Progress in nonparametric minimax estimation and high dimensional hypothesis testing

Yandi Shen

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Daniela Witten, Chair

Fang Han, Chair

Jon Wellner

Program Authorized to Offer Degree:
Department of Statistics

University of Washington

Abstract

Progress in nonparametric minimax estimation and high dimensional hypothesis testing

Yandi Shen

Co-Chairs of the Supervisory Committee:
Professor Daniela Witten
Statistics and Biostatistics

Professor Fang Han
Statistics

This dissertation is divided into two parts. In the first part, we study minimax estimation of functions and functionals in nonparametric regression models. The investigation of statistical limits in such models deepens theoretical understanding in related problems and leads to new probabilistic tools and methodologies of broader interest. In the second part, we study the asymptotics in some high dimensional testing problems involving the Gaussian distribution, such as the Gaussian sequence model with convex constraint and testing of covariance matrices. A general framework is developed to analyze the power behavior of test statistics via accurate non-asymptotic expansions.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisors, Prof. Daniela Witten and Prof. Fang Han, for their mentorship throughout my time at the University of Washington. They helped me grow both as a professional and as a person, for which I will always be grateful.

I would also like to thank the other two members of my committee: Prof. Jon Wellner and Prof. Yanqin Fan. I have had the greatest joy in discussing research problems with them at several occasions over the years, and I am grateful for all of the time they have put into assisting me with the editing process as well as their very thoughtful advice.

I would like to extend my sincerest gratitude to my wonderful collaborators over the course of my PhD: Prof. Chao Gao at University of Chicago, Prof. Qiyang Han at Rutgers University, Prof. Tiefeng Jiang at University of Minnesota, and Prof. Bodhisattva Sen at Columbia University. This dissertation would not have been complete without their greatly appreciated contributions.

Lastly, I would like to thank my friends and family, who have consistently supported and encouraged me throughout times both good and bad. They have helped to shape the value I place in all things, including the importance in the pursuit and exploration of truth, helping to make the world a better place, and honesty and humility. For that I will always be grateful.

DEDICATION

to my parents, Hua and Bin

Chapter 1

INTRODUCTION

This dissertation is mainly divided into two parts. The first part, to be detailed in Chapters 2 and 3, focuses on minimax estimation of functions and functionals in nonparametric regression models. The investigation of statistical limits in such models deepens theoretical understanding in related problems and leads to new probabilistic tools and methodologies of broader interest. This part is further divided into two sub-parts. In Chapter 2, we study the optimal estimation of quadratic-type functionals in the nonparametric regression model, which are considered to be fundamental objects in nonparametric minimax theory. We derive a variety of minimax rates and elaborate their detailed connections with existing results. These results help unify the minimax rates of estimating quadratic-type functionals under the nonparametric regression model with the other two benchmark models in nonparametric statistics - the density model and the white noise model.

In Chapter 3, we study the fundamental limit of general order spline estimation, both with and without shape constraints. These results significantly broaden and unify previous attempts and results in the literature focused on the special case of piecewise constant signals. They also provide theoretical support for developing ℓ_0-penalized spline regression as a useful alternative to ℓ_1- and ℓ_2-penalized ones.

The second part of the dissertation, to be detailed in Chapters 4 and 5, aims at developing a relatively general framework for power analysis in high dimensional testing problems involving the Gaussian distribution. We start in Chapter 4, where we study the high dimensional asymptotics of likelihood ratio tests in the Gaussian sequence model with convex constraint. Some general theory is developed, followed by detailed applications in the cases of orthant/circular cone, isotonic regression, Lasso, etc.

In Chapter 5, we move on to the study of covariance tests related to the high dimensional Gaussian distribution. We develop a general method for analyzing the power behavior of covariance test statistics via accurate non-asymptotic power expansions. These results apply to arbitrary alternative hypotheses under mild growth conditions on the dimension-to-sample ratio.

The main contents of this thesis are drawn, with minor modification, from the following four manuscripts: Chapter 2 is from "Optimal estimation of variance in nonparametric regression with random design", co-authored with Chao Gao, Daniela Witten, and Fang Han; Chapter 3 is from "A general method for power analysis in testing high dimensional covariance matrices", co-authored with Qiyang Han and Fang Han; Chapter 4 is from "High dimensional asymptotics of likelihood ratio tests in Gaussian sequence model under convex constraint", co-authored with Qiyang Han and Bodhisattva Sen; Chapter 5 is from "A general method for power analysis in testing high dimensional covariance matrices", co-authored with Qiyang Han and Tiefeng Jiang.

Chapter 2

VARIANCE ESTIMATION IN NONPARAMETRIC REGRESSION

2.1 Introduction

Consider the model

$$Y_i = f(X_i) + V^{1/2}(X_i)\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.1}$$

where $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) univariate random design points, and $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. with zero mean, unit variance, and are independent of $\{X_i\}_{i=1}^n$. In this paper, we study the optimal estimation of $V(\cdot)$ under both local and global squared risks. Variance estimation is a fundamental statistical problem (Von Neumann, 1941, 1942; Rice, 1984; Hall et al., 1990) with wide applications. It is useful in, for example, construction of confidence bands for the mean function, estimation of the signal-to-noise ratio (Verzelen and Gassiat, 2018), and selection of the optimal kernel bandwidth (Fan, 1992).

When $\{X_i\}_{i=1}^n$ are fixed, estimation of $V(\cdot)$ in (2.1) has been studied extensively in the literature via residual-based methods (Hall and Carroll, 1989; Ruppert et al., 1997; Härdle and Tsybakov, 1997; Fan and Yao, 1998) and difference-based methods (Muller and Stadtmuller, 1987; Müller et al., 2003; Brown and Levine, 2007; Wang et al., 2008). One important heuristic from previous studies is that, compared to residual-based methods, difference-based methods are able to achiever a smaller bias and subsequently a smaller mean squared error by avoiding direct estimation of the mean function. More precisely, when $X_i = i/n$, $i = 1, \ldots, n$ and $f(\cdot)$ and $V(\cdot)$ in (2.1) are α- and β-Hölder smooth, respectively, Wang et al. (2008) proposed a difference estimator which achieved the optimal rate of the order $n^{-4\alpha} \vee n^{-\frac{2\beta}{2\beta+1}}$

under both local and global squared risks.

In contrast, our study focuses on the case where $\{X_i\}_{i=1}^n$ are i.i.d. random design points on the real line. For this, we show that when $f(\cdot)$ and $V(\cdot)$ in (2.1) are α- and β-Hölder smooth, respectively, the minimax rate of estimating $V(\cdot)$ is of the order $n^{-\frac{8\alpha\beta}{4\alpha\beta+2\alpha+\beta}} \vee n^{-\frac{2\beta}{2\beta+1}}$ under both local and global squared risks. This result has several noteworthy implications:

- The minimax rates in random and fixed design settings share a common component, $n^{-\frac{2\beta}{2\beta+1}}$, as well as the same transition boundary $\alpha = \beta/(4\beta+2)$.
- For $\alpha < \beta/(4\beta+2)$, a faster rate is achievable with a random design.
- Unlike the fixed design setting, for $\alpha < \beta/(4\beta+2)$, α and β are now both present in the first term of the minimax rate in the random design case.

We now discuss in more detail this minimax rate. The upper bound of the minimax rate is achieved by smoothing pairwise differences via local polynomial regression, the former of which is formulated via U-statistics. Our analysis of this estimator hence relies on the four-term Bernstein inequality in Giné et al. (2000), and unlike classic kernel methods, requires no smoothness assumption on the design density.

For the lower bound, due to the appearances of both α and β in the non-trivial $n^{-\frac{8\alpha\beta}{4\alpha\beta+2\alpha+\beta}}$ part of the minimax rate and the additional randomness of $\{X_i\}_{i=1}^n$, the derivation is much more involved than its counterpart in the fixed design setting. We tackle the first difficulty of entangled α and β via a proper localization technique in the construction of the mean function $f(\cdot)$, depicted in Figure 2.2 in Section 2.3.2. The second difficulty caused by the randomness of $\{X_i\}_{i=1}^n$ is resolved with a new trapezoid-shaped construction of the mean $f(\cdot)$, aided by a result due to Kolchin et al. (1978) on the sparse multinomial distribution. This result helps characterize the asymptotic behavior of the locations of $\{X_i\}_{i=1}^n$ and plays a key role in our proof, but to our knowledge has not been well used in the nonparametric statistics literature.

In the special case of constant variance, (2.1) is reduced to

$$Y_i = f(X_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.2}$$

and the goal becomes estimation of σ^2. In this case, the problem is linked to estimation of a quadratic functional, which has been studied in depth in the other two benchmark nonparametric models, the density model (Bickel and Ritov, 1988; Laurent, 1996; Giné and Nickl, 2008) and the white noise model (Donoho and Nussbaum, 1990; Fan, 1991; Laurent and Massart, 2000). In the density model, one observes an i.i.d. univariate sequence $\{X_i\}_{i=1}^n$ from some unknown density $f(\cdot)$, and the goal is to estimate $\int f^2(x)dx$. In the white noise model, one observes a continuous-time process from $dY_t = f(t)dt + n^{-1/2}dW_t$ for $t \in [0, 1]$ with W_t a standard Wiener process. The goal is to estimate $\int_0^1 f^2(t)dt$. Under an α-smoothness condition on $f(\cdot)$, the minimax rate in both of the aforementioned two cases is $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ (cf. Theorem 1(ii) and 2(ii) in Bickel and Ritov (1988), Theorem 4 in Fan (1991)).

Following Doksum and Samarov (1995), a quadratic functional of interest under (2.2) with random design is

$$Q := \int f^2(x)p_X(x)w(x)dx, \tag{2.3}$$

where $p_X(\cdot)$ is the unknown design density and $w(\cdot) \geq 0$ is some known weight function. Assuming in (2.2) that f is α-Hölder smooth, we show that the minimax rate of estimating σ^2 and Q (when σ^2 is unknown) is $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$, thereby unifying the minimax rate of quadratic functional estimation in all three benchmark nonparametric models.

In this paper, we also provide extensions of (2.2) to multivariate cases, with a focus on the multivariate nonparametric regression model

$$Y_i = f(\boldsymbol{X}_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.4}$$

and the nonparametric additive model

$$Y_i = \sum_{k=1}^{d} f_k(X_{i,k}) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.5}$$

in both fixed and random designs. Here, $\boldsymbol{X}_i := (X_{i,1}, \ldots, X_{i,d})^\top$, $i = 1, \ldots, n$, for some fixed positive integer d. Regarding the fixed design, we consider two types, namely, the grid design (GD) and the diagonal design (DD). With a total of n design points, the former places them on a regular grid in the d-dimensional cube $[0,1]^d$ while the latter only places design points on the diagonal. Details are given in Sections 2.4.1 and 2.4.2.

We summarize the minimax rates in all of the aforementioned models in Table 2.1.

The rest of the paper is organized as follows. Section 2.2 presents the simple model (2.2) with constant variance. Section 2.3 discusses its heteroscedastic extension (2.1). Section 2.4 discusses the multivariate nonparametric regression model (2.4), the additive model (2.5), and several other extensions of our main results. The essential lower bound proof of the minimax rate $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ under model (2.2) is presented in Section A.1, with the rest of the proofs given in a supplement.

The notation used throughout the paper is as follows. For any positive integer n, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. For any real number a, we use $\lceil a \rceil$ to denote the smallest integer greater than or equal to a, and $\lfloor a \rfloor$ the largest integer strictly smaller than a. For any positive integer d, $\mathbf{0}_d$ denotes the zero vector of dimension d and $\mathbf{I}_d$ denotes the identity matrix of dimension d. For a real vector x, $\|x\|$ and $\|x\|_\infty$ denote its Euclidean and infinity norms, respectively. For a real matrix $\mathbf{A}$, we use $\|\mathbf{A}\|$, $\|\mathbf{A}\|_F$, and $|\mathbf{A}|$ to denote its spectral norm, Frobenius norm, and determinant, respectively. For an m-times differentiable function $f : \mathbb{R} \to \mathbb{R}$ with some positive integer m, we use $f^{(k)}$ to denote its kth derivative for $k = 1, 2, \ldots, m$. For identically distributed random variables X_i and X_j, we use $\mathbb{P}_{X_i}(\cdot)$ and $p_{X_i}(\cdot)$ to denote the distribution and density of X_i, $\widetilde{X}_{ij}$ to denote $X_i - X_j$, and $p_{\widetilde{X}_{ij}}(\cdot)$ to denote the density of $X_i - X_j$. Similar notation $\mathbb{P}_{\boldsymbol{X}_i}(\cdot), p_{\boldsymbol{X}_i}(\cdot), \widetilde{\boldsymbol{X}}_{ij}, p_{\widetilde{\boldsymbol{X}}_{ij}}(\cdot)$ applies to identically distributed random vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. For a positive integer d and $\boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the d-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. We will drop the subscript d for simplicity when $d = 1$. $\Phi(\cdot)$ and $\varphi(\cdot)$ represent the standard

Table 2.1: Summary of minimax rates in (2.1), (2.2), (2.4) and (2.5). The two types of fixed design considered, (GD) and (DD), are defined in (2.20) and (2.21), respectively. For a d-dimensional smoothness index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$, $\underline{\alpha} := d/(\sum_{k=1}^{d} 1/\alpha_k)$, $\alpha_{\min} := \min_{1\le k\le d} \alpha_k$, and $\alpha_{\max} := \max_{1\le k\le d} \alpha_k$. The respective sections contain the definition of the distribution class of $\{(X_i, \varepsilon_i)\}_{i=1}^n$ in the random design setting and distribution class of $\{\varepsilon_i\}_{i=1}^n$ in the fixed design setting. Our results include all of the random design rates and fixed design rates in (2.4) and (2.5). Note results for (2.4) and (2.5) have additional requirements; see Sections 2.4.1 and 2.4.2 for details.

	stated in	minimax rate	boundary
(2.1), fixed	Wang et al. (2008)	$n^{-4\alpha} \vee n^{-2\beta/(2\beta+1)}$	$\alpha = \beta/(4\beta+2)$
(2.1), random	Theorems 3, 4, 5	$n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}} \vee n^{-\frac{2\beta}{2\beta+1}}$	
(2.2), fixed	Wang et al. (2008)	$n^{-4\alpha} \vee n^{-1}$	$\alpha = 1/4$
(2.2), random	Theorems 1, 2	$n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$	
(2.4), fixed (GD)	Proposition 3	$n^{-4\alpha_{\max}/d} \vee n^{-1}$	$\alpha_{\max} = d/4$
(2.4), fixed (DD)	Proposition 4	$n^{-4\alpha_{\min}} \vee n^{-1}$	$\alpha_{\min} = 1/4$
(2.4), random	Propositions 1, 2	$n^{-8\underline{\alpha}/(4\underline{\alpha}+d)} \vee n^{-1}$	$\underline{\alpha} = d/4$
(2.5), fixed (GD)	Proposition 5	n^{-1}	
(2.5), fixed (DD)	Proposition 6	$n^{-4\alpha_{\min}} \vee n^{-1}$	$\alpha_{\min} = 1/4$
(2.5), random	Propositions 7, 8	$n^{-8\alpha_{\min}/(4\alpha_{\min}+1)} \vee n^{-1}$	

normal distribution and density. More generally, we will write $\varphi_{\mu,\sigma^2}(\cdot)$ as the density for the normal distribution with mean μ and variance σ^2. For two probability measures $\mathbb{P}, \mathbb{Q}$ defined on a common space $(\Omega, \mathcal{A})$, $\mathsf{TV}(\mathbb{P}, \mathbb{Q})$ denotes their total variation distance, that is, $\mathsf{TV}(\mathbb{P}, \mathbb{Q}) := \sup_{A\in\mathcal{A}} |*|\mathbb{P}(A) - \mathbb{Q}(A)$. For two real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim b_n$ if $|a_n| \le C|b_n|$ for some positive absolute constant C. We say $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

2.2 Homoscedastic case

To illustrate some of the main ideas developed in this paper, we begin with a discussion of the elementary univariate homoscedastic nonparametric regression model (2.2):

$$Y_i = f(X_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n.$$

Here, $\{X_i\}_{i=1}^n$ are i.i.d. copies of a univariate random variable X, $f(\cdot)$ belongs to an α-Hölder class that will be specified soon, and $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. copies of a variable ε with zero mean and unit variance and are independent of $\{X_i\}_{i=1}^n$. Both the mean function $f(\cdot)$ and the distribution of $\{X_i\}_{i=1}^n$ are assumed unknown.

Model (2.2) has been extensively studied using residual-based and difference-based methods; see, among many others, Von Neumann (1941), Von Neumann (1942), Rice (1984), Gasser et al. (1986), Hall et al. (1990), Hall and Marron (1990), Thompson et al. (1991), Müller et al. (2003), Wang et al. (2008). A related functional estimation problem has also been studied in semiparametric models (Robins et al., 2008, 2009). Most of the previous studies focus on the case of fixed design, especially the equidistant design with $X_i = i/n$, $i \in [n]$, for which the minimax rate of estimating σ^2 under an α-Hölder smoothness constraint on $f(\cdot)$ is known to be $n^{-4\alpha} \vee n^{-1}$ (cf. Theorems 1 and 2 in Wang et al. (2008)).

In detail, let I be a fixed (possibly infinite) interval on the real line. Define the Hölder class $\Lambda_{\alpha,I}(C_{\mathcal{F}})$ on I as follows:

$$\begin{aligned}\Lambda_{\alpha,I}(C_{\mathcal{F}}) := \Big\{ f : &\text{ for all } x, y \in I \text{ and } k = 0, \ldots, \lfloor \alpha \rfloor, \\ & |*|f^{(k)}(x) \le C_{\mathcal{F}} \text{ and } |*|f^{(\lfloor \alpha \rfloor)}(x) - f^{(\lfloor \alpha \rfloor)}(y) \le C_{\mathcal{F}} |x-y|^{\alpha'} \Big\},\end{aligned} \tag{2.6}$$

where $\alpha' := \alpha - \lfloor \alpha \rfloor$. Denote the support of X as $\mathrm{supp}(X)$.

Define the joint distribution class $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$ (where "cv" stands for "constant variance") with the following conditions:

(a) X satisfies $\mathrm{supp}(X) \subset I$.

(b) X has density $p_X(\cdot)$ and there exists a fixed positive constant C_0 such that

$$\sup_{x\in\mathbb{R}} p_X(x) \le C_0.$$

(c) There exist two fixed constants $\delta_0 > 0$ and $c_0 > 0$ such that for any $0 < \delta < \delta_0$, there exists a set $\mathcal{U}_\delta \subset [-1, 1]$ such that

$$\lambda(\mathcal{U}_\delta) \ge c_0 \quad \text{and} \quad \inf_{u\in\mathcal{U}_\delta} p_{\widetilde{X}_{ij}}(u\delta) \ge c_0,$$

where $\lambda(\cdot)$ represents the Lebesgue measure on the real line, and $\widetilde{X}_{ij} = X_i - X_j$.

(d) $\mathbb{E}\varepsilon^4 \le C_\varepsilon$ for some fixed positive constant C_ε.

Note that no smoothness condition is placed on the density of X. Condition (c) essentially requires the density $p_{\widetilde{X}_{ij}}$ to be "dense" around 0, and is strictly weaker than a uniform lower bound of $p_{\widetilde{X}_{ij}}$ over a fixed neighborhood of 0. It also follows from the following sufficient condition on the marginal density $p_X(\cdot)$ (see Lemma 11 in the supplement for the justification):

(c$'$) X is compactly supported (taken to be $[0, 1]$ without loss of generality). There exists some positive constant c_0 and subset $S \subset [-1, 1]$ with Lebesgue measure $\lambda(S) \ge 3/4$ such that $p_X(t) \ge c_0$ uniformly over $t \in S$.

In particular, (c$'$) covers the uniform distribution on $[0, 1]$ and the distribution of X in the lower bound construction in the proof of Theorem 2.

The rest of the section is devoted to proving, for any fixed positive constants $C_\mathcal{F}$ and C_σ, the following minimax rate:

$$\inf_{\widetilde{\sigma}^2} \sup_{f\in\Lambda_{\alpha,I}(C_\mathcal{F})} \sup_{\sigma^2\le C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \asymp n^{-8\alpha/(4\alpha+1)} \vee n^{-1}, \tag{2.7}$$

where $\mathbb{P}_{(X,\varepsilon)}$ denotes the joint distribution of (X, ε), and $\widetilde{\sigma}^2$ ranges over all estimators of σ^2.

2.2.1 Upper bound

The upper bound is achieved by a difference estimator based on U-statistics (with convention $0/0 = 0$):

$$\widehat{\sigma}^2 := \frac{\binom{n}{2}^{-1} \sum_{i<j} K_h(X_i - X_j)(Y_i - Y_j)^2/2}{\binom{n}{2}^{-1} \sum_{i<j} K_h(X_i - X_j)}. \tag{2.8}$$

Here, $K_h(\cdot) := K(\cdot/h)/h$, where $h = h_n$ is a bandwidth parameter satisfying $h_n \downarrow 0$ as $n \to \infty$, and $K(\cdot)$ is a symmetric density kernel supported on $[-1, 1]$ that satisfies

$$\underline{M}_K \le \inf_{|u|\le 1} K(u) \le \sup_{|u|\le 1} K(u) \le \overline{M}_K \tag{2.9}$$

for two fixed constants $\overline{M}_K$ and $\underline{M}_K$; one example is the box kernel $K(u) = \mathbb{1}\{|u| \le 1\}/2$ which satisfies (2.9) with $\overline{M}_K = \underline{M}_K = 1/2$.

The following error bound is derived via the exponential inequality for degenerate U-statistics due to Giné et al. (2000).

Theorem 1. *Suppose the kernel $K(\cdot)$ in $\widehat{\sigma}^2$ is chosen such that* (2.9) *is satisfied with constants $\overline{M}_K$ and $\underline{M}_K$, and the bandwidth h_n is chosen as*

$$h_n \asymp \begin{cases} n^{-2/(4\alpha+1)}, & 0 < \alpha < 1/4, \\ n^{-1}, & \alpha \ge 1/4. \end{cases} \tag{2.10}$$

Then, under (2.2) *with random design, it holds that*

$$\sup_{f \in \Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{\sigma^2 \le C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}_{cv,(X,\varepsilon)}} \mathbb{E}\big(\widehat{\sigma}^2 - \sigma^2\big)^2 \le C\big(n^{-8\alpha/(4\alpha+1)} \vee n^{-1}\big),$$

where C is some fixed positive constant that only depends on $\overline{M}_K, \underline{M}_K$, $\alpha, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{cv,(X,\varepsilon)}$.

Remark 1. *The error rate in Theorem 1 is achieved by choosing the optimal bandwidth* h_n *to balance the "bias-variance" decomposition:*

$$\left\{\mathbb{E}\big(\widehat{\sigma}^2 - \sigma^2\big)^2\right\}^{1/2} \lesssim h_n^{2(\alpha\wedge 1)} + \frac{1}{nh_n^{1/2}}, \tag{2.11}$$

where $a \wedge b := \min\{a, b\}$ *for any two real numbers* a, b*. The bias term* $h_n^{2(\alpha\wedge 1)}$ *reflects the second-order effect of the unknown mean on variance estimation, which has been noted by Hall and Carroll (1989) and Wang et al. (2008). The variance part follows from the fact that there is an average number of* $n^2 h_n$ *pairs of* (i, j) *such that* $|X_i - X_j| \leq h_n$*. We note that the same "bias-variance" decomposition has appeared in quadratic functional estimation in the density model and Gaussian sequence model (Bickel and Ritov, 1988; Fan, 1991; Giné and Nickl, 2008). See Section 2.4.3 for a more detailed discussion.*

Remark 2. *While most of the previous works are in the context of fixed design, Müller et al. (2003) considered constant variance estimation with random design, and their estimator (formula (1.4) therein) is almost identical to our* $\widehat{\sigma}^2$*. Under certain assumptions (Assumptions 1 and 2 and (2.4) - (2.7) therein), they show that their estimator is root-n consistent and asymptotically normal. However, as commented in the first paragraph on p. 184 of their paper, their condition (2.7) is only satisfied when the mean function smoothness* α *is strictly larger than* $1/4$*, and no analysis is provided below this threshold. Our minimax rate* $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ *therefore confirms that* $\alpha \geq 1/4$ *is indeed the minimal requirement for any variance estimator to be root-n consistent and we also demonstrate the optimality of* $\widehat{\sigma}^2$ *for* $0 < \alpha < 1/4$*.*

Finally, in (2.2), we have assumed that the smoothness index α is known. If it is unknown, then the variance can be estimated adaptively via Lepski-type methods (Lepski, 1991; Lepskiĭ, 1991). This is discussed in more detail in Section 2.4.5.

2.2.2 Lower bound

The derivation of the lower bound in (2.7) is much more involved. In particular, the construction in the fixed design setting (cf. Theorem 2 in Wang et al. (2008)) cannot be extended to the random design case, since the spike-type construction of $f(\cdot)$ located at each deterministic design point leads to a sub-optimal rate in the random design setting. To achieve a sharp rate, we have to exploit the randomness of $\{X_i\}_{i=1}^n$; this requires us to handle a highly convoluted alternative hypothesis that no longer leads to a product measure of $\{Y_i\}_{i=1}^n$ given each realization of $\{X_i\}_{i=1}^n$ in LeCam's two-point method. This calls for a careful analysis of the locations of $\{X_i\}_{i=1}^n$.

We now sketch a proof of the $n^{-8\alpha/(4\alpha+1)}$ component in (2.7) for $0 < \alpha < 1/4$, with a particular emphasis on where the difference arises with the fixed design setting. The proof can be roughly divided into two steps. In the first step, we construct a two-point testing problem with the null being a Gaussian (H_0) and the alternative a Gaussian location mixture ($\widetilde{H}_1$). In the second step, we approximate the Gaussian location mixture ($\widetilde{H}_1$) by a location mixture with compact support (H_1), which, unlike the alternative in the first step, belongs to the considered model class.

We start by introducing the construction of $f(\cdot)$, σ^2, ε, and X under the null H_0 and the alternative $\widetilde{H}_1$ in the first step. For each n, let

$$h_n \asymp n^{-2/(4\alpha+1)}, \quad \theta_n^2 \asymp h_n^{2\alpha}, \qquad \text{and} \qquad N := 1/(6h_n),$$

and divide the unit interval $[0, 1]$ into N intervals of length $6h_n$, with n large enough and h_n chosen such that N is a positive integer.

> *Choice of* $f(\cdot)$: Under H_0, let $f \equiv 0$. Under $\widetilde{H}_1$, let $f(\cdot)$ be a piecewise trapezoidal function on the N intervals. That is, for each $i \in [N]$, f takes on a value of $h_n^\alpha \widetilde{r}_i$ on the intervals $[(6i-5)h_n, (6i-1)h_n]$ and then linearly decreases to zero on the two endpoints $6(i-1)h_n$ and $6ih_n$, with $\{\widetilde{r}_i\}_{i=1}^N$ i.i.d. standard normal variables.

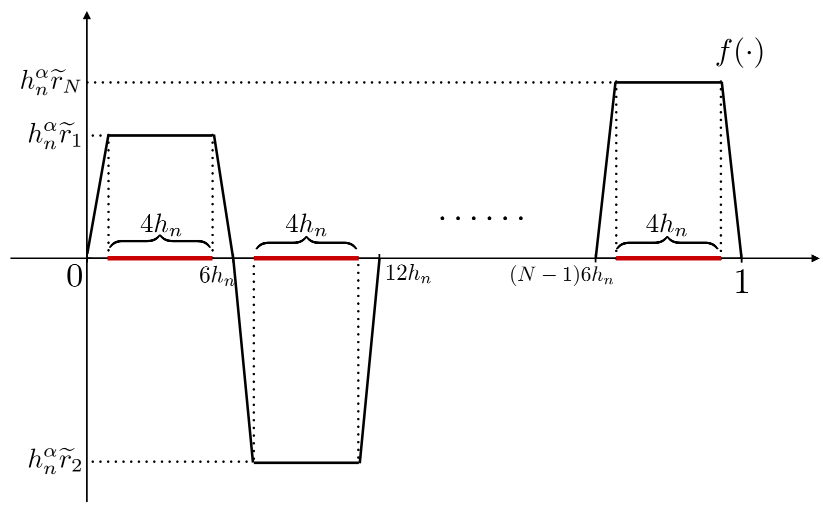

Figure 2.1: The black solid line represents the construction of $f(\cdot)$ under the alternative hypothesis $\widetilde{H}_1$. The thick red segments indicate the support of X under both H_0 and $\widetilde{H}_1$, on which X is uniformly distributed. Here, $h_n \asymp n^{-2/(4\alpha+1)}$ and is chosen such that $N := 1/(6h_n)$ is a positive integer. $\{\widetilde{r}_i\}_{i=1}^N$ are N i.i.d. standard normal variables.

Choice of σ^2: Under H_0, let $\sigma^2 = 1 + \theta_n^2$. Under $\widetilde{H}_1$, let $\sigma^2 = 1$.

Choice of ε: Under both H_0 and $\widetilde{H}_1$, let $\varepsilon \sim \mathcal{N}(0, 1)$.

Choice of X: Under both H_0 and $\widetilde{H}_1$, let $\{X_i\}_{i=1}^n$ be uniformly distributed over the union of the upper bases of the trapezoids, that is, over $\bigcup_{i=1}^N [(6i-5)h_n, (6i-1)h_n]$.

See Figure 2.1 for an illustration of the construction.

In contrast to the spike-type construction of $f(\cdot)$ in the fixed design setting, our construction is trapezoid-shaped, which guarantees a maximal variation in the mean to compensate for the difference in the variance under the null and alternative. This is unnecessary in the fixed design setting since the point of maximal variation in the mean (center of each spike) can be directly placed at each fixed $X_i = i/n$, resulting in n evenly spaced spikes in $f(\cdot)$.

Denote the joint distribution of $\{(X_i, Y_i)\}_{i=1}^n$ under H_0 and $\widetilde{H}_1$ by $\mathbb{P}_0$ and $\widetilde{\mathbb{P}}_1$ with respective density p_0 and $\widetilde{p}_1$. Under the above construction, conditional on $\{X_i\}_{i=1}^n$, $\{Y_i\}_{i=1}^n$ are distributed as

$$H_0 : p_0(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n) = \prod_{i=1}^n \varphi_{0,1+\theta_n^2}(Y_i)$$

and

$$\widetilde{H}_1 : \widetilde{p}_1(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n) = \prod_{j=1}^N \int \left(\prod_{\{i:b_i=j\}} \varphi_{h_n^\alpha v,1}(Y_i) \right) \varphi(v) dv,$$

where $\{b_i\}_{i=1}^n$ is the location index sequence of $\{X_i\}_{i=1}^n$ defined as

$$b_i := j \quad \text{if} \quad X_i \in [(6j-5)h_n, (6j-1)h_n],$$

which characterizes which trapezoid each X_i falls into. Using Lemma 6 that will be stated in Section A.1, one can then upper bound

$$\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) = \mathbb{E}\mathsf{TV}(\mathbb{P}_0(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n), \widetilde{\mathbb{P}}_1(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n)) \lesssim \theta_n^2 n h_n^{1/2},$$

which can be made smaller than a sufficiently small constant c by choosing h_n sufficiently small.

The second step of the proof aims to find a sequence of bounded random variables $\{r_i\}_{i=1}^N$ to replace the standard normal sequence $\{\widetilde{r}_i\}_{i=1}^N$ in $\widetilde{\mathbb{P}}_1$, so that for each realization of $\{r_i\}_{i=1}^N$, the corresponding $f(\cdot)$ in the alternative is α-Hölder smooth with a fixed constant. Then, denoting the distribution of $\{r_i\}_{i=1}^N$ as $\mathbb{G}$, one wishes to approximate the conditional distribution $\widetilde{\mathbb{P}}_1(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n)$ in $\widetilde{H}_1$ by $\mathbb{P}_1(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n)$ with density

$$p_1(\{Y_i\}_{i=1}^n \mid \{X_i\}_{i=1}^n) = \prod_{j=1}^N \int \left(\prod_{\{i:b_i=j\}} \varphi_{h_n^\alpha v,1}(Y_i) \right) \mathbb{G}(dv)$$

in H_1. Even with the aid of moment matching techniques already established in the literature, upper bounding $\mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1)$ is still nontrivial. Specifically, unlike in the fixed design setting, now with high probability the conditional distribution of $\{Y_i\}_{i=1}^n$ given $\{X_i\}_{i=1}^n$ is no longer a product measure. This is because multiple X_i's could fall into the same trapezoid in the construction of $f(\cdot)$. This can be handled relatively easily in the first step since there we only have to analyze the pairwise correlation of $Y_i \mid X_i$ and $Y_j \mid X_j$ depending on whether X_i and X_j fall into the same trapezoid, but it is much less tractable in the second step. More specifically, in order to match moments, we now have to divide the X_i's into groups based on their memberships among the trapezoids, which naturally requires us to monitor the locations of $\{X_i\}_{i=1}^n$, and in particular the number of X_i's that fall into the same trapezoid. This is possible by observing that the memberships of $\{X_i\}_{i=1}^n$ now follow a sparse multinomial distribution ($n^{2/(4\alpha+1)}$ bins, n balls) so that a result in Kolchin et al. (1978) can be applied. This allows us to show that with high probability the maximum number of X_i's in each trapezoid is bounded by a fixed constant, which, along with Lemma 5 in Section A.1, allows us to calculate

$$\mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) \lesssim n\theta_n^{2p}$$

for $p := 1 + \lceil 1/4\alpha \rceil$. This indicates that $\mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1)$ is smaller than some sufficiently small constant c. Then, by the triangle inequality,

$$\mathsf{TV}(\mathbb{P}_0, \mathbb{P}_1) \leq \mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) + \mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) \leq 2c.$$

Details of the above derivation will be given in Section A.1. The resulting lower bound is as follows.

Theorem 2. *Under* (2.2) *with random design, it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f \in \Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}_{cv,(X,\varepsilon)}} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \geq c\big(n^{-8\alpha/(4\alpha+1)} \vee n^{-1}\big),$$

where c is some fixed positive constant that only depends on $\alpha, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{cv,(X,\varepsilon)}$, and $\widetilde{\sigma}^2$ ranges over all estimators of σ^2.

Remark 3. *It remains an open problem to prove a lower bound rate that is strictly slower than n^{-1} over the sub-class of $\mathcal{P}_{cv,(X,\varepsilon)}$ with more regular designs, which includes in particular the uniform design on $[0,1]$. We conjecture that in this case, $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ is still the minimax rate in view of analogous results in quadratic functional estimation (Bickel and Ritov, 1988; Fan, 1991).*

2.3 Heteroscedastic case

We now study the heteroscedastic model (2.1),

$$Y_i = f(X_i) + V^{1/2}(X_i)\varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\{X_i\}_{i=1}^n$ are i.i.d. copies of X on the real line, $f(\cdot)$ and $V(\cdot)$ are α- and β-Hölder smooth on the fixed (possibly infinite) interval I, respectively, and $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. copies of ε with zero mean and unit variance and are independent of $\{X_i\}_{i=1}^n$. As in Section 2.2, smoothness indices α and β are assumed known, while $f(\cdot), V(\cdot)$, and the distribution of X are unknown. For any estimator $\widetilde{V}(\cdot)$, the estimation accuracy is measured both locally via

$$R_1(\widetilde{V}, V; x^*) := \left(\widetilde{V}(x^*) - V(x^*)\right)^2 \tag{2.12}$$

at a point x^* in the support of X, $\mathrm{supp}(X)$, and globally via

$$R_2(\widetilde{V}, V) := \int \left(\widetilde{V}(x) - V(x)\right)^2 \mathbb{P}_X(dx) \tag{2.13}$$

with $\mathbb{P}_X$ the distribution of X.

Model (2.1) has been studied in, for example, Muller and Stadtmuller (1987), Hall and Carroll (1989), Ruppert et al. (1997), Härdle and Tsybakov (1997), Fan and Yao (1998), Munk and Ruymgaart (2002), Brown and Levine (2007), Wang et al. (2008), with a focus mainly on the fixed design case. An exception is Munk and Ruymgaart (2002), with which we draw a detailed comparison in Remark 8 below. Theorems 1 and 2 in Wang et al. (2008)

established a minimax rate of the order $n^{-4\alpha} \vee n^{-2\beta/(2\beta+1)}$ under equidistance design $X_i = i/n$, $i \in [n]$ when $f(\cdot)$ and $V(\cdot)$ are α- and β-Hölder smooth on [0,1].

Define $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ (where "vf" stands for "variance function") as follows:

(a) X satisfies $\mathrm{supp}(X) \subset I$.

(b) X has density $p_X(\cdot)$, and there exists a fixed positive constant C_0 such that

$$\sup_{x \in \mathbb{R}} p_X(x) \leq C_0.$$

(c) There exist fixed positive constants c_0 and δ_0 such that

$$\inf_{x^* \in \mathrm{supp}(X)} p_X(x^*) \geq c_0 \quad \text{ and }$$
$$\inf_{0<\delta<\delta_0} \inf_{x^* \in \mathrm{supp}(X)} \lambda(\{u \in [-1,1] : x^* + \delta u \in \mathrm{supp}(X)\}) \geq c_0,$$

where $\lambda(\cdot)$ is the Lebesgue measure on the real line.

(d) $\mathbb{E}\varepsilon^4 \leq C_\varepsilon$ for some fixed positive constant C_ε.

One can readily verify that $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)} \subset \mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$, with the latter defined in the beginning of Section 2.2. Compared to $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$, Condition (c) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ is posed on the marginal density and support of X, since in the variance function case we require a sufficient number of close pairs (X_i, X_j) around each target x^*. We also note that, as in $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$, no smoothness assumption is posed on the design density in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$.

The rest of the section is devoted to proving, for any fixed positive constants $C_{\mathcal{F}}$ and $C_{\mathcal{V}}$, the following minimax rates

$$\begin{aligned}
\inf_{\widetilde{V}} \sup_{f \in \Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{V \in \Lambda_{\beta,I}(C_{\mathcal{V}})} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}_{\mathrm{vf},(X,\varepsilon)}} \sup_{x^* \in \mathrm{supp}(X)} \mathbb{E} R_1(\widetilde{V}, V; x^*) &\asymp n^{-\frac{8\alpha\beta}{4\alpha\beta+2\alpha+\beta}} \vee n^{-\frac{2\beta}{2\beta+1}}, \\
\inf_{\widetilde{V}} \sup_{f \in \Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{V \in \Lambda_{\beta,I}(C_{\mathcal{V}})} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}_{\mathrm{vf},(X,\varepsilon)}} \mathbb{E} R_2(\widetilde{V}, V) &\asymp n^{-\frac{8\alpha\beta}{4\alpha\beta+2\alpha+\beta}} \vee n^{-\frac{2\beta}{2\beta+1}},
\end{aligned} \tag{2.14}$$

where $\mathbb{P}_{(X,\varepsilon)}$ denotes the joint distribution of (X, ε), and $\widetilde{V}(\cdot)$ ranges over all estimators of $V(\cdot)$.

2.3.1 Upper bound

We now propose an estimator of $V(x^*)$ for some fixed $x^* \in \mathrm{supp}(X)$ by combining pairwise differences with local polynomial regression. We first introduce some notation. Let ℓ be the largest integer strictly smaller than β and

$$\boldsymbol{q}(u) := (1, u, u^2/2!, \ldots, u^\ell/\ell!)^\top.$$

For any $1 \le i < j \le n$, define

$$D_{ij} := (Y_i - Y_j)^2/2,\ X_{ij} := (X_i + X_j)/2,\ \text{and}\ K_{ij} := K_{h_1}(X_i - X_j)K_{h_2}(X_{ij} - x^*),$$

where h_1, h_2 are two bandwidths. Define an $(\ell+1)\times(\ell+1)$ matrix

$$\mathbf{B}_n := \binom{n}{2}^{-1} \sum_{i<j} \boldsymbol{q}\Big(\frac{X_{ij} - x^*}{h_2}\Big)\boldsymbol{q}^\top\Big(\frac{X_{ij} - x^*}{h_2}\Big)K_{ij}$$

and $\mathbf{B}_n^*$ as its adjugate such that $\mathbf{B}_n\mathbf{B}_n^* = \mathbf{B}_n^*\mathbf{B}_n = |\mathbf{B}_n|\mathbf{I}_{\ell+1}$. For example, when $\ell = 1$, we have

$$\mathbf{B}_n = \begin{bmatrix} s_0 & s_1 \\ s_1 & s_2 \end{bmatrix}, \quad \mathbf{B}_n^* = \begin{bmatrix} s_2 & -s_1 \\ -s_1 & s_0 \end{bmatrix}, \quad \text{and} \quad |\mathbf{B}_n| = s_0 s_2 - s_1^2,$$

where

$$s_k := \binom{n}{2}^{-1} \sum_{i<j} \Big(\frac{X_{ij} - x^*}{h_2}\Big)^k K_{ij}, \qquad k = 0, 1, 2.$$

Following Fan (1993), we propose a robust local polynomial estimator:

$$\widehat{V}_{\mathrm{LP}}(x^*) := \binom{n}{2}^{-1} \sum_{i<j} D_{ij}(|\mathbf{B}_n| + \tau_n)^{-1}\boldsymbol{q}^\top(0)\mathbf{B}_n^*\boldsymbol{q}\Big(\frac{X_{ij} - x^*}{h_2}\Big)K_{ij}, \tag{2.15}$$

where τ_n is some sufficiently small positive constant that decays to 0 polynomially with n. Let

$$w_{ij} := \binom{n}{2}^{-1} \boldsymbol{q}^\top(0)\mathbf{B}_n^*\boldsymbol{q}\Big(\frac{X_{ij} - x^*}{h_2}\Big)K_{ij} \quad \text{and} \quad \widetilde{w}_{ij} := w_{ij}/(|\mathbf{B}_n| + \tau_n).$$

Then, it holds that $\widehat{V}_{\mathrm{LP}}(x^*) = \sum_{i<j} \widetilde{w}_{ij} D_{ij}$, $\sum_{i<j} w_{ij} = |\mathbf{B}_n|$, and

$$\sum_{i<j} w_{ij}(X_{ij} - x^*)^k = \sum_{i<j} \widetilde{w}_{ij}(X_{ij} - x^*)^k = 0, \qquad k = 1, 2, \ldots, \ell. \tag{2.16}$$

The last property (2.16) is referred to as the *reproducing property* of local polynomial estimators (cf. Proposition 1.12 in Tsybakov (2009a)).

Theorem 3. *Suppose the kernel $K(\cdot)$ in $\widehat{V}_{LP}$ is chosen such that* (2.9) *holds with constants $\overline{M}_K$ and $\underline{M}_K$, $\tau_n \asymp n^{-\kappa}$ for some fixed constant $\kappa \geq 1$, and the bandwidths h_1, h_2 are chosen as*

$$(h_1, h_2) \asymp \begin{cases} \left(n^{-\frac{2\beta}{4\alpha\beta+\beta+2\alpha}}, n^{-\frac{4\alpha}{4\alpha\beta+\beta+2\alpha}}\right), & 0 < \alpha < \frac{\beta}{4\beta+2}, \\ \left(n^{-1}, n^{-\frac{1}{2\beta+1}}\right), & \alpha \geq \frac{\beta}{4\beta+2}. \end{cases} \tag{2.17}$$

Then, under (2.1) *with random design, it holds that*

$$\sup_{f\in\Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{V\in\Lambda_{\beta,I}(C_{\mathcal{V}})} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{vf,(X,\varepsilon)}} \sup_{x^*\in\mathrm{supp}(X)} \mathbb{E}R_1(\widehat{V}_{LP}, V; x^*) \leq C\left(n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}} \vee n^{-\frac{2\beta}{2\beta+1}}\right)$$

and

$$\sup_{f\in\Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{V\in\Lambda_{\beta,I}(C_{\mathcal{V}})} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{vf,(X,\varepsilon)}} \mathbb{E}R_2(\widehat{V}_{LP}, V) \leq C\left(n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}} \vee n^{-\frac{2\beta}{2\beta+1}}\right),$$

where C is some fixed positive constant that only depends on $\overline{M}_K, \underline{M}_K, \alpha, \beta, C_{\mathcal{F}}, C_{\mathcal{V}}$ and C_0, c_0, C_ε in $\mathcal{P}_{vf,(X,\varepsilon)}$.

Remark 4. *Variance function estimation in* (2.1) *with fixed design $X_i = i/n$, $i \in [n]$, has been studied in Wang et al. (2008). There the minimax rate is*

$$\inf_{\widetilde{V}} \sup_{f\in\Lambda_{\alpha,[0,1]}(C_{\mathcal{F}})} \sup_{V\in\Lambda_{\beta,[0,1]}(C_{\mathcal{V}})} \sup_{\mathbb{E}\varepsilon^4\leq C_\varepsilon} \sup_{x^*\in[0,1]} \mathbb{E}R_1(\widetilde{V}, V; x^*) \asymp n^{-4\alpha} \vee n^{-2\beta/(2\beta+1)},$$

$$\inf_{\widetilde{V}} \sup_{f\in\Lambda_{\alpha,[0,1]}(C_{\mathcal{F}})} \sup_{V\in\Lambda_{\beta,[0,1]}(C_{\mathcal{V}})} \sup_{\mathbb{E}\varepsilon^4\leq C_\varepsilon} \mathbb{E}R_2(\widetilde{V}, V) \asymp n^{-4\alpha} \vee n^{-2\beta/(2\beta+1)},$$

with the integral in R_2 under the Lebesgue measure on $[0,1]$. Comparing the above result with the error rate in Theorem 3, we see that the transition boundary in both the fixed and random design settings is $\alpha = \beta/(4\beta+2)$. When $\alpha \geq \beta/(4\beta+2)$, $V(\cdot)$ under both R_1 and R_2 can be estimated at the classic nonparametric rate $n^{-2\beta/(2\beta+1)}$ as if the mean function $f(\cdot)$ were known. When $\alpha < \beta/(4\beta+2)$, a faster rate can be achieved in the random design case. This can be intuitively understood by the fact that, by constrast to the fixed design case, a significant portion of pairs have distance smaller than $1/n$ in the random design setting.

Remark 5. *As has been noted in Wang et al. (2008), in the fixed design setting, estimating the variance (function) by smoothing the squared residuals obtained from pre-estimation of the mean function $f(\cdot)$ is sub-optimal. The same conclusion also applies to the random design setting. Since the design being fixed or random has no first-order effect on the estimation of the mean, the above method only achieves the rates $n^{-4\alpha/(2\alpha+1)} \vee n^{-1}$ in variance estimation and $n^{-4\alpha/(2\alpha+1)} \vee n^{-2\beta/(2\beta+1)}$ in variance function estimation, neither of which is minimax optimal.*

Remark 6. *Unlike in the fixed design case, once below the threshold $\alpha = \beta/(4\beta+2)$, α and β are now both present in the minimax rate in the random design case, suggesting that the smoothness of $V(\cdot)$ always has an effect on its estimation. This is because variance function estimation in the random design setting is essentially a "two-dimensional" problem, where we have to jointly choose two optimal neighborhood sizes to characterize the closeness between (i) each X_i and X_j; and (ii) every pair (X_i, X_j) and each target point x^*. By contrast, in the fixed design setting, the distance between X_i and X_j is constrained to be no smaller than $1/n$, and thus cannot be jointly optimized with the distance between (X_i, X_j) and x^*.*

Remark 7. *One might wonder whether the following Nadaraya-Watson type estimator can be used to establish the upper bound in Theorem 3:*

$$\widehat{V}_{NW}(x^*) := \frac{\sum_{i<j} K_{h_1}(X_i - X_j) K_{h_2}(X_{ij} - x^*) D_{ij}}{\sum_{i<j} K_{h_1}(X_i - X_j) K_{h_2}(X_{ij} - x^*)}, \tag{2.18}$$

where $K(\cdot)$ is now chosen to be a higher-order kernel to further reduce bias when $\beta > 1$. It turns out that the analysis of $\widehat{V}_{NW}$ requires an extra assumption on the smoothness of the density $p_X(\cdot)$ which can be completely avoided with $\widehat{V}_{LP}$. Moreover, it is well-known that local polynomial estimators have good finite sample properties and boundary performances when X is compactly supported (Fan and Gijbels, 1995).

Remark 8. *Munk and Ruymgaart (2002) considered minimax estimation of the variance function (and more generally, its derivatives) in the context of nonparametric regression with random design. We focus on the comparison of their results on variance function estimation with ours. Their lower bound (Theorem 1 therein) is proved independent of the smoothness level of the mean function and upper bound (Theorem 4 therein) is proved under sufficient smoothness on the mean function. Therefore their minimax rate is only comparable to the $n^{-2\beta/(2\beta+1)}$ component in ours. In this case, their lower bound of the order $n^{-(2\beta-1)/(2\beta)}$ is proved over the following class of variance function:*

$$\mathcal{S}_\beta := \left\{ 1 + \sum_{k=1}^{\infty} \delta_k e_k : |\delta_k| \lesssim k^{-\beta} \right\}$$

for any $\beta > 1$, where $\{e_k\}_{k=1}^{\infty}$ is an arbitrary basis on $L^2([-\pi, \pi])$. Moreover, continuous differentiability of the error density is required in their paper. In contrast, we pose no smoothness conditions on the error density, and neither $\mathcal{S}_\beta$ nor $\mathcal{S}_{\beta+1/2}$ can be embedded in the β-Hölder class Λ_β considered in our setting (e.g., $f(x) = |x|$ with domain $[-\pi, \pi]$ belongs to $\mathcal{S}_2$ but is not 1.5- or 2-Hölder smooth since it is not differentiable at the origin). In summary, the results in Munk and Ruymgaart (2002) neither imply nor contradict the $n^{-2\beta/(2\beta+1)}$ part in our minimax rate, and our results are more refined since they characterize the exact elbow $\alpha = \beta/(4\beta + 2)$ and also the minimax rate below this threshold.

2.3.2 Lower bound

The following are matching lower bounds to Theorem 3.

Theorem 4. *Under* (2.1) *with random design, for any* $x^* \in \operatorname{supp}(X)$,

$$\inf_{\widetilde{V}} \sup_{f \in \Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{V \in \Lambda_{\beta,I}(C_{\mathcal{V}})} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}_{vf,(X,\varepsilon)}} \mathbb{E} R_1(\widetilde{V}, V; x^*) \geq c\Big(n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}} \vee n^{-\frac{2\beta}{2\beta+1}}\Big),$$

where c is some fixed positive constant that only depends on $\alpha, \beta, C_{\mathcal{F}}, C_{\mathcal{V}}$ and C_0, c_0, C_ε in $\mathcal{P}_{vf,(X,\varepsilon)}$, and $\widetilde{V}$ ranges over all estimators of V.

Theorem 5. *Under* (2.1) *with random design,*

$$\inf_{\widetilde{V}} \sup_{f \in \Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{V \in \Lambda_{\beta,I}(C_{\mathcal{V}})} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}_{vf,(X,\varepsilon)}} \mathbb{E} R_2(\widetilde{V}, V) \geq c\Big(n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}} \vee n^{-\frac{2\beta}{2\beta+1}}\Big),$$

where c is some fixed positive constant that only depends on $\alpha, \beta, C_{\mathcal{F}}, C_{\mathcal{V}}$ and C_0, c_0, C_ε in $\mathcal{P}_{vf,(X,\varepsilon)}$, and $\widetilde{V}$ ranges over all estimators of V.

Due to the appearances of both α and β in the nontrivial $n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}}$ part of the minimax rate, proving the above two results is more involved than proving Theorem 2. In particular, it takes an extra step of localization in the construction of the mean function $f(\cdot)$ as well as $V(\cdot)$. More precisely, for the lower bound at a target point x^* in Theorem 4, our construction of both $f(\cdot)$ and $V(\cdot)$ only has variation within a small neighborhood of x^*. Such localized construction is not necessary in the fixed design setting, since when proving the $n^{-4\alpha}$ component therein (see Remark 4), the variance function can simply be taken as a constant.

In what follows, we give a proof sketch of the nontrivial $n^{-8\alpha\beta/(4\alpha\beta+\beta+2\alpha)}$ component of the lower bound in Theorem 4 for $\alpha < \beta/(4\beta+2)$; the proof of Theorem 5 can be seen as an extension of Theorem 4 via a standard construction of multiple hypotheses. We assume the support of X is contained in $I = [0,1]$, and for clarity of illustration, here we present the construction for an interior point $x^* \in (0,1) \bigcap \operatorname{supp}(X)$. The proof works for boundary points as well.

We continue to adopt the two-step approach introduced in the proof sketch of Theorem 2 in Section 2.2.2. The second step is very similar with the help of Lemmas 5 and 7, so we

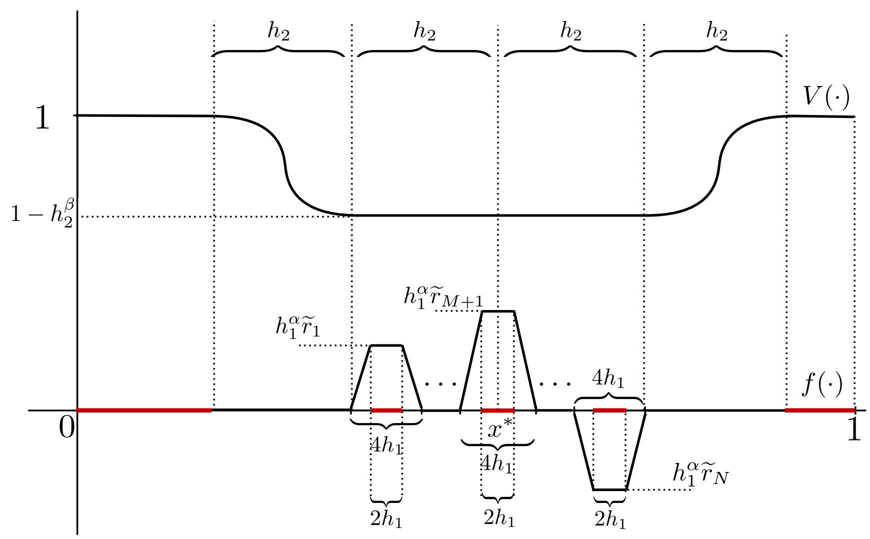

Figure 2.2: The black solid line on the top represents the variance function $V(\cdot)$ in the alternative $\widetilde{H}_1$, and the black solid line on the bottom represents the mean function $f(\cdot)$. The thick red segments mark the support of X under both H_0 and $\widetilde{H}_1$. Here, $h_1 \asymp n^{-\frac{2\beta}{4\alpha\beta+\beta+2\alpha}}$, $h_2 \asymp n^{-\frac{4\alpha}{4\alpha\beta+\beta+2\alpha}}$, and are chosen such that both $M := h_2/(4h_1) - 1/2$ and $N := 2M+1$ are positive integers. $\{\widetilde{r}_i\}_{i=1}^N$ are N i.i.d. standard normal variables.

will focus on the construction under the null H_0 and alternative $\widetilde{H}_1$ in the first step. Choose the parameters

$$h_1 \asymp n^{-\frac{2\beta}{4\alpha\beta+\beta+2\alpha}}, \quad h_2 \asymp n^{-\frac{4\alpha}{4\alpha\beta+\beta+2\alpha}}, \quad \text{and} \quad \theta_n^2 = h_1^{2\alpha} = h_2^\beta$$

so that $h_2/h_1 \to \infty$ as $n \to \infty$.

Choice of $V(\cdot)$: Under H_0 let $V \equiv 1$. Under $\widetilde{H}_1$, let $V(\cdot)$ be one minus a smooth bump function around x^* with width h_2 and height h_2^β so that $V(x^*) = 1 - \theta_n^2$.

Choice of $f(\cdot)$: Under H_0 let $f \equiv 0$. Under $\widetilde{H}_1$, let $f(\cdot)$ be a "local" version of the design in Theorem 2. That is, f takes on a value of 0 outside of $[x^* - h_2, x^* + h_2]$, and inside that h_2-neighborhood of x^*, f is piecewise trapezoidal with upper base length $2h_1$, lower base length $4h_1$ and height $\{h_1^\alpha \widetilde{r}_i\}_{i=1}^N$ for a standard normal sequence $\{\widetilde{r}_i\}_{i=1}^N$ with $N := h_2/(2h_1)$ a positive integer.

Choice of ε: Under both H_0 and $\widetilde{H}_1$, let $\varepsilon \sim \mathcal{N}(0, 1)$.

Choice of X: Under both H_0 and $\widetilde{H}_1$, let X be uniformly distributed on the union of $[0, 1]\backslash[x^* - h_2, x^* + h_2]$ and the upper bases of all the trapezoids inside $[x^* - h_2, x^* + h_2]$.

See Figure 2.2 for an illustration of $\widetilde{H}_1$.

Under the above construction, the squared distance between the null and alternative hypotheses $(1 - (1 - \theta_n^2))^2 = \theta_n^4 \asymp n^{-\frac{8\alpha\beta}{4\alpha\beta+\beta+2\alpha}}$ is the desired minimax rate. Using Lemma 6, we can show that

$$\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) \lesssim \theta_n^2 n h_1^{1/2} h_2^{1/2} \leq c$$

for some sufficiently small c, where $\mathbb{P}_0$ and $\widetilde{\mathbb{P}}_1$ represent the joint distribution of $\{(X_i, Y_i)\}_{i=1}^n$ under H_0 and $\widetilde{H}_1$, respectively. The detailed proof is presented in the supplement.

2.4 Discussion

The two univariate models (2.1) and (2.2) discussed in the previous two sections raise natural questions about possible extensions to the multivariate setting. In what follows, we first present some partial results in this direction in the sense of (2.4) and (2.5). We then establish some connections between our study and quadratic functional estimation and variance estimation in the linear model. Lastly, we discuss two more extensions of (2.2) in the direction of adaptive estimation and mean function with inhomogeneous smoothness. Throughout, consider $C_{\mathcal{F}}, C_\sigma, C_0, c_0, C_\varepsilon$ to be fixed positive constants.

2.4.1 Multivariate nonparametric regression

Consider the following multivariate version of (2.2):

$$Y_i = f(\boldsymbol{X}_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\{\boldsymbol{X}_i\}_{i=1}^n = \{(X_{i,1}, \ldots, X_{i,d})^\top\}_{i=1}^n$ are i.i.d. copies of $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ in $\mathbb{R}^d$ for some fixed positive integer d, $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. copies of ε with zero mean and unit variance and are independent of $\{\boldsymbol{X}_i\}_{i=1}^n$, and $f : \mathbb{R}^d \to \mathbb{R}$ belongs to a d-dimensional anisotropic Hölder class with smoothness index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$ defined below. The goal is to estimate σ^2 with $f(\cdot)$ and the distribution of $\boldsymbol{X}$ as nuisance parameters. This problem has been studied in Spokoiny (2002), Munk et al. (2005), Cai et al. (2009), to name a few, again with a focus on the fixed design setting.

Let $I_1, \ldots, I_d$ be d fixed (possibly infinite) intervals on $\mathbb{R}$ and let $\boldsymbol{I}$ be their Cartesian product $I_1 \times \ldots \times I_d \subset \mathbb{R}^d$. Following Barron et al. (1999) and Bhattacharya et al. (2014), we define an anisotropic Hölder class $\Lambda_{\boldsymbol{\alpha},\boldsymbol{I}}(C_{\mathcal{F}})$ on $\boldsymbol{I}$ as follows. For any $\boldsymbol{x} \in \boldsymbol{I}$ and $k \in [d]$, let $f_k(\cdot \mid \boldsymbol{x}_{-k})$ denote the univariate function $y \mapsto f(x_1, \ldots, x_{k-1}, y, x_{k+1}, \ldots, x_d)$, with $\boldsymbol{x}_{-k}$ defined as $\boldsymbol{x}$ without the kth component. Then, $\Lambda_{\boldsymbol{\alpha},\boldsymbol{I}}(C_{\mathcal{F}})$ is defined as all $f : \boldsymbol{I} \mapsto \mathbb{R}$ such that

$$\max_{1\le k\le d} \max_{0\le j\le \lfloor\alpha_k\rfloor} \sup_{\boldsymbol{x}\in\boldsymbol{I}} \Big\| f_k^{(j)}(\cdot \mid \boldsymbol{x}_{-k}) \Big\|_\infty \le C_{\mathcal{F}}$$

and

$$\max_{1\le k\le d} \sup_{x\in\boldsymbol{I}} \sup_{y_1,y_2\in I_k} \frac{|*|f_k^{(\lfloor\alpha_k\rfloor)}(y_1 \mid \boldsymbol{x}_{-k}) - f_k^{(\lfloor\alpha_k\rfloor)}(y_2 \mid \boldsymbol{x}_{-k})}{|y_1 - y_2|^{\alpha'_k}} \le C_{\mathcal{F}},$$

where again $\lfloor\alpha_k\rfloor$ is the largest integer strictly smaller than α_k and $\alpha'_k := \alpha_k - \lfloor\alpha_k\rfloor$. Let $\mathrm{supp}(\boldsymbol{X})$ be the support of $\boldsymbol{X}$.

Define $\mathcal{P}_{\mathrm{mcv},(\boldsymbol{X},\varepsilon)}$ (where "mcv" stands for "multivariate constant variance") as the multivariate counterpart of $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$:

(a) $\boldsymbol{X}$ satisfies $\text{supp}(\boldsymbol{X}) \subset \boldsymbol{I}$.

(b) $\boldsymbol{X}$ has density $p_{\boldsymbol{X}}(\cdot)$ and there exists a fixed positive constant C_0 such that

$$\sup_{\boldsymbol{u}\in\mathbb{R}^d} p_{\boldsymbol{X}}(\boldsymbol{u}) \le C_0.$$

(c) There exist two fixed constants $\delta_0 > 0$ and $c_0 > 0$ such that for any $\boldsymbol{\delta} \in \mathbb{R}^d$ that satisfies $\|\boldsymbol{\delta}\|_\infty < \delta_0$, there exists a set $\mathcal{U} := \mathcal{U}_{\boldsymbol{\delta}} \subset [-1,1]^d$ such that

$$\boldsymbol{\lambda}(\mathcal{U}_{\boldsymbol{\delta}}) \ge c_0 \quad \text{and} \quad \inf_{\boldsymbol{u}\in\mathcal{U}_{\boldsymbol{\delta}}} p_{\widetilde{\boldsymbol{X}}_{ij}}(u_1\delta_1,\dots,u_d\delta_d) \ge c_0,$$

where $\boldsymbol{\lambda}(\cdot)$ represents the Lebesgue measure on $\mathbb{R}^d$.

(d) $\mathbb{E}\varepsilon^4 \le C_\varepsilon$ for some fixed positive constant C_ε.

For an upper bound on the minimax risk, we propose the following multivariate extension of (2.8) via a product kernel (again with convention $0/0 = 0$):

$$\widehat{\sigma}_d^2 := \frac{\binom{n}{2}^{-1}\sum_{i<j}\Big(\prod_{k=1}^d K_{h_k}(X_{i,k}-X_{j,k})\Big)(Y_i-Y_j)^2/2}{\binom{n}{2}^{-1}\sum_{i<j}\Big(\prod_{k=1}^d K_{h_k}(X_{i,k}-X_{j,k})\Big)}, \tag{2.19}$$

where $K(\cdot)$ is a kernel chosen to satisfy (2.9), and $\{h_k\}_{k=1}^d$ is a kernel bandwidth sequence.

In the following results, we will use $\underline{\alpha}$ to denote the harmonic mean of the d-dimensional smoothness index $\boldsymbol{\alpha}$, i.e. $\underline{\alpha} := d/(\sum_{k=1}^d 1/\alpha_k)$. This quantity is known as the *effective smoothness* in classical problems such as anisotropic density estimation (Ibragimov and Khasminski, 1981; Birgé, 1986) and anisotropic function estimation (Nussbaum, 1986; Hoffman and Lepski, 2002).

Proposition 1. *Suppose* $0 < \alpha_k \le 1$, $k \in [d]$. *Suppose the kernel* $K(\cdot)$ *in* $\widehat{\sigma}_d^2$ *is chosen such that* (2.9) *is satisfied with constants* $\overline{M}_K$ *and* $\underline{M}_K$, *and the bandwidth sequence is chosen as* $h_k \asymp n^{-2\underline{\alpha}/(\alpha_k(4\underline{\alpha}+d))}$ *for all* $k \in [d]$. *Then, under* (2.4) *with random design, it holds that*

$$\sup_{f\in\Lambda_{\boldsymbol{\alpha},\boldsymbol{I}}(C_{\mathcal{F}})}\ \sup_{\sigma^2\le C_\sigma}\ \sup_{\mathbb{P}_{(\boldsymbol{X},\varepsilon)}\in\mathcal{P}_{mcv,(\boldsymbol{X},\varepsilon)}} \mathbb{E}\big(\widehat{\sigma}_d^2-\sigma^2\big)^2 \le C\big(n^{-8\underline{\alpha}/(4\underline{\alpha}+d)} \vee n^{-1}\big),$$

where C is some fixed positive constant that only depends on $\overline{M}_K, \underline{M}_K, \boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{mcv,(\boldsymbol{X},\varepsilon)}$.

Proposition 2. *Under (2.4) with random design, it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f \in \Lambda_{\boldsymbol{\alpha},\boldsymbol{I}}(C_{\mathcal{F}})} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{P}_{(\boldsymbol{X},\varepsilon)} \in \mathcal{P}_{mcv,(\boldsymbol{X},\varepsilon)}} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \geq c\big(n^{-8\underline{\alpha}/(4\underline{\alpha}+d)} \vee n^{-1}\big),$$

where c is some fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{mcv,(\boldsymbol{X},\varepsilon)}$, and $\widetilde{\sigma}^2$ ranges over all estimators of σ^2.

We note that Proposition 1 is only proved for $\alpha_k \in (0,1]$, $k \in [d]$. The general case when α_k is possibly larger than 1 is much more involved due to the difficulty in the random design analysis. Propositions 1 and 2, combined, imply that the minimax rate is $n^{-8\underline{\alpha}/(4\underline{\alpha}+d)} \vee n^{-1}$ for $\alpha_k \in (0,1]$, $k \in [d]$. In particular, when f is in an isotropic α-Hölder class ($0 < \alpha \leq 1$), this rate becomes $n^{-8\alpha/(4\alpha+d)} \vee n^{-1}$. We also remark that a different estimator achieving the rate $n^{-8\alpha/(4\alpha+d)} \vee n^{-1}$ over an isotropic α-Hölder class has been briefly sketched in Robins et al. (2008).

For completeness, we also state without proof some results for model (2.4) in the fixed design setting. In particular, we consider the following two types of fixed designs in the d-dimensional unit cube $[0,1]^d$, namely, the grid design (GD):

$$\begin{aligned} (X_{(i_1,\ldots,i_d),1}, \ldots, X_{(i_1,\ldots,i_d),d}) &= (i_1/n^{1/d}, \ldots, i_d/n^{1/d}), \\ (i_1, \ldots, i_d) &\in [n^{1/d}] \times \ldots \times [n^{1/d}] \end{aligned} \tag{2.20}$$

assuming $n^{1/d}$ is an integer, and the diagonal design (DD):

$$(X_{i,1}, \ldots, X_{i,d}) = (i/n, \ldots, i/n), \quad i \in [n]. \tag{2.21}$$

Here for any positive integer n, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. Let $\alpha_{\max} := \max_{k \in [d]} \alpha_k$ and $\alpha_{\min} := \min_{k \in [d]} \alpha_k$. The first result for (GD) is a simple modification of the isotropic result in Cai et al. (2009) by taking differences along the smoothest direction with index $\alpha_{\max}$. The second result can be readily deduced from the fact that $Y_i = \widetilde{f}(i/n) + \sigma\varepsilon_i$, $i \in [n]$, where $\widetilde{f}(x) := f(x, \ldots, x)$ is $\alpha_{\min}$-Hölder smooth.

Proposition 3. *Under* (2.4) *with fixed design (GD), it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f \in \Lambda_{\boldsymbol{\alpha},[0,1]^d}(C_{\mathcal{F}})} \sup_{\sigma^2 \le C_\sigma} \sup_{\mathbb{E}\varepsilon^4 \le C_\varepsilon} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \asymp n^{-4\alpha_{\max}/d} \vee n^{-1}$$

up to some fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma, C_\varepsilon$, *where* $\widetilde{\sigma}^2$ *ranges over all estimators of* σ^2.

Proposition 4. *Under* (2.4) *with fixed design (DD), it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f \in \Lambda_{\boldsymbol{\alpha},[0,1]^d}(C_{\mathcal{F}})} \sup_{\sigma^2 \le C_\sigma} \sup_{\mathbb{E}\varepsilon^4 \le C_\varepsilon} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \asymp n^{-4\alpha_{\min}} \vee n^{-1}$$

up to some fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma, C_\varepsilon$, *where* $\widetilde{\sigma}^2$ *ranges over all estimators of* σ^2.

When $f(\cdot)$ belongs to an isotropic α-Hölder class, Proposition 3 implies the minimax rate $n^{-4\alpha/d} \vee n^{-1}$ derived in Cai et al. (2009). Comparison with the random design rate $n^{-8\alpha/(4\alpha+d)} \vee n^{-1}$ thus shows that, for $0 < \alpha \le 1$, a faster rate is again achievable in the random design setting for $\alpha < d/4$.

2.4.2 Nonparametric additive model

Consider variance estimation in the additive model (2.5):

$$Y_i = \sum_{k=1}^{d} f_k(X_{i,k}) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n,$$

for some fixed integer $d \ge 2$, where $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. with zero mean and unit variance and are independent from $\{\boldsymbol{X}_i\}_{i=1}^n = \{(X_{i,1}, \ldots, X_{i,d})^\top\}_{i=1}^n$ in the random design setting. Unlike Section 2.4.1, we specify $d \ge 2$, since the minimax rate in the fixed design (GD) has completely different behavior for $d = 1$ and $d \ge 2$ (see Proposition 5 below).

Fixed design

We first consider the two fixed designs (GD) and (DD) defined in (2.20) and (2.21). For both designs, we consider an error distribution class with only a finite fourth moment condition. We start with (GD), where by iteratively taking pairwise differences, one is able to estimate the variance at the parametric rate n^{-1} without any smoothness assumption on the additive components $\{f_k\}_{k=1}^d$. For simplicity, we illustrate this idea with $d = 2$ with two additive components $f(\cdot)$ and $g(\cdot)$, and assume that $\sqrt{n}$ is an even number. In this case,

$$Y_{i,j} = f\left(\frac{i}{\sqrt{n}}\right) + g\left(\frac{j}{\sqrt{n}}\right) + \sigma\varepsilon_{i,j}, \quad (i,j) \in [\sqrt{n}] \times [\sqrt{n}],$$

where $\{\varepsilon_{i,j}\}_{i,j\in[\sqrt{n}]}$ are i.i.d. with zero mean and unit variance. By taking the pairwise difference in the first dimension, we have

$$Y_{(i_1,i_2),j} := Y_{i_1,j} - Y_{i_2,j} = f\left(\frac{i_1}{\sqrt{n}}\right) - f\left(\frac{i_2}{\sqrt{n}}\right) + \sigma(\varepsilon_{i_1,j} - \varepsilon_{i_2,j})$$

for all $j \in [\sqrt{n}]$ and $(i_1, i_2) \in [\sqrt{n}] \times [\sqrt{n}]$ such that $i_1 \neq i_2$. Taking again the pairwise difference in the second dimension, we have

$$Y_{(i_1,i_2),(j_1,j_2)} := Y_{(i_1,i_2),j_1} - Y_{(i_1,i_2),j_2} = \sigma(\varepsilon_{i_1,j_1} - \varepsilon_{i_2,j_1} - \varepsilon_{i_1,j_2} + \varepsilon_{i_2,j_2})$$

for all $(i_1, i_2, j_1, j_2) \in [\sqrt{n}] \times [\sqrt{n}] \times [\sqrt{n}] \times [\sqrt{n}]$ such that $i_1 \neq i_2$ and $j_1 \neq j_2$. Clearly, we have $\mathbb{E}Y_{(i_1,i_2),(j_1,j_2)} = 0$ and $\mathrm{Var}(Y_{(i_1,i_2),(j_1,j_2)}) = 4\sigma^2$. Let $m := \sqrt{n}/2$ and define $\mathcal{I} := \{(1,2),(3,4),\ldots,(2m-1,2m)\}$ with cardinality m. Then, for the set of data points $\{Y_{(i_1,i_2),(j_1,j_2)}\}_{(i_1,i_2),(j_1,j_2)\in}$ with cardinality $m^2 = n/4$, it can be readily verified that they are i.i.d. with mean 0 and variance $4\sigma^2$. Therefore, with $\overline{Y}$ defined as the sample average of $\{Y_{(i_1,i_2),(j_1,j_2)}\}_{(i_1,i_2),(j_1,j_2)\in\mathcal{I}}$, the sample variance estimator,

$$\widehat{\sigma}^2_{\text{add, GD}} := \frac{1}{n} \sum_{(i_1,i_2),(j_1,j_2)\in\mathcal{I}} \left(Y_{(i_1,i_2),(j_1,j_2)} - \overline{Y}\right)^2,$$

achieves the parametric rate n^{-1}. A similar derivation holds for general d.

Proposition 5. *Suppose $d \geq 2$. Under* (2.5) *with fixed design (GD), it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f_k, k \in [d]} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{E}\varepsilon^4 \leq C_\varepsilon} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \asymp n^{-1}$$

up to some fixed positive constant that only depends on C_σ and C_ε, where $\widetilde{\sigma}^2$ ranges over all estimators of σ^2, and the first supremum is taken over all functions defined on $[0,1]$ for each $k \in [d]$.

Now we move on to the design (DD), where we assume each additive component f_k in (2.5) is α_k-Hölder smooth on $[0,1]$ with some fixed constant $C_{\mathcal{F}}$. In this case, the model can equivalently be written as

$$Y_i = \widetilde{f}(i/n) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\widetilde{f} := \sum_{k=1}^d f_k$ is $\alpha_{\min}$-Hölder smooth. Therefore, the univariate estimator and lower bound in Wang et al. (2008) can be directly applied.

Proposition 6. *Under* (2.5) *with fixed design (DD), it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f_k \in \Lambda_{\alpha_k,[0,1]}(C_{\mathcal{F}}), k \in [d]} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{E}\varepsilon^4 \leq C_\varepsilon} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \asymp n^{-4\alpha_{\min}} \vee n^{-1}$$

up to some fixed positive constant that only depends on $C_{\mathcal{F}}, C_\sigma, C_\varepsilon$, where $\widetilde{\sigma}^2$ ranges over all estimators of σ^2.

Comparison of Propositions 6 and 4 shows that, in contrast to grid design (GD) and random design below, there is no gain from an additive structure in the mean function for the diagonal design (DD).

Random design

We now discuss (2.5) with a random design for $\{\boldsymbol{X}_i\}_{i=1}^n$ when f_k is α_k-Hölder smooth on some fixed set I_k for each $k \in [d]$. Since a shift in the mean does not affect the estimation of variance, we assume $\mathbb{E} f_k(X_{1,k}) = 0$ for each $k \in [d]$ for simplicity. Recall the definition of $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$ in the beginning of Section 2.2. Define the joint distribution class $\mathcal{P}_{\mathrm{add},(\boldsymbol{X},\varepsilon)}$ (where "add" stands for "additive") as:

> For each $k \in [d]$, the joint distribution of (X_k, ε) belongs to $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$ and the components of $\boldsymbol{X}$ are mutually independent.

In view of Theorem 2, the following lower bound is immediate.

Proposition 7. *Under* (2.5) *with random design, it holds that*

$$\inf_{\widetilde{\sigma}^2} \sup_{f_k \in \Lambda_{\alpha_k, I_k}(C_{\mathcal{F}}), k \in [d]} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{P}_{(\boldsymbol{X},\varepsilon)} \in \mathcal{P}_{add,(\boldsymbol{X},\varepsilon)}} \mathbb{E}\big(\widetilde{\sigma}^2 - \sigma^2\big)^2 \geq c\Big(n^{-\frac{8\alpha_{\min}}{4\alpha_{\min}+1}} \vee n^{-1}\Big),$$

where c is a fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{add,(\boldsymbol{X},\varepsilon)}$, and $\widetilde{\sigma}^2$ ranges over all estimators of σ^2.

We now describe a procedure that matches the lower bound in Proposition 7, but depends crucially on mutual independence. For illustrative purposes, we again consider the case of only two additive components $f(\cdot)$ and $g(\cdot)$, which are α- and β-Hölder smooth, respectively. Let X and W denote the two covariates. For each $i \in [n]$, define

$$\varepsilon_i^X := f(X_i) + \sigma\varepsilon_i \quad \text{ and } \quad \varepsilon_i^W := g(W_i) + \sigma\varepsilon_i,$$

and their corresponding variances

$$\sigma_X^2 := \mathbb{E} f^2(X) + \sigma^2 \quad \text{ and } \quad \sigma_W^2 := \mathbb{E} g^2(W) + \sigma^2.$$

Clearly, we have $\mathbb{E}\varepsilon_i^X = 0$ and $\mathbb{E}\varepsilon_i^W = 0$, and ε_i^X and ε_i^W are independent of $g(W_i)$ and $f(X_i)$, respectively. Now, notice that the additive model in (2.5) can be equivalently viewed as $Y_i = f(X_i) + \varepsilon_i^W$. Thus by applying the univariate kernel estimator defined in (2.8) to $\{(Y_i, X_i)\}_{i=1}^n$, which we denote as $\widehat{\sigma}_W^2$, one obtains

$$\mathbb{E}\big(\widehat{\sigma}_W^2 - \sigma_W^2\big)^2 \leq C(n^{-8\alpha/(4\alpha+1)} \vee n^{-1})$$

for some fixed positive constant C. Similarly, defining $\widehat{\sigma}_X^2$ as $\widehat{\sigma}_W^2$, one has

$$\mathbb{E}\big(\widehat{\sigma}_X^2 - \sigma_X^2\big)^2 \le C(n^{-8\beta/(4\beta+1)} \vee n^{-1}).$$

Lastly, under a finite fourth moment assumption on ε, a sample variance estimator of $\{Y_i\}_{i=1}^n$, denoted as $\widehat{\sigma}_Y^2$, achieves the parametric rate n^{-1} in estimating the total variance $\mathrm{Var}(Y)$, which can be decomposed as $\mathbb{E}f^2(X) + \mathbb{E}g^2(W) + \sigma^2$. Consequently, we have shown that the method-of-moments estimator

$$\widehat{\sigma}^2_{\text{moment},2} := \widehat{\sigma}_X^2 + \widehat{\sigma}_W^2 - \widehat{\sigma}_Y^2 \tag{2.22}$$

achieves the optimal rate in Proposition 7. We summarize the above derivation for the natural extension $\widehat{\sigma}^2_{\text{moment},d}$ to general d.

Proposition 8. *Under* (2.5) *with random design, it holds that*

$$\sup_{f_k \in \Lambda_{\alpha_k, I_k}(C_{\mathcal{F}}), k\in[d]} \ \sup_{\sigma^2 \le C_\sigma} \ \sup_{\mathbb{P}_{(\boldsymbol{X},\varepsilon)} \in \mathcal{P}_{add,(\boldsymbol{X},\varepsilon)}} \mathbb{E}\big(\widehat{\sigma}^2_{moment,d} - \sigma^2\big)^2 \le C\Big(n^{-\frac{8\alpha_{\min}}{4\alpha_{\min}+1}} \vee n^{-1}\Big),$$

where C is some fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{add,(\boldsymbol{X},\varepsilon)}$.

Propositions 7 and 8 together imply the minimax rate over $\mathcal{P}_{\text{add},(\boldsymbol{X},\varepsilon)}$, which further illustrates the fact that an additive structure in the mean function could possibly avoid the "curse of dimensionality" in variance estimation. However, we note that our results crucially rely on the mutual independence condition. It is still largely unclear if the same minimax rate could apply to the general case without this condition, though a discussion of an interesting connection to variance estimation under linear models shall be made in Section 2.4.4.

2.4.3 Connection to quadratic functional estimation

We now formally state the connection between quadratic functional estimation and variance estimation in (2.2), the first of which has been studied in, for example, Doksum and Samarov

(1995), Ruppert et al. (1995), Huang and Fan (1999), and Robins et al. (2009).

Recall the definition of Q in (2.3) with some non-negative weight function $w(\cdot)$. Squaring both sides of (2.2), multiplying by $w(X_i)$, and then taking the expectation, one has

$$\mathbb{E}\big(Y_i^2 w(X_i)\big) = \mathbb{E}\big(f^2(X_i)w(X_i)\big) + \sigma^2\mathbb{E}(w(X_i)\varepsilon_i^2) = Q + \sigma^2\mathbb{E}w(X_i).$$

Under a finite fourth moment assumption on ε, both $\mathbb{E}(Y_i^2 w(X_i))$ and $\mathbb{E}w(X_i)$ can be estimated at the parametric rate via the sample mean estimator, and σ^2 can be estimated via $\widehat{\sigma}^2$ in (2.8) with rate $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ under the quadratic risk. Therefore, the estimator

$$\widehat{Q} := \frac{1}{n}\sum_{i=1}^n Y_i^2 w(X_i) - \left(\frac{1}{n}\sum_{i=1}^n w(X_i)\right)\cdot\widehat{\sigma}^2$$

achieves the same rate $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$. In fact, it is not possible to improve upon this rate since if there exists an estimator $\widetilde{Q}$ with a faster convergence rate, then the "conjugate" estimator of σ^2 defined as

$$\widetilde{\sigma}^2 := \max\left\{\frac{\frac{1}{n}\sum_{i=1}^n Y_i^2 w(X_i) - \widetilde{Q}}{\frac{1}{n}\sum_{i=1}^n w(X_i)}, 0\right\}\cdot \mathbb{1}\left\{\frac{1}{n}\sum_{i=1}^n w(X_i) > 0\right\}$$

will also converge to σ^2 at a faster rate, violating the lower bound in Theorem 2.

The following result summarizes the derivation. Recall the definition of $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$ in the beginning of Section 2.2.

Proposition 9. *Suppose the weight function $w(\cdot)$ in the definition of Q is uniformly bounded on $\mathbb{R}$. Then, it holds that*

$$\inf_{\widetilde{Q}} \sup_{f\in\Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{\sigma^2\le C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{cv,(X,\varepsilon)}} \mathbb{E}\Big(\widetilde{Q} - Q\Big)^2 \asymp n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$$

up to some fixed positive constant that only depends on $w(\cdot)$, $\alpha, C_{\mathcal{F}}, C_\sigma$ and C_0, c_0, C_ε in $\mathcal{P}_{cv,(X,\varepsilon)}$, where $\widetilde{Q}$ ranges over all estimators of Q.

2.4.4 Connection to the linear model

Throughout this paper, we have treated the distribution of $\boldsymbol{X}$ as a nuisance parameter. Interestingly, when we do know the distribution of $\boldsymbol{X}$, variance estimation in nonparametric regression with random design becomes substantially easier with the aid of parallel work in the high-dimensional linear model (Verzelen and Villers, 2010; Dicker, 2014; Kong and Valiant, 2018; Verzelen and Gassiat, 2018). We first elaborate on this point using the simple model (2.2), and then formulate corresponding results for (2.4) and (2.5).

By applying the inverse of the distribution function F of X, (2.2) can be equivalently written as

$$Y_i = \overline{f}(U_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\{U_i\}_{i=1}^n = \{F(X_i)\}_{i=1}^n$ are i.i.d. uniform on $[0,1]$, and $\overline{f}(\cdot) := f \circ F^{-1}(\cdot)$ is still α-Hölder smooth under Lipschitz continuity on F^{-1}. Then, using a wavelet expansion for Hölder classes (cf. Proposition 2.5 in Meyer (1990)), one has

$$Y_i = \overline{f}_1(U_i) + \sum_{j=1}^{2^J} \psi_j(U_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.23}$$

where $\{\psi_j\}_{j=1}^\infty$ is an L_2-orthonormal wavelet basis under the Lebesgue measure on $[0,1]$, and $\overline{f}_1(\cdot)$ is the remainder term after truncation at resolution $J = J_n$ which satisfies $\|\overline{f}_1\|_\infty = O(2^{-\alpha J_n})$. Let $\boldsymbol{\psi} := (\psi_1, \ldots, \psi_{2^J})$ and assume without loss of generality that $\mathbb{E}\boldsymbol{\psi} = \mathbf{0}_{2^J}$, since a mean shift does not affect the estimation of variance. Moreover, due to the orthonormality of $\{\psi_j\}_{j=1}^\infty$, we have $\mathrm{Cov}(\boldsymbol{\psi}) = \mathbb{E}(\boldsymbol{\psi}\boldsymbol{\psi}^\top) = \mathbf{I}_{2^J}$. Following Verzelen and Gassiat (2018) and Kong and Valiant (2018), the estimator

$$\widehat{\sigma}^2_{\text{proj}} := \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2 - \binom{n}{2}^{-1} \sum_{i<j} Y_i Y_j \boldsymbol{\psi}^\top(U_i)\boldsymbol{\psi}(U_j)$$

has a variance term of the order $(2^{J_n} + n)/n^2$ and a bias term of the order $2^{-2\alpha J_n}$. Therefore, by choosing the optimal truncation level $2^{J_n} \asymp n^{2/(4\alpha+1)}$, $\widehat{\sigma}^2_{\text{proj}}$ recovers the optimal rate $n^{-8\alpha/(4\alpha+1)} \vee n^{-1}$ in Theorem 1.

Define $\widehat{\sigma}^2_{\mathrm{proj},d}$ (with tensor wavelet basis) and $\widehat{\sigma}^2_{\mathrm{proj,add}}$ as the natural extensions of $\widehat{\sigma}^2_{\mathrm{proj}}$ under (2.4) and (2.5), respectively (see the proofs of Propositions 10 and 11 in the supplement for exact definitions). In the wavelet expansion, we will use J_k to denote the truncation level for the kth component of $f(\cdot)$ in (2.4) and f_k in (2.5), and we use F_k to denote the marginal distribution of $X_{1,k}$. Recall that $\underline{\alpha} = d/(\sum_{k=1}^d 1/\alpha_k)$ for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$.

Proposition 10 (Multivariate nonparametric regression, design known)**.** *Suppose the distribution of $\boldsymbol{X}$ is known with $\mathrm{supp}(\boldsymbol{X}) \subset \boldsymbol{I}$ for some fixed set $\boldsymbol{I} \subset \mathbb{R}^d$, and $F_k^{-1}(\cdot)$ is Lipschitz continuous for all $k \in [d]$ with some fixed positive constant. Then, when 2^{J_k} is chosen to be of the order $n^{2\underline{\alpha}/(\alpha_k(4\underline{\alpha}+d))}$ for $k \in [d]$ in $\widehat{\sigma}^2_{proj,d}$, it holds that*

$$\sup_{f \in \Lambda_{\boldsymbol{\alpha},\boldsymbol{I}}(C_{\mathcal{F}})} \sup_{\sigma^2 \le C_\sigma} \sup_{\mathbb{E}\varepsilon^4 \le C_\varepsilon} \mathbb{E}\big(\widehat{\sigma}^2_{proj,d} - \sigma^2\big)^2 \le C\big(n^{-8\underline{\alpha}/(4\underline{\alpha}+d)} \vee n^{-1}\big),$$

where C is some fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma, C_\varepsilon$, and the distribution of $\boldsymbol{X}$.

Proposition 11 (Nonparametric additive model, design known)**.** *Suppose the distribution of $\boldsymbol{X}$ is known with $\mathrm{supp}(\boldsymbol{X}) \subset I_1 \times \ldots \times I_d$ for some fixed intervals $I_1, \ldots, I_d$ on the real line, and $F_k^{-1}(\cdot)$ is Lipschitz continuous for all $k \in [d]$ with some fixed positive constant. Then, when 2^{J_k} is chosen to be of the order $n^{2\alpha_k/(4\alpha_k+1)}$ for $k \in [d]$ in $\widehat{\sigma}^2_{proj,add}$, it holds that*

$$\sup_{f_k \in \Lambda_{\alpha_k,I_k}(C_{\mathcal{F}}), k \in [d]} \sup_{\sigma^2 \le C_\sigma} \sup_{\mathbb{E}\varepsilon^4 \le C_\varepsilon} \mathbb{E}\big(\widehat{\sigma}^2_{proj,add} - \sigma^2\big)^2 \le C\Big(n^{-\frac{8\alpha_{\min}}{4\alpha_{\min}+1}} \vee n^{-1}\Big),$$

where C is some fixed positive constant that only depends on $\boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma, C_\varepsilon$, and the distribution of $\boldsymbol{X}$.

As in the classical setting of mean function estimation via orthogonal series, the difference of the rates in Propositions 10 and 11 is clearly explained by the number of wavelet bases used to approximate f in (2.4) and $\{f_k\}_{k=1}^d$ in (2.5). We also note that, quite interestingly, Proposition 10 gives results beyond the case $0 < \alpha_1, \ldots, \alpha_d \le 1$ considered in Proposition 1, and Proposition 11 does not rely on the mutual independence of the components of $\boldsymbol{X}$.

2.4.5 Adaptive estimation of constant variance

In this subsection, we consider adaptive estimation of the variance σ^2 in model (2.2). This is achieved by a Lepski-type procedure (Lepski, 1991; Lepskiĭ, 1991). Let $\widehat{\sigma}^2(h)$ be the estimator in (2.8) with an explicit dependence on the bandwidth parameter h. For any given sample size n and fixed positive constant δ, define two positive integers m_1 and m_2 such that $2^{-m_1} \leq n^{-1} \leq 2^{-m_1+1}$ and $2^{-m_2-1} \leq n^{-(2-\delta)} \leq 2^{-m_2}$, and define the following dyadic grid

$$\mathcal{H}_\delta := \left\{2^{-j} : m_1 \leq j \leq m_2, j \in \mathbb{Z}\right\}.$$

Then, define the estimator $\widehat{\sigma}^2_{\text{adapt}} := \widehat{\sigma}^2\left(\widehat{h}_\delta\right)$ with

$$\widehat{h}_\delta := \max\left\{h \in \mathcal{H}_\delta : |*|\widehat{\sigma}^2(h) - \widehat{\sigma}^2(h') \leq \tau(\log n)^{1/2} n^{-1}(h')^{-1/2}, \forall h' \in \mathcal{H}_\delta, h' < h\right\}$$

for some sufficiently large positive constant τ. If the set being maximized is empty, we will take $\widehat{h}_\delta = n^{-(2-\delta)}$.

