

High-Dimensional Statistical Inference and False Discovery Rate Control with Covariates

by

Liyuan Zheng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2025

© Liyuan Zheng 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Dr. Ping-Shou Zhong
Professor, Department of Mathematics, Statistics, and Computer Science
University Of Illinois Chicago

Supervisors: Dr. Yingli Qin
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Dr. Kun Liang
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal Members: Dr. Pengfei Li
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Dr. Mu Zhu
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: Dr. Brendan McConkey
Associate Professor, Department of Biology
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, we focus on three statistical problems. First, we consider graph-based tests for differences of two high-dimensional distributions. Second, we investigate the estimation of multiple large covariance matrices and the application to high-dimensional quadratic discriminant analysis. Lastly, we focus on controlling the false discovery rate while incorporating complex auxiliary information.

Testing whether two samples are from a common distribution is an important problem in statistics. [Friedman & Rafsky \(1979\)](#) proposed a non-parametric multivariate distribution test based on the minimal spanning tree (MST). Recently, this test has been extended under various scenarios. However, as demonstrated in Chapter 2, these extensions are not sensitive to sparse alternatives. To address this, we propose a two-step testing procedure, IM-MST. Specifically, IM-MST incorporates marginal screening while accounting for the dependence structure via energy distance, followed by MST-based tests. IM-MST combines the strength of both non-parametric screening and MST-based tests. Simulation studies and real data applications are conducted to evaluate the numerical performance of the two-step procedure, demonstrating that IM-MST exhibits substantial power gains.

When estimating covariance matrices for data from two related categories, it is reasonable to assume that these covariance matrices share certain structural components. As a result, the precision matrix (the inverse of the covariance matrix) for each category can be decomposed into three parts: a common diagonal component, a common low-rank component, and a category-specific low-rank component. This decomposition can be motivated by a factor model, where some latent factors are common across two categories while others are specific to individual categories. In Chapter 3, we propose a consistent joint estimation method for two precision matrices building on the work of [Wu \(2017\)](#). Furthermore, these estimators are applied to formulate a high-dimensional quadratic discriminant analysis (QDA) rule, for which we derive the convergence rate for the classification error.

In many genetic multiple testing applications, the signs of the test statistics provide important directional information. For example, in RNA-seq data analysis, a negative sign could suggest that the expression of the corresponding gene is potentially suppressed, while a positive sign could indicate a potentially elevated expression level. However, most existing procedures that control the false discovery rate (FDR) ignore such valuable information. In Chapter 4, we extend the covariate and direction adaptive knockoff procedure ([Tian 2020](#)) by implementing powerful predictive functions. Through simulation studies and real data analysis, we show that our procedures are competitive to existing covariate-adaptive methods. The companion R package **Codak** is available.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Professors Yingli Qin and Kun Liang. I feel so grateful for the time and effort spent on me over the past several years. Thanks to their relentless support and sincere kindness, I have learned so much during this period. Without them, this thesis would not be possible.

I extend my sincere thanks to Professors Ping-Shou Zhong from UIC, Brendan McConkey, Pengfei Li, and Mu Zhu for dedicating their time to reviewing my thesis and providing valuable feedback. I am also grateful to Professor Weiming Li from SUFE, who offered me insightful suggestions for my research. Additionally, I deeply appreciate the faculty members and staff of the Department of Statistics and Actuarial Science for their support during my doctoral study.

I extend my heartfelt gratitude to my friends — Eric, Yifei, Jingyue, Jianchu, Bingying, Kecheng, Pusheng, Chi-kuang, Wei, Xiyue, Chi, Jiayue, Yiping, Yunfeng, Diribsa, Shaishav and Bonus — for filling my time at Waterloo with joy and memorable moments.

One change point in the time series of my life occurred when I met Xin. You bring brightness and happiness, serving as a positive signal that opens up countless new possibilities. My greatest hope is to embrace and realize all these possibilities together with you in the future.

Finally, I owe my deepest gratitude to my mother, the bravest person I know, whose unwavering support has made it possible for me to be where I am today.

Table of Contents

Examining Committee	ii
Author's Declaration	iii
Abstract	iv
Acknowledgements	v
List of Figures	x
List of Tables	xii
1 Introduction	1
1.1 Two-sample distribution tests	1
1.1.1 Minimum spanning tree	1
1.1.2 Three extensions of the original MST test	2
1.1.3 Energy-distance based marginal importance measure	6
1.2 High-dimensional quadratic discriminant analysis	8
1.3 Multiple testing and false discovery rate	10
1.3.1 False discovery rate	10
1.3.2 Existing false discovery rate control methods	11
1.4 Contributions and outline	14

2	Non-parametric two-step graph-based tests for high-dimensional distributions	15
2.1	Introduction	15
2.2	Literature review	16
2.2.1	Recent extensions of MST tests	16
2.2.2	Other similarity graph-based tests	17
2.3	Some limitations of MST-based tests	17
2.3.1	k -MST construction	17
2.3.2	Sparse alternatives	18
2.4	Main contributions	19
2.5	Two-step testing procedures	19
2.5.1	IM-MST	20
2.6	Simulation	23
2.6.1	Simulation settings	23
2.6.2	Simulation results	24
2.7	Real data analysis	30
2.7.1	Single-cell RNA sequencing data	30
2.7.2	Parkinson’s disease data	32
3	High-dimensional quadratic discriminant analysis with joint estimation of two large covariance matrices	35
3.1	Introduction	35
3.1.1	Factor model-based high-dimensional covariance matrix estimation	35
3.1.2	High-dimensional Quadratic Discriminant Analysis (QDA)	37
3.1.3	Contributions and outline	38
3.2	Problem set-up	39
3.2.1	“Partially common diagonal + low-rank” - pcDL	39
3.3	Adjusted Correlation Thresholding (ACT)	40

3.4	Precision matrix estimation using ACT	44
3.5	An application to Quadratic Discriminant Analysis (QDA)	47
3.6	Simulation	50
3.6.1	Rank recovery	50
3.6.2	QDA	51
3.7	Real data analysis	56
3.8	Technical proofs	57
3.8.1	Proofs of Theorems 3.3.2 and 3.3.4	57
3.8.2	Proofs of Theorems 3.5.1 and 3.5.2	62
4	Control the false discovery rate with complex auxiliary information	70
4.1	Introduction	70
4.2	Covariate and direction adaptive knockoff procedure (Codak)	72
4.3	Implementation	73
4.3.1	Two-group model	73
4.3.2	$\mathcal{F}_k$ -measurable EM-algorithm	75
4.3.3	Codak-gam	75
4.3.4	Codak-xg	76
4.4	Simulation	77
4.4.1	Candidate procedures	77
4.4.2	Simulation studies	78
4.5	Real data analysis	83
4.6	Technical detail	87
4.6.1	Codak-gam	87
4.6.2	Codak-xg	87

5	Conclusion and future work	88
5.1	Conclusion	88
5.2	Future work	89
5.2.1	Random projected QDA	89
5.2.2	Codak with neural network	90
	References	91
	APPENDICES	98
A	Tables of Chapter 2	99

List of Figures

1.1	Two examples of MSTs.	2
1.2	An example of an MST under scale alternatives.	3
2.1	Power comparison under Model 1 ($n = m = 80$).	26
2.2	Power comparison under Model 1 ($n = 80, m = 160$).	26
2.3	Power comparison under Model 2 ($n = m = 80$).	27
2.4	Power comparison under Model 2 ($n = 80, m = 160$).	27
2.5	Power comparison under Model 3 ($n = m = 80$).	28
2.6	Power comparison under Model 3 ($n = 80, m = 160$).	28
2.7	Power comparison under Model 4 ($n = m = 80$).	29
2.8	Power comparison under Model 4 ($n = 80, m = 160$).	29
2.9	Pathways being declared to have distributional differences between Tumor and Normal tissue types.	32
3.1	Average misclassification rates for Models 1 and 2.	54
3.2	Average misclassification rates for Models 3–6.	55
4.1	The realized FDR levels and power for sets (a) and (b); ∇ , Codak-gam; $\times$, Codak-xg; $\square$, ZAP; $\bigcirc$, AdaPT; $+$, AdaPT-GMM _g . The left column shows the realized FDR levels, and the right column shows the power. Rows (a) and (b) show the results for sets (a) and (b), respectively.	79
4.2	The realized FDR levels and power for sets (c) and (d); ∇ , Codak-gam; $\times$, Codak-xg; $\square$, ZAP; $\bigcirc$, AdaPT; $+$, AdaPT-GMM _g . The left column shows the realized FDR levels, and the right column shows the power. Rows (c) and (d) show the results for sets (c) and (d), respectively.	81

4.3	The realized FDR levels and power for sets (e) and (f); ∇ , Codak-gam; $\times$, Codak-xg; $\square$, ZAP; $\bigcirc$, AdaPT; $+$, AdaPT-GMM _g . The left column shows the realized FDR levels, and the right column shows the power. Rows (e) and (f) show the results for sets (e) and (f), respectively.	82
4.4	The number of rejections over 24,908 observations produced by different methods when controlling FDR at varying levels.	84
4.5	The feature importance for modeling $\pi_0(\mathbf{x}_i)$ (top left), $\lambda(\mathbf{x}_i)$ (top right), $\alpha(\mathbf{x}_i)$ (bottom left) and $\beta(\mathbf{x}_i)$ (bottom right) produced by Codak-xg. . . .	85
4.6	The estimated $\pi_0(\mathbf{x}_i)$ (left) and its logit transform (right) for varying <code>tval</code> and <code>logcpm</code> while holding other continuous covariates at their median values and categorical covariates at their most common categories.	86

List of Tables

1.1	Results of a multiple hypotheses testing problem.	10
2.1	Number of rejections (out of 100) by gMST and wMST with $k=1,3,5,7$. . .	18
2.2	Numbers of cells sequenced for NSCLC data.	30
2.3	Number of features for PD data.	33
2.4	p -values comparison for PD data.	33
3.1	Average rank estimates (s.e.) from 100 replications.	51
3.2	Average misclassification rates (s.e.) for prostate cancer data.	56
A.1	Size comparison under Model 1.	100
A.2	Size comparison under Model 3.	101
A.3	Size comparison under Model 4.	102
A.4	Power comparison under Model 1.	103
A.5	Power comparison under Model 2.	104
A.6	Power comparison under Model 3.	105
A.7	Power comparison under Model 4.	106

Chapter 1

Introduction

1.1 Two-sample distribution tests

In this section, we introduce the minimum spanning tree (MST)-based two-sample distribution test in Section 1.1.1. Section 1.1.2 discusses some recent extensions of MST-based distribution tests. Additionally, the energy-distance-based importance measure (IM) is presented in Section 1.1.3, serving as a marginal screening tool to enhance the performance of MST-based distribution tests.

1.1.1 Minimum spanning tree

[Friedman & Rafsky \(1979\)](#) introduced a multivariate test based on the minimum spanning tree (MST). Suppose that there are n observations from sample 1 and m observations from sample 2. The MST connects all $N = n + m$ observations into a unique graph with $N - 1$ edges by minimizing the total edge length. Once the MST is constructed, we count R_0 , the number of edges that connect observations from different samples (i.e., one observation from each sample). The original MST test rejects the null hypothesis $H_0 : F = G$ if R_0 is sufficiently small.

Figure 1.1 illustrates minimum spanning trees using 16 observations. In the left panel, the observations are mixed, which results in a large number of between sample edges ($R_0 = 10$). On the other hand, the right panel shows that the observations are well separated, leading to a small number of between-sample edges ($R_0 = 2$). For the right panel, We are likely to reject the null hypothesis, whereas rejecting the null is less likely

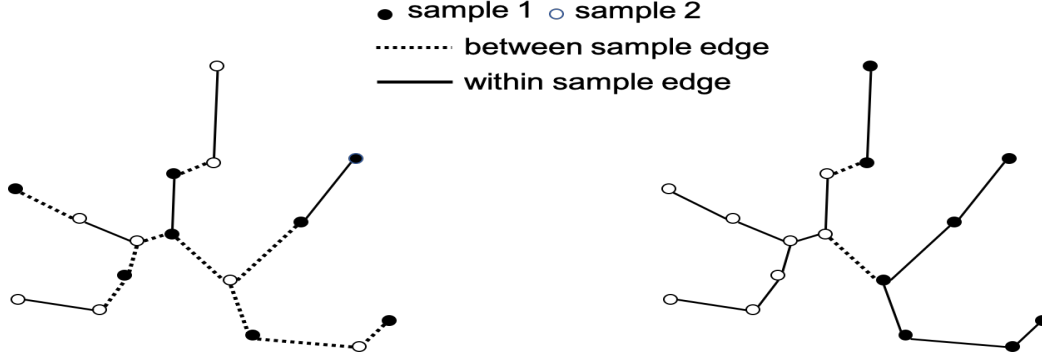


Figure 1.1: Two examples of MSTs.

with the observations in the left panel. The rationale is that if the two samples are drawn from different distributions, the pairwise distances between observations within the same sample are expected to be smaller than those between different samples. As a result, the MST is more likely to connect observations within the same sample, leading to a smaller between-sample edge count (R_0).

Under the null, the probability of a particular edge being a between-sample edge is $mn/[N(N-1)]$, and both expectation $\mathbb{E}(R_0)$ and variance $\mathbf{Var}(R_0)$ can be calculated. [Friedman & Rafsky \(1979\)](#) showed that under the null, the standardized test statistic converges weakly to the standard normal distribution,

$$\frac{R_0 - \mathbb{E}(R_0)}{\sqrt{\mathbf{Var}(R_0)}} \xrightarrow{D} N(0, 1).$$

Notice that this convergence result holds for any dimension p .

1.1.2 Three extensions of the original MST test

Generalized MST test (gMST)

Figure 1.1 illustrates the case of a locational difference, where observations within the same group are more likely to form connecting edges. Next, consider an alternative due to a scale difference, as shown in Figure 1.2. The large number of between-sample edges ($R_0 = 5$) suggests that the original MST test is likely to fail in rejecting the null hypothesis. This indicates that the original MST test has low power for detecting scale alternatives. Under such scale alternatives, observations from sample 1 (the outer group) are more likely

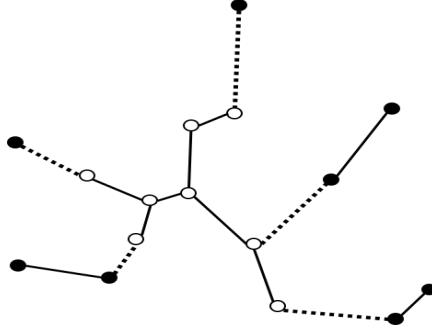


Figure 1.2: An example of an MST under scale alternatives.

to connect with observations from sample 2 (the inner group), while observations within sample 2 are more likely to connect with each other.

Instead of focusing on the number of edges connecting observations between samples, [Chen & Friedman \(2017\)](#) proposed to use the number of edges connecting observations within each sample to construct the test statistic. Intuitively, if observations from one sample are more (or less) likely to connect with each other than expected, we would reject the null hypothesis. As shown in Figure 1.2, the number of edges within sample 2 is large, while the number of edges within sample 1 is small. By considering both within-sample edge counts, we have stronger evidence to reject the null compared to relying solely on the number of between-sample edges.

Specifically, let R_1 represent the number of edges connecting observations within sample 1 and R_2 represent the number of edges connecting observations within sample 2. The generalized MST test statistic (gMST) is then defined as

$$R_g = \begin{pmatrix} R_1 - \mathbb{E}(R_1) \\ R_2 - \mathbb{E}(R_2) \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} R_1 - \mathbb{E}(R_1) \\ R_2 - \mathbb{E}(R_2) \end{pmatrix},$$

where $\Sigma = \mathbf{Var}((R_1, R_2)^T)$. Under permutation null, $\mathbb{E}(R_1), \mathbb{E}(R_2), \Sigma$ can be calculated through combinatorial analysis. [Chen & Friedman \(2017\)](#) showed that the asymptotic null distribution of the gMST test statistic is χ^2_2 . That is, under the null,

$$R_g \xrightarrow{D} \chi^2_2.$$

It's important to note that this convergence result also holds for any dimension p .

Compared to the original MST test, the gMST is effective for both locational and scale alternatives and is recommended when the type of alternative is unknown. However,

when the distribution difference is primarily locational, simulation results indicate that the original MST test is more powerful.

Weighted MST test (wMST)

Under locational alternatives, [Chen et al. \(2018\)](#) conducted simulation studies highlighting another power loss issue with the original MST test when one sample size is double that of the other ($m = 2n$). They found that, compared to the scenario with equal sample sizes ($m = n$), the power of the original MST test decreases. This finding is counterintuitive because increasing the sample size typically leads us to expect an increase in power. They demonstrated that this power loss is mainly due to the increasing variance of the original MST test statistic under unequal sample sizes. Specifically, as shown in [Friedman & Rafsky \(1979\)](#) and [Chen & Friedman \(2017\)](#), the variance of R_0 under the permutation null distribution is given by:

$$\begin{aligned} \mathbf{Var}(R_0) = & \frac{mn(m-1)(n-1)}{N(N-1)(N-2)(N-3)} \\ & \times \left(4|T| + \frac{(n-m)^2 - (N-2)}{(m-1)(n-1)} \left(\sum_{i=1}^N |T_i|^2 - \frac{4|T|^2}{N} \right) - \frac{8}{N(N-1)}|T|^2 \right), \end{aligned}$$

where $|T|$ is the total number of edges of the MST and $|T_i|$ is the number of edges connected to observation i . Notice that $\sum_{i=1}^N |T_i| = 2|T|$, and by the Cauchy-Schwarz inequality, $\sum_{i=1}^N |T_i|^2 - \frac{4|T|^2}{N}$ is non-negative. When the sample size of Sample 2 is double that of Sample 1, the number of edges connected to observation i (i.e., $|T_i|$) from Sample 1 is much smaller than that from Sample 2. In contrast, when both samples are of equal size, the values of $|T_i|$ are comparable across observations from the two samples. Therefore, under the unbalanced sample sizes, there exists greater variability in $|T_i|$, which increases $\mathbf{Var}(R_0)$ and reduces the power of the original MST test.

To address this, [Chen et al. \(2018\)](#) proposed the weighted MST test statistic (wMST), which assigns weights to R_1 and R_2 based on the reciprocal of their respective sample sizes, rather than treating them equally. To be specific, the wMST is defined as:

$$R_w = \frac{m-1}{N-2}R_1 + \frac{n-1}{N-2}R_2.$$

They demonstrated that formulating the test statistic this way minimizes its variance among all test statistics of the form $aR_1 + (1-a)R_2$, $0 \leq a \leq 1$. Notice that, the original

test statistics, which uses R_0 , is equivalent to using $R_1 + R_2$ with equal weights, given the relationship $R_0 + R_1 + R_2 = N - 1$. They then showed that under the null hypothesis,

$$\frac{R_w - \mathbb{E}(R_w)}{\sqrt{\mathbf{Var}(R_w)}} \xrightarrow{D} N(0, 1).$$

As discussed by [Chen & Friedman \(2017\)](#), under the alternative hypothesis, two scenarios can occur: (i) both samples are more likely to form connections within themselves, or (ii) one sample is more likely to form connections within itself while the other sample is less likely to do so. The gMST performs well under a broader range of alternatives compared to the wMST. However, for scenario (i), numerical studies in [Chen et al. \(2018\)](#) show that the wMST is slightly more powerful. Additionally, [Chu & Chen \(2019\)](#) proposed a max-type framework that combines the strengths of both gMST and wMST, offering a more comprehensive solution.

Ranked MST test (RISE)

Before we discuss the ranked MST test, we first introduce the concept of k -MST, which refers to the union of a series of disjoint MSTs. Specifically, the original MST is called the “first MST”. The “second MST” is constructed by minimizing the total edge length while excluding the edges appearing in the “first MST”. Then 2-MST is the union of the “first MST” and the “second MST”. Likewise, k -MST is the union of the “first MST”, “second MST”, and so forth, up to the “ k th MST”. With a larger k , a k -MST captures more similarity information between the two samples. [Friedman & Rafsky \(1979\)](#) first showed that the original MST test using a 3-MST is empirically more powerful than the original MST test using a 1-MST. Additionally, [Chen & Friedman \(2017\)](#) recommended using a 5-MST as a practical choice.

If a particular edge in the k -MST comes from the “ i th MST”, its graph-induced rank is assigned as $k - i + 1$. Notice that this graph-induced rank takes on a larger value when i is smaller. In other words, edges constructed from earlier MSTs represent higher similarity, so a higher rank should be assigned to them. Rather than treating all edges from different MSTs equally, assigning more weights to edges from earlier MSTs allows for better utilization of similarity information, thereby enhancing the performance of the test.

Instead of directly using the edge counts R_1 and R_2 , [Zhou & Chen \(2023\)](#) introduced R_1^r and R_2^r , which represent the total ranks of edges connecting observations within sample 1 and sample 2, respectively. Therefore, the Rank In Similarity graph Edge-count two-sample

test (RISE) statistic is defined as

$$R_r = \begin{pmatrix} R_1^r - \mathbb{E}(R_1^r) \\ R_2^r - \mathbb{E}(R_2^r) \end{pmatrix}^T \Sigma_r^{-1} \begin{pmatrix} R_1^r - \mathbb{E}(R_1^r) \\ R_2^r - \mathbb{E}(R_2^r) \end{pmatrix},$$

where $\Sigma_r = \mathbf{Var}((R_1^r, R_2^r)^T)$. They then showed that the asymptotic null of RISE is χ_2^2 . That is, under the null hypothesis,

$$R_r \xrightarrow{D} \chi_2^2.$$

The MST is a similarity graph constructed based on the pairwise similarities/distances between observations. RISE places more weights on the edges with higher similarity (smaller i), thereby incorporating more similarity information than both gMST and wMST. In addition, RISE benefits from the robustness property of ranks, making it less sensitive to outliers than the direct use of edge counts.

1.1.3 Energy-distance based marginal importance measure

However, when the signal is sparse, the performance of MST-based tests does not look promising, as discussed in detail in Section 2.3. A practical approach to address this challenge is to first identify informative features, thereby reducing the dimensionality of the data. Fan & Lv (2008) proposed a sure independence screening (SIS) procedure, which aims to filter out noisy features by ranking their marginal utility. The selected subset of features or dimensions contains all the informative ones with a probability approaching one. This idea has been applied to many important statistical problems, such as linear regression (Fan & Lv 2008), linear discrimination analysis (Fan & Fan 2008), and generalized linear regression (Fan & Song 2010). In addition, Li et al. (2012) proposed DC-SIS, based on marginal distance correlation, while Liu et al. (2022) proposed PC-Screen, which is based on projection correlations between features and response variables. Both methods are developed within the regression framework, and evaluate variable importance by comparing marginal utility. Unlike the methods discussed above, Peng et al. (2021) introduced a variable importance measure (IM) that accounts for the dependence structure among variables. This importance measure effectively quantifies the influence of each variable or dimension on the multivariate distributional difference between two groups.

Let X_i and Y_j be two p -dimensional vectors. Let ω be a p -dimensional weight vector. For $r = 1, \dots, p$, $\omega_r \geq 0$ is the weight for the r th dimension. Then the weighted Euclidean

distance between X_i, Y_j is

$$\|X_i - Y_j\|_\omega = \left\{ \sum_{r=1}^p \omega_r (X_{i,r} - Y_{j,r})^2 \right\}^{1/2}.$$

When $\omega = \mathbf{1}_p$, this becomes the regular unweighted Euclidean distance.

Suppose $X \sim F$ and $Y \sim G$ are two independent p -dimensional random vectors with finite means. Then, the weighted energy distance between F and G can be defined as

$$\mathcal{E}_\omega(X, Y) = 2\mathbb{E} \|X - Y\|_\omega - \mathbb{E} \|X - X'\|_\omega - \mathbb{E} \|Y - Y'\|_\omega,$$

where $X, X' \stackrel{\text{iid}}{\sim} F$ are independent of $Y, Y' \stackrel{\text{iid}}{\sim} G$. Let $\mathcal{E}(X, Y)$ denote the unweighted energy distance. [Rizzo & Székely \(2016\)](#) has stated that, in the univariate case, the squared energy distance is exactly twice the squared L^2 distance between two CDFs. Specifically,

$$\mathcal{E}(X, Y)^2 = 2 \int_{-\infty}^{\infty} (F(x) - G(x))^2 dx.$$

Although this explicit equivalence does not extend to higher dimensions, the property still holds that $\mathcal{E}(X, Y) = 0$ if and only if X and Y are identically distributed, or equivalently, $F = G$. ([Székely & Rizzo 2005](#))

As noted by [Peng et al. \(2021\)](#), the advantage of using weights is that they allow for the adjustment of weights ω_r , as a form of data perturbation to evaluate variable importance. Specifically, for the r th dimension, the variable importance measure (IM) based on the weighted energy distance can be derived from the partial derivative of $\mathcal{E}_\omega$ with respect to ω_r , evaluated at $\omega = \mathbf{1}_p$,

$$\tau_r(X, Y) \triangleq \left. \frac{\partial \mathcal{E}_\omega(X, Y)}{\partial \omega_r} \right|_{\omega=\mathbf{1}_p} = \mathbb{E} \left\{ \frac{(X_r - Y_r)^2}{\|X - Y\|} \right\} - \mathbb{E} \left\{ \frac{(X_r - X'_r)^2}{2\|X - X'\|} \right\} - \mathbb{E} \left\{ \frac{(Y_r - Y'_r)^2}{2\|Y - Y'\|} \right\}.$$

This partial derivative helps assess the impact of the r th dimension on the multivariate distributional difference. As pointed out by [Peng et al. \(2021\)](#), $\sum_{r=1}^p \tau_r(X, Y) = \mathcal{E}(X, Y)/2$. When the distributions F and G differ, there exists at least one dimension $r \in \{1, \dots, p\}$ such that $\tau_r(X, Y) > 0$. Conversely, if $F = G$, then $\tau_r(X, Y) = 0$ for every dimension $r \in \{1, \dots, p\}$.

A large positive value of τ_r indicates that the r th dimension plays an important role. Specifically, we can expect the weighted energy distance between groups to increase as ω_r

is increased, and decrease as ω_r is reduced. Therefore, we can rank the importance of each dimension based on its importance measure τ_r , and select the most important dimensions corresponding to the top-ranked τ_r values.

In practice, to estimate $\tau_r(X, Y)$, [Peng et al. \(2021\)](#) considered the following U-statistic,

$$\begin{aligned}\hat{\tau}_r(\mathcal{X}_n, \mathcal{Y}_m) = & \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \frac{(X_{i,r} - Y_{j,r})^2}{\|X_i - Y_j\|} \\ & - \frac{1}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \frac{(X_{i_1,r} - X_{i_2,r})^2}{\|X_{i_1} - X_{i_2}\|} - \frac{1}{m(m-1)} \sum_{1 \leq j_1 < j_2 \leq m} \frac{(Y_{j_1,r} - Y_{j_2,r})^2}{\|Y_{j_1} - Y_{j_2}\|}.\end{aligned}$$

[Peng et al. \(2021\)](#) showed in Theorem 5 that, under sub-Gaussian assumptions, for any constant $c > 0$, and $0 < \kappa < 1/2$, there exists a $c_1 > 0$ such that

$$P\left(\max_{1 \leq r \leq p} |\hat{\tau}_r(\mathcal{X}_n, \mathcal{Y}_m) - \tau_r(X, Y)| \geq cN^{-\kappa}\right) \leq O(pN^2 \exp\{-c_1 N^{(1-2\kappa)/3}\}).$$

This result justifies using $\hat{\tau}_r$ as an empirical importance measure for the r th dimension, allowing us to rank the importance of each dimension based on $\hat{\tau}_r$ before performing the MST-based tests.

1.2 High-dimensional quadratic discriminant analysis

Consider classifying a p -dimensional normally distributed vector $\mathbf{z}$ into one of two classes: $N(\boldsymbol{\mu}_1, \Sigma_1)$ and $N(\boldsymbol{\mu}_2, \Sigma_2)$, with prior probabilities π_1 and π_2 respectively. If both $\boldsymbol{\mu}_k$ and Σ_k , $k = 1, 2$, are known, then an optimal classification rule can be constructed. That is the oracle quadratic discriminant analysis (QDA) rule, where we classify $\mathbf{z}$ to class 1 when $Q(\mathbf{z}) > 0$, where

$$Q(\mathbf{z}) = (\mathbf{z} - \boldsymbol{\mu}_2)^T \Sigma_2^{-1} (\mathbf{z} - \boldsymbol{\mu}_2) - (\mathbf{z} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{z} - \boldsymbol{\mu}_1) - \log\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \log\left(\frac{\pi_1}{\pi_2}\right).$$

In practice, since π_k , $\boldsymbol{\mu}_k$ and Σ_k for $k = 1, 2$ are all unknown, we substitute their estimators into the oracle QDA rule, resulting in a data-driven QDA classifier.

For high-dimensional data, the dimensionality p may be larger than the number of observations n . While classical methods such as QDA are useful in lower-dimensional settings, they tend to perform poorly or even fail in high-dimensional settings.

The primary challenge in high-dimensional QDA lies in the complexity of estimating covariance matrices. Existing methods that address this issue include imposing structural assumptions on QDA and applying classical QDA after dimension reduction, as detailed in [Qin \(2018\)](#). We focus on the first type, and leave some discussions about the second type for future work in [Section 5.2.1](#).

One common structural assumption considered in high-dimensional QDA is the sparsity assumption. The sparse QDA approach has been extensively studied in the literature and has demonstrated efficiency in both simulation studies and real data analyses. Several sparse QDA rules have been proposed. In particular, two notable approaches are SQDA ([Li & Shao 2015](#)) and SDAR ([Cai & Zhang 2021](#)). SQDA pioneered the introduction of sparsity assumptions on $\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$, Σ_1 , Σ_2 and $\Sigma_1 - \Sigma_2$. By imposing sparsity assumptions, they constructed the consistent estimators of covariance matrices, and showed the convergence of classification error of the data-driven QDA rule. Furthermore, [Jiang et al. \(2018\)](#) introduced a direct estimation approach by assuming sparsity in $\Sigma_2^{-1} - \Sigma_1^{-1}$ and $\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$, and developed a consistent classification rule based on these assumptions. [Cai & Zhang \(2021\)](#) studied the necessity of imposing structural assumptions on parameters for high-dimensional QDA, and they developed an explicit convergence rate for the classification error of their proposed SDAR, providing a more rigorous theoretical foundation for high-dimensional QDA.

Instead of the sparsity assumptions, it is reasonable to assume that two classes may exhibit both common and class-specific characteristics. For example, in gene expression level data analysis, when comparing cancer and normal samples, the variance of the gene expression level may be partially explained by a common set of genes across two groups, while other genes may be regulated by distinct biological pathways unique to each group. In such cases, jointly estimating the two covariance matrices is often more effective than estimating them individually, as it captures shared information more efficiently and reduces the number of parameters that need to be estimated.

Within this framework, [Wu \(2017\)](#) assumed that the population precision matrix for group k can be decomposed into a diagonal matrix D , a shared low-rank matrix L , and a group-specific low-rank matrix $L^{(k)}$. They proposed a method that jointly estimates the precision matrices, where D and L are estimated using the entire dataset, while $L^{(k)}$ is estimated using data specific to group k . Under certain conditions on the population covariance matrices, they also established consistency results. Further details can be found in [Section 3.2](#).

1.3 Multiple testing and false discovery rate

In this section, we first introduce the false discovery rate (FDR) in Section 1.3.1. Then, we discuss several existing FDR control methods in Section 1.3.2.

1.3.1 False discovery rate

For a single hypothesis testing, we can control the probability of type I error by rejecting the null hypothesis with a p -value lower than a target level. When many null hypotheses are being tested simultaneously, referred to as a multiple hypotheses testing problem, the false discovery rate (FDR) is a common error metric (Benjamini & Hochberg 1995).

Suppose there are n null hypotheses being tested simultaneously, and $p_1, \dots, p_n$ are the associated p -values. Out of all n hypotheses, there are n_0 true null hypotheses and n_1 alternative hypotheses ($n_0 + n_1 = n$). A multiple testing procedure will reject a subset of the hypotheses based on the observed p -values. The results are summarized in Table 1.1. U is the number of true null hypotheses that we fail to reject; V represents the number of the true null hypotheses that are falsely rejected, i.e., the number of type I errors; T is the number of the alternative hypotheses that we (falsely) fail to reject, i.e., the number of type II errors, and S is the number of the alternative hypotheses that are (correctly) rejected. In total, there are R rejections and W non-rejections.

Table 1.1: Results of a multiple hypotheses testing problem.

	Fail to reject	Rejected	Total
Null	U	V	n_0
Alternative	T	S	n_1
Total	W	R	n

In multiple testing problems, if we only control the type I error rate individually, the expected number of false positives, $\mathbb{E}(V)$, will increase linearly with the total number of hypotheses n , which results in the multiplicity effect. Instead, let us consider the false discovery proportion (FDP), the proportion of type I errors among all rejections, that is,

$$\text{FDP} = \frac{V}{\max(R, 1)},$$

where $\max(R, 1)$ is used by contention to avoid 0 in the denominator. Then, the FDR is defined as the expectation of the FDP, i.e., $\text{FDR} = \mathbb{E}(\text{FDP})$, and it can be used as a measure of type I error in a multiple testing problem.

Under this notation, a multiple testing procedure with the rejection set denoted as $\mathcal{R}$, is said to control the FDR at level $\alpha \in (0, 1)$ if

$$\text{FDR}(\mathcal{R}) \leq \alpha.$$

Likewise, if a multiple testing procedure satisfies a weaker condition,

$$\limsup_{n \rightarrow \infty} \text{FDR}(\mathcal{R}) \leq \alpha,$$

then it is said to control the FDR asymptotically at level α .

Next, we discuss some methods in the literature that control the FDR in finite samples or in an asymptotic limit.

1.3.2 Existing false discovery rate control methods

[Benjamini & Hochberg \(1995\)](#) proposed a linear step-up procedure, commonly referred to as the Benjamini–Hochberg (BH) procedure. Denote the ordered p -values by $p_{(1)} \leq \dots \leq p_{(n)}$, and let $p_{(0)} = 0$ by convention. In particular, the BH procedure operates as follows. Let

$$R_\alpha^{\text{BH}} := \max \left\{ k \in \{0, \dots, n\} : p_{(k)} \leq \frac{\alpha k}{n} \right\},$$

and the rejection set is $\mathcal{R}_\alpha^{\text{BH}} := \{i : p_i \leq p_{(R_\alpha^{\text{BH}})}\}$. [Benjamini & Yekutieli \(2001\)](#) showed that under the assumption that the p -values from the true null hypotheses are independent of each other and also independent of the p -values from the alternative hypotheses, then the BH procedure controls the FDR at level $\pi_0 \alpha$, where $\pi_0 = n_0/n$ represents the proportion of the true null hypotheses.

