Robust Signal Detection with Quadratically Convex Orthosymmetric Constraints

Yikun Li and Matey Neykov

Department of Statistics and Data Science
Northwestern University
Evanston, IL 60208
yikunli2028@u.northwestern.edu   mneykov@northwestern.edu

Abstract

This paper is concerned with robust signal detection in Gaussian noise under quadratically convex orthosymmetric (QCO) constraints. Specifically, the null hypothesis H_0 assumes no signal, whereas the alternative H_1 considers signal that is separated in Euclidean norm from zero and belongs to a set K satisfying the QCO constraints. In addition, an adversary is allowed to inspect all the N samples and replace up to ϵN of them with arbitrary values, where $0 < \epsilon < c_0 < \frac{1}{2}$ is the corruption rate. Our main results establish the minimax rate of the separation radius ρ_{critical} between H_0 and H_1 purely in terms of the geometry of K, the corruption rate ϵ (up to logarithmic factors in $\frac{1}{\epsilon}$) and the scale of the noise σ. We argue that the Kolmogorov widths of the constraint play a central role in determining the minimax rate. This indicates similarity with the (uncorrupted) estimation problem under QCO constraints, which was first established by Donoho et al. (1990). Moreover, the minimax lower bound reveals interesting phase transitions of the testing problem regarding ϵ. Consistent with classic belief about testing and estimation, the testing problem is "easier" (even when one compares the results with recent papers studying the constrained robust estimation problem (Prasadan and Neykov, 2025)). In addition to the main results above, where the upper bound is achieved with an intractable algorithm, inspired by Canonne et al. (2023), we develop a polynomial-time algorithm which also nearly (up to logarithmic factors in N and $1/\epsilon$) achieves the minimax lower bound. In contrast to Canonne et al. (2023), our algorithm works for signals of arbitrary Euclidean length, and respects the QCO constraint. Finally, all the results above are naturally extended to the ℓ_p norm testing problem for $1 \le p < 2$.

Contents

1 Introduction

Data Contamination and Robust Statistics

One of the most basic tasks in statistics is to perform inference about the unknown parameters used in data generation. However, a variety of reasons can lead the observed data to deviate from the original samples of the true mechanism. These reasons could range from measurement errors in some percentage of the observations, to even malicious tampering with the data. For example, in certain biological datasets, inconsistencies in the data can be natural but beyond the level of white noise (see Martí, 2019; Fibbi et al., 2023, e.g.), while in machine learning security applications, contamination can be artificial or malicious, which is also called *data poisoning attacks* (see Ramirez et al., 2022, e.g.); see also Diakonikolas and Kane (2019) for other examples.

In this work, we take a stylized and classical statistical inference problem and study it from both constrained and robust perspectives. Specifically, we will study the Gaussian sequence model, where each coordinate X_{ij} of a single sample X_i is drawn independently from $\mathcal{N}\left(\mu_j, \sigma^2\right)$, and $X_i, i = 1, 2, \ldots$ are i.i.d. Moreover, a constraint K is imposed for the mean vector $\mu := (\mu_1, \mu_2, \ldots, \mu_d)$ s.t. $\mu \in K$, where K is some appropriate set that belongs to $\mathbb{R}^d$ (or $K \subseteq \ell_2$ in the infinite-dimensional case). This extra information about the parameter encodes the prior knowledge of the statistician, and facilitates the inference meanwhile. Finally, and importantly, we also assume that there is a potential source of contamination, which is called the adversary. It is powerful in the sense that it has the knowledge of the true model and the access to the original data. For some positive constant $0 \leq \epsilon < c_0 < \frac{1}{2}$ known to both the statistician and the adversary, the adversary is allowed to inspect the original data and modify up to ϵ fraction of the samples arbitrarily. The statistician, however, does not know which samples have been modified. The goal of the statistician is to perform inference about μ based on the contaminated data. A more formal explanation of the the problem we study can be found in Section 1.1.

Performing inference under constraints and even contamination could be a formidable task. Most of the related works focus on the estimation and testing problem under several simple constraints with no contamination (Donoho and Gasko (1992), Liu and Moitra (2021), Liu and Moitra (2023)). One crucial reason for this is that most of the methods and techniques heavily depend on the assumption of the true model behind the data (for example, likelihood ratio test and tests depending on the asymptotic distributions). Consider the following special case where $K = \mathbb{R}^d$ and $\epsilon = 0$. Given N samples $X_1, X_2, \ldots, X_N \in \mathbb{R}^d$ from $\mathcal{N}(\mu, \sigma^2 \mathbf{I}_d)$ with unknown μ and known $\sigma \in \mathbb{R}^+$. In the testing problem, we are asked to distinguish between two hypotheses:

$$
\begin{aligned}
H_0 &: \mu = 0, \\
H_1 &: \|\mu\|_2 \geq \rho.
\end{aligned}
\tag{1.1}
$$

In the estimation problem, we are asked to give an estimate $\widehat{\mu}$ of μ to achieve a small error under a certain loss function (a natural one here is the ℓ_2 loss). $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ is a sufficient statistic for the model. The testing and estimation based on $\bar{X}$ is widely known to be optimal for both problems. Specifically, to control errors within the level of (small) constant, one needs $\rho \gtrsim \frac{d^{1/4}}{\sqrt{N}} \sigma$ for the testing problem (in the minimax sense) (Ingster and Suslina (2003), Diakonikolas et al. (2022)) while $\inf_{\widehat{\mu}} \sup_{\mu \in \mathbb{R}^d} \mathbb{E} \|\widehat{\mu} - \mu\|_2^2 \gtrsim \frac{d}{N} \sigma^2$ if one wants to optimally estimate μ in ℓ_2 norm (LeCam (1973), also see Example 15.4 in Wainwright (2019)). However, as mentioned, the property of $\bar{X}$

is very sensitive to the authenticity of the samples. The method quickly breaks down with even modification on a single sample (the adversary may modify any single datum, thereby forcing the mean to be zero and causing considerable type-II error.). This is precisely the major reason why testing and estimation problems are very challenging when studied from a robust perspective.

Below, we provide an informal introduction to several key concepts that play a central role in our study — namely, the strong contamination model, the Kolmogorov k-width, and quadratically convex orthosymmetric (QCO) sets. Throughout this paper, we assume that the constraint set K is a QCO set, and we operate under the strong contamination model. As will be shown later, the Kolmogorov widths of K determine the critical radius of the testing problem, i.e., the minimal ρ under which the problem is testable in the minimax sense.

Strong Contamination Model

Throughout the years, different types of adversary models have been proposed in the field of robust statistics. One of the earliest ones, the *Huber contamination model*, was introduced by Huber (1964), where instead of directly sampling from the authentic distribution $\mathcal{P}$, it is assumed that there exists some error distribution $\mathcal{Q}$ selected by the adversary, and the data are sampled from the mixture $(1 - \epsilon)\mathcal{P} + \epsilon\mathcal{Q}$. Huber's contamination model can be regarded as a case of the larger *oblivious adversary*, where while the adversary has the adaptivity to some extent, the contamination process is guaranteed to be independent with the original samples. In contrast, the adversary in the *strong contamination model* (Diakonikolas and Kane (2019)) is granted with strong adaptivity, specifically, with knowledge of the original data, the true model and even the algorithms. The adversary is allowed to inspect and modify the original data based on its knowledge. In this model, the fraction of the contaminated samples must be controlled to obtain useful information. Note that this quantity is not bounded in the Huber contamination model (although the expected number of corruptions is ϵN). Other types of contamination are also studied, including the perturbation of the sampling distribution with bounded TV distance (which also includes Huber's contamination model). See Diakonikolas and Kane (2019) for an overview.

Kolmogorov k-Width

Introduced by Kolmogoroff (1936) in the study of function spaces and approximation theory, the Kolmogorov k-width (also known as the Kolmogorov N-width) characterizes how well an element can be reached by some linear approximations of intrinsic dimension k, given that the element belongs to a linear space and some restrictions are applied. For a given k, it selects the best k-dimensional projection operator under the measure of minimax error. Readers can refer to Definition 2.2 for a formal presentation. Kolmogorov k-width is a classic concept and has seen usage in various fields. See Floater et al. (2021) for a recent work in the problem of low-rank approximation of matrices using the Kolmogorov k-width. Evans et al. (2009) uses this concept to study the k-method in isogeometric analysis. Kolmogorov k-width is also popular in the problems related to partial differential equations (PDEs), where the decay rate of the Kolmogorov k-width is intensely studied (Cohen and DeVore (2015), Greif and Urban (2019)). See also Papapicco et al. (2022) for an application in neural networks and PDEs, and a very recent work, Florian Arbes and Urban (2025), for an application in optimal transportation and PDEs. Finally, see Greif and Urban (2019) and Liang and Parks (1996) for its power in the wave problem and wavelet representations. One can find more details and properties of the Kolmogorov k-width in Pinkus (2012).

QCO Sets

Quadratically convex orthosymmetric (QCO) sets were first introduced in Donoho et al. (1990) (see also (Johnstone, 2019, Chapter 4.8)), where the authors give many examples of such sets, such as hyperrectangles, ellipsoids (both aligned with the coordinate axes) and sets of the form $\{\theta \mid \sum_{i \in I} a_i \psi(\theta_i^2) \leq 1\}$, where ψ is a given convex function and $a_i \in \mathbb{R}^+, i = 1, 2, \ldots$. The last example implies that all unit balls in (weighted) ℓ_p norms, for $p \geq 2$ are QCO sets. The set of QCO constraints is therefore much larger than simply the set of all ellipsoids. See Definition 2.1 for the formal definition of a QCO set. In contrast to the present work, which focuses on testing, Donoho et al. (1990) characterized the minimax estimation rates in the corresponding Gaussian sequence model under an uncorrupted setting. It turns out that the Kolmogorov widths of the set K (see Definition 2.2) determine the minimax rate of estimation. Later, by taking $\epsilon = 0$, we will observe that the Kolmogorov widths also determine the minimax rate of testing with QCO constraints, though the resulting rate is strictly smaller. Furthermore, we shall compare the minimax testing rate under QCO constrains with corruption to the corresponding minimax estimation rate (recently derived by Prasadan and Neykov (2025)) and establish the same fundamental conclusion—testing is intrinsically easier than estimation.

1.1 Problem Formulation and Our Contributions

We now present the formal definition of the problem of interest. Specifically, we study the problem of minimax optimal signal detection in Gaussian noise (i.e., Gaussian mean testing) under the QCO constraints and strong ϵ-contamination model in both the high-dimensional and the classic scenarios. Let $\widetilde{\mathbf{X}} := \left\{\widetilde{X}_1, \ldots, \widetilde{X}_N\right\}$ be the set of original i.i.d. samples, where $\widetilde{X}_i := \left(\widetilde{X}_{i1}, \widetilde{X}_{i2}, \ldots, \widetilde{X}_{id}\right)^\top$, $\widetilde{X}_{ij} = \mu_j + \sigma \cdot \xi_{ij}$ and $\mu := (\mu_1, \mu_2, \ldots, \mu_d)^\top \in K \subset \mathbb{R}^d$ for a known QCO set K (i.e., the prior knowledge of the constraint). $\xi_{ij} \sim \mathcal{N}(0, 1)$ are i.i.d. standard Gaussian noises. Instead of directly observing the original samples, an unknown adaptive adversary $\mathcal{C}$ from the strong ϵ-contamination model (see Definition 2.8 for the formal definition) tampers with the data. The adversary swaps a fraction of the samples with arbitrary unknown values, where the contaminated fraction of the data is no more than a known, predetermined constant $\epsilon \in [0, c_0)$ for some small $c_0 < \frac{1}{2}$. The objective is to test the following null and alternative hypotheses with controlled type-I and type-II errors from the contaminated samples $\mathbf{X} = \{X_i, \ldots, X_N\} := \mathcal{C}(\widetilde{\mathbf{X}})$

$$
\begin{aligned}
&H_0 : \mu = 0, \\
&H_1 : \|\mu\|_2 \geq \rho, \mu \in K.
\end{aligned}
\tag{1.2}
$$

Intuitively, if $\|\mu\|_2$ is overly close to zero or the fraction of the contaminated samples is overly large, we are not able to detect the difference in the worst case. Therefore, we are interested in characterizing the asymptotically smallest value of ρ (depending on other parameters) for which the testing problem becomes feasible. In detail, for a given desired type-I and type-II errors α, define the set of acceptable tests under H_0 as,

$$
A_s(N, d, K, \epsilon, \alpha, \sigma) := \left\{\phi : \sup_{\mathcal{C}} \mathbb{P}_0\left(\phi(\mathcal{C}(\widetilde{\mathbf{X}})) = 1\right) \leq \alpha\right\},
\tag{1.3}
$$

where $\mathbb{P}_0$ denotes the law of $\widetilde{\mathbf{X}}$ when $\mu = \mathbf{0}$, and similar for $\mathbb{P}_\mu$. We write A_s and omit the dependency hereafter when the context is clear. Let

$$\rho(N, d, K, \epsilon, \alpha, \sigma) := \inf_\rho \left\{ \rho : \rho > 0, \sup_{\|\mu\|_2 \geq \rho, \mu \in K} \sup_{\mathcal{C}} \mathbb{P}_\mu \left(\phi(\mathcal{C}(\mathbf{X})) = 0 \right) \leq \alpha, \phi \in A_s \right\} \qquad (1.4)$$

be the smallest possible ρ such that the testing task with both errors less than α is doable. We also write it as ρ_{critical} for convenience when the context is clear. We are going to study the asymptotic behaviour of $\rho(N, d, K, \epsilon, \alpha, \sigma)$. See Figure 1 for an illustrative example of the problem.
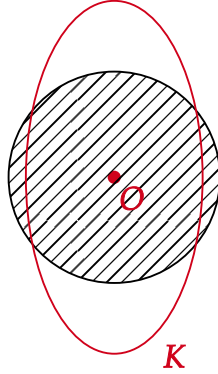

Figure 1: Testing $H_0 : \mu = 0$ vs $H_1 : \|\mu\|_2 \geq \rho, \mu \in K$ for the axis aligned ellipsoid (which is a special case of a QCO set) K based on a single observation X. Upon applying a rotation to the data, one can use the test we develop in Section 3.2 to solve the problems optimally for rotated (non-axis aligned) ellipsoids or more generally QCO sets (which are still centered around the origin). See our first remark in the discussion section for more details.

Our main result shows that ρ_{critical} is determined by three independently valid lower bounds. Moreover, the lower bounds are tight and optimal with only the exception of logarithmic terms. The first lower bound captures the geometric nature of the QCO set K via the Kolmogorov k-width. This lower bound is independent of the adversary, therefore also works in the uncorrupted setting, i.e., when $\epsilon = 0$. The second lower bound is obtained within the framework of *differential privacy*, which is introduced in Section C.3. The third lower bound is derived from a combined argument related to the identical mixture distributions, computation of the total variation and a designed strategy of the adversary. For the upper bounds, we present two algorithms which conclude whether H_0 holds or not while ensuring type-I and type-II errors both less than α. The requirements on $\|\mu\|_2$ imposed by the first algorithm are slightly closer to the theoretical lower bounds; however, this improvement comes at the cost of exponential time complexity, rendering the method computationally intractable for large N and d. The second algorithm, in contrast, provides a slightly weaker upper bound than the first algorithm (though still nearly optimal) while achieving polynomial-time complexity. Nevertheless, both upper bounds suggest that the three lower bounds previously discussed suffice to fully characterize the behavior of ρ_{critical} while also imply a phase transition phenomenon regarding ϵ for the contamination process. Finally, the results and algorithms derived in the ℓ_2 norm scenario can be naturally extended to the ℓ_p case for $1 \leq p < 2$. Section 6 contains such relevant discussions.

1.2 Related Literature

In this section, we discuss some relevant works. We first comment on the literature on minimax testing with uncorrupted data. In their seminal body of work on nonparametric testing, Ingster (1982, 1993a,b,c); Ingster and Suslina (2000) showed the difference between testing and estimation in multiple problems. See also Ingster and Suslina (2003) for a summary of the many existing results in the area. Despite various progress in *goodness-of-fit* testing in Gaussian models, we are not aware of a general result dedicated to QCO sets. For certain special types of ellipsoids, the problem was first studied by Ermakov (1991), where the author derived exact minimax rates. Suslina (1996) studied the asymptotics of the problem of signal detection in ℓ_q ellipsoids with ℓ_p ball removed. Since ℓ_q ellipsoid for $q \geq 2$ are QCO sets, when $p = 2$ we provide a generalization of these results. For general ellipsoids, Baraud (2002) studied the testing problem in $\mathbb{R}^n$ and ℓ_2 and obtained the minimax optimal critical radius. Our work can hence be seen as a generalization to part of Baraud (2002). In an interesting recent work, Wei and Wainwright (2020) studied local testing rates on ellipsoids. Here local means that the point of interest under the null is not necessarily the zero point, but can be an arbitrary point in the ellipsoid. Wei and Wainwright (2020) proved upper and lower bounds based on localized Kolmogorov widths, which are not guaranteed to match in general, but do match for certain examples with Sobolev ellipsoids. In contrast, the present work focuses on more general sets, but the point of interest under the null hypothesis is always anchored at the zero point, and the critical radius is determined based on the global (instead of local) Kolmogorov widths of the set.

Subsequently, we note that the recent literature on robust testing is relatively sparse. A notable exception is the paper Canonne et al. (2023), which studies a special case of the problem we consider, where $K = \mathbb{R}^d$, but considers a broader class of adversaries. The author also developed a polynomial-time algorithm, the framework of which we adapt to develop a more general algorithm suitable for our settings. In addition, Diakonikolas et al. (2017) develops a statistical query (SQ) lower bound for robust estimation problem in high-dimensional setting and extends the results to the testing problem.

1.3 Organization

This paper is structured as follows. Section 2 introduces the notations and several important concepts used through the paper. Section 3 presents the lower and upper bounds for the signal detection problem and contains the main results of the paper. It also contains the discussion of the phase transitions of the problem. Section 5 presents the experimental results of the polynomial-time algorithms related to Theorem 3.11 in Section 3. Section 6 contains the discussion which compares the minimax rates of testing to estimation and the extension of the main results to the ℓ_p norm separation. It also includes some open problems for future works. We attach necessary probabilistic tools and other lemmas in Appendix B. The proofs to the lower and upper bounds and other main corollaries in Section 3 and Section 6 are included in Appendix C. Appendix A presents the pseudo codes of the related algorithms.

2 Preliminary

2.1 Notations and Definitions

To simplify the analysis, we assume that $K \subset \mathcal{X} = \mathbb{R}^d$, and therefore $\mu \in \mathbb{R}^d, X_i \in \mathbb{R}^d$ through the paper. However, one can extend the results to the case when $K \subset \ell_2$ with the same techniques and algorithms with minimal efforts given the assumption $\lim_{k \to \infty} D_k(K) = 0$. See Definition 2.2 for further details. We use ρ to denote the value of the ℓ_2 norm of μ under H_1. $\|\cdot\|_p$ means the ℓ_p norm in $\mathbb{R}^d$ for $1 \le p \le \infty$. When A is a set, $|A|$ denotes the cardinality of A. We use calligraphic font $\mathcal{P}, \mathcal{Q}, \mathcal{R} \ldots$ to represent probability distributions. $\mathcal{P} \ll \mathcal{Q}$ means that $\mathcal{P}$ is absolutely continuous with respect to $\mathcal{Q}$. $\mathrm{TV}(\mathcal{P}, \mathcal{Q})$ denotes the total variation distance and $\mathrm{KL}(\mathcal{P}, \mathcal{Q})$ for the Kullback-Leibler divergence if $\mathcal{P} \ll \mathcal{Q}$. When the underlying probability distribution is not explicitly specified, we use $\mathbb{P}$ alternatively to denote its probability law in general. We use $\langle \cdot, \cdot \rangle$ and $(\cdot)^\top (\cdot)$ alternatively (possibly with a subscript when needed) to represent the inner product in a Hilbert space. $\mathrm{diag}\{a\}$ represents the diagonal matrix with diagonal entries $a_i, i = 1, 2, \ldots$. The convenient notation $[n]$ denotes the set $\{1, 2, \ldots, n\}$. When I is an index set, v_I represents the restriction of v on I if v is a vector and $\mathbf{Y}_I$ represents the sub-matrix of $\mathbf{Y}$ with rows selected according to I if $\mathbf{Y}$ is a matrix. $\mathcal{O}, \Theta$, Ω notations are used to denote the asymptotic order of rates. The lowercase $c_1, c_2, \ldots$ denote the universal constants and their specific values may differ from line to line.

In addition to these plain notations, we let $\widetilde{\mathbf{X}} := \left\{ \widetilde{X}_1, \ldots, \widetilde{X}_N \right\}$ represent the original samples, $\mathbf{X} := \mathcal{C}(\widetilde{\mathbf{X}}) = \{X_1, \ldots, X_N\}$ represent the data contaminated by an adversary $\mathcal{C}$. Let the uppercase C denote the unknown set of indices corresponding to the contaminated samples. By this notation, we have $\widetilde{\mathbf{X}}_{[N] \setminus C} = \mathbf{X}_{[N] \setminus C}$. Moreover, the following definitions formally introduce the concepts of QCO set and the Kolmogorov k-width used in this paper, which play crucial roles for our results.

Definition 2.1 (Quadratically convex orthosymmetric (QCO) set). *Given a set $K \subset \mathbb{R}^d$, we say K is a quadratically convex orthosymmetric (QCO) set if it satisfies the following conditions:*
(1), K is convex;
(2), K is quadratically convex, which means that K^2 is also convex, where K^2 is defined as

$$K^2 := \left\{ \left(\theta_1^2, \ldots, \theta_d^2 \right)^\top \mid (\theta_1, \ldots, \theta_d)^\top \in K \right\};$$

(3), K is orthosymmetric, which means that if $\theta = (\theta_1, \ldots, \theta_d)^\top \in K$, then $\theta_\eta := (\eta_1 \theta_1, \ldots, \eta_d \theta_d)^\top \in K$, where $\eta_i \in \{-1, 1\}, 1 \le i \le d$.

Remark. *The definition of QCO set can be naturally extended when $K \subset \ell_2$, as we only need to replace $[d]$ by the positive integers $\mathbb{N}^+$.*

Definition 2.2 (Kolmogorov k-width, also known as Kolmogorov N-width). *Let $\mathcal{X}$ be a Banach space equipped with the norm $\|\cdot\|$, and $K \subset \mathcal{X}$ is a subset. The Kolmogorov k-width is defined as*

$$D_k(K) = \inf_{P \in \mathcal{P}_k} \sup_{\theta \in K} \|\theta - P\theta\|, \tag{2.1}$$

where $\mathcal{P}_k$ is the set of all projection operators that project a vector onto some subspace of $\mathcal{X}$ with intrinsic dimension k.

Remark. *In this work, our analysis will be confined to orthogonal projections. Moreover, though $\mathcal{P}_k$ is the set of all orthogonal projection with intrinsic dimension k, we discretize this set — only consider the projections that are aligned with the coordinate axes without loss of generality, as shown by the analysis later. For now, let us first assume that the norm leveraged in the definition is the ℓ_2 norm.*

Based on the definition, one can derive several basic properties of the Kolmogorov widths. It is non-increasing with respect to k, meaning that $D_{k_1}(K) \geq D_{k_2}(K)$ for any $0 \leq k_1 \leq k_2$. In addition, for a finite-dimensional space $\mathcal{X}$ with full dimension d, we always have $D_0(K) = \sup_{\theta \in K} \|\theta\|_2$ and $D_d(K) = 0$ for any $K \subset \mathcal{X}$. For the more general case where $K \subset \ell_2$, we shall assume $\lim_{d \to \infty} D_d(K) = 0$. In this paper, we further define $D_k(K) = \infty$ when $k < 0$ and $D_k(K) = 0$ for $k > d$ for the convenience of the marginal cases.

The connection between the Kolmogorov widths and our main results is established through the following definition of the first and second dimensions of the optimal projections, which are designed to identify the optimal k satisfying certain marginal conditions determined by the geometric properties of the constraint set K and the corruption process.

Definition 2.3 (First optimal dimension and projection). *Given the noise scale σ, the number of samples N and the prior knowledge K, the first optimal dimension is an integer $k_1^\star(K, \sigma, N) \in [0, d]$ depending on K, σ, and N where it satisfies that $D_{k_1^\star - 1}(K) > \frac{(k_1^\star)^{1/4}}{\sqrt{N}}\sigma$ but $D_{k_1^\star}(K) \leq \frac{(k_1^\star + 1)^{1/4}}{\sqrt{N}}\sigma$. The projection operator corresponding to $k_1^\star(K, \sigma, N)$ is denoted as $P_1^\star(K, \sigma, N)$.*

Definition 2.4 (Second optimal dimension and projection). *Given the noise scale σ, the number of samples N, the prior knowledge K and the corruption rate ϵ, the second optimal dimension is an integer $k_2^\star(K, \sigma, N, \epsilon) \in [0, d]$ depending on K, σ, N, and ϵ where it satisfies that $D_{k_2^\star - 1}(K) > \frac{(k_2^\star)^{1/4}\sqrt{\epsilon}}{N^{1/4}}\sigma$ but $D_{k_2^\star}(K) \leq \frac{(k_2^\star + 1)^{1/4}\sqrt{\epsilon}}{N^{1/4}}\sigma$. The projection operator corresponding to $k_2^\star(K, \sigma, N, \epsilon)$ is denoted as $P_2^\star(K, \sigma, N, \epsilon)$.*

We note that from the monotone property of the Kolmogorov widths and the boundary values, such $k_1^\star(K, \sigma, N)$ and $k_2^\star(K, \sigma, N, \epsilon)$ must exist and are unique. In fact, it could be equally defined as

$$k_1^\star(K, \sigma, N) := \max\left\{ j \mid 0 \leq j \leq d, D_{j-1}(K) > \frac{j^{\frac{1}{4}}}{\sqrt{N}}\sigma \right\},$$

$$k_2^\star(K, \sigma, N, \epsilon) := \max\left\{ j \mid 0 \leq j \leq d, D_{j-1}(K) > \frac{j^{\frac{1}{4}}\sqrt{\epsilon}}{N^{\frac{1}{4}}}\sigma \right\},$$

or

$$k_1^\star(K, \sigma, N) := \min\left\{ j \mid 0 \leq j \leq d, D_j(K) \leq \frac{(j+1)^{\frac{1}{4}}}{\sqrt{N}}\sigma \right\},$$

$$k_2^\star(K, \sigma, N, \epsilon) := \min\left\{ j \mid 0 \leq j \leq d, D_j(K) \leq \frac{(j+1)^{\frac{1}{4}}\sqrt{\epsilon}}{N^{\frac{1}{4}}}\sigma \right\}.$$

In the general case where $K \subset \ell_2$, the requirement that $\lim_{d \to \infty} D_d(K) = 0$ ensures $k_1^\star(K, \sigma, N) < \infty$[1], which is important for the upper bounds and the implementation of the algorithms discussed later.

---

[1] $k_2^\star(K, \sigma, N, \epsilon)$ is possibly infinite. This is the case for example if $\epsilon = 0$, $D_k(K) > 0, \forall k \in \mathbb{N}$.

When the context is clear, we omit the dependency with K, σ, N, and ϵ, only writing them as $k_1^\star, k_2^\star$ and $P_1^\star, P_2^\star$.

Below are several concrete examples of the concepts introduced above, which serve to illustrate the definitions and provide motivation from practical applications.

Example 2.5 ($\mathbb{R}^d$). *The whole space $\mathbb{R}^d$ is trivially convex, quadratically convex and orthosymmetric. Hence, it is a QCO set. The Kolmogorov k-width can be easily calculated as*

$$D_k(\mathbb{R}^d) = \begin{cases} \infty, & (0 \leq k \leq d-1), \\ 0, & (k = d). \end{cases}$$

The Kolmogorov k-width is infinite until k reaches d, which implies one cannot uniformly approximate a vector in $\mathbb{R}^d$ by a linear subspace well without any additional constraint. The first and second optimal dimensions are $k_1^\star = k_2^\star = d, \forall \epsilon \in [0, c_0)$.

Example 2.6 (Hyperrectangles in $\mathbb{R}^d$). *The hyperrectangle in $\mathbb{R}^d$ is defined as*

$$H_d(a_1, \ldots, a_d) := \left\{ x \mid x = (x_1, \ldots, x_d) \in \mathbb{R}^d, |x_i| \leq a_i, a_i \geq 0, 1 \leq i \leq d \right\}.$$

Obviously, $H_d(a_1, \ldots, a_d)$ is convex and orthosymmetric. One can verify that $H_d^2(a_1, \ldots, a_d) = \{ x^2 \mid x \in H_d(a_1, \ldots, a_d) \}$ is also a hyperrectangle with each side being $[0, a_i^2], 1 \leq i \leq d$. Therefore, it is a QCO set. By reindexing if necessary, we may assume that the sequence $\{a_i\}_{i=1}^d$ is arranged in decreasing order. The Kolmogorov widths are

$$D_k(H_d(a_1, \ldots, a_d)) = \begin{cases} \sqrt{\sum_{i=k+1}^d a_i^2}, & (0 \leq k \leq d-1), \\ 0, & (k = d). \end{cases}$$

The decay rate of the Kolmogorov widths depends on the specific choice of a_i. The first optimal dimension is $k_1^\star = \max \left\{ j \mid 1 \leq j \leq d, \sqrt{\sum_{i=j}^d a_i^2} > \frac{j^{1/4}}{\sqrt{N}} \sigma \right\}$ when $\sqrt{\sum_{i=1}^d a_i^2} > \frac{1}{\sqrt{N}} \sigma$ and otherwise 0. The second optimal dimension is $k_2^\star = \max \left\{ j \mid 1 \leq j \leq d, \sqrt{\sum_{i=j}^d a_i^2} > \frac{j^{1/4} \sqrt{\epsilon}}{N^{1/4}} \sigma \right\}$ when $\sqrt{\sum_{i=1}^d a_i^2} > \frac{\sqrt{\epsilon}}{N^{1/4}} \sigma$ otherwise 0.

Example 2.7 (Ellipses in $\mathbb{R}^d$). *An ellipse in $\mathbb{R}^d$ that is centered at the origin and aligns with the axes can be represented as*

$$E_d(a_1, \ldots, a_d) = \left\{ x \mid x \in \mathbb{R}^d, \sum_{i=1}^d \frac{x_i^2}{a_i} \leq 1 \right\}.$$

Any such ellipse is convex and orthosymmetric. The square of the ellipse $E_d(a_1, \ldots, a_d)$ is

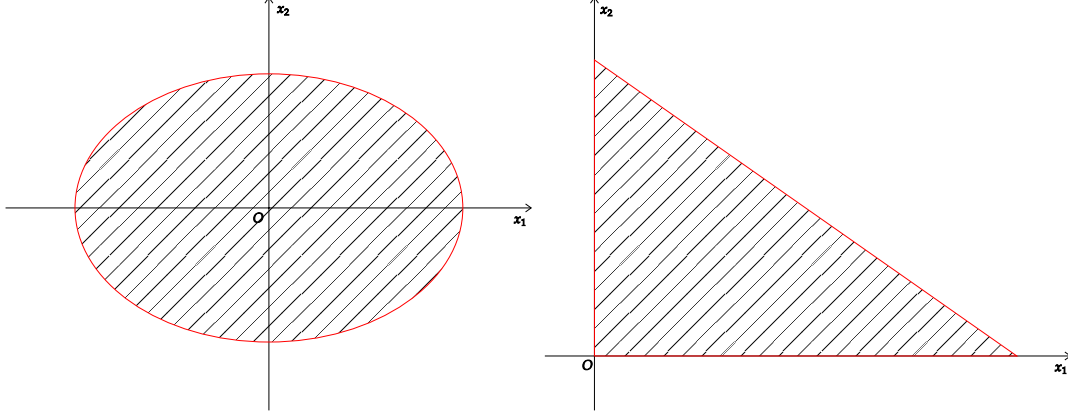$$E_d^2(a_1, \ldots, a_d) = \left\{ x \mid x \in \mathbb{R}_+^d, \sum_{i=1}^d \frac{x_i}{a_i} \leq 1 \right\},$$

Figure 2: An example of an ellipse and its square-set in $\mathbb{R}^2$. The square-set is a triangle in $\mathbb{R}^2$, which is convex.

which is part of a simplex (enclosed by several hyperplanes) in $\mathbb{R}^d$, and therefore convex. Figure 2 presents a depiction of $E_d^2(a_1, \ldots, a_d)$ when $d = 2$. Again, we assume $\{a_i\}_{i=1}^d$ is arranged in decreasing order, by which the Kolmogorov widths are

$$D_k(E_d(a_1, \ldots, a_d)) = \begin{cases} \sqrt{a_{k+1}}, & (0 \le k \le d-1), \\ 0, & (k = d). \end{cases}$$

The first optimal dimension is $k_1^\star = \max\left\{ j \mid 1 \le j \le d, a_j > \frac{\sqrt{j}}{N}\sigma^2 \right\}$ when $a_1 > \frac{1}{N}\sigma^2$ and 0 otherwise. The second optimal dimension $k_2^\star = \max\left\{ j \mid 1 \le j \le d, a_j > \frac{\sqrt{j}\epsilon}{\sqrt{N}}\sigma^2 \right\}$ when $a_1 > \frac{\epsilon}{\sqrt{N}}\sigma^2$ and 0 otherwise.

When $a_i = i^{-2s}$ for some $s > 1$, the corresponding ellipse in $\mathbb{R}^d$ can be regarded as a special case of the more general ellipses of Sobolev types, which is defined as $S(s) := \left\{ x \mid x \in \mathbb{R}^\infty, \sum\limits_{i=1}^\infty i^{2s}x_i^2 \le 1 \right\}$. This type of sets is closely related to the estimation and density approximation problem with the class of s-times differentiable functions in probability space. We refer readers to *Wainwright (2019)* and *Tsybakov (2009)* for more details.

Remark. When an ellipse is not axes-aligned: (1), the orthosymmetric property is not guaranteed, which can be easily checked; (2), the quadratically convex property is true for $d = 2$ and false for $d \ge 3$. To gain the idea of (2), consider an ellipse

$$E_d(A) = \left\{ x \mid x \in \mathbb{R}^d, x^\top A^{-1} x \le 1 \right\},$$

and its squared set denoted by $E_d^2(A)$. For any point y in $E_d^2(A)$, by definition there exists a sign set $\gamma \in \{-1, 1\}^d$ such that $\left(\gamma_1\sqrt{y_1}, \ldots, \gamma_d\sqrt{y_d} \right) \in E_d(A)$. Therefore, $E_d^2(A)$ can be expressed by

$$E_d^2(A) = \left\{ y \mid y \in \mathbb{R}_+^d, \min_{\gamma \in \{-1,1\}^d} \left\{ \sum_{i,j} A_{ij}\gamma_i\gamma_j\sqrt{y_iy_j} \right\} \le 1 \right\}.$$

When $d = 2$, by selecting γ_1 and γ_2 properly, $(y_1, y_2) \in \mathbb{E}_d^2(A)$ if and only if

$$A_{11}y_1 + A_{22}y_2 - 2\left|A_{12}\right|\sqrt{y_1y_2} \le 1.$$

Therefore, $E_2^2(A)$ is convex since it is a sublevel set of a convex function. However, for $d \geq 3$, it is impossible to let all coefficients of $\sqrt{y_i y_j}$ be negative, which leads to the non-convexity of the LHS.

Finally, we conclude this section with the definition of the strong ϵ-contamination model studied in this paper.

Definition 2.8 (Strong ϵ-contamination model). *Suppose that $\widetilde{\mathbf{X}} = \left\{ \widetilde{X}_1, \ldots, \widetilde{X}_N \right\}$ are independently drawn according to a distribution $\mathcal{P}$ under either H_0 or H_1 with unknown label. Let $\epsilon \in [0, c_0)$ be a predetermined known constant for some $c_0 < \frac{1}{2}$. An adversary $\mathcal{C}$ is assumed to have the full knowledge of $\mathcal{P}$. After the sampling step, $\mathcal{C}$ is allowed to inspect $\widetilde{\mathbf{X}}$ and arbitrarily replace up to ϵN of them before they are available for analysis. The resulting corrupted dataset $\mathbf{X} = \mathcal{C}(\widetilde{\mathbf{X}})$ is provided to the statistician while the corruption index set C is unknown.*

3 Main Results

3.1 Lower Bounds

In this section, we introduce three independent lower bounds, i.e., Theorem 3.2, 3.5, and 3.6, which jointly and fully determines the lower bound of ρ_{critical}.

3.1.1 Based on a Geometric Property

The following lemma serves as an important basis for the theoretical lower bounds in Theorems 3.2 and 3.5. The underlying intuition is that the orthosymmetry and convexity of the prior constraint K imply the existence of vectors with small individual coordinates that nonetheless yield a significant aggregate norm. Such "spread-out" vectors, and the maximal norm they can attain, are effectively characterized by the Kolmogorov widths of K. Specifically, Lemma 3.1 guarantees the existence of a vector with a large ℓ_2 norm but a small ℓ_∞ norm by leveraging the properties of QCO sets and the definition of the Kolmogorov widths.

Lemma 3.1. *Let $K \subset \mathcal{X}$ be a QCO set and c be a positive constant. Suppose that $D_{k-1}(K) > c\sigma$ for some $k \geq 1$. Then there exists a vector $\theta \in K$ such that $\|\theta\|_2 = c\sigma$ while $\|\theta\|_\infty \leq \frac{c}{\sqrt{k}}\sigma$.*

The proof of Lemma 3.1 is derived from the key properties of the QCO set K, combined with an application of the minimax theorem to a suitably truncated vector in K^2. The formal proof is provided in Appendix C.1. A similar argument also appears in Section 3.3.2 of Neykov (2022).