To state the error bound of $\widehat{\sigma}^2_{\text{adapt}}$, we need the following variant $\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}$ of the distribution class $\mathcal{P}_{\text{cv},(X,\varepsilon)}$ considered in Theorem 1, where we replace the finite fourth-moment assumption (d) therein by the stronger sub-Gaussian tail condition:

(d′) There exist some fixed positive constants $C_{1,\varepsilon}$ and $C_{2,\varepsilon}$ such that $\mathbb{E}\exp(t\varepsilon) \leq C_{1,\varepsilon}\exp(C_{2,\varepsilon}t^2)$ for any $t \in \mathbb{R}$.

A similar exponential moment assumption has been made in the context of adaptive estimation under fixed design (cf. Theorems 1 and 2 in Cai and Wang (2008)).

Proposition 12. *For any given sufficiently small fixed $\alpha_* > 0$, fix some $\delta_* \in (0, 8\alpha_*/(4\alpha_* + 1))$. Suppose the kernel $K(\cdot)$ in $\widehat{\sigma}^2_{adapt} = \widehat{\sigma}^2\left(\widehat{h}_{\delta_*}\right)$ is chosen such that* (2.9) *is satisfied with constants $\overline{M}_K$ and $\underline{M}_K$, and τ in $\widehat{h}_{\delta_*}$ is chosen to be sufficiently large (only depending on $\delta_*, C_{1,\varepsilon}, C_{2,\varepsilon}$). Then, under* (2.2) *with random design, it holds uniformly over all $\alpha \geq \alpha_*$ that*

$$\sup_{f \in \Lambda_{\alpha,I}(C_\mathcal{F})} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}^{adapt}_{cv,(X,\varepsilon)}} \mathbb{E}\left(\widehat{\sigma}^2_{adapt} - \sigma^2\right)^2 \leq C\left\{\left(\frac{\log n}{n^2}\right)^{4\alpha/(4\alpha+1)} \vee n^{-1}\right\},$$

where C is some fixed positive constant that only depends on $\delta_*, \overline{M}_K, \underline{M}_K, C_{\mathcal{F}}, C_\sigma,$ and $C_0, c_0, C_{1,\varepsilon}, C_{2,\varepsilon}$ in $\mathcal{P}^{adapt}_{cv,(X,\varepsilon)}$.

The following proposition shows that the extra poly-logarithmic term cannot be removed.

Proposition 13. *Let $\phi_{n,\alpha} := (\log n/n^2)^{2\alpha/(4\alpha+1)} \vee n^{-1/2}$ for any $\alpha > 0$ and positive integer n. Consider any fixed positive α_* and $\alpha_* \leq \alpha_1 < \alpha_2 < \infty$. Then, for any sufficiently large n and sufficiently small fixed positive constant c, any estimator $\widetilde{\sigma}^2$ will satisfy that, if*

$$\sup_{f \in \Lambda_{\alpha_2, I}(C_{\mathcal{F}})} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}^{adapt}_{cv,(X,\varepsilon)}} \mathbb{E}\big((\widetilde{\sigma}^2 - \sigma^2)/\phi_{n,\alpha_2}\big)^2 \leq c,$$

then

$$\sup_{f \in \Lambda_{\alpha_1, I}(C_{\mathcal{F}})} \sup_{\sigma^2 \leq C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)} \in \mathcal{P}^{adapt}_{cv,(X,\varepsilon)}} \mathbb{E}\big((\widetilde{\sigma}^2 - \sigma^2)/\phi_{n,\alpha_1}\big)^2 \geq c.$$

The above two results combined are in line with analogous adaptation results in quadratic functional estimation (Efromovich and Low, 1996; Cai and Low, 2006).

Chapter 3
GENERAL ORDER SPLINE ESTIMATION

3.1 Introduction

3.1.1 Overview

Consider the regression model

$$Y_i = f_0(i/n) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{3.1}$$

where $f_0 : [0,1] \to \mathbb{R}$ is an unknown function and ε_i's are independent normal random variables with mean zero and variance σ^2. Throughout the paper, we reserve the notation θ_0 for the truth in (3.1), i.e., $(\theta_0)_i \equiv f_0(i/n)$. The main goal of this paper is to study the approximation of θ_0 by splines with free knots.

Consider the (generalized) spline space with the following three parameters: d, the degree of the spline; d_0, the level of continuity; k, the maximal number of pieces. More formally, (d, d_0, k)-splines are defined as (exact definition in Section 3.2):

$$\Big\{ f : [0,1] \to \mathbb{R} : f \text{ has at most } k+1 \text{ knots, is a degree } d \text{ polynomial} \tag{3.2}$$
$$\text{between knots, and is } d_0\text{-times differentiable at each inner knot} \Big\}.$$

For any fixed degree d, d_0 takes value in $\{-1, 0, \dots, d-1\}$, with $d_0 = d-1$ being the smoothest case and $d_0 = -1$ allowing for discontinuity between pieces. To avoid degeneracy to global polynomials, we only consider the case $k \geq 2$ in this paper. The corresponding sequence space is defined as

$$\Theta(d, d_0, k) \equiv \Big\{ \theta \in \mathbb{R}^n : \theta_i = f(i/n) \text{ for some } (d, d_0, k)\text{-spline } f \Big\}. \tag{3.3}$$

Compared to splines in more classical settings de Boor (1978); Green and Silverman (1994); Wahba (1990), the above parameter space does not fix the knots a priori and thus provides more flexibility.

Splines of the forms (3.2) and (3.3) have frequently emerged in nonparametric curve estimation problems. For example, the classical smoothing splines Wahba (1990) arise from minimizing the least squares criterion with an ℓ_2 roughness penalty. In the ℓ_1 world, splines are closely related to total variation regularization or denoising studied in, e.g., Rudin et al. (1992); Mammen and van de Geer (1997); Chen et al. (2001); Davies and Kovac (2001); Tibshirani et al. (2005); Steidl et al. (2006); Rinaldo (2009); Harchaoui and Lévy-Leduc (2010); Hütter and Rigollet (2016); Dalalyan et al. (2017). In recent years, these methods with the spline space (3.3) received a revival of interest under the name *trend filtering*; cf. Kim et al. (2009); Tibshirani (2014); Wang et al. (2014); Guntuboyina et al. (2020).

Despite the long history and large volume of works related to the spline spaces (3.2)-(3.3), their fundamental statistical limits have remained largely unexplored. Our first main result in this paper reveals the following intriguing phase transition in the minimax rate of estimation error over $\Theta(d, d_0, k)$:

$$\inf_{\widetilde{\theta}} \sup_{\theta \in \Theta(d,d_0,k)} \mathbb{E}_\theta \|\widetilde{\theta} - \theta\|^2 \asymp_d \begin{cases} \sigma^2 k \log\log(16n/k), & 2 \le k \le k_0, \\ \sigma^2 k \log(en/k), & k \ge k_0 + 1. \end{cases} \tag{3.4}$$

Here, $\|\cdot\|$ denotes the Euclidean norm and $\asymp_d$ denotes equivalence in order up to some positive constant that only depends on d. The transition boundary k_0, which takes the form $\lfloor (d+1)/(d-d_0) \rfloor + 1$ with $\lfloor \cdot \rfloor$ denoting the floor function, governs the maximal number of pieces above which the optimal dependence of the estimation error on the sample size n changes from the faster $\log\log(16n)$ rate to the slower $\log(en)$ rate. Notably, for any fixed degree d, k_0 is an increasing function of the regularity parameter d_0. In the two extreme cases, we have $k_0 = d + 2$ if $d_0 = d - 1$ (smoothest) and $k_0 = 2$ if $d_0 = -1$ (roughest). In other words, *the driving factor behind the phase transition in (3.4) is the regularity due to* *the differentiability structure encoded in* d, d_0.

The minimax rate in (3.4) is achieved by the ℓ_0-constrained spline least squares estimator (LSE) $\widehat{\theta} \equiv \widehat{\theta}(\Theta(d, d_0, k), Y)$, with $Y \equiv (Y_1, \dots, Y_n)^\top$ and

$$\widehat{\theta}(\Theta, Y) \equiv \operatorname*{argmin}_{\theta \in \Theta} \|Y - \theta\|_2^2 \quad \text{for any } \Theta \subset \mathbb{R}^n. \tag{3.5}$$

In fact, a more general oracle inequality allowing for arbitrary model mis-specification can be proved for $\widehat{\theta}$. Due to the non-convexity of $\Theta(d, d_0, k)$, the solution to (3.5) with $\Theta = \Theta(d, d_0, k)$ may not be unique and we choose any $\widehat{\theta}$ that achieves the minimum. Among the three parameters, we take d and d_0 to be fixed in advance and consider k as a tuning parameter to balance the approximation error of θ_0 in (3.1) by $\Theta(d, d_0, k)$ and the complexity of the latter space. The estimator in (3.5) with $\Theta = \Theta(d, d_0, k)$ can therefore be viewed as a class of ℓ_0-splines in their constrained form.

The minimax rate in (3.4) and the rate-optimality of ℓ_0-constrained spline LSE are interesting from at least two very different angles. First, the minimax rate in (3.4) is particularly useful in penalty selection for the adaptive version of the ℓ_0-constrained spline LSE $\widehat{\theta}(\Theta(d, d_0, k), Y)$. Specifically, suppose $\theta_0 \in \Theta(d, d_0, k^*)$ in (3.1) with d and d_0 fixed in advance and an unknown k^* on the number of pieces. Our aim is to find an adaptive version of $\widehat{\theta}$ that does not require the knowledge of k^* but remains minimax optimal in estimation. Using the classical approach in Birgé and Massart (1993); Barron et al. (1999); Birgé and Massart (2001); Massart (2007), this can be done by resorting to the penalized spline LSE $\widehat{\theta}_{\text{adapt}}$, where

$$\widehat{\theta}_{\text{adapt}} \equiv \widehat{\theta}(\Theta(d, d_0, \widehat{k}), Y) \tag{3.6}$$

with some data-driven $\widehat{k}$:

$$\widehat{k} \equiv \operatorname*{argmin}_{1 \le k \le n} \left\{ \|Y - \widehat{\theta}(\Theta(d, d_0, k), Y)\|^2 + \text{pen}(k; d, d_0) \right\} \tag{3.7}$$

for some penalty function $\text{pen}(\cdot; d, d_0)$. The estimator $\widehat{\theta}_{\text{adapt}}$ can thus be viewed as a class of ℓ_0-penalized splines. Similar ℓ_0-penalized procedures have previously been studied in Kohler (1999); Boysen et al. (2009); Fan and Guan (2018); Jewell and Witten (2018). When the penalty $\text{pen}(\cdot; d, d_0)$ is chosen to be proportional to the minimax rate established in (3.4), $\widehat{\theta}_{\text{adapt}}$ is guaranteed to be adaptively minimax optimal over $\Theta(d, d_0, k)$ for all values of k.

Second, (3.4) suggests some interesting comparison between ℓ_0- and ℓ_1-regularizers in spline regression. For expository purpose, let us consider the simplest piecewise constant class $\Theta(0, -1, k)$, where the transition boundary is given by $k_0 = 2$. There, while the ℓ_0-constrained spline LSE, as defined in (3.5) with $\Theta = \Theta(0, -1, 2)$, is able to achieve the faster $\log\log(16n)$ rate with 2 pieces, the same rate has been proven to be un-attainable by the ℓ_1 trend filtering, even with an additional minimum spacing condition that could be substantially improved with ℓ_0-splines van de Geer (2018); Fan and Guan (2018); Guntuboyina et al. (2020). Computationally, unlike the context of sparse linear regression where the ℓ_0 problem of best-subset selection is provably NP-hard Natarajan (1995), efficient dynamic programming algorithms do exist for implementing (3.5), at least in the discontinuous case ($d_0 = -1$) Auger and Lawrence (1989); Winkler and Liebscher (2002); Jackson et al. (2005); Friedrich et al. (2008) and the first-order continuous case ($d_0 = 0$) Fearnhead et al. (2019). Our results hence suggest that *the ℓ_0-constrained spline LSE could be an attractive alternative to its ℓ_1 counterparts in spline regressions.*

To motivate the second main result of this paper, we recall the following minimax result from Gao et al. (2019): for all $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta^*(0,k)} \mathbb{E}_{\theta^*} \|\widetilde{\theta}^* - \theta^*\|^2 \asymp \sigma^2 k \log\log(16n/k), \tag{3.8}$$

where $\Theta^*(0, k)$ is the sub-class of $\Theta(0, -1, k)$ with non-decreasing signals. Comparing (3.4) with $d = 0, d_0 = -1$ and (3.8) above, we see that the phase transition from the faster rate $\log\log(16n)$ to the slower rate $\log(en)$ in (3.4) is eliminated in (3.8) under the additional monotonicity shape constraint. This raises the natural questions of whether a similar gain by shape constraints applies to higher-order splines, and if so, which type of shape constraints should be encouraged. As shape-constrained models repeatedly prove their usefulness in various applications, answering the above questions is of both practical and theoretical interests.

To this end, following Balabdaoui and Wellner (2007); Chatterjee et al. (2015), we consider the following sub-class of (d, d_0, k)-splines with an additional 'd-monotone' shape constraint (exact definition in Section 3.3):

$$\Big\{f : [0,1] \to \mathbb{R} : f \text{ is a } (d, d-1, k)\text{-spline with non-decreasing} \\ \text{highest-order polynomial coefficients}\Big\}. \tag{3.9}$$

Two canonical examples are $d = 0$ and 1, with the former corresponding to non-decreasing signals with at most k constant pieces, and the latter corresponding to convex signals with at most k linear pieces. Both classes have been extensively studied in the literature; cf. Zhang (2002); Chatterjee et al. (2015); Bellec (2018); Gao et al. (2019) for the case $d = 0$ and Guntuboyina and Sen (2015); Chatterjee et al. (2015); Bellec (2018) for the case $d = 1$. Define the sequence space corresponding to (3.9) as $\Theta^*(d, k)$.

As a special case of our second main result, we show an analogue of (3.8) under the convexity (=1-monotone) shape constraint: for all $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta^*(1,k)} \mathbb{E}_{\theta^*} \|\widetilde{\theta}^* - \theta^*\|^2 \asymp \sigma^2 k \log\log(16n/k). \tag{3.10}$$

The same upper bound actually holds for the general d-monotone class $\Theta^*(d, k)$, with a complementary lower bound showing that the $\log\log(16n)$ rate cannot be further improved even with only two pieces. Comparing (3.4) and (3.10), it is hence clear that *a higher-order 'd-monotonicity' shape constraint eliminates the phase transition in (3.4) for general d in that the faster $k \log\log(16n/k)$ rate can now be achieved for all k.* The d-monotonicity therefore offers an attractive non-parametric sub-class $\Theta^*(d, k)$ of the general $\Theta(d, d-1, k)$ over which additional gain can be obtained in estimating the underlying signal.

Finally, we remark on the technical challenges in proving (3.4) and (3.10). Unlike the relatively straightforward proof of the $\log(en)$ part in (3.4), the derivation of the correct transition boundary k_0 and the faster $\log\log(16n)$ rate requires non-trivial efforts from both analytical and probabilistic angles. The analytic step is to derive sharp enough controls for the magnitudes of the polynomial coefficients of signals in $\Theta(d, d_0, k)$ and $\Theta^*(d, k)$, which, in a certain sense, need be 'tied' to either the left-most or the right-most knot of the signal. This is possible either due to the strong regularity inherited in the differentiability structure of $\Theta(d, d_0, k)$ for $k \leq k_0$, or to the global regularity within the d-monotonicity shape constraint. Once the above controls are obtained, a generalized version of the law of iterated logarithm (LIL), which we will develop in Section 3.4, can be applied to obtain the iterated logarithmic rates in (3.4) and (3.10).

The rest of the paper is organized as follows. Sections 3.2 and 3.3 are devoted to the study of unshaped splines $\Theta(d, d_0, k)$ and shaped splines $\Theta^*(d, k)$, respectively. A general version of the LIL in expectation is developed in Section 3.4. Main proofs of the results are presented in Sections B.1 and B.2, with the remaining technical lemmas collected in the Appendix.

3.1.2 Notation

For any $x \in \mathbb{R}$, write $(x)_+ \equiv \max\{x, 0\}$. Let $\mathbf{1}.$ denote the indicator function. For any non-negative integers a, b, we use $[a; b]$ to denote the set $\{a, \ldots, b\}$ and $(a; b]$ to denote the set $\{a + 1, \ldots, b\}$. For any two positive integers a, b, let $\mathrm{Mod}(a; b)$ be the remainder of a divided by b. For any two real numbers a, b, define $a \vee b \equiv \max\{a, b\}$ and $a \wedge b \equiv \min\{a, b\}$. For any positive integers $m \geq n$, let $\underline{\odot}(m; n) \equiv m(m-1)\ldots(m-n+1)$ and $\overline{\odot}(m; n) \equiv m(m+1)\ldots(m+n-1)$. Let $\mathbb{Z}_+$ denote the set of positive integers and $\mathbb{Z}_{\geq 0} \equiv \mathbb{Z}_+ \cup \{0\}$. For any $d \in \mathbb{Z}_+$, let $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ stand for the unit sphere. We write $\mathbb{E}_{\theta_0}$ as expectation under the experiment (3.1) with truth θ_0.

Let $C^m([0,1])$ denote the set of all m-times differentiable functions on $[0,1]$. For any $f \in C^m([0,1])$ and integer $0 \le \ell \le m$, let $f^{(0)}(x) \equiv f(x)$ and $(D^{(\ell)}f)(x) \equiv f^{(\ell)}(x)$ be the ℓ-th derivative of f at point x. For any function f defined on $[0,1]$, $\tau \in [0,1]$, and real number c, define the first-order integral $(I^1_{c;\tau}f)(x) \equiv \int_\tau^x f(y)\,\mathrm{d}y + c$ for $x \in [0,1]$, and the m-th order integral iteratively as $(I^m_{c_0,\ldots,c_{m-1};\tau}f)(x) \equiv \big(I^1_{c_0;\tau}(I^{m-1}_{c_1,\ldots,c_{m-1};\tau}f)\big)(x)$ for any positive integer $m \ge 2$ and real sequence $\{c_\ell\}_{\ell=0}^{m-1}$. For any real function f, let $f(x_-)$ and $f(x_+)$ denote the left and right limits at x, respectively.

For two non-negative sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim_d b_n$ (resp. $a_n \gtrsim_d b_n$) if $a_n \le Cb_n$ (resp. $a_n \ge cb_n$) for some $C, c > 0$ that only depend on d. We also write $a_n \asymp_d b_n$ if both $a_n \lesssim_d b_n$ and $a_n \gtrsim_d b_n$ hold. In the following, we will suppress d in $\lesssim_d$, $\gtrsim_d$, and $\asymp_d$ when no confusion is possible. For any given constants $a_1, a_2, \ldots$, we write $C(a_1, a_2, \ldots)$ and $c(a_1, a_2, \ldots)$ to denote positive constants that only depend on $a_1, a_2, \ldots$.

3.2 General-order spline regression

We start with an exact definition of the general-order spline space in (3.2):

$$
\begin{aligned}
\mathcal{F}_n(d, d_0, k) \equiv \Big\{ & f : [0,1] \to \mathbb{R} : \text{there exist } 0 \equiv n_0 \le \ldots \le n_k \equiv n \text{ such that} \\
& n_0, \ldots, n_k \in \mathbb{Z}_{\ge 0}, \qquad n_i - n_{i-1} \ge (d+1)\mathbf{1}_{n_i > n_{i-1}}, \\
& f \text{ is a } d\text{-degree polynomial on each interval } (n_{i-1}/n, n_i/n], \text{ and} \\
& f^{(\ell)}\big((n_i/n)_-\big) = f^{(\ell)}\big((n_i/n)_+\big) \text{ for all } i \in [1; k-1] \text{ and } \ell \in [0; d_0] \Big\}.
\end{aligned}
$$

For any fixed degree $d \ge 0$, the range of d_0 is $[-1; d-1]$, with $d_0 = -1$ allowing the spline f to be completely discontinuous. The numbers $n_0/n, \ldots, n_k/n$ are the *knots* of f, with the middle $(k-1)$ ones as *inner knots*. Define the corresponding sequence space

$$
\Theta_n(d, d_0, k) \equiv \Big\{ \theta \in \mathbb{R}^n : \theta_i = f(i/n) \text{ for some } f \in \mathcal{F}_n(d, d_0, k) \Big\}; \tag{3.11}
$$

in what follows, we suppress the subscript n of $\Theta_n(d, d_0, k)$ when no confusion is possible and name $n_0, \ldots, n_k$ in its corresponding spline $f \in \mathcal{F}_n(d, d_0, k)$ the *knots* of θ.

Two remarks regarding the above spline class are in line.

(i) The function space $\mathcal{F}_n(d, d_0, k)$ enforces the inner knots of the spline to be positioned among the design points. This is due to two reasons. First, it ensures the existence of the LSE as defined in (3.5) with $\Theta = \Theta(d, d_0, k)$. Indeed, the minimization can be first taken over at most $(n+1)^{k-1}$ configurations of the inner knots, after which the problem becomes strictly convex with respect to the rest of the polynomial coefficients and thus has a unique solution. Second, it facilitates fast computation of the LSE via dynamic programming algorithms; see Fearnhead et al. (2019) for detailed illustration of the piecewise linear case.

(ii) The gap $d+1$ between n_i and n_{i-1} in the above definition is necessary for the identifiability of f in the discontinuous case. This minimum spacing condition improves substantially over existing ones made in a class of ℓ_1 methods; see Remark 11 ahead for more details.

For any fixed $d \in \mathbb{Z}_{\geq 0}$ and $d_0 \in [-1; d-1]$, let

$$k_0 \equiv k_0(d, d_0) \equiv \left\lfloor \frac{d+1}{d-d_0} \right\rfloor + 1. \tag{3.12}$$

Our first main result is the following oracle inequality. Recall that we only consider the case $k \geq 2$ in this paper and the analysis of global polynomials (corresponding to $k=1$) is rather straightforward.

Theorem 1. *Fix any $\theta_0 \in \mathbb{R}^n$. Let $\widehat{\theta} \equiv \widehat{\theta}(\Theta(d, d_0, k), Y)$ be the LSE as defined in* (3.5) *under the experiment* (3.1) *with truth θ_0. Then, for any $\delta > 0$, there exists some $C = C(d, \delta)$ such that the following statements hold for any $n \geq \underline{n}$ with some $\underline{n} = \underline{n}(d)$. If $2 \leq k \leq k_0$,*

$$\mathbb{E}_{\theta_0} \|\widehat{\theta} - \theta_0\|^2 \leq (1+\delta) \inf_{\theta \in \Theta(d, d_0, k)} \|\theta - \theta_0\|^2 + C\sigma^2 k \log\log(16n/k),$$

and if $k \geq k_0 + 1$,

$$\mathbb{E}_{\theta_0} \|\widehat{\theta} - \theta_0\|^2 \leq (1+\delta) \inf_{\theta \in \Theta(d,d_0,k)} \|\theta - \theta_0\|^2 + C\sigma^2 k \log(en/k).$$

The following lower bound result shows that Theorem 1 is optimal in the minimax sense.

Proposition 14. *Under the experiment* (3.1)*, there exists some* $c = c(d)$ *such that the following statements hold for all* $n \geq \underline{n}$ *with some* $\underline{n} = \underline{n}(d)$*. If* $2 \leq k \leq k_0$,

$$\inf_{\widetilde{\theta}} \sup_{\theta \in \Theta(d,d_0,k)} \mathbb{E}_\theta \|\widetilde{\theta} - \theta\|^2 \geq c\sigma^2 k \log\log(16n/k),$$

and if $k \geq k_0 + 1$,

$$\inf_{\widetilde{\theta}} \sup_{\theta \in \Theta(d,d_0,k)} \mathbb{E}_\theta \|\widetilde{\theta} - \theta\|^2 \geq c\sigma^2 k \log(en/k),$$

where the infimum over $\widetilde{\theta}$ *in both displays is taken over all measurable functions of* Y*.*

The proof of Theorem 1 is presented in Section B.1, and the proof of Proposition 14 can be found in Appendix B.3.1.

Remark 9. *The above two results imply, in particular, the minimax rates in* (3.4)*. There, the upper bound* $k\log(en/k)$ *above the transition boundary* k_0 *is not essentially new and can be proved via straightforward modifications of the classical arguments in, e.g., Donoho and Johnstone (1994); Birgé and Massart (2001). Rather, our main contribution lies in establishing the sharp transition boundary* k_0 *and the faster* $\log\log(16n)$ *rate below this boundary.*

In practice when the number of pieces k *is unknown, the minimax rates in* (3.4) *provide guidance for penalty selection in the adaptive version* (3.6) *of the* ℓ_0*-constrained spline LSE. Precisely, one can choose* $\widehat{k}$ *as in* (3.7) *with the penalty*

$$pen(k; d, d_0) \equiv \tau\sigma^2 \left[\mathbf{1}_{k=1} + k \log\log(16n/k) \cdot \mathbf{1}_{2 \leq k \leq k_0} + k\log(en/k) \cdot \mathbf{1}_{k > k_0} \right]$$

for some sufficiently large universal $\tau > 0$*. Then, standard arguments Birgé and Massart (1993); Barron et al. (1999); Birgé and Massart (2001); Massart (2007) guarantee that* $\widehat{\theta}_{adapt}$ *is adaptively minimax optimal over* $\Theta(d, d_0, k)$ *for all* $k \in \mathbb{Z}_+$*. Details are accordingly skipped.*

Remark 10. *It is important to mention here one crucial difference between our perspective for the phase transition results and the* $\log\log(16n)$ *rates and the one taken in Gao et al. (2019). There, the faster* $\log\log(16n)$ *rate for* $\Theta(0,-1,2)$ *follows immediately from the general iterated logarithmic rates for* $\Theta^*(0,k)$*, the class of piecewise constant and non-decreasing signals with at most* k *pieces (formally defined in Section 3.3). In other words, the* $\log\log(16n)$ *rate for* $\Theta(0,-1,2)$ *is perceived in Gao et al. (2019) as a consequence of the monotonicity shape constraint. In contrast, the* $\log\log(16n)$ *rate for* $\Theta(d,d_0,k)$ *in (3.4) in the regime* $k \le k_0$ *is inherited from the strong regularity in the signal parametrized by the degree* d *and the level of continuity* d_0*, rather than any explicit shape constraint. In the regime* $k > k_0$*, the* $\log\log(16n)$ *rate is not possible due to insufficient regularity in* $\Theta(d,d_0,k)$*, unless additional shape constraints are enforced; see Section 3.3 ahead for more details.*

Remark 11. *Recently, Guntuboyina et al. (2020) studied the theoretical properties of trend filtering (TF), a class of* ℓ_1*-regularized discrete spline methods. More precisely, under the experiment* (3.1)*, the* d*-th order TF estimator is*

$$\widehat{\theta}^d_{TF} \equiv \min_{\theta\in\mathbb{R}^n} \Big\{ \|Y-\theta\|^2 + \lambda \|D^{(d)}\theta\|_1 \Big\}, \tag{3.13}$$

where $\|\cdot\|_1$ *denotes the vector* ℓ_1 *norm,* $\lambda > 0$ *is a tuning parameter, and* $D^{(d)} : \mathbb{R}^n \to \mathbb{R}^{n-d}$*, when applied to vectors, represents the* d*-th order discrete difference operator defined as* $D^{(0)}\theta \equiv \theta$*,* $D^{(1)}\theta \equiv (\theta_2-\theta_1,\ldots,\theta_n-\theta_{n-1})^\top$*, and* $D^{(r)}\theta \equiv D^{(1)}(D^{(r-1)}\theta)$ *for* $r \ge 2$*. Equation* (3.13) *is a convex problem and can be solved efficiently via algorithms designed for lasso-type problems Tibshirani (2014).*

For any $\theta_0 \in \Theta(d,d-1,k)$ *in* (3.1)*, Corollary 2.11 in Guntuboyina et al. (2020) proved that, upon choosing the tuning parameter* λ *properly and assuming a minimum spacing condition to be detailed below,*

$$\mathbb{E}_{\theta_0}\|\widehat{\theta}^{d+1}_{TF} - \theta_0\|^2 \le C\sigma^2\Big(k\log(en/k) + k^{2(d+1)}\mathbf{1}_{d>0}\Big). \tag{3.14}$$

for some $C = C(d)$. Comparing (3.14) *with our Theorem 1 and Proposition 14, we see the following distinctions between ℓ_0-regularized splines and their ℓ_1 counterparts.*

(i) The bound (3.14) *requires a minimum spacing condition that regulates, for non-vanishing pieces $(n_i; n_{i+1}]$ between knots with different signs (see Page 210 of Guntuboyina et al. (2020) for their definition for the signs of knots), $n_{i+1} - n_i \geq cn/k$ for some $c = c(d)$. This is stronger than the constant gap condition assumed in $\Theta(d, d_0, k)$. Moreover, Theorem 4.2 in Fan and Guan (2018) suggests that this minimum spacing condition is essential to the TF estimators, namely, the performance of* (3.13) *could deteriorate to $\sqrt{n}$ (up to some polylogarithmic factors) without it.*

(ii) Over the class $\Theta(d, d-1, k)$ with transition boundary $k_0 = d+2$, the ℓ_1 TF estimator in (3.13) *is in general rate sub-optimal below the boundary, even with the additional minimum spacing condition mentioned above. Specifically, in the constant space $\Theta(0, -1, k)$, the minimax rate of estimation is $\log\log(16n)$ with $k = 2$ pieces, but the TF estimator in* (3.13) *with $d = 1$ can only achieve the slower $\log(en)$ rate in view of Lemma 2.4 of Guntuboyina et al. (2020).*

Remark 12. *For the computation of the ℓ_0-constrained spline LSE $\widehat{\theta}$ and its adaptive version* (3.6)*, the major difficulty in the development of efficient algorithms is measured by the regularity parameter d_0. For $d_0 = -1$, both estimators can be computed efficiently using standard dynamic programming algorithms Auger and Lawrence (1989); Winkler and Liebscher (2002); Jackson et al. (2005); Friedrich et al. (2008) along with more refined pruning arguments Killick et al. (2012); Maidstone et al. (2017). For the first-order continuous case $(d_0 = 0)$, Fearnhead et al. (2019) recently introduced for the linear case $(d = 1)$ a novel dynamic programming algorithm with linear to quadratic time complexity, which can be readily extended to arbitrary order $d \in \mathbb{Z}_+$. We expect that the above method could potentially be extended to the case of general d and d_0, but this will be left as the subject of future research.*

Lastly, we provide some intuition for the form of k_0 defined in (3.12). This will mostly be clear from the perspective of minimax lower bounds in Proposition 14. There, the situation is somewhat similar to the derivation of minimax lower bounds in the sparse linear regression setting Donoho and Johnstone (1994); Ye and Zhang (2010); Raskutti et al. (2011), in that we only have to find, for each fixed d and d_0, the minimum value of k such that a subset S of 1-sparse vectors can be constructed in $\Theta(d, d_0, k)$ with cardinality $|S| \geq cn$ for some $c = c(d)$. Heuristically, this value can be found via the following degree-of-freedom (DOF) calculation:

$$(k-2)(d+1) \geq (k-1)(d_0+1) + 1. \tag{3.15}$$

Here, the left-hand side is the DOF for any 1-sparse $\theta \in \Theta(d, d_0, k)$ with the two end pieces being constantly zero, as each of the middle $(k-2)$ pieces has $(d+1)$ DOF arising from the d-degree polynomial. On the right-hand side, the first term $(k-1)(d_0+1)$ results from the (d_0+1) continuity constraints at each of the $(k-1)$ inner knots, and the additional 1 DOF excludes the possibility that $\theta \equiv 0$. Solving (3.15) yields that $k \geq 1 + \lceil (d+2)/(d-d_0) \rceil$, which indeed holds for $k = k_0 + 1$ as defined in (3.12), with equality when $d_0 = d - 1$.

Figure 3.1 demonstrates the minimum number k of pieces needed for $d \in [0; 2]$ and $d_0 \in [-1; d-1]$ so that a 1-sparse vector can be constructed in general position. The minimum value of k in each scenario matches $k_0 + 1$ as defined in (3.12).

3.3 General-order splines with shape constraint

As mentioned in (3.8) in the introduction, in contrast to the phase transition in (3.4), the faster $\log\log(16n)$ rate of estimation becomes universal in the class $\Theta^*(0, k)$ that contains all piecewise constant non-decreasing signals. This section derives higher-order analogues of this result. We start with the convexity constraint in the linear case in Section 3.3.1, and then generalize to higher-order splines in Section 3.3.2.

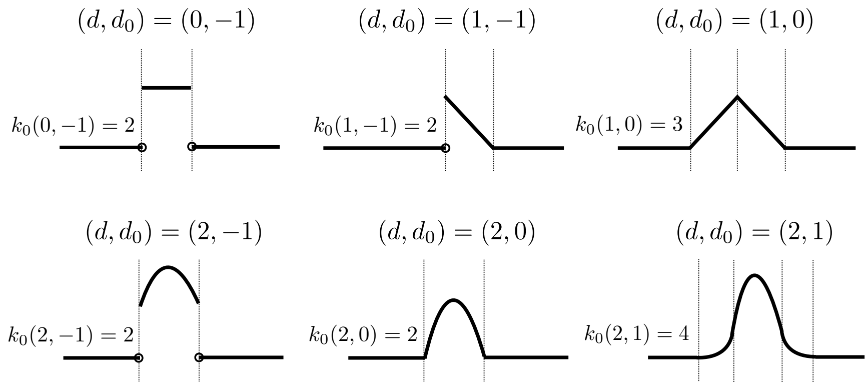

Figure 3.1: Minimum number of $k = k_0 + 1$ pieces required to construct 1-sparse vectors with general position in $\Theta(d, d_0, k)$ for $d \in [0; 2]$ and $d_0 \in [-1; d-1]$.

3.3.1 Convex piecewise linear regression

Convex regression is one of the central topics in shape constrained regression; see, e.g., Guntuboyina and Sen (2015); Chatterjee et al. (2015); Bellec (2018) for global risk bounds and adaptation properties of the convex LSE.

We start by defining the function space of convex piecewise linear functions:

$$\mathcal{F}_n^*(1, k) \equiv \Big\{ f \in \mathcal{F}_n(1, 0, k) : f \text{ has non-decreasing slopes on } [0, 1] \Big\}, \tag{3.16}$$

and the space on the sequence level:

$$\Theta_n^*(1, k) \equiv \{\theta^* \in \mathbb{R}^n : \theta_i^* = f^*(i/n) \text{ for some } f^* \in \mathcal{F}_n^*(1, k)\}, \tag{3.17}$$

with the subscript n in $\Theta_n^*(1, k)$ suppressed in the sequel. The following two results show that the convexity shape constraint eliminates the phase transition in $\Theta(1, 0, k)$.

Proposition 15. *Fix any $\theta_0 \in \mathbb{R}^n$. Let $\widehat{\theta}^* \equiv \widehat{\theta}(\Theta^*(1, k), Y)$ be the LSE as defined in* (3.5) *under the experiment* (3.1) *with truth θ_0. Then, for any $\delta > 0$, there exists some $C = C(\delta)$*

such that for any $n \geq \underline{n}$ with some universal $\underline{n}$ and $k \geq 2$,

$$\mathbb{E}_{\theta_0}\|\widehat{\theta}^* - \theta_0\|^2 \leq (1+\delta) \inf_{\theta^* \in \Theta^*(1,k)} \|\theta^* - \theta_0\|^2 + C\sigma^2 k \log\log(16n/k). \tag{3.18}$$

Proposition 16. *Under the experiment* (3.1)*, there exists some universal constant c such that for all $n \geq \underline{n}$ with some universal $\underline{n}$ and $k \geq 2$,*

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta^*(1,k)} \mathbb{E}_{\theta^*}\|\widetilde{\theta}^* - \theta^*\|^2 \geq c\sigma^2 k \log\log(16n/k),$$

where the infimum over $\widetilde{\theta}^*$ is taken over all measurable functions of Y.

Proposition 15 follows from its more general version in Theorem 2 ahead. The proof of Proposition 16 will be presented in Appedix B.3.2.

Remark 13. *The in-expectation version of Theorem 4.3 in Bellec (2018) proved a similar oracle inequality for the convex LSE:*

$$\mathbb{E}_{\theta_0}\|\widehat{\theta}(\Theta^*, Y) - \theta_0\|^2 \leq \inf_{\theta^* \in \Theta^*} \Big(\|\theta^* - \theta_0\|^2 + Ck(\theta^*)\log(en/k(\theta^*))\Big) \tag{3.19}$$

for some universal constant $C > 0$, where $\Theta^* \equiv \Theta^*(1,n)$ is the larger class of equispaced realizations of general convex functions on $[0,1]$, and $k(\theta^*)$ is the number of linear pieces of θ^*, i.e., $k(\theta^*) \equiv \sum_{i=2}^n \mathbf{1}_{2\theta_i^* < \theta_{i-1}^* + \theta_{i+1}^*}$. Note that Θ^*, as opposed to $\Theta^*(1,k)$, is a closed convex cone in $\mathbb{R}^n$. The bounds (3.18) *and* (3.19) *are complementary in nature: the bound* (3.19) *exploits the convexity of Θ^* to obtain a sharp oracle inequality (in the sense of leading constant 1 before $\inf_{\theta^* \in \Theta^*}\|\theta^* - \theta_0\|^2$), but only achieves a slower worst-case $k\log(en/k)$ rate over the smaller class $\Theta^*(1,k)$; the bound* (3.18)*, or its adaptive version modified in a similar way as* (3.6)*, is minimax optimal over $\Theta^*(1,k)$ but loses the sharp leading constant 1.*

3.3.2 General-order spline regression with shape constraint

Following Balabdaoui and Wellner (2007); Chatterjee et al. (2015), we consider the class of d-monotone splines defined as follows. Let

$$\mathcal{F}_n^*(0,k) \equiv \Big\{f \in \mathcal{F}_n(0,-1,k) : f \text{ is non-decreasing on } [0,1]\Big\}$$

be the 0-monotone class. Next, for any $d \in \mathbb{Z}_+$, define

$$\mathcal{F}_n^*(d,k) \equiv \Big\{f : [0,1] \to \mathbb{R} : \ f(x) = (I^d_{r_0,\ldots,r_{d-1};0} f_\circ)(x) \\ \text{for some } f_\circ \in \mathcal{F}_n^*(0,k) \text{ and real sequence } \{r_\ell\}_{\ell=0}^{d-1}\Big\}.$$

Define the sequence version of the above space as

$$\Theta_n^*(d,k) \equiv \Big\{\theta^* \in \mathbb{R}^n : \theta_i^* = f^*(i/n) \text{ for some } f^* \in \mathcal{F}_n^*(d,k)\Big\}, \tag{3.20}$$

shorthanded as $\Theta^*(d,k)$. One can readily check that for $d = 0$, $\Theta^*(0,k)$ is the class of k-piece isotonic signals studied in Gao et al. (2019); for $d = 1$, $\Theta^*(1,k)$ coincides with the convex piecewise linear class in (3.17). Moreover, two facts follow immediately from the above definitions: (i) For any $d \geq 1$, $f^* \in \mathcal{F}_n^*(d,k) \subset C^{d-1}([0,1])$ so that $\Theta^*(d,k) \subset \Theta(d,d-1,k)$ with the latter defined in (3.11); (ii) For any $d \geq 1$ and $\ell \in [1;d]$, it holds that $(f^*)^{(\ell)} \in \mathcal{F}_n^*(d-\ell,k)$.

The following result, with Proposition 15 as a special case, shows that d-monotonicity eliminates the phase transition in the general spline space $\Theta(d,d-1,k)$. Its proof is given in Section B.2.

Theorem 2. *Fix any $\theta_0 \in \mathbb{R}^n$. Let $\widehat{\theta}^* \equiv \widehat{\theta}(\Theta^*(d,k), Y)$ be the LSE as defined in* (3.5) *under the experiment* (3.1) *with truth θ_0. Then, for any $\delta > 0$, there exists some $C = C(d,\delta)$ such that for any $n \geq \underline{n}$ with some $\underline{n} = \underline{n}(d)$ and $k \geq 2$,*

$$\mathbb{E}_{\theta_0}\|\widehat{\theta}^* - \theta_0\|^2 \leq (1+\delta) \inf_{\theta^* \in \Theta^*(d,k)} \|\theta^* - \theta_0\|^2 + C\sigma^2 k \log\log(16n/k).$$

Moreover, there exists some $c = c(d)$ such that for all $n \geq \underline{n}$ and $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta^*(d,k)} \mathbb{E}_{\theta^*}\|\widetilde{\theta}^* - \theta^*\|^2 \geq c\sigma^2 \log\log(16n),$$

where the infimum over $\widetilde{\theta}^*$ is taken over all measurable functions of Y.

Remark 14. *The essential technical difficulties in proving Theorem 2 and Proposition 16 over the oracle inequality version of (3.8) (cf. Theorem 2.1 of Gao et al. (2019)) rest in the additional regularity of* $\Theta^*(d,k)$ *over* $\Theta^*(0,k)$*.*

(i) For the upper bound, Gao et al. (2019) made essential use of the fact that $\widehat{\theta}(\Theta^*(0,k))$ *is the sample average given the estimated knots; cf. Lemma 5.1 therein. The analogous property is, unfortunately, not true even for* $\widehat{\theta}(\Theta^*(1,k))$*. Instead, we provide a completely different proof which is based on a new parametrization for general-order splines with shape constraint (cf. Lemma 33 ahead). We further observe that this new proof technique, when applied to the setting of Gao et al. (2019), significantly simplifies their proof; see Section B.2.3 for details.*

(ii) For the lower bound in Proposition 16, the continuity constraint in $\Theta^*(1,k)$ *requires a much more delicate construction of least favorable signals that achieves the* $k\log\log(16n/k)$ *rate, compared to* $\Theta^*(0,k)$*; see Appendix B.3.2 for more details. This lower bound construction can actually be extended to yield the optimal* $k\log\log(16n/k)$ *rate over the quadratic class* $\Theta^*(2,k)$*, but a general lower bound of the order* $k\log\log(16n/k)$ *is still lacking for higher-order* d*-monotone splines.*

3.4 A generalized law of iterated logarithm

In this section, we present a generalized law of iterated logarithm (LIL) in expectation that underlies the $\log\log(16n)$ rates derived in Sections 3.2 and 3.3. Recall that a centered random variable X is said to be *sub-Gaussian* with parameter τ, if there exists some $K>0$ such that $\mathbb{E}\exp(\lambda X)\le K\exp(\lambda^2\tau^2/2)$ for any $\lambda\in\mathbb{R}$.

Theorem 3. *Fix positive integers* $d\ge 1$ *and* $n\ge 2$*. Let* $\{\varepsilon_i\}_{i=1}^n$ *be a sequence of independent and identically distributed centered sub-Gaussian random variables with parameter* 1*. Let*

$\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly increasing continuous function with inverse ψ^{-1}. Let

$$Z \equiv \max_{1\leq n_1<n_2\leq n} \frac{\big|\sum_{i\in(n_1;n_2]}(i-n_1)^d\varepsilon_i\big|}{(n_2-n_1)^d(n_2\wedge(n-n_1))^{1/2}}.$$

Then, provided that

$$\int_1^\infty e^{-c_0(\psi^{-1}(t))^2}\ \mathrm{d}t < \infty \tag{3.21}$$

for some sufficiently small $c_0 = c_0(d)$, there exist some $C_1 = C_1(\psi, d) > 0$ and $C_2 = C_2(d) > 0$ such that

$$\mathbb{E}\psi(Z) \leq C_1\big[\psi\big((C_2\log\log(16n))^{1/2}\big)\vee 1\big].$$

The proof of the above theorem can be found in Appendix B.4. Here are some choices of ψ's that will be relevant in the proofs of results in Sections 3.2 and 3.3.

Example 1. Let $\psi(t) = t^\alpha$ where $\alpha > 0$. Then $\psi^{-1}(t) = t^{1/\alpha}$, so clearly (3.21) holds.

Example 2. Let $\psi(t) = e^{ct^\alpha} - 1$ where $\alpha, c > 0$. Then $\psi^{-1}(t) = \big(\log(1+t)/c\big)^{1/\alpha}$. So

$$\int_1^\infty e^{-(c_0/c^{2/\alpha})(\log(1+t))^{2/\alpha}}\ \mathrm{d}t \begin{cases} < \infty, & \alpha\in(0,2], c\in(0, c_0\mathbf{1}_{\alpha=2}+\infty\mathbf{1}_{\alpha\in(0,2)}), \\ = \infty, & \text{otherwise}. \end{cases}$$

Note that a law of iterated logarithm in expectation fails in general for the choice $\psi(t) = e^{ct^\alpha} - 1$ whenever $\alpha > 2$, as $\alpha = 2$ corresponds to the maximal integrability of Gaussian random variables.

Chapter 4

ASYMPTOTICS OF LIKELIHOOD RATIO TESTS IN GAUSSIAN SEQUENCE MODEL WITH CONVEX CONSTRAINT

4.1 Introduction

4.1.1 The likelihood ratio test

Consider the Gaussian sequence model

$$Y = \mu + \xi, \tag{4.1}$$

where $\mu \in \mathbb{R}^n$ is unknown and $\xi = (\xi_1, \ldots, \xi_n)$ is an n-dimensional standard Gaussian vector. In a variety of applications, prior knowledge on the mean vector μ can be naturally translated into the constraint $\mu \in K$, where K is a closed convex set in $\mathbb{R}^n$. We refer the readers to Chatterjee (2014a); Guntuboyina and Sen (2018) and many references therein for a diverse list of concrete examples of K. In this paper, we will be interested in the following 'goodness-of-fit' testing problem:

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \in K, \tag{4.2}$$

where $\mu_0 \in K \subset \mathbb{R}^n$, and K is an arbitrary closed and convex subset of $\mathbb{R}^n$. Throughout the manuscript, the asymptotics will take place as $n \to \infty$, and the explicit dependence of μ, μ_0, K and related quantities on the dimension n will be suppressed for ease of notation.

Given observation Y generated from model (4.1), arguably the most natural and generic test for (4.2) is the likelihood ratio test (LRT). Under the Gaussian model (4.1), the log-likelihood ratio statistic (LRS) for (4.2) takes the form

$$\begin{aligned} T(Y) &\equiv \|Y-\mu_0\|^2 - \|Y-\widehat{\mu}_K\|^2 \\ &= \|\mu+\xi-\mu_0\|^2 - \|\mu+\xi-\Pi_K(\mu+\xi)\|^2 \geq 0. \end{aligned} \tag{4.3}$$

Here $\widehat{\mu}_K \equiv \Pi_K(Y) \equiv \arg\min_{\nu\in K}\|Y-\nu\|^2$ is the metric projection of Y onto the constraint set K with respect to the canonical Euclidean ℓ_2 norm $\|\cdot\|$ on $\mathbb{R}^n$. As K is both closed and convex, Π_K is well-defined, and the resulting $\widehat{\mu}_K$ is both the least squares estimator (LSE) and the maximum likelihood estimator for the mean vector μ under the Gaussian model (4.1). The risk behavior of $\widehat{\mu}_K$ is completely characterized in the recent work Chatterjee (2014a).

The LRT for (4.2) and its generalizations thereof have gained extensive attention in the literature, see e.g. Chernoff (1954); Bartholomew (1959a,b, 1961a,b); Kudô (1963); Barlow et al. (1972); Kudô and Choi (1975); Robertson and Wegman (1978); Warrack and Robertson (1984); Shapiro (1985); Raubertas et al. (1986); Robertson et al. (1988); Shapiro (1988); Menéndez and Salvador (1991); Menéndez et al. (1992a,b); Durot and Tocquet (2001); Meyer (2003); Sen and Meyer (2017); Wei et al. (2019) for an incomplete list. In our setting, an immediate way to use the LRS $T(Y)$ in (4.3) to form a test is to simulate the critical values of $T(Y)$ under H_0. More precisely, for any confidence level $\alpha \in (0,1)$, we may determine through simulations an acceptance region $\mathcal{I}_\alpha \subset \mathbb{R}$ such that the LRS satisfies $\mathbb{P}\big(T(Y)\in\mathcal{I}_\alpha\big) = 1-\alpha$ under H_0, and then formulate the LRT accordingly. In some special cases, including the classical setting where K is a subspace, the distribution of $T(Y)$ under the null is even known in closed form, so the simulation step can be skipped.

Clearly, the almost effortless LRT as described above already gives an exact type I error control at the prescribed level for a generic pair (μ_0, K). The equally important question of its power behavior, however, is more complicated and requires a much deeper level of investigation. In the classical setting of parametric models and certain semiparametric models, the power behavior of the LRT can be precisely determined, at least asymptotically, for contiguous alternatives in the corresponding parameter spaces, cf. van der Vaart (1998, 2002). An important and basic ingredient for the success of power analysis in these settings is the existence of a limiting distribution of the LRT under H_0 that can be 'perturbed' in a large number of directions of alternatives.

Unfortunately, the distribution of the LRS $T(Y)$ in (4.3) under the null, for both finite-sample and asymptotic regimes, is only understood in very few cases. One such case is, as mentioned above, the classical setting where K is a subspace of dimension $\dim(K)$. Then the null distribution of $T(Y)$ is a chi-squared distribution with $\dim(K)$ degrees of freedom. Another case is $\mu_0 = 0$ and K is a closed convex cone. In this case, the null distribution of $T(Y)$ is the *chi-bar squared distribution*, cf. Bartholomew (1961a); Kudô (1963); Barlow et al. (1972); Kudô and Choi (1975); Shapiro (1985); Robertson et al. (1988), which can be expressed as a finite mixture of chi-squared distributions. Apart from these special cases, little next to nothing is known about the distribution of the LRS $T(Y)$ for a general pair of (μ_0, K) under the null H_0, owing in large part to the fact that the null distribution of $T(Y)$ highly depends on the exact location of μ_0 with respect to K and is thus intractable in general. Consequently, the lack of such a general description of the limiting distribution of $T(Y)$ causes a fundamental difficulty in obtaining precise characterizations of the power behavior of the LRT for a general pair (μ_0, K).

4.1.2 Normal approximation and power characterization

The unifying theme of this paper starts from the simple observation that in the classical setting where K is a subspace, as long as $\dim(K)$ diverges as $n \to \infty$, the distribution of $T(Y)$ has a progressively Gaussian shape, under proper normalization. Such a normal approximation in 'high dimensions' also holds for the more complicated chi-bar squared distribution; see Dykstra (1991); Goldstein et al. (2017) for a different set of conditions. One may therefore hope that normal approximation of $T(Y)$ under the null would hold in a far more general context than just these cases. More importantly, such a distributional approximation would ideally form the basis for power analysis of the LRT.

Normal approximation

The first main result of this paper (see Theorem 4) shows that, although the exact distribution of $T(Y)$ under H_0 is highly problem-specific and depends crucially on the pair (μ_0, K) as described above, Gaussian approximation of $T(Y)$ indeed holds in a fairly general context, after proper normalization. More concretely, we show that under H_0,

$$\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}} \approx \mathcal{N}(0,1) \quad \text{in total variation} \tag{4.4}$$

holds in the high dimensional regime where the estimation error $\mathbb{E}_{\mu_0}\|\widehat{\mu}_K - \mu_0\|^2$ diverges in some appropriate sense; see Theorem 4 and the discussion afterwards for an explanation. Here and below, $\mathcal{N}(0,1)$ denotes the standard normal distribution, and we reserve the notation

$$m_\mu \equiv \mathbb{E}_\mu(T(Y)) \qquad \text{and} \qquad \sigma_\mu^2 \equiv \mathrm{Var}_\mu(T(Y)) \tag{4.5}$$

for the mean and variance of the LRS $T(Y)$ under (4.1) with mean μ, so that m_{μ_0} and $\sigma_{\mu_0}^2$ in (4.4) are the corresponding quantities under H_0. In a similar spirit, we use the subscript μ in $\mathbb{P}_\mu$ and other probabilistic notations to indicate that the evaluation is under (4.1) with mean μ.

When the normal approximation (4.4) holds, an asymptotically equivalent formulation of the previously mentioned finite-sample LRT is the following LRT using acceptance region determined by normal quantiles: For any $\alpha \in (0,1)$, let

$$\Psi(Y) \equiv \Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) \equiv \mathbf{1}\left(\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}} \in \mathcal{A}_\alpha^c\right), \tag{4.6}$$

where $\mathcal{A}_\alpha$ is a possibly unbounded interval in $\mathbb{R}$ such that $\mathbb{P}(\mathcal{N}(0,1) \in \mathcal{A}_\alpha) = 1 - \alpha$. Common choices of $\mathcal{A}_\alpha$ include: (i) $(-\infty, z_\alpha]$ for the one-sided LRT, and (ii) $[-z_{\alpha/2}, z_{\alpha/2}]$

for the two-sided LRT, where z_α, for any $\alpha \in (0,1)$, is the normal quantile defined by $\mathbb{P}(\mathcal{N}(0,1) \geq z_\alpha) = \alpha$. In what follows, we will focus on the LRT given by (4.6), and in particular its power behavior, when the normal approximation in (4.4) holds.

Power characterization

Using the normal approximation (4.4), our second main result (see Theorem 5) shows that under mild regularity conditions,

$$\mathcal{L}\bigg(\bigg\{\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}}\bigg\}\bigg) \subset \Delta_{\mathcal{A}_\alpha}^{-1}(\beta) \quad \Leftrightarrow \quad \mathbb{E}_\mu \Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) \to \beta \in [0,1]. \tag{4.7}$$

Here $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$, and for a sequence $\{w_n\} \subset \mathbb{R}$, $\mathcal{L}(\{w_n\})$ denotes the set of all limit points of $\{w_n\}$ in $\overline{\mathbb{R}}$, and $\Delta_{\mathcal{A}_\alpha} : \overline{\mathbb{R}} \to [0,1]$ is defined in (4.16) below. In particular, $\Delta_{\mathcal{A}_\alpha}(0) = \alpha$ and $\Delta_{\mathcal{A}_\alpha}(w) = 1$ only if $w \in \{\pm\infty\}$. Hence the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 1$, if and only if

$$\mathcal{L}\bigg(\bigg\{\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}}\bigg\}\bigg) \subset \Delta_{\mathcal{A}_\alpha}^{-1}(1) \subset \{\pm\infty\}. \tag{4.8}$$

The asymptotically exact power characterization (4.7) for the LRT is rather rare beyond the classical parametric and certain semiparametric settings under contiguous alternatives (cf. van der Vaart (1998, 2002)). The setting in (4.7) can therefore be viewed as a general nonparametric analogue of contiguous alternatives for the LRT in the Gaussian sequence model (4.1).

A notable implication of (4.8) is that for any alternative $\mu \in K$, the power characterization of the LRT depends on the quantity $m_\mu - m_{\mu_0}$, which cannot in general be equivalently reduced to a usual lower bound condition on $\|\mu - \mu_0\|$. This indicates the non-uniform power behavior of the LRT with respect to the Euclidean norm $\|\cdot\|$. As the LRT (with an optimal calibration) is known to be minimax optimal in terms of uniform separation under $\|\cdot\|$ in several examples (cf. Wei et al. (2019)), the non-uniform characterization (4.8) hints that the minimax optimality criteria can be too conservative and non-informative for evaluating the power behavior of the LRT.

Another implication of (4.8) is that it is possible in certain cases that the one-sided LRT (i.e., $\mathcal{A}_\alpha = (-\infty, z_\alpha])$ has an asymptotically vanishing power, whereas the two-sided LRT (i.e., $\mathcal{A}_\alpha = [-z_{\alpha/2}, z_{\alpha/2}])$ is power consistent. This phenomenon occurs when the limit point $-\infty$ in (4.8) is achieved for certain alternatives $\mu \in K$ in the high dimensional limit. See Remark 16 ahead for a detailed discussion.

4.1.3 Testing subspace versus closed convex cone

A particularly important special setting for (4.2) is the case of testing $H_0 : \mu = 0$ versus $H_1 : \mu \in K$, where K is assumed to be a closed convex cone in $\mathbb{R}^n$. We perform a detailed case study of the following slightly more general testing problem:

$$H_0 : \mu \in K_0 \qquad \text{versus} \qquad H_1 : \mu \in K, \tag{4.9}$$

where $K_0 \subset K \subset \mathbb{R}^n$ is a subspace, and K is a closed convex cone. The primary motivation to study (4.9) arises from the problem of testing a global polynomial structure versus its shape-constrained generalization; concrete examples include constancy versus monotonicity, linearity versus convexity, etc.; see Section 4.4.5 for details. The LRS for (4.9) takes the slightly modified form:

$$\begin{aligned} T(Y) \equiv T_{K_0}(Y) &\equiv \|Y - \widehat{\mu}_{K_0}\|^2 - \|Y - \widehat{\mu}_K\|^2 \\ &= \|\mu + \xi - \Pi_{K_0}(\mu + \xi)\|^2 - \|\mu + \xi - \Pi_K(\mu + \xi)\|^2. \end{aligned} \tag{4.10}$$

The dependence in the notation of the LRS $T(Y)$ on K_0 will usually be suppressed when no confusion could arise.

Specializing our first main result to this testing problem, we show in Theorem 6 that normal approximation of $T(Y)$ under H_0 holds essentially under the minimal growth condition that $\delta_K - \dim(K_0) \to \infty$, where δ_K is the statistical dimension of K (formally defined in Definition 1). Similar to (4.7), the normal approximation makes possible the following precise characterization of the power behavior of the LRT under the prescribed growth condition (see Theorem 7):

$$\mathcal{L}\bigg(\bigg\{\frac{\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\|^2 - \mathbb{E}\|\Pi_K(\xi)\|^2}{\sigma_0}\bigg\}\bigg) \subset \Delta^{-1}_{\mathcal{A}_\alpha}(\beta) \cap [0, +\infty]$$
$$\Leftrightarrow \quad \mathbb{E}_\mu \Psi(Y; m_0, \sigma_0) \to \beta \in [0, 1]. \tag{4.11}$$

As $\sigma_0^2 = \mathrm{Var}(T(\xi)) \asymp \delta_K - \dim(K_0)$ (cf. Lemma 2) for the modified LRS $T(Y)$ in (4.10), the LRT is power consistent under μ if and only if

$$\frac{\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\|^2 - \mathbb{E}\|\Pi_K(\xi)\|^2}{\big(\delta_K - \dim(K_0)\big)^{1/2}} \to +\infty. \tag{4.12}$$

Compared to the uniform $\|\cdot\|$-separation rate derived in the recent work Wei et al. (2019) (cf. (4.33) below), (4.11)-(4.12) provide asymptotically precise power characterizations of the LRT for a sequence of point alternatives. This difference is indeed crucial as (4.12), similar to (4.8), cannot be equivalently inverted into a lower bound on $\|\mu - \Pi_{K_0}(\mu)\|$ alone. This illustrates that the non-uniform power behavior of the LRT is not an aberration in certain artificial testing problems, but is rather a fundamental property of the LRT in the high dimensional regime that already appears in the special yet important setting of testing subspace versus a cone.

4.1.4 Examples

As an illustration of the scope of our theoretical results, we validate the normal approximation of the LRT and exemplify its power behavior in two classes of problems:

1. Testing in orthant/circular cone, isotonic regression and Lasso;

2. Testing parametric assumptions versus shape-constrained alternatives, e.g., constancy versus monotonicity, linearity versus convexity, and generalizations thereof.

Non-uniform power of the LRT

Some of the above problems give clear examples of the aforementioned non-uniform power behavior of the LRT: In the problem of testing $\mu = 0$ versus the orthant and (product) circular cone, the LRT is indeed powerful against most alternatives in the region where the uniform separation in $\|\cdot\|$ is not informative. More concretely:

- In the case of the orthant cone, the LRT is known to be minimax optimal (cf. Wei et al. (2019)) in terms of the uniform $\|\cdot\|$-separation of the order $n^{1/4}$. Our results show that the LRT is actually powerful for 'most' alternatives μ with $\|\mu\| = \mathcal{O}(n^{1/4})$, including some with $\|\cdot\|$-separation of the order n^{δ} for any $\delta > 0$. This showcases the conservative nature of the minimax optimality criteria. See Section 4.4.1 for details.

- In the case of (product) circular cone, the LRT is known to be minimax sub-optimal (cf. Wei et al. (2019)) with $\|\cdot\|$-separation of the order $n^{1/4}$ while the minimax separation rate is of the constant order. Our results show the minimax sub-optimality is witnessed only by a few unfortunate alternatives and the LRT is powerful within a large cylindrical set including many points of constant $\|\cdot\|$-separation order. This also identifies the minimax framework as too pessimistic for the sub-optimality results of the LRT; see Section 4.4.2 for details.

4.1.5 Organization

The rest of the paper is organized as follows. Section 4.2 reviews some basic facts on metric projection and conic geometry. Section 4.3 studies normal approximation for the LRS $T(Y)$ and the power characterizations of the LRT both in the general setting (4.2) and the more structured setting (4.9). Applications of the abstract theory to the examples mentioned above are detailed in Section 4.4. Proofs are collected in Sections C.1 and C.2 and the appendix.

4.1.6 Notation

For any positive integer n, let $[1:n]$ denote the set $\{1,\ldots,n\}$. For $a,b\in\mathbb{R}$, $a\vee b\equiv\max\{a,b\}$ and $a\wedge b\equiv\min\{a,b\}$. For $a\in\mathbb{R}$, let $a_\pm\equiv(\pm a)\vee 0$. For $x\in\mathbb{R}^n$, let $\|x\|_p$ denote its p-norm ($0\le p\le\infty$), and $B_p(r;x)\equiv\{z\in\mathbb{R}^n:\|z-x\|_p\le r\}$. We simply write $\|x\|\equiv\|x\|_2$, $B(r;x)\equiv B_2(r;x)$, and $B(r)\equiv B(r;0)$ for notational convenience. By $\mathbf{1}_n$ we denote the vector of all ones in $\mathbb{R}^n$. For a matrix $M\in\mathbb{R}^{n\times n}$, let $\|M\|$ and $\|M\|_F$ denote the spectral and Frobenius norms of M respectively.

For a multi-index $\boldsymbol{k}=(k_1,\ldots,k_n)\in\mathbb{Z}^n_{\ge 0}$, let $|\boldsymbol{k}|\equiv\sum_{i=1}^n k_i$. For $f:\mathbb{R}^n\to\mathbb{R}$, and $\boldsymbol{k}=(k_1,\ldots,k_n)\in\mathbb{Z}^n_{\ge 0}$, let $\partial_{\boldsymbol{k}}f(z)\equiv\frac{\partial^{|\boldsymbol{k}|}f(z)}{\partial_{k_1}z_1\cdots\partial_{k_n}z_n}$ for $z\in\mathbb{R}^n$ whenever definable. A vector-valued map $f:\mathbb{R}^n\to\mathbb{R}^m$ is said to have sub-exponential growth at ∞ if $\lim_{\|x\|\to\infty}\|f(x)e^{-\|x\|}\|=0$. For $f=(f_1,\ldots,f_n):\mathbb{R}^n\to\mathbb{R}^n$, let $J_f(z)\equiv(\partial f_i(z)/\partial z_j)_{i,j=1}^n$ denote the Jacobian of f and

$$\operatorname{div} f(z)\equiv\sum_{i=1}^n\frac{\partial}{\partial z_i}f_i(z)=\operatorname{Tr}(J_f(z))$$

for $z\in\mathbb{R}^n$ whenever definable.

We use C_x to denote a generic constant that depends only on x, whose numeric value may change from line to line unless otherwise specified. $a\lesssim_x b$ and $a\gtrsim_x b$ mean $a\le C_x b$ and $a\ge C_x b$ respectively, and $a\asymp_x b$ means $a\lesssim_x b$ and $a\gtrsim_x b$ ($a\lesssim b$ means $a\le Cb$ for some absolute constant C). For two nonnegative sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n\ll b_n$ (respectively $a_n\gg b_n$) if $\lim_{n\to\infty}(a_n/b_n)=0$ (respectively $\lim_{n\to\infty}(a_n/b_n)=\infty$). We follow the convention that $0/0=0$. $\mathcal{O}_{\mathbf{P}}$ and $\mathfrak{o}_{\mathbf{P}}$ denote the usual big and small O notation in probability.

We reserve the notation $\xi=(\xi_1,\ldots,\xi_n)$ for an n-dimensional standard normal random vector, and φ,Φ for the density and the cumulative distribution function of a standard normal random variable. For any $\alpha\in(0,1)$, let z_α be the normal quantile defined by $\mathbb{P}(\mathcal{N}(0,1)\ge z_\alpha)=\alpha$. For two random variables X,Y on $\mathbb{R}$, we use $d_{\mathrm{TV}}(X,Y)$ and $d_{\mathrm{Kol}}(X,Y)$

to denote their total variation distance and Kolmogorov distance defined respectively as

$$d_{\mathrm{TV}}(X,Y) \equiv \sup_{B\in\mathcal{B}(\mathbb{R})} \big|\mathbb{P}\big(X\in B\big) - \mathbb{P}\big(Y\in B\big)\big|,$$

$$d_{\mathrm{Kol}}(X,Y) \equiv \sup_{t\in\mathbb{R}} \big|\mathbb{P}\big(X\le t\big) - \mathbb{P}\big(Y\le t\big)\big|.$$

Here $\mathcal{B}(\mathbb{R})$ denotes all Borel measurable sets in $\mathbb{R}$.

4.2 Preliminaries: metric projection and conic geometry

In this section, we review some basic facts on metric projection and conic geometry. For any $x\in\mathbb{R}^n$, the metric projection of x onto a closed convex set $K\subset\mathbb{R}^n$ is defined by

$$\Pi_K(x) \equiv \operatorname*{argmin}_{y\in K} \|x-y\|^2.$$

It is a standard fact that the map Π_K is well-defined, 1-Lipschitz and hence absolutely continuous. The Jacobian J_{Π_K} is therefore almost everywhere (a.e.) well-defined.

Let $G:\mathbb{R}^n\to\mathbb{R}$ be defined by

$$G(y) \equiv \|y-\Pi_K(y)\|^2.$$

We summarize some useful properties of G and J_{Π_K} in the following lemma.

Lemma 1. *The following statements hold.*

1. *G is absolutely continuous and its gradient $\nabla G(y) = 2(y-\Pi_K(y))$ has sub-exponential growth at ∞.*
2. *For a.e. $y\in\mathbb{R}^n$, $\|J_{\Pi_K}(y)\|\vee\|I-J_{\Pi_K}(y)\|\le 1$ and $J_{\Pi_K}(y)^\top\Pi_K(y) = J_{\Pi_K}(y)^\top y$.*

Proof. (1) follows from (Goldstein et al., 2017, Lemma 2.2) and the proof of (Goldstein et al., 2017, Lemma A.2). The first claim of (2) is proved in (Goldstein et al., 2017, Lemma 2.1). For the second claim of (2), note that $\nabla G(y) = 2(I-J_{\Pi_K}(y))^\top(y-\Pi_K(y))$. By (1), $\nabla G(y) = 2(y-\Pi_K(y))$, so $J_{\Pi_K}(y)^\top(y-\Pi_K(y)) = 0$, proving the claim. $\square$

Recall that a closed and convex cone $K \subset \mathbb{R}^n$ is polyhedral if it is a finite intersection of closed half-spaces, and a face of K is a set of the form $K \cap H$, where H is a supporting hyperplane of K in $\mathbb{R}^n$. Let $\mathrm{lin}(F)$ denote the linear span of F. The dimension of a face F is $\dim F \equiv \dim(\mathrm{lin}(F))$, and the relative interior of F is the interior of F in $\mathrm{lin}(F)$.

The complexity of a closed convex cone K can be described by its statistical dimension defined as follows.

Definition 1. *The* statistical dimension δ_K *of a closed convex cone K is defined as* $\delta_K \equiv \mathbb{E}\|\Pi_K(\xi)\|^2$.

The statistical dimension δ_K has several equivalent definitions; see e.g. (Amelunxen et al., 2014, Proposition 3.1). In particular, $\delta_K = \mathbb{E}\sup_{\nu \in K \cap B(1)}(\langle \nu, \xi \rangle)^2$. For any polyhedral cone $K \subset \mathbb{R}^n$ and $j \in \{0, \ldots, n\}$, the j-th intrinsic volume of K is defined as

$$v_j(K) \equiv \mathbb{P}\big(\Pi_K(\xi) \in \text{relative interior of a } j\text{-dimensional face of } K\big). \tag{4.13}$$

More generally, the intrinsic volumes $\{v_j(K)\}_{j=0}^n$ for a closed convex cone $K \subset \mathbb{R}^n$ are defined as the limit of (4.13) using polyhedral approximation; see (McCoy and Tropp, 2014, Section 7.3). These quantities are well-defined and have been investigated in considerable depth; see e.g., Amelunxen et al. (2014); McCoy and Tropp (2014); Goldstein et al. (2017).

Definition 2. *For any closed convex cone $K \subset \mathbb{R}^n$, let V_K be a random variable taking values in $\{0, \ldots, n\}$ such that $\mathbb{P}(V_K = j) = v_j(K)$.*

We summarize some useful properties of δ_K and V_K in the following lemma. An elementary and self-contained proof is given in Appendix C.3.1.

Lemma 2. *Let K be a convex closed cone. Then*

1. $\delta_K = \mathbb{E}V_K$;

2. $\mathit{Var}(\|\Pi_K(\xi)\|^2) = \mathit{Var}(V_K) + 2\delta_K$;

3. $2\delta_K \leq Var(\|\Pi_K(\xi)\|^2) \leq 2\delta_K + 2\|\mathbb{E}\Pi_K(\xi)\|^2 \leq 4\delta_K$.

For any closed convex cone $K \subset \mathbb{R}^n$, its polar cone is defined as

$$K^* \equiv \{v \in \mathbb{R}^n : \langle v, u\rangle \leq 0, \text{ for all } u \in K\}. \tag{4.14}$$

With Π_{K^*} denoting the metric projection onto K^*, Moreau's theorem (Rockafellar, 1997, Theorem 31.5) states that for any $v \in \mathbb{R}^n$,

$$v = \Pi_K(v) + \Pi_{K^*}(v) \qquad \text{with } \langle \Pi_K(v), \Pi_{K^*}(v)\rangle = 0.$$

4.3 Theory

4.3.1 Normal approximation for $T(Y)$ and power characterizations

We start by presenting the normal approximation result for $T(Y)$ in (4.3) under the null hypothesis (4.2); see Section C.1.1 for a proof. This will serve as the basis for the size and power analysis of the LRT (4.6) in the testing problem (4.2).

Theorem 4. *Let $K \subset \mathbb{R}^n$ be a closed convex set and $\mu_0 \in K$. Then under H_0,*

$$d_{\mathrm{TV}}\left(\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}}, \mathcal{N}(0,1)\right) \leq \frac{8\sqrt{\mathbb{E}_{\mu_0}\|\widehat{\mu}_K - \mu_0\|^2}}{2\|\mathbb{E}_{\mu_0}\widehat{\mu}_K - \mu_0\|^2 + \|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_K}\|_F^2}. \tag{4.15}$$

Here $J_{\widehat{\mu}_K} \equiv J_{\widehat{\mu}_K}(\xi) \equiv J_{\Pi_K}(\mu_0 + \xi)$, and $m_{\mu_0}, \sigma_{\mu_0}$ are as defined in (4.5).

The bound (4.15) is obtained by a generalization of (Goldstein et al., 2017, Theorem 2.1) using the second-order Poincaré inequality Chatterjee (2009), together with a lower bound for $\sigma^2_{\mu_0}$ using Fourier analysis in the Gaussian space (Nourdin and Peccati, 2012, Section 1.5). The Fourier expansion can be performed up to the second order thanks to the absolute continuity of the first derivative of $T(Y)$ (cf. Lemma 1).

We now comment on the structure of (4.15). The first term $\|\mathbb{E}_{\mu_0}\widehat{\mu}_K - \mu_0\|^2$ in the denominator is the squared bias of the projection estimator $\widehat{\mu}_K$, while the second term $\|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_K}\|_F^2$, which depends on the magnitudes of the first-order partial derivatives of $\widehat{\mu}_K$, can be roughly understood as the 'variance' of $\widehat{\mu}_K$. Consequently, one may expect that the denominator is of the order $\mathbb{E}_{\mu_0}\|\widehat{\mu}_K - \mu_0\|^2$, so the overall bound scales as $\mathcal{O}\big(1/\sqrt{\mathbb{E}_{\mu_0}\|\widehat{\mu}_K - \mu_0\|^2}\big)$. As will be clear in Section 4.4, this is indeed the case in all the examples worked out, and the major step in applying (4.15) to concrete problems typically depends on obtaining sharp lower bounds for the 'variance' term $\|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_K}\|_F^2$, which may require non-trivial problem-specific techniques.

Using Theorem 4, we may characterize the size and power behavior of the LRT. For a possibly unbounded interval $I \subset \mathbb{R}$, let $\Delta_I : \overline{\mathbb{R}} \to [0,1]$ be defined as follows: For $w \in \mathbb{R}$,

$$\Delta_I(w) \equiv 1 - \mathbb{P}\big(\mathcal{N}(0,1) \in I - w\big) = \mathbb{P}\big(\mathcal{N}(0,1) \in I^c - w\big), \tag{4.16}$$

and $\Delta_I(\pm\infty) \equiv \lim_{w\to\pm\infty} \Delta_I(w)$, which is clearly well-defined. Δ_I is either monotonic or unimodal, so $\Delta_I^{-1}(\beta)$ contains at most two elements for any $\beta \in [0,1]$. Two primary examples of I are $\mathcal{A}_\alpha^{\mathrm{os}} \equiv (-\infty, z_\alpha]$ and $\mathcal{A}_\alpha^{\mathrm{ts}} \equiv [-z_{\alpha/2}, z_{\alpha/2}]$ — the acceptance regions for the one- and two-sided LRTs respectively, where we have

$$\Delta_{\mathcal{A}_\alpha^{\mathrm{os}}}(w) = \Phi(-z_\alpha + w), \;\; \Delta_{\mathcal{A}_\alpha^{\mathrm{ts}}}(w) = \Phi(-z_{\alpha/2} + w) + \Phi(-z_{\alpha/2} - w). \tag{4.17}$$

It is clear that $\Delta_{\mathcal{A}_\alpha^{\mathrm{os}}}(0) = \Delta_{\mathcal{A}_\alpha^{\mathrm{ts}}}(0) = \alpha$, $\Delta^{-1}_{\mathcal{A}_\alpha^{\mathrm{os}}}(1) = \{+\infty\}$, and $\Delta^{-1}_{\mathcal{A}_\alpha^{\mathrm{ts}}}(1) = \{\pm\infty\}$. Recall the definitions of m_μ and σ_μ^2 for general $\mu \in K$ in (4.5). The following result (see Section C.1.2 for a proof) characterizes the power behavior of the LRT.

Theorem 5. *Consider testing (4.2) using the LRT as in* (4.6). *There exists some constant* $C_{\mathcal{A}_\alpha} > 0$ *such that*

$$\begin{aligned}\bigg|\mathbb{E}_\mu \Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) - \Delta_{\mathcal{A}_\alpha}\bigg(\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}}\bigg)\bigg| \\ \leq 2 \cdot \mathrm{err}_{\mu_0} + C_{\mathcal{A}_\alpha} \cdot \mathscr{L}\bigg(1 \bigwedge \frac{\|\mu - \mu_0\|}{|m_\mu - m_{\mu_0}| \vee \sigma_{\mu_0}}\bigg).\end{aligned} \tag{4.18}$$

Here

$$\mathrm{err}_{\mu_0} \equiv d_{\mathrm{Kol}}\bigg(\frac{T(\mu_0+\xi)-m_{\mu_0}}{\sigma_{\mu_0}}, \mathcal{N}(0,1)\bigg) \leq \textit{right hand side of (4.15)},$$

and $\mathcal{L}(x) \equiv x\sqrt{1 \vee \log(1/x)}$ *for* $x > 0$ *and* $\mathcal{L}(0) \equiv 0$. *Consequently:*

1. *The LRT in* (4.6) *has size*

$$\big|\mathbb{E}_{\mu_0}\Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) - \alpha\big| \leq 2 \cdot \mathrm{err}_{\mu_0}.$$

2. *Suppose the normal approximation of* $T(Y)$ *holds under* H_0, *i.e.,* $\mathrm{err}_{\mu_0} \to 0$. *Then, for any* $\mu \in K$ *such that*

$$\|\mu - \mu_0\| \ll |m_\mu - m_{\mu_0}| \vee \sigma_{\mu_0}, \tag{4.19}$$

we have

$$\mathcal{L}\bigg(\bigg\{\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}}\bigg\}\bigg) \subset \Delta_{\mathcal{A}_\alpha}^{-1}(\beta) \;\Leftrightarrow\; \mathbb{E}_\mu\Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) \to \beta \in [0,1]. \tag{4.20}$$

Hence under (4.19), the LRT is power consistent under μ, *i.e.,* $\mathbb{E}_\mu\Psi(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 1$, *if and only if*

$$\mathcal{L}\bigg(\bigg\{\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}}\bigg\}\bigg) \subset \Delta_{\mathcal{A}_\alpha}^{-1}(1) \subset \{\pm\infty\}. \tag{4.21}$$

Remark 15. *The validity of the normal approximation in Theorem 5-(2) is imposed to express the exact power behavior (4.20) with the normal quantile. More generally, as long as the normalized LRS* $\big(T(Y) - m_{\mu_0}\big)/\sigma_{\mu_0}$ *has a distributional limit under* H_0, *(4.20) can be obtained accordingly with the corresponding quantiles.*

We now comment on conditions (4.19) and (4.21) in detail.

- Condition (4.19) centers around the key deviation quantity

$$\Delta T_{\mu,\mu_0}(\xi) \equiv T(\mu+\xi) - T(\mu_0+\xi), \tag{4.22}$$

which can be shown to satisfy

$$\mathbb{E}(\Delta T_{\mu,\mu_0}) = m_\mu - m_{\mu_0}, \qquad \mathrm{Var}(\Delta T_{\mu,\mu_0}) \le \|\mu-\mu_0\|^2.$$

Moreover, it can be shown that $\Delta T_{\mu,\mu_0}$ concentrates around its mean $m_\mu - m_{\mu_0}$ with sub-Gaussian tails (see Proposition 30). This concentration result allows us to connect the normal approximation under the null in Theorem 4 to the power behavior of the LRT under the alternative.

To further understand condition (4.19), note that in the small separation regime $\|\mu-\mu_0\| \ll \sigma_{\mu_0}$, (4.19) is automatically fulfilled; in the large separation regime where $\|\mu-\mu_0\| \gg \sigma_{\mu_0}$, (4.19) can typically be verified by establishing a *quadratic lower bound* $|m_\mu - m_{\mu_0}| \gtrsim \|\mu-\mu_0\|^2$. In this sense (4.19) excludes possibly ill-behaved alternatives that violate the prescribed quadratic lower bound in the critical regime $\|\mu-\mu_0\| \asymp \sigma_{\mu_0}$; see e.g., Example 4. Hence (4.19) cannot be removed in general for the validity of the power characterization (4.20).

- To verify (4.21), some problem specific understanding for m_μ and σ_{μ_0} is needed. As $\mathbb{E}_\mu\langle\xi,\widehat{\mu}_K\rangle = \mathbb{E}_\mu \operatorname{div}\widehat{\mu}_K$ by Stein's identity, we have

$$m_\mu = \|\mu-\mu_0\|^2 + 2\mathbb{E}_\mu \operatorname{div}\widehat{\mu}_K - \mathbb{E}_\mu\|\widehat{\mu}_K-\mu\|^2, \tag{4.23}$$

hence the numerator of (4.21) requires sharp estimates of the expected 'degrees of freedom' $\mathbb{E}_\mu \operatorname{div}\widehat{\mu}_K$ (cf. Meyer and Woodroofe (2000)), and the estimation error $\mathbb{E}_\mu\|\widehat{\mu}_K-\mu\|^2$. A (near) matching upper and lower bound for σ_{μ_0} will also be required to obtain necessary and sufficient characterizations. We mention that (4.21) cannot in general be equivalently inverted into a lower bound on $\|\mu-\mu_0\|$ only; see the remarks after Theorem 7 for a more detailed discussion.

Remark 16. *The LRT defined in (4.6) depends on the choice of the acceptance region $\mathcal{A}_\alpha$. Two obvious choices are:*

1. *(*One-sided LRT*). Let $\mathcal{A}_\alpha \equiv \mathcal{A}_\alpha^{os} = (-\infty, z_\alpha]$. This leads to the following one-sided LRT:*

$$\Psi_{os}(Y) \equiv \Psi_{os}(Y; m_{\mu_0}, \sigma_{\mu_0}) \equiv \mathbf{1}\bigg(\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}} > z_\alpha\bigg). \tag{4.24}$$

2. *(*Two-sided LRT*). Let $\mathcal{A}_\alpha \equiv \mathcal{A}_\alpha^{ts} = [-z_{\alpha/2}, z_{\alpha/2}]$. This leads to the following two-sided LRT:*

$$\Psi_{ts}(Y) \equiv \Psi_{ts}(Y; m_{\mu_0}, \sigma_{\mu_0}) \equiv \mathbf{1}\bigg(\bigg|\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}}\bigg| > z_{\alpha/2}\bigg). \tag{4.25}$$

In the classical case where K is a subspace of fixed dimension, the one-sided LRT is power consistent (under $\mu \in K$) if and only if the two-sided LRT is power consistent, so one can simply use the standard one-sided LRT. The situation can be rather different for certain high dimensional instances of K. Under the setting of Theorem 5-(2), as $\Delta^{-1}_{\mathcal{A}_\alpha^{os}}(1) = \{+\infty\}$ while $\Delta^{-1}_{\mathcal{A}_\alpha^{ts}}(1) = \{\pm\infty\}$, power consistency under μ for the one-sided LRT implies that for the two-sided LRT, but the converse fails when the $-\infty$ limit in (4.21) is achieved. See Example 5 ahead for a concrete example. However, in the special case where $\mu_0 = 0$ and K is a closed convex cone, $(m_\mu - m_{\mu_0})/\sigma_{\mu_0}$ can only diverge to $+\infty$ under mild growth condition on K, so in this case power consistency is equivalent for one- and two-sided LRTs. Also see Remark 17-(1).

As a simple toy example of Theorem 5, we consider the testing problem (4.2) in the linear regression case, where $K \equiv K_X \equiv \{X\theta : \theta \in \mathbb{R}^p\}$ for some fixed design matrix $X \in \mathbb{R}^{n\times p}$, with $p \leq n$. We will be interested in the high dimensional regime $\mathsf{rank}(X) \to \infty$ where the normal approximation for the LRT holds under the null.

Proposition 17. *Consider testing (4.2) with $K = K_X$. Suppose that $\mathsf{rank}(X) \to \infty$. Let $\Psi \in \{\Psi_{\mathrm{os}}, \Psi_{\mathrm{ts}}\}$.*

1. *If $\mu_0 \in K_X$, then*

$$d_{\mathrm{TV}}\left(\frac{T(Y)-m_{\mu_0}}{\sigma_{\mu_0}}, \mathcal{N}(0,1)\right) \leq \frac{8}{\sqrt{\mathsf{rank}(X)}}. \tag{4.26}$$

   Consequently the LRT is asymptotically size α with $\mathbb{E}_{\mu_0}\Psi(Y) = \alpha + \mathcal{O}(1/\sqrt{\mathsf{rank}(X)})$.

2. *For any $\mu \in K_X$, $m_\mu - m_{\mu_0} = \|\mu - \mu_0\|^2$, and the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi(Y) \to 1$, if and only if $\|\mu - \mu_0\| \gg (\mathsf{rank}(X))^{1/4}$.*

Proof. (1). Note that $\widehat{\mu}_{K_X} = \Pi_{K_X}(Y) = X(X^\top X)^- X^\top Y \equiv PY$, where A^- denotes the pseudo-inverse for A. Then $\mathbb{E}_{\mu_0}\widehat{\mu}_{K_X} = P\mu_0 = \mu_0$, $J_{\widehat{\mu}_{K_X}} = P$ and

$$\mathbb{E}_{\mu_0}\|\widehat{\mu}_{K_X} - \mu_0\|^2 = \|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_{K_X}}\|_F^2 = \mathrm{Tr}(PP^\top) = \dim(K_X) = \mathsf{rank}(X).$$

The claim (1) now follows from Theorem 4.
(2). By (4.23), for any $\mu \in K_X$,

$$\begin{aligned} m_\mu &= \|\mu - \mu_0\|^2 + \mathbb{E}\big[2\langle \xi, P\xi\rangle - \|P\xi\|^2\big] \\ &= \|\mu - \mu_0\|^2 + \mathbb{E}[\|P\xi\|^2] = \|\mu - \mu_0\|^2 + \mathsf{rank}(X), \end{aligned}$$

and with $\mu_0 \in K_X$,

$$\sigma_{\mu_0}^2 = \mathrm{Var}\big(\|P\xi\|^2\big) = \mathsf{rank}(X).$$

As $\|\mu - \mu_0\| \ll \|\mu - \mu_0\|^2 \vee \sqrt{\mathsf{rank}(X)}$ always holds, the claim follows from Theorem 5-(2). $\square$

More examples on testing in orthant/circular cone, isotonic regression and Lasso are worked out in Section 4.4.

4.3.2 Subspace versus closed convex cone

In this subsection, we study in detail the testing problem (4.9) as an important special case of (4.2). The additional subspace and cone structure will allow us to give more explicit characterizations of the size and the power of the LRT; note that here the LRS $T(Y)$ takes the modified form (4.10). We start with the following simple observation.

Lemma 3. *Let K be a closed convex set in $\mathbb{R}^n$. Then for μ such that $K - \mu \subset K$, we have*

$$\Pi_K(\mu + \xi) = \mu + \Pi_K(\xi), \quad \forall \xi \in \mathbb{R}^n.$$

Consequently,

$$\|\mu + \xi - \Pi_K(\mu + \xi)\|^2 = \|\xi - \Pi_K(\xi)\|^2.$$

Proof. By the definition of projection, we want to verify

$$\langle \mu + \xi - (\mu + \Pi_K(\xi)), \nu - (\mu + \Pi_K(\xi)) \rangle \leq 0, \quad \forall \nu \ \in K.$$

This amounts to verifying that

$$\langle \xi - \Pi_K(\xi), (\nu - \mu) - \Pi_K(\xi) \rangle \leq 0, \quad \forall \nu \in K.$$

As $\nu - \mu \in K$ by the condition $K - \mu \subset K$, the above inequality holds by the projection property for $\Pi_K(\xi)$. $\square$

Recall the definition of the statistical dimension δ_K in Definition 1. The above lemma provides us with simplifications of m_μ and σ_μ^2 as defined in (4.5): under the setting of (4.9), for any $\mu \in K_0$,

$$m_\mu \equiv m_0 = \delta_K - \delta_{K_0}, \qquad \sigma_\mu^2 \equiv \sigma_0^2 = \mathrm{Var}\big(\|\Pi_K(\xi)\|^2 - \|\Pi_{K_0}(\xi)\|^2\big). \tag{4.27}$$

Moreover, as K_0 is a subspace, we have $\delta_{K_0} = \dim(K_0)$. The following result (proved in Section C.1.3) derives the normal approximation of $T(Y)$ with an explicit error bound.

Theorem 6. *Suppose $K_0 \subset K \subset \mathbb{R}^n$ are such that K_0 is a subspace and K is a closed convex cone. Then for $\mu \in K_0$,*

$$d_{\mathrm{TV}}\left(\frac{T(Y) - m_0}{\sigma_0}, \mathcal{N}(0,1)\right) \leq \frac{8}{\sqrt{\delta_K - \delta_{K_0}}}.$$

It is easy to see from the above bound that under the growth condition $\delta_K - \delta_{K_0} \to \infty$, normal approximation of $T(Y)$ holds under the null. This growth condition cannot be improved in general: for a subspace K, $T(Y)$ follows a chi-squared distribution with $\delta_K - \delta_{K_0}$ degrees of freedom under the null, so normal approximation holds if and only if $\delta_K - \delta_{K_0} \to \infty$. The above theorem extends (Goldstein et al., 2017, Theorem 2.1) in which the case $K_0 = \{0\}$ is treated. Compared to classical results on the chi-bar squared distribution (Dykstra, 1991, Corollary 2.2), the growth condition here does not require exact knowledge for the mixing weights, and can be easily checked using Gaussian process techniques; see Section 4.4.5 for examples.

Using Theorem 6, we can prove sharp size and power behavior of the LRT; see Theorem 7 below (proved in Section C.1.4). For $p \geq 1$, let

$$\Gamma_{K,p}(\nu) \equiv \mathbb{E}\|\Pi_K\big(\nu + \xi\big)\|^p - \mathbb{E}\|\Pi_K(\xi)\|^p, \quad \nu \in \mathbb{R}^n.$$

We simply shorthand $\Gamma_{K,1}$ as Γ_K for notational convenience. Recall the definition of V_K in Definition 2 and that of the polar cone K^* in (4.14).

