Statistical Signal Detection in High Dimension *

Song Xi Chen †

Abstract. This paper offers an extended review on the development of high dimensional statistical signal detection in the last three decades. It mainly focuses on the signals in the means of the underlying source populations with independent data, while extensions to covariance signals or temporal dependent data are also entertained. We are specifically interested in methods which are suitable for general data distributions beyond Gaussian. Properties of the more easily formulated L_2 and L_∞ signal detection procedures are evaluated in terms of their control on the type I error probability and the power performance, and their limitation are also stated. To detect weak and sparse signals, several multi-thresholding tests that include the signature higher criticism test are discussed and compared in terms of their detection boundary and relative power performance. The performance of the L_2, L_∞ and the multi-thresholding tests are evaluated with respect to different range of signal sparsity level under the special Gaussian distribution. A power enhancement formulation of the aforementioned tests can attain the best performance over the entire range of signal sparsity under Gaussianity.

1 Introduction. Detection of differences (signals) between two or more statistical populations is an enduring problem in mathematical statistics, dating back to William Gosset's seminal small sample t-test [27] for the one dimensional Gaussian populations. Statistical signal detection is usually formulated in the form of statistical hypothesis testing. Hotelling [31] pioneered the multivariate signal detection for multivariate Gaussian populations. Neyman and Pearson (1933) [47] introduced the "most efficient tests" framework based on the likelihood ratio of two competing models. Comprehensive theory on hypothesis testing for fixed dimensional setting are well documented in Lehmman and Romano (2005)[43]. The asymptotic minimax theory for high dimensional Gaussian data has been well summarized in Ingster and Suslina (2003)[35].

As the modern data collection and storage techniques improve over the recent decades, data are increasingly high dimensional. High dimensional data are commonly collected in biological experiments, environmental studies, signal processing, marketing research and financial risk management, just to name a few. The phenomenon of high dimensionality was well reflected in genomic data collected in biological studies where the gene expression levels for tens of thousands of genes or the single nucleotide polymorphism (SNP) are measured for each study subject. It is also the case in studies of financial risk management and marketing research, where the number of the assets in a portfolio can range from hundreds to thousands. The number of items a consumer can purchase from a market can be as large as the number of items for sale in the market. While the data dimensions are very high, the number of observations or the sample size is relatively small due to limitations to replicate study objects. This results in the so called "large-p, small n" paradigm in modern statistical data. Here p, representing data dimension, is much larger than the sample size n so that $p/n \to \infty$.

Signal detection was a major focus of the conventional multivariate analysis as comprehensively summarized in Anderson (2003)[1] and Muirhead (1982)[45], which was mainly on two approaches. One was to assume the distribution of the random data vectors comes from a specific parametric distribution and derived the probability distributions of certain test statistics, which was used for signal detection in the multivariate data samples in the form of rejecting a no-signal null hypothesis. The Hotelling test based on the T^2-statistics for detecting multivariate differences between two population means is the typical representative. Another is the asymptotic approach by allowing the sample size n to be large when the finite sample distribution of an inferential statistic is not easily obtained. A common aspect in these two approaches is to treat the dimension of the multivariate observations as fixed. Hence, it is largely operated on the conventional "fixed-p, large n" paradigm.

The high dimensional data settings bring fresh mathematical statistical challenges and opportunities, as well documented by Donoho (2000) [20] in his centennial AMS meeting presentation. It had been found that some

*The full version of the paper can be accessed at https://arxiv.org/abs/0000.00000

†Department of Statistics and Data Science, Tsinghua University (sxchen@tsinghua.edu.cn).

of the classical multivariate methods either do not work well under high dimensionality or are not applicable at all. This had been demonstrated in Bai and Saradanasa (1996)[5] for Hotelling's T^2 test, in Ledoit and Wolf (2002, 2004)[41] and Chen, Zhang and Zhong (2010)[14] for the sphericity and identity tests on a high dimensional covariance matrix when the signals were in the form of covariances, and in Zhong and Chen (2010)[64] on the F-test for the linear regression coefficients. The last two decades had witnessed a steady stream of mathematical statistical works which tried to detect both dense and sparse signals by overcoming the high dimensional challenges.

This paper is designed to review the field of the high dimensional statistical signal detection in the last two decades. The review will mainly focus on the techniques and results for mean signals among independent and identically distributed (i.i.d.) high dimensional data, while developments in detecting of covariance signals for i.i.d data, and for the mean signals with temporal dependent stationary data will be covered as well.

2 L_2 Tests for Fixed and High Dimensional Means We outline a set of test procedures whose test statistics are based on cumulated empirical signals over the dimensions in the fashion of L_2.

2.1 Hotelling's Test for Fixed Dimensional Means We first review the Hotelling test [31] for fixed dimensional multivariate signal detection due to its key position in the journey of statistical signal detection since it inherited William Gosset's Student t-test [27] and was well connected to the high dimensional setting. Let $N_p(\mu, \Sigma)$ denote p-dimensional Gaussian distribution with mean vector μ and covariance matrix Σ.

Let $X_{11}, \ldots, X_{1n} \overset{i.i.d}{\sim} N_p(\mu_1, \Sigma)$ and $X_{21}, \ldots, X_{2n} \overset{i.i.d}{\sim} N_p(\mu_2, \Sigma)$ be two independent Gaussian samples with the same covariance matrix Σ of dimension p. The interest is in testing

$$H_0 : \mu_1 = \mu_2 \quad \text{v.s.} \quad H_1 : \mu_1 \neq \mu_2, \tag{2.1}$$

such that $\tau =: \mu_1 - \mu_2 \neq 0$ may be viewed as signals.

Let $m = n_1 n_2/(n_1 + n_2)$ and $q = 1/m$. Define $\bar{X}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ for $i = 1$ and 2, and the pooled covariance estimator

$$S_n = \frac{1}{n} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T. \tag{2.2}$$

Let

$$T^2_{n_1,n_2} = q(\bar{X}_1 - \bar{X}_2)^T S_n^{-1} (\bar{X}_1 - \bar{X}_2). \tag{2.3}$$

Hotelling (1931) [31] proved that $(np)^{-1}(n - p + 1)T^2_{n_1,n_2} \sim F_{p,n-p+1}(m\|\delta\|^2)$, where $F_{p,n-p+1}(\lambda)$ denotes the non-central Snedecor's F-distribution with the indicated degrees of freedom and the no-central component λ, $n = n_1 + n_2 - 2$ and $\delta = \Sigma^{-1/2}\tau$. This leads to the Hotelling's test that rejects the non-signal hypothesis H_0 if

$$\frac{n - p + 1}{np} T^2_{n_1,n_2} > F_{p,n-p+1,\alpha}, \tag{2.4}$$

where $F_{p,n-p+1,\alpha}$ is the upper α-quantile of the centered F distribution $F_{p,n-p+1} = F_{p,n-p+1}(0)$.

The power of a binary hypothesis test is the probability that the test rejects the null hypothesis (H_0) when a specific alternative hypothesis (H_1) is true. The minimum requirement for a test being consistent is that the power should approach to one as $n \to \infty$ under H_1. From (2.4), the power of Hotelling's test is given by

$$\mathcal{B}_{\mathrm{H}}(\delta) = P(F_{p,n-p+1}(m\|\delta\|^2) > F_{p,n-p+1,\,\alpha}). \tag{2.5}$$

When p is fixed, it can be shown that Hotelling's test is uniformly most powerful under normal distribution.

2.2 Asymptotic Power with Growing Dimensions For the high dimensional setting, p is allowed to grow as $n \to \infty$. Let z_α be the upper α-quantile of the standard univariate Gaussian distribution $N(0, 1)$ and Φ be the cumulative distribution function of $N(0, 1)$. Bai and Saranadasa (1996) [5] provided the following asymptotic power property of Hotelling's test under growing dimensions.

THEOREM 2.1. *For two Gaussian samples as outlined two lines above (2.1), if $p/n \to y \in (0, 1)$, $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$ and if $\|\delta\|^2 = o(1)$ as $n \to \infty$, then*

$$\mathcal{B}_{\mathrm{H}}(\delta) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-y)}{2y}} \kappa(1-\kappa)\|\delta\|^2\right) \to 0. \tag{2.6}$$

We can deduce the following comments from Theorem 2.1. If $\sqrt{n}\|\delta\|^2 \to \infty$, then $\mathcal{B}_{\mathrm{H}}(\delta) \to 1$. Hence the test is consistent in this case, which includes the fixed alternative case. If $\sqrt{n}\|\delta\|^2 \to 0$, namely $\|\delta\| = o(n^{-1/4})$, $\mathcal{B}_{\mathrm{H}}(\delta) \to \alpha$. Then the test is incapable of detecting the difference between H_0 and H_1 if H_1 is at a distance of a smaller order of $n^{-1/4}$ from H_0. If $\sqrt{n}\|\delta\|^2 \to a \in (0,\infty)$, the power is between α and 1 and is non-trivial. However, $(1-y)/y$ is a decreasing function of y and close to 0 if $y = \lim p/n$ is close to 1. Hence, an increase of the power as a increases is greatly hampered by $(1-y)/y$. This was first discovered in Bai and Saranadasa (1996) [5], and re-confirmed in Zhong and Chen (2011) [64] for the F-test for a growing dimensional linear regression coefficient vector.

2.3 Bai–Saranadasa Test for Moderately High Dimensional Means To rescue the power loss issue of Hotelling's test under the high dimensionality, Bai and Saranadasa (1996) [5] proposed a test statistic based on the L_2 norm of $\bar{X}_1 - \bar{X}_2$ given by

$$M_n = \|\bar{X}_1 - \bar{X}_2\|^2 - q\,\mathrm{tr}(S_n) \tag{2.7}$$

without the weighting by S_n^{-1} in the Hotelling formulation as S_n is no longer a consistent estimator of Σ when $p/n \to c \in (0,\infty)$, not mentioning for the case of $p/n \to \infty$. The second term $q\mathrm{tr}(S_n)$ is to center the statistic so that $E(M_n) = \|\tau\|^2$. Instead of the Gaussian assumption, they assumed that:

(BS1) For $i = 1, 2$ and $j = 1, \ldots, n_i$, $X_{ij} = \Gamma Z_{ij} + \mu_i$, where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma\Gamma^T = \Sigma$, and $\{Z_{ij}\}_{j=1}^{n_i}$ are m-variate i. i. d. random vectors with independent components satisfying $E(Z_{ij}) = 0$ and $\mathrm{var}(Z_{ij}) = I_m$. Furthermore, $E(Z_{ijk}^4) = 3 + \Delta < \infty$ and when $\sum_{k=1}^m v_k = 4$,

$$E\left(\prod_{k=1}^m Z_{ijk}^{v_k}\right) = \begin{cases} 0, & \text{if at least one } \ v_k = 1; \\ 1, & \text{if there are two } \ v_k\text{'s being 2.} \end{cases}$$

(BS2) $p/n \to y > 0$ and $n_1/(n_1+n_2) \to \kappa \in (0,1)$.

(BS3) $\tau^T\Sigma\tau = o(q\mathrm{tr}(\Sigma^2))$.

(BS4) $\lambda_{\max}(\Sigma) = o(\mathrm{tr}^{1/2}(\Sigma^2))$.

Let $\hat{\sigma}_M^2 = 2q^2(\mathrm{tr}(S_n^2) - n^{-1}\mathrm{tr}^2(S_n))$, which is shown to be a ratio-consistent estimator of $\mathrm{var}(M_n)$. Then, by using the martingale central limit theorem [28], it can be proven that under Conditions (BS1)–(BS4),

$$\frac{M_n - \mathrm{E}(M_n)}{\hat{\sigma}_M} \xrightarrow{d} N(0,1). \tag{2.8}$$

As $E(M_n) = 0$ under H_0, the Bai–Saranadasa test rejects H_0 if $M_n > \hat{\sigma}_M z_\alpha$. Bai and Saranadasa (1996) [5] provided its power function as follows.

THEOREM 2.2. *Under Conditions (BS1)–(BS4),*

$$\mathcal{B}_{\mathrm{BS}}(\tau) - \Phi\left(-z_\alpha + \frac{n\kappa(1-\kappa)\|\tau\|^2}{\sqrt{2\,\mathrm{tr}\,\Sigma^2}}\right) \to 0, \tag{2.9}$$

where $\beta_{\mathrm{BS}}(\tau) = P(M_n > \hat{\sigma}_M z_\alpha | H_1)$ is the power function of Bai–Saranadasa test.