Such existence in Lemma 3.1 is fundamental for the construction of the least favorable prior, where we define a uniform mixture over the coordinate-wise sign flips of this vector to induce a small total variation distance between the mixture and the null distribution H_0, thereby establishing one of the minimax lower bounds. This standard technique (see, e.g., LeCam (1973)) can be adapted to our current problem with the prior constraint K via Lemma 3.1.

Theorem 3.2 (First lower bound). *If $\rho \leq c(\alpha) \cdot \frac{(k_1^\star)^{1/4}}{\sqrt{N}}\sigma$, where c is a constant that only depends on α, or $k_1^\star = 0$, then*

$$\inf_{\psi:\mathbb{P}_0(\psi=1)\leq\alpha} \sup_{\theta\in K,\|\theta\|\geq\rho} \mathbb{P}_\theta(\psi = 0) \geq \alpha.$$

Moreover, one possible explicit expressions of $c(\alpha)$ is

$$c(\alpha) = \left[2\ln\left(1 + 4\left(1 - 2\alpha\right)^2\right)\right]^{\frac{1}{4}}.$$

For a fixed level α, Theorem 3.2 establishes that if $\rho \lesssim \frac{(k_1^\star)^{1/4}}{\sqrt{N}}\sigma$, no test can uniformly distinguish H_1 from H_0 with non-trivial power. As an immediate consequence, distinguishing H_0 and H_1 remains impossible in the presence of an adversary $\mathcal{C}$ under the same condition, since a valid strategy for $\mathcal{C}$ is simply applying the identity map (i.e., leaving the samples unperturbed).

Moreover, it is worth noting the marginal case when $k_1^\star = 0$. Informally speaking, $k_1^\star = 0$ means that the constraint set K is insufficiently rich to support a signal that is statistically distinguishable from H_0. Since the definition of $k_1^\star$ only relies on K, N, σ, and does not depend on the original samples $\widetilde{\mathbf{X}}$ or the observations $\mathbf{X}$, we shall assume in the following text that $D_0(K) = \sup_{\theta \in K} \|\theta\|_2 > \frac{1}{\sqrt{N}}\sigma$, and therefore $k_1^\star \geq 1$. Otherwise, the test will not be feasible in the first place. We kindly refer readers to Appendix C.2 for the proof of this argument and Theorem 3.2, which relies on the selected vector in Lemma 3.1 and Le Cam's method of "fuzzy hypothesis".

Recall that for uncontaminated observations without constraint, the critical value for $\|\mu\|_2$ is $\rho_{\text{critical}} \asymp \frac{d^{1/4}}{\sqrt{N}}\sigma$ (see Section 3.3.6 in Ingster and Suslina (2003)). Theorem 3.2 generalizes this classic result in the sense that additional prior knowledge of the mean is allowed and the lower bound is refined as a consequence. Indeed, Theorem 3.2 recovers the classic result when we set $\mathcal{X} = \mathbb{R}^d$ (where $k_1^\star = d$ according to Example 2.5). Moreover, the theorem implies that when we have non-trivial constraint K on μ, one of the essential dimensional attributes of the problem is the finite $k_1^\star$ instead of d. (Recall that $k_1^\star < \infty$ is ensured even if $K \subset \ell_2$.)

3.1.2 Based on the Framework of Differential Privacy

Differential privacy was first introduced by Dwork et al. (2006) in the context of private data analysis. Formally, it requires the output distribution of a randomized algorithm to be insensitive to small perturbations in the input dataset. A key insight in this field is the strong connection between robust statistics and differential privacy; specifically, under mild conditions, robust estimators can be converted into differentially private algorithms using the *propose-test-release* framework. Subsequently, differential privacy was applied to the goodness-of-fit testing (also called identity testing) in statistics. Pioneering works in this direction include Gaboardi et al. (2016), Rogers and Kifer (2017), Cai et al. (2017).

Definition 3.3 ((ϵ, δ)-differential privacy, Dwork et al. (2006)). *For a pair of non-negative real numbers (ϵ, δ), a randomized algorithm $\mathcal{A}$ defined on the sample space Ω is a (ϵ, δ)-differentially private algorithm if for any $S \subset \text{range}(\mathcal{A})$ and any $X, X' \in \Omega$ with the Hamming distance $d_H(X, X') \leq 1$, we have*

$$\mathbb{P}\left(\mathcal{A}(X) \in S\right) \leq e^\epsilon \mathbb{P}\left(\mathcal{A}(X') \in S\right) + \delta. \tag{3.1}$$

When $\delta = 0$, the notion is called pure differential privacy; when $\delta > 0$, it is referred to as approximate differential privacy.

Our second lower bound relies on a fundamental coupling lemma: if two datasets $\mathbf{X}$ and $\mathbf{X}'$, drawn from $\mathcal{U}$ and $\mathcal{V}$ respectively, can be coupled to overlap on a significant fraction of indices (i.e., $X_i = X_i'$ for most i), then distinguishing between $\mathcal{U}$ and $\mathcal{V}$ is statistically hard.

Lemma 3.4 (Core lemma). *Suppose* $\mathbf{X} = (X_1, X_2, \ldots, X_N)^\top$ *and* $\mathbf{X}' = (X_1', X_2', \ldots, X_N')^\top$ *are random matrices with rows* $X_i, X_i' \in \mathbb{R}^d$. *Assume the laws of* $\mathbf{X}$ *and* $\mathbf{X}'$ *are* $\mathcal{U}$ *and* $\mathcal{V}$, *respectively. Additionally, assume that there exists a coupling between* $\mathcal{U}$ *and* $\mathcal{V}$ *under which the expectation of the Hamming distance between* $\mathbf{X}$ *and* $\mathbf{X}'$ *is* $\mathcal{O}(\epsilon N)$, *i.e.,*

$$\mathbb{E}_{(\mathbf{X},\mathbf{X}')\sim(\mathcal{U},\mathcal{V})} d_H(\mathbf{X}, \mathbf{X}') \lesssim \epsilon N. \tag{3.2}$$

Then, there exists no robust test capable of distinguishing between $\mathcal{U}$ *and* $\mathcal{V}$ *under contamination level of* $\Omega(\epsilon)$ *while achieving type-I and type-II error probabilities below* 0.1 *simultaneously and uniformly.*

The proof of Lemma 3.4 relies on a result established by Acharya et al. (2018), which states that if (3.2) holds for a $(0, \delta)$-differentially private algorithm, then necessarily $\epsilon N = \Omega(1/\delta)$. Furthermore, for any robust test attempting to distinguish between $\mathcal{U}$ and $\mathcal{V}$, one can construct a $(0, \delta)$-differentially private algorithm via the *black-box robustness-to-privacy transformation* proposed in Canonne et al. (2023) and Hopkins et al. (2023). Combining these two arguments yields a necessary condition for the expected Hamming distance between $\mathbf{X}$ and $\mathbf{X}'$.

The following theorem, which establishes the second lower bound for the testing problem (1.2), builds upon Lemma 3.4 and the selected vector $\theta \in K$ associated with the second optimal dimension $k_2^\star$ (see Lemma 3.1). Specifically, let $\mathcal{U} = \mathbb{P}_0^{\otimes N}$ and $\mathcal{V} = \mathbb{P}_1^{\otimes N}$, where $\mathbb{P}_0$ denotes $\mathcal{N}(0, \sigma^2)$ and $\mathbb{P}_1$ is the mixture of $\mathcal{N}(v, \sigma^2)$ with v drawn uniformly from the set

$$V := \{v \mid v_i = \gamma_i \theta_i, \gamma_i \in \{-1, 1\}, i \geq 1, i \in I\}.$$

We construct a chain of distributions interpolating between $\mathcal{U}$ and $\mathcal{V}$ based on the optimal coupling (Lemma B.5). Consequently, the computation of the expected Hamming distance is reduced to the sum of the total variation distances between adjacent distributions in the chain. We kindly refer readers to Appendix C.3 for the formal proofs of Lemma 3.4 and Theorem 3.5.

Theorem 3.5 (Second lower bound). *Consider the testing problem* (1.2) *under the strong ϵ-contamination model with an adversary* $\mathcal{C}$. *If the signal strength satisfies* $\rho \lesssim \frac{(k_2^\star)^{1/4}\sqrt{\epsilon}\sigma}{N^{1/4}}$, *then no test can simultaneously achieve Type-I and Type-II errors bounded by* 0.1. *Formally, under this condition, we have*

$$\inf_{\phi \in A_s(0.1)} \sup_{\mu \in K} \sup_{\mathcal{C}} \mathbb{P}_\mu(\phi(\mathcal{C}(\mathbf{X}')) = 0) \geq 0.1,$$

where $A_s(0.1)$ *denotes the set of tests with Type-I error at most* 0.1, *as defined in* (1.3).

In Lemma 3.4 and Theorem 3.5, we fix the constant at 0.1 for simplicity, though it can be replaced by any sufficiently small universal constant. In the high-dimensional setting where d (and possibly $k_2^\star$) may substantially exceed N, one must ensure that

$$\frac{(k_2^\star)^{1/4}\sqrt{\epsilon}\sigma}{N^{1/4}} \leq D_0(K) = \sup_{\theta \in K} \|\theta\|_2.$$

If this condition is violated, the testing problem becomes statistically impossible.

3.1.3 Based on the Estimation Lower Bounds

The third lower bound, as established in Theorem 3.6 below, contrasts with Theorems 3.2 and 3.5 in that it is determined solely by the diameter of K (denoted as $D_0(K)$), the fraction of contamination ϵ, and the noise scale σ, while remaining independent of N.

Theorem 3.6 (Third lower bound). *For the signal detection problem* (1.2) *under the strong ϵ-contamination model with an adversary $\mathcal{C}$, if $\epsilon \gtrsim \frac{1}{\sqrt{N}}$ and $\rho \lesssim \epsilon\sigma$, we always have*

$$\inf_{\phi \in A_s} \sup_{\mu \in K, \|\mu\|_2 \geq \rho} \sup_{\mathcal{C}} \mathbb{P}_\mu(\phi(\mathcal{C}(\widetilde{\mathbf{X}})) = 0) = \Theta(1), \tag{3.3}$$

where A_s is the set of acceptable tests defined as (1.3).

Remark. *(1). This lower bound is notable as it coincides with the minimax rate for the estimation problem under the strong ϵ-contamination model (see, e.g., Prasadan and Neykov (2025)). This finding stands in contrast to the conventional wisdom that testing is typically statistically easier than estimation (see Ingster (1993a), Baraud (2002)). However, our subsequent analysis confirms that in this setting, the minimax rates for estimation and testing indeed match [or are of the same order] when certain conditions are met.*
(2). The condition $\epsilon \gtrsim \frac{1}{\sqrt{N}}$ can be safely removed. See our discussions in Section 3.3.

Recall that $D_0(K) = \sup_{\theta \in K} \|\theta\|_2$ represents the radius of K (i.e., the maximal distance from the origin). Theorem 3.6 implies that a necessary condition for the existence of a test with Type-I and Type-II errors uniformly bounded by α is that either $\rho \gtrsim \epsilon\sigma$ or $\epsilon = \mathcal{O}(1/\sqrt{N})$. Noting that the signal strength is inherently bounded by $\|\mu\|_2 \leq D_0(K)$, it follows that if $\epsilon \gtrsim 1/\sqrt{N}$ and $\epsilon\sigma \gtrsim D_0(K)$, consistent testing is statistically impossible.

The proof of Theorem 3.6 is inspired by a similar lower bound established for the estimation problem under star-shaped constraints in Prasadan and Neykov (2025). It relies on the construction of a specific distribution that admits two distinct mixture representations, with components corresponding to H_0 and H_1 respectively. Crucially, the adversary $\mathcal{C}$ can exploit this structural ambiguity to design a contamination strategy that renders the resulting distributions statistically indistinguishable when the lower bound condition is violated. The full proof is detailed in Appendix C.4.

3.2 Upper Bounds

The upper bounds are established constructively. Specifically, we design explicit testing procedures and identify the required signal strength $\rho = \|\mu\|_2$ that guarantees test validity (i.e., uniformly bounded Type-I and Type-II errors).

3.2.1 Theoretical Algorithm

Since the adversary $\mathcal{C}$ is allowed to modify only a limited number of samples, a natural strategy to mitigate the contamination is to identify subsets of $\mathbf{X}$ that exhibit a *consistency property*.

Definition 3.7 (Consistent subset). *A subset $S \subset \mathbf{X} = \{X_1, \ldots, X_N\}$ is called a consistent subset regarding the contamination fraction ϵ and a test $\phi : 2^S \mapsto \{0, 1\}$ if*
(i), $|S| \geq (1 - \epsilon)N$,
(ii), $\phi(S') = \phi(S)$ for any $S' \subset S$ with $|S'| \geq (1 - 2\epsilon)N$.

As shown in Appendix C.5, such subsets exist with high probability and behave as if they consisted entirely of authentic samples. To provide some intuition, consider a specific χ^2-type test ϕ_e. Assume that the original (unobserved) data $\widetilde{\mathbf{X}}$ is inherently consistent. This guarantees the existence of a consistent subset within $\mathbf{X}$ (at least for the set of uncorrupted samples $\mathbf{X}_{[N]\setminus C}$). Let $\mathbf{X}_S$ denote any such consistent subset. Consequently, we can recover the test result of the unobserved $\widetilde{\mathbf{X}}$ by establishing the following equivalence by the definition of the consistent subset:

$$\phi_e(\mathbf{X}_S) = \phi_e(\mathbf{X}_S \cap \mathbf{X}_{[N]\setminus C}) = \phi_e(\mathbf{X}_{[N]\setminus C}) = \phi_e(\widetilde{\mathbf{X}}).$$

The following theorem formally establishes the existence of a valid testing procedure under the condition (3.4), and consequently provides an upper bound for the testing problem (1.2). Specifically, we demonstrate that, when (3.4) holds, there exists a desired test ϕ_e and at least one consistent subset $\mathbf{X}_S$ with respect to ϕ_e, upon which the entire testing task becomes feasible. To achieve this, we explicitly define an event E_e of χ^2-type, on which ϕ_e is determined, and prove that the original samples $\widetilde{\mathbf{X}}$ is consistent with high probability. The detailed proof is provided in Appendix C.5.

Theorem 3.8 (Theoretical upper bound)**.** *Given a dataset* $\mathbf{X}$ *from the strong ϵ-contamination model and a fixed error tolerance* α*, for any* $k \in [1, d]^1$*, there exists a test* ϕ_e *for* (1.2) *such that the type-I and type-II errors are both uniformly less than* α *if* $\rho^2 \gtrsim E_{raw}^2(\epsilon, K, N, \sigma^2, k)$ *where* $E_{raw}^2(\epsilon, K, N, \sigma^2, k)$ *is defined as*

$$E_{raw}^2(\epsilon, K, N, \sigma^2, k) := D_k^2(K) + \sigma^2 \max \left\{ \frac{\sqrt{k}}{N}, \epsilon^2 \ln\left(\frac{1}{\epsilon}\right), \sqrt{\frac{\epsilon^2 \ln\left(\frac{1}{\epsilon}\right) k}{N}} \right\}. \qquad (3.4)$$

In other words, we have

$$\rho_{critical} \lesssim \min_k E_{raw}(\epsilon, K, N, \sigma^2, k).$$

We can further leverage the results of Theorem 3.8 by optimizing over k and applying the definition of $k_1^\star$ and $k_2^\star$.

Corollary 3.9. *Given the same setting as Theorem 3.8, we can alternatively define* $E(\epsilon, K, N, \sigma^2)$*, which is also a valid upper bound of* $\rho_{critical}$*, as*

$$E^2(\epsilon, K, N, \sigma^2) := \sigma^2 \max \left\{ \frac{\sqrt{\min\{k_1^\star, k_2^\star\}}}{N}, \epsilon^2 \ln\left(\frac{1}{\epsilon}\right), \sqrt{\frac{\epsilon^2 \ln\left(\frac{1}{\epsilon}\right) \min\{k_1^\star, k_2^\star\}}{N}} \right\} \qquad (3.5)$$

Proof. By the definition of $k_1^\star$, we know $D_{k_1^\star}(K) \leq \frac{\sqrt{k_1^\star+1}}{N}\sigma^2$. Therefore, if we set $k = k_1^\star$ in (3.4), $D_{k_1^\star}(K)$ can be absorbed into the other term involving the maximum. The logic also holds for $k_2^\star$. Finally, since (3.4) is always a valid upper bound, we select the smaller one, which leads to (3.5). $\qquad \square$

Corollary 3.9 effectively selects the smaller of the optimal dimensions $k_1^\star$ and $k_2^\star$, and consequently determines the optimal projection on which ϕ_e is based. We show in Section 3.3 that the upper bound (3.5) matches the theoretical lower bounds up to a logarithmic factor of $\ln(1/\epsilon)$.

---

[1]When $K \subset \ell_2$, we allow $1 \leq k < \infty$.

16

A significant improvement implied by Theorem 3.8 is that it does not require the asymptotic condition $\epsilon \to 0$, which is an essential assumption in many previous works. However, the test ϕ_e is limited by its reliance on an exhaustive search over all subsets of $\mathbf{X}$, resulting in exponential computational complexity with respect to N. This prohibitive cost renders the test computationally infeasible for large sample sizes. In the next section, motivated by the filtering method in Canonne et al. (2023), we propose a polynomial-time algorithm designed to handle the prior constraint K in the presence of the adversary $\mathcal{C}$.

3.2.2 Polynomial-Time Algorithm

In this section, we develop a polynomial-time algorithm[1] to detect whether $\mu = \mathbf{0}$ under a slightly stronger condition than the bound (3.4). In contrast to the theoretical algorithm which requires an exhaustive search, the proposed algorithm leverages weights to adaptively filter potentially corrupted samples. Specifically, this filtering strategy and the algorithm are built and developed upon the framework of ω-regularity, established in prior works (e.g., Canonne et al. (2023), Dong et al. (2019)).

Definition 3.10 (ω-regularity). *Given a weight vector $\omega = (\omega_1, \ldots, \omega_N)^\top$, $\mathbf{Y}$ is said to be $(\epsilon, \beta_1, \beta_2)$-regular if for all subsets $S \subset [N]$ with $|S| \le \epsilon N$, we have the following properties:*

(i), $\quad \left| \sum_{i \in S} \|Y_i\|_2^2 - |S| k^\star \right| \le c\beta_1,$

(ii), $\quad \left| \left\| \sum_{i \in S} \sqrt{\omega_i} Y_i \right\|_2^2 - \|\omega_S\|_1 k^\star \right| \le c\beta_2,$ *and*

(iii), $\quad \left| \left\langle \sum_{i \in S} \sqrt{\omega_i} Y_i, \sum_{j \in [N]} \sqrt{\omega_j} Y_j \right\rangle - \|\omega_S\|_1 k^\star \right| \le c\sqrt{N}\beta_1,$

where we recall that $\omega_S \in \mathbb{R}^N$ means the restriction of ω on the set S.

The intuition behind the polynomial-time algorithm parallels that of the theoretical construction, although establishing this connection relies on more advanced random matrix theory concerning (sub-)Gaussian random variables (see Appendix B for technical preliminaries). We first establish that the authentic samples satisfy the ω-regularity condition for specified parameters β_1, β_2 with high probability (setting the initial weights $\omega = \mathbf{1}_{[N]}$). Our analysis demonstrates that, under sufficient conditions on the signal strength $\rho = \|\mu\|_2$, the re-weighted samples effectively mimic the behavior of the unobserved authentic data $\widetilde{\mathbf{X}}$. Consequently, we are able to recover the test result corresponding to $\widetilde{\mathbf{X}}$. Such conditions are formalized in the following theorem.

Theorem 3.11 (Polynomial upper bound). *Assuming the same conditions as in Theorem 3.8, for any $1 \le k \le d^2$, as long as the following condition is satisfied, there exists a test ϕ_p to distinguish between H_0 and H_1 with type-I and type-II errors both uniformly less than α.*

$$\rho^2 \gtrsim \underbrace{D_k^2(K) + \sigma^2 \max\left\{ \frac{\epsilon \ln\left(\frac{N}{\alpha}\right)}{\sqrt{N}}, \epsilon^2 \ln\left(\frac{N}{\alpha}\right), \sqrt{\frac{\epsilon^2 k \ln\left(\frac{N}{\alpha}\right)}{N}}, \frac{\sqrt{k} \ln\left(\frac{1}{\alpha}\right)}{N} \right\}}_{:= P_{raw}^2(\epsilon, K, N, \sigma^2, k)}. \tag{3.6}$$

---

[1]By polynomial-time", we assume that computing the optimal projections (determining the optimal dimensions $k_1^\star$ and $k_2^\star$) can be performed in polynomial time, as is the case for any ellipsoid or hyperrectangle.

[2]When $K \subset \ell_2$, we allow $1 \le k < \infty$.

In other words, we have

$$\rho_{critical} \lesssim \min_k P_{raw}(\epsilon, K, N, \sigma^2, k)$$

In addition, the time and space complexity of the algorithm is at most polynomial in terms of the parameters $N, k, \frac{1}{\epsilon}, \frac{1}{\alpha}$.

Similar to Theorem 3.8, the following corollary characterizes the upper bound (3.6) with $k_1^\star$ and $k_2^\star$. The proof is omitted since it follows from Corollary 3.9 with minor modifications.

Corollary 3.12. *Given the same setting as Theorem 3.11, we can alternatively define $P(\epsilon, K, N, \sigma^2)$, which is also a valid upper bound of $\rho_{critical}$, as*

$$P^2(\epsilon, K, N, \sigma^2) = \sigma^2 \max \left\{ \frac{\epsilon \ln\left(\frac{N}{\alpha}\right)}{\sqrt{N}}, \epsilon^2 \ln\left(\frac{N}{\alpha}\right), \sqrt{\frac{\epsilon^2 \min\{k_1^\star, k_2^\star\} \ln\left(\frac{N}{\alpha}\right)}{N}}, \frac{\sqrt{\min\{k_1^\star, k_2^\star\}} \ln\left(\frac{1}{\alpha}\right)}{N} \right\}. \tag{3.7}$$

Analogous to the upper bound (3.5), (3.7) also nearly matches the theoretical lower bounds with only the exception of logarithmic terms. However, one can easily check that the upper bound (3.5) is generally tighter than (3.7), i.e., $E(\epsilon, K, N, \sigma^2) \leq P(\epsilon, K, N, \sigma^2)$.

We acknowledge that the proof of Theorem 3.11 and the associated algorithm build upon the framework established in Canonne et al. (2023). However, the results in Canonne et al. (2023) are restricted to the regime where $\|\mu\|_2 \lesssim 1$ in the unconstrained setting (i.e., $K = \mathbb{R}^d$). Consequently, we have significantly extended the algorithm to accommodate arbitrary signal magnitudes and to incorporate the prior constraint K via the optimal projection $P^\star$. Furthermore, our analysis refines the arguments of Canonne et al. (2023) to address certain technical issues. We kindly refer readers to Appendix C.6 for the complete proof of the properties regarding the ω-regularity and Theorem 3.11 and Appendix A for the pseudo code for the polynomial-time algorithm.

3.3 Main Results

In this section, we integrate the preceding theorems and bounds and summarize the main results. The lower bound is synthesized from Theorems 3.2, 3.5, 3.6. Although the bounds are derived through different approaches, an interesting phase transition phenomenon emerges, determined by the ranges of ϵ over which each term dominates others.

Theorem 3.13 (Main lower bound). *Given the mean testing problem (1.2) in the Gaussian sequence model, with potential contamination from a strong adversary $\mathcal{C}$ and a prior constraint on the mean μ induced by a QCO set K, the following condition is necessary to ensure the existence of a valid test whose type-I and type-II errors are both uniformly below a prescribed constant α*

$$\rho^2 \gtrsim \sigma^2 \max \left\{ \frac{\sqrt{k_1^\star}}{N}, \epsilon \sqrt{\frac{k_2^\star}{N}}, \epsilon^2 \right\}, \tag{3.8}$$

where N is the sample size, σ^2 is the variance of the noise, ϵ is the fraction of contamination, and $k_1^\star, k_2^\star$ are the optimal dimensions exclusively determined by N, σ, ϵ and the set K as in 2.3 and 2.4.

In the proof of Corollary 3.15 below, we show that the first term, which is independent with ϵ[1], dominates the other two if and only if $\epsilon \leq \frac{1}{\sqrt{N}}$. Similarly, the last term, which is independent with N, is the dominant if and only if $\epsilon \gtrsim \sqrt{\frac{k_2^\star}{N}}$. Therefore, the set $\left\{ \epsilon \,\middle|\, 0 \leq \epsilon < c_0, \epsilon \in \left[\frac{1}{\sqrt{N}}, \sqrt{\frac{k_2^\star}{N}} \right] \right\}$[2] serves as a threshold belt. When ϵ lies below this belt, it is too small for the adversary $\mathcal{C}$ to exert its influence, and the optimal minimax rate becomes essentially independent of ϵ. In other words, it is free to contaminate the dataset in this region. In contrast, if ϵ lies above the belt (for example, $\epsilon \asymp N^{-\frac{1}{4}}$), then the power of $\mathcal{C}$ can be fully utilized, to the extent that increasing the sample size N does not improve the problem's difficulty. When ϵ falls within the belt, the minimax rate is the result of the delicate balance between the adversary $\mathcal{C}$ and the testing procedures ϕ_e, ϕ_p. Based on this argument, (3.8) can be equivalently written as

$$
\rho \gtrsim
\begin{cases}
\frac{\sqrt{k_1^\star}}{N}\sigma^2, & \left(0 \leq \epsilon \lesssim \frac{1}{\sqrt{N}}\right) & \text{(Free contamination phase)}, \\[2mm]
\epsilon\sqrt{\frac{k_2^\star}{N}}\sigma^2, & \left(\frac{1}{\sqrt{N}} \lesssim \epsilon \lesssim \sqrt{\frac{k_2^\star}{N}}\right) & \text{(Balance phase)}, \\[2mm]
\epsilon^2\sigma^2, & \left(\epsilon \gtrsim \sqrt{\frac{k_2^\star}{N}}\right), & \text{(Overly contaminated phase)}.
\end{cases}
\tag{3.9}
$$

Informally, a message implied by (3.9) is that after the "safe threshold" $\frac{1}{\sqrt{N}}$, every single contaminated sample in the strong ϵ-contamination model invalidates the statistical testing use of $\sqrt{\frac{N}{k_2^\star}}$ uncontaminated samples on average, until the overly contaminated phase.

For the upper bound, Theorem 3.14 below is a natural downstream conclusion of Theorem 3.8 and 3.11.

Theorem 3.14 (Main upper bound). *Given the same setting as in Theorem 3.13, if $\rho \gtrsim E(\epsilon, K, N, \sigma^2)$, there exists a test ϕ_e achieving type-I and type-II errors uniformly less than α. Furthermore, if $\rho \gtrsim P(\epsilon, K, N, \sigma^2)$ and the determination of $k_1^\star$ and $k_2^\star$ can be completed within polynomial-time of $\left(N, d, \frac{1}{\epsilon}\right)$, there exists a test ϕ_p that is able to finish the testing process within polynomial-time of $\left(N, k_1^\star, k_2^\star, \frac{1}{\epsilon}, \frac{1}{\alpha}\right)$ while still achieving type-I and type-II errors uniformly less than α. $E(\epsilon, K, N, \sigma^2)$ and $P(\epsilon, K, N, \sigma^2)$ are defined in (3.5) and (3.7) respectively.*

We further establish that the main lower bound in Theorem 3.13 is minimax optimal and there is only minor difference between the conditions required by the two algorithms in Theorem 3.14. This is summarized in the following corollary.

Corollary 3.15 (Nearly matching between the lower bounds and the upper bounds). *The upper bound $E(\epsilon, K, N, \sigma^2)$ nearly matches the theoretical lower bound (3.8) with only an exception of the logarithmic factor $\ln\left(\frac{1}{\epsilon}\right)$. The upper bound $P(\epsilon, K, N, \sigma^2)$ of ϕ_p nearly matches $E(\epsilon, K, N, \sigma^2)$ with only exceptions of the logarithmic factors $\ln N$, $\ln \alpha$, and $\ln\left(\frac{1}{\epsilon}\right)$.*

For Corollary 3.15, it involves some refined analysis with the main lower bound (3.8) and $E(\epsilon, K, N, \sigma^2)$. We kindly refer readers to Appendix C.7 for the detailed arguments.

---

[1]We should recall that $k_1^\star$ is independent with ϵ

[2]We express in this form because $k_2^\star$ also depends on ϵ.

4 Extension to ℓ_p Norm Separation for $1 \le p < 2$

As an alternative, with the same assumptions on the samples $\widetilde{\mathbf{X}}$ and the adversary $\mathcal{C}$, we may consider the following ℓ_p robust testing problem.

$$
\begin{aligned}
H_0 &: \mu = 0, \\
H_1 &: \|\mu\|_p \ge \rho, \mu \in K.
\end{aligned}
\tag{4.1}
$$

The classic results for the testing problem (1.2) can be naturally generalized to (4.1) when $1 \le p < 2$ via the Hölder's inequality (see Lemma B.17) by Ingster and Suslina (2003). While Ingster's approach cannot be directly transferred to our settings, we establish similar generalized results with the p-convex condition and accordingly adjusted optimal dimensions and projections. Note that in this section, we shall substitute the ℓ_2 norm with the ℓ_p norm for the Kolmogorov k-width. Furthermore, to simplify the analysis, we only consider the discretized version (see the remark after Definition 2.2).

Definition 4.1 (Discretized Kolmogorov k-Width). *Let $\mathcal{X}$ be a Banach space equipped with the norm $\|\cdot\|$, and $K \subset \mathcal{X}$ is a subset. The discretized Kolmogorov k-width is defined as*

$$
\widetilde{D}_k(K) = \inf_{P \in \widetilde{\mathcal{P}}_k} \sup_{\theta \in K} \|\theta - P\theta\|,
\tag{4.2}
$$

where $\widetilde{\mathcal{P}}_k$ is the set of all projection operators that align with the axes and project a vector onto some subspace of $\mathcal{X}$ with intrinsic dimension k.

Though discretized, it inherits most of important properties from the original definition. When the norm is $\|\cdot\|_p$, let us denote the corresponding Kolmogorov k-width for K as $\widetilde{D}_{p,k}(K)$.

Definition 4.2 (p-convex set). *For $p \ge 1$, a set $K \subset \mathcal{X}$ is p-convex if*
(1). K is convex;
(2). K^p is convex, where K^p is defined as

$$
K^p := \left\{ (|\theta_1|^p, \ldots, |\theta_d|^p)^\top \mid (\theta_1, \ldots, \theta_d)^\top \in K \right\}.
$$

It is not difficult to verify that the whole space, hyperrectangles and ellipsoids are still all p-convex sets for $1 \le p < 2$.

Analogous to previous analysis with the ℓ_2 norm, the new lower bound relies on the following revised version of Lemma 3.1. It is not difficult to check that the proof is not essentially different.

Lemma 4.3. *Let $K \subset \mathcal{X}$ be a p-convex orthosymmetric set, and suppose that $\widetilde{D}_{p,k-1}(K) > c\sigma$. Then there exists a vector $\theta \in K$ such that $\|\theta\|_p = c\sigma$ and $\|\theta\|_\infty \le \frac{c}{k^{1/p}}\sigma$.*

We also require the following new optimal dimensions and projections for ℓ_p robust testing problem.

Definition 4.4 (ℓ_p optimal dimensions and projections). *Given the noise scale σ, the number of samples N, the prior knowledge K and the corruption rate ϵ, define the following optimal dimensions*

$$
k_{p,1}^\star := \max \left\{ j \mid 0 \le j \le d, \widetilde{D}_{p,j}(K) > \frac{j^{\frac{1}{p} - \frac{1}{4}}}{\sqrt{N}} \sigma \right\},
$$

$$
k_{p,2}^\star := \max \left\{ j \mid 0 \le j \le d, \widetilde{D}_{p,j}(K) > \frac{\sqrt{\epsilon} j^{\frac{1}{p} - \frac{1}{4}}}{N^{\frac{1}{4}}} \sigma \right\},
$$

$$
k_{p,3}^\star := \max \left\{ j \mid 0 \le j \le d, \widetilde{D}_{p,j}(K) > \epsilon \sigma j^{\frac{1}{p} - \frac{1}{2}} \right\}.
$$

Analogous to Definition 2.3 and 2.4, the corresponding optimal projections are defined accordingly and denoted as $P^\star_{p,1}, P^\star_{p,2}, P^\star_{p,3}$.

Given the modified lemma and definitions above, we present the lower bounds for the ℓ_p robust testing problem (4.1).

Theorem 4.5 (ℓ_p main lower bound). *Given the mean testing problem (4.1) in the Gaussian sequence model, with potential contamination from a strong adversary $\mathcal{C}$ and a prior constraint on the mean μ induced by a p-convex orthosymmetric set K, the following condition is necessary to ensure the existence of a valid test whose type-I and type-II errors are both uniformly below a prescribed constant α*

$$\|\theta\|_p \gtrsim \sigma \max \left\{ \frac{(k^\star_{p,1})^{\frac{1}{p}-\frac{1}{4}}}{\sqrt{N}}, \frac{\sqrt{\epsilon}(k^\star_{p,2})^{\frac{1}{p}-\frac{1}{4}}}{N^{\frac{1}{4}}}, \epsilon(k^\star_{p,3})^{\frac{1}{p}-\frac{1}{2}} \right\}. \tag{4.3}$$

where N is the sample size, σ^2 is the variance of the noise, ϵ is the fraction of contamination, and $k^\star_{p,1}, k^\star_{p,2}, k^\star_{p,3}$ are the optimal dimensions exclusively determined by N, σ, ϵ and the set K as in 4.4.

Undoubtedly, the lower bound (4.3) recovers (3.8) as expected when $p = 2$. A significant difference from the ℓ_2 norm separation is that we have another optimal dimension regarding the last term in (4.3). The intrinsic reason is that when covariance matrices are known and identical, the KL divergence between Gaussian distributions — an essential quantity we leverage in the proof of Theorem 3.6, depends on the ℓ_2 norm between the means. When considering the ℓ_p norm separation, we require $k^\star_{p,3}$ to adapt to this discrepancy. We defer the proof of Theorem 4.5 to Appendix C.8. In ℓ_2 norm separation, the upper bound (3.5) matches the lower bound (3.8) except for logarithmic factors. We identify similar results in ℓ_p norm separation, which is summarized as Theorem 4.6 below.

Theorem 4.6 (ℓ_p main upper bound). *If $\rho \gtrsim \sigma \max \left\{ \frac{(k^\star_{p,1})^{\frac{1}{p}-\frac{1}{4}}}{\sqrt{N}}, \frac{\sqrt{\epsilon}(k^\star_{p,2})^{\frac{1}{p}-\frac{1}{4}}}{N^{\frac{1}{4}}}, \epsilon(k^\star_{p,3})^{\frac{1}{p}-\frac{1}{2}} \right\}$, then there exists a test to distinguish between H_0 and H_1 with uniformly small type-I and type-II errors. Moreover, the time complexity of the algorithm is polynomial-time in $\left(N, d, \frac{1}{\epsilon}, \frac{1}{\alpha}\right)$. Note that in the expression above, we omit the polylogarithmic factors in $\left(N, k^\star_{p,1}, k^\star_{p,2}, k^\star_{p,3}, \frac{1}{\alpha}, \frac{1}{\epsilon}\right)$.*

At first examination, Theorem 4.6 appears to be a direct and natural corollary of Theorem 3.14. However, a closer inspection reveals subtleties in its derivation. A crucial observation is that, though one can select the minimum $k^\star_p := \min \left\{ k^\star_{p,1} k^\star_{p,2}, k^\star_{p,3} \right\}$ and its optimal projection $P^\star_p$, and is guaranteed that there exist three vectors $\theta_{p,1}, \theta_{p,2}, \theta_{p,3}$ such that $\left\| P^\star_p \theta_{p,1} \right\|^2_2 \gtrsim \frac{\sqrt{k^\star}}{N} \sigma^2, \left\| P^\star_p \theta_{p,2} \right\|^2_2 \gtrsim \epsilon \sqrt{\frac{k^\star_p}{N}} \sigma^2, \left\| P^\star_p \theta_{p,3} \right\|^2_2 \gtrsim \epsilon^2 \sigma^2$, $\theta_{p,1}, \theta_{p,2}, \theta_{p,3}$ are not necessary to be identical. In Theorem 3.8 or Theorem 3.11, however, we require a single vector satisfying all the conditions simultaneously. In fact, similar to the ℓ_2 norm case, we do need to select the minimal optimal dimension. To overcome the difficulty specified above, we again need to exploit the properties of the p-convex orthosymmetric set and the Kolmogorov widths. The proof of Theorem 4.6 is provided in Appendix C.9.

5 Experiments

In this section, we present the results of simulation to verify the correctness of the theorems above and algorithms in Appendix A on the application side. We generate artificial data, design different

strategies for the adversary to contaminate the data, and run the (polynomial-time) algorithm to check its effectiveness. The pseudo code is attached in Appendix A.

