Dimension independent error in polynomial time. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1279–1297. SIAM, 2021.
- Ainesh Bakshi, Pravesh Kothari, Goutham Rajendran, Madhur Tulsiani, and Aravindan Vijayaraghavan. Efficient certificates of anti-concentration beyond gaussians. In *Proceedings of the 65th IEEE Conference on Foundations of Computer Science (to appear)*, 2024. URL <https://arxiv.org/abs/2405.15084>.
- Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.
- Quentin Berthet and Philippe Rigollet. Computational lower bounds for sparse pca. *arXiv preprint arXiv:1304.0828*, 2013.
- Aditya Bhaskara, Aidao Chen, Aidan Perreault, and Aravindan Vijayaraghavan. Smoothed analysis in unsupervised learning via decoupling. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2019.
- Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny, and Klaus-Robert Müller. In search of non-gaussian components of a high-dimensional distribution. *Journal of Machine Learning Research*, 7(9):247–282, 2006. URL <http://jmlr.org/papers/v7/blanchard06a.html>.
- Matthew Brennan and Guy Bresler. Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. In *Conference on Learning Theory*, pages 469–470. PMLR, 2019.
- Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.
- Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 774–774. PMLR, 15–19 Aug 2021. URL <https://proceedings.mlr.press/v134/brennan21a.html>.
- Joan Bruna, Oded Regev, Min Jae Song, and Yi Tang. Continuous lwe. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 694–707, 2021.

- Rares-Darius Buhai, Pravesh K. Kothari, and David Steurer. Algorithms approaching the threshold for semi-random planted clique. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 1918–1926, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399135. doi: 10.1145/3564246.3585184. URL <https://doi.org/10.1145/3564246.3585184>.
- Rares-Darius Buhai, Jun-Ting Hsieh, Aayush Jain, and Pravesh K Kothari. The quasi-polynomial low-degree conjecture is false. *arXiv preprint arXiv:2505.17360*, 2025.
- A Carbery and J Wright. Distributional and l-q norm inequalities for polynomials over convex bodies in r-n. *Mathematical research letters*, 8(3):233–248, May 2001. ISSN 1073-2780.
- Ilias Diakonikolas and Daniel Kane. Non-gaussian component analysis via lattice basis reduction. In *Conference on Learning Theory*, pages 4535–4547. PMLR, 2022.
- Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.
- Ilias Diakonikolas, Daniel Kane, Lisheng Ren, and Yuxin Sun. Sq lower bounds for non-gaussian component analysis with weaker assumptions. *Advances in Neural Information Processing Systems*, 36:4199–4212, 2023.
- Yunzi Ding, Dmitriy Kunisky, Alexander S. Wein, and Afonso S. Bandeira. The average-case time complexity of certifying the restricted isometry property. *IEEE Trans. Inf. Theor.*, 67(11): 7355–7361, November 2021. ISSN 0018-9448. doi: 10.1109/TIT.2021.3112823. URL <https://doi.org/10.1109/TIT.2021.3112823>.
- Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2): 1–37, 2017.
- David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021. doi: 10.1073/pnas.2108492118. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2108492118>.
- Chao Gao, Liren Shan, Vaidehi Srinivas, and Aravindan Vijayaraghavan. Learning confidence ellipsoids and applications to robust subspace recovery, 2026. URL <https://arxiv.org/abs/2512.16875>.
- Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 354–375, Princeton, NJ, USA, 12–14 Jun 2013. PMLR. URL <https://proceedings.mlr.press/v30/Hardt13.html>.
- S Hayakawa, T Lyons, and H Oberhauser. Estimating the probability that a given vector is in the convex hull of a random sample. *Probability Theory and Related Fields*, 185(3-4):705–746, 2023.

- Justin Holmgren and Alexander S Wein. Counterexamples to the low-degree conjecture. *arXiv preprint arXiv:2004.08454*, 2020.
- Samuel B. Hopkins. *Statistical Inference and the Sum of Squares Method*. PhD thesis, Cornell University, 2018.
- Samuel B Hopkins and David Steurer. Bayesian estimation from few samples: community detection and related problems. *arXiv preprint arXiv:1710.00264*, 2017.
- Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.
- Jun-Ting Hsieh, Daniel M. Kane, Pravesh K. Kothari, Jerry Li, Sidhanth Mohanty, and Stefan Tiegel. Rigorous implications of the low-degree heuristic, 2026. URL <https://arxiv.org/abs/2601.05850>.
- Michael Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, November 1998. ISSN 0004-5411. doi: 10.1145/293347.293351. URL <https://doi.org/10.1145/293347.293351>.
- Dmitriy Kunisky. Low coordinate degree algorithms i: Universality of computational thresholds for hypothesis testing. *The Annals of Statistics*, 53, 04 2025. doi: 10.1214/24-AOS2484.
- Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.
- Gilad Lerman and Tyler Maunu. An overview of robust subspace recovery. *Proceedings of the IEEE*, 106(8):1380–1410, August 2018. ISSN 1558-2256. doi: 10.1109/jproc.2018.2853141. URL <http://dx.doi.org/10.1109/JPROC.2018.2853141>.
- Ankur Moitra and Alexander S. Wein. Precise error rates for computationally efficient testing. *The Annals of Statistics*, 2023. URL <https://api.semanticscholar.org/CorpusID:264832918>.
- Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM (JACM)*, 56(6):1–40, 2009.
- John W Tukey. Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, volume 2, pages 523–531. Vancouver, 1975.
- Alexander S Wein. Computational complexity of statistics: New insights from low-degree polynomials. *arXiv preprint arXiv:2506.10748*, 2025.
- Ilias Zadik, Min Jae Song, Alexander S Wein, and Joan Bruna. Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory*, pages 1247–1248. PMLR, 2022.

Appendix A. Additional Preliminaries

Lemma 7 *Let P and Q be two distributions over \mathbb{R}^n whose moments match up to degree k . Let m be the number of samples, and let $P^{\otimes m}$ and $Q^{\otimes m}$ be the product distributions for m i.i.d. samples. Then for any polynomial $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ of total degree at most k , we have*

$$\mathbb{E}_{(X_1, \dots, X_m) \sim P^{\otimes m}} [f(X_1, \dots, X_m)] = \mathbb{E}_{(Y_1, \dots, Y_m) \sim Q^{\otimes m}} [f(Y_1, \dots, Y_m)].$$

Proof By linearity of expectation, it suffices to prove the equality for any monomial $M(X_1, \dots, X_m)$ of total degree at most k . We can write any such monomial M as a product of per-sample monomials:

$$M(X_1, \dots, X_m) = \prod_{i=1}^m M_i(X_i)$$

where the total degree is $\sum_{i=1}^m \deg(M_i) \leq k$ and hence each $\deg(M_i) \leq k$. By independence of the samples,

$$\mathbb{E}_{P^{\otimes m}} [M] = \mathbb{E}_{P^{\otimes m}} \left[\prod_{i=1}^m M_i(X_i) \right] = \prod_{i=1}^m \mathbb{E}_{X_i \sim P} [M_i(X_i)].$$

By the moment matching assumption,

$$\prod_{i=1}^m \mathbb{E}_{X_i \sim P} [M_i(X_i)] = \prod_{i=1}^m \mathbb{E}_{Y_i \sim Q} [M_i(Y_i)] = \mathbb{E}_{Q^{\otimes m}} [M].$$

This proves the claim. ■

A.1. Hermite Polynomials

We use the standard probabilist's Hermite polynomials, denoted by $\text{He}_m(x)$. For our analysis, we will use the normalized Hermite polynomials, defined as:

$$h_m(x) = \frac{\text{He}_m(x)}{\sqrt{m!}}.$$

The most important property of these normalized polynomials is that they form a complete orthonormal basis for the space of square-integrable functions with respect to the standard Gaussian measure $\mathcal{N}(0, 1)$, i.e., for any $i, j \geq 0$,

$$\mathbb{E}_{x \sim \mathcal{N}(0,1)} [h_i(x) h_j(x)] = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

This orthonormality also extends to higher dimensions. For reference, these normalized Hermite polynomials also have the following standard identity, which relates them to the monomial basis:

$$h_m(x) = \sum_{j=0}^{\lfloor \frac{m}{2} \rfloor} a_{m,j} x^{m-2j}, \quad a_{m,j} = \frac{(-1)^j \sqrt{m!}}{2^j j! (m-2j)!}. \quad (4)$$

Appendix B. Technical Details in Section 3

This section contains the construction details for the moment matching lower bound.