The form of the power function in Theorem 2.2 shows that the test removes the dimensionality factor $(1-y)/y$ presented in the power of Hotelling's test. This indicates that the test can handle the high dimensional case better than the Hotelling's test. In the case of $p < n$, we need to compare $(n(1-y)/y)^{1/2}\|\delta\|^2$ with $(n/\mathrm{tr}^{1/2}\Sigma^2)\|\tau\|^2$ in order to see who is more powerful. When $\Sigma = I_p$, we have $(n/\mathrm{tr}^{1/2}\Sigma^2)\|\tau\|^2 \sim (n/y)^{1/2}\|\delta\|^2 > (n(1-y)/y)^{1/2}\|\delta\|^2$, which shows that the Bai–Saranadasa test is asymptotically more powerful than the Hotelling's test in this case.

2.4 Chen–Qin Test for Ultra High Dimensional Means Although Bai–Saranada test can be more powerful than the classical Hotelling's test, it is not suitable for the "large p, small n" paradigm of $p/n \to \infty$ as restricted by Condition (BS2). Chen and Qin (2010) [11] pointed out that the restriction comes from the need to control the terms $\sum_{j=1}^{n_i} X_{ij}^T X_{ij}$ for $i = 1$ and 2 in $\|\bar{X}_1 - \bar{X}_2\|^2$, which becomes burden in the testing. Instead, they proposed a U-statistic by removing the redundant terms given by

$$Q_n = \frac{1}{n_1(n_1-1)} \sum_{i \neq j}^{n_1} X_{1i}^T X_{1j} + \frac{1}{n_2(n_2-1)} \sum_{i \neq j}^{n_2} X_{2i}^T X_{2j} - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{1i}^T X_{2j}. \tag{2.10}$$

The term $q\,\mathrm{tr}(S_n)$ in the Bai–Saranadasa statistic (2.7) disappears in the U-statistic formulation. It can be shown that $E(Q_n) = \|\tau\|^2$, reflecting the L_2 signal strength level. Hence, Q_n is basically all we need for testing.

(CQ1) For $i = 1, 2$ and $j = 1, \ldots, n_i$, $X_{ij} = \Gamma_i Z_{ij} + \mu_i$, where each Γ_i is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i \Gamma_i^T = \Sigma_i$, and $\{Z_{ij}\}_{j=1}^{n_i}$ are m-variate i. i. d. random vectors satisfying $E(Z_{ij}) = 0$ and $\mathrm{var}(Z_{ij}) = I_m$. Furthermore, $E(Z_{ijk}^4) = 3 + \Delta < \infty$, and for a positive integer s such that $\sum_{l=1}^{s} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$,

$$E\left(Z_{ijl_1}^{\alpha_1} Z_{ijl_2}^{\alpha_2} \cdots Z_{ijl_s}^{\alpha_s}\right) = E(Z_{ijl_1}^{\alpha_1}) E(Z_{ijl_2}^{\alpha_2}) \cdots E(Z_{ijl_s}^{\alpha_s}).$$

(CQ2) $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$.

(CQ3) $\tau^T \Sigma_i \tau = o(n^{-1} \mathrm{tr}((\Sigma_1 + \Sigma_2)^2))$.

(CQ4) $\mathrm{tr}(\Sigma_i \Sigma_j \Sigma_l \Sigma_h) = o(\mathrm{tr}^2((\Sigma_1 + \Sigma_2)^2))$ for $i, j, l, h = 1$ or 2 as $p \to \infty$.

Note that the covariance matrices of the two samples can differ in Condition (CQ1). Let

$$\hat{\sigma}_Q^2 = \frac{2}{n_1(n_1-1)} \widehat{\mathrm{tr}(\Sigma_1^2)} + \frac{2}{n_2(n_2-1)} \widehat{\mathrm{tr}(\Sigma_2^2)} + \frac{4}{n_1 n_2} \widehat{\mathrm{tr}(\Sigma_1 \Sigma_2)} \tag{2.11}$$

be a ratio consistent estimator of $\mathrm{var}(Q_n)$. Here,

$$\widehat{\mathrm{tr}(\Sigma_i^2)} = \frac{1}{n_i(n_i-1)} \mathrm{tr}\left(\sum_{j \neq k}^{n_i} (X_{ij} - \bar{X}_{i(j,k)}) X_{ij}^T (X_{ik} - \bar{X}_{i(j,k)}) X_{ik}^T\right), \tag{2.12}$$

where $\bar{X}_{i(j,k)}$ is the i-th sample mean without X_{ij} and X_{ik}; and

$$\widehat{\mathrm{tr}(\Sigma_1 \Sigma_2)} = \frac{1}{n_1 n_2} \mathrm{tr}\left(\sum_{l=1}^{n_1} \sum_{k=1}^{n_2} (X_{1l} - \bar{X}_{1(l)}) X_{1l}^T (X_{2k} - \bar{X}_{2(k)}) X_{2k}^T\right), \tag{2.13}$$

where $\bar{X}_{i(l)}$ is the i-th sample mean without X_{il}. By using the martingale central limit theorem, it can be proven that under Conditions (CQ1)–(CQ4),

$$\frac{Q_n - \mathrm{E}(Q_n)}{\hat{\sigma}_Q} \xrightarrow{d} N(0,1) \quad \text{as} \quad \min\{n_1, n_2\} \to \infty. \tag{2.14}$$

Then, the Chen–Qin test rejects H_0 if $Q_n > \hat{\sigma}_Q z_\alpha$ with power function given below.

THEOREM 2.3. *Under Conditions (CQ1)–(CQ4), as $p, n \to \infty$,*

$$\mathcal{B}_{\mathrm{CQ}}(\tau) - \Phi\left(-z_\alpha + \frac{n\kappa(1-\kappa)\|\tau\|^2}{\sqrt{2\,\mathrm{tr}\,\tilde{\Sigma}^2(\kappa)}}\right) \to 0, \tag{2.15}$$

where $\mathcal{B}_{\mathrm{CQ}}(\tau) = P(Q_n > \hat{\sigma}_Q z_\alpha | H_1)$ is the power function of Chen–Qin test, and $\tilde{\Sigma}(\kappa) = (1-\kappa)\Sigma_1 + \kappa\Sigma_2$.

The power of Bai–Saranadasa test has the same form if $\Sigma_1 = \Sigma_2$ and if p and n are of the same order. Chen–Qin test can still be valid when $p/n \to \infty$, which is suitable for the "large p, small n" situation.

2.5 Extensions for Time Series Data Chen–Qin test can be extended for temporally dependent data to meet the need of real world data analysis. Suppose now that $\{X_{it}\}_{t=1}^{n_i}$ is a p dimensional stationary time series with mean μ_i and covariance matrix $\Sigma_{i,0}$ for $i = 1, 2$. Let $\Sigma_{i,k} = \text{cov}(X_{i,t+k}, X_{it})$ be the cross covariance matrix for $k = \pm 1, \ldots, \pm(n-1)$ and $\Sigma_{i,\infty} = \sum_{k=-\infty}^{\infty} \Sigma_{i,k}$ be the long-run covariance matrix of X_{it}, provided that the series converges. Zhang, Chen and Qiu (2025) [60] proposed the following band-excluded U-statistic

$$T_n(b) = \frac{1}{n_1(b)} \sum_{|s-t|\geq b}^{n_1} X_{1s}^T X_{1t} + \frac{1}{n_2(b)} \sum_{|s-t|\geq b}^{n_2} X_{2s}^T X_{2t} - \frac{2}{n_1 n_2} \sum_{s=1}^{n_1} \sum_{t=1}^{n_2} X_{1s}^T X_{2t}, \tag{2.16}$$

where $n_i(b) = (n_i - b)(n_i - b + 1)$ and b is a positive tuning parameter that defines a temporal exclusion band of width b to exclude products $X_{is}^T X_{it}$ with temporal distance less than the bandwidth. Note that $T_n(0)$ is the Bai–Saranadasa test statistic and $T_n(1)$ is the Chen–Qin test statistic. The removed pairs of X_{is} and X_{it} are likely to be more strongly correlated. Hence, the band exclusion with a large b effectively mitigates the bias induced by the temporal dependence in $T_n(0)$ or $T_n(1)$.

PROPOSITION 2.4. *Under the regularity conditions stated in Zhang, Chen and Qiu (2025) [60], if* $\log p = o(n)$, $b = o(n)$, *and* $b \geq c_1(\log n + \log p)$ *for a positive constant* c_1, *then as* $p, n \to \infty$,

$$E(T_n(b)) = \|\tau\|^2 + \sum_{i=1}^{2} \frac{2}{n_i(b)} \sum_{k=b}^{n_i-1} (n_i - k)\, \text{tr}(\Sigma_{i,k}) \;\; \textit{and} \;\; \text{var}(T_n(b)) = \left(\frac{2}{n^2} \text{tr}(M_\infty^2) + \frac{4}{n} \tau^T M_\infty \tau \right) (1 + o(1)), \tag{2.17}$$

where $M_\infty = \sum_{k=-\infty}^{\infty} M_i$ *and* $M_k = \kappa^{-1}\Sigma_{1,k} + (1-\kappa)^{-1}\Sigma_{2,k}$ *with* κ *defined in Condition (CQ2).*

Proposition 2.4 shows that the bias term of $T_n(b)$ is asymptotically equal to $2n^{-1} \sum_{k=b}^{\infty} \text{tr}(M_k)$, which diminishes to zero at a polynomial rate of p and n if $b \geq c_1(\log n + \log p)$ for a positive constant c_1 under an exponential β-mixing condition. Let $\hat{\sigma}_T^2$ be a ratio-consistent estimator of $T_n(b)$ as described in Zhang, Chen and Qiu (2025) [60]. Then the band-excluded U-test rejects H_0 if $T_n(b) > \hat{\sigma}_T z_\alpha$ with power function given below.

THEOREM 2.5. *Under the regularity conditions stated in Zhang, Chen and Qiu (2025) [60], as* $p, n \to \infty$,

$$\mathcal{B}_{\text{ZCQ}}(\tau) - \Phi\left(-z_\alpha + \frac{n\|\tau\|^2}{\sqrt{2\,\text{tr}(M_\infty^2)}} \right) \to 0, \tag{2.18}$$

where $\mathcal{B}_{\text{ZCQ}}(\tau) = P(T_n(b) > \hat{\sigma}_T z_\alpha | H_1)$ *is the power function of the band-excluded* U*-test.*

The power of the band-excluded U-test shares the same formulation with Chen–Qin test, except that the pooled covariance matrix $\tilde{\Sigma}(\kappa)/\kappa(1-\kappa)$ is replaced by its long-run version M_∞.

2.5.1 Self-normalization Formulation for One-sample Problem As Σ_1 and Σ_2 are typically unknown in practice, Chen–Qin type tests requires ratio-consistent estimators to estimate the variance of the test statistic. For one-sample hypotheses, Wang and Shao (2020) [55] proposed a self-normalization formulation that does not require such an estimation. Let $\{X_t\}_{t=1}^{N}$ be a p dimensional stationary time series with mean μ and long-run covariance matrix Σ_∞. Let b be a trimming parameter with $1/b + b/N = o(1)$ and $n = N - b$. For the partial sum process

$$S_n(r) = \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} X_{t+b}^T X_s, \quad r \in [0, 1], \tag{2.19}$$

Wang and Shao (2020) [55] proposed the self-normalized test statistic

$$S_n^* = \frac{S_n^2(1)}{W_n^2} \xrightarrow{d} \mathcal{K} = \frac{B^2(1)}{\int_0^1 (B(u^2) - u^2 B(1))^2 du}, \quad \text{where} \quad W_n^2 = \frac{1}{n} \sum_{k=1}^{n} \left(S_n(k/n) - \frac{k(k+1)}{n(n+1)} S_n(1) \right)^2, \tag{2.20}$$

$B(r), r \in [0, 1]$ denotes the standard Brownian motion, and the convergence is valid under the regularity conditions stated in the mentioned paper. Then, the self-normalized test rejects H_0 if $S_n^* > \mathcal{K}_\alpha$, where $\mathcal{K}_\alpha$ is the upper α quantile of $\mathcal{K}$ that can be obtained by the Monte Carlo method.