Furthermore, to increase the number of rejections, [Storey et al. \(2004\)](#) proposed an estimator of π_0 such that the BH procedure can be applied at a higher FDR level $\alpha/\hat{\pi}_0$, specifically,

$$\hat{\pi}_0(\lambda) = \frac{\#\{i : p_i \geq \lambda\} + 1}{m(1 - \lambda)},$$

for a tuning parameter $\lambda \in [0, 1)$. The rationale of such an estimator is that most of the p -values greater than λ correspond to the true null hypotheses. Since the estimator

$\widehat{\pi}_0(\lambda)$ is a conservative estimator of π_0 , the procedure does not achieve the exact FDR control at level α , but it can reject more null hypotheses than the BH procedure. [Liang & Nettleton \(2012\)](#) proposed a class of dynamic adaptive procedures with tuning parameters determined by the data, and they established the asymptotic FDR control. [MacDonald et al. \(2019\)](#) showed that these dynamic adaptive procedures control the FDR in finite samples.

These methods operate under the assumptions of the exchangeability of the null hypotheses. Under such circumstances, the hypothesis with a p -value p_1 will always be rejected if there exists another hypothesis in the rejection set with a p -value p_2 such that $p_1 \leq p_2$. However, it is desirable to develop more flexible FDR controlling procedures without the exchangeability assumptions. [Genovese et al. \(2006\)](#) proposed the weighted BH (wBH) with some prior weights such that $\sum_{i=1}^m w_i/n = 1$. Denote the weighted p -value by $q_i = p_i/w_i$, and the ordered weighted p -values $q_{(1)} \leq \dots \leq q_{(n)}$ with $q_{(0)} = 0$ by convention. Then, the wBH procedure operates as follows. Let

$$R_\alpha^{\text{wBH}} := \max \left\{ k \in \{0, \dots, n\} : q_{(k)} \leq \frac{\alpha k}{n} \right\},$$

and the rejection set is $\mathcal{R}_\alpha^{\text{wBH}} := \{i : q_i \leq q_{(R_\alpha^{\text{wBH}})}\}$. [Genovese et al. \(2006\)](#) showed that even with uninformative guessing of the weights, the wBH controls the FDR at level $\pi_0 \mu_0 \alpha$, where μ_0 represents the expectation of the weights under the null.

Furthermore, we consider the setting that auxiliary information, denoted by a single covariate X_i , is also available for each hypothesis H_i along with the p -value p_i . It is widely recognized that incorporating auxiliary information can enhance the power of FDR controlling procedures. [Ignatiadis & Huber \(2021\)](#) developed the independence hypothesis weighting (IHW) framework to use data-driven weights calculated as a function of the covariates. To be specific, we assign each hypothesis to a fold $I_\ell, \ell = 1, \dots, K$, and the weight function h_ℓ is determined by the data from all other folds, i.e.

$$W_i = h_\ell(X_i, \mathbf{X}_{-\ell}, \mathbf{p}_{-\ell}), \quad i \in I_\ell,$$

where $\mathbf{X}_{-\ell}$ and $\mathbf{p}_{-\ell}$ represent the set of the covariates and the p -values out of fold I_ℓ , respectively. [Ignatiadis & Huber \(2021\)](#) showed that applying the wBH procedure with such a cross-weighting approach controls the FDR in finite samples.

[Lei & Fithian \(2018\)](#) proposed AdaPT, an adaptive p -value thresholding procedure, to incorporate multiple covariates. It adaptively estimates a rejection threshold $s_t(x)$ for the p -values by utilizing both the p -values and covariates. At step t , the FDR is estimated by,

$$\widehat{\text{FDR}}_t = \frac{1 + A_t}{R_t}, \quad \text{where } R_t = |\{i : p_i \leq s_t(x_i)\}| \quad \text{and } A_t = |\{i : p_i \geq 1 - s_t(x_i)\}|.$$

If the estimated FDR falls below the target level α , AdaPT terminates and rejects every hypothesis with $p_i < s_t(x_i)$. To control the FDR, for each p_i , AdaPT generates its knockoff, $1 - p_i$, and defines a masked p -value, $g(p_i) = \min(p_i, 1 - p_i)$. If $g(p_i) \leq s_t(x_i)$ at step t , the true identity is masked. Once $g(p_i) > s_t(x_i)$, the hypothesis is removed from the rejection set, and the true identity is revealed. The next threshold $s_{t+1}(x)$ is then determined based on the partially observed p -values and all the auxiliary information from covariates. [Lei & Fithian \(2018\)](#) showed that AdaPT guarantees the finite-sample FDR control.

However, these methods may not efficiently utilize the sign information from the original statistics when transforming the original statistics into the p -values. In many genetic multiple testing applications, the signs of test statistics provide important directional information. For example, in RNA-seq data analysis, a negative sign suggests that the expression of the corresponding gene is potentially suppressed, while a positive sign indicates a potentially elevated expression level. To incorporate this directional information, [Tian et al. \(2021\)](#) introduced the signed p -values, which maintains a one-to-one correspondence with the original statistics. They developed the signed-knockoff procedure, which can be viewed as a special instance of AdaPT with the sign as a covariate. Instead of using the masked p -values, their procedure presents unordered pairs of the signed p -values and their knockoffs, with the true identities gradually revealed throughout the process. Furthermore, [Tian \(2020\)](#) extends the signed-knockoff procedure to incorporate generic covariates and handle missing data, and they proposed the covariate and direction adaptive knockoff procedure, known as Codak. They showed that Codak controls the FDR in finite samples.

Similar to Codak, AdaPT-GMM_g proposed by [Chao & Fithian \(2021\)](#) can address the directional information loss. AdaPT-GMM_g generalizes the original AdaPT framework to a masking algorithm based on the z -values ([Sun & Cai 2007](#)) and leverages a working Gaussian mixture model to reveal the masked z -values iteratively. The mixture component probabilities are modeled based on a neural network featuring a single hidden layer, and the Gaussian parameters need to be estimated.

[Leung & Sun \(2022\)](#) also argued that using the p -values is sub-optimal for two-sided testing. They developed a z -value covariate adaptive procedure called ZAP. They demonstrated that the interactive relationship between the z -values and the covariates can capture structural information that enhances testing power.

1.4 Contributions and outline

Chapters 2 – 4 are based on three independent research projects. In each chapter, we begin by reviewing the relevant literature, then propose an innovative method and study it in detail. The conclusion and extension are presented in Chapter 5. The notations are consistent within each chapter and its technical proofs, though for simplicity, some notations may be redefined across different chapters.

In Chapter 2, we propose a two-step testing procedure IM-MST for sparse alternatives. Specifically, IM-MST starts with marginal screening using the importance measure that also takes the dependence structures into consideration, followed by MST-based tests. The proposed procedures inherit the nice properties of both MST-based tests and non-parametric screening. We conduct simulation studies and real data applications to demonstrate the advantages of IM-MST over other existing methods.

In Chapter 3, we propose a method to estimate the number of common factors and the number of group-specific factors. We derive the consistency properties. By using our proposed method to pre-specify the ranks, we utilize a new method proposed by [Wu \(2017\)](#) to jointly estimate high-dimensional covariance/precision matrices from two groups under a “partially common diagonal + low-rank” (pcDL) decomposition assumption. We then apply the joint estimators of two precision matrices to QDA problems and establish an explicit convergence rate for the classification error.

In Chapter 4, we extend the Codak procedure proposed by [Tian \(2020\)](#) to incorporate covariate and directional information. We develop a computationally efficient algorithm to implement Codak by a generalized additive model (gam), which is most suitable for a small number of covariates. Additionally, we develop another implementation leveraging the flexibility of gradient-boosted trees, making it particularly effective for handling a large number of covariates with complex or unknown functional forms.

Chapter 2

Non-parametric two-step graph-based tests for high-dimensional distributions

2.1 Introduction

Suppose two independent random samples

$$\mathcal{X}_n = \{X_1, X_2, \dots, X_n\} \quad \text{and} \quad \mathcal{Y}_m = \{Y_1, Y_2, \dots, Y_m\},$$

follow multivariate distributions F and G , respectively. We are interested in testing $H_0 : F = G$ against $H_1 : F \neq G$. This is a fundamental problem in statistical inference which has been extensively studied. For parametric approaches, certain model assumptions are required to facilitate the estimation of parameters, such as mean vectors and covariance matrices. Such model assumptions about F and G are particularly hard to verify for modern high-dimensional data, which are widely available in fields such as bioinformatics, social networks, and genomics.

For non-parametric tests, the Wald-Wolfowitz runs test is a commonly used test. It operates by putting all observations into a sequence and counting the number of runs, which are defined as maximal non-empty segments of the sequence consisting of elements from within the same sample ([Wald & Wolfowitz 1940](#)). For example, if sample 1 contains observations $\{104, 197, 248, 253, 315, 331\}$, and sample 2 contains observations $\{184, 196, 300, 308, 323, 355\}$. The two samples are then combined and placed in order as

the following sequence, where one bracket represents one run,

$$\underbrace{104} \underbrace{184 \ 196} \underbrace{197 \ 248 \ 253} \underbrace{300 \ 308} \underbrace{315} \underbrace{323} \underbrace{331} \underbrace{335}.$$

The two samples are mixed together, with a large number of runs (8), and we are less likely to reject the null. The rationale is that if two samples arise from two different distributions, the number of runs is expected to be small. The null hypothesis H_0 should be rejected for a small number of runs. [Friedman & Rafsky \(1979\)](#) extended the univariate Wald-Wolfowitz runs test to multivariate data, and proposed a computationally efficient MST-based test. See [Section 1.1.1](#) and [1.1.2](#) for more details about some MST-based tests. However, these tests have some limitations, as we will discuss in [Section 2.3](#).

The rest of this chapter is organized as follows. We briefly review some existing graph-based tests in [Section 2.2](#). In [Section 2.3](#), we discuss some issues with these MST-based tests. Our main contributions are summarized in [Section 2.4](#). We introduce our proposed two-step testing procedures in [Section 2.5](#). Extensive simulation studies and real data applications are presented in [Sections 2.6](#) and [2.7](#), respectively.

2.2 Literature review

2.2.1 Recent extensions of MST tests

[Chen & Friedman \(2017\)](#) proposed the generalized MST (gMST) test, which is based on the number of edges connecting observations within each sample. They demonstrated that the gMST exhibits substantial power gains over the original MST when a scale difference is present. [Chen et al. \(2018\)](#) proposed the weighted MST (wMST) test for unbalanced sample sizes, which enhances power for high-dimensional location alternatives.

Additionally, to capture more similarity information between the two, particularly for high-dimensional data, [Chen & Friedman \(2017\)](#) introduced the idea of dense MSTs, referred to as k -MSTs, which are unions of multiple disjoint MSTs. Building on this idea, [Zhou & Chen \(2023\)](#) proposed the Rank In Similarity graph Edge-count (RISE) test, which is based on the k -MST. RISE assigns greater weights on edges from earlier constructed MSTs, allowing it to incorporate more similarity information compared to treating all edges equally across different MSTs. This approach exhibits robust performance compared to the gMST. Meanwhile, [Zhu & Chen \(2024\)](#) suggested using tests based on dense graphs (e.g., 50-MST) to further improve performance.

2.2.2 Other similarity graph-based tests

In the literature, there are also other types of similarity graphs. [Rosenbaum \(2005\)](#) introduced an edge-count test using the minimum distance non-bipartite pairing (MDP). MDP is a graph that contains exactly $N/2$ edges ($N = n + m$, assuming N is even), and each node/observation is connected to exactly one other node/observation. The graph with the smallest total edge length is the desired MDP. When N is odd, they suggested creating a ghost data point with a distance 0 to all observations and then ignored the edge connected to this ghost point. [Chen & Friedman \(2017\)](#) conducted a power comparison between the MST-based and MDP-based tests, showing that MST-based tests exhibit higher power for both location and scale alternatives. [Henze \(1988\)](#) and [Schilling \(1986\)](#) proposed an edge-count test using the K-nearest neighbor graph (KNN), which connects each observation to its K-nearest neighbors. [Zhou & Chen \(2023\)](#) mentioned that simulation results show that the MST-based and the KNN-based tests perform very similarly across various settings.

2.3 Some limitations of MST-based tests

In this section, we discuss some issues surrounding MST-based tests, particularly in the context of high-dimensional data. We present a simulated data example to demonstrate the power loss of the MST-based tests when signals are sparse. Specifically, the power loss occurs when the number of informative features/dimensions (true signals) is small, but the total number of features/dimensions is large.

2.3.1 k -MST construction

[Chen & Friedman \(2017\)](#) suggested using a denser MST to perform tests for high-dimensional data, as it can capture more similarity information. However, constructing a k -MST with a large value of k introduces challenges for MST-based tests. Specifically, while the total number of edges in a k -MST is fixed at $k(N - 1)$, the degree of individual nodes (i.e., the number of edges connected to a node) varies depending on the configuration of observations and can change from case to case. If a node with a large degree, known as a hub node, exists, the performance of MST-based tests would heavily depend on this hub node. Technically, this violates the technical assumption of no hub nodes required for Theorem 5.1.1 in [Chen & Friedman \(2017\)](#). In addition to the issue of hub nodes, [Petrie \(2016\)](#) pointed out that constructing a dense graph is computationally expensive. Through com-

prehensive simulation studies, they suggested that k_0 is typically smaller than $N/2$. They proposed constructing k -MSTs with $k \leq k_0 \approx 0.1N$.

2.3.2 Sparse alternatives

We provide the following example to investigate the performance of dense MSTs in the context of high-dimensional data with signal patterns ranging from dense to sparse. We consider sample sizes $m = n = 50$, dimensions $p = 50, 100, 150$, and two multivariate Gaussian distributions with $\Sigma_1 = \Sigma_2 = I_{p \times p}$. In terms of mean vectors, we set $\boldsymbol{\mu}_1 = 0$, and $d = 100\%, 50\%, 10\%$ of the p elements/dimensions of $\boldsymbol{\mu}_2$ are nonzero. These nonzero elements of $\boldsymbol{\mu}_2$ are generated from $N(1, 1)$ such that $\boldsymbol{\mu}_2^T \boldsymbol{\mu}_2 = \Delta^2$.

Table 2.1: Number of rejections (out of 100) by gMST and wMST with $k=1,3,5,7$.

p		50	100	150	150
Δ		1.4	1.4	1.4	2
$d = 100\%$	gMST(1,3,5,7)	47 62 59 58	45 45 55 47	45 46 58 68	54 58 77 72
	wMST(1,3,5,7)	60 80 84 73	47 60 71 73	53 57 68 71	55 69 76 89
$d = 50\%$	gMST(1,3,5,7)	6 4 7 10	5 4 5 32	4 4 4 45	3 60 73 73
	wMST(1,3,5,7)	5 5 33 41	4 5 34 45	4 4 35 61	36 59 72 77
$d = 10\%$	gMST(1,3,5,7)	3 3 5 10	5 4 5 7	3 4 4 5	4 44 68 70
	wMST(1,3,5,7)	3 3 4 14	4 6 12 25	2 18 17 27	24 33 44 50

Table 2.1 shows the number of rejections out of 100 at a significance level of $\alpha = 0.05$, by gMST and wMST under different settings. Table 2.1 demonstrates that (i) the performance of both tests deteriorates as d decreases from 100% to 50%, and further to 10%, when the overall signal size is fixed at $\Delta = 1.4$; (ii) as Δ increases from 1.4 to 2 with $p = 150$, the number of rejections trends upward across different values of d . In conclusion, both tests are sensitive to the value of Δ , while smaller proportions of nonzero elements (i.e., sparser signals) lead to fewer rejections for every fixed signal size Δ .

For high-dimensional data with sparse signals, the performance of both gMST and wMST is not promising, even when using k -MSTs with large k . Therefore, a natural approach to address this issue is to identify significant features first, thereby reducing the dimensionality. Fan & Lv (2008) introduced the Sure Independence Screening (SIS) to reduce computation costs in ultra-high dimensional variable selection by ranking features based on their marginal utility. In the context of linear regression, they ranked features

using the Pearson correlation. They suggested that such a sure screening method can effectively reduce dimensionality from a high scale to a moderate one, below the sample size, allowing for further statistical inference. In the context of classification, [Fan & Fan \(2008\)](#) showed that many noisy features do not contribute to reducing classification error. They further extended the SIS procedure to the context of classification. Therefore, when only a small subset of features accounts for most of the differences between groups, retaining informative features should be preferred.

2.4 Main contributions

We propose a two-step testing procedure, IM-MST, which begins with feature selection to retain informative features and only uses the selected features to construct the MST-based tests. For the first step, we adopt the importance measure (IM) proposed by [Peng et al. \(2021\)](#) to identify informative features. The benefit of IM screening is that due to the construction of IM, the dependence structure among all the features is taken into consideration. [Peng et al. \(2021\)](#) showed that the screening procedure based on IM possesses the sure screening property. Furthermore, under some mild assumptions, we establish the rank consistency property for IM screening. Therefore, our proposed two-step testing procedure retains informative features, enhancing the performance of MST-based tests.

2.5 Two-step testing procedures

When considering feature selection/screening, many existing procedures depend on a cut-off/threshold to separate informative features from non-informative ones, and the threshold can be decided by cross-validation or other tuning methods. In practice, one can select a reasonably conservative threshold to ensure that, with high probability, all informative features are included. However, we must also be cautious not to compromise the size of selected features for the sake of achieving the sure screening property by selecting an overly conservative threshold, as this is likely to result in an undesirably large model.

Several two-step procedures, involving screening in the first step followed by testing in the second, have been proposed. The first step typically incorporates FDR control for screening (see [Benjamini et al. \(2006\)](#) and [Yekutieli et al. \(2006\)](#)). These procedures focus on marginal differences in both the screening and testing steps. In contrast, our proposed two-step testing procedures focus on testing multivariate distributional differences between

groups. Specifically, it begins by selecting informative features in the first step using non-parametric importance measure, followed by applying non-parametric MST-based tests exclusively to the informative features identified.

2.5.1 IM-MST

Recall that $X \sim F$ and $Y \sim G$ are two independent p -dimensional random vectors with finite means. The importance measure (IM) for the r th dimension is defined as the partial derivative of the weighted energy distance $\mathcal{E}_\omega$ with respect to ω_r , evaluated at $\omega = \mathbf{1}_p$. That is,

$$\tau_r(X, Y) \triangleq \left. \frac{\partial \mathcal{E}_\omega(X, Y)}{\partial \omega_r} \right|_{\omega = \mathbf{1}_p} = \mathbb{E} \left\{ \frac{(X_r - Y_r)^2}{\|X - Y\|} \right\} - \mathbb{E} \left\{ \frac{(X_r - X'_r)^2}{2\|X - X'\|} \right\} - \mathbb{E} \left\{ \frac{(Y_r - Y'_r)^2}{2\|Y - Y'\|} \right\}.$$

As discussed in Section (1.1.3), $\tau_r(X, Y)$ can indeed measure the importance of the r th dimension by quantifying its influence on the multivariate distributional difference between two groups.

Given that $\mathcal{X}_n = \{X_i : 1 \leq i \leq n\}$ and $\mathcal{Y}_m = \{Y_j : 1 \leq j \leq m\}$ consist of independent and identically distributed copies of X and Y , respectively, Peng et al. (2021) considered a U-statistic, $\hat{\tau}_r$, as an estimator for $\tau_r(X, Y)$,

$$\begin{aligned} \hat{\tau}_r(\mathcal{X}_n, \mathcal{Y}_m) &= \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \frac{(X_{i,r} - Y_{j,r})^2}{\|X_i - Y_j\|} \\ &\quad - \frac{1}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \frac{(X_{i_1,r} - X_{i_2,r})^2}{\|X_{i_1} - X_{i_2}\|} - \frac{1}{m(m-1)} \sum_{1 \leq j_1 < j_2 \leq m} \frac{(Y_{j_1,r} - Y_{j_2,r})^2}{\|Y_{j_1} - Y_{j_2}\|}. \end{aligned}$$

They further showed that, under sub-Gaussian assumptions, for any constant $c > 0$, and $0 < \kappa < 1/2$, there exists a $c_1 > 0$ such that

$$P \left(\max_{1 \leq r \leq p} |\hat{\tau}_r(\mathcal{X}_n, \mathcal{Y}_m) - \tau_r(X, Y)| \geq cN^{-\kappa} \right) \leq O(pN^2 \exp \{-c_1 N^{(1-2\kappa)/3}\}).$$

This theorem provides a theoretical foundation for estimating the importance of all features simultaneously, enabling the selection of important features based on $\hat{\tau}_r$.

Denote $\mathcal{S}_0$ as the set of true important features, and assume that $\min_{r \in \mathcal{S}_0} \tau_r \geq cN^{-\kappa}$ for some constants $c > 0$ and $0 < \kappa < 1/2$. Under sub-Gaussian assumptions, and

assuming p satisfies $\log p = o(N^{(1-2\kappa)/3})$, Peng et al. (2021) showed that as $N \rightarrow \infty$, $P(\min_{r \in \mathcal{S}_0} \hat{\tau}_r \geq 0) \rightarrow 1$. This is the sure screen property, which ensures that all true informative features ($r \in \mathcal{S}_0$) would have non-negative $\hat{\tau}_r$, with high probability. This result indicates that feature selection based on $\hat{\tau}_r$ will include all true informative features, given a minimum signal strength condition.

Furthermore, we found that, under some mild conditions, feature selection based on $\hat{\tau}_r$ also enjoys the following rank consistency property.

Theorem 2.5.1. *Suppose $\min_{r \in \mathcal{S}_0} \tau_r \geq cN^{-\kappa}$ and $\max_{r \in \mathcal{S}_0^c} \tau_r < -cN^{-\kappa}$, for some constants $c > 0$ and $0 < \kappa < 1/2$. Thus, under sub-Gaussian assumptions and p satisfies that $\log p = o(N^{(1-2\kappa)/3})$, then, as $N \rightarrow \infty$,*

$$P\left(\min_{r \in \mathcal{S}_0} \hat{\tau}_r - \max_{r \in \mathcal{S}_0^c} \hat{\tau}_r > 0\right) \rightarrow 1.$$

Proof.

$$\begin{aligned} & P\left(\min_{r \in \mathcal{S}_0} \hat{\tau}_r - \max_{r \in \mathcal{S}_0^c} \hat{\tau}_r \leq 0\right) \\ & \leq P\left(\min_{r \in \mathcal{S}_0} \hat{\tau}_r - \max_{r \in \mathcal{S}_0^c} \hat{\tau}_r \leq \min_{r \in \mathcal{S}_0} \tau_r - \max_{k \in \mathcal{S}_0^c} \tau_r - 2cN^{-\kappa}\right) \\ & = P\left(\left(\min_{r \in \mathcal{S}_0} \tau_r - \min_{r \in \mathcal{S}_0} \hat{\tau}_r\right) + \left(\max_{r \in \mathcal{S}_0^c} \hat{\tau}_r - \max_{k \in \mathcal{S}_0^c} \tau_r\right) \geq 2cN^{-\kappa}\right) \\ & \leq P\left((\tau_{r_1} - \hat{\tau}_{r_1}) + (\hat{\tau}_{r_2} - \tau_{r_2}) \geq 2cN^{-\kappa}\right) \\ & \leq P\left(|\tau_{r_1} - \hat{\tau}_{r_1}| + |\hat{\tau}_{r_2} - \tau_{r_2}| \geq 2cN^{-\kappa}\right) \\ & \leq P\left(2 \max_{r=r_1, r_2} |\tau_r - \hat{\tau}_r| \geq 2cN^{-\kappa}\right) \\ & = P\left(\max_{r=r_1, r_2} |\tau_r - \hat{\tau}_r| \geq cN^{-\kappa}\right), \end{aligned}$$

where $\hat{\tau}_{r_1} = \min_{r \in \mathcal{S}_0} \hat{\tau}_r$ and $\hat{\tau}_{r_2} = \max_{r \in \mathcal{S}_0^c} \hat{\tau}_r$. Thus, $\tau_{r_1} \geq \min_{r \in \mathcal{S}_0} \tau_r$ and $\tau_{r_2} \leq \max_{r \in \mathcal{S}_0^c} \tau_r$. $\square$

This result guarantees that the informative features are distinguishable from the other features, given an assumption on the gap of signal strength between the informative and the other features.

Next, we introduce our IM-MST testing procedure, which employs IM-based selection in the first step with MST-based tests in the second step. In the screening step, we begin by including all dimensions in the selected set and sequentially remove the least informative dimension at each step. Let $\mathcal{S}$ denote the set of selected informative dimensions and $\mathcal{D}$ denote the set of deleted dimensions.

Algorithm 1: IM-MST.

Input: Data $\mathcal{X}_n$ and $\mathcal{Y}_m$, subsample sizes n_1, m_1 and the FDR level $0 < \alpha < 1$.

Initialization. Randomly sample an index set $\mathcal{I}_X$ with size n_1 from $\{1, \dots, n\}$ and an index set $\mathcal{I}_Y$ with size m_1 from $\{1, \dots, m\}$. Let $\mathcal{S}(0) = \{1, \dots, p\}$ and $\mathcal{D}(0) = \{\emptyset\}$. For $\ell \geq 1$, perform the following,

Step 1. Calculating IMs - $\Upsilon(\ell)$

For each $r \in \mathcal{S}(\ell - 1)$, calculate IM for each dimension based on $\{\mathcal{X}_{\mathcal{I}_X, \mathcal{S}(\ell-1)}, \mathcal{Y}_{\mathcal{I}_Y, \mathcal{S}(\ell-1)}\}$, and denote $\hat{\tau}_{\ell, r}$ be the IM for dimension r . Set $\Upsilon(\ell) = \{\hat{\tau}_{\ell, r} : r \in \mathcal{S}(\ell - 1)\}$.

Step 2. Deleting the least informative dimension

Find $\min \Upsilon(\ell)$, and let $d(\ell)$ be the corresponding index.

Set $\mathcal{S}(\ell) = \mathcal{S}(\ell - 1) \setminus d(\ell)$ and $\mathcal{D}(\ell) = \mathcal{D}(\ell - 1) \cup \{d(\ell)\}$.

- (a) If $\min \Upsilon(\ell) \geq 0$, set $\mathcal{S} = \mathcal{S}(\ell - 1)$, $\mathcal{D} = \mathcal{D}(\ell - 1)$, and stop iterating.
- (b) If $\min \Upsilon(\ell) < 0$, compute the MST p -value based on $\{\mathcal{X}_{\mathcal{I}_X, \mathcal{D}(\ell)}, \mathcal{Y}_{\mathcal{I}_Y, \mathcal{D}(\ell)}\}$, and denote it as p_ℓ . Perform a p -value adjustment procedure on $p_1, \dots, p_\ell$. If the adjusted $p_\ell \leq \alpha$, set $\mathcal{S} = \mathcal{S}(\ell - 1)$, $\mathcal{D} = \mathcal{D}(\ell - 1)$, and stop iterating. Otherwise, return to Step 1.

Step3. Implement MST-based tests on $\{\mathcal{X}_{\mathcal{I}_X^c, \mathcal{S}}, \mathcal{Y}_{\mathcal{I}_Y^c, \mathcal{S}}\}$, and calculate the p -value.

Output: Decision based on the p -value.

Ideally, each informative feature will have a non-negative IM at each iteration, but random variation in the importance measures can lead to negative signs in some iterations for some important features. To address this, in Step 2 (b), we test whether the feature with a negative sign is still informative. A p -value adjustment procedure is applied to account for multiple testing problems, with the Benjamini–Yekutieli procedure ([Benjamini & Yekutieli 2001](#)) being a common choice.

2.6 Simulation

In this section, we evaluate the performance of the proposed two-step procedures and compare them with existing methods using Monte Carlo simulations. Simulation results are presented for size and power comparisons across four different models, under various alternatives (sparse location differences, sparse scale differences, and sparse differences in general). We compare the following MST-based methods, with the first three being our proposed two-step testing procedures.

- IM-gMST, IM-wMST and IM-RISE (proposed two-step procedures);
- IM-MRPP by [Peng et al. \(2021\)](#), implemented using the R package `DistIM`;
- gMST by [Chen & Friedman \(2017\)](#), implemented using the R package `gTests`;
- wMST by [Chen et al. \(2018\)](#), implemented using the R package `gTests`;
- RISE by [Zhou & Chen \(2023\)](#);
- ORC-gMST: gMST with true informative features;
- ORC-wMST: wMST with true informative features;
- ORC-RISE: RISE with true informative features.

For the choice of k in MST-based tests, we follow the suggestion by [Chen & Friedman \(2017\)](#) and [Chen et al. \(2018\)](#) to use a small constant, such as $k = 5$. This setting is adopted for all MST-based tests in our simulation study.

2.6.1 Simulation settings

Throughout the simulation, we use both equal sample sizes ($n = m = 80$) and unequal sample sizes ($n = 80, m = 160$), with the dimensionality set to $p \in \{50, 100, 200, 300, 400, 500\}$. For our proposed two-step testing procedures, we randomly select three-eighths of the samples for screening, and use the remaining samples for testing. For each combination of (n, m, p) , we run 500 replications to calculate the empirical power and size, with each p -value being calculated based on 500 permutations. The significance level is set at 0.05.

We consider the following 4 models for power comparison. When evaluating empirical sizes, we use only Model 1, Model 3, and Model 4, generating observations for Sample 2 in the same manner as for Sample 1.

To quantify the difference between two multivariate Gaussian distributions, we define Δ as follows:

$$\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma_2^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Model 1: (Nor1) Sample 1 and Sample 2 are drawn from two multivariate Gaussian distributions with $\boldsymbol{\mu}_1 = 0$, $\Sigma_1 = \Sigma_2 = I$. and the first $0.05p$ dimensions of $\boldsymbol{\mu}_2$ take a common positive number such that $\Delta = 2$. As $\Delta = 2$ is fixed, increasing the number of informative features reduces the common positive number (i.e., signal strength) as the dimension p increases. This represents a sparse location difference.

Model 2: (Nor2) Sample 1 and Sample 2 are drawn from two multivariate Gaussian distributions. For Sample 1, $\boldsymbol{\mu}_1 = 0$ and $\Sigma_1 = I$. For Sample 2, $\Sigma_2 = \begin{bmatrix} \text{AR}(1) & 0 \\ 0 & I \end{bmatrix}$, where $\text{AR}(1)$ is a 10×10 autoregressive matrix of order 1 with $\rho = 0.5$. The first 10 dimensions of $\boldsymbol{\mu}_2$ are set to a common positive constant 0.55 such that $\Delta = 1.4$. This represents a combination of sparse location and scale differences.

Model 3: (t_5) Each dimension in Sample 1 is drawn from a central t_5 distribution. For Sample 2, the first 10 dimensions are drawn from a non-central t_5 with a non-centrality parameter $\delta = 0.8$, while the remaining dimensions are generated from a central t_5 .

Model 4: (Gamma) Each dimension in Sample 1 is drawn from $\text{Gamma}(2, 0.1)$. Each dimension in Sample 2 is initially drawn from $\text{Gamma}(2, 0.1)$, but its first 10 dimensions are scaled by a factor distributed as $\text{Lognormal}(\mu = 0, \sigma = \sqrt{0.6})$.

2.6.2 Simulation results

The empirical sizes for all tests under consideration are summarized in Tables A.1 - A.3.

- Under Model 1, gMST, wMST, IM-gMST, and IM-wMST maintain empirical sizes close to the nominal level of 0.05. However, the empirical sizes of RISE, IM-RISE, and IM-MRPP are slightly higher than the nominal level, particularly when the two groups have unequal sample sizes.

- Under Model 3, all tests maintain empirical sizes close to the nominal level of 0.05.
- Under Model 4, the empirical sizes of ORC-gMST and ORC-wMST are slightly above the nominal level of 0.05.

The empirical powers are summarized in Figures 2.1–2.8, and details can be found in Tables A.4 - A.7.

- Under Model 1, when there exists a location difference, the empirical powers of all tests, except ORC-gMST, ORC-wMST, and ORC-RISE, decrease as the dimension p increases. However, the power loss of our proposed two-step testing procedures is significantly smaller compared to other competing tests, with IM-wMST achieving the highest power.
- Under Model 2, when there exist both location and scale differences, IM-MSTs outperform other competing tests (excluding three original tests with true informative features), and IM-RISE achieves the highest power. Notably, the IM screening procedure accounts for the AR(1) dependence structure.
- Under Model 3, when there exists a location difference, IM-wMST outperforms all competing tests, with IM-RISE as the second-best procedure. The power difference between these two methods and others becomes more pronounced when the two samples have different sizes.
- Under Model 4, when there exists a scale difference, IM-gMST achieves the highest power, and the two-step testing procedures exhibit slightly higher power than other competing tests at high dimensions.
- With unbalanced sample sizes ($n = 80$ and $m = 160$), the IM-wMST outperforms other competing two-step MST-based testing procedures, while wMST also achieves higher power compared to other competing MST-based tests.

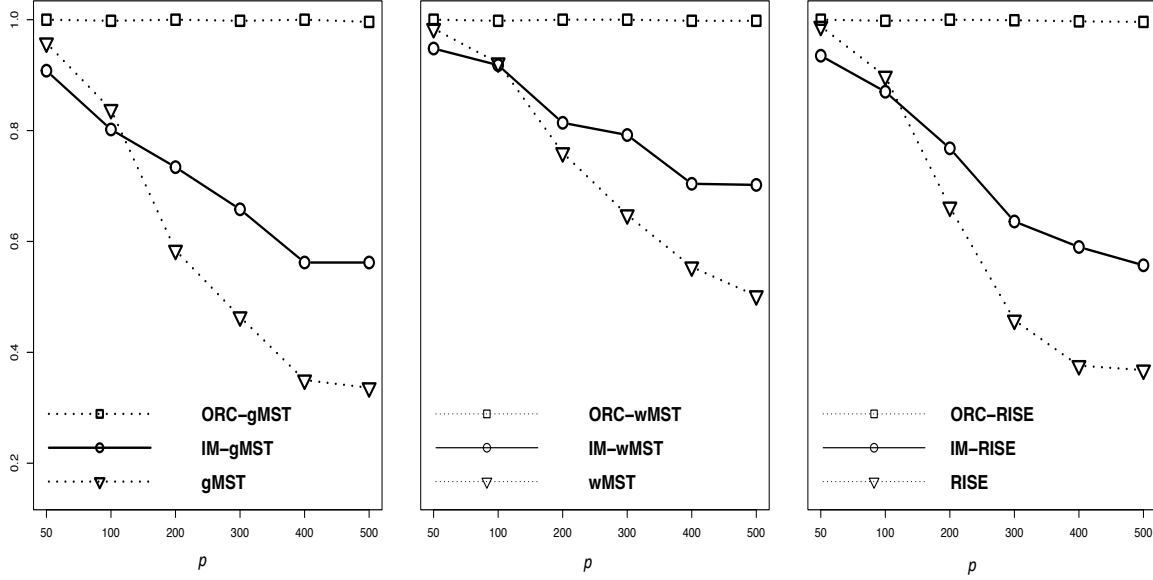


Figure 2.1: Power comparison under Model 1 ($n = m = 80$).