Theorem 9 *Let Q be a scale mixture of Gaussians with Gaussian scales $\mu = \mathcal{N}(0, 1)$ as in Definition 8. For any positive integer k , there exist $\alpha_k \in (0, 1)$ satisfying $\alpha_k \geq k^{-O(k^2)}$, and distributions $\mu_1(z), \mu_2(\lambda, z)$ that define a new distribution P , which has its first k moments matching those of Q . P is defined as follows for $X = (\bar{X}, X_n)$ over \mathbb{R}^n , where $\bar{X} \in \mathbb{R}^{n-1}, X_n \in \mathbb{R}$:*

- with probability α_k , $\bar{X} = 0, X_n \sim \mu_1$,
- with probability $1 - \alpha_k$, $(\lambda, z) \sim \mu_2; \bar{X} \mid \lambda \sim \mathcal{N}(0, \lambda^2 I_{n-1}), X_n = z$.

Proof [Proof of Theorem 9] We first establish the moment-matching conditions. We want to show that for any unit vector $v = (\bar{v}, v_n), \bar{v} \in \mathbb{R}^{n-1}, v_n \in \mathbb{R}$ and for any $\ell \in [k]$,

$$\mathbb{E}_{X \sim P}[\langle X, v \rangle^\ell] = \mathbb{E}_{X \sim Q}[\langle X, v \rangle^\ell].$$

Expanding the inner product gives:

$$\mathbb{E}[\langle X, v \rangle^\ell] = \mathbb{E}[(\bar{X} \cdot \bar{v} + X_n v_n)^\ell] = \sum_{\ell_1, \ell_2 \geq 0, \ell_1 + \ell_2 = \ell} \binom{\ell}{\ell_1} \mathbb{E}[(\bar{X} \cdot \bar{v})^{\ell_1} (X_n v_n)^{\ell_2}].$$

By definition, Q is a mixture of zero-centered Gaussians so it is rotationally invariant. The marginal of the first $n - 1$ coordinates of P is also rotationally invariant, so it suffices to show the equality for any \bar{v} . Without loss of generality, we may take $\bar{v} = e_1$. Therefore, it is equivalent to show that for any $\ell_1, \ell_2 \geq 0$ such that $\ell_1 + \ell_2 \leq k$,

$$\mathbb{E}_P[x_1^{\ell_1} x_n^{\ell_2}] = \mathbb{E}_Q[x_1^{\ell_1} x_n^{\ell_2}]. \quad (5)$$

Furthermore, both distributions are centrally symmetric. This implies that any odd moment is zero. Thus, we only need to match the moments where both ℓ_1 and ℓ_2 are even. The RHS of (5) is

$$\mathbb{E}_Q[x_1^{\ell_1} x_n^{\ell_2}] = \mathbb{E}_{\lambda \sim \mu} \left[\mathbb{E}_{x \sim \mathcal{N}(0, \lambda^2 I)} [x_1^{\ell_1} x_n^{\ell_2}] \right] = \mathbb{E}_{\lambda \sim \mu} [c_{\ell_1} c_{\ell_2} \lambda^{\ell_1 + \ell_2}] = c_{\ell_1} c_{\ell_2} c_{\ell_1 + \ell_2}$$

where c_ℓ is the ℓ -th moment of the standard Gaussian $\mathcal{N}(0, 1)$. When $\ell_1 = 0$, we have the LHS of (5) being

$$\mathbb{E}_P[x_1^{\ell_1} x_n^{\ell_2}] = \alpha_k \mathbb{E}_{z \sim \mu_1} [z^{\ell_2}] + (1 - \alpha_k) \mathbb{E}_{z \sim \mu_2} [z^{\ell_2}].$$

When $\ell_1 \neq 0$, we have

$$\mathbb{E}_P[x_1^{\ell_1} x_n^{\ell_2}] = (1 - \alpha_k) \mathbb{E}_{(\lambda, z) \sim \mu_2} \left[\mathbb{E}_{x_1 \sim \mathcal{N}(0, \lambda^2)} [x_1^{\ell_1} z^{\ell_2}] \right] = (1 - \alpha_k) c_{\ell_1} \mathbb{E}_{(\lambda, z) \sim \mu_2} [\lambda^{\ell_1} z^{\ell_2}].$$

Let $\mathcal{I} = \{(\ell_1, \ell_2) : \ell_1, \ell_2 \text{ even}, 0 < \ell_1 + \ell_2 \leq k\}$ and $D = |\mathcal{I}|$. Let $\nu(\lambda, z)$ be the distribution defined in Section 3.1. Define random variable $\Phi(\lambda, z) \in \mathbb{R}^D$ with coordinates

$$\Phi_{\ell_1, \ell_2}(\lambda, z) = \lambda^{\ell_1} z^{\ell_2}.$$

Then $\mathbb{E}_\nu[\Phi_{\ell_1, \ell_2}(\lambda, z)] = c_{\ell_2} c_{\ell_1 + \ell_2}$. Define our target vector $\theta \in \mathbb{R}^D$

$$\theta_{\ell_1, \ell_2} = \mathbb{E}_{(\lambda, z) \sim \mu_2}[\lambda^{\ell_1} z^{\ell_2}] = \begin{cases} \frac{c_{\ell_2}^2 - \alpha_k \mathbb{E}_{z \sim \mu_1}[z^{\ell_2}]}{1 - \alpha_k}, & \ell_1 = 0, \\ \frac{c_{\ell_2} c_{\ell_1 + \ell_2}}{1 - \alpha_k}, & \ell_1 \geq 1. \end{cases}$$

Let $\beta \in \mathbb{R}^D$ be arbitrary vector of coefficients with $\|\beta\| = 1$. Then

$$p_\beta(\lambda, z) := \langle \beta, \Phi \rangle = \sum_{\mathcal{I}} \beta_{\ell_1, \ell_2} \lambda^{\ell_1} z^{\ell_2}.$$

and $\mathbb{E}_\nu[p_\beta] = \sum_{\mathcal{I}} \beta_{\ell_1, \ell_2} c_{\ell_2} c_{\ell_1 + \ell_2}$. Let $\theta' \in \mathbb{R}^D$ be

$$\theta'_{\ell_1, \ell_2} = (1 + \delta) c_{\ell_2} c_{\ell_1 + \ell_2}.$$

So we have $\langle \beta, \theta' \rangle = (1 + \delta) \mathbb{E}_\nu[p_\beta(\lambda, z)]$. By Corollary 20, we know for any $\delta \leq 1/k^{Ck^2}$,

$$h_\Phi(\theta') = \inf_{\beta: \|\beta\|=1} \mathbb{P}[\langle \beta, \Phi - \theta' \rangle \geq 0] = \inf_{\beta: \|\beta\|=1} \mathbb{P}[p_\beta \geq (1 + \delta) \mathbb{E}_\nu[p]] \geq \eta_k(\delta) > 0.$$