THEOREM 2.6. *Under the regularity conditions stated in Wang and Shao (2020) [55], if $n\|\mu\|^2/(2\operatorname{tr}(\Sigma_\infty^2))^{1/2} \to c \in (0,\infty)$, then as $\min\{p,N\} \to \infty$,*

$$S_n^* \xrightarrow{d} \frac{(B(1)+c)^2}{\int_0^1 (B(u^2)-u^2B(1))^2 du} \quad \textit{and} \quad P(S_n^* > \mathcal{K}_\alpha) \to \mathcal{B}_{\mathrm{WS}}(c) \in (\alpha, 1). \tag{2.21}$$

Theorem 2.6 shows that the signal strength need to satisfy $n\|\mu\|^2/(2\operatorname{tr}(\Sigma_\infty^2))^{1/2} \to c$ to ensure a nontrivial asymptotic power $\mathcal{B}_{\mathrm{WS}}(c)$ for the self-normalized test. By Theorem 2.5, the nontrivial signal strength rate for the band-excluded test is essentially the same with $m\|\tau\|^2/(2\operatorname{tr}(M_\infty^2))^{1/2} \to c$, and the corresponding power limit is $\Phi(c - z_\alpha)$. Zhang, Chen and Qiu (2025) [60] numerically showed that $\Phi(c-z_\alpha) \geq \mathcal{B}_{\mathrm{WS}}(c)$ for the one-sample testing problem, indicating a power loss for the self-normalization formulation.

3 L_∞ Tests for Sparse and Strong Signals The power function of the L_2 tests listed above are all functions of $\|\tau\|^2 = \sum_{i=1}^p \tau_i^2$, where $\tau_i = \mu_{1i} - \mu_{2i}$ and those nonzero τ_i's are called signals. By this formulation, it can be seen that the L_2 test is good at accumulating signals spread out in all dimensions, which makes it powerful for detecting dense and weak signals. However, when signals are too sparse so that many dimensions are no signal bearing, the L_2 test will collect too many noises and have no power in this case. To appreciate this, we consider the case of $\Sigma_1 = \Sigma_2 = I_p$. Assume that there are $\sum_{i=1}^p 1(\tau_i \neq 0) = p^{1-\beta}$ signals for a signal sparsity parameter $\beta \in (0,1)$ and each signal satisfies $m\tau_i^2 = h^2$ for a signal strength parameter $h^2 > 0$. By Theorem 2.3, to ensure the Chen–Qin test has a nontrivial power, we must have $h^2 \asymp p^{\beta-1/2}$ as $p \to \infty$. Hence, the nontrivial signal strength can diminish to zero in a polynomial rate for dense signals with $\beta < 1/2$ but need to diverge rapidly for sparse signals with $\beta > 1/2$. As will be pointed out below, the optimal signal strength level for sparse signal regime is of order $h^2 \asymp \log p$, which implies that the L_2 test is unsuitable for this case.

3.1 L_∞ Test Based on Data Transformation To rescue the failure of the L_2 test for sparse signals, we need to find a test that is not sensitive to the noises and is able to focus on the most extreme candidate of signals. This is where the L_∞ test comes in. Let $X_{11},\ldots,X_{1n} \overset{i.i.d}{\sim} F_1$ and $X_{21},\ldots,X_{2n} \overset{i.i.d}{\sim} F_2$ be two independent samples with means μ_1 and μ_2 respectively and the same covariance matrix Σ of dimension p. Let A be a $p \times p$ invertible matrix. Cai, Liu and Xia (2014) [8] proposed a family of test statistics of the form

$$M(A) = \max_{1\leq i\leq p} \frac{m(\delta_i^A)^2}{b_{ii}}, \tag{3.1}$$

where $m = n_1n_2/(n_1+n_2)$, $\delta^A = (\delta_1^A,\ldots,\delta_p^A)^T = A(\bar{X}_1 - \bar{X}_2)$ and $B = (b_{ij})$ is the covariance matrix of X_{11}. Let $\Omega = \Sigma^{-1}$. They recommended the choices of $A = \Omega$, $\Omega^{1/2}$ and I_p and pointed out that $A = \Omega$ can be more powerful than $A = \Omega^{1/2}$ (resp. $A = I_p$) for sparse signals with $\beta > 6/7$ (resp. $\beta > 1/2$). When $\Sigma = (\sigma_{ij})$ is unknown, a consistent estimator $\hat{\Omega} = (\hat{\omega}_{ij})$ of the high dimensional precision matrix $\Omega = (\omega_{ij})$ is required.

(CLX1) (a) $C_0^{-1} \leq \lambda_{\min} \leq \lambda_{\max}(\Sigma) \leq C_0$ for a constant $C_0 > 0$. (b) $\max_{1\leq i<j\leq p} |\omega_{ij}/\omega_{ii}^{1/2}\omega_{jj}^{1/2}| \leq c_0 < 1$ for a constant $c_0 \in (0,1)$. (c) $\max_{1\leq j\leq p} \sum_{i=1}^p |\hat{\omega}_{ij} - \omega_{ij}| = o_P(1/\log p)$ and $\max_{1\leq i\leq p} |\hat{\omega}_{ii} - \omega_{ii}| = o_P(1/\log p)$.

(CLX2) Suppose that $\log p = o(n^{1/4})$. There exist some constants $\eta > 0$ and $K > 0$ such that $E\exp(\eta V_{ij}^2/\omega_{jj}) \leq K$ for $i = 1,2$ and $1 \leq j \leq p$, where $X_{i1} - \mu_i =_d V_i$ for $i = 1,2$.

(CLX3) Suppose that $p \leq c_1 n^{\gamma_0}$ for constants $\gamma_0, c_1 > 0$. There exist some constants $\eta > 0$ and $K > 0$ such that $E|V_{ij}/\omega_{jj}^{1/2}|^{2\gamma_0+2+\varepsilon} \leq K$ for $i = 1,2$ and $1 \leq j \leq p$.

(CLX4) (a) There are $\sum_{i=1}^p 1(\tau_i \neq 0) = p^{1-\beta}$ signals with $\beta > 3/4$ and the signals are randomly uniformly drawn from $\{1,\ldots,p\}$. (b) $\max_{1\leq i\leq p} m\tau_i^2/\sigma_{ii} \geq 2r\log p$ with $r \geq 1/\min_{1\leq i\leq p}(\sigma_{ii}\omega_{ii}) + \varepsilon$ for a constant $\varepsilon > 0$.

(CLX5) (a) There are $\sum_{i=1}^p 1(\tau_i \neq 0) = p^{1-\beta}$ signals with $\beta > 1/2$ and $\max_{1\leq j\leq p} \sum_{i=1}^p 1(\omega_{ij} \neq 0) = O(p^{\gamma(2\beta-1)})$ for a constant $\gamma \in (0,1)$. (b) There are $\sum_{i=1}^p 1(\tau_i \neq 0) = p^{1-\beta}$ signals with $\beta > 1/4$.

(CLX6) $\max_{1\leq j\leq p} \sum_{i=1}^p |\omega_{ij}| \leq M$ for a constant $M > 0$.

Theorem 3.1. *Under Conditions (CLX1) and (CLX2) (or (CLX3)), for any $x \in \mathbb{R}$, as $p, n \to \infty$,*

$$P(M(\hat{\Omega}) - 2\log p + \log\log p \leq x | H_0) \to \exp\left(-\frac{1}{\sqrt{\pi}} e^{-x/2}\right). \tag{3.2}$$

Suppose in addtion that Condition (CLX4) is valid. Then as $p, n \to \infty$,

$$P(M(\hat{\Omega}) > 2\log p - \log\log p + q_\alpha | H_1) \to 1, \tag{3.3}$$

where $q_\alpha = -\log \pi - 2\log\log(1/(1-\alpha))$ is the upper α quantile of Gumbel distribution.

The Cai–Liu–Xia test rejects H_0 if $M(\hat{\Omega}) > 2\log p - \log\log p + q_\alpha$. Condition (CLX4) in Theorem 3.1 also shows that the test is consistent for highly sparse signals with signal sparse parameter $\beta > 3/4$ and strong signals with signal strength level satisfying $\max_{1 \leq i \leq p} m\tau_i^2/\sigma_{ii} \geq 2r \log p$. Indeed, Cai, Liu and Xia (2014) [8] proved that the order for the signal strength level is minimax rate optimal.

Theorem 3.2. *Suppose that $F_1 = N(\mu_1, \Sigma)$ and $F_2 = N(\mu_2, \Sigma)$. Let $\alpha, \nu > 0$ and $\alpha + \nu < 1$. Under Conditions (CLX5)(a) (or (b)) and (CLX6), there is a constant $c > 0$ such that for all sufficiently large p and n,*

$$\sup_{\psi \in \Psi_\alpha} \inf_{\tau: \sum_{i=1}^p 1(\tau_i \neq 0) = p^{1-\beta}, \max_{1 \leq i \leq p} m\tau_i^2 \geq c \log p} P(\psi = 1) \leq 1 - \nu, \tag{3.4}$$

where $\Psi_\alpha = \{\psi : P(\psi = 1 | H_0) \leq \alpha\}$ is the collection of all level-α tests.

Theorem 3.2 shows that there is no consistent level-α test for c sufficiently small, which implies that the rate of $\max_{1 \leq i \leq p} m\tau_i^2/\sigma_{ii} \geq 2r \log p$ in (CLX4) cannot be improved.

From Conditions (CLX2) or (CLX3), it can be seen that Cai–Liu–Xia test requires either an exponential type moment condition or a polynomial rate restriction of p and n. For one-sample mean hypotheses, Qiu, Chen and Shao (2025) [49] proposed a variation of (3.1) based on a self-normalized Gaussian approximation result that can reduce those conditions to an 8/3-th moment condition and an ultra high dimensional setting of $\log p = o(n^{1/4})$ simultaneously. It is interesting to see if their results can be extended for two-sample or more general problem.

4 Thresholding Tests and Detection Boundary Accompanied with the high dimensional data and problems is often the sparse signal issue, which means that a large portion of the dimensions bear no signals despite the high dimensionality. There is a vast literature on estimating of sparse signals by the so-called regularization method such as the Lasso (Tibshirani, 1996[52]), smoothly clipped absolute deviation (SCAD; Fan & Li, 2001[26]), minimax concave penalty (MCP; Zhang, 2010[58]), adaptive Lasso (Zou, 2006[68]).

In contrast to estimation, statistical hypothesis testing and signal detection becomes considerably more challenging under sparse and faint signal settings, mainly due to the signals are present only in a very small subset of dimensions within a high-dimensional space, and the magnitude of each individual signal can be weak. This creates a fundamental difficulty in that when the signal-to-noise ratio is extremely low, reliable detection or statistical inference may become impossible. Signal detection problems of this nature have attracted substantial attention due to their relevance in diverse applications, including multi-channel detection, communication systems, and high-dimensional data analysis. The study of the fundamental limits of signal detection can be traced back to early work by [18], who analyzed the smallest detectable signal in Gaussian noise models. Later, [18] and [32, 33, 34] established the minimax framework for signal detection in Gaussian white noise, characterizing the critical signal strength required for reliable testing as the signal amplitude tends to zero. These pioneering works laid the foundation for the modern theory of sparse signal detection, where the focus is on understanding the lower bound of the detection boundary that separates the regimes of possible and impossible inference.

An important question in practice is how to design procedures that attain the lower bound of the detection boundary. While the classical likelihood ratio test (LRT) [47], studied in the foundational works mentioned above, provides a benchmark for optimality, it typically requires knowledge of unknown parameters such as the sparsity level or the signal intensity. This requirement makes the LRT impractical in many real-world scenarios, where such quantities are rarely available in advance. In practice, therefore, the goal is to construct procedures that are adaptive, capable of detecting signals automatically without prior specification of these unknown parameters.

