Optimally adaptive test for high dimensional hypotheses via minimax deficiency

Jingkun Qiu
Guanghua School of Management, Peking University,
Song Xi Chen
Department of Statistics and Data Science, Tsinghua University
and
Yumou Qiu
School of Mathematical Sciences and Center for Statistical Science, Peking University

Abstract

The detection boundary is a tool for power evaluation of a high dimensional test, which provides a binary phase transition of power in terms of signal density and strength. However, it cannot separate the L_2 and higher criticism (HC) tests under dense signals, and the L_∞ and HC tests under highly sparse signals as they share the same detection boundary. This paper proposes minimax relative deficiency and minimax absolute deficiency as sharper measures for power evaluation than the detection boundary, and develop an adaptive testing procedure by combining three basic tests via a power enhancement approach. The proposed test is robust to the unknown signal density and strength with sharp optimal relative deficiency and nearly optimal absolute deficiency over the whole signal density regime. A full comparison of the proposed test with the existing methods is provided using the minimax deficiency measures. Simulation studies and a real data application to climate change analysis are conducted to evaluate the proposed test and demonstrate its superiority.

Keywords: Detection boundary, high dimensionality, minimax optimality, power enhancement.

1 Introduction

Testing high dimensional hypotheses is a fundamental problem in modern data analysis. It has a wide range of applications, for example, gene differential expression analysis and detecting trend changes in spatial-temporal data. The last fifteen years have witnessed the development of statistical theory for testing high dimensional means based upon early works of Dempster (1958, 1960) and Bai and Saranadasa (1996). Existing high dimensional tests for means can be categorized into three basic types based on the metrics used to describe the difference between the mean, which leads to different power performance under different regimes of signals. They include the sum-of-square (L_2) test for dense signals (Chen and Qin, 2010), the maximum (L_∞) test for highly sparse signals (Cai et al., 2014), and the higher criticism (HC) test for sparse and weak signals (Donoho and Jin, 2004; Delaigle et al., 2011; Zhong et al., 2013; Chen et al., 2019). The three mean tests can be extended for testing linear regression coefficients (Ingster et al., 2010; Arias-Castro et al., 2011; Zhong and Chen, 2011; Xia et al., 2018; Qiu et al., 2018). Other tests can be formulated by different combinations or modifications of the aforementioned basic categories, for instance, the power enhancement tests by combining the L_2 and L_∞ tests (Fan et al., 2015; Yu et al., 2023) or combining the L_2 and HC tests (Ingster et al., 2010), the sum-of-power tests (Xu et al., 2016; He et al., 2021), and the projection test (Li and Li, 2022), among others. Recently, high dimensional hypothesis testing has also gained attention in text learning and signal detection in large language models (LLM), including watermark detection to distinguish LLM-generated text from its human-written counterpart (Li et al., 2025a,b).

The L_2 test statistic accumulates sample estimates in all dimensions, which is known to be powerful for dense signals but encounters a power loss for sparse signals when many dimensions bear no signal (Chen et al., 2019). The L_∞ test, on the other hand, is powerful for highly sparse signals (Donoho and Jin, 2004) but would lose power for dense signal settings (Cai et al., 2011), as it only uses the maximum sample difference. The HC test adaptively chooses a threshold to filter out the non-signal dimensions and accumulates only those with relatively large signals (Donoho and Jin, 2004; Cai et al., 2011). In this way, the thresholding statistics' variances are much reduced compared to the L_2 statistics.

The detection boundary provides an important way to evaluate the power of a high dimensional test under different signal density and strength. It characterizes a phase transition of the sum of type I and type II error rates converging to 0 or 1 with respect to signal density and strength (Ingster, 1997; Donoho and Jin, 2004). However, it cannot separate the L_2 (resp. L_∞) and HC tests under the dense (resp. highly sparse) signal regime as the detection boundaries for these tests are identical. This is because the detection boundary is a leading order performance measure, which cannot fully characterize the power performance of a high dimensional test. To fully evaluate the power performance of a test, one needs to consider higher order power information beyond the detection boundary.

In this paper, we propose two sharper minimax measures on testing power, the minimax relative deficiency (MRD) and the minimax absolute deficiency (MAD), which can quantitatively characterize an exact correspondence between the signal strength and the power function of a test. The MRD and MAD are defined as the ratio and difference of the squared signal strength between a test and the minimax optimal test under the same power and nominal size, respectively. Those two measures are called minimax deficiency because a larger value of the MRD or MAD indicates a less efficient test in the minimax sense. The classical detection boundary considered in Donoho and Jin (2004) is a degenerate case of MRD and MAD, which can be obtained by omitting the higher order information.

We develop an optimally adaptive test in terms of MRD and MAD over the entire regime of signal density (including both the dense and sparse signals), via a power enhancement formulation that combines the L_2, L_∞ and HC tests. It is shown that the proposed test is sharp optimal in terms of the MRD and outperforms all existing tests, which include the L_2, L_∞ and HC tests, and the power enhancement test by combining the L_2 and L_∞ tests in Fan et al. (2015) and Yu et al. (2023). Furthermore, the proposed test is nearly optimal in terms of the MAD, which is no larger than that of the power enhancement test that combines the L_2 and HC tests, with a strictly smaller MAD in a subset of the highly sparse signal regime. Using the proposed minimax deficiency measures, we also provide a unified framework to evaluate the power performance of a test over the entire range of signal density and conduct a full evaluation of different high dimensional tests, which draws sharper conclusions that is unable to do with the detection boundary. Those results demonstrate the necessity of combining all three basic tests for an optimally adaptive testing procedure.

The rest of this paper is organized as follows. Section 2 introduces the basic framework. Section 3 develops a new power enhancement test by combining the L_2, L_∞ and HC tests. Section 4 proposes the minimax deficiency measures MRD and MAD. Section 5 provides a full power evaluation by the proposed minimax measures. Section 6 extends the proposed method to non-Gaussian data with non-identity covariance matrices. Section 7 reports simulation studies, and Section 8 concludes the paper with discussions. A case study concerning the detection of the anthropogenic forcings to climate change and the technical proofs are relegated to the supplementary material (SM).

2 General high dimensional hypotheses

In order to include multiple high dimensional settings that encompass the one-sample and two-sample tests for means, and the tests regarding the linear regression coefficients that may involve nuisance parameters, we consider a general set-up as follows. Let $X_1, \dots, X_n$ be independent Gaussian random vectors in $\mathbb{R}^d$ with mean vectors $\theta_1, \dots, \theta_n$ and a common identity covariance matrix I_d. The different $\{\theta_i\}$ accommodates general settings including regression. We emphasize that those assumptions are made for deriving the most insightful and sharp minimax optimality properties for different types of tests. Extension of the proposed method to non-Gaussian data is discussed in Section 6.

Let $\theta = (\theta_1^{\mathrm{T}}, \dots, \theta_n^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nd}$. We consider testing general high dimensional hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta \setminus \Theta_0, \tag{1}$$

where Θ and Θ_0 are s and $(s-p)$ dimensional linear subspaces of $\mathbb{R}^{nd}$, respectively, satisfying $0 \leq s - p < s \leq nd$. The dimension of the parameter of interest is p, which is allowed to diverge to infinity and can be much larger than the sample size n. Note that $\Theta_0 = \{0\}$ if and only if the dimension of the nuisance parameter $s-p$ equals to zero. Most high dimensional testing problems can be written as a special case of the hypotheses in (1), including testing two-sample means and multi-response linear regression coefficients illustrated in Examples 2 and 3. In view of Shao and Zhang (2022) and Chernozhukov et al. (2023), most nonlinear testing problems can also be written as the hypotheses in (1) on score or influence functions.

Example 1 (One-sample mean hypotheses and watermark detection)**.** Let $X_1, \dots, X_n$ be independent from $\mathcal{N}(\mu, I_p)$ so that $\theta_i = \mu$ for $i = 1, \dots, n$ and thus $d = p = s$. We are interested in testing $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$ with the natural test statistic $Z_j = n^{-1/2} \sum_{i=1}^n X_{ij} \sim \mathcal{N}(\sqrt{n}\mu_j, 1)$ independently for $j = 1, \dots, p$. Similar hypotheses were considered in Li et al. (2025a,b) to detect watermarks for LLMs. They constructed a sequence of well-designed pivot statistics $V_1, \dots, V_p$, which are invariant to the next-token prediction distribution and independently follow a common known distribution ν_0 under the null hypotheses that an article is written by a human. They considered testing

$H_0 : V_1, \ldots, V_p \overset{i.i.d.}{\sim} \nu_0$ for the watermark detection problem. They used a test statistic $\sum_{j=1}^{p} h(V_j)$ for a given function $h(\cdot)$ in Li et al. (2025b) and the phi-divergence statistic, a generalization of the higher criticism, in Li et al. (2025a). As each V_j can be transformed to follow $\mathcal{N}(0,1)$ under H_0, our results are applicable to the watermark detection problem.

Example 2 (Two-sample mean hypotheses)**.** Let $Y_{11}, \ldots, Y_{1n_1}$ and $Y_{21}, \ldots, Y_{2n_2}$ be two independent Gaussian samples in $\mathbb{R}^p$ with means μ_1 and μ_2 respectively and a common identity covariance matrix I_p. The hypotheses of interest are $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$ (Bai and Saranadasa, 1996; Chen and Qin, 2010; Cai et al., 2014; Chen et al., 2019). This is a special setting of (1) with $X_i = Y_{1i}$, $\theta_i = \mu_1$ for $1 \leq i \leq n_1$ and $X_{n_1+i} = Y_{2i}$, $\theta_{n_1+i} = \mu_2$ for $1 \leq i \leq n_2$, $n = n_1 + n_2$, $d = p$ and $s = 2p$.