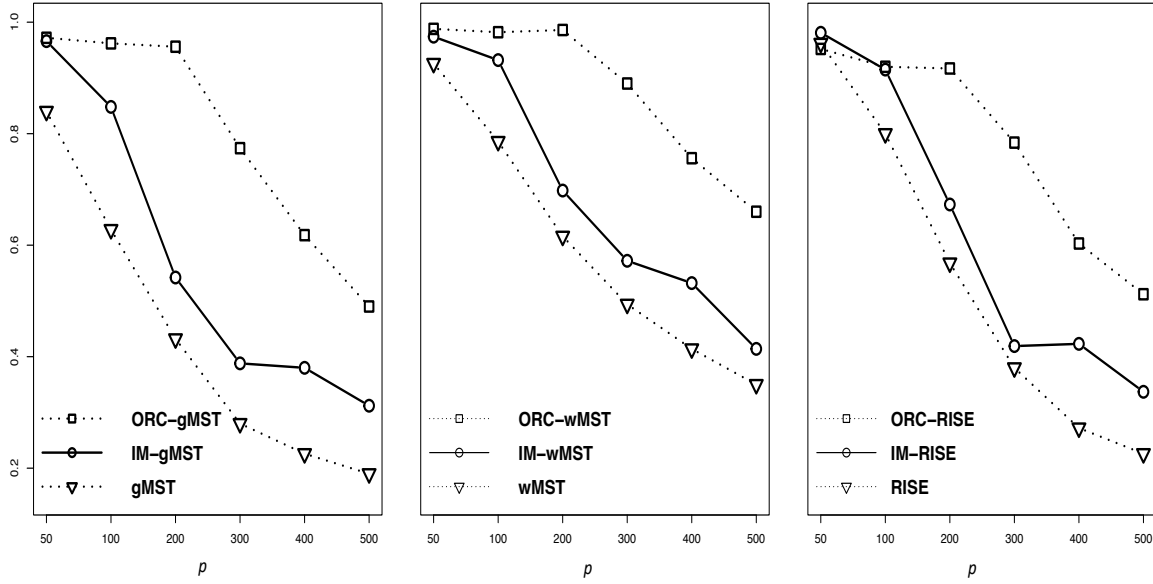


Figure 2.2: Power comparison under Model 1 ($n = 80, m = 160$).

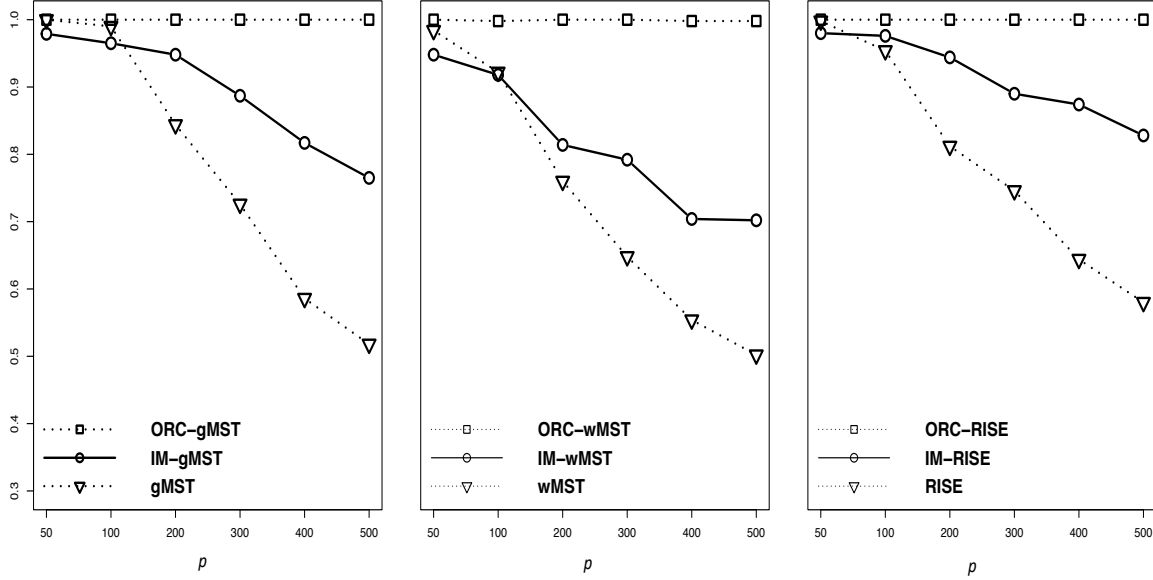


Figure 2.3: Power comparison under Model 2 ($n = m = 80$).

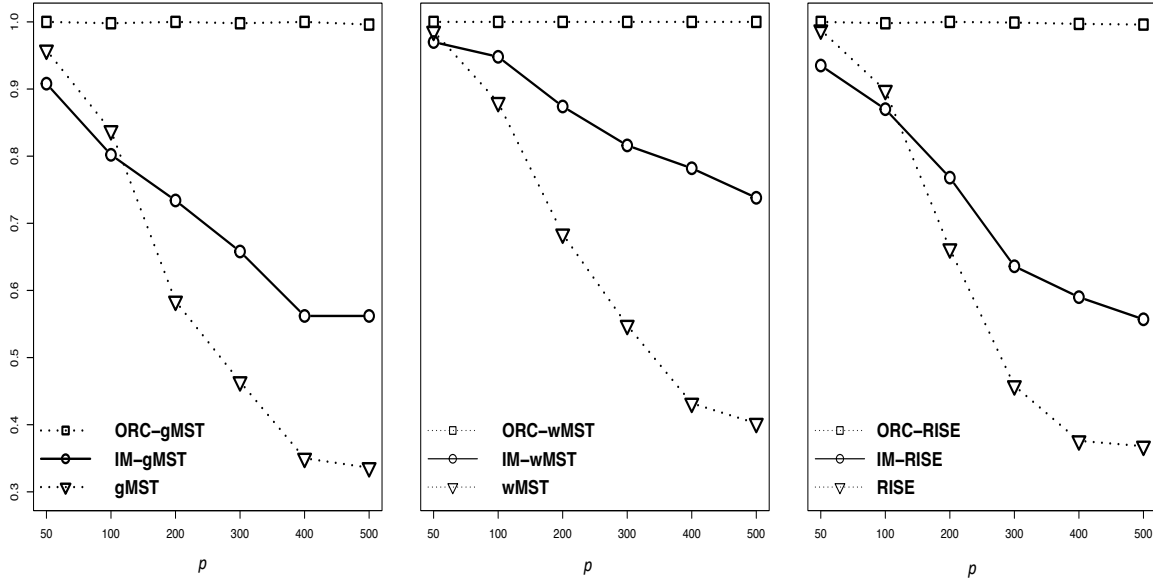


Figure 2.4: Power comparison under Model 2 ($n = 80, m = 160$).

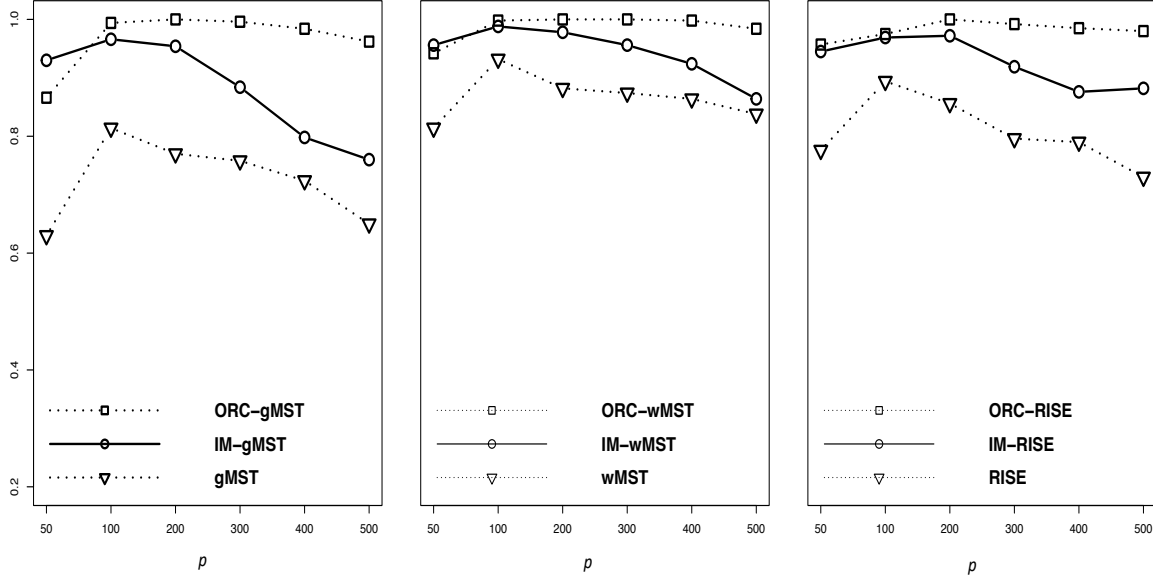


Figure 2.5: Power comparison under Model 3 ($n = m = 80$).

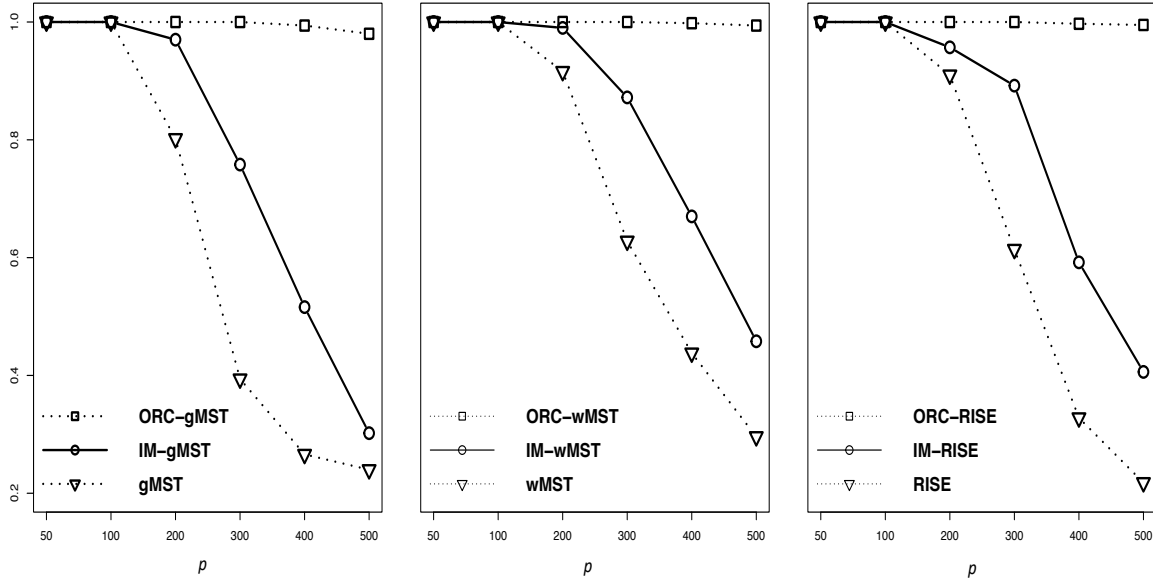


Figure 2.6: Power comparison under Model 3 ($n = 80, m = 160$).

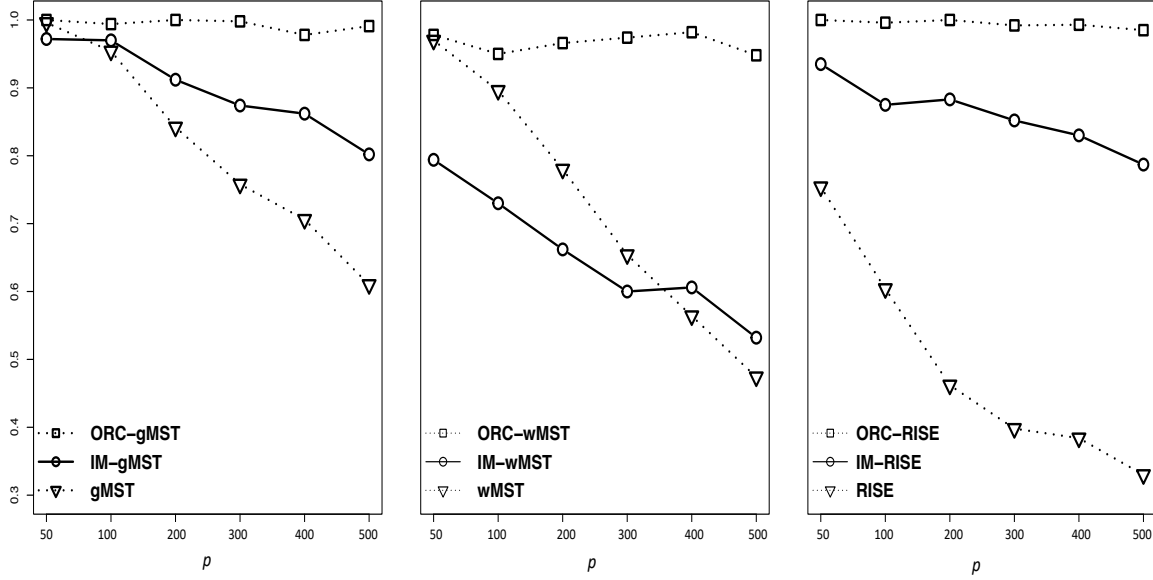


Figure 2.7: Power comparison under Model 4 ($n = m = 80$).

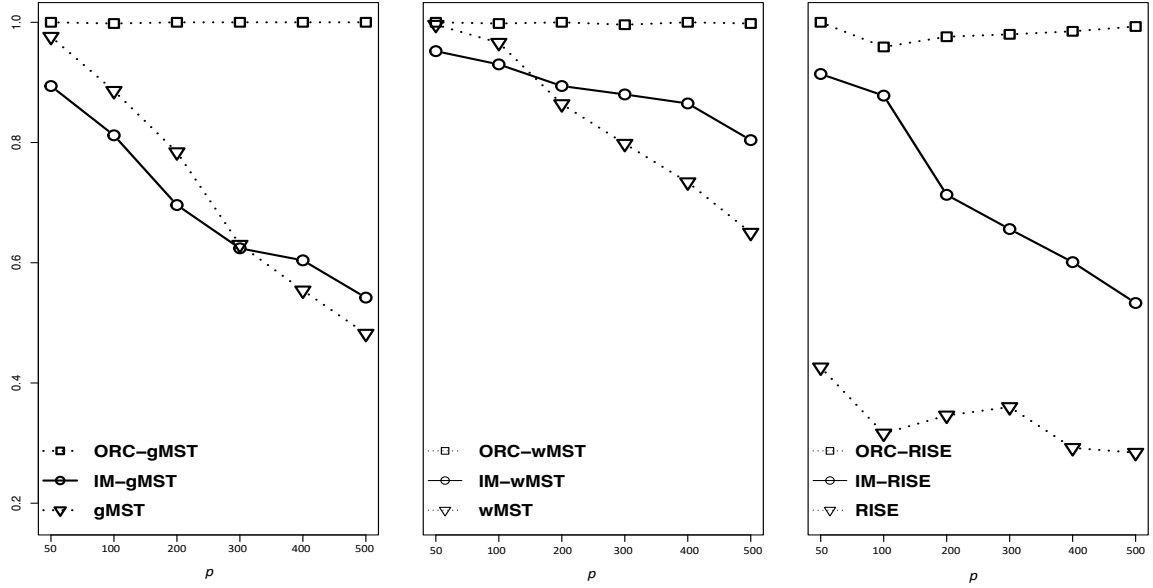


Figure 2.8: Power comparison under Model 4 ($n = 80, m = 160$).

2.7 Real data analysis

2.7.1 Single-cell RNA sequencing data

RNA sequencing quantifies gene expression levels by counting the number of sequenced reads mapped to specific genes. Single-cell RNA sequencing enables the measurement of gene expression at the cellular level, providing insights into the cellular heterogeneity of tissues. For a comprehensive review of single-cell RNA sequencing data, see [Mukherjee et al. \(2022\)](#). We analyzed single-cell transcriptomics data from patients with non-small cell lung cancer (NSCLC). The raw sequencing data are available under accession number GSE99254 in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo), one of the largest gene expression databases. Single-cell RNA sequencing was conducted on T cells isolated from **two distinct tissue types**: cancer tissue (**Tumor**) and tumor-adjacent normal tissue (**Normal**). T cells were further sorted by their surface markers into **four subtypes**: CD8+, CD4+, regulatory, and T helper cells. The number of cells sequenced for each subtype across the two tissue types is shown in Table 2.2. For this NSCLC data analysis, the number of genes per cell is 23,370.

Table 2.2: Numbers of cells sequenced for NSCLC data.

Tissue type	CD8+ T cells	CD4+ T cells	regulatory T cells	T helper cells
Normal	934	655	288	238
Tumor	2182	1591	1170	892

Diseases such as cancer may result from various combinations of perturbed gene activities. Grouping genes into functional sets, known as biological pathways, often provides valuable insights into the underlying biological systems, such as metabolic and signaling pathways ([Fey et al. 2015](#), [Salavert et al. 2016](#)). For each patient, identifying jointly perturbed gene activities within specific pathways could inform personalized diagnostic and treatment strategies. Thus, investigating the distributional differences in gene sets or biological pathways is a topic of important research interest.

We focused on 177 metabolic and signaling pathways described in the Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.genome.jp/kegg/>). A list of genes corresponding to each pathway was obtained, with genes assigned to 177 subsets, each representing a specific pathway. The number of genes per pathway ranges from 10 to 380. For each of the four T cell subtypes and each pathway, we performed IM-gMST to test for multivariate distributional differences in gene expression across the two tissue types.

- Among all 177 pathways, we failed to reject only 2 pathways in CD4⁺ regulatory T cells when comparing Normal and Tumor tissues.
- In contrast, the number of pathways that we failed to reject in CD8⁺ Cytotoxic cells, CD4⁺ Naive cells, and CD4⁺ Helper cells were 22, 50, and 74, respectively.

Based on the t-distributed stochastic neighbor embedding (tSNE) projection of single T cells by [Guo et al. \(2018\)](#), CD4⁺ regulatory T cells are visually more distinguishable from the other three subtypes. Our findings corroborate this observation, revealing notable differences in gene expression levels between CD4⁺ regulatory T cells and the other three subtypes.

Next, we focus on the three T cell subtypes (CD8⁺ cytotoxic cells, CD4⁺ naive cells, and CD4⁺ helper cells) and highlight overlapping pathways for which the null hypothesis of no distributional differences between Normal and Tumor tissue types was rejected. These results are illustrated in [Figure 2.9](#).

- These three T cell subtypes share 67 pathways for which the null hypothesis of no distributional differences was rejected. Notably, these 67 pathways include all signaling pathways previously identified as functionally related to NSCLC, such as the Ras, MAPK, and PI3K signaling pathways ([Brambilla & Gazdar 2009](#), [Santarpia et al. 2020](#)).
- The number of pathways uniquely rejected in only one of the three subtypes is relatively small. Specifically, only 5 pathways were unique to CD4⁺ naive cells, and another 5 were unique to CD4⁺ helper cells.

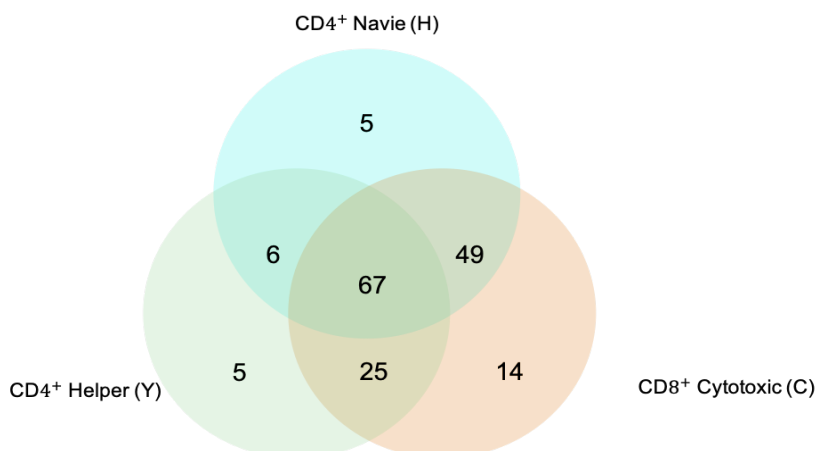


Figure 2.9: Pathways being declared to have distributional differences between Tumor and Normal tissue types.

2.7.2 Parkinson’s disease data

In recent years, Parkinson’s Disease (PD), a progressive neurodegenerative disorder, has become the second most common neurodegenerative disease among the elder individuals. This has sparked growing interest in developing telerdiagnosis and telemonitoring systems to detect PD in its early stages, offering a more convenient alternative to traditional physical clinical examinations. One of the most prominent early-stage symptoms of PD is vocal disorders or speech difficulties. As a result, vocal analysis has become a central focus in PD telemedicine research. It has been demonstrated that PD affects speech even at early stages, and speech features have been effectively used to assess PD and monitor its progression following surgical or pharmacological treatments. Various speech signal processing algorithms have been developed to extract meaningful information for PD analysis. Research has shown that the choice of feature extraction methods significantly impacts the accuracy of PD detection (Little et al. 2009, Sakar et al. 2013, 2019).

In this study, we examined whether sets of speech features exhibit distributional differences between PD patients and healthy individuals. The raw data is publicly available at <https://archive-beta.ics.uci.edu/dataset/470/parkinson+s+disease+classification>. The dataset consists of recordings from 188 PD patients and 64 healthy individuals. For each participant, sustained phonations of the vowel /a/ were recorded with 3 repetitions. Subsequently, 752 features were generated by processing these recordings through six speech signal processing algorithms or feature sets, as presented in Table 2.3.

Table 2.3: Number of features for PD data.

Feature set	Size
Baseline	21
Time Frequency	11
Vocal fold	22
Mel Frequency Cepstral Coefficients (MFCC)	84
Wavelet Transform Based (Wave)	182
Tunable Q-factor wavelet transform (TQWT)	432
Total	752

Since the sustained phonation for each individual were collected with 3 repetitions, we followed the convention of using their average for our analysis. Moreover, as suggested by [Sakar et al. \(2019\)](#), the feature vectors were standardized so that each feature has an overall zero mean and unit variance. We implemented the proposed IM-MSTs along with gMST, wMST, RISE, and IM-MRPP to compare their power. For IM screening, we selected $n_1 = 20$ out of $n = 64$ healthy individuals and $m_1 = 60$ out of $m = 188$ PD patients. The remaining samples were used for MST-based tests in the second step. To ensure a fair comparison, all MST-based tests were constructed with 5-MST, as discussed in Section 2.6.1. We report the asymptotic p -values in Table 2.4.

Table 2.4: p -values comparison for PD data.

	Baseline	Time	Vocal	MFCC	Wave	TQWT
gMST	4.18e-06	3.38e-09	1.30e-04	4.08e-14	4.15e-05	3.93e-09
wMST	3.16e-06	3.43e-10	1.17e-05	2.70e-14	5.09e-06	2.60e-09
RISE	1.29e-11	6.98e-16	1.13e-13	4.90e-36	4.97e-11	5.62e-14
IM-gMST	6.10e-04	3.67e-09	7.48e-02	4.89e-07	1.85e-07	9.40e-10
IM-wMST	3.77e-04	1.12e-08	4.08e-02	9.43e-08	1.75e-08	1.34e-09
IM-RISE	5.92e-08	8.54e-11	7.19e-06	3.19e-19	5.72e-12	4.71e-15
IM-MRPP	0.03	0.01	0.03	0.01	0.03	4.63e-02

- According to the p -values in Table 2.4, all six feature sets are declared to have distributional differences between PD patients and healthy individuals. This aligns

with the findings of [Sakar et al. \(2019\)](#), who concluded that each of these six feature sets could be used to achieve considerable classification accuracy when paired with methods such as Naive Bayes, Random Forest, and SVM.

- The p -values for all testing procedures are much smaller for MFCC and TQWT compared to the other four feature sets. Notably, these two feature sets also achieved the highest classification accuracy in the study by [Sakar et al. \(2019\)](#).
- Among all feature sets, RISE and IM-RISE produce the smallest p -values, while IM-MRPP yields the largest p -values.
- For Baseline (size 21), Time Frequency (size 11) and Vocal fold (size 22) feature sets, the p -values for gMST, wMST and RISE are smaller than those from their corresponding IM-MSTs.
- Conversely, for Wave (size 182) and TQWT (size 432) feature sets, IM-MSTs yield smaller p -values than the MST-based tests.
- In conclusion, for larger feature sets, IM-MSTs are preferred over MST-based tests. However, when the size of the feature set is relatively small, IM screening may be unnecessary, and MST-based tests perform well.

Chapter 3

High-dimensional quadratic discriminant analysis with joint estimation of two large covariance matrices

3.1 Introduction

Covariance matrix estimation plays a fundamental role in many multivariate statistical analyses, with applications in areas such as portfolio optimization, principal component analysis, and classification methods like quadratic discriminant analysis. In high-dimensional settings, where the number of variables p approaches or exceeds the number of observations n , classical methods such as the sample covariance matrix become unreliable due to singularity and the curse of dimensionality. To address these challenges, factor model assumptions have been extensively explored as a robust and efficient framework for high-dimensional covariance matrix estimation.

3.1.1 Factor model-based high-dimensional covariance matrix estimation

Specifically, factor models decompose the observable multivariate vector $\mathbf{y}$ as follows,

$$\mathbf{y} = B\mathbf{f} + \boldsymbol{\epsilon},$$

where

- $\mathbf{y}$ is the p -dimensional observable vector,
- $\mathbf{f}$ is the r -dimensional factor vector,
- $\boldsymbol{\epsilon}$ is the p -dimensional error vector, uncorrelated with $\mathbf{f}$,
- B is the $p \times r$ -dimensional loading matrix.

Assuming $\text{cov}(\mathbf{f}) = I_r$, where I_r is the $r \times r$ identity matrix, the factor model implies that the covariance matrix of $\mathbf{y}$ is:

$$\Sigma = BB^T + D, \quad (3.1)$$

where $\text{cov}(\boldsymbol{\epsilon}) = D$, typically assumed to be sparse or diagonal.

Fan et al. (2008) considered the scenario where the factors $\mathbf{f}$ are observable. They proposed estimating the loading matrix B using regression coefficients and approximating the error term covariance matrix D as a diagonal matrix of squared residuals. Fan et al. (2011) extended this approach to cases where the error term covariance matrix D may exhibit weak correlations. They modeled D as a sparse matrix and proposed an adaptive thresholding method for its estimation, using a predetermined threshold to enforce sparsity. Fan et al. (2013) addressed the more general scenario where the factors are latent and unobservable but the number of latent factors r is known. In this setting, the loading matrix B is estimated from the first r principal orthogonal components of the sample covariance matrix, while the error term covariance matrix D is estimated by thresholding the residual components of the sample covariance matrix.

In practice, the number of latent factors is typically unknown, making its determination a crucial step in facilitating covariance matrix estimation. Methods for estimating the number of factors often rely on the eigenvalues of the loading matrix or the population correlation matrix. Recently, Fan et al. (2022) demonstrated that the number of eigenvalues greater than 1 of the population correlation matrix is the same as the number of factors under mild conditions. They proposed a consistent, tuning-free method called Adjusted Correlation Thresholding (ACT) to estimate the number of factors. The ACT procedure begins with consistent estimators of the eigenvalues of the population correlation matrix, as introduced by Bai & Ding (2012). Then, ACT determines the number of factors by counting the estimated eigenvalues that are greater than $1 + \sqrt{p/(n-1)}$.

Wu et al. (2020) assumed a diagonal and low-rank decomposition of the covariance matrix and its inverse (the precision matrix). This assumption aligns with the factor

model decomposition in Equation (3.1), where D is a diagonal matrix. They emphasized estimating the precision matrix directly, as it can be applied to downstream tasks such as quadratic discriminant analysis. Assuming $\mathbf{y}$ follows a multivariate normal distribution, they proposed to estimate the precision matrix using a penalized likelihood approach. Building on this work, Wu (2017) extended the diagonal and low-rank decomposition framework to jointly estimate the covariance matrices of multiple groups. Specifically, the low-rank term of the precision matrix was further decomposed into a shared low-rank matrix and group-specific low-rank matrices. This extension allows for capturing both shared and unique structures across groups. We will discuss this assumption and their estimation approach in more detail later.

3.1.2 High-dimensional Quadratic Discriminant Analysis (QDA)

Quadratic Discriminant Analysis (QDA) is more flexible than Linear Discriminant Analysis (LDA), particularly in high-dimensional settings where testing for equality of covariance or precision matrices is challenging. Data-driven QDA involves estimating the mean and covariance or precision matrices for two groups and substituting these estimates into the QDA classification rule. The primary challenge lies in the complexity of estimating large covariance or precision matrices. To address this, many high-dimensional QDA approaches impose structural assumptions on these matrices to enhance estimation accuracy, classification performance, and computational feasibility.

Li & Shao (2015) emphasized the importance of structural assumptions for high-dimensional QDA and introduced a sparse QDA rule. This approach was the first to separately impose sparsity assumptions on covariance matrices, mean differences, and covariance differences. Later, Jiang et al. (2018) proposed a direct estimation method, assuming sparsity in mean differences and precision matrix differences, and developed a consistent QDA rule. However, they did not establish the convergence rate for the classification error. Cai & Zhang (2021) introduced sparsity assumptions on the difference between precision matrices and the projected difference between mean vectors, directly estimating these key quantities. They proposed the Sparse Discriminant Analysis by Regression method and provided explicit convergence rates for the classification error. Beyond sparse QDA, alternative structural assumptions have been explored to facilitate high-dimensional QDA. For instance, Wu et al. (2019) proposed ppQDA, which pools information about correlations among different dimensions alongside the diagonal elements of covariance matrices.

In addition to methods that rely on structural assumptions for covariance (or precision) matrix estimation, other approaches involve dimension reduction before applying QDA.

For instance, [Gaynanova & Wang \(2019\)](#) proposed the Discriminant Analysis via Projection (DAP) method, which incorporates QDA with linear dimension reduction. The DAP method simultaneously estimates two projection directions, $\Sigma_1(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ and $\Sigma_2(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$, and performs QDA after projecting the data onto these directions. [Wu & Hao \(2022\)](#) extended this concept by studying the optimal direction in which QDA achieves the least classification error. They proposed a method called QDAP (QDA by Projection). However, the optimal direction lacks a closed-form expression in general cases, and their approximation is valid only when $p < n < p^2$, limiting its applicability in higher-dimensional settings. Empirically, their method performs comparably to LDA for homoscedastic data and exhibits greater sensitivity to heteroscedastic data than DAP.

3.1.3 Contributions and outline

The contributions of this chapter are two-fold. Firstly, we extend the ACT method to estimate the number of common factors under a two-group factor model, and we derive the consistency properties of the estimator. Since the number of common factors coincides with the rank of the loading matrix, we use the estimator to pre-specify the ranks and jointly estimate high-dimensional precision matrices under the “partially common diagonal + low-rank” (pcDL) decomposition, as introduced by [Wu \(2017\)](#). Secondly, based on the pcDL decomposition, we use the joint estimators of the two precision matrices to implement a data-driven QDA classifier and establish an explicit convergence rate for the classification error. While [Cai & Zhang \(2021\)](#) also provided a convergence rate for their sparse QDA method, their approach relies on assuming sparsity in the projected difference of mean vectors and the difference between precision matrices. In contrast, our method assumes low-rank structures in the pcDL decomposition.

The rest of this chapter is organized as follows. In [Section 3.2](#), we introduce the notations and discuss the partially common diagonal + low-rank (pcDL) decomposition proposed by [Wu \(2017\)](#). [Section 3.3](#) extends the ACT method to estimate the number of common factors for two groups, with the consistency of the estimators established. In [Section 3.4](#), we examine the joint matrix estimation problem where the ranks are determined using the extended ACT method. [Section 3.5](#) introduces the resulting data-driven QDA rule and derives the convergence rate for the classification error. Simulation studies are presented in [Section 3.6](#), followed by real data analysis in [Section 3.7](#). All proofs are provided in [Section 3.8](#).

3.2 Problem set-up

We begin by introducing some notations. Let $\mathbf{R}^{p_1 \times p_2}$ denote the set of $p_1 \times p_2$ matrices, and $\mathbf{R}^p$ denote the set of p -dimensional vectors. The set of symmetric matrices in $\mathbf{R}^{p \times p}$ is denoted by $\mathbf{S}^p$, while $\mathbf{S}_+^p$ and $\mathbf{S}_{++}^p$ represent the sets of positive semi-definite and strictly positive definite matrices, respectively. Similarly, $\mathbf{D}^p$, $\mathbf{D}_+^p$, and $\mathbf{D}_{++}^p$ denote the sets of diagonal matrices with real, non-negative, and positive diagonal elements, respectively.

For $A \in \mathbf{S}^p$, $\text{tr}(A)$ denotes its trace, $|A|$ denotes its determinant, and $\lambda_1(A) \geq \dots \geq \lambda_p(A)$ denote its eigenvalues. Regarding matrix norms, $\|A\|_F = \{\text{tr}(A^T A)\}^{1/2}$ is the Frobenius norm, and $\|A\|_{op} = \{\lambda_1(AA^T)\}^{1/2}$ is the operator norm. For a vector a , $\|a\|_2$ denotes the ℓ_2 norm. Finally, I_p represents the $p \times p$ identity matrix, and the function $\text{diag}(\cdot)$ either converts a vector into a diagonal matrix or extracts the diagonal elements of a matrix, depending on the input. The indicator function is denoted by $\mathbb{1}\{\cdot\}$.

3.2.1 “Partially common diagonal + low-rank” - pcDL

Consider a dataset $Y = \{Y^{(1)}, Y^{(2)}\}$, where $Y^{(k)} = \{\mathbf{y}_1^{(k)}, \dots, \mathbf{y}_{n_k}^{(k)}\}$ for $k = 1, 2$, consists of n_k random vectors drawn independently and identically from a p -variate distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_0^{(k)}$. Let $n = n_1 + n_2$ be the total sample size. To facilitate estimation, we assume that each $\Sigma_0^{(k)}$ follows the decomposition:

$$\Sigma_0^{(k)} = D_{\Sigma_0} + L_{\Sigma_0} + L_{\Sigma_0}^{(k)}, \quad k = 1, 2, \quad (3.2)$$

where $D_{\Sigma_0} \in \mathbf{D}_{++}^p$, $L_{\Sigma_0} \in \mathbf{S}_+^p$ with $\text{rank}(L_{\Sigma_0}) = r_0$, and $L_{\Sigma_0}^{(k)} \in \mathbf{S}_+^p$ with $\text{rank}(L_{\Sigma_0}^{(k)}) = (v_{0k} - r_0)$. For brevity, we denote the vector of ranks as $v_0 = (v_{01}, v_{02})$. In this decomposition, D_{Σ_0} and L_{Σ_0} are shared or common across groups, while $L_{\Sigma_0}^{(k)}$ is group-specific. This decomposition is referred to as the “partially common diagonal + low-rank” (pcDL) assumption. Note that this decomposition assumption was first introduced by [Wu \(2017\)](#).

Similar to the prevalent sparsity assumption, this decomposition reduces the number of parameters to be estimated. The rank (i.e., the number of non-zero eigenvalues) of L_{Σ_0} and $L_{\Sigma_0}^{(k)}$ parallels the number of non-zero off-diagonal elements in Σ_0 under the sparsity assumption.