While L_2-based procedures are asymptotically optimal for dense and moderately sparse signals, they lose power in the sparse regime. To make this point more concrete, we consider the following example given in [10] about the asymptotic power of the Chen-Qin test discussed in Section 2.4. Assume the number of signals $\tau_k \neq 0$ for $k \in S_\beta$ is at the order of $|S_\beta| \asymp p^{1-\beta}$ and define $\bar{\tau}^2 = \sum_{k \in S_\beta} \tau_k^2 / p^{1-\beta}$ as the average signal size where τ_k denotes the signal magnitude as defined in Section 3. According to the power function given in Theorem 2.3, the power of the Chen-Qin's test is largely determined by the signal to noise ratio given below [10]:

$$\mathrm{SNR}_{CQ} = \frac{p^{1-\beta} n \bar{\tau}^2}{\sqrt{2p + 2\sum_{i \neq j} \rho_{ij}^2 + 4n \sum_{k,l \in S_\beta} \tau_k \tau_l \rho_{kl}}},$$

where ρ_{kl} is the covariance of the k-th and l-th components of $\sqrt{n}(\bar{X}_1 - \bar{X}_2)$, where $n = n_1 n_2/(n_1 + n_2)$. If $\Sigma_1 = \Sigma_2 = I$ and $\bar{\tau} = o(n^{-1/2} p^{\beta/2 - 1/4})$, then the above $\mathrm{SNR}_{CQ} = o(1)$ for the sparse case with $\beta > 1/2$. This means that the L_2 type Chen-Qin's test may have little power beyond the type I error. A similar result is obtained in [3], who demonstrated that an ANOVA test achieves full asymptotic power when $\max_j \tau^2 p^{1/2-\beta} \to \infty$, but it may lose power when $\beta > 1/2$. Furthermore, the optimality of the ANOVA tests have been investigated by [3] under a regression setting. They established that, under certain conditions, all tests are asymptotically powerless if $\max_j \tau_j^2 p^{1/2-\beta} \to 0$. Comparing these upper and lower bounds demonstrates that ANOVA is essentially optimal in the dense and moderately sparse regimes corresponding to $\beta \in [0, 1/2]$. By the same reasoning, the L_2- based test such as Chen-Qin's test introduced in Section 2 is also optimal in this setting for detecting dense or moderately sparse signals with $\beta \leq 1/2$.

In contrast, L_∞-based tests, such as those relying on the maximum statistic discussed in Section 3 or multiple testing procedures that control the false discovery rate, have been shown by [21] to be optimal when the signal count is on the order of $p^{1-\beta}$ with $\beta > 3/4$. Similar conclusions were later obtained by [3] in a regression framework, demonstrating that tests based on maximum-type statistics are optimal for very sparse signals. However, in the intermediate sparse regime, $\beta \in (1/2, 3/4)$, neither L_2-based methods nor L_∞-based methods achieve optimality.

These studies highlights a critical gap: the optimal choice of test depends heavily on the underlying sparsity level of the signals. In practice, however, the degree of sparsity is rarely known a priori. One does not typically know whether the signals are dense, moderately sparse, or extremely sparse. Consequently, it is of both theoretical and practical importance to design adaptive procedures that can automatically adjust to the unknown sparsity regime and achieve optimal performance across the full spectrum of sparsity levels.

4.1 Higher Criticism Test A pioneering and influential breakthrough was achieved by Donoho and Jin (2004) [21], who introduced a powerful adaptive procedure known as higher criticism. Originally proposed by [54] as an exploratory tool for large-scale multiple testing, higher criticism was rigorously developed by Donoho and Jin to a formal statistical test for detecting the presence of sparse signals in high-dimensional Gaussian mixture models. They showed that higher criticism can achieve the optimal detection boundary across a wide range of sparsity regimes, without requiring explicit knowledge of the sparsity level or signal strength. More specifically, they consider the detection problem under the following setting:

$$H_0 : X_i \overset{iid}{\sim} N(0,1) \quad \text{for } 1 \leq i \leq p, H_1 : X_i \overset{iid}{\sim} (1-\varepsilon_n) N(0,1) + \varepsilon_n N(\mu_n, 1) \quad \text{for } 1 \leq i \leq p, \tag{4.1}$$

where X_i may be considered as a test statistic constructed for the i-th dimension ($i = 1, \cdots, p$). They specify the sparse senario by setting $\varepsilon_n = p^{-\beta}$ for $\beta \in (1/2, 1)$ and the weak signal setting by setting $\mu_n = \sqrt{2r\log(p)}$. Here r is the parameter for the signal strength and β is the parameter for sparsity level. The smaller in r and bigger values in β imply more difficult detection or statistical inference. By translating the test statistics X_i into the corresponding pvalues $P_i = P(N(0,1) > X_i)$ and ranking them in an increasing order $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(p)}$, the higher criticism statistics is defined as

$$HC_n^* = \max_{1 \leq i \leq p\alpha_0} \frac{\sqrt{p}(i/p - P_{(i)})}{\sqrt{P_{(i)}(1 - P_{(i)})}}, \tag{4.2}$$

where α_0 is a pre-specified value. Donoho and Jin (2004) has proved that if $r > \rho(\beta)$, the procedure based on rejecting H_0 by comparing HC_n^* with $\sqrt{2\log\log(n)}$ can asymptotically distinguish H_0 and H_1 with probability 1, where $\rho(\beta)$ is the the optimal detection boundary given by:

$$\rho(\beta) = \begin{cases} \beta - 1/2, & \text{if } 1/2 < \beta \leq 3/4 \\ (1 - \sqrt{1-\beta})^2, & \text{if } 3/4 < \beta < 1. \end{cases} \tag{4.3}$$

The optimality is in the sense that if $r \leq \rho(\beta)$, there is no detection procedure that would be able to asymptotically distinguish the null H_0 and the alternative H_1. This detection boundary was first established by Ingster (1997)[32] for detecting sparse non-zero signals for Gaussian random vectors.

Following the seminal contribution of Donoho and Jin (2004) [21], a large body of research has emerged examining the performance of the higher criticism (HC) test across a wide range of statistical settings. These include for non-Gaussian models, feature selection, classification, regression, and large-scale multiple hypothesis testing; see, for example, [29, 30, 15, 16, 17, 19, 37, 40]. A comprehensive review of HC and its developments is provided in [22].

4.2 Multi-level Thresholding Test The idea of thresholding has appeared in [23, 25] in adaptive estimation and statistical inference for nonparametric functions. In the context of detecting and testing high dimensional mean vectors, [65] developed an adaptive thresholding test for detecting sparse and faint nonzero means for sub-Gaussian distributed data with unknown column-wise dependence. Consider iid p-dimensional random vectors $X_1, \cdots, X_n$ generated from an additive model:

$$X_i = W_i + \mu, \quad \text{for} \quad i = 1, \cdots, n,$$

where $\mu = (\mu_1, \cdots, \mu_p)^T$ and $W_i = (W_{i1}, \cdots, W_{ip})^T$ be iid random vectors with zero mean and common covariance, $\{W_{ij}\}_{j=1}^p$ is assumed to be a sequence of weakly stationary dependent random variables. The problem of interest was testing

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \text{nonzero } \mu_j \text{ are sparse and faint.} \tag{4.4}$$

Assume that the alternatives of interest have $p^{1-\beta}$ non-zero μ_j's for $\beta \in (1/2, 1)$ and these non-zero $\mu_j = \sqrt{2r\log(p)/n}$ for $r \in (0, 1)$. If the variance σ_j^2 of the j-th component X_j is known, a natural test statistic for testing the mean of the j-th component is based on the marginal test statistics $\sqrt{n}\bar{X}_j/\sigma_j$. If the test statistics has an absolute value larger than certain threshold, it is likely that the j-th component is non-zero. One natural approach is to aggregate the evidence across all components using the L_2-norm of the marginal test statistics, leading to an L_2-based global test. However, [11] demonstrated that such L_2-norm based tests can lose power, and may even be asymptotically powerless, when the alternatives are sparse as in (4.4). The fundamental difficulty lies in the accumulation of noise from the vast majority of null coordinates: the contribution of these noise terms can overwhelm the relatively weak signals, thereby erasing any detectable separation.

To overcome this problem and enhance the signal-to-noise ratio, thresholding methods have been proposed. The central idea is to eliminate coordinates that are unlikely to contain signals, thereby reducing the impact of pure noise. This is typically achieved via hard thresholding, where only components with sufficiently large marginal statistics are retained for further aggregation. More concretely, a single-level threshold test can be constructed by first applying a threshold to the marginal statistics and then aggregating the surviving components to form a global test statistic. Such thresholding tests preserve the contribution of strong signals while substantially reducing noise accumulation, leading to improved power in the sparse signal regime. More specifically, a single level threshold test can be constructed as

$$T_{\gamma n}(s) = \sum_{j=1}^{p} |\sqrt{n}\bar{X}_j/\sigma_j|^{\gamma} I\Big\{|\bar{X}_j| \geq \sigma_j\sqrt{2s\log(p)/n}\Big\}, \tag{4.5}$$

for some $s \in (0, 1)$ where $I(\cdot)$ is an indicator function and $\gamma = 0, 1$ and 2. When $\gamma = 0$, $T_{\gamma n}(s)$ corresponds to the HC method in [21]. When $\gamma = 1$, $T_{\gamma n}(s)$ corresponds to the hard-thresholding method in [23, 25]. It is clear that the performance of the single threshold test depends on the choice of the threshold level s. However, the choice of such threshold level depends on the unknown sparsity and signal strength. To construct an adaptive thresholding test, [65] considered maximizing the standardized $T_{\gamma n}(s)$ over a subset of $(0, 1)$. That is,

$$\widehat{\mathcal{M}}_{\gamma n} = \max_{s \in \mathcal{S}} \mathcal{T}_{\gamma n}(s), \tag{4.6}$$

where $\mathcal{T}_{\gamma n}(s) = \hat{\sigma}^{-1}_{T_{\gamma n},0}(s)\{T_{\gamma n}(s) - \hat{\mu}_{T_{\gamma n},0}(s)\}$. Here $\hat{\mu}_{T_{\gamma n},0}(s)$ and $\hat{\sigma}_{T_{\gamma n},0}(s)$ are, respectively, accurate enough estimators of mean $\mu_{T_{\gamma n},0}(s)$ and standard error $\sigma_{T_{\gamma n},0}(s)$ of $T_{\gamma n}(s)$ under the H_0. Under some regularity conditions, it has been shown in [65] that under H_0, for $\mathcal{S} = (0, 1-\eta]$ with a $\eta \in (0,1)$,

$$P\{a(\log p)\widehat{\mathcal{M}}_{\gamma n} - b(\log p, \eta) \le x\Big\} \to \exp(-e^{-x}), \tag{4.7}$$

where $a(y) = \sqrt{2\log(y)}$ and $b(y, \eta) = 2\log(y) + 2^{-1}\log\log(y) - 2^{-1}\log(4\pi/(1-\eta)^2)$.

To construct an adaptive thresholding test in (4.6), one needs to obtain accurate enough estimation for $\mu_{T_{\gamma n},0}(s)$ and $\sigma_{T_{\gamma n},0}(s)$. If the exact sampling distribution of $\bar{X}_j$ is known, one may obtain an estimator that is accurate enough. However, in most applications, the exact distribution of X_{ij} is unknown, approximations of $\mu_{T_{\gamma n},0}(s)$ and $\sigma_{T_{\gamma n},0}(s)$ are needed. Because the thresholding statistic considers the tail behavior of $\bar{X}_j$, the magnitude of $\mu_{T_{\gamma n},0}(s)$ is small, and hence it requires an approximation with an accuracy that is a smaller order of the magnitude of $\mu_{T_{\gamma n},0}(s)$. To obtain an accurate enough estimator, [65] has shown that large deviation results can be applied so that

$$\frac{\mu_{T_{2n},0}(s) - \hat{\mu}_{T_{2n},0}(s)}{\sigma_{T_{2n},0}(s)} = O\{L_p p^{(1-s)/2} n^{-1/2}\},$$

where L_p is a slowly varying function that is a function of $\log(p)$, and $\hat{\mu}_{T_{2n},0}(s)$ is obtained by assuming that W_{ij} is Gaussian distributed.

If one assumes that $p = n^{1/\theta}$ and $s \in ((1-\theta)_+, 1)$ for a positive θ, then the above ratio goes to zero. If $\theta \ge 1$, the above ratio holds for $s \in (0,1)$. Under this senario, [65] show that the muti-level thresholding statistics is able to achieve the detection boundary $\rho(\beta)$ given in (4.3) when $\theta \ge 1$. While for higher dimensions with $\theta < 1$, one has to restrict the range for $s \in (1-\theta, 1)$, the optimal detection boundary is modified correspondingly as

$$\rho_\theta(\beta) = \begin{cases} (\sqrt{1-\theta} - \sqrt{1-\beta-\theta/2})^2 & \text{if } 1/2 < \beta \le (3-\theta)/4 \\ \beta - 1/2, & \text{if } (3-\theta)/4 < \beta \le 3/4 \\ (1-\sqrt{1-\beta})^2, & \text{if } 3/4 < \beta < 1. \end{cases} \tag{4.8}$$

By choosing $\hat{\mu}_{T_{2n},0}(s)$ and $\hat{\sigma}_{T_{\gamma n},0}(s)$ using the corresponding approximation given by the large deviation results in [65], the computational burden of the maximal thresholding statistic can be significantly reduced by noting that $\hat{\mu}_{T_{\gamma n},0}(s)$ and and $\hat{\sigma}_{T_{\gamma n},0}(s)$ are decreasing function of s, hence the maximization can be obtained at a finite set of discrete thresholds.