By Theorem 18, θ' is in the convex set of moment vectors $\Phi(\lambda, z)$ of random $(\lambda, z) \sim \nu$. By Lemma 21, pick $r = r(h_\Phi(\theta')) > 0$ so that $B(\theta', r) \subseteq \text{conv}(\text{supp}(\Phi))$. Set $\frac{1}{1 - \alpha_k} = 1 + \delta$. For the $\ell_1 \geq 1$ coordinates, $\theta'_{\ell_1, \ell_2} = \theta_{\ell_1, \ell_2}$. For $\ell_1 = 0$, $\theta_{0, \ell_2} = \theta'_{0, \ell_2} - (1 + \delta) \alpha_k \mathbb{E}_{z \sim \mu_1}[z^{\ell_2}]$. Choose μ_1 so that $\mathbb{E}_{z \sim \mu_1}[z^{\ell_2}]$ moves the $\ell_1 = 0$ block by at most r and hence $\theta \subseteq B(\theta', r) \subseteq \text{conv}(\text{supp}(\Phi))$. Therefore, there exists a valid distribution μ_2 on (λ, z) with $\mathbb{E}_{\mu_2}[\Phi] = \theta$, i.e., all required mixed moments match as in (5). \blacksquare

Lemma 10 *Suppose $k \in \mathbb{N}$. Consider the degree k polynomial*

$$p(\lambda, z) = \sum_{\ell_1, \ell_2 \text{ even}, 0 < \ell_1 + \ell_2 \leq k} \beta_{\ell_1, \ell_2} \lambda^{\ell_1} z^{\ell_2}$$

where β is the vector of coefficients $\beta = (\beta_{\ell_1, \ell_2} : \ell_1, \ell_2 \in \{0, 1, \dots, k\} \text{ s.t. } \ell_1 + \ell_2 \leq k)$. Then there exists a universal constant C such that

$$\forall \delta > 0, \sup_{t \in \mathbb{R}} \mathbb{P}_{(\lambda, z) \sim \nu} \left[|p_\beta(\lambda, z) - t| \leq \delta \cdot |\mathbb{E}[p_\beta]| \right] \leq C \cdot k \cdot \left(\sqrt{k} \delta \right)^{\frac{1}{2k}}. \quad (3)$$

The proof of Lemma 10 is based on the following Carbery-Wright anti-concentration inequality for Gaussian distributions.

Theorem 16 (Carbery-Wright, Carbery and Wright (2001)) *Let $p : \mathbb{R}^n \rightarrow \mathbb{R}$ be a real-valued polynomial of total degree k , and let ν be a log-concave probability measure on \mathbb{R}^n . There exists a universal constant C such that for any $\delta > 0$,*

$$\sup_{t \in \mathbb{R}} \mathbb{P}_{x \sim \nu} \left[|p(x) - t| \leq \delta \sqrt{\text{Var}(p(x))} \right] \leq C \cdot k \cdot \delta^{1/k}.$$

To apply Carbery-Wright inequality, we first prove the following lemma that compares the mean and the variance of low-degree polynomials under a Gaussian. The univariate case where q is only a polynomial of λ is a special case and is also used in bounding the low-degree advantage in Theorem 12.

Lemma 11 *Let q be a polynomial in (λ, g) of degree at most k with zero constant term. Under $(\lambda, g) \sim N(0, I_2)$,*

$$|\mathbb{E}[q(\lambda, g)]| \leq C\sqrt{k\text{Var}(q(\lambda, g))}.$$

Proof Let

$$q(\lambda, g) = \sum_{\ell_1, \ell_2: 0 < \ell_1 + \ell_2 \leq k} \beta_{\ell_1, \ell_2} \lambda^{\ell_1} g^{\ell_2}.$$

We can also write $q(\lambda, g)$ using the normalized Hermite basis defined in Section A.1. Suppose γ_{i_1, i_2} are the coefficients corresponding to $h_{i_1}(\lambda)h_{i_2}(g)$, i.e.,

$$q(\lambda, g) = \sum_{0 \leq i_1, i_2 \leq k} \gamma_{i_1, i_2} h_{i_1}(\lambda) h_{i_2}(g).$$

Since $h_{i_1}(\lambda)h_{i_2}(g)$ are orthonormal over the standard Gaussian distribution,

$$\begin{aligned} \mathbb{E}_{N(0, I_2)}[q(\lambda, g)] &= \sum_{0 \leq i_1, i_2 \leq k} \gamma_{i_1, i_2} \mathbb{E}_{N(0, I_2)}[h_{i_1}(\lambda)h_{i_2}(g)] = \gamma_{0,0}, \\ \text{Var}_{N(0, I_2)}(q(\lambda, g)) &= \sum_{0 \leq i_1, i_2 \leq k} \gamma_{i_1, i_2}^2 \mathbb{E}_{N(0, I_2)}[h_{i_1}(\lambda)^2 h_{i_2}(g)^2] - \gamma_{0,0}^2 = \sum_{i_1, i_2 \in \mathcal{I}'} \gamma_{i_1, i_2}^2 \end{aligned}$$

where $\mathcal{I}' = \{(i_1, i_2) \neq 0 : i_1, i_2 \text{ even}, 0 \leq i_2 \leq i_1 \leq k\}$. By the formula (4), we have

$$\beta_{0,0} = \sum_{0 \leq i_1, i_2 \leq k, i_1, i_2 \text{ even}} a_{i_1, \frac{i_1}{2}} a_{i_2, \frac{i_2}{2}} \gamma_{i_1, i_2}, \quad a_{i, \frac{i}{2}} = \left(-\frac{1}{2}\right)^{\frac{i}{2}} \frac{\sqrt{i!}}{\left(\frac{i}{2}\right)!}.$$

Thus,

$$\beta_{0,0} = \gamma_{0,0} + \sum_{i_1, i_2 \in I^*} a_{i_1, \frac{i_1}{2}} a_{i_2, \frac{i_2}{2}} \gamma_{i_1, i_2}.$$

Given the constant term $\beta_{0,0} = 0$, we have the identity

$$\gamma_{0,0} = - \sum_{i_1, i_2 \in I^*} a_{i_1, \frac{i_1}{2}} a_{i_2, \frac{i_2}{2}} \gamma_{i_1, i_2}.$$

Then by Cauchy-Schwarz, we have

$$\frac{\gamma_{0,0}^2}{\sum_{i_1, i_2 \in I^*} \gamma_{i_1, i_2}^2} \leq \sum_{i_1, i_2 \in I^*} a_{i_1, \frac{i_1}{2}}^2 a_{i_2, \frac{i_2}{2}}^2.$$

By Stirling's approximation,

$$a_{i, \frac{i}{2}}^2 = \frac{i!}{2^i \left(\frac{i}{2}\right)! \left(\frac{i}{2}\right)!} \sim \sqrt{\frac{2}{\pi i}}.$$

Thus, we can get the following upper bound

$$\begin{aligned} \sum_{i_1, i_2 \in I^*} a_{i_1, \frac{i_1}{2}}^2 a_{i_2, \frac{i_2}{2}}^2 &\leq \left(\sum_{1 \leq i \leq k, i \text{ even}} a_{i, \frac{i}{2}}^2 \right)^2 - a_{0,0}^2 \\ &\leq \left(\sum_{1 \leq i \leq k, i \text{ even}} \sqrt{\frac{2}{\pi i}} + 1 \right)^2 - 1 \\ &\leq \frac{2k}{\pi} + 2\sqrt{\frac{2k}{\pi}} \end{aligned}$$

and then $|\mathbb{E}[q]| \leq C\sqrt{k\text{Var}(q)}$ for some constant C . \blacksquare

Proof [Proof of Lemma 10] Consider the joint distribution $(\lambda, g) \sim N(0, I_2)$. Then $(\lambda, \lambda g) \sim \nu$. Let $q(\lambda, g) = p(\lambda, \lambda g)$. It is equivalent to prove that

$$\forall \delta > 0, \sup_{t \in \mathbb{R}} \mathbb{P}_{(\lambda, g) \sim N(0,1)} \left[|q(\lambda, g) - t| \leq \delta \cdot |\mathbb{E}[q]| \right] \leq C \cdot k \cdot \left(\sqrt{k} \delta \right)^{\frac{1}{2k}}.$$

By Lemma 11, $|\mathbb{E}[q]| \leq C\sqrt{k\text{Var}(q)}$ for some constant C . Then by Carbery-Wright,

$$\begin{aligned} \forall \delta > 0, \sup_{t \in \mathbb{R}} \mathbb{P}_{(\lambda, z) \sim \nu} \left[|q(\lambda, z) - t| \leq \delta \cdot |\mathbb{E}[q]| \right] &\leq \sup_{t \in \mathbb{R}} \mathbb{P}_{(\lambda, z) \sim \nu} \left[|q(\lambda, z) - t| \leq \delta \cdot c\sqrt{k\text{Var}(q)} \right] \\ &\leq C \cdot k \cdot \left(\sqrt{k} \delta \right)^{\frac{1}{2k}}. \end{aligned}$$

\blacksquare

B.1. Tukey depth and Existence of Distribution Matching Moments

Next we connect the anti-concentration properties from Lemma 10 to the convex geometry required for our moment-matching proof in Section 3. Our goal is to prove that our target moment vector θ is in the convex hull of the support of $\Phi(\lambda, z)$.