5.1 Data Generation

We generate artificial data using the function `numpy.random.multivariate` in the `Numpy` package in `Python`. To fully verify the correctness of the algorithms, we consider both the classical and high-dimensional regimes separately, and explore a variety of parameter configurations. Specifically, in the classical scenario, we let the parameter tuple (N, d, ϵ) iterate over all possible combinations of $\{200, 2000\} \times \{5, 10, 20, 50\} \times \{0.01, 0.015, 0.02, 0.025\}$; in the high-dimensional case, we let the tuple iterate over all possible combinations of $\{100, 500\} \times \{100, 500\} \times \{0.01, 0.015, 0.02, 0.025\}$. We set $\sigma = 1$ in all generation procedures.

5.2 Choice of Constraints

For the choice of constraints, i.e., K, we select the common ellipsoidal constraint:

$$\mu \in K_e := \left\{ v \mid v \in \mathbb{R}^d, \sum_{i=1}^{n} \frac{v_i^2}{a_i} \leq 1, a_i = \sqrt{d} \cdot i^{-2}, 1 \leq i \leq d \right\}. \tag{5.1}$$

Note that our choice of the axes of the ellipse ensures that the maximal ℓ_2 norm of the vectors in K_e is greater than the theoretical lower bound (3.8). For a given positive value $c \leq \sqrt{a_1}$ as the ℓ_2 norm of μ under H_1, the following approach uniquely determines the specific vector μ_c in K_e such that $\|\mu_c\|_2 = c$.

$$\mu_c = \begin{cases} (\sqrt{\frac{c^2/a_{k^\star+1}-1}{1/a_{k^\star+1}-1/a_1}}, \underbrace{0,\ldots,0}_{k^\star-1 \text{ zeros}}, \sqrt{\frac{1-c^2/a_1}{1/a_{k^\star+1}-1/a_1}}, 0,\ldots,0), & \text{if } c^2 > a_{k^\star+1} \text{ and } k^\star < d, \\ \underbrace{(0,\ldots,0}_{k^\star \text{ zeros}}, c, 0,\ldots,0), & \text{if } c^2 \leq a_{k^\star+1} \text{ and } k^\star < d, \\ (c,0,\ldots,0), & \text{if } k^\star = d, \end{cases} \tag{5.2}$$

where $k^\star = \min\{k_1^\star, k_2^\star\}$. We verify in Appendix C.10 that the approach (5.2) ensures that the resulting vector μ_c achieves the minimal value of $\|P^\star \mu\|_2$ among the set $\{v \mid v \in \mathbb{R}^d, v \in K_e, \|v\|_2 = c\}$, where $P^\star$ is the optimal projection operator corresponding to $k^\star$. We believe that this property places μ_c in a near-worst-case position for the algorithm, and thus provides an appropriate setting to test its robustness.

5.3 Configurations of Adversary

For the adversary $\mathcal{C}$ in the strong ϵ-contamination model, we design the following two strategies. Note that in the strategies, μ_0 is an unknown fake mean vector determined by $\mathcal{C}$ that can depend on the constraint K.

In strategies 1 and 2, the adversary $\mathcal{C}$ and the resulting corruption are largely determined on the `arr` and `dist` arrays, which compute the distance between the original data and the origin, and between the original data and an imagined mean μ_0, respectively. Both strategies were implemented in our experiments, and the results indicate that the choice has minimal impact on the outcomes.

Algorithm 1: strategy 1 of the adversary $\mathcal{C}$
---
1 **import** $\widetilde{\mathbf{X}}, N, d, \sigma = 1$

2 Compute $\mathtt{arr} \in \mathbb{R}^N$, $\mathtt{arr}_i = \left\| \widetilde{X}_i \right\|_2, 1 \le i \le N$

3 **if** H_1 *is true* **then**

4 $\quad$ $C =$ indices of the largest ϵN entries in $\mathtt{arr}$

5 $\quad$ $\mathbf{X}_C = \epsilon N$ i.i.d. fake data from $\mathcal{N}\left(\mathbf{0}, \mathbf{I}_d\right)$

6 $\quad$ $\widetilde{\mathbf{X}}_C = \mathbf{X}_C$

7 **else**

8 $\quad$ **import** μ_0

9 $\quad$ $C =$ indices of the smallest ϵN entries in $\mathtt{arr}$

10 $\quad$ $\mathbf{X}_C = \epsilon N$ i.i.d. fake data from $\mathcal{N}\left(\mu_0, \mathbf{I}_d\right)$

11 $\quad$ $\widetilde{\mathbf{X}}_C = \mathbf{X}_C$

12 **return** $\mathbf{X} = \mathcal{C}(\widetilde{\mathbf{X}})$

Algorithm 2: strategy 2 of the adversary $\mathcal{C}$
---
1 **import** $\widetilde{\mathbf{X}}, N, d, \sigma = 1$

2 **if** H_1 *is true* **then**

3 $\quad$ Compute $\mathtt{dist} \in \mathbb{R}^N$, $\mathtt{dist}_i = \left\| \widetilde{X}_i \right\|_2, 1 \le i \le N$

4 $\quad$ $C =$ indices of the largest ϵN entries in $\mathtt{arr}$

5 $\quad$ $\mathbf{X}_C = \epsilon N$ i.i.d. fake data from $\mathcal{N}\left(\mathbf{0}, \mathbf{I}_d\right)$

6 $\quad$ $\widetilde{\mathbf{X}}_C = \mathbf{X}_C$

7 **else**

8 $\quad$ Compute μ_0 according to the constraint type

9 $\quad$ Compute $\mathtt{dist} \in \mathbb{R}^N$, $\mathtt{arr}_i = \left\| \widetilde{X}_i - \mu_0 \right\|_2$

10 $\quad$ $C =$ indices of the largest ϵN entries in $\mathtt{arr}$

11 $\quad$ $\mathbf{X}_C = \epsilon N$ i.i.d. fake data from $\mathcal{N}\left(\mu_0, \mathbf{I}_d\right)$

12 $\quad$ $\widetilde{\mathbf{X}}_C = \mathbf{X}_C$

13 **return** $\mathbf{X} = \mathcal{C}(\widetilde{\mathbf{X}})$

5.4 Experimental Results

Experimental results for $N = 200$ in the classical case and $N = 100$ in the high-dimensional case are presented in the following figures. Results for other parameter configurations are similar and are deferred to Appendix A. In our experiments, we maintained the Type-I error near zero while identifying the empirical $\widetilde{\rho}_{\text{critical}}$ under H_1 as the critical value of ρ when the corresponding type-II error exactly crosses $\beta = 0.05$.

We observed several notable phenomena from the simulation results. First, the algorithm performs reasonably well across all parameter configurations, even for the moderate sample sizes: it successfully rejects H_0 when $\|\mu\|_2$ is sufficiently large and fails to reject when the signal strength is too weak, in both the classical and the high-dimensional regimes. Second, focusing on ρ_{critical} — the critical norm of μ that yields a type-II error of $\alpha = 0.05$ — we find that ρ_{critical} decreases as N increases, and increases as d grows. This behavior is consistent with theoretical intuition: larger sample sizes improve detectability, whereas higher dimensionality makes the testing problem more challenging. However, a more refined observation is that the influence of d diminishes once the dimension becomes sufficiently large (e.g., around $d = 50$ in our experiments). The reason behind this phenomenon is that, although d increases to relatively large values, the optimal projection dimension $k^\star$ becomes increasingly insensitive to d under our setting of $N = 1, \sigma = 1$, and the ellipsoidal constraint. A direct calculation shows that when $N = 2000$, the optimal dimensions corresponding to full dimensions $[5, 10, 20, 50]$ are $[5, 9, 10, 12]$, respectively. This demonstrates that the gaps between the optimal dimensions can be substantially smaller than those between the actual dimensions, which leads to the phenomenon. Finally, the plots indicate that as N increases, the performance of the algorithm becomes less sensitive to variations in ϵ. Though not fully verified, we believe that this is attributed to the fact that for larger N, the eigensubspace structure of $\mathbf{X}\mathbf{X}^\top$ (or $\mathbf{X}^\top\mathbf{X}$) becomes more stable with respect to the corrupted observations.

We also remark that we computed the ratio between $\widetilde{\rho}_{\text{critical}}$ and the theoretical lower bound (3.8) in an attempt to assess the tightness of the bound. Although the preliminary results appeared encouraging, we chose not to report them or draw conclusions. The primary reason is that both (3.8) and its corresponding upper bound are derived under idealized minimax conditions that depend on the precise forms of K, μ, and the adversary $\mathcal{C}$. In practical settings, however, these worst case quantities (μ and $\mathcal{C}$) are inherently unknown to us. Consequently, despite our simulations incorporating designed mean structures and adversarial strategies, they cannot guarantee that the conditions required for demonstrating optimality are satisfied. Identifying these conditions poses a fundamental challenge, and we leave this direction to future work.

For the hyperrectangle-type constraint, we conducted the experiments with the same settings and procedures. The results exhibited behavior similar to those obtained under the ellipsoidal constraint. Therefore, we present only the results under the ellipsoidal case.

6 Discussion

6.1 Rotated QCO Sets

For completeness, we note that if the set K is an orthogonal rotation of a QCO set — i.e., $K = UK'$ for some orthogonal matrix U — then one may transform the data by considering $U^\top X$ and subsequently solve the corresponding problem on the QCO set K'. This follows from the fact that $U^\top \widetilde{X}_i = U^\top \mu + U^\top \xi_i$, where $U^\top \xi_i \sim N(0, \sigma^2 \mathbf{I})$ and $U^\top \mu \in K'$. Moreover, the computational cost
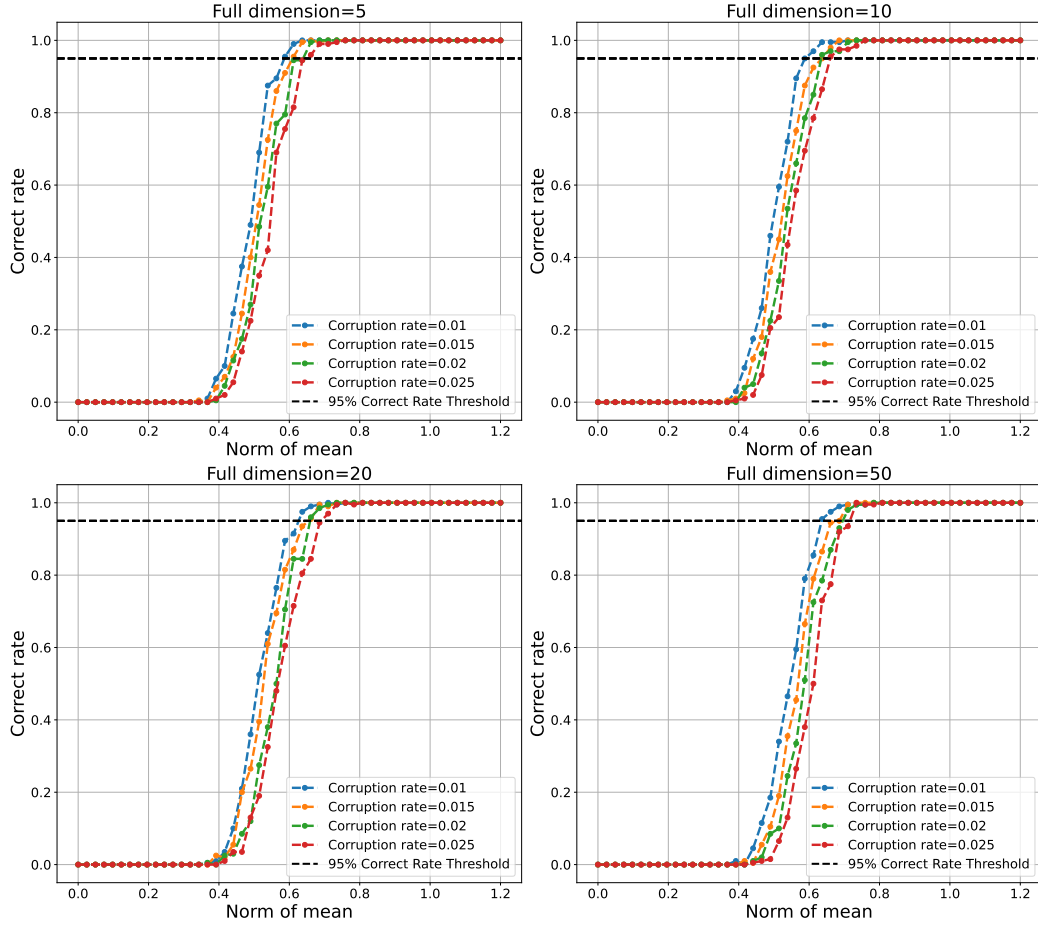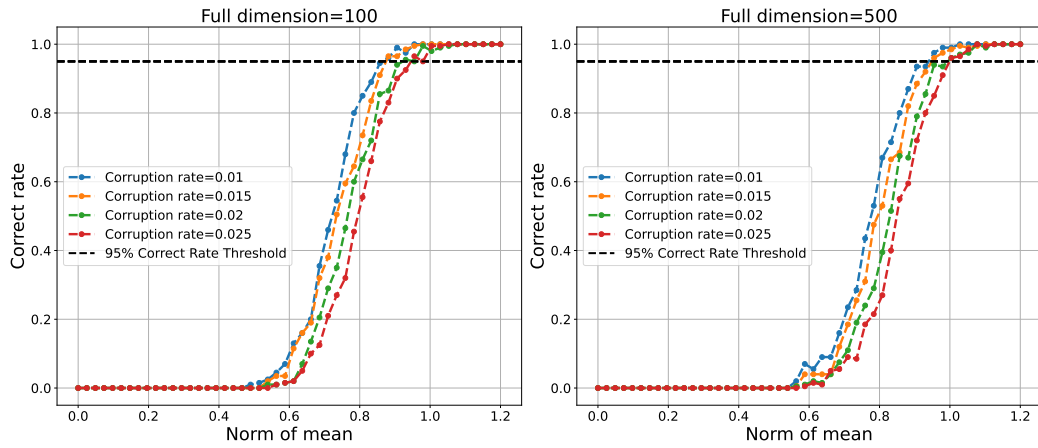
Figure 3: **Top**: Rejection rate (or Empirical power) versus the ℓ_2 norm of the mean under H_1 in the classic setting ($N = 200$). **Bottom**: Rejection rate versus the ℓ_2 norm of the mean in the high-dimensional setting ($N = 100$).

of determining the optimal dimensions for K and K' is essentially the same in most cases.

6.2 Equivalence in Optimal Dimensions

An alternative but equivalent way to define the optimal dimension is to simply select $k_{\mathrm{alt}}^{\star}$ satisfying $D_{k_{\mathrm{alt}}^{\star}-1}(K) > \left(\frac{k_{\mathrm{alt}}^{\star}}{N}\right)^{1/4} \max\left\{\sqrt{\epsilon}, N^{-\frac{1}{4}}\right\} \sigma$ and $D_{k_{\mathrm{alt}}^{\star}}(K) \leq \left(\frac{k_{\mathrm{alt}}^{\star}+1}{N}\right)^{1/4} \max\left\{\sqrt{\epsilon}, N^{-\frac{1}{4}}\right\} \sigma$. It also has the equivalent definitions as the discussion following Definition 2.3 and 2.4. Once $k_{\mathrm{alt}}^{\star}$ is determined, we simply project the data using $P_{\mathrm{alt}}^{\star}$, the optimal projection associated with $k_{\mathrm{alt}}^{\star}$. The remainder of the testing procedure is identical to that in Algorithms 3 and 8. By this new definition, the lower bound will be adjusted accordingly as

$$\rho^2 \gtrsim \sigma^2 \max\left\{\sqrt{\frac{k_{\mathrm{alt}}^{\star}}{N}} \max\left\{\epsilon, \frac{1}{\sqrt{N}}\right\}, \epsilon^2\right\}. \tag{6.1}$$

Similarly, the upper bounds $E_{\mathrm{alt}}^2(\epsilon, K, N, \sigma^2)$ and $P_{\mathrm{alt}}^2(\epsilon, K, N, \sigma^2)$ are obtained by substituting all $k_1^{\star}$ and $k_2^{\star}$ in $E^2(\epsilon, K, N, \sigma^2)$ and $P^2(\epsilon, K, N, \sigma^2)$ by $k_{\mathrm{alt}}^{\star}$. With the proof in Appendix C.7, one can verify that $k_{\mathrm{alt}}^{\star}, P_{\mathrm{alt}}^{\star}$ and its corresponding testing process are equivalent to that of $k_1^{\star}$, $k_2^{\star}$, $P_1^{\star}$, and $P_2^{\star}$. The new lower bound (6.1) and the upper bounds $E_{\mathrm{alt}}^2(\epsilon, K, N, \sigma^2)$ and $P^2(\epsilon, K, N, \sigma^2)$ do match the previous lower bound (3.8) and the upper bounds (3.5), (3.7) as well.

6.3 Testing-Estimation Relationship

The results developed in Section 3 should be compared and contrasted with the existing results for the corresponding estimation problem. Specifically, in the estimation setting — where the same distribution assumptions and corruption model are imposed, but the goal is to estimate the mean vector μ — the minimax rates are established by Donoho et al. (1990) and Neykov (2022). Notably, the Kolmogorov widths and the associated optimal dimensions again play a central role in these results.

Definition 6.1 (Optimal dimension and projection in estimation)**.** *Given the same assumptions on the distribution, the constraint K, and the corruption model as 2.3, 2.4, and 2.8, $k_e^{\star}$ is defined as the dimension that satisfies $D_{k_e^{\star}-1}(K) > \sqrt{\frac{k_e^{\star}}{N}}\sigma$ but $D_{k_e^{\star}} \leq \sqrt{\frac{k_e^{\star}+1}{N}}\sigma$. It can be equivalently defined as*

$$k_e^{\star} := \max_k \left\{ k \mid k \geq 0, D_{k-1}(K) \geq \sqrt{\frac{k}{N}}\sigma \right\}. \tag{6.2}$$

Theorem 6.2 (Robust Estimation Rate, Prasadan and Neykov (2025))**.** *Given the same assumptions on the distribution, the constraint K, and the corruption model as (1.2), the minimax rate for the estimation problem is given by*

$$\inf_{\widehat{\nu}} \sup_{\mu \in K} \mathbb{E}_\mu \|\widehat{\nu}(\mathbf{X}) - \mu\|_2^2 \asymp \sigma^2 \max\left\{\frac{k_e^{\star}}{N}, \epsilon^2\right\}, \tag{6.3}$$

where inf is taken with respect to all measurable functions $\widehat{\nu}$ of the data.

When $K = \mathbb{R}^d$ and there is no corruption, we have $k_e^{\star} = d$ and $\epsilon = 0$, in which case the rate above recovers the classic result that $\inf_{\widehat{\nu}} \sup_{\mu} \mathbb{E}_\mu \left\|\widehat{\nu}(\widetilde{\mathbf{X}}) - \mu\right\|_2^2 \asymp \frac{d}{N}\sigma^2$. It is also well known that, in this

setting, the estimation problem is generally more difficult than the testing problem. The reason is simple and intuitive — one is able to derive a valid test with small error probabilities whenever $\|\mu\|_2^2$ is asymptotically larger than the minimax lower bound for the estimation task. This relationship, unsurprisingly, is inherited by our more general problem.

Lemma 6.3 (Testing-estimation relationship). *Given the same assumptions on the distribution, the constraint K, and the corruption model above, we have*

$$\sigma^2 \max\left\{\frac{\sqrt{k_1^\star}}{N}, \epsilon\sqrt{\frac{k_2^\star}{N}}, \epsilon^2\right\} \lesssim \sigma^2 \max\left\{\frac{k_e^\star}{N}, \epsilon^2\right\} \tag{6.4}$$

We refer readers to Appendix C.11 for a straightforward computational verification of Lemma 6.3. Following the discussion after Corollary 3.15, we can also analyze the phase transition of the estimation problem. It occurs around $\epsilon \asymp \sqrt{\frac{k_e^\star}{N}}$. If $\epsilon \lesssim \sqrt{\frac{k_e^\star}{N}}$, then ϵ is too small to influence the minimax rate of the estimation error. Otherwise (for example, $\epsilon \asymp N^{-\frac{1}{4}}$), the adversary is strong enough to the extent that N is excluded and the minimax rate only relies on ϵ and σ.

Another interesting observation from Lemma 6.3 is that although estimation is generally more challenging than testing, the gap disappears as the corruption rate increases. In fact, both quantities stabilize at $\epsilon^2\sigma^2$ when ϵ is sufficient large. Similar phenomenon has also been observed in a recent work Kania et al. (2025). See Section 6.4 for additional details.

6.4 Open Problems and Future Work

An intriguing open direction is to investigate the localized minimax testing rates for QCO sets. A preliminary step in this direction was taken by Wei and Wainwright (2020), who studied local testing rates in the setting of ellipsoids. It would be interesting to understand the extent to which their results can be generalized to QCO sets. Moreover, from an intuitive standpoint, the origin appears to be the most challenging point for testing under QCO constraints, as it may correspond to the largest local volume relative to other points. This raises a natural question: does the global minimax testing rate coincide with the localized rate at the origin, and thus with the rates established in this paper?

A fundamental assumption of this paper is that the covariance matrix is fully known (and taken to be $\mathbf{I}_d$). This, however, is rarely guaranteed in real-world applications. An important direction for future work is therefore to understand whether — and to what extent — we can still distinguish between H_0 and H_1 when $\boldsymbol{\Sigma}$ must be estimated rather than treated as prior knowledge. A recent relevant work in robust covariance matrix estimation is Li and Ye (2020), where the authors proposed an algorithm achieving a near-optimal error rate $\mathcal{O}\left(\epsilon \ln\left(\frac{1}{\epsilon}\right)\right)$ in the Mahalanobis norm under ϵ-strong contamination model while remaining computationally efficient. Nevertheless, this algorithm is designed for zero-mean distributions and unconstrained setting. It remains unknown whether reliable testing is achieveble and how the constraint K can facilitate the problem when $\boldsymbol{\Sigma}$ is unknown.

In Kania et al. (2025), the authors consider a more general problem than (1.1):

$$\begin{aligned} H_0 &: \|\mu\|_2 \le \rho_0, \\ H_1 &: \|\mu\|_2 \ge \rho_1. \end{aligned} \tag{6.5}$$

In other words, rather than being fixed at the origin, the mean vector under the null is allowed to vary within a neighborhood of the origin. An interesting corollary of their work is that as the neighborhood expands, the minimax testing rate converges to the minimax estimation rate — an effect that parallels a key phenomenon in our setting. Such match in estimation and testing rate is also observed by other work (e.g., Ndaoud (2019)). Given the framework of their problem, it is natural to ask whether one can establish the minimax rate with constrained and adversarial setting of (6.5). From the perspective of the testing-estimation relationship, an even more important question is: under what conditions does the gap between testing and estimation disappear?

Another important avenue for future research will be to study ℓ_p norm separation for $p > 2$. We believe there are fundamental changes compared to the problem discussed in this work. The first reason is that χ^2-test is no longer suitable for the testing problem when $p > 2$. Instead, χ^p-test is a useful tool in this scenario, see the results by Ingster and Suslina (2003) when there is no constraint and adversary. The second reason is that the geometric properties of the ℓ_p balls when $p > 2$ are known to be dramatically different compared to those when $1 \leq p \leq 2$. This affects the selection of the extreme vector as in Lemma 3.1. It is also unclear at the moment whether the Kolmogorov widths will still drive the minimax rates in that case.

Finally, a more ambitious open problem is to characterize the minimax testing rates over general convex sets that contain the origin. It would be particularly exciting to determine whether an optimal test exists for every such convex set K, analogous to the result of Neykov (2022), which established the existence of a universally optimal estimation procedure for arbitrary convex sets in the corresponding estimation problem.

Acknowledgments

References

Acharya, J., Sun, Z., and Zhang, H. (2018). Differentially private testing of identity and closeness of discrete distributions. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, pages 577–606.

Bishop, A. N., Del Moral, P., and Niclas, A. (2018). An introduction to wishart matrix moments. *Foundations and Trends® in Machine Learning*, 11(2):97–218.

Cai, B., Daskalakis, C., and Kamath, G. (2017). Priv'IT: Private and sample efficient identity testing. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 635–644. PMLR.

Canonne, C., Hopkins, S. B., Li, J., Liu, A., and Narayanan, S. (2023). The Full Landscape of Robust Mean Testing: Sharp Separations between Oblivious and Adaptive Contamination . In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2159–2168, Los Alamitos, CA, USA. IEEE Computer Society.

Chen, M., Gao, C., and Ren, Z. (2017). Robust covariance and scatter matrix estimation under huber's contamination model.

Cohen, A. and DeVore, R. (2015). Kolmogorov widths under holomorphic mappings. *IMA Journal of Numerical Analysis*, 36(1):1–12.

Diakonikolas, I. and Kane, D. M. (2019). Recent advances in algorithmic high-dimensional robust statistics.

Diakonikolas, I., Kane, D. M., and Pensia, A. (2022). Gaussian mean testing made simple. *arXiv preprint arXiv:2210.13706*.

Diakonikolas, I., Kane, D. M., and Stewart, A. (2017). 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.

Dong, Y., Hopkins, S. B., and Li, J. (2019). Quantum entropy scoring for fast robust mean estimation and improved outlier detection.

Donoho, D. L. and Gasko, M. (1992). Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics*, 20(4):1803 – 1827.

Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, pages 1416–1437.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, page 265–284, Berlin, Heidelberg. Springer-Verlag.

Ermakov, M. S. (1991). Minimax detection of a signal in a gaussian white noise. *Theory of Probability & Its Applications*, 35(4):667–679.

Evans, J. A., Bazilevs, Y., Babuska, I., and Hughes, T. J. (2009). n-widths, sup–infs, and optimality ratios for the k-version of the isogeometric finite element method. *Computer Methods in Applied Mechanics and Engineering*, 198(21):1726–1741. Advances in Simulation-Based Engineering Sciences – Honoring J. Tinsley Oden.

Fibbi, E., Perrotta, D., Torti, F., and Stefan Van Aelst, T. V. (2023). Co-clustering contaminated data: a robust model-based approach. *Adv Data Anal Classif*, 18:121 – 161.

Floater, M. S., Manni, C., Sande, E., and Speleers, H. (2021). Best low-rank approximations and kolmogorov n-widths. *SIAM Journal on Matrix Analysis and Applications*, 42(1):330–350.

Florian Arbes, C. G. and Urban, K. (2025). The kolmogorov n-width for linear transport: exact representation and the influence of the data. *Adv Comput Math*, 51(13).

Gaboardi, M., Lim, H., Rogers, R., and Vadhan, S. (2016). Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In Balcan, M. F. and Weinberger, K. Q., editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2111–2120, New York, New York, USA. PMLR.

Greif, C. and Urban, K. (2019). Decay of the kolmogorov n-width for wave problems. *Applied Mathematics Letters*, 96:216–222.

Hanson, D. L. and Wright, F. T. (1971). A Bound on Tail Probabilities for Quadratic Forms in Independent Random Variables. *The Annals of Mathematical Statistics*, 42(3):1079 – 1083.

Hopkins, S. B., Kamath, G., Majid, M., and Narayanan, S. (2023). Robustness implies privacy in statistical estimation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 497–506, New York, NY, USA. Association for Computing Machinery.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

Ingster, Y. and Suslina, I. (2003). *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media.

Ingster, Y. I. (1982). On the minimax nonparametric detection of signals in white gaussian noise. *Problemy Peredachi Informatsii*, 18(2):61–73.

Ingster, Y. I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist*, 2(2):85–114.

Ingster, Y. I. (1993b). Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist*, 3(2):171–189.

Ingster, Y. I. (1993c). Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist*, 4(2):249–268.

Ingster, Y. I. and Suslina, I. A. (2000). Minimax nonparametric hypothesis testing for ellipsoids and besov bodies. *ESAIM: Probability and Statistics*, 4:53–135.

Johnstone, I. M. (2019). Gaussian estimation: Sequence and wavelet models. *Unpublished manuscript*.

Kania, L., Manole, T., Wasserman, L., and Balakrishnan, S. (2025). Testing imprecise hypotheses.

Kolmogoroff, A. (1936). Über die beste annäherung von funktionen einer gegebenen funktionenklasse. *Annals of Mathematics*, 37(1):107–110.

Kubokawa, T. and Srivastava, M. S. (2008). Estimation of the precision matrix of a singular wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99(9):1906–1928.

LeCam, L. (1973). Convergence of Estimates Under Dimensionality Restrictions. *The Annals of Statistics*, 1(1):38 – 53.

Li, J. and Ye, G. (2020). Robust gaussian covariance estimation in nearly-matrix multiplication time. In *Advances in Neural Information Processing Systems*.

Liang, J. and Parks, T. (1996). Kolmogorov n-widths and wavelet representations for signal classes. *IEEE Transactions on Signal Processing*, 44(7):1693–1703.

Liu, A. and Moitra, A. (2021). Settling the robust learnability of mixtures of gaussians. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 518–531, New York, NY, USA. Association for Computing Machinery.

Liu, A. and Moitra, A. (2023). Robustly learning general mixtures of gaussians. *J. ACM*, 70(3).

Martí, J. M. (2019). Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLoS Comput Biol*, 15(4).

Ndaoud, M. (2019). Interplay of minimax estimation and minimax support recovery under sparsity. In Garivier, A. and Kale, S., editors, *Algorithmic Learning Theory, ALT 2019, 22-24 March 2019, Chicago, Illinois, USA*, volume 98 of *Proceedings of Machine Learning Research*, pages 647–668. PMLR.

Neykov, M. (2022). On the minimax rate of the gaussian sequence model under bounded convex constraints. *arXiv preprint arXiv:2201.07329*.

Papapicco, D., Demo, N., Girfoglio, M., Stabile, G., and Rozza, G. (2022). The neural network shifted-proper orthogonal decomposition: A machine learning approach for non-linear reduction of hyperbolic equations. *Computer Methods in Applied Mechanics and Engineering*, 392:114687.

Pinkus, A. (2012). *N-widths in Approximation Theory*, volume 7. Springer Science & Business Media.

Prasadan, A. and Neykov, M. (2025). Information theoretic limits of robust sub-gaussian mean estimation under star-shaped constraints.

Ramirez, M. A., Kim, S.-K., Hamadi, H. A., Damiani, E., Byon, Y.-J., Kim, T.-Y., Cho, C.-S., and Yeun, C. Y. (2022). Poisoning attacks and defenses on artificial intelligence: A survey.

Rogers, R. and Kifer, D. (2017). A New Class of Private Chi-Square Hypothesis Tests. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 991–1000. PMLR.

Rudelson, M. and Vershynin, R. (2013). Hanson-wright inequality and sub-gaussian concentration.

Suslina, I. (1996). Minimax signal detection forl q-ellipsoids withl p-balls removed. *Journal of Mathematical Sciences*, 81(1):2442–2449.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, New York, NY, 1 edition.

Vershynin, R. (2024). *High-Dimensional Probability*. 1st edition.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

Wei, Y. and Wainwright, M. J. (2020). The local geometry of testing in ellipses: Tight control via localized kolmogorov widths. *IEEE Transactions on Information Theory*, 66(8):5110–5129.

A Theoretical and Polynomial-Time Algorithms

A.1 Theoretical Algorithm

According to Theorem 3.8 and Section C.5, we first attempt to search for a consistent set T with respect to ϕ_e. $\phi_e(\mathbf{X}_T)$ then determines whether we are to accept or reject H_0.

Algorithm 3: theoretical algorithm when $\rho \gtrsim E(\epsilon, K, N, \sigma^2)$

---

1 **import** $\mathbf{X}, N, K, \sigma$

2 Compute $k_1^\star, P_1^\star, k_2^\star, P_2^\star$ from N, K, σ, ϵ as 2.3 and 2.4

3 **if** $k_1^\star \leq k_2^\star$ **then**

4 $\quad \lfloor\ k^\star = k_1^\star, P^\star = P_1^\star$

5 **else**

6 $\quad \lfloor\ k^\star = k_2^\star, P^\star = P_2^\star$

7 $\mathbf{Y} = \mathbf{X}P^\star$

8 **for** T *in* {*all subsets with cardinality greater than* $(1-\epsilon)N$} **do**

9 $\quad$ consistent_flag = True

10 $\quad$ **if** $\left\| \sum_{i \in T} Y_i \right\|_2^2 - k^\star |T| \sigma^2 \geq c |T|^2 E^2(\epsilon, K, N, \sigma)$ **then**

11 $\quad\quad \lfloor$ subset_test = True

12 $\quad$ **else**

13 $\quad\quad \lfloor$ subset_test = False

14 $\quad$ **for** T' *in* {*all subsets of* T *with cardinality greater than* $(1-2\epsilon)N$} **do**

15 $\quad\quad$ **if** $\left\| \sum_{i \in T'} Y_i \right\|_2^2 - k^\star |T'| \sigma^2 \geq c |T'|^2 E^2(\epsilon, K, N, \sigma)$ **then**

16 $\quad\quad\quad \lfloor$ subsubset_test = True

17 $\quad\quad$ **else**

18 $\quad\quad\quad \lfloor$ subsubset_test = False

19 $\quad\quad$ **if** *subset_test XOR subsubset_test* **then**

20 $\quad\quad\quad$ consistent_flag = False

21 $\quad\quad\quad$ **break**

22 $\quad$ **if** *consistent_flag* **then**

23 $\quad\quad$ **if** *subset_test* **then**

24 $\quad\quad\quad$ Reject H_0

25 $\quad\quad\quad$ **return**

26 $\quad\quad$ **else**

27 $\quad\quad\quad$ Accept H_0

28 $\quad\quad\quad$ **return**

29 Accept H_0

30 **return**

A.2 Polynomial-Time Algorithm

The polynomial-time algorithm combines the three sub-algorithms in Section C.6 together to filter the observations $\mathbf{X}$ and through the weight ω. The theoretical guarantee is provided in Section C.6.

Algorithm 4: Prefiltering.

---

1 Set $\gamma_1 = c \left[\sqrt{k^\star \ln\left(\frac{N}{\alpha}\right)} + \ln\left(\frac{N}{\alpha}\right) \right]$, $\texttt{count} = 0, \texttt{i} = 0$

2 **while** $i < N$ **do**

3 $\quad$ **if** $\left| \|Y_i\|_2^2 - k^\star \right| > \gamma_1$ **then**

4 $\quad\quad$ $\texttt{count} = \texttt{count} + 1$

5 $\quad\quad$ **if** $count > \epsilon N$ **then**

6 $\quad\quad\quad$ **return** None

7 $\quad\quad$ Delete Y_i from $\mathbf{Y}$

8 $\quad$ $\texttt{i} = \texttt{i} + 1$

9 **return** $\mathbf{Y}$

Algorithm 5: Sample filtering when $N > k^\star$.

---

1 Set γ_2 as (C.13), and $\lambda = \left\| Y^\top D(\omega) Y - N \mathbf{I}_{k^\star} \right\|_2$. ($\omega$ is initialized as $\mathbf{1}$.)

2 **while** $\lambda \geq \gamma_2$ **do**

3 $\quad$ Set v to be the unit singular vector associated with λ

4 $\quad$ Compute $\tau_i = \langle v, Y_i \rangle^2 \, \mathbf{1}_{\{\omega_i > 0\}}$ for $1 \leq i \leq N$

5 $\quad$ Set I be the smallest index such that $\sum\limits_{i=1}^{I} \omega_i \geq 2\epsilon N$

6 $\quad$ Update w according to (C.14)

7 $\quad$ **if** $\|\omega\|_1 < N(1 - 2\epsilon)$ **then**

8 $\quad\quad$ **return** None

9 $\quad$ Set $\lambda = \left\| Y^\top D(\omega) Y - N \mathbf{I}_{k^\star} \right\|_2$

10 **return** ω

Algorithm 6: Sample filtering when $N \leq k^\star$.

---

1 Set γ_3 as (C.21), and $\lambda = \left\| \sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega) \right\|_2$. ($\omega$ is initialized as $\mathbf{1}$.)

2 while $\lambda \geq \gamma_3$ **do**

3   Set v to be the unit singular vector associated with λ

4   Compute $\tau_i = \frac{v_i^2}{\omega_i}\mathbf{1}_{\{w_i>0\}}$

5   Update w according to (C.22)

6   **if** $\|\omega\|_1 < N(1-6\epsilon)$ **then**

7   **return** None

8   Set $\lambda = \left\| \sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega) \right\|_2$

9 return ω

Algorithm 7: Weight filtering.

---

1 Compute $\tau_i = \left| \left\langle \sqrt{\omega_i}Y_i, \sum_{j=1}^{N}\sqrt{\omega_j}Y_j \right\rangle - \omega_i k^\star \right| \mathbf{1}_{\{\omega_i>0\}}, 1 \leq i \leq N$

2 Sort τ_i by desreasing order and find the indices $\{i_1, i_2, \ldots, i_{\epsilon N}\}$ corresponding to the first ϵN maximal τ_i

3 Set $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_{\epsilon N}}$ to zero

4 return ω

A.3   Deferred Experiment Results

The following figures display additional results deferred from Section 5.4.

Algorithm 8: polynomial-time algorithm when $\rho \gtrsim P(\epsilon, k^\star, N, 1)$.