Example 3 (Multi-response linear regression)**.** Let $X_i \sim \mathcal{N}(B^{\mathrm{T}} W_i, I_d)$ independently for $1 \leq i \leq n$, where $(W_1, \ldots, W_n) \in \mathbb{R}^{q \times n}$ is a fixed covariate matrix of rank $q \leq n$, and $B \in \mathbb{R}^{q \times d}$ is the regression coefficient matrix. The hypotheses of interest are $H_0 : AB = 0$ vs. $H_1 : AB \neq 0$ for a pre-specified matrix $A \in \mathbb{R}^{r \times q}$ of rank $r \leq q$ (Xia et al., 2018; Qiu et al., 2018; Li and Li, 2022). By taking $\theta_i = B^{\mathrm{T}} W_i$, this is a special case of (1) with $p = rd$ and $s = qd$. When $d = 1$, the setting reduces to the single-response linear regression considered in Ingster et al. (2010) and Arias-Castro et al. (2011).

To facilitate the analysis, we present the canonical form of the high dimensional hypotheses in (1) as that in Lehmann and Romano (2022, page 316). Let $Q = (Q_1, \ldots, Q_{nd})^{\mathrm{T}}$ be an $nd \times nd$ orthogonal matrix such that the first s rows $Q_1, \ldots, Q_s$ span Θ while the rows $Q_{p+1}, \ldots, Q_s$ span Θ_0. Let $X = (X_1^{\mathrm{T}}, \ldots, X_n^{\mathrm{T}})^{\mathrm{T}} \sim \mathcal{N}(\theta, I_{nd})$ be the vectorization in $\mathbb{R}^{nd}$ for the data. Let $\mu = (\mu_1, \ldots, \mu_{nd})^{\mathrm{T}} = Q\theta$ and $Z = (Z_1, \ldots, Z_{nd})^{\mathrm{T}} = QX \sim \mathcal{N}(\mu, I_{nd})$. As the first s rows in Q span the full parameter space Θ while the remaining $nd - s$ rows are orthogonal to Θ, we have $\mu_{s+1} = \cdots = \mu_{nd} = 0$ under both H_0 and H_1. Then, the canonical form of the hypotheses in (1) is written as

$$H_0 : \mu_1 = \cdots = \mu_p = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \mu_j \neq 0 \text{ for } 1 \leq j \leq p, \tag{2}$$

where $\mu_{p+1}, \ldots, \mu_s$ are nuisance parameters as they are unrestricted under H_0 and H_1.

The canonical form in (2) offers great convenience for the analysis. For the two-sample mean problem in Example 2, we may take

$$Q_j = \begin{cases} \sqrt{\dfrac{n_2}{n_1 n}} \displaystyle\sum_{i=1}^{n_1} e_{j+(i-1)p} - \sqrt{\dfrac{n_1}{n_2 n}} \displaystyle\sum_{i=1}^{n_2} e_{j+(n_1+i-1)p}, & j = 1, \ldots, p, \\ \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} e_{j+(i-2)p}, & j = p+1, \ldots, 2p, \end{cases} \tag{3}$$

and $Q_{2p+1}, \ldots, Q_{np}$ can be arbitrary vectors making Q orthogonal, where e_j is the j-th canonical basis vector in $\mathbb{R}^{np}$. Then it can be shown that $Z_j = Q_j^{\mathrm{T}} X = n_e^{1/2}(\bar{Y}_{1j} - \bar{Y}_{2j})$ for $j = 1, \ldots, p$, where $\bar{Y}_{1j}$ and $\bar{Y}_{2j}$ are the sample means of the j-th variable in the two samples and $n_e = n_1 n_2 / n$ is the effective sample size. For the multi-response linear model in Example 3, the orthogonal matrix Q can be found by considering the projection onto the corresponding column space of the covariates; see Anderson (2003, page 303) for a concrete construction. As the canonical form in (2) only focuses on the first p coordinates of μ, it suffices to use $Z_1, \ldots, Z_p$ to construct a high dimensional test.

3 Optimally adaptive test

Motivated by the power enhancement idea in Ingster et al. (2010) and Fan et al. (2015), we propose here a test that combines the three basic tests, namely the L_2, L_∞ and HC tests, together for adaptive signal detection for independent Gaussian data. We will demonstrate its optimality and adaptivity to the practically unknown signal density regime in the next two sections, and will extend the proposed test for non-Gaussian data in Section 6.

We first introduce the L_2, L_∞ and HC tests, respectively. Let z_α and g_α be, respectively, the upper α quantiles of the standard Gaussian distribution and the Gumbel distribution. The L_2 and L_∞ tests reject H_0 with limiting size α if

$$\sum_{j=1}^{p} Z_j^2 > p + z_\alpha\sqrt{2p} \quad \text{and} \quad \max_{1\le j\le p} |Z_j| > (2\log p - \log\log p - \log\pi + 2g_\alpha)^{1/2}, \tag{4}$$

respectively. Let $p_j = 2\{1-\Phi(|Z_j|)\}$ be the p-value of testing $\mu_j = 0$ by the test statistic Z_j for $j = 1,\ldots,p$ and $p_{(j)}$ be the ordered p-value such that $p_{(1)} \le \cdots \le p_{(p)}$. The higher criticism (HC) statistic in Donoho and Jin (2004) is

$$\mathrm{HC}^* = \max_{j:(\log p)/p \le p_{(j)} \le 1/2} \frac{\sqrt{p}(j/p - p_{(j)})}{\{p_{(j)}(1-p_{(j)})\}^{1/2}}, \tag{5}$$

and the HC test rejects H_0 with limiting size α if $\mathrm{HC}^* > \{2\log\log p + \log\log\log p - \log(4\pi) + 2g_\alpha\}^{1/2}$, where $\alpha \in (0,1)$ is a fixed constant.

The proposed $(L_2, L_\infty, \mathrm{HC})$ power enhancement test rejects H_0 at nominal level α if

$$T_* = T_2 + (T_\infty + T_{\mathrm{HC}})/2 > z_\alpha, \tag{6}$$

where $T_2 = (2p)^{-1/2}\sum_{j=1}^{p}(Z_j^2 - 1)$ is the L_2 test statistic as introduced in (4),

$$T_\infty = \sqrt{p}\mathbb{1}\Big[\max_{1\le j\le p}|Z_j| > \{2\log p + (2M-1)\log\log p\}^{1/2}\Big], \tag{7}$$

$$T_{\mathrm{HC}} = \sqrt{p}\mathbb{1}[\mathrm{HC}^* > \{2(M+1)\log\log p + \log\log\log p\}^{1/2}] \tag{8}$$

for a constant $M > 0$, and HC^* is the higher criticism statistic in (5). Here, T_∞ and T_{HC} in (6) are called the L_∞ and HC power enhancement components, boosting the power of the L_2 test statistic T_2 under the highly and moderately sparse signal regimes, respectively.

The test procedure (6) is based on the result $T_* \xrightarrow{d} N(0,1)$ under H_0 and the independent Gaussian setting in Section 2. This is a direct consequence of the following three facts:

$$T_2 \xrightarrow{d} N(0,1), \quad \mathbb{P}\Big(\max_{1\le j\le p} Z_j^2 - 2\log p + \log\log p \le x\Big) \to \exp\Big(-\frac{1}{\sqrt{\pi}}e^{-x/2}\Big) \quad \text{and} \tag{9}$$

$$\mathbb{P}\{(\text{HC}^*)^2 - 2\log\log p - \log\log\log p \le x\} \to \exp\Big(-\frac{1}{\sqrt{4\pi}}e^{-x/2}\Big), \tag{10}$$

where the facts in (9) follow from the central limit theorem and the limit theorem for maxima of independent Gaussian variables (Leadbetter et al., 1983), respectively, while (10) was shown in Jaeschke (1979).

Note that the cut-off value $\{2\log p + (2M-1)\log\log p\}^{1/2}$ for the L_∞ component in (7) is slightly larger than the critical value $(2\log p - \log\log p - \log\pi + 2g_\alpha)^{1/2}$ of the L_∞ test in (4). Similarly, the cut-off level $\{2(M+1)\log\log p + \log\log\log p\}^{1/2}$ of the HC component in (8) is also slightly larger than the critical value $\{2\log\log p + \log\log\log p - \log(4\pi) + 2g_\alpha\}^{1/2}$ of the HC test in (5). The additional $2M\log\log p$ term ensures that T_∞ and T_{HC} are negligible under H_0 such that $\mathbb{P}(T_\infty > 0 \mid H_0) = O\{(\log p)^{-M}\}$ and $\mathbb{P}(T_{\text{HC}} > 0 \mid H_0) = O\{(\log p)^{-M}\}$. When T_∞ or T_{HC} is nonzero, it will diverge to infinity rapidly due to the multiplicative factor of $\sqrt{p}$ in (7) and (8). As shown in the simulation studies in Section 7, it suffices to take $M = 2$ for good finite-sample performance under both H_0 and H_1.

As a comparison of the proposed test, we also consider (L_2, L_∞) and (L_2, HC) power enhancement tests that reject H_0 with limiting size α if, respectively,

$$T_2 + T_\infty > z_\alpha \quad \text{and} \quad T_2 + T_{\text{HC}} > z_\alpha. \tag{11}$$

A similar strategy that combines the L_2 and HC tests was discussed in Ingster et al. (2010) for testing a global hypothesis in the linear regression model. Note that the (L_2, L_∞) test is in the same formulation of the power enhancement test that rejects H_0 if $T_2 + \widetilde{T}_\infty > z_\alpha$ as proposed in Fan et al. (2015), where $\widetilde{T}_\infty = \sqrt{p}\mathbb{1}\{\max_{1\le j\le p}|Z_j| > (\log p)^{1/2}\log\log n\}$. The only difference is the cut-off value for the L_∞ component. Fan et al. (2015)'s proposal uses the cut-off value $(\log p)^{1/2}\log\log n$ for $\widetilde{T}_\infty$. This cut-off value is too large so that $\widetilde{T}_\infty$ cannot attain the minimax detection boundary in (13) below for the highly sparse signal regime of $\beta \in (3/4, 1)$. We therefore replace $(\log p)^{1/2}\log\log n$ with a smaller value for T_∞

in (7) to ensure that the detection boundary of T_∞ can match that of the L_∞ test in (4).