The decomposition (3.2) can also be motivated by a two-group latent factor model. Specifically,

$$\mathbf{y}_i^{(k)} = B\mathbf{f}_i^{(k)} + B^{(k)}\mathbf{g}_i^{(k)} + \boldsymbol{\epsilon}_i^{(k)}, \quad k = 1, 2, \quad i = 1, \dots, n_k, \quad (3.3)$$

where

- $B \in \mathbf{R}^{p \times r_0}$ is the common loading matrix,
- $B^{(k)} \in \mathbf{R}^{p \times (v_{0k} - r_0)}$ is the group-specific loading matrix,
- $\mathbb{E}(\mathbf{f}_i^{(k)}) = \mathbf{0}, \text{cov}(\mathbf{f}_i^{(k)}) = I_{r_0},$
- $\mathbb{E}(\mathbf{g}_i^{(k)}) = \mathbf{0}, \text{cov}(\mathbf{g}_i^{(k)}) = I_{v_{0k} - r_0},$
- $\boldsymbol{\epsilon}_i^{(k)} \in \mathbf{R}^p$ is an error vector with $\mathbb{E}(\boldsymbol{\epsilon}_i^{(k)}) = \mathbf{0}, \text{cov}(\boldsymbol{\epsilon}_i^{(k)}) = D$, where $D \in \mathbf{D}_{++}^p$.
- Additionally, $\mathbf{f}_i^{(k)}, \mathbf{g}_i^{(k)}$, and $\boldsymbol{\epsilon}_i^{(k)}$ are assumed to be independent.

The two-group latent factor model (3.3) implies that the covariance matrix of $\mathbf{y}_i^{(k)}$ adheres to the pcDL decomposition (3.2). This model can also be viewed as an extension of the framework proposed by Fan et al. (2022), which focuses on data in a single group. Therefore, the method developed by Fan et al. (2022) for estimating the number of latent factors (or equivalently, the rank of the corresponding loading matrix) can be adopted and modified to facilitate the estimation of precision or covariance matrices in our setting.

3.3 Adjusted Correlation Thresholding (ACT)

Using the adjusted correlation thresholding (ACT) method by Fan et al. (2022), applied separately to each group, we can estimate v_{01} and v_{02} . Additionally, by applying ACT to the pooled sample, we develop a new estimator for the total number of factors, v_{12} , which corresponds to the rank of the pooled loading matrix (B_{pool}) as defined below. Finally, using the relationship $r_0 = v_{01} + v_{02} - v_{12}$, we can determine the number of common factors, r_0 .

The latent factor model (3.3) implies the pooled covariance matrix is

$$\begin{aligned} \Sigma_0 &:= \gamma \Sigma_0^{(1)} + (1 - \gamma) \Sigma_0^{(2)} = BB^T + \gamma B^{(1)} B^{(1)T} + (1 - \gamma) B^{(2)} B^{(2)T} + D \\ &:= B_{pool} B_{pool}^T + D, \end{aligned}$$

where $n_1/n \rightarrow \gamma \in (0, 1)$, and

$$B_{pool} = \left(B, \sqrt{\gamma} B^{(1)}, \sqrt{1 - \gamma} B^{(2)} \right)$$

and the corresponding pooled correlation matrix is

$$R = [\text{diag}(\Sigma_0)]^{-1/2} \Sigma_0 [\text{diag}(\Sigma_0)]^{-1/2}.$$

We estimate v_{12} under the following assumptions.

- (A.1) The eigenvalues of Σ_0 are uniformly bounded both from below and above.
- (A.2) The overall loading matrix B_{pool} has full column rank, i.e., $\text{rank}(B_{pool}) = v_{12}$, where $v_{12} = o(p^{1/6})$.
- (A.3) Denote

$$X^{(1)} = \begin{pmatrix} \mathbf{f}_1^{(1)} & \dots & \mathbf{f}_{n_1}^{(1)} \\ \mathbf{g}_1^{(1)} & \dots & \mathbf{g}_{n_1}^{(1)} \\ \boldsymbol{\epsilon}_1^{(1)} & \dots & \boldsymbol{\epsilon}_{n_1}^{(1)} \end{pmatrix}_{(p+v_{01}) \times n_1} \quad \text{and} \quad X^{(2)} = \begin{pmatrix} \mathbf{f}_1^{(2)} & \dots & \mathbf{f}_{n_2}^{(2)} \\ \mathbf{g}_1^{(2)} & \dots & \mathbf{g}_{n_2}^{(2)} \\ \boldsymbol{\epsilon}_1^{(2)} & \dots & \boldsymbol{\epsilon}_{n_2}^{(2)} \end{pmatrix}_{(p+v_{02}) \times n_2}$$

where the entries $\{x_{ji}^{(k)}\}$ are independent random variables and for some constant $\delta_0 > 0$,

$$\sup_{j,k} \mathbb{E} \left| x_{j1}^{(k)} \right|^{6+\delta_0} < \infty.$$

- (A.4) R has v_{12} distinct spiked eigenvalues, and the population spectral distribution (PSD) of R , i.e.,

$$H_{p-v_{12}} := \frac{1}{p-v_{12}} \sum_{j=v_{12}+1}^p \delta_{\lambda_j(R)}$$

converges weakly to a proper probability distribution, denoted as H , where δ_a is the Dirac measure at a .

- (A.5) $p/n \rightarrow \rho \in (0, \infty)$ and $n_1/n \rightarrow \gamma \in (0, 1)$ as $p, n \rightarrow \infty$.

These assumptions, similar to those in [Fan et al. \(2022\)](#), are standard in random matrix theory.

Under assumptions (A.1) and (A.2), the conditions of Theorem 1 in [Fan et al. \(2022\)](#) are satisfied, leading to the following proposition.

Proposition 3.3.1. *For the latent factor model (3.3) satisfying assumptions (A.1) and (A.2), the following holds:*

$$\lambda_j(R) \leq 1, \quad j = v_{12} + 1, \dots, p, \quad \text{and } v_{12} = \max \{j : \lambda_j(R) > 1\}.$$

Proposition 3.3.1 establishes the relationship between v_{12} and the number of eigenvalues of the pooled correlation matrix that are greater than 1. This result provides a practical approach for estimating v_{12} by examining the eigenvalues of the correlation matrix derived from the pooled sample.

However, to achieve a consistent estimator of v_{12} , a minimal signal strength requirement must be satisfied. Let $\mathcal{F}(\nu_0)$ denote the set of correlation matrices where the v_{12} largest eigenvalues exceed ν_0 , i.e., $\mathcal{F}(\nu_0) = \{R : \lambda_{v_{12}}(R) > \nu_0\}$. Theorem 4 in Fan et al. (2022) shows that under the latent factor model (3.3), which satisfies assumptions (A.1)-(A.5), for any $R \in \mathcal{F}(\nu_0)$, if $\nu_0 < 1 + \sqrt{\rho}$, no method based on the sample eigenvalues of the correlation matrix can provide a consistent estimate of v_{12} . Consequently, $\nu_0 = 1 + \sqrt{\rho}$ represents the minimal signal strength required. For the remainder of this chapter, we consider the estimation of v_{12} under the assumption that $R \in \mathcal{F}(1 + \sqrt{\rho})$.

Next, we analyze the spectral properties and spiked eigenvalues of the pooled sample correlation matrix, $\hat{R}$, given by

$$\hat{R} = [\text{diag}(\hat{\Sigma})]^{-1/2} \hat{\Sigma} [\text{diag}(\hat{\Sigma})]^{-1/2},$$

where

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n_1} \left(\mathbf{y}_i^{(1)} - \bar{\mathbf{y}}^{(1)} \right) \left(\mathbf{y}_i^{(1)} - \bar{\mathbf{y}}^{(1)} \right)^T + \frac{1}{n} \sum_{j=1}^{n_2} \left(\mathbf{y}_j^{(2)} - \bar{\mathbf{y}}^{(2)} \right) \left(\mathbf{y}_j^{(2)} - \bar{\mathbf{y}}^{(2)} \right)^T,$$

$\bar{\mathbf{y}}^{(1)}$ and $\bar{\mathbf{y}}^{(2)}$ are the sample means of the two groups.

The spectral properties of $\hat{R}$ can be established by studying the corresponding empirical spectral distribution (ESD), F_n , through its Stieltjes transform, $m_n(z)$, where

$$F_n = \frac{1}{p - v_{12}} \sum_{j=v_{12}+1}^p \delta_{\lambda_j(\hat{R})} \quad \text{and} \quad m_n(z) = \int \frac{1}{t - z} dF_n(t), \quad z \in \mathcal{C}^+.$$

Furthermore, the companion versions of F_n and $m_n(z)$ are defined by

$$\underline{F}_n = (1 - \rho^*) \delta_0 + \rho^* F_n \quad \text{and} \quad \underline{m}_n(z) = \int \frac{1}{t - z} d\underline{F}_n(t),$$

where $\rho^* = (p - v_{12})/(n - 1)$. The following results provide the convergence of $\underline{F}_n$ and $\underline{m}_n(z)$.

Theorem 3.3.2. *Suppose that the assumptions (A.1)–(A.5) hold. Then, almost surely, the companion ESD $\underline{F}_n$ converges to a deterministic distribution $\underline{F}$, and the Stieltjes transform of $\underline{F}$ is the unique solution to*

$$z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1 + t\underline{m}(z)},$$

on the set $\{z : z \in \mathcal{C}^+, \underline{m}(z) \in \mathcal{C}^+\}$.

Theorem 3.3.2 establishes the convergence of $\underline{F}_n$ and implies the almost sure convergence of the Stieltjes transforms $\underline{m}_n(z)$, as follows:

$$\underline{F}_n(t) \xrightarrow{a.s.} \underline{F}(t), \quad \underline{m}_n(z) \xrightarrow{a.s.} \underline{m}(z) = \int \frac{1}{t - z} d\underline{F}(t), \quad \forall z \in \mathcal{C}^+.$$

This theorem extends Theorem 2 in Fan et al. (2022) to the two-group setting, and its proof is provided in Section 3.8.

Next, we discuss the estimation of the spiked eigenvalues of the pooled correlation matrix R . For simplicity, let λ_j and $\hat{\lambda}_j$ denote the j th largest eigenvalues of R and $\hat{R}$, respectively. Our corrected estimate of these eigenvalues, λ_j , is defined as follows:

$$\hat{\lambda}_j^C = -\frac{1}{\underline{m}_{n,j}(\hat{\lambda}_j)}, \quad j = 1, \dots, v_{12},$$

where

$$\underline{m}_{n,j}(z) = -\left(1 - \frac{p-j}{n-1}\right) \frac{1}{z} + \frac{1}{n-1} \sum_{\ell=j+1}^p \frac{1}{\hat{\lambda}_\ell - z}.$$

Proposition 3.3.3. *Under assumptions (A.1)–(A.5), if $\lambda_j \geq \lambda_{v_{12}+1}(R)(1 + \sqrt{\rho}) + \delta$ for some $\delta > 0$ and any $j \leq v_{12}$, we have*

$$\hat{\lambda}_j^C = \lambda_j + o_p(1) \quad \text{and} \quad \hat{\lambda}_j = \lambda_j + \rho \int \frac{\lambda_j t}{\lambda_j - t} dH(t) + o_p(1).$$

Proposition 3.3.3 shows that, in general, $\hat{\lambda}_j$ is not a consistent estimator of λ_j . This proposition can be established based on Theorem 3.3.2 and the consistency results discussed in Bai & Ding (2012) and Fan et al. (2022). Due to the similarity, the proof is omitted.

We propose the following estimate for v_{12} :

$$\widehat{v}_{12}(s) = \max\{j : \widehat{\lambda}_j^C > s\},$$

where the maximum of the empty set is defined as 0. We will demonstrate in Theorem 3.3.4 that $s = 1 + \sqrt{\rho}$ is the optimal choice. In addition, we can estimate v_{01} and v_{02} by directly applying the ACT method to each group separately:

$$\begin{aligned}\widehat{v}_{01}(s_1) &= \max\{i : \widehat{\lambda}_i^C(R_1) > s_1\}, \\ \widehat{v}_{02}(s_2) &= \max\{k : \widehat{\lambda}_k^C(R_2) > s_2\},\end{aligned}$$

where $\widehat{\lambda}_i^C(R_1)$ and $\widehat{\lambda}_k^C(R_2)$ are the corrected estimates of the spiked eigenvalues of the correlation matrices R_1 and R_2 , respectively. Therefore, the number r_0 can be estimated as:

$$\widehat{r}_0 = \widehat{v}_{01} + \widehat{v}_{02} - \widehat{v}_{12}.$$

Theorem 3.3.4. *Under assumptions (A.1)–(A.5), and suppose there exist three small positive constants δ_1 , δ_2 and δ such that $s_1 = 1 + \sqrt{\gamma\rho} + \delta_1 < \lambda_{v_{01}}(R_1)$, $s_2 = 1 + \sqrt{(1-\gamma)\rho} + \delta_2 < \lambda_{v_{02}}(R_2)$, and $s = 1 + \sqrt{\rho} + \delta < \lambda_{v_{12}}$. Then, the following results hold:*

$$\mathbb{P}(\widehat{v}_{01}(s_1) = v_{01}) \rightarrow 1, \quad \mathbb{P}(\widehat{v}_{02}(s_2) = v_{02}) \rightarrow 1, \quad \mathbb{P}(\widehat{v}_{12}(s) = v_{12}) \rightarrow 1, \quad \mathbb{P}(\widehat{r}_0 = r_0) \rightarrow 1.$$

The consistency of $\widehat{v}_{01}(s_1)$ and $\widehat{v}_{02}(s_2)$ was established in Fan et al. (2022). In Section 3.8, we prove the consistency of the remaining two terms using a similar technique.

3.4 Precision matrix estimation using ACT

Let us now focus on estimating precision matrices, which are more intuitive for optimization when maximizing the likelihood function and serve as a more “direct” component in QDA classifiers. For the remainder of this chapter, we assume $\mathbf{y}$ follows a multivariate normal distribution. Let $\Theta_0^{(k)} = (\Sigma_0^{(k)})^{-1}$ denote the precision matrix of group k and let $\Theta_0 = (\Theta_0^{(1)}, \Theta_0^{(2)})$ for simplicity. As discussed in Wu (2017) and based on the results of Henderson & Searle (1981), it follows that under (3.2), a similar pcDL decomposition applies to precision matrix as well,

$$\Theta_0^{(k)} \triangleq -L_0^{(k)} - L_0 + D_0. \tag{3.4}$$

Here, $D_0 \in \mathbf{D}_{++}^p$, $L_0 \in \mathbf{S}_+^p$, with $\text{rank}(L_0) = r_0$, and $L_0^{(k)} \in \mathbf{S}_+^p$, with $\text{rank}(L_0^{(k)}) = (v_{0k} - r_0)$. Wu (2017) then proposed that precision matrix estimators can be obtained by solving

$$\begin{aligned} \min_{\Theta} \quad & \sum_{k=1}^2 n_k \left\{ \text{tr}(\Theta^{(k)} \widehat{\Sigma}^{(k)}) - \log |\Theta^{(k)}| \right\}, \\ \text{subject to} \quad & \Theta^{(k)} = D - L - L^{(k)}, \\ & \Theta^{(k)} \in \mathbf{S}_+^p, \quad D \in \mathbf{D}_{++}^p, \\ & L \in \mathbf{S}_+^p, \quad \text{rank}(L) = r, \\ & L^{(k)} \in \mathbf{S}_+^p, \quad \text{rank}(L^{(k)}) = v_k - r, \end{aligned} \quad (3.5)$$

where r and $(v_k - r)$ are pre-selected ranks, as the true ranks r_0 and $(v_{0k} - r_0)$ are unknown; $\widehat{\Sigma}^{(k)}$ denotes the sample covariance matrix for group k , and the objective function corresponds to the negative log-likelihood of two independent multivariate normal samples. We denote the optimization variable as $\Theta = (\Theta^{(1)}, \Theta^{(2)})$ and the solution to (3.5) as $\widehat{\Theta}_{r,v} = (\widehat{\Theta}_{r,v}^{(1)}, \widehat{\Theta}_{r,v}^{(2)})$, where $v = (v_1, v_2)$.

Denote the search space of (3.5) as

$$\begin{aligned} \mathbf{F}_{r,v} = \quad & \{ \Theta = (\Theta^{(1)}, \Theta^{(2)}) \mid \Theta^{(k)} = D - L - L^{(k)}, \quad \Theta^{(k)} \in \mathbf{S}_+^p, \quad D \in \mathbf{D}_{++}^p, \\ & L \in \mathbf{S}_+^p, \quad \text{rank}(L) = r, \quad L^{(k)} \in \mathbf{S}_+^p, \quad \text{rank}(L^{(k)}) = v_k - r \}. \end{aligned}$$

The accuracy of $\widehat{\Theta}_{r,v}$ depends on (1) the estimation error, determined by the distance between $\widehat{\Theta}_{r,v}$ and Θ , and (2) the systematic error, determined by the distance between $\mathbf{F}_{r,v}$ and Θ_0 . Specifically, we define:

$$d_{r,v} = \min_{\Theta \in \mathbf{F}_{r,v}} \sum_{k=1}^2 \|\Theta^{(k)} - \Theta_0^{(k)}\|_F.$$

We restate Corollary 4.1 from Wu (2017) and present the consistency properties of $\widehat{\Theta}_{r,v}$ in the following proposition.

Proposition 3.4.1. (Corollary 4.1 in Wu (2017)) Under the pcDL decomposition (3.4), assuming (A1) and (A5) hold, and the pre-specified ranks satisfy $\max_k v_k = O\{\max_k v_{0k}, \log p\}$.

Then, the solution to (3.5), $\widehat{\Theta}_{r,v}$, satisfies the following the property:

$$\sum_{k=1}^2 \|\widehat{\Theta}_{r,v}^{(k)} - \Theta_0^{(k)}\|_F = O_p \{ \max(a_{n,p,v_0}, b_{n,p}, d_{r,v}) \},$$

in which

$$a_{n,p,v_0} = (\max_k v_{0k})^{1/2} (p/n)^{1/2}, \quad b_{n,p} = \{(p \log p)/n\}^{1/2}.$$

The convergence rate aligns with intuition, where a_{n,p,v_0} and $b_{n,p}$ represent the estimation error, and $d_{r,v}$ captures the systematic error. When the pre-specified ranks are greater than $\max \left(\max_k v_{0k}, \log p \right)$, Θ_0 is contained within $\mathbf{F}_{r,v}$, resulting in $d_{r,v} = 0$. Proposition 3.4.1 suggests that if these pre-specified ranks are not excessively large, the convergence rate behaves as though the ranks were chosen correctly. Conversely, if the ranks are chosen too small, Θ_0 is not contained in $\mathbf{F}_{r,v}$, leading to $d_{r,v} > 0$.

In the following, we adopt the ACT method introduced in Section 3.3. According to Theorem 3.3.4,

$$\hat{v}_{01} = v_{01} + o_p(1), \quad \hat{v}_{02} = v_{02} + o_p(1), \quad \text{and} \quad \hat{r}_0 = r_0 + o_p(1).$$

Thus, these estimated ranks do not exceed the true ranks in terms of order. Let $\hat{\Theta} = (\hat{\Theta}^{(1)}, \hat{\Theta}^{(2)})$ denote the solution to (3.5), with $\hat{r}_0$, $\hat{v}_{01}$, and $\hat{v}_{02}$ used as the pre-specified ranks. Consequently, we obtain the following results regarding the estimation of precision matrices.

Theorem 3.4.2. *Under the pcDL decomposition (3.4) and assuming (A.1)–(A.5) hold, then $\hat{\Theta}$ satisfies the following property:*

$$\sum_{k=1}^2 \|\hat{\Theta}^{(k)} - \Theta_0^{(k)}\|_F = O_p \{ \max(a_{n,p,v_0}, b_{n,p}, d_{r,v}) \},$$

in which

$$a_{n,p,v_0} = (\max_k v_{0k})^{1/2} (p/n)^{1/2}, \quad b_{n,p} = \{(p \log p)/n\}^{1/2}.$$

Wu (2017) instead proposed an approach that incorporates an AIC penalty with a tuning parameter to prevent ill-specified ranks. In contrast, our method is tuning-free and directly estimates the ranks, providing greater computational efficiency. A performance comparison of rank recovery between these two approaches is presented in Section 3.6.1.

3.5 An application to Quadratic Discriminant Analysis (QDA)

We now apply the pcDL-based (in conjunction with ACT) joint estimation of precision matrices to perform QDA. Recall that we have random samples from two classes, where n_1 samples belong to class 1 and n_2 samples belong to class 2:

$$\text{Class 1: } Y^{(1)} = \{\mathbf{y}_1^{(1)}, \dots, \mathbf{y}_{n_1}^{(1)}\} \stackrel{i.i.d.}{\sim} N_p(\boldsymbol{\mu}^{(1)}, (\Theta_0^{(1)})^{-1}).$$

$$\text{Class 2: } Y^{(2)} = \{\mathbf{y}_1^{(2)}, \dots, \mathbf{y}_{n_2}^{(2)}\} \stackrel{i.i.d.}{\sim} N_p(\boldsymbol{\mu}^{(2)}, (\Theta_0^{(2)})^{-1}).$$

A new data point $\mathbf{z}$ is then assumed to be drawn from one of the two classes with prior probabilities π_1 and π_2 , respectively. We denote the parameter set as

$$\boldsymbol{\theta} = (\pi_1, \pi_2, \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \Theta_0^{(1)}, \Theta_0^{(2)}).$$

When $\boldsymbol{\theta}$ is known, the oracle QDA rule (which coincides with the Bayes rule) minimizes the overall classification error and is given by $G_{\boldsymbol{\theta}}(\mathbf{z}) = 1 + \mathbb{1}\{Q(\mathbf{z}) \leq 0\}$, where

$$Q(\mathbf{z}) = 2 \log \pi_1 f_1(\mathbf{z}) - 2 \log \pi_2 f_2(\mathbf{z}), \quad (3.6)$$

and $f_k(\mathbf{z})$ represents the density function of class k . The classification error of the oracle rule can be expressed as

$$R_{\boldsymbol{\theta}}(G_{\boldsymbol{\theta}}(\mathbf{z})) = \mathbb{P}_{\boldsymbol{\theta}}\{G_{\boldsymbol{\theta}}(\mathbf{z}) \neq L(\mathbf{z})\},$$

where $L(\mathbf{z})$ denotes the true label of $\mathbf{z}$.

In practice, we begin by estimating $\boldsymbol{\theta}$, which leads to a data-driven QDA function $\widehat{Q}(\mathbf{z})$. The corresponding data-driven QDA rule is given by

$$\widehat{G}_{\boldsymbol{\theta}}(\mathbf{z}) = 1 + \mathbb{1}\{\widehat{Q}(\mathbf{z}) \leq 0\},$$

The classification error for this rule is denoted as $R_{\boldsymbol{\theta}}(\widehat{G}_{\boldsymbol{\theta}}(\mathbf{z}))$, and is given by

$$R_{\boldsymbol{\theta}}(\widehat{G}_{\boldsymbol{\theta}}(\mathbf{z})) = \mathbb{P}_{\boldsymbol{\theta}}\{\widehat{G}_{\boldsymbol{\theta}}(\mathbf{z}) \neq L(\mathbf{z})\}. \quad (3.7)$$

We say that $\widehat{G}_\theta(\mathbf{z})$ is consistent if the following excess risk satisfies

$$R_\theta \left(\widehat{G}_\theta(\mathbf{z}) \right) - R_\theta(G_\theta(\mathbf{z})) \xrightarrow{P} 0,$$

as the sample size n and the dimensionality p go to infinity. A similar definition of consistency can be found in [Li & Shao \(2015\)](#) and [Cai & Zhang \(2021\)](#). To further investigate our data-driven QDA rule and derive its excess risk, we rewrite the QDA function (3.6) in the same manner as [Cai & Zhang \(2021\)](#).

$$Q(\mathbf{z}) = (\mathbf{z} - \boldsymbol{\mu}^{(1)})^T \Omega (\mathbf{z} - \boldsymbol{\mu}^{(1)}) - 2\boldsymbol{\beta}^T (\mathbf{z} - \bar{\boldsymbol{\mu}}) - \log \left| \Omega \left(\Theta_0^{(1)} \right)^{-1} + I_p \right| + 2 \log \left(\frac{\pi_1}{\pi_2} \right),$$

where $\bar{\boldsymbol{\mu}} = (\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})/2$, $\Omega = \Theta_0^{(2)} - \Theta_0^{(1)}$, and $\boldsymbol{\beta} = \Theta_0^{(2)} (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})$. Note that the first three terms in $Q(\mathbf{z})$ depend on Ω and $\boldsymbol{\beta}$, while the ratio (π_1/π_2) in the last term can be estimated by $(\widehat{\pi}_1/\widehat{\pi}_2) = n_1/n_2$. Additionally, we estimate $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ using the sample means $\widehat{\boldsymbol{\mu}}^{(1)} = n_1^{-1} \sum_{i=1}^{n_1} \mathbf{y}_i^{(1)}$ and $\widehat{\boldsymbol{\mu}}^{(2)} = n_2^{-1} \sum_{j=1}^{n_2} \mathbf{y}_j^{(2)}$, respectively.

In the following, we show that, under the additional assumption below, Ω and $\boldsymbol{\beta}$ can be consistently estimated:

(A.6) There exist constants $M_1, M_2 > 0$ such that $\|\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\|_\infty \leq M_1$ and $\|\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\|_0 \leq M_2$, uniformly with respect to p .

Assumptions (A.1) and (A.7) imply an upper bound on $\|\boldsymbol{\beta}\|_2$, which is a general assumption in LDA and QDA; see [Cai & Liu \(2011\)](#), [Cai et al. \(2019\)](#), and [Cai & Zhang \(2021\)](#). The pcDL decomposition (3.4) specifies that $\Theta_0^{(k)} = D_0 - L_0 - L_0^{(k)}$ share the same diagonal matrix D and low-rank matrix L_0 across groups. This leads to

$$\Omega = \Theta_0^{(2)} - \Theta_0^{(1)} = L_0^{(1)} - L_0^{(2)},$$

and subsequently

$$\|\Omega\|_F \leq \|L_0^{(1)}\|_F + \|L_0^{(2)}\|_F \leq \lambda_1 \left(L_0^{(1)} \right) \sqrt{v_{01} - r_0} + \lambda_1 \left(L_0^{(2)} \right) \sqrt{v_{02} - r_0},$$

which is bounded. This boundedness is also a general assumption in QDA settings; see [Jiang et al. \(2018\)](#) and [Cai & Zhang \(2021\)](#).

Next, we establish the consistency properties of $\widehat{\Omega} = \widehat{\Theta}^{(2)} - \widehat{\Theta}^{(1)}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\Theta}^{(2)} \left(\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)} \right)$ for the pcDL-based estimators, given the ranks are estimated by the ACT method.

Theorem 3.5.1. *Under the pcDL decomposition (3.4) and suppose assumptions (A.1)–(A.6) hold. As n and p go to infinity,*

$$\|\widehat{\Omega} - \Omega\|_F = O_p \{ \max(a_{n,p,v_0}, b_{n,p}, d_{r,v}) \}, \quad \|\widehat{\beta} - \beta\|_2 = O_p \{ \max(a_{n,p,v_0}, b_{n,p}, d_{r,v}) \},$$

in which

$$a_{n,p,v_0} = (\max_k v_{0k})^{1/2} (p/n)^{1/2}, \quad b_{n,p} = \{(p \log p)/n\}^{1/2}.$$

Theorem 4.1 in Cai & Zhang (2021) established the convergence rates for the estimation errors of Ω and β under sparsity assumptions, which are $\sqrt{s_1 \log p/n}$ and $\sqrt{s_2 \log p/n}$, respectively. Here, s_1 and s_2 are the numbers of nonzero terms in Ω and β , respectively. In comparison, our convergence rates involve $v_{0k} = \text{rank}(L_{\Sigma_0}) + \text{rank}(L_{\Sigma_0}^{(k)})$.

By substituting $\widehat{\Theta} = (\widehat{\Theta}^{(1)}, \widehat{\Theta}^{(2)})$, the solution to (3.5), along with the sample means $\widehat{\mu}^{(1)}$ and $\widehat{\mu}^{(2)}$, we derive $\widehat{Q}(\mathbf{z})$ and, consequently, the data-driven QDA rule $\widehat{G}_{\text{pcDLQ}}$. Next, we introduce an additional assumption. Combined with Theorem 3.5.1, this allows us to establish the convergence rate of the excess error for $\widehat{G}_{\text{pcDLQ}}$.

(A.7) Let $f_{Q,\theta}(x)$ be the probability density of the QDA function $Q(\mathbf{z}; \theta)$, then there exists a constant $M_3 > 0$ such that $\sup_{|x| < \epsilon} f_{Q,\theta}(x) < M_3$, uniformly with respect to p .

Assumption (A.7) states that the density of the QDA function is bounded, a standard assumption commonly used in QDA; see Cai & Liu (2011), Li & Shao (2015), Jiang et al. (2018), and Cai & Zhang (2021).

Theorem 3.5.2. *Under the pcDL decomposition (3.4) and assuming that (A.1)–(A.7) hold, as n and p go to infinity, the proposed data-driven QDA rule $\widehat{G}_{\text{pcDLQ}}$ satisfies:*

$$\sup_{\theta} \mathbb{E} \left[R_{\theta}(\widehat{G}_{\text{pcDLQ}}) - R_{\theta}(G_{\theta}) \right] \lesssim [\max(a_{n,p,v_0}, b_{n,p}, d_{r,v})]^2 \cdot \log^2 n,$$

where

$$a_{n,p,v_0} = (\max_k v_{0k})^{1/2} (p/n)^{1/2}, \quad b_{n,p} = \{(p \log p)/n\}^{1/2}.$$

Under sparsity assumptions, Theorem 3 in Li & Shao (2015) showed that their sparse QDA is consistent but did not provide an explicit convergence rate for the excess error. Theorem 4.2 in Cai & Zhang (2021) established the convergence rate as $O((s_1 + s_2) \log p \cdot \log^2 n/n)$, where s_1 and s_2 are the sparsity parameters defined earlier. Their rate has a similar form to ours, with $\max\{v_{01}, v_{02}, \log p\} \cdot p$ in our setting being replaced by $(s_1 + s_2) \cdot \log p$. These two rates coincide in cases where $v_{0k} \leq \log p$ and the sparsity parameters satisfy $s_1 + s_2 = O(p)$. Furthermore, when this sparsity assumption is violated, our proposed QDA numerically outperforms their sparse QDA (see Section 3.6.2, Models 2–6).

3.6 Simulation

3.6.1 Rank recovery

In this section, we provide empirical evidence for rank recovery, specifically how the ACT method estimates the ranks v_{01} , v_{02} , and r_0 . We compare our method with the approach incorporating an AIC penalty proposed by [Wu \(2017\)](#), referred to here as AIC.

We consider various matrix structures. For clarity, we introduce the following notations: $U(\cdot, \cdot)$ denotes a Uniform distribution with the specified lower and upper bounds, $U(\cdot)$ denotes a Uniform distribution on the given interval, and $Ber(\cdot)$ represents a Bernoulli distribution with the specified success probability. The sample size is fixed at $n_k = 100$. Unless otherwise stated, values are drawn independently for different indices whenever random sampling is involved.

Example 1: The matrices are block diagonal with different numbers of blocks. Let $\Sigma_0^{(k)} = 0.5I_p + 0.25aa^T + 0.2a^{(k)}(a^{(k)})^T$, where $a = (b_1, b_2)$, $a^{(1)} = (b_3, b_4)$, $a^{(2)} = (b_5, b_6)$, and $b_i = (0_{q \times (i-1)q}, 1_{q \times q}, 0_{q \times (p-iq)})^T$. Here, $q = \lfloor p/6 \rfloor$.

Example 2: The matrices follow a “partially common diagonal + low-rank” structure and are also sparse. Let $\Sigma_0^{(k)} = 0.5I_p + 0.2aa^T + 0.2a^{(k)}(a^{(k)})^T$, where $a \in \mathbb{R}^{p \times 2}$, $a^{(k)} \in \mathbb{R}^p$, $a_j \sim Ber(20/p)$, and $a_{j_1, j_2}^{(k)} \sim Ber(20/p)$.

Example 3: The matrices have a perturbed “partially common diagonal + low-rank” structure. Let $\Sigma_{0k} = 0.5I_p + 0.2aa^T + 0.2a^{(k)}(a^{(k)})^T + P^{(k)}$, where $a \in \mathbb{R}^{p \times 2}$, $a^{(k)} \in \mathbb{R}^p$, $a_j \sim Ber(20/p)$, and $a_{j_1, j_2}^{(k)} \sim Ber(20/p)$. The perturbation matrix is defined as $P^{(k)} = 0.01\{P_0^{(k)} + (P_0^{(k)})^T\}/2$, where $P_0^{(k)} \in \mathbb{R}^{p \times p}$, and each element of $P_0^{(k)}$ is independently drawn from $U(0, 1)$ with probability 0.2 and set to zero otherwise. To ensure positive definiteness, we set $\Sigma_0^{(k)} = \Sigma_{0k} + \{|\min(\lambda_p(\Sigma_{0k}), 0)| + 0.05\}I_p$.

For Examples 1 and 2, using the information about D_{Σ_0} , L_{Σ_0} , and $L_{\Sigma_0}^{(k)}$ provided in the example setup, we derive the true ranks as follows: $v_{01} = 2$, $v_{02} = 2$, and $r_0 = 1$ for Example 1; $v_{01} = 3$, $v_{02} = 3$, and $r_0 = 1$ for Example 2. Example 3 only approximately satisfies the pcDL decomposition due to the inclusion of a perturbation matrix; however, the low-rank component is the same as in Example 2, with $v_{01} = 3$, $v_{02} = 3$, and $r_0 = 1$.

We calculate the average rank estimates with standard errors based on 100 replications, and Table 3.1 presents the results. On average, our ACT method achieves successful

recovery of the ranks. As expected, the recovery accuracy for Example 3 is slightly lower than that for Example 2 due to the added perturbation. Compared to the AIC method, our ACT method demonstrates slightly smaller bias and standard error. Moreover, our method is computationally efficient and does not require tuning. Specifically, for each rank estimation, our method takes only 0.01 seconds, whereas the AIC method, which depends on a penalty term with a tuning parameter, takes 1.8 seconds.