Theorem 7. *Consider testing (4.9) using the LRT $\Psi(Y; m_0, \sigma_0)$ with the modified LRS $T(Y)$ in (4.10). There exist constants $C_{\mathcal{A}_\alpha}, C'_{\mathcal{A}_\alpha} > 0$ such that*

$$\begin{aligned}
&\left|\mathbb{E}_\mu \Psi(Y; m_0, \sigma_0) - \Delta_{\mathcal{A}_\alpha}\left(\frac{\Gamma_{K,2}\big(\mu - \Pi_{K_0}(\mu)\big)}{\sigma_0}\right)\right| \\
&\qquad \leq 2 \cdot \mathrm{err}_0 + C_{\mathcal{A}_\alpha} \cdot \mathcal{L}\left(1 \bigwedge \frac{\|\mu - \Pi_{K_0}(\mu)\|}{\big|\Gamma_{K,2}\big(\mu - \Pi_{K_0}(\mu)\big)\big| \vee \sigma_0}\right) && (4.28) \\
&\qquad \leq C'_{\mathcal{A}_\alpha} \cdot \mathcal{L}\Big(\big(\delta_K - \delta_{K_0}\big)^{-1/4}\Big). && (4.29)
\end{aligned}$$

Here $\mathrm{err}_0, \mathcal{L}(\cdot)$ are defined in Theorem 5. Consequently:

1. *For $\mu \in K_0$, the LRT has size $\mathbb{E}_0\Psi(Y; m_0, \sigma_0)$, where*

$$\big|\mathbb{E}_0 \Psi(Y; m_0, \sigma_0) - \alpha\big| \leq \frac{16}{\sqrt{\delta_K - \delta_{K_0}}}.$$

2. Suppose further $\delta_K - \delta_{K_0} \to \infty$. Then for $\mu \in K$,

$$
\begin{aligned}
&\mathcal{L}\bigg(\bigg\{\frac{\Gamma_{K,2}\big(\mu - \Pi_{K_0}(\mu)\big)}{\sigma_0}\bigg\}\bigg) \subset \Delta^{-1}_{\mathcal{A}_\alpha}(\beta) \cap [0, +\infty] \\
\Leftrightarrow\quad &\mathcal{L}\bigg(\bigg\{\frac{2\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)}{\sqrt{2 + r(K, K_0)}\sqrt{1 - \delta_{K_0}/\delta_K}}\bigg\}\bigg) \subset \Delta^{-1}_{\mathcal{A}_\alpha}(\beta) \cap [0, +\infty] \\
\Leftrightarrow\quad &\mathbb{E}_\mu \Psi(Y; m_0, \sigma_0) \to \beta \in [0, 1], \qquad (4.30)
\end{aligned}
$$

where $r(K, K_0) \equiv Var(V_{K\cap K_0^*})/\delta_{K\cap K_0^*} \in [0, 2]$. Hence the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi(Y; m_0, \sigma_0) \to 1$, if and only if

$$
\frac{\Gamma_{K,2}\big(\mu - \Pi_{K_0}(\mu)\big)}{\big(\delta_K - \delta_{K_0}\big)^{1/2}} \to +\infty \quad \Leftrightarrow \quad \frac{\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)}{\sqrt{1 - \delta_{K_0}/\delta_K}} \to +\infty. \qquad (4.31)
$$

Remark 17. *1. By the proof of (Wei et al., 2019, Lemma E.1), $\Gamma_{K,2}(\nu) \geq \|\nu\|^2 \geq 0$ for all $\nu \in K$, so all the limit points in (4.30) are nonnegative. This leads to the equivalence of the power consistency property for the one-sided LRT (4.24) and the two-sided LRT (4.25).*

2. With the help of Lemma 3 and (4.27), which holds for any $\mu \in K_0$, some calculations yield that

$$
m_\mu - m_0 = \Gamma_{K,2}(\mu - \Pi_{K_0}(\mu)) \geq \|\mu - \Pi_{K_0}(\mu)\|^2. \qquad (4.32)
$$

Therefore, the counterpart of the generic condition (4.19) under (4.9)

$$
\|\mu - \Pi_{K_0}(\mu)\| \ll |m_\mu - m_0| \vee \sigma_0
$$

is automatically satisfied due to the global quadratic lower bound (4.32). In particular, (4.29) vanishes under the growth condition $\delta_K - \delta_{K_0} \to \infty$.

The power behavior of the LRT is characterized using $\Gamma_{K,2}$ and Γ_K in Theorem 7. The function $\Gamma_{K,2}$ is usually more amenable to explicit calculations in concrete examples, while the formulation using Γ_K allows us to recover the separation rate in $\|\cdot\|$ for the LRT derived in Wei et al. (2019) in the setting (4.9). We formally state this result below; see Section C.1.5 for a proof.

Corollary 1. *For $\Psi \in \{\Psi_{\mathrm{os}}, \Psi_{\mathrm{ts}}\}$, (4.31) is satisfied for any $\mu \in K$ such that*

$$\|\mu - \Pi_{K_0}(\mu)\| \gg \delta_K^{1/4} \bigwedge \left(\frac{\delta_K^{1/2}}{0 \bigvee \inf_{\eta \in K \cap B(1)} \langle \eta, \mathbb{E}\Pi_K(\xi) \rangle} \right). \tag{4.33}$$

Below we give a detailed comparison of (4.31) and its sufficient condition (4.33) due to Wei et al. (2019):

- (*Optimality*) By Wei et al. (2019), condition (4.33) cannot be further improved in the *worst* case in the sense that for every fixed pair (K_0, K), there exists some $\mu \in K$ violating (4.33) that invalidates (4.31). Furthermore, the same work also shows that the uniform $\|\cdot\|$-separation rate in (4.33) is minimax optimal in many cone testing problems.

- (*Non-uniform power*) On the other hand, it is important to mention that (4.31) is not equivalent to (4.33). In fact, as we will see in the example of testing 0 versus the orthant cone K_+ and the product circular cone $K_{\times,\alpha}$ (to be detailed in Corollary 2 and Theorem 9), the worst case condition (4.33) in terms of a separation in $\|\cdot\|$ is too conservative: condition (4.31) allows natural configurations of $\mu \in \{K_+, K_{\times,\alpha}\}$ whose separation rate in $\|\cdot\|$ can be n^δ for any $\delta \in (0, 1/4)$, while (4.33) necessarily requires a separation rate in $\|\cdot\|$ of order at least $n^{1/4}$. Therefore, although (4.33) gives the best possible inversion of (4.31) in terms of uniform separation in $\|\cdot\|$, condition (4.31) can be much weaker than (4.33), and characterizes the non-uniform power behavior of the LRT.

To give a better sense of the results in Theorem 7, we consider a toy example where K is also a subspace.

Proposition 18. *Let $\Psi \in \{\Psi_{\mathrm{os}}, \Psi_{\mathrm{ts}}\}$. Suppose $\delta_K - \delta_{K_0} \to \infty$.*

1. *If $\mu \in K_0$, the LRT is asymptotically size α with $\mathbb{E}_\mu \Psi(Y; m_0, \sigma_0) = \alpha + \mathcal{O}\big((\delta_K - \delta_{K_0})^{-1/2}\big)$.*

2. *For $\mu \in K$, the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi(Y; m_0, \sigma_0) \to 1$, if and only if $\|\mu - \Pi_{K_0}(\mu)\| \gg \big(\delta_K - \delta_{K_0}\big)^{1/4}$.*

Proof. (1) is a direct consequence of Theorem 7-(1). (2) follows from Theorem 7-(2) upon noting that

$$\begin{aligned}
&\Gamma_{K,2}(\mu - \Pi_{K_0}(\mu)) = \mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\|^2 - \mathbb{E}\|\Pi_K(\xi)\|^2 = \|\mu - \Pi_{K_0}(\mu)\|^2,\\
&\sigma_0^2 = \mathrm{Var}\big(\|\Pi_{K\cap K_0^*}(\xi)\|^2\big) = 2\delta_{K\cap K_0^*} = 2(\delta_K - \delta_{K_0}).
\end{aligned}$$

The second line of the above display uses Lemma 2-(3). □

More examples on testing parametric assumptions versus shape-constrained alternatives will be detailed in Section 4.4.

4.4 Examples

This section is organized as follows. Sections 4.4.1-4.4.4 study the generic testing problem (4.2) in the context of orthant/circular cones, isotonic regression, and Lasso, respectively. Section 4.4.5 specializes the subspace versus cone testing problem (4.9) to the setting of testing parametric assumptions versus shape-constrained alternatives. For simplicity of presentation, we will focus on the two-sided LRT (4.25), and simply call it the LRT unless otherwise specified.

4.4.1 Testing in orthant cone

Consider the orthant cone

$$K_+ \equiv \{\nu = (\nu_1, \ldots, \nu_n) \in \mathbb{R}^n : \nu_i \geq 0, i \in [1:n]\}.$$

We are interested in the testing problem (4.2) with $K = K_+$. Testing in the orthant cone has previously been studied by Kudô (1963); Raubertas et al. (1986); Wei et al. (2019). The following result (see Section C.2.1 for a proof) gives the limiting distribution of the LRS and characterizes the power behavior of the LRT in this example.

Theorem 8. *1. There exists a universal constant $C > 0$ such that for $\mu_0 \in K_+$,*

$$d_{\mathrm{TV}}\left(\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}}, \mathcal{N}(0,1)\right) \le \frac{C}{\sqrt{n}}.$$

Consequently the LRT is asymptotically size α with $\mathbb{E}_{\mu_0}\Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) = \alpha + \mathcal{O}(n^{-1/2})$.

2. For any $\mu \in K_+$, the LRT is power consistent under μ, i.e., $\mathbb{E}_{\mu}\Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 1$, if and only if

$$\left| \sum_{i=1}^{n} \left\{ \overline{S}_+(\mu_i) - \overline{S}_+((\mu_0)_i) \right\} + \|\mu - \mu_0\|^2 \right| \gg n^{1/2}.$$

Here, $\overline{S}_+$ is an increasing, concave and bounded function on $[0,\infty)$ with $\overline{S}_+(0) = 0$ and defined as

$$\overline{S}_+(x) \equiv \Phi(x) + x\varphi(x) - x^2(1 - \Phi(x)) - \frac{1}{2}, \quad x \ge 0. \tag{4.34}$$

Let us further investigate the special case $\mu_0 = 0$ to illustrate the non-uniform power behavior of the LRT mentioned after Theorem 7. In other words, we consider testing $\mu = 0$ versus the orthant cone K_+. Let

$$S_+(x) \equiv \overline{S}_+(x) + x^2 = \Phi(x) + x\varphi(x) + x^2\Phi(x) - \frac{1}{2}, \quad x \ge 0.$$

As $S_+'(x) = 2\big[\varphi(x) + x\Phi(x)\big]$, $S_+'(0) = 2\varphi(0) > 0$, and $S_+''(x) = 2\Phi(x) \ge 0$, S_+ is a strictly increasing and convex function on $[0,\infty)$ with $S_+(0) = 0$. Furthermore, it can be verified via direct calculation that uniformly over $x \ge 0$, $S_+(x) \asymp x \vee x^2$. Theorem 8 immediately yields the following corollary.

Corollary 2. *1. For $\mu = 0$, the LRT is asymptotically size α with $\mathbb{E}_0 \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) = \alpha + \mathcal{O}(n^{-1/2})$.*

2. For $\mu \in K_+$, the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) \to 1$, if and only if $\|\mu\|_1 \vee \|\mu\|^2 \gg n^{1/2}$.

The results in (Wei et al., 2019, Section 3.1.5), or equivalently, condition (4.33) show that the type II error of an optimally calibrated LRT vanishes uniformly for $\mu \in K_+$ such that $\|\mu\| \gg n^{1/4}$. Our results above indicate that the regime where the LRT has asymptotic power 1, for the orthant cone K_+, is actually characterized by the condition $\|\mu\|_1 \vee \|\mu\|^2 \gg n^{1/2}$ and is hence non-uniform with respect to $\|\cdot\|$. We give two concrete examples below.

Example 3. Let $q \in (0, 1/2)$ and $\tau_1, \tau_2 > 0$ be two fixed positive constants. Consider the following alternatives: (1) $\mu = (\tau_1 n^{-q})\mathbf{1}_n \in K_+$, and (2) $\mu = (\tau_2 i^{-q})_{i=1}^n \in K_+$. In both cases, $\|\mu\|_1 \asymp n^{1-q}$ and $\|\mu\|^2 \asymp n^{1-2q}$. The above corollary then yields that the LRT is power consistent under μ if and only if $q \in (0, 1/2)$, while the characterization of Wei et al. (2019) guarantees power consistency of the LRT only for $q \in (0, 1/4)$. In particular, as $q \to 1/2$, the LRT is power consistent for certain alternative μ with $\|\mu\| \asymp n^{\delta}$ for any $\delta > 0$. See Section 4.4.1 ahead for some simulation evidence.

One may further wonder whether the above examples only highlight 'exceptional' alternatives in the regime where the uniform separation in $\|\cdot\|$ fails to be informative, i.e., with $M_n \equiv \{\mu \in K_+ : \|\mu\|^2 \leq C n^{1/2}\}$ for some large enough absolute constant $C > 0$, whether the above examples only constitute a small fraction of M_n. To this end, let $A_n \equiv \{\mu \in M_n : \|\mu\|_1 \vee \|\mu\|^2 \geq C n^{1/2}\}$ be the region in M_n in which the LRT is indeed powerful. By a standard volumetric calculation, it is easy to see that $A_n / M_n \to 1$. In other words, the LRT is indeed powerful for 'most' alternatives in the region where the uniform separation in $\|\cdot\|$ is not informative as $n \to \infty$. Hence the non-uniform characteri-

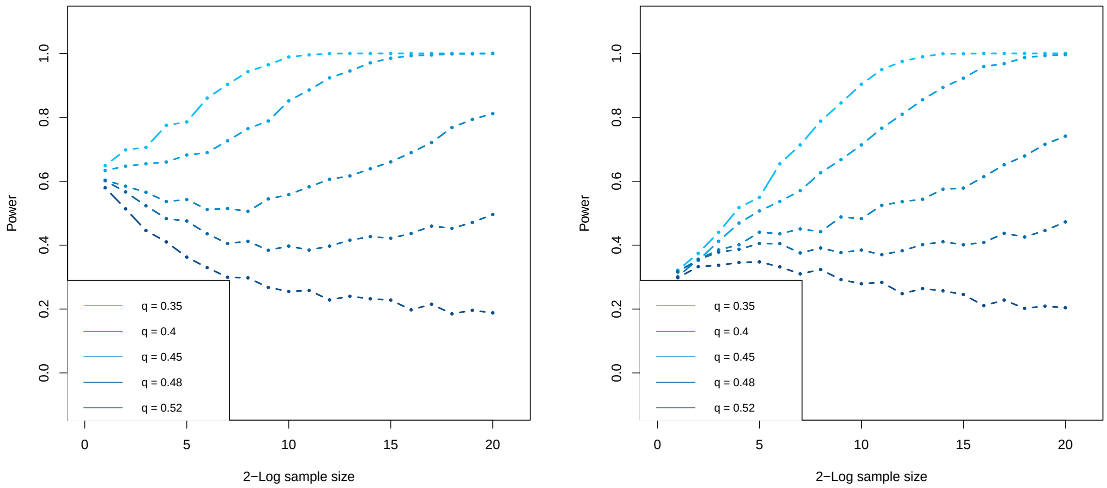

Figure 4.1: The power curves for the alternatives $\mu = (2n^{-q})_{i=1}^{n}$ (in the left panel) and $\mu = (i^{-q})_{i=1}^{n}$ (in the right panel) as q varies, for sample sizes $n \in \{2^{\ell}, \ell \in [1:20]\}$. The plots illustrate that the LRT has power in the range $q \in (0, 1/2)$ in both the examples.

zation in Corollary 2-(2) is essential for determining whether the LRT is powerful for a given alternative $\mu \in K_+$ in the regime $\|\mu\| = \mathcal{O}(n^{1/4})$.

As the separation rate $n^{1/4}$ in $\|\cdot\|$ is minimax optimal for testing 0 versus K_+ (cf. (Wei et al., 2019, Proposition 1)), the discussion above also illustrates the conservative nature of the minimax formulation in this testing problem.

An illustrative simulation study

Below we present simulation results under the two settings considered in Example 3. The confidence level will be taken as $\alpha = 0.05$. The power of the LRT in both the simulations below is calculated using an average of 2000 replications.

- In Figure 4.1, we take $\tau_1 = 2, \tau_2 = 1$ and examine the sharpness of the power char-

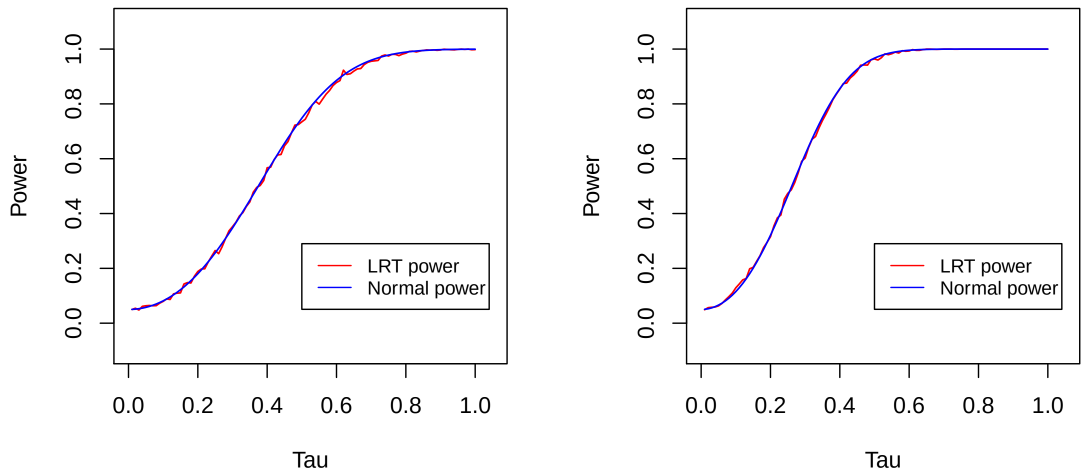

Figure 4.2: Fixed $q = 0.3$ and $n = 20000$. The alternatives are $\mu = (\tau_1 n^{-0.3})$ in left panel and $\mu = (\tau_2 i^{-0.3})$ in the right panel with $\tau_1, \tau_2 \in \{0.01, 0.02, \ldots, 1\}$. The red line denotes the power curve of LRT, i.e., $\{\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) : \tau_1\}$, while the blue line denotes the theoretical power curve via normal approximation, i.e., $\{\Delta_{\mathcal{A}_\alpha}\big(\Gamma_{K,2}(\mu)/\sigma_0\big) : \tau_2\}$.

acterization $q \in (0, 1/2)$ predicted by Corollary 2-(2). Clearly, Figure 4.1 shows that $q \in (0, 1/2)$ is the correct range where the LRT is powerful in both the settings of Example 3, rather than $q \in (0, 1/4)$ as predicted by Wei et al. (2019).

- In Figure 4.2, we fix $q = 0.3$, $n = 20000$, and examine the validity of the normal power expansion (4.28) in Theorem 7 along the alternatives considered in Example 3 with $\tau_1, \tau_2 \in \{0.01, 0.02, \ldots, 1\}$. Formally, we consider two power curves: (i) the power of the LRT, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0)$, (ii) theoretical power given by the normal approximation, i.e., $\Delta_{\mathcal{A}_\alpha}\big(\Gamma_{K,2}(\mu)/\sigma_0\big)$, for alternatives of the form $\mu = (\tau_1 n^{-0.3})_{i=1}^n$ and $\mu = (\tau_2 i^{-0.3})_{i=1}^n$ with the prescribed τ_1, τ_2's. Figure 4.2 clearly shows that the two power curves are very close to each other.

Counter-examples

Let $\mu_0 = \mathbf{1}_n \in K_+$, and $\mu = c\mathbf{1}_n$ for some fixed $c > 0$ to be determined. As long as $c \neq 1$, we have $\|\mu - \mu_0\|^2 = n(c-1)^2 \asymp n$. We also have $\sigma_\mu^2 = n \cdot \mathrm{Var}\big[(c+\xi_1-1)^2 - (c+\xi_1)_-^2\big] \equiv n\rho^2(c) \asymp n$, and

$$m_\mu - m_{\mu_0} = \|\mu - \mu_0\|^2 + \sum_{i=1}^n \big(\overline{S}_+(c) - \overline{S}_+(1)\big) = n\big\{(c-1)^2 + \overline{S}_+(c) - \overline{S}_+(1)\big\},$$

where $\overline{S}_+$ is defined in (4.34). Let $F(c) \equiv (c-1)^2 + \overline{S}_+(c) - \overline{S}_+(1)$. Then $F(1) = 0$, $F(0) = 0.5753...$, and $F'(1) = \overline{S}_+'(1) = 0.1666... > 0$.

We first present a choice of c that leads to an example showing the necessity of (4.19) for the power characterization (4.20).

Example 4. By the previous discussion, F must admit a zero in the open interval $(0, 1)$, which we denote as c_0. With $c = c_0$, we then have $m_\mu = m_{\mu_0}$. Moreover, as $\sigma_\mu^2 = n\rho^2(c_0) \neq n\rho^2(1) = \sigma_{\mu_0}^2$, so by Theorem 8-(1),

$$\frac{T(\mu+\xi) - m_{\mu_0}}{\sigma_{\mu_0}} = \frac{T(\mu+\xi) - m_\mu}{\sigma_\mu} \cdot \frac{\sigma_\mu}{\sigma_{\mu_0}} \to_d \mathcal{N}\bigg(0, \frac{\rho^2(c_0)}{\rho^2(1)}\bigg) \neq_d \mathcal{N}(0,1).$$

This means (4.20) fails.

Next we present a choice of c that leads to an example showing the necessity of considering two-sided LRT.

Example 5. By the previous discussion, $F(c) < 0$ for $c \in (0, 1)$ near 1. Pick any $c_1 \in (0, 1)$ such that $F(c_1) < 0$ and consider $c = c_1$. Let $\mu = c_1\mathbf{1}_n$. As $\sigma_{\mu_0} \asymp n^{1/2}$, $(m_\mu - m_{\mu_0})/\sigma_{\mu_0} \asymp -n^{1/2}$, so by Theorem 8-(1),

$$\frac{T(\mu+\xi) - m_{\mu_0}}{\sigma_{\mu_0}} = \frac{T(\mu+\xi) - m_\mu}{\sigma_\mu} \cdot \frac{\sigma_\mu}{\sigma_{\mu_0}} + \frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}} \to -\infty$$

in probability. This means that the two-sided LRT in (4.25) is powerful under $\mu = c_1\mathbf{1}_n$, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 1$, but the one-sided LRT in (4.24) is not powerful under μ, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{os}}(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 0$.

4.4.2 Testing in circular cone

For any $\alpha \in (0, \pi/2)$, let the α-circular cone be defined by

$$K_\alpha \equiv \left\{ \nu \in \mathbb{R}^{n-1} : \nu_1 \geq \|\nu\| \cos(\alpha) \right\},$$

and let $K_{\times,\alpha} \equiv K_\alpha \times \mathbb{R} \subset \mathbb{R}^n$. Consider the testing problem (4.2) with $\mu_0 = 0$ and $K \in \{K_\alpha, K_{\times,\alpha}\}$. The circular cone has recently been used in modeling by Besson (2006); Greco et al. (2008). The following result (see Section C.2.2 for a proof) gives the limiting distribution of the LRS and characterizes the power behavior of the LRT in this example.

Theorem 9. *1. Let $K \in \{K_\alpha, K_{\times,\alpha}\}$. There exists some universal constant $C > 0$ such that,*

$$d_{\mathrm{TV}}\left(\frac{T(Y) - m_0}{\sigma_0}, \mathcal{N}(0,1) \right) \leq \frac{C}{\sqrt{n}}.$$

Consequently the LRT is asymptotically size α with $\mathbb{E}_{\mu_0} \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) = \alpha + \mathcal{O}(n^{-1/2})$.

2. (a) For any $\mu \in K_\alpha$, the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) \to 1$, if and only if $\|\mu\| \gg 1$.

(b) For any $\mu \in K_{\times,\alpha}$, the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) \to 1$, if and only if $\|\mu^1\| \gg 1$ or $|\mu^2| \gg n^{1/4}$.

Here for any $\mu \in \mathbb{R}^n$, $\mu = (\mu^1, \mu^2) \in \mathbb{R}^{n-1} \times \mathbb{R}$ with $\mu^1 \in \mathbb{R}^{n-1}$ denoting the first $n-1$ components of μ and $\mu^2 \in \mathbb{R}$ denoting the last.

Regarding the two cones $\{K_\alpha, K_{\times,\alpha}\}$, Wei et al. (2019) showed the following:

- For K_α, an optimally calibrated LRT is powerful for $\mu \in K_\alpha$ such that $\|\mu\| \gg 1$. The minimax $\|\cdot\|$-separation rate is of the same constant order, so the LRT is minimax optimal.

- For $K_{\times,\alpha}$, an optimally calibrated LRT is powerful for $\mu \in K_{\times,\alpha}$ such that $\|\mu\| \gg n^{1/4}$, while the minimax $\|\cdot\|$-separation rate is of constant order, so the LRT is strictly minimax sub-optimal.

Theorem 9-(2) is rather interesting compared to the above results of Wei et al. (2019):

- For K_α, Theorem 9-(2)(a) shows that the power behavior of LRT is uniform with respect to $\|\cdot\|$ for K_α. In other words, for any $\mu \in K_\alpha$ with $\|\mu\| = \mathcal{O}(1)$ the LRT is necessarily not powerful.

- For $K_{\times,\alpha}$, Theorem 9-(2)(b) shows that the only bad alternatives that drive the uniform separation rate $n^{1/4}$ in $\|\cdot\|$ are those $\mu = (\mu^1, \mu^2) \in K_{\times,\alpha}$ lying in the narrow cylinder $\|\mu^1\| = \mathcal{O}(1)$ and $|\mu^2| = \mathcal{O}(n^{1/4})$, and the LRT will be powerful for points of the form, e.g. $(\mu^1, 0)$ as soon as $\|\mu^1\| \gg 1$. This is in line with the result of Theorem 9-(2)(a), and provides another example where the LRT exhibits non-uniform power behavior with respect to $\|\cdot\|$.

Similar to the LRT in the orthant cone, one may easily see that the conservative uniform separation rate (i.e., $\|\mu\| \gg n^{1/4}$) in $\|\cdot\|$ for $K_{\times,\alpha}$ fails to detect 'most' alternatives where the LRT is powerful, as $n \to \infty$. In this sense, the minimax sub-optimality of LRT for testing 0 versus $K_{\times,\alpha}$ is also conservative as the LRT behaves badly for only a few alternatives with large separation rate in $\|\cdot\|$.

The phenomenon observed above for the product circular cone can be easily extended as follows. For some positive integer m and generic closed convex cones $K_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, let $K_\times \equiv \times_{i=1}^m K_i \subset \mathbb{R}^{\sum_{i=1}^m n_i}$ be the associated product cone. Then the LRT for testing 0 versus $K_\times$ is power consistent under $\mu = (\mu^i)_{i=1}^m \in \times_{i=1}^m K_i = K$ if and only if

$$\frac{\sum_{i=1}^m \Gamma_{K_i,2}\big(\mu^i\big)}{\big(\sum_{i=1}^m \delta_{K_i}\big)^{1/2}} \to \infty.$$

The proof is largely similar to Theorem 9-(2)(b) so we omit the details.

4.4.3 Testing in isotonic regression

Let the monotone cone be defined by

$$K_\uparrow \equiv K_{\uparrow,0} = \{\nu = (\nu_1, \ldots, \nu_n) \in \mathbb{R}^n : \nu_1 \leq \ldots \leq \nu_n\}.$$

We consider the testing problem (4.2) with $K = K_\uparrow$ using the two-sided LRT (as in (4.25)). The following result (see Section C.2.3 for a proof) gives the limiting distribution of the LRS and characterizes the power behavior of the LRT in this example.

Theorem 10. *1. Suppose $\mu_0 \in K_\uparrow$, and for a universal constant $L > 1$,*

$$\frac{1}{L} \leq \min_{1\leq i\leq n-1} n\big((\mu_0)_{i+1} - (\mu_0)_i\big) \leq \max_{1\leq i\leq n-1} n\big((\mu_0)_{i+1} - (\mu_0)_i\big) \leq L. \tag{4.35}$$

Then

$$d_{\mathrm{TV}}\bigg(\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}}, \mathcal{N}(0,1)\bigg) \leq \frac{C}{n^{1/6}}.$$

Here $C > 0$ is a constant depending on L only. Consequently the LRT is asymptotically size α with $\mathbb{E}_{\mu_0}\Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) = \alpha + \mathcal{O}(n^{-1/6})$.

2. Let $\mu_0 = \big(f_0(i/n)\big)_{i=1}^n$ and $\mu = \big(f(i/n)\big)_{i=1}^n$, where $f, f_0 : [0,1] \to \mathbb{R}$ are C^2 monotone functions related by $f_0 = f + \rho_n\delta$ for some C^1 function $\delta : [0,1] \to \mathbb{R}$ with $\int \delta^2 = 1$ and δ' is bounded away from 0 and ∞. Suppose the first derivatives f', f_0' are bounded away from 0 and ∞ and second derivatives f'', f_0'' are bounded away from ∞. Then the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu\Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 1$, if and only if $\rho_n \gg n^{-5/12}$, if and only if $\|\mu - \mu_0\| \gg n^{1/12}$.

A few remarks are in order.

- (*Normal approximation*) The normal approximation in Theorem 10-(1) settles the problem of the limiting distribution for the LRT used in the simulation in (Durot and Tocquet, 2001, Section 4). There the LRT is compared to a goodness-of-fit test based on the central limit theorem (CLT) for the ℓ_1 estimation error of isotonic LSE (cf. Groeneboom (1985); Groeneboom et al. (1999); Durot (2007)). We note that condition (4.35) on the sequence μ_0 is equivalent to a bounded first derivative away from 0 and ∞ at the function level. This condition is commonly adopted in global CLTs for ℓ_p type losses of isotonic LSEs, cf. Durot (2007). In fact, the condition in Durot (2007) is stronger than (4.35) to guarantee a CLT for ℓ_p estimation error of the isotonic LSE.

- (*Rate of normal approximation*) We conjecture that the error rate $\mathcal{O}(n^{-1/6})$ in the above normal approximation is optimal based on the following heuristics. Writing $\widehat{\mu}$ as a shorthand for $\widehat{\mu}_{K_\uparrow}$, the LRS $T(Y)$ can be written, under H_0, as

$$
\begin{aligned}
T(Y) &= 2\langle \xi, \widehat{\mu} - \mu_0 \rangle - \|\widehat{\mu} - \mu_0\|^2 \\
&= \sum_{i=1}^{n} \Big(2\xi_i(\widehat{\mu}_i - (\mu_0)_i) - (\widehat{\mu}_i - (\mu_0)_i)^2 \Big).
\end{aligned}
$$

  Under the regularity condition (4.35), the isotonic LSE $\widehat{\mu}$ is localized in the sense that each $\widehat{\mu}_i$ roughly depends on μ_0 and ξ only via indices in a local neighborhood of i that contains $\mathcal{O}(n^{2/3})$ many points. So one may naturally view $T(Y)$ as roughly a summation of $\mathcal{O}(n^{1/3})$ 'independent' blocks, each of which roughly has variance of constant order. This naturally leads to the $\mathcal{O}(1/\sqrt{n^{1/3}}) = \mathcal{O}(n^{-1/6})$ rate in the Berry-Esseen bound of Theorem 10-(1). Our Theorem 10-(1) formalizes this intuition, but the proof is along a completely different line.

- (*Local power analysis*) The 'local alternative' setting in Theorem 10-(2) follows that of Durot and Tocquet (2001). In particular, the separation rate in Theorem 10-(2) is reminiscent of (Durot and Tocquet, 2001, Theorem 3.1). Durot and Tocquet (2001) obtained, under similar configurations and regularity conditions, a separation rate for a goodness-of-fit test based on the CLT for the ℓ_1 estimation error of the isotonic LSE of order $\rho_n \gg n^{-5/12} \vee n^{-1/2}\delta_n^{-1/2}$, where δ_n is the length of the support of the function δ. Our results here show that the LRT has a sharp separation rate $\rho_n \gg n^{-5/12}$ under the prescribed configuration, which is no worse than the one derived in Durot and Tocquet (2001) based on ℓ_1 estimation error.

In the isotonic regression example above, the main challenge in deriving the normal approximation for $T(Y)$ is to lower bound the quantity $\|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_{K_\uparrow}}\|_F^2$ in (4.15). We detail this intermediate result in the following proposition, which may be of independent interest (see Section C.2.3 for a proof).

Proposition 19. *Under the setting of Theorem 10-(1), there exists a small enough constant $\kappa > 0$, depending on L only, such that*

$$\left(\mathbb{E}_{\mu_0} J_{\widehat{\mu}_{K_\uparrow}}\right)_{ij} \geq \kappa n^{-2/3}$$

for $\{(i,j) : |i-j| \leq \kappa n^{2/3}, 0.1n \leq i,j \leq 0.9n\}$ for n large enough.

The above proposition is proved via exploiting the min-max representation of the isotonic LSE, a property not shared by general shape-constrained LSEs. We conjecture that results analogous to Theorem 10 hold for the general k-monotone cone $K_{\uparrow,k}$, to be formally defined in Section 4.4.5, but an analogue to Proposition 19 above is not yet available for general $K_{\uparrow,k}$.

4.4.4 Testing in Lasso

Consider the linear regression model

$$Y = \mu + \xi \equiv X\theta + \xi,$$

where $X \in \mathbb{R}^{n\times p}$ is a fixed design matrix with $p \le n$ and full column rank. Let $\Sigma \equiv X^\top X/n$ be the Gram matrix. Let $\widehat{\theta}^0 \equiv (X^\top X)^{-1}X^\top Y$ be the ordinary LSE, $\widehat{\theta} \equiv \widehat{\theta}(\lambda)$ be the constrained Lasso solution defined as

$$\widehat{\theta}(\lambda) \equiv \mathop{\mathrm{argmin}}_{\theta\in\mathbb{R}^p} \frac{1}{2}\|Y - X\theta\|^2 \qquad \text{s.t. } \|\theta\|_1 \le \lambda, \tag{4.36}$$

and $\widehat{\mu} \equiv \widehat{\mu}(\lambda) \equiv X\widehat{\theta}(\lambda)$. The setting here fits into our general framework by letting

$$K \equiv K_{X,\lambda} \equiv \{\mu = X\theta : \|\theta\|_1 \le \lambda\}$$

and $\widehat{\mu}_K \equiv \widehat{\mu}$. We are be interested in the testing problem (4.2), i.e., $H_0 : \mu = \mu_0$ versus $H_1 : \mu \in K_{X,\lambda}$, where $\mu_0 = X\theta_0 \in K_{X,\lambda}$ with $\|\theta_0\|_1 \le \lambda$. Such a goodness-of-fit test and the related problem of constructing confidence sets for the Lasso estimator has previously been studied in Verzelen and Villers (2010); Chatterjee and Lahiri (2011); Nickl and van de Geer (2013); Shah and Bühlmann (2018). In the following, we use the two-sided LRT $\Psi_{\mathrm{ts}}(Y; m_0, \sigma_0)$ (as in (4.25)) to test (4.2) and study its power characterization (see Section C.2.4 for a proof).

Theorem 11. *Suppose $p \to \infty$. For $\mu \in K_{X,\lambda}$, let*

$$\mathfrak{p}_{\lambda,\mu} \equiv \mathbb{P}_\mu\big(\|\widehat{\theta}^0\|_1 \ \ge \ \lambda\big).$$

1. *There exists a universal constant $C > 0$ such that, for $\mu_0 \in K_{X,\lambda}$,*

$$d_{\mathrm{TV}}\bigg(\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}}, \mathcal{N}(0,1)\bigg) \le \frac{C\sqrt{p + n\mathfrak{p}_{\lambda,\mu_0}^{1/2}}}{\big(p - C(n\mathfrak{p}_{\lambda,\mu_0})^2\big)_+}.$$

   Consequently the LRT is asymptotically size α with $\mathbb{E}_{\mu_0}\Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) = \alpha + \mathcal{O}(p^{-1/2})$, provided that $n\mathfrak{p}_{\lambda,\mu_0}^{1/2} = \mathfrak{o}(1)$.

2. *Suppose $n \cdot (\mathfrak{p}_{\lambda,\mu}^{1/2} \vee \mathfrak{p}_{\lambda,\mu_0}^{1/2}) = \mathfrak{o}(1)$. For any $\mu \in K_{X,\lambda}$, the LRT is power consistent, i.e., $\mathbb{E}_\mu\Psi_{\mathrm{ts}}(Y; m_{\mu_0}, \sigma_{\mu_0}) \to 1$, if and only if $\|\mu - \mu_0\| \gg p^{1/4}$.*

A few remarks are in order.

- (*Choice of* λ) To apply Theorem 11, we need to control the probability term $\mathfrak{p}_{\lambda,\mu}$ for a generic $\mu = X\theta \in K_{X,\lambda}$. This can be done via the following exponential inequality (see Lemma 48): for any $t \geq 1$,
$$\mathbb{P}_\mu\bigg(\|\widehat{\theta}^0\|_1 \geq \|\theta\|_1 + t\sqrt{\frac{p}{n\lambda_{\min}(\Sigma)}}\bigg) \leq e^{-t^2/C}.$$
Here $C > 0$ is a universal constant and $\lambda_{\min}(\Sigma)$ is the smallest eigenvalue of Σ. Therefore, for any $\lambda \geq \|\theta_0\|_1 + r_n$ with $r_n \equiv C\sqrt{p \log n / \big(n\lambda_{\min}(\Sigma)\big)}$ for a large enough constant $C > 0$, $n \cdot (\mathfrak{p}_{\lambda,\mu}^{1/2} \vee \mathfrak{p}_{\lambda,\mu_0}^{1/2}) = \mathfrak{o}(1)$ uniformly in $\mu \in K_{X,\lambda - r_n}$, and hence the LRT is asymptotically size α and power consistent for all such prescribed μ's if and only if $\|\mu - \mu_0\| \gg p^{1/4}$.

- (*Lasso in penalized form*) Theorem 11 is applicable for Lasso in its constrained form as defined in (4.36). We note here that this result does not yet extend to Lasso in the penalized form (cf. Tibshirani (1996); Chen et al. (2001)) as the equivalence between the two forms is random in nature.

The proof of Theorem 11 relies on the following proposition, which may be of independent interest (see Section C.2.4 for a proof).

Proposition 20. *The following hold:*

1. $\|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_{K_{X,\lambda}}}\|_F^2 \geq p/2 - 4\big(n\mathfrak{p}_{\lambda,\mu_0}\big)^2$.

2. *For any* $\mu \in K_{X,\lambda}$, $\big|\mathbb{E}_\mu \operatorname{div} \widehat{\mu}_{K_{X,\lambda}} - p\big| \leq 2p \cdot \mathfrak{p}_{\lambda,\mu}$.

3. *For any* $\mu \in K_{X,\lambda}$, $\big|\mathbb{E}_\mu \|\widehat{\mu}_{K_{X,\lambda}} - \mu\|^2 - p\big| \leq Cn\mathfrak{p}_{\lambda,\mu}^{1/2}$.

Here $C > 0$ *is an absolute constant.*

The proof of the above proposition makes essential use of an explicit representation of the Jacobian $J_{\widehat{\mu}_{K_{X,\lambda}}}$ derived in Kato (2009), which complements its analogues for Lasso in the penalized form derived in Zou et al. (2007); Tibshirani and Taylor (2012).

4.4.5 Testing parametric assumptions versus shape-constrained alternatives

For fixed $k \in \mathbb{Z}_{\geq 0}$ and $n \geq k+2$, and consider the testing problem

$$H_0 : \mu \in K_{0,k} \quad \text{versus} \quad H_1 : \mu \in K_{\uparrow,k}. \tag{4.37}$$

Here $K_{\uparrow,k} \equiv \{\mu \in \mathbb{R}^n : \nabla^{k+1}\mu \geq 0\}$ and $K_{0,k} \equiv \{\mu \in \mathbb{R}^n : \nabla^{k+1}\mu = 0\}$, with $\nabla : \mathbb{R}^n \to \mathbb{R}^{n-1}$ denoting the difference operator defined by $\nabla(\mu_i)_{i=1}^n \equiv (\mu_{i+1} - \mu_i)_{i=1}^{n-1}$, and $\nabla^{k+1} \equiv \nabla \circ \cdots \circ \nabla : \mathbb{R}^n \to \mathbb{R}^{n-k-1}$ with $k+1$ compositions. It can be readily verified that $K_{0,k}$ is a subspace of dimension $k+1$, $K_{\uparrow,k}$ is a closed and convex cone, and $K_{0,k} \subset K_{\uparrow,k} \subset \mathbb{R}^n$. Hence (4.37) is a special case of the general testing problem (4.9).

Testing a parametric model against a nonparametric alternative has previously been studied in Cox et al. (1988); Eubank and Spiegelman (1990); Azzalini and Bowman (1993); Härdle and Mammen (1993); Stute (1997); Fan and Huang (2001); Guerre and Lavergne (2005); Christensen and Sun (2010); Neumeyer and Van Keilegom (2010); Sen and Meyer (2017) among which the shape-constrained alternatives in (4.37) are sometimes preferred since the model fits therein usually do not involve the choice of tuning parameters. In particular:

1. When $k = 0$, (4.37) becomes:

$$H_0 : \mu \text{ is `constant'}, \qquad \text{versus} \qquad H_1 : \mu \text{ is `monotone'}.$$

2. When $k = 1$, (4.37) becomes:

$$H_0 : \mu \text{ is `linear'}, \qquad \text{versus} \qquad H_1 : \mu \text{ is `convex'}.$$

The above two settings have previously been considered in Bartholomew (1959a,b); Robertson et al. (1988); Sen and Meyer (2017).

Theorem 12. *Fix $k \in \mathbb{Z}_{\geq 0}$. Consider testing* (4.37) *using the two-sided LRT $\Psi_{\mathrm{ts}}(Y; m_0, \sigma_0)$, as in* (4.25).

1. *There exists a constant $C > 0$, depending on k only, such that for $\mu \in K_{0,k}$,*

$$d_{\mathrm{TV}}\left(\frac{T(Y) - m_0}{\sigma_0}, \mathcal{N}(0,1)\right) \leq \frac{C}{\mathbf{1}_{k=0}\sqrt{\log(en)} + \mathbf{1}_{k\geq 1}\sqrt{\log\log(16n)}}.$$

   Consequently for $\mu \in K_{0,k}$, the LRT is asymptotically size α with $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) = \alpha + \mathcal{O}\big(\mathbf{1}_{k=0}(\log(en))^{-1/2} + \mathbf{1}_{k\geq 1}(\log\log(16n))^{-1/2}\big)$.

2. *For $\mu \in K_{\uparrow,k}$ with $\|\mu - \Pi_{K_{0,k}}(\mu)\| \gg \log^{1/4}(en)$, the LRT is power consistent under μ, i.e., $\mathbb{E}_\mu \Psi_{\mathrm{ts}}(Y; m_0, \sigma_0) \to 1$.*

The key step in the proof of Theorem 12 (proved in Section C.2.5) is to obtain the correct order of the statistical dimension $\delta_{K_{\uparrow,k}}$. The discrepancy between $k = 0$ and $k \geq 1$ in claim (1) is due to the fact that while a universal upper bound of the order $\log(en)$ can be proved for any fixed $k \geq 0$, only a lower bound of the order $\log\log(16n)$ can be proved for $k \geq 1$. We conjecture that the correct order of $\delta_{K_{\uparrow,k}}$ should be $\log(en)$ for all fixed $k \geq 0$.

The above theorem can be easily extended to the multi-dimensional analogue of (4.37) in the context of, e.g., testing constancy versus coordinate-wise monotonicity, linearity versus multi-dimensional convexity, by using results of Han et al. (2019); Kur et al. (2020); we omit the details here.

Chapter 5

HIGH DIMENSIONAL TESTING OF COVARIANCE MATRICES

5.1 Introduction

5.1.1 Problem setup

Let $X_1, \ldots, X_n$ be i.i.d. samples from a p-variate normal distribution $\mathcal{N}_p(\mu, \Sigma)$. We are interested in the following general testing problem:

$$H_0 : (\mu, \Sigma) \in \mathcal{H}_0 \quad \text{versus} \quad H_1 : H_0 \text{ does not hold} \tag{5.1}$$

for certain classes $\mathcal{H}_0$ to be specified in later sections. For the most of the paper, we will focus on the marginal testing of the covariance matrix Σ.

Covariance matrix testing is a fundamental problem in multivariate statistical analysis. Departing from the classical low-dimensional setting where the dimension p is fixed, e.g. Anderson (1958); Muirhead (1982); Eaton (1983), the majority of recent works have been devoted to (5.1) in the high-dimensional setting where p is allowed to grow proportionally or even polynomially with n; see e.g., Ledoit and Wolf (2002); Birke and Dette (2005); Bai et al. (2009); Jiang et al. (2012); Cai and Ma (2013); Jiang and Yang (2013); Jiang and Qi (2015); Chen and Jiang (2018), for an incomplete list. To facilitate discussions below, let $X = [X_1, \ldots, X_n]^\top \in \mathbb{R}^{n \times p}$ be the data matrix, and let $T(X)$ be a generic test statistic whose distribution is invariant under H_0, i.e., the law of $T(X)$ remains the same for any $(\mu, \Sigma) \in \mathcal{H}_0$ in (5.1). Denote (throughout the paper we use the symbol $\equiv$ for definition)

$$m_{(\mu,\Sigma)} \equiv \mathbb{E}_{(\mu,\Sigma)} T(X), \quad \sigma^2_{(\mu,\Sigma)} \equiv \mathrm{Var}_{(\mu,\Sigma)}\big(T(X)\big) \tag{5.2}$$

for the mean and variance of $T(X)$ under $\mathcal{N}_p(\mu, \Sigma)$, respectively. We always assume that the two quantities in (5.2) are finite. In a similar spirit, we use the subscript (μ, Σ) in $\mathbb{E}_{(\mu,\Sigma)}$ and other probabilistic notations to indicate that the evaluation is under measure $\mathcal{N}_p(\mu, \Sigma)$. Due to the distributional invariance of $T(X)$, its mean and variance under the null

$$m_{H_0} \equiv m_{(\mu_0,\Sigma_0)}, \quad \sigma^2_{H_0} \equiv \sigma^2_{(\mu_0,\Sigma_0)} \tag{5.3}$$

are well-defined for any specification of $(\mu_0, \Sigma_0) \in \mathcal{H}_0$.

A common theme of the aforementioned works is a central limit theorem (CLT) for the normalized test statistic $T(X)$ under the null H_0: under the assumption $\min\{n, p\} \to \infty$ along with some other case-specific growth conditions on (n, p), it holds that

$$\frac{T(X) - m_{H_0}}{\sigma_{H_0}} \text{ converges in distribution to } \mathcal{N}(0, 1) \text{ under } H_0. \tag{5.4}$$

Hereafter $\mathcal{N}(0, 1)$ denotes the standard normal distribution.

The persistence of the universal CLT (5.4) in a wide class of covariance test statistics $T(X)$ in the high-dimensional regime, as cited above, is not a mere coincidence: it is known that Gaussian approximation holds when $T(X)$ depends on 'sufficient average' of eigenvalues of the sample covariance matrix, for instance when $T(X)$ can be written as its linear spectral statistic Bai and Silverstein (2004). From a statistical point of view, the validity of CLT (5.4) immediately leads to the construction of an asymptotically exact test: for any prescribed size $\alpha \in (0, 1)$,

$$\Psi(X) \equiv \Psi(X; m_{H_0}, \sigma_{H_0}) \equiv \mathbf{1}\Big(\frac{T(X) - m_{H_0}}{\sigma_{H_0}} > z_\alpha\Big). \tag{5.5}$$

Here z_α is the normal quantile such that $\mathbb{P}(\mathcal{N}(0, 1) > z_\alpha) = \alpha$. The quantities m_{H_0} and $\sigma^2_{H_0}$ are usually known in closed forms, at least asymptotically, in the above cited works to carry out the tests. Even not amenable to exact expression, these quantities can be simulated easily as well.

To assess the quality of a generic test $\Psi(X)$ and facilitate comparison between different tests, a more subtle and difficult question is to study the power behavior, ideally asymptotically exact, for each and every test. Although of fundamental importance, existing technical devices for asymptotically exact power analysis of covariance tests are rather limited. Roughly speaking, these techniques fall into the following two main categories:

1. Establish directly a central limit theorem for $T(X)$ under the alternative Wang and Yao (2013); Cai and Ma (2013); Chen and Jiang (2018); Jiang (2019). A number of case-specific techniques, e.g., random matrix theory Wang and Yao (2013), moment calculations Chen and Jiang (2018), martingale theory Cai and Ma (2013), have been used along this line for different tests.

2. Use contiguity theory in conjunction with Le Cam's third lemma. This program is carried in Onatski et al. (2013, 2014) in the spiked covariance alternative with a fixed number of spikes.

In addition to the case-specific nature of the techniques involved, a common downside of these methods lies in the imposition of rather restrictive conditions on both the growth of (n, p) and the alternative Σ under which the power analysis is valid. These restrictions may sometimes be more fundamental than technical. For instance, the method (2) works only for spiked alternatives in the sub-critical regime below the Baik-Ben Arous-Péché (BBP) phase transition Baik et al. (2005), as the log likelihood ratio process could become singular in the super-critical regime above the BBP phase transition. See Section 5.5 for a detailed technical comparison.

5.1.2 A new method of power analysis

In this paper, we develop a general method for analyzing the power behavior of a generic test statistic $T(X)$ when the CLT (5.4) under the null holds. For the test $\Psi(X)$ in (5.5) built from a generic test statistic $T(X)$ whose distribution remains invariant under H_0, the general power formula (see Theorem 13) takes the following form: for any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$, where $\mathcal{M}_p$ is the set of all $p \times p$ covariance matrices,

$$\left|\mathbb{E}_{(\mu,\Sigma)}\Psi(X) - \left[1 - \Phi\left(z_\alpha - \frac{m_{(\mu,\Sigma)} - m_{H_0}}{\sigma_{H_0}}\right)\right]\right| \leq \mathrm{err}_{H_0} + \mathrm{err}_{(\mu,\Sigma)}. \tag{5.6}$$

Here $m_{(\mu,\Sigma)}$ is defined in (5.2), err_{H_0} is the normal approximation error of $T(X)$ under H_0 in Kolmogorov distance [formally defined in (5.13) ahead], and

$$\mathrm{err}_{(\mu,\Sigma)} = \Big(\frac{V_{(\mu,\Sigma)}}{\max\big\{|m_{(\mu,\Sigma)} - m_{H_0}|, \sigma_{H_0}\big\}}\Big)^{2/3} \tag{5.7}$$

characterizes the departure of (μ, Σ) from the null (so $\mathrm{err}_{(\mu_0,\Sigma_0)} = 0$ for any $(\mu_0, \Sigma_0) \in \mathcal{H}_0$). The 'variance' parameter $V^2_{(\mu,\Sigma)}$, formally defined in (5.12) ahead, characterizes the order of stochastic fluctuation of the test statistic $T(X)$ under the alternative compared to that under the null.

From (5.6), it is clear that when the CLT (5.4) under the null holds, the power of $\Psi(X)$ under the alternative (μ, Σ) has an asymptotically exact expression via the parameter $\big(m_{(\mu,\Sigma)} - m_{H_0}\big)/\sigma_{H_0}$, provided that $\mathrm{err}_{(\mu,\Sigma)} \to 0$. Consequently, the key step in applying (5.6) rests in the validation of the condition $\mathrm{err}_{(\mu,\Sigma)} \to 0$. Informally, this condition is satisfied as long as the distribution of $T(X)$ 'stabilizes' under the prescribed alternative in an appropriate sense. More precisely, $\mathrm{err}_{(\mu,\Sigma)}$ vanishes as long as the order of stochastic fluctuation $V_{\mu,\Sigma}$ is smaller compared either to the mean difference $|m_{(\mu,\Sigma)} - m_{H_0}|$, or to the standard deviation σ_{H_0} of the test statistic $T(X)$ under the null. In typical applications that will be detailed below, the former case corresponds to the large departure regime of the alternative from the null, i.e., (μ, Σ) is sufficiently away from $\mathcal{H}_0$, while the latter to the small departure regime, so usually the validity of the power expansion (5.6) holds for the entire regime of alternatives.

From a different angle, our theory (5.6) is reminiscent of the classical Le Cam's contiguity theory in parametric LAN models. There if an estimator sequence is asymptotically normally distributed under the null, then it is again asymptotically normal under the alternative, but with a mean shift whose exact value depends on the magnitude of the local alternative, cf. van der Vaart (1998). Therefore the power of the corresponding test based on such an estimator sequence is determined completely by this mean shift parameter. Our theory (5.6) suggests a similar paradigm in the context of high-dimensional covariance testing (5.1), in that the power of a test statistic with a null CLT (5.4) is determined by the (normalized) mean shift parameter $\big(m_{(\mu,\Sigma)} - m_{H_0}\big)/\sigma_{H_0}$.

5.1.3 Two testing cases: testing identity and sphericity

To demonstrate the versatility of the general principle described above, we apply it to a number of test statistics in two benchmark special cases of (5.1).

The first application is the test for identity $\Sigma = I$. In the growing p setting, this problem has been extensively studied in the literature, see e.g., Ledoit and Wolf (2002); Srivastava (2005); Bai et al. (2009); Chen et al. (2010); Jiang et al. (2012); Cai and Ma (2013); Jiang and Yang (2013); Zheng et al. (2015); Chen and Jiang (2018). Among the tests studied in the above works, we apply our general theory (5.6) to the following three tests: Likelihood Ratio Test (LRT) (see Section 5.3.1), Ledoit-Nagao-Wolf's test Nagao (1973); Ledoit and Wolf (2002) (see Section 5.3.3), and Cai-Ma's test Cai and Ma (2013) (see Section 5.3.4). Compared to previous results where power analysis is either missing or requires restrictive conditions on the alternative, our results pose no assumptions on the alternative Σ and only mild conditions on the growth of (n, p). As an example, the LRT, denoted by $\Psi_{\mathrm{LRT}}(X)$, is shown to admit the following asymptotic power formula (see Theorem 15): under $\min\{n, p\} \to \infty$ with $\limsup(p/n) < 1$,

$$\mathbb{E}_{(\mu,\Sigma)}\Psi_{\mathrm{LRT}}(X) \sim 1 - \Phi\left(z_\alpha - \frac{d_S(\Sigma, I)}{\sqrt{2\big(-\frac{p}{n-1} - \log\big(1 - \frac{p}{n-1}\big)\big)}}\right). \tag{5.8}$$

Here $a \sim b$ stands for $a/b \to 1$ under the prescribed asymptotics.

To give a flavor of how (5.8) follows from our general theory (5.6), recall that the key step in applying (5.6) is to establish that $\mathrm{err}_{(\mu,\Sigma)} \to 0$. In the LRT setting, a much stronger estimate can be proved in that $\mathrm{err}_{(\mu,\Sigma)} \leq Cp^{-1/3}$ holds for some absolute constant $C > 0$. This key estimate follows by a series of algebraic manipulations, upon calculating that $V^2_{(\mu,\Sigma)} = (n-1)\|\Sigma - I\|_F^2$, $m_{(\mu,\Sigma)} - m_{H_0} = [(n-1)/2]d_S(\Sigma, I)$, and $\sigma^2_{H_0} \geq cp^2$ for some absolute constant $c > 0$[1]. See Proposition 21 and its proof for more details.

The second application is the sphericity test $\Sigma = \lambda I$ for some unspecified $\lambda > 0$. In the growing p setting, this problem has previously been studied in Ledoit and Wolf (2002); Srivastava (2005); Chen et al. (2010); Jiang et al. (2012); Jiang and Yang (2013); Jiang and Qi (2015). We study in this paper the following two widely-used tests: LRT for sphericity (Section 5.4.1), John's test John (1971) (Section 5.4.2), both invariant under H_0. Similar to the previous case, our results on the power behavior of these tests do not pose any assumption on the alternative Σ. As an example, the LRT for sphericity, denoted by $\Psi_{\mathrm{LRT},s}(X)$, is shown to admit the following asymptotic power formula (see Theorem 23): under $\min\{n, p\} \to \infty$ with $\limsup(p/n) < 1$,

$$\mathbb{E}_{(\mu,\Sigma)}\Psi_{\mathrm{LRT},s}(X) \sim 1 - \Phi\bigg(z_\alpha - \frac{-\log\det(\Sigma \cdot b^{-1}(\Sigma))}{\sqrt{2\big(-\frac{p}{n-1} - \log\big(1 - \frac{p}{n-1}\big)\big)}}\bigg). \tag{5.9}$$

Here $\det(\cdot)$ is the matrix determinant and $b(\Sigma) \equiv \mathrm{Tr}(\Sigma)/p$ with $\mathrm{Tr}(\cdot)$ denoting the trace. To the best of our knowledge, the above power formula for the LRT in the sphericity is new in the literature.

It should be mentioned that although we state (5.8)-(5.9) in asymptotic formulae for simplicity of representation in the introduction, these results actually hold with explicit non-asymptotic error bounds due to the intrinsic finite-sample nature of our theory (5.6). The non-asymptotic error bounds of the power expansion of all the aforementioned tests require

[1]Here $\|\cdot\|_F$ is the matrix Frobenius norm and $d_S(\cdot,\cdot)$ is the matrix Stein loss to be defined in (5.20) ahead.

quantitative normal approximation error bounds err_{H_0} for the corresponding test statistics, whose proofs depend on several spectral estimates for a class of special high dimensional matrices that will be detailed in Section D.1. These results, proved using techniques from (second-order) Poincaré inequalities, random matrices and zonal polynomials, are new and of independent interest (and sometimes even improve significantly known asymptotic results).

We conclude this introduction by noting that in contrast to (5.1) which targets at general alternatives, several previous works Onatski et al. (2013); Wang and Yao (2013); Onatski et al. (2014) obtained power expansions similar to (5.6) within a special class of alternatives known as the *spiked covariance model* Johnstone (2001). We draw detailed comparisons with these results in Section 5.5. In particular, as will be clear in Section 5.5, although Onatski et al. (2013); Wang and Yao (2013); Onatski et al. (2014) showed that some of the aforementioned tests have asymptotically equivalent power behavior under the spiked covariance alternative with a fixed number of spikes, our new power characterizations indicate that such equivalence in general fails when many spikes exist.

5.1.4 Organization

The rest of the paper is organized as follows. We detail the general principle described above in Section 5.2. Sections 5.3 and 5.4 are devoted to testing $\Sigma = I$ and $\Sigma = \lambda I$ respectively. Section 5.5 focuses on the case study of spike alternatives. Some concluding remarks are in Section 5.6 followed by some key spectral estimates in Section D.1. Sections D.2 - D.7 contain the main proofs of results in Sections 5.3 and 5.4, with the rest of technical details deferred to the appendices.

5.1.5 Notation

For any positive integer n, let $[n]$ denote the set $\{1,\ldots,n\}$. For $a,b\in\mathbb{R}$, $a\vee b\equiv\max\{a,b\}$ and $a\wedge b\equiv\min\{a,b\}$. For $a\in\mathbb{R}$, let $a_+\equiv a\vee 0$ and $a_-\equiv(-a)\vee 0$. For $x\in\mathbb{R}^n$, let $\|x\|_p = \|x\|_{\ell_p(\mathbb{R}^n)}$ denote its p-norm ($0 \leq p \leq \infty$) with $\|x\|_2$ abbreviated as $\|x\|$. Let $B_p(r;x) \equiv \{z \in \mathbb{R}^p : \|z - x\| \leq r\}$ be the unit ℓ_2 ball in $\mathbb{R}^p$. By $\mathbf{1}_n$ we denote the vector of all ones in $\mathbb{R}^n$. For a matrix $M \in \mathbb{R}^{n\times n}$, let $\|M\|_{\mathrm{op}}$ and $\|M\|_F$ denote the spectral and Frobenius norms of M respectively. We use $\{e_j\}$ to denote the canonical basis, whose dimension should be self-clear from the context.

We use C_x to denote a generic constant that depends only on x, whose numeric value may change from line to line unless otherwise specified. Notations $a \lesssim_x b$ and $a \gtrsim_x b$ mean $a \leq C_x b$ and $a \geq C_x b$ respectively, and $a \asymp_x b$ means $a \lesssim_x b$ and $a \gtrsim_x b$. The symbol $a \lesssim b$ means $a \leq Cb$ for some absolute constant C. For two nonnegative sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \ll b_n$ (respectively $a_n \gg b_n$) if $\lim_{n\to\infty}(a_n/b_n) = 0$ (respectively $\lim_{n\to\infty}(a_n/b_n) = \infty$). We write $a_n \sim b_n$ if $\lim_{n\to\infty}(a_n/b_n) = 1$. We follow the convention that $0/0 = 0$.

Let φ, Φ be the density and the cumulative distribution function of a standard normal random variable. For any $\alpha \in (0,1)$, let z_α be the normal quantile defined by $\mathbb{P}(\mathcal{N}(0,1) > z_\alpha) = \alpha$. For two random variables X, Y on $\mathbb{R}$, we use $d_{\mathrm{TV}}(X,Y)$ and $d_{\mathrm{Kol}}(X,Y)$ to denote their total variation distance and Kolmogorov distance defined respectively by

$$\begin{aligned} d_{\mathrm{TV}}(X,Y) &\equiv \sup_{B\in\mathcal{B}(\mathbb{R})} \big|\mathbb{P}\big(X \in B\big) - \mathbb{P}\big(Y \in B\big)\big|, \\ d_{\mathrm{Kol}}(X,Y) &\equiv \sup_{t\in\mathbb{R}} \big|\mathbb{P}\big(X \leq t\big) - \mathbb{P}\big(Y \leq t\big)\big|. \end{aligned} \tag{5.10}$$

Here $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra of $\mathbb{R}$.

Let γ_d be the standard Gaussian measure on $\mathbb{R}^d$, and for $r \geq 1$ let $W^{r,2}(\gamma_d)$ be the completion of $C_0^\infty(\mathbb{R}^d)$, the space of smooth and compactly supported functions in $\mathbb{R}^d$, with respect to the norm

$$\|f\|_r \equiv \bigg[\sum_{|\boldsymbol{\alpha}|\leq r} \int \big(\partial^{\boldsymbol{\alpha}} f(x)\big)^2 \gamma_d(\mathrm{d}x)\bigg]^{1/2}. \tag{5.11}$$

In other words, $W^{r,2}(\gamma_d)$ is the Sobolev space with respect to the Gaussian measure γ_d.

5.2 A general principle

Consider a generic test statistic $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ whose law is invariant under H_0, i.e., for any $(\mu, \Sigma) \in \mathcal{H}_0$, the law of $T(X)$ remains the same. For any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$, let $\mathcal{T}_{(\mu,\Sigma)} : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$ be defined by

$$\mathcal{T}_{(\mu,\Sigma)}(z) \equiv \nabla T\big(z\Sigma^{1/2} + \mathbf{1}_n\mu^\top\big)\Sigma^{1/2}, \quad z \in \mathbb{R}^{n\times p}.$$

Here $\mathbf{1}_n$ is the n-vector of all ones, and $\nabla T : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$ is the map with $\big(\nabla T(z)\big)_{ij} = \partial T(z)/\partial z_{ij}$. Recall that $Z_1, \ldots, Z_n$ are i.i.d. random variables with a standard p-variate normal distribution $\mathcal{N}(0, I_p)$. For any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$, define the quantity

$$V^2_{(\mu,\Sigma)} \equiv \inf_{(\mu_0,\Sigma_0)\in\mathcal{H}_0} \mathbb{E}\big\|\mathcal{T}_{(\mu,\Sigma)}(Z) - \mathcal{T}_{(\mu_0,\Sigma_0)}(Z)\big\|_F^2. \tag{5.12}$$

The infimum in the above definition is usually dummy as in many cases T itself is invariant over $\mathcal{H}_0$ in the sense that for any $(\mu_0, \Sigma_0), (\mu_1, \Sigma_1) \in \mathcal{H}_0$ and any $z \in \mathbb{R}^{n\times p}$, $T(z\Sigma_0^{1/2} + \mathbf{1}_n\mu_0^\top) = T(z\Sigma_1^{1/2} + \mathbf{1}_n\mu_1^\top)$, so $\mathcal{T}_{(\mu_0,\Sigma_0)}(z) = \mathcal{T}_{(\mu_1,\Sigma_1)}(z)$.

Equipped with the above definitions, the following result provides a general recipe of analyzing the behavior of the statistic $T(X)$ whenever normal approximation under the null is possible; its proof is presented later in this section. Recall the quantities $m_{(\mu,\Sigma)}, m_{H_0}, \sigma^2_{(\mu,\Sigma)}, \sigma^2_{H_0}$ defined in (5.2)-(5.3) and that $\gamma_{n\times p}$ denotes the standard Gaussian measure in $\mathbb{R}^{n\times p}$.

Theorem 13. *Suppose that $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ is an element of $W^{1,2}(\gamma_{n\times p})$, and the law of $T(X)$ is invariant under H_0. Then for any $(\mu, \Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$ and $t \in \mathbb{R}$,*

$$\begin{aligned}&\left|\mathbb{P}_{(\mu,\Sigma)}\bigg(\frac{T\big(X\big) - m_{H_0}}{\sigma_{H_0}} > t\bigg) - \mathbb{P}\bigg(\mathcal{N}\bigg(\frac{m_{(\mu,\Sigma)} - m_{H_0}}{\sigma_{H_0}}, 1\bigg) > t\bigg)\right| \\ &\leq \mathrm{err}_{H_0} + C \cdot \bigg(\frac{(1 + |t|)V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)} - m_{H_0}|} \bigwedge \frac{V_{(\mu,\Sigma)}}{\sigma_{H_0}}\bigg)^{2/3}.\end{aligned}$$

Here $C > 0$ is a universal constant, and

$$\mathrm{err}_{H_0} \equiv d_{\mathrm{Kol}}\bigg(\frac{T\big(X\big) - m_{H_0}}{\sigma_{H_0}}, \mathcal{N}(0, 1)\bigg) \quad \textit{under } H_0 \tag{5.13}$$

is the normal approximation error of $T(X)$ under H_0 in Kolmogorov distance as defined in (5.10).

Remark 18. *The comparison with normal distributions in the above theorem could be extended to more general distributions. We refrain from such extensions because all of the tests statistics considered in this paper (see Sections 5.3 and 5.4 ahead) have a normal limit under the null.*

Theorem 13 unifies and broadens substantially the scope of power analysis in the current covariance testing literature. In particular, it poses no apriori assumptions on the alternative and applies to both contiguous and non-contiguous one, while most of the current literature focus on the behavior of the statistic under contiguous alternatives (with only known exceptions in Chen and Jiang (2018) for the LRT).

Recall the generic test $\Psi(X)$ defined in (5.5). The following result is an immediate consequence of Theorem 13.

Corollary 3. *Suppose that $T : \mathbb{R}^{n\times p} \to \mathbb{R}$ is an element of $W^{1,2}(\gamma_{n\times p})$ and the law of $T(X)$ is invariant under H_0. For any $\alpha \in (0,1)$, there exists some $C_\alpha > 0$ such that*

$$\begin{aligned}\left|\mathbb{E}_{(\mu,\Sigma)}\Psi(X) - \left[1 - \Phi\left(z_\alpha - \frac{m_{(\mu,\Sigma)} - m_{H_0}}{\sigma_{H_0}}\right)\right]\right| \\ \leq \mathrm{err}_{H_0} + C_\alpha\left(\frac{V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)} - m_{H_0}| \vee \sigma_{H_0}}\right)^{2/3}\end{aligned}$$

holds for any $(\mu,\Sigma) \in \mathbb{R}^p \times \mathcal{M}_p$. Here err_{H_0} is defined in (5.13).

The above result reduces the analysis of the power behavior of $\Psi(X)$ in (5.5) into essentially the following two steps:

1. (*Normal approximation under H_0*) Show that

$$\mathrm{err}_{H_0} = d_{\mathrm{Kol}}\left(\frac{T(X) - m_{H_0}}{\sigma_{H_0}}\right) \to 0, \qquad \text{under } H_0.$$

Asymptotic normality has been derived for a variety of statistics in the high dimensional covariance testing literature, using mostly case-specific techniques; see e.g., Ledoit and Wolf (2002); Birke and Dette (2005); Bai et al. (2009); Jiang et al. (2012); Cai and Ma (2013); Jiang and Yang (2013); Jiang and Qi (2015); Chen and Jiang (2018) for an incomplete list. In this paper, we show $\mathrm{err}_{H_0} \to 0$ and recover many of the results cited above in a unified manner via Chatterjee's second-order Poincaré inequality Chatterjee (2009). This approach has two advantages: (i) it allows us to derive a rate of convergence of the normal approximation, which is conjectured to be optimal in many examples (see Sections 5.3 and 5.4 ahead for details); (ii) in a similar spirit to its application in the original work Chatterjee (2009), it only requires good enough estimates of the gradient and Hessian of $T(X)$, as opposed to the rather "hard" calculation techniques adopted in previous literature.

2. (*Ratio control*) Show that

$$\frac{V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)} - m_{H_0}| \vee \sigma_{H_0}} \to 0. \tag{5.14}$$

This requires upper bounds on $V_{(\mu,\Sigma)}$ and lower bounds for $|m_{(\mu,\Sigma)} - m_{H_0}|$ and σ_{H_0}. The general strategy for these bounds are: (i) the term $V_{(\mu,\Sigma)}$ defined in (5.12) can be evaluated by computing the gradient of T, (ii) the null variance σ_{H_0} has asymptotically exact formula for many test statistics in the literature, and can sometimes be directly evaluated via Fourier expansion in the Gaussian space; (iii) the mean difference term $|m_{(\mu,\Sigma)} - m_{H_0}|$ requires a near closed-form formula for the mean of T under the alternative. As will be seen in the examples in Sections 5.3 and 5.4, evaluation of these quantities typically requires certain case-specific techniques.

The remainder of this section is devoted to the proof of Theorem 13, which utilizes the following lemma.

Lemma 4. *For any $t \in \mathbb{R}$ and $u \in \mathbb{R}$,*

$$\Big|\mathbb{P}\big(\mathcal{N}(u,1) \le t\big) - \mathbb{P}\big(\mathcal{N}((1+\eta)u,1) \le t\big)\Big| \le 2(1+|t|)\cdot|\eta|.$$

Proof. This result strengthens (Han et al., 2020, Lemma 5.4). We assume without loss generality $\eta \in [-1/2, 1/2]$ because otherwise the right hand side of the desired display is greater than or equal to 1. Note that the left hand side is bounded by

$$\left|\int_{t-u}^{t-(1+\eta)u} \varphi(z)\,\mathrm{d}z\right| \le |\eta|\cdot\left[\sup_{v\in[(t-u)-|\eta u|,(t-u)+|\eta u|]} \varphi(v)|u|\right] \equiv |\eta|\cdot M_t(u).$$

Here $\varphi(\cdot)$ is the normal density. First consider $u \ge 0$. Then $M_t(u) \le \sup_{v\in[t-3u/2,t-u/2]} \varphi(v)u$.

- If $t - u/2 \le 0$, then

$$\begin{aligned} M_t(u) &\le \varphi\Big(t-\frac{u}{2}\Big)u = \varphi\Big(t-\frac{u}{2}\Big)(u-2t) + 2t\varphi\Big(t-\frac{u}{2}\Big) \\ &\le 2\sup_{x\in\mathbb{R}}|x|\varphi(x) + \frac{2}{\sqrt{2\pi}}|t| = \frac{2}{\sqrt{2\pi e}} + \frac{2}{\sqrt{2\pi}}|t|. \end{aligned}$$

  Here we used the readily verified fact that $\sup_{x\in\mathbb{R}}|x|\varphi(x) = 1/\sqrt{2\pi e}$.

- If $t - 3u/2 \ge 0$, then

$$\begin{aligned} M_t(u) &\le \varphi\Big(t-\frac{3u}{2}\Big)u = \varphi\Big(t-\frac{3u}{2}\Big)\Big(u-\frac{2t}{3}\Big) + \frac{2}{3}t\varphi\Big(t-\frac{3u}{2}\Big) \\ &\le \frac{2}{3}\Big(\frac{1}{\sqrt{e}} + \frac{1}{\sqrt{2\pi}}|t|\Big). \end{aligned}$$

- Otherwise $(2/3)t \le u \le 2t$, so $M_t(u) \le |u| \le 2|t|$.

The case $u < 0$ can be handled similarly, so we have $\sup_u M_t(u) \le 2(1+|t|)$. □

Proof of Theorem 13. Let $Z \in \mathbb{R}^{n\times p}$ be a matrix generated by n i.i.d. samples from $\mathcal{N}(0, I_p)$. Let $X^{(\mu,\Sigma)} \equiv Z\Sigma^{1/2} + \mathbf{1}_n\mu^\top$. Without loss of generality, let $(\mu_0, \Sigma_0) \in \mathcal{H}_0$ be the pair achieving the infimum in (5.12). Then,

$$\begin{aligned}\frac{T(X)-m_{H_0}}{\sigma_{H_0}} &\stackrel{d}{=} \frac{T\big(X^{(\mu,\Sigma)}\big)-T\big(X^{(\mu_0,\Sigma_0)}\big)}{\sigma_{H_0}}+\frac{T\big(X^{(\mu_0,\Sigma_0)}\big)-m_{H_0}}{\sigma_{H_0}}\\ &= \frac{m_{(\mu,\Sigma)}-m_{H_0}}{\sigma_{H_0}}+\frac{W(Z)}{\sigma_{H_0}}+\frac{T\big(X^{(\mu_0,\Sigma_0)}\big)-m_{H_0}}{\sigma_{H_0}}. \end{aligned} \tag{5.15}$$

Here $W(Z)$ is the centered variable defined by

$$W(Z)\equiv T\big(Z\Sigma^{1/2}+\mathbf{1}_n\mu^\top\big)-T\big(Z\Sigma_0^{1/2}+\mathbf{1}_n\mu_0^\top\big)-\big(m_{(\mu,\Sigma)}-m_{H_0}\big).$$

Using the chain rule,

$$\begin{aligned}&\partial_{(ij)}W(Z)\\ &=\sum_{(i'j')}\left[\frac{\partial}{\partial X^{(\mu,\Sigma)}_{(i'j')}}T\big(X^{(\mu,\Sigma)}\big)\cdot\frac{\partial X^{(\mu,\Sigma)}_{(i'j')}}{\partial Z_{(ij)}}-\frac{\partial}{\partial X^{(\mu_0,\Sigma_0)}_{(i'j')}}T\big(X^{(\mu_0,\Sigma_0)}\big)\cdot\frac{\partial X^{(\mu_0,\Sigma_0)}_{(i'j')}}{\partial Z_{(ij)}}\right]\\ &=\sum_{j'}\left[\big(\nabla T(X^{(\mu,\Sigma)})\big)_{ij'}\cdot(\Sigma^{1/2})_{j'j}-T(X^{(\mu_0,\Sigma_0)})\big)_{ij'}\cdot(\Sigma_0^{1/2})_{j'j}\right]\\ &=\Big(\nabla T\big(X^{(\mu,\Sigma)}\big)\Sigma^{1/2}-\nabla T\big(X^{(\mu_0,\Sigma_0)}\big)\Sigma_0^{1/2}\Big)_{ij}\\ &=\big(\mathfrak{T}_{(\mu,\Sigma)}(Z)-\mathfrak{T}_{(\mu_0,\Sigma_0)}(Z)\big)_{ij}.\end{aligned}$$

By the Gaussian-Poincaré inequality (Boucheron et al., 2013, Theorem 3.20),

$$\mathrm{Var}\big(W(Z)\big)\le\mathbb{E}\bigg[\sum_{(ij)}\big(\partial_{(ij)}W(Z)\big)^2\bigg]=\mathbb{E}\big\|\mathfrak{T}_{(\mu,\Sigma)}(Z)-\mathfrak{T}_{(\mu_0,\Sigma_0)}(Z)\big\|_F^2=V^2_{(\mu,\Sigma)}.$$

This means for any $u>0$, on an event E with probability at least $1-u^{-2}$,

$$\big|W(Z)\big|\le u\cdot V_{(\mu,\Sigma)}.$$

Hence for any $t\in\mathbb{R}$, the decomposition (5.15) entails that [recall the definition of err_{H_0} in

(5.13)]

$$
\begin{aligned}
&\mathbb{P}\bigg(\frac{T\big(X^{(\mu,\Sigma)}\big)-m_{H_0}}{\sigma_{H_0}}>t\bigg)\\
=\ &\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}+W(Z)}{\sigma_{H_0}}+\frac{T\big(X^{(\mu_0,\Sigma_0)}\big)-m_{H_0}}{\sigma_{H_0}}>t\bigg)\\
\le\ &\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}+u\cdot V_{(\mu,\Sigma)}}{\sigma_{H_0}}+\frac{T\big(X^{(\mu_0,\Sigma_0)}\big)-m_{H_0}}{\sigma_{H_0}}>t\bigg)+\frac{1}{u^2}\\
\le\ &\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}+u\cdot V_{(\mu,\Sigma)}}{\sigma_{H_0}}+\mathcal{N}(0,1)>t\bigg)+\frac{1}{u^2}+\mathrm{err}_{H_0}\equiv\mathfrak{p}(u)+\mathrm{err}_{H_0}.
\end{aligned}
$$

Next we bound $\mathfrak{p}(\cdot)$ using two different ways. First by Lemma 45, we have

$$
\begin{aligned}
\inf_{u>0}\mathfrak{p}(u)&\le\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}}{\sigma_{H_0}}+\mathcal{N}(0,1)>t\bigg)\\
&\qquad+\inf_{u>0}\bigg[2(1+|t|)u\cdot\frac{V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)}-m_{H_0}|}+\frac{1}{u^2}\bigg]\\
&\le\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}}{\sigma_{H_0}}+\mathcal{N}(0,1)>t\bigg)+C\bigg(\frac{(1+|t|)V_{(\mu,\Sigma)}}{|m_{(\mu,\Sigma)}-m_{H_0}|}\bigg)^{2/3}.
\end{aligned}
$$

On the other hand, by anti-concentration of the standard normal distribution, i.e., $|\mathbb{P}\big(\mathcal{N}(0,1)\le a\big)-\mathbb{P}\big(\mathcal{N}(0,1)\le b\big)|\le|a-b|$ for any $a,b\in\mathbb{R}$,

$$
\begin{aligned}
\inf_{u>0}\mathfrak{p}(u)&\le\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}}{\sigma_{H_0}}+\mathcal{N}(0,1)>t\bigg)+\inf_{u>0}\bigg[\frac{u\cdot V_{(\mu,\Sigma)}}{\sigma_{H_0}}+\frac{1}{u^2}\bigg]\\
&\le\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}}{\sigma_{H_0}}+\mathcal{N}(0,1)>t\bigg)+C\bigg(\frac{V_{(\mu,\Sigma)}}{\sigma_{H_0}}\bigg)^{2/3}.
\end{aligned}
$$

Collecting the bounds completes the proof for one direction. For the other direction, we have

$$
\begin{aligned}
&\mathbb{P}\bigg(\frac{T\big(X^{(\mu,\Sigma)}\big)-m_{H_0}}{\sigma_{H_0}}>t\bigg)\\
=\ &\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}+W(Z)}{\sigma_{H_0}}+\frac{T\big(X^{(\mu_0,\Sigma_0)}\big)-m_{H_0}}{\sigma_{H_0}}>t\bigg)\\
\ge\ &\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}-u\cdot V_{(\mu,\Sigma)}}{\sigma_{H_0}}+\frac{T\big(X^{(\mu_0,\Sigma_0)}\big)-m_{H_0}}{\sigma_{H_0}}>t\bigg)-\frac{1}{u^2}\\
\ge\ &\mathbb{P}\bigg(\frac{m_{(\mu,\Sigma)}-m_{H_0}-u\cdot V_{(\mu,\Sigma)}}{\sigma_{H_0}}+\mathcal{N}(0,1)>t\bigg)-\frac{1}{u^2}-\mathrm{err}_{H_0}.
\end{aligned}
$$

The rest of the proof follows from similar arguments as in the previous direction by invoking the two different bounds. □

5.3 *Testing identity* $\Sigma = I$

We introduce some additional notation. Based on i.i.d. samples $X_1, \ldots, X_n$ from $\mathcal{N}(\mu, \Sigma)$, the sample covariance matrix and its unbiased modification are given by

$$
\begin{aligned}
S_* &\equiv n^{-1} \sum_{k=1}^{n} \big(X_k - \overline{X}\big)\big(X_k - \overline{X}\big)^\top \quad \text{with} \quad \overline{X} \equiv n^{-1} \sum_{i=1}^{n} X_i, \\
S &\equiv \frac{n}{N} S_* \overset{d}{=} \frac{1}{N} \sum_{k=1}^{N} (X_k - \mu)^\top (X_k - \mu).
\end{aligned} \tag{5.16}
$$

Here

$$
N = n - 1 \tag{5.17}
$$

and the equal in distribution in (5.16) follows from (Muirhead, 1982, Theorem 3.1.2). Throughout the rest of the paper, we will mainly work with S for mathematical simplicity (unless otherwise specified), and adopt the right most expression of (5.16) as its definition whenever no confusion could arise.

5.3.1 LRT

Consider the testing problem:

$$
H_0 : \Sigma = I \quad \text{versus} \quad H_1 : H_0 \text{ does not hold.} \tag{5.18}
$$

This is a special case of (5.1) by taking $\mathcal{H}_0 = \mathbb{R}^p \times \{I\}$, and has been extensively studied in the literature; see Ledoit and Wolf (2002); Srivastava (2005); Bai et al. (2009); Chen et al. (2010); Jiang et al. (2012); Cai and Ma (2013); Jiang and Yang (2013); Zheng et al. (2015); Chen and Jiang (2018) for an incomplete list.

This subsection studies the behavior of the LRT for testing (5.18). The modified log-likelihood ratio statistic $T_{\mathrm{LRT}} : \mathbb{R}^{N\times p} \to \mathbb{R}$ (cf. (Muirhead, 1982, Theorem 8.4.2)) is defined as

$$T_{\mathrm{LRT}}(X) \equiv \frac{N}{2}\big[\mathrm{Tr}(S) - \log\det S - p\big]. \tag{5.19}$$

Trivially, the law of $T_{\mathrm{LRT}}(X)$ is invariant under H_0. The general principle in Theorem 13 thereby applies in view of the regularity of T_{LRT} (see Appendix D.9). We will use the set of notation $\big(m_{\Sigma;\mathrm{LRT}}, \sigma_{\Sigma;\mathrm{LRT}}, V_{\Sigma;\mathrm{LRT}}\big)$ to represent their generic versions defined in (5.2) and (5.12).

Following the discussion after Corollary 3, we start by establishing a quantitative CLT for $T_{\mathrm{LRT}}(X)$ under H_0; its proof is presented in Section D.2.2.

Theorem 14. *Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some constant $C = C(\varepsilon) > 0$, such that under H_0,*

$$d_{\mathrm{TV}}\bigg(\frac{T_{\mathrm{LRT}}(X) - m_{I;\mathrm{LRT}}}{\sigma_{I;\mathrm{LRT}}}, \mathcal{N}(0,1)\bigg) \le \frac{C}{p}.$$

Below we make some comments on Theorem 14:

- (*(n,p)-condition*) It is clear from the definition of $T_{\mathrm{LRT}}(X)$ that if $p \ge n$ then S is singular and the log-likelihood ratio statistic $T_{\mathrm{LRT}} \equiv -\infty$ is degenerate. The CLT for the log-likelihood ratio statistic $T_{\mathrm{LRT}}(X)$ under H_0 was first derived in Bai et al. (2009) using random matrix theory under the assumption that $p/n \to y$ for some $y \in (0,1)$. This result was then improved in Jiang et al. (2012) and Chen and Jiang (2018) to hold under the condition $n > p+1$ and $p \to \infty$, and in Zheng et al. (2015) to relax the Gaussian assumption. The condition $p/N \le 1-\varepsilon$ in Theorem 14 is used to derive the stable estimate $\mathbb{E}\|S^{-1}\|_{\mathrm{op}} \le C$ for some constant $C = C(\varepsilon) > 0$; see Lemma 50 for details.

- (*Rate of normal approximation*) As introduced in Section 5.2, Theorem 14 and other CLT's to follow are proved via Chatterjee's second-order Poincaré inequality Chatterjee (2009), which provides the rate p^{-1} of normal approximation. As for fixed p, $2T_{\mathrm{LRT}}(X)$ converges weakly under H_0 to a chi-squared distribution with $p(p+1)/2$ degrees of freedom (cf. (Muirhead, 1982, Theorem 8.4.9)), we conjecture that the rate p^{-1} cannot be further improved.

The following result establishes the ratio control (5.14) for the log-likelihood ratio statistic $T_{\mathrm{LRT}}(X)$; its proof is presented in Section D.2.3. For p.s.d. Σ_1 and p.d. Σ_2, let

$$d_S(\Sigma_1, \Sigma_2) \equiv \mathrm{Tr}(\Sigma_1\Sigma_2^{-1}) - \log\det\left(\Sigma_1\Sigma_2^{-1}\right) - p \tag{5.20}$$

be the Stein loss with the convention that $d_S(\Sigma_1, \Sigma_2) \equiv \infty$ if Σ_1 is singular.

Proposition 21. *Suppose Σ is non-singular. The following hold:*

1. $V^2_{\Sigma;\mathrm{LRT}} = N\|\Sigma - I\|_F^2$.

2. $m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}} = (N/2)d_S(\Sigma, I)$.

3. *In the asymptotic regime $N \geq p+1$ with $p \to \infty$,*
$$\sigma^2_{I;\mathrm{LRT}} \sim \frac{N^2}{2}\left[-\frac{p}{N} - \log\left(1 - \frac{p}{N}\right)\right].$$
In particular, $\sigma^2_{I;\mathrm{LRT}} \geq cp^2$ for some universal constant $c > 0$.

4. *There exists some universal constant $C > 0$ such that*
$$\frac{V_{\Sigma;\mathrm{LRT}}}{|m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}}| \vee \sigma_{I;\mathrm{LRT}}} \leq \frac{C}{p^{1/2}}.$$

The above proposition gives a prototypical example of how to proceed with the ratio control (5.14). For the log-likelihood ratio statistic $T_{\mathrm{LRT}}(X)$ defined in (5.19), both $V_{\Sigma;\mathrm{LRT}}$

and the mean difference $m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}}$ admit easy-to-handle closed-form formulae. To give some insights for the bound obtained in Proposition 21-(4), let us consider the 'local regime' of alternatives in which $d_S(\Sigma, I) \approx \|\Sigma - I\|_F^2$. Then (5.14) can be bounded, up to a constant, by

$$\frac{\sqrt{N\|\Sigma - I\|_F^2}}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LRT}}} \leq \sup_{x\geq 0} \frac{x}{x^2 \vee \sigma_{I;\mathrm{LRT}}} = \frac{1}{\inf_{x\geq 0}\big(x \vee \frac{\sigma_{I;\mathrm{LRT}}}{x}\big)} = \frac{1}{\sigma_{I;\mathrm{LRT}}^{1/2}},$$

in the prescribed local regime of alternatives. The above simple reasoning exemplifies the essential reason why the ratio (5.14) must be small: if Σ is sufficiently away from I, then the mean difference $m_{\Sigma;\mathrm{LRT}} - m_{I;\mathrm{LRT}}$ is substantially larger than $V_{\Sigma;\mathrm{LRT}}$, but would otherwise be compensated by the diverging nature of $\sigma_{I;\mathrm{LRT}}$.

Let $\Psi_{\mathrm{LRT}}(X)$ be the LRT built from the generic test (5.1) and the log-likelihood ratio statistic $T_{\mathrm{LRT}}(X)$. Combining the above results with the generic Theorem 13, we obtain the following asymptotic formula for the power behavior of $\Psi_{\mathrm{LRT}}(X)$. Recall that z_α is the normal quantile defined under (5.5).

Theorem 15. *Suppose $p/N \leq 1 - \varepsilon$ for some $\varepsilon \in (0, 1)$. Then there exists some constant $C = C(\varepsilon, \alpha) > 0$ such that*

$$\bigg|\mathbb{E}_\Sigma \Psi_{\mathrm{LRT}}(X) - \mathbb{P}\bigg(\mathcal{N}\bigg(\frac{N \cdot d_S(\Sigma, I)}{2\sigma_I}, 1\bigg) > z_\alpha\bigg)\bigg| \leq C \cdot p^{-1/3}. \tag{5.21}$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\limsup(p/N) < 1$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{LRT}}(X) \sim 1 - \Phi\Bigg(z_\alpha - \frac{d_S(\Sigma, I)}{\sqrt{2\big(-\frac{p}{N} - \log\big(1 - \frac{p}{N}\big)\big)}}\Bigg).$$

Remark 19. *Note that the right hand side of the above asymptotic expression is bounded below by $1 - \Phi(z_\alpha) = \alpha > 0$, hence (5.21) along with $p \to \infty$ suffice for the above asymptotic equivalence to hold.*

Below we make some comments on Theorem 15:

- (*Case of singular* Σ) Rigorously speaking, Theorem 13 only holds for the case where m_Σ is finite, which excludes the case of singular Σ. However, in the latter case, there exists some $a \in \mathbb{R}^p$ such that $a^\top X_1 \equiv 0$ a.s., and hence S is necessarily singular as well, thus rendering the test $\Psi_{\mathrm{LRT}}(X)$ always rejecting the null.

- (*Power comparison*) To the best of our knowledge, Chen and Jiang (2018) is the only work that contains a formal theory on the power behavior of the LRT $\Psi_{\mathrm{LRT}}(X)$ targeting at general alternatives. A related but different LRT was considered in Onatski et al. (2013, 2014) which was targeted at the special class of spike alternatives; see more details in Section 5.5. Comparing to (Chen and Jiang, 2018, Theorem 1), Theorem 15 removes their condition $\sup_n \|\Sigma\|_{\mathrm{op}} < \infty$ and applies to arbitrary alternatives Σ.

- (*Minimax optimality*) It was established in (Cai and Ma, 2013, Theorem 1) that when $\limsup(p/n) < \infty$, the minimax rate of testing (5.18) is $\sqrt{p/n}$ under the $\|\cdot\|_F$ norm. Using the relation [recall (5.20)] $d_S(\Sigma, I) = \sum_{j=1}^p (\lambda_j - 1 - \log \lambda_j) \gtrsim \sum_{j=1}^p (\lambda_j - 1)^2 \wedge |\lambda_j - 1|$ with $\{\lambda_j\}_{j=1}^p \geq 0$ being the eigenvalues of Σ, it is easy to deduce from Theorem 15 that over the class of non-singular Σ, the LRT $\Psi_{\mathrm{LRT}}(X)$ is minimax rate optimal in the asymptotic regime $N \wedge p \to \infty$ with $\limsup(p/N) < 1$; see Sections 5.3.4 and 5.5 for more refined comparison.

5.3.2 LRT: simultaneous testing of mean and covariance

Consider the following variant of (5.18):

$$H_0 : \mu = 0, \Sigma = I \quad \text{versus} \quad H_1 : H_0 \text{ does not hold.} \tag{5.22}$$

This is a special case of (5.1) by taking $\mathcal{H}_0 = \{0\} \times \{I\}$, and has previously been studied in Jiang and Yang (2013); Jiang and Qi (2015); Chen and Jiang (2018).