---

1 **import X**, $N, K, \sigma = 1$

2 **import** Algorithm 4, 5, 6, 7

3 $\mathbf{X} = \mathbf{X}/\sigma$

4 Compute $k_1^\star, P_1^\star, k_2^\star, P_2^\star$ from $N, K, \sigma = 1, \epsilon$ as 2.3 and 2.4

5 **if** $k_1^\star \leq k_2^\star$ **then**

6 $\quad \lfloor \ k^\star = k_1^\star, P^\star = P_1^\star$

7 **else**

8 $\quad \lfloor \ k^\star = k_2^\star, P^\star = P_2^\star$

9 $\mathbf{Y} = \mathbf{X}P^\star$

10 Execute Algorithm 4, and record the returning value $\mathtt{R_1}$

11 **if** R_1 *is None* **then**

12 $\quad$ Reject H_0

13 $\quad$ **return**

14 **else**

15 $\quad \lfloor \ \mathbf{Y} = \mathtt{R_1}$

16 Set $\omega = \mathbf{1}_{[k^\star]}$

17 **if** $N > k^\star$ **then**

18 $\quad \lfloor$ Execute Algorithm 5, and record the returning value $\mathtt{R_2}$

19 **else**

20 $\quad \lfloor$ Execute Algorithm 6, and record the returning value $\mathtt{R_2}$

21 **if** R_2 *is None* **then**

22 $\quad$ Reject H_0

23 $\quad$ **return**

24 **else**

25 $\quad \lfloor \ \omega = \mathtt{R_2}$

26 Execute Algorithm 7, and record the returning value $\mathtt{R_3}$

27 $\omega = \mathtt{R_3}$

28 **if** $\left| \left\| \sum_{i=1}^{N} \sqrt{\omega_i} Y_i \right\|_2^2 - k^\star \|\omega\|_1 \right| \geq c_2 N^2 P(\epsilon, k^\star, N, 1)$ **then**

29 $\quad \lfloor$ Reject H_0

30 **else**

31 $\quad \lfloor$ Accept H_0

32 **return**
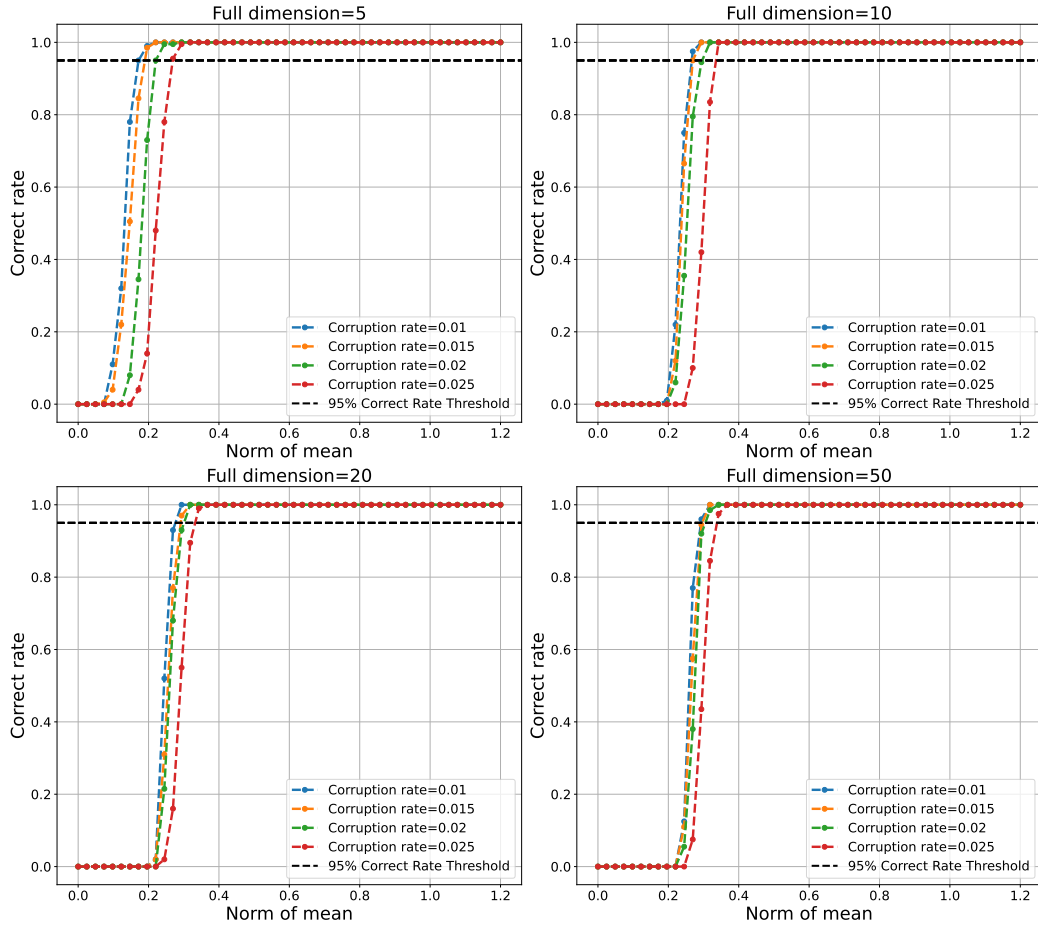
Figure 4: **Top**: Rejection rate (or Empirical power) versus the ℓ_2 norm of the mean under H_1 in the classic setting ($N = 2000$). **Bottom**: Rejection rate versus the ℓ_2 norm of the mean in the high-dimensional setting ($N = 500$).

B Some Basic Probabilistic Concepts and Useful Lemmas

In this section, we recall some basic probabilistic concepts and useful lemmas that are frequently mentioned and leveraged in this paper. Most of them are related to concentrations and the measures of distance between distributions.
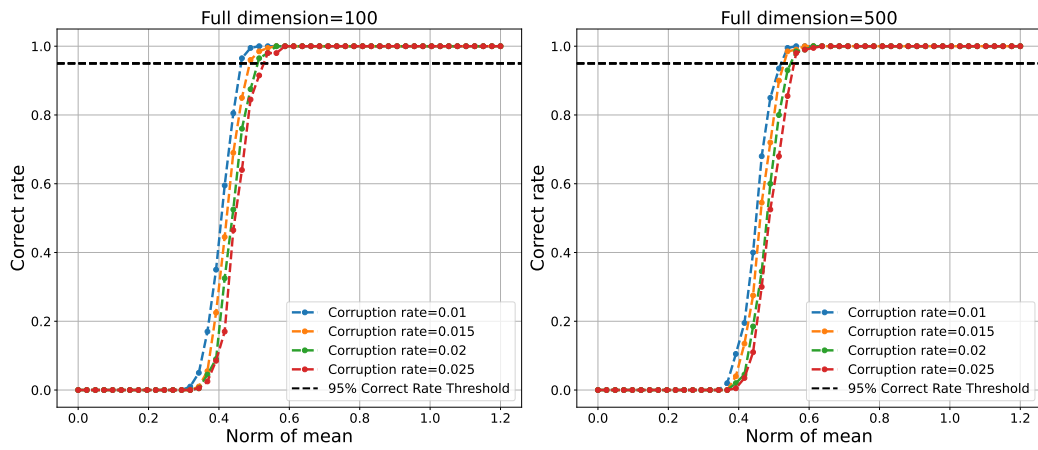
Definition B.1 (Total variation distance). *For a σ-algebra $(\Omega, \mathscr{F})$ and the probabilities $\mathcal{P}, \mathcal{Q}$ defined on $\mathscr{F}$, the total variation distance between $\mathcal{P}$ and $\mathcal{Q}$ is defined as*

$$TV(\mathcal{P}, \mathcal{Q}) = \sup_{E \in \mathscr{F}} |\mathcal{P}(E) - \mathcal{Q}(E)|.$$

From the definition, we immediately know that $\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = \mathrm{TV}(\mathcal{Q}, \mathcal{P})$ and $0 \leq \mathrm{TV}(\mathcal{P}, \mathcal{Q}) \leq 1$. $\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = 0$ if and only if $\mathcal{P} = \mathcal{Q}$ with only exception of zero-measure sets. When $\mathcal{P}, \mathcal{Q}$ has densities p, q with respect to some measure ν and space Ω_ν, it can be also easily shown that

$$\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = \int_{\Omega_\nu} \min \{p, q\} \, d\nu = \frac{1}{2} \int_{\Omega_\nu} |p - q| \, d\nu. \tag{B.1}$$

Definition B.2 (the Kullback-Leibler divergence (KL divergence)). *Given the same σ-algebra and probabilities, additionally assume that $\mathcal{P} \ll \mathcal{Q}$. The KL divergence between $\mathcal{P}$ and $\mathcal{Q}$ is defined as*

$$KL(\mathcal{P}, \mathcal{Q}) = \int_\Omega \ln \frac{d\mathcal{P}}{d\mathcal{Q}} d\mathcal{P}.$$

KL divergence is always non-negative, with $\mathrm{KL}(\mathcal{P}, \mathcal{Q}) = 0$ implying $\mathcal{P} = \mathcal{Q}$ with only exception of zero-measure sets. Contrasted with total variation distance, KL divergence is not symmetric, meaning that $\mathrm{KL}(\mathcal{P}, \mathcal{Q}) \neq \mathrm{KL}(\mathcal{Q}, \mathcal{P})$ in general, even if $\frac{d\mathcal{P}}{d\mathcal{Q}}$ and $\frac{d\mathcal{Q}}{d\mathcal{P}}$ both exist.

Lemma B.3 (Pinsker's inequality). *For the probability distributions $\mathcal{P}, \mathcal{Q}$ with $\mathcal{P} \ll \mathcal{Q}$, we have*

$$TV(\mathcal{P}, \mathcal{Q}) \leq \sqrt{2KL(\mathcal{P}, \mathcal{Q})}.$$

When $\mathcal{P} \ll \mathcal{Q}$ and $\mathcal{Q} \ll \mathcal{P}$, it can be refined as

$$TV(\mathcal{P}, \mathcal{Q}) \leq \sqrt{2 \min \{KL(\mathcal{P}, \mathcal{Q}), KL(\mathcal{Q}, \mathcal{P})\}}.$$

There are many proofs for this lemma. One can refer Tsybakov (2009) for a quick proof.

Definition B.4 (Coupling). *Given probability distributions $\mathcal{P}, \mathcal{Q}$, a valid coupling of $\mathcal{P}$ and $\mathcal{Q}$ is any pair of random variables (X, Y) where the marginal distribution of X ang Y is $\mathcal{P}, \mathcal{Q}$ respectively.*

The following classic lemma help us connect the total variation distance with the Hamming distance in the proof of Theorem 3.5.

Lemma B.5. *For any probability distributions $\mathcal{P}$ and $\mathcal{Q}$, their total variation distance equals to the infimum over the probability that X is not equal to Y, where (X, Y) is a valid coupling of $\mathcal{P}$ and $\mathcal{Q}$.*

$$TV(P, Q) = \inf_{(X, Y) \text{ is a valid coupling of } P \text{ and } Q} \mathbb{P}(X \neq Y). \tag{B.2}$$

The coupling that achieves the equality above is referred as the optimal coupling between $\mathcal{P}$ and $\mathcal{Q}$.

The following definitions and lemmas concern the sub-Gaussian and sub-exponential random variables, which are closely related to the concentration bounds in the proof of the main theorems.

Definition B.6 (Sub-Gaussian random variables)**.** *For a univariate random variable X, define its sub-Gaussian norm as $\|X\|_{\psi_2} := \inf\left\{t : t > 0, \mathbb{E}\exp\left\{\frac{X^2}{t^2}\right\} \leq 2\right\}$. If such $\|X\|_{\psi_2}$ does not exist, then define $\|X\|_{\psi_2} = \infty$. X is said to be sub-Gaussian if $\|X\|_{\psi_2} < \infty$.*

There are many equivalent definitions for sub-Gaussian random variables. The most frequently mentioned one is $\mathbb{P}(|X| \geq t) \leq 2\exp\left\{-\frac{t^2}{c\|X\|_{\psi_2}^2}\right\}$.

Definition B.7 (Sub-exponential random variables)**.** *For a univariate random variable X, define its sub-exponential norm as $\|X\|_{\psi_1} := \inf\left\{t : t > 0, \mathbb{E}\exp\left\{\frac{|X|}{t}\right\} \leq 2\right\}$. If such $\|X\|_{\psi_1}$ does not exist, then define $\|X\|_{\psi_1} = \infty$. X is said to be sub-exponential if $\|X\|_{\psi_1} < \infty$.*

There are also many equivalent definitions for sub-exponential random variables. A popular one is $\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2\exp\left\{-c\min\left\{\frac{t^2}{\|X\|_{\psi_1}^2}, \frac{t}{\|X\|_{\psi_1}}\right\}\right\}$.

The requirement of sub-exponential random variables is weaker than that of sub-Gaussian in the sense that if X is sub-Gaussian, it must be sub-exponential. Moreover, if X is sub-Gaussian, then by definition X^2 is sub-exponential.

The following is the famous Bernstein's inequality, which handles multiple independent sub-Gaussian (sub-exponential) random variables. It also generalizes the equivalent definitions mentioned above.

Lemma B.8 (Bernstein's inequality)**.** *Assume $X_1, \ldots, X_n$ are independent, zero-mean random variables. If $X_1, \ldots, X_n$ are sub-Gaussian, then for any $t \geq 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left\{-ct^2 / \sum_{i=1}^n \|X_i\|_{\psi_2}^2\right\}. \tag{B.3}$$

If $X_1, \ldots, X_n$ are sub-exponential, then we have

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left\{-c\min\left\{t^2 / \sum_{i=1}^n \|X_i\|_{\psi_1}^2, t / \max_i\left\{\|X_i\|_{\psi_1}\right\}\right\}\right\}. \tag{B.4}$$

The next lemma is particularly useful when we control the ℓ_2 norm of a random vector with independent sub-Gaussian coordinates.

Lemma B.9. *Assume $X = (X_1, \ldots, X_n) \in \mathbb{R}^d$ where $X_1, \ldots, X_n$ are independent, $\mathbb{E}X_i = 0$, $Var(X_i) = 1$, and each X_i is sub-Gaussian. Then $\|X\|_2 - \sqrt{n}$ is sub-Gaussian with $\left\|\|X\|_2 - \sqrt{n}\right\|_{\psi_2} \leq c\max_i\left\{\|X_i\|_{\psi_2}^2\right\}$.*

The definitions and lemmas above can be found in many modern statistics textbook. For an example, see Vershynin (2024).

The famous Hanson-Wright inequality is a powerful tool to bound the tail probability of quadratic forms of the sub-Gaussian random variables. It can also be viewed as a generalization of the Bernstein's inequality (B.4). See Hanson and Wright (1971) for the initial paper and Rudelson and Vershynin (2013) for a modern proof.

Lemma B.10 (Hanson-Wright inequality). *Assume that $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ is a random vector with independent, zero-mean and sub-Gaussian coordinates. Let A be an $n \times n$ matrix and $K = \max_i \|X_i\|_{\psi_2}$. Then for any $t \geq 0$, we have*

$$\mathbb{P}\left(|X^\top A X - \mathbb{E} X^\top A X| > t\right) \leq 2 \exp\left\{-c \min\left\{\frac{t^2}{K^4 \|A\|_F^2}, \frac{t}{K^2 \|A\|_2}\right\}\right\}, \qquad \text{(B.5)}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, and $\|\cdot\|_2$ is the ℓ_2 operator norm of a matrix.

We now turn to the Wishart matrix, which plays a central role in covariance estimation in high-dimensional Gaussian models. Additional properties of the Wishart distribution and its applications in estimation can be found in Kubokawa and Srivastava (2008) and Bishop et al. (2018).

Definition B.11 (Wishart matrix). *Given a matrix $X \in \mathbb{R}^{n \times d}$ where each row is independently drawn from $\mathcal{N}(0, \boldsymbol{\Sigma})$ for some $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$. Then, the Wishart matrix with scale matrix $\boldsymbol{\Sigma}$ and dimension (d, n), denoted as $W_d(\boldsymbol{\Sigma}, n)$, is defined as*

$$W_d(\boldsymbol{\Sigma}, n) = \mathbf{X}^\top \mathbf{X}.$$

Specifically, when $\boldsymbol{\Sigma} = \mathbf{I}_d$ is the identity matrix, the notation is simplified to $W_d(n)$. The distribution of $W_d(\boldsymbol{\Sigma}, n)$ is called Wishart distribution with parameter $d, n, \boldsymbol{\Sigma}$. When the meaning is clear from the context, we also use notation $W_d(\boldsymbol{\Sigma}, n)$ (and also $W_d(n)$) as Wishart distribution.

For the Gaussian random matrix X, the following lemma can be used to bound its maximum and minimum singular values, which can be further used to establish the bound for the ℓ_2 operator norm of $W_d(\boldsymbol{\Sigma}, n)$. One can find a proof of this lemma in Wainwright (2019) by the concentration bound for Lipschitz functions and the Sudakov-Fernique comparison inequality.

Lemma B.12. *Assume $X \in \mathbb{R}^{n \times d}$, and each row is independently drawn from $\mathcal{N}(0, \boldsymbol{\Sigma})$, then for any $\epsilon > 0$, we have*

$$\mathbb{P}\left(\frac{\sigma_{max}(X)}{\sqrt{n}} \geq \gamma_{max}(\sqrt{\boldsymbol{\Sigma}})(1 + \epsilon) + \sqrt{\frac{tr(\boldsymbol{\Sigma})}{n}}\right) \leq \exp\left\{-\frac{n\epsilon^2}{2}\right\}.$$

When $n \geq d$, we further have the following control on minimum singular value of X:

$$\mathbb{P}\left(\frac{\sigma_{min}(X)}{\sqrt{n}} \leq \gamma_{min}(\sqrt{\boldsymbol{\Sigma}})(1 - \epsilon) - \sqrt{\frac{tr(\boldsymbol{\Sigma})}{n}}\right) \leq \exp\left\{-\frac{n\epsilon^2}{2}\right\}.$$

Here $\sigma_{max}(X)$ (or $\sigma_{min}(X)$) is the max (or min) singular value of X and $\gamma_{max}(\sqrt{\boldsymbol{\Sigma}})$ (or $\gamma_{min}(\sqrt{\boldsymbol{\Sigma}})$) is the max (or min) eigenvalue of $\sqrt{\boldsymbol{\Sigma}}$.

We can leverage Lemma B.12 to bound the ℓ_2 operator norm of $W_d(n)$.

Corollary B.13. *Assume $W = \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{d \times d}$ is drawn from $W_d(n)$, and $n \geq d$, then for any $\delta > 0$, with probability higher than $1 - 2\delta$, we have*

$$\|W - n\mathbf{I}_d\|_2 \leq C\left(\sqrt{nd} + \sqrt{n \ln\left(\frac{1}{\delta}\right)} + \ln\left(\frac{1}{\delta}\right)\right)$$

Proof. Set $\mathbf{\Sigma} = \mathbf{I}_d$ in the first inequality of Lemma B.12, we have:

$$\mathbb{P}\left(\frac{\sigma_{\max}(X)}{\sqrt{n}} \geq (1+\epsilon) + \sqrt{\frac{d}{n}}\right) \leq \exp\left\{-\frac{n\epsilon^2}{2}\right\}.$$

Using $\sigma_{\max}(X) = \sqrt{\gamma_{\max}(W_d(n))}$ and after some calculation, we have

$$\mathbb{P}\left(\gamma_{\max}(W_d(n)) - n \geq d + 2\sqrt{nd} + n\epsilon^2 + 2n\epsilon + 2\sqrt{nd}\epsilon\right) \leq \exp\left\{-\frac{n\epsilon^2}{2}\right\}.$$

Setting $\delta = \sqrt{\frac{\ln\left(\frac{1}{\epsilon}\right)}{n}}$ yields

$$\mathbb{P}\left(\gamma_{\max}(W_d(n)) - n \geq d + 2\sqrt{nd} + \ln\left(\frac{1}{\delta}\right) + 2\sqrt{n\ln\left(\frac{1}{\delta}\right)} + 2\sqrt{d\ln\left(\frac{1}{\delta}\right)}\right) \leq \delta.$$

The lower bound follows from the condition that $n \geq d$. The upper bound is derived with a similar argument. The proof is completed. $\qquad\square$

Under our problem setting, the following corollary is also required. It allows the mean vector μ of the normal distribution to be nonzero, but at the expense of introducing a dependence on $\|\mu\|_2$.

Corollary B.14. *Assume* $\mathbf{X} \in \mathbb{R}^{n \times d}$ *and each row is independently drawn from* $\mathcal{N}(\mu, \mathbf{I}_d)$ *with* $\rho := \|\mu\|_2$, *then for* $\delta > 0$, *with probability higher than* $1 - 2\delta$, *we have:*

(i), $n \geq d$: $\left\|\mathbf{X}^\top\mathbf{X} - n\mathbf{I}_d\right\|_2 \leq C_1\left((1+\rho)\left(\sqrt{nd} + \sqrt{n\ln\left(\frac{1}{\delta}\right)}\right) + \ln\left(\frac{1}{\delta}\right) + n\rho^2\right)$,

(ii), $n < d$: $\left\|\mathbf{X}\mathbf{X}^\top - d\mathbf{I}_n\right\|_2 \leq C_2\left((1+\rho)\left(\sqrt{nd} + \sqrt{d\ln\left(\frac{1}{\delta}\right)}\right) + \ln\left(\frac{1}{\delta}\right) + n\rho^2\right)$.

Proof. When $n \geq d$, let $Z_i = X_i - \mu$, we have

$$\left\|\mathbf{X}^\top\mathbf{X} - n\mathbf{I}_d\right\|_2 = \left\|\sum_{i=1}^n X_i X_i^\top - n\mathbf{I}_d\right\|_2 \leq \left\|\sum_{i=1}^n Z_i Z_i^\top - n\mathbf{I}_d\right\|_2 + 2\left\|\sum_{i=1}^n \mu Z_i^\top\right\|_2 + n\left\|\mu\mu^\top\right\|_2.$$

By Corollary B.13, we already have $\left\|\sum_{i=1}^n Z_i Z_i^\top - n\mathbf{I}_d\right\|_2 \leq C_1\left(\sqrt{nd} + \sqrt{n\ln\left(\frac{1}{\delta}\right)} + \ln\left(\frac{1}{\delta}\right)\right)$. We also have $n\left\|\mu\mu^\top\right\|_2 = n\alpha^2$. It remains to bound $\left\|\sum_{i=1}^n \mu Z_i^\top\right\|_2$.

$$\left\|\sum_{i=1}^n \mu Z_i^\top\right\|_2 \leq \|\mu\|_2 \cdot \left\|\sum_{i=1}^n Z_i\right\|_2.$$

Since $Z_i \sim \mathcal{N}(0, \mathbf{I}_d)$ for each i, from B.9 we know that $\left\|\sum_{i=1}^n \frac{Z_i}{\sqrt{n}} - d\right\|_2$ is a sub-Gaussian random variable and therefore with probability higher than $1 - \delta$ we have:

$$\left\|\sum_{i=1}^n \frac{Z_i}{\sqrt{n}} - d\right\|_2 \leq C\sqrt{\ln\left(\frac{1}{\delta}\right)},$$

which leads to $\left\|\sum_{i=1}^{n} Y_i\right\|_2 \leq C\left(\sqrt{nd} + \sqrt{n \ln\left(\frac{1}{\delta}\right)}\right)$. Combined with previous results on $\left\|\sum_{i=1}^{n} Z_i Z_i^\top - n\mathbf{I}_d\right\|_2$ and $n\left\|\mu\mu^\top\right\|_2$, we complete the proof of (i).

To bound $\left\|\mathbf{X}\mathbf{X}^\top - d\mathbf{I}_n\right\|_2$, we notice that $\mathbf{X}\mathbf{X}^\top$ has the same non-trivial eigenvalues with $\mathbf{X}^\top\mathbf{X} = \mathbf{Z}^\top\mathbf{Z} + 2\sum_{i=1}^{n} \mu Z_i^\top + n\mu\mu^\top$, in this case, we only need to address the ℓ_2 operator norm of $\mathbf{Z}^\top\mathbf{Z}$, which again equals to $\mathbf{Z}\mathbf{Z}^\top$. Since $\mathbf{Z}\mathbf{Z}^\top$ actually follows the distribution of $W_n(d)$, the results follow from the proof of (i). $\qquad\square$

The next lemma is from Fact 4.2 in Dong et al. (2019). It is an analog of Corollary B.13, but within the field when the ℓ_1 norm of a weight ω is small compared to n.

Lemma B.15. *Assume* $\mathbf{X} \in \mathbb{R}^{n \times d}$ *and each row is independently drawn from* $\mathcal{N}(0, \mathbf{I}_d)$. *Then, with some constant* $C_1 > 0$ *and another sufficiently large constant* C_2, *as long as* $n \geq C_2 \cdot \frac{1}{\epsilon}$, *the following holds with probability at least* $1 - \delta$ *for any* v, ω *such that* $\|v\|_2 = 1$, $\|\omega\|_1 \leq C_2 \cdot \epsilon n$:

$$\left\|\sum_{i=1}^{n} \omega_i X_i X_i^\top\right\|_2 \leq C_1\left(\epsilon n \ln\left(\frac{1}{\epsilon}\right) + d + \ln\left(\frac{1}{\delta}\right)\right).$$

Corollary B.16. *Assuming the same conditions as Lemma B.15, we further allow* X_i *from* $\mathcal{N}(\mu, \mathbf{I}_d)$ *with* $\|\mu\|_2 \leq \alpha$. *As a consequence, the inequality will track the dependence on* α:

$$\left\|\sum_{i=1}^{n} \omega_i X_i X_i^\top\right\|_2 \leq C\left[\epsilon n\left(\ln\left(\frac{1}{\epsilon}\right) + \alpha^2\right) + \sqrt{\epsilon n}\alpha + d + \ln\left(\frac{1}{\delta}\right)\right].$$

Proof. The proof is basically from Lemma B.15 the standard argument of concentration. For any v such that $\|v\|_2 = 1$, we have:

$$v^\top\left(\sum_{i=1}^{n} \omega_i X_i X_i^\top\right)v = v^\top\left(\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^\top\right)v + \sum_{i=1}^{n} \omega_i(v^\top\mu)^2 + 2(v^\top\mu)\sum_{i=1}^{n} \omega_i(v^\top(X_i - \mu))$$

Bound of the first term can be obtained from Lemma B.15. For the second term, by the Cauchy-Schwarz inequality we can bound it by $2\|\omega\|_1 \cdot \alpha^2 \leq C \cdot \epsilon n\alpha^2$. For the third term, since $\sum_{i=1}^{n} \omega_i v^\top(X_i - \mu)$ follows from $\mathcal{N}(0, \|\omega\|_2^2)$, we know we can bound it by $C \cdot \alpha\|\omega\|_2 \leq C \cdot \alpha\sqrt{\epsilon n}$. Combining them together yields the result. $\qquad\square$

Hölder's inequality is a fundamentally important tool in probability, mathematical analysis and functional analysis. We briefly review it here.

Lemma B.17 (Hölder's inequality). *Let* $(\Omega, \mathscr{F})$ *be a* σ-*algebra equipped with a measure* μ. *For a pair of conjugate numbers* p, q: $\frac{1}{p} + \frac{1}{q}, 1 \leq p, q \leq \infty$, *and* $f \in L_p(\Omega, \mu)$, $g \in L_q(\Omega, \mu)$, *we have*

$$\|fg\|_1 = \int_\Omega |fg|\, d\mu \leq \left(\int_\Omega |f|^p\, d\mu\right)^{\frac{1}{p}}\left(\int_\Omega |g|^q\, d\mu\right)^{\frac{1}{q}} = \|f\|_p \|g\|_q.$$

The equality holds if and only if $a|f|^p = b|g|^q$ *a.s. for some constants* a, b.

The next lemma is a famous result for the combination number.

Lemma B.18. *For any $0 \leq k \leq n$, we have $\sum_{i=0}^{k} \binom{n}{i} \leq \left(\frac{en}{k}\right)^k$.*

Proof. By the Taylor's expansion of e^k, we know $e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} \geq \sum_{i=0}^{k} \frac{k^i}{i!}$. Therefore, it suffices to show

that $\frac{k^k}{n^k}\binom{n}{i} \leq \frac{k^i}{i!}$ for any $0 \leq i \leq k$, and this is equivalent to showing that $\frac{k^k}{n^k}\prod_{j=0}^{i-1}(n-j) \leq k^i$.

However, the last statement can be directly obtained from the fact that $\frac{k^{k-i}}{n^{k-i}} \leq 1 \leq \frac{n^i}{\prod_{j=0}^{i-1}(n-j)}$. $\qquad\square$

Finally, we present an intuitive and useful result at the intersection of convexity and combinatorics. While this result was originally introduced in Canonne et al. (2023), their proof contains a technical gap which we address below.

Lemma B.19. *Denote $[0,1]^N$ to be the set $\{(x_1, x_2, \ldots, x_N) : 0 \leq x_i \leq 1, \forall i \in [N]\}$. Given any nonnegative integer $k \leq N$ and any vector $x \in [0,1]^N$ with condition $\|x\|_1 \leq k$, there exists a positive integer M, sets $T_1, T_2, \ldots, T_M \subset [N]$, and positive numbers $a_1, a_2, \ldots, a_M$ such that the following statements hold:*

(i), $|T_i| \leq k, \forall i \in [M]$;

(ii), $0 \leq a_i \leq 1, \forall i \in [M], \sum_{i=1}^{M} a_i \leq 1$;

(iii), $x = \sum_{i=1}^{M} a_i \mathbf{1}_{T_i}$,

where same as before $\mathbf{1}_{T_i}$ is the indicator vector of the set T_i.

Proof. We first show the following result: for any vector $(x_1, \ldots, x_N) \in [0,1]^n$ and an given integer $k \leq N$ such that $\|x\|_1 = k$, and $0 \leq x_i \leq 1$ there exist sets T_i of cardinality $|T_i| = k$ and weights $a_i \geq 0$ with $\sum_{i=1}^{M} a_i = 1$ such that $x = \sum_{i=1}^{M} a_i \mathbf{1}_{T_i}$.

The proof of this statement uses Birkhoff-von Neumann's theorem. Construct the following $N \times N$ matrix

$$
W := \begin{pmatrix}
\frac{x_1}{k} & \frac{x_2}{k} & \cdots & \frac{x_N}{k} \\
\frac{x_1}{k} & \frac{x_2}{k} & \cdots & \frac{x_N}{k} \\
\vdots & \vdots & \cdots & \vdots \\
\frac{1-x_1}{N-k} & \frac{1-x_2}{N-k} & \cdots & \frac{1-x_N}{N-k} \\
\vdots & \vdots & \cdots & \vdots \\
\frac{1-x_1}{N-k} & \frac{1-x_2}{N-k} & \cdots & \frac{1-X_N}{N-k}
\end{pmatrix},
$$

where the W_{ij} is defined as

$$
W_{ij} = \begin{cases} \frac{x_j}{k}, & 1 \leq i \leq k, \\ \frac{1-x_j}{N-k}, & k+1 \leq i \leq N. \end{cases}
$$

It is easy to check that this is a doubly stochastic matrix. By Birkhoff-von Neumann's theorem, there exists a set of permutation matrices $P_1, \ldots, P_M$ such that

$$W = \sum_{i=1}^{M} \omega_i P_i.$$

Since the sum of the first k rows of W is x, now summing all the first k rows of the matrices P_i shows the desired result.

For the general case, suppose that $k - 1 < \|x\|_1 < k$. Consider the augmented vector $\widetilde{x} = (x_1, \ldots, x_N, k - \sum_{i=1}^{N} x_i)$, and apply the logic above to that vector. This gives us a representation $\widetilde{x} = \sum_{i=1}^{\widetilde{M}} w_i v_i$, where v_i consists of k-ones and $(N + 1 - k)$-zeros. Erasing the last entry of the v_i's, we obtain a representation of x with desired vectors and weights. The proof is completed. $\qquad\square$

C Deferred Proofs of the Lemmas and Main Theorems

C.1 Proof of Lemma 3.1

Proof of Lemma 3.1. By the conditions and the definition of the Kolmogorov k-width, the following inequality holds:

$$\inf_{P \in \mathcal{P}_{k-1}} \sup_{\theta \in K} \|\theta - P\theta\|_2^2 > c^2 \sigma^2.$$

If we only consider the projections aligned with the coordinate axes, there are $n_k := \binom{d}{k}$ such projections (and countably many when $\mathcal{X} \subset \ell_2$). Denote the index set of the whole space $\mathcal{X}$ by I — i.e., $I = [d]$ when $\mathcal{X} = \mathbb{R}^d$ and $\mathbb{N}$ when $\mathcal{X} \subset \ell_2$ — and the index sets of such projections by J. We have

$$\inf_{P \in \mathcal{P}_{k-1}} \sup_{\theta \in K} = \|\theta - P\theta\|_2^2 \leq \inf_{S \in J} \sup_{\theta \in K} \sum_{i \in I} \theta_i^2 - \sum_{i \in S} \theta_i^2,$$

where the minimum over S is taken with respect to all elements in J. Since K is quadratically convex, the RHS above can be written as

$$\inf_{S \in J} \sup_{\theta \in K} \sum_{i \in I} \theta_i^2 - \sum_{i \in S} \theta_i^2 \leq \inf_{S \in J} \sup_{t \in K^2} \sum_{i \in I} t_i - \sum_{i \in S} t_i = \inf_{w \in S_k} \sup_{t \in K^2} \|t\|_1 - \langle w, t \rangle,$$

where w ranges in the set $S_k := \{ e \mid e \in \mathbb{R}^d \text{ (or } \mathbb{R}^\infty), e_i \in \{0, 1\}, \|e\|_1 = k - 1 \}$. It follows that for each small enough $\varepsilon > 0$ and each $w_i \in S_k$ there exists a t_i such that $\|t_i\|_1 - \langle w_i, t_i \rangle \geq c^2 \sigma^2 - \varepsilon$. Since K is convex and orthosymmetric, we may assume without loss of generality that $t_i \geq 0$ for all $i \in I$, and $t_i = 0$ on the support of w_i. In this way, after appropriately scaling down the coordinates of t_i (which is permissible since K is convex), we obtain

$$\|t_i\|_1 = \|t_i\|_1 - \langle w_i, t_i \rangle = c^2 \sigma^2 - \epsilon,$$

where the second equality follows from the fact that $\mathbf{0} \in K$. We now argue that there exists a convex combination $t_\alpha := \sum_{i \in J} \alpha_i t_i$ such that $\|t_\alpha\|_\infty \leq \frac{c^2 \sigma^2}{k}$. First, we observe that since all t_i have nonnegative entries, we have $\|t_\alpha\|_\infty = \sup_{r : r \geq 0, \|r\|_1 = 1} \langle r, t_\alpha \rangle$. Hence, it suffices to show that

$$\inf_{\alpha : \alpha \geq 0, \|\alpha\|_1 = 1} \sup_{r : r \geq 0, \|r\|_1 = 1} \langle r, t_\alpha \rangle \leq \frac{c^2 \sigma^2}{k}.$$

Since both sets over which the optimization is performed are convex, and the function $\langle r, t_\alpha \rangle = \sum_{i \in J} \alpha_i \langle r, t_i \rangle$ is convex-concave (in fact, it is linear in both arguments), by the minimax theorem, we have

$$\inf_{\alpha : \alpha \geq 0, \|\alpha\|_1 = 1} \sup_{r : r \geq 0, \|r\|_1 = 1} \langle r, t_\alpha \rangle = \sup_{r : r \geq 0, \|r\|_1 = 1} \inf_{\alpha : \alpha \geq 0, \|\alpha\|_1 = 1} \langle r, t_\alpha \rangle = \sup_{r : r \geq 0, \|r\|_1 = 1} \min_{i \in J} \langle r, t_i \rangle.$$

Let U denote the index set of the largest $k - 1$ entries of r. Observe that $\min_{i \in J} \langle r, t_i \rangle \leq \langle r, t_r \rangle$, where t_r is chosen such that $t_{r,U} = \mathbf{0}$. Consequently,

$$\langle r, t_r \rangle = \sum_{i \notin U} r_i t_{r,i} \leq \|r_{\setminus U}\|_\infty \|t_r\|_1 = \|r_{\setminus U}\|_\infty (c^2 \sigma^2 - \varepsilon),$$

where $r_{\backslash U}$ denotes the vector obtained by removing the largest $k-1$ coordinates of r. Finally, observe that $\left\|r_{\backslash U}\right\|_\infty \leq \frac{1}{k}$ since $1 = \|r\|_1 \geq k\left\|r_{\backslash U}\right\|_\infty$. Hence, we conclude that there exists $t^\star = t_{\alpha^\star}$ such that

$$\|t^\star\|_\infty \leq \frac{c^2\sigma^2 - \epsilon}{k} < \frac{c^2\sigma^2}{k}.$$

In addition, since t^* is a convex combination of vectors t_i, we must have $\|t^\star\|_1 = c^2\sigma^2 - \varepsilon$. Taking $\varepsilon \to 0$ and noticing $\|t^\star\|_\infty = \|\theta^\star\|_\infty^2$, $\|t^\star\|_1 = \|\theta^\star\|_2^2$ complete the proof. □

C.2 Proof of Theorem 3.2

Proof of Theorem 3.2. We first assume $k^\star \geq 1$, and start by a classical treatment of minimax testing lower bounds which we take from Baraud (2002). Let ν denote an arbitrary prior on the set $\{\mu \mid \mu \in K, \|\mu\|_2 \geq \rho\}$, and $\mathbb{P}_\nu$ denote the marginal joint distribution of $X_1, X_2, \ldots, X_N$ given the prior ν. We have that

$$
\begin{aligned}
\inf_{\psi : \mathbb{P}_0(\psi=1)\leq\alpha} \sup_{\mu\in K, \|\mu\|_2\geq\rho} \mathbb{P}_\mu(\psi=0) &\geq \inf_{\psi : \mathbb{P}_0(\psi=1)\leq\alpha} \mathbb{P}_\nu(\psi=0) \\
&\geq \inf_{\psi : \mathbb{P}_0(\psi=1)\leq\alpha} 1 - \mathbb{P}_\nu(\psi=1) \\
&\geq \inf_{\psi : \mathbb{P}_0(\psi=1)\leq\alpha} 1 - \mathbb{P}_0(\psi=1) + (\mathbb{P}_0(\psi=1) - \mathbb{P}_\nu(\psi=1)) \\
&\geq 1 - \alpha - \mathrm{TV}\left(\mathbb{P}_0, \mathbb{P}_\nu\right) \\
&\overset{(i)}{=} 1 - \alpha - \frac{1}{2}\left\|\mathbb{P}_0 - \mathbb{P}_\nu\right\|_1,
\end{aligned}
$$

where (i) uses (B.1). Now assuming $\mathbb{P}_\nu \ll \mathbb{P}_0$, and let $L_\nu = \frac{d\mathbb{P}_\nu}{d\mathbb{P}_0}$ be the Radon-Nikodym derivative of $\mathbb{P}_\nu$ to $\mathbb{P}_0$. By Cauchy-Schwartz inequality, we have

$$\left\|\mathbb{P}_0 - \mathbb{P}_\nu\right\|_1 = \mathbb{E}_0\left|L_\nu - 1\right| \leq \sqrt{\mathbb{E}_0\left(L_\nu - 1\right)^2} = \sqrt{\mathbb{E}_0\left[L_\nu^2\right] - 1}.$$

Hence we conclude

$$\inf_{\psi : \mathbb{P}_0(\psi=1)\leq\alpha} \sup_{\mu\in K, \|\mu\|_2\geq\rho} \mathbb{P}_\mu(\psi=0) \geq 1 - \alpha - \frac{1}{2}\sqrt{\mathbb{E}_0\left[L_\nu^2\right] - 1}.$$

Let θ be the vector from Lemma 3.1 when $k = k_1^\star$, which means $\|\theta\|_2 = \frac{\left(k_1^\star\right)^{1/4}}{\sqrt{N}}\sigma$ and $\|\theta\|_\infty \leq \frac{1}{\sqrt{N}\left(k_1^\star\right)^{1/4}}\sigma$. We will use the prior $(\kappa\gamma_i\theta_i)_{i\in I} \in K$, where γ_i are i.i.d. Rademacher random variables for $i \in I$[1] and $0 < \kappa < 1$ is some small constant. Observe that all these vectors have squared Euclidean norm equal to $\rho^2 = \frac{\kappa^2\sqrt{k_1^\star}}{N}\sigma^2$.