Table 3.1: Average rank estimates (s.e.) from 100 replications.

			v_{01}	v_{02}	r_0
Example 1	$p = 50$	ACT	2.01(0.10)	2.01(0.10)	1.02(0.14)
		AIC	2.09(0.32)	2.10(0.30)	0.97(0.22)
$v_{01} = 2, v_{02} = 2$	$p = 100$	ACT	1.99(0.33)	1.93(0.29)	1.03(0.35)
		AIC	1.86(0.78)	1.72(0.57)	0.83(0.40)
$r_0 = 1$	$p = 200$	ACT	1.83(0.39)	1.85(0.39)	1.05(0.51)
		AIC	1.79(0.46)	1.74(0.43)	0.72(0.14)
Example 2	$p = 50$	ACT	3.01(0.01)	3.00(0.01)	1.03(0.06)
		AIC	3.02(0.14)	3.04(0.10)	0.94(0.17)
$v_{01} = 3, v_{02} = 3$	$p = 100$	ACT	2.99(0.10)	3.01(0.17)	0.97(0.18)
		AIC	3.02(0.14)	3.06(0.34)	0.95(0.22)
$r_0 = 1$	$p = 200$	ACT	3.01(0.10)	3.03(0.17)	1.04(0.27)
		AIC	3.07(0.26)	3.10(0.41)	0.89(0.31)
Example 3	$p = 50$	ACT	3.02(0.07)	3.04(0.05)	0.96(0.30)
		AIC	3.02(0.14)	3.06(0.34)	0.90(0.37)
$v_{01} = 3, v_{02} = 3$	$p = 100$	ACT	3.02(0.10)	3.02(0.08)	1.06(0.29)
		AIC	3.03(0.17)	3.07(0.35)	0.88(0.14)
$r_0 = 1$	$p = 200$	ACT	3.04(0.12)	3.05(0.17)	1.15(0.29)
		AIC	3.12(0.20)	3.12(0.38)	0.78(0.48)

3.6.2 QDA

In this section, we evaluate the empirical performance of our proposed pcDLQ rule and compare it to several competing methods, including:

1. Sparse QDA (SQDA) by [Cai & Zhang \(2021\)](#);
2. Sparse LDA (SLDA) by [Cai & Liu \(2011\)](#);
3. Classical QDA;

4. Classical LDA;
5. QDA using the plain DL estimators (DLQ) by [Wu et al. \(2019\)](#);
6. The oracle procedure.

The oracle procedure leverages the true underlying model and serves as a benchmark for comparison. Regarding sparse QDA methods, [Cai & Zhang \(2021\)](#) demonstrated that their proposed sparse QDA method outperforms other existing sparse QDA methods in the literature (e.g., [Li & Shao \(2015\)](#), [Jiang et al. \(2018\)](#)). Therefore, we only include theirs and denote it as SQDA in our simulations.

The training and testing sample size are set to be $n_1 = n_2 = 100$, while the dimension p varies among 50, 75, 100 to 200. We consider various structures, including sparse covariance matrices, non-sparse covariance matrices with sparse differential graphs, and covariance matrices similar to those in Section 3.6.1. To overcome the singularity of sample covariance matrices when $p > n_1$ or $p > n_2$, we employ $\hat{\Sigma}^{(k)} + \sqrt{\log p / (n_1 + n_2)} I_p$ as the estimator in the classical QDA.

Model 1 is identical to Model 1 in [Cai & Zhang \(2021\)](#). The mean of class 1 is set to $\boldsymbol{\mu}^{(1)} = (0, \dots, 0)^T$. The differential graph Ω is constructed as a random sparse symmetric matrix with 10 nonzero entries in total, independently generated from $N(0, 1)$. Finally, we set $\Theta_0^{(2)} = \Omega + \Theta_0^{(1)}$, $\boldsymbol{\mu}^{(2)} = \Sigma_0^{(2)} \boldsymbol{\beta}$, where

$$\boldsymbol{\beta} = (\underbrace{1, \dots, 1}_{10}, 0, \dots, 0)^T.$$

The covariance matrices are given by $\Sigma_0^{(1)} = (\Theta_0^{(1)})^{-1}$ and $\Sigma_0^{(2)} = (\Theta_0^{(2)})^{-1}$.

For Models 2-6, we directly generate covariance matrices $\Sigma^{(1)}$ and $\Sigma^{(2)}$. Model 3 and 4 use the covariance matrices defined in Examples 2 and 3 from Section 3.6.1, with

$$\boldsymbol{\mu}^{(1)} = (0, \dots, 0)^T \text{ and } \boldsymbol{\mu}^{(2)} = (\underbrace{0.6, \dots, 0.6}_{10}, 0, \dots, 0)^T.$$

Model 1: Block Sparse model

$\Theta_0^{(1)} = K^T \text{diag}(d) K$, where $d \in \mathbf{R}^p$, $d_i \sim U(0, 1)$, and $K \in \mathbf{R}^{p \times p}$ is a random matrix with entries independently drawn from $N(0, 1)$.

Model 2: Diagonal + Low-rank

$\Sigma_0^{(k)} = \text{diag}(d^{(k)}) + 0.2 a^{(k)} (a^{(k)})^T$, where $d^{(k)} \in \mathbf{R}^p$, $d_i^{(k)} \sim U(0.2, 0.6)$, $a^{(k)} \in \mathbf{R}^{p \times 2}$, $a_{i,j}^{(k)} \sim \text{Ber}(0.4)$.

Model 3: Partially Common Diagonal + Low-rank

$\Sigma_0^{(k)} = 0.5I_p + 0.2aa^T + 0.2a^{(k)}(a^{(k)})^T$, where $a \in \mathbf{R}^{p \times 2}$, $a^{(k)} \in \mathbf{R}^p$, $a_j \sim \text{Ber}(20/p)$ and $a_{j_1, j_2}^{(k)} \sim \text{Ber}(20/p)$.

Model 4: Perturbed Partially Common Diagonal + Low-rank

$\Sigma_{0k} = 0.5I_p + 0.2aa^T + 0.2a^{(k)}(a^{(k)})^T + P^{(k)}$, where $a \in \mathbf{R}^{p \times 2}$, $a^{(k)} \in \mathbf{R}^p$, $a_j \sim \text{Ber}(20/p)$ and $a_{j_1, j_2}^{(k)} \sim \text{Ber}(20/p)$. The perturbation matrix $P^{(k)} = 0.01\{P_0^{(k)} + (P_0^{(k)})^T\}/2$. Here, $P_0^{(k)} \in \mathbf{R}^{p \times p}$ has entries independently drawn from $U(0, 1)$ with probability 0.2 and set to zero otherwise. To ensure positive definiteness, we set $\Sigma_0^{(k)} = \Sigma_{0k} + \{|\min(\lambda_p(\Sigma_{0k}), 0)| + 0.05\}I_p$.

Model 5: Partially Common Sparse

$B_0^{(k)} = 0.2a + 0.2a^{(k)}$, where $a, a^{(k)} \in \mathbf{R}^{p \times p}$, $a_{j_1, j_2} \sim \text{Ber}(0.2)$, and $a_{j_1, j_2}^{(k)} \sim \text{Ber}(0.2)$. Define $B^{(k)} = 0.5I_p + \{B_0^{(k)} + (B_0^{(k)})^T\}/2$. To ensure positive definiteness, let $\Sigma_0^{(k)} = [B^{(k)} + \{|\min(\lambda_p(B^{(k)}), 0)| + 0.05\}I_p]^{-1}$.

Model 6: Partially Common Networks

Let $A \in \mathbf{R}^{p \times p}$ be the adjacency matrix of a network with 10 unconnected sub-networks, each following a power-law degree distribution. Each element of A is then multiplied by an independent random variable $a_{j_1, j_2} \sim U\{(-0.4, -0.1) \cup (0.4, 0.1)\}$. Define $B_1 = I_p + (A + A^T)/2$. To ensure positive definiteness, let

$$B_2 = [B_1 + \{|\min(\lambda_p(B_1), 0)| + 0.05\}I_p]^{-1}.$$

Define each element of B as follows,

$$B_{j_1, j_2} = \begin{cases} 1 & j_1 = j_2 \\ 0.6(B_2)_{j_1, j_2} / \sqrt{(B_2)_{j_1, j_1}(B_2)_{j_2, j_2}} & j_1 \neq j_2. \end{cases}$$

Finally, define $\Sigma_0^{(1)} = B$ and $\Sigma_0^{(2)}$ as identical to $\Sigma_0^{(1)}$, except that one sub-network block is replaced with the identity matrix. This structure is similar to that used in [Danaher et al. \(2014\)](#).

The performance of the classification rule is measured by the average misclassification rate based on 100 replications. Figures 3.1 and 3.2 present the results.

- **Model 1:** SQDA achieves the lowest misclassification rate, outperforming other methods. However, the gap between SQDA and pcDLQ narrows as p increases.

- **Model 2:** DLQ achieves the lowest misclassification rate, as expected, while pcDLQ performs slightly worse.
- **Models 3 and 4:** Our method correctly estimates the number of latent factors. pcDLQ achieves the lowest misclassification rate, while LDA and SQDA also perform comparably.
- **Models 5 and 6:** When the pcDL decomposition is not strictly satisfied but the two groups share some common factors, pcDLQ still outperforms other methods.

In summary, classical LDA and QDA fail when $p > n$, making structural assumptions crucial for achieving low misclassification rate. Both the sparsity assumption on the differential graph (Ω) and the partially common diagonal+low-rank assumption on covariance matrices are effective. Among these, the proposed pcDLQ demonstrated robustness across various settings.

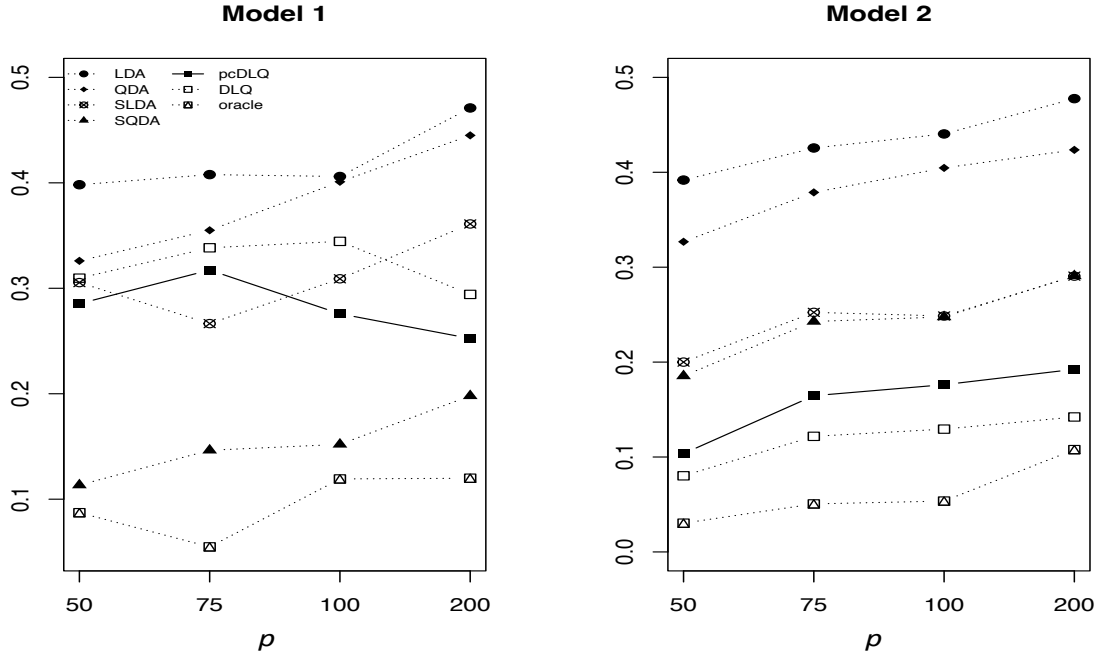


Figure 3.1: Average misclassification rates for Models 1 and 2.

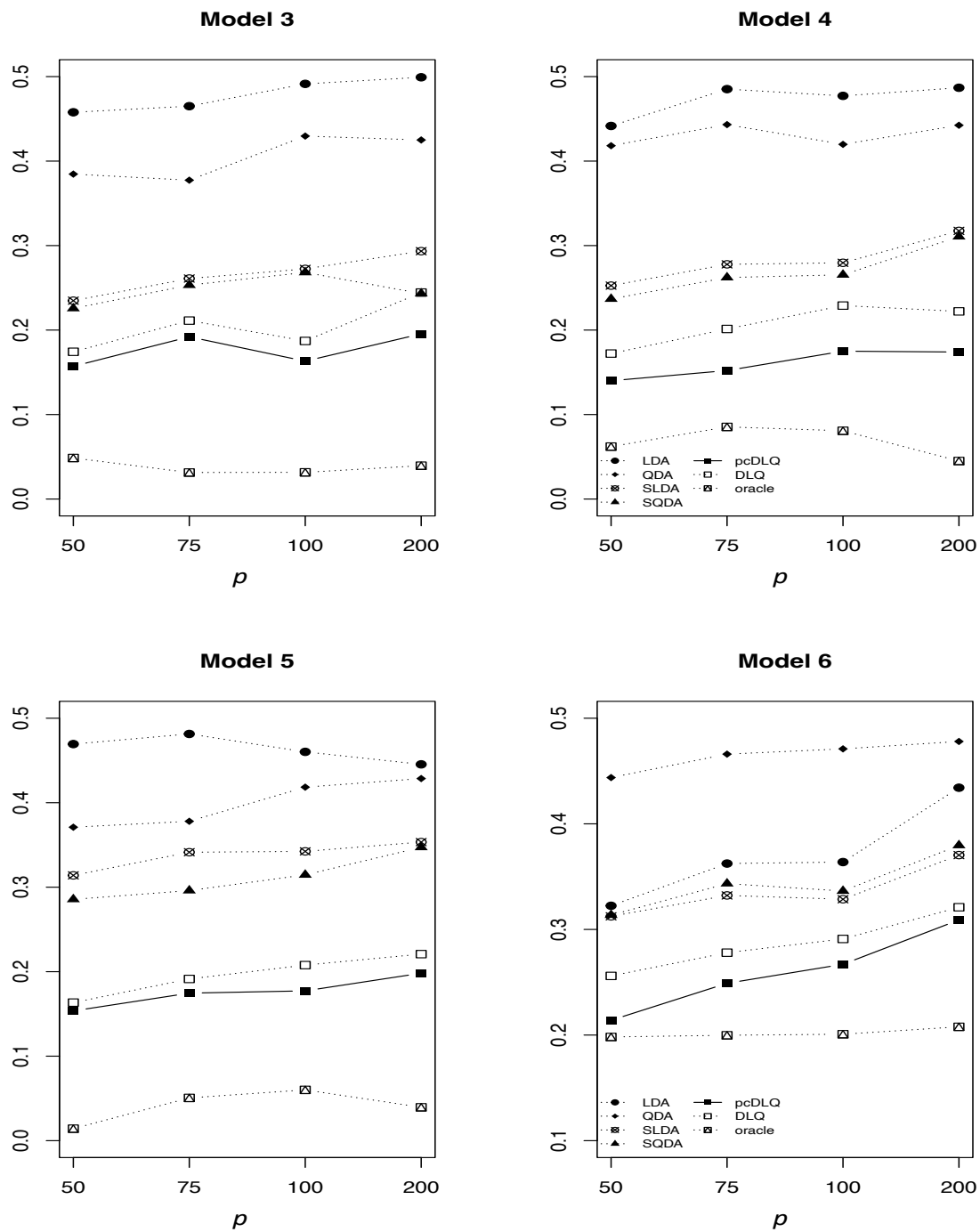


Figure 3.2: Average misclassification rates for Models 3–6.

3.7 Real data analysis

In addition to the simulation studies, we demonstrate the performance of the pcDLQ classifier through the analysis of a real dataset. The prostate cancer dataset by [Singh et al. \(2002\)](#) is publicly available at <ftp://stat.ethz.ch/Manuscripts/dettling/prostate.rda>.

The prostate cancer data consists of gene expression levels of 6033 genes from 102 individuals (50 normal and 52 prostate cancer patients). For this analysis, we retain only the top p genes with the largest absolute values of the two samples t -statistic. The dataset is randomly split into a training set (75%) and a testing set (the remaining 25%). All classifiers are trained on the training data and evaluated on the testing set. We repeat this process 100 times for $p = 50, 75, 100$.

Table 3.2 reports the average misclassification errors along with their standard rates. The pcDLQ achieves the best results across all three values of p , with SQDA and SLDA performing slightly worse. As expected, the classical QDA and LDA perform no better than random guessing.

Table 3.2: Average misclassification rates (s.e.) for prostate cancer data.

p	pcDLQ	SQDA	QDA	SLDA	LDA	DLQ
50	0.228(0.077)	0.267(0.104)	0.424(0.097)	0.295(0.105)	0.424(0.099)	0.300(0.077)
75	0.165(0.051)	0.265(0.092)	0.439(0.090)	0.287(0.109)	0.446(0.089)	0.288(0.079)
100	0.192(0.077)	0.256(0.079)	0.440(0.085)	0.260(0.082)	0.442(0.182)	0.212(0.108)

3.8 Technical proofs

3.8.1 Proofs of Theorems 3.3.2 and 3.3.4

Following a similar proof as in [Fan et al. \(2022\)](#), which established the rank estimator for a single correlation matrix, we extend their results to the rank estimator for the pooled correlation matrix of two groups.

Proof of Theorem 3.3.2. Let $S_n = [\text{diag}(\Sigma_0)]^{-1/2} \hat{\Sigma} [\text{diag}(\Sigma_0)]^{-1/2} = (\hat{\sigma}_{\ell_1, \ell_2})$. Its corresponding correlation matrix can be rewritten as

$$\hat{R} = [\text{diag}(S_n)]^{-1/2} S_n [\text{diag}(S_n)]^{-1/2}.$$

We first show that $\max_{k=v_{12}+1, \dots, p} |\lambda_k(\hat{R}) - \lambda_k(S_n)| \rightarrow 0$, a.s in step 1. Then, in step 2, we show the almost sure convergence of $m_n(z)$ to $m^{S_n}(z)$, the Stieljes transform of the ESD of S_n . In step 3, we show the almost sure convergence of $m^{S_n}(z)$ to $m(z)$, which finalizes the proof.

Step 1. By Lemma S.2 in [Fan et al. \(2022\)](#), we have

$$\begin{aligned} \min_{j \leq p} \hat{\sigma}_{jj} &= \min_{j \leq p} \lambda_j(\text{diag}(S_n)) \leq \frac{\lambda_k(S_n)}{\lambda_k(S_n [\text{diag}(S_n)]^{-1})}, \\ \frac{\lambda_k(S_n)}{\lambda_k(S_n [\text{diag}(S_n)]^{-1})} &\leq \max_{j \leq p} \lambda_j(\text{diag}(S_n)) = \max_{j \leq p} \hat{\sigma}_{jj}, \end{aligned}$$

for $k \leq p$. Thus,

$$\max_{k=1, \dots, p} \left| \frac{\lambda_k(S_n)}{\lambda_k(S_n [\text{diag}(S_n)]^{-1})} - 1 \right| \leq \max \left\{ \left| \min_{j \leq p} \hat{\sigma}_{jj} - 1 \right|, \left| \max_{j \leq p} \hat{\sigma}_{jj} - 1 \right| \right\} = \max_{j \leq p} |\hat{\sigma}_{jj} - 1|,$$

By Lemma S.8 in [Fan et al. \(2022\)](#), $\max |\hat{\sigma}_{jj} - 1| = o_{a.s.}(1)$, then we have

$$\max_{k \leq p} \left| \frac{\lambda_k(S_n)}{\lambda_k(S_n [\text{diag}(S_n)]^{-1})} - 1 \right| = o_{a.s.}(1). \quad (3.8)$$

Therefore,

$$\frac{\max_{k=v_{12}+1, \dots, p} |\lambda_k(S_n [\text{diag}(S_n)]^{-1}) - \lambda_k(S_n)|}{\lambda_{v_{12}+1}(S_n [\text{diag}(S_n)]^{-1})} \leq \max_{k=v_{12}+1, \dots, p} \left| \frac{\lambda_k(S_n)}{\lambda_k(S_n [\text{diag}(S_n)]^{-1})} - 1 \right| = o_{a.s.}(1).$$

Then, if $\lambda_{v_{12}+1}(S_n [\text{diag}(S_n)]^{-1}) = O(1)$, we have $\max_{k=v_{12}+1, \dots, p} |\lambda_k(\widehat{R}) - \lambda_k(S_n)| = o_{a.s.}(1)$.

By Lemma S.2 in [Fan et al. \(2022\)](#) again, we have

$$\lambda_{v_{12}+1}(S_n [\text{diag}(S_n)]^{-1}) \leq \frac{\lambda_{v_{12}+1}(S_n)}{\min_j \widehat{\sigma}_{jj}},$$

$$\lambda_{v_{12}+1}(S_n) \leq \lambda_{v_{12}+1}(QQ^T) \lambda_1 \left(\frac{1}{n} \sum_{i=1}^{n_1} (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)})^T + \frac{1}{n} \sum_{i=1}^{n_2} (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)})^T \right),$$

where $Q = [\text{diag}(\Sigma_0)]^{-1/2}(B, B^{(1)}, B^{(2)}, D^{1/2})$, and by Proposition 3.3.1, $\lambda_{v_{12}+2}(QQ^T) \leq 1$.

Note that,

$$\begin{aligned} & \lambda_1 \left(\frac{1}{n} \sum_{i=1}^{n_1} (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)})^T + \frac{1}{n} \sum_{i=1}^{n_2} (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)})^T \right) \\ & \leq \lambda_1 \left(\frac{1}{n} \sum_{i=1}^{n_1} (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)})^T \right) + \lambda_1 \left(\frac{1}{n} \sum_{i=1}^{n_2} (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)})^T \right) \\ & = \frac{n_1}{n} \lambda_1 \left(\frac{1}{n_1} \sum_{i=1}^{n_1} (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(1)})^T \right) + \frac{n_2}{n} \lambda_1 \left(\frac{1}{n_2} \sum_{i=1}^{n_2} (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)}) (\mathbf{x}_i - \bar{\mathbf{x}}^{(2)})^T \right) \\ & \leq \left(\sqrt{\frac{n_1}{n}} + \sqrt{\rho_n} \right)^2 + \left(\sqrt{\frac{n_2}{n}} + \sqrt{\rho_n} \right)^2 + \epsilon_0 + o_{a.s.}(1), \end{aligned}$$

where the last inequality holds by applying Lemma S.6 in [Fan et al. \(2022\)](#), $\rho_n = p/n$, and ϵ_0 is an arbitrary positive constant. Therefore, we have

$$\lambda_{v_{12}+1}(S_n [\text{diag}(S_n)]^{-1}) \leq \left(\sqrt{\frac{n_1}{n}} + \sqrt{\rho_n} \right)^2 + \left(\sqrt{\frac{n_2}{n}} + \sqrt{\rho_n} \right)^2 + \epsilon_0 + o_{a.s.}(1).$$

Thus, $\lambda_{v_{12}+1}(S_n [\text{diag}(S_n)]^{-1}) = O(1)$, and,

$$\max_{k=v_{12}+1, \dots, p} |\lambda_k(\widehat{R}) - \lambda_k(S_n)| = o_{a.s.}(1). \quad (3.9)$$

Step 2. Define the Stieljes transform of the ESD $F^{S_n}(t) = p^{-1} \sum_{j=1}^p \mathbb{1}(\lambda_j(S_n) \leq t)$ as

$$m_n^{S_n}(z) = \int (t - z)^{-1} dF^{S_n}(t).$$

Next, we show the convergence of $|m_n(z) - m_n^{S_n}(z)|$. For $v_{12} = o(p)$,

$$p^{-1} \left| \sum_{j=1}^{v_{12}} (\lambda_j(S_n) - z)^{-1} \right| \leq v_{12} p^{-1} v^{-1} \rightarrow 0,$$

for $z = u + iv$ with $v > 0$. Therefore,

$$m_n^{S_n}(z) = (p - v_{12})^{-1} \sum_{j=v_{12}+1}^p (\lambda_j(S_n) - z)^{-1} + o(1), \quad z \in \mathcal{C}^+.$$

For any $z = u + iv \in \mathcal{C}^+$, we have

$$\begin{aligned} m_n(z) - m_n^{S_n}(z) &= (p - v_{12})^{-1} \sum_{j=v_{12}+1}^p \left[(\lambda_j(\hat{R}) - z)^{-1} - (\lambda_j(S_n) - z)^{-1} \right] + o(1) \\ &= (p - v_{12})^{-1} \sum_{j=v_{12}+1}^p \frac{\lambda_j(S_n) - \lambda_j(\hat{R})}{(\lambda_j(\hat{R}) - z)(\lambda_j(S_n) - z)} + o(1). \end{aligned}$$

Thus, we have

$$\begin{aligned} &|m_n(z) - m_n^{S_n}(z)| \\ &\leq \frac{\max_{j=v_{12}+1, \dots, p} |\lambda_j(S_n) - \lambda_j(\hat{R})|}{\min_{j=v_{12}+1, \dots, p} \sqrt{(\lambda_j(\hat{R}) - u)^2 + v^2} \sqrt{(\lambda_j(S_n) - u)^2 + v^2}} + o(1) \\ &\leq \frac{\max_{j=v_{12}+1, \dots, p} |\lambda_j(S_n) - \lambda_j(\hat{R})|}{v^2} + o(1). \end{aligned}$$

By (3.9), for $z = u + iv \in \mathcal{C}^+$, we have

$$|m_n(z) - m_n^{S_n}(z)| = o_{a.s.}(1). \quad (3.10)$$

Step 3. Let $\bar{\epsilon}^{(1)} = n_1^{-1} \sum_{i=1}^{n_1} \epsilon_i$, $\bar{\epsilon}^{(2)} = n_2^{-1} \sum_{i=1}^{n_2} \epsilon_i$, denote

$$S_{44} = S_{44}^{(1)} + S_{44}^{(2)},$$

where

$$S_{44}^{(1)} = \frac{1}{n} Q_4 \sum_{i=1}^{n_1} (\epsilon_i - \bar{\epsilon}^{(1)}) (\epsilon_i - \bar{\epsilon}^{(1)})^T Q_4^T, \quad S_{44}^{(2)} = \frac{1}{n} Q_4 \sum_{i=1}^{n_2} (\epsilon_i - \bar{\epsilon}^{(2)}) (\epsilon_i - \bar{\epsilon}^{(2)})^T Q_4^T,$$

and $Q_4 = [\text{diag}(\Sigma_0)]^{-1/2} D^{1/2}$.

Let $m_n^{S_{44}}(z)$ be the Stieltjes transform of S_{44} ,

$$m_n^{S_{44}}(z) = \int (t - z)^{-1} dF^{S_{44}}(t) = p^{-1} \sum_{j=1}^p (\lambda_j(S_{44}) - z)^{-1}, \quad z \in \mathcal{C}^+,$$

where $F^{S_{44}}(t) = p^{-1} \sum_{j=1}^p \mathbb{1}(\lambda_j(S_{44}) \leq t)$. Then, consider $|(t - z)^{-1}| \leq v^{-1}$ with $z = u + iv, v > 0$

$$|m_n^{S_n}(z) - m_n^{S_{44}}(z)| = \left| \int (t - z)^{-1} d(F^{S_n}(t) - F^{S_{44}}(t)) \right| \leq \frac{2 \|F^{S_n} - F^{S_{44}}\|}{v},$$

where $\|F^{S_n} - F^{S_{44}}\| = \sup_t |F^{S_n}(t) - F^{S_{44}}(t)|$. By Theorem A.43 in [Bai & Silverstein \(2010\)](#), we have

$$\|F^{S_n} - F^{S_{44}}\| \leq \frac{\text{rank}(S_n - S_{44})}{p} \leq \frac{v_{12}}{p} = o(1).$$

Therefore, we conclude that for any $z = u + iv$ with $v > 0$,

$$|m_n^{S_n}(z) - m_n^{S_{44}}(z)| = o(1). \quad (3.11)$$

By (3.10) and (3.11), we have $|m_n(z) - m_n^{S_{44}}(z)| = o_{a.s.}(1)$.

Write $R = A_1 A_1^T + Q_4 Q_4^T$, where $A_1 = [\text{diag}(\Sigma_0)]^{-1/2} \left(B, \sqrt{n_1/n} B^{(1)}, \sqrt{n_2/n} B^{(2)} \right)$ with $\text{rank}(A_1) = v_{12}$. By Theorem A.43 in [Bai & Silverstein \(2010\)](#) again, we have

$$\sup_t |F^{Q_4 Q_4^T}(t) - F^R(t)| \leq v_{12} p^{-1} = o(1).$$

Note that $\sup_t |F^R(t) - H_{p-v_{12}}(t)| = o(1)$, and $H(t)$ is the limit of $H_{p-v_{12}}(t)$ by assumption (A.4). Thus, $H(t)$ is also the limit of $F^{Q_4 Q_4^T}(t)$. By [Silverstein & Choi \(1995\)](#), we have $|m_n^{S_{44}}(z) - m(z)| = o_{a.s.}(1)$, where $m(z)$ and $\underline{m}(z)$ are related such that

$$\underline{m}(z) = -\frac{1 - \rho}{z} + \rho m(z),$$

and $\underline{m}(z)$ is the unique solution to

$$z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1 + t\underline{m}(z)}, \quad z \in \mathcal{C}^+.$$

Let $\psi(x) = 1 + \rho \int \frac{t}{x-t} dH(t)$, we have $z = -\underline{m}^{-1}(z) \psi(-\underline{m}^{-1}(z))$.

□

Proof of Theorem 3.3.4. [Fan et al. \(2022\)](#) show the first two convergence. For the third one, note that

$$\lambda_1 \geq \dots \geq \lambda_{v_{12}} > 1 + \sqrt{\rho} + \epsilon_0 > 1 \geq \lambda_{v_{12}+1} \geq \dots \geq \lambda_p$$

for a very small positive number $\epsilon_0 > \delta$. By Theorem 3.3.2, we have

$$\widehat{\lambda}_{v_{12}+1} = \widehat{\lambda}_{v_{12}+1}^C + \rho \int \frac{\widehat{\lambda}_{v_{12}+1}^C t}{\widehat{\lambda}_{v_{12}+1}^C - t} dH(t) + o_p(1) = \widehat{\lambda}_{v_{12}+1}^C \psi(\widehat{\lambda}_{v_{12}+1}^C) + o_p(1).$$

Moreover, by [Fan et al. \(2022\)](#), $\widehat{\lambda}_{v_{12}+1} \leq (1 + \sqrt{\rho})\psi(1 + \sqrt{\rho}) + o_p(1)$. Thus, we have

$$\widehat{\lambda}_{v_{12}+1}^C \leq 1 + \sqrt{\rho} + o_p(1).$$

By Proposition 3.3.3, we have

$$\widehat{\lambda}_j^C = \lambda_j + o_p(1) \geq 1 + \sqrt{\rho} + \epsilon_0 + o_p(1), \quad j \leq v_{12}.$$

Therefore, when $s = 1 + \sqrt{\rho} + \delta$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\widehat{v}_{12}^C(s) = v_{12}) = 1.$$

Since $r_0 = v_{01} + v_{02} - v_{12}$, then the last convergence also holds. □

3.8.2 Proofs of Theorems 3.5.1 and 3.5.2

In this section, we begin by presenting some technical lemmas that will be used in the proofs. Following a similar approach to Cai & Zhang (2021), we establish the convergence of the estimates for two key components in Theorem 3.5.1. Subsequently, we establish the convergence of the classification error for our proposed pcDLQ classifier in Theorem 3.5.2.

Lemma 3.8.1. *For any matrices $A \in \mathbf{S}_+^p$, $B, C \in \mathbf{R}^{p \times p}$, with $A = UDU^T$, $D \in \mathbf{D}_+^p$, and a vector $\mathbf{a} \in \mathbf{R}^p$,*

$$\|A\mathbf{a}\|_2 = \|UDU^T\mathbf{a}\|_2 = \|D\mathbf{a}\|_2 \leq \|\text{diag}(D)\|_2 \|\mathbf{a}\|_2 = \|A\|_F \|\mathbf{a}\|_2, \quad (3.12)$$

$$\|BC\|_F \leq \|B\|_F \max_{1 \leq i, j \leq p} |C_{ij}| \leq \|B\|_F \|C\|_{op}, \quad (3.13)$$

$$\|C\|_{op} = \lambda_1(D) \leq \sqrt{\sum_{i=1}^p \lambda_i^2(D)} = \|D\|_F \leq \sqrt{p} \|D\|_{op}. \quad (3.14)$$

Lemma 3.8.2. *For any two matrices $A, B \in \mathbf{S}_+^p$*

$$\log |A| - \log |B| \leq \left| \text{tr} (B^{-1}(A - B)) \right|.$$

Proof. Consider an arbitrary positive semidefinite matrix C with eigenvalues $\lambda_1(C), \dots, \lambda_p(C)$, then

$$\log |I + C| = \sum_{i=1}^p \log(1 + \lambda_i(C)) \leq \sum_{i=1}^p \lambda_i(C) = \text{tr}(C).$$

Hence,

$$\log |A| - \log |B| = \log |B^{-1}A| = \log |I + B^{-1}(A - B)| \leq \text{tr} (B^{-1}(A - B)).$$

□

Next, we present proofs of Theorems 3.5.1 and 3.5.2.

Proof of Theorem 3.5.1.

$$\begin{aligned} \left\| \widehat{\Omega} - \Omega \right\|_F &= \left\| (\widehat{\Theta}^{(2)} - \widehat{\Theta}^{(1)}) - (\Theta_0^{(2)} - \Theta_0^{(1)}) \right\|_F \\ &= \left\| (\widehat{\Theta}^{(2)} - \Theta_0^{(2)}) - (\widehat{\Theta}^{(1)} - \Theta_0^{(1)}) \right\|_F \\ &\leq \left\| \widehat{\Theta}^{(2)} - \Theta_0^{(2)} \right\|_F + \left\| \widehat{\Theta}^{(1)} - \Theta_0^{(1)} \right\|_F \\ &= O_p\{\max(a_{n,p,v_0}, b_{n,p}, d_{r,v})\} \quad (\text{by Proposition 3.4.1}). \end{aligned}$$

Denote $\boldsymbol{\delta} = \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}$. To obtain the convergence rate of $\boldsymbol{\beta}$, we start from

$$\begin{aligned}
\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 &= \|\hat{\Theta}^{(2)}\hat{\boldsymbol{\delta}} - \Theta_0^{(2)}\boldsymbol{\delta}\|_2 \\
&\leq \|\Theta_0^{(2)}\boldsymbol{\delta} - \Theta_0^{(2)}\hat{\boldsymbol{\delta}}\|_2 + \|\Theta_0^{(2)}\hat{\boldsymbol{\delta}} - \hat{\Theta}^{(2)}\hat{\boldsymbol{\delta}}\|_2 \\
&= \|\Theta_0^{(2)}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\|_2 + \|(\hat{\Theta}^{(2)} - \Theta_0^{(2)})\hat{\boldsymbol{\delta}}\|_2 \\
&\leq \|\Theta_0^{(2)}\|_F \|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|_2 + \|(\hat{\Theta}^{(2)} - \Theta_0^{(2)})\|_F \|\hat{\boldsymbol{\delta}}\|_2 \quad (\text{by Lemma 3.8.1}). \tag{3.15}
\end{aligned}$$

Next, we discuss (3.15) by parts. Based on the assumption of matrix structure,

$$\|\Theta_0^{(2)}\|_F \leq \|D\|_F + \|L_0^{(2)}\|_F. \tag{3.16}$$

Then for diagonal matrix D ,

$$\|D\|_F \leq \sqrt{p}\lambda_1(D). \tag{3.17}$$

For low-rank matrix $L_0^{(2)}$,

$$\begin{aligned}
\|L_0^{(2)}\|_F &= \|D^{1/2}UVU^TD^{1/2}\|_F \\
&\leq \sqrt{v_{02}} \|D^{1/2}UVU^TD^{1/2}\|_{op} \\
&= \sqrt{v_{02}}\lambda_1(DV) \\
&\leq \sqrt{v_{02}}\lambda_1(D). \tag{3.18}
\end{aligned}$$

where $L_0^{(2)} = D^{1/2}UVU^TD^{1/2}$ is the analytic solution to the objective function, as discussed in Proposition 5.1. Eigen-decompose $D^{1/2}\Sigma D^{1/2}$, let $w_1, \dots, w_p$ be the eigenvalues in descending order and $u_1, \dots, u_p$ be the associated eigenvectors. $U = (u_1, \dots, u_p)$, orthogonal, and $V = \text{Diag}\{1 - 1/\max(w_1, 1), \dots, 1 - 1/\max(w_p, 1)\} \prec I_p$. The third equation holds because $\|X\|_{op} = \lambda_1(X)$ for a symmetric matrix X , and $\lambda_1(X) = \lambda_1(UXU^T)$ for any orthogonal matrix U . In addition, since both D and $L_0^{(2)}$ are positive definite, then by assumption (A.1),

$$\lambda_1(D) \leq \lambda_1(\Theta_0^{(2)}) \leq c_1^{-1}. \tag{3.19}$$

Finally, by assumption (A.5),

$$\begin{aligned}
\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|_2 &\leq \|\hat{\boldsymbol{\mu}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_2 + \|\hat{\boldsymbol{\mu}}^{(1)} - \boldsymbol{\mu}^{(1)}\|_2 \\
&\leq M_2 \left(\|\hat{\boldsymbol{\mu}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_\infty + \|\hat{\boldsymbol{\mu}}^{(1)} - \boldsymbol{\mu}^{(1)}\|_\infty \right). \tag{3.20}
\end{aligned}$$

Note that $\left\|\hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu}^{(k)}\right\|_{\infty} \lesssim \sqrt{\frac{\log p}{n}}$, see Lemma 8.5 in [Cai & Zhang \(2021\)](#) for detail, then by (3.16)-(3.20), (3.15) becomes

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2 &\leq c_1^{-1} M_2 \sqrt{\frac{\max(v_{02}, p) \log p}{n}} + M_1 M_2 \left\|\hat{\boldsymbol{\Theta}}^{(2)} - \boldsymbol{\Theta}_0^{(2)}\right\|_F \\ &= O_p\{\max(a_{n,p,v_0}, b_{n,p}, d_{r,v})\}.\end{aligned}$$

□

Proof of Theorem 3.5.2. We express the the excess risk $R_{\theta}(\hat{G}_{\text{pcDLQ}}) - R_{\theta}(G_{\theta})$ as the integral of the QDA function over the normal distribution, and then discuss the upper bound of the integral by splitting it into various parts.