We will study the behavior of the LRT for (5.22). Using the vanilla version S_* of the sample covariance [recall (5.16)], the log-likelihood ratio statistic takes the form (cf. (Muirhead, 1982, Theorem 8.5.1))

$$T_{(\mu,\Sigma);\mathrm{LRT}}(X) \equiv \frac{n}{2}\big(\mathrm{Tr}(S_*) - \log\det S_* - p + \overline{X}^\top\overline{X}\big). \tag{5.23}$$

We will use $\big(m_{(\mu,\Sigma);\mathrm{LRT}}, \sigma_{(\mu,\Sigma);\mathrm{LRT}}, V_{(\mu,\Sigma);\mathrm{LRT}}\big)$ to represent their generic versions defined in (5.2) and (5.12).

The next theorem establishes a quantitative CLT for $T_{(\mu,\Sigma);\mathrm{LRT}}(X)$; its proof is given in Section D.3.2.

Theorem 16. *Suppose $p/n \le 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some constant $C = C(\varepsilon) > 0$, such that under H_0,*

$$d_{\mathrm{TV}}\left(\frac{T_{(\mu,\Sigma);\mathrm{LRT}}(X) - m_{(0,I);\mathrm{LRT}}}{\sigma_{(0,I);\mathrm{LRT}}}, \mathcal{N}(0,1)\right) \le \frac{C}{p}.$$

The following result establishes the ratio control in (5.14) for $T_{(\mu,\Sigma);\mathrm{LRT}}(X)$; its proof is presented in Section D.3.3. Recall that $d_S(\Sigma_1,\Sigma_2)$ is the Stein loss defined in (5.20).

Proposition 22. *Suppose Σ is non-singular. The following hold:*

1. $V^2_{(\mu,\Sigma);\mathrm{LRT}} = n\big(\|\Sigma - I\|_F^2 + \mu^\top\Sigma\mu\big)$.

2. $m_{(\mu,\Sigma);\mathrm{LRT}} - m_{(0,I);\mathrm{LRT}} = (n/2)\big(d_S(\Sigma, I) + \|\mu\|^2\big)$.

3. *In the asymptotic regime $n \ge p+2$ with $p \to \infty$,*

$$\sigma^2_{(0,I);\mathrm{LRT}} \sim \frac{n^2}{2}\left[-\frac{p}{n-1} - \log\left(1 - \frac{p}{n-1}\right)\right].$$

   In particular, $\sigma^2_{(0,I);\mathrm{LRT}} \ge cp^2$ for some universal constant $c > 0$.

4. *There exists some universal constant $C > 0$ such that*

$$\frac{V_{(\mu,\Sigma);\mathrm{LRT}}}{|m_{(\mu,\Sigma);\mathrm{LRT}} - m_{(0,I)}| \vee \sigma_{(0,I);\mathrm{LRT}}} \le \frac{C}{p^{1/2}}.$$

The above proposition is similar to Proposition 21. It should however be mentioned that the neat closed-form formula in (2) is available due to the fact we use the vanilla version S_* in (5.23).

Let $\Psi_{\mathrm{LRT};m}(X)$ be the LRT built from the generic test (5.1) and the log-likelihood ratio statistic $T_{\mathrm{LRT};m}$. Combining the above results with the generic Theorem 13, we obtain the following asymptotic power formula for $\Psi_{\mathrm{LRT};m}(X)$.

Theorem 17. *Suppose $p/n \leq 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some constant $C = C(\varepsilon, \alpha) > 0$ such that*

$$\left|\mathbb{E}_{(\mu,\Sigma)}\Psi_{\mathrm{LRT};m}(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{n\cdot\big(d_S(\Sigma,I)+\|\mu\|^2\big)}{2\sigma_{(0,I)}},1\right) > z_\alpha\right)\right| \leq C\cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $n \wedge p \to \infty$ with $\limsup(p/n) < 1$,

$$\mathbb{E}_{(\mu,\Sigma)}\Psi_{\mathrm{LRT};m}(X) \sim 1 - \Phi\left(z_\alpha - \frac{d_S(\Sigma,I)+\|\mu\|^2}{\sqrt{2\big(-\frac{p}{n-1} - \log\big(1-\frac{p}{n-1}\big)\big)}}\right).$$

The law of $T_{(\mu,\Sigma);\mathrm{LRT}}(X)$ under the alternative is previously derived in (Chen and Jiang, 2018, Theorem 2) under several regularity conditions on (μ,Σ). We remove those conditions completely in Theorem 17.

5.3.3 Ledoit-Nagao-Wolf's test

This subsection studies testing (5.18) using the (rescaled) modified Nagao's trace statistic Nagao (1973) by Ledoit and Wolf Ledoit and Wolf (2002):

$$T_{\mathrm{LNW}}(X) \equiv \frac{N}{4}\left[\mathrm{Tr}\big(S-I\big)^2 - \frac{1}{N}\mathrm{Tr}^2(S)\right]. \tag{5.24}$$

An asymptotically equivalent statistic as an unbiased estimator of $\|\Sigma - I\|_F^2$ has also been studied in Srivastava (2005). One advantage of using (5.24) is that it applies to the case $p > n$ where the LRT in Section 5.3.1 becomes degenerate.

We will use $\big(m_{\Sigma;\mathrm{LNW}}, \sigma_{\Sigma;\mathrm{LNW}}, V_{\Sigma;\mathrm{LNW}}\big)$ to represent their generic versions defined in (5.2) and (5.12). The next theorem establishes a quantitative CLT for $T_{\mathrm{LNW}}(X)$ under H_0; its proof is presented in Section D.4.2.

Theorem 18. *There exists an absolute constant $C > 0$ such that under H_0,*

$$d_{\mathrm{TV}}\bigg(\frac{T_{\mathrm{LNW}}(X) - m_{I;\mathrm{LNW}}}{\sigma_{I;\mathrm{LNW}}}, \mathcal{N}(0,1)\bigg) \leq \frac{C}{N \wedge p}.$$

The CLT for $T_{\mathrm{LNW}}(X)$ was first derived in (Ledoit and Wolf, 2002, Proposition 7) under the condition that $p/N \to y \in (0, \infty)$, which was later improved in (Birke and Dette, 2005, Theorem 3.6) to include the case $y \in \{0, \infty\}$. Here we give explicit error bounds in the normal approximation.

The following result establishes the ratio control (5.14) for T_{LNW}; its proof is presented in Section D.4.3.

Proposition 23. *Suppose $p/N \leq M$ for some $M > 0$. Then the following hold:*

1. $V^2_{\Sigma;\mathrm{LNW}} \leq C_1 N\big(\|\Sigma\|^2_{\mathrm{op}} \vee 1\big)\|\Sigma - I\|_F^2$ *for some constant $C_1 = C_1(M) > 0$.*

2. *With $Q_{\mathrm{LNW}}(\Sigma) \equiv (N^{-1} - 2N^{-2})\,\mathrm{Tr}(\Sigma^2 - I)$,*

$$m_{\Sigma;\mathrm{LNW}} - m_{(0,I)} = \frac{N}{4}\big[\|\Sigma - I\|_F^2 + Q_{\mathrm{LNW}}(\Sigma)\big].$$

3. *In the asymptotic regime $N \wedge p \to \infty$,*

$$\sigma^2_{I;\mathrm{LNW}} \sim \frac{p^2}{4}.$$

4. *There exists some constant $C_2 = C_2(M) > 0$ such that*

$$\frac{V_{\Sigma;\mathrm{LNW}}}{|m_{\Sigma;\mathrm{LNW}} - m_{I;\mathrm{LNW}}| \vee \sigma_{I;\mathrm{LNW}}} \leq \frac{C_2}{p^{1/2}}.$$

There are two significant structural differences in the above proposition compared to Proposition 21. First, compared to $V_{\Sigma;\mathrm{LRT}}$, $V_{\Sigma;\mathrm{LNW}}$ comes with an additional multiplicative factor $\|\Sigma\|_{\mathrm{op}}^2 \vee 1$. Second, although a closed-form formula is available for $m_{\Sigma;\mathrm{LNW}}$, a somewhat undesirable 'residual term' $Q_{\mathrm{LNW}}(\Sigma)$ exists. Removing the effect of these terms in the ratio control (4) requires additional technicalities that will be detailed in Section D.4.3.

We note that the variance formula for $\sigma^2_{I;\mathrm{LNW}}$ in the above proposition is particularly easy to derive from scratch due to the polynomial structure of $T_{\mathrm{LNW}}(X)$ in (5.24). This result will also be useful in the variance formula for John's test to be studied in Section 5.4.2.

Let $\Psi_{\mathrm{LNW}}(X)$ be the test built from (5.1) and the statistic in (5.24). Combining the above results with Theorem 13 and some additional efforts to remove the residual term $Q_{\mathrm{LNW}}(\Sigma)$ in the mean difference formula (2) in the above proposition, we have the following asymptotic power formula for $\Psi_{\mathrm{LNW}}(X)$; see Section D.4.4 for its proof.

Theorem 19. *Suppose $p/N \le M$ for some $M > 0$. Then there exists some constant $C = C(\alpha, M) > 0$ such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{LNW}}(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \|\Sigma - I\|_F^2}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right)\right| \le C \cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\limsup(p/N) < \infty$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{LNW}}(X) \sim 1 - \Phi\left(z_\alpha - \frac{\|\Sigma - I\|_F^2}{2(p/N)}\right).$$

The asymptotic behavior of T_{LNW} under the alternative is previously only known in (Srivastava, 2005, Theorem 4.1) under rather restrictive conditions on both Σ and growth of p. Theorem 19 only requires p/N to be bounded and makes no assumptions on Σ.

5.3.4 Cai-Ma's test

In this subsection, we assume additionally that the data matrix X is centered, i.e., $X_1, \ldots, X_n$ are i.i.d. from $\mathcal{N}(0, \Sigma)$. Consider testing (5.18) using the (rescaled) U-statistic by Cai and

Ma Cai and Ma (2013):

$$T_{\mathrm{CM}}(X) \equiv \frac{1}{(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j). \tag{5.25}$$

Here $h : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is the symmetric kernel defined by

$$h(X_1, X_2) \equiv (X_1^\top X_2)^2 - (X_1^\top X_1 + X_2^\top X_2) + p.$$

Some direct calculation yields the mean and variance of $T_{\mathrm{CM}}(X)$ (cf. (Cai and Ma, 2013, Equation (8), (9))):

$$\begin{aligned} m_{\Sigma;\mathrm{CM}} &= \frac{n}{2}\|\Sigma - I\|_F^2, \\ \sigma^2_{\Sigma;\mathrm{CM}} &= \frac{n}{n-1}\big[\mathrm{Tr}^2(\Sigma^2) + \mathrm{Tr}(\Sigma^4)\big] + 2n\mathrm{Tr}\big(\Sigma^2(\Sigma - I)^2\big). \end{aligned} \tag{5.26}$$

A closely related statistic built from a higher order U-statistic was first studied in Chen et al. (2010). Similar to (5.24) in the previous subsection, (5.25) is applicable in the case $p > n$.

We will use $\big(m_{\Sigma;\mathrm{CM}}, \sigma_{\Sigma;\mathrm{CM}}, V_{\Sigma;\mathrm{CM}}\big)$ to represent their generic versions defined in (5.2) and (5.12). The next theorem establishes a quantitative CLT for $T_{\mathrm{CM}}(X)$ under H_0; its proof is given in Section D.5.2.

Theorem 20. *There exists some absolute constant $C > 0$ such that under H_0,*

$$d_{\mathrm{TV}}\bigg(\frac{T_{\mathrm{CM}}(X)}{\sigma_{I;\mathrm{CM}}}, \mathcal{N}(0,1)\bigg) \le C \cdot \bigg(\frac{\log n}{n} \bigvee \frac{1}{p}\bigg).$$

The above theorem improves (Cai and Ma, 2013, Proposition 3) under the null in two directions: (1) the above Berry-Esseen bound holds in the stronger total variation distance, and (2) the normal approximation rate is improved from $(n \wedge p)^{-1/5}$ to $(n \wedge p)^{-1}$ (ignoring the logarithmic factor).

The following result establishes the ratio control (5.14) for T_{CM}; its proof is given in Section D.5.3.

Proposition 24. *Let $y = p/n$. Then the following hold:*

1. *There exists some absolute constant $C_1 > 0$ such that*

$$V_{\Sigma;\mathrm{CM}}^2 \le C_1\big[(1 \vee y^3) + (y^2 \vee y^3)\log^2 n\big] \cdot n\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.$$

2. $m_{\Sigma;\mathrm{CM}} - m_{I;\mathrm{CM}} = (n/2)\|\Sigma - I\|_F^2$.

3. $\sigma_{I;\mathrm{CM}}^2 = (p^2 + p) \cdot n/(n-1)$*, so in the asymptotic regime* $n \wedge p \to \infty$,

$$\sigma_{I;\mathrm{CM}}^2 \sim p^2.$$

4. *There exists some absolute constant $C_2 > 0$ such that*

$$\frac{V_{\Sigma;\mathrm{CM}}}{|m_{\Sigma;\mathrm{CM}} - m_{I;\mathrm{CM}}| \vee \sigma_{I;\mathrm{CM}}} \le \frac{C_2\big[(1 \vee y)^{3/2} + y(1 \vee y)^{1/2}\log n\big]}{(n \wedge p)^{1/2}}.$$

Remark 20. *We keep the ratio term $y = p/n$ in the above result (in particular (4)) to cope with the additional* $\log n$ *factor, so we may remove lower bound conditions for p.*

Let $\Psi_{\mathrm{CM}}(X)$ be the test built from (5.1) and the statistic in (5.25). Combining the above results with the generic Theorem 13, we obtain the following asymptotic power formula for $\Psi_{\mathrm{CM}}(X)$; some details are provided in Section D.5.4.

Theorem 21. *Suppose $p/n \le M$ for some $M > 0$. Then there exists some constant $C = C(\alpha, M) > 0$ such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{CM}}(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{n\|\Sigma - I\|_F^2}{2\sigma_{I;\mathrm{CM}}}, 1\right) > z_\alpha\right)\right| \le C\left(\frac{\log^{2/3} n}{n^{1/3}} \bigvee \frac{1}{p^{1/3}}\right).$$

Consequently, in the asymptotic regime $n \wedge p \to \infty$ with $\limsup(p/n) < \infty$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{CM}}(X) \sim 1 - \Phi\left(z_\alpha - \frac{\|\Sigma - I\|_F^2}{2(p/n)}\right).$$

(Cai and Ma, 2013, Equation (27)) proved that the exact power expansion above holds uniformly for all alternatives Σ such that $b\sqrt{p/n} \le \|\Sigma - I\|_F \le B\sqrt{p/n}$ for some fixed

$0 < b \leq B < \infty$. Their proof is based on a Berry-Esseen bound (cf. (Cai and Ma, 2013, Proposition 3)) for the CLT of $T_{\mathrm{CM}}(X)$ under general alternatives using the test-specific martingale difference representation. Theorem 21 holds uniformly for all possible alternatives Σ when additionally $p/n \leq M$, using the general method of Theorem 13.

5.4 *Testing sphericity* $\Sigma = \lambda I$

5.4.1 Likelihood ratio test

Consider the testing problem:

$$H_0 : \Sigma = \lambda I \quad \text{versus} \quad H_1 : H_0 \text{ does not hold} \tag{5.27}$$

for some un-specified $\lambda > 0$. This is a special case of (5.1) by taking $\mathcal{H}_0 = \mathbb{R}^p \times \{\lambda I : \lambda > 0\}$, and has been extensively studied previously in Ledoit and Wolf (2002); Srivastava (2005); Chen et al. (2010); Jiang et al. (2012); Jiang and Yang (2013).

This subsection studies the LRT for (5.27). The (re-scaled) log-likelihood ratio statistic for (5.27) is defined by (cf. (Muirhead, 1982, Theorem 8.3.2)):

$$T_{\mathrm{LRT},s}(X) \equiv \frac{N}{2}\big(p \log \mathrm{Tr}(S) - \log\det S - p\log p\big). \tag{5.28}$$

Evidently, the law of $T_{\mathrm{LRT},s}(X)$ does not depend on the λ in (5.27) and hence is invariant under H_0. Thus the general principle in Theorem 13 applies due to regularity of $T_{\mathrm{LRT};s}$ (see Appendix D.9). We will use $\big(m_{\Sigma;\mathrm{LRT},s}, \sigma_{\Sigma;\mathrm{LRT},s}, V_{\Sigma;\mathrm{LRT},s}\big)$ to represent their generic versions defined in (5.2) and (5.12).

For a symmetric $p \times p$ matrix M, let

$$b_\ell(M) \equiv p^{-1}\mathrm{Tr}(M^\ell), \quad b(M) \equiv b_1(M). \tag{5.29}$$

The next theorem establishes a quantitative CLT for $T_{\mathrm{LRT},s}(X)$; its proof is presented in Section D.6.2. Recall that $T_{\mathrm{LRT},s}$ is only non-degenerate if $p \leq n - 1 = N$.

Theorem 22. *Suppose $p/N \leq 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Then there exists some $C = C(\varepsilon) > 0$ such that under H_0,*

$$d_{\mathrm{TV}}\bigg(\frac{T_{\mathrm{LRT},s}(X) - m_{I;\mathrm{LRT},s}}{\sigma_{I;\mathrm{LRT},s}}, \mathcal{N}(0,1)\bigg) \leq \frac{C}{p}.$$

The CLT for $T_{\mathrm{LRT},s}(X)$ was previously derived in (Jiang and Yang, 2013, Theorem 1) under the asymptotics $y \in (0,1]$. The quantitative CLT above does not require p to grow proportionally to N but excludes the boundary case $y = 1$.

The following result establishes the ratio control (5.14) for $T_{\mathrm{LRT},s}$; see Section D.6.3 for its proof.

Proposition 25. *Suppose Σ is non-singular. The following hold:*

1. *There exists some absolute constant $C_1 > 0$ such that*

$$V^2_{\Sigma;\mathrm{LRT},s} \leq C_1 N \|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2$$

   holds for N, p large enough.

2. *The mean difference is given by*

$$m_{\Sigma;\mathrm{LRT},s} - m_{I;\mathrm{LRT},s} = \frac{N}{2}\big[-\log\det(\Sigma \cdot b^{-1}(\Sigma)) + Q_{\mathrm{LRT},s}(\Sigma \cdot b^{-1}(\Sigma))\big].$$

   Here

$$\big|Q_{\mathrm{LRT},s}\big(\Sigma \cdot b^{-1}(\Sigma)\big)\big| \leq C_2 N^{-1} b\big[\big(\Sigma \cdot b^{-1}(\Sigma)\big)^2\big] \tag{5.30}$$

   for some absolute constant $C_2 > 0$.

3. *In the asymptotic regime $N \wedge p \to \infty$ with $\limsup(p/N) < 1$,*

$$\sigma^2_{I;\mathrm{LRT},s} \sim \frac{N^2}{2}\bigg[-\frac{p}{N} - \log\bigg(1 - \frac{p}{N}\bigg)\bigg].$$

4. There exists some absolute constant $C_3 > 0$ such that

$$\frac{V_{\Sigma;\mathrm{LRT},s}}{|m_{\Sigma;\mathrm{LRT},s} - m_{I;\mathrm{LRT},s}| \vee \sigma_{I;\mathrm{LRT},s}} \leq \frac{C_3}{(\sigma_{I;\mathrm{LRT},s} \wedge N)^{1/2}}.$$

There is a genuine difference between the above ratio control result and the previous ones studied in Section 5.3, in that a closed-form formula for the mean difference $m_{\Sigma;\mathrm{LRT},s} - m_{I;\mathrm{LRT},s}$ is no longer available. One therefore has to work with strong enough upper bounds for the 'residual term' $Q_{\mathrm{LRT},s}(\Sigma \cdot b^{-1}(\Sigma))$, the removal of which constitutes the main technicalities in the proofs; see Section D.6.3 for details.

Let $\Psi_{\mathrm{LRT},s}(X)$ be the test built from (5.1) and the statistic in (5.28). Combining the above results with Theorem 13 and some additional efforts to remove the residual term $Q_{\mathrm{LRT},s}(\Sigma \cdot b^{-1}(\Sigma))$, we have the following asymptotic power formula for $\Psi_{\mathrm{LRT},s}(X)$; see Section D.6.4 for its proof.

Theorem 23. *Suppose $p/N \leq 1 - \varepsilon$ for some $\varepsilon \in (0, 1)$. Then there exists some constant $C = C(\varepsilon, \alpha) > 0$ such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{LRT},s}(X) - \mathbb{P}\bigg(\mathcal{N}\bigg(- \frac{N \log\det\big(\Sigma \cdot b^{-1}(\Sigma)\big)}{2\sigma_{I;s}}, 1\bigg) > z_\alpha\bigg)\right| \leq C \cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\limsup(p/N) < 1$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{LRT},s}(X) \sim 1 - \Phi\Bigg(z_\alpha - \frac{-\log\det\big(\Sigma \cdot b^{-1}(\Sigma)\big)}{\sqrt{2\big(-\frac{p}{N} - \log\big(1 - \frac{p}{N}\big)\big)}}\Bigg).$$

To the best of our knowledge, in the high dimensional regime $N \wedge p \to \infty$, the LRT for (5.27) was only studied in Jiang and Yang (2013); Jiang and Qi (2015), where formal theory was missing on the power behavior of $\Psi_{\mathrm{LRT},s}$. Theorem 23 fills this gap.

5.4.2 John's test

Consider testing (5.27) using the (rescaled) John's trace statistic John (1971):

$$T_{\mathrm{J}}(X) \equiv \frac{N}{4}\mathrm{Tr}\bigg[\bigg(\frac{S}{p^{-1}\mathrm{Tr}(S)} - I\bigg)^2\bigg]. \tag{5.31}$$

Clearly the law of $T_J(X)$ is invariant under H_0, and the above statistic is non-degenerate for all configurations of (n, p). The general principle in Theorem 13 thereby applies in view of the regularity of T_{J} (see Appendix D.9). We will use $\big(m_{\Sigma;\mathrm{J}}, \sigma_{\Sigma;\mathrm{J}}, V_{\Sigma;\mathrm{J}}\big)$ to represent their generic versions defined in (5.2) and (5.12).

The next theorem establishes a quantitative CLT for $T_{\mathrm{J}}(X)$ under H_0; its proof is given in Section D.7.2.

Theorem 24. *There exists some absolute constant $C > 0$, such that under H_0,*

$$d_{\mathrm{TV}}\bigg(\frac{T_{\mathrm{J}}(X) - m_{I;\mathrm{J}}}{\sigma_{I;\mathrm{J}}}, \mathcal{N}(0,1)\bigg) \leq \frac{C}{N \wedge p}.$$

Central limit theorems for $T_{\mathrm{J}}(X)$ under H_0 in high dimensions are first obtained in Ledoit and Wolf (2002). We improve these results both in terms of non-asymptotic normal approximation bound and the removal of the condition $0 < \liminf(p/N) \leq \limsup(p/N) < \infty$.

The following result establishes the ratio control (5.14) for T_{J}; its proof is presented in Section D.7.3. Recall the definition of $b(\Sigma)$ in (5.29).

Proposition 26. *Suppose $p/N \leq M$ for some $M > 1$. Then the following hold for N larger than a big enough absolute constant:*

1. *There exists some constant $C_1 = C_1(M) > 0$ such that*

$$V_{\Sigma;\mathrm{J}}^2 \leq C_1 \cdot N\big(\|\Sigma \cdot b^{-1}(\Sigma)\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2.$$

2. *The mean difference is given by*

$$m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}} = \frac{N}{4}\big[\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2 + Q_{\mathrm{J}}\big(\Sigma \cdot b^{-1}(\Sigma)\big)\big].$$

   Here

$$\big|Q_{\mathrm{J}}\big(\Sigma \cdot b^{-1}(\Sigma)\big)\big| \leq C_2 \cdot N^{-1/2}\big(p^{-1}\|\Sigma \cdot b^{-1}(\Sigma)\|_F^2 + 1\big)\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F$$

   for some $C_2 = C_2(M) > 0$.

3. In the asymptotic regime $N \wedge p \to \infty$,

$$\sigma^2_{I;\mathrm{J}} \sim \frac{p^2}{4}.$$

4. There exists some $C_3 = C_3(M) > 0$ such that

$$\frac{V_{\Sigma;\mathrm{J}}}{|m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}}| \vee \sigma_{I;\mathrm{J}}} \leq \frac{C_3}{p^{1/2}}.$$

The proof of the above ratio control result is the most complicated among results of this type studied in this paper. The main complication is due to the existence of the $\mathrm{Tr}(S)$ term in the denominator in (5.31), which leads to the complications both in the control of $V^2_{\Sigma;J}$ and the 'residual term' $Q_{\mathrm{J}}(\Sigma \cdot b^{-1}(\Sigma))$. On the other hand, similar to Proposition 25-(3), the asymptotic formula for $\sigma^2_{I;J}$ in the above proposition also removes the condition $0 < \liminf(p/N) \leq \limsup(p/N) < \infty$ that is required in (Ledoit and Wolf, 2002, Proposition 3), via a comparison to $\sigma^2_{I;\mathrm{LNW}}$ studied in Proposition 23-(3).

Let $\Psi_{\mathrm{J}}(X)$ be the test built from (5.1) and the statistic in (5.31). Combining the above results with Theorem 13 and some additional efforts to remove the residual term $Q_{\mathrm{J}}(\Sigma \cdot b^{-1}(\Sigma))$, we have the following asymptotic power formula for $\Psi_{\mathrm{J}}(X)$; see Section D.7.4 for its proof.

Theorem 25. *Suppose $p/N \leq M$ for some $M > 1$. Then there exists some constant $C = C(\alpha, M) > 0$ such that*

$$\left|\mathbb{E}_\Sigma \Psi_{\mathrm{J}} - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2}{4\sigma_{I;\mathrm{J}}}, 1\right) > z_\alpha\right)\right| \leq C \cdot p^{-1/3}.$$

Consequently, in the asymptotic regime $N \wedge p \to \infty$ with $\limsup(p/N) < \infty$,

$$\mathbb{E}_\Sigma \Psi_{\mathrm{J}} \sim 1 - \Phi\left(z_\alpha - \frac{\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F^2}{2(p/N)}\right).$$

The power behavior for John's test is previous studied in Onatski et al. (2013, 2014); Wang and Yao (2013) for a special class of alternatives under the spiked covariance model with a fixed number of spikes; see Section 5.5 ahead for a detailed discussion. To the best of our knowledge, the theorem above gives the first complete characterization of the power behavior for John's test for arbitrary alternatives in the high-dimensional regime $N \wedge p \to \infty$ with $\limsup(p/N) < \infty$.

5.5 Spiked covariance models

In this section, we consider a special class of alternatives known as the spiked covariance model Johnstone (2001):

$$\Sigma(a) = \mathrm{diag}\big(1 + a_1, \ldots, 1 + a_p\big), \tag{5.32}$$

where $a = (a_1, \ldots, a_p) \in (-1, \infty)^p$. Write $\overline{a} = \sum_{j=1}^p a_j/p$. Specializing the results obtained in Sections 5.3 and 5.4, we have the following.

Corollary 4. *The following hold.*

1. *The power for the likelihood ratio test of $\Sigma = I$ satisfies*

$$\mathbb{E}_{\Sigma(a)}\Psi_{\mathrm{LRT}} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p \big(a_j - \log(1 + a_j)\big)}{\sqrt{2\big(-\frac{p}{N} - \log\big(1 - \frac{p}{N}\big)\big)}}\right) \equiv \beta_{\mathrm{LRT}}(a),$$

   under $N \wedge p \to \infty$ with $\limsup(p/N) < 1$.

2. *The powers for Ledoit-Nagao-Wolf and Cai-Ma's tests of $\Sigma = I$ satisfy*

$$\mathbb{E}_{\Sigma(a)}\Psi_{\mathrm{LNW}} \sim \mathbb{E}_{\Sigma(a)}\Psi_{\mathrm{CM}} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p a_j^2}{2(p/N)}\right) \equiv \beta_{\mathrm{LNW,CM}}(a),$$

   under $N \wedge p \to \infty$ with $\limsup(p/N) < \infty$.

3. *The power for the likelihood ratio test of $\Sigma = \lambda I$ satisfies*

$$\mathbb{E}_{\Sigma(a)}\Psi_{\mathrm{LRT};s} \sim 1 - \Phi\left(z_\alpha - \frac{\sum_{j=1}^p \log \frac{1+\overline{a}}{1+a_j}}{\sqrt{2\big(-\frac{p}{N} - \log\big(1 - \frac{p}{N}\big)\big)}}\right) \equiv \beta_{\mathrm{LRT};s}(a),$$

   under $N \wedge p \to \infty$ with $\limsup(p/N) < 1$.

4. *The power for John's test of* $\Sigma = \lambda I$ *satisfies*

$$\mathbb{E}_{\Sigma(a)}\Psi_{\mathrm{J}} \sim 1 - \Phi\bigg(z_\alpha - \frac{\sum_{j=1}^p (a_j - \overline{a})^2/(1+\overline{a})^2}{2(p/N)}\bigg) \equiv \beta_{\mathrm{J}}(a),$$

under $N \wedge p \to \infty$ *with* $\limsup(p/N) < \infty$.

(1), (2) and (4) above recover (Onatski et al., 2014, Proposition 8 (i)-(ii)), while (3)-(4) above recover (Wang and Yao, 2013, Equations (4.5) and (4.8)). Both Onatski et al. (2013); Wang and Yao (2013) considered the case where $r \equiv \|a\|_0$ and the non-zero elements of a are fixed. The techniques in Onatski et al. (2014) work with a further restriction $\|a\|_\infty < \sqrt{y}$ where y is the limiting value of the ratio p/N. This restriction coincides with the Baik-Ben Arous-Péché (BBP) phase transition Baik et al. (2005), and is essential for the techniques of Onatski et al. (2014), due to the singular nature of the likelihood ratio process when $\|a\|_\infty > \sqrt{y}$ already in the case $r = 1$, see (Onatski et al., 2013, Theorem 8). The restriction $\|a\|_\infty < \sqrt{y}$ is removed in Wang and Yao (2013) for the likelihood ratio test $\Psi_{\mathrm{LRT};s}$ and John's test Ψ_{J} for sphericity, by variations of Bai-Silverstein techniques developed in Bai and Silverstein (2004); Bai et al. (2009). Both results in Wang and Yao (2013) and Onatski et al. (2014) also hold beyond Gaussian distributions.

It is easy to see that in the setting of Onatski et al. (2013); Wang and Yao (2013) with a fixed number of spikes as described above, the asymptotic powers are the same for the following two group of tests:

1. Likelihood ratio tests $\Psi_{\mathrm{LRT}}, \Psi_{\mathrm{LRT};s}$: $\beta_{\mathrm{LRT}} = \beta_{\mathrm{LRT};s}$.

2. Ledoit-Nagao-Wolf, Cai-Ma and John's tests: $\beta_{\mathrm{LNW,CM}} = \beta_{\mathrm{J}}$.

Clearly, neither group of tests universally dominates the other in terms of the power behavior. For instance, the power of tests in (1) dominates that of (2) when some of a_j's are close to -1 (i.e., Σ is near singular), while the reversed phenomenon occurs when some of a_j's are close to ∞.

In general, the asymptotic power equivalence of the above two groups may not hold when the number of spikes are no longer fixed. Instead, we have the following power ordering within each group.

Corollary 5. *1. Likelihood ratio tests $\Psi_{\mathrm{LRT}}, \Psi_{\mathrm{LRT};s}$ have the power ordering:*

$$\beta_{\mathrm{LRT}}(a) \geq \beta_{\mathrm{LRT};s}(a).$$

2. Ledoit-Nagao-Wolf, Cai-Ma and John's tests $\Psi_{\mathrm{LNW}}, \Psi_{\mathrm{CM}}, \Psi_{\mathrm{J}}$ have the power ordering:

$$\beta_{\mathrm{LNW,CM}}(a) \begin{cases} \geq \beta_{\mathrm{J}}(a), & \overline{a^2}\big(1-(1+\overline{a})^2\big) \leq \overline{a}^2; \\ < \beta_{\mathrm{J}}(a), & \overline{a^2}\big(1-(1+\overline{a})^2\big) > \overline{a}^2. \end{cases}$$

Here $\overline{a^2} \equiv \sum_{j=1}^p a_j^2/p$.

Proof. (1) follows from the inequality $\sum_{j=1}^p \log(1+\overline{a}) \leq \sum_{j=1}^p \overline{a} = \sum_{j=1}^p a_j$. (2) follows by the following calculation:

$$\frac{\sum_{j=1}^p (a_j - \overline{a})^2}{(1+\overline{a})^2} = \frac{\sum_{j=1}^p (a_j^2 - \overline{a}^2)}{(1+\overline{a})^2} = \sum_{j=1}^p a_j^2 + \frac{\sum_{j=1}^p a_j^2 \cdot \big(1-(1+\overline{a})^2\big) - p\overline{a}^2}{(1+\overline{a})^2}.$$

The proof is complete. □

Note that $\{\overline{a} \geq 0\} \subsetneq \{\overline{a^2}\big(1-(1+\overline{a})^2\big) \leq \overline{a}^2\}$ (the inclusion is in fact proper), so if $\overline{a} \geq 0$, John's test Ψ_{J} will be less powerful than Ledoit-Nagao-Wolf and Cai-Ma's tests $\Psi_{\mathrm{LNW}}, \Psi_{\mathrm{CM}}$. Furthermore, both inequalities in the above corollary can be strict asymptotically, and similar to the discussion above, there are no universal power dominance relationships between the tests in the two groups.

5.5.1 An illustrative simulation study

Below we present some simulation results in the current spiked model setting to support the findings of Corollary 4 and 5. The confidence level will be taken to be $\alpha = 0.05$. The power of the considered tests in all the simulations below is calculated under the configuration $(n, p) = (300, 200)$ using an average of 1000 replications. Theoretical normal power curves predicted by Corollary 4 are plotted in solid lines and the empirical ones are plotted in dotted lines.

- (*Fixed number of spikes, near singular* Σ) In the top left panel of Figure 5.1, we fix a number of $r = 5$ spikes with the same magnitude $a_j = -1 + \tau^{-1}$ with $\tau \in \{1, 2, \dots, 10\}$. Note that $\tau = 1$ corresponds to the null hypothesis with no spikes, and as τ grows Σ becomes increasingly singular. As predicted by Corollary 4 and the discussion thereafter, in this case of near singular alternatives Σ, the power of the tests in the first group $\{\Psi_{\mathrm{LRT}}, \Psi_{\mathrm{LRT};s}\}$ dominates that of the second group $\{\Psi_{\mathrm{LNW}}, \Psi_{\mathrm{CM}}, \Psi_{\mathrm{J}}\}$.

- (*Fixed number of spikes, non-singular* Σ) In the top right panel of Figure 5.1, we again fix a number of $r = 5$ spikes with the same magnitude $a_j = 0.3(\tau - 1)$ with $\tau \in \{1, 2, \dots, 10\}$. Again $\tau = 1$ corresponds to the null case. Compared to the previous case, we have the same power grouping effect but the dominance is reversed.

- (*Growing number of spikes*) In the bottom panel of Figure 5.1, we fix a number of $r = 50$ spikes with the same magnitude $a_j = 0.08(\tau - 1)$ with $\tau \in [10]$. This exemplifies the case of a growing number of spikes where, as predicted by Corollary 5, we see: (i) The power of LRT Ψ_{LRT} for testing identity dominates its counterpart $\Psi_{\mathrm{LRT};s}$ for testing sphericity; (ii) the powers of the $\{\Psi_{\mathrm{LNW}}, \Psi_{\mathrm{CM}}\}$ and Ψ_{J} are no longer equivalent, with the former having a strictly larger power as $\overline{a} > 0$.

5.6 Concluding remarks

In this paper, we develop a general method for power analysis in high dimensional covariance testing problems when a CLT holds for the test statistic under the null. We apply the new method to a number of tests in two prototypical problems of testing identity $\Sigma = I$ and sphericity $\Sigma = \lambda I$ with unspecified $\lambda > 0$. The key technical step is to control the ratio (5.14) which typically requires case-specific techniques. A strong enough control of the ratio (5.14), as demonstrated in many examples in the paper, leads to a sharp asymptotic power expansion of the test that holds for arbitrary alternatives. If normal approximation of the test statistic under the null can be quantified non-asymptotically, then the full finite-sample strength of our method can be utilized. For the tests studied in this paper, this is achieved via non-trivial applications of Chatterjee's second-order Poincaré inequality Chatterjee (2009) (see Lemma 58).

Below we sketch some directions for future research:

1. (*Upper limit condition on* (n, p)) Other than the minimal condition $n \wedge p \to \infty$, two additional conditions on (n, p) made in this paper are: (i) in results for the LRT (Sections 5.3.1, 5.3.2, 5.4.1), we assume that $\limsup(p/n) < 1$ which precludes the boundary case $p/n \to 1$; (ii) for the asymptotic power formulae of all other test statistics (Theorems 19, 21, 25), we require that $\limsup(p/n) < \infty$. Some degree of relaxation is certainly possible, but it remains an open question to obtain the best possible upper limit condition for p/n under which the power expansion of these tests is still valid.

2. (*Gaussian assumption*) Throughout the paper we have worked with Gaussian observations to obtain complete characterizations for the power behavior of various tests. Gaussianity enters at a technical level via the Gaussian-Poincaré inequality, the second-order Poincaré inequality Chatterjee (2009), Fourier expansion in the Gaussian space Nourdin and Peccati (2012). It is of great interest to extend our results to non-Gaussian observations. The main technical hurdles seem however to be sharp estimates for intermediate terms like $V_{(\mu,\Sigma)}$, $|m_{(\mu,\Sigma)} - m_{H_0}|$, and $\sigma^2_{H_0}$.

3. (*Other testing problems*) The general method developed in Section 5.2 could potentially be applied to other testing problems, including but not limited to, (block) independence

test, two- and multi-sample (joint) testing of μ and Σ, testing of regression coefficients in multivariate linear regression. These problems require further technical works and will therefore be pursued elsewhere.

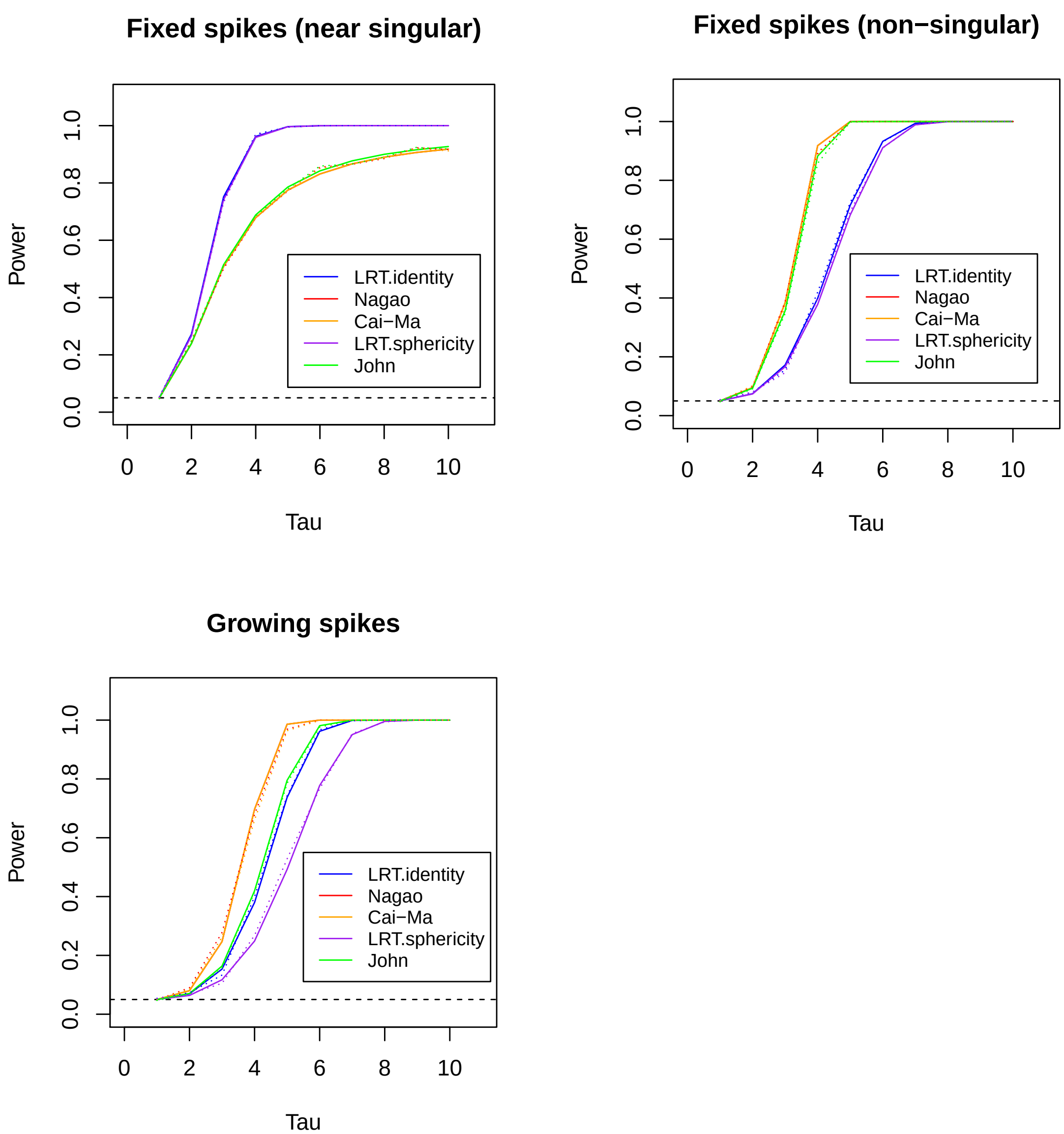

Figure 5.1: The power curves for the tests $\{\Psi_{\mathrm{LRT}}, \Psi_{\mathrm{LNW}}, \Psi_{\mathrm{CM}}, \Psi_{\mathrm{LRT};s}, \Psi_{\mathrm{J}}\}$ in the spiked covariance model (5.32). τ on the x-axis indexes the spike magnitude, with $\tau = 1$ corresponding to the null case $\Sigma = I$. The dashed horizontal line marks the prescribed size $\alpha = 0.05$. Note that the red solid line for Ψ_{LNW} is not visible as it coincides with the orange one for Ψ_{CM} predicted by Corollary 4-(2).

BIBLIOGRAPHY

Amelunxen, D., Lotz, M., McCoy, M. B., and Tropp, J. A. (2014). Living on the edge: phase transitions in convex programs with random data. *Inf. Inference*, 3(3):224–294.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. Wiley Publications in Statistics. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London.

Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.*, 51(1):39–54.

Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B*, 55(2):549–557.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.*, 37(6B):3822–3840.

Bai, Z. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605.

Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697.

Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Statist.*, 35(6):2536–2564.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. John

Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics.

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413.

Bartholomew, D. J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika*, 46(1-2):36–48.

Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives. II. *Biometrika*, 46:328–335.

Bartholomew, D. J. (1961a). Ordered tests in the analysis of variance. *Biometrika*, 48:325–332.

Bartholomew, D. J. (1961b). A test of homogeneity of means under restricted alternatives. *J. Roy. Statist. Soc. Ser. B*, 23:239–281.

Bellec, P. C. (2018). Sharp oracle inequalities for Least Squares estimators in shape restricted regression. *Ann. Statist.*, 46(2):745–780.

Bellec, P. C. and Zhang, C.-H. (2018). Second order stein: Sure for sure and other applications in high-dimensional inference. *arXiv preprint arXiv:1811.04121*.

Besson, O. (2006). Adaptive detection of a signal whose signature belongs to a cone. In *Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006.*, pages 409–413. IEEE.

Bhattacharya, A., Pati, D., and Dunson, D. (2014). Anisotropic function estimation using multi-bandwidth Gaussian processes. *The Annals of Statistics*, 42(1):352–381.

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā: The Indian Journal of Statistics, Series A*, 50(3):381–393.

Birgé, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probability Theory and Related Fields*, 71(2):271–291.

Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields*, 97(1-2):113–150.

Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc. (JEMS)*, 3(3):203–268.

Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.*, 74(3):281–289.

Bogachev, V. I. (1998). *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, Oxford.

Bousquet, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 213–247. Birkhäuser, Basel.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Boysen, L., Kempe, A., Liebscher, V., Munk, A., and Wittich, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.*, 37(1):157–183.

Brown, L. D. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *The Annals of Statistics*, 35(5):2219–2232.

Brown, L. D. and Low, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *The Annals of Statistics*, 24(6):2524–2535.

Cai, T. T., Levine, M., and Wang, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *Journal of Multivariate Analysis*, 100(1):126–136.

Cai, T. T. and Low, M. G. (2006). Optimal adaptive estimation of a quadratic functional. *The Annals of Statistics*, 34(5):2298–2325.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388.

Cai, T. T. and Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics*, 36(5):2025–2054.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.*, 106(494):608–625.

Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40.

Chatterjee, S. (2014a). A new perspective on least squares under convex constraint. *Ann. Statist.*, 42(6):2340–2381.

Chatterjee, S. (2014b). *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham.

Chatterjee, S., Guntuboyina, A., and Sen, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.*, 43(4):1774–1800.

Chen, H. and Jiang, T. (2018). A study of two high-dimensional likelihood ratio tests under alternative hypotheses. *Random Matrices Theory Appl.*, 7(1):1750016, 23.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159. Reprinted from SIAM J. Sci. Comput. **2**0 (1998), no. 1, 33–61 (electronic).

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, 105(490):810–819.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statistics*, 25:573–578.

Christensen, R. and Sun, S. K. (2010). Alternative goodness-of-fit tests for linear models. *J. Amer. Statist. Assoc.*, 105(489):291–301.

Cox, D., Koh, E., Wahba, G., and Yandell, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.*, 16(1):113–119.

Dalalyan, A. S., Hebiri, M., and Lederer, J. (2017). On the prediction performance of the Lasso. *Bernoulli*, 23(1):552–581.

Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.*, 29(1):1–65. With discussion and rejoinder by the authors.

de Boor, C. (1978). *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin.

de la Peña, V. H. and Giné, E. (1999). *Decoupling.* Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

Devroye, L., Mehrabian, A., and Reddad, T. (2018). The total variation distance between high-dimensional Gaussians. *arXiv preprint arXiv:1810.08693*.

Dicker, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika*, 101(2):269–284.

Doksum, K. and Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, 23(5):1443–1473.

Donoho, D. L. and Johnstone, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields*, 99(2):277–303.

Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Journal of Complexity*, 6(3):290–323.

Durot, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.*, 35(3):1080–1104.

Durot, C. and Tocquet, A.-S. (2001). Goodness of fit test for isotonic regression. *ESAIM Probab. Statist.*, 5:119–140.

Dykstra, R. (1991). Asymptotic normality for chi-bar-square distributions. *Canad. J. Statist.*, 19(3):297–306.

Eaton, M. L. (1983). *Multivariate statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. A vector space approach.

Efromovich, S. and Low, M. (1996). On optimal adaptive estimation of a quadratic functional. *The Annals of Statistics*, 24(3):1106–1125.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.*, 85(410):387–392.

Fan, J. (1991). On the estimation of quadratic functionals. *The Annals of Statistics*, 19(3):1273–1294.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21(1):196–216.

Fan, J. and Gijbels, I. (1995). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J. and Huang, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.*, 96(454):640–652.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660.

Fan, Z. and Guan, L. (2018). Approximate ℓ_0-penalized estimation of piecewise-constant signals on graphs. *Ann. Statist.*, 46(6B):3217–3245.

Fearnhead, P., Maidstone, R., and Letchford, A. (2019). Detecting changes in slope with an L_0 penalty. *J. Comput. Graph. Statist.*, 28(2):265–275.

Friedrich, F., Kempe, A., Liebscher, V., and Winkler, G. (2008). Complexity penalized M-estimation: fast computation. *J. Comput. Graph. Statist.*, 17(1):201–224.

Gao, C., Han, F., and Zhang, C.-H. (2019+). On estimation of isotonic piecewise constant signals. *Ann. Statist. (to appear). Available at arXiv:1705.06386.*

Gao, C. and Zhou, H. H. (2016). Rate exact Bayesian adaptation with modified block priors. *Ann. Statist.*, 44(1):318–345.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73(3):625–633.

Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA.

Giné, E. and Nickl, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli*, 14(1):47–61.

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York.

Goldstein, L., Nourdin, I., and Peccati, G. (2017). Gaussian phase transitions and conic intrinsic volumes: Steining the Steiner formula. *Ann. Appl. Probab.*, 27(1):1–47.

Greco, M., Gini, F., and Farina, A. (2008). Radar detection and classification of jamming signals belonging to a cone class. *IEEE transactions on signal processing*, 56(5):1984–1993.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. A roughness penalty approach.

Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 539–555. Wadsworth, Belmont, CA.

Groeneboom, P., Hooghiemstra, G., and Lopuhaä, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.*, 27(4):1316–1347.

Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.*, 33(2):840–870.

Guntuboyina, A., Lieu, D., Chatterjee, S., and Sen, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist.*, 48(1):205–229.

Guntuboyina, A. and Sen, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields*, 163(1-2):379–411.

Guntuboyina, A. and Sen, B. (2018). Nonparametric shape-restricted regression. *Statist. Sci.*, 33(4):568–594.

Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: the effect of estimating the mean. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1):3–14.

Hall, P., Kay, J. W., and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77(3):521–528.

Hall, P. and Marron, J. (1990). On variance estimation in nonparametric regression. *Biometrika*, 77(2):415–419.

Han, Q., Sen, B., and Shen, Y. (2020). High dimensional asymptotics of likelihood ratio tests in gaussian sequence model under convex constraint. *arXiv preprint arXiv:2010.03145*.

Han, Q., Wang, T., Chatterjee, S., and Samworth, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.*, 47(5):2440–2471.

Han, Q. and Zhang, C.-H. (2019+). Limit distribution theory for block estimators in multiple isotonic regression. *Ann. Statist. (to appear). Available at arXiv:1905.12825*.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.*, 105(492):1480–1493.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, 81(1):223–242.

Hoffman, M. and Lepski, O. (2002). Random rates in anisotropic regression (with discussion). *The Annals of Statistics*, 30(2):325–396.

Huang, L.-S. and Fan, J. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli*, 5(5):927–949.

Hütter, J.-C. and Rigollet, P. (2016). Optimal rates for total variation denoising. In *Conference on Learning Theory*, pages 1115–1146.

Ibragimov, I. and Khasminski, R. (1981). More on estimation of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.(LOMI)*, 108:72–88.

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, 12(2):105–108.

Jewell, S. and Witten, D. (2018). Exact spike train inference via ℓ_0 optimization. *Ann. Appl. Stat.*, 12(4):2457–2482.

Jiang, D., Jiang, T., and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference*, 142(8):2241–2256.

Jiang, T. (2019). Determinant of sample correlation matrix with application. *Ann. Appl. Probab.*, 29(3):1356–1397.

Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.*, 42(4):988–1009.

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.*, 41(4):2029–2074.

John, S. (1971). Some optimal multivariate tests. *Biometrika*, 58:123–127.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327.

Kato, K. (2009). On the degrees of freedom in shrinkage estimation. *J. Multivariate Anal.*, 100(7):1338–1352.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.*, 107(500):1590–1598.

Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009). l_1 trend filtering. *SIAM Rev.*, 51(2):339–360.

Kohler, M. (1999). Nonparametric estimation of piecewise smooth regression functions. *Statist. Probab. Lett.*, 43(1):49–55.

Kolchin, V. F., Sevastyanov, B. A., and Chistyakov, V. P. (1978). *Random Allocations*. Winston.

Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133.

Kong, W. and Valiant, G. (2018). Estimating learnability in the sublinear data regime. *arXiv preprint arXiv:1805.01626*.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, 50:403–418.

Kudô, A. and Choi, J. R. (1975). A generalized multivariate analogue of the one sided test. *Mem. Fac. Sci. Kyushu Univ. Ser. A*, 29(2):303–328.

Kur, G., Gao, F., Guntuboyina, A., and Sen, B. (2020). Convex regression in multidimensions: Suboptimality of least squares estimators. *arXiv preprint arXiv:2006.02044*.

Laurent, B. (1996). Efficient estimation of integral functionals of a density. *The Annals of Statistics*, 24(2):659–681.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4):1081–1102.

Lepski, O. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and Its Applications*, 35(3):454–466.

Lepskiĭ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 36(4):645–659.

Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Stat. Comput.*, 27(2):519–533.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.*, 25(1):387–413.

Massart, P. (2007). *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

Maz'ya, V. (2011). *Sobolev spaces with applications to elliptic partial differential equations*, volume 342. Springer, Heidelberg.

McCoy, M. B. and Tropp, J. A. (2014). From Steiner formulas for cones to concentration of intrinsic volumes. *Discrete Comput. Geom.*, 51(4):926–963.

Menéndez, J. A., Rueda, C., and Salvador, B. (1992a). Dominance of likelihood ratio tests under cone constraints. *Ann. Statist.*, 20(4):2087–2099.

Menéndez, J. A., Rueda, C., and Salvador, B. (1992b). Testing nonoblique hypotheses. *Comm. Statist. Theory Methods*, 21(2):471–484.

Menéndez, J. A. and Salvador, B. (1991). Anomalies of the likelihood ratio tests for testing restricted hypotheses. *Ann. Statist.*, 19(2):889–898.

Meyer, M. and Woodroofe, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.*, 28(4):1083–1104.

Meyer, M. C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika*, 90(1):223–232.

Meyer, Y. (1990). *Ondelettes et Opérateurs I: Ondelettes*. Hermann, Paris.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Muller, H.-G. and Stadtmuller, U. (1987). Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics*, 15(2):610–625.

Müller, U. U., Schick, A., and Wefelmeyer, W. (2003). Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. *Statistics: A Journal of Theoretical and Applied Statistics*, 37(3):179–188.

Munk, A., Bissantz, N., Wagner, T., and Freitag, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(1):19–41.

Munk, A. and Ruymgaart, F. (2002). Minimax rates for estimating the variance and its derivatives in non-parametric regression. *Australian and New Zealand Journal of Statistics*, 44(4):479–488.

Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, 1:700–709.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234.

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.*, 101(5):1067–1078.

Nickl, R. and van de Geer, S. (2013). Confidence sets in sparse regression. *Ann. Statist.*, 41(6):2852–2876.

Nourdin, I. and Peccati, G. (2012). *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge. From Stein's method to universality.

Nussbaum, M. (1986). On nonparametric estimation of a regression function, being smooth on a domain in $\mathbb{R}^k$. *Theory of Probability and its Applications*, 31:118–125.

Onatski, A., Moreira, M. J., and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.*, 41(3):1204–1231.

Onatski, A., Moreira, M. J., and Hallin, M. (2014). Signal detection in high dimension: the multispiked case. *Ann. Statist.*, 42(1):225–254.

Pillai, N. S. and Yin, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001.

Raskutti, G., Wainwright, M. J., and Yu, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inform. Theory*, 57(10):6976–6994.

Raubertas, R. F., Lee, C.-I. C., and Nordheim, E. V. (1986). Hypothesis tests for normal means constrained by linear inequalities. *Comm. Statist. A—Theory Methods*, 15(9):2809–2833.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.*, 12(4):1215–1230.

Rinaldo, A. (2009). Properties and refinements of the fused lasso. *Ann. Statist.*, 37(5B):2922–2952.

Robertson, T. and Wegman, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.*, 6(3):485–505.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester.

Robins, J., Li, L., Tchetgen, E., and van der Vaart, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics: Essays in Honor of David A. Freedman*, pages 335–421. Institute of Mathematical Statistics.

Robins, J., Tchetgen, E. T., Li, L., and van der Vaart, A. (2009). Semiparametric minimax rates. *Electronic Journal of Statistics*, 3:1305–1321.

Rockafellar, R. T. (1997). *Convex Analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ. Reprint of the 1970 original, Princeton Paperbacks.

Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. volume 60, pages 259–268. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

Ruiz, S. M. (1996). An algebraic identity leading to Wilson's theorem. *The Mathematical Gazette*, 80(489):579–582.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

Ruppert, D., Wand, M. P., Holst, U., and Hösjer, O. (1997). Local polynomial variance-function estimation. *Technometrics*, 39(3):262–273.

Sen, B. and Meyer, M. (2017). Testing against a linear regression model using ideas from shape-restricted estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(2):423–448.

Shah, R. D. and Bühlmann, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80(1):113–135.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72(1):133–144.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Internat. Statist. Rev.*, 56(1):49–62.

Shen, Y., Han, Q., and Han, F. (2020). On a phase transition in general order spline regression. *arXiv preprint arXiv:2004.10922*.

Spokoiny, V. (2002). Variance estimation for high-dimensional regression models. *Journal of Multivariate Analysis*, 82(1):111–133.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, 35(2):251–272.

Steidl, G., Didas, S., and Neumann, J. (2006). Splines in higher order tv regularization. *International Journal of Computer Vision*, 70(3):241–255.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, 25(2):613–641.

Thompson, A., Kay, J., and Titterington, D. (1991). Noise estimation in signal restoration using regularization. *Biometrika*, 78(3):475–488.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(1):91–108.

Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.*, 42(1):285–323.

Tibshirani, R. J. and Taylor, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.*, 40(2):1198–1232.

Tsybakov, A. B. (2009a). *Introduction to Nonparametric Estimation*. Springer, New York.

Tsybakov, A. B. (2009b). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

van de Geer, S. (2018). On tight bounds for the Lasso. *J. Mach. Learn. Res.*, 19:Paper No. 46, 48.

van der Vaart, A. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

van der Vaart, A. (2002). Semiparametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1999)*, volume 1781 of *Lecture Notes in Math.*, pages 331–457. Springer, Berlin.

van Lint, J. H. (1999). *Introduction to coding theory*, volume 86 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge.

Verzelen, N. and Gassiat, E. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli*, 24(4B):3683–3710.

Verzelen, N. and Villers, F. (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist.*, 38(2):704–752.

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics*, 12(4):367–395.

Von Neumann, J. (1942). A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics*, 13(1):86–88.

Wahba, G. (1990). *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Wang, L., Brown, L. D., Cai, T. T., and Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, 36(2):646–664.

Wang, Q. and Yao, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.*, 7:2164–2192.

Wang, Y.-X., Smola, A., and Tibshirani, R. (2014). The falling factorial basis and its statistical applications. In *International Conference on Machine Learning*, pages 730–738.

Warrack, G. and Robertson, T. (1984). A likelihood ratio test regarding two nested but oblique order-restricted hypotheses. *J. Amer. Statist. Assoc.*, 79(388):881–886.

Wei, Y., Wainwright, M. J., and Guntuboyina, A. (2019). The geometry of hypothesis testing over convex cones: generalized likelihood ratio tests and minimax radii. *Ann. Statist.*, 47(2):994–1024.

Winkler, G. and Liebscher, V. (2002). Smoothers for discontinuous signals. volume 14, pages 203–222. Statistical models and methods for discontinuous phenomena (Oslo, 1998).

Ye, F. and Zhang, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.*, 11:3519–3540.

Zhang, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.*, 30(2):528–555.

Zheng, S., Bai, Z., and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.*, 43(2):546–591.

Zou, H., Hastie, T., and Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *Ann. Statist.*, 35(5):2173–2192.

Appendix A

APPENDIX OF CHAPTER 2

A.1 Proof of Theorem 2

Proof. We will only prove the lower bound $n^{-8\alpha/(4\alpha+1)}$ in the regime $0 < \alpha < 1/4$ since for $\alpha \geq 1/4$, the rate reduces to the parametric rate n^{-1} and the proof is straightforward. Throughout the proof, C represents a generic sufficiently large positive constant and c represents a generic sufficiently small positive constant always taken to be smaller than $1/4$. Both C and c only depend on $\alpha, C_{\mathcal{F}}, C_\sigma, C_\varepsilon, C_0, c_0$ and might have different values for each occurrence. By appropriately rescaling the parameters in the lower bound construction, without loss of generality, we assume that the sample size n and the constants $C_{\mathcal{F}}, C_\sigma, C_\varepsilon, C_0$ are sufficiently large, c_0 is sufficiently small, and $[0,1] \subset I$.

We will make use of Le Cam's two point method. Introduce the following constants:

$$\theta_n^2 := h_n^{2\alpha} := cn^{-4\alpha/(4\alpha+1)} \quad \text{and} \quad N := N_n := 1/(6h_n), \tag{A.1}$$

where we tune the constant c in h_n so that N is a positive integer. We now specify $f(\cdot)$, distribution of X and distribution of ε in the null and alternative hypotheses, H_0 and H_1, respectively.

Choice of σ^2: Under H_0, let $\sigma^2 = 1 + \theta_n^2$. Under H_1, let $\sigma^2 = 1$.

Choice of ε: Under both H_0 and H_1, let $\varepsilon \sim \mathcal{N}(0,1)$.

Choice of X: Under both H_0 and H_1, let X be uniformly distributed on the union of the intervals $[(6i-5)h_n, (6i-1)h_n]$ for $i \in [N]$.

Choice of $f(\cdot)$: Under H_0, let $f \equiv 0$. Under H_1, let f take the value $h_n^\alpha r_i$ on $[(6i-5)h_n, (6i-1)h_n]$, where $\{r_i\}_{i=1}^N$ are N i.i.d. symmetric and bounded random variables with distribution $\mathbb{G}$ satisfying

$$\int_{-\infty}^{\infty} x^j \mathbb{G}(dx) = \int_{-\infty}^{\infty} x^j \varphi(x)dx, \quad j = 1, \ldots, q, \tag{A.2}$$

where q is some fixed odd integer strictly larger than $1 + 1/(2\alpha)$. Let f be 0 at points $6(i-1)h_n$ for $i \in [N]$, and then linearly interpolate f for the rest of the unspecified points on $[0, 1]$.

See Figure 2.1 for an illustration. In the definition of $f(\cdot)$ under H_1, the existence of the distribution $\mathbb{G}$ is guaranteed by Lemma 5, and the range of $\{r_i\}_{i=1}^N$, which we denote as B, only depends on α.

Clearly, $\sigma^2 \leq C_\sigma$ under both H_0 and H_1. Moreover, $f(\cdot)$ under both H_0 and H_1 belongs to $\Lambda_{\alpha,[0,1]}(C_{\mathcal{F}})$ due to the boundedness of $\{r_i\}_{i=1}^N$ in H_1. Next, we show that the joint distribution of (X, ε) belongs to $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$. Condition (d) clearly holds and Condition (a) holds with $I = [0, 1]$. Condition (b) holds as well by the fact that $p_X(u) = 3/2$ for $u \in [(6i-5)h_n, (6i-1)h_n]$ for $i \in [N]$ and $p_X(u) = 0$ otherwise. Lastly, for Condition (c), it holds by the convolution formula that for any $0 < u < 1/2$

$$\begin{aligned} p_{\widetilde{X}_{ij}}(u) = \int_u^1 p_X(t)p_X(t-u)dt &\geq \sum_{i=\lceil u/(6h_n)\rceil+1}^{N} \int_{(6i-5)h_n}^{(6i-1)h_n} p_X(t)p_X(t-u)dt \\ &\geq \sum_{i=\lceil u/(6h_n)\rceil+1}^{N} \frac{3}{2} \cdot \frac{3}{2} \cdot 2h_n \geq \frac{3}{8} - 9h_n \geq \frac{1}{4} \end{aligned}$$

for sufficiently large n. Here, the second inequality follows from the fact that for any fixed $t \in [(6i-5)h_n, (6i-1)h_n]$, $p_X(t) = 3/2$ and $p_X(t-u) = 0$ on a subset with Lebesgue measure at most $2h_n$. By symmetry of $\widetilde{X}_{ij}$, Condition (c) also holds with $\delta_0 = 1/2$ and $\mathcal{U}_\delta \equiv [-1, 1]$.

Denote by $\sigma_i^2, f_i, \mathbb{P}_{i,(X,\varepsilon)}$, $i = 0, 1$, the choice of σ^2, f, and $\mathbb{P}_{(X,\varepsilon)}$ under H_0 and H_1, respectively. Let π be the distribution on $\Lambda_{\alpha,I}(C_{\mathcal{F}})$ such that $f_1 \sim \pi$. Moreover, let $\mathbb{E}_{\sigma^2,f,\mathbb{P}_{(X,\varepsilon)}}$

represent the expectation with respect to the model (2.2) with parameters $\sigma^2, f, \mathbb{P}_{(X,\varepsilon)}$. Then, we have

$$\begin{aligned}
&\inf_{\widetilde{\sigma}^2} \sup_{f\in\Lambda_{\alpha,I}(C_{\mathcal{F}})} \sup_{\sigma^2\le C_\sigma} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}} \mathbb{E}\big(\widetilde{\sigma}^2-\sigma^2\big)^2 \\
&\ge \inf_{\widetilde{\sigma}^2}\bigg\{\frac{1}{2}\mathbb{E}_{\sigma_0^2,f_0,\mathbb{P}_{0,(X,\varepsilon)}}\big(\widetilde{\sigma}^2-\sigma^2\big)^2+\frac{1}{2}\int \mathbb{E}_{\sigma_1^2,f,\mathbb{P}_{1,(X,\varepsilon)}}\big(\widetilde{\sigma}^2-\sigma^2\big)^2 d\pi(f)\bigg\} \\
&\ge \inf_{\widetilde{\sigma}^2}\bigg\{\frac{1}{2}\mathbb{E}_{\sigma_0^2,f_0,\mathbb{P}_{0,(X,\varepsilon)}}\big(\widetilde{\sigma}^2-\sigma^2\big)^2+\frac{1}{2}\mathbb{E}_{\sigma_1^2,f_1,\mathbb{P}_{1,(X,\varepsilon)}}\big(\widetilde{\sigma}^2-\sigma^2\big)^2\bigg\},
\end{aligned}$$

where the first inequality follows by lower bounding the maximum risk with Bayes risk with prior π. In what follows, we will use $\mathbb{P}_0$ and $\mathbb{P}_1$ to denote the joint distribution of $\{Y_i, X_i\}_{i=1}^n$ under H_0 and H_1, respectively. Note that the choice of θ_n^2 in (A.1) leads to the desired lower bound under the quadratic loss. Therefore, adopting the standard reduction scheme with Le Cam's two point method (cf. Theorem 2.2 in Tsybakov (2009a)), it suffices to show that $\mathsf{TV}(\mathbb{P}_0,\mathbb{P}_1)\le c<1$. To show this, let $\{\widetilde{r}_i\}_{i=1}^N$ be N i.i.d. standard normal random variables, and $\widetilde{\mathbb{P}}_1$ be the joint distributions of $\{X_i, Y_i\}_{i=1}^n$ under H_1 with $\{r_i\}_{i=1}^N$ replaced by $\{\widetilde{r}_i\}_{i=1}^N$. Then, by triangle inequality, we have

$$\mathsf{TV}(\mathbb{P}_0,\mathbb{P}_1)\le \mathsf{TV}(\mathbb{P}_0,\widetilde{\mathbb{P}}_1)+\mathsf{TV}(\mathbb{P}_1,\widetilde{\mathbb{P}}_1).$$

We will show $\mathsf{TV}(\mathbb{P}_0,\widetilde{\mathbb{P}}_1)\le c$ and $\mathsf{TV}(\mathbb{P}_1,\widetilde{\mathbb{P}}_1)\le c$ seperately.

For the first inequality, define $\boldsymbol{x}:=(x_1,\ldots,x_n)$, $d\boldsymbol{x}:=dx_1\ldots dx_n$ and similarly for $\boldsymbol{y}$ and $d\boldsymbol{y}$. Denote p_0, p_1, and $\widetilde{p}_1$ as the densities of $\mathbb{P}_0$, $\mathbb{P}_1$, and $\widetilde{\mathbb{P}}_1$ with respect to the Lebesgue measure. Then, we have

$$\begin{aligned}
\mathsf{TV}(\mathbb{P}_0,\widetilde{\mathbb{P}}_1) &= \frac{1}{2}\int\int |*| p_0(\boldsymbol{x},\boldsymbol{y})-\widetilde{p}_1(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} \\
&= \int p(\boldsymbol{x})d\boldsymbol{x}\bigg\{\frac{1}{2}\int |*| p_0(\boldsymbol{y}\mid\boldsymbol{x})-\widetilde{p}_1(\boldsymbol{y}\mid\boldsymbol{x})d\boldsymbol{y}\bigg\} \qquad (A.3)\\
&= \int p(\boldsymbol{x})d\boldsymbol{x}\mathsf{TV}(\mathbb{P}_0(\boldsymbol{y}\mid\boldsymbol{x}),\widetilde{\mathbb{P}}_1(\boldsymbol{y}\mid\boldsymbol{x})),
\end{aligned}$$

where $p(\boldsymbol{x}):=\prod_{i=1}^n p_X(X_i)$ stands for the common density of $\{X_i\}_{i=1}^n$ under $\mathbb{P}_0$ and $\widetilde{\mathbb{P}}_1$. Note that under $\mathbb{P}_0$, $\boldsymbol{y}\mid\boldsymbol{x}\sim\mathcal{N}_n(0,\boldsymbol{\Sigma_0})$, with $\boldsymbol{\Sigma_0}=(1+\theta_n^2)\mathbf{I}_n$. Define $\{b_i\}_{i=1}^n$ to be the location index sequence of $\{X_i\}_{i=1}^n$ taking values in $[N]$, that is,

$$b_i = j \quad \text{if} \quad X_i \in [(6j-5)h_n, (6j-1)h_n].$$

Then, due to the symmetry of $\{r_i\}_{i=1}^N$ and design of the nonparametric component f, it holds that under $\widetilde{\mathbb{P}}_1$, $\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}_n(0, \boldsymbol{\Sigma_1})$, with $(\boldsymbol{\Sigma_1})_{ii} = 1 + h_n^{2\alpha} = 1 + \theta_n^2$ and $(\boldsymbol{\Sigma_1})_{ij} = h_n^{2\alpha}\mathbb{1}\{b_i = b_j\}$ for $i \neq j$. Define $N_0 := \sum_{i \neq j} \mathbb{1}\{b_i = b_j\}$. Since $\boldsymbol{\Sigma_1}$ is positive definite (see Lemma 12 in the supplement), we have by Lemma 6 that

$$\mathsf{TV}(\mathbb{P}_0(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x})) \leq C\frac{\theta_n^2}{1+\theta_n^2} N_0^{1/2} \leq C\theta_n^2 N_0^{1/2}.$$

Note that N_0 is a random variable that depends on $\{X_i\}_{i=1}^n$, and by (A.3) and Jensen's inequality we have

$$\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) \leq C\theta_n^2 \mathbb{E} N_0^{1/2} \leq C\theta_n^2 (\mathbb{E}N_0)^{1/2}.$$

Some simple algebra shows that $\mathbb{E}N_0 \leq Cn^2 h_n$, thus by choosing a sufficiently small c in the definition of h_n in (A.1), we have

$$\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) \leq C\theta_n^2 n h_n^{1/2} \leq c.$$

To complete the proof, we now show that $\mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) \leq c$. Consider an arbitrary realization of $\{X_i\}_{i=1}^n$, and assume that based on their location indices $\{b_i\}_{i=1}^n$, $\{X_i\}_{i=1}^n$ is partitioned into L clusters with corresponding cardinality s_ℓ so that the X_i's in the same cluster have the same value b_i. Apparently, we have the relations $1 \leq L \leq n$ and $\sum_{\ell=1}^L s_\ell = n$. Let $m_{\max}$ be the maximum cluster size, and define the "good event" $\Omega_n := \{m_{\max} \leq K\}$, where $K := \lfloor 2/(1-4\alpha)) \rfloor + 2$. Then, it holds that

$$\begin{aligned}
\mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) &= \mathbb{E}\Big(\mathbb{1}_{\Omega_n} \mathsf{TV}(\mathbb{P}_1(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x})\Big) + \mathbb{E}\Big(\mathbb{1}_{\Omega_n^c} \mathsf{TV}(\mathbb{P}_1(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x}))\Big) \\
&\leq \mathbb{E}\Big(\mathbb{1}_{\Omega_n} \mathsf{TV}(\mathbb{P}_1(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x})\Big) + \mathbb{P}(\Omega_n^c).
\end{aligned}$$

Under the choice of h_n in (A.1), N is of the order $n^{2/(4\alpha+1)}$, and

$$\lambda_K := \lim_{n\to\infty} \frac{n^K}{K!N^{K-1}} = 0.$$

Thus by Lemma 7 (and continuity), it holds that Ω_n has asymptotic probability 1 under both $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$. As a result, it suffices to upper bound $\mathsf{TV}(\mathbb{P}_1(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x}))$ for each realization $\boldsymbol{x}$ in Ω_n, where the maximum cluster size $m_{\max}$ is bounded by a fixed constant.

Denoting p_{1,π_ℓ} and $\widetilde{p}_{1,\pi_\ell}$ for each $\ell \in [L]$ as the joint density of those y_i's in the ℓth cluster π_ℓ conditioning on the given realization of $\{X_i\}_{i=1}^n$ under $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$, we obtain that

$$p_1(\boldsymbol{y} \mid \boldsymbol{x}) - \widetilde{p}_1(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{\ell=1}^{L} p_{1,\pi_\ell} - \prod_{\ell=1}^{L} \widetilde{p}_{1,\pi_\ell}.$$

The above inequality further implies by telescoping that

$$|*|p_1(\boldsymbol{y} \mid \boldsymbol{x}) - \widetilde{p}_1(\boldsymbol{y} \mid \boldsymbol{x}) \le \sum_{\ell=1}^{L} |*|p_{1,\pi_\ell} - \widetilde{p}_{1,\pi_\ell}.$$

For each $\ell \in [L]$, $|*|p_{1,\pi_\ell} - \widetilde{p}_{1,\pi_\ell}$ only depends on the ℓth cluster through its cardinality, which we now control for a general cluster size $d \ge 1$. Without loss of generality, we assume that $\ell = 1$ and the y_i's in this cluster are $\{y_1, \ldots, y_d\}$ with common location index $b_i = 1$ for $i \in [d]$. Then, under the choice of θ_n^2 in (A.1), we clearly have $Y_i = \theta_n r_1 + \varepsilon_i$ under $\mathbb{P}_1$ and $Y_i = \theta_n \widetilde{r}_1 + \varepsilon_i$ under $\widetilde{\mathbb{P}}_1$ for $i \in [d]$, where the sequence $\{\varepsilon_i\}_{i=1}^d$ follows the standard normal distribution under both $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$. Therefore it holds that

$$\begin{aligned} p_{1,\pi_1}(y_1, \ldots, y_d) &= \int_{-\infty}^{\infty} \varphi(y_1 - \theta_n v) \ldots \varphi(y_d - \theta_n v) \mathbb{G}(dv), \\ \widetilde{p}_{1,\pi_1}(y_1, \ldots, y_d) &= \int_{-\infty}^{\infty} \varphi(y_1 - \theta_n v) \ldots \varphi(y_d - \theta_n v) \varphi(v) dv, \end{aligned}$$

where $\mathbb{G}$ is the distribution of $\{r_i\}_{i=1}^N$ specified in (A.2). Using the well-known equality $\varphi(t - \theta_n v) = \varphi(t)(\sum_{k=0}^{\infty} v^k \theta_n^k H_k(t)/k!)$ for any t, v, where H_k is the kth order Hermite polynomial, it holds that

$$
\begin{aligned}
&\varphi(y_1-\theta_n v)\dots\varphi(y_d-\theta_n v)\\
&=\varphi(y_1)\dots\varphi(y_d)\sum_{k_1,\dots,k_d=0}^{\infty} v^{\sum_{i=1}^d k_i}\theta_n^{\sum_{i=1}^d k_i}\frac{H_{k_1}(y_1)}{k_1!}\dots\frac{H_{k_d}(y_d)}{k_d!}\\
&=\varphi(y_1)\dots\varphi(y_d)\sum_{k=0}^{\infty} v^k\theta_n^k\sum_{k_1+\dots+k_d=k}\frac{H_{k_1}(y_1)}{k_1!}\dots\frac{H_{k_d}(y_d)}{k_d!}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&p_{1,\pi_1}(y_1,\dots,y_d)-\widetilde{p}_{1,\pi_1}(y_1,\dots,y_d)\\
&=\varphi(y_1)\dots\varphi(y_d)\sum_{k=0}^{\infty}\theta_n^k\sum_{k_1+\dots+k_d=k}\frac{H_{k_1}(y_1)}{k_1!}\dots\frac{H_{k_d}(y_d)}{k_d!}\int v^k(\mathbb{G}-\Phi)(dv)\\
&=\varphi(y_1)\dots\varphi(y_d)\sum_{k=p}^{\infty}\theta_n^{2k}\sum_{k_1+\dots+k_d=2k}\frac{H_{k_1}(y_1)}{k_1!}\dots\frac{H_{k_d}(y_d)}{k_d!}\int v^{2k}(\mathbb{G}-\Phi)(dv),
\end{aligned}
$$

where the second equality follows by the symmetry and moment matching property of $\mathbb{G}$ in (A.2) and $p := (q+1)/2$ is a positive integer. This further yields

$$
\begin{aligned}
&|*|p_{1,\pi_1}(y_1,\dots,y_d)-\widetilde{p}_{1,\pi_1}(y_1,\dots,y_d)\\
&\le\varphi(y_1)\dots\varphi(y_d)\sum_{k=p}^{\infty}\theta_n^{2k}\sum_{k_1+\dots+k_d=2k}\frac{|*|H_{k_1}(y_1)}{k_1!}\dots\frac{|*|H_{k_d}(y_d)}{k_d!}\int v^{2k}\mathbb{G}(dv)+\\
&\quad\varphi(y_1)\dots\varphi(y_d)\sum_{k=p}^{\infty}\theta_n^{2k}\sum_{k_1+\dots+k_d=2k}\frac{|*|H_{k_1}(y_1)}{k_1!}\dots\frac{|*|H_{k_d}(y_d)}{k_d!}\int v^{2k}\varphi(v)dv\\
&:= I+II.
\end{aligned}
$$

For term I, since $\mathbb{G}$ is compactly supported on $[-B,B]$, one clearly has

$$
I\le\varphi(y_1)\dots\varphi(y_d)\sum_{k=p}^{\infty}\theta_n^{2k}B^{2k}\sum_{k_1+\dots+k_d=2k}\frac{|*|H_{k_1}(y_1)}{k_1!}\dots\frac{|*|H_{k_d}(y_d)}{k_d!}.
$$

For term II, using the equality $\int\varphi(v)v^{2k}dv=(2k-1)!!$, with $(2k-1)!!:=(2k-1)(2k-$

3) . . . 1, we obtain

$$II = \varphi(y_1)\dots\varphi(y_d)\sum_{k=p}^{\infty}\theta_n^{2k}(2k-1)!!\sum_{k_1+\ldots+k_d=2k}\frac{|*|H_{k_1}(y_1)}{k_1!}\cdots\frac{|*|H_{k_d}(y_d)}{k_d!}.$$

We now upper bound $\int_{-\infty}^{\infty}|*|H_k(t)\varphi(t)dt$ for an arbitrary positive integer k. When k is even, as has been calculated in Wang et al. (2008) (cf. chain of inequality after Equation (19) on Page 662), $\int_{-\infty}^{\infty}|*|H_k(t)\varphi(t)dt \le 2^{k/2}(k-1)!!$. When k is odd, set $k = 2\widetilde{k}+1$, then we have

$$\begin{aligned}
\int_{-\infty}^{\infty}|*|H_k(t)\varphi(t)dt &= \int_{-\infty}^{\infty}\varphi(t)|*|(2\widetilde{k}+1)!\sum_{m=0}^{\widetilde{k}}\frac{(-1)^m t^{2\widetilde{k}+1-2m}}{m!(2\widetilde{k}+1-2m)!2^m}dt \\
&\le \sum_{m=0}^{\widetilde{k}}\frac{(2\widetilde{k}+1)!}{m!(2\widetilde{k}+1-2m)!2^m}\int_{-\infty}^{\infty}|t|^{2\widetilde{k}+1-2m}\varphi(t)dt \\
&= \sqrt{\frac{2}{\pi}}\sum_{m=0}^{\widetilde{k}}\frac{(2\widetilde{k}+1)!(2\widetilde{k}-2m)!!}{m!(2\widetilde{k}+1-2m)!2^m} \\
&= \sqrt{\frac{2}{\pi}}\sum_{m=0}^{\widetilde{k}}\frac{(2\widetilde{k}+1)!(2m)!!}{(\widetilde{k}-m)!(2m+1)!2^{\widetilde{k}-m}} \\
&= \sqrt{\frac{2}{\pi}}(2\widetilde{k}+1)!!\sum_{m=0}^{\widetilde{k}}\frac{\widetilde{k}!}{m!(\widetilde{k}-m)!}\frac{(m!)^2 2^{2m}}{(2m+1)!} \\
&\le (2\widetilde{k}+1)!!\sum_{m=0}^{\widetilde{k}}\frac{\widetilde{k}!}{m!(\widetilde{k}-m)!} \\
&= (2\widetilde{k}+1)!!2^{\widetilde{k}},
\end{aligned}$$

where in the third line we use the fact that $\int_{-\infty}^{\infty}|t|^{2\ell+1}\varphi(t)dt = \sqrt{2/\pi}(2\ell)!!$ for any positive integer ℓ. Define for any positive integer k: $[k]_1 := k-1$ if k is even and k if k is odd, and $[k]_2 := k/2$ if k is even and $(k-1)/2$ if k is odd. Then, the above calculation implies that $\int_{-\infty}^{\infty}|*|H_k(t)\varphi(t)dt \le ([k]_1)!!2^{[k]_2}$ for any k, and moreover, it can be readily checked that

$([k]_1)!!/(k!) = 1/(2^{[k]_2}([k]_2)!)$. Therefore, for term $I = I(y_1, \ldots, y_d)$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d} I(y_1, \ldots, y_d) dy_1 \ldots dy_d \\
&\leq \sum_{k=p}^{\infty} \theta_n^{2k} (B^2)^k \sum_{k_1+\ldots+k_d=2k} \frac{1}{(k_1)! \ldots (k_d)!} ([k_1]_1)!! 2^{[k_1]_2} \ldots ([k_d]_1)!! 2^{[k_d]_2} \\
&= \sum_{k=p}^{\infty} \theta_n^{2k} (B^2)^k \sum_{k_1+\ldots+k_d=2k} \frac{1}{([k_1]_2)! \ldots ([k_d]_2)!}.
\end{aligned}$$

Now note that the number of d-tuple $(k_1, \ldots, k_d)$ such that $k_1 + \ldots + k_d = 2k$ is upper bounded by $(Ck)^d$, which is further bounded by C^k for every $k \geq 0$ with some sufficiently large C that only depends on d, and for each such tuple, it holds that

$$k - \frac{d}{2} = \sum_{i=1}^{d} \frac{k_i - 1}{2} \leq \sum_{i=1}^{d} [k_i]_2 \leq \sum_{i=1}^{d} \frac{k_i}{2} = k,$$

thus we have

$$\sum_{k_1+\ldots+k_d=2k} \{([k_1]_2)! \ldots ([k_d]_2)!\}^{-1} \leq C^k \sum_{k-d/2 \leq \bar{k}_1+\ldots+\bar{k}_d \leq k} \{(\bar{k}_1)! \ldots (\bar{k}_d)!\}^{-1}.$$

For the latter quantity, we have by the multinomial identity

$$\sum_{x_1+\ldots+x_{d+1}=k} k!/(x_1! \ldots x_{d+1}!)(d+1)^{-k} = 1$$

that

$$\begin{aligned}
\frac{(d+1)^k}{k!} &= \sum_{\bar{k}_1+\ldots+\bar{k}_{d+1}=k} \frac{1}{(\bar{k}_1)! \ldots (\bar{k}_{d+1})!} \\
&= \sum_{\bar{k}_1+\ldots+\bar{k}_d \leq k} \frac{1}{(\bar{k}_1)! \ldots (\bar{k}_d)!(k - (\bar{k}_1 + \ldots + \bar{k}_d))!} \\
&\geq \sum_{k-d/2 \leq \bar{k}_1+\ldots+\bar{k}_d \leq k} \frac{1}{(\bar{k}_1)! \ldots (\bar{k}_d)!(k - (\bar{k}_1 + \ldots + \bar{k}_d))!} \\
&\geq \left(\left(\frac{d}{2}\right)!\right)^{-1} \sum_{k-d/2 \leq \bar{k}_1+\ldots+\bar{k}_d \leq k} \frac{1}{(\bar{k}_1)! \ldots (\bar{k}_d)!}.
\end{aligned}$$

This concludes that

$$\int_{\mathbb{R}^d} I(y_1,\ldots,y_d)dy_1\ldots dy_d \leq \theta_n^{2p}\sum_{k=p}^{\infty}\frac{(CB^2)^k}{k!} \leq \theta_n^{2p}e^{CB^2}.$$

Using a similar argument for $II = II(y_1,\ldots,y_d)$, we obtain

$$\int_{\mathbb{R}^d} II(y_1,\ldots,y_d)dy_1\ldots dy_d \leq \sum_{k=p}^{\infty}\frac{(2k-1)!!}{k!}\theta_n^{2k}C^k = \sum_{k=p}^{\infty}\frac{(2k-1)!!}{(2k)!!}\theta_n^{2k}(2C)^k \leq \theta_n^{2p}C^p \quad \text{(A.4)}$$

since $\theta_n^2 < 1/C$ for sufficiently large n.

Putting together the pieces, we have for every realization $\boldsymbol{x}$ in Ω_n

$$\begin{aligned}\int_{\mathbb{R}^n} |*|p_1(\boldsymbol{y} \mid \boldsymbol{x}) - \widetilde{p}_1(\boldsymbol{y} \mid \boldsymbol{x})d\boldsymbol{y} &\leq \sum_{\ell=1}^{L}\int_{\mathbb{R}^{|\pi_\ell|}} |*|p_{1,\pi_\ell} - \widetilde{p}_{1,\pi_\ell} \leq L \max_{1\leq d\leq K}\theta_n^{2p}(e^{CB^2}+C^p) \\ &\leq n\theta_n^{2p}(e^{CB^2}+C^p) \leq c.\end{aligned}$$

Here, the second inequality follows since every $|*|p_{1,\pi_\ell} - \widetilde{p}_{1,\pi_\ell}$ depends on the ℓth cluster only through its cardinality, the third inequality follows since $L \leq n$ and K is a fixed absolute constant that only depends on α, and the last inequality follows due to the choice $\theta_n^2 = h_n^{2\alpha} = cn^{-4\alpha/(4\alpha+1)}$ and the value of p. This completes the proof. $\square$

Lemma 5 (Lemma 1, Wang et al. (2008)). *For any fixed positive integer q, there exist a $B < \infty$ and a symmetric distribution $\mathbb{G}$ on $[-B,B]$ such that $\mathbb{G}$ and the standard normal distribution have the same first q moments, that is,*

$$\int_{-B}^{B} x^j\mathbb{G}(dx) = \int_{-\infty}^{\infty} x^j\varphi(x)dx, \quad j = 1,\ldots,q.$$

Lemma 6 (Theorem 1.1, Devroye et al. (2018)). *If $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are positive definite $d \times d$ matrices, then*

$$\frac{1}{100} \leq \frac{TV(\mathcal{N}_d(\boldsymbol{\mu},\boldsymbol{\Sigma}_1),\mathcal{N}_d(\boldsymbol{\mu},\boldsymbol{\Sigma}_2))}{\min\{1,\|\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2 - \mathbf{I}_d\|_F\}} \leq \frac{3}{2}.$$

For the following lemma, we first introduce some terminology regarding the multinomial distribution. Let m, M be two positive integers, and the random vector $(f_1, \ldots, f_M)$ be the multinomial count with total count m and equal probability $(1/M, 1/M, \ldots, 1/M)$. Define $\rho := m/M$. For any positive integer $r \geq 2$, define $\lambda := \lambda_r := \lim_{m\to\infty} m^r/(r!M^{r-1})$. Following Kolchin et al. (1978) (Chapter 2, Equation (11)), we will call the domain of variation $m, M \to \infty$, in which

$$\rho \to 0, \quad 0 < \lambda_r < \infty$$

the *left-hand r-domain*. The following lemma characterizes the asymptotic behavior of the maximum frequency $f_{\max}$ defined as $\max_{1\leq j\leq M} f_j$.

Lemma 7 (Theorem 1 of Section 2.6, Kolchin et al. (1978)). *Suppose the multinomial distribution with total count m and equal probability $(1/M, \ldots, 1/M)$ is in the left-hand r-domain for some positive integer $r \geq 2$ with limit λ_r, then it holds that*

$$\mathbb{P}(f_{max} = r-1) \to e^{-\lambda_r} \text{ and } \quad \mathbb{P}(f_{max} = r) \to 1 - e^{-\lambda_r},$$

i.e., the maximum frequency converges asymptotically to a two-point distribution.

A.2 Proofs of results in Section 2.2

A.2.1 Proof of Theorem 1

Proof. Throughout the proof, we will use C, c to denote two generic fixed positive constants that only depend on $\overline{M}_K, \underline{M}_K, \alpha, C_{\mathcal{F}}, C_\sigma, C_\varepsilon, C_0, c_0$. C and c might have different values at each occurrence. We also use the notation $\widetilde{W}_{ij} := W_i - W_j$ for a generic random variable W.