4 Minimax deficiency

To demonstrate the superiority of the proposed $(L_2, L_\infty, \mathrm{HC})$ test over the L_2, L_∞, HC, (L_2, L_∞) and (L_2, HC) test introduced in Section 3, we shall propose new power performance measures in this section with a thorough analysis. We emphasize that the new measures are able to provide more power performance information than the existing detection boundary.

4.1 Detection boundary

We first evaluate the aforementioned six tests by using the classical detection boundary. For the hypotheses in (2), the alternative hypotheses for the minimax analysis are

$$H_1(\beta, h) : (\mu_1, \ldots, \mu_p)^{\mathrm{T}} \in \mathcal{U}_p(\beta, h) \quad \text{and}$$

$$\mathcal{U}_p(\beta, h) = \Big\{(\mu_1, \ldots, \mu_p)^{\mathrm{T}} : \sum_{j=1}^{p} \mathbb{1}(\mu_j \neq 0) \geq p^{1-\beta} \text{ and } |\mu_j| \geq h \text{ for } \mu_j \neq 0\Big\}, \tag{12}$$

where $\beta \in (0,1)$ is the signal sparsity parameter, $h > 0$ is the signal strength parameter and $\mathbb{1}(\cdot)$ denotes the indicator function. Under the class $\mathcal{U}_p(\beta, h)$, there are at least $p^{1-\beta}$ nonzero coordinates with strength being at least h. The cases of $\beta \in (0, 1/2)$, $\beta \in (1/2, 3/4)$ and $\beta \in (3/4, 1)$ divide the whole signal density regime into the dense, moderately sparse and highly sparse signal regimes, respectively. Let Ψ_α be the collection of all level-α tests,

$$h_\beta(r) = \begin{cases} p^{r/2}, & 0 < \beta < \dfrac{1}{2}, \\ \sqrt{2r\log p}, & \dfrac{1}{2} < \beta < 1; \end{cases} \qquad \rho^*(\beta) = \begin{cases} \beta - \dfrac{1}{2}, & 0 < \beta < \dfrac{3}{4}, \\ (1 - \sqrt{1-\beta})^2, & \dfrac{3}{4} < \beta < 1. \end{cases} \tag{13}$$

Similar to the results in Ingster (1997), Donoho and Jin (2004) and Cai et al. (2011) for one-sample mean hypotheses without nuisance parameters, the following proposition shows that $r = \rho^*(\beta)$ is the minimax detection boundary for the canonical hypotheses in (2) with $(s-p)$-dimensional nuisance parameters over the class $\mathcal{U}_p(\beta, h_\beta(r))$.

Proposition 1. *Given a level $\alpha \in (0,1)$, a signal sparsity parameter $\beta \in (0,1)$, the signal strength specification $h_\beta(r)$ and the detection boundary $\rho^*(\beta)$ specified in (13) for a constant $r \in \mathbb{R}$, the maximin power function satisfies that as $p \to \infty$,*

$$\sup_{\psi \in \Psi_\alpha} \inf_{H_1(\beta, h_\beta(r))} \mathbb{P}(\psi \text{ rejects } H_0) \to \begin{cases} 1, & r > \rho^*(\beta), \\ \alpha, & r < \rho^*(\beta). \end{cases} \tag{14}$$

Since Donoho and Jin (2004), the detection boundary has been served as a benchmark in evaluating the power performance of high dimensional tests. In view of the minimax detection boundary in Proposition 1, a powerful test ψ with nominal size $\alpha \in (0,1)$ should achieve the detection boundary in the sense that as $p \to \infty$,

$$\inf_{H_1(\beta, h_\beta(r))} \mathbb{P}(\psi \text{ rejects } H_0) \to 1 \quad \text{for} \quad r > \rho^*(\beta). \tag{15}$$

The next proposition evaluates the L_2, L_∞ and HC tests, and the three power enhancement (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \text{HC})$ tests by the detection boundary criterion (15).

Proposition 2. *Given a nominal size $\alpha \in (0,1)$, a signal sparsity parameter $\beta \in (0,1)$, the signal strength specification $h_\beta(r)$ and the detection boundary $\rho^*(\beta)$ specified in (13) for a constant $r \in \mathbb{R}$, the criterion (15) is satisfied, respectively, for the L_2 test when $\beta \in (0, 1/2)$, for the L_∞ test when $\beta \in (3/4, 1)$, for the HC test when $\beta \in (0,1)$, for the (L_2, L_∞) test when $\beta \in (0, 1/2) \cup (3/4, 1)$, for the (L_2, HC) test when $\beta \in (0,1)$ and for the $(L_2, L_\infty, \text{HC})$ test when $\beta \in (0,1)$, as $p \to \infty$.*

Proposition 2 demonstrates that the detection boundary cannot fully distinguish the power of different tests. It is noted that both the L_2 and HC tests attain the detection boundary under the dense signal regime of $\beta \in (0, 1/2)$, and that both the L_∞ and HC tests attain the detection boundary under the highly sparse signal regime of $\beta \in (3/4, 1)$. Furthermore, the HC, (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests can all attain the detection boundary for the whole signal density regime of $\beta \in (0, 1)$. This reveals the limitation of the detection boundary in differentiating power performance and describing nontrivial power performance among different tests. In the following, we will show that the L_2, HC and L_∞ tests are the most powerful in the dense, moderately sparse and highly sparse regimes, respectively, by developing sharper minimax measures beyond the detection boundary. We will also show that the proposed $(L_2, L_\infty, \mathrm{HC})$ test is adaptive to different signal regimes and better than these three basic tests over the entire range of signal density.

4.2 Least strength function

The detection boundary only provides the leading order information of the signal strength with respect to the signal density for the separation of the power tending to α or 1. It fails to consider the nontrivial power between α and 1, which motivates us to investigate finer specifications for the signal strength parameter h in (12) that include the omitted higher order terms to distinguish the powers of two competing tests sharing the same detection boundary. To find an appropriate specification of the signal strength h of a test ψ corresponding to a nontrivial power $\delta \in (\alpha, 1)$, we solve the value of h such that $\inf_{H_1(\beta,h)} \mathbb{P}(\psi \text{ rejects } H_0) = \delta$. The solution, denoted by $\mathcal{H}(\psi; \beta, \delta)$, is called the *least strength function* of ψ, as it gives the smallest signal strength to guarantee the worst power of ψ equals to the nontrivial value δ over the alternative hypothesis $H_1(\beta, h)$. As such a solution does not always exist, the least strength function is formally defined as

$$\mathcal{H}(\psi; \beta, \delta) = \inf \Big\{ h > 0 : \inf_{H_1(\beta,h)} \mathbb{P}(\psi \text{ rejects } H_0) \geq \delta \Big\}. \tag{16}$$

Let ψ_A and ψ_B be two tests satisfying ψ_A is more powerful than ψ_B in the sense that $\inf_{H_1(\beta,h)} \mathbb{P}(\psi_A \text{ rejects } H_0) \geq \inf_{H_1(\beta,h)} \mathbb{P}(\psi_B \text{ rejects } H_0)$ for $h > 0$. Then it can be shown that $\mathcal{H}(\psi_A; \beta, \delta) \leq \mathcal{H}(\psi_B; \beta, \delta)$ for $\delta \in (\alpha, 1)$. Namely, a test with a larger power function will have a smaller least strength function. This motivates us to introduce the *minimax strength function*, which is defined as the "lower envelope" of the least strength functions over all level-α tests Ψ_α. Formally,

$$\mathcal{H}(\Psi_\alpha; \beta, \delta) = \inf_{\psi \in \Psi_\alpha} \mathcal{H}(\psi; \beta, \delta) \equiv \inf \left\{ h > 0 : \sup_{\psi \in \Psi_\alpha} \inf_{H_1(\beta,h)} \mathbb{P}(\psi \text{ rejects } H_0) \geq \delta \right\}, \tag{17}$$

where the last equality implies that the minimax strength function $\mathcal{H}(\Psi_\alpha; \beta, \delta)$ is the least strength function for the minimax optimal test. By Lemma S3 in the SM, $\mathcal{H}(\Psi_\alpha; \beta, \delta)$ can be uniquely determined by the maximin power function.

Theorem 1. *Given $\alpha, \beta \in (0, 1)$, $\delta \in (\alpha, 1)$, the squared minimax strength function satisfies*

$$\mathcal{H}^2(\Psi_\alpha; \beta, \delta) = \begin{cases} \sqrt{2}\{z_\alpha - z_\delta + o(1)\} \exp\{\rho^*(\beta) \log p\}, & 0 < \beta < 1/2, \\ 2\rho^*(\beta) \log p + \log 2 + 2 \log\{z_\alpha - z_\delta + o(1)\}, & 1/2 < \beta < 3/4, \\ 2\rho^*(\beta) \log p + \dfrac{c_1^2(\beta)}{c_2(\beta)}(\log \log p + \log \pi) + c_{\alpha,\beta}(\delta) + o(1), & 3/4 < \beta < 1 \end{cases} \tag{18}$$

as $p \to \infty$, where $c_1(\beta) = 1 - \sqrt{1 - \beta}$, $c_2(\beta) = \sqrt{1 - \beta}$, $\rho^*(\beta)$ is the minimax optimal detection boundary in (13), and $c_{\alpha,\beta}(\cdot)$ is a strictly increasing bijection from $(\alpha, 1)$ to $\mathbb{R}$ depending only on α and β.

The theorem implies that the detection boundary $r = \rho^*(\beta)$ is the leading order term of the minimax strength function $\mathcal{H}(\Psi_\alpha; \beta, \delta)$. The terms associated with the nontrivial power $\delta \in (\alpha, 1)$ appear in the higher order terms. The detection boundary only considers the binary phase transition of $r > \rho^*(\beta)$ and $r < \rho^*(\beta)$ in Proposition 1, which corresponds to the limiting power of 1 or α, respectively. It is insufficient to quantify nontrivial power performance on the detection boundary. In contrast, the minimax strength function contains the higher order power information not shown in the detection boundary. The derivation of the expression of $\mathcal{H}(\Psi_\alpha; \beta, \delta)$ at the boundary of the signal sparsity levels, say $\beta = 0, 1/2, 3/4, 1$, is similar and is omitted for simplicity.