Recall that $f_1(\mathbf{z})$ and $f_2(\mathbf{z})$ are the density functions of class 1 and class 2, respectively. $Q(\mathbf{z})$ represents QDA function with known parameters, and it can be written as

$$Q(\mathbf{z}) = 2(\log \pi_1 f_1(\mathbf{z}) - \log \pi_2 f_2(\mathbf{z})).$$

Note that $\hat{Q}(\mathbf{z})$ represents the sample-driven QDA function by pcDLQ, and we denote the difference as

$$M(\mathbf{z}) = Q(\mathbf{z}) - \hat{Q}(\mathbf{z}).$$

Furthermore, let $B_1(\mathbf{z}) := \{\mathbf{z} | Q(\mathbf{z}) > 0\}$ represent the set of observations classified to class 1 by the Bayes rule, and $\hat{B}_1(\mathbf{z}) := \{\mathbf{z} | \hat{Q}(\mathbf{z}) > 0\}$ represent the set of observations classified to class 1 by the pcDLQ rule. Then,

$$B_1 \cap \hat{B}_1^C = \{\mathbf{z} | Q(\mathbf{z}) > 0, \hat{Q}(\mathbf{z}) \leq 0\} = \{\mathbf{z} | 0 < Q(\mathbf{z}) \leq Q(\mathbf{z}) - \hat{Q}(\mathbf{z}) = M(\mathbf{z})\}$$

is the set of observations classified to class 1 by the Bayes rule but class 2 by the pcDLQ rule. For simplicity, we denote $s_{p,r} := \max(a_{n,p,v_0}, b_{n,p}, d_{r,v})$, in which

$$a_{n,p,v_0} = (\max_k v_{0k})^{1/2} (p/n)^{1/2}, \quad b_{n,p} = \{(p \log p)/n\}^{1/2}.$$

Therefore, the excess risk,

$$\begin{aligned}
& R_\theta(\widehat{G}_{\text{pcDLQ}}) - R_\theta(G_\theta) \\
&= \mathbb{P}_\theta \{G_\theta(\mathbf{z}) = L(\mathbf{z})\} - \mathbb{P}_\theta \{\widehat{G}_{\text{pcDLQ}}(\mathbf{z}) = L(\mathbf{z})\} \\
&= \int_{B_1(\mathbf{z})} \pi_1 f_1(\mathbf{z}) d\mathbf{z} + \int_{B_1^C(\mathbf{z})} \pi_2 f_2(\mathbf{z}) d\mathbf{z} - \int_{\widehat{B}_1(\mathbf{z})} \pi_1 f_1(\mathbf{z}) d\mathbf{z} - \int_{\widehat{B}_1^C(\mathbf{z})} \pi_2 f_2(\mathbf{z}) d\mathbf{z} \\
&= \int_{B_1(\mathbf{z})} \pi_1 f_1(\mathbf{z}) d\mathbf{z} + \pi_2 \left(1 - \int_{B_1(\mathbf{z})} f_2(\mathbf{z}) d\mathbf{z}\right) - \int_{\widehat{B}_1(\mathbf{z})} \pi_1 f_1(\mathbf{z}) d\mathbf{z} - \pi_2 \left(1 - \int_{\widehat{B}_1(\mathbf{z})} f_2(\mathbf{z}) d\mathbf{z}\right) \\
&= \int_{B_1(\mathbf{z})} (\pi_1 f_1(\mathbf{z}) - \pi_2 f_2(\mathbf{z})) d\mathbf{z} - \int_{\widehat{B}_1(\mathbf{z})} (\pi_1 f_1(\mathbf{z}) - \pi_2 f_2(\mathbf{z})) d\mathbf{z} \\
&= \int_{B_1(\mathbf{z})} \pi_1 f_1(\mathbf{z}) (1 - e^{-Q(\mathbf{z})/2}) d\mathbf{z} - \int_{\widehat{B}_1(\mathbf{z})} \pi_1 f_1(\mathbf{z}) (1 - e^{-Q(\mathbf{z})/2}) d\mathbf{z} \\
&\leq \int_{B_1(\mathbf{z}) \cap \widehat{B}_1^C(\mathbf{z})} f_1(\mathbf{z}) (1 - e^{-Q(\mathbf{z})/2}) d\mathbf{z} \\
&= \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} [(1 - e^{-Q(\mathbf{z})/2}) \mathbb{1}\{0 < Q(\mathbf{z}) \leq M(\mathbf{z})\}] \\
&= \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} [(1 - e^{-Q(\mathbf{z})/2}) \mathbb{1}\{0 < Q(\mathbf{z}) \leq M(\mathbf{z})\} \cdot \mathbb{1}\{M(\mathbf{z}) < M \log n \cdot s_{p,r}\}] \\
&\quad + \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} [(1 - e^{-Q(\mathbf{z})/2}) \mathbb{1}\{0 < Q(\mathbf{z}) \leq M(\mathbf{z})\} \cdot \mathbb{1}\{M(\mathbf{z}) \geq M \log n \cdot s_{p,r}\}] \\
&\leq \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} [(1 - e^{-Q(\mathbf{z})/2}) \mathbb{1}\{0 < Q(\mathbf{z}) \leq M(\mathbf{z})\} \cdot \mathbb{1}\{M(\mathbf{z}) < M \log n \cdot s_{p,r}\}] \\
&\quad + \mathbb{P}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} (M(\mathbf{z}) \geq M \log n \cdot s_{p,r}). \tag{3.21}
\end{aligned}$$

We now consider the last two terms in (3.21), separately. First, we show that there exist some constant $c, C > 0$, such that for any $M > 0$,

$$\mathbb{P}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} (M(\mathbf{z}) > M s_{p,r}) \leq e^{-cM} + Cp^{-1}. \tag{3.22}$$

To bound the tail probability of $M(\mathbf{z}) = Q(\mathbf{z}) - \widehat{Q}(\mathbf{z})$, we split $Q(\mathbf{z})$ so as $\widehat{Q}(\mathbf{z})$ into

four parts as following,

$$\begin{aligned}
Q(\mathbf{z}) &= (\mathbf{z} - \boldsymbol{\mu}^{(1)})^T \Omega (\mathbf{z} - \boldsymbol{\mu}^{(1)}) - 2\boldsymbol{\beta}^T (\mathbf{z} - \boldsymbol{\mu}^{(1)}) + \boldsymbol{\beta}^T (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}) \\
&\quad - \log (|\Omega \Sigma^{(1)} + I_p|) + 2 \log \left(\frac{\pi_1}{\pi_2} \right) \\
&= \underbrace{\boldsymbol{\beta}^T (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})}_{\text{I}} - \log (|\Omega \Sigma^{(1)} + I_p|) - \underbrace{2\boldsymbol{\beta}^T (\mathbf{z} - \boldsymbol{\mu}^{(1)})}_{\text{II}} \\
&\quad + \underbrace{(\mathbf{z} - \boldsymbol{\mu}^{(1)})^T \Omega (\mathbf{z} - \boldsymbol{\mu}^{(1)})}_{\text{III}} + \underbrace{2 \log \left(\frac{\pi_1}{\pi_2} \right)}_{\text{IV}}, \tag{3.23}
\end{aligned}$$

$$\begin{aligned}
\widehat{Q}(\mathbf{z}) &= \underbrace{\widehat{\boldsymbol{\beta}}^T (\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)})}_{\widehat{\text{I}}} - \log (|\widehat{\Omega} \widehat{\Sigma}^{(1)} + I_p|) - \underbrace{2\widehat{\boldsymbol{\beta}}^T (\mathbf{z} - \widehat{\boldsymbol{\mu}}^{(1)})}_{\widehat{\text{II}}} \\
&\quad + \underbrace{(\mathbf{z} - \widehat{\boldsymbol{\mu}}^{(1)})^T \widehat{\Omega} (\mathbf{z} - \widehat{\boldsymbol{\mu}}^{(1)})}_{\widehat{\text{III}}} + \underbrace{2 \log \left(\frac{\widehat{\pi}_1}{\widehat{\pi}_2} \right)}_{\widehat{\text{IV}}}. \tag{3.24}
\end{aligned}$$

Part I: We first consider $(\text{I} - \widehat{\text{I}})$.

$$\begin{aligned}
&\left| \boldsymbol{\beta}^T (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}) - \widehat{\boldsymbol{\beta}}^T (\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)}) \right| \\
&\leq \left| \widehat{\boldsymbol{\beta}}^T (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)} + \widehat{\boldsymbol{\mu}}^{(1)}) \right| + \left| (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}) \right| \quad (\text{by triangle inequality}) \\
&\leq \left\| \widehat{\boldsymbol{\beta}} \right\|_1 \cdot \left\| \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)} + \widehat{\boldsymbol{\mu}}^{(1)} \right\|_\infty + \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_2 \left\| \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} \right\|_2 \quad (\text{by Holder's inequality}) \\
&\lesssim \left\| \boldsymbol{\beta} \right\|_1 \cdot \left\| \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)} + \widehat{\boldsymbol{\mu}}^{(1)} \right\|_\infty + \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_2 \left\| \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} \right\|_2 \\
&\lesssim \sqrt{p} \left\| \boldsymbol{\beta} \right\|_2 \cdot \left\| \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)} + \widehat{\boldsymbol{\mu}}^{(1)} \right\|_\infty + \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_2 \left\| \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} \right\|_2 \\
&\lesssim_{S_{p,r}}.
\end{aligned}$$

Notice that $(\Omega\Sigma^{(1)} + I_p)^{-1} = \Theta_0^{(1)}\Sigma^{(2)} = (\Theta_0^{(2)} - \Omega)\Sigma^{(2)} = I_p - \Omega\Sigma^{(2)}$, then

$$\begin{aligned}
& \log |\Omega\Sigma^{(1)} + I_p| - \log |\hat{\Omega}\hat{\Sigma}^{(1)} + I_p| \\
& \leq \text{tr} \left((\Omega\Sigma^{(1)} + I_p)^{-1} (\hat{\Omega}\hat{\Sigma}^{(1)} - \Omega\Sigma^{(1)}) \right) \quad (\text{by Lemma 3.8.2}) \\
& = \text{tr} \left((-\Omega\Sigma^{(2)} + I_p) (\hat{\Omega}\hat{\Sigma}^{(1)} - \Omega\Sigma^{(1)}) \right) \\
& = \text{tr} \left((-\Omega\Sigma^{(2)}) (\hat{\Omega}\hat{\Sigma}^{(1)} - \Omega\Sigma^{(1)}) \right) + \text{tr} \left(\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)} \right) \\
& \leq \|\Omega\Sigma^{(2)}\|_F \cdot \|\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)}\|_F + \text{tr} \left(\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)} \right) \quad (\text{by Cauchy-Schwarz inequality}) \\
& \leq \|\Omega\|_F \|\Sigma^{(2)}\|_{op} \cdot \|\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)}\|_F + \text{tr} \left(\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)} \right) \quad (\text{by Lemma 3.8.1}) \\
& \leq \|\Omega\|_F \|\Sigma^{(2)}\|_{op} \cdot \|\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)}\|_F + \text{tr} \left(\hat{\Omega}\hat{\Sigma}^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)} \right) + \text{tr} \left(\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)} \right). \quad (3.25)
\end{aligned}$$

In addition, we separately discuss these three components in (3.25)

$$\begin{aligned}
\|\Omega\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)}\|_F & \leq \|\Omega\Sigma^{(1)} - \hat{\Omega}\Sigma^{(1)}\|_F + \|\hat{\Omega}(\Sigma^{(1)} - \hat{\Sigma}^{(1)})\|_F \\
& \leq \|\Sigma^{(1)}\|_{op} \|\hat{\Omega} - \Omega\|_F + \|\hat{\Omega}\|_F \|\Sigma^{(1)} - \hat{\Sigma}^{(1)}\|_{op} \quad (\text{by Lemma 3.8.1}) \\
& \lesssim s_{p,r} + \|\Omega\|_F \|\Sigma^{(1)} - \hat{\Sigma}^{(1)}\|_{op} \\
& \lesssim s_{p,r}.
\end{aligned}$$

For the second term,

$$\text{tr} \left(\hat{\Omega}\Sigma^{(1)} - \hat{\Omega}\hat{\Sigma}^{(1)} \right) \leq \|\hat{\Omega}\|_F \|\Sigma^{(1)} - \hat{\Sigma}^{(1)}\|_F \lesssim \|\Omega\|_F \|\Sigma^{(1)} - \hat{\Sigma}^{(1)}\|_F \lesssim s_{p,r}.$$

The remaining term $\text{tr} \left(\Omega\Sigma^{(1)} - \hat{\Omega}\Sigma^{(1)} \right)$ will be discussed when we derive the distribution of the term involving $\mathbf{z}$ in part III. Therefore, refer to lemma 7.4 in [Cai & Zhang \(2021\)](#), there exists a constant M_1 such that

$$\mathbb{P} \left(|\mathbf{I} - \hat{\mathbf{I}}| > M_1 s_{p,r} \right) \leq p^{-1}. \quad (3.26)$$

Part II: Next, for the second term $(\mathbf{II} - \hat{\mathbf{II}})$, notice that

$$\hat{\boldsymbol{\beta}}^T (\mathbf{z} - \hat{\boldsymbol{\mu}}^{(1)}) - \boldsymbol{\beta}^T (\mathbf{z} - \boldsymbol{\mu}^{(1)}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\mathbf{z} - \boldsymbol{\mu}^{(1)}) + \hat{\boldsymbol{\beta}}^T (\boldsymbol{\mu}^{(1)} - \hat{\boldsymbol{\mu}}^{(1)}).$$

For the term $\widehat{\boldsymbol{\beta}}^T (\boldsymbol{\mu}^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)})$, we have discussed in the part I. Then, when $\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})$, we have

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\mathbf{z} - \boldsymbol{\mu}^{(1)}) \sim N\left(\mathbf{0}, (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \Sigma^{(1)} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right).$$

Additionally,

$$\left| (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \Sigma^{(1)} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right| \leq \|\Sigma^{(1)}\|_{op} \cdot \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2 \leq s_{p,r}^2.$$

Therefore, for some c' , $M_2 > 0$ for tail distribution,

$$\mathbb{P}\left(\left|\text{II} - \widehat{\text{II}}\right| > M_2 s_{p,r}\right) = \mathbb{P}\left(\left|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\mathbf{z} - \boldsymbol{\mu}^{(1)})\right| > M_2 s_{p,r}\right) \leq e^{-c' M_2^2}. \quad (3.27)$$

Part III: For the term $(\text{III} - \widehat{\text{III}})$ and the last term in (S.32) involving $\mathbf{z}$, where $\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})$, we have

$$\begin{aligned} & (\mathbf{z} - \boldsymbol{\mu}^{(1)})^T \Omega (\mathbf{z} - \boldsymbol{\mu}^{(1)}) - (\mathbf{z} - \boldsymbol{\mu}^{(1)})^T \widehat{\Omega} (\mathbf{z} - \boldsymbol{\mu}^{(1)}) - \text{tr}(\Omega \Sigma^{(1)} - \widehat{\Omega} \Sigma^{(1)}) \\ &= (\mathbf{z} - \boldsymbol{\mu}^{(1)})^T (\Omega - \widehat{\Omega}) (\mathbf{z} - \boldsymbol{\mu}^{(1)}) - \text{tr}(\Omega \Sigma^{(1)} - \widehat{\Omega} \Sigma^{(1)}) \\ &= \mathbf{z}_0^T (\Sigma^{(1)})^{1/2} (\widehat{\Omega} - \Omega) (\Sigma^{(1)})^{1/2} \mathbf{z}_0 - \text{tr}((\Sigma^{(1)})^{1/2} (\widehat{\Omega} - \Omega) (\Sigma^{(1)})^{1/2}) \\ &= \sum_{i=1}^p \lambda_i (\mathbf{z}_{0i}^2 - 1), \end{aligned} \quad (3.28)$$

where λ_i are the eigenvalues of $(\Sigma^{(1)})^{1/2} (\widehat{\Omega} - \Omega) (\Sigma^{(1)})^{1/2}$ and $\mathbf{z}_0$ is standard normal. In fact, the combination of variables from a normal distribution is still a variable from the normal distribution forms the final step. Furthermore,

$$\begin{aligned} \sqrt{\sum_{i=1}^p \lambda_i^2} &= \left\| (\Sigma^{(1)})^{1/2} (\widehat{\Omega} - \Omega) (\Sigma^{(1)})^{1/2} \right\|_F \leq \|\Sigma^{(1)}\|_{op} \|\widehat{\Omega} - \Omega\|_F \lesssim s_{p,r}, \\ \max_i |\lambda_i| &\leq \left\| (\Sigma^{(1)})^{1/2} (\widehat{\Omega} - \Omega) (\Sigma^{(1)})^{1/2} \right\|_2 \leq \|\Sigma^{(1)}\|_{op} \|\widehat{\Omega} - \Omega\|_{op} \lesssim s_{p,r}. \end{aligned}$$

Then, by Bernstein type inequality for sub-exponential random variable (see chapter 2 in [Wainwright \(2019\)](#) for detail), regarding (3.28), for some c'' , M_3 , $C_1 > 0$, we have,

$$\mathbb{P}\left(\left|\text{III} - \widehat{\text{III}}\right| > M_3 s_{p,r}\right) = \mathbb{P}\left(\left|\sum_{i=1}^p \lambda_i (\mathbf{z}_{0i}^2 - 1)\right| \geq M_3 s_{p,r}\right) \leq e^{-c'' M_3} + C_1 p^{-1}. \quad (3.29)$$

Part IV: Then we consider $(\text{IV} - \widehat{\text{IV}})$. Note that,

$$\left| 2 \log \left(\frac{\pi_1}{\pi_2} \right) - 2 \log \left(\frac{\widehat{\pi}_1}{\widehat{\pi}_2} \right) \right| \lesssim |\widehat{\pi}_1 - \pi_1| + |\widehat{\pi}_2 - \pi_2|.$$

And by Hoeffding inequality, for $k = 1, 2$, for constant $c''' > 0$, such that

$$\mathbb{P}(|\widehat{\pi}_k - \pi_k| > t) \lesssim e^{-c''' \cdot nt^2}.$$

Therefore, we have some constant $c''', M_4 > 0$,

$$\begin{aligned} & \mathbb{P} \left(\left| \text{IV} - \widehat{\text{IV}} \right| > M_4 \sqrt{\frac{1}{n}} \right) \\ &= \mathbb{P} \left(\left| 2 \log \left(\frac{\pi_1}{\pi_2} \right) - 2 \log \left(\frac{\widehat{\pi}_1}{\widehat{\pi}_2} \right) \right| > M_4 \sqrt{\frac{1}{n}} \right) \lesssim e^{-c''' M_4^2}. \end{aligned} \quad (3.30)$$

By combining (3.26) (3.27) (3.29) and (3.30), there exists $c, C > 0$, such that for $M > 0$,

$$\mathbb{P}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})}(|M(\mathbf{z})| > Ms_{p,r}) \leq e^{-cM} + Cp^{-1}.$$

Since $M(\mathbf{z})$ is the excess risk and is non-negative, then we complete the proof of (3.22).

Finally, (3.21) follows,

$$\begin{aligned} & R_\theta(G_{\text{pcDLQ}}) - R_\theta(G_\theta^*) \\ & \leq \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} \left[(1 - e^{-Q(\mathbf{z})/2}) \mathbb{1}\{0 < Q(\mathbf{z}) \leq M(\mathbf{z})\} \cdot \mathbb{1}\{M(\mathbf{z}) < M \log ns_{p,r}\} \right] \\ & + \mathbb{P}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} (M(\mathbf{z}) \geq M \log ns_{p,r}) \\ & \lesssim \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} \left[(1 - e^{-Q(\mathbf{z})/2}) \mathbb{1}\{0 < Q(\mathbf{z}) \leq M(\mathbf{z})\} \cdot \mathbb{1}\{M(\mathbf{z}) < M \log ns_{p,r}\} \right] \\ & + n^{-1} + p^{-1} \\ & \lesssim \log n \cdot s_{p,r} \cdot \mathbb{E}_{\mathbf{z} \sim N_p(\boldsymbol{\mu}^{(1)}, \Sigma^{(1)})} [\mathbb{1}\{0 < Q(\mathbf{z}) < M \log ns_{p,r}\}] + n^{-1} + p^{-1} \\ & \lesssim \log^2 n \cdot s_{p,r}^2, \end{aligned}$$

where the last inequality uses the assumption (A.6) that $\sup_{|x| < \delta} f_{Q, \boldsymbol{\theta}}(x) < M_3$.

□

Chapter 4

Control the false discovery rate with complex auxiliary information

4.1 Introduction

High-throughput genetic technology has progressed rapidly in recent years. For example, RNA sequencing (RNA-seq) has emerged as the next-generation high-throughput transcriptome technology capable of measuring the expression levels of tens of thousands of genes simultaneously. The high-throughput studies can be used to identify differentially expressed (DE) genes, which is important for understanding disease mechanisms. Typically, DE genes are identified through hypothesis testing for each gene, resulting in testing hundreds to thousands of hypotheses simultaneously. In this multiple testing problem, one common error metric is the false discovery rate (FDR), first introduced by [Benjamini & Hochberg \(1995\)](#), which represents the expected proportion of false positives among all rejected null hypotheses.

In many multiple testing applications, auxiliary information from related studies is often available in addition to a simple list of the p -values ([Ferkingstad et al. 2008](#), [Fortney et al. 2015](#)). For example, in the analysis of RNA-seq data, test statistics from related microarray studies can serve as valuable auxiliary information. It is widely recognized that incorporating such auxiliary information can enhance the power of FDR controlling procedures, but we also need to balance power improvement with the risk of false positives. Ensuring reliable results by controlling FDR in finite samples is particularly important, especially given that the number of hypotheses tested can vary significantly across different analyses. For example, RNA-seq datasets typically involve tens of thousands of genes, while

some protein experiments may only measure a few dozen proteins. As such, in this paper, we focus on methods for covariate-assisted multiple testing that guarantee finite-sample FDR control. Leveraging auxiliary information, [Ignatiadis & Huber \(2021\)](#) proposed a method that includes a single covariate, while [Lei & Fithian \(2018\)](#) introduced an adaptive p -value thresholding procedure (AdaPT) that incorporates multiple covariates. Both above methods ensure finite-sample FDR control and are p -value based.

However, when transforming the original statistics into the p -values, the signs of the statistics are typically ignored, symmetrically treating both positive and negative deviations between the test and control groups. The signs often contain valuable directional information. For example, in RNA-seq data analysis, a negative sign could indicate gene downregulation, while a positive sign could suggest upregulation. To retain directional information, [Tian et al. \(2021\)](#) introduced the signed p -values, which maintain a one-to-one correspondence with the original statistics. They developed the signed-knockoff procedure, which generates knockoffs of the signed p -values. The signed-knockoff procedure presents unordered pairs of the signed p -values and their knockoffs at the beginning, and the true identities are gradually revealed throughout the process. The number of knockoffs in the rejection region then serves to estimate the number of false discoveries. Similarly, [Chao & Fithian \(2021\)](#) used the z -value ([Sun & Cai 2007](#)), and proposed the AdaPT-GMM_g method, while [Leung & Sun \(2022\)](#) introduced an alternative z -value-based procedure called ZAP to retain directional information.

While these existing methods ensure finite-sample FDR control, their power can be limited, especially when dealing with complex interaction effects between covariates. In this chapter, we extend the covariate and direction adaptive knockoff procedure (Codak) by [Tian \(2020\)](#) to incorporate generic covariates. We show that the proposed methods control the FDR and exhibit competitive power performance in numerical studies.

The rest of this chapter is organized as follows. In [Section 4.2](#), we review the Codak procedure. [Section 4.3](#) details the implementation of Codak. In [Section 4.4](#), we compare Codak with existing methods through simulation studies. [Section 4.5](#) applies Codak to a gene expression data analysis. We defer some technical details to [Section 4.6](#).

4.2 Covariate and direction adaptive knockoff procedure (Codak)

Codak is an iterative framework for multiple testing while incorporating covariate and directional information. Codak starts with a relatively large rejection set and then sequentially accepts null hypotheses, one at a time, while estimating the FDR at each step. If the FDR estimate falls below the target level α , the procedure stops and rejects all remaining hypotheses in the rejection set. Otherwise, the procedure continues. Any data-adaptive method can be used to select the next null hypothesis to accept. Mostly, we choose the one with the largest local FDR, the posterior probability that the null hypothesis is true, given the test statistic and auxiliary information. In the ideal scenario where the data-generating model is fully known, the Bayesian decision-theoretic framework establishes that the local FDR is the optimal ranking statistic for maximizing power while controlling the FDR (Sun & Cai 2007, Lei & Fithian 2018). In practice, the local FDR is estimated with respect to an empirical Bayesian model called the conditional two-group model, which we will discuss in Section 4.3.

Now, we formalize the Codak procedure. Consider testing n null hypotheses simultaneously: H_{0i} , $i = 1, \dots, n$. For the i th null hypothesis H_{0i} , let t_i denote the test statistic, p_i the corresponding p -value, and $\mathbf{x}_i$ the vector of auxiliary variables. To incorporate the directional information, we define the signed p -value of the null hypothesis H_{0i} as

$$q_i = \text{sign}(t_i)(1 - p_i),$$

and the knockoff of a signed p -value q_i as

$$\tilde{q}_i = \text{sign}(q_i) - q_i.$$

Let $q_i^* = q_i \wedge \tilde{q}_i$ be the minimum of q_i and $\tilde{q}_i$, which can be viewed as a masked version of the signed p -value q_i .

At the beginning of the Codak procedure, all the signed p -values are masked, and the available information is represented as the filtration:

$$\mathcal{F}_0 = \sigma(\{q_i^*\}_{i=1}^n, \{\mathbf{x}_i\}_{i=1}^n).$$

Through the process, the true identity of a masked signed p -value is revealed after its corresponding null hypothesis is accepted. Assuming that k null hypotheses have been accepted, let R_k represent the current rejection set, and the available information to select the next null hypothesis to accept is contained in the filtration:

$$\mathcal{F}_k = \sigma(\mathcal{F}_0, \{b_i\}_{i \notin R_k}),$$

where $b_i = I(|q_i| > 1/2)$, used to determine the true identity of a q_i^* . We estimate the FDR as follows,

$$\widehat{\text{FDR}}_k = \frac{1 + \#\{i : i \in R_k \& |q_i| < 1/2\}}{\#\{i : i \in R_k \& |q_i| > 1/2\} \vee 1}.$$

Since the signed p -values follow a uniform distribution, the numerator of $\widehat{\text{FDR}}_k$ serves as a slightly conservative estimator for the number of false rejections.

In conclusion, the Codak procedure (Tian 2020) is defined as follows,

Algorithm 2: Covariate and direction adaptive knockoff procedure (Codak)

- Step1.** Set $k = 0$. Set the $\mathcal{F}_0$ -measurable initial rejection set $R_0 \subseteq \{1, 2, \dots, n\}$.
Step2. Choose an index $i \in R_k$ and remove it from R_k . Set $R_{k+1} = R_k / \{i\}$. The index choice must be $\mathcal{F}_k$ -measurable. Increase k by 1.
Step3. If $\widehat{\text{FDR}}_k \leq \alpha$ or $R_k = \emptyset$, stop and reject any H_{i_0} with $i \in R_k$ and $|q_i| > 1/2$. Otherwise, return to Step 2.
-

In our implementation, we set the initial rejection set $R_0 = \{i : |q_i| < 0.3 \text{ or } |q_i| > 0.7\}$. Therefore, in all cases, null hypotheses with the p -values exceeding 0.3 will be accepted. Notice that Codak can be viewed as a specific instance of the AdaPT framework, and Theorem 4.1 in Tian (2020) shows that Codak controls the FDR in finite samples.

4.3 Implementation

In Step 2 of the Codak procedure, we only require the choice of the next null hypothesis to accept to be $\mathcal{F}_k$ -measurable but leave the details unspecified. As we discussed before, mostly in practice, we estimate the local FDR with the limited information available in $\mathcal{F}_k$ under a conditional two-group model, and select the next hypothesis with the largest local FDR.

4.3.1 Two-group model

Suppose that all the signed p -values are observable, consider the following two-group model. For $i = 1, \dots, n$,

$$\begin{cases} H_i \mid \mathbf{x}_i \sim \text{Bernoulli} \{ \pi_0(\mathbf{x}_i) \}, \\ q_i \mid H_i \sim H_i f_0 + (1 - H_i) f_{1\mathbf{x}_i}, \end{cases}$$

where H_i is the true null indicator, f_0 is the pdf of $\text{Uniform}(-1, 1)$, and $f_{1\mathbf{x}_i} = f_1(\cdot | \mathbf{x}_i)$ is the alternative pdf conditioning on $\mathbf{x}_i$. Both the probability of $H_i = 1$, $\pi_0(\mathbf{x}_i)$, and $f_{1\mathbf{x}_i}$ can depend on the covariates $\mathbf{x}_i$ and need to be estimated. Under the two-group model, the overall density function f conditional on $\mathbf{x}_i$ is

$$f(q_i | \mathbf{x}_i) = \pi_0(\mathbf{x}_i) / 2 + \{1 - \pi_0(\mathbf{x}_i)\} f_1(q_i | \mathbf{x}_i),$$

and the local FDR can be expressed as

$$\text{Lfdr}_i = \pi_0(\mathbf{x}_i) / \{2f(q_i | \mathbf{x}_i)\}.$$

The following parametric model of $f_{1\mathbf{x}_i}$ is used:

$$f_1(q_i | \mathbf{x}_i) = \lambda(\mathbf{x}_i) \frac{\alpha(\mathbf{x}_i)}{2} \left(\frac{q_i + 1}{2} \right)^{\alpha(\mathbf{x}_i) - 1} + \{1 - \lambda(\mathbf{x}_i)\} \frac{\beta(\mathbf{x}_i)}{2} \left(\frac{1 - q_i}{2} \right)^{\beta(\mathbf{x}_i) - 1}.$$

We see that $f_{1\mathbf{x}_i}$ is a mixture of two components with transformed beta distributions, where $\lambda(\mathbf{x}_i)$ is the mixing proportion. We denote the two components as the α component and the β component, respectively, since $\alpha(\mathbf{x}_i)$ and $\beta(\mathbf{x}_i)$ are the shape parameter functions in the two components.

As both $\pi_0(\mathbf{x}_i)$ and $\lambda(\mathbf{x}_i)$ represent proportions, we model them through the logistic regression:

$$\begin{aligned} \log \left\{ \frac{\pi_0(\mathbf{x}_i)}{1 - \pi_0(\mathbf{x}_i)} \right\} &= \boldsymbol{\theta}_\pi^\top \boldsymbol{\Phi}_\pi(\mathbf{x}_i), \\ \log \left\{ \frac{\lambda(\mathbf{x}_i)}{1 - \lambda(\mathbf{x}_i)} \right\} &= \boldsymbol{\theta}_\lambda^\top \boldsymbol{\Phi}_\lambda(\mathbf{x}_i), \end{aligned}$$

where $\boldsymbol{\theta}_\pi$ and $\boldsymbol{\theta}_\lambda$ are two parameter vectors, and $\boldsymbol{\Phi}_\pi(\mathbf{x}_i) = (\Phi_{\pi 1}(\mathbf{x}_i), \dots, \Phi_{\pi m_\pi}(\mathbf{x}_i))^\top$ and $\boldsymbol{\Phi}_\lambda(\mathbf{x}_i) = (\Phi_{\lambda 1}(\mathbf{x}_i), \dots, \Phi_{\lambda m_\lambda}(\mathbf{x}_i))^\top$ are two featurizations of $\mathbf{x}_i$. For computational consideration, we model $\alpha(\mathbf{x}_i)$ and $\beta(\mathbf{x}_i)$ as

$$\begin{aligned} -\log \{\alpha(\mathbf{x}_i)\} &= \boldsymbol{\theta}_\alpha^\top \boldsymbol{\Phi}_\alpha(\mathbf{x}_i), \\ -\log \{\beta(\mathbf{x}_i)\} &= \boldsymbol{\theta}_\beta^\top \boldsymbol{\Phi}_\beta(\mathbf{x}_i), \end{aligned}$$

where $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\beta$ are two parameter vectors, and $\boldsymbol{\Phi}_\alpha(\mathbf{x}_i)$ and $\boldsymbol{\Phi}_\beta(\mathbf{x}_i)$ are two featurizations.

After we estimate these parameter vectors, for any $i \in R_k$, we use the following local FDR estimate:

$$\widehat{\text{Lfdr}}_i = \frac{\widehat{\pi}_0(\mathbf{x}_i) + \{1 - \widehat{\pi}_0(\mathbf{x}_i)\} \left[\widehat{\lambda}(\mathbf{x}_i) \widehat{\alpha}(\mathbf{x}_i) + \{1 - \widehat{\lambda}(\mathbf{x}_i)\} \widehat{\beta}(\mathbf{x}_i) \right]}{\widehat{\pi}_0(\mathbf{x}_i) + \{1 - \widehat{\pi}_0(\mathbf{x}_i)\} \widehat{f}_1(q_i^*; \mathbf{x}_i) + \{1 - \widehat{\pi}_0(\mathbf{x}_i)\} \widehat{f}_1(\widetilde{q}_i^*; \mathbf{x}_i)}.$$

More details about this expression can be found in Section 4.3.3 in [Tian \(2020\)](#).