We first introduce the concept of Tukey depth, a geometric measure of data-centrality. We then use a Carathéodory-style theorem (Theorem 18) that formally links positive Tukey depth to the existence of a vector within the convex hull of a set of random samples.

Definition 17 (Tukey depth) *Given a random variable Φ in \mathbb{R}^D and a point $\theta \in \mathbb{R}^D$, the Tukey-depth of θ is given by*

$$h_{\Phi}(\theta) = \inf_{u: \|u\|_2=1} \mathbb{P} \left[\langle u, \Phi - \theta \rangle \geq 0 \right]. \quad (6)$$

The above notion introduced by Tukey [Tukey \(1975\)](#) plays an important role in statistics. The Tukey depth is also related to Carathéodory-style bounds for random polytopes as follows. To prove that $\theta \in \text{conv}(\{\Phi_1, \dots, \Phi_m\})$, it suffices by the hyperplane separator theorem to show that for every direction u , there exists $i \in [m]$ such that $\langle u, \Phi_i \rangle \geq \langle u, \theta \rangle$. Hence, if we sample $m = \tilde{O}(D/h_{\Phi}(\theta))$ random points Φ_1, \dots, Φ_m i.i.d. from the distribution of Φ , one can use Chernoff-Hoeffding bounds along with an ϵ -net over unit vectors in \mathbb{R}^D to argue that $\theta \in \text{conv}\{(\Phi_1, \dots, \Phi_m)\}$ w.h.p. The following theorem in [Hayakawa et al. \(2023\)](#) gives a tighter bound.

Theorem 18 (Carathéodory bound for random polytopes based on Tukey depth, Hayakawa et al. (2023))

Consider a random vector Φ in \mathbb{R}^D and let $\theta \in \mathbb{R}^D$ have Tukey depth $h_\Phi(\theta)$. Then given $m = (3D + 1)/h_\Phi(\theta)$ i.i.d. samples Φ_1, \dots, Φ_m from the distribution of Φ , we have

$$\mathbb{P} \left[\theta \in \text{conv} \left(\{ \Phi_1, \dots, \Phi_m \} \right) \right] \geq \frac{1}{2}.$$

We will apply the above theorem with the random vector θ being the candidate moment vector of interest, and Φ being random samples from the distribution ν to show the existence of a valid distribution that matches the moments given by θ .

We first show the following lemma that translates the two-sided concentration bound (Lemma 10) into one-sided concentration bound, which then gives a lower bound on the Tukey depth of our target point.

Lemma 19 Suppose ν is a probability measure over \mathbb{R}^n that satisfies the following anti-concentration statements: for any multivariate polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$ of degree at most k ,

$$\forall \delta > 0, \mathbb{P}_{x \sim \nu} \left[|p(x) - \mathbb{E}_\nu[p]| \leq \delta \cdot \sqrt{\text{Var}(p(x))} \right] \leq \eta_k(\delta).$$

Then for any $\delta > 0$ and any $\delta_1 > 0$ such that $\eta_k(\delta_1) \leq 1/2$, we have

$$\mathbb{P}_{x \sim \nu} \left[p(x) \geq \mathbb{E}_\nu[p(x)] + \delta \cdot \sqrt{\text{Var}(p(x))} \right] \geq \frac{\delta_1^2}{16} - \eta_k(\delta).$$

In particular, there exists $\delta_0 > 0$ such that

$$\forall \delta < \delta_0, \mathbb{P}_{x \sim \nu} \left[p(x) \geq \mathbb{E}_\nu[p(x)] + \delta \cdot \sqrt{\text{Var}(p(x))} \right] \geq 3\eta_k(\delta),$$

by setting $\delta_1 = 8\sqrt{\eta_k(\delta)}$. Here $\delta_0 > 0$ a constant chosen such that $\eta_k(8\sqrt{\eta_k(\delta_0)}) < \frac{1}{2}$.

Proof Consider the r.v. $Z = p(x) - \mathbb{E}[p(x)]$. We have $\mathbb{E}[Z] = 0$. By the anti-concentration property of p under measure ν , we have

$$\mathbb{P}[|Z| \leq \delta \sqrt{\text{Var}[Z]}] \leq \eta_k(\delta), \text{ and } \mathbb{P}[|Z| \leq \delta_1 \sqrt{\text{Var}[Z]}] \leq \eta_k(\delta_1). \quad (7)$$

Furthermore since $\mathbb{E}[Z] = 0$, by Cauchy-Schwarz,

$$\frac{1}{2} \mathbb{E}[|Z|] = \mathbb{E} \left[Z \mathbf{1}[Z \geq 0] \right] \leq \sqrt{\text{Var}[Z]} \sqrt{\mathbb{P}[Z \geq 0]}.$$

Then

$$\mathbb{P} \left[Z \geq \mathbb{E}[Z] + \delta \sqrt{\text{Var}[Z]} \right] \geq \mathbb{P}[Z \geq 0] - \eta_k(\delta) \geq \frac{\mathbb{E}[|Z|]^2}{4\text{Var}[Z]} - \eta_k(\delta). \quad (8)$$

Moreover

$$\mathbb{E}[|Z|] \geq \mathbb{P}[|Z| \geq \delta_1 \sqrt{\text{Var}[Z]}] \cdot \delta_1 \sqrt{\text{Var}[Z]} \geq (1 - \eta_k(\delta_1)) \delta_1 \sqrt{\text{Var}[Z]}. \quad (9)$$

Combining (8), (9), we get

$$\mathbb{P} \left[Z \geq \mathbb{E}[Z] + \delta \sqrt{\text{Var}[Z]} \right] \geq \frac{(1 - \eta_k(\delta_1))^2 \delta_1^2}{4} - \eta_k(\delta) \geq \frac{\delta_1^2}{16} - \eta_k(\delta),$$

for our choice of parameters. ■

Corollary 20 *Let $\Phi \in \mathbb{R}^d$ be a random variable with coordinates $\Phi_{\ell_1, \ell_2} = \lambda^{\ell_1} z^{\ell_2}$ where $(\ell_1, \ell_2) \in \mathcal{I} = \{(i_1, i_2) \neq 0 : i_1, i_2 \text{ even}, 0 \leq i_2 \leq i_1 \leq k\}$ and $(\lambda, z) \sim \nu$. Then for any $\delta \leq 1/k^{Ck^2}$,*

$$h_\Phi((1 + \delta)\mathbb{E}[\Phi]) \geq 2\eta_k(\sqrt{k}\delta) > 0.$$

Proof Let $\beta \in \mathbb{R}^D$ be an arbitrary vector of coefficients with $\|\beta\| = 1$. Then

$$p(\lambda, z) := \langle \beta, \Phi \rangle = \sum_{\mathcal{I}} \beta_{\ell_1, \ell_2} \lambda^{\ell_1} z^{\ell_2}.$$