4.3 Minimax deficiency measures

For a high dimensional test ψ, we could compare its least strength function $\mathcal{H}(\psi; \beta, \delta)$ with the minimax strength function $\mathcal{H}(\Psi_\alpha; \beta, \delta)$ for power evaluation. In order to make full use of higher order power information to evaluate the power performance of different high dimensional tests, we propose the *minimax relative deficiency (MRD)*,

$$\mathrm{MRD}(\psi; \alpha, \beta, \delta) = \frac{\mathcal{H}^2(\psi; \beta, \delta)}{\mathcal{H}^2(\Psi_\alpha; \beta, \delta)}, \tag{19}$$

and the *minimax absolute deficiency (MAD)*,

$$\mathrm{MAD}(\psi; \alpha, \beta, \delta) = \mathcal{H}^2(\psi; \beta, \delta) - \mathcal{H}^2(\Psi_\alpha; \beta, \delta). \tag{20}$$

Compared to the efficiency measures in the fixed dimensional settings (Pitman, 1949; Hodges and Lehmann, 1970), the proposed deficiency measures are defined under a non-asymptotic minimax framework in the high dimensional setting in order to more effectively evaluate the power performance of high dimensional tests beyond the detection boundary.

The two deficiency measures satisfy $\mathrm{MRD}(\psi; \alpha, \beta, \delta) \geq 1$ and $\mathrm{MAD}(\psi; \alpha, \beta, \delta) \geq 0$ for any level-α test $\psi \in \Psi_\alpha$ at any signal sparsity β and nontrivial power δ. A larger value of the MRD or MAD indicates a less efficient test in the minimax sense, which is the reason for the term deficiency. The MRD is a relative measure that compares the multiplicative factor on the leading orders of $\mathcal{H}^2(\psi; \beta, \delta)$ and $\mathcal{H}^2(\Psi_\alpha; \beta, \delta)$, which is suitable for the dense signal regime of $\beta \in (0, 1/2)$. The MAD is a contrast measure that further compares the additive higher order terms in the least strength functions, designed for the sparse signal regime of $\beta \in (1/2, 1)$. To compare two competing tests, we first compare their MRD and favor the one with the smaller MRD. If both tests' MRDs converge to 1 as $p \to \infty$, we then compare their MADs and favor the one with the smaller value.

Note that a test attaining the minimax detection boundary does not necessarily imply its MRD converging to 1 for the dense signal regime of $\beta \in (0, 1/2)$, as will be shown below for the comparison between the L_2 and HC tests. Therefore, MRD is a sharper measure than the detection boundary in this case. However, attaining the minimax detection boundary is equivalent to MRD converging to 1 for the sparse signals of $\beta \in (1/2, 1)$. In this case, MAD provides a sharper measure for comparing two tests, as will be shown below as well for the comparison between the L_∞ and HC tests.

5 Power comparison

In this section, we provide a full power evaluation of the L_2, L_∞ and HC tests, and the three power enhancement (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests by the proposed MRD and MAD for the hypotheses in (2) and (12). The following theorem provides a thorough evaluation of the two minimax measures of the six tests. The explicit expressions of their least strength functions are provided in Propositions S1–S6 in the SM.

Theorem 2. *Given a nominal size $\alpha \in (0, 1)$, a signal sparsity parameter $\beta \in (0, 1)$ and a nontrivial power $\delta \in (\alpha, 1)$, the MRDs and MADs of the L_2, L_∞ and HC tests, and the three power enhancement (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests for the hypotheses in (2) and (12) are given in Table 1, as $p \to \infty$.*

From the first three blocks in Table 1, we have the following findings about the power comparison among the L_2, L_∞ and HC tests. Under the dense signal regime of $\beta \in (0, 1/2)$,

Table 1: Minimax relative deficiencies (MRDs) and minimax absolute deficiencies (MADs) of the L_2, L_∞, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests, where $c_0(\beta) = c_1(\beta)/c_2(\beta)$, $c_1(\beta) = 1 - \sqrt{1-\beta}$, $c_2(\beta) = \sqrt{1-\beta}$, and $\tau_M = 1 - 1/(2M+2)^2 \in (3/4, 1)$ for $M > 0$.

Test	β	MRD	MAD
L_2	$(0, 1/2)$	$\to 1$	$\to 0$
	$(1/2, 3/4)$	$= O(p^{\beta-1/2}/\log p)$	$= O(p^{\beta-1/2})$
	$(3/4, 1)$		
L_∞	$(0, 1/2)$	$= O(p^{1/2-\beta}\log p)$	$= O(\log p)$
	$(1/2, 3/4)$	$= O(1)$	
	$(3/4, 1)$	$\to 1$	$= O(1)$
HC	$(0, 1/2)$	$= O((\log\log p)^{1/2})$	$\to 0$
	$(1/2, 3/4)$	$\to 1$	$= O(\log\log p)$
	$(3/4, 1)$		$\leq c_0(\beta)\{3c_1(\beta) - 1\}\log\log p + O(\log\log\log p)$
(L_2, L_∞)	$(0, 1/2)$	$\to 1$	$\to 0$
	$(1/2, 3/4)$	$= O(1)$	$= O(\log p)$
	$(3/4, 1)$	$\to 1$	$\leq c_0(\beta)\{c_1(\beta) + 2Mc_2(\beta)\}\log\log p + O(1)$
(L_2, HC)	$(0, 1/2)$	$\to 1$	$\to 0$
	$(1/2, 3/4)$		$= O(\log\log p)$
	$(3/4, 1)$		$\leq c_0(\beta)\{3c_1(\beta) - 1\}\log\log p + O(\log\log\log p)$
$(L_2, L_\infty, \mathrm{HC})$	$(0, 1/2)$	$\to 1$	$\to 0$
	$(1/2, 3/4)$		$= O(\log\log p)$
	$(3/4, \tau_M)$		$\leq c_0(\beta)\{3c_1(\beta) - 1\}\log\log p + O(\log\log\log p)$
	$[\tau_M, 1)$		$\leq c_0(\beta)\{c_1(\beta) + 2Mc_2(\beta)\}\log\log p + O(1)$

the L_2 test is sharp optimal with MRD converging to 1 and MAD converging to 0. The MRD of the HC test diverges at a slow rate of $O((\log\log p)^{1/2})$, which is inferior to the L_2 test in this regime. Under the moderately sparse signal regime of $\beta \in (1/2, 3/4)$, only the HC test has the MRD converging to 1. The MRD of the L_2 test diverges at a polynomial rate of p, and that of the L_∞ test is bounded above and does not converge to 1. These show that the HC test is optimal in the regime of $\beta \in (1/2, 3/4)$. Under the highly sparse signal regime of $\beta \in (3/4, 1)$, the MRD and MAD of the L_2 test are the largest, diverging at polynomial rates of p. Although the MRDs of both L_∞ and HC tests converge to 1, the MAD criterion shows that the L_∞ test is superior to the HC test in this regime. The MAD of the L_∞ test is bounded above, while the MAD of the HC test is at the order $O(\log\log p)$.

In summary, the proposed minimax deficiency measures, MRD and MAD, can differentiate the relative superiority of the L_2, HC and L_∞ tests under the dense, moderately sparse and highly sparse signal regimes, respectively, which provide sharper conclusions than what the detection boundary can offer in Proposition 2. However, the MRD of any one of the three tests would not converge to 1 over the whole signal density regime.

From the last three blocks in Table 1, we see that only the MRDs of the (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests converge to 1 over the entire signal regimes, indicating their optimality and adaptation to different signal density regimes. The MRD of the (L_2, L_∞) test does not converge to 1 for $\beta \in (1/2, 3/4)$, since the L_∞ test is sub-optimal with an MRD not converging to 1 in this regime. This shows that the (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests are superior to the (L_2, L_∞) power enhancement test in terms of the MRD.

We further compare the MADs of the (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests in the highly sparse signal regime of $\beta \in (3/4, 1)$. Since $2Mc_2(\beta) \leq 2c_1(\beta) - 1$ is equivalent to $\beta \geq \tau_M := 1 - 1/(2M+2)^2 \in (3/4, 1)$, the MAD of the $(L_2, L_\infty, \mathrm{HC})$ test is smaller than that of the (L_2, HC) test for $\beta \in [\tau_M, 1)$, and the MAD of those two tests are equivalent for $\beta \in (3/4, \tau_M)$. This shows that the $(L_2, L_\infty, \mathrm{HC})$ test is superior to the (L_2, HC) test.

We emphasize that the proposed $(L_2, L_\infty, \mathrm{HC})$ test is also superior to the HC test with a smaller MRD for the dense signal regime of $\beta \in (0, 1/2)$ and a smaller MAD for $\beta \in [\tau_M, 1)$.

Although the HC test can attain the detection boundary as shown in Proposition 2, the proposed minimax deficiency measures MRD and MAD in the last section are helpful to recognize more powerful testing procedures. We also point out that the power enhancement tests have a larger MAD than the L_∞ test in the highly sparse signal regime, due to the additional $2M \log \log p$ term used in the cut-off value of T_∞ for size control, which indicates a mild power loss compared to the L_∞ test in this regime.

We also notice that a smaller M leads to a smaller τ_M and thus a smaller MAD for the $(L_2, L_\infty, \mathrm{HC})$ test over $\beta \in (3/4, 1)$. However, it does not necessarily imply that M can be arbitrarily small in practice because the parameter M is introduced for finite-sample size control as discussed in Section 3. When the sample size is sufficiently large, a smaller M can be chosen and thus a better power performance can be expected.