The concepts of detectability and non-detectability concern the limiting behavior of statistical signal detection, where probability of the type I error tends to zero and the power (one minus probability of the type II error) tends to one. From a practical perspective, sometimes, it maybe more relevant to evaluate procedures under a given nominal type I error level. In such cases, comparing the power of different detection methods becomes both interesting and important, though technically more delicate and challenging. In this direction, [65] compared the asymptotic power $\mathcal{B}_\gamma(r,\beta)$ of three multi-level thresholding tests using $\mathcal{M}_{\gamma n}$ as the test statistics for $\gamma = 0, 1, 2$ with a given signal size r and sparsity level β, and found that, asymptotically,

$$\mathcal{B}_0(r,\beta) \le \mathcal{B}_1(r,\beta) \le \mathcal{B}_2(r,\beta),$$

for $r > 2\beta - 1$ and $\mathcal{B}_\gamma(r,\beta)$ ($\gamma = 0,1,2$) are asymptotically equivalent when $r \in (\rho(\beta), 2\beta-1]$. Note that $\mathcal{B}_0(r,\beta)$ is the asymptotic power of the test based on the HC test. This means that the multi-L_2 thresholding is more powerful than the HC test when the signal strength r is above the detection boundary.

4.3 Multi-Thesholding after Signal Enhancement An important direction concerns the behavior of HC for $\Sigma \ne I_p$ (the so-called dependent case for Gaussian data) was considered in Hall and Jin (2008,2010) [29, 30]. They proposed the innovated higher criticism (iHC), which adapts HC to the dependent data. Let $X = (X_1, \ldots, X_p)^T$ and recall that the standard mixture model in (4.1) assumes $X = \mu + Z$ where $Z \sim N(0, I_p)$ where μ is an unknown mean vector. The iHC generalizes this framework by allowing $Z \sim N(0, \Sigma)$, where Σ has unit diagonal entries and off-diagonal entries that decay at a polynomial rate [30]. Let $\Omega = \Sigma^{-1} = (\omega_{kl})$ denote the precision matrix. The key idea of iHC is to transform the data to $\Omega X = \Omega\mu + \Omega Z$, which can potentially increase the size of signals in $\Omega\mu$. The standard HC statistic HC_n^* can then be applied to the transformed data ΩX. A surprising and important finding in [30] is that iHC can attain a lower detection boundary than standard HC for certain classes of covariance structures. Specifically, define $\underline{\omega} = \liminf_{p\to\infty}(\min_k \omega_{kk})$ and

$\bar{\omega} = \limsup_{p\to\infty}(\max_k \omega_{kk})$. They show that iHC is powerful whenever $r \geq \underline{\omega}^{-1}\rho(\beta)$, where $\rho(\beta)$ is the detection boundary given in (4.3) for the HC test. More importantly, in the matrix classes considered in [30], both $\underline{\omega}$ and $\bar{\omega}$ exceed 1. This implies that iHC can successfully detect signals weaker than those detectable by standard HC. In the special case where $\underline{\omega} = \bar{\omega}$, the detection boundary achieved by iHC is in fact optimal.

The innovated HC developed in Hall and Jin (2010) [30] enjoys the benefit of cultivating dependence to improve the detection boundary. However, these results are obtained under the assumption that the precision matrix Ω is known. In practice, however, Ω is typically unknown and needs to be estimated, and the impact of the estimated Ω and the Gaussian assumption in iHC were also not clear. Motivated by this, Chen, Li and Zhong (2019) [10] proposed an improved multi-level thresholding test for two-sample and ANOVA tests for non-Gaussian distributions where Ω is estimated using regularization method proposed by [6].

Let $X_{i1}, \cdots, X_{in_i}$ be an iid p-dimensional sample drawn from a p-dimensional distribution with mean μ_i and variance Σ_i for $i = 1, \cdots, m$ and $m \geq 2$. [10] considered the following ANOVA test problem

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m \quad \text{and} \quad H_1 : \mu_k \neq \mu_l \text{ for some } 1 \leq k \neq l \leq m. \tag{4.9}$$

For easy presentation, we focus on the two-sample case with $m = 2$ below. Let $T_{nk} = n(\bar{X}_{1,k} - \bar{X}_{2,k})^2$ be an estimator of $(\mu_{1,k} - \mu_{2,k})^2$ for $k = 1, \cdots, p$ and $n = n_1 n_2/(n_1 + n_2)$. Let $\kappa = \lim_{n\to\infty} n_1/(n_1 + n_2)$. Let $\Omega = \Sigma_w^{-1}$ where $\Sigma_w = (1 - \kappa)\Sigma_1 + \kappa\Sigma_2$ is the variance of $\sqrt{n}(\bar{X}_{1,k} - \bar{X}_{2,k})$. To incorporate the dependence into the thresholding test statistic, consider transforming data by multiplying Ω on the data points X_{ik}. Because Ω is unknown, [10] considers estimating it by a banded estimator $\hat{\Omega}_\tau$ proposed by [6], where τ is a chosen banding width. It is noted that the banding or other regularization is needed to ensure consistent estimation of a high dimensional Σ as the sample covariance matrix is not longer consistent.

The transformed data are $\{\hat{Z}_{1i} = \hat{\Omega}_\tau X_{1i} : 1 \leq i \leq n_1\}$ and $\{\hat{Z}_{2i} = \hat{\Omega}_\tau X_{2i} : 1 \leq i \leq n_2\}$. Based on the transformed data, the thresholding test statistic is

$$J_n(s,\tau) = \sum_{k=1}^{p} \left\{ \frac{n(\bar{\hat{Z}}_{1,k} - \bar{\hat{Z}}_{2,k})^2}{\hat{\omega}_{kk}} - 1 \right\} I \left\{ \frac{n(\bar{\hat{Z}}_{1,k} - \bar{\hat{Z}}_{2,k})^2}{\hat{\omega}_{kk}} > \sqrt{2s\log(p)} \right\}, \tag{4.10}$$

where $\bar{\hat{Z}}_{l,k}$ is the sample mean of $Z_{li,k}$ for $l = 1, 2$, and $\hat{\omega}_{kk}$ is the k-th diagonal component of $\hat{\Omega}_\tau$. A test based on $L_n(s)$ without the data transformation can be proposed in an analogy to $J_n(s,\tau)$. Define $\text{SNR}_{L_n} = (\mu_{L_n(s),1} - \mu_{L_n(s),0})/\sigma_{L_n(s),1}$ and $\text{SNR}_{J_n} = (\mu_{J_n(s),1} - \mu_{J_n(s),0})/\sigma_{J_n(s),1}$ as the SNR for the test based on $L_n(s)$ and $J_n(s)$, respectively. The power of the single threshold tests are determined by their corresponding SNRs. It was shown in [10], $\text{SNR}_{J_n} \geq \text{SNR}_{L_n}$ with probability one, which indicates that the transformed thresholding test has a better power than that of the thresholding test based on the untransformed data.

Assume the banding width τ in $\hat{\Omega}_\tau$ is at the order of $(n^{-1}\log p)^{-1/\{2(\nu+1)\}}$ and p is at the order of $n^{1/\theta}$ for $\theta \in (0, 1)$. An adaptive version of the single thresholding statistic (4.10) proposed in Chen, Li and Zhong (2019) [10] is given by

$$\widehat{\mathcal{M}}_{\hat{J}_n} = \max_{s\in\Lambda_n} \frac{\hat{J}_n(s,\tau) - \hat{\mu}_{J_n(s,\tau),0}}{\hat{\sigma}_{J_n(s,\tau),0}}, \tag{4.11}$$

where $\Lambda_n = (1 - \nu\theta/(\nu + 1), 1 - \eta)$. It was shown in [10], under some regularity conditions, the asymptotic distribution of $\widehat{\mathcal{M}}_{\hat{J}_n}$ under H_0 given in (4.9) is

$$P\Big\{a(\log p)\widehat{\mathcal{M}}_{\hat{J}_n} - b(\log p, \nu\theta/(\nu + 1) - \eta) \leq x\Big\} \to \exp(-e^{-x}), \tag{4.12}$$

where the functions $a(\cdot)$ and $b(\cdot)$ are defined in (4.7). It was further shown in [10], the adaptive test statistic $\mathcal{M}_{\hat{J}_n}$ has the sum of the type I and II errors goes to zero if $r > \underline{\omega}^{-1}\rho_{\nu,\theta}(\beta)$ where

$$\rho_{\nu,\theta}(\beta) = \begin{cases} \left\{\sqrt{1 - \nu\theta/(\nu + 1)} - \sqrt{1 - \beta - \nu\theta/2(\nu + 1)}\right\}^2 & \text{if } 1/2 < \beta \leq 3/4 - \nu\theta/4(\nu + 1) \\ \beta - 1/2, & \text{if } 3/4 - \nu\theta/4(\nu + 1) < \beta \leq 3/4 \\ (1 - \sqrt{1 - \beta})^2, & \text{if } 3/4 < \beta < 1. \end{cases} \tag{4.13}$$

Compared with the detection boundary obtained by iHC discussed above, the detection boundary in (4.13) not only confirms the advantage of transformation, namely, lowering the boundary by a factor of $\underline{\omega}^{-1}$, but also makes explicit the impact of $\hat{\Omega}\tau$. For instance, when $1/2 < \beta \le \frac{3}{4} - \frac{\nu\theta}{4(\nu+1)}$, it can be shown that $\rho_{\nu,\theta}(\beta) \ge \rho(\beta)$. The above mentioned results for two-sample test problems have also been generalized to ANOVA test problems in [10] where there are more than two sources or populations.

5 Higher Order Power Comparison and Combination Tests We have categorized the aforementioned high dimensional tests into three basic classes, namely L_2, L_∞ and thresholding tests. Let $\sum_{i=1}^p 1(\tau_i \neq 0) \ge p^{1-\beta}$ and $m\tau_i^2 \ge h^2$ with $h = p^{r/2}$ for $0 < \beta < 1/2$ and $h = \sqrt{2r\log p}$ for $1/2 < \beta < 1$. Results in Ingster (1997) [32], Donoho and Jin (2004) [21], Cai, Jeng and Jin (2011) [53] and Qiu, Chen and Qiu (2025) [48] showed that $\rho(\beta)$ in (4.3) is the detection boundary for Gaussian data with $\Sigma_1 = \Sigma_2 = I_p$, where the detection boundary $\rho(\beta) = \beta - 1/2$ is extended for the dense signal regime of $\beta \in (0, 1/2)$. The maximin power based on the likelihood function has a trivial limit α when $r < \rho(\beta)$ while the worst case power of the L_2, L_∞ and thresholding tests can converge to 1 when $r > \rho(\beta)$ with $\beta \in (0, 1/2)$, $\beta \in (3/4, 1)$ and $\beta \in (0, 1)$, respectively. As a leading order performance measure that focuses on the trivial power limit α or 1, the detection boundary fails to differentiate the power performance between the L_2 and thresholding tests for dense signals with $\beta \in (0, 1/2)$ and the L_∞ and thresholding tests for highly sparse signals with $\beta \in (3/4, 1)$.

5.1 Higher Order Power Comparison For Gaussian data with $\Sigma_1 = \Sigma_2 = I_p$, Qiu, Chen and Qiu (2025) [48] suggested to use the least strength function

$$\mathcal{H}(\psi; \beta, \delta) = \inf\Big\{h > 0 : \inf_{H_1(\beta,h)} \mathbb{P}(\psi \text{ rejects } H_0) \ge \delta\Big\} \tag{5.1}$$

and the minimax strength function

$$\mathcal{H}(\Psi_\alpha; \beta, \delta) = \inf_{\psi\in\Psi_\alpha} \mathcal{H}(\psi; \beta, \delta) = \inf\Big\{h > 0 : \sup_{\psi\in\Psi_\alpha} \inf_{H_1(\beta,h)} \mathbb{P}(\psi \text{ rejects } H_0) \ge \delta\Big\} \tag{5.2}$$

to compare the higher order power information of a test ψ of interest and the minimax optimal test, where $H_1(\beta, h)$ denotes the alternative hypothesis with $\sum_{i=1}^p 1(\tau_i \neq 0) \ge p^{1-\beta}$ and $m\tau_i^2 \ge h^2$.