Denote the two U-statistics on the numerator and denominator of $\widehat{\sigma}^2$ respectively as U_1, U_2, with corresponding mean values θ_1, θ_2. That is, with $i \neq j$,

$$\theta_1 := \mathbb{E}\big\{K_h(X_i - X_j)(Y_i - Y_j)^2/2\big\} \quad \text{and} \quad \theta_2 := \mathbb{E}K_h(X_i - X_j).$$

Define the "good" event $\mathcal{E} := \{U_2 \geq \theta_2/2\}$ and $\mathcal{E}^c$ as its complement, then it holds that

$$\mathbb{E}\big(\widehat{\sigma}^2 - \sigma^2\big)^2 = \mathbb{E}\bigg\{\Big(\frac{U_1 - U_2\sigma^2}{U_2}\Big)^2 \mathbb{1}\{\mathcal{E}\}\bigg\} + \mathbb{E}\bigg\{\Big(\frac{U_1 - U_2\sigma^2}{U_2}\Big)^2 \mathbb{1}\{\mathcal{E}^c\}\bigg\}. \tag{A.5}$$

By definition of $\mathcal{E}$, the first term satisfies that

$$\mathbb{E}\bigg\{\Big(\frac{U_1 - U_2\sigma^2}{U_2}\Big)^2 \mathbb{1}\{\mathcal{E}\}\bigg\} \leq \frac{4}{\theta_2^2}\mathbb{E}\big(U_1 - U_2\sigma^2\big)^2.$$

For θ_2, we have

$$\begin{aligned}\theta_2 = \mathbb{E}K_h(X_i - X_j) &= \int \frac{1}{h}K\Big(\frac{v}{h}\Big)p_{\widetilde{X}_{ij}}(v)dv = \int_{-1}^{1} K(u)p_{\widetilde{X}_{ij}}(uh)du \\ &\geq \int_{\mathcal{U}_h} K(u)p_{\widetilde{X}_{ij}}(uh)du \geq \inf_{u\in\mathcal{U}_h} p_{\widetilde{X}_{ij}}(uh) \inf_{u\in[-1,1]} K(u)\lambda(\mathcal{U}_h) \geq \underline{M}_K c_0^2.\end{aligned}$$

Here, the third equality follows from the fact that $K(\cdot)$ is supported in $[-1,1]$, and $\mathcal{U}_h$ starting from the first inequality is defined in Condition (c) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ (note that for any fixed $\delta_0 > 0$ given therein, $h \leq \delta_0$ for sufficiently large n). Moreover, it holds that

$$\mathbb{E}\big(U_1 - U_2\sigma^2\big)^2 \leq 3\Big\{\mathbb{E}(U_1 - \theta_1)^2 + \sigma^4\mathbb{E}(U_2 - \theta_2)^2 + \big(\theta_1 - \theta_2\sigma^2\big)^2\Big\}.$$

By Lemmas 8 and 9 and the fact that $\sigma^4 \leq C_\sigma^2$, we have

$$\mathbb{E}(U_1 - \theta_1)^2 + \sigma^4\mathbb{E}(U_2 - \theta_2)^2 \leq C(n^{-1} + n^{-2}h^{-1}).$$

For the third term $(\theta_1 - \theta_2\sigma^2)^2$, we have

$$\theta_1 = \mathbb{E}\big\{K_h(X_i - X_j)(Y_i - Y_j)^2/2\big\} = \mathbb{E}\big\{K_h(X_i - X_j)(f(X_i) - f(X_j))^2/2\big\} + \theta_2\sigma^2$$

and

$$\begin{aligned}&\mathbb{E}\big\{K_h(X_i - X_j)(f(X_i) - f(X_j))^2/2\big\} \leq C\mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)|*|\widetilde{X}_{ij}^{2(\alpha\wedge1)}\right\} \\ &= C\int \frac{1}{h}K\Big(\frac{u}{h}\Big)|*|u^{2(\alpha\wedge1)}p_{\widetilde{X}_{ij}}(u)du = C\int K(v)h^{2(\alpha\wedge1)}|v|^{2(\alpha\wedge1)}p_{\widetilde{X}_{ij}}(vh)dv \\ &\leq Ch^{2(\alpha\wedge1)}\sup_{u\in\mathbb{R}} p_{\widetilde{X}_{ij}}(u)\int K(v)|v|^{2(\alpha\wedge1)}dv \leq Ch^{2(\alpha\wedge1)}.\end{aligned}$$

Here, the first inequality follows since $f \in \Lambda_\alpha(C_{\mathcal{F}})$, and the last inequality follows from Condition (b) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ and the convolution formula. Putting together the pieces, the choice of h in (2.10) in the main paper yields

$$\mathbb{E}\left\{\left(\frac{U_1 - U_2\sigma^2}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}\}\right\} \le C(h^{4(\alpha\wedge 1)} + n^{-1} + n^{-2}h^{-1}) \le C(n^{-8\alpha/(4\alpha+1)} \vee n^{-1}).$$

For the second term in (A.5), we have

$$\mathbb{E}\left\{\left(\frac{U_1 - U_2\sigma^2}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}^c\}\right\} \le 2\sigma^4\mathbb{P}(\mathcal{E}^c) + 2\mathbb{E}\left\{\left(\frac{U_1}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}^c\}\right\}.$$

Direct calculation shows that

$$\begin{aligned}
\left(\frac{U_1}{U_2}\right)^2 &= \frac{\sum_{i<j,i'<j'} K_h\left(\widetilde{X}_{ij}\right) K_h\left(\widetilde{X}_{i'j'}\right)(Y_i - Y_j)^2(Y_{i'} - Y_{j'})^2/4}{\sum_{i<j,i'<j'} K_h\left(\widetilde{X}_{ij}\right) K_h\left(\widetilde{X}_{i'j'}\right)} \\
&\le \frac{\sum_{i<j,i'<j'} K_h\left(\widetilde{X}_{ij}\right) K_h\left(\widetilde{X}_{i'j'}\right)\left\{(f(X_i) - f(X_j))^2 + \sigma^2\widetilde{\varepsilon}_{ij}^2\right\}\left\{(f(X_{i'}) - f(X_{j'}))^2 + \sigma^2\widetilde{\varepsilon}_{i'j'}^2\right\}}{\sum_{i<j,i'<j'} K_h\left(\widetilde{X}_{ij}\right) K_h\left(\widetilde{X}_{i'j'}\right)} \\
&\le C\frac{\sum_{i<j,i'<j'} K_h\left(\widetilde{X}_{ij}\right) K_h\left(\widetilde{X}_{i'j'}\right)\left\{|*|\widetilde{X}_{ij}^{2(\alpha\wedge 1)} + \sigma^2\widetilde{\varepsilon}_{ij}^2\right\}\left\{|*|\widetilde{X}_{i'j'}^{2(\alpha\wedge 1)} + \sigma^2\widetilde{\varepsilon}_{i'j'}^2\right\}}{\sum_{i<j,i'<j'} K_h\left(\widetilde{X}_{ij}\right) K_h\left(\widetilde{X}_{i'j'}\right)} \\
&\le C\left(h^{4(\alpha\wedge 1)} + \sigma^2 h^{2(\alpha\wedge 1)} \max_{i<j} \widetilde{\varepsilon}_{ij}^2 + \sigma^4 \max_{i<j,i'<j'} \widetilde{\varepsilon}_{ij}^2\widetilde{\varepsilon}_{i'j'}^2\right),
\end{aligned}$$

where the last inequality follows by the support of $K(\cdot)$. By the condition $\sigma^2 \le C_\sigma$ and the independence of $\{\varepsilon_i\}_{i=1}^n$ and $\mathbb{1}\{\mathcal{E}\}$, this implies that

$$\mathbb{E}\left\{\left(\frac{U_1 - U_2\sigma^2}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}^c\}\right\} \le C\mathbb{P}(\mathcal{E}^c)\left\{1 + \mathbb{E}\max_{i<j} \widetilde{\varepsilon}_{ij}^2 + \mathbb{E}\max_{i<j,i'<j'} \widetilde{\varepsilon}_{ij}^2\widetilde{\varepsilon}_{i'j'}^2\right\}.$$

Applying the first part of Lemma 9 with $v = n\theta_2^2/16$, $u = n^2h\theta_2^2/16$ with the condition $h = \Omega(n^{-(2-\delta)})$ being satisfied with $\delta = 8\alpha/(4\alpha+1)$ and $\delta = 1$ for $\alpha \le 1/4$ and $\alpha > 1/4$ respectively, it holds that

$$\mathbb{P}(\mathcal{E}^c) = \mathbb{P}(|U_2 - \theta_2| \ge \theta_2/2) \le C\left\{\exp(-\theta_2^2 n/16) + \exp(-\theta_2^2 n^2 h/16)\right\}.$$

Moreover, by Condition (d) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$, there exists some fixed positive constant η such that

$$1 + \mathbb{E}\max_{i<j}\widetilde{\varepsilon}_{ij}^2 + \mathbb{E}\max_{i<j,i'<j'}\widetilde{\varepsilon}_{ij}^2\widetilde{\varepsilon}_{i'j'}^2 \le Cn^{\eta}.$$

Putting together the pieces and using the fact that $n^2h \to \infty$ as $n \to \infty$, it yields

$$\mathbb{E}\left\{\left(\frac{U_1 - U_2\sigma^2}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}^c\}\right\} = o(n^{-8\alpha/(4\alpha+1)} \vee n^{-1}).$$

This completes the proof. $\square$

A.2.2 Supporting lemmas

Lemma 8. *Suppose $f \in \Lambda_\alpha(C_{\mathcal{F}})$ and $\sigma^2 \le C_\sigma$ for some fixed constants $C_{\mathcal{F}}, C_\sigma$ and the joint distribution of (X,ε) satisfies Conditions (a), (b) and (d) in $\mathcal{P}_{cv,(X,\varepsilon)}$ with constants C_0, C_ε. Then, the U-statistic U_1 defined in the proof of Theorem 1 satisfies*

$$\mathbb{E}(U_1 - \theta_1)^2 \le C\big(n^{-1} \vee n^{-2}h^{-1}\big),$$

where C is some fixed positive constant that only depends on $\overline{M}_K, \underline{M}_K, \alpha, C_{\mathcal{F}}, C_\sigma, C_\varepsilon, C_0$.

Proof. Denote g as the kernel of U_1, that is,

$$g(\boldsymbol{D}_i, \boldsymbol{D}_j) := K_h(X_i - X_j)(Y_i - Y_j)^2/2, \quad \boldsymbol{D}_i := (X_i, \varepsilon_i)^\top.$$

Recall that $\theta_1 = \mathbb{E}g(\boldsymbol{D}_i, \boldsymbol{D}_j)$ for $i \ne j$. Then, it holds that

$$\mathbb{E}(U_1 - \theta_1)^2 = \binom{n}{2}^{-2} \sum_{i<j,i'<j'} \mathbb{E}\{(g(\boldsymbol{D}_i, \boldsymbol{D}_j) - \theta_1)(g(\boldsymbol{D}_{i'}, \boldsymbol{D}_{j'}) - \theta_1)\}. \tag{A.6}$$

When i, j, i', j' take four different values, the expectation is zero. When they take three values, say, $i = i' < j < j'$, by writing $\mathbb{E}_\varepsilon$ as the conditional expectation given $\{X_i\}_{i=1}^n$, we have

$$
\begin{aligned}
&\mathbb{E}(g(\boldsymbol{D}_i,\boldsymbol{D}_j)g(\boldsymbol{D}_i,\boldsymbol{D}_{j'}))\\
&=\frac{1}{4}\mathbb{E}\bigg\{\frac{1}{h^2}K\bigg(\frac{X_i-X_j}{h}\bigg)K\bigg(\frac{X_i-X_{j'}}{h}\bigg)(Y_i-Y_j)^2(Y_i-Y_{j'})^2\bigg\}\\
&\lesssim \mathbb{E}\bigg\{\frac{1}{h^2}K\bigg(\frac{X_i-X_j}{h}\bigg)K\bigg(\frac{X_i-X_{j'}}{h}\bigg)((f(X_i)-f(X_j))^2+\sigma^2\widetilde{\varepsilon}_{ij}^2)((f(X_i)-f(X_{j'}))^2+\sigma^2\widetilde{\varepsilon}_{ij'}^2)\bigg\}\\
&\lesssim \mathbb{E}\bigg\{\frac{1}{h^2}K\bigg(\frac{X_i-X_j}{h}\bigg)K\bigg(\frac{X_i-X_{j'}}{h}\bigg)\mathbb{E}_\varepsilon\Big\{\Big(|*|\widetilde{X}_{ij}^{2(\alpha\wedge 1)}+\sigma^2\widetilde{\varepsilon}_{ij}^2\Big)\Big(|*|\widetilde{X}_{ij'}^{2(\alpha\wedge 1)}+\sigma^2\widetilde{\varepsilon}_{ij'}^2\Big)\Big\}\bigg\}\\
&\lesssim \mathbb{E}\bigg\{\frac{1}{h^2}K\bigg(\frac{X_i-X_j}{h}\bigg)K\bigg(\frac{X_i-X_{j'}}{h}\bigg)\Big(|*|\widetilde{X}_{ij}\widetilde{X}_{ij'}{}^{2(\alpha\wedge 1)}+2\sigma^2\Big(|*|\widetilde{X}_{ij}^{2(\alpha\wedge 1)}+|*|\widetilde{X}_{ij'}^{2(\alpha\wedge 1)}\Big)+\mathbb{E}\varepsilon^4\cdot\sigma^4\Big)\bigg\}\\
&\lesssim \mathbb{E}\bigg\{\frac{1}{h^2}K\bigg(\frac{X_i-X_j}{h}\bigg)K\bigg(\frac{X_i-X_{j'}}{h}\bigg)\bigg\}\\
&=\int K(v)K(w)p_X(u)p_X(u+hv)p_X(u+hw)dudvdw\lesssim 1.
\end{aligned}
$$

In the last line, we invoke Conditions (b) and (d) in $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$. Moreover, it can be readily calculated that $\theta_1=O(1)$. This concludes that the summand in (A.6) is bounded by a fixed constant when i,j,i',j' take three different values. Lastly, performing a similar analysis, we obtain that $\mathbb{E}\{(g(\boldsymbol{D}_i,\boldsymbol{D}_j)-\theta_1)(g(\boldsymbol{D}_{i'},\boldsymbol{D}_{j'})-\theta_1)\}=O(1/h)$ when $i=i'$ and $j=j'$. We therefore conclude that

$$
\mathrm{Var}(U_1)\lesssim\frac{n^3+n^2h^{-1}}{n^4}\asymp n^{-1}+n^{-2}h^{-1}.
$$

This completes the proof. □

Lemma 9. *Suppose $h_n\gtrsim n^{-(2-\delta)}$ for some $0<\delta<2$. Then, assuming Condition (b) in $\mathcal{P}_{cv,(X,\varepsilon)}$ with constant C_0, the U-statistic U_2 defined in the proof of Theorem 1 satisfies*

$$
\mathbb{P}\Big(|*|U_2-\theta_2\geq C(v^{1/2}n^{-1/2}+u^{1/2}n^{-1}h^{-1/2})\Big)\leq C(\exp(-u)+\exp(-v))
$$

for any $u,v>0$, and

$$
\mathbb{E}(U_2-\theta_2)^2\leq C(n^{-1}\vee n^{-2}h^{-1}),
$$

where C is some fixed positive constant that only depends on $\overline{M}_K,\underline{M}_K,\alpha,C_0$.

Proof. We first prove the concentration inequality by upper bounding the 5 quantities in Lemma 13. Denote g as the kernel of U_2 and g_1 as its linear part, that is, for some $i \neq j$,

$$g_1(X_i) := \mathbb{E}(g(X_i, X_j) \mid X_i) := \mathbb{E}(K_h(X_i - X_j) \mid X_i).$$

For B_1, we have

$$g_1(X_i) = \int \frac{1}{h} K\left(\frac{u - X_i}{h}\right) p_X(u) du = \int K(u) p_X(uh + X_i) du \lesssim 1$$

due to Condition (b) in $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$. Thus it also holds $\nu_1^2 \lesssim 1$. For B_2, we have

$$\begin{aligned} B_2^2 = n \sup_{X_i} \mathbb{E}\big\{g^2(X_i, X_j) \mid X_i\big\} &= n \sup_{X_i} \int \frac{1}{h^2} K^2\left(\frac{u - X_i}{h}\right) p_X(u) du \\ &\lesssim \frac{n}{h} \sup_{X_i} \int K(u) p_X(uh + X_i) du \lesssim nh^{-1}, \end{aligned}$$

where in the first inequality we use the condition that $K(\cdot)$ is bounded by $\overline{M}_K$. Moreover, we clearly have $B_3 \lesssim h^{-1}$. Lastly, for ν_2^2, it holds that

$$\nu_2^2 = \int \frac{1}{h^2} K^2\left(\frac{u}{h}\right) p_{\widetilde{X}_{ij}}(u) du \lesssim \frac{1}{h} \int K(u) p_{\widetilde{X}_{ij}}(uh) du \lesssim \frac{1}{h},$$

where the last inequality follows by Condition (b) in $\mathcal{P}_{\mathrm{cv},(X,\varepsilon)}$ and the convolution formula

$$p_{\widetilde{X}_{ij}}(u) = \int p_X(t) p_X(t - u) dt \le \sup_{u \in \mathbb{R}} p_X(u) \int p_X(t) dt = \sup_{u \in \mathbb{R}} p_X(u).$$

Therefore, Lemma 13 yields that

$$\mathbb{P}(|U_2 - \theta_2| \ge a_1 v^{1/2} + a_2 v + b_1 u^{1/2} + b_2 u + b_3 u^{3/2} + b_4 u^2) \le C(\exp(-v) + \exp(-u)),$$

where $a_1 \lesssim n^{-1/2}, a_2 \lesssim n^{-1}, b_1 \lesssim n^{-1} h^{-1/2}, b_2 \lesssim n^{-1}, b_3 \lesssim n^{-3/2} h^{-1/2}, b_4 \lesssim n^{-2} h^{-1}$. Under the condition that $h \gtrsim n^{-(2-\delta)}$ for some $\delta > 0$ and n is sufficiently large, the dominant terms in the above inequality are a_1 and b_1, that is,

$$n^{-1/2} \vee n^{-1} h^{-1/2}.$$

This proves the first part of the theorem. The expectation version follows by Lemma 10. $\square$

Lemma 10. *Suppose a random variable X satisfies the tail condition $\mathbb{P}(|X| \geq a_1 t^{1/2} + a_2 t + a_3 t^{3/2} + a_4 t^2) \leq C_1 \exp(-C_2 t)$ for any $t > 0$ and some positive constants $a_1, a_2, a_3, a_4, C_1, C_2$. Then, for any positive integer p, it holds that*

$$\mathbb{E}(|X|^p)^{1/p} \leq C_3 p^{1/p}(a_1 + a_2 + a_3 + a_4)$$

for some positive constant C_3 that only depends on C_1, C_2.

Proof. We use C_3 to denote a positive constant that only depends on C_1 and C_2, which might have different values at each occurrence. The tail condition in the assumption is equivalent to

$$\mathbb{P}(|X| \geq t) \leq C_3 \exp\left\{-C_3\left(\frac{t^2}{a_1^2} \wedge \frac{t}{a_2} \wedge \frac{t^{2/3}}{a_3^{2/3}} \wedge \frac{t^{1/2}}{a_4^{1/2}}\right)\right\}.$$

Let I_1-I_4 be a partition of $(0, +\infty)$ such that for $t \in I_1$, $t^2/a_1^2 = \min\left\{\frac{t^2}{a_1^2} \wedge \frac{t}{a_2} \wedge \frac{t^{2/3}}{a_3^{2/3}} \wedge \frac{t^{1/2}}{a_4^{1/2}}\right\}$, and similarly for I_2, I_3, I_4. Then, we have

$$\begin{aligned}
&\mathbb{E}(|X|^p)\\
&= \int_0^\infty p t^{p-1} \mathbb{P}(|X| \geq t) dt\\
&\leq C_3 \Big\{ \int_{I_1} p t^{p-1} \exp\left(-C_3 \frac{t^2}{a_1^2}\right) dt + \int_{I_2} p t^{p-1} \exp\left(-C_3 \frac{t}{a_2}\right) dt + \int_{I_3} p t^{p-1} \exp\left(-C_3 \frac{t^{2/3}}{a_3^{2/3}}\right) dt +\\
&\quad \int_{I_4} p t^{p-1} \exp\left(-C_3 \frac{t^{1/2}}{a_4^{1/2}}\right) dt \Big\}\\
&\leq C_3 \Big\{ \int_0^\infty p t^{p-1} \exp\left(-C_3 \frac{t^2}{a_1^2}\right) dt + \int_0^\infty p t^{p-1} \exp\left(-C_3 \frac{t}{a_2}\right) dt + \int_0^\infty p t^{p-1} \exp\left(-C_3 \frac{t^{2/3}}{a_3^{2/3}}\right) dt +\\
&\quad \int_0^\infty p t^{p-1} \exp\left(-C_3 \frac{t^{1/2}}{a_4^{1/2}}\right) dt \Big\}\\
&= C_3 p \Big\{ a_1^p \int_0^\infty t^{p-1} \exp\big(-C_3 t^2\big) dt + a_2^p \int_0^\infty t^{p-1} \exp(-C_3 t) dt + a_3^p \int_0^\infty t^{p-1} \exp\big(-C_3 t^{2/3}\big) dt +\\
&\quad a_4^p \int_0^\infty t^{p-1} \exp\big(-C_3 t^{1/2}\big) dt \Big\}\\
&\leq C_3 p (a_1^p + a_2^p + a_3^p + a_4^p).
\end{aligned}$$

This completes the proof. $\square$

Lemma 11. *Recall the Condition (c) in the definition of $\mathcal{P}_{cv,(X,\varepsilon)}$ in the main paper, and Condition (c′) in the subsequent paragraph. We have (c′) ⇒ (c).*

Proof. Choose some $\delta_0 < 1/8$ and fix any $\delta \leq \delta_0$ and $u \in [-1, 1]$. By the convolution formula, we have

$$p_{\widetilde{X}_{ij}}(u\delta) = \int_S p_X(s)p_X(s - u\delta)ds \geq c_0 \int_S p_X(s - u\delta)ds.$$

Therefore, it suffices to show that $\lambda(\{S - u\delta\} \bigcap [0, 1] \bigcap S)$ is lower bounded by some fixed constant, say, 1/8. Assume this does not hold, then we have

$$1 = \lambda([0, 1]) \geq \lambda\Big(\{S - u\delta\} \bigcap [0, 1]\Big) + \lambda(S) - 1/8 \geq \lambda(S) - \delta + \lambda(S) - 1/8 \geq 5/4,$$

which is a contradiction. $\square$

Lemma 12. *The covariance $\mathbf{\Sigma}_1$ defined in the proof of Theorem 2 in the main paper is positive definite.*

Proof. We will prove that for any $\boldsymbol{a} \in \mathbb{R}^n$ with $\|\boldsymbol{a}\| = 1$, it holds that $\boldsymbol{a}^\top \mathbf{\Sigma}_1 \boldsymbol{a} > 0$. For each realization of $\{X_i\}_{i=1}^n$, partition the set $[n]$ into L clusters for some positive integer $1 \leq L \leq n$ such that the Y_i's in each cluster fall into the same trapezoid in the lower bound construction. Denote these L clusters as $\mathcal{M}_1, \ldots, \mathcal{M}_L$. Then, it follows that

$$\mathrm{Var}\left(\sum_{i=1}^n a_i Y_i \mid \{X_i\}_{i=1}^n\right) = \sum_{\ell=1}^L \mathrm{Var}\left(\sum_{i \in \mathcal{M}_\ell} a_i Y_i \mid \{X_i\}_{i=1}^n\right).$$

Since $\|\boldsymbol{a}\| = 1$, there exists some $\ell_0 \in [L]$ such that $A := \sum_{i \in \mathcal{M}_{\ell_0}} a_i^2 > 0$. Partition $\{i : i \in \mathcal{M}_{\ell_0}\}$ according to the sign:

$$\mathcal{A}_+ := \{i \in \mathcal{M}_{\ell_0} : a_i \geq 0\}, \quad \mathcal{A}_- := \{i \in \mathcal{M}_{\ell_0} : a_i < 0\}$$

and define $S_+ := \sum_{i\in\mathcal{A}_+} a_i$ and $A_+ := \sum_{i\in\mathcal{A}_+} a_i^2$, and S_- and A_- similarly. Then, $A = A_+ + A_-$. Moreover, it holds that

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{i\in\mathcal{M}_{\ell_0}} a_i Y_i \mid \{X_i\}_{i=1}^n\right) \\
&= (1+h_n^{2\alpha}) \sum_{i\in\mathcal{M}_{\ell_0}} a_i^2 + h_n^{2\alpha}\left(\sum_{i,j\in\mathcal{A}_+;i\neq j} a_i a_j + \sum_{i,j\in\mathcal{A}_-;i\neq j} a_i a_j + 2\sum_{i\in\mathcal{A}_+,j\in\mathcal{A}_-} a_i a_j\right) \\
&= A(1+h_n^{2\alpha}) + h_n^{2\alpha}\big(S_+^2 - A_+ + S_-^2 - A_- + 2S_+S_-\big) \\
&= A + h_n^{2\alpha}(S_+ + S_-)^2 \geq A > 0.
\end{aligned}$$

This completes the proof. □

Lemma 13 (Theorem 3.3, Giné et al. (2000))**.** *Let $Z_1,\ldots,Z_n,Z \in \mathcal{Z}$ be i.i.d., and $g : \mathcal{Z}^2 \to \mathbb{R}$ be a symmetric measurable function with $\mathbb{E}\{g(Z_1,Z_2)\} < \infty$. Write $U_n(g) := \sum_{i<j} g(Z_i,Z_j)$ and $g_1(z) := \mathbb{E}\{g(Z,z)\}$. Define*

$$B_1 := \sup_{Z_2} \mathbb{E}\{|*|g(Z_1,Z_2) \mid Z_2\}, \quad B_2 := \left(n\sup_{Z_2} \mathbb{E}\big\{g^2(Z_1,Z_2) \mid Z_2\big\}\right)^{1/2}, \quad B_3 := \|g\|_\infty$$

and

$$\nu_1^2 := \mathbb{E}\big\{g_1^2(Z_2)\big\}, \quad \nu_2^2 := \mathbb{E}\big\{g^2(Z_1,Z_2)\big\}.$$

Then, it holds that

$$\begin{aligned}
&\mathbb{P}\big(|*|U_n(g) - \mathbb{E}\{U_n(g)\} \geq t + C_1 n\nu_2 u^{1/2} + C_2 nB_1 u + C_3 B_2 u^{3/2} + C_4 B_3 u^2\big) \\
&\leq 2\exp\left(\frac{-t^2/n^2}{8n\nu_1^2 + 4B_1\cdot t/n}\right) + C_5 e^{-u},
\end{aligned}$$

where C_1-C_5 are absolute constants.

A.3 Proofs of results in Section 2.3

A.3.1 Proof of Theorem 3

Proof. Throughout the proof, C and c will denote two generic positive constants that do not depend on n and might have different values at each occurrence. We only prove the case for the pointwise error and the result for the integrated error will follow. Consider a fixed $x^* \in \operatorname{supp}(X)$. We will continue to use the notation $\ell, \boldsymbol{q}(\cdot), \mathbf{B}_n, X_{ij}, K_{ij}$ introduced in Section 2.3.1 in the main paper. We will drop the subscript in $\widehat{V}_{\mathrm{LP}}(x^*)$ for notational simplicity. Recall the choice of (h_1, h_2) in (2.17) in the main paper.

Define $\mathbf{B} := \mathbb{E}\mathbf{B}_n$ and the good event $\Omega_n := \{\|\mathbf{B}_n - \mathbf{B}\| \le 1/(2\|\mathbf{B}^{-1}\|)\}$. Note that Ω_n is well-defined as we now prove $\mathbf{B}$ is indeed invertible. For any $\boldsymbol{a} \in \mathbb{S}^{\ell}$, it holds that

$$
\begin{aligned}
&\boldsymbol{a}^\top \mathbf{B}\boldsymbol{a}\\
&= \int\int \left\{\boldsymbol{a}^\top \boldsymbol{q}\left(\frac{(u+v)/2 - x^*}{h_2}\right)\right\}^2 \frac{1}{h_1} K\left(\frac{u-v}{h_1}\right)\frac{1}{h_2}K\left(\frac{(u+v)/2 - x^*}{h_2}\right) p_X(u)p_X(v)dudv\\
&= \int\int \left\{\boldsymbol{a}^\top \boldsymbol{q}(v)\right\}^2 K(u)K(v)p_X(x^* + h_2 v + h_1 u/2)p_X(x^* + h_2 v - h_1 u/2)dudv\\
&= \int\int \left\{\boldsymbol{a}^\top \boldsymbol{q}(v - h_1 u/(2h_2))\right\}^2 K(u)K(v - h_1 u/(2h_2))p_X(x^* + h_2 v)p_X(x^* + h_2 v - h_1 u)dudv\\
&= \int_{-1}^{1}\int_{-1+h_1u/(2h_2)}^{1+h_1u/(2h_2)} \left\{\boldsymbol{a}^\top \boldsymbol{q}\left(v - \frac{h_1 u}{2h_2}\right)\right\}^2 K(u)K\left(v - \frac{h_1 u}{2h_2}\right)p_X(x^* + h_2 v)p_X(x^* + h_2 v - h_1 u)dvdu\\
&\ge \underline{M}_K^2 \int_{-1}^{1}\int_{-1+h_1u/(2h_2)}^{1+h_1u/(2h_2)} \left\{\boldsymbol{a}^\top \boldsymbol{q}\left(v - \frac{h_1 u}{2h_2}\right)\right\}^2 p_X(x^* + h_2 v)p_X(x^* + h_2 v - h_1 u)dvdu,
\end{aligned}
$$

where in the last inequality we use the lower bound $\underline{M}_K$ on $K(\cdot)$. Note that the first term of the integrand $\left\{\boldsymbol{a}^\top \boldsymbol{q}\left(v - \frac{h_1 u}{2h_2}\right)\right\}^2$ is a polynomial of variables u, v and thus only takes zero value with Lebesgue measure at most 0. By the second part of Condition (c) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ with $\delta = h_2$ (note that $h_2 \le \delta_0$ for any fixed $\delta_0 > 0$ and sufficiently large n), for the given x^*, there exists a set $\mathcal{A}_{x^*} \subset [-1, 1]$ with Lebesgue measure at least c_0 such that for all $v \in \mathcal{A}_{x^*}$, $x^* + h_2 v \in \operatorname{supp}(X)$, and moreover, for each $v \in \mathcal{A}_{x^*}$, the second part of Condition (c) with

$\delta = h_1$ again implies the existence of a set $\mathcal{A}_{x^*,v} \subset [-1,1]$ with Lebesgue measure at least c_0 such that for all $u \in \mathcal{A}_{x^*,v}$, it holds that $x^* + h_2 v - h_1 u \in \operatorname{supp}(X)$. Therefore, by the first part of Condition (c) and the fact that $h_1/h_2 \to 0$, there exists a fixed positive constant c such that

$$\lambda_{\min} := \inf_{\boldsymbol{a} \in \mathbb{S}^{\ell}} \boldsymbol{a}^\top \mathbf{B} \boldsymbol{a} \geq c > 0.$$

This concludes that Ω_n is well-defined. By triangle inequality, we have

$$\begin{aligned}
&\mathbb{E}\Big(\widehat{V}(x^*) - V(x^*)\Big)^2 \\
\lesssim\ &\mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n} - V(x^*)\Big)^2 + \mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n^c}\Big)^2 \\
\lesssim\ &\Big\{\mathbb{E}\Big((\widehat{V}(x^*) - V(x^*))\mathbb{1}_{\Omega_n}\Big)\Big\}^2 + \mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n} - \mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n}\Big)\Big)^2 + \\
&\mathbb{E}\Big(\Big(V^2(x^*) + \widehat{V}^2(x^*)\Big)\mathbb{1}_{\Omega_n^c}\Big).
\end{aligned} \tag{A.7}$$

By Lemma 14, we have for the first term

$$\Big\{\mathbb{E}\Big((\widehat{V}(x^*) - V(x^*))\mathbb{1}_{\Omega_n}\Big)\Big\}^2 \leq C\Big(h_1^{4(\alpha\wedge 1)} + h_2^{2\beta} + h_1^{2(\beta\wedge 1)} + \tau_n^2\Big).$$

By Lemma 15 with conditions $nh_2 \to \infty$ and $n^2 h_1 h_2 \to \infty$ satisfied with the choices of (h_1, h_2) in (2.17), we have for the second term

$$\mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n} - \mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n}\Big)\Big)^2 \leq C\big(n^{-1}h_2^{-1} + n^{-2}(h_1 h_2)^{-1} + \tau_n^2\big).$$

Plugging in the values of (h_1, h_2) as in (2.17) and choosing $\tau_n \asymp n^{-\kappa}$ for some fixed $\kappa \geq 1$, we obtain that

$$\mathbb{E}\Big(\widehat{V}(x^*) - V(x^*)\Big)^2 \leq C(n^{-\frac{8\alpha\beta}{4\alpha\beta + 2\alpha + \beta}} + n^{-\frac{2\beta}{2\beta+1}}) + \mathbb{E}\Big(\Big(V^2(x^*) + \widehat{V}^2(x^*)\Big)\mathbb{1}_{\Omega_n^c}\Big).$$

Lastly, note that $\widehat{V}(x^*) = \sum_{i<j} D_{ij} w_{ij} / (\sum_{i<j} w_{ij} + \tau_n)$ is a linear estimator with weight $w_{ij} = \binom{n}{2}^{-1} \boldsymbol{q}^\top(0) \mathbf{B}_n^* \boldsymbol{q}((X_{ij} - x^*)/h_2) K_{ij}$. By definition of $\mathbf{B}_n^*$, w_{ij} is thus a weighted polynomial of $K_{ij}(X_{ij} - x^*)/h_2$ up to some order that only depends on ℓ. Therefore, in view of the choice

of τ_n (decaying to 0 polynomially with n), h_1, h_2, there exists some sufficiently large constant η (only depending on α, β, κ) such that

$$\mathbb{E}\Big(\mathbb{E}\Big(\widehat{V}^2(x^*) \mid \{X_i\}_{i=1}^n\Big)\Big)^2 \lesssim n^{\eta},$$

and thus by Cauchy-Schwarz and the exponential inequality in Lemma 17, it holds that

$$\begin{aligned}
\mathbb{E}\Big(\Big(V^2(x^*) + \widehat{V}^2(x^*)\Big)\mathbb{1}_{\Omega_n^c}\Big) &= \mathbb{E}\Big((\mathbb{E}\widehat{V}^2(x^*) \mid \{X_i\}_{i=1}^n + V^2(x^*))\mathbb{1}_{\Omega_n^c}\Big) \\
&\lesssim (\mathbb{E}(\mathbb{E}\widehat{V}^2(x^*) \mid \{X_i\}_{i=1}^n)^2 + V^4(x^*))^{1/2}\mathbb{P}^{1/2}(\Omega_n^c) \\
&= o\Big(n^{-\frac{8\alpha\beta}{4\alpha\beta+2\alpha+\beta}} + n^{-\frac{2\beta}{2\beta+1}}\Big).
\end{aligned}$$

This completes the proof. □

A.3.2 Proof of Theorem 4

Proof. Note that the boundary of $n^{-8\alpha\beta/(4\alpha\beta+\beta+2\alpha)}$ and $n^{-2\beta/(2\beta+1)}$ lies at $\alpha = \beta/(4\beta + 2)$. When $\alpha \geq \beta/(4\beta+2)$, the statement can be proved using a slight variation of the proof of Theorem 4.2 in Brown and Levine (2007), and we omit the details here. Next, we will focus on the case where $\alpha < \beta/(4\beta + 2)$. Consider a fixed point $x^* \in \mathrm{supp}(X)$. Throughout the proof, C and c represent two generic positive constants which only depend on $\alpha, \beta, C_{\mathcal{F}}, C_{\mathcal{V}}, C_\sigma, C_0, c_0, C_\varepsilon$ and might have different values at each occurrence, but like in the proof of Theorem 2, let c be always smaller than 1/4. Also, without loss of generality, assume that the sample size n and $C_{\mathcal{F}}, C_{\mathcal{V}}, C_\sigma, C_\varepsilon, C_0$ are sufficiently large, c_0 is sufficiently small, and $[0, 1] \subset I$.

We will make use of Le Cam's two point method. Introduce the constants

$$\theta_n^2 := h_1^{2\alpha} := h_2^{\beta} := cn^{-\frac{4\alpha\beta}{4\alpha\beta+\beta+2\alpha}}, \quad M := h_2/(4h_1) - 1/2, \quad N := 2M + 1 = h_2/(2h_1), \quad \text{(A.8)}$$

where we tune the constant c in θ_n^2 so that M is a positive integer. Note that under the above choice, $h_2/h_1 \to \infty$ as $n \to \infty$. We now specify $f(\cdot), V(\cdot)$, distribution of X and distribution of ε in the null and alternative hypotheses, H_0 and H_1, respectively.

Choice of ε: Under both H_0 and H_1, let $\varepsilon \sim \mathcal{N}(0,1)$.

Choice of $V(\cdot)$: Under H_0, let $V \equiv 1$. Under H_1, let $V = 1 - \theta_n^2 H((x - x^*)/h_2)$, where $H(\cdot)$ is β-Hölder smooth, infinitely differentiable, compactly supported on $[-2,2]$, and takes value 1 on $[-1,1]$.

Choice of $f(\cdot)$: Under H_0, let $f \equiv 0$. Under H_1, let f be zero outside $[x^* - h_2, x^* + h_2]$, and inside this interval, the linear interpolant of the function that takes value r_i on $[x^* - h_2 + (4i-3)h_1, x^* - h_2 + (4i-1)h_1]$ and zero at $x^* - h_2 + 4(i-1)h_1$ for all $i \in [N]$, where $\{r_i\}_{i=1}^N$ is an i.i.d. sequence of symmetric and compactly supported random variables with distribution $\mathbb{G}$ satisfying

$$\int_{-\infty}^{\infty} x^j \mathbb{G}(dx) = \int_{-\infty}^{\infty} x^j \varphi(x) dx, \quad j = 1, \ldots, q,$$

where q is some fixed odd integer strictly larger than $1 + (\beta + 2\alpha)/(2\alpha\beta)$.

Choice of X: Under both H_0 and H_1, let X be uniformly distributed on the union of the intervals

$$[0,1] \bigcap \left([0, x^* - 2h_2] \bigcup [x^* + 2h_2, 1] \bigcup_{i=1}^{N} [x^* - h_2 + (4i-3)h_1, x^* - h_2 + (4i-1)h_1] \right).$$

See Figure 2.2 for an illustration. We now make a few remarks about the above construction. For the design of $V(\cdot)$ under H_1, one example of the smooth bump function $H(\cdot)$ is $(\mathbb{1}_{[-3/2,3/2]} * \varphi_{1/2})(\cdot)$, where $\varphi_\varepsilon(x) := \varphi(x/\varepsilon)/\varepsilon$ with $\varphi(x) := \exp(-1/(1-x^2))\mathbb{1}\{|x| \leq 1\}$ being a smooth and compactly supported mollifier. The design of $f(\cdot)$ under H_1 is a "localized" version of $f(\cdot)$ in the proof of Theorem 2. The existence of $\{r_i\}_{i=1}^N$ is again guaranteed by Lemma 5, and their range, which we denote as B, only depends on α and β and is thus fixed. Lastly, we indeed have $x^* \in \operatorname{supp}(X)$ since it is in the $(M+1)$th interval in the N intervals specified in the support of X. Moreover, under H_1, conditioning on the event that $X_i \in [x^* - h_2, x^* + h_2]$ and any realization of $\{r_i\}_{i=1}^N$, $f(X_i)$ is uniformly distributed over $\{h_1^\alpha r_1, \ldots, h_1^\alpha r_N\}$.

Clearly, under both the null and the alternative hypotheses, $V(\cdot)$ is β-Hölder smooth, and under H_1, $f(\cdot)$ is α-Hölder smooth for each realization of $\{r_i\}_{i=1}^N$ due to their compact support. Next, we show that the joint distribution of (X, ε) satisfies the three conditions in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$. Condition (d) clearly holds and Condition (a) holds with $I = [0,1]$. Condition (b) holds as well since for any u in the support of X, $p_X(u) = 1/(1-3h_2)$ for $x^* \in (0,1)$ and $p_X(u) = 2/(2-3h_2)$ for $x^* \in \{0,1\}$, both of which are smaller than 2 for sufficiently large n. Lastly, for Condition (c), the first part clearly holds since $\inf_{u \in \mathrm{supp}(X)} p_X(u) \geq 1$. For the second part, define $\mathcal{A}_{x,\delta} := \{u \in [-1,1] : x + \delta u \in \mathrm{supp}(X)\}$. Then, for any $x^* \in (0,1)$ and any $0 < \delta \leq 1/2$, we have $\lambda(\mathcal{A}_{x,\delta}) \geq 1/2$ if $x \in (0, x^* - 2h_2] \bigcup [x^* + 2h_2, 1)$ and $\lambda(\mathcal{A}_{x,\delta}) \geq 1/4$ if $x \in \bigcup_{i=1}^N [x^* - h_2 + (4i-3)h_1, x^* - h_2 + (4i-1)h_1]$. A similar statement holds for $x^* \in \{0,1\}$. We therefore conclude that Condition (c) also holds.

Denote $\mathbb{P}_0$ and $\mathbb{P}_1$ to be the joint distributions of $\{X_i, Y_i\}_{i=1}^n$ under H_0 and H_1, then the pointwise squared distance between $\mathbb{P}_0$ and $\mathbb{P}_1$ $(V_0(x^*) - V_1(x^*))^2 \asymp \theta_n^4$ is the desired minimax rate. Further define $\widetilde{\mathbb{P}}_1$ as the corresponding joint distributions of $\{X_i, Y_i\}_{i=1}^n$ under H_1 with $\{r_i\}_{i=1}^N$ replaced by an i.i.d. standard normal sequence $\{\widetilde{r}_i\}_{i=1}^N$. Then, following the same line of proof of Theorem 2, it suffices to show that $\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) \leq c$ and $\mathsf{TV}(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) \leq c$.

For the first inequality, in view of (A.3) in the proof of Theorem 2, it suffices to upper bound $\mathsf{TV}(\mathbb{P}_0(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x}))$ for each realization $\{x_i\}_{i=1}^n$. Note that under $\mathbb{P}_0$, $\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}_n(0, \boldsymbol{\Sigma_0})$, with $\boldsymbol{\Sigma_0} = \mathbf{I}_n$. Denote $\{b_i\}_{i=1}^n$ as the location index sequence of $\{X_i\}_{i=1}^n$ taking values in $\{0, 1, \ldots, N\}$, that is, $b_i = 0$ if $X_i \notin [x^* - h_2, x^* + h_2]$ and $b_i = j$ if $X_i \in [x^* - h_2 + (4j-3)h_1, x^* - h_2 + (4j-1)h_1]$ for $j \in [N]$. Then, due to the symmetry of $\{r_i\}_{i=1}^N$, the design of the nonparametric component f, and the fact that $H(\cdot)$ takes value 1 on $[-1,1]$, it holds that under $\widetilde{\mathbb{P}}_1$, $\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}_n(0, \boldsymbol{\Sigma_1})$, with

$$(\boldsymbol{\Sigma_1})_{ii} = 1 - \theta_n^2 \mathbb{1}\{|*|X_i - x^* \leq h_2\} + h_1^{2\alpha} \mathbb{1}\{|*|X_i - x^* \leq h_2\} = 1$$

and $(\boldsymbol{\Sigma_1})_{ij} = h_1^{2\alpha} \mathbb{1}\{b_i = b_j, b_i \geq 1, b_j \geq 1\}$ for $i \neq j$. Define $N_0 := \sum_{i \neq j} \mathbb{1}\{b_i = b_j, b_i \geq$

$1, b_j \geq 1\}$. Then, we have by Lemma 6 that

$$\mathsf{TV}(\mathbb{P}_0(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x})) \leq C\big(h_1^{4\alpha} N_0\big)^{1/2} = C\theta_n^2 N_0^{1/2}.$$

Note that N_0 is a random variable that depends on $\{X_i\}_{i=1}^n$, and by (A.3) in the proof of Theorem 2,

$$\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) \leq C\theta_n^2 \mathbb{E}(N_0^{1/2}) \leq C\theta_n^2 (\mathbb{E}N_0)^{1/2}.$$

Since direct calculation implies that $\mathbb{E}(N_0) \leq Cn^2 N h_1^2 = Cn^2 h_1 h_2$, we have $\mathsf{TV}(\mathbb{P}_0, \widetilde{\mathbb{P}}_1) \lesssim 1$ under the given choice of h_1, h_2 and θ_n.

Using a conditioning argument, the second part of proving $\mathsf{TV}(\widetilde{\mathbb{P}}_1, \mathbb{P}_1) \lesssim 1$ follows similarly from that of Theorem 2 by noting that $n^2 h_1 \to \infty$ and $nh_1 \to 0$ as $n \to \infty$ under the constraint $\alpha < \beta/(4\beta + 2)$ so that Lemma 7 can be similarly applied. The proof is complete. □

A.3.3 Proof of Theorem 5

Proof. As in the proof of Theorem 4, we focus on the regime $\alpha < \beta/(4\beta+2)$. We will couple the proof of Theorem 4 with a standard technique via multiple hypotheses in the classic setting of mean function estimation.

Introduce the following notation:

$$\theta_n^2 := h_1^{2\alpha} := h_2^{\beta} := cn^{-\frac{4\alpha\beta}{4\alpha\beta+\beta+2\alpha}}, \quad N_2 := 1/(4h_2), \quad N_1 := h_2/(2h_1),$$
$$\text{and} \quad x_i^* := 2h_2 + (i-1)4h_2, \quad i \in [N_2],$$

where we tune the constant c in θ_n^2 so that N_1 and N_2 are both positive integers. Note that under the above choice, $h_2/h_1 \to \infty$ as $n \to \infty$. By the renowned Varshamov-Gilbert bound (cf. Lemma 2.8 in Tsybakov (2009a)), there exists a set of length-N_2 binary sequences $\{\boldsymbol{\Delta}_j\}_{j=0}^M$ with $M \geq 2^{N_2/8}$ such that $\boldsymbol{\Delta}_0 = \mathbf{0}_{N_2}$ and for any $0 \leq k < \ell \leq M$, it holds that

$\rho(\mathbf{\Delta}_k, \mathbf{\Delta}_\ell) \geq N_2/8$, where ρ is the Hamming distance. We now choose a number of $M+1$ hypotheses with $\{\mathbf{\Delta}_j\}_{j=0}^M$ satisfying the above property, which we denote as $\mathbb{P}_0, \mathbb{P}_1, \ldots, \mathbb{P}_M$. We now specify $f(\cdot), V(\cdot)$, distribution of X and distribution of ε under each hypothesis.

Choice of ε: Under $\mathbb{P}_0$ and $\mathbb{P}_j$ for all $j \in [M]$, let $\varepsilon \sim \mathcal{N}(0,1)$.

Choice of $V(\cdot)$: Under $\mathbb{P}_0$, let $V_0 \equiv 1$. Under $\mathbb{P}_j$, let $V_j(x) := 1 - \sum_{i=1}^{N_2} \Delta_{j,i} \theta_n^2 H((x - x_i^*)/h_2)$, where $H(\cdot)$ is infinitely differentiable, compactly supported on $[-2, 2]$ and takes value 1 on $[-1, 1]$.

Choice of $f(\cdot)$: Under $\mathbb{P}_0$, let $f_0 \equiv 0$. Under $\mathbb{P}_j$, for all $i \in [N_2]$ such that $\Delta_{j,i} = 1$, let f be the linear interpolation of the function that takes value $r_{i,k}^{(j)}$ on the interval $[x_i^* - h_2 + (4k-3)h_1, x_i^* - h_2 + (4k-1)h_1]$ and value zero at $x_i^* - h_2 + 4(k-1)h_1$ for all $k \in [N_1]$, where by denoting $m_j := \|\Delta_j\|_0$, $\{r_{i,k}^{(j)}\}_{j\in[M], i\in[m_j], k\in[N_1]}$ is an i.i.d. sequence of symmetric and compactly supported random variables with distribution $\mathbb{G}$ satisfying

$$\int_{-\infty}^{\infty} x^j \mathbb{G}(dx) = \int_{-\infty}^{\infty} x^j \varphi(x) dx, \quad j = 1, \ldots, q,$$

where q is some fixed odd integer that only depends on α and β.

Choice of X: Under $\mathbb{P}_0$ and $\mathbb{P}_j$ for all $j \in [M]$, let X be uniformly distributed on the union of the disjoint intervals

$$\bigcup_{i=1}^{N_2} \bigcup_{k=1}^{N_1} [x_i^* - h_2 + (4k-3)h_1, x_i^* - h_2 + (4k-1)h_1].$$

The existence of $H(\cdot)$ in the design of $V(\cdot)$ and variables $\{r_{i,k}^{(j)}\}_{j\in[M], i\in[m_j], k\in[N_1]}$ is as argued in the proof of Theorem 4. Moreover, one can readily check that for each $0 \leq k < \ell \leq M$, the integrated squared distance between each $\mathbb{P}_k$ and $\mathbb{P}_\ell$ satisfies

$$d(\mathbb{P}_k, \mathbb{P}_\ell) := \int (V_k(x) - V_\ell(x))^2 p_X(x) dx \gtrsim h_2^{2\beta} \asymp n^{-\frac{8\alpha\beta}{4\alpha\beta + \beta + 2\alpha}},$$

which is the desired lower bound.

Clearly, under each $\mathbb{P}_j$, $0 \leq j \leq M$ and for each realization of $\{r_{i,k}^{(j)}\}_{j\in[M],i\in[m_j],k\in[N_1]}$, $f_j(\cdot)$ and $V_j(\cdot)$ are α- and β-Hölder smooth, respectively, due to the compact support of $\{r_{i,k}^{(j)}\}_{j\in[M],i\in[m_j],k\in[N_1]}$. Moreover, the joint distribution of (X, ε) (same in all hypothese) satisfies the conditions in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ with a similar argument as in Theorem 4.

We now proceed with the proof. Note that under the above design, the support of X is segmented into $N_3 := N_1 \times N_2$ intervals, and we let $\{b_i\}_{i=1}^n$ be the location index of $\{X_i\}_{i=1}^n$, taking values in $[N_2]\times[N_1]$, that is, $b_i = (k, \ell)$ if $X_i \in [x_k^* - h_2 + (4\ell-3)h_1, x_k^* - h_2 + (4\ell-1)h_1]$. As in the proof of Theorem 2, define the event $\Omega_n := \big\{\max_{(k,\ell)\in\mathbb{P}[N_2]\times[N_1]} \#\{b_i = (k,\ell)\} \leq K\big\}$, where K is the smallest integer strictly larger than $2\beta/(\beta - 4\alpha\beta - 2\alpha)$. Then, by Lemma 7, it holds that Ω_n has asymptotic probability 1 under all of $\mathbb{P}_j$ and $\widetilde{\mathbb{P}}_j$ for $0 \leq j \leq M$. Now, by a standard reduction scheme with multiple hypotheses (cf. Chapter 2.2 in Tsybakov (2009a)) and Lemma 19, it suffices to show that

$$\frac{1}{M}\sum_{j=1}^{M} K(\mathbb{P}_j, \mathbb{P}_0; \Omega_n) \leq c\log(M) \tag{A.9}$$

for some $0 < c < 1/8$, where $K(\mathbb{P}, \mathbb{Q}; \Omega_n)$ is the "conditional" Kullback divergence between probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined as $K(\mathbb{P}, \mathbb{Q}; \mathcal{E}) := \int_{\mathcal{E}} \log(d\mathbb{P}/d\mathbb{Q})d\mathbb{P}$ for any measurable set $\mathcal{E}$. In order to show (A.9), it further suffices to show that $K(\mathbb{P}_j, \mathbb{P}_0; \Omega_n) \leq \log(M)$ for all $j \in [M]$. We now focus on a particular $j \in [M]$. For notational brevity, we will drop the superscript (j) in the sequence of variables $\{r_{i,k}^{(j)}\}_{i\in[m_j],k\in[N_1]}$ for this particular j. Note that $N_2/8 \leq m_j \leq N_2$ by the property that $\rho(\mathbf{\Delta}_0, \mathbf{\Delta}_j) \geq N_2/8$. Moreover, by the design of f, there are a total of $m_j N_1$ trapezoids in the union of the intervals $[x_i^* - h_2, x_i^* + h_2]$ for those i such that $\Delta_{j,i} = 1$. Define $\widetilde{\mathbb{P}}_j$ as the joint distribution of $\{(X_i, Y_i)\}_{i=1}^n$ under $\mathbb{P}_j$ but with $\{r_{i,k}\}_{i\in[m_j],k\in[N_1]}$ replaced by a sequence of i.i.d. standard normal variables denoted as

$\{\widetilde{r}_{i,k}\}_{i\in[m_j],k\in[N_1]}$. By definition, it holds that

$$\begin{aligned}
K(\mathbb{P}_j, \mathbb{P}_0; \Omega_n) &= \int_{\Omega_n} p_j \log \frac{p_j}{p_0} \\
&= \int_{\Omega_n} p_j \log \frac{p_j}{\widetilde{p}_j} + \int_{\Omega_n} \widetilde{p}_j \log \frac{\widetilde{p}_j}{p_0} + \int_{\Omega_n} (p_j - \widetilde{p}_j) \log \frac{\widetilde{p}_j}{p_0} \\
&= K(\mathbb{P}_j, \widetilde{\mathbb{P}}_j; \Omega_n) + K(\widetilde{\mathbb{P}}_j, \mathbb{P}_0; \Omega_n) + \int_{\Omega_n} (p_j - \widetilde{p}_j) \log \frac{\widetilde{p}_j}{p_0}
\end{aligned}$$

for density functions with respect to some common dominating measure. Next, we will show respectively that, by matching the moments of $\{r_{i,k}\}_{i\in[m_j],k\in[N_1]}$ and the standard Gaussian random variable up to some sufficiently high order, it holds that

$$K(\mathbb{P}_j, \widetilde{\mathbb{P}}_j; \Omega_n) \lesssim 1, \quad K(\widetilde{\mathbb{P}}_j, \mathbb{P}_0; \Omega_n) \lesssim \log(M), \quad \text{and} \quad \int_{\Omega_n} (p_j - \widetilde{p}_j) \log(\widetilde{p}_j / p_0) \lesssim 1.$$

First note that, by denoting $\boldsymbol{x} := (x_1, \ldots, x_n)$, $d\boldsymbol{x} := dx_1 \ldots dx_n$ and similarly for $\boldsymbol{y}$ and $d\boldsymbol{y}$, we have

$$\begin{aligned}
K(\mathbb{P}_j, \widetilde{\mathbb{P}}_j; \Omega_n) &= \int \mathbb{1}\{\Omega_n\} p(\boldsymbol{x}) d\boldsymbol{x} \int \log\left(\frac{d\mathbb{P}_j(\boldsymbol{y} \mid \boldsymbol{x})}{d\widetilde{\mathbb{P}}_j(\boldsymbol{y} \mid \boldsymbol{x})}\right) \mathbb{P}_j(d\boldsymbol{y} \mid \boldsymbol{x}) \\
&= \mathbb{E}\Big\{\mathbb{1}\{\Omega_n\} K\Big(\mathbb{P}_j(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_j(\boldsymbol{y} \mid \boldsymbol{x})\Big)\Big\} \\
&\le \mathbb{E}\Big\{\mathbb{1}\{\Omega_n\} \chi^2\Big(\mathbb{P}_j(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_j(\boldsymbol{y} \mid \boldsymbol{x})\Big)\Big\} \\
&:= \chi^2\Big(\mathbb{P}_j, \widetilde{\mathbb{P}}_j; \Omega_n\Big),
\end{aligned}$$

where the inequality follows by Lemma 2.7 in Tsybakov (2009a). Therefore the first inequality $K(\mathbb{P}_j, \widetilde{\mathbb{P}}_j; \Omega_n) \lesssim 1$ holds by Lemma 18.

Next we prove $K(\widetilde{\mathbb{P}}_j, \mathbb{P}_0; \Omega_n) \lesssim \log(M)$. Again, it suffices to prove that for any realization of $\{X_i\}_{i=1}^n$ in Ω_n, it holds that $K(\widetilde{\mathbb{P}}_j(\boldsymbol{y} \mid \boldsymbol{x}), \mathbb{P}_0(\boldsymbol{y} \mid \boldsymbol{x})) \lesssim \log(M) \asymp N_2$. Note that under $\mathbb{P}_0$, $\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}_n(0, \boldsymbol{\Sigma_0})$, with $\boldsymbol{\Sigma_0} = \mathbf{I}_n$. Recall that the location index sequence $\{b_i\}_{i=1}^n = \{(k_i, \ell_i)\}_{i=1}^n$ takes the value $(k_i, \ell_i) = (k, \ell)$ if $X_i \in [x_k^* - h_2 + (4\ell - 3)h_1, x_k^* - h_2 + (4\ell - 1)h_1]$. Then, due to the symmetry of $\{r_{i,k}\}_{i\in[m_j],k\in[N_1]}$, the design of the nonparametric component

f, and the fact that $K(\cdot)$ takes value 1 on $[-1,1]$, it holds that under $\widetilde{\mathbb{P}}_j$, $\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}_n(0, \boldsymbol{\Sigma_1})$, with

$$(\boldsymbol{\Sigma_1})_{ii} = 1 - \theta_n^2 \mathbb{1}\{\Delta_{j,k_i} = 1\} + h_1^{2\alpha} \mathbb{1}\{\Delta_{j,k_i} = 1\} = 1$$

and $(\boldsymbol{\Sigma_1})_{i_1 i_2} = h_1^{2\alpha} \mathbb{1}\big\{\Delta_{j,k_{i_1}} = 1, (k_{i_1}, \ell_{i_1}) = (k_{i_2}, \ell_{i_2})\big\}$ for $i_1 \neq i_2$. Define

$$N_0 := \sum_{i_1 \neq i_2} \mathbb{1}\big\{\Delta_{j,k_{i_1}} = 1, (k_{i_1}, \ell_{i_1}) = (k_{i_2}, \ell_{i_2})\big\}.$$

Then, by the proof of Lemma 3.6 in Gao and Zhou (2016), it holds that

$$K(\widetilde{\mathbb{P}}_j(\boldsymbol{y} \mid \boldsymbol{x}), \mathbb{P}_0(\boldsymbol{y} \mid \boldsymbol{x})) \leq C h_1^{4\alpha} N_0 = C\theta_n^4 N_0.$$

Note that N_0 is a random variable that depends on $\{X_i\}_{i=1}^n$, and by direct calculation we have

$$\mathbb{E}(N_0) \leq n^2 m_j h_2 h_1 \asymp n^2 N_2 h_1 h_2.$$

Putting together the pieces, we have $K(\widetilde{\mathbb{P}}_j, \mathbb{P}_0; \Omega_n) \leq \theta_n^4 n^2 h_1 h_2 N_2 \lesssim N_2$. This completes the proof of the second inequality.

Lastly, we show that $\int_{\Omega_n} (p_j - \widetilde{p}_j) \log(\widetilde{p}_j / p_0) \lesssim 1$. First note that

$$\begin{aligned}
&\int_{\Omega_n} (p_j - \widetilde{p}_j) \log(\widetilde{p}_j / p_0) \\
&\leq \int_{\Omega_n} |*| p_j - \widetilde{p}_j |*| \log(\widetilde{p}_j / p_0) \\
&\leq \left(\int_{\Omega_n} |*| p_j - \widetilde{p}_j \right)^{1/2} \left(\int |*| p_j - \widetilde{p}_j \log^2(\widetilde{p}_j / p_0) \right)^{1/2} \\
&\leq \left(\int_{\Omega_n} |*| p_j - \widetilde{p}_j \right)^{1/2} \left\{ \left(\int p_j \log^2(\widetilde{p}_j / p_0) \right)^{1/2} + \left(\int \widetilde{p}_j \log^2(\widetilde{p}_j / p_0) \right)^{1/2} \right\}.
\end{aligned}$$

By Lemmas 21 and 18, by matching moments up to some sufficiently high order, the first term above can be upper bounded (up to some constant) by $n^{-\eta}$ for any $\eta > 0$, therefore it suffices to show that both $\int p_j \log^2(\widetilde{p}_j/p_0)$ and $\int \widetilde{p}_j \log^2(\widetilde{p}_j/p_0)$ can be upper bounded by some polynomial of n of fixed order. Consider any realization of $\{X_i\}_{i=1}^n$ in Ω_n, and assume that based on their location indices $\{b_i\}_{i=1}^n$, the n data points are partitioned into $L_1 + L_2$ clusters with cardinality s_ℓ such that the X_i's in the same cluster have the same value b_i. Moreover, for each data point in the first L_1 clusters, the location index $b_i = (k_i, \ell_i)$ satisfies that $\Delta_{j,k_i} = 1$ while for the data points in the last L_2 clusters, it holds that $\Delta_{j,k_i} = 0$. Apparently, we have the relations $1 \leq L_1 + L_2 \leq n$, $\sum_{\ell=1}^{L_1+L_2} s_\ell = n$ and $1 \leq s_\ell \leq K$ for $\ell \in [L_1 + L_2]$. Moreover, denoting $\widetilde{\mathbb{P}}_{j,\pi_\ell}$ and $\mathbb{P}_{0,\pi_\ell}$ (resp. $\widetilde{p}_{j,\pi_\ell}$ and p_{0,π_ℓ}) for each $\ell \in [L_1 + L_2]$ as the joint distribution (resp. density) of those Y_i's in the ℓth cluster conditioning on the given realization $\{X_i\}_{i=1}^n$ under $\widetilde{\mathbb{P}}_j$ and $\mathbb{P}_0$, we have

$$\widetilde{p}_j = \prod_{\ell=1}^{L_1+L_2} \widetilde{p}_{j,\pi_\ell} \quad \text{and} \quad p_0 = \prod_{\ell=1}^{L_1+L_2} p_{0,\pi_\ell}.$$

Moreover, for any $L_1 + 1 \leq \ell \leq L_1 + L_2$, it holds that $\widetilde{p}_{j,\pi_\ell} = p_{0,\pi_\ell}$, therefore it holds that

$$\log^2\left(\frac{\widetilde{p}_j}{p_0}\right) = \left(\sum_{\ell=1}^{L_1} \log(\widetilde{p}_{j,\pi_\ell}) - \log(p_{0,\pi_\ell})\right)^2 \lesssim n \sum_{\ell=1}^{L_1} \log^2(\widetilde{p}_{j,\pi_\ell}/p_{0,\pi_\ell}).$$

Now consider any $\ell \in [L_1]$ and assume that $s_\ell = d$ for some positive integer d. Without loss of generality, assume the y_i's in this cluster are $\{y_1, \ldots, y_d\}$, and they take the form $Y_i = h_1^\alpha \widetilde{r}_{1,1} + (1 - h_2^\beta)^{1/2}\varepsilon_i = \theta_n \widetilde{r}_{1,1} + (1 - h_2^\beta)^{1/2}\varepsilon_i$ under $\widetilde{\mathbb{P}}_j$ and $Y_i = \varepsilon_i$ under $\mathbb{P}_0$. Define $\sigma^2 := (1 - h_2^\beta)$ which is positive for large enough n. Then, the previous equalities imply that

$$p_{0,\pi_\ell} = \varphi(y_1)\ldots\varphi(y_d) = (2\pi)^{-d/2}\exp\left(-\frac{\sum_{i=1}^d y_i^2}{2}\right)$$

and

$$\begin{aligned}
\widetilde{p}_{j,\pi_\ell} &= \int \frac{1}{\sigma}\varphi\left(\frac{y_1 - \theta_n v}{\sigma}\right)\cdots\frac{1}{\sigma}\varphi\left(\frac{y_d - \theta_n v}{\sigma}\right)\varphi(v)dv \\
&= (2\pi)^{-d/2}\frac{1}{\sigma^{d-1}(d\theta_n^2 + \sigma^2)^{1/2}}\exp\left(-\frac{\sum_{i=1}^d y_i^2}{2\sigma^2} + \frac{(\sum_{i=1}^d y_i\theta_n)^2}{2\sigma^2(d\theta_n^2 + \sigma^2)^2}\right).
\end{aligned}$$

Putting together the pieces, we obtain that

$$\log^2(\widetilde{p}_{j,\pi_\ell}/p_{0,\pi_\ell}) \lesssim d^2\log^2(1/\sigma) + \left(\sum_{i=1}^d y_i^2\right)^2 + \left(\sum_{i=1}^d y_i\theta_n\right)^4 \lesssim 1 + \sum_{i=1}^d y_i^4.$$

Therefore we have

$$\int p_j \log^2(\widetilde{p}_j/p_0) \lesssim n\sum_{\ell=1}^{L_1}\int p_j(1+\sum_{i=1}^d y_i^4) \lesssim n\sum_{\ell=1}^{L_1}\sum_{i=1}^d \int y_i^4 \mathbb{P}_j(dy_i) \lesssim n^2,$$

where we use the fact that $L_1 \leq n$. Similarly, we have $\int \widetilde{p}_j \log^2(\widetilde{p}_j/p_0) \lesssim n^2$. The proof is thus complete. $\square$

A.3.4 Supporting lemmas

Lemma 14. *Suppose $f \in \Lambda_\alpha(C_{\mathcal{F}})$, $V \in \Lambda_\beta(C_{\mathcal{V}})$, $\sigma^2 \leq C_\sigma$ for some fixed constants $C_{\mathcal{F}}, C_{\mathcal{V}}, C_\sigma$, and the joint distribution of (X,ε) belongs to $\mathcal{P}_{vf,(X,\varepsilon)}$. Then, with Ω_n defined in the proof of Theorem 3, it holds that*

$$|*|\mathbb{E}\Big\{\Big(\widehat{V}(x^*) - V(x^*)\Big)\mathbb{1}_{\Omega_n}\Big\} \leq C\Big(h_1^{2(\alpha\wedge 1)} + h_2^\beta + h_1^{\beta\wedge 1} + \tau_n\Big)$$

for some fixed positive constant C that only depends on $\alpha, \beta, C_{\mathcal{F}}, C_{\mathcal{V}}, C_\sigma, C_0, C_\varepsilon$.

Proof. We adopt the notation $\ell, \boldsymbol{q}(\cdot), \mathbf{B}_n, X_{ij}$, and K_{ij} from the proof of Theorem 3. Also recall the definition of $\mathbf{B}_n^*$, D_{ij}, w_{ij}, and $\widetilde{w}_{ij}$ from the definition of $\widehat{V}(x^*)$. Writing $\mathbb{E}_\varepsilon$ as the conditional expectation given $\{X_i\}_{i=1}^n$, it holds that

$$|*|\mathbb{E}\Big\{\Big(\widehat{V}(x^*) - V(x^*)\Big)\mathbb{1}_{\Omega_n}\Big\} = |*|\mathbb{E}\Big\{\mathbb{1}_{\Omega_n}\mathbb{E}_\varepsilon\Big(\widehat{V}(x^*) - V(x^*)\Big)\Big\}.$$

Then $\widehat{V}(x^*) = \sum_{i<j}\widetilde{w}_{ij}D_{ij}$ and

$$\begin{aligned} &|*|\mathbb{E}\Big\{\mathbb{1}_{\Omega_n}\mathbb{E}_\varepsilon\Big(\widehat{V}(x^*) - V(x^*)\Big)\Big\} \\ \leq\ & |*|\mathbb{E}\left\{\mathbb{1}_{\Omega_n}\sum_{i<j}\widetilde{w}_{ij}(\mathbb{E}_\varepsilon D_{ij} - V(x^*))\right\} + V(x^*)\tau_n|*|\mathbb{E}\big\{\mathbb{1}_{\Omega_n}(|\mathbf{B}_n| + \tau_n)^{-1}\big\}. \end{aligned}$$

By definition, it holds on Ω_n that $\|\mathbf{B}_n - \mathbf{B}\| \leq \lambda_{\min}(\mathbf{B})/2$, where $\lambda_{\min}(\mathbf{B})$ is the smallest eigenvalue of $\mathbf{B}$. Thus by Weyl's inequality, it holds that $\lambda_{\min}(\mathbf{B}_n) \geq \lambda_{\min}(\mathbf{B})/2 \geq c$ for some fixed positive constant c due to the invertibility of $\mathbf{B}$ as proved in Theorem 3 under Condition (c) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$. Then, using the fact that $|\mathbf{B}_n| \geq \lambda_{\min}^{\ell+1}(\mathbf{B}_n)$ and the boundedness of $V(\cdot)$, it holds that

$$|*|\mathbb{E}\Big\{\mathbb{1}_{\Omega_n}\mathbb{E}_\varepsilon\Big(\widehat{V}(x^*) - V(x^*)\Big)\Big\} \leq |*|\mathbb{E}\left\{\mathbb{1}_{\Omega_n}\sum_{i<j}\widetilde{w}_{ij}(\mathbb{E}_\varepsilon D_{ij} - V(x^*))\right\} + C\tau_n.$$

Direct calculation shows that $\mathbb{E}_\varepsilon D_{ij} - V(x^*) = (f(X_i) - f(X_j))^2/2 + (V(X_i) - V(x^*))/2 + (V(X_j) - V(x^*))/2$. Due to symmetry, we only need to control the first two terms. For the first term, using the fact $\mathbf{B}_n^* = |\mathbf{B}_n|\mathbf{B}_n^{-1}$ on Ω_n ($\mathbf{B}_n$ invertible on Ω_n), we have

$$\begin{aligned}
&\mathbb{E}\left\{\mathbb{1}_{\Omega_n}\sum_{i<j}(|\mathbf{B}_n| + \tau_n)^{-1}w_{ij}(f(X_i) - f(X_j))^2\right\} \\
&= \mathbb{E}\left\{\mathbb{1}_{\Omega_n}\binom{n}{2}^{-1}\sum_{i<j}(|\mathbf{B}_n| + \tau_n)^{-1}\boldsymbol{q}^\top(0)\mathbf{B}_n^*\boldsymbol{q}\left(\frac{X_{ij} - x^*}{h_2}\right)K_{ij}(f(X_i) - f(X_j))^2\right\} \\
&= \mathbb{E}\left\{(|\mathbf{B}_n| + \tau_n)^{-1}\boldsymbol{q}^\top(0)\mathbf{B}_n^*\boldsymbol{q}\left(\frac{X_{ij} - x^*}{h_2}\right)K_{ij}(f(X_i) - f(X_j))^2\mathbb{1}_{\Omega_n}\right\} \\
&\lesssim h_1^{2(\alpha\wedge 1)}\mathbb{E}\big\{(|\mathbf{B}_n| + \tau_n)^{-1}\|\mathbf{B}_n^*\|K_{ij}\mathbb{1}_{\Omega_n}\big\} = h_1^{2(\alpha\wedge 1)}\mathbb{E}\left\{\frac{|\mathbf{B}_n|}{|\mathbf{B}_n| + \tau_n}\|\mathbf{B}_n^{-1}\|K_{ij}\mathbb{1}_{\Omega_n}\right\} \\
&\lesssim Ch_1^{2(\alpha\wedge 1)}\mathbb{E}K_{ij} \lesssim Ch_1^{2(\alpha\wedge 1)}.
\end{aligned}$$

Here, the second inequality follows from the fact that $\|\mathbf{B}_n^{-1}\| = \lambda_{\min}(\mathbf{B}_n)^{-1}$ is bounded by some fixed constant on Ω_n, and in the last inequality we use the fact that $\mathbb{E}K_{ij}$ is bounded Condition (b) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$. For the second term, it holds that

$$\begin{aligned}
&|*|\mathbb{E}(\mathbb{1}_{\Omega_n}\sum_{i<j}\widetilde{w}_{ij}(V(X_i) - V(x^*))) \\
&\leq |*|\mathbb{E}(\mathbb{1}_{\Omega_n}\sum_{i<j}\widetilde{w}_{ij}(V(X_{ij}) - V(x^*))) + |*|\mathbb{E}(\mathbb{1}_{\Omega_n}\sum_{i<j}\widetilde{w}_{ij}(V(X_i) - V(X_{ij}))) := I + II.
\end{aligned}$$

For I, using the Hölder property of $V(\cdot)$ and the reproducing property of local polynomial estimators (see (2.16) in the main paper), that is,

$$\sum_{i<j} w_{ij}(X_{ij}-x^*)^k = \sum_{i<j} \widetilde{w}_{ij}(X_{ij}-x^*)^k = 0, \quad k=1,2,\ldots,\ell,$$

it holds that

$$\begin{aligned}
&|*|\mathbb{E}\Big\{ \mathbb{1}_{\Omega_n} \sum_{i<j} \widetilde{w}_{ij}(V(X_{ij}) - V(x^*)) \Big\} \\
&= |*|\mathbb{E}\Big\{ \mathbb{1}_{\Omega_n} \sum_{i<j} \widetilde{w}_{ij} \Big(\sum_{k=1}^{\ell-1} \frac{V^{(k)}(x^*)}{k!}(X_{ij}-x^*)^k + \frac{V^{(\ell)}(x^*+\tau(X_{ij}-x^*))}{\ell!}(X_{ij}-x^*)^\ell \Big) \Big\} \\
&= |*|\mathbb{E}\Big\{ \mathbb{1}_{\Omega_n} \sum_{i<j} \widetilde{w}_{ij} \frac{(X_{ij}-x^*)^\ell}{\ell!} \big(V^{(\ell)}(x^*+\tau(X_{ij}-x^*)) - V^{(\ell)}(x^*) \big) \Big\} \\
&= |*|\mathbb{E}\Big\{ \mathbb{1}_{\Omega_n} (|\mathbf{B}_n| + \tau_n)^{-1} \boldsymbol{q}^\top(0) \mathbf{B}_n^* \boldsymbol{q}\Big(\frac{X_{ij}-x^*}{h_2} \Big) K_{ij} \frac{(X_{ij}-x^*)^\ell}{\ell!} \big(V^{(\ell)}(x^*+\tau(X_{ij}-x^*)) - V^{(\ell)}(x^*) \big) \Big\} \\
&\lesssim \mathbb{E}\big\{ \mathbb{1}_{\Omega_n} |X_{ij}-x^*|^\beta (|\mathbf{B}_n| + \tau_n)^{-1} \|\mathbf{B}_n^*\| K_{ij} \big\} \\
&\lesssim h_2^\beta \mathbb{E}\big(\mathbb{1}_{\Omega_n} \|\mathbf{B}_n^{-1}\| K_{ij} \big) \lesssim h_2^\beta,
\end{aligned}$$

where in the second line we use the Taylor expansion of $V(\cdot)$ around x^* with some $\tau \in [0,1]$, in the fifth line we use the fact that $\boldsymbol{q}((X_{ij}-x^*)/h_2)$ has bounded ℓ_2 norm due to the compact support of $K(\cdot)$, and in the last inequality we use again the fact that $\|\mathbf{B}_n^{-1}\|$ is bounded by some fixed constant on Ω_n. With a similar calculation and the fact that β-smooth functions are Lipschitz for $\beta \geq 1$, we have $II \lesssim h_1^{\beta\wedge 1}$. Therefore, putting together the pieces, we obtain

$$|*|\mathbb{E}\Big\{ \Big(\widehat{V}(x^*) - V(x^*) \Big) \mathbb{1}_{\Omega_n} \Big\} \leq C\Big(h_1^{2(\alpha\wedge 1)} + h_2^\beta + h_1^{\beta\wedge 1} + \tau_n \Big).$$

This completes the proof. □

Lemma 15. *Suppose $f \in \Lambda_\alpha(C_\mathcal{F})$, $V \in \Lambda_\beta(C_\mathcal{V})$, $\sigma^2 \leq C_\sigma$ for some fixed constants $C_\mathcal{F}, C_\mathcal{V}, C_\sigma$, and the joint distribution of (X,ε) belongs to $\mathcal{P}_{vf,(X,\varepsilon)}$. Assume that $nh_2 \to \infty$ and $n^2 h_1 h_2 \to \infty$ as $n \to \infty$. Then, with Ω_n defined in the proof of Theorem 3, it holds that*

$$Var\Big(\widehat{V}(x^*) \mathbb{1}_{\Omega_n} \Big) \leq C\big(n^{-1} h_2^{-1} + n^{-2}(h_1 h_2)^{-1} + \tau_n^2 \big)$$

for some fixed positive constant C that only depends on $\alpha, \beta, C_{\mathcal{F}}, C_{\mathcal{V}}, C_\sigma, C_0, C_\varepsilon$.

Proof. We adopt the notation $\ell, \boldsymbol{q}(\cdot), \mathbf{B}_n, X_{ij}$, and K_{ij} from the proof of Theorem 3. Also recall the definition of $\mathbf{B}_n^*$ and D_{ij} from the definition of $\widehat{V}(x^*)$. Define the vector-valued U-statistic

$$\boldsymbol{U}_n := \binom{n}{2}^{-1} \boldsymbol{g}(X_i, X_j) := \binom{n}{2}^{-1} \sum_{i<j} D_{ij} \boldsymbol{q}\left(\frac{X_{ij} - x^*}{h_2}\right) K_{ij}$$

and let $\boldsymbol{\theta} := \mathbb{E}\boldsymbol{g}(X_i, X_j) \in \mathbb{R}^{\ell+1}$. Then, for each $j \in [\ell+1]$, writing $\mathbb{E}_\varepsilon$ as the conditional expectation given $\{X_i\}_{i=1}^n$, we have

$$\begin{aligned}
\theta_j &= \mathbb{E}\left(D_{ij} \frac{((X_{ij} - x^*)/h_2)^j}{j!} K_{ij}\right) = \mathbb{E}\left(\frac{((X_{ij} - x^*)/h_2)^j}{j!} K_{ij} \mathbb{E}_\varepsilon D_{ij}\right) \\
&= \mathbb{E}\left(\frac{((X_{ij} - x^*)/h_2)^j}{2j!} K_{ij}((f(X_i) - f(X_j))^2 + V(X_i) + V(X_j))\right) \lesssim \mathbb{E}K_{ij} \lesssim 1,
\end{aligned}$$

where we have used Condition (b) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$, boundedness of $V(\cdot)$ and compact support of $K(\cdot)$. Therefore we have $\|\boldsymbol{\theta}\| = O(1)$. With the above notation, and using the fact that $\mathbf{B}_n$ is invertible and satisfies $\mathbf{B}_n^* = |\mathbf{B}_n| \mathbf{B}_n^{-1}$ on Ω_n, we have

$$\widehat{V}(x^*) \mathbb{1}_{\Omega_n} = \frac{|\mathbf{B}_n|}{|\mathbf{B}_n| + \tau_n} \boldsymbol{q}^\top(0) \mathbf{B}_n^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n} = \boldsymbol{q}^\top(0) \mathbf{B}_n^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n} - \frac{\tau_n}{|\mathbf{B}_n| + \tau_n} \boldsymbol{q}^\top(0) \mathbf{B}_n^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n}.$$

Thus in order to upper bound $\mathrm{Var}\left(\widehat{V}(x^*)\mathbb{1}_{\Omega_n}\right)$, it suffices to upper bound the variances of the two terms in the above display. For the second term, using the fact that $|\mathbf{B}_n|$ is bounded away from zero on Ω_n under Condition (c) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$, we have

$$\mathrm{Var}\left(\frac{\tau_n}{|\mathbf{B}_n| + \tau_n} q^\top(0) \mathbf{B}_n^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n}\right) \lesssim \tau_n^2 \mathbb{E}\big(q^\top(0) \mathbf{B}_n^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n}\big)^2 \lesssim \tau_n^2 \mathbb{E}\big(\|\mathbf{B}_n^{-1} \mathbb{1}_{\Omega_n}\|^2 \|\boldsymbol{U}_n\|^2\big) \lesssim \tau_n^2,$$

where the last inequality follows since $\|\boldsymbol{\theta}\| = O(1)$ and $\boldsymbol{U}_n$ concentrates to θ by Lemma 16 since $nh_2 \to \infty$ and $n^2 h_1 h_2 \to \infty$. The first term can be decomposed as

$$\boldsymbol{q}^\top(0) \mathbf{B}_n^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n} = \boldsymbol{q}^\top(0)(\mathbf{B}_n^{-1} - \mathbf{B}^{-1}) \boldsymbol{U}_n \mathbb{1}_{\Omega_n} + \boldsymbol{q}^\top(0) \mathbf{B}^{-1} \boldsymbol{U}_n \mathbb{1}_{\Omega_n} := I + II.$$

For the first term, it holds on the event Ω_n that

$$\|\mathbf{B}_n^{-1} - \mathbf{B}^{-1}\| = \|\mathbf{B}^{-1}\|\|\mathbf{B}_n^{-1}\|\|\mathbf{B}_n - \mathbf{B}\| \le \|\mathbf{B}^{-1}\|(\|\mathbf{B}_n^{-1} - \mathbf{B}^{-1}\| + \|\mathbf{B}^{-1}\|)\|\mathbf{B}_n - \mathbf{B}\|.$$

Thus on the event Ω_n, it holds that $\|\mathbf{B}_n^{-1} - \mathbf{B}^{-1}\| \le \|\mathbf{B}^{-1}\|^2\|\mathbf{B}_n - \mathbf{B}\|/(1 - \|\mathbf{B}^{-1}\|\|\mathbf{B}_n - \mathbf{B}\|) \le 2\|\mathbf{B}^{-1}\|^2\|\mathbf{B}_n - \mathbf{B}\|$. This implies that

$$\begin{aligned}\mathrm{Var}(I) &\le \mathbb{E}\big(\boldsymbol{q}^\top(0)(\mathbf{B}_n^{-1} - \mathbf{B}^{-1})\boldsymbol{U}_n\mathbb{1}_{\Omega_n}\big)^2 \le \mathbb{E}\big(\|(\mathbf{B}_n^{-1} - \mathbf{B}^{-1})\mathbb{1}_{\Omega_n}\|^2\|\boldsymbol{U}_n\|^2\big) \\ &\lesssim \mathbb{E}\big(\|\mathbf{B}_n - \mathbf{B}\|^2\|\boldsymbol{U}_n\|^2\big) \le (\mathbb{E}\|\mathbf{B}_n - \mathbf{B}\|^4)^{1/2}(\mathbb{E}\|\boldsymbol{U}_n\|^4)^{1/2}.\end{aligned}$$

Clearly, $(\mathbb{E}\|\boldsymbol{U}_n\|^4)^{1/2} = O(1)$ since $\boldsymbol{U}_n$ concentrates to $\boldsymbol{\theta}$ and $\|\boldsymbol{\theta}\| = O(1)$, and by Lemmas 17 and 10, it holds that $(\mathbb{E}\|\mathbf{B}_n - \mathbf{B}\|^4)^{1/2} \lesssim n^{-1}h_2^{-1} + n^{-2}h_1^{-1}h_2^{-1}$. This concludes that

$$\mathrm{Var}(I) \lesssim n^{-1}h_2^{-1} + n^{-2}h_1^{-1}h_2^{-1}.$$

Lastly, for II, writing $Z := \boldsymbol{q}^\top(0)\mathbf{B}^{-1}\binom{n}{2}^{-1}\sum_{i<j}\boldsymbol{g}(X_i, X_j)$, we have

$$\mathrm{Var}(II) = \mathbb{E}(Z\mathbb{1}_{\Omega_n} - \mathbb{E}(Z\mathbb{1}_{\Omega_n}))^2 = \mathbb{E}((Z - \mathbb{E}Z) + \mathbb{E}(Z\mathbb{1}_{\Omega_n^c}) - Z\mathbb{1}_{\Omega_n^c})^2 \lesssim \mathrm{Var}(Z) + \mathbb{E}(Z\mathbb{1}_{\Omega_n^c})^2.$$

By Lemma 16, it holds that

$$\mathrm{Var}(Z) \lesssim n^{-1}h_2^{-1} + n^{-2}h_1^{-1}h_2^{-1}.$$

Lastly, by Cauchy's inequality and Lemma 17,

$$\mathbb{E}(Z\mathbb{1}_{\Omega_n^c})^2 \le (\mathbb{E}Z^4)^{1/2}\mathbb{P}^{1/2}(\Omega_n^c) \lesssim \big(\|\mathbf{B}^{-1}\|^4\mathbb{E}\|\boldsymbol{U}_n\|^4\big)^{1/2}\mathbb{P}^{1/2}(\Omega_n^c) = o(n^{-1}h_2^{-1} + n^{-2}h_1^{-1}h_2^{-1}).$$

Thus we conclude that

$$\mathrm{Var}(II) \lesssim n^{-1}h_2^{-1} + n^{-2}h_1^{-1}h_2^{-1}.$$

Putting together the pieces, we have proved

$$\mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n} - \mathbb{E}\Big(\widehat{V}(x^*)\mathbb{1}_{\Omega_n}\Big)\Big)^2 \lesssim n^{-1}h_2^{-1} + n^{-2}(h_1h_2)^{-1} + \tau_n^2.$$

This completes the proof. $\square$

Lemma 16. *Consider the term Z defined in the proof of Lemma 15:*

$$Z = \boldsymbol{q}^\top(0)\mathbf{B}^{-1}\binom{n}{2}^{-1}\sum_{i<j} D_{ij}\boldsymbol{q}\left(\frac{X_{ij}-x^*}{h_2}\right)K_{ij}.$$

Then, under the same conditions of Lemma 15, it holds that

$$Var(Z) \le C\big(n^{-1}h_2^{-1} + n^{-2}h_1^{-1}h_2^{-1}\big)$$

for some fixed positive constant C that only depends on α, β, C_0.

Proof. Denote $\boldsymbol{g}(X_i, X_j) := D_{ij}\boldsymbol{q}\big(\frac{X_{ij}-x^*}{h_2}\big)K_{ij}$ and $\boldsymbol{\theta} := \mathbb{E}\boldsymbol{g}(X_i, X_j)$. Then, it holds that

$$\begin{aligned}
\mathrm{Var}(Z) &= \mathbb{E}\left(\boldsymbol{q}^\top(0)\mathbf{B}^{-1}\binom{n}{2}^{-1}\sum_{i<j}(\boldsymbol{g}(X_i, X_j)-\boldsymbol{\theta})\right)^2\\
&\le \|\mathbf{B}^{-1}\|^2\mathbb{E}\left\|\binom{n}{2}^{-1}\sum_{i<j}(\boldsymbol{g}(X_i, X_j)-\boldsymbol{\theta})\right\|^2\\
&\lesssim n^{-4}\sum_{i<j, i'<j'}\mathbb{E}\big\{(\boldsymbol{g}(X_i, X_j)-\boldsymbol{\theta})^\top(\boldsymbol{g}(X_{i'}, X_{j'})-\boldsymbol{\theta})\big\},
\end{aligned}$$

where the last inequality follows since $\mathbf{B}^{-1}$ is invertible under Condition (c) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ as proved in Theorem 3. Apparently, when i, j, i', j' are all different, the summand equals to zero. When i, j, i', j' take three different values, say, $i = i' < j < j'$, we have

$$\begin{aligned}
&\mathbb{E}\big\{(\boldsymbol{g}(X_i, X_j)-\boldsymbol{\theta})^\top(\boldsymbol{g}(X_i, X_{j'})-\boldsymbol{\theta})\big\}\\
&= \mathbb{E}\left\{D_{ij}D_{ij'}\boldsymbol{q}^\top\left(\frac{X_{ij}-x^*}{h_2}\right)\boldsymbol{q}^\top\left(\frac{X_{ij'}-x^*}{h_2}\right)K_{ij}K_{ij'}\right\} - \|\boldsymbol{\theta}\|^2.
\end{aligned}$$

Let $\boldsymbol{Z}_1 := \boldsymbol{g}(X_i, X_j)$ and $\boldsymbol{Z}_2 := \boldsymbol{g}(X_i, X_{j'})$. Then, for any $k \in [\ell+1]$, it holds that $|*|Z_{1,k} \lesssim D_{ij}K_{h_1}(X_i - X_j)K_{h_2}(X_{ij} - x^*)$ and similarly for $Z_{2,k}$. Therefore, using the finite fourth moment of ε in Condition (d) of $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ in the calculation of $\mathbb{E}_\varepsilon(D_{ij}D_{ij'})$ and the fact that both $f(\cdot)$ and $V(\cdot)$ are bounded, we have

$$
\begin{aligned}
&\mathbb{E}(|Z_{1,k}Z_{2,k}|)\\
\lesssim\ &\mathbb{E}\{D_{ij}K_{h_1}(X_i-X_j)K_{h_2}(X_{ij}-x^*)D_{ij'}K_{h_1}(X_i-X_{j'})K_{h_2}(X_{ij'}-x^*)\}\\
\lesssim\ &\mathbb{E}\{K_{h_1}(X_i-X_j)K_{h_2}(X_{ij}-x^*)K_{h_1}(X_i-X_{j'})K_{h_2}(X_{ij'}-x^*)\}\\
=\ &\int_{\mathbb{R}^3}\frac{1}{h_1^2h_2^2}K\Big(\frac{v-u}{h_1}\Big)K\Big(\frac{\frac{u+v}{2}-x^*}{h_2}\Big)K\Big(\frac{w-u}{h_1}\Big)K\Big(\frac{\frac{u+w}{2}-x^*}{h_2}\Big)p_X(u)p_X(v)p_X(w)dudvdw\\
=\ &\frac{1}{h_2}\int_{\mathbb{R}^3}K(\widetilde{u})K(\widetilde{v})K(\widetilde{w})K\Big(\frac{\widetilde{u}h_2+h_1(\widetilde{w}-\widetilde{v})/2}{h_2}\Big)p_X(s_1)p_X(s_2)p_X(s_3)d\widetilde{u}d\widetilde{v}d\widetilde{w}\\
\lesssim\ &1/h_2,
\end{aligned}
$$

where we again invoke Condition (b) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ and the compact support of $K(\cdot)$, and

$$
s_1=\widetilde{u}h_2+x^*-\frac{h_1\widetilde{v}}{2},\quad s_2=\widetilde{u}h_2+x^*+\frac{h_1\widetilde{v}}{2},\quad s_3=\widetilde{u}h_2+x^*+\widetilde{w}h_1-\frac{\widetilde{v}h_1}{2}.
$$

This, along with the fact that $\|\boldsymbol{\theta}\|=O(1)$, concludes that

$$
\mathbb{E}\big\{(\boldsymbol{g}(X_i,X_j)-\boldsymbol{\theta})^\top(\boldsymbol{g}(X_{i'},X_{j'})-\boldsymbol{\theta})\big\}\lesssim h_2^{-1}
$$

when i,j,i',j' take three different values. Similarly, one can prove that when i,j,i',j' take two different values, that is $i=i'$ and $j=j'$,

$$
\mathbb{E}\big\{(\boldsymbol{g}(X_i,X_j)-\boldsymbol{\theta})^\top(\boldsymbol{g}(X_{i'},X_{j'})-\boldsymbol{\theta})\big\}\lesssim (h_1h_2)^{-1}.
$$

Putting together the pieces, we obtain that

$$
\mathrm{Var}(Z)\lesssim\frac{n^3h_2^{-1}+n^2(h_1h_2)^{-1}}{n^4}=n^{-1}h_2^{-1}+n^{-2}h_1^{-1}h_2^{-1}.
$$

This completes the proof. $\square$

Lemma 17. *Suppose Condition (b) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$ holds. Assume $n^2h_1h_2\to\infty$, $nh_2\to\infty$, and $h_1/h_2\to 0$. Then, for any $u,v>0$, the matrices $\mathbf{B}_n$ and $\mathbf{B}$ defined in the proof of Theorem 3 satisfy*

$$
\mathbb{P}\Big(\|\mathbf{B}_n-\mathbf{B}\|\geq C(v^{1/2}n^{-1/2}h_2^{-1/2}+u^{1/2}n^{-1}h_1^{-1/2}h_2^{-1/2})\Big)\leq C(\exp(-u)+\exp(-v))
$$

for some fixed positive constant C.