In summary, the proposed $(L_2, L_\infty, \mathrm{HC})$ power enhancement test is sharp optimal in the leading order term of its least strength function with the MRD converging to 1 over the whole signal density regime. For the higher order terms of the least strength function, the proposed test is also nearly optimal with the MAD bounded by $O(\log \log p)$ over the whole regime. Based on the L_2 test, it takes advantage of both L_∞ and HC tests and therefore has an MAD no worse than the other two power enhancement tests, (L_2, L_∞) and (L_2, HC) tests. With the help of the newly developed minimax deficiency measures, we establish the superiority of the proposed $(L_2, L_\infty, \mathrm{HC})$ test over the existing high dimensional tests.

6 Non-Gaussianity and non-identity covariance

In the previous sections, we consider the independent Gaussian setting for concise and clear minimax power comparison without the burden of involved conditions and estimation techniques which are nonessential to the revelation of the core issues in high dimensional testing. This has been a well adopted approach, for instance in Fan (1996); Donoho and Jin (2004); Hall and Jin (2010). Li et al. (2025a,b) also adopted a similar independent assumption to detect watermarks for large language models, as mentioned in Example 1. Having said these, analogous tests can be formulated and evaluated in more general settings.

This section considers one-sample mean hypotheses for non-Gaussian data with non-identity covariances. Let $X_1, \ldots, X_n$ be independent and identically distributed random vectors in $\mathbb{R}^p$ with mean μ and covariance matrix Σ, which may not be Gaussian distributed. Let $\sigma_j^2 = \mathrm{Var}(X_{ij})$ be the variance of the j-th variable. To test the hypotheses

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu \neq 0, \tag{21}$$

we replace the original L_2 test statistic T_2 by a standardized U-statistic $T_{2,t} = \{2n(n-1)\widehat{\mathrm{tr}(\Sigma^2)}\}^{-1/2} \sum_{i \neq j}^{n} X_i^{\mathrm{T}} X_j$ in Chen and Qin (2010), where $n(n-1)\widehat{\mathrm{tr}(\Sigma^2)} = \sum_{i \neq j}^{n} X_i^{\mathrm{T}}(X_j - \bar{X}_{(i,j)}) X_j^{\mathrm{T}}(X_i - \bar{X}_{(i,j)})$ and $\bar{X}_{(i,j)}$ is the sample mean after excluding X_i and X_j. Similarly, we replace $Z_1, \ldots, Z_p$ in the original formulations of the L_∞ and HC tests by corresponding t-statistics $t_{1,n}, \ldots, t_{p,n}$ with asymptotic p-values $p_{j,t} = 2\{1 - \Phi(t_{j,n})\}$ (Qiu et al., 2025; Qiu, 2025), where $t_{j,n} = \sqrt{n}\bar{X}_{j,n}/\hat{\sigma}_{j,n}$, $\bar{X}_{j,n} = n^{-1}\sum_{i=1}^{n} X_{ij}$ and $\hat{\sigma}_{j,n}^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_{ij} - \bar{X}_{j,n})^2$. In particular, the HC test statistic is replaced by $\mathrm{HC}_t^* = \max_{j:(\log p)^c/p \leq p_{(j),t} \leq 1/p^d} \{p_{(j),t}(1 - p_{(j),t})/p\}^{-1/2}(j/p - p_{(j),t})$, where $c \geq 1, d \in (0,1)$ are two constants and $p_{(j),t}$'s are the ordered p-values based on t-statistics. This leads to the non-Gaussian version of the proposed $(L_2, L_\infty, \mathrm{HC})$ statistic

$$T_{*,t} = T_{2,t} + (T_{\infty,t} + T_{\mathrm{HC},t})/2, \tag{22}$$

where, for a constant $M > 0$, the power enhancement components are given by

$$T_{\infty,t} = \sqrt{p}\mathbb{1}\Big[\max_{1\leq j\leq p} |t_{j,n}| > \{2\log p + (2M-1)\log\log p\}^{1/2}\Big] \quad \text{and} \tag{23}$$

$$T_{\mathrm{HC},t} = \sqrt{p}\mathbb{1}[\mathrm{HC}_t^* > \{2(M+1)\log\log p + \log\log\log p\}^{1/2}]. \tag{24}$$

Under H_0 of (21), Proposition S7 in the SM shows that as $p, n \to \infty$,

$$T_{2,t} \xrightarrow{d} N(0,1), \quad \mathbb{P}\Big(\max_{1\le j\le p} t_{j,n}^2 - 2\log p + \log\log p \le x\Big) \to \exp\Big(-\frac{1}{\sqrt{\pi}}e^{-x/2}\Big) \quad \text{and} \tag{25}$$

$$\mathbb{P}\{(\mathrm{HC}_t^*)^2 - 2\log\log p - \log\log\log p \le x\} \to \exp\Big(-\frac{1-d}{\sqrt{4\pi}}e^{-x/2}\Big) \tag{26}$$

under a non-Gaussian distribution assumption on the eighth moment of the data and a mixing dependent condition among variables up to an unknown permutation (Chen and Qin, 2010; Zhong et al., 2013). Note that (25) and (26) generalize the limit results in (9) and (10) beyond the independent Gaussian assumption and that the condition $n \to \infty$ is not required in (9) and (10) but in (25) and (26) due to the non-Gaussianity.

Based on those results, Corollary S1 shows that $T_{*,t} \xrightarrow{d} N(0,1)$ under H_0 of (21) and the conditions in Section S6 in the SM. Thus, the non-Gaussian $(L_2, L_\infty, \mathrm{HC})$ test rejects H_0 with asymptotic size α if $T_{*,t} > z_\alpha$. Similar to (11), the non-Gaussian (L_2, L_∞) and (L_2, HC) tests can be constructed as $T_{2,t} + T_{\infty,t} > z_\alpha$ and $T_{2,t} + T_{\mathrm{HC},t} > z_\alpha$, respectively.

In the following, we generalize the MRD and MAD in (19) and (20) to non-Gaussian data. As the conditions required for evaluating worst-case power in Section S2 may not be satisfied for non-Gaussian tests, we consider a set of specific alternative hypotheses. Let

$$\mathcal{H}_t(\psi;\beta,\delta) = \inf\big\{h > 0 : \mathbb{P}(\psi \text{ rejects } H_0 | \nu = \nu^*) \ge \delta\big\} \tag{27}$$

be the generalized least strength function of a test ψ, where $\nu = (\sqrt{n}\mu_1/\sigma_1, \ldots, \sqrt{n}\mu_p/\sigma_p)^{\mathrm{T}}$, and $\nu_j^* = h$ for $1 \le j \le \lceil p^{1-\beta}\rceil$ and $\nu_j^* = 0$ for $\lceil p^{1-\beta}\rceil + 1 \le j \le p$. Compared to the original least strength function $\mathcal{H}(\psi;\beta,\delta)$ in (16), $\mathcal{H}_t(\psi;\beta,\delta)$ does not take the minimum power over the class $H_1(\beta, h)$ of alternative hypotheses, as its analytic expression is generally not available under the non-Gaussian data. In view of the analysis in Section S3 of the SM, we note that $\mathcal{H}(\psi;\beta,\delta) = \mathcal{H}_t(\psi;\beta,\delta)$ under the independent Gaussian assumption when ψ satisfies regularity conditions listed in Section S2. Compared (27) to the minimax strength function in (17), we define the generalized MRD and MAD, respectively, as

$$\text{GMRD}(\psi;\alpha,\beta,\delta) = \frac{\mathcal{H}_t^2(\psi;\beta,\delta)}{\mathcal{H}^2(\Psi_\alpha;\beta,\delta)} \text{ and} \tag{28}$$

$$\text{GMAD}(\psi;\alpha,\beta,\delta) = \mathcal{H}_t^2(\psi;\beta,\delta) - \mathcal{H}^2(\Psi_\alpha;\beta,\delta). \tag{29}$$

We emphasize that $\text{GMRD}(\psi;\alpha,\beta,\delta) \geq 1$ and $\text{GMAD}(\psi;\alpha,\beta,\delta) \geq 0$ are not necessarily valid in general. This is because $\mathcal{H}(\Psi_\alpha;\beta,\delta)$ is constructed under the independent Gaussian assumption while $\mathcal{H}_t(\psi;\beta,\delta)$ is constructed under the more general setting. However, the original ideal minimax strength function $\mathcal{H}(\Psi_\alpha;\beta,\delta)$ can still serve as a baseline for power performance comparison so that we can focus on the higher order power information via the generalized MRD and the generalized MAD. This is still meaningful as the goal is to compare the power performances among competing tests.

The following theorem generalizes the theoretical results in Theorem 2 for MRD and MAD to non-Gaussian distributed data with non-identity covariance matrices.

Theorem 3. *Given a nominal size $\alpha \in (0,1)$, a signal sparsity parameter $\beta \in (0,1)$ and a nontrivial power $\delta \in (\alpha,1)$, the generalized MRDs and the generalized MADs of the L_2, L_∞, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \text{HC})$ tests introduced in this section for non-Gaussian data are given in Table 2, as $p, n \to \infty$ under Conditions S5–S11 in the SM.*

From Table 2, we notice that the theoretical results for the generalized MRD and the generalized MAD are still informative, although not as sharp as those in the independent Gaussian case due to the exact data distribution is no longer available. We can still conclude that the L_2 test performs better than the L_∞ and HC tests under the dense signal regime of $\beta \in (0,1/2)$ with a better MRD bound and the MAD converging to 0. Under the moderately sparse signal regime, the HC test has a bounded MRD for $\beta \in (1/2,(2+d)/4)$ and has the MRD converging to 1 for $\beta \in ((2+d)/4,3/4)$, which is still better than

Table 2: Generalized minimax relative deficiencies (GMRDs) and generalized minimax absolute deficiencies (GMADs) of the non-Gaussian L_2, L_∞, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \text{HC})$ tests, where $c_\lambda = \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma) < \infty$ is the condition number of Σ.