The strength function approach can be more informative than the detection boundary since $\rho(\beta)$ is essentially a leading order term of $\mathcal{H}(\Psi_\alpha; \beta, \delta)$. Indeed,

$$\mathcal{H}^2(\Psi_\alpha; \beta, \delta) = \begin{cases} \sqrt{2}\{z_\alpha - z_\delta + o(1)\}\exp\{\rho(\beta)\log p\}, & 0 < \beta < 1/2, \\ 2\rho(\beta)\log p + \log 2 + 2\log\{z_\alpha - z_\delta + o(1)\}, & 1/2 < \beta < 3/4, \\ 2\rho(\beta)\log p + c_1(\beta)(\log_2 p + \log \pi) + c_2(\alpha, \beta, \delta) + o(1), & 3/4 < \beta < 1, \end{cases} \tag{5.3}$$

where $c_1(\cdot)$ and $c_2(\cdot,\cdot,\cdot)$ are some functions. Then, the power performance of a test of interest can be reflected by the difference between its least strength function and the minimax strength function. To this end, Qiu, Chen and Qiu (2025) [48] proposed the minimax relative deficiency (MRD) and the minimax absolute deficiency (MAD) given, respectively, by

$$\text{MRD}(\psi; \alpha, \beta, \delta) = \frac{\mathcal{H}^2(\psi; \beta, \delta)}{\mathcal{H}^2(\Psi_\alpha; \beta, \delta)} \quad \text{and} \quad \text{MAD}(\psi; \alpha, \beta, \delta) = \mathcal{H}^2(\psi; \beta, \delta) - \mathcal{H}^2(\Psi_\alpha; \beta, \delta). \tag{5.4}$$

The both measures negatively reflect the power performance of a test with the optimal MRD being 1 and the optimal MAD being 0. As presented in Table 5.1, they showed that the L_2 test is optimal in terms of the MRD and the MAD for dense signals with $\beta \in (0, 1/2)$, the L_∞ test is optimal in MRD and nearly optimal up to an $O(1)$ term in MAD for highly sparse signals with $\beta \in (3/4, 1)$, and the HC test is optimal in MRD and nearly optimal up to an $O(\log\log p)$ term in MAD for sparse signal regime of $\beta \in (1/2, 1)$.

5.2 Combination Tests The minimax deficiency measures differentiate the power performances of the L_2, L_∞ and thresholding tests, quantitatively showing that each test can be powerful for some particular signal sparsity regime and may be powerless for other regions. A natural idea is to combine any two or all of the three tests to construct a more powerful one. Ingster, Tsybakov and Verzelen (2010) [36] proposed an L_2 and HC

Table 5.1: MRDs and MADs of the L_2, L_∞ and HC tests

<table>
<tr><th>Test</th><th>β</th><th>MRD</th><th>MAD</th></tr>
<tr><td rowspan="3">L_2</td><td>$(0, 1/2)$</td><td>$\to 1$</td><td>$\to 0$</td></tr>
<tr><td>$(1/2, 3/4)$</td><td rowspan="2">$= O(p^{\beta-1/2}/\log p)$</td><td rowspan="2">$= O(p^{\beta-1/2})$</td></tr>
<tr><td>$(3/4, 1)$</td></tr>
<tr><td rowspan="3">L_∞</td><td>$(0, 1/2)$</td><td>$= O(p^{1/2-\beta}\log p)$</td><td rowspan="2">$= O(\log p)$</td></tr>
<tr><td>$(1/2, 3/4)$</td><td>$= O(1)$</td></tr>
<tr><td>$(3/4, 1)$</td><td>$\to 1$</td><td>$= O(1)$</td></tr>
<tr><td rowspan="3">HC</td><td>$(0, 1/2)$</td><td>$= O((\log\log p)^{1/2})$</td><td>$\to 0$</td></tr>
<tr><td>$(1/2, 3/4)$</td><td rowspan="2">$\to 1$</td><td rowspan="2">$= O(\log\log p)$</td></tr>
<tr><td>$(3/4, 1)$</td></tr>
</table>

combined test for the regression problem. Fan, Liao and Yao (2015) [24] proposed a general power enhancement framework that combines the L_2 and L_∞ tests to gain a higher power performance under both dense and sparse signal regimes, which was adopted by Yu, Li, Xue and Li (2023) [57] for testing high dimensional mean vectors and covariance matrices simultaneously. Qiu, Chen and Qiu (2025) [48] further proposed an L_2, L_∞ and HC combined power enhancement test for general high dimensional hypotheses that is optimal in MRD for the whole region of $\beta \in (0, 1)$ and nearly optimal in MAD up to an $O(\log\log p)$ term.

6 Detection for Changes in High Dimensional Covariances

6.1 Testing Homogeneity of Several Covariances In multivariate analysis, testing the equality of several covariance matrices is a classical problem. Assume $X_{g1}, \cdots, X_{gn_g} \overset{i.i.d}{\sim} N_p(\mu_g, \Sigma_g)$ are p-dim random vectors from $m \geq 2$ populations ($1 \leq g \leq m$). The problem of interest is to test the changes among covariance matrices Σ_g. That is

$$H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_m \text{ versus } H_1 : \Sigma_k \neq \Sigma_l \text{ for some } 1 \leq k \neq l \leq m. \tag{6.1}$$

When the dimension p is fixed and the sample sizes n_g are larger than p, the likelihood ratio statistic for testing (6.1) is given by [1]

$$\lambda_1 = \frac{\prod_{g=1}^m |A_g|^{n_g/2}}{|A|^{N/2}} \frac{N^{pN/2}}{\prod_{g=1}^m n_g^{pn_g/2}}, \tag{6.2}$$

where $N = \sum_{g=1}^m n_g$, $A_g = \sum_{j=1}^{n_g}(X_{gj} - \bar{X}_g)(X_{gj} - \bar{X}_g)^T$, $\bar{X}_g = \sum_{k=1}^{n_g} X_{gk}/n_g$, and $A = \sum_{g=1}^m A_g$. When p is fixed and the sample sizes n_g go to infinity, the asymptotic distribution of $-2\log(\lambda_1)$ is χ^2 with degree of freedom $\{(m-1)p(p+1)\}/2$. It is clear that when the data dimension is larger than the sample size, the likelihood ratio statistic λ_1 is degenerate since A_g has determinant 0. Besides the likelihood ratio tests, [46] proposed another invariance test based on the criterion $\frac{1}{2}\sum_{g=1}^m \mathrm{tr}\{(S_g^{-1}S - I)^2\}$, where $S_g = A_g/n_g$ and $S = A/N$. Because the sample covariances are degenerated, similar to the likelihood ratio test, the criterion in [46] is not applicable because the inverses of the sample covariances do not exist when p is larger than the sample sizes. Under the growing dimension setting where the data dimension p is growing with the sample size but $p < \min_{1\leq g\leq m} n_g$, [4] has shown that the likelihood ratio statistic λ_1 diverges to infinity almost surely, so the classical χ^2 approximation could lead to many false rejections under H_0.

To address the challenges of distributional approximation when the dimension grows with the sample size, two main approaches have been developed in the literature. The first approach derives the limiting distribution using tools from large-dimensional random matrix theory. For example, [4] proposed a corrected likelihood ratio statistic based on λ_1, introducing new correction terms and establishing a revised asymptotic distribution for λ_1 under $p/n \to c \in (0, 1)$. Their results showed that, under H_0, the asymptotic distribution is standard normal rather than chi-square. Further contributions along this line include [61, 69, 59, 62], where the growth condition on the dimension relative to the sample size was relaxed to $p/n \to c \in (0, \infty)$ [59, 62, 56]. The second approach, developed by [38, 39], establishes a central limit theorem (CLT) for the likelihood ratio statistics under H_0 in the asymptotic regime $p/n \to c \in (0, 1)$ using moment generating functions. However, these results either rely on the assumption of multivariate normality or are not applicable in ultra-high dimensional settings where p is much larger than the sample size.

In addition to correcting likelihood ratio statistics, another broad class of methods introduces new criteria that are robust to high dimensionality, thereby avoiding the singularity issues caused by degenerate sample covariance matrices. For testing $H_0 : \Sigma = I_p$, Ledoit and Wolf (2002) [42] proposed a modified statistic,

$$W = \frac{1}{p}\mathrm{tr}\{(S-I)^2\} - \frac{1}{p}\mathrm{tr}\{(S-I)^2\} + \frac{p}{n},$$

to replace the earlier criterion $\frac{1}{p}\mathrm{tr}(S-I)^2$ suggested by [46], which was found not robust and not applicable in the high dimensional regime by [42]. The test for identity problem has also been investigated by Chen et al. (2010) [13] for the ultra-high dimensional case without distributional assumption. A signal detection problem for $\Sigma = I_p$ was also considered in [2]. For testing the homogeneity of covariance matrices, [50] developed a test based on an unbiased estimator of $\sum_{i<j}\mathrm{tr}(\Sigma_i - \Sigma_j)^2$, which was constructed using the traces of the sample covariance matrices $\mathrm{tr}(S_i)$. Both [42] and [50] derived asymptotic distributions of their proposed tests under the regime $p/n \to c \in (0,\infty)$. Along similar lines, [51] proposed estimators for the sum of weighted pairwise Frobenius norm distances between covariance matrices, also leveraging trace-based functions of the sample covariances.

There are few test procedures for detecting the changes among covariance matrices for the ultra-high dimensional scenarios with $p/n_g \to \infty$. We focus on introducing a procedure for detecting the changes among two covariances developed in Li and Chen (2012)[44]. Their test statistic was constructed based on an unbiased estimation of the squared Frobinuous norm of the difference of two covariance matrices, that is $\mathrm{tr}\{(\Sigma_1-\Sigma_2)^2\}$. By noting that $\mathrm{tr}\{(\Sigma_1-\Sigma_2)^2\} = \mathrm{tr}(\Sigma_1^2) + \mathrm{tr}(\Sigma_2^2) - 2\mathrm{tr}(\Sigma_1^2)$, Li and Chen (2012)[44] constructed unbiased estimators for each term. Using a method in Chen et al. (2010) [13], the estimator for $\mathrm{tr}(\Sigma_g^2)$ for $g = 1, 2$ is given by

$$\begin{aligned}\widehat{\mathrm{tr}(\Sigma_g^2)} = {} & \frac{1}{n_g(n_g-1)}\sum_{i\neq j}(X_{gi}^T X_{gj})^2 - \frac{2}{n_g(n_g-1)(n_g-2)}\sum_{i,j,k}^{*} X_{gi}^T X_{gj} X_{gj}^T X_{gk} \\ & + \frac{1}{n_g(n_g-1)(n_g-2)(n_g-3)}\sum_{i,j,k,l}^{*} X_{gi}^T X_{gj} X_{gk}^T X_{gl},\end{aligned} \tag{6.3}$$

where $\sum^*$ denotes summation over mutually distinct indices. For example, $\sum^*_{i,j,k}$ means summation over $\{(i,j,k) : i \neq j, j \neq k, k \neq i\}$. The second and third terms in the above expression are used to account for non-zero means μ_1 and μ_2. Applying a similar idea, an estimator for $\mathrm{tr}(\Sigma_1\Sigma_2)$ can be constructed as

$$\begin{aligned}\mathrm{tr}\widehat{(\Sigma_1\Sigma_2)} = {} & \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(X_{1i}^T X_{2j})^2 - \frac{1}{n_1(n_1-1)n_2}\sum_{i,k}^{*}\sum_{j} X_{1i}^T X_{2j} X_{2j}^T X_{1k} \\ & - \frac{1}{n_2(n_2-1)n_1}\sum_{i,k}^{*}\sum_{j} X_{2i}^T X_{1j} X_{1j}^T X_{2k} + \frac{1}{n_1 n_2(n_1-1)(n_2-1)}\sum_{i,j}^{*}\sum_{k,l}^{*} X_{1i}^T X_{2j} X_{1k}^T X_{2l}.\end{aligned} \tag{6.4}$$

The test statistic proposed by Li and Chen (2012)[44] is

$$T_{n_1,n_2} = \widehat{\mathrm{tr}(\Sigma_1^2)} + \widehat{\mathrm{tr}(\Sigma_2^2)} - 2\mathrm{tr}\widehat{(\Sigma_1\Sigma_2)}. \tag{6.5}$$

Under the assumptions (CQ1), (CQ2) in Section 2.4 and the assumption that as $\min\{n_1, n_2\} \to \infty$, $p = p(n_1, n_2) \to \infty$, for any $k, l \in \{1, 2\}$, $\mathrm{tr}(\Sigma_k\Sigma_l) \to \infty$ and

$$\mathrm{tr}\{\Sigma_i\Sigma_j\Sigma_k\Sigma_l\} = o\{\mathrm{tr}(\Sigma_i\Sigma_j)\mathrm{tr}(\Sigma_k\Sigma_l)\}.$$

Li and Chen (2012) [44] proved that

$$\sigma_{n_1,n_2}^{-1}\left[T_{n_1,n_2} - \mathrm{tr}\{(\Sigma_1-\Sigma_2)^2\}\right] \xrightarrow{d} N(0,1), \tag{6.6}$$

where $\sigma^2_{n_1,n_2}$ is the leading order variance of T_{n_1,n_2}. Given the above asymptotic normality results, an asymptotic α level test rejects H_0 if $T_{n_1,n_2} > \hat\sigma_{n_1,n_2,0} z_\alpha$ where z_α is the upper-α quantile of N(0,1) and $\hat\sigma^2_{n_1,n_2,0} = 2\widehat{\mathrm{tr}(\Sigma_1^2)}/n_2 + 2\widehat{\mathrm{tr}(\Sigma_2^2)}/n_1$.