Proof. Using a standard entropy argument (see, for example, Lemma 5.3 in Vershynin (2012)), it holds that for any $t > 0$,

$$\mathbb{P}(\|\mathbf{B}_n - \mathbf{B}\| \geq t) \leq N \max_{1 \leq i \leq N} \mathbb{P}\big(|*|\boldsymbol{a}_i^\top(\mathbf{B}_n - \mathbf{B})\boldsymbol{a}_i \geq t/2\big),$$

where $N := 5^{\ell+1}$ and $\{\boldsymbol{a}_i\}_{i=1}^N$ is a 1/2-net on the unit sphere $\mathbb{S}^\ell$. We now upper bound $\mathbb{P}\big(|*|\boldsymbol{a}^\top(\mathbf{B}_n - \mathbf{B})\boldsymbol{a} \geq t/2\big)$ for an arbitrary $\boldsymbol{a} \in \mathbb{S}^\ell$ with the help of Lemma 13. For this, we upper bound the five quantities $B_1, B_2, B_3, \nu_1^2, \nu_2^2$ therein. Denote the kernel of $\boldsymbol{a}^\top \mathbf{B}_n \boldsymbol{a}$ as g and its linear part as g_1, that is,

$$g(X_i, X_j) := \bigg(\boldsymbol{a}^\top \boldsymbol{q}\Big(\frac{X_{ij} - x^*}{h_2}\Big)\bigg)^2 K_{ij} \text{ and } g_1(X_i) = \mathbb{E}(g(X_i, X_j) \mid X_i).$$

Then, we have

$$\begin{aligned} g_1(X_i) &= \int \bigg(\boldsymbol{a}^\top \boldsymbol{q}\Big(\frac{(u + X_i)/2 - x^*}{h_2}\Big)\bigg)^2 \frac{1}{h_1} K\Big(\frac{u - X_i}{h_1}\Big) \frac{1}{h_2} K\Big(\frac{(u + X_i)/2 - x^*}{h_2}\Big) p_X(u) du \\ &\lesssim \frac{1}{h_2} \int K(\widetilde{u}) K\Big(\frac{X_i + h_1 \widetilde{u}/2 - x^*}{h_2}\Big) p_X(X_i + \widetilde{u} h_1) d\widetilde{u} \lesssim h_2^{-1}, \end{aligned}$$

where in the first inequality we use the fact that $\boldsymbol{q}\Big(\frac{(u+X_i)/2 - x^*}{h_2}\Big)$ has bounded ℓ_2 norm due to the compact support of $K(\cdot)$, and in the last inequality we apply Condition (b) in $\mathcal{P}_{\mathrm{vf},(X,\varepsilon)}$. Therefore $B_1 \lesssim h_2^{-1}$ and one can similarly show that $\nu_1^2 \lesssim h_2^{-1}$.

For B_2, we have

$$\begin{aligned} B_2^2 &= n \sup_{X_i} \mathbb{E}\bigg\{\bigg(\boldsymbol{a}^\top \boldsymbol{q}\Big(\frac{X_{ij} - x^*}{h_2}\Big)\bigg)^4 \frac{1}{h_1^2} K^2\Big(\frac{X_i - X_j}{h_1}\Big) \frac{1}{h_2^2} K^2\Big(\frac{X_{ij} - x^*}{h_2}\Big) \mid X_i\bigg\} \\ &\lesssim \frac{n}{h_1 h_2^2} \sup_{X_i} \mathbb{E}\bigg\{\frac{1}{h_1} K\Big(\frac{X_i - X_j}{h_1}\Big) \mid X_i\bigg\} \\ &= \frac{n}{h_1 h_2^2} \sup_{X_i} \int K(u) p_X(X_i + u h_1) du \lesssim n h_1^{-1} h_2^{-2}. \end{aligned}$$

This concludes that $B_2 \lesssim n^{1/2} h_1^{-1/2} h_2^{-1}$. Moreover, one can easily show that $\nu_2^2 \lesssim (h_1 h_2)^{-1}$ and $B_3 \lesssim (h_1 h_2)^{-1}$. Putting together the pieces and applying Lemma 13, we obtain that for any $u, v > 0$,

$\mathbb{P}\big(|*|\boldsymbol{a}^\top(\mathbf{B}_n - \mathbf{B})\boldsymbol{a} \ge a_1 v^{1/2} + a_2 v + b_1 u^{1/2} + b_2 u + b_3 u^{3/2} + b_4 u^2\big) \le C(\exp(-v) + \exp(-u))$,

where $a_1 \lesssim n^{-1/2}h_2^{-1/2}, a_2 \lesssim n^{-1}h_2^{-1}$ and $b_1 \lesssim n^{-1}h_1^{-1/2}h_2^{-1/2}, b_2 \lesssim n^{-1}h_2^{-1}, b_3 \lesssim n^{-3/2}h_1^{-1/2}h_2^{-1}, b_4 \lesssim n^{-2}h_1^{-1}h_2^{-1}$. Under the conditions $n^2h_1h_2 \to \infty$, $nh_2 \to \infty$ and $h_1/h_2 \to 0$ as $n \to \infty$, the dominant terms are a_1 and b_1, that is,

$$n^{-1/2}h_2^{-1/2} \vee n^{-1}h_1^{-1/2}h_2^{-1/2}.$$

This completes the proof. □

Lemma 18. *Under the setting and conditions of Theorem 5, for any positive $\eta > 0$, there exists an i.i.d. sequence $\{r_{i,k}\}_{i\in[m_j],k\in[N_1]}$ (with m_j, N_1 defined in Theorem 5) with range contained in $[-B, B]$ for some B only depending on α, β, η, such that the probability measures $\mathbb{P}_j$ and $\widetilde{\mathbb{P}}_j$ defined therein satisfy that*

$$\chi^2\Big(\mathbb{P}_j, \widetilde{\mathbb{P}}_j; \Omega_n\Big) \lesssim n^{-\eta},$$

where, for any measurable subset $\mathcal{E}$ and two probability measures $\mathbb{P}$ and $\mathbb{Q}$, $\chi^2(\mathbb{P}, \mathbb{Q}; \mathcal{E})$ is the conditional χ^2-distance defined as

$$\chi^2(\mathbb{P}, \mathbb{Q}; \mathcal{E}) := \int_{\mathcal{E}} \frac{(p-q)^2}{q}$$

with p, q being the densities of $\mathbb{P}$ and $\mathbb{Q}$ with respect to some common dominating measure.

Proof. Without loss of generality, let $j = 1$. First note that the conditional χ^2-distance can be written as

$$\begin{aligned}
\chi^2\Big(\mathbb{P}_1, \widetilde{\mathbb{P}}_1; \Omega_n\Big) &= \int \mathbb{1}\{\Omega_n\}\frac{(p_1(\boldsymbol{x},\boldsymbol{y}) - \widetilde{p}_1(\boldsymbol{x},\boldsymbol{y}))^2}{\widetilde{p}_1(\boldsymbol{x},\boldsymbol{y})}d\boldsymbol{x}d\boldsymbol{y} \\
&= \int \mathbb{1}\{\Omega_n\}p(\boldsymbol{x})d\boldsymbol{x}\int \frac{(p_1(\boldsymbol{y}\mid\boldsymbol{x}) - \widetilde{p}_1(\boldsymbol{y}\mid\boldsymbol{x}))^2}{\widetilde{p}_1(\boldsymbol{y}\mid\boldsymbol{x})}d\boldsymbol{y} \\
&= \mathbb{E}\Big\{\mathbb{1}\{\Omega_n\}\chi^2\Big(\mathbb{P}_1(\boldsymbol{y}\mid\boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y}\mid\boldsymbol{x})\Big)\Big\},
\end{aligned}$$

where we have used $p(\cdot)$ to represent the density of $\{X_i\}_{i=1}^n$ under both $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$.

Recall the definition of the location index sequence $\{b_i\}_{i=1}^n$ in Theorem 5. Consider any realization of $\{X_i\}_{i=1}^n$ in Ω_n, and assume that based on their location indices $\{b_i\}_{i=1}^n$, $\{X_i\}_{i=1}^n$ is partitioned into $L_1 + L_2$ clusters with cardinality s_ℓ such that the X_i's in the same cluster have the same value b_i. Moreover, for those data points in the first L_1 clusters, the location index $b_i = (k_i, \ell_i)$ satisfies that $\Delta_{1,k_i} = 1$ while for the data points in the last L_2 clusters, it holds that $\Delta_{1,k_i} = 0$ (recall the definition of $\boldsymbol{\Delta}_1 = (\Delta_{1,1}, \ldots, \Delta_{1,N_2})$ in the lower bound design of Theorem 5. Apparently, we have the relations $1 \leq L_1 + L_2 \leq n$, $\sum_{\ell=1}^{L_1+L_2} s_\ell = n$ and $1 \leq s_\ell \leq K$ for $\ell \in [L_1 + L_2]$ (recall the definition of K in Theorem 5). Moreover, denoting $\mathbb{P}_{1,\pi_\ell}$ and $\widetilde{\mathbb{P}}_{1,\pi_\ell}$ (resp. p_{1,π_ℓ} and $\widetilde{p}_{1,\pi_\ell}$) for each $\ell \in [L_1 + L_2]$ as the joint distribution (resp. density) of those Y_i's in the ℓth cluster conditioning on the given realization $\{X_i\}_{i=1}^n$ under $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$, we have

$$
\begin{aligned}
\chi^2\Big(\mathbb{P}_1(\boldsymbol{y} \mid \boldsymbol{x}), \widetilde{\mathbb{P}}_1(\boldsymbol{y} \mid \boldsymbol{x})\Big) &= \prod_{\ell=1}^{L_1+L_2} \Big(1 + \chi^2\Big(\mathbb{P}_{1,\pi_\ell}, \widetilde{\mathbb{P}}_{1,\pi_\ell}\Big)\Big) - 1 \\
&= \prod_{\ell=1}^{L_1} \Big(1 + \chi^2\Big(\mathbb{P}_{1,\pi_\ell}, \widetilde{\mathbb{P}}_{1,\pi_\ell}\Big)\Big) - 1 \\
&\leq \exp\left\{\sum_{\ell=1}^{L_1} \chi^2\Big(\mathbb{P}_{1,\pi_\ell}, \widetilde{\mathbb{P}}_{1,\pi_\ell}\Big)\right\} - 1
\end{aligned}
$$

where the first equality follows by mutual independence of data points in each cluster, the second inequality follows from the fact that for those data points Y_i's in the latter L_2 clusters, each $Y_i \mid X_i \sim \mathcal{N}(0, 1)$ under both $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$. Since $L_1 \leq n$, it suffices to show that for any realization of $\{X_i\}_{i=1}^n$ in Ω_n, by matching enough moments, we have $\chi^2\Big(\mathbb{P}_{1,\pi_\ell}, \widetilde{\mathbb{P}}_{1,\pi_\ell}\Big) \leq n^{-\eta}$ for any $\eta > 0$.

For each $\ell \in [L_1]$, $|*|p_{1,\pi_\ell} - \widetilde{p}_{1,\pi_\ell}$ only depends on the ℓth cluster via its cardinality, which we now control for a general cluster size $1 \leq d \leq K$. Without loss of generality, we assume that $\ell = 1$ and the y_i's in this cluster are $\{y_1, \ldots, y_d\}$ with common location index $b_i = (1, 1)$. In this case, under the choice of θ_n^2 and h_1 given in Theorem 5, we clearly have

$Y_i = \theta_n r_{1,1} + (1-h_2^\beta)^{1/2}\varepsilon_i$ under $\mathbb{P}_1$ and $Y_i = \theta_n \widetilde{r}_{1,1} + (1-h_2^\beta)^{1/2}\varepsilon_i$ under $\widetilde{\mathbb{P}}_1$ for $i \in [d]$, where the sequence $\{\varepsilon_i\}_{i=1}^d$ follows the standard normal distribution under both $\mathbb{P}_1$ and $\widetilde{\mathbb{P}}_1$. Define $\sigma^2 := 1 - h_2^\beta = 1 - \theta_n^2$. Then, it holds that

$$p_{1,\pi_\ell}(y_1,\dots,y_d) = \int_{-\infty}^{\infty} \frac{1}{\sigma}\varphi\Big(\frac{y_1-\theta_n v}{\sigma}\Big)\dots\frac{1}{\sigma}\varphi\Big(\frac{y_d-\theta_n v}{\sigma}\Big)\mathbb{G}(dv),$$
$$\widetilde{p}_{1,\pi_\ell}(y_1,\dots,y_d) = \int_{-\infty}^{\infty} \frac{1}{\sigma}\varphi\Big(\frac{y_1-\theta_n v}{\sigma}\Big)\dots\frac{1}{\sigma}\varphi\Big(\frac{y_d-\theta_n v}{\sigma}\Big)\varphi(v)dv,$$

where $\mathbb{G}$ is the distribution of $\{r_{i,k}\}_{i\in[m_j],k\in[N_1]}$. Using the well-known equality $\varphi(t-\theta_n v) = \varphi(t)\big(\sum_{k=0}^\infty v^k\theta_n^k H_k(t)/k!\big)$ for any t, v, where H_k is the kth order Hermite polynomial, it holds that

$$\begin{aligned}
&\varphi\Big(\frac{y_1-\theta_n v}{\sigma}\Big)\dots\varphi\Big(\frac{y_d-\theta_n v}{\sigma}\Big)\\
&= \varphi\Big(\frac{y_1}{\sigma}\Big)\dots\varphi\Big(\frac{y_d}{\sigma}\Big)\sum_{k_1,\dots,k_d=0}^{\infty}\Big(\frac{\theta_n v}{\sigma}\Big)^{k_1+\dots+k_d}\frac{H_{k_1}(y_1/\sigma)}{k_1!}\dots\frac{H_{k_d}(y_d/\sigma)}{k_d!}\\
&= \varphi\Big(\frac{y_1}{\sigma}\Big)\dots\varphi\Big(\frac{y_d}{\sigma}\Big)\sum_{k=0}^{\infty}\Big(\frac{\theta_n v}{\sigma}\Big)^{k}\sum_{k_1+\dots+k_d=k}\Big(\frac{\theta_n v}{\sigma}\Big)^{k}\frac{H_{k_1}(y_1/\sigma)}{k_1!}\dots\frac{H_{k_d}(y_d/\sigma)}{k_d!},
\end{aligned}$$

and therefore by the symmetry of $\mathbb{G}$ and by matching the moments of $\mathbb{G}$ and the standard normal distribution up to order $2p$ for some positive integer p to be chosen later, we obtain

$$\begin{aligned}
&p_{1,\pi_\ell}(y_1,\dots,y_d) - \widetilde{p}_{1,\pi_\ell}(y_1,\dots,y_d)\\
&= \varphi\Big(\frac{y_1}{\sigma}\Big)\dots\varphi\Big(\frac{y_d}{\sigma}\Big)\sum_{k=0}^{\infty}\Big(\frac{\theta_n}{\sigma}\Big)^{k}\sum_{k_1+\dots+k_d=k}\frac{H_{k_1}(y_1/\sigma)}{k_1!}\dots\frac{H_{k_d}(y_d/\sigma)}{k_d!}\int v^k(\mathbb{G}-\Phi)(dv)\\
&= \varphi\Big(\frac{y_1}{\sigma}\Big)\dots\varphi\Big(\frac{y_d}{\sigma}\Big)\sum_{k=p}^{\infty}\Big(\frac{\theta_n}{\sigma}\Big)^{2k}\sum_{k_1+\dots+k_d=2k}\frac{H_{k_1}(y_1/\sigma)}{k_1!}\dots\frac{H_{k_d}(y_d/\sigma)}{k_d!}\int v^{2k}(\mathbb{G}-\Phi)(dv).
\end{aligned}$$

Define $\delta_{2k} := \int v^{2k}(\mathbb{G}-\Phi)(dv)$. Then, the above inequality further implies that

$$\begin{aligned}
&(p_{1,\pi_\ell}(y_1,\dots,y_d) - \widetilde{p}_{1,\pi_\ell}(y_1,\dots,y_d))^2\\
&= \varphi^2\Big(\frac{y_1}{\sigma}\Big)\dots\varphi^2\Big(\frac{y_d}{\sigma}\Big)\sum_{k,\ell=p}^{\infty}\Big(\frac{\theta_n}{\sigma}\Big)^{2k+2\ell}\sum_{\substack{k_1+\dots+k_d=2k\\ \ell_1+\dots+\ell_d=2\ell}}\frac{H_{k_1}(y_1/\sigma)}{k_1!}\frac{H_{\ell_1}(y_1/\sigma)}{\ell_1!}\dots\frac{H_{k_d}(y_d/\sigma)}{k_d!}\frac{H_{\ell_d}(y_d/\sigma)}{\ell_d!}
\end{aligned} \tag{A.10}$$

On the other hand, letting $Z \sim \mathbb{G}$, we have

$$
\begin{aligned}
\widetilde{p}_{1,\pi_\ell}(y_1,\dots,y_d) &= \int \frac{1}{\sigma}\varphi\left(\frac{y_1-\theta_n v}{\sigma}\right)\dots\frac{1}{\sigma}\varphi\left(\frac{y_d-\theta_n v}{\sigma}\right)\mathbb{G}(dv)\\
&= \frac{1}{\sigma^d}\varphi\left(\frac{y_1}{\sigma}\right)\dots\varphi\left(\frac{y_d}{\sigma}\right)\int \exp\left\{-\frac{d}{2\sigma^2}(\theta_n v)^2+\frac{\sum_{i=1}^d y_i\theta_n v}{\sigma^2}\right\}\mathbb{G}(dv)\\
&= \frac{1}{\sigma^d}\varphi\left(\frac{y_1}{\sigma}\right)\dots\varphi\left(\frac{y_d}{\sigma}\right)\mathbb{E}\left\{\exp\left\{-\frac{d}{2\sigma^2}\theta_n^2 Z^2+\frac{\sum_{i=1}^d y_i\theta_n}{\sigma^2}Z\right\}\right\}\\
&\geq \frac{1}{\sigma^d}\varphi\left(\frac{y_1}{\sigma}\right)\dots\varphi\left(\frac{y_d}{\sigma}\right)\exp\left(-\frac{d\theta_n^2}{2\sigma^2}\right),
\end{aligned}
\tag{A.11}
$$

where the last inequality follows from Jensen's inequality and the fact $\mathbb{E}Z = 0$, $\mathbb{E}Z^2 = 1$ from moment matching. Combining (A.10) and (A.11), we obtain that

$$
\begin{aligned}
\chi^2(\mathbb{P}_{1,\pi_\ell},\widetilde{\mathbb{P}}_{1,\pi_\ell}) &= \int \frac{\left(p_{1,\pi_\ell}(y_1,\dots,y_d)-\widetilde{p}_{1,\pi_\ell}(y_1,\dots,y_d)\right)^2}{\widetilde{p}_{1,\pi_\ell}(y_1,\dots,y_d)}dy_1\dots dy_d\\
&\leq \sigma^d\sum_{k,\ell=p}^{\infty}\delta_{2k}\delta_{2\ell}\sum_{\substack{k_1+\ldots+k_d=2k\\ \ell_1+\ldots+\ell_d=2\ell}}\left(\frac{\theta_n}{\sigma}\right)^{2k+2\ell}\prod_{j=1}^d\int\varphi\left(\frac{y_j}{\sigma}\right)\frac{H_{k_j}(y_j/\sigma)}{k_j!}\frac{H_{\ell_j}(y_j/\sigma)}{\ell_j!}dy_j\\
&= \sigma^{2d}\sum_{k,\ell=p}^{\infty}\delta_{2k}\delta_{2\ell}\sum_{\substack{k_1+\ldots+k_d=2k\\ \ell_1+\ldots+\ell_d=2\ell}}\left(\frac{\theta_n}{\sigma}\right)^{2k+2\ell}\prod_{j=1}^d\int\varphi(y_j)\frac{H_{k_j}(y_j)}{k_j!}\frac{H_{\ell_j}(y_j)}{\ell_j!}dy_j\\
&\leq \sum_{k=p}^{\infty}\left(\frac{\theta_n}{\sigma}\right)^{4k}\delta_{2k}^2\sum_{k_1+\ldots+k_d=2k}\frac{1}{k_1!}\dots\frac{1}{k_d!},
\end{aligned}
$$

where the last inequality follows from the fact that $\sigma \leq 1$ and $\int\varphi(t)H_k(t)H_\ell(t)dt = k!\mathbb{1}(k=\ell)$. Now, using the multinomial identity

$$
\sum_{k_1+\ldots+k_d=2k}\frac{(2k)!}{k_1!\dots k_d!}\left(\frac{1}{d}\right)^{2k} = 1,
$$

we obtain that $\sum_{k_1+\ldots+k_d=2k} 1/(k_1!\dots k_d!) = d^{2k}/(2k)!$, therefore it holds that

$$
\chi^2(\mathbb{P}_{1,\pi_\ell},\widetilde{\mathbb{P}}_{1,\pi_\ell}) \leq \sum_{k=p}^{\infty}\left(\frac{\theta_n\sqrt{d}}{\sigma}\right)^{4k}\frac{1}{(2k)!}\delta_{2k}^2.
$$

Now, by Lemma 5, for any positive integer p, we can find a symmetric distribution $\mathbb{G}$ that has the same first p moments as the standard normal distribution and is compactly supported on $[-B, B]$ for some B that only depends on p. This combined with the fact that $\int t^{2k}\varphi(t)dt = (2k-1)!!$ implies that $\delta_{2k}^2 \lesssim B^{4k} + (2k)!$. We therefore obtain

$$\chi^2(\mathbb{P}_{1,\pi_\ell}, \widetilde{\mathbb{P}}_{1,\pi_\ell}) \lesssim \sum_{k=p}^{\infty}\left(\frac{\theta_n\sqrt{d}}{\sigma}\right)^{4k} \lesssim n^{-\eta}$$

by choosing a sufficiently large p that only depends on α, β and η, where we also use the fact that d is bounded by an absolute constant and for sufficiently large n, it holds that $\sigma > 1/2$. This completes the proof. $\square$

Lemma 19. *For some $M \geq 2$, let $\mathbb{P}_0, \mathbb{P}_1, \ldots, \mathbb{P}_M$ be $M+1$ hypotheses on some measurable space $(\mathcal{X}, \mathcal{A})$ such that for each $0 \leq i \neq j \leq M$, $\mathbb{P}_i$ and $\mathbb{P}_j$ are mutually absolutely continuous. Let Ω be a measurable subset of $\mathcal{X}$ such that $\mathbb{P}_j(\Omega)$ is identical for all $0 \leq j \leq M$. Define the "conditional" version of Kullback divergence as*

$$K(\mathbb{P}, \mathbb{Q}; \Omega) := \int_\Omega \log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{P}. \tag{A.12}$$

Then, if

$$\frac{1}{M}\sum_{j=1}^{M} K(\mathbb{P}_j, \mathbb{P}_0; \Omega) \leq c^* \log(M)$$

for some $0 < c^* < 1/8$, the following statement holds

$$\inf_\psi p_{e,M}(\psi) \geq \frac{\sqrt{M}}{1+\sqrt{M}}\left(\mathbb{P}_0(\Omega) - 2c^* - \sqrt{\frac{2c^*}{\log(M)}}\right),$$

where the infimum ranges over all tests taking values in $\{0, 1, \ldots, M\}$ and $p_{e,M}(\psi) := \max_{0\leq j\leq M} \mathbb{P}_j(\psi \neq j)$.

Proof. This is exactly the conditional version of Theorem 2.5 in Tsybakov (2009a). We first show that subject to the condition

$$\frac{1}{M}\sum_{j=1}^{M} K(\mathbb{P}_j, \mathbb{P}_0; \Omega) \leq c$$

for some $c > 0$, for all $0 < \tau < 1$, it holds that

$$\frac{1}{M}\sum_{j=1}^{M}\mathbb{P}_j\left(\frac{d\mathbb{P}_0}{d\mathbb{P}_j} \geq \tau\right) \geq \mathbb{P}_0(\Omega) - c', \tag{A.13}$$

where $c' := -(c + \sqrt{c/2})/\log(\tau)$. For this, we have for each $j \in [M]$

$$\begin{aligned}
\mathbb{P}_j\left(\frac{d\mathbb{P}_0}{d\mathbb{P}_j} \geq \tau\right) &= \mathbb{P}_j\left(\frac{d\mathbb{P}_j}{d\mathbb{P}_0} \leq \frac{1}{\tau}\right) \\
&= 1 - \left\{\mathbb{P}_j\left(\left\{\frac{d\mathbb{P}_j}{d\mathbb{P}_0} \geq \frac{1}{\tau}\right\}\bigcap \Omega\right) + \mathbb{P}_j\left(\left\{\frac{d\mathbb{P}_j}{d\mathbb{P}_0} \geq \frac{1}{\tau}\right\}\bigcap \Omega^c\right)\right\} \\
&\geq \mathbb{P}_0(\Omega) - \mathbb{P}_j\left(\left\{\frac{d\mathbb{P}_j}{d\mathbb{P}_0} \geq \frac{1}{\tau}\right\}\bigcap \Omega\right) \\
&= \mathbb{P}_0(\Omega) - \mathbb{P}_j\left(\left\{\log\left(\frac{d\mathbb{P}_j}{d\mathbb{P}_0}\right) \geq \log\left(\frac{1}{\tau}\right)\right\}\bigcap \Omega\right) \\
&\geq \mathbb{P}_0(\Omega) - (\log(1/\tau))^{-1}\mathbb{E}_{\mathbb{P}_j}\left(\log\left(\frac{d\mathbb{P}_j}{d\mathbb{P}_0}\right)_+ \mathbb{1}\{\Omega\}\right),
\end{aligned}$$

where in the third line we use the fact that $\mathbb{P}_j(\Omega) = \mathbb{P}_0(\Omega)$, and for a real number a, $a_+ := \max\{0, a\}$. Let p_0 and p_j be the densities of $\mathbb{P}_0$ and $\mathbb{P}_j$ with respect to some common dominating measure. Then, by definition of the conditional Kullback divergence, Lemma 20, and Lemma 21, it holds that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_j}\left(\log\left(\frac{d\mathbb{P}_j}{d\mathbb{P}_0}\right)_+ \mathbb{1}\{\Omega\}\right) &= \int_\Omega p_j\left(\log\frac{p_j}{p_0}\right)_+ \\
&= K(\mathbb{P}_j, \mathbb{P}_0; \Omega) + \int_\Omega p_j\left(\log\frac{p_j}{p_0}\right)_- \\
&\leq K(\mathbb{P}_j, \mathbb{P}_0; \Omega) + \mathsf{TV}(\mathbb{P}_j, \mathbb{P}_0; \Omega) \\
&\leq K(\mathbb{P}_j, \mathbb{P}_0; \Omega) + \sqrt{K(\mathbb{P}_j, \mathbb{P}_0; \Omega)/2}.
\end{aligned}$$

Now, by the condition $\sum_{j=1}^{M} K(\mathbb{P}_j, \mathbb{P}_0; \Omega)/M \leq c$ and Cauchy's inequality, it holds that

$$\frac{1}{M}\sum_{j=1}^{M}\sqrt{K(\mathbb{P}_j, \mathbb{P}_0; \Omega)} \leq \left\{\frac{1}{M}\sum_{j=1}^{M}K(\mathbb{P}_j, \mathbb{P}_0; \Omega)\right\}^{1/2} \leq \sqrt{c}.$$

We therefore conclude that (A.13) is true. Next, by Proposition 2.2 in Tsybakov (2009a), we obtain that

$$\inf_{\psi} p_{e,M}(\psi) \geq \sup_{0<\tau<1} \frac{\tau M}{\tau M + 1} \left\{ \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_j \left(\frac{d\mathbb{P}_0}{d\mathbb{P}_j} \geq \tau \right) \right\} \geq \sup_{0<\tau<1} \frac{\tau M}{\tau M + 1} \left(\mathbb{P}_0(\Omega) + \frac{c + \sqrt{c/2}}{\log \tau} \right).$$

Lastly, by choosing $c = c^* \log M$ and $\tau = 1/\sqrt{M}$, we obtain

$$\begin{aligned} \inf_{\psi} p_{e,M}(\psi) &\geq \sup_{0<\tau<1} \frac{\tau M}{\tau M + 1} \left\{ \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_j \left(\frac{d\mathbb{P}_0}{d\mathbb{P}_j} \geq \tau \right) \right\} \\ &\geq \frac{\sqrt{M}}{1 + \sqrt{M}} \left(\mathbb{P}_0(\Omega) - 2c^* - \sqrt{\frac{2c^*}{\log M}} \right). \end{aligned}$$

This completes the proof. □

Lemma 20. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ such that $\mathbb{P} \ll \mathbb{Q}$, and Ω be a measurable subset of $\mathcal{X}$. Define the conditional version of the total variation distance as follows*

$$\mathsf{TV}(\mathbb{P}, \mathbb{Q}; \Omega) := \sup_{A \in \mathcal{A}} |*| \mathbb{P}(A \bigcap \Omega) - \mathbb{Q}(A \bigcap \Omega). \tag{A.14}$$

Then, it holds that

$$\int_{\Omega} \left(\log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right) \right)_{-} \leq \mathsf{TV}(\mathbb{P}, \mathbb{Q}; \Omega),$$

where $a_{-} := \max\{0, -a\}$.

Proof. Let p and q be the densities of $\mathbb{P}$ and $\mathbb{Q}$ with respect to some common dominating measure, and define $A := \{q \geq p > 0\}$. Then, it holds that

$$\int_{\Omega} \left(\log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right) \right)_{-} d\mathbb{P} = \int_{\Omega \bigcap \{p,q>0\}} p \left(\log \frac{p}{q} \right)_{-} = \int_{A \bigcap \Omega} p \log \frac{q}{p} \leq \int_{A \bigcap \Omega} q - p \leq \mathsf{TV}(\mathbb{P}, \mathbb{Q}; \Omega).$$

This completes the proof. □

Lemma 21. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ such that $\mathbb{P} \ll \mathbb{Q}$, and Ω be a measurable subset of $\mathcal{X}$ such that $\mathbb{P}(\Omega) = \mathbb{Q}(\Omega)$. For the conditional version of the Kullback divergence (defined in* (A.12)*) and total variation distance (defined in* (A.14)*), it holds that*

$$\mathsf{TV}(\mathbb{P}, \mathbb{Q}; \Omega) \leq \sqrt{K(\mathbb{P}, \mathbb{Q}; \Omega)/2}.$$

Proof. Firstly, using the condition $\mathbb{P}(\Omega) = \mathbb{Q}(\Omega)$, it can be readily verified that the conditional total variation distance can be equivalently written as

$$\mathsf{TV}(\mathbb{P}, \mathbb{Q}; \Omega) = \frac{1}{2} \int_{\Omega} |p - q|,$$

where p and q are the densities of $\mathbb{P}$ and $\mathbb{Q}$ with respect to some common dominating measure (cf. Lemma 2.1 in Tsybakov (2009a)). Then, following the proof of the first Pinsker's inequality in Lemma 2.5 in Tsybakov (2009a), it holds that

$$\begin{aligned}
\mathsf{TV}(\mathbb{P}, \mathbb{Q}; \Omega) &= \frac{1}{2} \int_{\Omega} |*| p - q \\
&= \frac{1}{2} \int_{\Omega \bigcap \{q > 0\}} |*| \frac{p}{q} - 1 q \\
&\leq \frac{1}{2} \int_{\Omega \bigcap \{q > 0\}} q \sqrt{\left(\frac{4}{3} + \frac{2p}{3q}\right) \psi(\frac{p}{q})} \\
&\leq \frac{1}{2} \left\{ \int_{\Omega} \left(\frac{4q}{3} + \frac{2p}{3} \right) \right\}^{1/2} \left\{ \int_{\Omega} q \left(\frac{p}{q} \log \frac{p}{q} - \frac{p}{q} + 1 \right) \right\}^{1/2} \\
&= \sqrt{\frac{\mathbb{P}(\Omega)}{2}} K^{1/2}(\mathbb{P}, \mathbb{Q}; \Omega) \\
&\leq \sqrt{K(\mathbb{P}, \mathbb{Q}; \Omega)/2},
\end{aligned}$$

where $\psi(x) := x \log x - x + 1$. This completes the proof. □

A.4 Proofs of results in Section 2.4

We only provide the proofs for Propositions 1, 2, 5, 8, 11-13. The proofs of Propositions 4, 6, 7, and 9 are straightforward, and the proof of Proposition 10 is similar to that of Proposition

11.

A.4.1 Proof of Proposition 1

Proof. Given the proof of Theorem 1 and its supporting lemmas, the proof here is relatively straightforward. We only provide here a sketched version for completeness. For simplicity, we only prove the case with $d = 2$, and we will show that the desired upper bound can be achieved with the bandwidths choices

$$h_1 \asymp n^{-2\alpha_2/(4\alpha_1\alpha_2+\alpha_1+\alpha_2)} \text{ and } h_2 \asymp n^{-2\alpha_1/(4\alpha_1\alpha_2+\alpha_1+\alpha_2)}.$$

C and c still represent two generic fixed positive constants whose values may change at each occurrence.

Define $U_1, U_2, \theta_1, \theta_2$ and the "good" event $\mathcal{E}$ the same way as in Theorem 1. Following its proof, we now lower bound θ_2 and upper bound the term $|*|\theta_1 - \theta_2\sigma^2$. For θ_2, we have

$$\begin{aligned}
\theta_2 &= \mathbb{E}\{K_{h_1}(X_{i,1} - X_{j,1})K_{h_2}(X_{i,2} - X_{j,2})\} \\
&= \int_{\mathbb{R}^2} \frac{1}{h_1h_2} K\left(\frac{u}{h_1}\right) K\left(\frac{v}{h_2}\right) p_{\widetilde{\boldsymbol{X}}_{ij}}(u, v) du dv \\
&= \int_{\mathbb{R}^2} K(u)K(v) p_{\widetilde{\boldsymbol{X}}_{ij}}(uh_1, vh_2) du dv \\
&= \int_{-1}^{1} \int_{-1}^{1} K(u)K(v) p_{\widetilde{\boldsymbol{X}}_{ij}}(uh_1, vh_2) du dv \\
&\geq \int_{\mathcal{U}_{(h_1,h_2)}} K(u)K(v) p_{\widetilde{\boldsymbol{X}}_{ij}}(uh_1, vh_2) du dv \\
&\geq \boldsymbol{\lambda}(\mathcal{U}_{(h_1,h_2)}) \inf_{\boldsymbol{u} \in \mathcal{U}_{(h_1,h_2)}} p_{\widetilde{\boldsymbol{X}}_{\boldsymbol{ij}}}(u_1h_1, u_2h_2) \underline{M}_K \\
&\geq c_0^2 \underline{M}_K.
\end{aligned}$$

Here, the fourth equality follows from the kernel condition that $K(\cdot)$ is supported in $[-1, 1]$, and the set $\mathcal{U}_{(h_1,h_2)}$ starting from the first inequality follows from Condition (b) in $\mathcal{P}_{\text{mcv},(\boldsymbol{X},\varepsilon)}$

since for any fixed $\delta_0 > 0$ chosen therein, $\|\boldsymbol{\delta}\|_\infty := \|(h_1, h_2)\|_\infty \le \delta_0$ for sufficiently large n. For $|*|\theta_1 - \theta_2\sigma^2$, using the condition $\alpha_i \in (0,1]$, $i = 1,2$, it holds that

$$
\begin{aligned}
&|*|\theta_1 - \theta_2\sigma^2 \\
&= \mathbb{E}\Big\{K_{h_1}(\widetilde{X}_{ij,1})K_{h_2}(\widetilde{X}_{ij,2})(f(\boldsymbol{X}_i) - f(\boldsymbol{X}_j))^2/2\Big\} \\
&\lesssim \mathbb{E}\left\{\frac{1}{h_1h_2}K\left(\frac{\widetilde{X}_{ij,1}}{h_1}\right)K\left(\frac{\widetilde{X}_{ij,2}}{h_2}\right)\left(|*|\widetilde{X}_{ij,1}^{2\alpha_1} + |*|\widetilde{X}_{ij,2}^{2\alpha_2}\right)\right\} \\
&= \int \frac{1}{h_1h_2}K\left(\frac{u}{h_1}\right)K\left(\frac{v}{h_1}\right)(|*|u^{2\alpha_1} + |*|v^{2\alpha_2})p_{\widetilde{\boldsymbol{X}}_{ij}}(u,v)dudv \\
&= h_1^{2\alpha_1}\int K(u)K(v)|u|^{2\alpha_1}p_{\widetilde{\boldsymbol{X}}_{ij}}(uh_1, vh_2)dudv + \\
&\quad h_2^{2\alpha_2}\int K(u)K(v)|v|^{2\alpha_2}p_{\widetilde{\boldsymbol{X}}_{ij}}(uh_1, vh_2)dudv \\
&\le C_0\left(h_1^{2\alpha_1}\int K(u)K(v)|u|^{2\alpha_1}dudv + h_2^{2\alpha_2}\int K(u)K(v)|v|^{2\alpha_2}dudv\right) \\
&\le C(h_1^{2\alpha_1} + h_2^{2\alpha_2}),
\end{aligned}
$$

where we have applied Condition (a) in $\mathcal{P}_{\mathrm{mcv},(\boldsymbol{X},\varepsilon)}$ and the compact support of $K(\cdot)$. Therefore, Lemmas 22 and 23 and the above estimates imply that

$$
\mathbb{E}\left\{\left(\frac{U_1 - U_2\sigma^2}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}\}\right\} \lesssim (h_1^{4\alpha_1} + h_2^{4\alpha_2} + n^{-1} + n^{-2}(h_1h_2)^{-1}) \asymp (n^{-\frac{8\alpha_1\alpha_2}{4\alpha_1\alpha_2+\alpha_1+\alpha_2}} + n^{-1}).
$$

Moreover, using the same argument as in the proof of Theorem 1, it holds that

$$
\mathbb{E}\left\{\left(\frac{U_1 - U_2\sigma^2}{U_2}\right)^2 \mathbb{1}\{\mathcal{E}^c\}\right\} = o(n^{-\frac{8\alpha_1\alpha_2}{4\alpha_1\alpha_2+\alpha_1+\alpha_2}} + n^{-1}).
$$

This completes the proof for $d = 2$. In the case of general dimension d with heterogeneous smoothness index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$, the upper bound takes the form

$$
\mathbb{E}\big(\widehat{\sigma}_d^2 - \sigma^2\big)^2 \lesssim n^{-1} + n^{-2}\left(\prod_{k=1}^d h_k\right)^{-1} + \sum_{k=1}^d h_k^{4\alpha_k}
$$

and we choose $h_k \asymp n^{-2\underline{\alpha}/(\alpha_k(4\underline{\alpha}+d))}$. This completes the proof. $\square$

A.4.2 Proof of Proposition 2

Proof. Given the proof of Theorem 2, the proof here is relatively straightforward. We will thus only present the construction of the hardest sub-problem. For simplicity, we will only prove the case for $d = 2$. We also only consider the regime of (α_1, α_2) in which the lower bound is sub-parametric: $4\alpha_1\alpha_2 < \alpha_1 + \alpha_2$. Throughout the proof, C represents some generic positive constant and does not depend on n, and c represents a generic sufficiently small positive constant which also does not depend on n. In particular, c is always taken to be smaller than 1. Both C and c might have different values for each occurrence.

Introduce the following constants:

$$\theta_n^2 := h_1^{2\alpha_1} := h_2^{2\alpha_2} := cn^{-\frac{4\alpha_1\alpha_2}{4\alpha_1\alpha_2+\alpha_1+\alpha_2}}, \quad N_1 := 1/(6h_1), \quad N_2 := 1/(6h_2), \tag{A.15}$$

where we tune the constant c in θ_n^2 so that N_1 and N_2 are both positive integers. We now specify $f(\cdot)$, distribution of $\boldsymbol{X}$, σ^2 and distribution of ε in the null and alternative hypotheses, H_0 and H_1, respectively.

Choice of σ^2: Under H_0, let $\sigma^2 = 1 + \theta_n^2$. Under H_1, let $\sigma^2 = 1$.

Choice of ε: Under both H_0 and H_1, let $\varepsilon \sim \mathcal{N}(0, 1)$.

Choice of $\boldsymbol{X}$: Under both H_0 and H_1, let $\boldsymbol{X}$ be uniformly distributed on the union of the rectangles $[(6i_1 - 5)h_1, (6i_1 - 1)h_1] \times [(6i_2 - 5)h_2, (6i_2 - 1)h_2]$ for $i_1 \in [N_1]$ and $i_2 \in [N_2]$.

Choice of $f(\cdot)$: Under H_0, let $f \equiv 0$. Under H_1, let f be a smooth bump function that takes value $\theta_n r_{i_1,i_2}$ on the rectangle $[(6i_1-5)h_1, (6i_1-1)h_1]\times[(6i_2-5)h_2, (6i_2-1)h_2]$, and then smoothly decays to 0 on the union of the segments $\{x_1 = 6(i_1 - 1)h_1, 0 \leq x_2 \leq 1\}$ for $i_1 \in [N_1]$ and $\{0 \leq x_1 \leq 1, x_2 = 6(i_2 - 1)h_2\}$ for $i_2 \in [N_2]$. Here, the double indexed sequence $\{r_{i_1,i_2}\}_{i_1\in[N_1],i_2\in[N_2]}$ are $N_1 \times N_2$ i.i.d. symmetric and compactly supported random variables with distribution $\mathbb{G}$ satisfying

$$\int_{-\infty}^{\infty} x^j \mathbb{G}(dx) = \int_{-\infty}^{\infty} x^j \varphi(x) dx, \quad j = 1, \ldots, q,$$

where q is some odd integer strictly larger than $1 + (\alpha_1 + \alpha_2)/(2\alpha_1\alpha_2)$.

In the definition of $f(\cdot)$ under H_1, the existence of the distribution G is guaranteed by Lemma 5 and its range only depends on α_1 and α_2. The smoothness property of $f(\cdot)$ can be achieved by mollifying an indicator function.

We only verify Condition (c) in $\mathcal{P}_{\mathrm{mcv},(\boldsymbol{X},\varepsilon)}$, which holds by the convolution formula that for any $0 \le u \le 1/2$ and $0 \le v \le 1/2$

$$\begin{aligned}
p_{\widetilde{\boldsymbol{X}}_{ij}}(u,v) &= \int_u^1 \int_v^1 p_{\boldsymbol{X}}(t_1,t_2) p_{\boldsymbol{X}}(t_1-u, t_2-v) dt_1 dt_2 \\
&\ge \sum_{i_1=\lceil u/(6h_1)\rceil+1}^{N_1} \sum_{i_2=\lceil v/(6h_2)\rceil+1}^{N_2} \int_{(6i_1-5)h_1}^{(6i_1-1)h_1} \int_{(6i_2-5)h_2}^{(6i_2-1)h_2} p_{\boldsymbol{X}}(t_1,t_2) p_{\boldsymbol{X}}(t_1-u, t_2-v) dt_1 dt_2 \\
&\ge \sum_{i_1=\lceil u/(6h_1)\rceil+1}^{N_1} \sum_{i_2=\lceil v/(6h_2)\rceil+1}^{N_2} (2h_1)(2h_2) \cdot \frac{9}{4} \cdot \frac{9}{4} \\
&\ge \frac{9}{256}
\end{aligned}$$

for sufficiently large n. A similar calculation holds for all $|u| \le 1/2$ and $|v| \le 1/2$. Therefore, Condition (c) holds with $\delta_0 = 1/2$ and $\mathcal{U}_{\boldsymbol{\delta}} \equiv [-1,1]^2$. □

A.4.3 Proof of Proposition 5

Proof. We employ an iterative usage of pairwise difference. Under the regular design, we have $Y_{i_1,\ldots,i_d} = \sum_{k=1}^d f_k(i_k/n^{1/d}) + \sigma\varepsilon_{i_1,\ldots,i_d}$ for $(i_1,\ldots,i_d) \in [n^{1/d}] \times \ldots \times [n^{1/d}]$, where we assume without loss of generality that $n^{1/d}$ is an even integer. Let $m := n^{1/d}/2$ and define $\mathcal{I} := \{(1,2),\ldots,(2m-1,2m)\}$ with cardinality m. For all index pairs $(i_k^{(1)}, i_k^{(2)}) \in \mathcal{I}$, $k \in [d]$, we have

$$Y_{(i_1^{(1)},i_1^{(2)}),\ldots,(i_d^{(1)},i_d^{(2)})} := \sum_{j_k\in\{i_k^{(1)},i_k^{(2)}\},k\in[d]} Y_{j_1,\ldots,j_d}(-1)^{\sum_{k=1}^d \mathbb{1}\{j_k=i_k^{(1)}\}}$$
$$= \sum_{j_k\in\{i_k^{(1)},i_k^{(2)}\},k\in[d]} \sigma\varepsilon_{j_1,\ldots,j_d}(-1)^{\sum_{k=1}^d \mathbb{1}\{j_k=i_k^{(1)}\}}.$$

Clearly, we have $\mathbb{E}\Big(Y_{(i_1^{(1)},i_1^{(2)}),\ldots,(i_d^{(1)},i_d^{(2)})}\Big) = 0$ and $\mathrm{Var}\Big(Y_{(i_1^{(1)},i_1^{(2)}),\ldots,(i_d^{(1)},i_d^{(2)})}\Big) = 2^d\sigma^2$. More importantly, the newly formed data sequence $\{Y_{(i_1^{(1)},i_1^{(2)}),\ldots,(i_d^{(1)},i_d^{(2)})}\}_{(i_k^{(1)},i_k^{(2)})\in\mathcal{I},k\in[d]}$ with cardinality $m^d = n/2^d$ is i.i.d. with mean 0 and variance $2^d\sigma^2$. Therefore, by defining $\overline{Y}$ to be the average of this newly formed data sequence and $\widehat{\sigma}^2$ to be

$$\widehat{\sigma}^2 := \frac{1}{n}\sum_{(i_k^{(1)},i_k^{(2)})\in\mathcal{I},k\in[d]}\Big(Y_{(i_1^{(1)},i_1^{(2)}),\ldots,(i_d^{(1)},i_d^{(2)})} - \overline{Y}\Big)^2,$$

we clearly have $\mathbb{E}(\widehat{\sigma}^2-\sigma^2)^2 \lesssim n^{-1}$ for some absolute positive constant C under a finite fourth moment assumption, which is clearly not improvable. Thus the proof is complete. □

A.4.4 Proof of Proposition 8

Proof. Throughout the proof, C represents a positive constant that only depends $\boldsymbol{\alpha}$ and C_0. Following the argument before the statement of Proposition 8, define

$$\varepsilon_i^{(\ell)} := \sum_{k\in[d],k\neq\ell} f_k(X_{i,k}) + \varepsilon_i$$

and its variance

$$\sigma_{(\ell)}^2 := \sum_{k\in[d],k\neq\ell} \mathbb{E}f_k^2(X_{i,k}) + \sigma^2$$

for all $\ell\in[d]$. Clearly, under the mutual independence of the components of $(X_{i,1},\ldots,X_{i,d})$, it holds that $\mathbb{E}\varepsilon_i^{(\ell)} = 0$ and $\varepsilon_i^{(\ell)}$ is independent of $f_\ell(X_{i,\ell})$. For each $\ell\in[d]$, by viewing the model equivalently as $Y_i = f_\ell(X_{i,\ell}) + \varepsilon_i^{(\ell)}$ for $i\in[n]$ and then applying the univariate kernel smoother defined in (2.8), which renders an estimator which we denote as $\widehat{\sigma}^2_{(\ell)}$, we obtain by Theorem 1

$$\mathbb{E}\big(\widehat{\sigma}^2_{(\ell)} - \sigma^2_{(\ell)}\big)^2 = \mathbb{E}\left(\widehat{\sigma}^2_{(\ell)} - \sum_{k\in[d],k\neq\ell} \mathbb{E}f_k^2(W_{i,k}) - \sigma^2\right)^2 \leq Cn^{-\frac{8\alpha_\ell}{4\alpha_\ell+1}}.$$

Moreover, letting $\overline{Y}$ be the average of $\{Y_i\}_{i=1}^n$ and $\widehat{\sigma}^2_Y$ be the sample variance estimator, that is, $\widehat{\sigma}^2_Y := \sum_{i=1}^n (Y_i - \overline{Y})^2/n$, it holds that

$$\mathbb{E}\big(\widehat{\sigma}^2_Y - \mathrm{Var}(Y)\big)^2 = \mathbb{E}\left(\widehat{\sigma}^2_Y - \sum_{k\in[d]} \mathbb{E}f_k^2(X_{i,k}) - \sigma^2\right)^2 \leq Cn^{-1}.$$

Since $\sigma^2 = \sum_{\ell=1}^d \sigma^2_{(\ell)} - (d-1)\mathrm{Var}(Y)$, thus by defining $\widehat{\sigma}^2 := \sum_{\ell=1}^d \widehat{\sigma}^2_{(\ell)} - (d-1)\widehat{\sigma}^2_Y$, we obtain that

$$\begin{aligned}\mathbb{E}\big(\widehat{\sigma}^2 - \sigma^2\big)^2 &= \mathbb{E}\big\{\big(\widehat{\sigma}^2_{(1)} - \sigma^2_{(1)}\big) + \ldots + \big(\widehat{\sigma}^2_{(d)} - \sigma^2_{(d)}\big) + (d-1)\big(\widehat{\sigma}^2_Y - \mathrm{Var}(Y)\big)\big\}^2 \\ &\leq C\Big(n^{-\frac{8\alpha_{\min}}{4\alpha_{\min}+1}} + n^{-1}\Big).\end{aligned}$$

This completes the proof. □

A.4.5 Proof of Proposition 11

Proof. For simplicity, we only prove the case with two additive components $f(X)$ and $g(W)$ which are α- and β-Hölder smooth, respectively. Throughout the proof, C represents a generic fixed positive constant that only depends on α, β, C_0 and the joint distribution of (X, W). Denote the marginal distribution of X and W as F_X and F_W. Since the transition boundary for both α and β is $1/4$, we may assume without of loss of generality that $0 < \alpha, \beta < 1$. As a result, since F_X^{-1} and F_W^{-1} are both Lipschitz with fixed positive constants, $\overline{f} := f \circ F_X^{-1}$ and $\overline{g} := g \circ F_W^{-1}$ are still α- and β-Hölder smooth. With a standard wavelet expansion (cf. Proposition 2.5 in Meyer (1990)), we can write the model equivalently as

$$Y_i = \overline{f}_1(U_{1,i}) + \sum_{j=1}^{2^{J_1}} \psi_j(U_{1,i})\gamma_{1,j} + \overline{g}_1(U_{2,i}) + \sum_{j=1}^{2^{J_2}} \varphi_j(U_{2,i})\gamma_{2,j} + \sigma\varepsilon_i,$$

where $\{\psi_j\}_{j=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ are two sets of orthonormal wavelet basis (with respect to the Lebesgue measure on $[0,1]$), $\{U_{1,i}\}_{i=1}^{n} = \{F_X(X_i)\}_{i=1}^{n}$ and $\{U_{2,i}\}_{i-=1}^{n} = \{F_W(W_i)\}_{i=1}^{n}$ are two uniform $[0,1]$ sequences, and $\|\overline{f}_1\|_\infty \le C(2^{-\alpha J_1})$ and $\|\overline{g}_1\|_\infty \le C(2^{-\beta J_2})$. Define $\boldsymbol{U}_i := (\psi_1(U_{1,i}), \ldots, \psi_{2^{J_1}}(U_{1,i}), \varphi_1(U_{2,i}), \ldots, \varphi_{2^{J_2}}(U_{2,i}))$ as the new feature vector of length $2^{J_1} + 2^{J_2}$. Without loss of generality, we assume $\mathbb{E}\boldsymbol{U}_i = 0$ (a mean shift does not affect the estimation of variance) and let $\boldsymbol{\Sigma} := \mathrm{Cov}(\boldsymbol{U}_i)$. Without loss of generality, we can assume $\boldsymbol{\Sigma}$ is strictly positive definite (otherwise we can orthogonalize with respect to the linear span of $(\psi_1(U_{1,i}), \ldots, \psi_{2^{J_1}}(U_{2,i}), \varphi_1(U_{1,i}), \ldots, \varphi_{2^{J_2}}(U_{2,i}))$ in (A.16) below), and thus it holds that

$$Y_i = \boldsymbol{V}_i^\top \boldsymbol{\gamma} + \overline{f}_1(U_{1,i}) + \overline{g}_1(U_{2,i}) + \sigma\varepsilon_i, \tag{A.16}$$

where $\boldsymbol{\gamma} := \boldsymbol{\Sigma}^{1/2}(\gamma_{1,1}, \ldots, \gamma_{1,2^{J_1}}, \gamma_{2,1}, \ldots, \gamma_{2,2^{J_2}})$, and $\boldsymbol{V}_i := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{U}_i$.

We now calculate the bias and variance of the estimator $\widehat{\sigma}^2_{\text{proj,add}}$ defined as

$$\widehat{\sigma}^2_{\text{proj,add}} := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 - \binom{n}{2}^{-1}\sum_{i<j} Y_i Y_j \boldsymbol{V}_i^\top \boldsymbol{V}_i.$$

Direct calculation shows that

$$\begin{aligned}
|*|\widehat{\sigma}^2_{\text{proj,add}} - \sigma^2 &= |*|\mathbb{E}\overline{f}_1^2(U_1) + \mathbb{E}\overline{g}_1^2(U_2) + 2\mathbb{E}\big(\overline{f}_1(U_1)\overline{g}_1(U_2)\big) - \big\|\mathbb{E}\big((\overline{f}_1(U_1) - \overline{g}_1(U_2))\boldsymbol{V}\big)\big\|_2^2 \\
&\lesssim \mathbb{E}\overline{f}_1^2(U_1) + \mathbb{E}\overline{g}_1^2(U_2) + \big\|\mathbb{E}\big(\overline{f}_1(U_1)\boldsymbol{V}\big)\big\|_2^2 + \|\mathbb{E}(\overline{g}_1(U_2)\boldsymbol{V})\|_2^2 \\
&\lesssim 2^{-2\alpha J_1} + 2^{-2\beta J_2} + \big\|\mathbb{E}\big(\overline{f}_1(U_1)\boldsymbol{V}\big)\big\|_2^2 + \|\mathbb{E}(\overline{g}_1(U_2)\boldsymbol{V})\|_2^2.
\end{aligned}$$

Moreover, we have

$$\big\|\mathbb{E}\big(\overline{f}_1(U_1)\boldsymbol{V}\big)\big\|_2^2 = \sup_{\|\boldsymbol{a}\|\le 1}\big(\mathbb{E}\big(\overline{f}_1(U_1)\boldsymbol{V}^\top\boldsymbol{a}\big)\big)^2 \le \mathbb{E}\overline{f}_1^2(U_1)\cdot\sup_{\|\boldsymbol{a}\|\le 1}\big\{\boldsymbol{a}^\top\mathbb{E}(\boldsymbol{V}\boldsymbol{V}^\top)\boldsymbol{a}\big\} \lesssim 2^{-2\alpha J_1},$$

where the last inequality again follows by the identity covariance of $\boldsymbol{V}$. Similarly, it holds that $\|\mathbb{E}(\overline{g}_1(U_2)\boldsymbol{V})\|_2^2 \lesssim 2^{-2\beta J_2}$. We therefore conclude that the bias of $\widehat{\sigma}^2_{\text{proj,add}}$ is smaller than the order $2^{-2\alpha J_1} + 2^{-2\beta J_2}$.

Next, we calculate the variance of $\widehat{\sigma}^2_{\text{proj,add}}$. For this, it suffices to upper bound the variance of $\sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n-1)$ and $\binom{n}{2}^{-1}\sum_{i<j} Y_i Y_j \boldsymbol{V}_i^\top \boldsymbol{V}_j$. The first variance is clearly

of the order n^{-1} under the boundedness of $f(\cdot)$ and $g(\cdot)$ and the fact that $\mathbb{E}\varepsilon_i^4 \le C_\varepsilon$, so we focus on the second variance. Direct calculation shows that

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i<j} Y_i Y_j \boldsymbol{V}_i^\top \boldsymbol{V}_j\right) \\
&= \sum_{i<j, i'<j'} \big(\mathbb{E}\big(Y_i Y_j Y_{i'} Y_{j'} (\boldsymbol{V}_i^\top \boldsymbol{V}_j)(\boldsymbol{V}_{i'}^\top \boldsymbol{V}_{j'})\big) - \mathbb{E}\big(Y_i Y_j \boldsymbol{V}_i^\top \boldsymbol{V}_j\big)\mathbb{E}\big(Y_{i'} Y_{j'} \boldsymbol{V}_{i'}^\top \boldsymbol{V}_{j'}\big)\big). \qquad \text{(A.17)}
\end{aligned}$$

When i, j, i', j' take four different values, the above summand is clearly 0. When they take 3 values ($i = i' < j < j'$), denoting $z_i := \overline{f}_1(U_{1,i}) + \overline{g}_1(U_{2,i})$, $i \in [n]$, we have

$$\begin{aligned}
&\mathbb{E}\big(Y_i^2 Y_j Y_{j'} (\boldsymbol{V}_i^\top \boldsymbol{V}_j)(\boldsymbol{V}_i^\top \boldsymbol{V}_{j'})\big) \\
&= \mathbb{E}\big((z_i + \boldsymbol{V}_i^\top \boldsymbol{\gamma} + \sigma\varepsilon_i)^2 (z_j + \boldsymbol{V}_j^\top \boldsymbol{\gamma} + \sigma\varepsilon_j)(z_{j'} + \boldsymbol{V}_{j'}^\top \boldsymbol{\gamma} + \sigma\varepsilon_{j'})(\boldsymbol{V}_i^\top \boldsymbol{V}_j)(\boldsymbol{V}_i^\top \boldsymbol{V}_{j'})\big).
\end{aligned}$$

Next, we expand the above display and upper bound each term individually. For simplicity, we only show the calculation for the following dominating term and the other terms follow similarly.

$$\begin{aligned}
&\mathbb{E}\big((\boldsymbol{V}_i^\top \gamma)^2 (\boldsymbol{V}_j^\top \boldsymbol{\gamma})(\boldsymbol{V}_{j'}^\top \boldsymbol{\gamma})(\boldsymbol{V}_i^\top \boldsymbol{V}_j)(\boldsymbol{V}_i^\top \boldsymbol{V}_{j'})\big) \\
&= \mathbb{E}\big((\boldsymbol{V}_j^\top \boldsymbol{\gamma})\boldsymbol{V}_j^\top \big((\boldsymbol{V}_i^\top \boldsymbol{\gamma})^2 \boldsymbol{V}_i \boldsymbol{V}_i^\top\big) \boldsymbol{V}_{j'} (\boldsymbol{V}_{j'}^\top \boldsymbol{\gamma})\big) \\
&= \mathbb{E}\big(\boldsymbol{\gamma}^\top \boldsymbol{V}_j \boldsymbol{V}_j^\top\big)\mathbb{E}\big((\boldsymbol{V}_i^\top \boldsymbol{\gamma})^2 \boldsymbol{V}_i \boldsymbol{V}_i^\top\big)\mathbb{E}\big(\boldsymbol{V}_{j'} \boldsymbol{V}_{j'}^\top \boldsymbol{\gamma}\big) \\
&= \boldsymbol{\gamma}^\top \mathbb{E}\big((\boldsymbol{V}_i^\top \boldsymbol{\gamma})^2 \boldsymbol{V}_i \boldsymbol{V}_i^\top\big)\boldsymbol{\gamma} \\
&= \mathbb{E}(\boldsymbol{V}_i^\top \boldsymbol{\gamma})^4,
\end{aligned}$$

where in the second equality we use the independence of $\boldsymbol{V}_i, \boldsymbol{V}_j, \boldsymbol{V}_{j'}$ and in the third equality we use the fact that $\mathbb{E}(\boldsymbol{V}_i^\top \boldsymbol{V}_i) = \operatorname{Cov}(\boldsymbol{V}_i) = \mathbf{I}_L$ since $\mathbb{E}\boldsymbol{V}_i = \boldsymbol{\Sigma}^{-1/2}\mathbb{E}\boldsymbol{U}_i = 0$, where $L := 2^{J_1} + 2^{J_2}$. Moreover, by definition, it holds that $|*|\boldsymbol{V}_i^\top \boldsymbol{\gamma} = |*|f(X_i) + g(W_i) - z_i \le \|f\|_\infty + \|g\|_\infty + |z_i| \le C$ due to the boundedness of $f(\cdot), g(\cdot)$ and $|z_i|$. This concludes that when i, j, i', j' take three different values, the summand in (A.17) can be upper bounded by a fixed constant. When they take two different values ($i = i', j = j'$), we have

$$\mathbb{E}\big(Y_i^2 Y_j^2 (\boldsymbol{V}_i^\top \boldsymbol{V}_j)^2\big) = \mathbb{E}\big((z_i + \boldsymbol{V}_i^\top \boldsymbol{\gamma} + \sigma\varepsilon_i)^2 (z_j + \boldsymbol{V}_j^\top \boldsymbol{\gamma} + \sigma\varepsilon_j)^2 (\boldsymbol{V}_i^\top \boldsymbol{V}_j)^2\big).$$

We again upper bound the dominating term in the expansion of the above display.

$$\mathbb{E}\big((\boldsymbol{V}_i^\top\boldsymbol{\gamma})^2(\boldsymbol{V}_j^\top\boldsymbol{\gamma})^2(\boldsymbol{V}_i^\top\boldsymbol{V}_j)^2\big) \le C\mathbb{E}(\boldsymbol{V}_i^\top\boldsymbol{V}_j)^2 = C\mathrm{Tr}(\mathbf{I}_L) = CL,$$

where we again use the fact that $(\boldsymbol{V}_i^\top\boldsymbol{\gamma})$ and $(\boldsymbol{V}_j^\top\boldsymbol{\gamma})$ are bounded by a fixed constant. Putting together the pieces, we obtain that

$$\mathrm{Var}\big(\widehat{\sigma}^2_{\text{proj,add}}\big) \le C(n^{-1} + n^{-4}(n^3 + n^2 L)) = C\frac{n + 2^{J_1} + 2^{J_2}}{n^2}.$$

Optimal choice of $2^{J_1} \asymp n^{2/(4\alpha+1)}$ and $2^{J_2} \asymp n^{2/(4\beta+1)}$ then gives the desired error bound. $\square$

A.4.6 Proof of Proposition 12

Proof. We use $\mathcal{H}_*$ as a shorthand for $\mathcal{H}_{\delta^*}$. Fix any $\alpha \ge \alpha_*$. Define the oracle bandwidth

$$h^* := \begin{cases} n^{-1}, & \alpha > 1/4, \\ \max\Big\{h \in \mathcal{H}_* : h^{2\alpha} \le c\sqrt{\log n/(n^2 h)}\Big\}, & 0 < \alpha \le 1/4, \end{cases}$$

for some positive constant c to be specified later. When $0 < \alpha \le 1/4$, h^* is taken to be $n^{-2/(4\alpha+1)}$ if the set being maximized is empty. If not, then it holds that $(2h^*)^{2\alpha} > c\sqrt{\log n/(2n^2h^*)}$, and thus $(h^*)^{2\alpha} \asymp \sqrt{\log n/(n^2h^*)}$, or $h^* \asymp (\log n/n^2)^{1/(4\alpha+1)}$.

We first prove that with high probability, we have $\widehat{h}_{\delta_*} \ge h^*$. For this, we have

$$\begin{aligned}
\mathbb{P}\Big(\widehat{h}_{\delta_*} < h^*\Big) &\le \mathbb{P}\Big(\exists h \in \mathcal{H}_*, h \le h^*, |*|\widehat{\sigma}^2(h) - \widehat{\sigma}^2(h^*) \ge \tau\sqrt{\log n/(n^2h)}\Big) \\
&\le \sum_{h\in\mathcal{H}_*, h\le h^*} \mathbb{P}\Big(|*|\widehat{\sigma}^2(h) - \widehat{\sigma}^2(h^*) \ge \tau\sqrt{\log n/(n^2h)}\Big) \\
&\le \sum_{h\in\mathcal{H}_*, h\le h^*} \mathbb{P}\Big(|*|\widehat{\sigma}^2(h) - \sigma^2 \ge \frac{\tau}{2}\sqrt{\log n/(n^2h)}\Big) + |*|\mathcal{H}_* \cdot \mathbb{P}\Big(|*|\widehat{\sigma}^2(h^*) - \sigma^2 \ge \frac{\tau}{2}\sqrt{\log n/(n^2h^*
\end{aligned}$$

We now upper bound each probability in the above summation for any $h \le h^*$. As in the proof of Theorem 1, denote the two U-statistics on the numerator and denominator of $\widehat{\sigma}^2(h)$ as U_1,U_2, with corresponding mean values θ_1, θ_2. That is,

$$\theta_1 := \mathbb{E}\big\{K_h(X_i - X_j)(Y_i - Y_j)^2/2\big\} \text{ and } \theta_2 := \mathbb{E}K_h(X_i - X_j).$$

Define the "good" event $\mathcal{E} := \{U_2 \geq \theta_2/2\}$ and $\mathcal{E}^c$ as its complement, then it holds that

$$\begin{aligned}
&\mathbb{P}\Big(|*|\widehat{\sigma}^2(h) - \sigma^2 \geq \frac{\tau}{2}\sqrt{\log n/(n^2 h)} \bigcap \mathcal{E}\Big) \\
&\leq \mathbb{P}\bigg(|*|U_1 - U_2\sigma^2 \geq \frac{\tau}{\theta_2}\sqrt{\log n/(n^2 h)}\bigg) \\
&= \mathbb{P}\bigg(|*|(U_1-\theta_1) + (\theta_1 - \theta_2\sigma^2) + (U_2 - \theta_2) \geq \frac{\tau}{\theta_2}\sqrt{\log n/(n^2 h)}\bigg) \\
&\leq \mathbb{P}\bigg(|*|(U_1-\theta_1) \geq \frac{\tau}{4\theta_2}\sqrt{\log n/(n^2 h)}\bigg) + \mathbb{P}\bigg(|*|(U_2-\theta_2) \geq \frac{\tau}{4\theta_2}\sqrt{\log n/(n^2 h)}\bigg),
\end{aligned}$$

where the last inequality follows from the fact $h \leq h^*$ and the bound $|*|\theta_1 - \theta_2\sigma^2 \lesssim h^{2(\alpha\wedge 1)}$ calculated in the proof Theorem 1. By choose $u \asymp \log n$ and $v \asymp \log n/(nh)$ in Lemma 9 and τ to be sufficiently large, it holds that

$$\mathbb{P}\bigg(|*|U_2 - \theta_2 \geq \frac{\tau}{4\theta_2}\sqrt{\log n/(n^2 h)}\bigg) \lesssim n^{-C}$$

for arbitrarily large C. Furthermore, for sufficiently large τ and η in Lemma 24 below, choosing the same u and v yields that

$$\mathbb{P}\bigg(|*|U_1 - \theta_1 \geq \frac{\tau}{4\theta_2}\sqrt{\log n/(n^2 h)}\bigg) \lesssim n^{-C}$$

for arbitrarily large C. This, combined with the calculation

$$\mathbb{P}(\mathcal{E}^c) \lesssim \exp(-\theta_2^2 n/16) + \exp(-\theta_2^2 n^2 h/16)$$

in the proof of Theorem 1, concludes that $\mathbb{P}(\widetilde{\mathcal{E}}^c) \lesssim n^{-C}$, where $\widetilde{\mathcal{E}} := \Big\{\widehat{h}_{\delta_*} < h^*\Big\}$. Therefore, we have

$$\begin{aligned}
\mathbb{E}\big(\widehat{\sigma}^2_{\text{adapt}} - \sigma^2\big)^2 &\lesssim \mathbb{E}\Big\{\big(\widehat{\sigma}^2_{\text{adapt}} - \sigma^2\big)^2 \mathbb{1}\Big\{\widetilde{\mathcal{E}}^c\Big\}\Big\} + \mathbb{E}\Big\{\big(\widehat{\sigma}^2_{\text{adapt}} - \widehat{\sigma}^2(h^*)\big)^2 \mathbb{1}\Big\{\widetilde{\mathcal{E}}\Big\}\Big\} + \mathbb{E}\Big\{\big(\widehat{\sigma}^2(h^*) - \sigma^2\big)^2\Big\} \\
&\lesssim n^{-C} + \frac{\log n}{n^2 h^*} + \big(n^{-1} + (h^*)^{4(\alpha\wedge 1)} + (n^2 h^*)^{-1}\big) \\
&\lesssim \left(\frac{\log n}{n^2}\right)^{4\alpha/(4\alpha+1)} + n^{-1}.
\end{aligned}$$

This completes the proof. $\square$

A.4.7 Proof of Proposition 13

Proof. Note that the desired result is equivalent to the following statement:

$$\inf_{\widetilde{\sigma}^2} \max\left\{ \sup_{\substack{f\in\Lambda_{\alpha_1}(C_{\mathcal{F}})\\ \sigma^2\le C_\sigma, \mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}}} \mathbb{E}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha_1}\big)^2, \sup_{\substack{f\in\Lambda_{\alpha_2}(C_{\mathcal{F}})\\ \sigma^2\le C_\sigma, \mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}}} \mathbb{E}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha_2}\big)^2 \right\} \ge c$$

for sufficiently large n and sufficiently small c. By applying Lemma 25 with $\mathcal{A}=\{\alpha_1,\alpha_2\}$ with $\alpha_*\le\alpha_1<\alpha_2$, it suffices to lower bound the adaptive minimax rate under measure $\widetilde{\mathbb{P}}$ defined therein. More precisely, we will prove that for $n\ge n_0$ with some sufficiently large n_0,

$$\inf_{\widetilde{\sigma}^2} \max\left\{ \sup_{\substack{f\in\Lambda_{\alpha_1}(C_{\mathcal{F}})\\ \sigma^2\le C_\sigma, \mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}}} \mathbb{E}_{\widetilde{\mathbb{P}}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha_1}\big)^2, \sup_{\substack{f\in\Lambda_{\alpha_2}(C_{\mathcal{F}})\\ \sigma^2\le C_\sigma, \mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}}} \mathbb{E}_{\widetilde{\mathbb{P}}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha_2}\big)^2 \right\} > c$$

for some sufficiently small positive c. In order to show this, we will prove that, for any $n\ge n_0$ and any estimator $\widetilde{\sigma}^2$, if

$$\sup_{\sigma^2\le C_\sigma} \sup_{f\in\Lambda_{\alpha_2}(C_{\mathcal{F}})} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha_2}\big)^2 \le c, \tag{A.18}$$

then it holds that

$$\sup_{\sigma^2\le C_\sigma} \sup_{f\in\Lambda_{\alpha_1}(C_{\mathcal{F}})} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}^{\text{adapt}}_{\text{cv},(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha_1}\big)^2 > c. \tag{A.19}$$

If $\alpha_2>1/4$, then $\phi_{n,\alpha_2}\asymp n^{-1/2}$, and we can choose a sufficiently small c such that (A.18) never holds. Therefore, in what follows, we will assume $\alpha_*\le\alpha_1<\alpha_2\le 1/4$, in which case $\phi_{n,\alpha_i}\asymp(\log n/n^2)^{2\alpha_i/(4\alpha_i+1)}$ for $i=1,2$.

We will now apply Lemma 26. To this end, we adopt a two-point method and introduce the two probability measures $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$ (conditioning on a fixed realization of $m\sim\text{Poi}(2n)$) as follows. Introduce

$$h_n = c\bigg(\frac{\log n}{n^2}\bigg)^{1/(4\alpha_1+1)} \quad \text{and } N := N_n := h_n^{-1},$$

where c is some sufficiently small constant tuned such that N is a positive integer. Let Q be a discrete distribution that takes value 0 with probability 1/2, -1 with probability 1/4 and 1 with probability 1/4. It then can be readily checked that $m_1(Q) = 0$ and $m_2(Q) = 1/2$.

Choice of ε: Under H_0, let $\varepsilon \sim (1 + h_n^{2\alpha_1}/2)^{-1/2}((h_n^{\alpha_1} Q) * \mathcal{N}(0,1))$. Under H_1, let $\varepsilon \sim \mathcal{N}(0,1)$.

Choice of σ^2: Under H_0, let $\sigma^2 = 1 + h_n^{2\alpha_1}/2$. Under H_1, let $\sigma^2 = 1$.

Choice of X: Under both H_0 and H_1, let X be uniformly distributed on the union of the intervals $[(6i-5)h_n, (6i-1)h_n]$ for $i \in [N]$.

Choice of $f(\cdot)$: Under H_0, let $f \equiv 0$. Under H_1, let f take the value $h_n^{\alpha_1} r_i$ on $[(6i - 5)h_n, (6i-1)h_n]$, where $\{r_i\}_{i=1}^N$ are N i.i.d. variables with law Q.

Clearly, by the boundedness of Q, H_0 belongs to the model class indexed by the smoothness index α_2, and H_1 belongs to the model class indexed by α_1. Moreover, the absolute difference in σ^2 under H_0 and H_1 is lower bounded by the order $h_n^{2\alpha_1} \asymp (\log n/n^2)^{2\alpha_1/(4\alpha_1+1)}$.

Denote $\widetilde{p}_0$ and $\widetilde{p}_1$ as the densities of $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$. Define $f_{\max} := \lceil 2/(1-4\alpha_1) \rceil + 1$ and let d_i be the number of X's that fall into $[(6i-5)h_n, (6i-1)h_n]$ for each $i \in [N]$. Consider the following event

$$\mathcal{E} := \left\{ \{m \le 3n\} \bigcap \{ \max_{1 \le i \le N} d_i \le f_{\max} \} \right\}.$$

Note that under both $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$, the sequence $\{d_i\}_{i=1}^N$ are i.i.d. Poisson variables with mean $2n/N$. Thus, a standard Poisson tail estimate and Lemma 27 imply that the event $\mathcal{E}$ has asymptotic probability 1 under both $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$. Next, we calculate the χ^2 distance between

$\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$ conditioning on the event $\mathcal{E}$. First, we have

$$\begin{aligned}\int \frac{\widetilde{p}_1^2}{\widetilde{p}_0}\mathbb{1}\{\mathcal{E}\} &= \int p(m)p(x_1,\ldots,x_m)\mathbb{1}\{\mathcal{E}\}\int\prod_{j=1}^N \frac{p_{1,j}^2}{p_{0,j}} \\ &= \int p(m)p(x_1,\ldots,x_m)\mathbb{1}\{\mathcal{E}\}\int\prod_{j=1}^N \big(1+\chi^2(p_{1,j},p_{0,j})\big),\end{aligned}$$

where $p(m)$ and $p(x_1,\ldots,x_m)$ are the pmf and pdf of m and $\{X_i\}_{i=1}^m$ under both $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_1$, and $p_{0,j}$ is the conditional density of all those Y_i's with corresponding $X_i \in [(6j-5)h_n,(6j-1)h_n]$ and similarly for $p_{1,j}$.

We now upper bound each $\chi^2(p_{0,j},p_{1,j})$, where we assume that there are d_j X_i's that belong to $[(6j-5)h_n,(6j-1)h_n]$. Write d instead of d_j for short. Here, d is a random variable that only depends on m and $\{X_i\}_{i=1}^m$, and on the event $\mathcal{E}$, we have $d \le f_{\max}$. Clearly, if $d=0$ or 1, then $\chi^2(p_{0,j},p_{1,j})=0$. Assume $d\ge 2$. Assume for simplicity that the d data points are $y_1,\ldots,y_d$. Then, by definition of $\widetilde{\mathbb{P}}_0$, we have

$$\widetilde{p}_{0,j} \ge (1/2)^{f_{\max}}\varphi(y_1)\ldots\varphi(y_d) \ge c\varphi(y_1)\ldots\varphi(y_d).$$

On the other hand, we have by direct calculation

$$\begin{aligned}\int \widetilde{p}_{0,j}^2/(\varphi(y_1)\ldots\varphi(y_d)) &= \mathbb{E}_{\substack{s_1,\ldots,s_d\sim h^{\alpha_1}Q \\ \widetilde{s}_1,\ldots,\widetilde{s}_d\sim h^{\alpha_1}Q}}\exp\left(\sum_{i=1}^d s_i\widetilde{s}_i\right), \\ \int \widetilde{p}_{1,j}^2/(\varphi(y_1)\ldots\varphi(y_d)) &= \mathbb{E}_{t,\widetilde{t}\sim h^{\alpha_1}Q}\exp(dt\widetilde{t}), \\ \int \widetilde{p}_{0,j}\widetilde{p}_{1,j}/(\varphi(y_1)\ldots\varphi(y_d)) &= \mathbb{E}_{t,s_1,\ldots,s_d\sim h^{\alpha_1}Q}\exp\left(\sum_{i=1}^d ts_i\right).\end{aligned}$$

We therefore conclude that

$$\begin{aligned}\chi^2(p_{0,j},p_{1,j}) &\lesssim \sum_{k=1}^\infty \frac{1}{k!}\sum_{1\le i_1,\ldots,i_k\le d}\big(\mathbb{E}_{s_1,\ldots,s_d\sim h^{\alpha_1}Q}(s_{i_1}\ldots s_{i_k}) - \mathbb{E}_{t\sim h^{\alpha_1}Q}t^k\big)^2 \\ &:= \sum_{k=1}^\infty \frac{1}{k!}\sum_{1\le i_1,\ldots,i_k\le d}\Delta_{i_1,\ldots,i_k}^2.\end{aligned}$$

For any $k \geq 1$, if $(i_1, \ldots, i_k)$ are all identical, then $\Delta_{i_1,\ldots,i_k} = 0$. More generally, if $(i_1, \ldots, i_k)$ take ℓ different values, where $\ell \leq d \leq f_{\max}$ on the event $\mathcal{E}$, there are $\binom{d}{\ell}$ ways of choosing ℓ different values among $[d]$, and there are a total of $\binom{k-1}{\ell-1}$ ways to distribute ℓ values in $(i_1, \ldots, i_k)$, thus we obtain the estimate

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{1}{k!} \sum_{1 \leq i_1,\ldots,i_k \leq d} \Delta^2_{i_1,\ldots,i_k} &\leq \sum_{k=2}^{\infty} \frac{1}{k!} \sum_{\ell=2}^{k} h^{2\alpha_1 k} \binom{d}{\ell} \binom{k-1}{\ell-1} \\ &= \sum_{\ell=2}^{f_{\max}} \binom{d}{\ell} \sum_{k=\ell}^{\infty} \frac{1}{k!} h^{2\alpha_1 k} \binom{k-1}{\ell-1}.\end{aligned}$$

For each $2 \leq \ell \leq f_{\max}$, by the Stirling's formula, we have

$$\sum_{k=\ell}^{\infty} \frac{1}{k!} h^{2\alpha_1 k} \binom{k-1}{\ell-1} \lesssim \sum_{k=\ell}^{\infty} \frac{h^{2\alpha_1 k}}{k^{k+1/2} e^{-k}} \left(\frac{ek}{\ell}\right)^k \lesssim \sum_{k=\ell}^{\infty} \frac{(e^2 h^{2\alpha_1})^k}{k^{1/2} \ell^k} \lesssim h^{2\alpha_1 \ell}.$$

Next, using the trivial bound $\binom{d}{\ell} \leq d^2 f_{\max}^{\ell-2}$, we obtain that $\chi^2(\widetilde{p}_{0,j}, \widetilde{p}_{1,j}) \lesssim d^2 h^{4\alpha_1}$ if $d \geq 2$. This implies that

$$\int \frac{\widetilde{p}_1^2}{\widetilde{p}_0} \mathbb{1}\{\mathcal{E}\} \lesssim \int p(m) p(x_1, \ldots, x_m) \prod_{j=1}^{N} (1 + \chi_j^2),$$

where $\chi_j^2 = 0$ for $d_j = 0, 1$ and $\chi_j^2 \leq d_j^2 h^{4\alpha_1}$ for $d_j \geq 2$. Next,

$$\prod_{j=1}^{N} (1 + \chi_j^2) = 1 + \sum_{j=1}^{N} \chi_j^2 + \sum_{1 \leq i < j \leq N} \chi_i^2 \chi_j^2 + \ldots + \sum_{1 \leq i_1 < \ldots < i_N \leq N} \chi_{i_1}^2 \ldots \chi_{i_N}^2.$$

Consider the kth term in the above display. Note that on the event $\mathcal{E}$, we have $\sum_{j=1}^{N} d_j \leq 3n$, thus there are at most $(3n/2)$ j's with $d_j \geq 2$. Then, using the estimate $\binom{n}{k} \leq n^k / k!$, we have

$$\sum_{1 \leq i_1 < \ldots < i_k \leq N} \chi_{i_1}^2 \ldots \chi_{i_k}^2 \leq \binom{(3n/2)}{k} h^{4k\alpha_1} d_{i_1}^2 \ldots d_{i_k}^2 \leq \frac{C^k n^k}{k!} h^{4k\alpha_1} d_{i_1}^2 \ldots d_{i_k}^2.$$

We therefore conclude that

$$\int \frac{\widetilde{p}_1^2}{\widetilde{p}_0} \mathbb{1}\{\mathcal{E}\} \leq \mathbb{E}_{d_1,\ldots,d_N \sim \mathrm{Poi}(2n/N)} \left(1 + \sum_{k=1}^{N} \frac{C^k n^k h^{4\alpha_1 k}}{k!} d_1^2 \ldots d_k^2\right) \leq 1 + \sum_{k=1}^{\infty} \frac{(c \log n)^k}{k!} = n^c$$

for some sufficiently small constant c. Then, by Lemma 26 and the calculation

$$\frac{\varepsilon\sqrt{I_{\mathcal{E}}}}{\Delta} \lesssim \frac{n^{c/2}(\log n/n^2)^{4\alpha_2/(4\alpha_2+1)}}{(\log n/n^2)^{4\alpha_1/(4\alpha_1+1)}} \to 0,$$

where Δ, I, ε are defined as in Lemma 26, we conclude that for sufficiently large n and any considered estimator $\widetilde{\sigma}^2$, if (A.18) holds, then (A.19) will follow. This shows that, even over two smooth classes $\alpha \in \{\alpha_1, \alpha_2\}$, the adaptive minimax rate can be no faster than $\phi_{n,\alpha}$. $\square$

A.4.8 Supporting Lemmas

Lemma 22. *Suppose $f \in \Lambda_{\boldsymbol{\alpha}}(C_{\mathcal{F}})$ and $\sigma^2 \leq C_\sigma$for some fixed constants $C_{\mathcal{F}}, C_\sigma$, and the joint distribution of $(\boldsymbol{X}, \varepsilon)$ satisfies the conditions in $\mathcal{P}_{mcv,(\boldsymbol{X},\varepsilon)}$. Then, the U-statistic U_1 defined in the proof of Proposition 1 satisfies*

$$\mathbb{E}(U_1 - \theta_1)^2 \leq C(n^{-1} \vee n^{-2}(h_1h_2)^{-1})$$

for some positive constant C that only depends on $\overline{M}_K, \underline{M}_K, \boldsymbol{\alpha}, C_{\mathcal{F}}, C_\sigma, C_0, C_\varepsilon$.

Proof. Denote g as the kernel of U_1, that is,

$$g(\boldsymbol{D}_i, \boldsymbol{D}_j) := K_{h_1}(X_{i,1} - X_{j,1})K_{h_2}(X_{i,2} - X_{j,2})(Y_i - Y_j)^2/2, \quad \boldsymbol{D}_i := (\boldsymbol{X}_i, \varepsilon_i)^\top.$$

Then, it holds that

$$\mathrm{Var}(U_1) = \binom{n}{2}^{-1} \sum_{i<j, i'<j'} \mathbb{E}\{(g(\boldsymbol{D}_i, \boldsymbol{D}_j) - \theta_1)(g(\boldsymbol{D}_{i'}, \boldsymbol{D}_{j'}) - \theta_1)\}.$$

When i, j, i', j' take four different values, the expectation is zero. Using a similar argument as in the proof of Lemma 8, when i, j, i', j' take three different values, it holds that

$$\mathbb{E}\{(g(\boldsymbol{D}_i, \boldsymbol{D}_j) - \theta_1)(g(\boldsymbol{D}_{i'}, \boldsymbol{D}_{j'}) - \theta_1)\} = O(1).$$

When they take two different values,

$$\mathbb{E}\{(g(\boldsymbol{D}_i, \boldsymbol{D}_j) - \theta_1)(g(\boldsymbol{D}_{i'}, \boldsymbol{D}_{j'}) - \theta_1)\} = O((h_1h_2)^{-1}).$$

We therefore conclude that

$$\mathrm{Var}(U_1) \lesssim \frac{n^3 + n^2(h_1h_2)^{-1}}{n^4} \asymp n^{-1} + n^{-2}(h_1h_2)^{-1}.$$

This completes the proof. □

Lemma 23. *Suppose $h_1h_2 \gtrsim n^{-(2-\delta)}$ for some $0 < \delta < 2$, and the joint distribution of $(\boldsymbol{X}, \varepsilon)$ satisfies the conditions in $\mathcal{P}_{mcv,(\boldsymbol{X},\varepsilon)}$ Then, for any $u, v > 0$, the U-statistic U_2 defined in the proof of Theorem 1 satisfies*

$$\mathbb{P}\big(|*|U_2 - \theta_2 \geq C(v^{1/2}n^{-1/2} + u^{1/2}n^{-1}(h_1h_2)^{-1/2})\big) \leq C(\exp(-u) + \exp(-v))$$

for sufficiently large n and

$$\mathbb{E}(U_2 - \theta_2)^2 \leq C(n^{-1} \vee n^{-2}(h_1h_2)^{-1}),$$

where C is some positive constant that only depends on $\overline{M}_K, \underline{M}_K, \boldsymbol{\alpha}, C_0$.

Proof. The proof is similar to that of Lemma 9. In the application of Lemma 13, the five quantities are of the order $B_1 \lesssim 1$, $B_2 \lesssim (h_1h_2)^{-1}$, $B_3 \leq n^{1/2}(h_1h_2)^{-1/2}$, $\nu_1^2 \lesssim 1$, $\nu_2^2 \lesssim (h_1h_2)^{-1}$. Therefore, for any $u, v > 0$, it holds that

$$\mathbb{P}(|U_2 - \theta_2| \geq a_1v^{1/2} + a_2v + b_1u^{1/2} + b_2u + b_3u^{3/2} + b_4u^2) \leq C(\exp(-v) + \exp(-u)),$$

where $a_1 \lesssim n^{-1/2}, a_2 \lesssim n^{-1}, b_1 \lesssim n^{-1}(h_1h_2)^{-1/2}, b_2 \lesssim n^{-1}, b_3 \lesssim n^{-3/2}(h_1h_2)^{-1/2}, b_4 \lesssim n^{-2}(h_1h_2)^{-1}$. Under the condition that $h_1h_2 = \Omega(n^{-(2-\delta)})$ for some $\delta > 0$ and n is sufficiently large, the dominant terms in the above inequality are a_1 and b_1, that is,

$$n^{-1/2} \vee n^{-1}(h_1h_2)^{-1/2}.$$

This proves the first part of the theorem. The expectation version follows by Lemma 10. □

Lemma 24. *Suppose the conditions of Proposition 12 hold. For the U-statistic U_1 defined therein, suppose $h \gtrsim n^{-(2-\delta)}$ for some $0 \leq \delta \leq 2$. Then, for any $\eta > 0$, there exists some positive constant $C = C(\overline{M}_K, \underline{M}_K, \alpha, \eta)$ such that*

$$\mathbb{P}\big(|*|U_1 - \theta_1 \geq C(v^{1/2}n^{-1/2} + u^{1/2}n^{-1}h^{-1/2})\big) \leq C(\exp(-u) + \exp(-v)) + n^{-\eta}.$$

Proof. Denote g_1 as the kernel of U_1, and consider its truncated version of defined as

$$\overline{g}_1(D_i, D_j) := \frac{1}{2h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)\{(f(X_i) - f(X_j)) + \widetilde{\varepsilon}_{ij}\}^2 \mathbb{1}\{|\varepsilon_i| \leq \kappa_n\}\mathbb{1}\{|\varepsilon_j| \leq \kappa_n\},$$

where $D_i = (X_i, \varepsilon_i)$ and κ_n is some truncation parameter satisfying $\kappa_n \uparrow \infty$ as $n \to \infty$ to be specified later. We first consider the concentration of $\overline{g}_1$ around its mean value $\overline{\theta}_1 := \mathbb{E}\{\overline{g}(D_i, D_j)\}$. For this, we will make use of Lemma 13 by upper bounding the 5 quantities $B_1, B_2, B_3, \nu_1^2, \nu_2^2$ therein.

For B_1, denoting $\widetilde{g}_1(D) = \mathbb{E}\{\overline{g}_1(D, D_j) \mid D\}$, it holds that

$$\begin{aligned}
\widetilde{g}_1(D_i) &= \mathbb{E}\left\{\frac{1}{2h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)((f(X_i) - f(X_j)) + \widetilde{\varepsilon}_{ij})^2 \mathbb{1}\{|\varepsilon_i| \leq \kappa_n\}\mathbb{1}\{|\varepsilon_j| \leq \kappa_n\} \mid D_i\right\} \\
&\lesssim \mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)(|X_i - X_j|^{2\alpha} + \widetilde{\varepsilon}_{ij}^2)\mathbb{1}\{|\varepsilon_i| \leq \kappa_n\}\mathbb{1}\{|\varepsilon_j| \leq \kappa_n\} \mid D_i\right\} \\
&\lesssim \mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)|\widetilde{X}_{ij}|^{2\alpha} \mid X_i\right\} + \mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)\widetilde{\varepsilon}_{ij}^2 \mathbb{1}\{|\varepsilon_i| \leq \kappa_n\} \mid X_i, \varepsilon_i\right\}.
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
&\mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)|\widetilde{X}_{ij}|^{2\alpha} \mid X_i\right\} = \int \frac{1}{h}K\left(\frac{u - X_i}{h}\right)|u - X_i|^{2\alpha}p_X(u)du \\
&= \int K(v)|vh|^{2\alpha}p_X(X_i + vh)dv \leq \sup_{u \in \mathbb{R}} p_X(u)h^{2\alpha}\int K(v)|v|^{2\alpha}dv \lesssim h^{2\alpha}.
\end{aligned}$$

For the second term, we obtain similarly that

$$\mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)\widetilde{\varepsilon}_{ij}^2 \mathbb{1}\{|\varepsilon_i| \leq \kappa_n\} \mid X_i, \varepsilon_i\right\} \lesssim (\varepsilon_i^2 + \sigma^2)\mathbb{1}\{|\varepsilon_i| \leq \kappa_n\} \lesssim \kappa_n^2 + \sigma^2.$$

Putting together the pieces and using the fact that $\kappa_n \uparrow \infty$ as $n \to \infty$, we obtain that

$$B_1 = \|\widetilde{g}_1\|_\infty \lesssim \kappa_n^2.$$

Moreover, with similar analysis, it can be readily checked that $\nu_1^2 \lesssim 1$.