Test	β	GMRD	GMAD
L_2	$(0, 1/2)$	$\in [c_\lambda^{-1}, c_\lambda] + o(1)$	$\to 0$
	$(1/2, 3/4)$	$= O(p^{\beta-1/2}/\log p)$	$= O(p^{\beta-1/2})$
	$(3/4, 1)$		
L_∞	$(0, 1/2)$	$= O(p^{1/2-\beta}\log p)$	$= O(\log p)$
	$(1/2, 3/4)$	$= O(1)$	
	$(3/4, 1)$	$\in [1 + o(1), O(1)]$	$\in [O(1), O(\log p)]$
HC	$(0, (1-d)/2)$	$= O(p^{d/2}(\log\log p)^{1/2}/\log p)$	$\to 0$
	$((1-d)/2, 1/2)$	$= O(p^{1/2-\beta}\log p)$	$= O(\log p)$
	$(1/2, (2+d)/4)$	$= O(1)$	
	$((2+d)/4, 1)$	$\to 1$	$= O(\log\log p)$
(L_2, L_∞)	$(0, 1/2)$	$\in [c_\lambda^{-1}, c_\lambda] + o(1)$	$\to 0$
	$(1/2, 3/4)$	$= O(1)$	$= O(\log p)$
	$(3/4, 1)$	$\in [1 + o(1), O(1)]$	$\in [O(\log\log p), O(\log p)]$
(L_2, HC) and $(L_2, L_\infty, \text{HC})$	$(0, 1/2)$	$\in [c_\lambda^{-1}, c_\lambda] + o(1)$	$\to 0$
	$(1/2, (2+d)/4)$	$= O(1)$	$= O(\log p)$
	$((2+d)/4, 1)$	$\to 1$	$= O(\log\log p)$

the L_2 and L_∞ tests. Likewise, the L_∞ test has more attractive MRD and MAD lower bounds under the highly sparse signal regime of $\beta \in (3/4, 1)$. For the $(L_2, L_\infty, \text{HC})$ power enhancement test, it has a better MRD than the L_∞ and HC tests for dense signals, and a better MAD than the L_2 test for sparse signals. The (L_2, HC) and $(L_2, L_\infty, \text{HC})$ tests are still more powerful than the (L_2, L_∞) test with smaller MRD. Thus, we can still conclude that the proposed $(L_2, L_\infty, \text{HC})$ test incorporates the strength of the three basic tests and is adaptive to different signal regimes under the non-Gaussian data.

7 Simulation studies

We report results from simulation experiments which were designed to examine the empirical sizes and powers of the three basic tests, namely the L_2, L_∞ and HC tests, and the three power enhancement tests, namely the (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests for both the Gaussian version in Section 3 and the non-Gaussian version in Section 6. For the Gaussian data, we also compared the proposed test with the existing combination tests in the literature, including the Cauchy combination test (Liu and Xie, 2020), the minimum p-value (min-p) test of the L_2 and L_∞ tests (Feng et al., 2024) and the Fisher combination test (Yu et al., 2024). Although the original Fisher combination test in Yu et al. (2024) was designed for testing covariance matrices, it can be ready modified for testing means by the asymptotic results in Feng et al. (2024) under the independent Gaussian assumption.

7.1 Power evaluation for Gaussian data

We first considered the two-sample mean hypotheses in Example 2, and generated independent Gaussian data $Y_{1i} \sim \mathcal{N}(\mu_1, I_p)$ for $i = 1, \dots, n_1$ and $Y_{2i} \sim \mathcal{N}(\mu_2, I_p)$ for $i = 1, \dots, n_2$. All tests were formulated based on normalized sample differences $Z_j = n_e^{1/2}(\bar{Y}_{1j} - \bar{Y}_{2j})$ for $j = 1, \dots, p$, where $n_e = n_1 n_2/(n_1 + n_2)$ is the effective sample size. We set the dimension $p = 10^3, 10^4$ and 10^5, $n_1 = n_2 = 200$ so that $n_e = 100$, and $M = 2$ for the cut-off values in the power enhancement components in (7) and (8). To mimic the null hypothesis, the means μ_1 and μ_2 were set to zero, and the nominal level was $\alpha = 0.05$. Simulation results in this study were based on 10^4 replications.

Table 3 reports the empirical sizes of the nine tests, which shows that all tests could

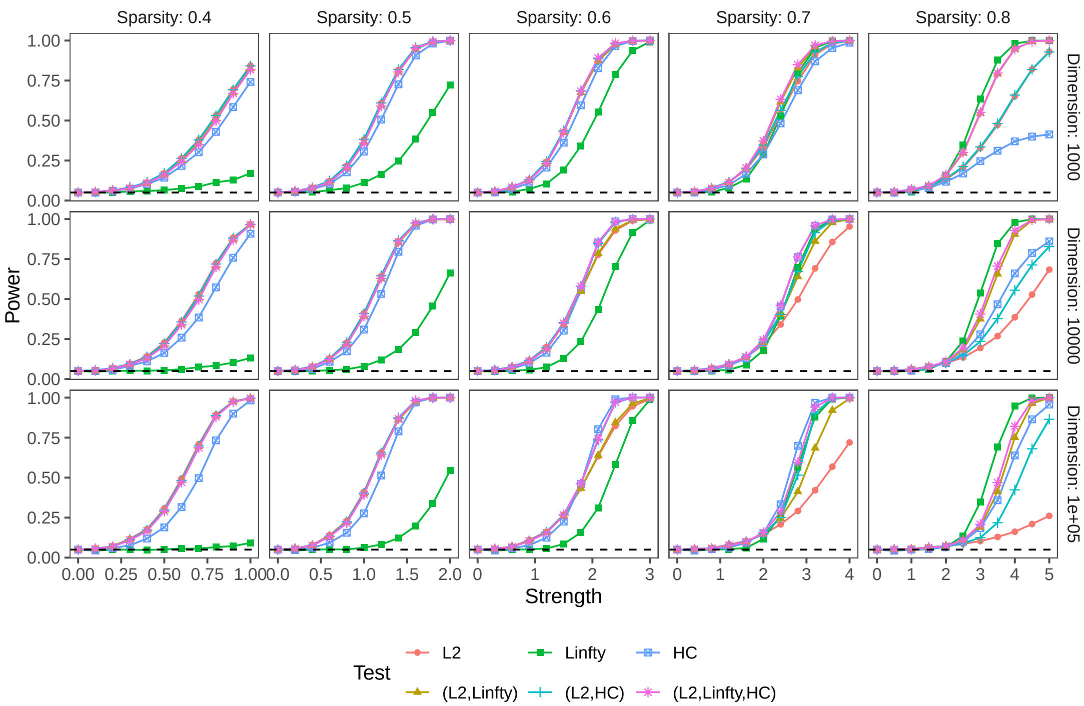
Table 3: Empirical sizes of the L_2, L_∞, HC, (L_2, L_∞), (L_2, HC), $(L_2, L_\infty, \text{HC})$ tests, and the Cauchy, min-p and Fisher combination tests at the nominal level $\alpha = 0.05$ under $H_0 : \mu_1 = \mu_2$ with $n_1 = n_2 = 200$ and $p = 10^3, 10^4$ and 10^5.

p	L_2	L_∞	HC	(L_2, L_∞)	(L_2, HC)	$(L_2, L_\infty, \text{HC})$	Cauchy	min-p	Fisher
10^3	0.055	0.045	0.066	0.064	0.058	0.067	0.052	0.049	0.059
10^4	0.050	0.048	0.060	0.055	0.053	0.058	0.052	0.051	0.054
10^5	0.049	0.051	0.062	0.052	0.052	0.055	0.055	0.051	0.054

control their empirical sizes close to the nominal level $\alpha = 0.05$. The three power enhancement tests did not have much size distortion compared to the L_2 test as $M = 2$ in the power enhancement components (7) and (8) guarantees $\mathbb{P}(T_\infty > 0) = O((\log p)^{-2})$ and $\mathbb{P}(T_{\text{HC}} > 0) = O((\log p)^{-2})$ under H_0. Although the empirical sizes of the six tests were close to the nominal 5%, for fair comparison, we adjusted the critical values based on the null simulation so that the modified empirical sizes of all the six tests were exactly 0.05.

To evaluate the power of the tests over different signal density regimes under the alternative hypotheses, we set $\mu_{1k} - \mu_{2k} = hn_e^{-1/2}$ for $1 \le k \le \lfloor p^{1-\beta} \rfloor$, $\mu_{1k} - \mu_{2k} = -hn_e^{-1/2}$ for $\lfloor p^{1-\beta} \rfloor + 1 \le k \le 2\lfloor p^{1-\beta} \rfloor$ and $\mu_{1k} - \mu_{2k} = 0$ for $k \ge 2\lfloor p^{1-\beta} \rfloor + 1$, where β ranged from 0.4 to 0.8 with an increment of 0.1. Under this setting, the number of nonzero $\mu_{1k} - \mu_{2k}$ was equal to $2\lfloor p^{1-\beta} \rfloor$, and they took the values of $hn_e^{-1/2}$ and $-hn_e^{-1/2}$ with an equal proportion. The cases of $\beta = 0.4, 0.5$, $\beta = 0.6, 0.7$, and $\beta = 0.8$ represented the dense, moderately sparse, and highly sparse signal regimes, respectively. Let $h_0 = \beta - 0.3$. For each signal sparsity level β, the signal strength was set as $h = h_0, 2h_0, \ldots, 10h_0$.