6.2 Adaptive Tests and Detection Boundary It has been shown in Li and Chen (2012), the power function $\mathcal{B}(\Sigma_1, \Sigma_2)$ of the test based on T_{n_1,n_2} is bounded below by

$$\mathcal{B}(\Sigma_1, \Sigma_2) \geq \Phi\left(-\frac{z_\alpha}{k_n(1-k_n)} + \frac{\mathrm{tr}\{(\Sigma_1-\Sigma_2)^2\}}{\sigma_{n_1,n_2}}\right).$$

If the all the eigenvalues of Σ_1 and Σ_2 are bounded away from zero and infinity, and $\mathrm{tr}\{(\Sigma_1-\Sigma_2)^2\} = O(n^{-1}p)$, the test T_{n_1,n_2} will be powerful. Let $\delta_\beta = p^{-1}\sqrt{\mathrm{tr}\{(\Sigma_1-\Sigma_2)^2\}}$ be the average signal and δ_β is at least at the order of $n^{-1/2}p^{-1/2}$, the test T_{n_1,n_2} has nontrivial power. When all the signals have strength larger than $1/\sqrt{n}$ and the number of signals is larger than p, the Frobenius distance $\|\Sigma_1-\Sigma_2\|_F$ is larger than $(p/n)^{1/2}$, and the test proposed by Li and Chen (2012) [44] is powerful. Let $q = p(p+1)/2$ and $\delta_{jk} = \sigma_{1,jk} - \sigma_{2,jk}$ be the difference between jk component of $\Sigma_1 - \Sigma_2$. If $\delta_{jk} \neq 0$, then it is a signal and denote $S_\beta = \{(j,k) : \delta_{jk} \neq 0\}$ the set the signals. Chen et al. (2023)[9, 12] considered the family of covariance class as following:

$$\mathscr{C}(\beta, r_{0,jk}) = \left\{(\Sigma_1, \Sigma_2) : |S_\beta| \asymp q^{1-\beta} \text{ and } \delta_{jk}^2 = 2r_{0,jk}\log p/n \text{ for } (j,k) \in S_\beta\right\}. \tag{6.7}$$

In the covariance class $\mathscr{C}(\beta, r_{0,jk})$, Li and Chen (2012)[44]'s test is powerful for the dense case with $\beta \in [0, 1/2]$. However, the L_2 based test such as [44] may lose power for the sparse case with $\beta > 1/2$.

For the sparse alternatives with $\beta > 1/2$, Cai et al. (2013) [7] developed a maximal type of test statistic. Recall that S_1 and S_2 are, respectively, sample covariance estimators for Σ_1 and Σ_2. Define $S_1 = (\hat\sigma_{1,jk})$ and $S_2 = (\hat\sigma_{2,jk})$, and

$$\hat\theta_{g,jk} = \frac{1}{n_g}\sum_{i=1}^{n_g}\left[(X_{ij}-\bar X_j)(X_{ik}-\bar X_k) - \hat\sigma_{2,jk}\right]^2,$$

as an estimator for $\mathrm{Var}\big((X_{gj}-\mu_{gj})(X_{gk}-\mu_{gk})\big)$. Cai et al. (2013) [7] proposed the following maximal type test statistic

$$M_n = \max_{1\leq j\leq k\leq p} M_{n,jk} \quad \text{where} \quad M_{n,jk} = \frac{(\hat\sigma_{1,jk} - \hat\sigma_{1,jk})^2}{\hat\theta_{1,jk}/n_1 + \hat\theta_{2,jk}/n_2}. \tag{6.8}$$

They have shown that under H_0, the test statistic $M_n - 4\log p + \log\log p$ converges to a type I extreme value distribution with distribution function $\exp\{-\frac{1}{\sqrt{8\pi}}\exp(-\frac{x}{2})\}$. Based on the study in [7], the maximal type of test statistic M_n is powerful when the standardized signal $r_{0,jk} \geq 4$ in the covariance class (6.7) when $\beta > 1/2$ [9].

To detect changes in covariance matrices under both sparse and dense alternatives, [56] proposed a weighted statistic that combines the empirical characteristic function of the empirical spectral distribution with the maximum of the squared differences between entries of two sample covariance matrices. Under the assumption $p/n \to c \in (0,\infty)$, they showed that the weighted test statistic follows a weighted chi-square distribution under H_0. Building on the idea of [24], [62] developed a power-enhanced approach that combines the weighted Frobenius norm of the differences between two sample covariance matrices with an additional term involving the maximal norm of the standardized differences. Under the same asymptotic regime as [56], they derived the asymptotic null distribution and compared the power performance with existing tests. However, these combination tests are not designed for ultra-high-dimensional settings, and it remains unclear whether they can attain full power across the entire sparsity spectrum $\beta \in (1/2, 1)$ for the covariance class $\mathscr{C}(\beta, r_{0,jk})$.

Chen et al. (2023) [9, 12] developed a multi-level thresholding test for detecting signals in covariance matrices that is adaptive to both unknown sparsity and signal strength under ultra-high-dimensional settings. The proposed single-level thresholding statistic is

$$T_n(s) = \sum_{1\leq j\leq k\leq p} M_{n,jk} I(M_{n,jk} > 4s\log p),$$

for some $s \in (0,1)$. To accommodate unknown sparsity levels and signal strengths, Chen et al. (2023)[9, 12] introduced a multi-level thresholding method, inspired by the higher criticism statistic [21] and earlier multi-thresholding approaches for mean detection [65, 10]. The statistic is defined as

$$\nu_n(s_0, \eta) = \max_{s\in(s_0, 1-\eta)} \hat\sigma^{-1}_{T_n,0}(s)\{T_n(s) - \hat\mu_{T_n,0}(s)\},$$

where s_0 is a lower bound for the threshold and η is an arbitrarily small positive constant, $\hat{\sigma}_{T_n,0}(s)$ and $\hat{\mu}_{T_n,0}(s)$ are estimators obtained by large deviation results (see [9, 12] for details). An asymptotic α-level test rejects the null hypothesis whenever $\nu_n(s_0,\eta) > (q_\alpha +)$ $\nu_n(s_0,\eta) > [q\alpha + b\{\log(p), s_0, \eta\}]/a\{\log(p)\}$ where $a(y) = \{2\log(y)\}^{1/2}$ and $b(y, s_0, \eta) = 2\log(y) + \log\log(y)/2 - \log(\pi)/2 + \log(1 - s_0 - \eta)$, and q_α is the upper α quantile of the Gumbel distribution. Assume $p = n^{1/\theta}$ and $r_{jk} = r_{0,jk}/\{(1-\kappa)\theta_{1,jk} + \kappa\theta_{2,jk}\}$. Chen et al. (2023) [9, 12] has shown that the multi-level thresholding statistic is powerful whenever $\min_{(j,k)\in S_\beta} r_{jk} > \rho^*_\theta(\beta)$ where

$$\rho^*_\theta(\beta) = \begin{cases} (\sqrt{4-2\theta} - \sqrt{6-8\beta-\theta})^2 & \text{if } 1/2 < \beta \leq 5/8 - \theta/16 \\ \beta - 1/2, & \text{if } 5/8 - \theta/16 < \beta \leq 3/4 \\ (1 - \sqrt{1-\beta})^2, & \text{if } 3/4 < \beta < 1. \end{cases} \tag{6.9}$$

In addition, $\rho^*_0(\beta)$ is the detection boundary when dimension p grows exponentially fast with n by considering the degenerated polynomial growth case with $\theta = 0$ [9, 12]. Given the above detection boundary, [9, 12] noted that the multi-level thresholding test is more powerful than the maximal test proposed by [7] in the sense that the signal strength required by the maximal test is stronger than that required by the multi-level thresholding test.

6.3 Related Work for Detecting Changes in Mean and Covariances High-dimensional longitudinal or functional data refer to datasets in which high-dimensional measurements are repeatedly collected for a large number of features from a relatively small number of subjects over multiple time points. When the number of time points is small, the data are typically referred to as high-dimensional longitudinal data; when the number of time points is large, they are instead described as high-dimensional functional data [66, 67, 63]. Statistical inference in these settings must account for both spatial and temporal dependence: spatial dependence arises among the components of each high-dimensional measurement at a given time point, while temporal dependence arises among the repeated high-dimensional measurements taken across different time points. Let $X_{it} = (X_{it1}, \cdots, X_{itp})^T$ be a $p-$dimensional random vector observed for the i-th subject for $i = 1, \cdots, n$ at time t for $t = 1, \cdots, T$. Assume $E(X_{it}) = \mu_t$, [66] considered the change-point detection and estimation problem for mean vector μ_t by testing $H_0 : \mu_1 = \cdots = \mu_T$ versus $H_1 : \mu_1 = \cdots = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \cdots = \mu_{\tau_q} \neq \mu_{\tau_q+1} = \cdots = \mu_T$, where $1 \leq \tau_1 < \cdots < \tau_q < T$ are q ($q < \infty$) locations of unknown change points. If the null hypothesis H_0 is rejected, the locations of the change points are further estimated. For a given $1 \leq t \leq T-1$, to quantify the difference of two sets of mean vectors $\{\mu_t\}_{s_1=1}^t$ and $\{\mu_t\}_{s2=t+1}^T$, [66] defined a measure $M_t = \sum_{s_1=1}^t \sum_{s_2=t+1}^T \|\mu_{s_1} - \mu_{s_2}\|^2$ and then constructed a Chen-Qin type statistics $\widehat{M}_t$ to estimate M_t. Let $\hat{\sigma}_{nt,0}$ be an estimator of the standard error of the estimator $\widehat{M}_t$ under the H_0. A test statistic $\mathscr{M} = \max_{0<t/T<1} \widehat{M}_t/\hat{\sigma}_{nt,0}$ was developed in [66] to test the existence of the change points. If the existence of the change points is confirmed, an estimation procedure by $\hat{\tau} = \arg\max_{0<t/T<1} \widehat{M}_t$ was proposed by [66] to estimate the first change point, and then a binary segmentation approach was developed to identify all the change points. The change points estimator were shown to be consistent under mild conditions [66]. Moreover, [67, 63] further considered change points detection problem for covariance matrices for high dimensional longitudinal and functional data.

Acknowledgments. The acknowledgments section comes immediately before the references and after any appendices. It should be declared by `\section*{Acknowledgments}` so that it is not numbered.

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, Hoboken, NJ, 3rd ed., 2003.

[2] E. Arias-Castro, S. Bubeck, and G. Lugosi, *Detection of correlations*, The Annals of Statistics, 40 (2012), pp. 412–435, http://www.jstor.org/stable/41713640 (accessed 2025-09-30).

[3] E. Arias-Castro, E. J. Candès, and Y. Plan, *Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism*, The Annals of Statistics, 39 (2011), pp. 2533–2556, https://doi.org/10.1214/11-AOS910.

[4] Z. Bai, D. Jiang, J.-F. Yao, and S. Zheng, *Corrections to LRT on large-dimensional covariance matrix by RMT*, The Annals of Statistics, 37 (2009), pp. 3822 – 3840, https://doi.org/10.1214/09-AOS694.

[5] Z. D. Bai and M. Saranadasa, *Effect of high dimension: By an example of a two sample problem*, Statistica Sinica, 6 (1996), pp. 311–329, https://www3.stat.sinica.edu.tw/statistica/j6n2/j6n21/j6n21.htm.

[6] P. J. Bickel and E. Levina, *Regularized estimation of large covariance matrices*, The Annals of Statistics, 36 (2008), pp. 199–227, https://doi.org/10.1214/009053607000000758.