---

[1] $I = [d]$ when $K \subset \mathbb{R}^d$ and $I = \mathbb{N}$ when $K \subset \ell_2$.

By the definition of $\mathbb{P}_\nu$, calculating the χ^2-divergence of this mixture yields:

$$\mathbb{E}_0(L_\nu)^2 - 1 = \mathbb{E}_{X_1,X_2,\ldots,X_N \sim \text{i.i.d.} \mathcal{N}(0,\sigma^2 \mathbf{I}_d)} \left(\frac{d\mathbb{P}_\nu}{d\mathbb{P}_0}\right)^2 - 1$$

$$= \mathbb{E}_{\mu,\mu'} \mathbb{E}_{X_1,X_2,\ldots,X_N \sim \text{i.i.d.} \mathcal{N}(0,\sigma^2 \mathbf{I}_d)} \frac{\prod_{i=1}^N \exp\left\{\frac{-\|X_i-\mu\|_2^2}{2\sigma^2}\right\} \prod_{i=1}^N \exp\left\{-\frac{\|X_i-\mu'\|_2^2}{2\sigma^2}\right\}}{\prod_{i=1}^N \exp\left\{-\frac{\|X_i\|_2^2}{\sigma^2}\right\}} - 1$$

$$= \mathbb{E}_{\mu,\mu'} \exp\left\{\frac{N\mu^\top \mu'}{\sigma^2}\right\} - 1,$$

where by $\mathbb{E}_{\mu,\mu'}$ we mean expectation over two i.i.d. draws from the prior distribution ν that we specified above. Using the symmetry of Rademacher random variable, we have

$$\mathbb{E}_{\mu,\mu'} \exp\left\{\frac{N\mu^\top \mu'}{\sigma^2}\right\} - 1 = \mathbb{E}_{\mu,\mu'} \exp\left\{N \sum_{i \in I} \frac{\mu_i \mu_i'}{\sigma^2}\right\} - 1$$

$$= \prod_{i \in I} \cosh\left(\frac{N\kappa^2 \theta_i^2}{\sigma^2}\right) - 1$$

$$\leq \exp\left\{\sum_{i \in I} \frac{N^2 \theta_i^4 \kappa^4}{2\sigma^4}\right\} - 1$$

$$\leq \exp\left\{\frac{N^2 \|\theta\|_\infty^2 \|\theta\|_2^2 \kappa^4}{2\sigma^4}\right\} - 1$$

$$\leq \exp\left\{\frac{\kappa^4}{2}\right\} - 1,$$

where we used the well known inequality $\cosh(x) \leq \exp\left\{\frac{x^2}{2}\right\}$, which can be checked simply by comparing the Taylor series, and the properties of θ from Lemma 3.1. It follows that the minimax risk for testing is lower bounded by

$$1 - \alpha - \frac{1}{2}\sqrt{\exp\left\{\frac{\kappa^4}{2}\right\} - 1}.$$

Recall that the ℓ_2 norm of the selected μ is $\rho = \|\mu\|_2 = \frac{(k_1^\star)^{1/4} \kappa}{\sqrt{N}} \sigma$. If we let the expression above equal to α and solve for κ, we obtain $\kappa = c(\alpha)$. Consequently, we conclude that if $\kappa \leq c(\alpha)$, and therefore $\rho \leq c(\alpha) \cdot \frac{(k_1^\star)^{1/4}}{\sqrt{N}} \sigma$, testing with uniform power α is not possible.

The remaining corner case is when $k_1^\star = 0$, or equivalently $D_0(K) = \sup_{\theta \in K} \|\theta\|_2 \leq \frac{1}{\sqrt{N}} \sigma$. In this case, any point $\theta \in K$ satisfies $\|\theta\|_2 \leq \frac{1}{\sqrt{N}} \sigma$. Taking a point θ_0 that (nearly) achieves the equality, and considering the mixture formed by $\gamma \theta_0$, where γ is a Rademacher random variable, one obtains that the resulting χ^2-divergence is bounded by $\cosh\left(\frac{N\|\theta\|_2^2}{\sigma^2}\right) - 1 \leq \exp\left\{\frac{N^2 \|\theta\|_2^4}{2\sigma^4}\right\} - 1 \leq \sqrt{e} - 1$ and consequently

$$\inf_{\psi : \mathbb{P}_0(\psi=1) \leq \alpha} \sup_{\mu \in K} \mathbb{P}_\mu(\psi = 0) \geq 1 - \alpha - \frac{1}{2}\sqrt{\sqrt{e} - 1},$$

47

implying that it is impossible to test the problem. Hence, $k_1^\star$ must be at least positive for sufficiently small α. □

C.3   Proof of Theorem 3.5

In this section, we formally prove Theorem 3.5. We first recall the core lemma. To establish Lemma C.1, two auxiliary lemmas, Lemma C.3 and Lemma C.4, are presented. Based on these tools, we then construct a symmetric set served as the prior for the mean μ via the QCO properties and compute the Hamming distance as an application of Lemma B.5 via a sequential optimal coupling. Finally, the lower bound is yielded by Pinsker's inequality and a concrete calculation of the KL divergence.

Lemma C.1 (Core lemma). *Suppose* $\mathbf{X} = (X_1, X_2, \ldots, X_N)^\top$ *and* $\mathbf{X}' = (X_1', X_2', \ldots, X_N')^\top$ *are random matrices with rows* $X_i, X_i' \in \mathbb{R}^d$. *Assume the laws of* $\mathbf{X}$ *and* $\mathbf{X}'$ *are* $\mathcal{U}$ *and* $\mathcal{V}$, *respectively. Additionally, assume that there exists a coupling between* $\mathcal{U}$ *and* $\mathcal{V}$ *under which the expectation of the Hamming distance between* $\mathbf{X}$ *and* $\mathbf{X}'$ *is* $\mathcal{O}(\epsilon N)$, *i.e.,*

$$\mathbb{E}_{(\mathbf{X},\mathbf{X}')\sim(\mathcal{U},\mathcal{V})} d_H(\mathbf{X}, \mathbf{X}') \lesssim \epsilon N. \tag{C.1}$$

Then, there exists no robust test capable of distinguishing between $\mathcal{U}$ *and* $\mathcal{V}$ *under contamination level of* $\Omega(\epsilon)$ *while achieving type-I and type-II error probabilities below* 0.1 *simultaneously and uniformly.*

We start with a review of the notion of (ϵ, δ)-differential privacy for convenience.

Definition C.2 ((ϵ, δ)-differential privacy, Dwork et al. (2006)). *For a pair of non-negative real numbers* (ϵ, δ), *a randomized algorithm* $\mathcal{A}$ *defined on the sample space* Ω *is an* (ϵ, δ)-*differentially private algorithm if for any* $S \subset range(\mathcal{A})$ *and any* $X, X' \in \Omega$ *with the Hamming distance* $d_H(X, X') \leq 1$, *we have*

$$\mathbb{P}\left(\mathcal{A}(X) \in S\right) \leq e^\epsilon \mathbb{P}\left(\mathcal{A}(X') \in S\right) + \delta. \tag{C.2}$$

When $\delta = 0$, *the notion is called pure differential privacy; when* $\delta > 0$, *it is referred to as approximate differential privacy.*

Lemma C.3 is motivated by Theorem 11 in Acharya et al. (2018), and shares a nearly identical conclusion with the core lemma.

Lemma C.3. *Assume that there is a coupling between* $\mathcal{U}$ *and* $\mathcal{V}$ *under which that the expectation of the Hamming distance between* $\mathbf{X}$ *and* $\mathbf{X}'$ *is smaller or equal than* D, *i.e.,*

$$\mathbb{E}_{(\mathbf{X},\mathbf{X}')\sim(\mathcal{U},\mathcal{V})} d_H(\mathbf{X}, \mathbf{X}') \leq D. \tag{C.3}$$

Then, if a test $\mathcal{A}$ *to distinguish between* $\mathcal{U}$ *and* $\mathcal{V}$: $\Omega \mapsto \{0, 1\}$ *is a* $(0, \delta)$-*differentially private algorithm, with type-I and type-II error probabilities both below* 0.1, *then we must have* $D = \Omega(\frac{1}{\delta})$.

Proof. By the assumption of the lemma, suppose that $(\mathbf{X}, \mathbf{X}')$ is the random variable from the coupling. We have $\mathbb{P}\left(\mathcal{A}(\mathbf{X}) = 0\right) \geq 0.9$ and $\mathbb{P}\left(\mathcal{A}(\mathbf{X}') = 1\right) \geq 0.9$. By Markov's inequality, we have $\mathbb{P}\left(d_H(\mathbf{X}, \mathbf{X}') > 10D\right) \leq \frac{\mathbb{E}d_H(\mathbf{X},\mathbf{X}')}{10D} \leq 0.1$. From the union bound, we obtain that

$$\mathbb{P}\left(\mathcal{A}(\mathbf{X}) = 0 \cap \mathcal{A}(\mathbf{X}') = 1 \cap d_H(\mathbf{X}, \mathbf{X}') < 10D\right) \geq 0.7. \tag{C.4}$$

By the definition of differential privacy, we can construct a chain from $\mathbf{X}$ to $\mathbf{X}'$ such that each neighbour points differ on at most one coordinate, by which we have

$$|\mathbb{P}\left(\mathcal{A}(\mathbf{X}) = 0\right) - \mathbb{P}\left(\mathcal{A}(\mathbf{X}') = 0\right)| \leq \delta \cdot d_H(\mathbf{X}, \mathbf{X}'). \tag{C.5}$$

It is worth noting that the probability in (C.4) includes two parts of randomness: the randomness of $\mathbf{X}$ and $\mathbf{X}'$ from Ω, and the randomness of $\mathcal{A}$. Now take a realization of $\mathbf{X}$ and $\mathbf{X}'$, denoted by $\mathbf{x}$ and $\mathbf{x}'$, that satisfy (C.4), we have

$$1 - 0.7 \geq \mathbb{P}(\mathcal{A}(\mathbf{x}') = 0) \geq \mathbb{P}(\mathcal{A}(\mathbf{x}) = 0) - \delta \cdot d_H(\mathbf{x}, \mathbf{x}') \geq 0.7 - \delta \cdot 10D, \tag{C.6}$$

from which we obtain that $10D \geq 0.4\frac{1}{\delta}$. The proof is completed. $\qquad\square$

Lemma C.4 shows how to obtain a $(0, \delta)$-differentially private test from a given robust test between $\mathcal{U}$ and $\mathcal{V}$ using the black-box robustness-to-privacy transformation. The techniques are from Canonne et al. (2023) and Hopkins et al. (2023).

Lemma C.4. *Suppose $\mathcal{A}$ is a robust test for the mean testing problem* (1.2) *achieving type-I and type-II errors at most* 0.1 *under a contamination level at most ϵ. Then, for $\delta = \frac{1}{\epsilon N}$, there exists a $(0, \delta)$-differentially private algorithm $\widetilde{\mathcal{A}}$ whose error probabilities are also at most* 0.1 *for any dataset with size N.*

Proof. For any algorithm $\mathcal{A}$ that attempts to distinguish between H_0 and H_1 based on $\mathbf{X}$, define $\mathcal{S}(X, \mathcal{A})$ as the smallest number such that there exists another dataset $\mathbf{X}'$ with $\mathcal{A}(\mathbf{X}') = 1$ and $d_H(\mathbf{X}, \mathbf{X}') = \mathcal{S}(\mathbf{X}, \mathcal{A})$. Using this quantity, we construct a new algorithm $\widetilde{\mathcal{A}}$ that outputs 1 with probability $\max\left(0, 1 - \delta \cdot \mathcal{S}(X, \mathcal{A})\right)$. We now prove that $\widetilde{\mathcal{A}}$ is the desired algorithm.

First, we notice that for any two datasets $\mathbf{X}'$, $\mathbf{X}''$ that satisfy $d_H(\mathbf{X}', \mathbf{X}'') \leq 1$, we always have $|\mathcal{S}(X', \mathcal{A}) - \mathcal{S}(X'', \mathcal{A})| \leq 1$ for any algorithm $\mathcal{A}$. This is foreseen since if we denote $\mathbf{X}^\dagger$ as the specific dataset that satisfies $\mathcal{A}(\mathbf{X}^\dagger) = 1$ and $d_H(\mathbf{X}', \mathbf{X}^\dagger) = \mathcal{S}(\mathbf{X}', \mathcal{A})$, then by the triangular inequality we have $d_H(\mathbf{X}'', \mathbf{X}^\dagger) \leq \mathcal{S}(\mathbf{X}', \mathcal{A}) + 1$, and consequently $\mathcal{S}(\mathbf{X}'', \mathcal{A}) \leq \mathcal{S}(\mathbf{X}', \mathcal{A}) + 1$. In the same way we also have $\mathcal{S}(\mathbf{X}'', \mathcal{A}) \geq \mathcal{S}(\mathbf{X}', \mathcal{A}) - 1$. The statement is proved.

By the statement above we know that the constructed algorithm $\widetilde{\mathcal{A}}$ is a $(0, \delta)$-differentially private algorithm since for any two dataset $\mathbf{X}$ and $\mathbf{X}'$ with $d_H(\mathbf{X}, \mathbf{X}') \leq 1$, we have

$$\begin{aligned} |\mathbb{P}\left(\widetilde{\mathcal{A}}(\mathbf{X}) = 1\right) - \mathbb{P}\left(\widetilde{\mathcal{A}}(\mathbf{X}') = 1\right)| &= |\max\left\{0, 1 - \delta \cdot \mathcal{S}(\mathbf{X}, \mathcal{A})\right\} - \max\left\{0, 1 - \delta \cdot \mathcal{S}(\mathbf{X}', \mathcal{A})\right\}| \\ &\leq |\delta \cdot \mathcal{S}(\mathbf{X}, \mathcal{A}) - \delta \cdot \mathcal{S}(\mathbf{X}', \mathcal{A})| \\ &\leq \delta. \end{aligned}$$

Now assume $\widetilde{\mathbf{X}}$ is from H_0. By the definition of robust testing, with probability at least 0.9, we can arbitrarily modify $\widetilde{\mathbf{X}}$ and obtain the same result as long as the number of modifications does not exceed ϵN. This equivalently says that $\mathcal{S}(\widetilde{\mathbf{X}}, \mathcal{A}) \geq \epsilon N$, which, by the construction of $\widetilde{\mathcal{A}}$, implies that $\widetilde{\mathcal{A}}$ will output 1 with probability 0 in this case. On the other hand, if $\widetilde{\mathbf{X}}$ is from H_1, then with probability at least 0.9, we will have $\mathcal{A}(X) = 1$, in which case $\mathcal{S}(X, \mathcal{A}) = 0$ and consequently $\widetilde{\mathcal{A}}$ will output 1 with probability 1. This shows that $\widetilde{\mathcal{A}}$ is a $(0, \delta)$-differential privacy with error probabilities at most 0.1 under both hypotheses. $\qquad\square$

We are now prepared to prove the core lemma and Theorem 3.5.

Proof of Lemma C.1. Suppose we already have a robust test $\mathcal{A}$ for the mean testing problem 1.2 with the fraction of contamination at most ϵ. Then by Lemma C.4, we can contruct a $(0, \delta)$-differential privacy $\widetilde{\mathcal{A}}$ with errors at most 0.1. Therefore, the constructed algorithm $\widetilde{\mathcal{A}}$ satifies the condition of Lemma C.3. Take $D = \mathbb{E}d_H(\mathbf{X}, \mathbf{X}')$ in Lemma C.3, and by the conclusion of Lemma C.3 we should have $\mathbb{E}d_H(\mathbf{X}, \mathbf{X}') \gtrsim \frac{1}{1/\epsilon N} = \epsilon N$. This necessary condition finishes the proof. $\qquad\square$

Proof of Theorem 3.5. Take $\mathcal{U}$ to be the law of N i.i.d. samples from $\mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$ and $\mathcal{V}$ to be the law of N i.i.d. samples from $\mathcal{N}\left(v, \sigma^2 \mathbf{I}\right)$ with v drawn from the uniform distribution on the set

$$V := \{v \mid v_i = \gamma_i \theta_i, \gamma_i \in \{-1, 1\}, i \geq 1, i \in I\},$$

where θ is the vector from Lemma 3.1 setting $c = \frac{(k_2^\star)^{1/4}\sqrt{\epsilon}}{N^{1/4}}\sigma$, i.e., $\|\theta\|_2 = \frac{(k_2^\star)^{1/4}\sqrt{\epsilon}}{N^{1/4}}\sigma$ and $\|\theta\|_\infty \leq \frac{\sqrt{\epsilon}}{(k_2^\star)^{1/4}N^{1/4}}\sigma$. We are going to compute $d_H(\mathbf{X}, \mathbf{X}')$ using Lemma B.5. Specifically, we compare (in terms of total variation) a distribution of k exchangeable variables generated from $\mathcal{V}$ to a joint distribution of $k - 1$ exchangeable variables generated from $\mathcal{N}\left(v, \sigma^2 \mathbf{I}\right)$ and the last one from $\mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$ for $1 \leq k \leq N$. We approach this by decomposing both $\mathbf{X}$ and $\mathbf{X}'$ by their sequence of conditional probabilities and constructing the optimal coupling between X_k and X_k' conditional on $X_1, \ldots, X_{k-1}$ using Lemma B.5 such that

$$\mathbb{P}\left(X_k \neq X_k' | X_1, \ldots, X_{k-1}\right) = \mathrm{TV}(p(x_k | x_1, \ldots, x_{k-1}), q(x_k)),$$

where $p(x_k | x_1, \ldots, x_{k-1})$ is the density of conditional distribution of the first k coordinates of X' and $q(x_k)$ is the density of standard normal distribution. If we denote the optimal coupling above by $\mathcal{P}_k = \mathcal{P}_k(X_k, X_k' | X_1, ..., X_{k-1})$, it is not hard to verify that

$$\widetilde{\mathcal{P}}(\mathbf{X}, \mathbf{X}') := \mathcal{P}_1(X_1, X_1') \cdot \prod_{i=2}^{N} \mathcal{P}_i(X_k, X_k' | X_1, \ldots, X_{k-1})$$

is a valid coupling between $\mathcal{U}$ and $\mathcal{V}$. By the property of the optimal coupling and the definition of total variation distance, we have under $\widetilde{\mathcal{P}}$ that

$$\mathbb{E}_{\mathbf{X},\mathbf{X}' \sim \widetilde{\mathcal{P}}} \mathbf{1}_{\left\{X_k \neq X_k'\right\}} = \mathbb{E}\mathrm{TV}(p(x_k | x_1, \ldots, x_{k-1}), q(x_k))$$
$$= \mathrm{TV}(p(x_1, \ldots, x_k), q(x_1, \ldots, x_k)),$$

where $p(x_1, \ldots, x_k)$ is the joint distribution of the first k coordinates of X' and $q(x_1, \ldots, x_k) := p(x_1, \ldots, x_{k-1}) \cdot q(x_k)$. Since

$$\mathbb{E}_{(\mathbf{X},\mathbf{X}') \sim \widetilde{\mathcal{P}}} d_H(\mathbf{X}, \mathbf{X}') = \mathbb{E}_{\mathbf{X},\mathbf{X}' \sim \widetilde{\mathcal{P}}} \sum_{i=1}^{N} \mathbf{1}_{\left\{X_i \neq X_i'\right\}},$$

to evaluate $\mathbb{E}_{(\mathbf{X},\mathbb{X}') \sim \widetilde{\mathcal{P}}} d_H(\mathbf{X}, \mathbf{X}')$, it suffices to evaluate

$$\sum_{k=1}^{N} \mathrm{TV}(p(x_1, \ldots, x_k), q(x_1, \ldots, x_k)).$$

We bound each individual term by Pinsker's inequality, where we have

$$2\mathrm{TV}^2(p(x_1,\ldots,x_k),q(x_1,\ldots,x_k)) \leq \mathrm{KL}(p(x_1,\ldots,x_k)|q(x_1,\ldots,x_k))$$

$$\overset{(i)}{=} \mathbb{E}\ln\left(\frac{\frac{1}{2^{|V|}}\sum_{v\in V}\exp\left\{\left(\sum_{i=1}^{k}X_i\right)^{\top}v/\sigma^2\right\}}{\frac{1}{2^{|V|}}\sum_{v\in V}\exp\left\{\left(\sum_{i=1}^{k-1}X_i\right)^{\top}v/\sigma^2\right\}}\right) - \frac{\|v\|_2^2}{2\sigma^2}$$

$$\overset{(ii)}{=} \mathbb{E}\ln\left(\frac{\prod_{j=1}^{d}\cosh\left\{\left(\sum_{i=1}^{k}X_i\right)_j v_j/\sigma^2\right\}}{\prod_{j=1}^{d}\cosh\left\{\left(\sum_{i=1}^{k-1}X_i\right)_j v_j/\sigma^2\right\}}\right) - \frac{\|v\|_2^2}{2\sigma^2},$$

where $(\cdot)_j$ denotes the j-th coordinate, (i) is from that $\|v\|_2^2$ does not depend on the specific choice of $v\in V$ by the construction of V and (ii) is from an important observation that

$$\frac{1}{2^{|V|}}\sum_{v\in V}\exp\left\{\left(\sum_{i=1}^{k}X_i\right)^{\top}v\right\} = \prod_{j=1}^{d}\cosh\left\{\left(\sum_{i=1}^{k}X_i\right)_j v_j\right\}.$$

Therefore, we have

$$\sum_{k=1}^{N}\mathrm{KL}(p(x_1,\ldots,x_k)|q(x_1,\ldots,x_k)) = \sum_{k=1}^{N}\mathbb{E}\left[\ln\left(\frac{\prod_{j=1}^{d}\cosh\left\{\left(\sum_{i=1}^{k}X_i\right)_j v_j/\sigma^2\right\}}{\prod_{j=1}^{d}\cosh\left\{\left(\sum_{i=1}^{k-1}X_i\right)_j v_j/\sigma^2\right\}}\right) - \frac{\|v\|_2^2}{2\sigma^2}\right]$$

$$\overset{(iii)}{=} \sum_{j=1}^{d}\mathbb{E}\ln\left(\cosh\left\{\left(\sum_{i=1}^{N}X_i\right)_j v_j/\sigma^2\right\}\right) - N\cdot\frac{\|v\|_2^2}{2\sigma^2},$$

where (iii) is from telescoping and the exchange of summation and expectation. Using the inequality that $\ln(\cosh x) \leq \frac{x^2}{2}, \forall x\in\mathbb{R}$, which can be easily proved by inspecting the Taylor's expansion of $\ln(\cosh x)$, and the fact that $\left(\sum_{i=1}^{N}X_i\right)_j \sim \mathcal{N}\left(Nv_j, N\sigma^2\right)$ for any j, we can bound the expression above by

$$\sum_{k=1}^{N}\mathrm{KL}(p(x_1,\ldots,x_k)|q(x_1,\ldots,x_k)) \leq \frac{1}{2}\sum_{j=1}^{d}\mathbb{E}\left[\left(\sum_{i=1}^{N}X_i\right)_j v_j/\sigma^2\right]^2 - N\cdot\frac{\|v\|_2^2}{2\sigma^2}$$

$$\leq \frac{1}{2}\sum_{j=1}^{d}\frac{v_j^2}{\sigma^4}\left(N^2v_j^2 + N\sigma^2\right) - N\cdot\frac{\|v\|_2^2}{2\sigma^2}$$

$$= \frac{N^2\|v\|_4^4}{2\sigma^4}.$$

Combining all of the expressions and inequalities above together, we arrive at the following

$$\frac{1}{N}\sum_{k=1}^{N} \mathrm{TV}(p(x_1,\ldots,x_k)|q(x_1,\ldots,x_k)) \lesssim \frac{1}{N}\sum_{k=1}^{N}\sqrt{\mathrm{KL}(p(x_1,\ldots,x_k)|q(x_1,\ldots,x_k))}$$

$$\overset{(i)}{\leq} \sqrt{\frac{1}{N}\cdot\sum_{k=1}^{N}\mathrm{KL}(p(x_1,\ldots,x_k)|q(x_1,\ldots,x_k))}$$

$$\lesssim \frac{\sqrt{N}\,\|v\|_4^2}{\sigma^2},$$

where (i) follows from the Jensen's inequality. Recall that by Lemma 3.1 we have $\|\theta\|_2 = \frac{(k_2^\star)^{1/4}\sqrt{\epsilon}}{N^{1/4}}\sigma$ and $\|\theta\|_\infty \leq \frac{\sqrt{\epsilon}}{(k_2^\star)^{1/4}N^{1/4}}\sigma$. Now that $\theta_i^4 \leq \theta_i^2 \cdot \|\theta\|_\infty^2$ for all $i \in \{1,2,\ldots,d\}$, we know $\|\theta\|_4^4 \leq \|\theta\|_2\,\|\theta\|_\infty \leq \frac{\epsilon}{\sqrt{N}}\sigma^2$. By the definition of V, we know that $\|v\|_4^2 = \|\theta\|_4^2$. Therefore, we have $\frac{\sqrt{N}\|v\|_4^2}{\sigma^2} \leq \frac{\sqrt{N}}{\sigma^2}\cdot\frac{\epsilon}{\sqrt{N}}\sigma^2 \leq \epsilon$. By the core lemma, we have

$$\inf_{\phi\in A_s(0.1)} \sup_{\mu\in K} \sup_{\mathcal{C}} \mathbb{P}_\mu(\phi(\mathcal{C}(\widetilde{\mathbf{X}})) = 0) \geq 0.1.$$

The proof is completed. $\qquad\square$

C.4 Proof of Theorem 3.6

In this section, we prove Theorem 3.6. We reach the result by constructing two mixture distributions based on H_0 and H_1 which are, in fact, identical. We then evaluate the total variation distance between two truncated binomial distributions, which enables us to construct an adaptive adversary, leading high type-II error. Similar techniques also appear in Prasadan and Neykov (2025) and Chen et al. (2017).

Proof of Theorem 3.6. We prove the theorem by several steps.

Step I: control the TV distance. Set $\epsilon' = \epsilon - 1/\sqrt{2N}$. Since $\epsilon \gtrsim 1/\sqrt{N}$, we have $\epsilon \asymp \epsilon'$. Now fix a $\mu_0 \in K$ such that $\|\mu_0\|_2^2 \leq \frac{4\sigma^2\epsilon'^2}{(1-\epsilon')^2}$. Consider two Gaussian measures $\mathbb{P}_0 \sim \mathcal{N}(0,\sigma^2\mathbf{I}_d)$ and $\mathbb{P}_{\mu_0} \sim \mathcal{N}(\mu_0,\sigma^2\mathbf{I}_d)$. By the Pinsker's inequality we have $\mathrm{TV}(\mathbb{P}_0,\mathbb{P}_{\mu_0}) \leq \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\mathbb{P}_0\,\|\,\mathbb{P}_{\mu_0})} = \frac{\epsilon'}{1-\epsilon'}$. Since $r(x) := \frac{x}{1-x}$ is an increasing function for $x \in (0,1)$, there exists some $\epsilon'' \in (0,\epsilon')$ such that $\mathrm{TV}(\mathbb{P}_0,\mathbb{P}_{\mu_0}) = \frac{\epsilon''}{1-\epsilon''}$.

Step II: construct a pair of identical mixtures. Let p_0 and p_{μ_0} be the probability density functions of $\mathbb{P}_0$ and $\mathbb{P}_{\mu_0}$, construct two distributions Q_1, Q_2 with density

$$q_1 = \frac{(p_{\mu_0} - p_0)\mathbf{1}_{\{p_{\mu_0}\geq p_0\}}}{\mathrm{TV}(\mathbb{P}_0,\mathbb{P}_{\mu_0})},\ \ q_2 = \frac{(p_0 - p_{\mu_0})\mathbf{1}_{\{p_0>p_{\mu_0}\}}}{\mathrm{TV}(\mathbb{P}_0,\mathbb{P}_{\mu_0})}. \tag{C.7}$$

Consider the mixture measures

$$R_1 := (1-\epsilon'')\mathbb{P}_0 + \epsilon''Q_1,\ \ R_2 := (1-\epsilon'')\mathbb{P}_{\mu_0} + \epsilon''Q_2. \tag{C.8}$$

It is not hard to verify that actually $R_1 = R_2$, i.e., they represent the same distribution. Consequently $\mathrm{TV}\left(R_1^{\otimes N}, R_2^{\otimes N}\right) = 0$, where $R_j^{\otimes N}, j \in \{1, 2\}$ is the N-fold product of R_j.

Step III: expand the TV distance. Define

$$\binom{[N]}{s} := \{(i_1, \ldots, i_s) : i_1, \ldots, i_s \in \{1, 2, \ldots, N\}, 1 \leq i_1 < i_2 < \cdots < i_s \leq N\}.$$

Further define the quantities

$$S_1^L := \sum_{s=0}^{N\left(\epsilon'' + \frac{1}{\sqrt{2N}}\right)} \sum_{I \in \binom{[N]}{s}} (1 - \epsilon'')^{N-s} \epsilon''^s \prod_{i \in [N] \setminus I} p_0(x_i) \prod_{i \in I} q_1(x_i),$$

$$S_2^L := \sum_{s=0}^{N\left(\epsilon'' + \frac{1}{\sqrt{2N}}\right)} \sum_{I \in \binom{[N]}{s}} (1 - \epsilon'')^{N-s} \epsilon''^s \prod_{i \in [N] \setminus I} p_{\mu_0}(x_i) \prod_{i \in I} q_2(x_i),$$

$$S_1^U := \sum_{s=N\left(\epsilon'' + \frac{1}{\sqrt{2N}}\right)+1}^{N} \sum_{I \in \binom{[N]}{s}} (1 - \epsilon'')^{N-s} \epsilon''^s \prod_{i \in [N] \setminus I} p_0(x_i) \prod_{i \in I} q_1(x_i),$$

$$S_2^U := \sum_{s=N\left(\epsilon'' + \frac{1}{\sqrt{2N}}\right)+1}^{N} \sum_{I \in \binom{[N]}{s}} (1 - \epsilon'')^{N-s} \epsilon''^s \prod_{i \in [N] \setminus I} p_{\mu_0}(x_i) \prod_{i \in I} q_2(x_i),$$

where without loss of generality we assume that all bounds of summation are integers. For S_1^U and S_2^U, we have

$$\frac{1}{2} \int_{\mathbb{R}^d} \left| S_1^U - S_2^U \right| d\mathbf{x} \leq \frac{1}{2} \int_{\mathbb{R}^d} (S_1^U + S_2^U) d\mathbf{x} \leq \mathbb{P}\left(B > N\left(\epsilon'' + \frac{1}{\sqrt{2N}}\right) \right) \overset{(i)}{\leq} e^{-1}.$$

Here B is a binomial random variable with parameters N and ϵ'', and (i) follows from the Hoeffding's inequality.

By expanding the TV distance between the N-fold products of R_1 and R_2 and after some calculation, we have

$$0 = \mathrm{TV}\left(R_1^{\otimes N}, R_2^{\otimes N}\right) \overset{(ii)}{\geq} \frac{1}{2} \int_{\mathbb{R}^d} \left| S_1^L - S_2^L \right| d\mathbf{x} - \frac{1}{2} \int_{\mathbb{R}^d} \left| S_1^U - S_2^U \right| d\mathbf{x},$$

where (ii) is from the definition of TV distance and the triangle inequality. Combining with previous argument, we obtain

$$\frac{1}{2} \int_{\mathbb{R}^d} \left| S_1^L - S_2^L \right| d\mathbf{x} \leq e^{-1}.$$

Step IV: conditional argument. Consider a adversary $\mathcal{C}_{B,\epsilon''}$ with following strategy. Upon observing the original data $\widetilde{\mathbf{X}}$ and its underlying law (which is either $\mathbb{P}_0$ or $\mathbb{P}_{\mu_0}$ in our case), the adversary first samples a number n_c from the conditional law of $B | B \leq N(\epsilon'' + 1/\sqrt{2N})$, then sample n_c replacements from the mixture distribution corresponding to the true underlying law.

Specifically, if $\widetilde{\mathbf{X}}$ is from $\mathbb{P}_0$, then replacements will be from Q_1, otherwise the replacements will be from Q_2. In this way, if we denote $\widetilde{\mathbb{P}}_0$ to be the law of the contaminated samples from $\mathbb{P}_0$ and Q_1, and $\widetilde{\mathbb{P}}_{\mu_0}$ to be the law of the contaminated samples from $\mathbb{P}_{\mu_0}$ and Q_2, then with minor calculation we have

$$\mathrm{TV}(\widetilde{\mathbb{P}}_0, \widetilde{\mathbb{P}}_{\mu_0}) = \frac{\frac{1}{2}\int_{\mathbb{R}^d} \left|S_1^L - S_2^L\right| d\mathbf{x}}{\mathbb{P}\left(B \leq N\left(\epsilon'' + \frac{1}{\sqrt{2N}}\right)\right)} \overset{(iii)}{\leq} \frac{e^{-1}}{1 - e^{-1}} < 1,$$

where (iii) is again from the Hoeffding's inequality and the argument from the step III.

Step V: finish the proof. Now for any $\widetilde{\mu}$ that $\|\widetilde{\mu}\|_2 \leq \|\mu_0\|_2 = \frac{2\sigma\epsilon'}{1-\epsilon'}$. Following a similar argument in the proof of Theorem 3.2, for any $\phi \in A_s$, we have

$$\sup_{\mu \in K, \|\mu\|_2 \geq \rho} \sup_{\mathcal{C}} \mathbb{P}_\mu\left(\phi(\mathcal{C}(\mathbf{X}')) = 0\right) \geq \mathbb{P}_{\widetilde{\mu}}\left(\phi(\mathcal{C}_{B,\epsilon''}(\mathbf{X}')) = 0\right)$$

$$= 1 - \mathbb{P}_{\widetilde{\mu}}\left(\phi(\mathcal{C}_{B,\epsilon''}(\mathbf{X}')) = 1\right) + \mathbb{P}_0\left(\phi(\mathcal{C}_{B,\epsilon''}(\mathbf{X}')) = 1\right)$$
$$- \mathbb{P}_0\left(\phi(\mathcal{C}_{B,\epsilon''}(\mathbf{X}')) = 1\right)$$
$$\geq 1 - \alpha - \left(\mathbb{P}_{\widetilde{\mu}}\left(\phi(\mathcal{C}_{B,\epsilon''}(\mathbf{X}')) = 1\right) - \mathbb{P}_0\left(\phi(\mathcal{C}_{B,\epsilon''}(\mathbf{X}')) = 1\right)\right)$$
$$= 1 - \alpha - \left(\widetilde{\mathbb{P}}_{\widetilde{\mu}}(\phi = 1) - \widetilde{\mathbb{P}}_0(\phi = 1)\right)$$
$$\geq 1 - \alpha - \mathrm{TV}\left(\widetilde{\mathbb{P}}_0, \widetilde{\mathbb{P}}_{\widetilde{\mu}}\right)$$
$$\geq 1 - \alpha - \mathrm{TV}\left(\widetilde{\mathbb{P}}_0, \widetilde{\mathbb{P}}_{\mu_0}\right) = \Theta(1).$$

Taking infimum over $\phi \in A_s$ on both sides completes the proof. $\qquad\square$

C.5 Proof of Theorem 3.8

We start with a recall of the definition of the consistent subset, which plays a central role in the proof.