Next, we focus on how to estimate these parameter vectors using an EM-algorithm.

4.3.2 $\mathcal{F}_k$ -measurable EM-algorithm

After k hypotheses have been accepted, we are allowed to estimate the parameters only based on the information in $\mathcal{F}_k$, which are $\{q_i^*\}_{i=1}^n, \{\mathbf{x}_i\}_{i=1}^n$ and $\{b_i = I(|q_i| > 1/2)\}_{i \notin R_k}$. For each $i \in R_k$, we only observe q_i^* , and it can be shown that the density function of q_i^* is

$$\begin{aligned} g(q_i^* | \mathbf{x}_i) &= f(q_i^* | \mathbf{x}_i) + f(\tilde{q}_i^* | \mathbf{x}_i) \\ &= \pi_0(\mathbf{x}_i) + \{1 - \pi_0(\mathbf{x}_i)\} f_1(q_i^* | \mathbf{x}_i) + \{1 - \pi_0(\mathbf{x}_i)\} f_1(\tilde{q}_i^* | \mathbf{x}_i). \end{aligned}$$

On the other hand, for each $i \notin R_k$, we observe q_i through q_i^* and b_i . As a result, the log-likelihood function of $\{\{q_i^*\}_{i=1}^n, \{b_i\}_{i \notin R_k}\}$ is

$$\begin{aligned} l^{(k)}(\Theta | \mathcal{F}_k) &= \sum_{i \notin R_k} \log f(q_i | \mathbf{x}_i) + \sum_{i \in R_k} \log g(q_i^* | \mathbf{x}_i) \\ &= \sum_{i \notin R_k} \log [\pi_0(\mathbf{x}_i)/2 + \{1 - \pi_0(\mathbf{x}_i)\} f_1(q_i | \mathbf{x}_i)] \\ &\quad + \sum_{i \in R_k} \log [\pi_0(\mathbf{x}_i) + \{1 - \pi_0(\mathbf{x}_i)\} f_1(q_i^* | \mathbf{x}_i) + \{1 - \pi_0(\mathbf{x}_i)\} f_1(\tilde{q}_i^* | \mathbf{x}_i)]. \end{aligned}$$

where $\Theta = (\boldsymbol{\theta}_\pi, \boldsymbol{\theta}_\lambda, \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)$ is the combination of all four parameter vectors.

Maximizing the log-likelihood function $l^{(k)}$ directly with respect to Θ is computationally challenging, and we develop an EM-algorithm. In the E-step, based on the current parameter vectors Θ , we compute the mixture component probabilities, such as the probability of being null and the probability of being α component for the revealed. In the M-step, these probabilities are used as weights, and we estimate Θ through different regression methods. We consider two options: a generalized additive model and gradient-boosted trees, referring to them as Codak-gam and Codak-xg, respectively. In contrast, the predictive function in [Leung & Sun \(2022\)](#) is implemented as an additive model, which ignores the potential interaction effects among covariates.

4.3.3 Codak-gam

Similar to AdaPT ([Lei & Fithian 2018](#)), we utilize the `gam` function from the `mgcv` package ([Wood 2017](#)) to fit a generalized additive model (gam) during each iteration of the M-step. We observe that in the EM-algorithm, the only change between iterations is the updated weights, while the rest of the gam setup remains unchanged. Unlike AdaPT, which reinitializes the model from scratch at each iteration, we set up the gam model just

once and reuse it in subsequent iterations with the updated weights. We also observe that the estimated smoothing parameters change only slightly between iterations. Therefore, we calculate these smoothing parameters in the first iteration and then reuse them for the remaining iterations. These optimizations enable our procedure to execute in nearly half the time compared to reinitializing and recalculating everything in each iteration. We defer the implementation details to Section 4.6.1.

4.3.4 Codak-xg

Similar to Yurko et al. (2020), we employ gradient-boosted trees through the XGboost library (Chen & Guestrin 2016). XGboost constructs a flexible predictive function as a weighted sum of numerous simple trees, optimized via a gradient descent procedure that minimizes a specific objective function. Codak-xg leverages several key advantages of XGBoost, including invariance to monotonic transformations of variables and automatic incorporation of important variable interactions. Therefore, Codak-xg is particularly effective when dealing with large numbers of covariates with unknown functional forms. To further maximize power gains and minimize computational complexity, we incorporate incremental learning and consider a “coarse + fine” tuning procedure.

Incremental learning

Following the initial iteration of the EM-algorithm, XGBoost is enabled to incrementally enhance its model by setting the argument `xgb_model` to the model obtained from the preceding iteration, thereby avoiding the need to start from scratch. In each follow-up iteration, the training model inherits the decision trees from previous iterations and constructs further decision trees based on them. Consequently, the training model adjusts its internal structure and parameters based on the updated weights, which is the only difference between iterations, similar to Codak-gam.

Parameter tuning

To maximize the power of gradient-boosted trees, we tune parameters including the learning rate (`eta`), the maximum depth of each tree (`max_depth`), and the number of iterations (`nrounds`). We use the cross-validation (CV) approach, inspired by Yurko et al. (2020). Technically, we could tune these parameters for modeling $\pi_0(\mathbf{x}_i)$, $\lambda(\mathbf{x}_i)$, $\alpha(\mathbf{x}_i)$ and $\beta(\mathbf{x}_i)$ separately. However, we only consider matching settings for all models (i.e., all models

have the same tuning parameters) to reduce computational complexity. Additionally, incorporating the CV procedure is computationally expensive, making it difficult to apply at each model fitting step. Instead, we perform the CV procedure once at the beginning, using a large and rough parameter space. Then, through the process, as the estimated FDR drops just below 0.15, we switch to a narrower and more precise parameter space and conduct another CV procedure. That is what we call a “coarse + fine” tuning procedure. This method ensures that the initial coarse tuning captures the general behavior of the model while the subsequent fine tuning adjusts the parameters more precisely to maximize predictive accuracy. We defer the details about the potential parameter choices to Section 4.6.2.

4.4 Simulation

We evaluate the realized FDR levels and the power of various procedures incorporating covariates. The results presented in this section are based on 400 replications, with a default target FDR level set at $\alpha = 0.1$.

4.4.1 Candidate procedures

We consider the following procedures:

- Codak-gam: our proposed procedure using gam.
- Codak-xg: our proposed procedure using XGBoost.
- ZAP: the adaptive z -value based procedure with side information by [Leung & Sun \(2022\)](#).
- AdaPT: the adaptive p -value thresholding procedure by [Lei & Fithian \(2018\)](#).
- AdaPT-GMM_g: the refined AdaPT method under a Gaussian mixture model using a two-layer neural network by [Chao & Fithian \(2021\)](#).

For the AdaPT, we use `adapt_gam` function from R package `adaptMT`. ZAP is implemented by `ZAP.finite` function from R package `ZAP`, and we utilize `adapt_gmm` from R package `AdaPTGMM` to fit AdaPT-GMM_g. All methods are implemented with the default setting.

4.4.2 Simulation studies

We generate test statistics from the mixture model, $f(t|\mathbf{x}) = \pi_0(\mathbf{x})f_0(t) + (1 - \pi_0(\mathbf{x}))f_1(t|\mathbf{x})$, where $\mathbf{x}$ is a covariate vector. For scenarios where $\mathbf{x} = (x_1, x_2)$ is two-dimensional, as in Sets (a), (b), (c), and (d), we use $s(x_1, x_2)$ as the smoothing function in both Codak-gam and AdaPT. In cases where $\mathbf{x} = (x_1, \dots, x_k)$, $k > 2$, is multi-dimensional, as in Sets (e) and (f), we apply the additive smoothing function $s(x_1) + \dots + s(x_k)$ in Codak-gam and AdaPT.

Sets (a) and (b): two-dimensional covariate with one-sided signal

In Set (a), we generate 2500 hypotheses, where covariates $\mathbf{x}$ are from the equally spaced 50×50 grid in the area $[-100, 100] \times [-100, 100]$. Set the null probability $\pi_0(\mathbf{x}) = I(x_1^2 + x_2^2 \geq \delta)$, where we consider varying $\delta = 1500, 2500, 3500, 4500, 5500$, to evaluate the difference of null proportion. Set f_0 and f_1 as the pdfs of $N(0, 1)$ and $N(2, 1)$, respectively. This case is adopted from Example 1 in [Lei & Fithian \(2018\)](#), but we set different δ levels instead of different α levels.

In Set (b), we generate 4000 hypotheses, where covariates x_1 and x_2 are generated independently from $N(0, 1/2)$. Set the null probability $\pi_0(\mathbf{x}) = 1 - (1 + \exp(2 - x_1 - x_2))^{-1}$. Set f_0 and f_1 as the pdfs of $N(0, 1)$ and $N(\mu(\mathbf{x}), 1)$, where $\mu(\mathbf{x}) = 2\delta / (1 + \exp(-x_1 - x_2))$ with varying $\delta = 1.3, 1.5, 1.7, 1.9, 2.1$, to evaluate the difference of effect size. This case is adopted from Setup 1 in [Leung & Sun \(2022\)](#).

We report the average realized FDR levels and power in Figure 4.1. The FDR levels of AdaPT-GMM_g can be slightly inflated in some scenarios of Set (a) and (b), while all other procedures control the FDR properly. In Set (a), Codak-gam demonstrates superior power compared to other methods when δ is small and maintains stability as δ increases. Codak-xg performs similarly to AdaPT at smaller values of δ and gains an advantage as δ increases. In Set (b), Codak-gam shows power comparable to ZAP, while Codak-xg is slightly less powerful. These results indicate that, with one-sided signals, both Codak-gam and Codak-xg are competitive and can achieve power gains in certain scenarios. Conversely, ZAP and AdaPT-GMM_g underperform in Set (a), and AdaPT is less powerful in Set (b).

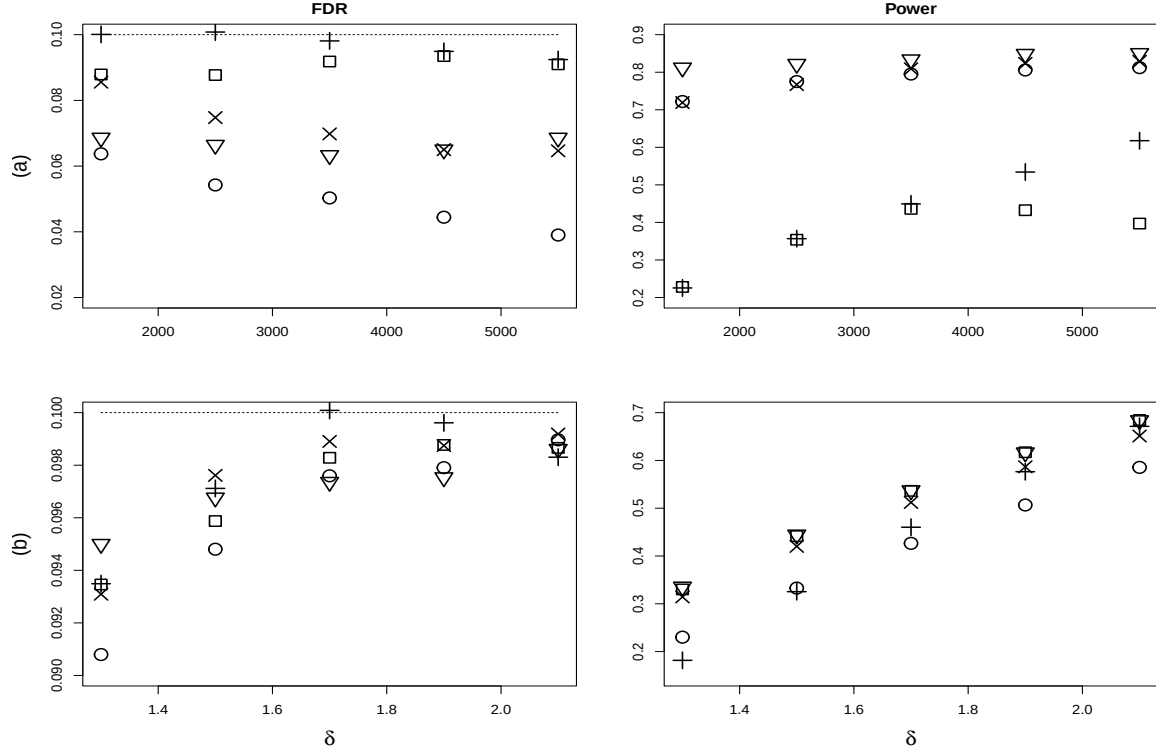


Figure 4.1: The realized FDR levels and power for sets (a) and (b); ∇ , Codak-gam; $\times$, Codak-xg; $\square$, ZAP; $\circ$, AdaPT; $+$, AdaPT-GMM $_g$. The left column shows the realized FDR levels, and the right column shows the power. Rows (a) and (b) show the results for sets (a) and (b), respectively.

Sets (c) and (d): two-dimensional covariate with two-sided signal

In Set (c), we generate 4000 hypotheses, where covariates $\mathbf{x}$ are uniformly from the area $[-1, 1] \times [-1, 1]$. Set the null probability $\pi_0(\mathbf{x}) = 0.6 + 0.3(1 - \omega)(x_1^2 + x_2^2 - 2/3) + 0.375\omega x_1 x_2$, where we consider varying $\omega = 0, 0.25, 0.5, 0.75, 1$, to evaluate the difference of joint effect. Set f_0 as the pdf of $N(0, 1)$. We consider the mixture model of t under the alternatives,

$$f_1(t|\mathbf{x}) = \pi_1(\mathbf{x})g_1(t) + (1 - \pi_1(\mathbf{x}))g_2(t),$$

where $\pi_1(\mathbf{x}) = (1 + x_1 x_2)/2$, and g_1 and g_2 are the pdfs of $N(-1.5, 1)$ and $N(3, 1)$, respectively. The terms x_1^2 and x_2^2 in the $\pi_0(\mathbf{x})$ represent the additive effect, and the term $x_1 x_2$ in both $\pi_0(\mathbf{x})$ and $\pi_1(\mathbf{x})$ represent the interaction effect between x_1 and x_2 . As ω increases,

the interaction effect contributes more to the variation of the true null probabilities. This case is adopted from Set 1 in [Liang \(2019\)](#).

Similarly, in Set (d), we generate 4000 hypotheses, with $\mathbf{x}$ and $\pi_0(\mathbf{x})$ from the same setup in Set (c). Compared with Set (c), the test statistic generated in Set (d) is now from t distribution. Under the null, set f_0 as the pdf of t distribution with 8 degrees of freedom, t_8 . Under the alternatives, we consider a similar mixture model of t as Set (c) but set $\pi_1(\mathbf{x}) = (2 + x_1 + x_2)/4$, and g_1 and g_2 as the pdfs of non-central t distributions $t_{8, -\sqrt{5/2}}$ and $t_{8, \sqrt{5/2}}$. This case is adopted from Set 2 in [Liang \(2019\)](#).

Two left panels in Figure 4.2 plot the average realized FDR levels for Sets (c) and (d), respectively, and they are all below the target level. In terms of power, as shown in the top right panel of Figure 4.2, Codak-gam still outperforms others in Set (c), with Codak-xg performing only slightly below it. As ω increases, the interaction effect becomes more dominant than the additive effect, resulting in a greater difference in power between ZAP and Codak-gam/Codak-xg. This highlights Codak's improved performance by effectively capturing interactions between covariates through gam and gradient-boosted trees. The realized power of Set (d) is plotted in the bottom right panel of Figure 4.2. The power rankings across all methods remain consistent with those in Set (c), except that ZAP outperforms AdaPT when the additive effect contributes more than the interaction effect.

Sets (e) and (f): multi-dimensional covariate with two-sided signal

In Set (e), we generate 4000 hypotheses, where covariates $\mathbf{x}$ are uniformly from the 6-dimensional area $[-1, 1]^6$. Set the null probability $\pi_0(\mathbf{x}) = 0.6 + 0.05(x_1^2 \times \text{sign}(x_5) + x_2^2 \times \text{sign}(x_6)) + 0.05(x_1x_3 \times \text{sign}(x_5) + x_2x_4 \times \text{sign}(x_6))$. Set f_0 as the pdf of $N(0, 1)$. We consider the mixture model of t under the alternatives,

$$f_1(t|\mathbf{x}) = \pi_1(\mathbf{x})g_1(t) + (1 - \pi_1(\mathbf{x}))g_2(t),$$

where $\pi_1(\mathbf{x}) = 0.5 + 0.125(x_1 \times \text{sign}(x_5) + x_2 \times \text{sign}(x_6) + x_1x_3 \times \text{sign}(x_5))$, g_1 is the pdf of $N(-1.5 - \omega, 1)$, with varying $\omega = 0, 0.25, 0.5, 0.75, 1$, and g_2 is the pdf of $N(2.5, 1)$. Other than the previous settings, as ω increases, the signal strength becomes stronger.

In this scenario, with a multi-dimensional covariate, implementing gam in Codak-gam and AdaPT using the joint smoothing function is computationally expensive, so we employ an additive smoothing function $s(x_1) + \dots + s(x_6)$ instead. However, the interaction terms present in both $\pi_0(\mathbf{x})$ and $\pi_1(\mathbf{x})$ functions, and the terms $\text{sign}(x_5)$ and $\text{sign}(x_6)$ may negatively affect the estimation of $\pi_0(\mathbf{x})$ and $\pi_1(\mathbf{x})$ functions when using additive smoothing function in Codak-gam and AdaPT.

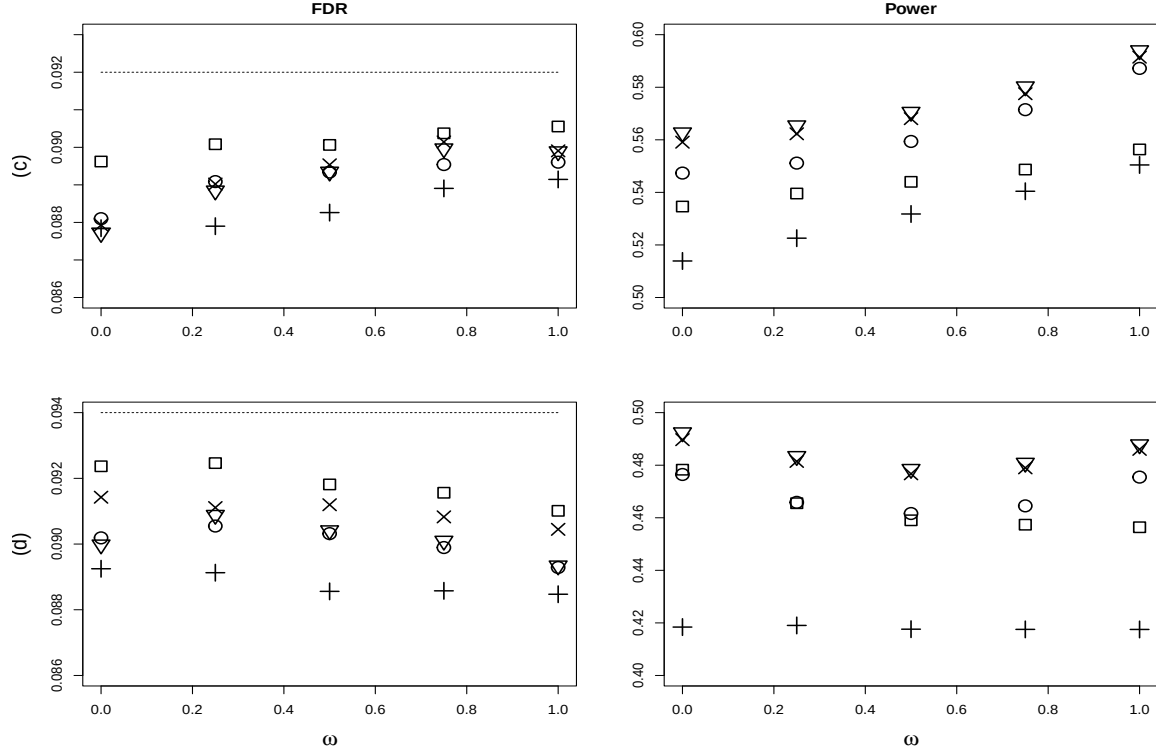


Figure 4.2: The realized FDR levels and power for sets (c) and (d); ∇ , Codak-gam; $\times$, Codak-xg; $\square$, ZAP; $\bigcirc$, AdaPT; $+$, AdaPT-GMM_g. The left column shows the realized FDR levels, and the right column shows the power. Rows (c) and (d) show the results for sets (c) and (d), respectively.

To further investigate the performance of our methods in multi-dimensional cases with interaction effects, we consider the following Set (f). Specifically, we generate 4000 hypotheses, where covariates $\mathbf{x}$ are uniformly from the 9-dimensional area $[-\pi, \pi]^9$. Set the null probability $\pi_0(\mathbf{x}) = 0.4 + (\sin(x_1) \times \sin(x_2) + \sin(x_3) \times \sin(x_4))/5$. Set f_0 as the pdf of $N(0, 1)$. Under the alternatives, we consider a similar mixture model of t as Set (e) but let $\pi_1(\mathbf{x}) = 0.5 + \sin(x_5) \times \sin(x_6) \times \sin(x_7)/3$, g_1 be the pdf of $N(-1 - \omega, 1)$, with varying $\omega = 0, 0.25, 0.5, 0.75, 1$, and g_2 be the pdf of $N(2 + \sin(x_8) \times \sin(x_9)/3, 1)$. In this case, the covariate contributes to interaction effects in the $\pi_0(\mathbf{x})$, $\pi_1(\mathbf{x})$, and g_2 functions, but does not produce any marginal effects. The additive smoothing function $s(x_1) + \dots + s(x_9)$ is used to implement gam in Codak-gam and AdaPT.

We plot the realized FDR levels in the two left panels of Figure 4.3 for these two settings, respectively. The results show that all procedures control the FDR properly in

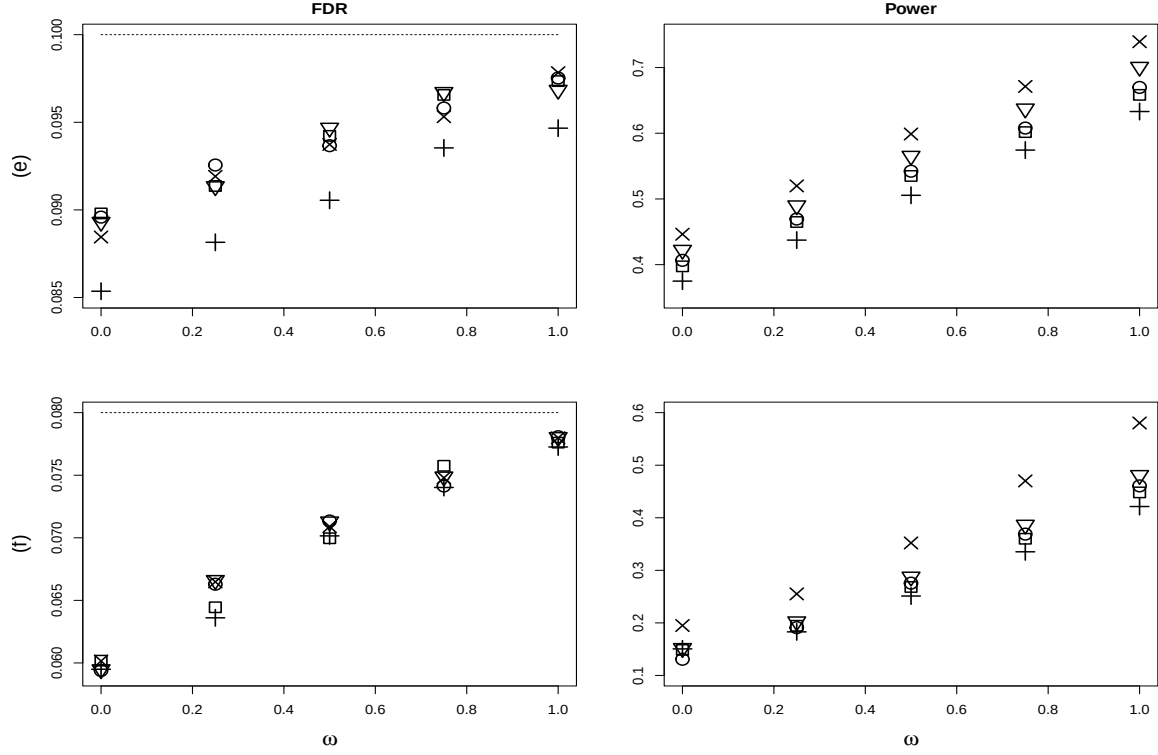


Figure 4.3: The realized FDR levels and power for sets (e) and (f); ∇ , Codak-gam; $\times$, Codak-xg; $\square$, ZAP; $\circ$, AdaPT; $+$, AdaPT-GMM_g. The left column shows the realized FDR levels, and the right column shows the power. Rows (e) and (f) show the results for sets (e) and (f), respectively.

both settings. As shown in the two right panels of Figure 4.3, the Codak-xg demonstrates superior performance over all other methods, particularly in scenarios with larger ω . These two settings illustrated the ability of Codak-xg to exhibit power gains by employing flexible predictive functions in the XGboost framework. Additionally, using additive smoothing functions results in minimal differences between Codak-gam, AdaPT, and ZAP, with Codak-gam demonstrating slightly higher power.

4.5 Real data analysis

In an RNA-seq study of psoriasis, [Jabbari et al. \(2012\)](#) compared three pairs of lesional and nonlesional skin samples, with the read count data accessible via the recount2 project ([Collado-Torres et al. 2017](#)) under accession number SRP016583. The dataset includes 45,735 genes with at least one nonzero count across all samples. We analyzed RNA-seq data using the limma-voom method ([Law et al. 2014](#)), which returns a t -statistic for each gene. The limma-voom method was selected due to its superior false discovery rate (FDR) control in small sample size experiments, as demonstrated in simulation studies by [Law et al. \(2014\)](#) and [Benidt & Nettleton \(2015\)](#).

For each gene, the length of the coding region in nucleotides (`len`) is also available from the recount2 project, where longer genes generally show higher observed counts and greater power for detecting differential expressions. Additionally, a related microarray study by [Gudjonsson et al. \(2010\)](#), which analyzed 58 pairs of lesional and nonlesional skin samples, is publicly accessible via the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) with accession number GSE13355. This microarray study included 24,908 of the 45,735 genes due to its design, and the t -statistics for these genes, obtained using the limma method ([Smyth 2004](#)), are referred to as `tval` in this study. For illustration purposes, we focus on these 24,908 genes in this study.

Each gene is also characterized by several other covariates. Categorical variables include chromosome (`chr`), identifying the specific chromosome where the gene is located; strand location (`str`), indicating whether the gene is transcribed from the positive or negative DNA strand; and gene type (`type`), distinguishing between 17,752 protein-coding genes, responsible for producing proteins, and 7,156 non-coding genes, which often play regulatory roles. Continuous variables include start (`pos1`) and end positions (`pos2`), which provide the genomic coordinates of the gene, important for mapping its location within the genome. The percentage of guanine and cytosine bases (`gc`) reflects the GC content of the gene, affecting the stability and the expression level of the gene. The gene transcript count (`trans`) represents the number of distinct RNA molecules transcribed from the gene, offering insights into gene activity. Finally, the log-transformed counts per million reads (`logcpm`) measure normalized gene expression levels, where a log transformation is applied to stabilize variance, facilitating comparisons across genes. For ease of interpretability, we use the normalized rank of each continuous covariate. This transformation, one-to-one and monotonic, allowed us to interpret the covariates in terms of their quantiles.

Codak-gam and AdaPT require the specification of an underlying model, which is challenging given that we have 10 covariates for this study and no prior information about

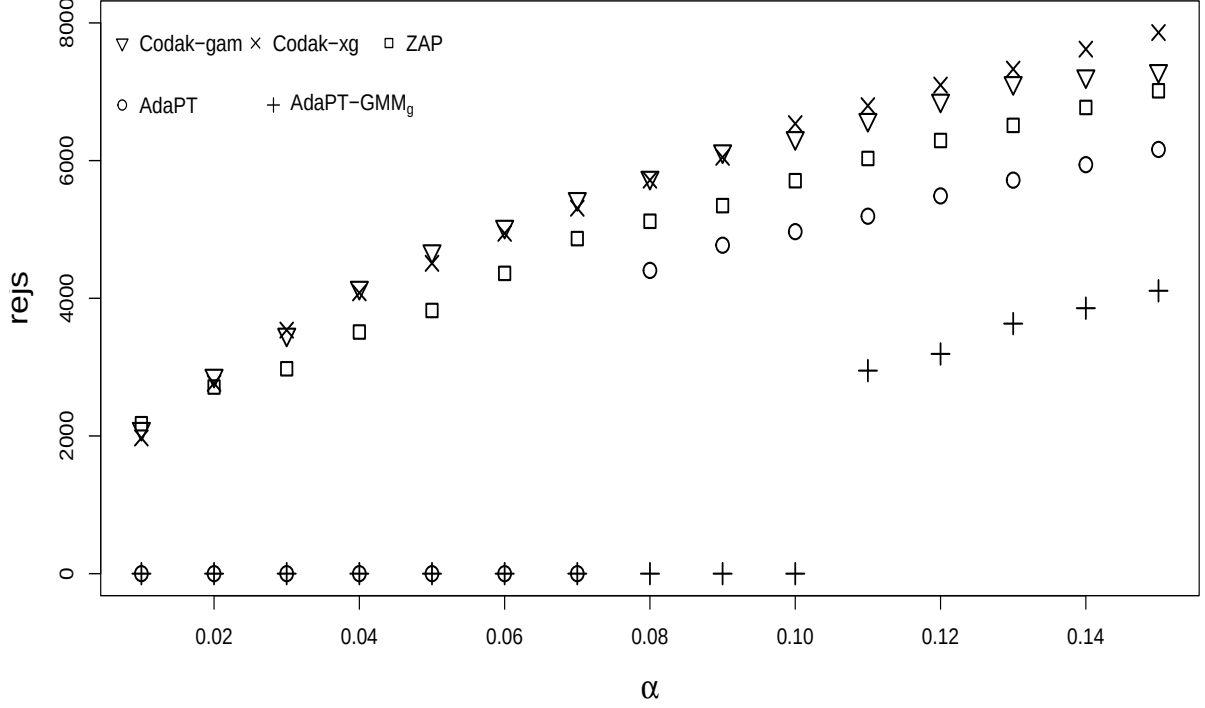


Figure 4.4: The number of rejections over 24,908 observations produced by different methods when controlling FDR at varying levels.

their functional forms. Therefore, both methods are modeled using additive smoothing functions. On the other hand, Codak-xg doesn't require the specification of the functional forms.

Figure 4.4 illustrates the relationship between the number of rejections and the nominal FDR levels. For relatively large α , (i.e., $\alpha > 0.03$), Codak-gam and Codak-xg show the highest number of rejections across all α levels. However, the difference between these two methods is less evident, and the reasons for their close results will be explored later. ZAP, another directional method, consistently produces slightly fewer rejections compared to Codak-gam and Codak-xg. We also observe that AdaPT and AdaPT-GMM_g make no rejections when $\alpha \leq 0.07$, which suggests an inherent conservatism in the methods.

We summarize the feature importance for modeling $\pi_0(\mathbf{x}_i)$, $\lambda(\mathbf{x}_i)$, $\alpha(\mathbf{x}_i)$ and $\beta(\mathbf{x}_i)$ produced by Codak-xg in Figure 4.5. Specifically, for each feature, we count how many times it is used for splitting across all trees in the model and then we report its relative

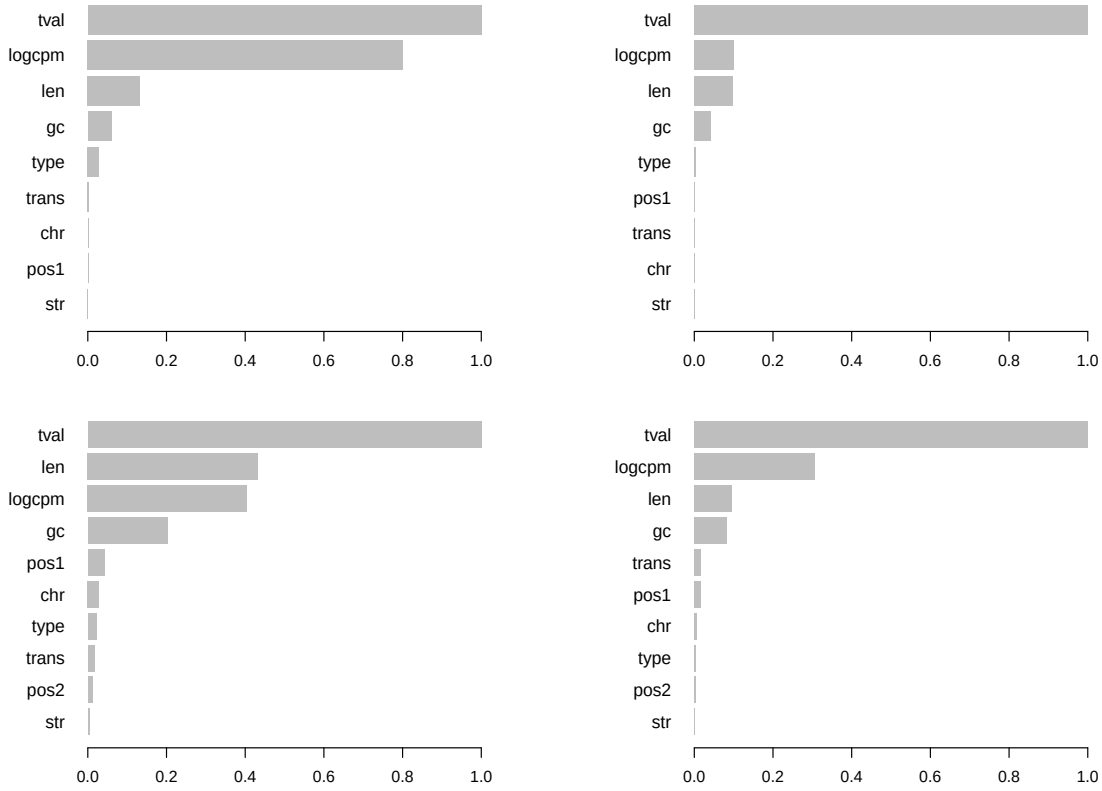


Figure 4.5: The feature importance for modeling $\pi_0(\mathbf{x}_i)$ (top left), $\lambda(\mathbf{x}_i)$ (top right), $\alpha(\mathbf{x}_i)$ (bottom left) and $\beta(\mathbf{x}_i)$ (bottom right) produced by Codak-xg.

frequency to the highest one. The feature importance plots show that across all four models, **tval** consistently appears as the most important feature. Other features like **logcpm** and **len** are also important but to a lesser extent, varying slightly across the different models, while features like **trans**, **chr**, and **str** generally have little importance in these models.