By Lemma 19, we know that

$$\forall \delta < \delta_0, \quad \mathbb{P}_{x \sim \nu} \left[p(x) \geq \mathbb{E}_\nu[p(x)] + \delta \cdot \sqrt{\text{Var}(p(x))} \right] \geq 3\eta_k(\delta)$$

where $\delta_0 > 0$ is chosen such that $\eta_k(8\sqrt{\eta_k(\delta_0)}) < \frac{1}{2}$. Since $(\lambda, g) \sim \mathcal{N}(0, I_2)$ is log-concave, Carbery-Wright gives $\eta_k(\delta) = C_0 k \delta^{1/k}$ for $q(\lambda, g)$. Because $q(\lambda, g) = p(\lambda, \lambda g)$, the distributional statements for q under $\mathcal{N}(0, I_2)$ transfer verbatim to p under ν . Hence we can choose $\delta_0 = 1/k^{Ck^2}$ for some constant C so that it satisfies the condition in Lemma 19. From the argument in the proof of Lemma 10, we have $|\mathbb{E}[p]| \leq c\sqrt{k\text{Var}(p)}$ for some constant c . So

$$\begin{aligned} h_\Phi((1 + \delta)\mathbb{E}[\Phi]) &= \inf_{\beta: \|\beta\|_2=1} \mathbb{P}_\nu \left[\langle \beta, \Phi - (1 + \delta)\mathbb{E}[\Phi] \rangle \geq 0 \right] \\ &= \inf_{\beta: \|\beta\|_2=1} \mathbb{P}_\nu [p \geq (1 + \delta)\mathbb{E}[p]] \\ &\geq \inf_{\beta: \|\beta\|_2=1} \mathbb{P}_\nu \left[p \geq \mathbb{E}[p] + c\delta\sqrt{k\text{Var}(p)} \right] \\ &\geq 3\eta_k(c\sqrt{k}\delta) > 0. \end{aligned}$$

■

Corollary 20 is the crucial link back to the proof of Theorem 9. It confirms that the specific moment vector $\Phi(\lambda, z)$ with $(\lambda, z) \sim \nu$ from our construction is sufficiently anti-concentrated (via Lemma 10 and Lemma 19) to guarantee that a point slightly perturbed from its mean has a non-zero Tukey depth. The following lemma shows that this non-zero depth implies the point is in the interior of the convex hull, which provides the “room” needed to complete the moment-matching argument in the proof of Theorem 9.

Lemma 21 *Let Φ be a random vector in \mathbb{R}^D such that every affine hyperplane has zero Φ -measure. If $h_\Phi(\theta) > 0$, then θ is in the interior of $\text{conv}(\text{supp}(\Phi))$. In particular, there exists $r = r(h_\Phi(\theta)) > 0$ such that $B(\theta, r) \subseteq \text{conv}(\text{supp}(\Phi))$.*

Proof We will prove by contradiction. Suppose θ is on the boundary of $\text{conv}(\text{supp}(\Phi))$, then a supporting hyperplane u satisfies $\langle u, \phi \rangle \leq \langle u, \theta \rangle$ for all ϕ in the support, which implies $h_\Phi(\theta) = 0$. Therefore, θ is in the interior of $\text{conv}(\text{supp}(\Phi))$. Since the interior of $\text{conv}(\text{supp}(\Phi))$ is open, there exists $r > 0$ such that $B(\theta, r) \subseteq \text{conv}(\text{supp}(\Phi))$. ■

Appendix C. Proofs in Section 4

Lemma 13 *Let P, Q be distributions on \mathbb{R}^n with finite moments up to degree $2k$. Suppose the single-sample low-degree advantage is $\text{LDA}_{\leq k}^{(1)}(P, Q) = \delta$. Then for any m ,*

$$\text{LDA}_{\leq k}^{(m)}(P, Q)^2 \leq (1 + \delta^2)^m - 1.$$

Proof Let $\{\psi_j : j \in \mathcal{J}\}$ be an orthonormal basis of polynomials of degree at most k under Q with $\psi_0 = 1$. For a multi-index $\gamma = (\gamma_1, \dots, \gamma_m)$, define

$$\Psi_\gamma(X_1, \dots, X_m) = \prod_{i=1}^m \psi_{\gamma_i}(X_i).$$

Because $Q^{\otimes m}$ is a product measure, Ψ_γ are orthonormal with respect to $Q^{\otimes m}$. For any polynomial f of degree at most k , subtracting the constant $\mathbb{E}_{Q^{\otimes m}}[f]$ does not change LDA so we may assume $\mathbb{E}_{Q^{\otimes m}}[f] = 0$. Then f can be written as

$$f = \sum_{\gamma \neq 0} c_\gamma \Psi_\gamma$$

where $\gamma \neq 0$ means not all coordinates of γ are zero. Let $s(\gamma)$ be the number of nonzero entries in γ . Then $s(\gamma) \leq k$ since the degree of f is at most k . By orthonormality,

$$\text{Var}_{Q^{\otimes m}}(f) = \mathbb{E}_{Q^{\otimes m}}[f^2] = \sum_{\gamma \neq 0} c_\gamma^2.$$

Let $\mu_j = \mathbb{E}_P[\psi_j]$. We have

$$\mathbb{E}_{P^{\otimes m}}[\Psi_\gamma] = \prod_{i=1}^m \mu_{\gamma_i}.$$

Since $\mathbb{E}_Q[\psi_j] = 0$ for $j \neq 0$,

$$\mathbb{E}_{P^{\otimes m}}[f] - \mathbb{E}_{Q^{\otimes m}}[f] = \sum_{\gamma \neq 0} c_\gamma \prod_{i=1}^m \mu_{\gamma_i}.$$

By Cauchy-Schwarz,

$$\text{LDA}_{\leq k}^{(m)}(P, Q)^2 = \frac{(\mathbb{E}_{P^{\otimes m}}[f] - \mathbb{E}_{Q^{\otimes m}}[f])^2}{\text{Var}_{Q^{\otimes m}}(f)} = \frac{\left(\sum_{\gamma \neq 0} c_\gamma \prod_{i=1}^m \mu_{\gamma_i}\right)^2}{\sum_{\gamma \neq 0} c_\gamma^2} \leq \sum_{\gamma \neq 0} \prod_{i=1}^m \mu_{\gamma_i}^2.$$

Similarly, when $m = 1$, we have

$$\text{LDA}_{\leq k}^{(1)}(P, Q)^2 = \frac{\left(\sum_j c_j \mu_j\right)^2}{\sum_j c_j^2} \leq \sum_j \mu_j^2.$$

When c and μ are linearly dependent, the equation holds. Thus $\text{LDA}_{\leq k}^{(1)}(P, Q)^2 = \sum_j \mu_j^2 = \delta^2$. Fix a subset $S \subseteq [m]$ of size s . Group γ by the nonzero coordinates by the set S ,

$$\sum_{\substack{\gamma: \{i: \gamma_i \neq 0\} = S, \\ \sum_{i \in S} \deg(\psi_{\gamma_i}) \leq k}} \prod_{i \in S} \mu_{\gamma_i}^2 \leq \sum_{j \in \mathcal{J}} \prod_{i \in S} \mu_j^2 = \prod_{i \in S} \left(\sum_{j \in \mathcal{J}} \mu_j^2 \right) = \left(\sum_{j \in \mathcal{J}} \mu_j^2 \right)^s = \delta^{2s}.$$

Sum over all possible S with size $s \leq k$,

$$\text{LDA}_{\leq k}^{(m)}(P, Q)^2 \leq \sum_{s=1}^k \binom{m}{s} \delta^{2s} \leq (1 + \delta^2)^m - 1.$$

■

C.1. Generalizing the scale distribution: the Christoffel function

We now revisit the proof of Theorem 12 and explain which structural property of the scale distribution governs the single-sample LDA bound.

In the proof, we crucially used the property that $X \sim Q$ can be expressed as the product $X = \lambda g$ of two independent random variables, where $\lambda \in \mathbb{R}$ represents a scale distribution drawn from μ , and $g \in \mathbb{R}^n$ is drawn from any rotationally invariant distribution. The specific choices of g and λ being Gaussian were for convenience. We now explain precisely which structural property of the scale distribution μ governs the LDA bound, and how the argument generalizes.