Definition C.5 (Consistent subset). *A subset $S \subset \mathbf{X} = \{X_1, \ldots, X_N\}$ is called a consistent subset regarding the contamination fraction ϵ and a test $\phi : 2^S \mapsto \{0, 1\}$ if*
(i), $|S| \geq (1 - \epsilon)N$,
(ii), $\phi(S') = \phi(S)$ for any $S' \subset S$ with $|S'| \geq (1 - 2\epsilon)N$.

We are now ready to prove Theorem 3.8. To provide a high-level outline, we first establish the high-probability existence of a consistent subset with respect to ϕ_e implied by E_e defined in (C.9). This is done by expanding the ℓ_2 norm of the projection vector in E_e and bounding each component of the expansion. The rationale of ϕ_e then follows from the nature of the consistent subset. In the proof, $k^\star \in [1, d]$[1] is any fixed dimension that does not necessary to be $k_1^\star$ or $k_2^\star$, and $P^\star$ is its corresponding projection operator determined by the Kolmogorov widths. The pseudo code of ϕ_e is attached in Appendix A.

Proof of Theorem 3.8. We will first prove the fact that for a properly fixed error tolerance α, the consistent subset exists with probability at least $1 - \alpha$. Specifically, consider the following χ^2-type

---

[1] $1 \leq k^\star < \infty$ if $K \subset \ell_2$.

event

$$E_e := \left\{ \left\| P^\star \sum_{i \in T} X_i \right\|_2^2 - k^\star |T| \sigma^2 \geq c |T|^2 E^2(\epsilon, K, N, \sigma^2) \right\}, \tag{C.9}$$

where T is any subset of $\mathbf{X}$, and $c \in \mathbb{R}$ is some universal, properly chosen constant. On the subset T, the test ϕ_e rejects H_0 when E_e holds and fails to reject otherwise.

For the rest of the proof, let us first focus on $\widetilde{\mathbf{X}}$ and any its subset $\widetilde{\mathbf{X}}_{S'}$ with $|S'| \geq (1 - 2\epsilon)N$. We will prove that with $\widetilde{\mathbf{X}}$ is consistent with all such S' if (3.4) is satisfied. The key to the proof is carefully analyzing the ℓ_2 norm of the sum of the projected original samples

$$\left\| P^\star \sum_{i \in S'} \widetilde{X}_i \right\|_2^2 = \underbrace{\left\| P^\star \sum_{i \in S'} (\widetilde{X}_i - \mu) \right\|_2^2}_{:=\mathrm{I}} + \underbrace{|S'|^2 \|P^\star \mu\|_2^2}_{:=\mathrm{II}} + \underbrace{2 |S'| \sum_{i \in S'} \left[P^\star (\widetilde{X}_i - \mu) \right]^\top \mu}_{:=\mathrm{III}}. \tag{C.10}$$

We attempt to show that the parts I and III, which are ℓ_2 norm of the centralized projected vectors and the crossing term, are both asymptotically less than the part II under the condition (3.4), which vanishes under the null.

Part I. By representing $S' = [N] \setminus ([N] \setminus S')$, we have

$$\mathrm{I} = \underbrace{\left\| P^\star \sum_{i \in [N]} (\widetilde{X}_i - \mu) \right\|_2^2}_{:=\mathrm{I}_1} + \underbrace{\left\| P^\star \sum_{i \in [N] \setminus S'} (\widetilde{X}_i - \mu) \right\|_2^2}_{:=\mathrm{I}_{2,S'}} - \underbrace{2 \sum_{i \in [N]} (\widetilde{X}_i - \mu)^\top P^\star \sum_{i \in [N] \setminus S'} (\widetilde{X}_i - \mu)}_{:=\mathrm{I}_{3,S'}}.$$

For I_1, from the property of the Gaussian distribution, we know $\mathrm{I}_1 \sim N\sigma^2 \cdot \chi^2_{k^\star}$. Therefore $\mathbb{E}\mathrm{I}_1 = k^\star N \sigma^2$, $\mathrm{Var}\,(\mathrm{I}_1) = 2k^\star N^2 \sigma^4$. By the Chebyshev's inequality we know that

$$\left| \mathrm{I}_1 - k^\star N \sigma^2 \right| \leq \sqrt{\frac{2k^\star}{\alpha}} N\sigma^2 := t_{\mathrm{I}_1}.$$

with probability greater than $1 - \alpha$.

For $\mathrm{I}_{2,S'}$, since we allow S' to be any subset of S, refined analysis is needed. Using the Hanson-Wright inequality B.10 and observing that $\|P^\star\|_F = \sqrt{k}, \|P^\star\|_2 \leq 1$, for each fixed $S' \subset S$, we have

$$\mathbb{P}\left(\left| \mathrm{I}_2 - k^\star \left| [N] \setminus S' \right| \sigma^2 \right| \geq t \right) \leq 2 \exp \left\{ -c \min \left\{ \frac{t^2}{k^\star \left| [N] \setminus S' \right|^2 \sigma^4}, \frac{t}{\left| [N] \setminus S' \right| \sigma^2} \right\} \right\}.$$

Since $|S'| \geq (1 - 2\epsilon)N$, we have $\sum_{i=0}^{2\epsilon N} \binom{N}{i} = \left(\frac{e}{2\epsilon} \right)^{2\epsilon N}$ possible $[N] \setminus S'$ by Lemma B.18. Using the union bound, we have $\mathrm{I}_{2,S'} - k^\star \left| [N] \setminus S' \right| \sigma^2 \geq t$ holds for some S' with probability less than

$$\sum_{S'} \exp \left\{ -c \min \left\{ \frac{t^2}{k^\star \left| [N] \setminus S' \right|^2 \sigma^4}, \frac{t}{\left| [N] \setminus S' \right| \sigma^2} \right\} \right\} \leq \left(\frac{e}{2\epsilon} \right)^{2\epsilon N} \cdot \exp \left\{ -c \min \left\{ \frac{t^2}{k^\star \epsilon^2 N^2 \sigma^4}, \frac{t}{\epsilon N \sigma^2} \right\} \right\}.$$

If we select t_{I_2} to be

$$t_{\mathrm{I}_2} = \max \left\{ \sqrt{\frac{4 \left[2\epsilon N \ln \left(\frac{e}{2\epsilon} \right) + \ln \left(\frac{2}{\alpha} \right) \right] k^\star \epsilon^2 N^2 \sigma^4}{c}}, \frac{2 \left[2\epsilon N \ln \left(\frac{e}{\epsilon} \right) + \ln \left(\frac{2}{\alpha} \right) \right] \epsilon N \sigma^2}{c} \right\}$$

then it can be verified that $\left| \mathrm{I}_{2,S'} - k^\star \left|[N]\backslash S'\right| \sigma^2 \right| \le t_{\mathrm{I}_2}$ holds for all S' with probability greater than $1 - \alpha$.

For $\mathrm{I}_{3,S'}$, we use the following argument of conditional probability to effectively bound its tail. First, observe that the joint distribution of $\sum_{i \in [N]\backslash S'} (\widetilde{X}_i - \mu)$ and $\sum_{i \in [N]} (\widetilde{X}_i - \mu)$ is a multivariate Gaussian

$$\left(\sum_{i \in [N]\backslash S'} (\widetilde{X}_i - \mu)^\top, \sum_{i \in [N]} (\widetilde{X}_i - \mu)^\top \right) \sim \mathcal{N} \left(0, \begin{pmatrix} |[N]\backslash S'| \, \mathbf{I}_d & |[N]\backslash S'| \, \mathbf{I}_d \\ |[N]\backslash S'| \, \mathbf{I}_d & N\mathbf{I}_d \end{pmatrix} \sigma^2 \right).$$

By the property of multivariate Gaussian distribution, conditional on the event $\left\{ \sum_{i \in [N]} (\widetilde{X}_i - \mu) = x \right\}$, the conditional distribution of $\sum_{i \in [N]\backslash S'} (\widetilde{X}_i - \mu)$ is

$$\left[\sum_{i \in [N]\backslash S'} (\widetilde{X}_i - \mu) \,|\, \sum_{i \in [N]} (\widetilde{X}_i - \mu) = x \right] \sim \mathcal{N} \left(\frac{|[N]\backslash S'|}{N} x, \sigma^2 \left(|[N]\backslash S'| - \frac{|[N]\backslash S'|^2}{N} \right) \mathbf{I}_d \right).$$

By the maximal inequality for (sub-)Gaussian random variable that $\mathbb{P} \left(\max_{1 \le i \le N} \widetilde{X}_i \ge \sqrt{2\sigma^2 (\ln N + t)} \right) \le e^{-t}$, and conditional on the event $\left\{ \sum_{i \in [N]} (\widetilde{X}_i - \mu) = x \right\}$, we have the probability of the following event

$$\left\{ \mathrm{I}_{3,S'} \ge \frac{|[N]\backslash S'|}{N} x^\top P^\star x + \sqrt{2 x^\top P^\star x \left(|[N]\backslash S'| - \frac{|[N]\backslash S'|^2}{N} \right) \sigma^2 \cdot \left(2\epsilon N \ln \left(\frac{e}{2\epsilon} \right) + t \right)} \text{ for some } S' \subset S \right\}$$

less than e^{-t}. We also need the following basic inequality of conditional probability for any random variables X and Y

$$\begin{aligned}
\mathbb{P}\left(Y \ge a\right) &= \mathbb{E}\mathbf{1}_{\{Y \ge a\}} \\
&= \mathbb{E}\left[\mathbf{1}_{\{Y \ge a\}} \cdot \left(\mathbf{1}_{\{f(X) \ge a\}} + \mathbf{1}_{\{f(X) < a\}} \right) \right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{Y \ge a\}} \cdot \mathbf{1}_{\{f(X) \ge a\}} \,|\, X \right] + \mathbb{E}\left[\mathbf{1}_{\{Y \ge a\}} \cdot \mathbf{1}_{\{f(X) < a\}} \,|\, X \right] \right] \\
&\le \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{f(X) \ge a\}} \,|\, X \right] + \mathbb{E}\left[\mathbf{1}_{\{Y \ge f(X)\}} \cdot \mathbf{1}_{\{f(X) < a\}} \,|\, X \right] \right] \\
&\le \mathbb{P}\left(f(X) \ge a \right) + \mathbb{E}\left[\mathbb{P}\left(Y \ge f(X) \,|\, X \right) \right].
\end{aligned}$$

Now observe that $|[N]\backslash S'| \le 2\epsilon N$ for all S', and take $Y = \max_{S'} \{\mathrm{I}_{3,S'}\}$, $X = \sum_{i \in [N]} (\widetilde{X}_i - \mu)$, $f(x) = g(x^\top P^\star x)$ where $g(x) = \frac{|[N]\backslash S'|}{N} x + \sqrt{4x \cdot \epsilon N \left(\epsilon N \ln \left(\frac{e}{\epsilon} \right) + t \right)}$ and $a = g(Nk^\star \sigma^2 + \sqrt{\frac{4k^\star}{\alpha} N\sigma^2})$. By the argument above and the fact that $g(\cdot)$ is an increasing function on the positive real line, we know

$$\mathbb{P}\left(f(X) \ge a \right) = \mathbb{P}\left(x^\top P^\star x \ge Nk^\star \sigma^2 + \sqrt{\frac{4k^\star}{\alpha} N\sigma^2} \right) \le \frac{\alpha}{2}.$$

From the maximal inequality above, we also have $\mathbb{E}\left[\mathbb{P}\left(Y \geq f(X) \mid X\right)\right] \leq \mathbb{E}e^{-t} = e^{-t}$. Taking $t = \ln\left(\frac{2}{\alpha}\right)$, and combining the results above, we conclude that with probability greater than $1 - \alpha$ the following event holds for all S'.

$$
\left\{ \mathrm{I}_{3,S'} \leq \left|[N]\backslash S'\right| k^{\star}\sigma^2 + \left|[N]\backslash S'\right| \sqrt{\frac{4k^{\star}}{\alpha}}\sigma^2 + \sqrt{8\left(Nk^{\star}\sigma^2 + \sqrt{\frac{4k^{\star}}{\alpha}}N\sigma^2\right)\epsilon N \left(2\epsilon N \ln\left(\frac{e}{2\epsilon}\right) + \ln\left(\frac{2}{\alpha}\right)\right)\sigma^2} \right\}.
$$

By the symmetrical nature of $\mathrm{I}_{3,S'}$, same technique and argument can also be applied to the probability of the lower tail of $\mathrm{III}_{3,S'}$.

Part III. We rewrite the part III as

$$
\mathrm{III} = \underbrace{\left|S'\right| \sum_{i \in [N]} (\widetilde{X}_i - \mu)^\top P^{\star}\mu}_{\mathrm{III}_{1,S'}} - \underbrace{\left|S'\right| \sum_{i \in [N]\backslash S'} (\widetilde{X}_i - \mu)^\top P^{\star}\mu}_{\mathrm{III}_{2,S'}}.
$$

Dealing with $\mathrm{III}_{1,S'}$ is relatively easier. Observe that $\mathrm{III}_{1,S'}$ actually follows a Gaussian distribution with $\mathbb{E}\mathrm{III}_{1,S'} = 0$ and $\mathrm{Var}\left(\mathrm{III}_{1,S'}\right) = \left|S'\right|^2 N\sigma^2 \|P^{\star}\mu\|_2^2 \leq N^3\sigma^2 \|P^{\star}\mu\|_2^2$. By the Chebyshev's inequality, we have

$$
\mathbb{P}\left(\left|\mathrm{III}_{1,S'}\right| \leq \sqrt{\frac{N^3\sigma^2 \|P^{\star}\mu\|_2^2}{\alpha}}\right) \geq 1 - \alpha.
$$

For $\mathrm{III}_{2,S'}$, since the summation involves S', we again need to exploit the union bound (or equivalently the maximal inequality). $\mathrm{III}_{2,S'}$ also follows a Gaussian distribution with $\mathbb{E}\mathrm{III}_{2,S'} = 0$ and $\mathrm{Var}\left(\mathrm{III}_{2,S'}\right) = \left|S'\right|^2 \left|[N]\backslash S'\right| \|P^{\star}\mu\|_2^2 \sigma^2 \leq 2\epsilon N^3 \|P^{\star}\mu\|_2^2 \sigma^2$. By the Chebyshev's inequality and the union bound, we know that the following holds.

$$
\mathbb{P}\left(\left|\mathrm{III}_{2,S'}\right| \leq \sqrt{2\epsilon N^3 \|P^{\star}\mu\|_2^2 \sigma^2} \cdot \sqrt{2\epsilon N \ln\left(\frac{e}{2\epsilon}\right) + \ln\left(\frac{1}{\alpha}\right)}, \forall S' \subset S\right) \geq 1 - \alpha.
$$

In summary, each tail behaviour of the term above consists of two parts: the mean of the term and the variation around the mean. To clearly display the results above, the following table can be helpful. To complete the proof, we finally need to compare the variation of each term above with the part II.

I_1. Require: $\left|S'\right|^2 \|P^{\star}\mu\|_2^2 \gtrsim \sqrt{\frac{2k^{\star}}{\alpha}}N\sigma^2$. This is equivalent to $\|P^{\star}\mu\|_2^2 \gtrsim \frac{\sqrt{k^{\star}}}{N}\sigma^2$.

$\mathrm{I}_{2,S'}$. Require: $\left|S'\right|^2 \|P^{\star}\mu\|_2^2 \gtrsim \max\left\{\sqrt{\frac{4\left[2\epsilon N \ln\left(\frac{e}{\epsilon}\right) + \ln\left(\frac{2}{\alpha}\right)\right]k^{\star}\epsilon^2 N^2\sigma^4}{c}}, \frac{2\left[2\epsilon N \ln\left(\frac{e}{\epsilon}\right) + \ln\left(\frac{2}{\alpha}\right)\right]\epsilon N\sigma^2}{c}\right\}$. This is equivalent to $\|P^{\star}\mu\|_2^2 \gtrsim \sqrt{\frac{\epsilon^3 \ln\left(\frac{1}{\epsilon}\right)k^{\star}\sigma^4}{N}}$ and $\|P^{\star}\mu\|_2^2 \gtrsim \epsilon^2 \ln\left(\frac{1}{\epsilon}\right)\sigma^2$.

$\mathrm{I}_{3,S'}$. Require: $\left|S'\right|^2 \|P^{\star}\mu\|_2^2 \gtrsim \left|[N]\backslash S'\right| \sqrt{\frac{4k^{\star}}{\alpha}}\sigma^2 \vee \sqrt{8\left(Nk^{\star} + \sqrt{\frac{4k^{\star}}{\alpha}}N\right)\epsilon N \left(2\epsilon N \ln\left(\frac{1}{2\epsilon}\right) + \ln\left(\frac{2}{\alpha}\right)\right)\sigma^2}$.

This is equivalent to $\|P^{\star}\mu\|_2^2 \gtrsim \frac{\epsilon\sqrt{k^{\star}}}{N}\sigma^2$ and $\|P^{\star}\mu\|_2^2 \gtrsim \sqrt{\frac{\epsilon^2 \ln\left(\frac{1}{\epsilon}\right)k^{\star}\sigma^4}{N}}$.

$\mathrm{III}_{1,S'}$. Require: $\left|S'\right|^2 \|P^{\star}\mu\|_2^2 \gtrsim \sqrt{\frac{N^3\sigma^2 \|P^{\star}\mu\|_2^2}{\alpha}}$. This is equivalent to $\|P^{\star}\mu\|_2^2 \gtrsim \frac{\sigma^2}{N}$.

Terms	Mean	Variation	Tail
I_1	$k^\star N\sigma^2$	$\sqrt{\frac{2k^\star}{\alpha}}N\sigma^2$	α
$\mathrm{I}_{2,S'}$	$k^\star\,\lvert[N]\backslash S'\rvert\,\sigma^2$	$\max\left\{\sqrt{\frac{4\left[2\epsilon N\ln\left(\frac{e}{2\epsilon}\right)+\ln\left(\frac{2}{\alpha}\right)\right]k^\star\epsilon^2 N^2\sigma^4}{c}},\ \frac{2\left[2\epsilon N\ln\left(\frac{e}{\epsilon}\right)+\ln\left(\frac{2}{\alpha}\right)\right]\epsilon N\sigma^2}{c}\right\}$	α
$\mathrm{I}_{3,S'}$	$k^\star\,\lvert[N]\backslash S'\rvert\,\sigma^2$	$\left[\lvert[N]\backslash S'\rvert\sqrt{\frac{4k^\star}{\alpha}}+\sqrt{8\left(Nk^\star+\sqrt{\frac{4k^\star}{\alpha}}N\right)\epsilon N\left(2\epsilon N\ln\left(\frac{e}{2\epsilon}\right)+\ln\left(\frac{2}{\alpha}\right)\right)}\right]\sigma^2$	2α
$\mathrm{III}_{1,S'}$	0	$\sqrt{\frac{N^3\sigma^2\lVert P^\star\mu\rVert_2^2}{\alpha}}$	α
$\mathrm{III}_{2,S'}$	0	$\sqrt{2\epsilon N^3\lVert P^\star\mu\rVert_2^2\sigma^2}\cdot\sqrt{2\epsilon N\ln\left(\frac{e}{2\epsilon}\right)}$	α

Table 1: corresponding means and variations of the terms from (C.10).

$\mathrm{III}_{2,S'}$. Require: $\lvert S'\rvert^2\lVert P^\star\mu\rVert_2^2\gtrsim\sqrt{2\epsilon N^3\lVert P^\star\mu\rVert_2^2\sigma^2}\cdot\sqrt{2\epsilon N\ln\left(\frac{e}{2\epsilon}\right)+\ln\left(\frac{1}{\alpha}\right)}$. This is equivalent to $\lVert P^\star\mu\rVert_2^2\gtrsim\max\left\{\epsilon^2\ln\left(\frac{1}{\epsilon}\right)\sigma^2,\frac{\epsilon\sigma^2}{N}\right\}$.

In the argument above, we have used the assumption that $(1-2\epsilon)N\leq\lvert S'\rvert\leq N$ and therefore $0\leq\lvert[N]\backslash S'\rvert\leq 2\epsilon N$. Summarizing the table and the argument above, it is required that

$$\lVert P^\star\mu\rVert_2^2\gtrsim\sigma^2\max\left\{\frac{\sqrt{k^\star}}{N},\epsilon^2\ln\left(\frac{1}{\epsilon}\right),\sqrt{\frac{\epsilon^2\ln\left(\frac{1}{\epsilon}\right)k^\star}{N}}\right\}.$$

Finally, by the Pythagorean theorem, we have $\lVert P^\star\mu\rVert_2^2=\rho^2-\lVert\mu-P^\star\mu\rVert_2^2\geq\rho^2-D_{k^\star}^2(K)$. Therefore, in order that $\lVert P_k^\star\mu\rVert_2^2$ satifies the condition above, it suffices that

$$\rho^2\gtrsim D_{k^\star}^2(K)+\sigma^2\max\left\{\frac{\sqrt{k^\star}}{N},\epsilon^2\ln\left(\frac{1}{\epsilon}\right),\sqrt{\frac{\epsilon^2\ln\left(\frac{1}{\epsilon}\right)k^\star}{N}}\right\}=E_{\mathrm{raw}}^2(\epsilon,K,N,\sigma^2).$$

In fact, we have proved that there exists consistent subsets within $\mathbf{X}$ (at least $\mathbf{X}_{[N]\backslash C}$) for ϕ_e with high probability. The remaining part goes naturally. For the observed dataset $\mathbf{X}$, we scan through all $S\subset[N]$ with $\lvert S\rvert\geq(1-\epsilon)N$ and perform ϕ_e on $\mathbf{X}_S$ and $\mathbf{X}_{S'}$ with $S'\subset S,\lvert S\rvert'\geq(1-2\epsilon)N$ to check if the subset is consistent. The proof above guarantees that we can eventually find a consistent subset S_0 (with high probability), though $\mathbf{X}_{S_0}$ may contain corrupted samples. However, once found, we simply perform the test ϕ_e on $\mathbf{X}_{S_0}$ and accept or reject H_0 depending on the testing result. By the assumption on the cardinality, we know $\lvert S_0\cap[N]\backslash C\rvert\geq(1-2\epsilon)N$. The proof and the assumption above show that with probability greater than $1-\alpha$, we have $\phi_e(\mathbf{X}_{S_0})=\phi_e(\mathbf{X}_{S_0\cap[N]\backslash C})=\phi_e(\mathbf{X}_{[N]\backslash C})=\phi_e(\widetilde{\mathbf{X}})$. Therefore, testing on $\mathbf{X}_{S_0}$ is as testing on $\widetilde{\mathbf{X}}$. Since ϕ_e itself is a valid test for the uncontaminated testing problem (1.1), the proof is completed. $\square$

C.6 Proof of Theorem 3.11

Motivated by Canonne et al. (2023), we demonstrate the rationale of Algorithm 8 in this section. Same as the proof of Theorem 3.8, $k^\star$ is any fixed dimension and $P^\star$ is its corresponding projection operator. For the conciseness of the notation, we define $\widetilde{Y}_i:=P^\star\widetilde{X}_i$, $Y_i:=P^\star X_i$, and $\nu:=P^\star\mu$. We

assume without loss of generality that $\widetilde{Y}_i, Y_i \in \mathbb{R}^{k^\star}$ and $\mathrm{Var}\left(\widetilde{Y}_i\right) = \mathbf{I}_{k^\star}$ with the fact in mind that any projected multivariate Gaussian random variable can be regarded as embedded into a lower dimension subspace with covariance matrix $\mathrm{diag}\{(1, ..., 1, 0, \ldots, 0)\}$ by rotating the axes. We also assume $\sigma = 1$ in this section (one can scale the results by σ anytime if needed).

To clarify the structure of this section and highlight how the definitions, lemmas, and algorithms fit together, we provide the following roadmap. We depart from the definition of regularity, and arrive at the final complete testing procedure, Algorithm 8. The pseudo code is provided in Appendix A.
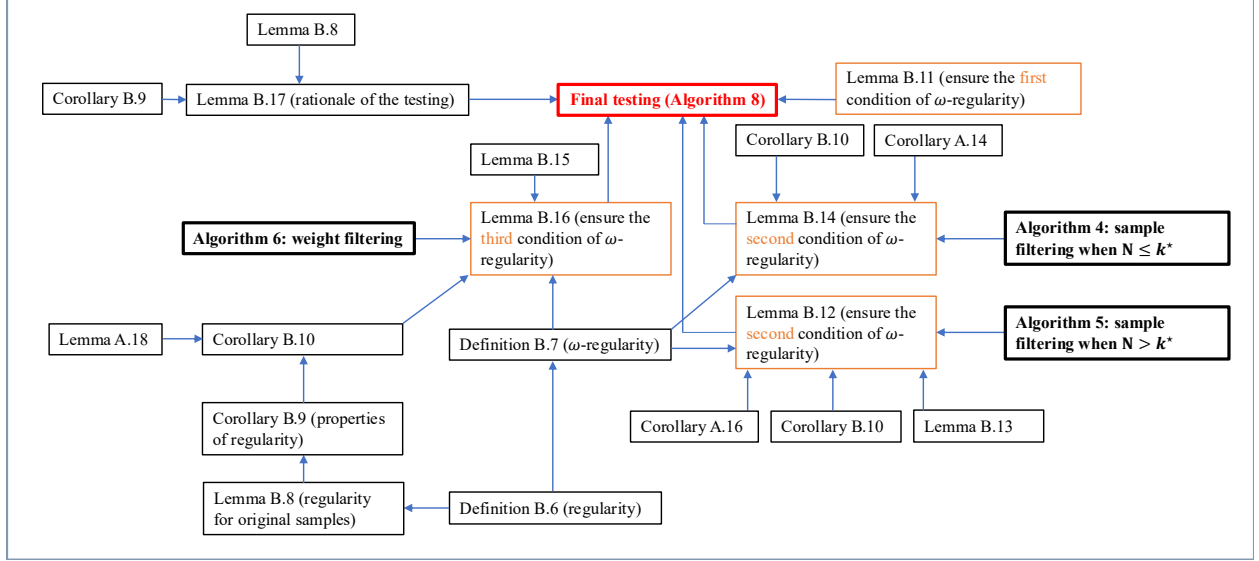

Figure 5: roadmap of this section.

We start with the definition of regularity and ω-regularity, which controls the behaviour of the ℓ_2 norm of the individual and sum, and the inner product.

Definition C.6 (regularity). $\mathbf{Y} = \{Y_1, \ldots, Y_N\}$ *is said to be* $(\epsilon, \beta_1, \beta_2)$-*regular if for all subsets* $S \subset [N]$ *with* $|S| \leq \epsilon N$, *we have the following properties:*

(i), $\left| \sum_{i \in S} \|Y_i\|_2^2 - |S|\, k^\star \right| \leq c\beta_1,$

(ii), $\left| \left\| \sum_{i \in S} Y_i \right\|_2^2 - |S|\, k^\star \right| \leq c\beta_2,$ *and*

(iii), $\left| \left\langle \sum_{i \in S} Y_i, \sum_{j \in [N]} Y_j \right\rangle - |S|\, k^\star \right| \leq c\sqrt{N}\beta_1.$

ω-regularity is a natural generalization of the definition above, which involves the interaction of a weight vector ω with $\mathbf{Y}$.

Definition C.7 (ω-regularity). *Given a weight vector* $\omega = (\omega_1, \ldots, \omega_N)^\top$, $\mathbf{Y}$ *is said to be* $(\epsilon, \beta_1, \beta_2)$-*regular if for all subsets* $S \subset [N]$ *with* $|S| \leq \epsilon N$, *we have the following properties:*

59

(i), $\left| \sum_{i \in S} \|Y_i\|_2^2 - |S| \, k^\star \right| \leq c \beta_1,$

(ii), $\left| \left\| \sum_{i \in S} \sqrt{\omega_i} Y_i \right\|_2^2 - \|\omega_S\|_1 \, k^\star \right| \leq c \beta_2, \text{ and}$

(iii), $\left| \left\langle \sum_{i \in S} \sqrt{\omega_i} Y_i, \sum_{j \in [N]} \sqrt{\omega_j} Y_j \right\rangle - \|\omega_S\|_1 \, k^\star \right| \leq c \sqrt{N} \beta_1.$

Recall that $\omega_S \in \mathbb{R}^N$ means the restriction of ω on the set S.

C.7 immediately recovers C.6 if $\omega = \mathbf{1}_S$.

The next lemma states that any uncontaminated $\mathbf{Y}$ from $\mathcal{N}(\nu, \mathbf{I}_{k^\star})$ is in fact $(\epsilon, \beta_1, \beta_2)$-regular with high probability and specified β_1 and β_2 below depending on $\epsilon, \|\nu\|_2, N, k^\star$.

Lemma C.8. *Assume* $Y_1, \ldots, Y_N$ *are i.i.d. samples from* $\mathcal{N}(\nu, \mathbf{I}_{k^\star})$, *then with probability greater than* $1 - \alpha$, $\mathbf{Y} = \{Y_1, \ldots, Y_N\}$ *is* $(\epsilon, \beta_1, \beta_2)$-regular with

$$\beta_1 = \epsilon N \left(\sqrt{k^\star} + \sqrt{N} \|\nu\|_2 \right) \sqrt{\ln\left(\frac{N}{\alpha}\right)} + \epsilon N \ln\left(\frac{N}{\alpha}\right) + \epsilon N \sqrt{N} \|\nu\|_2^2,$$

$$\beta_2 = \epsilon N \sqrt{\epsilon N k^\star \ln\left(\frac{1}{\epsilon}\right) + (\epsilon N)^2 \ln\left(\frac{1}{\epsilon}\right)} + \|\nu\|_2 (\epsilon N)^2 \sqrt{\ln\left(\frac{1}{\epsilon}\right)} + \|\nu\|_2^2 (\epsilon N)^2,$$

(C.11)

given the condition that $N \geq \frac{\ln(1/\alpha)}{\epsilon \ln(1/\epsilon)}$.

Proof. The last condition is of no importance since it only depends on the predetermined constants α, ϵ, and the fact that $\epsilon \geq \frac{1}{N}$. We prove this lemma by direct check of the definition.

Check of (i). By adding and subtracting ν, for a single sample Y, we have

$$\|Y\|_2^2 - k^\star = \|Y - \nu\|_2^2 - k^\star + \|\nu\|_2^2 + 2\nu^\top (Y - \nu).$$

By assumption of the distribution of Y, $\|Y - \mu\|_2^2$ follows a chi-square distribution with degree of freedom $k^\star$. By standard concentration bound, we have:

$$\mathbb{P}\left(\left| \|Y - \nu\|_2^2 - k^\star \right| \geq t \right) \leq 2 \exp\left\{ -C \min\left\{ \frac{t^2}{k^\star}, t \right\} \right\}.$$

Now set the probability to be $\frac{\alpha}{6N}$, we have $t = c \max\left\{ \sqrt{k^\star \ln\left(\frac{N}{\alpha}\right)}, \ln\left(\frac{N}{\alpha}\right) \right\} \leq c \left(\sqrt{k^\star \ln\left(\frac{N}{\alpha}\right)} + \ln\left(\frac{N}{\alpha}\right) \right)$ for some appropriate c.

$\nu^\top (Y - \nu)$ follows a Gaussian distribution with mean zero and variance $\|\nu\|_2^2$. By the concentration bound for the Gaussian, we have

$$\mathbb{P}\left(\left| 2\nu^\top (Y - \nu) \right| \geq t \right) \leq 2 \exp\left\{ -\frac{t^2}{8 \|\nu\|_2^2} \right\}.$$

Setting the probability to be $\frac{\alpha}{6N}$ as the same, we find $t = c \sqrt{\|\nu\|_2^2 \ln\left(\frac{N}{\alpha}\right)}$. Combining the results above, we have

$$\left| \|Y_i\|_2^2 - k^\star \right| \leq c \left(\left(\sqrt{k^\star} + \|\nu\|_2 \right) \sqrt{\ln\left(\frac{N}{\alpha}\right)} + \ln\left(\frac{N}{\alpha}\right) + \|\nu\|_2^2 \right)$$

60

with probability greater than $1 - \frac{\alpha}{3N}$. Consequently, this holds for all $Y_i, 1 \leq i \leq N$ with probability greater than $1 - \frac{\alpha}{3}$ by the union bound. Therefore, we have

$$\left| \sum_{i \in S} \|Y_i\|_2^2 - |S| \, k^\star \right| \leq c \left[\epsilon N \sqrt{(k^\star + \|\nu\|_2^2) \ln\left(\frac{N}{\alpha}\right)} + \epsilon N \ln\left(\frac{N}{\alpha}\right) + \epsilon N \|\nu\|_2^2 \right] \tag{C.12}$$

Check of (ii). By the same trick, we have $\left\| \sum_{i \in S} Y_i \right\|_2^2 = \left\| \sum_{i \in S} (Y_i - \nu) \right\|_2^2 + \|\nu\|_2^2 |S|^2 + 2 |S| \, \nu^\top \sum_{i \in S} (Y_i - \nu)$. Similarly, we have $\left\| \sum_{i \in S} (Y_i - \nu) \right\|_2^2 = |S| \left\| \frac{1}{\sqrt{|S|}} \sum_{i \in S} (Y_i - \nu) \right\|_2^2$, and

$$\mathbb{P} \left(\left| \left\| \frac{1}{\sqrt{|S|}} \sum_{i \in S} (Y_i - \nu) \right\|_2^2 - k^\star \right| \geq t \right) \leq 2 \exp \left\{ -c \min \left\{ \frac{t^2}{k^\star}, t \right\} \right\}.$$

By Lemma B.18, we have at most $\left(\frac{e}{\epsilon}\right)^{\epsilon N}$ different possible S in total. Setting the probability to be $\frac{\alpha}{6} \left(\frac{\epsilon}{e}\right)^{\epsilon N}$ and solving for t, we obtain

$$t = c \max \left\{ \sqrt{k^\star \left[\ln\left(\frac{1}{\alpha}\right) + \epsilon N \ln\left(\frac{1}{\epsilon}\right) \right]}, \ln\left(\frac{1}{\alpha}\right) + \epsilon N \ln\left(\frac{1}{\epsilon}\right) \right\}$$

$$\leq c \left[\sqrt{k^\star \left[\ln\left(\frac{1}{\alpha}\right) + \epsilon N \ln\left(\frac{1}{\epsilon}\right) \right]} + \ln\left(\frac{1}{\alpha}\right) + \epsilon N \ln\left(\frac{1}{\epsilon}\right) \right]$$

$$\leq c \left[\sqrt{k^\star \epsilon N \ln\left(\frac{1}{\epsilon}\right)} + \epsilon N \ln\left(\frac{1}{\epsilon}\right) \right],$$

where the last step use the condition that $N \geq \frac{\ln(1/\alpha)}{\epsilon \ln(1/\epsilon)}$.

For $\nu^\top \sum_{i \in S} (Y_i - \nu)$, by the same concentration bound and technique, we will obtain

$$\mathbb{P} \left(\left| \nu^\top \sum_{i \in S} (Y_i - \nu) \right| \geq c \|\nu\|_2 \sqrt{\epsilon N} \cdot \sqrt{\epsilon N \ln\left(\frac{1}{\epsilon}\right)} \right) \leq \frac{\alpha}{6} \left(\frac{\epsilon}{e}\right)^{\epsilon N}.$$

Combining the results together, and using the union bound over all possible S, we have

$$\left| \left\| \sum_{i \in S} Y_i \right\|_2^2 - |S| \, k^\star \right| \leq c \left[\epsilon N \sqrt{\epsilon N k^\star \ln\left(\frac{1}{\epsilon}\right)} + (\epsilon N)^2 \ln\left(\frac{1}{\epsilon}\right) + \|\nu\|_2 (\epsilon N)^2 \sqrt{\ln\left(\frac{1}{\epsilon}\right)} + \|\nu\|_2^2 (\epsilon N)^2 \right]$$

with probability greater than $1 - \frac{\alpha}{3}$. The RHS is exactly β_2 in (C.11).