[7] T. Cai, W. Liu, and Y. Xia, *Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings*, Journal of the American Statistical Association, 108 (2013), pp. 265–277, https://doi.org/10.1080/01621459.2012.758041.

[8] T. T. Cai, W. Liu, and Y. Xia, *Two-sample test of high dimensional means under dependence*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 76 (2014), pp. 349–372, https://doi.org/10.1111/rssb.12034.

[9] S. X. Chen, B. Guo, and Y. Qiu, *Testing and signal identification for two-sample high-dimensional covariances via multi-level thresholding*, Journal of Econometrics, 235 (2023), pp. 1337–1354, https://doi.org/10.1016/j.jeconom.2022.10.008.

[10] S. X. Chen, J. Li, and P.-S. Zhong, *Two-sample and ANOVA tests for high dimensional means*, The Annals of Statistics, 47 (2019), pp. 1443 – 1474, https://doi.org/10.1214/18-AOS1720.

[11] S. X. Chen and Y.-L. Qin, *A two-sample test for high-dimensional data with applications to gene-set testing*, The Annals of Statistics, 38 (2010), pp. 808 – 835, https://doi.org/10.1214/09-AOS716.

[12] S. X. Chen, Y. Qiu, and S. Zhang, *Sharp optimality for high-dimensional covariance testing under sparse signals*, The Annals of Statistics, 51 (2023), pp. 1921 – 1945, https://doi.org/10.1214/23-AOS2310.

[13] S. X. Chen, L.-X. Zhang, and P.-S. Zhong, *Tests for high-dimensional covariance matrices*, Journal of the American Statistical Association, 105 (2010), pp. 810–819, https://doi.org/10.1198/jasa.2010.tm09560.

[14] S. X. Chen, Y. Zhang, and P. Zhong, *Tests for high-dimensional covariance matrices*, Journal of the American Statistical Association, 105 (2010), pp. 810–819.

[15] A. Delaigle and P. Hall, *Higher criticism in the context of unknown distribution, non-independence and classification*, in Perspectives in Mathematical Sciences I: Probability and Statistics, Platinum Jubilee Proceedings of the Indian Statistical Institute, N. S. N. Sastry, T. S. S. R. K. Rao, M. Delampady, and B. Rajeev, eds., vol. 7 of Statistical Science and Interdisciplinary Research, World Scientific, Hackensack, NJ, 2009, pp. 109–138, https://doi.org/10.1142/9789814273633_0006.

[16] A. Delaigle, P. Hall, and J. Jin, *Robustness and accuracy of methods for high dimensional data analysis based on student's t-statistic*, Journal of the Royal Statistical Society Series B: Statistical Methodology, 73 (2011), pp. 283–301, https://doi.org/10.1111/j.1467-9868.2010.00761.x.

[17] M. Ditzhaus and A. Janssen, *Detectability of nonparametric signals: higher criticism versus likelihood ratio*, Electronic Journal of Statistics, 12 (2018), pp. 4094 – 4137, https://doi.org/10.1214/18-EJS1502.

[18] R. L. Dobrusin, *A statistical problem arising in the theory of detection of signals in the presence of noise in a multi-channel system and leading to stable distribution laws*, Theory of Probability & Its Applications, 3 (1958), pp. 161–173, https://doi.org/10.1137/1103015.

[19] D. Donoho and J. Jin, *Higher criticism thresholding: Optimal feature selection when useful features are rare and weak*, Proceedings of the National Academy of Sciences, 105 (2008), pp. 14790–14795, https://doi.org/10.1073/pnas.0807471105.

[20] D. L. Donoho, *High-dimensional data analysis: The curses and blessings of dimensionality*, Proceedings of the American Mathematical Society, 127 (2000), pp. 1801–1820.

[21] D. L. Donoho and J. Jin, *Higher criticism for detecting sparse heterogeneous mixtures*, The Annals of Statistics, 32 (2004), pp. 962–994.

[22] D. L. Donoho and J. Jin, *Higher Criticism for Large-Scale Inference, Especially for Rare and Weak Effects*, Statistical Science, 30 (2015), pp. 1 – 25, https://doi.org/10.1214/14-STS506.

[23] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455, https://doi.org/10.1093/biomet/81.3.425.

[24] J. Fan, Y. Liao, and J. Yao, *Power enhancement in high-dimensional cross-sectional tests*, Econometrica, 83 (2015), pp. 1497–1541, https://doi.org/10.3982/ECTA12749.

[25] J. Q. Fan, *Test of significance based on wavelet thresholding and neyman's truncation*, Journal of the American Statistical Association, 91 (1996), pp. 674–688, https://doi.org/10.1080/01621459.1996.10476936.

[26] J. Q. Fan and R. Z. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, 96 (2001), pp. 1348–1360.

[27] W. Gosset, *The probable error of a mean*, Biometrika, 6 (1908), pp. 1–25.

[28] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*, Probability and Mathematical Statistics, Academic Press, New York, 1980.

[29] P. Hall and J. Jin, *Properties of higher criticism under strong dependence*, The Annals of Statistics, 36 (2008), pp. 381 – 402, https://doi.org/10.1214/009053607000000767.

[30] P. Hall and J. Jin, *Innovated higher criticism for detecting sparse signals in correlated noise*, The Annals of Statistics, 38 (2010), pp. 1686 – 1732, https://doi.org/10.1214/09-AOS764.

[31] H. Hotelling, *The generalization of Student's ratio*, Ann. Math. Statist., 2 (1931), pp. 360–378.

[32] Y. Ingster, *Some problems of hypothesis testing leading to infinitely divisible distributions*, Mathematical Methods of Statistics, 6 (1997), p. 47–69.

[33] Y. Ingster, *Minimax detection of a signal in ℓ_p metrics.*, J Math Sci, 68 (1999), p. 503–515, https://doi.org/10.1007/BF01254275.

[34] Y. Ingster, *Adaptive detection of a signal of growing dimension*, Mathematical Methods of Statistics, 10 (2001).

[35] Y. I. Ingster and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, vol. 169 of Lecture Notes in Statistics, Springer, 2003, https://doi.org/10.1007/978-0-387-21580-8.

[36] Y. I. Ingster, A. B. Tsybakov, and N. Verzelen, *Detection boundary in sparse regression*, Electron. J. Stat., 4 (2010), pp. 1476–1526, https://doi.org/10.1214/10-EJS589.

[37] P. Ji and J. Jin, *UPS delivers optimal phase diagram in high-dimensional variable selection*, The Annals of Statistics, 40 (2012), pp. 73 – 103, https://doi.org/10.1214/11-AOS947.

[38] D. Jiang, T. Jiang, and F. Yang, *Likelihood ratio tests for covariance matrices of high-dimensional normal distributions*, Journal of Statistical Planning and Inference, 142 (2012), pp. 2241–2256, https://doi.org/10.1016/j.jspi.2012.02.057.

[39] T. Jiang and F. Yang, *Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions*, The Annals of Statistics, 41 (2013), pp. 2029 – 2074, https://doi.org/10.1214/13-AOS1134.

[40] J. Jin and T. T. Cai, *Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons*, Journal of the American Statistical Association, 102 (2007), pp. 495–506, https://doi.org/10.1198/016214507000000167.

[41] O. Ledoit and M. Wolf, *Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size*, The Annals of Statistics, 30 (2002), pp. 1081–1102.

[42] O. Ledoit and M. Wolf, *Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size*, The Annals of Statistics, 30 (2002), pp. 1081 – 1102, https://doi.org/10.1214/aos/1031689018.

[43] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, New York, 3rd ed., 2005.

[44] J. Li and S. X. Chen, *Two sample tests for high-dimensional covariance matrices*, The Annals of Statistics, 40 (2012), pp. 908 – 940, https://doi.org/10.1214/12-AOS993.

[45] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, Hoboken, NJ, 1982.

[46] H. Nagao, *On Some Test Criteria for Covariance Matrix*, The Annals of Statistics, 1 (1973), pp. 700 – 709, https://doi.org/10.1214/aos/1176342464.

[47] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London, Series A, 231 (1933), pp. 289–337.

[48] J. Qiu, S. X. Chen, and Y. Qiu, *Optimally adaptive test for high dimensional hypotheses via minimax deficiency*, (2025). Manuscript.

[49] J. Qiu, S. X. Chen, and Q.-M. Shao, *Self-normalized Cramér type moderate deviation theorem for Gaussian approximation*, Ann. Statist., 53 (2025), pp. 1319–1346, https://doi.org/10.1214/25-aos2507.

[50] J. R. Schott, *A test for the equality of covariance matrices when the dimension is large relative to the sample sizes*, Computational Statistics & Data Analysis, 51 (2007), pp. 6535–6542, https://doi.org/10.1016/j.csda.2007.03.004.

[51] M. S. Srivastava and H. Yanagihara, *Testing the equality of several covariance matrices with fewer observations than the dimension*, Journal of Multivariate Analysis, 101 (2010), pp. 1319–1329, https://doi.org/10.1016/j.jmva.2009.12.010.

[52] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B, 58 (1996), pp. 267–288.

[53] T. Tony Cai, X. Jessie Jeng, and J. Jin, *Optimal detection of heterogeneous and heteroscedastic mixtures*, Journal of the Royal Statistical Society Series B: Statistical Methodology, 73 (2011), pp. 629–662, https://doi.org/10.1111/j.1467-9868.2011.00778.x.

[54] J. W. Tukey, *T13 n: The higher criticism.* Course Notes, Statistics 411, Princeton University, 1976. Unpublished manuscript.

[55] R. Wang and X. Shao, *Hypothesis testing for high-dimensional time series via self-normalization*, Ann. Statist., 48 (2020), pp. 2728–2758, https://doi.org/10.1214/19-AOS1904.

[56] Q. Yang and G. Pan, *Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices*, Journal of the American Statistical Association, 112 (2017), pp. 188–200, http://www.jstor.org/stable/45027912 (accessed 2025-09-29).

[57] X. Yu, D. Li, L. Xue, and R. Li, *Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing*, J. Amer. Statist. Assoc., 118 (2023), pp. 2548–2561, https://doi.org/10.1080/01621459.2022.2061354.

[58] C.-H. Zhang, *Nearly unbiased variable selection under minimax concave penalty*, The Annals of Statistics, 38 (2010), pp. 894–942.

[59] Q. Zhang, J. Hu, and Z. Bai, *Invariant test based on the modified correction to LRT for the equality of two high-dimensional covariance matrices*, Electronic Journal of Statistics, 13 (2019), pp. 850 – 881, https://doi.org/10.1214/19-EJS1542.

[60] S. Zhang, S. X. Chen, and Y. Qiu, *Mean tests for high-dimensional time series*, Statist. Sinica, 35 (2025), pp. 171–201.

[61] S. Zheng, Z. Bai, and J. Yao, *Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing*, The Annals of Statistics, 43 (2015), pp. 546 – 591, https://doi.org/10.1214/14-AOS1292.

[62] S. Zheng, R. Lin, J. Guo, and G. Yin, *Testing homogeneity of high-dimensional covariance matrices*, Statistica Sinica, 30 (2020), pp. 35–53.

[63] P.-S. Zhong, *Homogeneity tests of covariance for high-dimensional functional data with applications to event segmentation*, Biometrics, 79, pp. 3332–3344, https://doi.org/10.1111/biom.13844.

[64] P.-S. Zhong and S. X. Chen, *Tests for high-dimensional regression coefficients with factorial designs*, Journal of the American Statistical Association, 106 (2011), pp. 260–274.

[65] P.-S. Zhong, S. X. Chen, and M. Xu, *Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence*, The Annals of Statistics, 41 (2013), pp. 2820 – 2851, https://doi.org/10.1214/13-AOS1168.

[66] P.-S. Zhong, J. Li, and P. Kokoszka, *Multivariate analysis of variance and change points estimation for high-dimensional longitudinal data*, Scandinavian Journal of Statistics, 48, pp. 375–405, https://doi.org/10.1111/sjos.12460.

[67] P.-S. Zhong, R. Li, and S. Santo, *Homogeneity tests of covariance matrices with high-dimensional longitudinal data*, Biometrika, 106 (2019), pp. 619–634, https://doi.org/10.1093/biomet/asz011.

[68] H. Zou, *The adaptive lasso and its oracle properties*, Journal of the American Statistical Association, 101 (2006), pp. 1418–1429.

[69] T. Zou, R. Lin, S. Zheng, and G.-L. Tian, *Two-sample tests for high-dimensional covariance matrices using both difference and ratio*, Electronic Journal of Statistics, 15 (2021), pp. 135 – 210, https://doi.org/10.1214/20-EJS1783.