The central step in the proof reduces the n -dimensional single-sample advantage to a bound on the ratio $|\mathbb{E}_\mu[h(\lambda)]|/\sqrt{\text{Var}_\mu(h(\lambda))}$ for univariate degree- k polynomials $h(\lambda)$ satisfying $h(0) = 0$. The ratio $|\mathbb{E}_\mu[h]|/\sqrt{\text{Var}_\mu(h)}$ has a clean characterization via the orthonormal polynomials $\{p_j\}_{j \geq 0}$ for the measure μ . Expanding $h(\lambda) = \sum_{j=0}^k c_j p_j(\lambda)$, orthonormality gives $\mathbb{E}_\mu[h] = c_0$ and $\text{Var}_\mu(h) = \sum_{j=1}^k c_j^2$. The constraint $h(0) = 0$ yields

$$c_0 = - \sum_{j=1}^k c_j p_j(0).$$

By Cauchy–Schwarz,

$$c_0^2 \leq \left(\sum_{j=1}^k c_j^2 \right) \left(\sum_{j=1}^k p_j(0)^2 \right) = \text{Var}_\mu(h) \cdot \sum_{j=1}^k p_j(0)^2,$$

so that

$$\frac{|\mathbb{E}_\mu[h(\lambda)]|}{\sqrt{\text{Var}_\mu(h(\lambda))}} \leq \sqrt{\sum_{j=1}^k p_j(0)^2}, \quad (10)$$

with equality achieved when $c_j \propto p_j(0)$ for $j \geq 1$. The quantity

$$\mathcal{K}_k(\mu, 0) := \sum_{j=0}^k p_j(0)^2$$

is the Christoffel sum for μ at 0 (the inverse of the Christoffel function $\lambda_k(\mu, 0)^{-1}$). Since $p_0 \equiv 1$, the bound (10) becomes $|\mathbb{E}_\mu[h]|/\sqrt{\text{Var}_\mu(h)} \leq \sqrt{\mathcal{K}_k(\mu, 0) - 1}$. The growth rate of $\mathcal{K}_k(\mu, 0)$ with k is therefore the fundamental quantity controlling the LDA bound. In particular, any scale distribution for which this quantity grows slowly enough with k yields the same kind of bound on the single-sample low-degree advantage, and hence the same lifting argument to m samples.

Appendix D. Simple Noise-Tolerant Algorithm

In this section we show that there are simple efficient algorithms that are not based on the low-degree polynomial method that can solve this problem efficiently. Consider the setting when the $\alpha = 1/\text{poly}(n)$ fraction lies on a $d \leq 1$ dimensional subspace. On the one hand, the low-degree method fails even when an α fraction of points lie on the subspace. On the other hand, our simple algorithm samples points and checks whether there are $(d + 1)$ points (*one* point if $d = 0$) that are close to a d -dimensional subspace (close to 0 if $d = 0$). Moreover, it can achieve tolerance to worst-case errors, and rerandomization.

An elegant and clever algorithm for robust subspace recovery due to Hardt and Moitra (Hardt and Moitra, 2013) uses anti-concentration properties of $(1 - \alpha)$ points not on the subspace S : it gives efficient algorithms when $\alpha \geq \frac{\dim(S)}{n}$. Subsequent work by Bhaskara, Chen, Perreault and VijayaragHAVAN (Bhaskara et al., 2019) gave an improved algorithm in the smoothed analysis setting where the points not on the subspace are randomly perturbed (by a random vector of average length $\rho = 1/\text{poly}(n)$), and gave guarantees of recovery when $\alpha \gtrsim (d/n)^\ell$ for any constant $\ell > 0$ in $(nd)^{O(\ell)}$ time. In our setting where $d = O(1)$ and $\alpha = 1/n^{O(1)}$, we can pick an appropriate $\ell = O(1)$ to get algorithmic guarantees. However this algorithm is only moderately noise tolerant: every point can be adversarially perturbed (moved) by $1/n^{O(\ell)}$. See also (Bakshi and Kothari, 2021; Gao et al., 2026) for algorithmic results in other parameter regimes. We will instead give a simple algorithm specialized for the $d = O(1)$ setting that is elementary and also allows points to be moved by an amount $\epsilon = \Omega(1)$.

D.1. Robust Subspace Recovery with ϵ Relative Error Perturbations.

We define the hypothesis testing version of the problem below. We are given samples from a distribution \mathcal{D} over \mathbb{R}^n , and the goal is to distinguish between the following two hypothesis:

- **NULL:** $\mathcal{D} = Q_{\text{rot}}$ where Q_{rot} is the rotationally invariant distribution defined in Section 3.
- **PLANTED:** \mathcal{D} is *any* distribution with at least an α probability mass supported on a planted subspace S of dimension d .

Note that the PLANTED case above is more general than the planted distribution in the lower bound example in Section 3. Moreover, our algorithm will work even when the samples drawn from \mathcal{D} are then adversarially perturbed as follows.

Definition 22 (Adversarial (ϵ, p) -noisy sample from \mathcal{D}) Draw m samples $x_1, \dots, x_m \in \mathbb{R}^n$ drawn i.i.d. from \mathcal{D} . Then an adversary can perform the following for $p \in [0, 1)$:

1. Re-randomize to draw from Q_{rot} : For each $i \in [m]$, with probability p we redraw $x_i \sim Q_{\text{rot}}$, and with probability $(1 - p)$ it remains unaltered.

2. Adversarial ϵ -perturbation to each point: for each $i \in [m]$ the sample x_i is adversarially perturbed to form $\tilde{x}_i = x_i + z_i$ where $\|z_i\|_2 \leq \epsilon \|x_i\|_2$.

The input are the corrupted samples $\tilde{x}_1, \dots, \tilde{x}_m$.

Note that the adversary observes the m samples x_1, \dots, x_m , and the corruptions to the samples can be correlated arbitrarily. We remark that the low-degree polynomial method continues to fail in this model.

We analyze the following simple algorithm.

Noise Tolerant Algorithm

Input: Given m samples $\tilde{x}_1, \dots, \tilde{x}_m$, dimension d , lower bound $\alpha_0 \in (0, 1)$ for parameter α , upper bound $p_0 \in [0, 1)$ on parameter p

Output: NULL, or PLANTED (along with a subspace S of dimension d).

1. Normalize the vectors $a_i = \tilde{x}_i / \|\tilde{x}_i\|$ for all $i \in [m]$.
2. Run over all $\binom{m}{d+1}$ choices $T \subset [m]$ of size $|T| = d + 1$
3. (i) Let A_T be the matrix formed by the corresponding points $A_T = (a_i : i \in T) \in \mathbb{R}^{n \times (d+1)}$.
 (ii) If $\sigma_{d+1}(A_T) \leq \frac{1}{2}$, output PLANTED and return.
 % (If the subspace is desired, output the top- d left singular space of A_T).
4. Output NULL.

Figure 2: Noise Tolerant Algorithm for detecting if an α fraction of the points lie close to a $d = O(1)$ -dimensional subspace

Theorem 14 (Noise tolerant algorithm with relative error perturbations) *There exists a universal constants $c, C > 0$ such that the following holds when $m \geq C(d + 1)/(\alpha(1 - p))$ and any $0 \leq \epsilon \leq \frac{1}{8(d+1)}$. There is an algorithm (given in Figure 2) that runs in time $O(\binom{m}{d+1})$ that given as input noisy samples $\tilde{x}_1, \dots, \tilde{x}_m$ as described above with $\|\tilde{x}_i - x_i\| \leq \epsilon \|x_i\|$, can solve with high probability the testing problem.*

The argument for the NULL setting (soundness) follows from the following lemma.