Check of (iii). Finally, we are going to evaluate the term $\left\langle \sum_{i \in S} Y_i, \sum_{j \in [N]} Y_j \right\rangle$. Since this is equal to

$\sum_{i \in S} \|Y_i\|_2^2 + \sum_{i \in S} \left\langle Y_i, \sum_{j \in [N] \setminus i} Y_j \right\rangle$. We have obtained the bound for $\|Y_i\|_2^2$ in (i). For the term of inner

product, we have

$$\left\langle Y_i, \sum_{j\in[N]\setminus i} Y_j \right\rangle = \left\langle Y_i - \nu, \sum_{j\in[N]\setminus i} (Y_j - \nu) \right\rangle + \langle Y_i - \nu, (N-1)\nu \rangle + \left\langle \nu, \sum_{j\in[N]\setminus i} (Y_j - \nu) \right\rangle + (N-1)\|\nu\|_2^2.$$

Applying the sub-exponential concentration bound for the first term and sub-Gaussian concentration bound for the second as the previous handling, this is bounded by

$$c\left[\sqrt{Nk^\star \ln\left(\frac{N}{\alpha}\right)} + \ln\left(\frac{N}{\alpha}\right) + (N-1)\|\nu\|_2 \sqrt{\ln\left(\frac{N}{\alpha}\right)} + (N-1)\|\nu\|_2^2 \right].$$

With the bound of $\|Y_i\|_2^2$ in (i), the whole thing is bounded by

$$c_1 \left[\epsilon N \left(\sqrt{k^\star} + \|\nu\|_2\right) \sqrt{\ln\left(\frac{N}{\alpha}\right)} + \epsilon N \ln\left(\frac{N}{\alpha}\right) + \epsilon N \|\nu\|_2^2 \right]$$

$$+ c_2 \left[\epsilon N \sqrt{Nk^\star \ln\left(\frac{N}{\alpha}\right)} + \epsilon N \ln\left(\frac{N}{\alpha}\right) + \epsilon N^2 \|\nu\|_2 \sqrt{\ln\left(\frac{N}{\alpha}\right)} + \epsilon N^2 \|\nu\|_2^2 \right]$$

$$\leq c \left[\epsilon N \left(\sqrt{Nk^\star} + N\|\nu\|_2\right) \sqrt{\ln\left(\frac{N}{\alpha}\right)} + \epsilon N \ln\left(\frac{N}{\alpha}\right) + \epsilon N^2 \|\nu\|_2^2 \right]$$

Comparing with (C.12) and Definition C.6, we find that the required β_1 is

$$\epsilon N \left(\sqrt{k^\star} + \sqrt{N}\|\nu\|_2\right) \sqrt{\ln\left(\frac{N}{\alpha}\right)} + \epsilon N \ln\left(\frac{N}{\alpha}\right) + \epsilon N \sqrt{N}\|\nu\|_2^2,$$

which is exactly β_1 in (C.11). The proof is completed. $\square$

By Lemma C.8, we should be able to control the inner product between the sum of samples on two sets S and S' by the polarization identity. The next corollary exactly states this benefit.

Corollary C.9. *Assume* $\mathbf{Y}$ *is* $(2\epsilon, \beta_1, \beta_2)$*-regular. Then for any subsets* $S, S' \subset [N]$ *with* $|S| \leq \epsilon N, |S'| \leq \epsilon N$, *we have with probability greater than* $1 - \alpha$:

$$\left| \left\langle \sum_{i\in S} Y_i, \sum_{j\in S'} Y_j \right\rangle - |S \cap S'| k^\star \right| \leq c\beta_2.$$

Proof. By the polarization identity, we know

$$\left\langle \sum_{i\in S} Y_i, \sum_{j\in S'} Y_j \right\rangle = \frac{1}{2} \left[\left\| \sum_{i\in S\cup S'} Y_i \right\|_2^2 + \left\| \sum_{i\in S\cap S'} Y_i \right\|_2^2 - \left\| \sum_{i\in S\setminus S'} Y_i \right\|_2^2 - \left\| \sum_{i\in S'\setminus S} Y_i \right\|_2^2 \right].$$

Since each of $S\cup S', S\cap S', S\setminus S', S'\setminus S$ has cardinality no more than $2\epsilon N$ by assumption, by Definition C.6 and Lemma C.8, we know

$$\left| \left\langle \sum_{i\in S} Y_i, \sum_{j\in S'} Y_j \right\rangle - \frac{1}{2} \left(|S \cup S'| + |S \cap S'| - |S\setminus S'| - |S'\setminus S| \right) k^\star \right| \leq c\beta_2.$$

Finally, since $|S \cup S'| - |S\setminus S'| - |S'\setminus S| = |S \cap S'|$, the proof is completed. $\square$

With Lemma B.19, it is not hard to extend the result in Corollary C.9 to the case dealing with the convex combination of the indicator vectors.

Corollary C.10. *Assume* $\mathbf{Y}$ *is* $(2\epsilon, \beta_1, \beta_2)$*-regular, then for any vectors* $a, b \in [0,1]^N$ *with* $\|a\|_1 \leq \epsilon N, \|b\|_1 \leq \epsilon N$, *we have with probability greater than* $1 - \alpha$:

(i), $\left| \left\langle \sum_{i=1}^{N} a_i Y_i, \sum_{j=1}^{N} b_j Y_j \right\rangle - \langle a, b \rangle \, k^\star \right| \leq c_1 \beta_2$;

(ii), $\left| \left\langle \sum_{i=1}^{N} a_i Y_i, \sum_{j=1}^{N} Y_j \right\rangle - \|a\|_1 \, k^\star \right| \leq c_2 \sqrt{N} \beta_1$

In Corollary C.9, there is no corresponding (ii) since that is exactly the third definition of regularity.

Proof. We proof by direct calculation. By Lemma B.19, the vectors a and b can be represent as the convex combination of some indicator vectors, respectively. Let $a = \sum_{k=1}^{M_a} \xi_k \mathbf{1}_{S_k}, b = \sum_{l=1}^{M_b} \eta_l \mathbf{1}_{S_l}$, where $S_k, S_l \subset [N], 1 \leq k \leq N, 1 \leq l \leq N$. We have:

$$\left\langle \sum_{i=1}^{N} a_i Y_i, \sum_{j=1}^{N} b_j Y_j \right\rangle = \left\langle \sum_{i=1}^{N} \sum_{k=1}^{M_a} \xi_k \mathbf{1}_{S_k,i} Y_i, \sum_{j=1}^{N} \sum_{l=1}^{M_b} \eta_l \mathbf{1}_{S_l,j} Y_j \right\rangle$$

$$= \sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l \left\langle \sum_{i \in S_k} Y_i, \sum_{j \in S_l} Y_j \right\rangle$$

$$\overset{(1)}{=} \sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l \left(|S_k \cap S_l| \, k^\star + R_{k,l} \right)$$

$$= k^\star \cdot \sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l \sum_{i=1}^{N} \mathbf{1}_{S_k,i} \cdot \mathbf{1}_{S_l,i} + \sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l R_{k,l}$$

$$= k^\star \cdot \sum_{i=1}^{N} \left(\sum_{k=1}^{M_a} \xi_k \mathbf{1}_{S_k,i} \right) \left(\sum_{l=1}^{M_b} \eta_l \mathbf{1}_{S_l,i} \right) + \sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l R_{k,l}$$

$$= k^\star \cdot \sum_{i=1}^{N} a_i b_i + \sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l R_{k,l}.$$

Here in the equation, (1) is from Corollary C.9, and $R_{k,l} \leq c\beta_2$ for some constant c. Now that by Lemma B.19, $0 \leq \sum_{k=1}^{M_a} \xi_k \leq 1, 0 \leq \sum_{l=1}^{M_b} \eta_l \leq 1$, we know $\sum_{k=1}^{M_a} \sum_{l=1}^{M_b} \xi_k \eta_l R_{k,l} \leq c\beta_2$. Therefore, (i) is checked. With very similar argument, we can also check (ii). The proof is completed. $\square$

At this stage, we have all tools needed for the proof. The goal of the following Algorithms 9, 10, 11 is to filter the weight ω and the samples $\mathbf{Y}$ such that the resulting ω and $\mathbf{Y}$ meet the requirements of ω-regularity. We start with the first requirement.

Lemma C.11. *Upon finishing Algorithm 9, all of samples satisfy*

$$\left| \|Y_i\|_2^2 - k^\star \right| \leq \gamma_1.$$

Algorithm 9: Prefiltering.

---

1   Set $\gamma_1 = c\left[\sqrt{k^\star \ln\left(\frac{N}{\alpha}\right)} + \ln\left(\frac{N}{\alpha}\right)\right]$, $\mathtt{count} = 0$, $\mathtt{i} = 0$

2   **while** $i < N$ **do**

3   **if** $\left|\|Y_i\|_2^2 - k^\star\right| > \gamma_1$ **then**

4       $\mathtt{count} = \mathtt{count} + 1$

5   **if** $count > \epsilon N$ **then**

6   **return** None

7   Delete Y_i from $\mathbf{Y}$

8     $\mathtt{i} = \mathtt{i} + 1$

9   **return** $\mathbf{Y}$

Moreover, under the null hypothesis H_0, the algorithm will terminate without rejecting with probability at least $1 - \alpha$.

Proof. The satisfaction of the inequality comes directly from the definition of Algorithm 9. According to Lemma C.8 where setting $\|\nu\|_2 = 0$, all uncontaminated samples will satisfy the inequality with probability at least $1 - \alpha$ and therefore will not be deleted. The remaining follows. $\qquad\square$

Remark. *Rigorously speaking, after the prefiltering, the number of samples left should be adjusted to $N - \mathtt{count}$, where $\mathtt{count}$ is possible to be positive. However, if the prefiltering finishes without rejection, $\mathtt{count}$ is at most ϵN, upon which, and for the conciseness of the notations, we continue to use N in the proofs later. One can verify that this does not affect the correctness of the following proofs.*

The next two algorithms both deal with the second requirement of ω-regularity. However, different techniques are required for the classic scenario $N > k^\star$ and the high-dimensional scenario $N \leq k^\star$. Assume first $N > k^\star$.
Let

$$\gamma_2 := c\left[\sqrt{Nk^\star} + \sqrt{N \ln\left(\frac{1}{\alpha}\right)} + \ln\left(\frac{1}{\alpha}\right) + \epsilon N \ln\left(\frac{1}{\epsilon}\right)\right]. \tag{C.13}$$

Denote $D(\omega) = k^\star \cdot \mathrm{diag}\{\omega\}$, and $\lambda(\omega) = \left\|Y^\top D(\omega) Y - N\mathbf{I}_{k^\star}\right\|_2$. The major part of the algorithm is a "while" iteration, whose control condition is $\lambda(\omega^{(t)}) \geq \gamma_2$, where $\omega(t)$ represents the value of ω in the t-th iteration. Inside each iteration, let $v^{(t)}$ denote the singular vector associated with $\lambda(\omega^{(t)})$. We first compute $\tau_i^{(t)} = \langle v, Y_i\rangle^2 \mathbf{1}_{\left\{\omega_i^{(t)} > 0\right\}}$, then sort the samples according to the decreasing order of $\tau_i^{(t)}$. (Without any loss of generality, we assume $\tau_i^{(t)}$ are already sorted.) Next, let I be the smallest index such that $\sum\limits_{i=1}^{I} \omega_i^{(t)} \geq 2\epsilon N$ and update $\omega_i^{(t+1)}$ as following:

$$\omega_i^{(t+1)} = \begin{cases} \left(1 - \dfrac{\tau_i^{(t)}}{\tau_1^{(t)}}\right)\omega_i^{(t)} & \text{if } i \leq I, \\ \omega_i^{(t)} & \text{if } i > I. \end{cases} \tag{C.14}$$

Finally, we check the condition whether $\|\omega\|_1 \leq N(1 - 2\epsilon)$ and reject the null hypothesis directly if this holds and pass to the following step if not.

Algorithm 10: Sample filtering when $N > k^\star$.

1 Set γ_2 as (C.13), and $\lambda = \left\| Y^\top D(\omega) Y - N\mathbf{I}_{k^\star} \right\|_2$. ($\omega$ is initialized as $\mathbf{1}$.)

2 while $\lambda \geq \gamma_2$ **do**

3 $\quad$ Set v to be the unit singular vector associated with λ

4 $\quad$ Compute $\tau_i = \langle v, Y_i \rangle^2 \mathbf{1}_{\{\omega_i > 0\}}$ for $1 \leq i \leq N$

5 $\quad$ Set I be the smallest index such that $\sum\limits_{i=1}^{I} \omega_i \geq 2\epsilon N$

6 $\quad$ Update w according to (C.14)

7 $\quad$ **if** $\|\omega\|_1 < N(1 - 2\epsilon)$ **then**

8 $\quad\quad$ $\lfloor$ **return** None

9 $\quad$ Set $\lambda = \left\| Y^\top D(\omega) Y - N\mathbf{I}_{k^\star} \right\|_2$

10 return ω

We are going to prove that if the algorithm finishes without rejecting the null, the second condition of ω-regularity is guaranteed with high probability. This is formalized as the following lemma.

Lemma C.12. *Given the algorithm 10, we have the following facts:*

(i), the algorithm will terminate within finite rounds,

(ii), under H_0, the algorithm will not output rejection with high probability;

(iii), once finished without rejection, for any $S \subset [N]$ with $|S| \leq \epsilon N$, we have $\left\| \sum\limits_{i \in S} \sqrt{\omega_i} Y_i \right\|_2^2 \leq$ $2\gamma_2 \epsilon N$ with probability higher than $1 - \alpha$.

Proof. Proof of (i) is relatively easy. By the update (C.14), the ω_i corresponding to the maximal τ_i will be set to zero in each iteration. Consequently, $\|\omega\|_1$ will be less than $N(1 - 2\epsilon)$ in at most $2\epsilon N$ rounds, followed by the rejection.

For (ii), define $\Lambda_1 = \left\{ \omega : \omega \in [0, 1]^N, \left\| \mathbf{1}_{[N]\setminus C} - \omega_{[N]\setminus C} \right\|_1 \leq \|\mathbf{1}_C - \omega_C\|_1 \right\}$, and denote $\omega^{(t)}, \tau^{(t)}$ as the value of ω, τ in the t-th iteration. We attempt to prove that under H_0, if $\omega^{(t)} \in \Lambda_1$ and we are still in the iteration, then so does $\omega^{(t+1)}$. Suppose this holds, we have $\left\| \mathbf{1} - \omega^{(t+1)} \right\|_1 \leq 2 \left\| \mathbf{1}_C - \omega_C^{(t+1)} \right\|_1 \leq 2\epsilon N$, therefore the condition for rejection will never be triggered.

Now it remains to prove the statement. Suppose $\omega^{(t)} \in \Lambda_1$ and we are still in the iteration. By the definition of Λ_1 and the update (C.14), it suffices to prove $\sum\limits_{i \in [I] \cap C} \tau_i^{(t)} \omega_i^{(t)} \geq \sum\limits_{i \in [I] \cap ([N]\setminus C)} \tau_i^{(t)} \omega_i^{(t)}$, i.e.,

$$v^\top \left(\sum_{i \in [I] \cap C} \omega_i^{(t)} Y_i Y_i^\top \right) v \geq v^\top \left(\sum_{i \in [I] \cap ([N]\setminus C)} \omega_i^{(t)} Y_i Y_i^\top \right) v.$$

By the control condition, we have

$$v^\top \left(\sum_{i=1}^{N} \omega_i^{(t)} Y_i Y_i^\top - N\mathbf{I}_{k^\star} \right) v \geq \gamma_2.$$

By Corollary B.14 (set $\|\nu\|_2 = 0$) and the definition of γ_2 (C.13), with high probability, we have

$$v^\top \left(\sum_{i \in C} \omega_i^{(t)} Y_i Y_i^\top \right) v = v^\top \left(\sum_{i=1}^{N} \omega_i^{(t)} Y_i Y_i^\top - N\mathbf{I}_{k^\star} \right) v - v^\top \left(\sum_{i \in [N] \setminus C} \omega_i^{(t)} Y_i Y_i^\top - N\mathbf{I}_{k^\star} \right) v$$

$$\geq v^\top \left(\sum_{i=1}^{N} \omega_i^{(t)} Y_i Y_i^\top - N\mathbf{I}_{k^\star} \right) v - v^\top \left(\sum_{i=1}^{N} \widetilde{Y}_i \widetilde{Y}_i^\top - N\mathbf{I}_{k^\star} \right) v$$

$$\geq (1-c)\gamma_2.$$

By the definition of I we know $2\epsilon N \leq \sum_{i \in [I]} \omega_i < 2\epsilon N + 1$. Because $\omega^{(t)} \in \Lambda_1$, we know $\left\| \mathbf{1} - \omega^{(t)} \right\|_1 \leq 2 \left\| \mathbf{1}_C - \omega_C^{(t)} \right\|_1 \leq 2\epsilon N$. Hence $I - (2\epsilon N + 1) \leq \sum_{i \in [I]} (1 - \omega_i^{(t)}) \leq 2\epsilon N$, which leads to $I \leq 4\epsilon N + 1$. Therefore, according to Lemma B.15, with high probability, we have

$$v^\top \left(\sum_{i \in [I] \cap ([N] \setminus C)} \omega_i^{(t)} Y_i Y_i^\top \right) v \leq c\gamma_2. \tag{C.15}$$

Now that we sorted τ_i by decreasing order, $\sum_{i \in [I]} \omega_i \geq 2\epsilon N$, and $|C| \leq \epsilon N$, we have

$$\sum_{i \in C} \omega_i^{(t)} \tau_i^{(t)} - \sum_{i \in [I]} \omega_i^{(t)} \tau_i^{(t)} \leq \sum_{i \in C \setminus [I]} \omega_i^{(t)} \tau_i^{(t)} \leq \frac{1}{2} \sum_{i \in [I]} \omega_i^{(t)} \tau_i^{(t)}.$$

Therefore $\sum_{i \in [I]} \omega_i^{(t)} \tau_i^{(t)} \geq \frac{2}{3} \sum_{i \in C} \omega_i^{(t)} \tau_i^{(t)} \geq \frac{2(1-c)}{3} \gamma_2$. Note that by selecting the constant in (C.13) sufficiently large, we can let c sufficiently small. Hence, we can assume $c \leq \frac{1}{4}$, by which we have $\sum_{i \in [I]} \omega_i^{(t)} \tau_i^{(t)} \geq \frac{1}{2}\gamma_2$. Combine with (C.15), we find

$$\sum_{i \in [I] \cap C} \omega_i^{(t)} \tau_i^{(t)} \geq \frac{1}{4}\gamma_2 \geq \sum_{i \in [I] \cap ([N] \setminus C)} \omega_i^{(t)} \tau_i^{(t)}.$$

Hence, (ii) is proved.

To prove (iii), we need the following lemma, which exploits the property of the norm of the mean from the algorithm.

Lemma C.13. *If $\|\nu\|_2 \geq c$ for some sufficient large but universal constant c, then Algorithm 10 will reject H_0 with high probability.*

Proof. First, we fix the vector $v = \frac{\nu}{\|\nu\|_2}$. If the algorithm exits without rejection, we have

$$\left\| \sum_{i=1}^{N} \omega_i Y_i Y_i^\top - N\mathbf{I}_{k^\star} \right\|_2 < \gamma_2.$$

66

This suggests that $\sum_{i=1}^{N} \omega_i (v^\top Y_i)^2 - N < \gamma_2$, based on which we have

$$\frac{\sum_{i \in [N] \backslash C} \omega_i (v^\top Y_i)^2}{N} - 1 \leq \frac{\sum_{i=1}^{N} \omega_i (v^\top Y_i)^2}{N} - 1 < \frac{\gamma_2}{N} \leq c,$$

where the last step is from the definition of γ_2 (C.13). If we expand $v^\top Y_i$, this merely says

$$\sum_{i \in [N] \backslash C} \omega_i \left[v^\top (Y_i - \nu) \right]^2 + \sum_{i \in [N] \backslash C} \omega_i \|\nu\|_2^2 + 2 \sum_{i \in [N] \backslash C} \omega_i \left[v^\top (Y_i - \nu) \right] \|\nu\|_2 \leq cN. \qquad (C.16)$$

By the fact that $ab \geq -\frac{1}{4}a^2 - b^2$ for $a, b \in \mathbb{R}$, we know

$$2 \sum_{i \in [N] \backslash C} \omega_i \left[v^\top (Y_i - \nu) \right] \|\nu\|_2 \geq -\frac{1}{2} \sum_{i \in [N] \backslash C} \omega_i \|\nu\|_2^2 - 2 \sum_{i \in [N] \backslash C} \omega_i \left[v^\top (Y_i - \nu) \right]^2.$$

Adding both sides with $\sum_{i \in [N] \backslash C} \omega_i \|\nu\|_2^2$ and plugging (C.16) in, we have

$$\frac{1}{2} \sum_{i \in [N] \backslash C} \omega_i \|\nu\|_2^2 \leq cN + \sum_{i \in [N] \backslash C} \omega_i \left[v^\top (Y_i - \nu) \right]^2 \leq cN + \sum_{i \in [N] \backslash C} \left[v^\top (Y_i - \nu) \right]^2.$$

Now since $\|v\|_2 = 1$, we know that $\left[v^\top (Y_i - \nu) \right]^2$ is subject to a chi-square distribution with degree of freedom 1. By the standard argument for the sub-exponential random variables (see B.7), we know that $\sum_{i \in [N] \backslash C} \left[v^\top (Y_i - \nu) \right]^2 \leq N + c\sqrt{N}$ with (high) constant probability. On the other hand, by the definition of the algorithm we have $\left\| \omega_{[N] \backslash C} \right\|_1 \geq \|\omega\|_1 - \epsilon N \geq (1 - 3\epsilon)N$. Therefore, we have with (high) constant probability,

$$\frac{1}{2}(1 - 3\epsilon)N \|\nu\|_2^2 \leq c_1 \cdot N + c_2 \cdot \sqrt{N}.$$

Now dividing both sides by $\frac{1}{2}(1 - 3\epsilon)N$, we see that Lemma C.13 holds if we take $c = \frac{2(c_1 + c_2)}{1 - 3\epsilon}$ $\square$

We are going to prove (iii) by contradiction. Suppose that at the end of the algorithm without rejection, we have $\left\| \sum_{i \in S} \sqrt{\omega_i} Y_i \right\|_2^2 > 2\gamma_2 \epsilon N$ for some $S \subset [N]$ with $|S| \leq \epsilon N$. This indicates that there exists a vector $v \in \mathbb{R}^{k^\star}$ with $\|v\|_2 = 1$ such that $\sum_{i \in S} \sqrt{\omega_i} v^\top Y_i > \sqrt{2\gamma_2 \epsilon N}$. Therefore, by the Cauchy-Schwarz inequality, we have $\sum_{i \in S \cup C} \omega_i (v^\top Y_i)^2 \geq 2\gamma_2$. This implies that

$$\left\| \sum_{i \in S \cup C} \omega_i Y_i Y_i^\top + |[N] \backslash (S \cup C)| \nu \nu^\top \right\|_2 \geq \left\| \sum_{i \in S \cup C} \omega_i Y_i Y_i^\top \right\|_2 \geq 2\gamma_2. \qquad (C.17)$$

67

Now consider the quantity $\left\| \sum_{i \in [N] \setminus (S \cup C)} \omega_i Y_i Y_i^\top - |[N] \setminus (S \cup C)| \nu\nu^\top - N\mathbf{I}_{k^\star} \right\|$, by the triangular inequality, we have

$$
\left\| \sum_{i \in [N] \setminus (S \cup C)} \omega_i Y_i Y_i^\top - |[N] \setminus (S \cup C)| \nu\nu^\top - N\mathbf{I}_{k^\star} \right\| \leq \left\| \sum_{i=1}^N \omega_i \widetilde{Y}_i \widetilde{Y}_i^\top - N\nu\nu^\top - N\mathbf{I}_{k^\star} \right\|_2
$$

$$
+ \left\| \sum_{i \in S \cup C} \omega_i \widetilde{Y}_i \widetilde{Y}_i^\top - |S \cup C| \nu\nu^\top \right\|_2
$$

$$
\leq \mathrm{I} + \mathrm{II} + \mathrm{III} + |S \cup C| \, \|\nu\|_2^2.
$$

where I, II and III are defined as

$$
\mathrm{I} = \left\| \sum_{i=1}^N \widetilde{Y}_i \widetilde{Y}_i^\top - N\nu\nu^\top - N\mathbf{I}_{k^\star} \right\|_2, \, \mathrm{II} = \left\| \sum_{i=1}^N (1 - \omega_i) \widetilde{Y}_i \widetilde{Y}_i^\top \right\|_2, \, \mathrm{III} = \left\| \sum_{i \in S \cup C} \omega_i \widetilde{Y}_i \widetilde{Y}_i^\top \right\|.
$$

By Lemma C.13, without loss of generality we can assume $\|\nu\|_2 \leq c_0$ for some constant c_0. For the last term, since $|S \cup C| \leq 2\epsilon N$, obviously it can be bounded by $c\gamma_2$. We are going to deal with the other three parts respectively.

Bound on I. We have:

$$
\mathrm{I} \leq \left\| \sum_{i=1}^N (\widetilde{Y}_i - \nu)(\widetilde{Y}_i - \nu)^\top - N\mathbf{I}_{k^\star} \right\|_2 + 2 \left\| \sum_{i=1}^N (\widetilde{Y}_i - \nu)^\top \nu \right\|_2.
$$

Set $\|\nu\|_2 = 0$ in Corollary B.14, we have

$$
\left\| \sum_{i=1}^N (\widetilde{Y}_i - \nu)(\widetilde{Y}_i - \nu)^\top - N\mathbf{I}_{k^\star} \right\|_2 \leq C_1 \left(\sqrt{Nk^\star} + \sqrt{N \ln\left(\frac{1}{\alpha}\right)} + \ln\left(\frac{1}{\alpha}\right) \right). \tag{C.18}
$$

By normalization and standard concentration tools (see B.6), we have

$$
\left\| \sum_{i=1}^N (\widetilde{Y}_i - \nu)^\top \nu \right\|_2 = \|\nu\|_2 \cdot \left\| \sum_{i=1}^N (\widetilde{Y}_i - \nu) \right\|_2 = C_0 \cdot \sqrt{Nk^\star} \cdot \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\widetilde{Y}_i - \nu}{\sqrt{k^\star}} \right\|_2 \leq c_2 \sqrt{Nk^\star} \tag{C.19}
$$

with (high) constant probability. Now combine (C.18) and (C.19) to get

$$
\mathrm{I} \leq (c_1 + 2c_2) \left(\sqrt{Nk^\star} + \sqrt{N \ln\left(\frac{1}{\alpha}\right)} + \ln\left(\frac{1}{\alpha}\right) \right) \leq c\gamma_2.
$$

Bound on II. Since upon the finish the algorithm, we have $\|\mathbf{1} - \omega\|_1 \leq 2\epsilon N$, by Corollary B.16, we know

$$
\mathrm{II} \leq c_3 \left[\epsilon N \ln\left(\frac{1}{\epsilon}\right) + k^\star + \ln\left(\frac{1}{\alpha}\right) \right] \leq c\gamma_2.
$$

Bound on III. Since $|S \cup C| \leq 2\epsilon N$, by Corollary B.16 and an argument similar to the one in II, we also have

$$
\mathrm{III} \leq c\gamma_2.
$$

Therefore, we conclude that

$$\left\|\sum_{i\in[N]\backslash(S\cup C)}\omega_i Y_i^\top Y_i - |[N]\backslash(S\cup C)|\,\nu\nu^\top - N\mathbf{I}_{k^\star}\right\|_2 \leq 4c\gamma_2. \tag{C.20}$$

Now combine (C.17) and (C.20) and use the triangular inequality, we find that

$$\left\|\sum_{i=1}^{N}\omega_i Y_i^\top Y_i - N\mathbf{I}_{k^\star}\right\|_2 \geq (2-4c)\gamma_2.$$

Again, by choosing the constant in the definition of γ_2 (C.13) sufficiently large, we can make c above sufficiently small. Therefore, we can assume $c < \frac{1}{4}$, which leads to $\left\|\sum_{i=1}^{N}\omega_i Y_i^\top Y_i - N\mathbf{I}_{k^\star}\right\|_2 > \gamma_2$, a contradiction of the control condition of Algorithm 10. The proof is completed. $\qquad\square$

We now turn to the high-dimensional scenario where $N \leq k^\star$. In this case, we let

$$\gamma_3 = c\left(\sqrt{Nk^\star} + \sqrt{k^\star \ln\left(\frac{1}{\alpha}\right) + \ln\left(\frac{1}{\alpha}\right)} + \epsilon N \ln\left(\frac{1}{\epsilon}\right)\right) \tag{C.21}$$

for some universal constant c. The control condition of the iteration, in this case, is

$$\left\|\sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega)\right\|_2 \geq \gamma_3,$$

where $D(\omega) = k^\star \cdot \mathrm{diag}\{\omega\}$. Inside each iteration, similar as before, we find v, which is the eigenvector associated with λ, and compute $\tau_i = \frac{v_i^2}{\omega_i}\mathbf{1}_{\{\omega_i>0\}}, 1 \leq i \leq N$. Next, we update ω as following

$$\omega_i = \left(1 - \frac{\tau_i}{\max_i \tau_i}\right)\omega_i, 1 \leq i \leq N. \tag{C.22}$$

Finally, we check the condition whether $\|\omega\|_1 \leq N(1-6\epsilon)$ and reject the null directly if this holds and pass to the following step if not.

Similar to the algorithm 10, we will have the following results to guarantee the satisfaction of the second condition of ω-regularity.

Lemma C.14. *Given the algorithm 11, the following statements hold:*
(i), the algorithm will terminate within finite rounds;
(ii), under H_0, the algorithm will not output rejection with high probability;
(iii), upon the finish of the algorithm, for any $S \subset [N]$ with $|S| \leq \epsilon N$, we have

$$\left|\left\|\sum_{i\in S}\sqrt{\omega_i}Y_i\right\|_2^2 - \|\omega_S\|_1\,k^\star\right| \leq c\epsilon N\gamma_3$$

with probability higher than $1-\alpha$.

69

Algorithm 11: Sample filtering when $N \leq k^\star$.

---

1 Set γ_3 as (C.21), and $\lambda = \left\| \sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega) \right\|_2$. ($\omega$ is initialized as $\mathbf{1}$.)

2 while $\lambda \geq \gamma_3$ **do**

3 $\quad$ Set v to be the unit singular vector associated with λ

4 $\quad$ Compute $\tau_i = \frac{v_i^2}{\omega_i}\mathbf{1}_{\{w_i > 0\}}$

5 $\quad$ Update w according to (C.22)

6 $\quad$ **if** $\|\omega\|_1 < N(1 - 6\epsilon)$ **then**

7 $\quad\quad$ $\lfloor$ **return** None

8 $\quad$ Set $\lambda = \left\| \sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega) \right\|_2$

9 return ω

Proof. Proof of (i) is same as the one in the proof of Lemma C.12.

To prove (ii), for the fixed samples S, define $\Lambda_2 = \left\{ \omega : \omega \in [0,1]^N, \left\|\mathbf{1}_{[N]\backslash C} - \omega_{[N]\backslash C}\right\|_1 \leq 5\left\|\mathbf{1}_C - \omega_C\right\|_1 \right\}$. We will prove that $\omega^{(t)} \in \Lambda_2$ before exiting the iteration. Suppose first this holds, we have $\|\mathbf{1} - \omega\|_1 \leq 6\|\mathbf{1}_C - \omega_C\|_1 \leq 6\epsilon N$. This implies that we will always exit the iteration before the algorithm has chance to reject.

Now assume $\omega^{(t)} \in \Lambda_2$ and we are still in the iteration, we will prove $\omega^{(t+1)} \in \Lambda_2$. By the definition, we only need to prove $5\sum_{i \in C} v_i^2 \geq \sum_{i \in [N]\backslash C} v_i^2$. Suppose this does not hold, then $\|v_C\|_2^2 < \frac{1}{6}$. By Corollary B.14 (set $\|\nu\|_2 = 0$), we have $\left\| \sqrt{D(\omega_{[N]\backslash C}^{(t)})}Y_{[N]\backslash C}Y_{[N]\backslash C}^\top\sqrt{D(\omega_{[N]\backslash C}^{(t)})} - D(\omega_{[N]\backslash C}^{(t)}) \right\|_2 \leq c\gamma_3 \leq c\lambda^{(t)}$ with high probability. However, since v is the eigenvector corresponding to the maximal eigenvalue, we have

$$\lambda^{(t)} = \left| v^\top \left(\sqrt{D(\omega^{(t)})}Y^\top Y\sqrt{D(\omega^{(t)})} - D(\omega^{(t)}) \right) v \right|$$
$$\leq c\lambda^{(t)}\left\|v_{[N]\backslash C}\right\|_2^2 + 2\lambda^{(t)}\left\|v_{[N]\backslash C}\right\|_2\|v_C\|_2 + \lambda^{(t)}\|v_C\|_2^2.$$

Noting that $\|v_C\|_2^2 + \left\|v_{[N]\backslash C}\right\|_2^2 = 1$, define $f_c(x) = c + (1-c)x + 2\sqrt{x(1-x)}, 0 \leq x < \frac{1}{6}$. Note that by selecting the constant in (C.21) sufficiently large, we can let $c \to 0$. It can be proved that in this case $0 \leq f_0(x) < 1$, which is a contradiction. (ii) is proved.

Upon the termination of the algorithm without rejection, we have $\left\| \sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega) \right\|_2 < \gamma_3$ and $\|\omega\|_1 \geq (1 - 6\epsilon)N$. Therefore, we have

$$\left| \left\| \sum_{i \in S} \sqrt{\omega_i}Y_i \right\|_2^2 - k^\star\|\omega_S\|_1 \right| = \left| \sum_{i \in S}\sum_{j \in S} \sqrt{\omega_i\omega_j}Y_i^\top Y_j - k^\star\|\omega_S\|_1 \right|$$
$$= \left| \mathbf{1}_S^\top \left[\sqrt{D(\omega)}YY^\top\sqrt{D(\omega)} - D(\omega) \right] \mathbf{1}_S \right| \overset{(i)}{\leq} c\epsilon N\gamma_3.$$

Here (i) uses the condition that $\left\| \sqrt{D(\omega)}\mathbf{Y}\mathbf{Y}^\top\sqrt{D(\omega)} - D(\omega) \right\|_2 \leq \gamma_3$. The proof is completed. $\quad\square$

After Algorithms 10 or 11, we have filtered the samples to ensure the second condition of ω-regularity. Now we turn to the check of the final condition.

Actually, the weight ω output by Algorithms 10 or 11 already provides us with some useful insight.

Lemma C.15. *Set β_1, β_2 according to (C.11) and $\gamma_i, i \in \{2,3\}$ according to (C.13) or (C.21) depending on whether $N \geq k^\star$ or not. For the weight ω output by the Algorithms 10 or 11 and any $S \subset [N]\backslash C$ with $|S| \leq \epsilon N$, we have*

$$\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle - k^\star \|\omega_S\|_1 \right| \leq c(\sqrt{N}\beta_1 + \beta_2 + \epsilon N \gamma)$$

with high probability.

Proof. By Lemma C.8, with high probability the original samples $\widetilde{\mathbf{Y}}$ are $(2\epsilon, \beta_1, \beta_2)$-regular. By spliting $[N] = ([N]\backslash C) \cup C$, we have

$$\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle - k^\star \|\omega_S\|_1 \right| = \underbrace{\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, \widetilde{Y}_j \rangle - k^\star \|\omega_S\|_1 \right|}_{\text{I}}$$

$$+ \underbrace{\left| \sum_{i \in S} \sum_{j \in C} \sqrt{\omega_i \omega_j} \langle Y_i, \widetilde{Y}_j \rangle \right|}_{\text{II}} + \underbrace{\left| \sum_{i \in S} \sum_{j \in C} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle \right|}_{\text{III}}$$

For the term I, we further have

$$\text{I} \leq \underbrace{\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i} \langle Y_i, \widetilde{Y}_j \rangle - k^\star \|\sqrt{\omega_S}\|_1 \right|}_{\text{I}_1} + \underbrace{\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i}(1 - \sqrt{\omega_j}) \langle Y_i, \widetilde{Y}_j \rangle - k^\star \|\sqrt{\omega_S}(1 - \sqrt{\omega_S})\|_1 \right|}_{\text{I}_2}$$

We now bound the terms $\text{I}_1, \text{I}_2, \text{II}, \text{III}$ respectively.

Bound on I_1. Because $|S| \leq \epsilon N$, from ω-regularity and Corollary C.10, we know $\text{I}_1 \leq c_1 \sqrt{N}\beta_1$.

Bound on I_2. Because ω is output by the Algorithms 10 or 11, we have $\|1 - \sqrt{\omega}\|_1 \leq \|1 - \omega\|_1 \leq \epsilon N$. By Corollary C.10 again, we have $\text{I}_2 \leq c_2 \beta_2$.

Bound on II. From Corollary C.10 and the fact $S \cap C = \varnothing$, we know $\text{II} \leq c_3 \beta_2$.

Bound on III. From Lemma C.12, C.14, and the polarization equality, we have $\text{III} \leq c_4 \epsilon N \gamma$.

Combining them together, we come to the conclusion of the lemma. $\qquad \square$

We are now prepared for the final algorithm, which is designed to ensure the third condition of ω-regularity. Given samples S and weight ω by the algorithms 10 or 11, we compute $\tau_i = \left| \left\langle \sqrt{\omega_i} Y_i, \sum_{j=1}^{N} \sqrt{\omega_j} Y_j \right\rangle - \omega_i k^\star \right| \mathbf{1}_{\{\omega_i > 0\}}$. Then, we set ω_i to zero for the indices i corresponding to the ϵN maximal τ_i.

Algorithm 12: Weight filtering.