For B_2, it holds that

$$\begin{aligned}
|*|\overline{g}_1(D_i, D_j) &\lesssim \frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)\big\{(f(X_i) - f(X_j))^2 + \widetilde{\varepsilon}_{ij}^2\big\}\mathbb{1}\{|\varepsilon_i| \le \kappa_n\}\mathbb{1}\{|\varepsilon_j| \le \kappa_n\} \\
&\lesssim \frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)\Big\{|\widetilde{X}_{ij}|^{2\alpha} + \widetilde{\varepsilon}_{ij}^2\Big\}\mathbb{1}\{|\varepsilon_i| \le \kappa_n\}\mathbb{1}\{|\varepsilon_j| \le \kappa_n\} \\
&\lesssim \frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)|*|\frac{\widetilde{X}_{ij}}{h}^{2\alpha} h^{2\alpha} + \frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)\kappa_n^2 \lesssim \frac{1}{h}\kappa_n^2.
\end{aligned}$$

We therefore conclude that $B_2 = \|\overline{g}_1\|_\infty \lesssim h^{-1}\kappa_n^2$.

For B_3, we have

$$\begin{aligned}
B_3^2 &= n \sup_{D_i} \mathbb{E}\big\{\overline{g}_1^2(D_i, D_j) \mid D_i\big\} \\
&= n \sup_{D_i} \mathbb{E}\left\{\frac{1}{h^2} K^2\left(\frac{\widetilde{X}_{ij}}{h}\right)(f(X_i) - f(X_j) + \widetilde{\varepsilon}_{ij})^4 \mathbb{1}\{|*|\varepsilon_i \le \kappa_n\}\mathbb{1}\{|*|\varepsilon_j \le \kappa_n\} \mid X_i, \varepsilon_i\right\} \\
&\lesssim \frac{nM_K}{h}\mathbb{E}\left\{\frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)\big\{(f(X_i) - f(X_j))^4 + \widetilde{\varepsilon}_{ij}^4\big\}\mathbb{1}\{|*|\varepsilon_i \le \kappa_n\}\mathbb{1}\{|*|\varepsilon_j \le \kappa_n\} \mid X_i, \varepsilon_i\right\} \\
&\lesssim \frac{nM_K}{h}\mathbb{E}\left\{\frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)|*|\widetilde{X}_{ij}^{4\alpha} \mid X_i, \varepsilon_i\right\} + \frac{nM_K}{h}\mathbb{E}\left\{\frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)\widetilde{\varepsilon}_{ij}^4 \mathbb{1}\{|*|\varepsilon_i \le \kappa_n\}\mathbb{1}\{|*|\varepsilon_j \le \kappa_n\} \mid X_i, \varepsilon_i\right\}.
\end{aligned}$$

Now, using similar calculation in the analysis of B_1, it holds that

$$\begin{aligned}
&\mathbb{E}\left\{\frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)|*|\widetilde{X}_{ij}^{4\alpha} \mid X_i, \varepsilon_i\right\} \lesssim h^{4\alpha}, \\
&\mathbb{E}\left\{\frac{1}{h} K\left(\frac{\widetilde{X}_{ij}}{h}\right)\widetilde{\varepsilon}_{ij}^4 \mathbb{1}\{|*|\varepsilon_i \le \kappa_n\}\mathbb{1}\{|*|\varepsilon_j \le \kappa_n\} \mid X_i, \varepsilon_i\right\} \lesssim \kappa_n^4.
\end{aligned}$$

Putting together the pieces, we conclude that $B_3 \lesssim \kappa_n^4 n h^{-1}$.

Lastly, for ν_2^2, we have

$$\begin{aligned}
\nu_2^2 &= \mathbb{E}\{\overline{g}_1^2(D_i, D_j)\} \lesssim \mathbb{E}\left\{\frac{1}{h^2}K^2\left(\frac{\widetilde{X}_{ij}}{h}\right)(f(X_i) - f(X_j) + \widetilde{\varepsilon}_{ij})^4\right\} \\
&\lesssim h^{-1}\mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)|*|f(X_i) - f(X_j)^4\right\} + \frac{M_K}{h}\mathbb{E}\left\{\frac{1}{h}K\left(\frac{\widetilde{X}_{ij}}{h}\right)\widetilde{\varepsilon}_{ij}^4\right\} \\
&\lesssim h^{-1}(h^{4\alpha} + \mathbb{E}(\varepsilon_i^4)) \lesssim h^{-1}.
\end{aligned}$$

Define $\overline{U}_1$ to be the U-statistic generated by the kernel $\overline{g}_1$, that is, $\overline{U}_1 := \binom{n}{2}^{-1}\sum_{i<j}\overline{g}_1(D_i, D_j)$, and define $\overline{\theta}_1$ to be the mean value $\mathbb{E}\{\overline{g}_1(D_i, D_j)\}$. Define the event $\mathcal{E} := \{|*|\varepsilon_i \leq \kappa_n \text{ for all } i \in [n]\}$. Then, we have for any $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(|*|U_1 - \theta_1 \geq t) &= \mathbb{P}\Big(|*|\overline{U}_1 - \theta_1 \geq t \bigcap \mathcal{E}\Big) + \mathbb{P}(\mathcal{E}^c) \\
&\leq \mathbb{P}\big(|*|\overline{U}_1 - \overline{\theta}_1 \geq t - |*|\theta_1 - \overline{\theta}_1\big) + \mathbb{P}(\mathcal{E}^c).
\end{aligned}$$

For the first term, we have by Lemma 13 that, for any given $u, v > 0$, it holds that

$$\mathbb{P}\big(|*|\overline{U}_1 - \overline{\theta}_1 \geq a_1 v^{1/2} + a_2 v + b_1 u^{1/2} + b_2 u + b_3 u^{3/2} + b_4 u^2\big) \leq C(\exp(-u) + \exp(-v)),$$

where $a_1 \lesssim n^{-1/2}, a_2 \lesssim \kappa_n^2/n, b_1 \lesssim n^{-1}h^{-1/2}, b_2 \lesssim n^{-1}\kappa_n^2, b_3 \lesssim \kappa_n^2 n^{-3/2}h^{-1/2}, b_4 \lesssim n^{-2}h^{-1}\kappa_n^2$. Choosing $\kappa_n = \kappa\sqrt{\log n}$ for some sufficiently large constant κ, then as long as $h = \Omega(n^{-(2-\delta)})$ for some $\delta > 0$, then the dominant terms in the above inequality are a_1 and b_1, that is,

$$n^{-1/2} \vee n^{-1}h^{-1/2}.$$

Therefore, Lemma 10 implies that

$$\mathbb{E}\big(|*|\overline{U}_1 - \overline{\theta}_1\big) \leq C(n^{-1/2} \vee n^{-1}h^{-1/2}).$$

Now we calculate the difference between θ_1 and $\overline{\theta}_1$. By definition, we have

$$
\begin{aligned}
|*|\theta_1 - \overline{\theta}_1 &= |*|\mathbb{E}\bigg\{\frac{1}{h}K\bigg(\frac{\widetilde{X}_{ij}}{h}\bigg)(f(X_i) - f(X_j) + \widetilde{\varepsilon}_{ij})^2 \mathbb{1}\Big\{|\varepsilon_i| \geq \kappa_n \bigcup |\varepsilon_j| \geq \kappa_n\Big\}\bigg\} \\
&\lesssim |*|\mathbb{E}\bigg\{\frac{1}{h}K\bigg(\frac{\widetilde{X}_{ij}}{h}\bigg)(f(X_i) - f(X_j) + \widetilde{\varepsilon}_{ij})^2 \mathbb{1}\{|\varepsilon_i| \geq \kappa_n\}\bigg\} \\
&\lesssim |*|\mathbb{E}\bigg\{\frac{1}{h}K\bigg(\frac{\widetilde{X}_{ij}}{h}\bigg)(f(X_i) - f(X_j))^2 \mathbb{1}\{|\varepsilon_i| \geq \kappa_n\}\bigg\} + |*|\mathbb{E}\bigg\{\frac{1}{h}K\bigg(\frac{\widetilde{X}_{ij}}{h}\bigg)\widetilde{\varepsilon}_{ij}^2 \mathbb{1}\{|\varepsilon_i| \geq \kappa_n\}\bigg\} \\
&\lesssim h^{2\alpha}\mathbb{P}(|\varepsilon_i| \geq \kappa_n) + \mathbb{E}\big\{\varepsilon_i^2 \mathbb{1}(|\varepsilon_i| \geq \kappa_n)\big\} \\
&\lesssim h^{2\alpha} n^{-\kappa^2/(2\kappa_\varepsilon^2)} + \kappa_\varepsilon^2 n^{-\kappa^2/(4\kappa_\varepsilon^2)},
\end{aligned}
$$

where the second line is by symmetry, and in the last line we use the sub-Gaussianity of ε_i. Therefore, as long as $h = \Omega(n^{-(2-\delta)})$ for some $\delta > 0$, by choosing κ large enough (depending only on $\delta, \kappa_\varepsilon, u, v$), it holds that

$$
|*|\theta_1 - \overline{\theta}_1 = o(a_1 v^{1/2} + a_2 v + b_1 u^{1/2} + b_2 u + b_3 u^{3/2} + b_4 u^2).
$$

Lastly, by the sub-Gaussanity of ε_i, it holds that

$$
\mathbb{P}(\mathcal{E}^c) \leq n\mathbb{P}(|\varepsilon_i| \geq \kappa_n) \lesssim n^{-\eta}
$$

for sufficiently large η by choosing κ correspondingly large enough. This completes the proof. □

The following Poissonization lemma reduces the original problem of Proposition 13 into the case with a random sample size, which facilitates the calculation of χ^2 distance. We introduce some notation. Consider the following experiment: for any given positive integer n, $f(\cdot)$, σ, distribution $p_X(\cdot)$ of X, and distribution $p_\varepsilon(\cdot)$ of ε,

- generate $m \sim \mathrm{Poi}(2n)$;
- generate $X_1, \ldots, X_m \sim p_X$ and $\varepsilon_1, \ldots, \varepsilon_m \sim p_\varepsilon$;

- generate $Y_i = f(X_i) + \sigma\varepsilon_i$ for each $i \in [m]$.

Denote the original experiment and the above experiment as $\mathbb{P}$ and $\widetilde{\mathbb{P}}$, respectively, where we omit the dependence on n, $f(\cdot)$, σ, distribution $p_X(\cdot)$ of X, and distribution $p_\varepsilon(\cdot)$ of ε.

Lemma 25 (Poissonization)**.** *Let $\phi_{n,\alpha}$ be defined as in Proposition 13. For any fixed $\alpha_* > 0$ and set $\mathcal{A} \subset [\alpha_*, \infty)$, the following inequality holds:*

$$\begin{aligned}
&\inf_{\widetilde{\sigma}^2} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f\in\Lambda_\alpha(C_{\mathcal{F}})} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{cv,(X,\varepsilon)}} \mathbb{E}_{\mathbb{P}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2 \\
\ge\ &\inf_{\widetilde{\sigma}^2} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f\in\Lambda_\alpha(C_{\mathcal{F}})} \sup_{\mathbb{P}_{(X,\varepsilon)}\in\mathcal{P}_{cv,(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2 - 4nC_\sigma^2\exp(-n/6).
\end{aligned}$$

Proof. Define the event $\mathcal{E} := \{m \ge n\}$, where $m \sim \text{Poi}(2n)$. Then, a standard tail estimate has $\mathbb{P}(\mathcal{E}) \ge 1 - e^{-n/6}$. For the adaptive minimax rate under $\widetilde{\mathbb{P}}$, we have

$$\begin{aligned}
&\inf_{\widetilde{\sigma}^2} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2 \\
=\ &\inf_{\widetilde{\sigma}^2} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\Big\{\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2(\mathbb{1}\{\mathcal{E}\}+\mathbb{1}\{\mathcal{E}^c\})\Big\} \\
=\ &\inf_{\widetilde{\sigma}^2\le C_\sigma} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\Big\{\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2(\mathbb{1}\{\mathcal{E}\}+\mathbb{1}\{\mathcal{E}^c\})\Big\} \\
\le\ &\inf_{\widetilde{\sigma}^2\le C_\sigma} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\Big\{\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2\mathbb{1}\{\mathcal{E}\}\Big\} + 4nC_\sigma^2\cdot \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \widetilde{\mathbb{P}}(\mathcal{E}^c) \\
\le\ &\inf_{\widetilde{\sigma}^2} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\Big\{\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2\mathbb{1}\{\mathcal{E}\}\Big\} + 4nC_\sigma^2\exp(-n/6) \\
\le\ &\inf_{\widetilde{\sigma}^2=\widetilde{\sigma}^2\left(\{(X_i,Y_i)\}_{i=1}^n\right)} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\widetilde{\mathbb{P}}}\Big\{\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2\mathbb{1}\{\mathcal{E}\}\Big\} + 4nC_\sigma^2\exp(-n/6) \\
\le\ &\inf_{\widetilde{\sigma}^2} \sup_{\alpha\in\mathcal{A}} \sup_{\sigma^2\le C_\sigma} \sup_{f(\cdot),\mathbb{P}_{(X,\varepsilon)}} \mathbb{E}_{\mathbb{P}}\Big\{\big((\widetilde{\sigma}^2-\sigma^2)/\phi_{n,\alpha}\big)^2\Big\} + 4nC_\sigma^2\exp(-n/6).
\end{aligned}$$

This completes the proof. $\square$

The following lemma will also be used in the proof of Proposition 13, and is a slight variation of the constrained risk inequality derived in Brown and Low (1996) (see Theorem 1 therein). We first introduce some notation. Consider some measurable space equipped with a class of probability measures $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$, where (Θ, d) is a metric space. For each $\theta \in \Theta$, let f_θ be the density of $\mathbb{P}_\theta$ with respect to some common dominating measure ν, and denote by $\mathbb{E}_\theta$ the expectation under the measure $\mathbb{P}_\theta$. For any estimator T of θ, define its risk as

$$R_\theta := R(\theta, T) := \mathbb{E}_\theta(T-\theta)^2 = \int (T(x)-\theta)^2 f_\theta(x)\nu(dx).$$

Now, fix two measures $\mathbb{P}_{\theta_1}$ and $\mathbb{P}_{\theta_2}$, and let $\mathcal{E}$ be a measurable set. Define

$$I_{\mathcal{E}} := I(\theta_1, \theta_2, \mathcal{E}) = \mathbb{E}_{\theta_1}(q^2(X)\mathbb{1}\{X \in \mathcal{E}\}),$$

where $q(x) := f_{\theta_2}(x)/f_{\theta_1}(x)$.

Lemma 26. *Let $\Delta := d(\theta_1, \theta_2)$ and assume that, for certain estimator T, $R(\theta_1, T) \le \varepsilon^2$ and $0 < \varepsilon < \Delta((\mathbb{P}_{\theta_2}(\mathcal{E})/\sqrt{I_{\mathcal{E}}}) \wedge 1)$. Then,*

$$R(\theta_2, T) \ge \Delta^2 \mathbb{P}_{\theta_2}(\mathcal{E})^2 \left(1 - \frac{2\varepsilon\sqrt{I_{\mathcal{E}}}}{\mathbb{P}_{\theta_2}(\mathcal{E})\Delta}\right).$$

Proof. We follow the proof of Theorem 1 in Brown and Low (1996) by considering the same estimator T therein which minimizes $R(\theta_2, T)$ subject to the condition $R(\theta_1, T) \le \varepsilon^2$. Then, with ρ defined therein, we have

$$T(x) = \frac{\rho\Delta q(x)}{1+\rho q(x)}$$

and $R(\theta_1, T) = \varepsilon^2$, so that (see Equation (2.6) in the proof of Theorem 1 in Brown and Low (1996))

$$\varepsilon^2 = \Delta^2 \int \left(\frac{\rho q(x)}{1+\rho q(x)}\right)^2 f_{\theta_1}(x)\nu(dx)$$

under the condition $\varepsilon < \Delta$. Then, by Cauchy-Schwarz,

$$\begin{aligned}
\varepsilon\sqrt{I_{\mathcal{E}}} &= \Delta\left(\int \left(\frac{\rho q(x)}{1+\rho q(x)}\right)^2 f_{\theta_1}(x)\nu(dx)\right)^{1/2} \left(\int q^2(x) f_{\theta_1}(x)\mathbb{1}_{\mathcal{E}}(x)\nu(dx)\right)^{1/2} \\
&\ge \Delta \int \frac{\rho q(x)}{1+\rho q(x)} f_{\theta_2}(x)\mathbb{1}_{\mathcal{E}}(x)\nu(dx).
\end{aligned}$$

Then, under the condition $\varepsilon\sqrt{I_{\mathcal{E}}} \leq \Delta\mathbb{P}_{\theta_2}(\mathcal{E})$, we have

$$\begin{aligned}
\left(\Delta\mathbb{P}_{\theta_2}(\mathcal{E}) - \varepsilon\sqrt{I_{\mathcal{E}}}\right)^2 &\leq \Delta^2\left(\mathbb{P}_{\theta_2}(\mathcal{E}) - \int \frac{\rho q(x)}{1+\rho q(x)} f_{\theta_2}(x)\mathbb{1}_{\mathcal{E}}(x)\nu(dx)\right)^2 \\
&= \Delta^2\left(\int \frac{1}{1+\rho q(x)} f_{\theta_2}(x)\mathbb{1}_{\mathcal{E}}(x)\nu(dx)\right)^2 \\
&\leq \Delta^2\left(\int \frac{1}{1+\rho q(x)} f_{\theta_2}(x)\nu(dx)\right)^2 \\
&\leq \Delta^2 \int \left(\frac{1}{1+\rho q(x)}\right)^2 f_{\theta_2}(x)\nu(dx) \\
&= R_{\theta_2},
\end{aligned}$$

where the last equality is true due to Equation (2.5) in the proof of Theorem 1 in Brown and Low (1996). The statement then follows from $(a-b)^2 \geq a^2(1-2b/a)$ for $a, b > 0$. $\square$

Lemma 27. *Suppose $X_1, \ldots, X_N$ are i.i.d. Poisson variables with mean value n/N, where $N = N_n \geq Cn^{1+\delta}$ for some positive constant δ and absolute constant C. Then, there exists some positive integer $f_{\max}$ that only depends on δ such that*

$$\mathbb{P}\left(\max_{1\leq i\leq N} X_i \leq f_{\max}\right) \to 1$$

as $n \to \infty$.

Proof. Let $\lambda := n/N$. We will show that the above statement holds for $f_{\max} = \lceil (1+\delta)/\delta \rceil$. For each variable X_i and positive integer k, we have

$$\mathbb{P}(X_i \leq k) = \sum_{\ell=0}^{k} \frac{\lambda^\ell}{\ell!} e^{-\lambda} = e^{-\lambda}(1 + \sum_{\ell=1}^{k} (n/N)^\ell/\ell!).$$

Therefore, it holds that

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq i\leq N} X_i \leq k\right) &= \exp(-n)\exp\left(N\log\left(1 + \sum_{\ell=1}^{k}(n/N)^\ell/\ell!\right)\right) \\
&= \exp(-n)\exp\left(N\sum_{m=1}^{\infty}\frac{1}{m}(-1)^{m-1}\sum_{\ell_1,\ldots,\ell_m=1}^{k}\frac{1}{\ell_1!\ldots\ell_m!}(n/N)^{\ell_1+\ldots+\ell_m}\right).
\end{aligned}$$

The exponent in the above display is a polynomial function of (n/N), and clearly the coefficient for $(n/N)^1$ is 1. Next, we will show next that the coefficients corresponding to $(n/N)^\ell$ for $\ell = 2, \dots, k$ are all zero. For simplicity, we will show this for $\ell = k$. By Lemma 28, the coefficient for $(n/N)^k$ is

$$\frac{1}{k!}\sum_{m=1}^{k}(-1)^{m-1}\frac{1}{m}\sum_{\ell=1}^{m}\ell^k(-1)^{m-\ell}\binom{m}{\ell} = \frac{1}{k!}\sum_{\ell=1}^{k}(-1)^{\ell}\ell^{k-1}\sum_{m=\ell}^{k}\binom{m-1}{\ell-1} = \frac{1}{k!}\sum_{\ell=1}^{k}(-1)^{\ell}\ell^{k-1}\binom{m}{\ell} = 0,$$

where the first identity is by direct calculation, the second is the Hockey-Stick identity, and the third is proved in Ruiz (1996).

Next, we consider the coefficient of $(n/N)^p$ for some general $p \geq k+1$, which takes the form

$$\sum_{m=1}^{\infty}\frac{1}{m}(-1)^{m-1}\sum_{\substack{1\leq \ell_1,\dots,\ell_m\leq k\\ \ell_1+\dots+\ell_m=p}}\frac{1}{\ell_1!\dots\ell_m!} = \sum_{m=1}^{p}\frac{1}{m}(-1)^{m-1}\sum_{\substack{1\leq \ell_1,\dots,\ell_m\leq k\\ \ell_1+\dots+\ell_m=p}}\frac{1}{\ell_1!\dots\ell_m!} \leq \sum_{m=1}^{p}\frac{1}{m}\frac{m^p}{p!}$$
$$\lesssim \frac{1}{p}\frac{p^p}{p!} \lesssim e^p p^{-3/2}.$$

Thus, in view of the fact that $N \geq Cn^{1+\delta}$, we have

$$\sum_{p=k+1}^{\infty}(n/N)^p\sum_{m=1}^{\infty}\frac{1}{m}(-1)^{m-1}\sum_{\substack{1\leq \ell_1,\dots,\ell_m\leq k\\ \ell_1+\dots+\ell_m=p}}\frac{1}{\ell_1!\dots\ell_m!} \lesssim \sum_{p=k+1}^{\infty}\left(\frac{en}{N}\right)^p p^{-3/2} \lesssim (en/N)^{k+1}.$$

By definition of $f_{\max}$, we have $(n/N)^{f_{\max}+1} \to 0$ as $n \to \infty$, thus $\mathbb{P}(\max_{1\leq i\leq N} X_i \leq f_{\max}) \to 1$ as $n \to \infty$. $\square$

Lemma 28. *Fix any positive integer k. Then, for any positive integer $1 \leq m \leq k$, the following identity holds:*

$$\sum_{\substack{1\leq \ell_1,\dots,\ell_m\leq k\\ \ell_1+\dots+\ell_m=k}}\frac{1}{\ell_1!\dots\ell_m!} = \frac{1}{k!}\sum_{\ell=1}^{m}(-1)^{m-\ell}\ell^k\binom{m}{\ell}.$$

Proof. We will prove by induction. Denote the LHS by $C(m)$. Suppose the statement holds up to m. Using the identity

$$\sum_{\substack{0\le \ell_1,\ldots,\ell_{m+1}\le k\\ \ell_1+\ldots+\ell_{m+1}=k}} \frac{k!}{\ell_1!\ldots \ell_{m+1}!}(m+1)^{-k} = 1,$$

we obtain the recursive equation

$$C(m+1) + \binom{m+1}{1}C(m) + \ldots + \binom{m+1}{m}C(1) = \frac{(m+1)^k}{k!}.$$

Therefore, plugging in the equation for $C(\ell)$, $\ell = 1, \ldots, m$, we obtain that

$$\begin{aligned} C(m+1) &= \frac{(m+1)^k}{k!} - \sum_{\ell=1}^{m} \frac{\binom{m+1}{\ell}}{k!} \sum_{j=0}^{\ell-1} (\ell-j)^k (-1)^j \binom{\ell}{j} \\ &= \frac{1}{k!}\left[(m+1)^k - \sum_{\ell=1}^{m} \binom{m+1}{\ell} \sum_{j=1}^{\ell} j^k (-1)^{\ell-j} \binom{\ell}{j}\right] \\ &= \frac{1}{k!}\left[(m+1)^k - \sum_{j=1}^{m} j^k \sum_{\ell=j}^{m} (-1)^{\ell-j} \binom{\ell}{j}\binom{m+1}{\ell}\right]. \end{aligned}$$

Thus it suffices to show that

$$\sum_{\ell=j}^{m} (-1)^{\ell} \binom{\ell}{j}\binom{m+1}{\ell} = (-1)^m \binom{m+1}{j}.$$

This is indeed true since the LHS equals

$$\begin{aligned} &\binom{m+1}{j} \sum_{\ell=j}^{m} \binom{m+1-j}{\ell-j} (-1)^{\ell} = \sum_{\ell=0}^{m-j} \binom{m+1}{j}\binom{m-j+1}{\ell}(-1)^{\ell+j} \\ &= \binom{m+1}{j}(-1)^j \left(\sum_{\ell=0}^{m-j+1} \binom{m-j+1}{\ell}(-1)^{\ell} + (-1)^{m-j}\right) = \binom{m+1}{j}(-1)^m, \end{aligned}$$

which completes the proof. $\square$

Appendix B

APPENDIX OF CHAPTER 3

B.1 Proof of Theorem 1

Starting from this section, unless otherwise specified, we will focus on the case $\sigma^2 = 1$; the extension to an arbitrary $\sigma^2 > 0$ is straightforward and hence not recorded here. We will also omit the proof for the $k \log(en/k)$ part of Theorem 1 as it follows essentially from the classical arguments in Donoho and Johnstone (1994); Birgé and Massart (2001) by completely ignoring the regularity constraints. For the rest of the section, we focus on illustrating the form of k_0 in (3.12) from the upper bound perspective and proving the faster $\log\log(16n)$ rate below the transition boundary. Section B.1.1 provides a proof outline with illustrative simple cases discussed at first. Section B.1.2 reduces the proof of Theorem 1 to the bound of complexity width in Proposition 28. The key ingredients to the proof of this proposition will be presented in Sections B.1.3 and B.1.4, followed by the main proof in Section B.1.5.

B.1.1 Proof outline

Piecewise linear case

We first consider the piecewise linear case $d = 1, d_0 \in \{-1, 0\}$, and assume $\theta_0 = 0$ in (3.1) for simplicity of discussion. Here, the transition boundary in (3.12) is $k_0 = 2$ for $d_0 = -1$ and $k_0 = 3$ for $d_0 = 0$, beyond which the $\log\log(16n)$ rate cannot be attained. We focus on the case of $k = 3$ pieces and illustrate the difference between $d_0 = -1$ and $d_0 = 0$. To start, a standard reduction to complexity width in Proposition 27 ahead yields that for some universal constant $C > 0$,

$$\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 - C\|\theta_{\mathrm{oracle}} - \theta_0\|^2 \leq C \cdot \mathbb{E} \sup_{\theta \in \Theta(1,d_0,3): \|\theta\| \leq 1} \big(\varepsilon \cdot \theta\big)^2 \equiv C \cdot \mathbb{E}Z^2,$$

where θ_{oracle} is any oracle in $\Theta(1, d_0, 3)$ such that $\inf_{\theta \in \Theta(1,d_0,3)} \|\theta - \theta_0\|^2$ is achieved, and $\mathbb{E}Z^2$ is termed the 'complexity width' of $\Theta(1, d_0, 3)$. To bound $\mathbb{E}Z^2$, we use the following parametrization for any given $f \in \mathcal{F}_n(1, d_0, 3)$ with knots $0 = n_0/n \leq n_1/n \leq n_2/n \leq n_3/n = 1$: for $i \in \{0, 1, 2\}$,

$$f(x) = a_i + b_i(x - n_i/n), \quad x \in \left(\frac{n_i}{n}, \frac{n_{i+1}}{n}\right]. \tag{B.1}$$

Under the additional continuity constraint when $d_0 = 0$, one has

$$a_1 = a_0 + b_0(n_1 - n_0)/n \quad \text{and} \quad a_2 = a_1 + b_1(n_2 - n_1)/n. \tag{B.2}$$

Under the parametrization (B.1), the supremum within the complexity width can be bounded by

$$Z \leq \sup_{\theta \in \Theta(1,d_0,3): \|\theta\| \leq 1} \sum_{i=0}^{2} \left(|a_i| \left| \sum_{j \in (n_i; n_{i+1}]} \varepsilon_j \right| + \left| \frac{b_i}{n} \right| \left| \sum_{j \in (n_i; n_{i+1}]} (j - n_i)\varepsilon_j \right| \right).$$

The magnitudes of $\{a_i\}$ and $\{b_i\}$ can be drastically different for $d_0 = -1$ and $d_0 = 0$. We illustrate this on the middle piece $(n_1; n_2]$.

- ($d_0 = -1$). The constraint $1 \geq \|\theta\| \geq \|\theta\|_{(n_1;n_2]}$ directly yields the following estimates for a_1 and b_1 with some universal $C > 0$:

$$|a_1| \leq C(n_2 - n_1)^{-1/2} \quad \text{and} \quad |b_1/n| \leq C(n_2 - n_1)^{-3/2}. \tag{B.3}$$

  Such estimates cannot be improved for, e.g., $f(x) = c(L^{-1/2} - nL^{-3/2}(x - 1/2))\mathbf{1}_{(1/2, 1/2+L/n]}(x)$ for small $c > 0$ and $L \geq 2$.

- ($d_0 = 0$). With the additional continuity constraint in (B.2), refined estimates can be obtained:

$$|a_1| \leq C n_2^{-1/2} \quad \text{and} \quad |b_1/n| \leq C(n_2 \wedge (n - n_1))^{-3/2}. \tag{B.4}$$

  These estimates only hold up to $k = 3$ pieces. For $k \geq 4$, the best possible estimates are of type (B.3) by considering, e.g., $f(x) = c\Big(nL^{3/2}\big(x - (1/2 - L/n)\big)\mathbf{1}_{(1/2-L/n,1/2]}(x) - nL^{3/2}\big(x - (1/2 + L/n)\big)\mathbf{1}_{(1/2,L/n+1/2]}(x)\Big)$ for small $c > 0$ and $L \geq 2$.

The crucial difference here is that estimates of type (B.4) enable a law of iterated logarithm (cf. Theorem 3) with $\mathbb{E}Z^2 \lesssim \log\log(16n)$, while those of (B.3) correspond to the maxima of $O(n)$ independent Gaussian random variables with $\mathbb{E}Z^2 \lesssim \log(en)$.

General case

Similar to the linear case discussed above, the key step is to prove

$$\mathbb{E} \sup_{\theta \in \Theta(d,d_0,k_0), \|\theta\| \leq 1} (\varepsilon \cdot \theta)^2 \leq C \log\log(16n), \tag{B.5}$$

and we need to obtain estimates of type (B.4). For simplicity, we consider the smoothest case $d_0 = d - 1$ so that $k_0 = d + 2$.

Fix a degree d, and any $f \in \mathcal{F}_n(d, d-1, d+2)$ along with the corresponding $\theta \in \Theta(d, d-1, d+2)$ of unit norm and knots $0 = n_0 \leq n_1 \leq \ldots \leq n_{d+2} = n$. We use the following parametrization of f:

$$f(x) = \sum_{\ell=1}^{d+1} a_\ell^i \left(x - \frac{n_i}{n}\right)^{\ell-1}, \quad x \in \left(\frac{n_i}{n}, \frac{n_{i+1}}{n}\right], \tag{B.6}$$

and focus on a generic piece $(n_i; n_{i+1}]$ at the sequence level. Here the superscript i represents 'the $(i+1)$-th piece $(n_i; n_{i+1}]$' and the subscript ℓ represents 'the ℓ-th coefficient' in the polynomial. We aim at obtaining the following estimates:

$$1 \geq c \cdot (a_\ell^i)^2((n_{i+1} - n_i)/n)^{2(\ell-1)}(n_{i+1} \wedge (n - n_i)), \quad \ell \in [1; d+1], \tag{B.7}$$

with some $c = c(d)$. Once these estimates are obtained, one can immediately apply Theorem 3 to obtain a $\log\log(16n)$ bound on the complexity width on $(n_i; n_{i+1}]$.

In (B.7), the $(d-1)$-th order differentiability at each inner knot naturally divides the coefficients into two groups, the 'shared coefficients' $\{a_\ell^i\}_{\ell\in[1;d]}$ and the 'nuisance coefficient' a_{d+1}^i. This suggests the following two-step proof strategy:

(i) First, we show that the estimate for the second group, a_{d+1}^i, follows from that of the first group; cf. Lemma 31 ahead.

(ii) Second, we obtain estimates in (B.7) for $\ell \in [1; d]$ with the choice $k_0 = d+2$; cf. Lemma 32 ahead.

In the proof below, we will see clearly why $k_0 = d+2$ is the maximal number of pieces where the estimates in (B.7) are achievable. At a high level, the coefficient estimates $\{a_\cdot^i\}$ on the piece $(n_i; n_{i+1}]$ necessarily depend on coefficient estimates at locations to the both sides of i. The passage of such information, for example from the rightmost knot, is precisely characterized in Lemma 30 ahead through a set of quadratic forms, which are obtained via 'iterative cancellation' to be detailed in Section B.1.3. The transition boundary k_0 is then determined via 'counting of quadratic forms' (cf. (B.15) in the main proof ahead) that mirrors the DOF calculation in (3.15), thereby unifying the heuristics in the upper and lower bounds.

B.1.2 Reduction to complexity width

We first introduce some notation. For any fixed $\theta_0 \in \mathbb{R}^n$, let $\theta_{\mathrm{oracle}} \equiv \theta_{\mathrm{oracle}}(\theta_0) \in \Theta(d, d_0, k)$ be an oracle such that $\inf_{\theta\in\Theta(d,d_0,k)}\|\theta - \theta_0\|$ is achieved, with knots $0 = n_0 \le n_1 \le \ldots \le n_k = n$. For each $\theta \in \mathbb{R}^n$, define $\theta_{[j]}$ as the sub-vector $(\theta_i)_{i\in(n_j;n_{j+1}]}$ and $v_j(\theta) \equiv v_j(\theta; \theta_{\mathrm{oracle}}) \equiv (\theta - \theta_{\mathrm{oracle}})_{[j]}/\|(\theta - \theta_{\mathrm{oracle}})_{[j]}\|$.

The following result is a standard reduction principle for the LSE tailored to the class of splines. Its proof can be found in Appendix B.5.

Proposition 27. *Fix any* $\theta_0 \in \mathbb{R}^n$. *Let* $\widehat{\theta} \equiv \widehat{\theta}(\Theta(d, d_0, k), Y)$ *be the LSE as defined in* (3.5) *under the experiment* (3.1) *with truth* θ_0. *Then, for any* $\delta > 0$, *there exists some* $C = C(\delta) > 0$ *such that*

$$\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 \le (1+\delta)\|\theta_{\text{oracle}} - \theta_0\|^2 + C \cdot \mathbb{E} \sup_{\theta \in \Theta(d, d_0, k)} \sum_{j=0}^{k-1} \big(\varepsilon_{[j]} \cdot v_j(\theta)\big)^2.$$

Now, note that each $v_j(\theta)$ is also a spline with unit norm and the same parameters (d, d_0, k) (rigorously speaking, the two end pieces of $v_j(\theta)$ may have length smaller than $d+1$, but these pieces are negligible since there are at most $2k$ of them and each only contributes a constant (up to d) factor to the complexity width). Therefore, in view of Proposition 27, the $\log\log(16n)$ part of Theorem 1 for $k \le k_0$ is immediately implied by the following result by noticing that $\Theta(d, d_0, k) \subset \Theta(d, d_0, k_0)$ for all $k \le k_0$.

Proposition 28. *There exists some* $C = C(d)$ *such that*

$$\mathbb{E} \sup_{\theta \in \Theta(d, d_0, k_0), \|\theta\| \le 1} (\varepsilon \cdot \theta)^2 \le C \log\log(16n).$$

The following two subsections present the main ingredients to the proof of Proposition 28, whose details will be presented in Section B.1.5.

B.1.3 Groundwork

Fix any $f \in \mathcal{F}_n(d, d_0, k_0)$ with knots $0 = n_0/n \le n_1/n \le \ldots \le n_{k_0}/n = 1$ and recall the parametrization (B.6). Due to the regularity constraints, similar relations as the linear equations of the type (B.2) exist between adjacent knots. We use the notation $\mathrm{Coef}[a_p^i; a_q^{i-1}]$ to denote the coefficient of a_q^{i-1} in the linear equation of a_p^i, i.e., $a_p^i = \sum_q \mathrm{Coef}[a_p^i; a_q^{i-1}] a_q^{i-1}$.

The following lemma makes explicit this dependence. Its proof and proofs for other lemmas in this subsection are contained in Appendix B.5. We write

$$n_{i;j} \equiv (n_i - n_j)/n. \tag{B.8}$$

Lemma 29. *For any $i \in [1; k_0 - 1]$, $p \in [1; d_0 + 1]$, and $q \in [1; d + 1]$,*

$$\mathrm{Coef}[a_p^i; a_q^{i-1}] = \binom{q-1}{p-1} n_{i;i-1}^{q-p} \mathbf{1}_{q \geq p}.$$

The next Lemma 30 provides, as described in the proof outline in Section B.1.1, the exact forms of the quadratic forms obtained by 'iterative cancellation' from right. These quadratic forms lay the foundation for coefficient estimates of type (B.7). For the rest of this section, we reserve the notation s for the number of 'iterative cancellation' performed.

Before stating the general formulation in Lemma 30, we first present the illustrative case of cubic spline ($d = 3, k_0 = 5$) in the sequence space with unit norm. We detail below the starting point ($s = 0$) and the first two steps of cancellation ($s \in \{1, 2\}$). Following the proof outline in Section B.1.1, we separate the quadratic forms that only involve the 'shared coefficients' $\{a_\ell^i\}_{\ell \in [1;3]}$ and those that also involve the 'nuisance coefficient' a_4^i.

- ($s = 0$). The ℓ_2 constraint on $(n_4; n_5]$ for the signal ($\|\theta\|_{(n_4;n_5]} \leq \|\theta\| = 1$) provides control on the following 4 quadratic forms of length 1:

$$1 \geq c \cdot \left[\left\{ (n - n_4)(a_1^4)^2 + \frac{(n-n_4)^3}{n^2}(a_2^4)^2 + \frac{(n-n_4)^5}{n^4}(a_3^4)^2 \right\} + \frac{(n-n_4)^7}{n^6}(a_4^4)^2 \right].$$

- ($s = 1$). For the first cancellation, we have, by Lemma 29,

$$\begin{pmatrix} a_1^4 \\ a_2^4 \\ a_3^4 \end{pmatrix} = \begin{pmatrix} 1 & n_{4;3} & n_{4;3}^2 & n_{4;3}^3 \\ 0 & 1 & 2n_{4;3} & 3n_{4;3}^2 \\ 0 & 0 & 1 & 3n_{4;3} \end{pmatrix} \begin{pmatrix} a_1^3 \\ a_2^3 \\ a_3^3 \\ a_4^3 \end{pmatrix}. \tag{B.9}$$

The identity (B.9) enables us to first find a linear combination of (a_2^4, a_3^4) to cancel a_4^3, and then to find another linear combination of (a_1^4, a_2^4, a_3^4) to cancel both a_3^3 and a_4^3.

These, along with direct expansion of the term $(a_3^4)^2(n-n_4)^5/n^4$ using (B.9), leave us with 3 quadratic forms of length 2:

$$1 \geq c \cdot \left[\left\{ (n-n_4)\big(3a_1^3 + n_{4;3}a_2^3\big)^2 + \frac{(n-n_4)^3}{n^2}\big(a_2^3 + n_{4;3}a_3^3\big)^2 \right\} + \frac{(n-n_4)^5}{n^4}\big(a_3^3 + 3n_{4;3}a_4^3\big)^2 \right].$$

- $(s=2)$. For the second cancellation, we have, by Lemma 29 again,

$$\begin{pmatrix} a_1^3 \\ a_2^3 \\ a_3^3 \end{pmatrix} = \begin{pmatrix} 1 & n_{3;2} & n_{3;2}^2 & n_{3;2}^3 \\ 0 & 1 & 2n_{3;2} & 3n_{3;2}^2 \\ 0 & 0 & 1 & 3n_{3;2} \end{pmatrix} \begin{pmatrix} a_1^2 \\ a_2^2 \\ a_3^2 \\ a_4^2 \end{pmatrix}.$$

Then, finding a linear combination of (a_1^3, a_2^3, a_3^3) to cancel a_4^2 and directly expanding $\big(a_2^3 + n_{4;3}a_3^3\big)^2(n-n_4)^3/n^2$, we obtain 2 quadratic forms of length 3:

$$\begin{aligned} 1 \geq c \cdot \Bigg[&(n-n_4)\Big(3a_1^2 + (2n_{3;2} + n_{4;3})a_2^2 + (n_{3;2}^2 + n_{3;2}n_{4;3})a_3^2\Big)^2 \\ &+ \frac{(n-n_4)^3}{n^2}\Big(a_2^2 + (2n_{3;2} + n_{4;3})a_3^2 + (3n_{3;2}^2 + 3n_{3;2}n_{4;3})a_4^2\Big)^2 \Bigg]. \end{aligned}$$

To state the above cancellation scheme for general d and d_0, some further notation is introduced. Fix d, d_0, and the resulting k_0 as defined in (3.12). Define the sequence $\{\overline{\beta}_j^s\}$, $s \in [0; \lfloor (d_0+1)/(d-d_0) \rfloor]$ recursively as follows. Let $\overline{\beta}_0^s \equiv 1$,

$$\overline{\beta}_j^s \equiv \sum_{\ell=0}^{j} \binom{s(d-d_0)-\ell}{j-\ell} n_{k_0-s;k_0-1-s}^{j-\ell} \overline{\beta}_\ell^{s-1} \tag{B.10}$$

for $j \in [1; s(d-d_0)]$, and $\overline{\beta}_j^s \equiv 0$ for $j > s(d-d_0)$. Further define, for every $i \in [1; (s+1)d_0 - sd + 1]$ and $j \in [0; s(d-d_0)]$,

$$D(i,0) \equiv 1, \quad D(i,j) \equiv \frac{\overline{\odot}(i;j)}{\underline{\odot}(d+1-i;j)} \quad \text{for } j \geq 1.$$

Lastly, let $\overline{\beta}_{i,j}^s \equiv D(i,j)\overline{\beta}_j^s$.

We work under the extra condition that

$$n_{1;0} \wedge n_{k_0;k_0-1} \geq \max\{n_{2;1}, \dots, n_{k_0-1;k_0-2}\}. \tag{B.11}$$

We remark that condition (B.11) is made merely for presentational simplicity; see the comments after Lemma 32 ahead for detailed discussion of this condition.

Lemma 30. *Suppose (B.11) holds. Fix d, d_0, and k_0 as defined in* (3.12)*, and any $\theta \in \Theta(d, d_0, k_0)$ such that $\|\theta\| \leq 1$. Then, there exists some $c = c(d)$ such that, for any $s \in [0; \lfloor (d_0+1)/(d-d_0) \rfloor]$,*

$$1 \geq c\left\{ \sum_{i=1}^{(s+1)d_0-sd+1} + \sum_{i=(s+1)d_0-sd+2}^{sd_0-(s-1)d+1} \right\} \frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}} \left(\sum_{j=0}^{s(d-d_0)} \overline{\beta}_{i,j}^{s} a_{i+j}^{k_0-1-s} \right)^2. \tag{B.12}$$

Remark 21. *Several remarks for the quadratic forms above are in order.*

(i) The quadratic forms in (B.12) *are obtained via iterative cancellation from knot n_{k_0-1}.*

(ii) In a generic $\overline{\beta}_{i,j}^{s}$, the superscript s marks the counts of cancellations already performed, i indicates the i-th quadratic form, and j indicates the coefficient for the j-th component in this quadratic form.

(iii) In (B.12)*, we intentionally separate the indices $i \in [1; (s+1)d_0 - sd + 1]$ and $i \in [(s+1)d_0 - sd + 2; sd_0 - (s-1)d + 1]$ since the first set of quadratic forms only involves the 'shared coefficients' $a_j^{\cdot}$ with $j \in [1; d_0 + 1]$.*

(iv) Every time s grows by 1, the first summand of (B.12) *has $(d - d_0)$ fewer quadratic forms with each one comprising of $(d - d_0)$ more components.*

B.1.4 Key estimates

Recall the coefficient sequence $\{a_\ell^i\}_{i\in[0;k_0-1],\ell\in[1;d+1]}$ defined in (B.6). As described in Section B.1.1, we aim to obtain sharp estimates of type (B.7). For any $a, b \in [1; n]$, define

$$M(a, b) \equiv (a \wedge (n-b))^{1/2}.$$

The first result below reduces the task of obtaining (B.7) for all the coefficients down to estimating only the 'shared coefficients' $\{a_\ell^\cdot\}_{\ell\in[1;d_0+1]}$, from which the estimates for 'nuisance coefficients' $\{a_\ell^\cdot\}_{\ell\in[d_0+2;d+1]}$ can be derived. Its proof can be found in Appendix B.5.

Lemma 31. *Fix any $i \in [1; k_0 - 2]$. Suppose there exists some $c = c(d)$ such that for every $\ell \in [1; d_0 + 1]$, it holds that $1 \geq c(a_\ell^i)^2 n_{i+1,i}^{2(\ell-1)} M^2(n_{i+1}, n_i)$. Then, there exists some $c' = c'(d)$ such that*

$$1 \geq c'(a_\ell^i)^2 n_{i+1,i}^{2(\ell-1)} M^2(n_{i+1}, n_i)$$

for every $\ell \in [d_0 + 2; d + 1]$.

Following the preceding lemma, the next result, which builds on the groundwork derived in Lemma 30, makes use of an inductive argument to derive sharp estimates of the type (B.7) for $\{a_\ell^{i+1}\}_{\ell\in[1;d_0+1]}$ on a fixed target piece $(n_{i+1}; n_{i+2}]$. To make the notation more accessible, we present here the special case $d_0 = d-1$ (so that $k_0 = d+2$) and defer the case of general d_0 to Appendix B.5.5.

Lemma 32. *Suppose $d_0 = d-1$ and (B.11) holds. Fix $i \in [0; d-1]$. For some $c = c(d)$, the following estimates hold for all locations $1 \leq j \leq i+1$:*

$$1 \geq c \max_{1\leq\ell\leq d} \left\{ (a_\ell^j)^2 \cdot n_{i+2;j}^{2\{(d-i)\wedge(\ell-1)\}} \cdot \left(\prod_{k=d-i+2}^{(d-j+2)\wedge\ell} n_{d+3-k;j}^2 \right) \cdot n_{j+1;j}^{2(\ell-(d-j+2))_+} \cdot M^2(n_{j+1}, n_j) \right\}.$$

Here $\prod_{k=k_1}^{k_2} \equiv 1$ for $k_2 < k_1$. In particular, for $j = i+1$:

$$1 \geq c \max_{1\leq\ell\leq d} \left\{ (a_\ell^{i+1})^2 \cdot n_{i+2;i+1}^{2(\ell-1)} \cdot M^2(n_{i+2}, n_{i+1}) \right\}. \tag{B.13}$$

The proof of the above lemma is presented in the next subsection. We emphasize that the condition (B.11) is made only for presentational simplicity, as we explain below. If it does not hold, we can adopt the following partition of the pieces $\{(n_0; n_1], \ldots, (n_{d+1}; n_{d+2}]\}$ via general length constraints. Fix a target piece $(n_{i+1}; n_{i+2}]$ with $i \in [0; d-1]$.

S1. First locate among all pieces the longest one denoted as $(n_{i_1^*}; n_{i_1^*+1}]$ with $i_1^* \in [0; d+1]$. If this is the target piece, then we can directly apply Lemma 37 in Appendix B.7 to this piece to obtain the desired estimates in (B.13).

S2. If not, assume without loss of generality that the target piece is to the left of this longest piece, i.e., $i{+}1 < i_1^*$. Then, we can locate the longest piece among $\{(n_0; n_1], \ldots, (n_{i_1^*-1}; n_{i_1^*}]\}$, which we denote as $(n_{i_2^*}; n_{i_2^*+1}]$ with $i_2^* \in [0; i_1^* - 1]$. If the target piece is among $\{(n_{i_2^*}; n_{i_2^*+1}], \ldots, (n_{i_1^*-1}; n_{i_1^*}]\}$, we can then make the following two modifications of Lemmas 30 and 32: (i) choose location $n_{i_1^*}$ (instead of the current n_{d+1}) as the starting point for the cancellation of the quadratic forms; (ii) choose location $i_2^* + 1$ (instead of the current location 1) as the starting point for the induction in Lemma 32. These two modifications will yield the desired estimates for $\{a_\ell^{i+1}\}$ in (B.13).

S3. If this is not the case, i.e., $(n_{i+1}; n_{i+2}] \in \{(n_1; n_2], \ldots, (n_{i_2^*-1}; n_{i_2^*}]\}$, we can then iterate S2 with i_2^* in place of i_1^*. This partitioning will terminate in a finite number of steps.

Condition (B.11) (with $n_{1;0} \le n_{d+2;d+1}$), along with the current versions of Lemmas 30 and 32, correspond to the above partitioning scheme with an early stop at S2 with $i_1^* = d+1$ and $i_2^* = 0$. On the other hand, condition (B.11) represents the most difficult case in the sense that the maximal gap $i_1^* - i_2^* = d+1$ activates the condition $k \le k_0 = d+2$ as seen in (B.15) in the proof ahead.

B.1.5 Main proof

The main step in the proof of Proposition 28 is the set of coefficient estimates in Lemma 32, with its more general version stated in Appendix B.5.5. We present the proof of this lemma in the special case $d_0 = d - 1$; the proof for the general case is completely analogous.

Proof of Lemma 32. Let

$$Q_j^2(\ell) = n_{i+2;j}^{2\{(d-i)\wedge(\ell-1)\}} \cdot \left(\prod_{k=d-i+2}^{(d-j+2)\wedge\ell} n_{d+3-k;j}^2 \right) \cdot n_{j+1;j}^{2(\ell-(d-j+2))_+}.$$

For the rest of the proof, empty $\prod$ is to be understood as 1 and empty $\bigvee$ is to be understood as 0. We will prove (a slightly stronger version with $M(n_1, n_0)$ instead of $M(n_{j+1}, n_j)$)

$$1 \gtrsim \max_{1\le\ell\le d} \left\{ (a_\ell^j)^2 Q_j^2(\ell) \right\} \cdot M^2(n_1, n_0) \tag{B.14}$$

by induction on $j \in [1; i+1]$. The baseline case $j = 1$ clearly holds by the condition (B.11) and application of Lemma 37 to the piece $(n_0; n_1]$. Now, suppose the induction holds up to some location $j \in [1; i]$, and we will prove the iteration at location $j+1$.

(Part I). We deal with $\{a_\ell^{j+1}\}_{\ell=1}^{d-j+1}$ in this part. For this, we first obtain estimates for a_{d+1}^j and then use triangle inequality. Applying Lemma 30 with $d_0 = d-1$ and $s = d-j$, the j-th term in the first summand therein yields that

$$\begin{aligned} 1 &\gtrsim \frac{(n - n_{d+1})^{2j-1}}{n^{2(j-1)}} \left(\sum_{\ell=0}^{d-j} \overline{\beta}_{j,\ell}^{d-j} a_{j+\ell}^{j+1} \right)^2 \\ &= \frac{(n - n_{d+1})^{2j-1}}{n^{2(j-1)}} \left(\sum_{\ell=0}^{d-j} \overline{\beta}_{j,\ell}^{d-j} \sum_{k=\ell}^{d-j+1} \binom{k+j-1}{\ell+j-1} n_{j+1;j}^{k-\ell} a_{k+j}^j \right)^2 \\ &\equiv \frac{(n - n_{d+1})^{2j-1}}{n^{2(j-1)}} \left(\sum_{k=0}^{d-j+1} \overline{\gamma}_{j,k}^{d-j+1} a_{k+j}^j \right)^2, \end{aligned}$$

where we used Lemma 29 and $\overline{\gamma}_{j,k}^{d-j+1} \equiv \sum_{q=0}^{(d-j)\wedge k} \overline{\beta}_{j,q}^{d-j} \binom{k+j-1}{q+j-1} n_{j+1;j}^{k-q}$. Note that for a generic number of k pieces, when $j = 1$, we need to take $d_0 = d-1$ and $s = (k-1)-(j+1) = k-3$ in Lemma 30, in which case the first summand is non-void if and only if

$$d - s = d - k + 3 \ge 1 \iff k \le d+2. \tag{B.15}$$

This explains the transition boundary $k_0 = d+2$ as in (3.12).

Combining the above estimate with the estimates for $\{a_k^j\}_{k=j}^d$ from the induction assumption, and using Lemma 38 to cancel everything but a_{d+1}^j, we have

$$\begin{aligned}
1 &\gtrsim \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}}\left(\sum_{k=0}^{d-j+1}\overline{\gamma}_{j,k}^{d-j+1}a_{k+j}^j\right)^2+\sum_{k=j}^{d}\left[(a_k^j)^2\cdot Q_j^2(k)\cdot M^2(n_1,n_0)\right]\\
&\gtrsim \left(a_{d+1}^j\right)^2\left\{\frac{(n-n_{d+1})^{2j-1}(\overline{\gamma}_{j,d-j+1}^{d-j+1})^2}{n^{2(j-1)}}\wedge\bigwedge_{k=j}^{d}\left[Q_j^2(k)M^2(n_1,n_0)\frac{(\overline{\gamma}_{j,d-j+1}^{d-j+1})^2}{(\overline{\gamma}_{j,k-j}^{d-j+1})^2}\right]\right\}\\
&\equiv \left(a_{d+1}^j\right)^2\left\{A_j\wedge\bigwedge_{k=j}^{d}B_{j,k}\right\}.
\end{aligned}$$

As $A_j/(\overline{\gamma}_{j,d-j+1}^{d-j+1})^2=n_{d+2;d+1}^{2(j-1)}(n-n_{d+1})\gtrsim B_{j,j}/(\overline{\gamma}_{j,d-j+1}^{d-j+1})^2$ by the assumption that the two end pieces are longer than any middle pieces, we only need to bound from below $\wedge_{k=j}^d B_{j,k}$. By definition of $\overline{\gamma}_{\cdot,\cdot}^{\cdot}$ and non-negativity of $\overline{\beta}_{\cdot,\cdot}^{\cdot}$, for any $j\le k\le d$,

$$\begin{aligned}
\frac{(\overline{\gamma}_{j,d-j+1}^{d-j+1})^2}{(\overline{\gamma}_{j,k-j}^{d-j+1})^2} &\asymp \frac{\left(\sum_{q=0}^{d-j}\overline{\beta}_{j,q}^{d-j}n_{j+1;j}^{(d-j+1)-q}\right)^2}{\left(\sum_{q=0}^{k-j}\overline{\beta}_{j,q}^{d-j}n_{j+1;j}^{k-j-q}\right)^2}\quad\text{(by definition)}\\
&\asymp \bigvee_{p=0}^{d-k}\frac{\bigvee_{q=0}^{k-j}(\overline{\beta}_{j,p+q}^{d-j})^2 n_{j+1;j}^{2\{(d-j+1)-(p+q)\}}}{\bigvee_{q=0}^{k-j}(\overline{\beta}_{j,q}^{d-j})^2 n_{j+1;j}^{2(k-j-q)}}\quad\text{(by rearranging the numerator)}\\
&\ge \bigvee_{p=0}^{d-k}\left\{n_{j+1;j}^{2(d-k+1-p)}\bigwedge_{q=0}^{k-j}\left(\frac{\overline{\beta}_{j,p+q}^{d-j}}{\overline{\beta}_{j,q}^{d-j}}\right)^2\right\}\quad\text{(by Lemma 38)}\\
&\gtrsim \bigvee_{p=0}^{d-k}\left\{n_{j+1;j}^{2(d-k+1-p)}\bigwedge_{q=0}^{k-j}\prod_{r=1+q}^{p+q}n_{d+2-r;j+1}^2\right\}\quad\text{(by Lemma 43)}\\
&= \bigvee_{p=0}^{d-k}\left\{n_{j+1;j}^{2(d-k+1-p)}\prod_{r=1+k-j}^{p+k-j}n_{d+2-r;j+1}^2\right\}\quad\text{(minimum at } q=k-j).
\end{aligned}$$

Hence

$$1\gtrsim (a_{d+1}^j)^2\left[\bigwedge_{k=j}^{d}Q_j^2(k)\bigvee_{p=0}^{d-k}\left\{n_{j+1;j}^{2(d-k+1-p)}\prod_{r=1+k-j}^{p+k-j}n_{d+2-r;j+1}^2\right\}\right]M^2(n_1,n_0).$$

This implies that for $1 \le \ell \le d$, by taking $p = (\ell - k)_+$ above and Lemma 29,

$$\begin{aligned}
&M(n_1, n_0)|a_\ell^{j+1}| \lesssim M(n_1, n_0)\Bigg[\sum_{k=\ell}^{d} n_{j+1;j}^{k-\ell}|a_k^j| + n_{j+1;j}^{d+1-\ell}|a_{d+1}^j|\Bigg] \\
&\lesssim \sum_{k=\ell}^{d} n_{j+1;j}^{k-\ell} Q_j^{-1}(k) + \bigvee_{k=j}^{d}\Bigg\{Q_j^{-1}(k) n_{j+1;j}^{k-\ell+(\ell-k)_+}\prod_{r=1}^{(\ell-k)_+} n_{d+2+j-k-r;j+1}^{-1}\Bigg\} \\
&= \sum_{k=\ell}^{d} n_{j+1;j}^{k-\ell} Q_j^{-1}(k) + \bigvee_{j \le k < \ell \vee j}\Bigg\{Q_j^{-1}(k)\prod_{r=1}^{\ell\vee j-k} n_{d+2+j-k-r;j+1}^{-1}\Bigg\} + \bigvee_{\ell\vee j \le k \le d}\Bigg\{Q_j^{-1}(k) n_{j+1;j}^{k-\ell}\Bigg\}.
\end{aligned}$$

Using that $k \mapsto n_{j+1;j}^{k-\ell} Q_j^{-1}(k)$ is non-increasing, the first and third terms in the above display are on the same order as $Q_j^{-1}(\ell) + Q_j^{-1}(\ell \vee j) n_{j+1;j}^{\ell\vee j-\ell} \asymp Q_j^{-1}(\ell)$. Hence we only need to verify for all $1 \le \ell \le d - j + 1$, $1 \le j \le i$,

$$\mathfrak{Q}_{j,1}(\ell) + \mathfrak{Q}_{j,2}(\ell) \equiv Q_j^{-1}(\ell) + \bigvee_{j \le k < \ell \vee j}\Bigg\{Q_j^{-1}(k)\prod_{r=1}^{\ell\vee j-k} n_{d+2+j-k-r;j+1}^{-1}\Bigg\} \lesssim Q_{j+1}^{-1}(\ell). \tag{B.16}$$

(Case 1). If $1 \le \ell \le d - i + 1$, $Q_j^{-1}(\ell) = n_{i+2;j}^{-(\ell-1)}$ and $Q_j^{-1}(\ell) = n_{i+2;j+1}^{-(\ell-1)}$, so:

- (first term) $\mathfrak{Q}_{j,1}(\ell) = n_{i+2;j}^{-(\ell-1)} \le n_{i+2;j+1}^{-(\ell-1)} = Q_{j+1}^{-1}(\ell)$.

- (second term) without loss of generality we assume $\ell > j$ (otherwise this term does not exist):

$$\begin{aligned}
\mathfrak{Q}_{j,2}(\ell) &= \bigvee_{j \le k < \ell}\Bigg\{n_{i+2;j}^{-(k-1)}\prod_{r=1}^{\ell-k} n_{d+2+j-k-r;j+1}^{-1}\Bigg\} \\
&\le \bigvee_{j \le k < \ell}\Bigg\{n_{i+2;j}^{-(k-1)} n_{d+2+j-\ell;j+1}^{-(\ell-k)}\Bigg\} \\
&\le \bigvee_{j \le k < \ell}\Bigg\{n_{i+2;j+1}^{-(k-1)} n_{i+2;j+1}^{-(\ell-k)}\Bigg\} = n_{i+2;j+1}^{-(\ell-1)} = Q_{j+1}^{-1}(\ell),
\end{aligned}$$

  where the first equality follows since $k < \ell \le d - i + 1$ so that $Q_j^{-1}(k) = n_{i+2;j}^{-(k-1)}$, and the second inequality follows by noting that $\ell \le d - i + 1$ implies $d + 2 + j - \ell \ge i + 2$.

(Case 2). If $d-i+2 \le \ell \le d-j+1$, $Q_j^{-1}(\ell) = n_{i+2;j}^{-(d-i)} \prod_{s=d-i+2}^{\ell} n_{d+3-s;j}^{-1}$ and $Q_{j+1}^{-1}(\ell) = n_{i+2;j+1}^{-(d-i)} \prod_{s=d-i+2}^{\ell} n_{d+3-s;j+1}^{-1}$, so:

- (first term) similarly as above,

$$\mathfrak{Q}_{j,1}(\ell) = n_{i+2;j}^{-(d-i)} \prod_{s=d-i+2}^{\ell} n_{d+3-s;j}^{-1} \le n_{i+2;j+1}^{-(d-i)} \prod_{s=d-i+2}^{\ell} n_{d+3-s;j+1}^{-1} = Q_{j+1}^{-1}(\ell).$$

- (second term) similarly as above we assume $\ell > j$, then

$$\begin{aligned} \mathfrak{Q}_{j,2}(\ell) &= \bigvee_{j \le k < \ell} \left\{ n_{i+2;j}^{-(d-i)} \prod_{s=d-i+2}^{k} n_{d+3-s;j}^{-1} \prod_{r=1}^{\ell-k} n_{d+2+j-k-r;j+1}^{-1} \right\} \\ &\le n_{i+2;j+1}^{-(d-i)} \bigvee_{j \le k < \ell} \left\{ \prod_{u=d+3-k}^{i+1} n_{u;j+1}^{-1} \prod_{u=d+2+j-\ell}^{d+1+j-k} n_{u;j+1}^{-1} \right\}. \end{aligned}$$

  Note that $\prod_{u=d+2+j-\ell}^{d+1+j-k} n_{u;j+1}^{-1} \le \prod_{u=d+2+j-\ell-(j-1)}^{d+1+j-k-(j-1)} n_{u;j+1}^{-1} = \prod_{u=d+3-\ell}^{d+2-k} n_{u;j+1}^{-1}$, where the inequality follows by $j \ge 1$ and $\ell \le d-j+1$, so the above display can be further bounded by

$$\mathfrak{Q}_{j,2}(\ell) \le n_{i+2;j+1}^{-(d-i)} \prod_{u=d+3-\ell}^{i+1} n_{u;j+1}^{-1} = Q_{j+1}^{-1}(\ell).$$

Hence (B.16) is verified and we have finished the proof for Part I.

(Part II). We deal with $\{a_\ell^{j+1}\}_{\ell=d-j+2}^{d}$ in this step. Applying Lemma 30 with $d_0 = d-1$ and $s = d-j$, the last $(j-1)$ terms in the first summand therein take the form

$$1 \gtrsim \frac{(n-n_{d+1})^3}{n^2} \left(\overline{\beta}_{2,0}^{d-j} a_2^{j+1} + \ldots + \overline{\beta}_{2,d-j}^{d-j} a_{d-j+2}^{j+1} \right)^2 \tag{R.2}$$

$$+ \frac{(n-n_{d+1})^5}{n^4} \left(\overline{\beta}_{3,0}^{d-j} a_3^{j+1} + \ldots + \overline{\beta}_{3,d-j}^{d-j} a_{d-j+3}^{j+1} \right)^2 \tag{R.3}$$

$$\ldots$$

$$+ \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}} \left(\overline{\beta}_{j,0}^{d-j} a_j^{j+1} + \ldots + \overline{\beta}_{j,d-j}^{d-j} a_d^{j+1} \right)^2. \tag{R.j}$$

Combining (R.2) with the estimates for $\{a_\ell^{j+1}\}_{\ell=2}^{d-j+1}$ obtained in Part I, and using Lemma 39 iteratively to cancel everything but a_{d-j+2}^{j+1}, we obtain

$$\begin{aligned}
1 &\gtrsim \frac{(n-n_{d+1})^3}{n^2}\bigg(\sum_{k=0}^{d-j}\overline{\beta}_{2,k}^{d-j}a_{k+2}^{j+1}\bigg)^2+\sum_{k=2}^{d-j+1}\Big[(a_k^{j+1})^2\cdot Q_{j+1}^2(k)\cdot M^2(n_1,n_0)\Big]\\
&\gtrsim \big(a_{d-j+2}^{j+1}\big)^2\bigg\{\frac{(n-n_{d+1})^3(\overline{\beta}_{2,d-j}^{d-j})^2}{n^2}\wedge\bigwedge_{k=2}^{d-j+1}\bigg[Q_{j+1}^2(k)M^2(n_1,n_0)\frac{(\overline{\beta}_{2,d-j}^{d-j})^2}{(\overline{\beta}_{2,k-2}^{d-j})^2}\bigg]\bigg\}\\
&\equiv \big(a_{d-j+2}^{j+1}\big)^2\bigg\{A_j^{(2)}\wedge\bigwedge_{k=2}^{d-j+1}B_{j,k}^{(2)}\bigg\}.
\end{aligned}$$

Similar to Part I, we only need to get a lower bound for $\bigwedge_{k=2}^{d-j+1}B_{j,k}^{(2)}$. As $(\overline{\beta}_{2,d-j}^{d-j})^2/(\overline{\beta}_{2,k-2}^{d-j})^2\gtrsim\prod_{r=k-1}^{d-j}n_{d+2-r;j+1}^2$ by Lemma 43, it follows that

$$1\gtrsim\big(a_{d-j+2}^{j+1}\big)^2\bigwedge_{k=2}^{d-j+1}\bigg[Q_{j+1}^2(k)\prod_{r=k-1}^{d-j}n_{d+2-r;j+1}^2\bigg]M^2(n_1,n_0).$$

As $k\mapsto Q_{j+1}^2(k)\prod_{r=k-1}^{d-j}n_{d+2-r;j+1}^2=Q_{j+1}^2(d-j+1)n_{d+3-k;j+1}^2$ is non-increasing on $k\in[2;d-j+1]$, the minimum is taken at $k=d-j+1$ in the above display. Since $Q_{j+1}^2(d-j+1)n_{j+2;j+1}^2=Q_{j+1}^2(d-j+2)$, we arrive at

$$1\gtrsim\big(a_{d-j+2}^{j+1}\big)^2Q_{j+1}^2(d-j+2)M^2(n_1,n_0),$$

which is the desired estimate for a_{d-j+2}^{j+1}. Now iterate along (R.3)-(R.j) to complete the proof for Part II. This completes the proof. □

Proof of Proposition 28. We shorthand $\Theta(d,d_0,k_0)$ as Θ, and the sample points will be indexed using ι. For any $\theta\in\Theta$, let $\{n_j\}_{j=0}^{k_0}$ be its knots: $0=n_0\le n_1\le\ldots\le n_{k_0}=n$. The overall complexity width can then be bounded piece by piece:

$$\mathbb{E}\sup_{\theta\in\Theta}(\varepsilon\cdot\theta)^2=\mathbb{E}\sup_{\theta\in\Theta}\bigg(\sum_{i=1}^{k_0}(\varepsilon\cdot\theta)_{(n_{i-1};n_i]}\bigg)^2\le C\sum_{i=1}^{k_0}\mathbb{E}\sup_{\theta\in\Theta}(\varepsilon\cdot\theta)_{(n_{i-1};n_i]}^2.$$

We will prove that each summand in the above display can be bounded by a constant multiple of $\log\log(16n)$.

We start with the first piece $(n_0; n_1]$. Let $f \in \mathcal{F}_n(d, d_0, k)$ be a generating spline of θ, i.e., $\theta_\iota = f(\iota/n)$ for $\iota \in [1; n]$. For this piece, we use the following parametrization of $f(\cdot)$ slightly different from (B.6): for any $x \in (0, n_1/n]$,

$$f(x) = \sum_{\ell=1}^{d_0+1} \widetilde{a}_\ell^1 \left(x - \frac{n_1}{n}\right)^{\ell-1} + \sum_{\ell=d_0+2}^{d+1} a_\ell^0 \left(x - \frac{n_1}{n}\right)^{\ell-1}. \tag{B.17}$$

Then, the complexity width in question can be written as

$$(\varepsilon \cdot \theta)_{(n_0;n_1]} = \sum_{\ell=1}^{d_0+1} \sum_{\iota \in (n_0;n_1]} \widetilde{a}_\ell^1 \left(\frac{\iota - n_1}{n}\right)^{\ell-1} \varepsilon_\iota + \sum_{\ell=d_0+2}^{d+1} \sum_{\iota \in (n_0;n_1]} a_\ell^0 \left(\frac{\iota - n_1}{n}\right)^{\ell-1} \varepsilon_\iota.$$

Applying Lemma 37 to the piece $(n_0; n_1]$, we have

$$\sum_{\ell=1}^{d_0+1} (\widetilde{a}_\ell^1)^2 \frac{n_1^{2\ell-1}}{n^{2(\ell-1)}} + \sum_{\ell=d_0+2}^{d+1} (a_\ell^0)^2 \frac{n_1^{2\ell-1}}{n^{2(\ell-1)}} \lesssim 1. \tag{B.18}$$

Thus the complexity width over the first piece $(n_0; n_1]$ can be bounded by

$$\begin{aligned}
&\mathbb{E} \sup_{\theta \in \Theta} (\varepsilon \cdot \theta)^2_{(n_0;n_1]} \\
&\lesssim \sum_{\ell=1}^{d_0+1} \mathbb{E} \sup_{1 \le n_1 \le n} \sup_{(\widetilde{a}_\ell^1)^2 \frac{n_1^{2\ell-1}}{n^{2(\ell-1)}} \le 1} \frac{(\widetilde{a}_\ell^1)^2}{n^{2(\ell-1)}} \left(\sum_{\iota \in (n_0;n_1]} (\iota - n_1)^{\ell-1} \varepsilon_\iota \right)^2 \\
&+ \sum_{\ell=d_0+2}^{d+1} \mathbb{E} \sup_{1 \le n_1 \le n} \sup_{(a_\ell^0)^2 \frac{n_1^{2\ell-1}}{n^{2(\ell-1)}} \le 1} \frac{(a_\ell^0)^2}{n^{2(\ell-1)}} \left(\sum_{\iota \in (n_0;n_1]} (\iota - n_1)^{\ell-1} \varepsilon_\iota \right)^2 \\
&\le C \log\log(16n),
\end{aligned}$$

where the second inequality is due to Theorem 3 with $\psi(x) = x^2$ therein. The complexity width over the last piece $(n_{k_0-1}; n_{k_0}]$ can be handled similarly.

Starting from the second until the second last piece, we use the parametrization (B.6) on the piece $(n_{i+1}; n_{i+2}]$, yielding

$$(\varepsilon \cdot \theta)_{(n_{i+1};n_{i+2}]} = \sum_{\ell=1}^{d+1} \sum_{\iota \in (n_{i+1};n_{i+2}]} a_\ell^{i+1} \left(\frac{\iota - n_{i+1}}{n} \right)^{\ell-1} \varepsilon_\iota.$$

Thus the complexity width in question can be bounded by

$$\begin{aligned}
&\mathbb{E} \sup_{\theta \in \Theta} (\varepsilon \cdot \theta)^2_{(n_{i+1};n_{i+2}]} \lesssim \sum_{\ell=1}^{d+1} \mathbb{E} \sup_{\theta \in \Theta} \frac{(a_\ell^{i+1})^2}{n^{2(\ell-1)}} \Bigg(\sum_{\iota \in (n_{i+1};n_{i+2}]} (\iota - n_{i+1})^{\ell-1} \varepsilon_\iota \Bigg)^2 \\
&\lesssim \sum_{\ell=1}^{d+1} \mathbb{E} \sup_{\substack{n_{i+1} < n_{i+2}, \\ (a_\ell^{i+1})^2 n_{i+2;i+1}^{2(\ell-1)} M^2(n_{i+2}, n_{i+1}) \le 1}} \frac{\left(a_\ell^{i+1}\right)^2}{n^{2(\ell-1)}} \Bigg(\sum_{\iota \in (n_{i+1};n_{i+2}]} (\iota - n_{i+1})^{\ell-1} \varepsilon_\iota \Bigg)^2 \\
&\le C \log\log(16n),
\end{aligned}$$

where the second inequality is by plugging in the estimates a_ℓ^{i+1}, $\ell \in [1; d+1]$ from Lemma 35 (the general version of Lemma 32 with general $d_0 \in [-1; d-1]$), and the third inequality is by applying Theorem 3 with $\psi(x) = x^2$ therein. The proof is thus complete. □

B.2 Proof of Theorem 2, upper bound

B.2.1 Proof outline

For expository purpose, we focus on the convex linear case $\Theta^*(1, k)$ with truth $\theta_0 = 0$ in (3.1). Using the reduction Proposition 27, the key ingredient is to show

$$\mathbb{E} \sup_{\theta^* \in \Theta^*(1,k): \|\theta^*\| \le 1} \left(\varepsilon \cdot \theta^*\right)^2 \le C \log\log(16n). \tag{B.19}$$

To control the complexity width, we may parametrize any $\theta^* \in \Theta^*(1, k)$ by

$$\theta_i^* = c_0 + \sum_{j=1}^{j^*} a_j \left(\frac{n_j - i}{n} \right)_+ + \sum_{j=j^*}^{k-1} b_j \left(\frac{i - n_j}{n} \right)_+, \tag{B.20}$$

where

- j^* is the index of the knot where the slope of the underlying convex function f^* crosses zero if it does, and is otherwise set to be k;

- $\{a_j\}$ and $\{b_j\}$ are two *non-negative* real sequences parametrizing the *change of slope*, in the two regions where f^* has negative and positive slopes, respectively.

With the parametrization (B.20), proving (B.19) then reduces to obtaining sharp estimates for $\{a_j\}, \{b_j\}$, and c_0. These estimates are obtained in rather different ways:

- For the coefficients $\{a_j\}, \{b_j\}$, the non-negativity property turns out to be the key in obtaining sharp estimates for their magnitudes. Combined with the LIL (cf. Theorem 3), these coefficients contribute the desired $\log\log(16n)$ factor to the complexity width (B.19).

- For the coefficient c_0, an *a priori* estimate $|c_0| \le C/\sqrt{n}$ is obtained (cf. Lemma 34) under the assumed (convexity) shape constraint and the ℓ_2 constraint on the signal. This means that the coefficient c_0 only contributes a constant factor to the complexity width (B.19).

It should be noted that for the larger class $\Theta(1, 0, k)$ without the convexity shape constraint, a parametrization in the form of (B.20) still holds but *without the non-negativity constraint on* $\{a_j\}, \{b_j\}$. The lack of such sign constraints unfortunately makes this representation not quite useful in obtaining LIL for $\Theta(1, 0, 3)$, so a different representation (cf. (B.1)) and a different proof strategy (cf. Section B.1.1) are adopted for $\Theta(1, 0, 3)$.

B.2.2 Groundwork

The first result establishes a canonical parametrization for general-order d-monotone splines. By definition, the polynomial coefficient of the highest order for a d-monotone spline is increasing and thus crosses zero at most once. In the following parametrization, we choose this cross point as the pivot.

Lemma 33. *For any $f^* \in \mathcal{F}_n^*(d,k)$, there exists some integer $j^* \in [0;k]$ and real sequences $\{a_j\}_{j=1}^{j^*}$, $\{b_j\}_{j=j^*}^{k-1}$, and $\{c_\ell\}_{\ell=0}^{d-1}$ such that $a_j(-1)^{d+1} \geq 0$, $b_j \geq 0$, and*

$$f^*(x) = \sum_{j=1}^{j^*} a_j \Big(\frac{n_j}{n} - x\Big)_+^d + \sum_{j=j^*}^{k-1} b_j \Big(x - \frac{n_j}{n}\Big)_+^d + \sum_{\ell=0}^{d-1} \frac{c_\ell}{\ell!} x^\ell \tag{B.21}$$

for $x \in (0,1]$, where $\{n_j/n\}_{j=0}^{k}$ are the knots of f^*. On the sequence level, we have for every $\theta^* \in \Theta^*(d,k)$:

$$\theta_i^* = \sum_{j=1}^{j^*} a_j \left(\frac{n_j - i}{n}\right)_+^d + \sum_{j=j^*}^{k-1} b_j \left(\frac{i - n_j}{n}\right)_+^d + \sum_{\ell=0}^{d-1} \frac{c_\ell}{\ell!} (i/n)^\ell. \tag{B.22}$$

The next result generalizes the bound $|c_0| \leq C/\sqrt{n}$ in the previous proof outline, indicating that all lower-order polynomial coefficients of a d-monotone spline can be well-controlled.

Lemma 34. *For any $\theta^* \in \Theta^*(d,k)$ with $\|\theta^*\|^2 \leq 1$, there exists some $C = C(d)$ such that, in its canonical form* (B.22)*, $|c_\ell| \leq C/\sqrt{n}$ for every $\ell \in [0; d-1]$.*

The proof of the above lemmas can be found in Appendix B.6.

B.2.3 Main proof

Proof of Theorem 2 (upper bound). Throughout the proof, we will shorthand $\Theta^*(d,k)$ as Θ^*. We start with a slight modification of the reduction principle in Proposition 27.

Let $L_0 \equiv n/k$ be an integer without loss of generality. Let θ^*_{oracle} be an oracle in Θ^* that achieves the infimum. Let $n_j \equiv n_j(\theta^*_{\text{oracle}})$, $0 \leq j \leq k$ be the knots of θ^*_{oracle}: $0 = n_0 \leq n_1 \leq \ldots \leq n_k = n$. For each $j \in [0; k-1]$, let $m_j \equiv m_j(\theta^*_{\text{oracle}}) \equiv \lceil (n_{j+1} - n_j)/L_0 \rceil$, $n_{j,p} \equiv n_{j,p}(\theta^*_{\text{oracle}}) \equiv n_j + p \cdot L_0$ for $p \in [0; m_j - 1]$ so that $n_{j,0} = n_j$ and $n_{j,m_j} \equiv n_{j,m_j}(\theta^*_{\text{oracle}}) \equiv n_{j+1}$. Lastly, for any $\theta^* \in \Theta^*$, let $s_{j,p} \equiv s_{j,p}(\theta^*, \theta^*_{\text{oracle}})$ be the number of knots of $\theta^* - \theta^*_{\text{oracle}}$ on the segment $(n_{j,p}, n_{j,p+1}]$, so that $\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1} s_{j,p} \leq k$. Under the above notation, define, for each $\theta \in \mathbb{R}^n$, $(\theta)_{[j,p]}$ as the sub-vector $(\theta_i)_{i\in(n_{j,p},n_{j,p+1}]}$.

Following the same line of proof as Proposition 27 on this finer resolution $\{n_{j,p}\}$, we have, for any $\delta > 0$ and then some $C = C(\delta)$,

$$\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 \leq (1+\delta)\|\theta^*_{\text{oracle}} - \theta_0\|^2 + C \cdot \mathbb{E} \sup_{\theta^*\in\Theta^*} \sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1} \big(\varepsilon_{[j,p]} \cdot v_{j,p}(\theta^*)\big)^2,$$

where $v_{j,p}(\theta^*) \equiv v_{j,p}(\theta^*; \theta^*_{\text{oracle}}) \equiv (\theta^* - \theta^*_{\text{oracle}})_{[j,p]}/\|(\theta^* - \theta^*_{\text{oracle}})_{[j,p]}\|$.

We now prove that the second term on the right side can be bounded by a constant multiple of $k \log\log(16n/k)$. Some extra notation is hence needed. For any $\theta^* \in \Theta^*$, denote the set of $s_{j,p}$ knots of $v_{j,p}(\theta^*)$ as $n_{j,p,1}, \ldots, n_{j,p,s_{j,p}}$. Also define $n_{j,p,0} \equiv n_{j,p,0}(\theta^*_{\text{oracle}}) \equiv n_{j,p}$ and $n_{j,p,s_{j,p}+1} \equiv n_{j,p,s_{j,p}+1}(\theta^*_{\text{oracle}}) \equiv n_{j,p+1}$. Moreover, in view of the canonical parametrization of shape-constrained splines in Lemma 33, let for each fixed $j \in [0; k-1]$ and $p \in [0; m_j]$ the index $q \equiv q(\theta^*, \theta^*_{\text{oracle}}) \in [0; s_{j,p}]$ be such that, on $(n_{j,p}, n_{j,p+1}]$, $(n_{j,p,q^*-1}, n_{j,p,q^*}]$ is the last piece on which the sign of the highest order polynomial component of $\theta^* - \theta^*_{\text{oracle}}$ is negative.

Under the above notation, we have $v_{j,p}(\theta^*) \in \Theta^*_{n_{j,p+1}-n_{j,p}}(d, s_{j,p}+1)$ (here we assume without loss of generality that the two end pieces of $\theta^* - \theta^*_{\text{oracle}}$ adjacent to $n_{j,p}$ and $n_{j,p+1}$ also have length at least $d+1$ since there are at most $2k$ such pieces and each only contributes a constant factor to the complexity width). Thus Lemma 33 entails that there exist real sequences $\{c_{j,p,\ell}\} \equiv \{c_{j,p,\ell}(\theta^*, \theta^*_{\text{oracle}})\}$, and some $q^* \in [1; s_{j,p}]$ along with sequences of equal sign $\{a_{j,p,q}\}_{q=1}^{q^*} \equiv \{a_{j,p,q}(\theta^*, \theta^*_{\text{oracle}})\}_{q=1}^{q^*}$, $\{b_{j,p,q}\}_{q=q^*}^{s_{j,p}} \equiv \{b_{j,p,q}(\theta^*, \theta^*_{\text{oracle}})\}_{q=q^*}^{s_{j,p}}$ such that

$$\begin{aligned}\big(v_{j,p}(\theta^*)\big)_i &= \sum_{q=1}^{q^*} a_{j,p,q}\left(\frac{n_{j,p,q} - (i + n_{j,p})}{n}\right)_+^d + \sum_{q=q^*}^{s_{j,p}} b_{j,p,q}\left(\frac{(i+n_{j,p}) - n_{j,p,q}}{n}\right)_+^d \\ &\quad + \sum_{\ell=0}^{d-1} \frac{c_{j,p,\ell}}{\ell!}\left(\frac{i - n_{j,p}}{n}\right)^\ell \equiv \big(v^1_{j,p}(\theta^*)\big)_i + \big(v^2_{j,p}(\theta^*)\big)_i, \end{aligned} \tag{B.23}$$

where $(v^2_{j,p}(\theta^*))_i \equiv \sum_{\ell=0}^{d-1} c_{j,p,\ell}\big((i-n_{j,p})/n\big)^\ell/\ell!$. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v_{j,p}(\theta^*)\big)^2\\
&\le 2\bigg(\mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v^1_{j,p}(\theta^*)\big)^2+\mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v^2_{j,p}(\theta^*)\big)^2\bigg)\\
&\equiv 2\Big((I)+(II)\Big).
\end{aligned}
$$

We first upper bound (II). Since for each j,p and $\theta^*\in\Theta^*$, $v_{j,p}(\theta^*)\in\Theta^*_{n_{j,p+1}-n_{j,p}}(d,s_{j,p}+1)$ and has unit norm, Lemma 34 entails that there exists some $C=C(d)$ such that $|c_{j,p,\ell}|\le C/\sqrt{n_{j,p+1}-n_{j,p}}$ for $j\in[0;k-1]$, $p\in[0;m_j]$, and $\ell\in[0;d-1]$. Let $\Delta n_{j,p}\equiv n_{j,p+1}-n_{j,p}$. Then, we have

$$
\begin{aligned}
(II)&\le C\cdot\mathbb{E}\sup_{|c_{j,p,\ell}|\le C/\sqrt{\Delta n_{j,p}}}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=0}^{d-1}\frac{c^2_{j,p,\ell}}{n^{2\ell}(\ell!)^2}\bigg(\sum_{i\in(n_{j,p};n_{j,p+1}]}(i-n_{j,p})^\ell\varepsilon_i\bigg)^2\\
&\le C\cdot\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=0}^{d-1}(\Delta n_{j,p})^{-1}\frac{\mathbb{E}\big[\sum_{i\in(n_{j,p};n_{j,p+1}]}(i-n_{j,p})^\ell\varepsilon_i\big]^2}{n^{2\ell}}\\
&\le C\cdot\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}1=C\cdot\sum_{j=0}^{k-1}m_j\le Ck.
\end{aligned}
$$

Next, we bound (I). Some extra notation is needed. Define the following partition of $(n_{j,p};n_{j,p+1}]$ with intervals

$$
I^B_{j,p,\ell}\equiv\Big(n_{j,p}+\big\lceil(1-2^{-(\ell-1)})\Delta n_{j,p}\big\rceil\,;n_{j,p}+\big\lceil(1-2^{-\ell})\Delta n_{j,p}\big\rceil\Big]
$$

for $\ell\in[1;t_{j,p}]$ and $t_{j,p}\equiv\lceil\log_2\Delta n_{j,p}\rceil$, and similarly,

$$
I^A_{j,p,\ell}\equiv\Big(n_{j,p+1}-\big\lceil(1-2^{-\ell})\Delta n_{j,p}\big\rceil\,;n_{j,p+1}-\big\lceil(1-2^{-(\ell-1)})\Delta n_{j,p}\big\rceil\Big].
$$

From this definition, we immediately have (with analogous conclusions for $I^A_{j,p,\ell}$): (i) $|I^B_{j,p,\ell}|\le$

$\lceil 2^{-\ell}\Delta n_{j,p}\rceil$; (ii) $2(\sum_{\ell>\ell_0}|I^B_{j,p,\ell}|+1)\geq\sum_{\ell\geq\ell_0}|I^B_{j,p,\ell}|$ for any $\ell_0\in[1;t_{j,p}]$. Then, let

$$\begin{aligned}
B_{j,p,\ell} &\equiv B_{j,p,\ell}(\theta^*,\theta^*_{\text{oracle}}) \equiv \sum_{q=q^*}^{s_{j,p}} b_{j,p,\ell}\mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}},\\
\delta^B_{j,p,\ell} &\equiv \delta^B_{j,p,\ell}(\theta^*,\theta^*_{\text{oracle}}) \equiv \max\Big\{q^*\leq q\leq s_{j,p}:\mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}}\Big\}.
\end{aligned}$$

In words, $\delta^B_{j,p,\ell}$ equals to 1 if and only if among the knots $\{n_{j,p,q}\}_{q=q^*}^{s_{j,p}}$, there is at least one that lies in the interval $I^B_{j,p,\ell}$, and if such is the case, $B_{j,p,\ell}$ returns the block sum. We omit the similar definitions for $A_{j,p,\ell}$ and $\delta^A_{j,p,\ell}$. By definition, we immediately have $\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\leq s_{j,p}$.

In the parametrization (B.23), using the constraint $\|v_{j,p}(\theta^*)\|\leq 1$ and the bounds $|c_{j,p,\ell}|\leq C/\sqrt{n_{j,p+1}-n_{j,p}}$ for $\ell\in[0;d-1]$, we have $\|v^1_{j,p}(\theta^*)\|\leq C$ (recall the definition of $v^1_{j,p}$ in (B.23)) for some $C=C(d)$. Hence for some sufficiently small $c=c(d)$,

$$\begin{aligned}
1 &\geq c\cdot\sum_{i\in(n_{j,p};n_{j,p+1}]}\Bigg[\sum_{q=1}^{q^*}a_{j,p,q}\bigg(\frac{n_{j,p,q}-i}{n}\bigg)_+^d+\sum_{q=q^*}^{s_{j,p}}b_{j,p,q}\bigg(\frac{i-n_{j,p,q}}{n}\bigg)_+^d\Bigg]^2\\
&\geq c\cdot\sum_{i\in(n_{j,p};n_{j,p+1}]}\Bigg[\sum_{q=1}^{q^*}a_{j,p,q}\bigg(\frac{n_{j,p,q}-i}{n}\bigg)_+^d\Bigg]^2\vee\Bigg[\sum_{q=q^*}^{s_{j,p}}b_{j,p,q}\bigg(\frac{i-n_{j,p,q}}{n}\bigg)_+^d\Bigg]^2,
\end{aligned}$$

where the second inequality follows from the fact that the interaction term between the two summands in the first inequality is 0 for each i.

Now, starting from the constraint $1\geq c\cdot\sum_{i\in(n_{j,p};n_{j,p+1}]}\big[\sum_{q=q^*}^{s_{j,p}}b_{j,p,q}\big(\frac{i-n_{j,p,q}}{n}\big)_+^d\big]^2$, we will obtain estimates for $B_{j,p,\ell}$. Fix j,p. By the disjointness of $I^B_{j,p,\ell}$ and the non-negativeness of

$\{b_{j,p,q}\}$, we have

$$
\begin{aligned}
1 &\geq c \cdot \sum_{i\in(n_{j,p};n_{j,p+1}]} \left[\sum_{\ell=1}^{t_{j,p}} \sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}} \left(\frac{i-n_{j,p,q}}{n} \right)_+^d \right]^2 \\
&\geq c \cdot \sum_{i\in(n_{j,p};n_{j,p+1}]} \left[\sum_{\ell=1}^{t_{j,p}} \sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}} \left(\frac{i-(I^B_{j,p,\ell})_+}{n} \right)_+^d \right]^2 \\
&= c \cdot \sum_{i\in(n_{j,p};n_{j,p+1}]} \left[\sum_{\ell=1}^{t_{j,p}} B_{j,p,\ell} \left(\frac{i-(I^B_{j,p,\ell})_+}{n} \right)_+^d \right]^2 \\
&\geq c \cdot \sum_{\ell=1}^{t_{j,p}} B^2_{j,p,\ell} \sum_{i\in(n_{j,p};n_{j,p+1}]} \left(\frac{i-(I^B_{j,p,\ell})_+}{n} \right)_+^{2d} \\
&\geq c \cdot \sum_{\ell=1}^{t_{j,p}} B^2_{j,p,\ell} \frac{(n_{j,p+1}-(I^B_{j,p,\ell})_+)^{2d+1}}{n^{2d}} \geq c \cdot \sum_{\ell=1}^{t_{j,p}} B^2_{j,p,\ell} \frac{(n_{j,p+1}-(I^B_{j,p,\ell})_-)^{2d+1}}{n^{2d}},
\end{aligned}
\tag{B.24}
$$

where $(I^B_{j,p,\ell})_+$ $((I^B_{j,p,\ell})_-)$ is defined to be the right (left) endpoint of $I^B_{j,p,\ell}$, and the last inequality follows from property (ii) of the partition $I^B_{j,p,\ell}$.