Lemma 23 (NULL setting) *In the setting of Theorem 14, there exists a constant $c' > 0$ such that with probability at least $1 - m^2 \exp(-c'n)$ we have*

$$\forall T \subset [m] \text{ with } |T| = d + 1, \sigma_{d+1}(A_T) > \frac{1}{2}.$$

Proof The proof follows from the *incoherence* of the matrix A .

Claim 24 *There exists universal constants $c, c' > 0$ such that the set of vectors $x_1, \dots, x_m \sim Q_{\text{rot}}$, the (corrupted) normalized vectors $\{a_i = \frac{(x_i + z_i)}{\|x_i + z_i\|} : i \in [m]\}$ satisfies with probability at least $1 - O(m^2 \exp(-c'n))$ that*

$$\forall i \neq j \in [m], |\langle a_i, a_j \rangle| \leq \frac{1}{(1-\epsilon)^2} \left(\frac{c\sqrt{\log m}}{\sqrt{n}} + 2\epsilon + \epsilon^2 \right). \quad (11)$$

Proof of the Claim. From standard concentration properties of random unit vectors we have for some universal constant $c > 0$,

$$\begin{aligned} \mathbb{P}_{\hat{x}, \hat{y} \sim_{\text{unif}} \mathbb{S}^{n-1}} \left[|\langle \hat{x}, \hat{y} \rangle| \geq \frac{c\sqrt{\log m}}{\sqrt{n}} \right] &\leq \frac{1}{m^3}, \quad \text{and hence by a union bound,} \\ \forall i \neq j \in [m], \mathbb{P}_{x_i, x_j \sim Q_{\text{rot}}} \left[\left| \left\langle \frac{x_i}{\|x_i\|}, \frac{x_j}{\|x_j\|} \right\rangle \right| \leq \frac{c\sqrt{\log m}}{\sqrt{n}} \right] &\geq 1 - \frac{1}{m}. \end{aligned} \quad (12)$$

Let us condition on the event in (12). Also note that $\|x_i + z_i\| \geq (1-\epsilon)\|x_i\|$. Then we have for all $i \neq j \in [m]$

$$\begin{aligned} |\langle a_i, a_j \rangle| &= \left| \left\langle \frac{x_i + z_i}{\|x_i + z_i\|}, \frac{x_j + z_j}{\|x_j + z_j\|} \right\rangle \right| \leq \frac{|\langle x_i, x_j \rangle|}{(1-\epsilon)^2 \|x_i\| \|x_j\|} + \frac{|\langle z_i, x_j \rangle| + |\langle x_i, z_j \rangle| + |\langle z_i, z_j \rangle|}{(1-\epsilon)^2 \|x_i\| \|x_j\|} \\ &\leq \frac{c\sqrt{\log m}}{(1-\epsilon)^2 \sqrt{n}} + \frac{\|z_i\| \|x_j\| + \|x_i\| \|z_j\| + \|z_i\| \|z_j\|}{\|x_i\| \|x_j\|} \leq \frac{c\sqrt{\log m}}{(1-\epsilon)^2 \sqrt{n}} + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2}, \end{aligned}$$

where the last line used (12) and $\|z_i\| \leq \epsilon \|x_i\|$.

Now to complete the rest of the proof, consider any $T \subset [m]$ with $|T| = d+1$, and a test unit vector $\beta \in \mathbb{R}^{d+1}$ for the matrix A_T . We have

$$\begin{aligned} \|A_T \beta\|_2^2 &= \left\| \sum_{i \in T} \beta_i a_i \right\|_2^2 \geq \sum_{i \in T} \beta_i^2 \|a_i\|_2^2 - \sum_{i \neq j} |\beta_i| |\beta_j| |\langle a_i, a_j \rangle| \\ &\geq \sum_{i \in T} \beta_i^2 - \sum_{i \neq j \in T} |\beta_i| |\beta_j| \left(\frac{c\sqrt{\log m}}{\sqrt{n}(1-\epsilon)^2} + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2} \right) \\ &\geq 1 - \|\beta\|_1^2 \left(\frac{c\sqrt{\log m}}{\sqrt{n}(1-\epsilon)^2} + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2} \right) \geq 1 - \frac{(d+1)}{(1-\epsilon)^2} \left(\frac{c\sqrt{\log m}}{\sqrt{n}} + 2\epsilon + \epsilon^2 \right) \\ &> \frac{1}{2} \quad \text{since } \epsilon < \frac{1}{8(d+1)}, \text{ and } m < 2^{o(n)}. \end{aligned}$$

■

Proof To prove the algorithm works, we will show that there is (a) an approximate linear dependence between $d+1$ of the points sampled from the α portion in the PLANTED case, and (b) no approximate linear dependence between any set of $d+1$ points in the NULL case.

PLANTED case (completeness): Let x_i be the uncorrupted sample from \mathcal{D} corresponding to \tilde{x}_i , and let $\tilde{x}_i = x_i + z_i$ with $\|z_i\| \leq \epsilon \|x_i\|$. Given $m \geq C(d+1)/(\alpha(1-p))$, there exists in expectation at least Cd points from \mathcal{D} lying in a subspace S ; hence with high probability there

exists $T^* \subset [m]$ with $|T^*| = d + 1$ corresponding to points $\{x_i : i \in T^*\}$ that belong to a subspace S of dimension d . Recall that $a_i = \tilde{x}_i / \|x_i\|$.

The submatrix $M_{T^*} = (\frac{x_i}{\|x_i\|} : i \in T^*)$ has $\sigma_{d+1}(M_{T^*}) = 0$. Let $\beta \in \mathbb{R}^{d+1}$ be a unit vector satisfying $M_{T^*}\beta = 0$. Then

$$\begin{aligned}
 A_{T^*}\beta &= \sum_{i \in T^*} \beta_i a_i = \sum_{i \in T^*} \beta_i \frac{x_i + z_i}{\|x_i + z_i\|} = \sum_{i \in T^*} \beta_i \frac{x_i}{\|x_i + z_i\|} + \beta_i \frac{z_i}{\|x_i + z_i\|} \\
 &= \sum_{i \in T^*} \beta_i \left(\frac{x_i}{\|x_i + z_i\|} - \frac{x_i}{\|x_i\|} \right) + \sum_{i \in T^*} \beta_i \frac{z_i}{\|x_i + z_i\|} && \text{(since } M_{T^*}\beta = 0 \text{)} \\
 \|A_{T^*}\beta\|_2 &\leq \sum_{i \in T^*} |\beta_i| \left(\frac{\|x_i\|}{\|x_i + z_i\|} - 1 \right) + \epsilon \sum_{i \in T^*} |\beta_i| \frac{\|x_i\|}{\|x_i + z_i\|} && \text{(since } \|z_i\| \leq \epsilon \|x_i\| \text{)} \\
 &\leq \left(\sum_{i \in T^*} |\beta_i| \right) \left(\frac{2\epsilon}{1 - \epsilon} \right) \leq \frac{2\epsilon\sqrt{d+1}}{1 - \epsilon} < \frac{1}{2} && \text{(since } \epsilon < \frac{1}{8(d+1)} \text{)}.
 \end{aligned}$$

NULL case (soundness): This follows from Lemma 23. This completes the proof. ■

D.2. Model with Additive Perturbations

In this section, we gave a variant of the algorithm that also works with worst-case additive perturbations of bounded magnitude.

We consider the same hypothesis testing problem as before: We are given samples from a distribution \mathcal{D} over \mathbb{R}^n . The goal is to distinguish between the following two hypothesis:

- **NULL:** $\mathcal{D} = Q_{\text{rot}}$ where Q_{rot} is the rotationally invariant distribution defined in Section 3.
- **PLANTED:** \mathcal{D} is *any* distribution with at least an α probability mass supported on a planted subspace S of dimension d .

We now consider the setting where the samples drawn from \mathcal{D} are then adversarially perturbed as follows.