1 Compute $\tau_i = \left| \left\langle \sqrt{\omega_i} Y_i, \sum_{j=1}^{N} \sqrt{\omega_j} Y_j \right\rangle - \omega_i k^\star \right| \mathbf{1}_{\{\omega_i > 0\}}, 1 \le i \le N$

2 Sort τ_i by desreasing order and find the indices $\{i_1, i_2, \ldots, i_{\epsilon N}\}$ corresponding to the first ϵN maximal τ_i

3 Set $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_{\epsilon N}}$ to zero

4 return ω

Lemma C.16. *For the weight output by Algorithm 12 and any $S \subset [N]$ with $|S| \le \epsilon N$, we have*

$$\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle - k^\star \|\omega_S\|_1 \right| \le c(\sqrt{N}\beta_1 + \beta_2 + \epsilon N \gamma).$$

This lemma is very similar to Lemma C.15 but stronger in the sense that we allow any $S \subset [N]$ rather than $S \subset [N] \backslash C$. This advantage is exactly granted by the sorting and pruning processes in Algorithm 12.

Proof. The proof of this lemma is not hard but depends on the following fact. Because we have proved in Lemma C.15 that for any $S \subset [N] \backslash C$ and $|S| \le \epsilon N$, we have $\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle - k^\star \|\omega_S\|_1 \right| \le c(\sqrt{N}\beta_1 + \beta_2 + \epsilon N \gamma)$, we know that this bound also holds for any $S \subset [N] \backslash \{1, 2, \ldots, \epsilon N\}$. With the proning on ω to remove the first ϵN coordinates, we can assume without loss of generality that $S \subset [N] \backslash \{1, 2, \ldots \epsilon N\}$. We have

$$\left| \sum_{i \in S} \sum_{j=1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle - k^\star \|\omega_T\|_1 \right| = \left| \sum_{i \in S} \sum_{j=\epsilon N+1}^{N} \sqrt{\omega_i \omega_j} \langle Y_i, Y_j \rangle - k^\star \|\omega_T\|_1 \right| \overset{(1)}{\le} c(\sqrt{N}\beta_1 + \beta_2 + \epsilon N \gamma).$$

Here (1) can be obtained by considering the subtraction of summation on the original weight with indices between 1 and ϵN from the summation on the whole original weight, and using the fact that $S \cap \{1, 2, \ldots, \epsilon N\} = \varnothing$. $\qquad \square$

We have finished the proofs of Lemma C.11, C.12, C.14, and C.16, which means we have verified the correctness of ω-regularity for our samples with high probability. The final step is to exploit the property of ω-regularity to distinguish between H_0 and H_1.

Lemma C.17. *Let ω be the weights after running through the algorithm 9, 10 or 11, and 12, and β_1, β_2 seleced as in Lemma C.8. Then, with high probability we have*

$$\left| \left\| \sum_{i=1}^{N} \sqrt{\omega_i} Y_i \right\|_2^2 - k^\star \|\omega\|_1 - N^2 \|\nu\|_2^2 \right| \le c \left(\sqrt{N}\beta_1 + \beta_2 + \epsilon N \gamma + \underbrace{\left(N^{\frac{3}{2}} \|\nu\|_2 + \sqrt{N k^\star} \right) \sqrt{\ln\left(\frac{1}{\alpha}\right)} + \ln\left(\frac{1}{\alpha}\right)}_{:=R} \right).$$

$$\text{(C.23)}$$

Proof. The proof of this lemma is from the definition of ω-regularity and straightforward calculations. To evaluate $\left\|\sum_{i=1}^{N}\sqrt{\omega_i}Y_i\right\|_2^2$, we write it as:

$$\left\|\sum_{i=1}^{N}\sqrt{\omega_i}Y_i\right\|_2^2 = \left\|\sum_{i\in[N]\setminus C}\sqrt{\omega_i}Y_i + \sum_{i\in C}\sqrt{\omega_i}Y_i\right\|_2^2$$

$$= \underbrace{\left\|\sum_{i\in[N]\setminus C}\sqrt{\omega_i}Y_i\right\|_2^2}_{\text{I}} - \underbrace{\left\|\sum_{i\in C}\sqrt{\omega_i}Y_i\right\|_2^2}_{\text{II}} + \underbrace{2\sum_{i=1}^{N}\sum_{j\in C}\sqrt{\omega_i\omega_j}Y_i^\top Y_j}_{\text{III}}$$

We are going to bound the terms individually.

Bound on I. Define the vector $r \in \mathbb{R}^N$ where $r_i = 1$ for $i \in C$ and $1 - \sqrt{\omega_i}$ for $i \in [N]\setminus C$. By Algorithm 10 or 11, we have $\|r\|_1 \le \epsilon N + 6\epsilon N = 7\epsilon N$. For the term I, we have

$$\left\|\sum_{i\in[N]\setminus C}\sqrt{\omega_i}Y_i\right\|_2^2 = \underbrace{\left\|\sum_{i=1}^{N}\widetilde{Y}_i\right\|_2^2}_{\text{I}_1} + \underbrace{\left\|\sum_{i=1}^{N}r_i\widetilde{Y}_i\right\|_2^2}_{\text{I}_2} - \underbrace{2\sum_{i=1}^{N}\sum_{j=1}^{N}r_j\widetilde{Y}_i^\top\widetilde{Y}_j}_{\text{I}_3}$$

For I_1, we have

$$\text{I}_1 = \sum_{i=1}^{N}\left\|\widetilde{Y}_i - \nu\right\|_2^2 + N^2\|\nu\|_2^2 + 2N\nu^\top\sum_{i=1}^{N}(\widetilde{Y}_i - \nu).$$

By the standard concentration bounds (see B.6 and B.7), with high probability we have

$$\left|\text{I}_1 - Nk^\star - N^2\|\nu\|_2^2\right| \le c_1\left(N^{\frac{3}{2}}\|\nu\|_2\sqrt{\ln\left(\frac{1}{\alpha}\right)} + \sqrt{Nk^\star\ln\left(\frac{1}{\alpha}\right)} + \ln\left(\frac{1}{\alpha}\right).\right)$$

For I_2, by the argument above, Lemma C.8, and Corollary C.10 with appropriate choice of ϵ, with high probability we have

$$\left|\text{I}_2 - k^\star\|r\|_2^2\right| \le c_2\beta_2.$$

Again by Lemma C.8 and Corollary C.10, with high probability we have

$$|\text{I}_3 - 2k^\star\|r\|_1| \le c_3\sqrt{N}\beta_1.$$

Combining the results above and using the fact that $Nk^\star + k^\star\|r\|_2^2 + 2k^\star\|r\|_1 = k^\star\|\mathbf{1} - r\|_2^2 = k^\star\|\omega_{[N]\setminus C}\|_1$, we have

$$\left|\text{I} - k^\star\|\omega_{[N]\setminus C}\|_1 - N^2\|\nu\|_2^2\right| \le c\left[\sqrt{N}\beta_1 + \beta_2 + \left(N^{\frac{3}{2}}\|\nu\|_2 + \sqrt{Nk^\star}\right)\sqrt{\ln\left(\frac{1}{\alpha}\right)} + \ln\left(\frac{1}{\alpha}\right)\right].$$

Bound on II. By Lemma C.12 (recall the definition of γ_2 (C.13) and the fact that $N > k^\star$ in this case) or C.14, with high probability we have

$$|\text{II} - k^\star\|\omega_C\|_1| \le c\epsilon N\gamma.$$

Bound on III. By Lemma C.16, with high probability we have

$$\left| \mathrm{III} - 2k^\star \left\| \omega_C \right\|_1 \right| \le c \left(\sqrt{N} \beta_1 + \beta_2 + \epsilon N \gamma \right).$$

Now combining the bounds of I, II, and III together, we have

$$\left| \left\| \sum_{i=1}^N \sqrt{\omega_i} Y_i \right\|_2^2 - k^\star \left\| \omega \right\|_1 - N^2 \left\| \nu \right\|_2^2 \right| \le c \left(\sqrt{N} \beta_1 + \beta_2 + \epsilon N \gamma + \left(N^{\frac{3}{2}} \left\| \nu \right\|_2 + \sqrt{N k^\star} \right) \sqrt{\ln \left(\frac{1}{\alpha} \right)} + \ln \left(\frac{1}{\alpha} \right) \right),$$

which is the desired results. $\qquad\square$

Proof of Theorem 3.11. Substituting ν with $P^\star \mu$, the key is to require each term on the right-hand-side in (C.23) to be asymptotically smaller than the term $N^2 \left\| P^\star \mu \right\|_2^2$. For convenience, we rewrite the definition of β_1 and β_2 here.

$$\beta_1 = \epsilon N \left(\sqrt{k^\star} + \sqrt{N} \left\| P^\star \mu \right\|_2 \right) \sqrt{\ln \left(\frac{N}{\alpha} \right)} + \epsilon N \ln \left(\frac{N}{\alpha} \right) + \epsilon N \sqrt{N} \left\| P^\star \mu \right\|_2^2,$$

$$\beta_2 = \epsilon N \sqrt{\epsilon N k^\star \ln \left(\frac{1}{\epsilon} \right)} + (\epsilon N)^2 \ln \left(\frac{1}{\epsilon} \right) + \left\| P^\star \mu \right\|_2 (\epsilon N)^2 \sqrt{\ln \left(\frac{1}{\epsilon} \right)} + \left\| P^\star \mu \right\|_2^2 (\epsilon N)^2.$$

The table below presents the terms and corresponding required conditions for $\left\| P^\star \mu \right\|_2^2$. Although there are many terms, only a few of them are useful indeed.

Inspecting the column "**Required Conditions**" and computing the maximal, we conclude the final requirement for $\left\| P^\star \mu \right\|_2$:

$$\left\| P^\star \mu \right\|_2^2 \gtrsim \max \left\{ \frac{\epsilon \ln \left(\frac{N}{\alpha} \right)}{\sqrt{N}}, \epsilon^2 \ln \left(\frac{N}{\alpha} \right), \sqrt{\frac{\epsilon^2 k^\star \ln \left(\frac{N}{\alpha} \right)}{N}}, \frac{\sqrt{k^\star} \ln \left(\frac{1}{\alpha} \right)}{N} \right\}. \tag{C.24}$$

Once this requirement is satisfied, by Lemma C.17 and the inequality (C.23), we know that $\left| \left\| \sum_{i=1}^N \sqrt{\omega_i} P^\star X_i \right\|_2^2 - k^\star \left\| \omega \right\|_1 \right|$ will be less than $c_1 N^2 P(\epsilon, k^\star, N, 1)$ under H_0 and greater than $c_2 N^2 P(\epsilon, k^\star, N, 1)$ under H_1 for some $0 < c_1 < c_2 < 1$, and both with high probability. c_1 and c_2 are universal and their values can be determined following proper choices of $\epsilon, \alpha, N, k^\star$. From the argument at the end of the proof of Theorem 3.8, we know (C.24) leads to (3.6). The proof is completed. $\qquad\square$

C.7 Proof of Corollary 3.15

Proof. By Definition 2.3 and 2.4, we can rewrite $k_1^\star$ and $k_2^\star$ as

$$k_1^\star := \max_k \left\{ k \mid k \ge 0, \frac{D_{k-1}(K)}{k^{1/4}} \ge \frac{\sigma}{\sqrt{N}} \right\},$$

$$k_2^\star := \max_k \left\{ k \mid k \ge 0, \frac{D_{k-1}(K)}{k^{1/4}} \ge \frac{\sqrt{\epsilon}}{N^{\frac{1}{4}}} \sigma \right\}.$$

For a fixed K, define function $g(k) = \frac{D_{k-1}(K)}{k^{1/4}}$. By the property of Kolmogorov widths, we know g is a decreasing function. Therefore, in the alternative definition above, if $\frac{\sigma}{\sqrt{N}} \ge \frac{\sqrt{\epsilon}}{N^{1/4}} \sigma$, i.e., $\epsilon \le \frac{1}{\sqrt{N}}$,

Sources	No.	Terms	Required Conditions for $\|P^\star\mu\|_2$
$\sqrt{N}\beta_1$	1	$\epsilon N\sqrt{Nk^\star}\sqrt{\ln\left(\frac{N}{\alpha}\right)}$	$\|P^\star\mu\|_2^2 \gtrsim \frac{\epsilon\sqrt{k^\star\ln\left(\frac{N}{\alpha}\right)}}{\sqrt{N}}$
	2	$\epsilon N^2\,\|P^\star\mu\|_2\,\sqrt{\ln\left(\frac{N}{\alpha}\right)}$	$\|P^\star\mu\|_2 \gtrsim \epsilon\sqrt{\ln\left(\frac{N}{\alpha}\right)}$
	3	$\epsilon N\sqrt{N}\ln\left(\frac{N}{\alpha}\right)$	$\|P^\star\mu\|_2^2 \gtrsim \frac{\epsilon\ln\left(\frac{N}{\alpha}\right)}{\sqrt{N}}$
	4	$\epsilon N^2\,\|P^\star\mu\|_2^2$	$\epsilon \lesssim 1$
β_2	5	$\epsilon N\sqrt{\epsilon Nk^\star\ln\left(\frac{1}{\epsilon}\right)}$	$\|P^\star\mu\|_2^2 \gtrsim \frac{\epsilon\sqrt{\epsilon k^\star\ln\left(\frac{1}{\epsilon}\right)}}{\sqrt{N}}$
	6	$\epsilon^2 N^2\ln\left(\frac{1}{\epsilon}\right)$	$\|P^\star\mu\|_2^2 \gtrsim \epsilon^2\ln\left(\frac{1}{\epsilon}\right)$
	7	$\|P^\star\mu\|_2\,\epsilon^2 N^2\sqrt{\ln\left(\frac{1}{\epsilon}\right)}$	$\|P^\star\mu\|_2^2 \gtrsim \epsilon^2\sqrt{\ln\left(\frac{1}{\epsilon}\right)}$
	8	$\|P^\star\mu\|_2^2\,\epsilon^2 N^2$	$\epsilon \lesssim 1$
$\epsilon N\gamma$	9	$\epsilon N\sqrt{Nk^\star}$	$\|P^\star\mu\|_2^2 \gtrsim \frac{\epsilon\sqrt{k^\star}}{\sqrt{N}}$
	10	$\epsilon N\sqrt{N\ln\left(\frac{1}{\alpha}\right)}$	$\|P^\star\mu\|_2^2 \gtrsim \frac{\epsilon\sqrt{\ln\left(\frac{1}{\alpha}\right)}}{\sqrt{N}}$
	11	$\epsilon N\ln\left(\frac{1}{\alpha}\right)$	Dominated by the term 3
	12	$\epsilon^2 N^2\ln\left(\frac{1}{\epsilon}\right)$	$\|P^\star\mu\|_2^2 \gtrsim \epsilon^2\ln\left(\frac{1}{\epsilon}\right)$
R	13	$N^{3/2}\,\|P^\star\mu\|_2\,\sqrt{\ln\frac{1}{\alpha}}$	$\|P^\star\mu\|_2 \gtrsim \frac{\sqrt{\ln\left(\frac{1}{\alpha}\right)}}{\sqrt{N}}$
	14	$N\sqrt{k^\star}\ln\left(\frac{1}{\alpha}\right)$	$\|P^\star\mu\|_2^2 \gtrsim \frac{\sqrt{k^\star}\ln\left(\frac{1}{\alpha}\right)}{N}$
	15	$\ln\left(\frac{1}{\alpha}\right)$	Dominated by the term 3

Table 2: terms from (C.23) and corresponding required conditions on $\|P^\star\mu\|_2$.

we have $k_1^\star \leq k_2^\star$; similarly, if $\epsilon \geq \frac{1}{\sqrt{N}}$, we have $k_1^\star \geq k_2^\star$. We discuss the two possibilities in the following.

(i): $\epsilon \leq \frac{1}{\sqrt{N}}$. In this case, the dominant term in $E^2(\epsilon, K, N, \sigma^2)$ is $\frac{\sqrt{k_1^\star}}{N}\sigma^2$. If $k_1^\star = k_2^\star$. Obviously, we know $\frac{\sqrt{k_1^\star}}{N}\sigma^2$ in (3.8) will also be the dominant. Otherwise, we have $k_1^\star + 1 \leq k_2^\star$. By the definition and the monotone property of Kolmogorov widths, we have

$$\frac{(k_2^\star)^{\frac{1}{4}}\sqrt{\epsilon}}{N^{\frac{1}{4}}}\sigma \leq D_{k_2^\star - 1} \leq D_{k_1^\star}(K) \leq \frac{(k_1^\star + 1)^{\frac{1}{4}}}{\sqrt{N}}\sigma.$$

This leads to

$$\epsilon \leq \frac{1}{\sqrt{N}}\sqrt{\frac{k_1^\star + 1}{k_2^\star}} \lesssim \frac{1}{\sqrt{N}}\sqrt{\frac{k_1^\star}{k_2^\star}}.$$

With a little calculation, one can verify that this is equivalent to saying $\frac{\sqrt{k_1^\star}}{N}\sigma^2$ is the dominant in (3.8) again. Hence, (3.8) matches $E(\epsilon, K, N, \sigma^2)$ in this possibility.

(ii): $\epsilon \geq \frac{1}{\sqrt{N}}$. In this case, the dominant term in $E^2(\epsilon, K, N, \sigma^2)$ is either $\epsilon\sqrt{\ln\left(\frac{1}{\epsilon}\right)\frac{k_2^\star}{N}}\sigma^2$ or $\epsilon^2 \ln\left(\frac{1}{\epsilon}\right)\sigma^2$. If $k_1^\star = k_2^\star$, then $\epsilon\sqrt{\frac{k_2^\star}{N}}\sigma^2$ will dominate over $\frac{\sqrt{k_1^\star}}{N}\sigma^2$ and consequently the dominant term in (3.8) is either $\epsilon\sqrt{\frac{k_2^\star}{N}}\sigma^2$ or $\epsilon^2\sigma^2$, which nearly matches $E^2(\epsilon, K, N, \sigma^2)$ with only an exception of a $\ln\left(\frac{1}{\epsilon}\right)$ factor. Otherwise, we have $k_1^\star \geq k_2^\star + 1$. Again, by the definition and the monotone property of Kolmogorov widths, we have

$$\frac{(k_1^\star)^{\frac{1}{4}}}{\sqrt{N}}\sigma \leq D_{k_1^\star - 1}(K) \leq D_{k_2^\star}(K) \leq \frac{(k_2^\star + 1)^{\frac{1}{4}}\sqrt{\epsilon}}{N^{\frac{1}{4}}}\sigma,$$

which leads to

$$\epsilon \geq \frac{1}{\sqrt{N}}\sqrt{\frac{k_1^\star}{k_2^\star + 1}}.$$

If $k_2^\star = 0$, then the dominant terms in (3.8) and $E^2(\epsilon, K, N, \sigma^2)$ are $\epsilon^2\sigma^2$ and $\epsilon^2 \ln\left(\frac{1}{\epsilon}\right)\sigma^2$, respectively, which match except for a $\ln\left(\frac{1}{\epsilon}\right)$ term. Otherwise, we will have

$$\epsilon \gtrsim \frac{1}{\sqrt{N}}\sqrt{\frac{k_1^\star}{k_2^\star}},$$

which again can be used to verify that $\epsilon\sqrt{\frac{k_2^\star}{N}}\sigma^2$ dominates over $\frac{\sqrt{k_1^\star}}{N}\sigma^2$ in (3.8). In both cases, (3.8) and $E^2(\epsilon, K, N, \sigma^2)$ match with only an exception of a $\ln\left(\frac{1}{\epsilon}\right)$ factor. The other part of the corollary is directly from the expression of $E^2(\epsilon, K, N, \sigma^2)$ and $P^2(\epsilon, K, N, \sigma^2)$. The proof is completed. $\square$

C.8   Proof of Theorem 4.5

Proof. The techniques are parallel to the proof of the lower bounds above. By Lemma 4.3, for $k = k_{p,1}^\star$, $c = \frac{(k_{p,1}^\star)^{1/p - 1/4}}{\sqrt{N}}\sigma$, there exists a $\theta \in K$ such that $\|\theta\|_p = \frac{(k_{p,1}^\star)^{1/p - 1/4}}{\sqrt{N}}\sigma$ and $\|\theta\|_\infty \leq \frac{1}{(k_{p,1}^\star)^{1/4}\sqrt{N}}\sigma$. Again, we impose a uniform prior on the set $\{\mu \mid \mu \in K, \mu_i = \kappa\gamma_i\theta_i, \gamma_i \in \{-1, 1\}, i \in I\}$ for some

positive constant κ. By the same arguments of Appendix C.2, if $\mathbb{E}_{\mu,\mu'} \exp\left\{\frac{N\mu^\top \mu'}{\sigma^2}\right\} - 1$, or equivalently $\frac{N^2 \|\theta\|_4^4 \kappa^4}{2\sigma^4}$ is sufficiently small, then there is no test with uniformly small type-I and type-II errors. We have

$$\frac{N^2 \|\theta\|_4^4 \kappa^4}{2\sigma^4} \leq \frac{N^2 \sum_{i=1}^{d} \|\theta\|_\infty^{4-p} |\theta_i|^p \kappa^4}{2\sigma^4} = \frac{N^2 \|\theta\|_\infty^{4-p} \|\theta\|_p^p \kappa^4}{2\sigma^4} \leq \kappa^4.$$

This implies that it is necessary to ensure $\kappa \gtrsim 1$, from which we establish the first lower bound that $\rho_{\text{critical}} \gtrsim \frac{(k_{p,1}^\star)^{1/p-1/4}}{\sqrt{N}} \sigma$.

Similarly, it suffices to bound $\frac{\sqrt{N}\|v\|_4^2}{\sigma^2}$ for the second term, where v is uniform from the set

$$V := \{v \mid v_i = \gamma_i \theta_i, \gamma_i \in \{-1,1\}, i \in I\},$$

and $\theta \in K$ from Lemma 4.3 satisfies $\|\theta\|_p = \frac{\sqrt{\epsilon}(k_{p,2}^\star)^{1/p-1/4}}{N^{1/4}} \sigma$ and $\|\theta\|_\infty \leq \frac{\sqrt{\epsilon}}{(Nk_{p,2}^\star)^{1/4}} \sigma$. Consequently, it is not difficult to verify that $\frac{\sqrt{N}\|v\|_4^2}{\sigma^2} \leq \epsilon$. By Lemma C.1, we establish the second lower bound that $\rho_{\text{critical}} \gtrsim \frac{\sqrt{\epsilon}(k_{p,2}^\star)^{1/p-1/4}}{N^{1/4}} \sigma$.

Suppose now that θ is the vector from Lemma 4.3 when $k = k_{p,3}^\star$. Subsequently, we know $\|\theta\|_2^2 \leq \|\theta\|_p^p \cdot \|\theta\|_\infty^{2-p} \leq \epsilon^2 \sigma^2$. Following the same reasoning in Appendix C.4, we know there always exists a strategy for $\mathcal{C}$ to discourage the testing. Therefore, we establish the third lower bound. The proof is completed. $\qquad\square$

C.9 Proof of Theorem 4.6

Before we prove Theorem 4.6, we need the following observation of the p-convex orthosymmetric set.

Lemma C.18 (Truncation of a p-convex orthosymmetric set). *Assume $K \subset \mathbb{R}^d$ is a p-convex orthosymmetric set. For any $1 \leq j \leq d$ and $J := \{i_1, \ldots, i_j\} \subset \{1, \ldots, d\}$, define the Truncation of K on J as*

$$K_J := \{\theta_J \mid \theta_J = (\theta_{i_1}, \ldots, \theta_{i_j}) \in \mathbb{R}^j, \exists \beta \in K, \beta_J = \theta_J\},$$

where $\beta_J = (\beta_{i_1}, \ldots, \beta_{i_j})$ is the truncated β on the index set J. Then $K_J \subset \mathbb{R}^j$ is also p-convex orthosymmetric.

Proof. Lemma C.18 is straightforward from the definition. For any $\theta_{J,1}, \theta_{J,2} \in K_J$, suppose $\theta_{J,1} = \beta_{1,J}, \theta_{J,2} = \beta_{2,J}, \beta_1, \beta_2 \in K$, then we know for any $\lambda \in [0,1]$, we have $\lambda \theta_{J,1}^p + (1-\lambda)\theta_{J,2}^p = \lambda \beta_{1,J}^p + (1-\lambda)\beta_{2,J}^p = (\lambda \beta_1^p + (1-\lambda)\beta_2^p)_J \overset{\text{(i)}}{=} \widetilde{\beta}_J^p \in K_J^p$, where $\widetilde{\beta} \in K^p$ and (i) is from that K is p-convex. Consequently, K_J is p-convex.

Suppose $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d), \xi_i \in \{-1,1\}$, then for any $\theta_J \in K_J$, we have $\boldsymbol{\xi}_J \odot \theta_J = \boldsymbol{\xi}_J \odot \beta_J = (\boldsymbol{\xi} \odot \beta)_J \overset{\text{(ii)}}{=} \bar{\beta}_J \in K_J$, where $\bar{\beta} \in K$, $\odot$ is element-wise multiplication and (ii) is from that K is orthosymmetric. Consequently, K_J is also orthosymmetric. $\qquad\square$

We are now prepared to prove the original theorem.

Proof. Suppose first $\epsilon \leq \frac{1}{\sqrt{N}}$. In this case, we have $\frac{j^{1/p-1/4}}{\sqrt{N}}\sigma \geq \frac{\sqrt{\epsilon}j^{1/p-1/4}}{N^{1/4}}\sigma$ for any $0 \leq j \leq d$. By the monotonicity of Kolmogorov widths and the definitions of $k_{p,1}^\star$ and $k_{p,2}^\star$, we know $k_{p,1}^\star \leq k_{p,2}^\star$. Later calculations show that in this case we also have $k_{p,2}^\star \leq k_{p,3}^\star$. This implies $k_{p,1}^\star \leq k_{p,2}^\star \leq k_{p,3}^\star$. For $k_{p,1}^\star$ and its corresponding $P_{p,1}^\star$, without loss of generality we assume $P_{p,1}^\star\theta = (\theta_1,\ldots,\theta_{k_{p,1}^\star},0,\ldots,0)$. Regard $P_{p,1}^\star\theta$ as a vector in $\mathbb{R}^{k_{p,1}^\star}$, by the Hölder's inequality B.17, we know $\|P_{p,1}^\star\theta\|_2 \gtrsim \frac{(k_{p,1}^\star)^{1/4}}{\sqrt{N}}\sigma$ for any $\theta \in K$. Further define $k_{p,1,1}^\star := \max\left\{j \mid 0 \leq j \leq k_{p,1}^\star, D_{j-1}(K) > \frac{j^{1/4}}{\sqrt{N}}\sigma\right\}$. Since $\epsilon \leq \frac{1}{\sqrt{N}}$, by the [proof](#) of Corollary 3.15 and Lemma C.18, we know that if $\|P_{p,1}^\star\theta\|_2 \gtrsim \frac{(k_{p,1,1}^\star)^{1/4}}{\sqrt{N}}\sigma$, then there exists a desired test. However, this is already guaranteed since $\|P_{p,1}^\star\theta\|_2 \gtrsim \frac{(k_{p,1}^\star)^{1/4}}{\sqrt{N}}\sigma$ and $k_{p,1}^\star \geq k_{p,1,1}^\star$.

Suppose now $\epsilon \geq \frac{1}{\sqrt{N}}$. Analogous to the discussion above, we know that $k_{p,2}^\star \leq k_{p,1}^\star$. If $k_{p,2}^\star = k_{p,3}^\star$, then we can assume $P_{p,2}^\star = P_{p,3}^\star$, and we have

$$\left\|P_{p,2}^\star\theta\right\|_2 \gtrsim \frac{\sqrt{\epsilon}(k_{p,2}^\star)^{\frac{1}{4}}}{N^{\frac{1}{4}}}\sigma,$$

$$\left\|P_{p,2}^\star\theta\right\|_2 \gtrsim \epsilon\sigma.$$

Since $\epsilon \geq \frac{1}{\sqrt{N}}$, we also have

$$\left\|P_{p,2}^\star\theta\right\|_2 \gtrsim \frac{(k_{p,2}^\star)^{1/4}}{\sqrt{N}}\sigma.$$

Similarly, we can regard $P_{p,2}^\star\theta$ as a vector in $\mathbb{R}^{k_{p,2}^\star}$ and define $k_{p,2,1}^\star$ and $k_{p,2,2}^\star$ accordingly as we did for $k_1^\star$ and $k_2^\star$. Since $k_{p,2}^\star \geq \max\left\{k_{p,2,1}^\star, k_{p,2,2}^\star\right\}$, we are again guaranteed to have a desired test by the inequalities above and the results in the ℓ_2 norm case.

Assume now $k_{p,2}^\star \neq k_{p,3}^\star$. Suppose we have $k_{p,2}^\star \geq k_{p,3}^\star + 1$. By the monotonicity of Kolmogorov widths and the definitions of $k_{p,2}^\star$ and $k_{p,3}^\star$, we have

$$\frac{\sqrt{\epsilon}(k_{p,2}^\star)^{\frac{1}{p}-\frac{1}{4}}}{N^{\frac{1}{4}}}\sigma \leq \widetilde{D}_{p,k_{p,2}^\star-1}(K) \leq \widetilde{D}_{p,k_{p,3}^\star}(K) < \epsilon(k_{p,3}^\star)^{\frac{1}{p}-\frac{1}{2}}\sigma,$$

from which we obtain

$$\epsilon \gtrsim \frac{(k_{p,2}^\star)^{\frac{2}{p}-\frac{1}{2}}}{(k_{p,3}^\star)^{\frac{2}{p}-1}} \cdot \frac{1}{\sqrt{N}} \gtrsim \sqrt{\frac{k_{p,3}^\star}{N}}.$$

For $k_{p,3}^\star$, we already have $\left\|P_{p,3}^\star\theta\right\|_2 \gtrsim \epsilon\sigma$. By the inequalities above, it is not hard to verify that $\left\|P_{p,3}^\star\theta\right\|_2 \gtrsim \frac{\sqrt{\epsilon}(k_{p,3}^\star)^{1/4}}{N^{1/4}}\sigma$ and $\left\|P_{p,3}^\star\theta\right\|_2 \gtrsim \frac{(k_{p,3}^\star)^{1/4}}{\sqrt{N}}\sigma$. Therefore, again by the discussion above and the results in the ℓ_2 norm case, we are guaranteed to have a desired test in $\mathbb{R}^{k_{p,3}^\star}$.

The last case is $k_{p,2}^\star \leq k_{p,3}^\star - 1$. The analysis is essentially the same as the case above. We have

$$\epsilon(k_{p,3}^\star)^{\frac{1}{p}-\frac{1}{2}}\sigma \leq \widetilde{D}_{p,k_{p,3}^\star-1}(K) \leq \widetilde{D}_{p,k_{p,2}^\star}(K) \leq \frac{\sqrt{\epsilon}(k_{p,2}^\star+1)^{\frac{1}{p}-\frac{1}{4}}}{N^{\frac{1}{4}}}\sigma,$$

which leads to

$$\epsilon \lesssim \frac{(k_{p,2}^\star)^{\frac{2}{p}-\frac{1}{2}}}{\sqrt{N}(k_{p,3}^\star)^{\frac{2}{p}}-1} \lesssim \sqrt{\frac{k_{p,2}^\star}{N}}.$$

Since we have $\left\|P^\star_{p,2}\theta\right\|_2 \gtrsim \frac{\sqrt{\epsilon}(k^\star_{p,2})^{\frac{1}{4}}}{N^{\frac{1}{4}}}\sigma$, we also have $\left\|P^\star_{p,2}\theta\right\|_2 \gtrsim \frac{(k^\star_{p,2})^{1/4}}{\sqrt{N}}\sigma$ and $\left\|P^\star_{p,2}\theta\right\|_2 \gtrsim \epsilon\sigma$, which are sufficient for us to have a desired test in the space $\mathbb{R}^{k^\star_{p,2}}$. $\qquad\square$

C.10   Verification of the Minimality in Section 5.2

Proof. Without any loss of generality we assume a_i's are sorted in decreasing order. If $k^\star = d$, the assertion is obvious since $\|P^\star\mu\|_2 = \|\mu\|_2 = c$. Suppose $k^\star < d$ in the following. If $c^2 \le a_{k^\star+1}$, then the vector

$$\mu = (\underbrace{0,\ldots,0}_{k^\star \text{ zeros}}, c, 0, \ldots, 0)$$

satisfies $P^\star\mu = \mathbf{0}$. Therefore, the assertion holds again. Let us now assume that $c^2 > a_{k^\star+1}$. Since $\|\mu\|_2 = c$, we have $\sum_{i=k^\star+1}^{d}\mu_i^2 = c^2 - \sum_{i=1}^{k^\star}\mu_i^2$. On the other hand, since $\sum_{i=1}^{d}\frac{\mu_i^2}{a_i} \le 1$ and a_i's are decreasing, we have

$$\frac{c^2 - \sum_{i=1}^{k^\star+1}\mu_i^2}{a_{k^\star+1}} \le \sum_{i=k^\star+1}^{d}\frac{\mu_i^2}{a_i} \le 1 - \sum_{i=1}^{k^\star}\frac{\mu_i^2}{a_i}.$$

Rearranging gives

$$\frac{c^2}{a_{k^\star+1}} - 1 \le \sum_{i=1}^{k^\star}\mu_i^2\left(\frac{1}{a_{k^\star+1}} - \frac{1}{a_i}\right) \le \sum_{i=1}^{k^\star}\mu_i^2\left(\frac{1}{a_{k^\star+1}} - \frac{1}{a_1}\right).$$

Hence, we conclude that $\|P^\star\mu\|_2^2 = \sum_{i=1}^{k^\star}\mu_i^2 \ge \frac{c^2/a_{k^\star+1}-1}{1/a_{k^\star+1}-1/a_1}$. It is clear that the vector given in Section 5.2 achieves this bound. The only remaining thing is to see that the vector indeed has ℓ_2 norm c and it belongs to K, which can be verified by simple algebra. The proof is completed. $\qquad\square$

C.11   Proof of Lemma 6.3

Proof. By the definition of $k^\star_1$ and $k^\star_e$ and the decreasing property of Kolmogorov widths, since $\frac{k^{1/4}}{\sqrt{N}} \le \sqrt{\frac{k}{N}}$ for any k, N, we have $k^\star_1 \ge k^\star_e$. On the other side, for any $j + 1 \ge (k^\star_e + 1)^2$, we have $D_j(K) \le D_{k^\star_e}(K) \le \sqrt{k^\star_e + 1}\sigma \le (j+1)^{\frac{1}{4}}\sigma$. Therefore, we know $k^\star_1 \lesssim (k^\star_e)^2$. To finish the proof, we discuss different cases of ϵ.

(i): $0 \le \epsilon \le \frac{1}{\sqrt{N}}$. In this case, the dominant terms in (3.8) and (6.3) are $\frac{\sqrt{k^\star_1}}{N}\sigma^2$ and $\frac{k^\star_e}{N}\sigma^2$, respectively. Since $k^\star_1 \lesssim (k^\star_e)^2$, the claim follows.

(ii): $\epsilon \ge \frac{1}{\sqrt{N}}$. In this case, we need to further discuss the relationship between $k^\star_2$ and $k^\star_e$.

(1): $k^\star_2 = k^\star_e$. When $\frac{1}{\sqrt{N}} \le \epsilon \le \sqrt{\frac{k^\star_2}{N}}$, the dominant terms in (3.8) and (6.3) are $\epsilon\sqrt{\frac{k^\star_2}{N}}\sigma^2$ and $\frac{k^\star_e}{N}\sigma^2$, respectively. However, since $\epsilon \le \frac{k^\star_2}{N}$, we know $\epsilon\sqrt{\frac{k^\star_2}{N}}\sigma^2 \le \frac{k^\star_e}{N}\sigma^2$. Otherwise, the dominant terms will be both $\epsilon^2\sigma^2$. Therefore, the claim holds for this case.

(2): $k^\star_2 + 1 \le k^\star_e$. We have

$$\sqrt{\frac{k^\star_e}{N}}\sigma \le D_{k^\star_e-1} \le D_{k^\star_2} \le \left(\frac{k^\star_2+1}{N}\right)^{\frac{1}{4}}\sqrt{\epsilon}\sigma,$$

79

which leads to

$$\epsilon \geq \frac{1}{\sqrt{N}} \cdot \frac{k_e^\star}{\sqrt{k_2^\star + 1}} \geq \sqrt{\frac{k_e^\star}{N}}.$$

Once this holds, the dominant terms will be both $\epsilon^2 \sigma^2$. Therefore, the claim holds for this case.
(3): $k_e^\star + 1 \leq k_2^\star$. We have

$$\left(\frac{k_2^\star}{N}\right)^{\frac{1}{4}} \sqrt{\epsilon}\sigma \leq D_{k_2^\star-1} \leq D_{k_e^\star} \leq \sqrt{\frac{k_e^\star + 1}{N}}\sigma,$$

which leads to

$$\frac{1}{\sqrt{N}} \leq \epsilon \leq \frac{1}{\sqrt{N}} \cdot \frac{k_e^\star + 1}{\sqrt{k_2^\star}} \lesssim \sqrt{\frac{k_e^\star}{N}}.$$

Similarly, this implies that the dominant terms in (3.8) and (6.3) are $\epsilon\sqrt{\frac{k_2^\star}{N}}\sigma^2$ and $\frac{k_e^\star}{N}\sigma^2$, respectively, and $\epsilon^2\sigma^2$ will never dominate. Since $\epsilon \lesssim \frac{1}{\sqrt{N}} \cdot \frac{k_e^\star}{\sqrt{k_2^\star}}$ by the inequalities above, we know $\epsilon\sqrt{\frac{k_2^\star}{N}}\sigma^2 \lesssim \frac{k_e^\star}{N}\sigma^2$.

The discussion above shows that the minimax rate of estimation is greater than the minimax rate of testing in all possible cases. The proof is completed. □