Figure 1 summarizes the empirical powers of the nine tests. Among the three basic tests, the L_2 test was the most powerful for the dense signal regime ($\beta = 0.4, 0.5$) and was powerless when the signals became sparse. In contrast, the L_∞ test was the most powerful



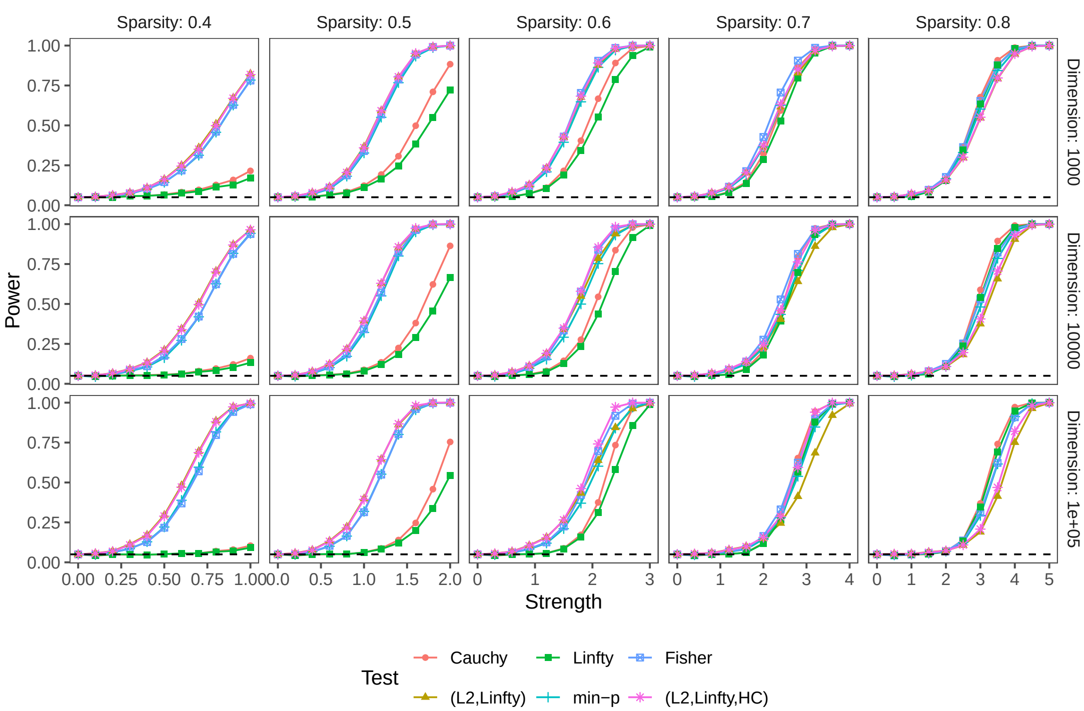


Figure 1: Empirical powers of the L_2, L_∞, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests at the nominal level $\alpha = 0.05$ under $H_1 : \mu_1 \neq \mu_2$ (top); and empirical powers of the Cauchy, min-p and Fisher combination, L_∞, (L_2, L_∞) and $(L_2, L_\infty, \mathrm{HC})$ tests (bottom).

among the three basic tests for the highly sparse signal regime ($\beta = 0.8$) and was powerless when the signals became dense and weak. The HC test was inferior to the L_2 and L_∞ tests under the dense and highly sparse signal regimes, respectively. It was the most powerful one for the moderately sparse signal regime ($\beta = 0.6, 0.7$) and when the dimension p was sufficiently large ($p = 10^4, 10^5$). These empirical findings coincide with the theoretical results shown in Table 1 using the proposed MRD and MAD and demonstrate that any of the three basic tests suffered from power loss under certain signal density regimes.

The powers of all of the three power enhancement tests were almost identical to that of the L_2 test under the dense signal regime of $\beta = 0.4, 0.5$, which were higher than those of the L_∞ and HC tests. Under this regime, the L_2 test is already sharp optimal as shown in Table 1, and the T_∞ and T_{HC} power enhancement components did not gain extra power. For the moderately sparse signal regime of $\beta = 0.6, 0.7$, the powers of the (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests were similar to that of the HC test, and they were more powerful than the (L_2, L_∞) test as well as the L_2 and L_∞ tests. For the highly sparse signal regime of $\beta = 0.8$, the proposed $(L_2, L_\infty, \mathrm{HC})$ test was the most powerful among the three power enhancement tests, followed by the (L_2, L_∞) and (L_2, HC) tests in sequence. Although the $(L_2, L_\infty, \mathrm{HC})$ test was slightly inferior to the L_∞ test under this regime, as expected from their difference in the MAD, the proposed test was at least the second best among all six tests. Given the superior performance of the proposed $(L_2, L_\infty, \mathrm{HC})$ test under the dense and moderately sparse signals, the overall rank of its power performance over all signal regimes we considered was the highest among the six tests in the top panel of Figure 1.

For the three combination tests, we noticed that the Cauchy combination test had similar but slightly larger empirical powers than the L_∞ test, and the Fisher combination test had similar but slightly larger empirical powers than the min-p test. For the dense signal regime of $\beta = 0.4, 0.5$, the empirical powers of the (L_2, L_∞) and $(L_2, L_\infty, \text{HC})$ tests were higher than the other four tests in the bottom panel of Figure 1, followed by the min-p and Fisher tests, and then followed by the Cauchy and L_∞ tests. For $\beta = 0.6$, the proposed $(L_2, L_\infty, \text{HC})$ test was still most powerful, followed by the Fisher, (L_2, L_∞) and min-p tests, and then followed by the Cauchy and L_∞ tests. For $\beta = 0.7$, all the six tests in the bottom of Figure 1 had similar power behaviors, except that the (L_2, L_∞) power enhancement test was slightly powerless when $p = 10^5$. For the highly sparse signal regime of $\beta = 0.8$, the Cauchy and L_∞ tests were most powerful, but all other tests were comparable. The overall rank of the power performance of the proposed $(L_2, L_\infty, \text{HC})$ test was still the best among the six tests presented in the bottom panel of Figure 1.

We therefore conclude that the aforementioned simulation studies supported the theoretical findings in Section 5, and demonstrated the superiority of the proposed $(L_2, L_\infty, \text{HC})$ test over the existing high dimensional testing procedures.

7.2 Non-Gaussian tests and nuisance parameter estimation

To evaluate the proposed non-Gaussian power enhancement test in Section 6, we generated independent random vectors $Y_{ki} = \mu_k + \Sigma_k^{1/2} X_{ki}$ for $i = 1, \ldots, n_k$ and $k = 1, 2$, where $\Sigma_1 = (\sigma_{1,j_1j_2}) = (0.3^{|j_1-j_2|})$ and $\Sigma_2 = (\sigma_{2,j_1j_2}) = (0.5^{|j_1-j_2|})$ so that the components of Y_{ki} were correlated, and X_{kij}'s were generated from independent centered Gamma random variables with shape parameter 4 and scale parameter 2. The hypotheses of interest were $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ and the two-sample extensions of the non-Gaussian tests presented in Section 6 were examined in this simulation study. For H_1, we set $\mu_2 = 0$, and μ_1 had $2\lfloor p^{1-\beta} \rfloor$ nonzero components with value $hn_e^{-1/2}$ and other components of μ_1 was zero, where $n_e = n_1 n_2/(n_1 + n_2)$ and the configurations of β and h were the same as those

Table 4: Empirical sizes of the non-Gaussian L_2, L_∞, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests for the raw data Y_{ki} and the transformed data ΩY_{ki} and $\hat{\Omega}_\tau Y_{ki}$ with the oracle and estimated Ω, respectively, at the nominal level $\alpha = 0.05$ under $H_0 : \mu_1 = \mu_2$.

Data	L_2	L_∞	HC	(L_2, L_∞)	(L_2, HC)	$(L_2, L_\infty, \mathrm{HC})$
Raw	0.066	0.048	0.080	0.075	0.071	0.080
Oracle transformation	0.058	0.040	0.058	0.064	0.059	0.065
Estimated transformation	0.021	0.039	0.040	0.029	0.021	0.029

in the Gaussian setting. The nonzero components of μ_1 were uniformly allocated among $\{1, \ldots, p\}$ and were kept fixed once generated. We set $n_1 = n_2 = 200$ and $p = 10^3$, $M = 2$, and the simulation results in this study were based on 10^3 replications. For the tuning parameters in the HC_t^* statistic in Section 6, we set $c = 1$ and $d = 1 - (\log n)/(\log p)$.

We also considered the tests based on the transformed data ΩY_{ki} using the oracle transformation matrix Ω and $\hat{\Omega}_\tau Y_{ki}$ using the estimated transformation matrix $\hat{\Omega}_\tau$, where $\Omega = (n_1 + n_2)(n_2\Sigma_1 + n_1\Sigma_2)^{-1}$ and $\hat{\Omega}_\tau$ is the banding Cholesky estimator of Ω in Chen et al. (2019) with a bandwidth parameter τ, which was set as 1 in this simulation study. The data transformation approach was suggested to further enhance the power performance in Cai et al. (2014) for the L_∞ test and Hall and Jin (2010) and Chen et al. (2019) for HC type tests. As Ω is typically unknown in practice and therefore a high dimensional nuisance parameter, we compared the results for ΩY_{ki} with those for $\hat{\Omega}_\tau Y_{ki}$ to examine the effect of nuisance parameter estimation on the power performances of non-Gaussian tests.

Table 4 reports the empirical sizes of the six non-Gaussian tests for the original data and the transformed data with the true Ω and its estimate $\hat{\Omega}_\tau$. We noticed that the non-Gaussian L_2, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \mathrm{HC})$ tests under the raw data experienced slight size distortion comparing to their respective independent Gaussian counterparts in Table 3. The size distortion was reduced under the oracle data transformation and all of the six tests became slightly conservative under the estimated transformation. For a fair power comparison, we adjusted the critical values as in previous simulation studies so that the modified empirical sizes of all tests were exactly equal to the nominal level 0.05.

Figure 2 summarizes the empirical powers of the six non-Gaussian tests. By comparing the first row of Figure 2 with the first row in the top panel of Figure 1, we found that all of the six non-Gaussian tests behaved similarly to their counterparts under the independent Gaussian assumption. In particular, the proposed $(L_2, L_\infty, \mathrm{HC})$ test still performed among the best of the six tests. By comparing the first and second rows of Figure 2, we noticed that all of the six tests under data transformation had slightly stronger powers than those under the original data, confirming the findings in Hall and Jin (2010), Cai et al. (2014) and Chen et al. (2019). Finally, by comparing the second and last rows of Figure 2, we concluded that all of the six tests performed robustly to the estimation of Ω in this study. The estimation of the high dimensional nuisance parameters in the transformation matrix Ω did not perceptibly affect the power performance of the proposed tests.