Definition 25 (Adversarial additive (η, p) -noisy sample from \mathcal{D}) Draw m samples $x_1, \dots, x_m \in \mathbb{R}^n$ drawn i.i.d. from \mathcal{D} . Then an adversary can perform the following for $\eta, p \in [0, 1]$:

1. Re-randomize to draw from Q_{rot} : For each $i \in [m]$, with probability p we redraw $x_i \sim Q_{\text{rot}}$, and with probability $(1 - p)$ it remains unaltered.
2. Adversarial η -perturbation to each point: for each $i \in [m]$ the sample x_i is adversarially perturbed to form $\tilde{x}_i = x_i + z_i$ where $\|z_i\|_2 \leq \eta$.

The input are the corrupted samples $\tilde{x}_1, \dots, \tilde{x}_m$.

We analyze the following simple algorithm.

Noise Tolerant Algorithm for Additive Perturbations

Input: Given m samples $\tilde{x}_1, \dots, \tilde{x}_m$, dimension d , parameter $\alpha \in (0, 1)$, parameter $p \in (0, 1)$, and threshold $\tau > 0$

Output: NULL, or PLANTED(along with a subspace S of dimension d).

1. Subsample $J \subseteq [m]$ of size $|J| = 2 \log(4/\delta)(d + 1)/(\alpha(1 - p))$.
2. Run over all $\binom{m}{d+1}$ choices $T \subset J$ of size $|T| = d + 1$
3. (i) Let A_T be the matrix formed by the corresponding points $A_T = (\tilde{x}_i : i \in T) \in \mathbb{R}^{n \times (d+1)}$.
 (ii) If $\sigma_{d+1}(A_T) \leq \tau$, output PLANTED.
 % (If the subspace is desired, output the top- d left singular space of A_T).
4. Output NULL

Figure 3: Algorithm for detecting even under *additive perturbations* if an α fraction of the points lie close to a $d = O(1)$ -dimensional subspace

Theorem 26 (Noise tolerant algorithm with additive error perturbations) *There exists a universal constants $c, c' > 0, C \geq 1$ such that the following holds when $m \geq 2(d+1) \log(4/\delta)/(\alpha(1-p))$ and any $\delta \in (0, 1), 0 \leq \eta \leq \frac{\alpha(1-p)\sqrt{n}}{C(d+1)\log(4/\delta)^2}$. Then there is an algorithm (given in Figure 3 with $\tau = \frac{\alpha(1-p)\sqrt{n}}{c \log^2(4/\delta)}$) running in time $O(\binom{m}{d+1})$ that given as input $\tilde{x}_1, \dots, \tilde{x}_m$ generated as described above with $\|\tilde{x}_i - x_i\|_2 \leq \eta$ can solve the testing problem with probability at least $1 - \delta$.*

In this model, the average length of a point drawn from Q_{rot} is $\mathbb{E}[\|X\|_2^2] = \Theta(n)$. Moreover, an α fraction of the points drawn from Q_{rot} have length $\|X\|_2 \leq c_* \alpha \sqrt{n}$ for some absolute constant $c_* > 0$ — hence when $\eta \geq c_* \alpha \sqrt{n}$, the hypothesis testing problem becomes statistically impossible. Theorem 26 shows that the algorithm handles η that is within a constant factor (for $d = O(1)$) of the best possible bound of $c_* \alpha \sqrt{n}$.

The following claim shows that for most sample points drawn from Q_{rot} , their length is at least 2τ .

Claim 27 *There exists a universal constant $c > 0$ such that for any $\gamma \in (0, 1/2)$, we have*

$$\mathbb{P}_{X \sim Q_{\text{rot}}} \left[\|X\|_2 \leq \gamma \sqrt{n} \right] \leq \min\{c\gamma, 1\}. \quad (13)$$

In particular, by setting $\gamma = \alpha(1-p)/(4c \log^2(4/\delta)(d+1))$, we get that the fraction of points drawn from Q_{rot} whose length $\|X\|_2 \leq \frac{\alpha(1-p)\sqrt{n}}{2c \log(4/\delta)^2}$ is at most $\frac{\alpha(1-p)}{4 \log(4/\delta)^2}$.

Proof This follows directly from the small ball probability bound for a standard Gaussian i.e., $\mathbb{P}_{g \sim N(0,1)}[|g| \leq \delta/2] \leq c\delta/2$, along with length concentration of a standard spherical Gaussian. ■

We now proceed to the proof of Theorem 26.

Proof To prove the algorithm works, we will show that there is (a) an approximate linear dependence between many $(d + 1)$ -tuple of the points sampled from the α portion in the subspace in the PLANTED case, and (b) approximate linear dependence between much fewer set of $(d + 1)$ -tuples of the samples in the NULL case.

PLANTED case (completeness): Let x_i be the uncorrupted sample from \mathcal{D} corresponding to \tilde{x}_i , and let $\tilde{x}_i = x_i + z_i$ with $\|z_i\| \leq \eta$. Given $m \geq 2 \log(4/\delta)(d + 1)/(\alpha(1 - p))$, there exists in expectation $(1 - p)\alpha|J| \geq 2(d + 1) \log(4/\delta)$ points from \mathcal{D} lying in a subspace S . Hence with probability $1 - \delta/4$ there are at least $d + 1$ samples $J_S \subseteq J$ lying in subspace S . For any tuple $T \subseteq J_S$ with $|T| = d + 1$, we have that $\sigma_{d+1}((x_i : i \in T)) = 0$. By Weyl's inequality, we have $\sigma_{d+1}(A_T) \leq \eta\sqrt{d + 1} \leq \tau/2$, as required.

NULL case (soundness): The soundness analysis is very similar to the soundness analysis of Theorem 14. We know from Claim 27, that for at least $1 - \alpha/4$ fraction of the samples $\{x_i : i \in J\}$ drawn from Q_{rot} , we have $\|x_i\|_2 \geq 2\tau$. For these samples which constitute at least $(1 - \alpha/4)$ fraction, the additive perturbations can be viewed as multiplicative perturbations, and we can apply Claim 24 to get a bound.

Formally, among the m samples, let $J_{\text{big}} = \{i \in J : \|x_i\|_2 \geq 2\tau\}$. Consider a modified set of points x'_1, \dots, x'_m where for each $i \in J$ we have $x'_i = \tilde{x}_i$ if $i \in J_{\text{big}}$ and $x'_i = x_i$ if $i \notin J_{\text{big}}$. First we observe that these points $\{x'_i : i \in J\}$ satisfy the conditions of the Lemma 23 as the perturbations satisfy $\|x'_i - x_i\|_2 < \frac{1}{8(d+1)} \|x_i\|_2$. Let A' be the matrix with its i th column of A' equal to $x'_i/\|x'_i\|$. Hence we can now apply Lemma 23 to conclude that for some absolute constant $c' > 0$, we have with probability at least $1 - O(m^2 \exp(-c'n))$ that

$$\forall T \subseteq [J] \text{ s.t. } |T| = d + 1, \sigma_{d+1}(A'_T) > 1/2.$$

Moreover, we have from Claim 27 that

$$\mathbb{E}[|J \setminus J_{\text{big}}|] \leq \frac{\alpha(1 - p)}{4(d + 1) \log(4/\delta)^2} \cdot \frac{2 \log(4/\delta)(d + 1)}{\alpha(1 - p)} \leq \frac{1}{2 \log(4/\delta)}.$$

Hence with probability at least $1 - \delta/2$, we have that $J_{\text{big}} = J$. In other words, with probability at least $1 - \delta/2$, $x'_i = \tilde{x}_i \forall i \in J$. Hence we have with probability at least $1 - \delta/2 - m^2 \exp(-c'n)$,

$$\forall T \subseteq J \text{ s.t. } |T| = d + 1, \sigma_{d+1}(A_T) \geq \min_{i \in T} \|\tilde{x}_i\|_2 \cdot \sigma_{d+1}(A'_T) \geq \frac{1}{2} \cdot 2\tau \geq \tau.$$

This concludes the proof. ■