To further explore the potential joint effect between **tval** and **logcpm** in modeling $\pi_0(\mathbf{x}_i)$, we modify the smoothing term of **tval** and **logcpm** to $s(\mathbf{tval}, \mathbf{logcpm})$ while keeping the smoothing terms of other covariates additive in the smoothing function. By Codak-gam, we estimate $\pi_0(\mathbf{x}_i)$ for varying **tval** and **logcpm** while holding other continuous covariates at their median values and categorical covariates at their most common categories. Figure 4.6 presents the 3-D plots of estimated $\pi_0(\mathbf{x}_i)$ and its logit transform. The $\pi_0(\mathbf{x}_i)$ on the logit scale shown on the right panel reveals a noticeable joint effect between **tval** and **logcpm**. It is clear that the logit of $\pi_0(\mathbf{x}_i)$ is small when both **tval** and

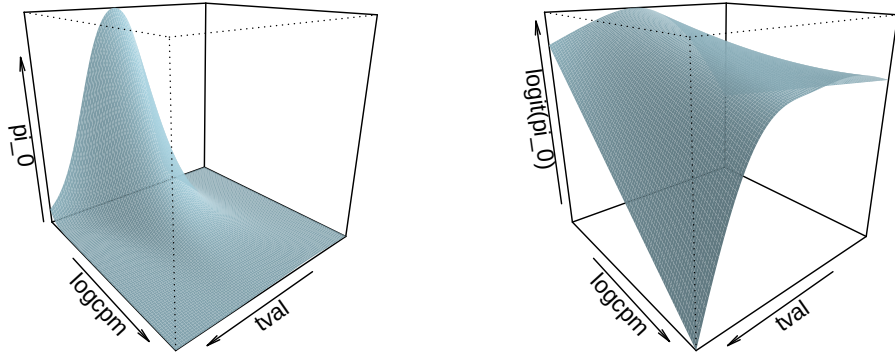


Figure 4.6: The estimated $\pi_0(\mathbf{x}_i)$ (left) and its logit transform (right) for varying `tval` and `logcpm` while holding other continuous covariates at their median values and categorical covariates at their most common categories.

`logcpm` have higher values. However, as shown on the left panel, the joint effect appears to be less evident for the $\pi_0(\mathbf{x}_i)$ on the original scale. The difference between the two 3-D plots indicates that the joint effect is present when modeling the logit of $\pi_0(\mathbf{x}_i)$, and Codak-xg has the ability to capture the joint effect. However, this advantage does not always lead to a significantly smaller $\pi_0(\mathbf{x}_i)$ and a higher number of rejections compared to Codak-gam with an additive smoothing function.

4.6 Technical detail

Recall that in Step 2 of Codak, after k hypotheses have been accepted, we estimate the local FDR by a $\mathcal{F}_k$ -measurable EM-algorithm. In practice, we reestimate the local FDR after every $\lceil n/20 \rceil$ hypotheses have been accepted to save computational time. Additionally, we limit the EM-algorithm to a maximum of 10 iterations. Next, we present some technical details about the Codak-gam and Codak-xg.

4.6.1 Codak-gam

Recall that we use the `gam` function from the `mgcv` package (Wood 2017) to fit a generalized additive model (gam) during each iteration of the M-step. To avoid redundant computations, we initialize the gam model setup before each reveal by setting `fit=FALSE` in the `gam` function and assigning a random weight. The `gam` function then returns an object `G` containing all the necessary setup information. In subsequent iterations, we simply update the weights in `G`, bypassing the need to reinitialize the entire gam model setup. Additionally, to reduce the computational complexity of repeatedly calculating the smoothing parameters in each gam model, we determine them during the first iteration and use these values in all subsequent iterations.

4.6.2 Codak-xg

Next, we present our parameter choices for the “coarse + fine” tuning procedures in Codak-xg. In our default setting, the coarse tuning step generates 15 learning rates equally spaced from $[0.001, 0.281]$ and sets the candidate maximum depths to 1, 3, 5, 7. From the coarse tuning step, we select the top three combinations with the highest log likelihood. For each selected combination (say $\text{eta}_i, \text{max_depth}_i$), the fine tuning step evaluates candidate learning rates within the range $[\text{eta}_i - 0.01, \text{eta}_i + 0.01]$ and maximum depths of $\text{max_depth}_i - 1, \text{max_depth}_i$, and $\text{max_depth}_i + 1$. Additionally, the default number of iterations (`nrounds`) is set as a function of the learning rate, $\text{round}(-5 \times \log(\text{eta} - 0.00009))$, see discussion in Yurko et al. (2020).

Chapter 5

Conclusion and future work

5.1 Conclusion

In Chapter 2, we propose a two-step testing procedure called IM-MST. In the first step, we use the importance measure proposed by [Peng et al. \(2021\)](#) to identify a set of informative features, taking into account the dependence structure in the data. In the second step, MST-based tests are applied exclusively to the selected features. This approach leverages the strengths of both MST and IM, ensuring the retention of informative features and improving the power of MST-based tests, especially in scenarios with sparse multivariate distributional differences.

In Chapter 3, building on the framework of [Wu \(2017\)](#), we focus on the joint estimation of precision matrices under the “partially common diagonal + low-rank” (pcDL) decomposition assumption. We extend the adjusted correlation thresholding (ACT) method to estimate both the number of common factors and the number of group-specific factors, establishing the consistency properties of our approach. These joint estimators are then applied to quadratic discriminant analysis (QDA), where we establish an explicit convergence rate for the classification error. Finally, we assess the performance of our method through empirical studies, demonstrating its accuracy in estimating the number of latent factors and the effectiveness of the proposed pcDLQ classifier using both simulated and real datasets.

In Chapter 4, we extend the covariate and direction adaptive knockoff (Codak) procedure by implementing a generalized additive model (gam) and gradient-boosted trees, leading to the Codak-gam and Codak-xg variants, respectively. For Codak-gam, we simplify the initialization and enhance the computational efficiency. For Codak-xg, we leverage

the flexibility of XGBoost and include covariates without requiring predefined functional forms. Additionally, we incorporate incremental learning within Codak-xg and also propose a “coarse + fine” tuning approach to enhance its predictive capabilities and save computational time. Our numerical studies demonstrate that both Codak-gam and Codak-xg achieve power gain over other existing methods while maintaining the FDR control. Specifically, when a small number of covariates are present, Codak-gam can model their joint effects and is more powerful. However, when dealing with a large number of covariates, Codak-xg outperforms Codak-gam, as Codak-gam must limit itself to only additive effects due to no prior information about the function forms of covariates. Lastly, we apply our methods to an RNA-seq study of psoriasis, where our proposed methods outperform other methods in terms of the number of rejections and can effectively model the joint effect between two important covariates.

5.2 Future work

5.2.1 Random projected QDA

As discussed in 3.1.2, an alternative approach to the high-dimensional QDA problem is to perform QDA after projecting the data onto a lower-dimensional subspace. Rather than searching for an optimal projection direction, recent works propose ensembling classifiers on randomly projected subspaces. These methods combine a collection of classifiers operating on subspaces, potentially improving performance when a single optimal projection direction is unavailable. Ensemble classifiers also enhance stability by aggregating the outcomes of multiple classifiers, often achieving better performance than individual classifiers.

Cannings & Samworth (2017) proposed a general framework for random projection ensembles with pre-selection (RPEn). RPEn reduces the original feature space into lower-dimensional subspaces using random projections and trains base classifiers within these subspaces. Next, RPEn selects the best classifiers, defined as those with the lowest leave-one-out (LOO) cross-validation errors, and makes predictions based on a majority vote among these selected classifiers. The choice of base classifier is flexible, with QDA being one of the provided options. However, the projected dimensions must be kept small enough to control computational costs.

Inspired by the ensembling approach and the optimality of the QDA rule, we may consider random projections of p -dimensional data into k -dimensional subspaces, where $k < p$. In each subspace, we apply the QDA rule and use either a majority voting

scheme or an averaged QDA function approach. Following [Durrant & Kabán \(2013\)](#), we will initially perform 500 random projections and set the dimension of the subspaces to $k = \min\{\lfloor n_1/2 \rfloor, \lfloor n_2/2 \rfloor\}$, where n_1 and n_2 are the training sample sizes. Based on the convergence results for sample covariance matrices established by [Bai & Silverstein \(2004\)](#), we can derive the asymptotic properties of the estimators for each term involving Σ_i in the projected QDA function. For instance,

$$\frac{1}{k} \bar{x}_i^T S_i^{-1} \bar{x}_i - \frac{1}{k} \left(\frac{c_i}{1 - c_i} + \frac{\mu_i^T \Sigma_i^{-1} \mu_i}{1 - c_i} \right) \xrightarrow{\text{a.s.}} 0, i = 1, 2,$$

where $k/n_i \rightarrow c_i \in (0, 1)$, and $\bar{x}_i$ and S_i represent the projected sample mean and projected sample covariance matrix for the i th group, respectively.

[Cannings & Samworth \(2017\)](#) observed empirically that most randomly projected QDA functions cluster around 0, making majority voting susceptible to random guessing. Alternatively, directly averaging the QDA functions may help aggregate contributions from the more informative ones (i.e., those with larger absolute values). However, both majority voting and averaging QDA functions make it challenging to explicitly characterize the convergence rate of the classification errors. A rigorous investigation of this issue is left for future research.

5.2.2 Codak with neural network

As discussed in [4.3.2](#), we can estimate the parameters through different methods, and we consider two options: gam and gradient-boosted trees. In the future, we might be interested in implementing it using a neural network, as suggested in [Leung & Sun \(2022\)](#). Similar to gradient-boosted trees, neural networks can incorporate covariates without requiring the specification of the functional forms. Neural networks can capture more complex relationships among covariates due to the ability to approximate arbitrary functions, especially employing multiple hidden layers. However, neural networks demand more computational resources compared to gradient-boosted trees because neural networks require more careful tuning of parameters such as the learning rate, the number of layers, and the number of nodes per layer. Poorly optimized parameters can lead to suboptimal performance. Given these considerations, implementing neural networks and comparing their performance against gradient-boosted trees remains an important direction for future research.

References

- Bai, Z. & Ding, X. (2012), ‘Estimation of spiked eigenvalues in spiked models’, *Random Matrices: Theory and Applications* **1**(02), 1150011.
- Bai, Z. & Silverstein, J. W. (2004), ‘CLT for linear spectral statistics of large-dimensional sample covariance matrices’, *The Annals of Probability* **32**(1A), 553 – 605.
- Bai, Z. & Silverstein, J. W. (2010), *Spectral analysis of large dimensional random matrices*, Springer.
- Benidt, S. & Nettleton, D. (2015), ‘SimSeq: a nonparametric approach to simulation of rna-sequence datasets’, *Bioinformatics* **31**(13), 2131–2140.
- Benjamini, Y. & Hochberg, Y. (1995), ‘Controlling the false discovery rate: a practical and powerful approach to multiple testing’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **57**(1), 289–300.
- Benjamini, Y., Krieger, A. M. & Yekutieli, D. (2006), ‘Adaptive linear step-up procedures that control the false discovery rate’, *Biometrika* **93**(3), 491–507.
- Benjamini, Y. & Yekutieli, D. (2001), ‘The control of the false discovery rate in multiple testing under dependency’, *The Annals of Statistics* **29**(4), 1165–1188.
- Brambilla, E. & Gazdar, A. (2009), ‘Pathogenesis of lung cancer signalling pathways: roadmap for therapies’, *European Respiratory Journal* **33**(6), 1485–1497.
- Cai, T. T. & Liu, W. (2011), ‘A direct estimation approach to sparse linear discriminant analysis’, *Journal of the American Statistical Association* **106**(496), 1566–1577.
- Cai, T. T., Ma, J., Zhang, L. et al. (2019), ‘CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality’, *The Annals of Statistics* **47**(3), 1234–1267.

- Cai, T. T. & Zhang, L. (2021), ‘A convex optimization approach to high-dimensional sparse quadratic discriminant analysis’, *The Annals of Statistics* **49**(3), 1537–1568.
- Cannings, T. I. & Samworth, R. J. (2017), ‘Random-projection ensemble classification’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**(4), 959–1035.
- Chao, P. & Fithian, W. (2021), ‘AdaPT-GMM: Powerful and robust covariate-assisted multiple testing’, *arXiv preprint arXiv:2106.15812*.
- Chen, H., Chen, X. & Su, Y. (2018), ‘A weighted edge-count two-sample test for multivariate and object data’, *Journal of the American Statistical Association* **113**(523), 1146–1155.
- Chen, H. & Friedman, J. H. (2017), ‘A new graph-based two-sample test for multivariate and object data’, *Journal of the American Statistical Association* **112**(517), 397–409.
- Chen, T. & Guestrin, C. (2016), ‘Xgboost: A scalable tree boosting system’, *Proceedings of the 22nd ACM Sigkdd international conference on knowledge discovery and data mining* pp. 785–794.
- Chu, L. & Chen, H. (2019), ‘Asymptotic distribution-free change-point detection for multivariate and non-euclidean data’, *The Annals of Statistics* **47**(1), 382–414.
- Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S. E., Taub, M. A., Hansen, K. D., Jaffe, A. E., Langmead, B. & Leek, J. T. (2017), ‘Reproducible RNA-seq analysis using recount2’, *Nature biotechnology* **35**(4), 319–321.
- Danaher, P., Wang, P. & Witten, D. M. (2014), ‘The joint graphical lasso for inverse covariance estimation across multiple classes’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 373–397.
- Durrant, R. & Kabán, A. (2013), Random projections as regularizers: Learning a linear discriminant ensemble from fewer observations than dimensions, *in* ‘Asian conference on machine learning’, PMLR, pp. 17–32.
- Fan, J. & Fan, Y. (2008), ‘High dimensional classification using features annealed independence rules’, *The Annals of Statistics* **36**(6), 2605–2637.
- Fan, J., Fan, Y. & Lv, J. (2008), ‘High dimensional covariance matrix estimation using a factor model’, *Journal of Econometrics* **147**(1), 186–197.

- Fan, J., Guo, J. & Zheng, S. (2022), ‘Estimating number of factors by adjusted eigenvalues thresholding’, *Journal of the American Statistical Association* **117**(538), 852–861.
- Fan, J., Liao, Y. & Mincheva, M. (2011), ‘High dimensional covariance matrix estimation in approximate factor models’, *The Annals of Statistics* **39**(6), 3320–3356.
- Fan, J., Liao, Y. & Mincheva, M. (2013), ‘Large covariance estimation by thresholding principal orthogonal complements’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(4), 603–680.
- Fan, J. & Lv, J. (2008), ‘Sure independence screening for ultrahigh dimensional feature space’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(5), 849–911.
- Fan, J. & Song, R. (2010), ‘Sure independence screening in generalized linear models with NP-dimensionality’, *The Annals of Statistics* **38**(6), 3567–3604.
- Ferkingstad, E., Frigessi, A., Rue, H., Thorleifsson, G. & Kong, A. (2008), ‘Unsupervised empirical Bayesian multiple testing with external covariates’, *The Annals of Applied Statistics* **2**(2), 714–735.
- Fey, D., Halasz, M., Dreidax, D., Kennedy, S. P., Hastings, J. F., Rauch, N., Munoz, A. G., Pilkington, R., Fischer, M., Westermann, F. et al. (2015), ‘Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients’, *Science signaling* **8**(408), ra130–ra130.
- Fortney, K., Dobriban, E., Garagnani, P., Pirazzini, C., Monti, D., Mari, D., Atzmon, G., Barzilai, N., Franceschi, C., Owen, A. B. et al. (2015), ‘Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity’, *PLoS genetics* **11**(12), e1005728.
- Friedman, J. H. & Rafsky, L. C. (1979), ‘Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests’, *The Annals of Statistics* **7**(4), 697–717.
- Gaynanova, I. & Wang, T. (2019), ‘Sparse quadratic classification rules via linear dimension reduction’, *Journal of multivariate analysis* **169**, 278–299.
- Genovese, C. R., Roeder, K. & Wasserman, L. (2006), ‘False discovery control with p-value weighting’, *Biometrika* **93**(3), 509–524.

- Gudjonsson, J. E., Ding, J., Johnston, A., Tejasvi, T., Guzman, A. M., Nair, R. P., Voorhees, J. J., Abecasis, G. R. & Elder, J. T. (2010), ‘Assessment of the psoriatic transcriptome in a large sample: additional regulated genes and comparisons with in vitro models’, *Journal of Investigative Dermatology* **130**(7), 1829–1840.
- Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., Kang, B., Liu, Z., Jin, L., Xing, R. et al. (2018), ‘Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing’, *Nature medicine* **24**(7), 978–985.
- Henderson, H. V. & Searle, S. R. (1981), ‘On deriving the inverse of a sum of matrices’, *SIAM Review* **23**(1), 53–60.
- Henze, N. (1988), ‘A multivariate two-sample test based on the number of nearest neighbor type coincidences’, *The Annals of Statistics* **16**(2), 772–783.
- Ignatiadis, N. & Huber, W. (2021), ‘Covariate powered cross-weighted multiple testing’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **83**(4), 720–751.
- Jabbari, A., Suárez-Fariñas, M., Dewell, S. & Krueger, J. G. (2012), ‘Transcriptional profiling of psoriasis using RNA-seq reveals previously unidentified differentially expressed genes’, *The Journal of investigative dermatology* **132**(1), 246.
- Jiang, B., Wang, X. & Leng, C. (2018), ‘A direct approach for sparse quadratic discriminant analysis’, *The Journal of Machine Learning Research* **19**(1), 1098–1134.
- Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. (2014), ‘voom: Precision weights unlock linear model analysis tools for RNA-seq read counts’, *Genome biology* **15**, 1–17.
- Lei, L. & Fithian, W. (2018), ‘AdaPT: an interactive procedure for multiple testing with side information’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(4), 649–679.
- Leung, D. & Sun, W. (2022), ‘ZAP: Z-value adaptive procedures for false discovery rate control with side information’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **84**(5), 1886–1946.
- Li, Q. & Shao, J. (2015), ‘Sparse quadratic discriminant analysis for high dimensional data’, *Statistica Sinica* **40**(2), 457–473.
- Li, R., Zhong, W. & Zhu, L. (2012), ‘Feature screening via distance correlation learning’, *Journal of the American Statistical Association* **107**(499), 1129–1139.

- Liang, K. (2019), ‘Empirical Bayes analysis of RNA sequencing experiments with auxiliary information’, *The Annals of Applied Statistics* **13**(4), 2452–2482.
- Liang, K. & Nettleton, D. (2012), ‘Adaptive and dynamic adaptive procedures for false discovery rate control and estimation’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**(1), 163–182.
- Little, M., McSharry, P., Hunter, E., Spielman, J. & Ramig, L. (2009), ‘Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease’, *IEEE Transactions on Biomedical Engineering* **56**(4), 1015–1022.
- Liu, W., Ke, Y., Liu, J. & Li, R. (2022), ‘Model-free feature screening and fdr control with knockoff features’, *Journal of the American Statistical Association* **117**(537), 428–443.
- MacDonald, P. W., Liang, K. & Janssen, A. (2019), ‘Dynamic adaptive procedures that control the false discovery rate’, *Electronic Journal of Statistics* **13**(2), 3009–3024.
- Mukherjee, S., Agarwal, D., Zhang, N. R. & Bhattacharya, B. B. (2022), ‘Distribution-free multisample tests based on optimal matchings with applications to single cell genomics’, *Journal of the American Statistical Association* **117**(538), 627–638.
- Peng, L., Qu, L. & Nettleton, D. (2021), ‘Variable importance assessments and backward variable selection for multi-sample problems’, *Journal of Multivariate Analysis* **186**, 104807.
- Petrie, A. (2016), ‘Graph-theoretic multisample tests of equality in distribution for high dimensional data’, *Computational Statistics and Data Analysis* **96**, 145–158.
- Qin, Y. (2018), ‘A review of quadratic discriminant analysis for high-dimensional data’, *Wiley Interdisciplinary Reviews: Computational Statistics* **10**(4), e1434.
- Rizzo, M. L. & Székely, G. J. (2016), ‘Energy distance’, *wiley interdisciplinary reviews: Computational statistics* **8**(1), 27–38.
- Rosenbaum, P. R. (2005), ‘An exact distribution-free test comparing two multivariate distributions based on adjacency’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(4), 515–530.
- Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gorgen, F., Delil, S., Apaydin, H. & Kursun, O. (2013), ‘Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings’, *IEEE Journal of Biomedical and Health Informatics* **17**(4), 828–834.

- Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Tutuncu, M., Aydin, T., Isenkul, M. E. & Apaydin, H. (2019), ‘A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform’, *Applied Soft Computing* **74**, 255–263.
- Salavert, F., Hidago, M. R., Amadoz, A., Cubuk, C., Medina, I., Crespo, D., Carbonell-Caballero, J. & Dopazo, J. (2016), ‘Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models’, *Nucleic acids research* **44**(W1), W212–W216.
- Santarpia, M., Aguilar, A., Chaib, I., Cardona, A. F., Fancelli, S., Laguia, F., Bracht, J. W. P., Cao, P., Molina-Vila, M. A., Karachaliou, N. et al. (2020), ‘Non-small-cell lung cancer signaling pathways, metabolism, and PD-1/PD-L1 antibodies’, *Cancers* **12**(6), 1475.
- Schilling, M. F. (1986), ‘Multivariate two-sample tests based on nearest neighbors’, *Journal of the American Statistical Association* **81**(395), 799–806.
- Silverstein, J. W. & Choi, S.-I. (1995), ‘Analysis of the limiting spectral distribution of large dimensional random matrices’, *Journal of Multivariate Analysis* **54**(2), 295–309.
- Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D’Amico, A. V., Richie, J. P. et al. (2002), ‘Gene expression correlates of clinical prostate cancer behavior’, *Cancer cell* **1**(2), 203–209.
- Smyth, G. K. (2004), ‘Linear models and empirical bayes methods for assessing differential expression in microarray experiments’, *Statistical applications in genetics and molecular biology* **3**(1), 3–25.
- Storey, J. D., Taylor, J. E. & Siegmund, D. (2004), ‘Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(1), 187–205.
- Sun, W. & Cai, T. T. (2007), ‘Oracle and adaptive compound decision rules for false discovery rate control’, *Journal of the American Statistical Association* **102**(479), 901–912.
- Székel, G. J. & Rizzo, M. L. (2005), ‘A new test for multivariate normality’, *Journal of Multivariate Analysis* **93**(1), 58–80.

- Tian, Z. (2020), *Nonparametric estimation in a compound mixture model and false discovery rate control with auxiliary information*, University of Waterloo Ph.D. Thesis.
- Tian, Z., Liang, K. & Li, P. (2021), ‘A powerful procedure that controls the false discovery rate with directional information’, *Biometrics* **77**(1), 212–222.
- Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press.
- Wald, A. & Wolfowitz, J. (1940), ‘On a test whether two samples are from the same population’, *The Annals of Mathematical Statistics* **11**(2), 147–162.
- Wood, S. N. (2017), *Generalized additive models: an introduction with R*, Chapman and Hall/CRC.
- Wu, R. & Hao, N. (2022), ‘Quadratic discriminant analysis by projection’, *Journal of Multivariate Analysis* **190**, 104987.
- Wu, Y. (2017), *High-dimensional discriminant analysis and covariance matrix estimation*, University of Waterloo Ph.D. Thesis.
- Wu, Y., Qin, Y. & Zhu, M. (2019), ‘Quadratic discriminant analysis for high-dimensional data’, *Statistica Sinica* **29**(2), 939–960.
- Wu, Y., Qin, Y. & Zhu, M. (2020), ‘High-dimensional covariance matrix estimation using a low-rank and diagonal decomposition’, *Canadian Journal of Statistics* **48**(2), 308–337.
- Yekutieli, D., Reiner-Benaim, A., Benjamini, Y., Elmer, G. I., Kafkafi, N., Letwin, N. E. & Lee, N. H. (2006), ‘Approaches to multiplicity issues in complex research in microarray analysis’, *Statistica Neerlandica* **60**(4), 414–437.
- Yurko, R., G’Sell, M., Roeder, K. & Devlin, B. (2020), ‘A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk’, *Proceedings of the National Academy of Sciences* **117**(26), 15028–15035.
- Zhou, D. & Chen, H. (2023), ‘A new ranking scheme for modern data and its application to two-sample hypothesis testing’, *The Thirty Sixth Annual Conference on Learning Theory* pp. 3615–3668.
- Zhu, Y. & Chen, H. (2024), ‘Limiting distributions of graph-based test statistics on sparse and dense graphs’, *Bernoulli* **30**(1), 770–796.

APPENDICES

Appendix A

Table A.1: Size comparison under Model 1.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	0.046	0.076	0.044	0.042	0.038	0.034
wMST	0.066	0.054	0.048	0.050	0.024	0.042
RISE	0.070	0.072	0.044	0.046	0.038	0.042
IM-gMST	0.046	0.062	0.046	0.044	0.044	0.054
IM-wMST	0.050	0.052	0.046	0.068	0.034	0.056
IM-RISE	0.061	0.058	0.045	0.051	0.045	0.053
IM-MRPP	0.062	0.070	0.057	0.056	0.046	0.051
ORC-gMST	0.040	0.048	0.064	0.072	0.032	0.050
ORC-wMST	0.052	0.052	0.060	0.048	0.048	0.054
ORC-RISE	0.049	0.052	0.061	0.050	0.049	0.051
$n = 80, m = 160$						
gMST	0.052	0.064	0.058	0.044	0.032	0.022
wMST	0.052	0.060	0.060	0.046	0.052	0.040
RISE	0.056	0.050	0.052	0.058	0.030	0.052
IM-gMST	0.040	0.054	0.056	0.038	0.056	0.036
IM-wMST	0.052	0.050	0.046	0.057	0.048	0.040
IM-RISE	0.049	0.048	0.049	0.055	0.051	0.041
IM-MRPP	0.064	0.072	0.049	0.048	0.053	0.047
ORC-gMST	0.038	0.054	0.062	0.064	0.046	0.050
ORC-wMST	0.042	0.048	0.058	0.062	0.044	0.048
ORC-RISE	0.041	0.047	0.053	0.061	0.044	0.050

Table A.2: Size comparison under Model 3.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	0.042	0.034	0.058	0.062	0.042	0.038
wMST	0.060	0.042	0.072	0.064	0.038	0.044
RISE	0.048	0.038	0.044	0.050	0.048	0.042
IM-gMST	0.048	0.050	0.048	0.030	0.060	0.034
IM-wMST	0.042	0.058	0.052	0.046	0.044	0.048
IM-RISE	0.043	0.035	0.045	0.043	0.049	0.051
IM-MRPP	0.051	0.052	0.052	0.049	0.057	0.050
ORC-gMST	0.050	0.056	0.048	0.050	0.052	0.048
ORC-wMST	0.040	0.052	0.064	0.050	0.060	0.046
ORC-RISE	0.047	0.050	0.055	0.047	0.048	0.057
$n = 80, m = 160$						
gMST	0.036	0.050	0.046	0.058	0.042	0.058
wMST	0.064	0.062	0.052	0.066	0.048	0.064
RISE	0.044	0.060	0.070	0.056	0.048	0.048
IM-gMST	0.054	0.044	0.074	0.046	0.036	0.044
IM-wMST	0.054	0.046	0.052	0.042	0.056	0.048
IM-RISE	0.050	0.044	0.057	0.052	0.048	0.041
IM-MRPP	0.051	0.047	0.049	0.050	0.051	0.048

Table A.3: Size comparison under Model 4.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	0.048	0.062	0.052	0.044	0.056	0.034
wMST	0.042	0.048	0.038	0.058	0.042	0.054
RISE	0.050	0.054	0.038	0.060	0.040	0.042
IM-gMST	0.046	0.050	0.048	0.050	0.040	0.060
IM-wMST	0.056	0.068	0.036	0.036	0.060	0.062
IM-RISE	0.051	0.055	0.048	0.052	0.050	0.048
IM-MRPP	0.047	0.048	0.051	0.047	0.055	0.049
ORC-gMST	0.044	0.054	0.056	0.062	0.042	0.060
ORC-wMST	0.070	0.058	0.062	0.056	0.062	0.060
ORC-RISE	0.042	0.052	0.058	0.046	0.049	0.058
$n = 80, m = 160$						
gMST	0.044	0.044	0.026	0.042	0.046	0.056
wMST	0.042	0.042	0.052	0.062	0.052	0.050
RISE	0.040	0.050	0.052	0.050	0.032	0.050
IM-gMST	0.052	0.044	0.048	0.038	0.054	0.036
IM-wMST	0.054	0.046	0.062	0.044	0.056	0.042
IM-RISE	0.050	0.051	0.042	0.039	0.057	0.049
IM-MRPP	0.048	0.050	0.053	0.047	0.055	0.050
ORC-gMST	0.052	0.060	0.058	0.056	0.052	0.044
ORC-wMST	0.048	0.066	0.072	0.070	0.048	0.064
ORC-RISE	0.069	0.053	0.049	0.050	0.051	0.047

Table A.4: Power comparison under Model 1.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	0.732	0.532	0.294	0.202	0.208	0.148
wMST	0.862	0.732	0.486	0.372	0.356	0.278
RISE	0.852	0.574	0.336	0.210	0.170	0.170
IM-gMST	0.938	0.698	0.550	0.416	0.404	0.314
IM-wMST	0.958	0.818	0.710	0.612	0.624	0.506
IM-RISE	0.942	0.723	0.611	0.567	0.517	0.463
IM-MRPP	0.893	0.747	0.582	0.512	0.301	0.197
ORC-gMST	0.904	0.908	0.892	0.692	0.540	0.388
ORC-wMST	0.964	0.960	0.950	0.834	0.686	0.546
ORC-RISE	0.917	0.908	0.887	0.752	0.555	0.423
$n = 80, m = 160$						
gMST	0.840	0.628	0.432	0.280	0.226	0.190
wMST	0.926	0.786	0.616	0.494	0.414	0.350
RISE	0.962	0.800	0.568	0.380	0.272	0.226
IM-gMST	0.966	0.848	0.542	0.388	0.380	0.312
IM-wMST	0.974	0.932	0.698	0.572	0.532	0.414
IM-RISE	0.981	0.915	0.673	0.419	0.423	0.337
IM-MRPP	0.887	0.764	0.678	0.423	0.326	0.292
ORC-gMST	0.972	0.962	0.956	0.774	0.618	0.490
ORC-wMST	0.988	0.982	0.986	0.890	0.756	0.660
ORC-RISE	0.952	0.920	0.917	0.784	0.603	0.512

Table A.5: Power comparison under Model 2.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	1.000	0.990	0.844	0.726	0.586	0.518
wMST	0.984	0.922	0.760	0.648	0.554	0.502
RISE	0.998	0.954	0.812	0.746	0.644	0.580
IM-gMST	0.979	0.965	0.948	0.887	0.817	0.765
IM-wMST	0.948	0.918	0.814	0.792	0.704	0.702
IM-RISE	0.980	0.976	0.944	0.890	0.874	0.828
IM-MRPP	0.806	0.618	0.506	0.497	0.432	0.402
ORC-gMST	1.000	0.998	1.000	0.998	1.000	0.996
ORC-wMST	1.000	0.998	1.000	1.000	0.998	0.998
ORC-RISE	1.000	0.998	1.000	0.999	0.997	0.996
$n = 80, m = 160$						
gMST	0.958	0.838	0.584	0.464	0.350	0.336
wMST	0.986	0.880	0.684	0.548	0.432	0.402
RISE	0.988	0.898	0.662	0.458	0.376	0.368
IM-gMST	0.908	0.802	0.734	0.658	0.562	0.562
IM-wMST	0.970	0.948	0.874	0.816	0.782	0.738
IM-RISE	0.935	0.870	0.768	0.636	0.590	0.557
IM-MRPP	0.890	0.805	0.746	0.637	0.542	0.539
ORC-gMST	1.000	1.000	1.000	1.000	1.000	1.000
ORC-wMST	1.000	1.000	1.000	1.000	1.000	1.000
ORC-RISE	1.000	1.000	1.000	1.000	1.000	1.000

Table A.6: Power comparison under Model 3.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	0.630	0.814	0.770	0.758	0.724	0.650
wMST	0.814	0.932	0.882	0.874	0.864	0.838
RISE	0.776	0.894	0.856	0.796	0.790	0.730
IM-gMST	0.930	0.966	0.954	0.884	0.798	0.760
IM-wMST	0.956	0.988	0.978	0.956	0.924	0.864
IM-RISE	0.945	0.969	0.972	0.919	0.876	0.882
IM-MRPP	0.850	0.783	0.639	0.520	0.487	0.442
ORC-gMST	0.866	0.994	1.000	0.996	0.984	0.962
ORC-wMST	0.942	0.998	1.000	1.000	0.998	0.984
ORC-RISE	0.957	0.975	1.000	0.992	0.985	0.980
$n = 80, m = 160$						
gMST	1.000	1.000	0.802	0.394	0.266	0.240
wMST	1.000	1.000	0.916	0.628	0.438	0.296
RISE	1.000	1.000	0.910	0.614	0.328	0.218
IM-gMST	1.000	1.000	0.970	0.758	0.516	0.302
IM-wMST	1.000	1.000	0.990	0.872	0.670	0.458
IM-RISE	1.000	1.000	0.957	0.892	0.592	0.406
IM-MRPP	0.795	0.738	0.717	0.589	0.456	0.339
ORC-gMST	1.000	1.000	1.000	1.000	0.994	0.980
ORC-wMST	1.000	1.000	1.000	1.000	0.998	0.994
ORC-RISE	1.000	1.000	1.000	1.000	0.997	0.995

Table A.7: Power comparison under Model 4.

$n = m = 80$	$p = 50$	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
gMST	0.996	0.954	0.842	0.758	0.706	0.610
wMST	0.970	0.896	0.780	0.654	0.564	0.474
RISE	0.754	0.604	0.462	0.398	0.384	0.330
IM-gMST	0.972	0.970	0.912	0.874	0.862	0.802
IM-wMST	0.794	0.730	0.662	0.600	0.606	0.532
IM-RISE	0.935	0.875	0.883	0.852	0.830	0.787
IM-MRPP	0.742	0.783	0.639	0.525	0.487	0.453
ORC-gMST	1.000	0.994	1.000	0.998	0.978	0.991
ORC-wMST	0.978	0.950	0.966	0.974	0.982	0.948
ORC-RISE	1.000	0.996	1.000	0.992	0.993	0.985
$n = 80, m = 160$						
gMST	0.976	0.886	0.784	0.630	0.554	0.482
wMST	0.996	0.966	0.864	0.798	0.734	0.650
RISE	0.426	0.316	0.346	0.360	0.292	0.284
IM-gMST	0.894	0.812	0.696	0.624	0.604	0.542
IM-wMST	0.952	0.930	0.894	0.880	0.865	0.804
IM-RISE	0.914	0.878	0.713	0.656	0.601	0.533
IM-MRPP	0.836	0.865	0.798	0.817	0.683	0.595
ORC-gMST	1.000	0.998	1.000	1.000	1.000	1.000
ORC-wMST	1.000	0.998	1.000	0.996	1.000	0.998
ORC-RISE	1.000	0.959	0.976	0.980	0.985	0.993