8 Discussion

We develop new shaper minimax measures, the MRD and MAD, than the detection boundary to fully quantify the power performance of a high dimensional test. It is revealed that the existing L_2, L_∞ and HC tests are not adaptively optimal over the entire range of the signal density as each one of the tests is superior to the other two in one segment of the signal density. We propose a power enhancement test by combining those three basic tests, which is adaptively optimal with MRD converging to 1 and MAD nearly comparable to the minimax optimal likelihood ratio test.

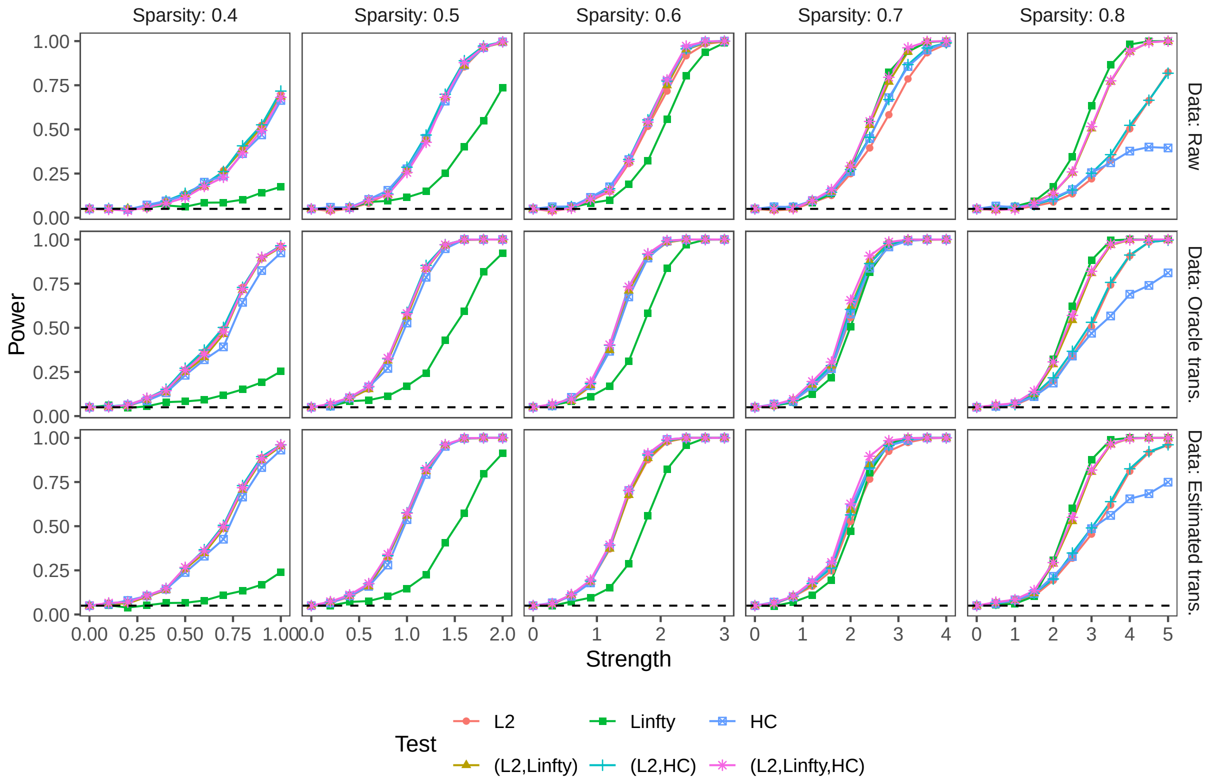


Figure 2: Empirical powers of the non-Gaussian L_2, L_∞, HC, (L_2, L_∞), (L_2, HC) and $(L_2, L_\infty, \text{HC})$ tests for the raw data Y_{ki} and the transformed data ΩY_{ki} and $\hat{\Omega}_\tau Y_{ki}$ with the oracle and estimated Ω, respectively, at the nominal level $\alpha = 0.05$ under $H_1 : \mu_1 \neq \mu_2$, where $\mu_2 = 0$, μ_1 had $2\lfloor p^{1-\beta} \rfloor$ nonzero components with value $hn_e^{-1/2}$, $\beta = 0.4, 0.5, 0.6, 0.7, 0.8$, and $h = h_0, 2h_0, \ldots, 10h_0$ for $h_0 = \beta - 0.3$.

The MRD (resp. the MAD) can be equivalently explained as the ratio of (resp. the difference between) the required sample size to ensure that the test can have a pre-specified power under the fixed alternative hypothesis to (resp. and) that of the minimax optimal likelihood ratio test with the same nominal size. Those two measures are the most informative for power evaluation, gathering higher order information in nontrivial powers ignored by the detection boundary. Our power analysis illustrate the utilities of the minimax deficiency measures over the detection boundary.

The proposed power enhancement test is formulated based on the L_2 test by adding the L_∞ and HC for power enhancement, which makes the testing procedure adaptive to the practically unknown signal density. As a compromise, the proposed test is slightly not as powerful as the L_∞ test for the highly sparse signal regime of $\beta \in (3/4, 1)$ with a slowly diverging MAD shown in Table 1. An alternative formulation can be made by exchanging the positions of the L_2 and L_∞ components so that the new formulation is based on the L_∞ statistic with the power enhancement from the L_2 and HC tests. The latter formulation would be more powerful for sparse signals as measured by the MAD, while having slightly inferior MRD in the dense range of the signal sparsity.

Similar to other combination methods, the proposed test cannot do better than an oracle testing procedure that uses the L_2, HC and L_∞ test when $\beta \in (0, 1/2)$, $\beta \in (1/2, 3/4)$ and $\beta \in (3/4, 1)$ respectively, as if the signal density β were known a prior. This is the price of the adaptive inference. On the other hand, the oracle test (and thus the proposed test) cannot achieve the sharp optimality with MAD converging to 0 uniformly for $\beta \in (0, 1)$ because none of the three basic tests has this property for any particular signal density level of $\beta \in (1/2, 1)$ as shown in Table 1. To our knowledge, the only test that can achieve the sharp optimality with a zero MAD uniformly for $\beta \in (0, 1)$ is the minimax optimal likelihood ratio test introduced in Section S4 in the SM, which requires the knowledge of both signal density β and signal strength h, and is thus not practical or adaptive.

Acknowledgments. The authors thank anonymous reviewers for their helpful comments that have improved the presentation of the paper. This work was partially supported by Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM801), National Natural Science Foundation of China grants 12292980, 12292983 and 92358303. There are no competing interests to declare.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley Series in Probability and Statistics. John Wiley & Sons, NJ.

Arias-Castro, E., Candès, E. J., and Plan, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.*, 39(5):2533–2556.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica*, 6(2):311–329.

Cai, T. T., Jeng, X. J., and Jin, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(5):629–662.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 76(2):349–372.

Chen, S. X., Li, J., and Zhong, P.-S. (2019). Two-sample and ANOVA tests for high dimensional means. *Ann. Statist.*, 47(3):1443–1474.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, 38(2):808–835.

Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2023). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.*, 10:427–449.

Delaigle, A., Hall, P., and Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(3):283–301.

Dempster, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.*, 29:995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16:41–50.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.*, 91(434):674–688.

Fan, J., Liao, Y., and Yao, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4):1497–1541.

Feng, L., Jiang, T., Li, X., and Liu, B. (2024). Asymptotic independence of the sum and maximum of dependent random variables with applications to high-dimensional tests. *Statist. Sinica*, 34(3):1745–1763.

Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.*, 38(3):1686–1732.

He, Y., Xu, G., Wu, C., and Pan, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *Ann. Statist.*, 49(1):154–181.

Hodges, Jr., J. L. and Lehmann, E. L. (1970). Deficiency. *Ann. Math. Statist.*, 41:783–801.

Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.*, 6(1):47–69.

Ingster, Y. I., Tsybakov, A. B., and Verzelen, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.*, 4:1476–1526.

Jaeschke, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.*, 7(1):108–115.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, NY.

Lehmann, E. L. and Romano, J. P. (2022). *Testing Statistical Hypotheses*. Springer, NY.

Li, C. and Li, R. (2022). Linear hypothesis testing in linear models with high-dimensional responses. *J. Amer. Statist. Assoc.*, 117(540):1738–1750.

Li, X., Ruan, F., Wang, H., Long, Q., and Su, W. J. (2025a). Robust detection of watermarks for large language models under human edits. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, qkaf056.

Li, X., Ruan, F., Wang, H., Long, Q., and Su, W. J. (2025b). A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules. *Ann. Statist.*, 53(1):322–351.

Liu, Y. and Xie, J. (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Amer. Statist. Assoc.*, 115(529):393–402.

Pitman, E. J. G. (1949). *Lecture Notes on Nonparametric Statistical Inference*. Manuscript.

Qiu, J. (2025). Asymptotics of higher criticism via Gaussian approximation. Available at https://arxiv.org/abs/2511.06789.

Qiu, J., Chen, S. X., and Shao, Q.-M. (2025). Self-normalized Cramér type moderate deviation theorem for Gaussian approximation. *Ann. Statist.*, 53(3):1319–1346.

Qiu, Y., Chen, S. X., and Nettleton, D. (2018). Detecting rare and faint signals via thresholding maximum likelihood estimators. *Ann. Statist.*, 46(2):895–923.

Shao, Q.-M. and Zhang, Z.-S. (2022). Berry-Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms. *Bernoulli*, 28(3):1548–1576.

Xia, Y., Cai, T. T., and Li, H. (2018). Joint testing and false discovery rate control in high-dimensional multivariate regression. *Biometrika*, 105(2):249–269.

Xu, G., Lin, L., Wei, P., and Pan, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika*, 103(3):609–624.

Yu, X., Li, D., and Xue, L. (2024). Fisher's combined probability test for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, 119(545):511–524.

Yu, X., Li, D., Xue, L., and Li, R. (2023). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. *J. Amer. Statist. Assoc.*, 118(544):2548–2561.

Zhong, P.-S. and Chen, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.*, 106(493):260–274.

Zhong, P.-S., Chen, S. X., and Xu, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist.*, 41(6):2820–2851.