We are now ready to bound the term (I). First by the vanishing of interaction terms, we have $(I) = (I_1) + (I_2)$, where

$$
\begin{aligned}
(I_1) &\equiv \mathbb{E} \sup_{\theta^*\in\Theta^*} \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} \left[\sum_{q=0}^{q^*} a_{j,p,q} \left(\sum_{i\in(n_{j,p};n_{j,p,q}]} \left(\frac{n_{j,p,q}-i}{n} \right)_+^d \varepsilon_i \right) \right]^2, \\
(I_2) &\equiv \mathbb{E} \sup_{\theta^*\in\Theta^*} \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} \left[\sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \left(\sum_{i\in(n_{j,p,q};n_{j,p+1}]} \left(\frac{i-n_{j,p,q}}{n} \right)_+^d \varepsilon_i \right) \right]^2.
\end{aligned}
$$

Due to symmetry, we only bound (I_2) as follows:

$$
\begin{aligned}
(I_2) &= \mathbb{E}\sup_{\theta^*}\sum_{j,p}\Bigg[\sum_{q=q^*}^{s_{j,p}}\sum_{\ell=1}^{t_{j,p}}\mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}}b_{j,p,q}\Bigg(\sum_{i\in(n_{j,p,q};n_{j,p+1}]}\left(\frac{i-n_{j,p,q}}{n}\right)^d\varepsilon_i\Bigg)\Bigg]^2\\
&\le \mathbb{E}\sup_{\theta^*}\sum_{j,p}\Bigg[\sum_{\ell=1}^{t_{j,p}}\Bigg\{\sum_{q=q^*}^{s_{j,p}}\mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}}b_{j,p,q}\Bigg\}\max_{\tau\in I^B_{j,p,\ell}}\Bigg|\sum_{i\in(\tau;n_{j,p+1}]}\left(\frac{i-\tau}{n}\right)^d\varepsilon_i\Bigg|\Bigg]^2\\
&= \mathbb{E}\sup_{\theta^*}\sum_{j,p}\Bigg[\sum_{\ell=1}^{t_{j,p}}B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\Bigg|\sum_{i\in(\tau;n_{j,p+1}]}\left(\frac{i-\tau}{n}\right)^d\varepsilon_i\Bigg|\Bigg]^2\\
&\le \mathbb{E}\max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;n_{j,p+1}]}(i-\tau)^d\varepsilon_i\right)^2}{(n_{j,p+1}-(I^B_{j,p,\ell})_-)^{2d+1}}.
\end{aligned}
$$

Here, the first inequality follows from the non-negativity of $\{b_{j,p,q}\}$, the second equality follows from the definition of $B_{j,p,\ell}$, and the last inequality follows from Cauchy-Schwarz along with the estimates for $B_{j,p,\ell}$ in (B.24). Furthermore, we define

$$
\Delta^B \equiv \left\{\{\delta^B_{j,p,\ell}\} : \delta^B_{j,p,\ell}\in\{0,1\}, \sum_{j=1}^{k}\sum_{p=1}^{m_j}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\le k\right\}
$$

to be the admissible set for the sequence $\{\delta^B_{j,p,\ell}\}$. As $\sum_{j=1}^k\sum_{p=1}^{m_j}\sum_{\ell=1}^{t_{j,p}}1=\sum_{j=1}^k\sum_{p=1}^{m_j}\lceil\log_2(n_{j,p+1}-n_{j,p})\rceil\le Ck\lceil\log_2(n/k)\rceil$, a combinatorial estimate yields that $|\Delta^B|\le\binom{Ck\lceil\log_2(n/k)\rceil}{k}\le(Ce\lceil\log_2(n/k)\rceil)^k$.

Now, using the basic inequality $(a+b)^2\le 2(a^2+b^2)$, it suffices to bound by the order $k\log\log(16n/k)$ the following two terms:

$$
\mathbb{E}\max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}} \tag{B.25}
$$

and

$$
\mathbb{E}\max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in((I^B_{j,p,\ell})_+;n_{j,p+1}]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}}. \tag{B.26}
$$

From here on, in view of Theorem 3, the proof is essentially the same as that of Lemma 5.2 in Gao et al. (2019) (our (B.25) and (B.26) correspond to their (42) and (43)). For the sake of completeness, we will present the proof for the bound of (B.25); the bound for (B.26) follows from essentially the proof of (43) in Gao et al. (2019).

Denote the variable in (B.25) as Z, i.e.,

$$Z \equiv \max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B} \sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}} \delta^B_{j,p,\ell} \max_{\tau\in I^B_{j,p,\ell}} \frac{\Big(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\Big)^2}{\Big(n_{j,p+1}-(I^B_{j,p,\ell})_-\Big)^{2d+1}}.$$

We bound the tail probability of Z as follows. For any $u\geq 0$ and small enough $c>0$,

$$\begin{aligned}
&\mathbb{P}(Z>u)\\
&\leq \sum_{\{\delta^B_{j,p,\ell}\}\in\Delta^B} \mathbb{P}\Bigg[\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}} \delta^B_{j,p,\ell} \max_{\tau\in I^B_{j,p,\ell}} \frac{\Big(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\Big)^2}{\Big(n_{j,p+1}-(I^B_{j,p,\ell})_-\Big)^{2d+1}} > u\Bigg]\\
&\leq \sum_{\{\delta^B_{j,p,\ell}\}\in\Delta^B} e^{-cu}\prod_{j,p,\ell}\mathbb{E}\exp\Bigg[c\delta^B_{j,p,\ell} \max_{\tau\in I^B_{j,p,\ell}} \frac{\Big(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\Big)^2}{\Big(n_{j,p+1}-(I^B_{j,p,\ell})_-\Big)^{2d+1}}\Bigg]\\
&\lesssim \sum_{\{\delta^B_{j,p,\ell}\}\in\Delta^B} e^{-cu}\cdot\exp\Bigg(\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}} C\delta^B_{j,p,\ell}\log\log\big(16(n_{j,p+1}-n_{j,p})\big)\Bigg)\\
&\leq \exp\big(\log|\Delta^B|-cu+Ck\log\log(16n/k)\big)\leq \exp\big(-cu+Ck\log\log(16n/k)\big).
\end{aligned}$$

Here, the second inequality follows from the independence of the partial sum processes over the partition $\{I^B_{j,p,\ell}\}$, the third inequality follows by choosing c to be sufficiently small and then applying Theorem 3 with $\psi(x)=\exp(cx^2)-1$ therein, and the fourth inequality follows from the fact that $n_{j,p+1}-n_{j,p}\leq n/k$ and that $\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\leq k$ for any $\{\delta^B_{j,p,\ell}\}\in\Delta^B$. The proof is now complete by integrating the tail estimate. $\square$

B.3 Proof of lower bounds

B.3.1 Lower bound in Section 3.2

Proof of Proposition 14. We start with the first claim. In view of the fact that minimax rate over $\Theta(d, d_0, k)$ is non-decreasing in k and $\Theta(d, d-1, k) \subset \Theta(d, d_0, k)$ for any $d_0 \in [-1; d-1]$, it suffices to show that

$$\inf_{\widetilde{\theta}} \sup_{\theta \in \Theta(d,d-1,2)} \mathbb{E}_\theta \|\widetilde{\theta} - \theta\|^2 \geq c \log\log(16n).$$

For this, we will apply a standard reduction argument to multiple hypothesis testing (cf. Theorem 2.5 of Tsybakov (2009b)). Define the following series of splines. Let $M \equiv \lfloor \log_2(n/(d+1)) \rfloor$, and for each $\ell \in [1; M]$, $\tau_\ell \equiv \lfloor (1 - 2^{-\ell})n \rfloor$ and $f^\ell(x) \equiv \alpha_\ell (x - \tau_\ell/n)_+^d$ with $\alpha_\ell \equiv c(2^\ell)^{(2d+1)/2}\sqrt{\log\log(16n)/n}$ for some sufficiently small c. Further define $f^0(x) \equiv 0$ on $[0, 1]$, and the induced vectors $\theta_i^\ell \equiv f^\ell(i/n)$ for $i \in [1; n]$ and $\ell \in [0; M]$. Denote the corresponding joint distribution of $\{Y_i\}_{i=1}^n$ under the experiment (3.1) with truth θ^ℓ as P_ℓ, $\ell \in [0; M]$. It can be readily verified that $\theta^\ell \in \Theta(d, d-1, 2)$, and the Kullback-Leibler divergence between P_0 and each P_ℓ, denoted as $\mathrm{KL}(P_0, P_\ell)$, satisfies

$$\mathrm{KL}(P_0, P_\ell) = \|\theta^0 - \theta^\ell\|^2/2 = \|\theta^\ell\|^2/2 \asymp \log\log(16n)$$

for every $\ell \in [1; M]$. Moreover, for any $1 \leq j < k \leq M$, it holds by direct calculation that

$$\begin{aligned} d(P_j, P_k) \equiv \|\theta^j - \theta^k\|^2 &\geq \sum_{i \in (\tau_j, \tau_k]} (\theta_i^j - \theta_i^k)^2 \asymp \alpha_j^2 \frac{(\tau_k - \tau_j)^{2d+1}}{n^{2d}} \\ &\asymp \alpha_j^2 \frac{(n - \tau_j)^{2d+1}}{n^{2d}} \asymp \alpha_j^2 \frac{2^{-j(2d+1)}}{n} \asymp \log\log(16n). \end{aligned}$$

Theorem 2.5 in Tsybakov (2009b) therefore entails the desired lower bound.

Next, we prove the second claim. By following the same reduction as in the previous claim, it suffices to show that for any $k \geq k_0 + 1$, there exists some nonzero $f \in \mathcal{F}_n(d, d_0, k_0 + 1)$

such that $f(x) = 0$ for $x \in [0, c] \cup [1 - c, 1]$ with some universal c. Take $c = 1/3$. Let $\tau_0 \equiv 0$, $\tau_j \equiv 1/3 + (j - 1)/(3(k_0 - 1))$ for $j \in [1; k_0]$, and $\tau_{k_0+1} \equiv 1$. Define

$$f(x) \equiv \Big(\sum_{j=1}^{k_0-1} \sum_{\ell=d_0+1}^{d} c_\ell^j (x - \tau_j)_+^\ell \Big) \cdot \mathbf{1}_{[1/3,2/3]}(x), \quad x \in [0, 1].$$

By definition, f vanishes on $[0, 1/3] \cup [2/3, 1]$. Moreover, it can be readily checked that, for any real sequence $\{c_\ell^j\}_{j\in[1;k_0-1],\ell\in[d_0+1;d]}$, $f^{(\ell)}((\tau_j)_-) = f^{(\ell)}((\tau_j)_+)$ for $j \in [1; k_0 - 1]$ and $\ell \in [0; d_0]$. Therefore, in order to show that $f \in \mathcal{F}_n(d, d_0, k_0 + 1)$ and is non-zero, it suffices to show that there exists a non-zero realization of the sequence $\{c_\ell^j\}_{j\in[1;k_0-1],\ell\in[d_0+1;d]}$ such that $f^{(\ell)}((\tau_{k_0)})_-) = f^{(\ell)}((\tau_{k_0})_+) = 0$ for all $\ell \in [0; d_0]$. This is equivalent to finding a non-zero solution for the homogeneous linear system $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{b}$, where $\boldsymbol{c} \equiv \{c_\ell^j\}_{j\in[1;k_0-1],\ell\in[d_0+1;d]} \in \mathbb{R}^{(k_0-1)(d-d_0)}$, $\boldsymbol{b} \equiv \mathbf{0}_{(k_0-1)(d-d_0)}$, and

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots \boldsymbol{A}_{k_0-1} \end{bmatrix}$$

with

$$\boldsymbol{A}_j \equiv \begin{bmatrix} \underline{\odot}(d_0+1;0)\tau_{k_0,j}^{d_0+1} & \underline{\odot}(d_0+2;0)\tau_{k_0,j}^{d_0+2} & \ldots & \underline{\odot}(d;0)\tau_{k_0,j}^{d} \\ \underline{\odot}(d_0+1;1)\tau_{k_0,j}^{d_0} & \underline{\odot}(d_0+2;1)\tau_{k_0,j}^{d_0+1} & \ldots & \underline{\odot}(d;1)\tau_{k_0,j}^{d-1} \\ & \ldots & & \\ \underline{\odot}(d_0+1;d_0)\tau_{k_0,j} & \underline{\odot}(d_0+2;d_0)\tau_{k_0,j}^{2} & \ldots & \underline{\odot}(d;d_0)\tau_{k_0,j}^{d-d_0} \end{bmatrix}$$

and $\tau_{j_1,j_2} \equiv \tau_{j_1} - \tau_{j_2}$. Note that the coefficient matrix $\boldsymbol{A}$ has $d_0 + 1$ rows and $(k_0 - 1)(d - d_0)$ columns, where, by definition of k_0,

$$(k_0 - 1)(d - d_0) \geq d_0 + 2 \iff \left\lfloor \frac{d+1}{d-d_0} \right\rfloor + 1 \geq \frac{d+2}{d-d_0}.$$

The above equivalence indeed holds since if $(d + 1)/(d - d_0)$ is an integer, then

$$\left\lfloor \frac{d+1}{d-d_0} \right\rfloor + 1 = \frac{d+1+(d-d_0)}{d-d_0} \geq \frac{d+2}{d-d_0},$$

and if not

$$\left\lfloor \frac{d+1}{d-d_0} \right\rfloor + 1 \geq \left\lceil \frac{d+1}{d-d_0} \right\rceil \geq \frac{d+2}{d-d_0}.$$

This entails that the solution space of the linear system $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{b}$ is of dimension at least one and thus the system is guaranteed to have a non-trivial solution. The proof is thus complete. □

B.3.2 Lower bound in Section 3.3

Proof of Proposition 16. We will continue to adopt the standard reduction to multiple testing (cf. Theorem 2.5 of Tsybakov (2009b)) as in the proof of Proposition 14. We first introduce a set of basis functions. Let $\widetilde{k} \equiv k/3$ which we assume without loss of generality to be an integer, $\ell_0 \equiv \left\lfloor \log_2(n/(2\widetilde{k})) \right\rfloor$, and $\tau_\ell \equiv (1 - 2^{-(\ell-1)})/\widetilde{k}$ for $\ell \in [1; \ell_0 + 1]$. Next, for $x \in [0, 1/\widetilde{k}]$, let $\widetilde{f}_\ell(x) \equiv c(2^{\ell-1})^{3/2}\sqrt{\log\log(16n/k)/n}(x - \tau_\ell)_+$ for $\ell \in [1; \ell_0]$ and $f_{\mathrm{ref}}(x) \equiv c(2^{\ell_0})^{3/2}\sqrt{\log\log(16n/k)/n}(x - \tau_{\ell_0+1})_+$ (here the subscript "ref" stands for "reference" and f_{ref} will be pieced together later to be the true signal underlying the distribution P_0 in Theorem 2.5 of Tsybakov (2009b)). Then let $f_\ell(x) \equiv \widetilde{f}_\ell(x) \vee f_{\mathrm{ref}}(x)$, and it can be verified that $f_\ell(x) = \widetilde{f}_\ell(x)$ on $[0, \tau_{\ell_0+1}]$. The above set of functions resembles those constructed in the proof of Proposition 14, and satisfies the similar properties

$$\sum_{i:(i/n)\in(0,1/\widetilde{k}]} \big(f_\ell(i/n) - f_{\ell'}(i/n)\big)^2 \geq c\log\log(16n/k) \tag{B.27}$$

for any $1 \leq \ell \neq \ell' \leq \ell_0$, and

$$\begin{aligned}
\sum_{i:(i/n)\in(0,1/\widetilde{k}]} \big(f_\ell(i/n) - f_{\mathrm{ref}}(i/n)\big)^2 &\leq \sum_{i:(i/n)\in(0,1/\widetilde{k}]} \big(f_\ell(i/n)\big)^2 \\
&\leq 2\bigg(\sum_{i:(i/n)\in(0,1/\widetilde{k}]} \big(\widetilde{f}_\ell(i/n)\big)^2 + \sum_{i:(i/n)\in(0,1/\widetilde{k}]} \big(f_{\mathrm{ref}}(i/n)\big)^2\bigg) \\
&\leq C\log\log(16n/k).
\end{aligned} \tag{B.28}$$

We now construct the hypotheses in the multiple testing framework. For $j \in [1; \widetilde{k}]$, let $f_\ell^j(\cdot), f_{\mathrm{ref}}^j(\cdot)$ be a set of functions defined on $[(j-1)/\widetilde{k}, j/\widetilde{k}]$ as follows. Let $f_\ell^1(x) \equiv f_\ell(x)$ and $f_{\mathrm{ref}}^1(x) \equiv f_{\mathrm{ref}}(x)$ as defined above. Next, for $j \in [2; k]$, we define inductively $f_{\mathrm{ref}}^j(x) \equiv$

$f_{\mathrm{ref}}^{j-1}(x) + f_{\mathrm{ref}}(x-(j-1)/\widetilde{k})$, where $f_{\mathrm{ref}}^{j-1}(x)$ for $x \in [(j-1)/\widetilde{k}, j/\widetilde{k}]$ is to be understood as the extension from $[(j-2)/\widetilde{k}, (j-1)/\widetilde{k}]$. Also define $f_{\ell}^{j}(x) \equiv f_{\mathrm{ref}}^{j}(x) + f_{\ell}(x-(j-1)/\widetilde{k})$. Lastly, we piece them together as

$$f^0(x) \equiv \sum_{j=1}^{\widetilde{k}} f_{\mathrm{ref}}^{j}(x)\mathbf{1}_{((j-1)/\widetilde{k}, j/\widetilde{k}]}(x)$$

and

$$f^{\boldsymbol{\ell}}(x) \equiv \sum_{j=1}^{\widetilde{k}} f_{\ell_j}^{j}(x)\mathbf{1}_{((j-1)/\widetilde{k}, j/\widetilde{k}]}(x),$$

where $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_{\widetilde{k}})^\top \in [1;\ell_0]^{\widetilde{k}}$. One can readily verify that all of the f^0 and $f^{\boldsymbol{\ell}}$ belong to the class $\mathcal{F}_n^*(1,k)$. Indeed, continuity follows directly from the construction and since there are at most 3 pieces on each of $[(j-1)/\widetilde{k}, j/\widetilde{k}]$, there will be at most $3\widetilde{k} = k$ pieces in total. Therefore, the sequence counterparts $\theta^0 \equiv (f^0(i/n))_i$ and $\theta^{\boldsymbol{\ell}} \equiv (f^{\boldsymbol{\ell}}(i/n))_i$ belong to $\Theta^*(1,k)$.

Let $\rho(\cdot,\cdot)$ denote the Hamming distance. Then, the Gilbert-Varshamov bound (cf. Theorems 5.1.7 and 5.1.9 in van Lint (1999)) entails that with some small $c > 0$, there exists a subset $\mathcal{S} \subset [1;\ell_0]^{\widetilde{k}}$ with cardinality $|\mathcal{S}| \asymp \ell_0^{c\widetilde{k}}$ such that $\rho(\boldsymbol{\ell}, \boldsymbol{\ell}') \geq c\widetilde{k}$ for any $\boldsymbol{\ell} \neq \boldsymbol{\ell}' \in \mathcal{S}$. Adopting those in $\mathcal{S}$ as the truth in the experiment (3.1), we obtain a total of $M \equiv 1 + |\mathcal{S}| \asymp \ell_0^{c\widetilde{k}}$ hypotheses, which we denote as P^0 and $P^{\boldsymbol{\ell}}$, $\boldsymbol{\ell} \in \mathcal{S}$.

It remains to verify: (i) $\|\theta^{\boldsymbol{\ell}} - \theta^{\boldsymbol{\ell}'}\|^2 \geq ck\log\log(16n/k)$ for any $\boldsymbol{\ell} \neq \boldsymbol{\ell}' \in \mathcal{S}$; (ii) $\mathrm{KL}(P^0, P^{\boldsymbol{\ell}}) \leq C\log|\mathcal{S}|$ for any $\boldsymbol{\ell} \in \mathcal{S}$. We first verify (i). By definition of $\theta^{\boldsymbol{\ell}}$ and $\theta^{\boldsymbol{\ell}'}$, on each $[(j-1)/\widetilde{k}, j/\widetilde{k}]$ such that $\ell_j \neq \ell_j'$, we have by (B.27),

$$\begin{aligned}\sum_{i:\frac{i}{n}\in(\frac{j-1}{\widetilde{k}},\frac{j}{\widetilde{k}}]} (\theta_i^{\boldsymbol{\ell}} - \theta_i^{\boldsymbol{\ell}'})^2 &= \sum_{i:\frac{i}{n}\in(\frac{j-1}{\widetilde{k}},\frac{j}{\widetilde{k}}]} \left[f_{\ell_j}\left(\frac{i}{n} - \frac{j-1}{\widetilde{k}}\right) - f_{\ell_j'}\left(\frac{i}{n} - \frac{j-1}{\widetilde{k}}\right)\right]^2 \\ &= \sum_{i:\frac{i}{n}\in(0,\frac{1}{\widetilde{k}}]} \left[f_{\ell_j}\left(\frac{i}{n}\right) - f_{\ell_j'}\left(\frac{i}{n}\right)\right]^2 \geq c\log\log(16n/k).\end{aligned}$$

This entails that

$$\|\theta^{\boldsymbol{\ell}} - \theta^{\boldsymbol{\ell}'}\|^2 \geq \rho(\boldsymbol{\ell}, \boldsymbol{\ell}')c\log\log(16n/k) \geq ck\log\log(16n/k).$$

Similarly, for (ii), we have by (B.28)

$$\mathrm{KL}(P^0, P^{\boldsymbol{\ell}}) = \|\theta^0 - \theta^{\boldsymbol{\ell}}\|^2/2 \le C\widetilde{k} \cdot \sum_{i:\frac{i}{n}\in(0,\frac{1}{\widetilde{k}}]} \left[f_{\ell_j}\left(\frac{i}{n}\right) - f_{\mathrm{ref}}\left(\frac{i}{n}\right)\right]^2$$
$$\le C\widetilde{k} \log\log(16n/k) \asymp \log|\mathcal{S}|.$$

Application of Theorem 2.5 in Tsybakov (2009b) then completes the proof. □

Proof of Theorem 2 (lower bound). This is immediate by realizing that $\Theta^*(d,2) \subset \Theta^*(d,k)$ for $k \ge 2$ and the lower bound construction in the first part of the proof of Proposition 14 can be directly applied to establish a lower bound for $\Theta^*(d,2)$. □

B.4 Proof of Theorem 3

Proof of Theorem 3. We first claim that there exists some $c = c(d)$ such that for any $t > 0$, the event

$$\mathcal{E}_1 \equiv \left\{ \max_{1\le n_1<n_2\le n} (n_2 - n_1)^{-d}(n_2 \wedge (n-n_1))^{-1/2} \left| \sum_{(n_1;n_2]} (i-n_1)^d \varepsilon_i \right| \ge t \right\}$$

is contained in the event

$$\mathcal{E}_2 \equiv \left\{ \max_{1\le n_1<n_2\le n} (n_2 \wedge (n-n_1))^{-1/2} \left| \sum_{(n_1;n_2]} \varepsilon_i \right| \ge ct \right\}.$$

On $\mathcal{E}_2^c$, for any $1 \le n_1 < n_2 \le n$, it holds that $\big|\sum_{(n_1;n_2]} \varepsilon_i\big| \le c(n_2 \wedge (n-n_1))^{1/2} t$. Then,

$$
\begin{aligned}
\Big|\sum_{(n_1;n_2]} \varepsilon_i (i-n_1)^d\Big| &= \Big|\sum_{i\in(n_1;n_2]} \varepsilon_i \sum_{j=1}^{i-n_1} \big(j^d - (j-1)^d\big)\Big| \\
&= \Big|\sum_{j=1}^{n_2-n_1} \big(j^d - (j-1)^d\big) \sum_{i\in[n_1+j;n_2]} \varepsilon_i\Big| \\
&\le \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{j=1}^{n_2-n_1} (j-1)^\ell \Big|\sum_{i\in[n_1+j;n_2]} \varepsilon_i\Big| \\
&\le ct \cdot \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{j=1}^{n_2-n_1} (j-1)^\ell \Big(\sqrt{n_2} \wedge \sqrt{n-n_1-(j-1)}\Big) \\
&\le 2ct \cdot \sum_{\ell=0}^{d-1} \binom{d}{\ell} \int_0^{n_2-n_1} x^\ell \big(\sqrt{n_2} \wedge \sqrt{n-n_1-x}\big) \mathrm{d}x \qquad \text{(B.29)} \\
&\le 4ct \cdot \sum_{\ell=0}^{d-1} \binom{d}{\ell} (n_2-n_1)^{\ell+1} (n_2 \wedge (n-n_1))^{1/2} \qquad \text{(B.30)} \\
&\le c2^{d+2} t \cdot (n_2-n_1)^d (n_2 \wedge (n-n_1))^{1/2},
\end{aligned}
$$

where the inequality (B.29) follows from the fact that the map $x \mapsto x^\ell(\sqrt{n_2} \wedge \sqrt{n-n_1-x})$ first increases and then decreases on $[0, n-n_1]$, and the inequality (B.30) follows from a separate discussion of $n_2 \le n-n_1$ and $n_2 > n-n_1$ and the following two bounds: $\int_0^{n_2-n_1} x^\ell \,\mathrm{d}x = (\ell+1)^{-1}(n_2-n_1)^{\ell+1}$ and

$$
\begin{aligned}
&\int_0^{n_2-n_1} x^\ell \sqrt{n-n_1-x}\,\mathrm{d}x \le (n_2-n_1)^\ell \int_{n_1}^{n_2} \sqrt{n-x}\,\mathrm{d}x \\
&= (n_2-n_1)^\ell \int_{n-n_2}^{n-n_1} \sqrt{x}\,\mathrm{d}x = (n_2-n_1)^\ell \cdot \frac{2}{3}\big((n-n_1)^{3/2} - (n-n_2)^{3/2}\big) \\
&= (n_2-n_1)^\ell \cdot \frac{2}{3} \frac{(n_2-n_1)\big[(n-n_1)^2 + (n-n_1)(n-n_2) + (n-n_2)^2\big]}{(n-n_1)^{3/2} + (n-n_2)^{3/2}} \\
&\le 2(n_2-n_1)^{\ell+1}(n-n_1)^{1/2}.
\end{aligned}
$$

Therefore the claim holds by choosing $c = 2^{-(d+2)}$. This entails that, for any $t > 0$,

$$
\begin{aligned}
\mathbb{P}(Z \geq t) &\leq \mathbb{P}\bigg(\max_{1 \leq n_1 < n_2 \leq n} (n_2 \wedge (n - n_1))^{-1/2} \bigg| \sum_{(n_1;n_2]} \varepsilon_i \bigg| \geq ct \bigg) \\
&\leq \mathbb{P}\bigg(\max_{n_1 < n_2} \frac{\big| \sum_{(n_1;n_2]} \varepsilon_i \big|}{(n - n_1)^{1/2}} \geq ct \bigg) + \mathbb{P}\bigg(\max_{n_1 < n_2} \frac{\big| \sum_{(n_1;n_2]} \varepsilon_i \big|}{n_2^{1/2}} \geq ct \bigg) \\
&\equiv (I) + (II).
\end{aligned}
$$

Due to symmetry, we only bound (I). By the triangle inequality,

$$
(I) \leq \mathbb{P}\bigg(\sup_{n_1 < n_2} \frac{\big| \sum_{i=n_1+1}^{n} \varepsilon_i \big|}{(n - n_1)^{1/2}} > ct/2 \bigg) + \mathbb{P}\bigg(\sup_{n_1 < n_2} \frac{\big| \sum_{i=n_2+1}^{n} \varepsilon_i \big|}{(n - n_1)^{1/2}} > ct/2 \bigg).
$$

By Lévy's maximal inequality (cf. Theorem 1.1.5 of de la Peña and Giné (1999)), the first probability is bounded by

$$
\sum_{r=1}^{\lceil \log_2 n \rceil} \mathbb{P}\bigg(\sup_{2^{r-1} \leq (n - n_1) < 2^r} 2^{-(r-1)/2} \bigg| \sum_{i=n_1+1}^{n} \varepsilon_i \bigg| \geq ct/2 \bigg) \leq 9 \lceil \log_2 n \rceil e^{-c' t^2}.
$$

Similarly, the second inequality is bounded by

$$
\begin{aligned}
&\sum_{r=1}^{\lceil \log_2 n \rceil} \mathbb{P}\bigg(\sup_{\substack{2^{r-1} \leq (n - n_1) < 2^r \\ 1 \leq n_1 < n_2 \leq n}} (n - n_1)^{-1/2} \bigg| \sum_{i=n_2+1}^{n} \varepsilon_i \bigg| \geq ct/2 \bigg) \\
&\leq \sum_{r=1}^{\lceil \log_2 n \rceil} \mathbb{P}\bigg(\sup_{n - 2^r < n_2 \leq n} 2^{-(r-1)/2} \bigg| \sum_{i=n_2+1}^{n} \varepsilon_i \bigg| \geq ct/2 \bigg) \leq 9 \lceil \log_2 n \rceil e^{-c' t^2}.
\end{aligned}
$$

Putting together the pieces, it holds that $\mathbb{P}(Z \geq t) \leq 18 \lceil \log_2 n \rceil e^{-c'' t^2}$, where we take $c'' < c_0$ without loss of generality. Now, if $\psi(\cdot)$ is bounded on $[0, \infty)$ by some C, then the result holds trivially. Otherwise, $\psi(x) \uparrow \infty$ as $x \to \infty$, and integration by parts yields that for any $x_0 \geq 0$,

$$
\begin{aligned}
\mathbb{E}\psi(Z) &= \int_0^\infty \mathbb{P}(\psi(Z) \geq t)\ \mathrm{d}t = \int_0^\infty \mathbb{P}(Z \geq \psi^{-1}(t))\ \mathrm{d}t \\
&\leq \int_0^\infty \Big\{ 1 \wedge \big[C \log(16n) \cdot e^{-c''(\psi^{-1}(t))^2} \big] \Big\}\ \mathrm{d}t \\
&\leq x_0 + C \cdot \int_{x_0}^\infty \log(16n) \cdot e^{-c''(\psi^{-1}(t))^2}\ \mathrm{d}t.
\end{aligned}
$$

By monotonicity of ψ^{-1}, for any $t \geq x_0$, $\psi^{-1}(t) \geq \psi^{-1}(t)/2 + \psi^{-1}(x_0)/2$, so the integral above can be further bounded by

$$\int_{x_0}^{\infty} \big[\log(16n) \cdot e^{-(c''/4)(\psi^{-1}(x_0))^2}\big] e^{-(c''/4)(\psi^{-1}(t))^2} \ \mathrm{d}t \leq \int_{1}^{\infty} e^{-(c''/4)(\psi^{-1}(t))^2} \ \mathrm{d}t,$$

provided that $x_0 \geq 1$ and $\log(16n) \cdot e^{-(c''/4)(\psi^{-1}(x_0))^2} \leq 1$, or equivalently, $x_0 \geq 1 \vee \psi\big(\sqrt{(4/c'') \log\log(16n)}\big)$. The claim now follows from the condition (3.21). □

B.5 Proofs for technical results in Section B.1

B.5.1 Proof of Proposition 27

Proof of Proposition 27. The basic inequality $\|Y - \widehat{\theta}\|^2 \leq \|Y - \theta_{\text{oracle}}\|^2$ entails that

$$\|\widehat{\theta} - \theta_0\|^2 \leq \|\theta_{\text{oracle}} - \theta_0\|^2 + 2\varepsilon \cdot \big(\widehat{\theta} - \theta_{\text{oracle}}\big).$$

Then we have, for any $\eta > 0$,

$$\begin{aligned}
\varepsilon \cdot \big(\widehat{\theta} - \theta_{\text{oracle}}\big) &= \sum_{j=0}^{k-1} (\varepsilon_{[j]} \cdot (\widehat{\theta} - \theta_{\text{oracle}})_{[j]}) = \sum_{j=0}^{k-1} (\varepsilon_{[j]} \cdot v_j(\widehat{\theta})) \|(\widehat{\theta} - \theta_{\text{oracle}})_{[j]}\| \\
&\leq \eta^{-1} \cdot \sum_{j=0}^{k-1} \big(\varepsilon_{[j]} \cdot v_j(\widehat{\theta})\big)^2 + \eta \cdot \sum_{j=0}^{k-1} \|(\widehat{\theta} - \theta_{\text{oracle}})_{[j]}\|^2 \\
&= \eta^{-1} \cdot \sum_{j=0}^{k-1} \big(\varepsilon_{[j]} \cdot v_j(\widehat{\theta})\big)^2 + \eta \cdot \|\widehat{\theta} - \theta_{\text{oracle}}\|^2.
\end{aligned}$$

Applying the inequality $\|\widehat{\theta} - \theta_{\text{oracle}}\|^2 \leq 2\big(\|\widehat{\theta} - \theta_0\|^2 + \|\theta_{\text{oracle}} - \theta_0\|^2\big)$ then yields that

$$\|\widehat{\theta} - \theta_0\|^2 \leq \frac{1 + 2\eta}{1 - 2\eta} \|\theta_{\text{oracle}} - \theta_0\|^2 + \frac{1}{\eta(1 - 2\eta)} \sum_{j=0}^{k-1} (\varepsilon_{[j]} \cdot v_j(\widehat{\theta}))^2.$$

For any given $\delta > 0$, choosing $\eta = \delta/(2\delta + 4)$, upper bounding the right-hand side by the supremum over $\Theta(d, d_0, k)$, and then taking expectation on both sides yield the desired result. □

B.5.2 Proof of Lemma 29

Proof of Lemma 29. On the pieces $(n_{i-1}/n, n_i/n]$ and $(n_i/n, n_{i+1}/n]$, the function f can be parametrized as

$$f_{i-1}(x) \equiv \sum_{q=1}^{d+1} a_q^{i-1}\left(x - \frac{n_{i-1}}{n}\right)^{q-1}, \quad f_i(x) \equiv \sum_{q=1}^{d+1} a_q^{i}\left(x - \frac{n_{i}}{n}\right)^{q-1}.$$

By the fact that $0 \leq p-1 \leq d_0$ and thus the continuity of the $(p-1)$th derivative at knot n_i/n, it holds that $f_{i-1}^{(p-1)}(n_i/n) = f_i^{(p-1)}(n_i/n)$. But

$$f_{i-1}^{(p-1)}\left(\frac{n_i}{n}\right) = \sum_{q=1}^{d+1} a_q^{i-1}\frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}}\left(x - \frac{n_{i-1}}{n}\right)^{q-1}\Bigg|_{x=\frac{n_i}{n}} = \sum_{q=p}^{d+1} a_q^{i-1}\underline{\odot}(q-1;p-1)n_{i;i-1}^{q-p},$$

$$f_{i}^{(p-1)}\left(\frac{n_i}{n}\right) = \sum_{q=1}^{d+1} a_q^{i}\frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}}\left(x - \frac{n_{i}}{n}\right)^{q-1}\Bigg|_{x=\frac{n_i}{n}} = (p-1)!a_p^i.$$

This entails that

$$(p-1)!a_p^i = \sum_{q=p}^{d+1} \underline{\odot}(q-1;p-1)a_q^{i-1}n_{i;i-1}^{q-p} = \sum_{q=p}^{d+1}\frac{(q-1)!}{(q-p)!}a_q^{i-1}n_{i;i-1}^{q-p}.$$

This implies that $\mathrm{Coef}[a_p^i; a_q^{i-1}] = (q-1)!/((q-p)!(p-1)!)n_{i;i-1}^{q-p} = \binom{q-1}{p-1}n_{i;i-1}^{q-p}$ if $q \geq p$; otherwise it is 0. $\square$

B.5.3 Proof of Lemma 30

Proof of Lemma 30. The baseline case $s = 0$ follows from the condition $\|\theta\| \leq 1$ and application of Lemma 37 to the piece $(n_{k_0-1}; n_{k_0}]$. The iteration from s to $s+1$ then follows from Lemma 41, which is to be stated and proved in Appendix B.7 with its conditions satisfied since $n_{k_0;k_0-1} \geq \max\{n_{2;1}, n_{3;2}, \ldots, n_{k_0-1;k_0-2}\}$ by (B.11). $\square$

B.5.4 Proof of Lemma 31

Proof of Lemma 31. Fix $i \leq k_0 - 2$ as in the lemma statement. For simplicity, we again work under the condition $n_{k_0;k_0-1} = \max\{n_{2;1}, \ldots, n_{k_0;k_0-1}\}$. We will prove by induction: suppose the desired estimates hold for a_ℓ^i, $\ell \in [d_0+1;\ell_0]$ for some $\ell_0 \in [d_0+1;d]$ and we will prove that the estimate also holds for $a_{\ell_0+1}^i$. The condition of the lemma serves as the baseline $\ell_0 = d_0+1$. For the general induction from ℓ_0 to ℓ_0+1, let $L \equiv 1+(d-d_0)(k_0-1-i)$. Then, Lemma 30 entails that

$$1 \gtrsim \frac{(n-n_{k_0-1})^{2(\ell_0+1-L)+1}}{n^{2(\ell_0+1-L)}} \bigg(\sum_{\ell=\ell_0+2-L}^{\ell_0+1} \overline{\beta}_{\ell_0+2-L,\ell-(\ell_0+2-L)}^{k_0-1-i} a_\ell^i \bigg)^2.$$

On the other hand, we have

$$1 \gtrsim \sum_{\ell=\ell_0+2-L}^{\ell_0+1} n_{i+1;i}^{2(\ell-1)} M^2(n_{i+1},n_i)(a_\ell^i)^2,$$

where the summands with $\ell \in [\ell_0+2-L; d_0+1]$ are from the condition of the lemma and those with $\ell \in [d_0+2;\ell_0+1]$ are from the induction assumption. Now, combining the above two estimates and applying Lemma 39 iteratively to cancel every a_ℓ^i, $\ell \in [\ell_0+2-L;\ell_0]$, we have

$$1 \gtrsim (a_{\ell_0+1}^i)^2((I) \wedge (II)),$$

where

$$\begin{aligned}
(I) &\equiv \frac{(n-n_{k_0-1})^{2(\ell_0+1-L)+1}}{n^{2(\ell_0+1-L)}} (\overline{\beta}_{\ell_0+2-L,L-1}^{k_0-1-i})^2, \\
(II) &\equiv \bigwedge_{\ell=\ell_0+2-L}^{\ell_0} n_{i+1;i}^{2(\ell-1)} M(n_{i+1},n_i) \frac{(\overline{\beta}_{\ell_0+2-L,L-1}^{k_0-1-i})^2}{(\overline{\beta}_{\ell_0+2-L,\ell-(\ell_0+2-L)}^{k_0-1-i})^2}.
\end{aligned}$$

By Lemma 43 and the condition $n_{k_0;k_0-1} = \max\{n_{2;1},\ldots,n_{k_0;k_0-1}\}$, we obtain that $(I) \gtrsim n_{i+1;i}^{2\ell_0} M(n_{i+1},n_i)$. Similarly, by Lemma 43, as the factors $n_{\cdot;\cdot}$'s in the lower bound of $(\overline{\beta}_{\ell_0+2-L,L-1}^{k_0-1-i})^2/(\overline{\beta}_{\ell_0+2-L,\ell-(\ell_0+2-L)}^{k_0-1-i})^2$ can all be further bounded below by $n_{i+1;i}$, we obtain by direct calculation that $(II) \gtrsim n_{i+1;i}^{2\ell_0} M(n_{i+1},n_i)$. Putting together the lower bounds for (I), (II) completes the induction. $\square$

B.5.5 General statement of Lemma 32

We restate here Lemma 32 for the case of general $d_0 \in [-1; d-1]$. Introduce the following notation:

$$\underline{\odot} n_{\cdot;j}(a,b,c) \equiv n^c_{a;j} \cdot n^c_{a-1;j} \dots n^{\mathrm{Mod}(b;c)}_{a-\lfloor b/c \rfloor - 1;j}$$

for positive integers a, b, c. Fix $i \geq 2$. Recall the definition $M(a,b) = (a \wedge (n-b))^{1/2}$ for $a, b \in [1; n]$ and the condition (B.11).

Lemma 35. *The following estimates hold for all locations $1 \leq j \leq i+1$:*

$$\begin{aligned}
1 \gtrsim \max_{1 \leq \ell \leq d_0+1} &\Big\{ n^{2\ell \vee 2(d-d_0)(k_0-i-2)}_{i+2;j} \\
&\times \underline{\odot} n^2_{\cdot;j}\Big(i+1, \big\{\ell - (d-d_0)(k_0-i-2) - 1\big\} \wedge \big\{(d-d_0)(i+1-j)\big\}, d-d_0\Big) \\
&\times n^{2(\ell-1-(d-d_0)(k_0-j-1))_+}_{j+1;j} \Big\} \cdot M^2(n_{j+1}, n_j).
\end{aligned}$$

In particular, for $j = i+1$:

$$1 \geq c \max_{1 \leq \ell \leq d_0+1} \Big\{ (a^{i+1}_\ell)^2 \cdot n^{2(\ell-1)}_{i+2;i+1} \cdot M^2(n_{i+2}, n_{i+1}) \Big\}.$$

The proof for this general case is completely analogous to the one presented in Section B.1.5.

B.6 Proofs for technical results in Section B.2

B.6.1 Proof of Lemma 33

Proof of Lemma 33. For any $f \in \mathcal{F}^*_n(d,k)$, let $f_\circ \equiv f_\circ(f) \in \mathcal{F}^*_n(0,k)$ be such that $f = (I^d_{r_0,\dots,r_{d-1};0} f_\circ)$ for some real sequence $\{r_\ell\}_{\ell=0}^{d-1}$, with corresponding knots $\{n_j\}_{j=1}^{k-1} = \{n_j(f_\circ)\}_{j=1}^{k-1}$ and magnitudes $\{\mu_j\}_{j=1}^k = \{\mu_j(f_\circ)\}_{j=1}^k$ between $(n_{j-1}/n, n_j/n]$, i.e., $f_\circ(x) = \sum_{j=1}^k \mu_j \mathbf{1}_{(n_{j-1}/n, n_j/n]}(x)$ for $x \in (0,1]$. Then $\mu_1 \leq \ldots \leq \mu_k$. Let

$$j^* \equiv j^*(f_\circ) \equiv \max\{1 \leq j \leq k : \mu_j \leq 0\}.$$

Define two sequences $\{\widetilde{a}_j\}_{j=1}^{j^*}$ and $\{\widetilde{b}_j\}_{j=j^*}^{k-1}$ as follows: $\widetilde{a}_{j^*} \equiv \mu_{j^*} \le 0$ and $\widetilde{a}_j \equiv \mu_j - \mu_{j+1} \le 0$ for $j \in [1; j^* - 1]$, $\widetilde{b}_{j^*} \equiv \mu_{j^*+1} \ge 0$ and $\widetilde{b}_j \equiv \mu_{j+1} - \mu_j \ge 0$ for $j \in [j^* + 1; k - 1]$. Then, letting $\tau_j \equiv n_j/n$, $f_\circ$ can be re-parametrized as

$$f_\circ(x) = \sum_{j=1}^{j^*} \widetilde{a}_j \mathbf{1}_{(0,\tau_j]}(x) + \sum_{j=j^*}^{k-1} \widetilde{b}_j \mathbf{1}_{(\tau_j,1]}(x), \quad x \in (0, 1].$$

Define the function $g_\ell^-(x;\tau) \equiv (\tau - x)_+^\ell$ with any parameter $\tau \in [0, 1]$. Then, direct calculation shows that

$$\begin{aligned}\int_0^x g_\ell^-(u;\tau)\ \mathrm{d}u &= \int_0^{x\wedge\tau} (\tau - u)^\ell\ \mathrm{d}u = \int_{\tau - x\wedge\tau}^{\tau} u^\ell\ \mathrm{d}u \\ &= \frac{\tau^{\ell+1}}{\ell+1} - \frac{(\tau - x)_+^{\ell+1}}{\ell+1} = \frac{\tau^{\ell+1}}{\ell+1} + \frac{(-1)}{\ell+1} \cdot g_{\ell+1}^-(x;\tau).\end{aligned}$$

Similarly, with $g_\ell^+(x;\tau) \equiv (x-\tau)_+^\ell$, it holds that $\int_0^x g_\ell^+(u;\tau)\ \mathrm{d}u = \int_\tau^{x\vee\tau} (u-\tau)^\ell\ \mathrm{d}u = \int_0^{x\vee\tau-\tau} u^\ell\ \mathrm{d}u = g_{\ell+1}^+(x;\tau)/(\ell+1)$. This entails that

$$(I^d_{r_0,\ldots,r_{d-1};0} f_\circ)(x) = \sum_{j=1}^{j^*} (-1)^d \frac{\widetilde{a}_j}{d!} (\tau_j - x)_+^d + \sum_{j=j^*}^{k-1} \frac{\widetilde{b}_j}{d!} (x - \tau_j)_+^d + P_{d-1}(x),$$

where $P_{d-1}(x)$ is some polynomial of order $d-1$. The proof is then complete by noting that $\{(-1)^d \widetilde{a}_j/d!\}_{j=1}^{j^*}$ has sign $(-1)^{d+1}$ and $\{\widetilde{b}_j/d!\}_{j=j^*}^{k-1}$ is non-negative. □

B.6.2 Proof of Lemma 34

We need the following simple fact that translates the ℓ_2 constraint on θ^* at the sequence level to an integral L_2 constraint on f^* at the underlying function level. Its proof can be found after the proof of Lemma 34.

Lemma 36. *Let $f^* \in \mathcal{F}_n^*(d,k)$ and $(\theta^*)_i \equiv (f^*(i/n))_i$. Then, if $\|\theta^*\|^2 \le 1$, there exists some $c = c(d)$ such that $1 \ge c \cdot n \int_0^1 (f^*)^2(x)\ \mathrm{d}x$. Actually, this inequality holds for the larger unshaped spline space $\mathcal{F}_n(d, d_0, k)$.*

Proof of Lemma 34. Fix any $\theta \in \Theta^*(d,k)$ and its generating spline $f \in \mathcal{F}_n^*(d,k)$. Then, under the condition $\|\theta\|^2 \le 1$, Lemma 36 entails that there exists some $K = K(d) > 0$ such that $\int_0^1 f^2(x)\,\mathrm{d}x \le K/n$. Due to scale invariance, it suffices to prove that $|c_\ell(f)| \le C$ for $\ell \in [0; d-1]$ for some $C = C(d)$ under the condition $\|f\|_2^2 = \int_0^1 f^2(x)\,\mathrm{d}x \le 1$.

For $f \in \mathcal{F}_n^*(d,k)$, let $\{n_j = n_j(f)\}_{j\in[0;k]}$ be its knots and $j^* = j^*(f)$ be as in its canonical form in Lemma 33. Let $\tau_j \equiv \tau_j(f) \equiv n_j(f)/n$ for $j \in [1;k]$ and $\tau^* \equiv \tau^*(f) \equiv \tau_{j^*(f)}(f)$. We will prove that for some $K = K(d) > 0$,

$$\int_0^1 f^2(x)\,\mathrm{d}x \ge K \cdot \max_{0\le\ell\le d-1} c_\ell^2(f), \quad \text{for any } f \in \mathcal{F}_n^*(d,k).$$

We focus on the case $\tau^*(f) \in [0, 1/2]$ and prove that

$$\int_{\tau^*(f)}^1 f^2(x)\,\mathrm{d}x \ge K \cdot \max_{0\le\ell\le d-1} c_\ell^2(f), \quad \text{for any } f \in \mathcal{F}_n^*(d,k).$$

We present the proof for $c_{d-1}(f)$ whenever $c_{d-1}(f) \neq 0$; the bounds for $\{c_\ell(f)\}_{\ell\in[0;d-2]}$ follow from completely analogous arguments. Below we omit notational dependence on f if no confusion could arise. On $[\tau^*, 1]$, f has the canonical form

$$f(x) = \sum_{j=j^*}^{k-1} b_j (x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} \frac{c_\ell}{\ell!} x^\ell.$$

This can be alternatively parametrized as $f(x) = \sum_{\ell=0}^{d-1} c_\ell x^\ell/\ell! + (I^d_{0,\ldots,0;\tau^*(f)} f_\circ)(x)$, where $f_\circ(x) \equiv \sum_{j=j^*}^{k-1} (b_j \cdot d!)\mathbf{1}_{x>\tau_j} \in \mathcal{F}_n^*(0,k)$, and $\tau^*(f) = \tau^*(f_\circ)$. Therefore, we have

$$\begin{aligned}
1 &\ge \int_{\tau^*(f)}^1 \Big(\sum_{\ell=0}^{d-1} c_\ell x^\ell/\ell! + (I^d_{0,\ldots,0;\tau^*(f)} f_\circ)(x)\Big)^2\,\mathrm{d}x\\
&= c_{d-1}^2 \int_{\tau^*(f_\circ)}^1 \bigg[\sum_{\ell=0}^{d-2} \frac{c_\ell}{|c_{d-1}|}\frac{x^\ell}{\ell!} + \mathrm{sgn}(c_{d-1})\frac{x^{d-1}}{(d-1)!} + \Big(I^d_{0,\ldots,0;\tau^*(f_\circ)}\frac{f_\circ}{|c_{d-1}|}\Big)(x)\bigg]^2\,\mathrm{d}x\\
&\ge c_{d-1}^2 \inf_{\substack{c'_0,\ldots,c'_{d-2}\in\mathbb{R}, c'_{d-1}\in\{\pm1\}\\ \widetilde{f}_\circ\in\cup_n\mathcal{F}_n^*(0,k),\tau^*(\widetilde{f}_\circ)\le 1/2}} \int_{\tau^*(\widetilde{f}_\circ)}^1 \bigg[\sum_{\ell=0}^{d-1}\frac{c'_\ell x^\ell}{\ell!} + (I^d_{0,\ldots,0;\tau^*(\widetilde{f}_\circ)}\widetilde{f}_\circ)(x)\bigg]^2\,\mathrm{d}x\\
&= c_{d-1}^2 \inf_{\substack{c''_0,\ldots,c''_{d-2}\in\mathbb{R}, c''_{d-1}\in\{\pm1\}\\ \widetilde{f}_\circ\in\cup_n\mathcal{F}_n^*(0,k),\tau^*(\widetilde{f}_\circ)\le 1/2}} \int_{\tau^*(\widetilde{f}_\circ)}^1 \bigg[\sum_{\ell=0}^{d-1}\frac{c''_\ell (x-\tau^*(\widetilde{f}_\circ))^\ell}{\ell!} + (I^d_{0,\ldots,0;\tau^*(\widetilde{f}_\circ)}\widetilde{f}_\circ)(x)\bigg]^2\,\mathrm{d}x,
\end{aligned}$$

where in the third line we use the fact that $\widetilde{f}_\circ = f_\circ/|c_{d-1}| \in \mathcal{F}_n^*(0,k)$ and satisfies $\tau^*(\widetilde{f}_\circ) = \tau^*(f_\circ) \le 1/2$. Thus, to prove the desired result, it suffices to show that there exists some $K = K(d) > 0$ such that

$$\inf_{\substack{\widetilde{c}_0,\ldots,\widetilde{c}_{d-2}\in\mathbb{R},\widetilde{c}_{d-1}\in\{\pm1\}\\ \widetilde{f}_\circ\in\cup_n\mathcal{F}_n^*(0,k),\tau^*(\widetilde{f}_\circ)\le 1/2,k\in\mathbb{Z}_+}} \int_{\tau^*(\widetilde{f}_\circ)}^1 \bigg[\sum_{\ell=0}^{d-1} \frac{\widetilde{c}_\ell(x-\tau^*(\widetilde{f}_\circ))^\ell}{\ell!} + (I^d_{0,\ldots,0;\tau^*(\widetilde{f}_\circ)}\widetilde{f}_\circ)(x)\bigg]^2 \,\mathrm{d}x \ge K. \tag{B.31}$$

Suppose this is not true, then there exist a function sequence $\{\widetilde{f}_{n,\circ}\}_n \subset \cup_{n',k'}\mathcal{F}^*_{n'}(0,k')$ with $\tau_n^* \equiv \tau_n^*(\widetilde{f}_{n,\circ}) \subset [0,1/2]$ and real sequences $\{\widetilde{c}_{n,\ell}\}_{n,\ell}$ with $\widetilde{c}_{n,d-1} \in \{\pm1\}$, such that

$$\int_0^1 \mathbf{1}_{[\tau_n^*,1]}(x)\Big[\sum_{\ell=0}^{d-1}\frac{\widetilde{c}_{n,\ell}}{\ell!}(x-\tau_n^*)^\ell + (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x)\Big]^2\,\mathrm{d}x \to 0.$$

Since L_2 convergence implies almost everywhere (a.e.) convergence, it follows that

$$\mathbf{1}_{[\tau_n^*,1]}(x)\cdot\Big[\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(x-\tau_n^*)^\ell/\ell! + (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x)\Big] \to 0, \quad \text{a.e. on } [0,1].$$

Since the sequence $\{\tau_n^*\} \subset [0,1/2]$ is bounded, $\tau_n^* \to \tau^*$ along some subsequence for some $\tau^* \in [0,1/2]$, and we work with this subsequence below. As $\mathbf{1}_{[\tau_n^*,1]}(x) \to 1$ for any fixed $x \in (\tau^*,1]$, the sequence of functions in the brackets in the above display converges a.e. to 0 on $(\tau^*,1]$. In other words,

$$\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(x-\tau_n^*)^\ell/\ell! + (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x) \to 0, \quad \text{a.e. on } (\tau^*,1]. \tag{B.32}$$

We first prove that under (B.32), $\{\widetilde{c}_{n,\ell}\}_n$ is necessarily bounded for each $\ell \in [0;d-1]$. Since $\{\widetilde{c}_{n,d-1}\}_n \subset \{-1,+1\}$ is already bounded, it suffices to prove the claim for $\ell \in [0;d-2]$. If this is not the case, then there exists some nonempty subset $\mathcal{L} \subset [0;d-2]$ such that for every $\ell \in \mathcal{L}$, $\{\widetilde{c}_{n,\ell}\}_n$ is divergent, i.e., $\limsup_n |\widetilde{c}_{n,\ell}| = +\infty$. As $\tau_n^* \to \tau^*$, we may find some slowly decaying $\varepsilon_n \downarrow 0$ such that (i) $\varepsilon_n > (\tau^* - \tau_n^*)_+$, (ii) $\{\widetilde{c}_{n,\ell}\varepsilon_n^\ell\}_n$ is still divergent for every $\ell \in \mathcal{L}$, and (B.32) holds with $x_n \equiv \tau_n^* + \varepsilon_n > \tau^*$. Now, by definition of $\widetilde{f}_{n,\circ}(\cdot)$, there exist some k_n, $j_n^* \in [1;k_n]$, $0 \equiv \tau_{n,0} \le \ldots \le \tau_{n,k_n} \equiv 1$, and non-negative sequence $\{\mu_{n,j}\}_{j=j_n^*}^{k_n-1}$ such that $\tau_n^* \equiv \tau_{n,j_n^*} \leq 1/2$ and for $x \in [\tau_n^*, 1]$, $\widetilde{f}_{n,\circ}(x) = \sum_{j=j_n^*}^{k_n-1} \mu_{n,j} \mathbf{1}_{x>\tau_{n,j}}$. Thus by a direct calculation, we have for $x \in [\tau_n^*, 1]$

$$(I^d_{0,\ldots,0;\tau_n^*} \widetilde{f}_{n,\circ})(x) = \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!} \big(x - \tau_n^*\big)_+^d.$$

So by (B.32) and definition of $\{x_n\}$,

$$\sum_{\ell=0}^{d-1} \frac{\widetilde{c}_{n,\ell}}{\ell!} \varepsilon_n^\ell + \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!} \varepsilon_n^d \to 0.$$

Let $\ell_0 \in \mathcal{L}$ be the index such that $\{\widetilde{c}_{n,\ell} \varepsilon_n^\ell\}_n$ has the fastest divergence rate, i.e., $\limsup_n |\widetilde{c}_{n,\ell_0}| \varepsilon_n^{\ell_0} / (|\widetilde{c}_{n,\ell}| \varepsilon_n^\ell) \geq \alpha$ for some positive α and every $\ell \in \mathcal{L}$. Without loss of generality, we further choose $\{\varepsilon_n\}$ such that the maximal divergence rate and the index that achieves this rate are unique, i.e., ℓ_0 is unique and satisfies $\limsup_n |\widetilde{c}_{n,\ell_0}| \varepsilon_n^{\ell_0} / (|\widetilde{c}_{n,\ell}| \varepsilon_n^\ell) = \infty$ for every $\ell \in \mathcal{L} \setminus \{\ell_0\}$. This then entails that

$$B_n \equiv \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!} \varepsilon_n^d \gtrsim |\widetilde{c}_{n,\ell_0}| \varepsilon_n^{\ell_0} \tag{B.33}$$

and is positive and divergent. Next, for the chosen sequence $\{\varepsilon_n\}$, choose $\{\eta_n\} \subset [1, \infty)$ as some slowly growing sequence such that (B.32) holds with the sequence $x_n' \equiv \tau_n^* + \varepsilon_n \eta_n \geq \tau_n^* + \varepsilon_n > \tau^*$, i.e.,

$$\sum_{\ell=0}^{d-1} \frac{\widetilde{c}_{n,\ell}}{\ell!} (\eta_n \varepsilon_n)^\ell + \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!} (\eta_n \varepsilon_n)^d \to 0, \tag{B.34}$$

and that $\{\varepsilon_n \eta_n\} \downarrow 0$ and $\{\widetilde{c}_{n,\ell_0} (\eta_n \varepsilon_n)^{\ell_0}\}$ remains to be the fastest divergent sequence among $\mathcal{L}$, i.e., $\limsup_n |\widetilde{c}_{n,\ell_0}| (\varepsilon_n \eta_n)^{\ell_0} / (|\widetilde{c}_{n,\ell}| (\varepsilon_n \eta_n)^\ell) = \infty$ for every $\ell \in \mathcal{L} \setminus \{\ell_0\}$. Similar to (B.33), we have $\sum_{j=j_n^*}^{k_n-1} \mu_{n,j} (\eta_n \varepsilon_n)^d / d! \gtrsim |\widetilde{c}_{n,\ell_0}| (\eta_n \varepsilon_n)^{\ell_0}$ and is positive and divergent. But this is impossible since

$$\begin{aligned}
\sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!} (\eta_n \varepsilon_n)^d &= (\eta_n)^d B_n \gtrsim (\eta_n)^{d-\ell_0} (\widetilde{c}_{n,\ell_0} (\varepsilon_n \eta_n)^{\ell_0}) \\
&\asymp (\eta_n)^{d-\ell_0} \bigg| \sum_{\ell=0}^{d-1} \widetilde{c}_{n,\ell} (\eta_n \varepsilon_n)^\ell / \ell! \bigg|,
\end{aligned}$$

where the first inequality is by (B.33) and the last relation is by the maximal divergence rate of $\{\widetilde{c}_{n,\ell_0}(\eta_n\varepsilon_n)^{\ell_0}\}$, and thus

$$\begin{aligned}
&\sum_{\ell=0}^{d-1}\frac{\widetilde{c}_{n,\ell}}{\ell!}(\eta_n\varepsilon_n)^\ell+\sum_{j=j^*}^{k_n-1}\frac{\mu_{n,j}}{d!}(\eta_n\varepsilon_n)^d\\
&\gtrsim\big[(\eta_n)^{d-\ell_0}-1\big]\bigg|\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(\eta_n\varepsilon_n)^\ell/\ell!\bigg|\geq\big[\eta_n-1\big]\bigg|\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(\eta_n\varepsilon_n)^\ell/\ell!\bigg|\to\infty,
\end{aligned}$$

a contradiction to (B.34). This concludes that $\{\widetilde{c}_{n,\ell}\}_n$ are necessarily bounded for every $\ell\in[0;d-1]$. Thus there exists a real sequence $\{c^*_\ell\}_{\ell=0}^{d-1}$ with $c^*_{d-1}\in\{\pm1\}$ such that $\widetilde{c}_{n,\ell}\to c^*_\ell$ along some subsequence for each $\ell\in[0;d-1]$. Coming back to (B.32) and noting that $\tau^*_n\to\tau^*$ along some subsequence, we then conclude that

$$h_n(x)\equiv(I^d_{0,\ldots,0;\tau^*_n}\widetilde{f}_{n,\circ})(x)\to\sum_{\ell=0}^{d-1}\frac{-c^*_\ell}{\ell!}(x-\tau^*)^\ell\equiv h^*(x)\tag{B.35}$$

a.e. on $(\tau^*,1]$ as $n\to\infty$. We will now prove that $\{c^*_\ell\}_{\ell=0}^{d-1}$ are necessarily non-positive. Fix some positive integer $m>d$ and define a regular grid on $(\tau^*,1]$: $t_i\equiv\tau^*+i(1-\tau^*)/m$ for $i\in[0;m]$. Without loss of generality, assume that $\{t_i\}_{i=1}^m$ belongs to the set with full Lebesgue measure such that (B.35) holds. Define $(\xi_{n,i})_{i=1}^m\equiv(h_n(t_i))_{i=1}^m$ (resp. $(\xi^*_i)_{i=1}^m\equiv(h^*(t_i))_{i=1}^m$) to be the realization of $h_n(\cdot)$ (resp. $h^*(\cdot)$) on this grid. Define ∇ to be the finite difference operator that maps $(y_1,\ldots,y_m)^\top\in\mathbb{R}^m$ to $(y_2-y_1,\ldots,y_m-y_{m-1})^\top\in\mathbb{R}^{m-1}$. Then, since $\lim_n\min_{\ell\in[0;d]}h_n^{(\ell)}(x)\geq0$ for $x\in(\tau^*,1]$, it holds that for each fixed $m\geq d+1$, $\nabla^\ell\xi_n\in\mathbb{R}^{m-\ell}_{\geq0}$ holds for all $\ell\in[0;d]$ for n large enough. On the other hand, for each $\ell\in[0;d-1]$ and $p\in[\ell;d-1]$, there exists some positive constant $L_{p,\ell}>0$ for such that

$$\big(\nabla^\ell\xi^*\big)_1=\Bigg(\nabla^\ell\bigg(\sum_{p=0}^{d-1}\frac{-c^*_p}{p!}(t_j-\tau^*)^p\bigg)_{j=1}^m\Bigg)_1=\sum_{p=\ell}^{d-1}-c^*_pL_{p,\ell}((1-\tau^*)/m)^p.$$

Since for each fixed $m\geq d+1$, $\nabla^\ell\xi_n\to\nabla^\ell\xi^*$ as $n\to\infty$ by (B.35) and $\nabla^\ell\xi_n\in\mathbb{R}^{m-\ell}_{\geq0}$ for n large enough, it holds that $\big(\nabla^\ell\xi^*\big)_1\geq0$ for each fixed $m\geq d+1$. Multiplying by m^ℓ on both sides of the above equation and letting $m \to \infty$ we conclude that $c_\ell^* \le 0$ for $\ell \in [0; d-2]$ and $c_{d-1}^* = -1$.

With $\{c_\ell^*\}_{\ell=0}^{d-1} \in \mathbb{R}_{\le 0}^d$, h_n, h^* have the property that their derivatives up to order $d-1$ are all convex functions, so on arbitrary compact interval contained in $(\tau^*, 1)$, $D^{(d-1)}h_n$ converges uniformly to $D^{(d-1)}h^* \equiv 1$ (cf. Theorem 25.7 of Rockafellar (1997) and the remark after its proof). This cannot happen as $D^{(d-1)}h_n(\tau_n^*) = 0$, $\tau_n^* \to \tau^*$ and $D^{(d-1)}h_n$ is convex. We have therefore established the contradiction and proved (B.31). ☐

Proof of Lemma 36. By Lemma 33, any $f \in \mathcal{F}_n^*(d,k)$ has the canonical parametrization

$$f(x) = \sum_{j=1}^{j^*} a_j(\tau_j - x)_+^d + \sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell,$$

where $\{\tau_j\}_{j=1}^{k-1} \equiv \{n_j/n\}_{j=1}^{k-1} \subset [0,1]$. Let $\tau^* \equiv \tau_{j^*}$. Then, it holds that $\int_0^1 f^2(x)\ \mathrm{d}x = (I) + (II)$, where

$$(I) \equiv \int_0^{\tau^*} \Big(\sum_{j=1}^{j^*} a_j(\tau_j - x)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x,$$

$$(II) \equiv \int_{\tau^*}^1 \Big(\sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x.$$

We now upper bound (II) by its sequence counterpart; the bound for (I) is similar. Since

$$\begin{aligned}(II) &= \sum_{m=j^*}^{k-1} \int_{\tau_m}^{\tau_{m+1}} \Big(\sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x \\ &= \sum_{m=j^*}^{k-1} \int_{\tau_m}^{\tau_{m+1}} \Big(\sum_{j=j^*}^{m} b_j(x-\tau_j)^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x,\end{aligned}$$

we may bound the integral piece by piece. More generally, we show that there exists some $K = K(d) > 0$ such that for any $a, b \in [0; n]$ with $b - a \ge d+1$ and d-degree polynomial $P(x) \equiv \sum_{\ell=0}^d c_\ell x^\ell$,

$$\int_{a/n}^{b/n} P^2(x)\ \mathrm{d}x \le K \cdot n^{-1} \sum_{i \in (a;b]} P^2(i/n). \tag{B.36}$$

The above display holds because

$$\int_{a/n}^{b/n} P^2(x)\ \mathrm{d}x = \int_{a/n}^{b/n} \Big(\sum_{\ell=0}^{d} c_\ell x^\ell\Big)^2\ \mathrm{d}x \lesssim_d \sum_{\ell=0}^{d} c_\ell^2 \cdot \int_{a/n}^{b/n} x^{2\ell}\ \mathrm{d}x$$
$$\leq \sum_{\ell=0}^{d} \frac{c_\ell^2}{n} \sum_{i\in(a;b]} \left(\frac{i}{n}\right)^{2\ell} = \frac{1}{n} \sum_{i\in(a;b]} \sum_{\ell=0}^{d} \left(c_\ell \left(\frac{i}{n}\right)^{\ell}\right)^2 \lesssim_d \frac{1}{n} \sum_{i\in(a;b]} \left(\sum_{\ell=0}^{d} c_\ell \left(\frac{i}{n}\right)^{\ell}\right)^2,$$

where the last inequality is due to Lemma 37 and the condition $b-a \geq (d+1)$. Then for every $\theta \in \Theta(d, d_0, k)$ with unit norm constraint and the corresponding $f \in \mathcal{F}_n(d, d_0, k)$, by (B.36) we have

$$1 \geq \|\theta\|^2 \geq \|\theta\|^2_{(n_{j^*};n]} = \sum_{m=j^*}^{k-1} \|\theta\|^2_{(n_m;n_{m+1}]} = \sum_{m=j^*}^{k-1} \sum_{i\in(n_m;n_{m+1}]} f^2(i/n)$$
$$\gtrsim n \sum_{m=j^*}^{k-1} \int_{\tau_m}^{\tau_{m+1}} f^2(x)\ \mathrm{d}x = n \int_{\tau^*}^{1} f^2(x)\ \mathrm{d}x.$$

The bound for (II) is thus complete. $\square$

B.7 Auxiliary lemmas

Lemma 37. *Fix any positive integer d. There exists some $c = c(d)$ such that for any integers $n \geq 0$, $m \geq d+1$, and real sequence $\{a_\ell\}_{\ell=1}^{d+1}$,*

$$\sum_{i=1}^{m} \left[a_1 + a_2\left(\frac{i}{n}\right) + \ldots + a_{d+1}\left(\frac{i}{n}\right)^d\right]^2 \geq c \sum_{\ell=1}^{d+1} a_\ell^2 \frac{m^{2\ell-1}}{n^{2(\ell-1)}}.$$

Proof of Lemma 37. As the left hand side of the above inequality equals

$$\sum_{i=1}^{n} \left(\sum_{\ell=1}^{d+1} a_\ell (i/n)^{\ell-1}\right)^2 = \sum_{1\leq \ell,\ell'\leq d+1} a_\ell a_{\ell'} \sum_{i=1}^{m} (i/n)^{\ell+\ell'-2}$$
$$= \sum_{1\leq \ell,\ell'\leq d+1} a_\ell (m/n)^{\ell-1} m^{1/2} \cdot a_{\ell'} (m/n)^{\ell'-1} m^{1/2} \cdot \left[m^{-(\ell+\ell'-1)} \sum_{i=1}^{m} i^{\ell+\ell'-2}\right],$$

using matrix notation, it can be written as $x^\top A x$, where $x \equiv (a_\ell (m/n)^{\ell-1} m^{1/2})_{\ell=1}^{d+1} \in \mathbb{R}^{d+1}$, and the matrix $(A)_{ij} \equiv (A(m,d))_{ij} \equiv (m^{-(i+j-1)} \sum_{k=1}^{m} k^{i+j-2})_{ij} \in \mathbb{R}^{(d+1)\times(d+1)}$.

We first show that A is strictly positive-definite for the fixed d and any $m \geq d+1$. Note that A is actually a moment matrix and can be written as $A_{ij} = \mathbb{E}(X^{i-1} \cdot X^{j-1})$, where X is uniformly distributed on the set $\{1/m, \ldots, m/m\}$. Therefore, for any $c \in \mathbb{S}^d$, writing, with a slight abuse of notation, $Z \equiv \sum_{i=1}^{d+1} c_i X^{i-1}$, it holds that

$$\begin{aligned} c^\top Ac &= \sum_{1 \leq i,j \leq d+1} c_i c_j A_{ij} = \sum_{1 \leq i,j \leq d+1} c_i c_j \mathbb{E}(X^{i-1} \cdot X^{j-1}) = \mathbb{E}\biggl(\sum_{i=1}^{d+1} c_i X^{i-1}\biggr)^2 \\ &= \mathbb{E}Z^2 = (\mathbb{E}Z)^2 + \mathrm{Var}(Z). \end{aligned}$$

If $\mathrm{Var}(Z) = 0$, then $Z \equiv \alpha$ almost surely for some constant α, which is equivalent to that the polynomial

$$T(x) \equiv (c_0 - \alpha) + c_1 x + \ldots + c_{d+1} x^d$$

having distinct roots $\{1/m, \ldots, m/m\}$. If $c_1 = \ldots = c_{d+1} = 0$, then $c_0 = \pm 1$ since $\|c\| = 1$, which implies that $Z = \pm 1$, and thus $c^\top Ac \geq (\mathbb{E}Z)^2 = 1$. Otherwise, we have $c_i \neq 0$ for some $i \in [1; d]$, and hence $T(x)$ is not a constant and thus has at most d roots, which contradicts the condition that $m \geq d+1$. So we conclude that $c^\top Ac > 0$ for any $c \in \mathbb{S}^d$ and thus A is strictly positive-definite.

Next, we show that for any $i \in [1; d+1]$, the $(-i, -i)$-minor of A (i.e. A minus the ith row and column) is also strictly positive-definite. For this, define Q_i as the permutation matrix that switches row i with row $i+1$, and define $P_i \equiv Q_i Q_{i+1} \ldots Q_d$ for $i \leq d$ and $P_{d+1} \equiv I_{d+1}$, the $(d+1)$-dimensional identity matrix. Further define $B \equiv P_i^\top A P_i$. Then, the $(-i, -i)$-minor of A is the $(-(d+1), -(d+1))$-minor of B. By Sylvester's criterion, it suffices to show that B is strictly positive-definite, but for any $c \in \mathbb{S}^d$, it holds that

$$c^\top Bc = c^\top P_i^\top A P_i c \equiv \widetilde{c}^\top A \widetilde{c} > 0,$$

where in the last inequality we have used the fact that

$$\widetilde{c}^\top \widetilde{c} = c^\top P_i^\top P_i c = c^\top Q_d \ldots Q_i Q_i \ldots Q_d c = c^\top c = 1.$$

Next, we show that $x^\top A x \geq c a_d^2 m^{2d+1}/n^{2d}$ for some $c = c(d)$; bounds involving $a_0, \ldots, a_{d-1}$ can be similarly obtained. For this, write A in the block form $[A_{11}, A_{12}; A_{21}, A_{22}]$, where $A_{12} \in \mathbb{R}^{d\times 1}$. Writing y as the first d components of x, i.e. $y \equiv (a_0 m^{1/2}, a_1 m^{3/2}/n, \ldots, a_{d-1} m^{(2d-1)/2}/n^{d-1})^\top$, we have

$$\begin{aligned} x^\top A x &= (y, a_d m^{(2d+1)/2}/n^d)^\top \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} (y, a_d m^{(2d+1)/2}/n^d) \\ &= y^\top A_{11} y + 2y^\top A_{21} a_d m^{(2d+1)/2}/n^d + A_{22}(a_d m^{(2d+1)/2}/n^d)^2. \end{aligned}$$

This is a quadratic form in y and achieves its minimum at $y^* = -A_{11}^{-1} A_{12} a_d m^{(2d+1)/2}/n^d$ (note that A_{11}, the $(-(d+1), -(d+1))$-minor of A, is indeed invertible as proved before), which implies that

$$x^\top A x \geq a_d^2 \frac{m^{2d+1}}{n^{2d}} (A_{22} - A_{21} A_{11}^{-1} A_{12}).$$

Therefore if we can show that $A_{22} \geq (1+\varepsilon) A_{21} A_{11}^{-1} A_{12}$ for some positive $\varepsilon = \varepsilon(d)$, then we have

$$\begin{aligned} x^\top A x &\geq \frac{\varepsilon}{1+\varepsilon} a_d^2 \frac{m^{2d+1}}{n^{2d}} A_{22} = \frac{\varepsilon}{1+\varepsilon} a_d^2 \frac{m^{2d+1}}{n^{2d}} m^{-(2d+1)} \sum_{k=1}^m k^{2d} \\ &\geq \frac{\varepsilon}{1+\varepsilon} a_d^2 \frac{m^{2d+1}}{n^{2d}} m^{-(2d+1)} \int_0^m x^{2d}\, \mathrm{d}x = \frac{\varepsilon}{(2d+1)(1+\varepsilon)} a_d^2 \frac{m^{2d+1}}{n^{2d}}. \end{aligned}$$

Using the block matrix inverse formula $(A^{-1})_{d+1,d+1} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$ and the fact that $(A^{-1})_{d+1,d+1} \leq \|A^{-1}\|_2 = \lambda_{\min}^{-1}(A)$ ($\lambda_{\min}$ takes the smallest eigenvalue), we have

$$\begin{aligned} & A_{22} \geq (1+\varepsilon) A_{21} A_{11}^{-1} A_{12} \iff (1+\varepsilon)(A_{22} - A_{21} A_{11}^{-1} A_{12}) \geq \varepsilon A_{22} \\ \iff & (A^{-1})_{d+1,d+1} \leq \frac{1+\varepsilon}{\varepsilon} A_{22}^{-1} \Leftarrow \lambda_{\min}^{-1}(A) \leq \frac{1+\varepsilon}{\varepsilon} \min_{1\leq j\leq d+1} A_{jj}^{-1}, \end{aligned}$$

which is further implied by

$$\lambda_{\min}(A) \geq \frac{\varepsilon}{1+\varepsilon} \max_{1\leq j\leq d+1} A_{jj}. \tag{B.37}$$

For this, we have, for every $j \in [1; d+1]$,

$$\begin{aligned} A_{jj} &= m^{-(2j-1)} \sum_{k=1}^{m} k^{2j-2} \leq m^{-(2j-1)} \int_{1}^{m+1} x^{2j-2}\,\mathrm{d}x \\ &\leq \frac{1}{2j-1}\left(1+\frac{1}{m}\right)^{2j-1} \leq 2^{2d+1}. \end{aligned}$$

It remains to show that there exists some sufficiently small $c^* = c^*(d)$ such that $\lambda_{\min}(A) \geq c^* > 0$, then we can take $\varepsilon = c^*/(2^{2d+1} - c^*)$ in (B.37). For this, let U be a random variable uniformly distributed on $[0,1]$ and define matrix $\overline{A}$ as $\overline{A}_{i,j} \equiv \mathbb{E}(U^{i-1} \cdot U^{j-1})$. Then, since d is fixed, it holds by the definition of $A, \overline{A}$, and the Portmanteau theorem that $A \to \overline{A}$ in the matrix spectral norm as $m \to \infty$. By Weyl's inequality, there exists some positive integer $N = N(d)$ such that for $m \geq N$, $\lambda_{\min}(A) \geq \lambda_{\min}(\overline{A})/2$. On the other hand, a similar argument that establishes the positive definiteness of A yields that $\lambda_{\min}(\overline{A}) \geq c > 0$ for some $c = c(d)$. Therefore we can take $c^* = c^*(d) = \min_{d+1 \leq m \leq N} \lambda_{\min}(A(m,d)) \wedge (c/2)$. This completes the proof. □

Lemma 38. *Let $\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m$ be two non-negative sequences. Then, it holds that $(\bigwedge_{i=1}^m a_i) \cdot (\bigvee_{i=1}^m b_i) \geq \bigwedge_{i=1}^m a_i b_i$.*

Proof of Lemma 38. Without loss of generality, let a_1 be the smallest value among $\{a_i\}_{i=1}^m$. Then, it holds that $\left(\bigwedge_{i=1}^m a_i\right) \cdot \left(\bigvee_{i=1}^m b_i\right) = a_1 \cdot \left(\bigvee_{i=1}^m b_i\right) \geq a_1 b_1 \geq \left(\bigwedge_{i=1}^m a_i b_i\right)$. □

Lemma 39. *Let $\alpha_1, \alpha_2 > 0$ and β_1, β_2 be real numbers. Then, for any $x \in \mathbb{R}$, it holds that*

$$\alpha_1(x+\beta_1)^2 + \alpha_2(x+\beta_2)^2 \geq (\alpha_1 \wedge \alpha_2)(\beta_1 - \beta_2)^2/2.$$

Proof of Lemma 39. At $x^* \equiv -(\alpha_1/(\alpha_1+\alpha_2) \cdot \beta_1 + \alpha_2/(\alpha_1+\alpha_2) \cdot \beta_2)$, the quadratic form achieves it minimum value $\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}(\beta_1 - \beta_1)^2$, which is further lower bounded by $(\alpha_1 \wedge \alpha_2)(\beta_1 - \beta_2)^2/2$. □

Lemma 40. *Let n be any positive integer. Then, for any polynomial $P(\cdot)$ of degree strictly smaller than n, it holds that*

$$\sum_{j=0}^{n} \binom{n}{j} P(j)(-1)^j = 0.$$

Proof of Lemma 40. We prove by induction. The claim clearly holds for $n = 1$. Suppose the claim holds for some n, we will prove that it also holds for $n+1$. Let d be the degree of $P(\cdot)$. We will prove that the claim holds for all monomials $P(x) \equiv x^d$ where $0 \leq d \leq n = (n+1)-1$. The case $d = 0$ follows from the binomial identity:

$$\sum_{j=0}^{n+1} \binom{n+1}{j} (-1)^j = (1 + (-1))^{n+1} = 0.$$

Next, for any $1 \leq d \leq n$, it holds that

$$\begin{aligned}\sum_{j=0}^{n+1} \binom{n+1}{j} j^d(-1)^j &= \sum_{j=1}^{n+1} \binom{n+1}{j} j^d(-1)^j = (n+1) \sum_{j=1}^{n+1} \binom{n}{j-1} j^{d-1}(-1)^j \\ &= (n+1) \sum_{j=0}^{n} \binom{n}{j} (j+1)^{d-1}(-1)^j = 0,\end{aligned}$$

where the last identity follows from the claim for n and the fact that $0 \leq d-1 \leq n-1 < n$. $\square$

For the following lemma, recall the definition of the sequence $\{\overline{\beta}^{\cdot}_{\cdot,\cdot}\}$ defined before Lemma 30.

Lemma 41. *Fix d, d_0, k_0 as defined in (3.12), and any $s \in [0; \lfloor (d_0+1)/(d-d_0) \rfloor - 1]$. Suppose there exists some $c_1 = c_1(d)$ such that*

$$1 \geq c_1 \cdot \sum_{k=1}^{(s+1)d_0 - sd+1} \frac{(n - n_{k_0-1})^{2k-1}}{n^{2(k-1)}} \left(\sum_{\ell=0}^{s(d-d_0)} \overline{\beta}^{s}_{k,\ell} a^{k_0-1-s}_{k+\ell} \right)^2. \tag{B.38}$$

Furthermore, assume that $n_{k_0;k_0-1} \geq n_{k_0-1-s;k_0-2-s}$. Then, there exists some positive constant $c_2 = c_2(d)$ such that

$$1 \geq c_2 \cdot \sum_{k=1}^{(s+1)d_0 - sd+1} \frac{(n - n_{k_0-1})^{2k-1}}{n^{2(k-1)}} \left(\sum_{\ell=0}^{(s+1)(d-d_0)} \overline{\beta}^{s+1}_{k,\ell} a^{k_0-2-s}_{k+\ell} \right)^2.$$

Note that in the above lemma the hypothesis involves only quadratic forms with 'shared coefficients' $\{a_\ell^\cdot\}_{\ell\in[1;d_0+1]}$, while the conclusion involves the ones with both 'shared coefficients' $\{a_\ell^\cdot\}_{\ell\in[1;d_0+1]}$ and 'nuisance coefficients' $\{a_\ell^\cdot\}_{\ell\in[d_0+2;d+1]}$.

Before the proof of Lemma 41, we need one further result. For this, some extra notation is needed:

$$\begin{aligned}
\overline{v}_{i,j}^s &\equiv \frac{\overline{\odot}(i+d-d_0-j); s(d-d_0))}{j!\overline{\odot}(i+d-d_0; s(d-d_0))}(-1)^j n_{k_0-s;k_0-1-s}^j \\
&\qquad \times \prod_{m=1}^{j}\Big(d-\big(i+(d-1-d_0)\big)-s(d-d_0)+m\Big), \\
T_k &\equiv \sum_{\ell=0}^{s(d-d_0)} \overline{\beta}_{k,\ell}^s a_{k+\ell}^{k_0-1-s}.
\end{aligned}$$

Lemma 42. *Fix d, d_0, and s. It holds for $i\in[1;(s+1)d_0-sd+1]$ that*

$$M \equiv \sum_{k=i}^{(s+1)d_0-sd+1} \overline{v}_{\overline{i},k-i}^{s+1}\cdot T_k = \sum_{k=0}^{(d-d_0)(s+1)} \overline{\beta}_{i,k}^{s+1} a_{i+k}^{k_0-2-s}.$$

Proof. In order to prove the desired result, we need to show the following two claims:

- The coefficient of $a_{i+j}^{k_0-2-s}$ in M equals 0 for $(s+1)(d-d_0)+1\le j\le d-i+1$;
- The coefficient of $a_{i+j}^{k_0-2-s}$ in M equals $\overline{\beta}_{i,j}^{s+1}$ for $0\le j\le (s+1)(d-d_0)$.

Let

$$\begin{aligned}
i_0 &\equiv i_0(d,d_0,s,i) \equiv (s+1)d_0-sd+1-i, \\
\overline{i} &\equiv \overline{i}(d,d_0,s,i) \equiv (s+2)d_0-(s+1)d+2-i = i_0-(d-1-d_0), \\
\Delta n &\equiv n_{k_0-1-s;k_0-2-s}.
\end{aligned}$$

By definition of M and Lemma 29, we have

$$
\begin{aligned}
\mathrm{Coef}[M; a_{i+j}^{k_0-2-s}] &= \sum_{k=i}^{(s+1)d_0-sd+1} \overline{v}^{s+1}_{\bar{i},k-i}\mathrm{Coef}[T_k; a_{i+j}^{k_0-2-s}] \\
&= \sum_{k=i}^{(s+1)d_0-sd+1} \frac{\overline{\odot}(\bar{i}+d-d_0-(k-i)); (s+1)(d-d_0))(-1)^{k-i}(\Delta n)^{k-i}}{(k-i)!\overline{\odot}(\bar{i}+d-d_0; (s+1)(d-d_0))} \\
&\qquad \times \prod_{m=1}^{k-i}(d-i_0-(s+1)(d-d_0)+m)\cdot \mathrm{Coef}\left[\sum_{\ell=0}^{s(d-d_0)} \overline{\beta}^s_{k,\ell} a^{k_0-1-s}_{k+\ell}; a^{k_0-2-s}_{i+j}\right] \\
&= \sum_{k=0}^{i_0} \frac{\overline{\odot}(i_0+1-k; (s+1)(d-d_0))(-1)^k(\Delta n)^k}{k!\overline{\odot}(i_0+1; (s+1)(d-d_0))}\overline{\odot}(i;k) \\
&\qquad \times \left(\sum_{\ell=0}^{s(d-d_0)} \overline{\beta}^s_{i+k,\ell}\binom{i+j-1}{i+k+\ell-1}(\Delta n)^{j-k-\ell}\right) \\
&\equiv \sum_{\ell=0}^{s(d-d_0)} (\Delta n)^{j-\ell}\overline{\beta}^s_{\ell}\cdot A_\ell,
\end{aligned}
$$

where

$$
\begin{aligned}
A_\ell \equiv \sum_{k=0}^{i_0}(-1)^k \frac{\overline{\odot}(i_0+1-k; (s+1)(d-d_0))}{k!\overline{\odot}(i_0+1; (s+1)(d-d_0))}\overline{\odot}(i;k) \\
\times \binom{i+j-1}{i+k+\ell-1}\frac{\overline{\odot}(i+k;\ell)}{\underline{\odot}(d+1-i-k;\ell)},
\end{aligned}
$$

and we used $\overline{\beta}^s_{i+k,\ell} = D(i+k,\ell)\overline{\beta}^s_\ell = \frac{\overline{\odot}(i+k;\ell)}{\underline{\odot}(d+1-i-k;\ell)}\overline{\beta}^s_\ell$, with $\overline{\beta}^s_\ell$ defined in (B.10). Let $C(i,j,\ell) \equiv \binom{i+j-1}{i+\ell-1}\cdot\overline{\odot}(i;\ell)$. Then $C(i,j,\ell)\binom{j-\ell}{k} = \overline{\odot}(i;k)\binom{i+j-1}{i+k+\ell-1}\overline{\odot}(i+k;\ell)/k!$. So A_ℓ equals

$$
\begin{aligned}
& C(i,j,\ell)\sum_{k=0}^{i_0}\binom{j-\ell}{k}\frac{\overline{\odot}(i_0+1-k; (s+1)(d-d_0))}{\overline{\odot}(i_0+1; (s+1)(d-d_0))}(-1)^k\frac{1}{\underline{\odot}(d-i-k+1;\ell)} \\
&= C(i,j,\ell)\sum_{k=0}^{i_0+(s+1)(d-d_0)}\binom{j-\ell}{k}\frac{\overline{\odot}(i_0+1-k; (s+1)(d-d_0))}{\overline{\odot}(i_0+1; (s+1)(d-d_0))}(-1)^k\frac{1}{\underline{\odot}(d-i-k+1;\ell)} \\
&= C(i,j,\ell)\sum_{k=0}^{j-\ell}\binom{j-\ell}{k}\frac{\overline{\odot}(i_0+1-k; (s+1)(d-d_0))}{\overline{\odot}(i_0+1; (s+1)(d-d_0))}(-1)^k\frac{1}{\underline{\odot}(d-i-k+1;\ell)} \\
&= \frac{C(i,j,\ell)}{\overline{\odot}(i_0+1; (s+1)(d-d_0))}\sum_{k=0}^{j-\ell}\binom{j-\ell}{k}(-1)^k\underline{\odot}(d-i-k+1-\ell; (s+1)(d-d_0)-\ell),
\end{aligned}
$$

where the first identity follows from the fact that $\overline{\odot}(i_0+1-k;(s+1)(d-d_0)) = 0$ for any $i_0+1 \le k \le i_0+(s+1)(d-d_0)$, the second identity follows from the fact that $i_0+(s+1)(d-d_0) = d-i+1 \ge j \ge j-\ell$, the third identity follows from the fact that $\ell \le s(d-d_0) < (s+1)(d-d_0)$.

For the first claim, as $\underline{\odot}(d-i-k+1-\ell;(s+1)(d-d_0)-\ell)$ is a polynomial of degree at most $(s+1)(d-d_0)-\ell < j-\ell$, Lemma 40 entails that $A_\ell = 0$ for all $0 \le \ell \le s(d-d_0)$, thus proving the first claim. We now prove the second claim under the condition $j \le (s+1)(d-d_0)$. By definition of the $\{\overline{\beta}^{\cdot}_{\cdot,\cdot}\}$ sequence, we have

$$\overline{\beta}^{s+1}_{i,j} = D(i,j)\overline{\beta}^{s+1}_{j} = D(i,j)\left\{\sum_{\ell=0}^{j}\binom{(s+1)(d-d_0)-\ell}{j-\ell}(\Delta n)^{j-\ell}\overline{\beta}^{s}_{\ell}\right\}.$$

Therefore, to prove the claim, it suffices to match the coefficients of $\overline{\beta}^{s}_{\ell}$ for $0 \le \ell \le s(d-d_0)$, as $\overline{\beta}^{s}_{\ell} = 0$ for $\ell > s(d-d_0)$ from the definition of $\overline{\beta}^{s}_{\cdot}$, and $A_\ell = 0$ for $\ell \ge j$. In other words, we only need to show $A_\ell = D(i,j)\binom{(s+1)(d-d_0)-\ell}{j-\ell}$. By using iteratively the identity $\binom{n}{k} = \binom{n}{k-1} + \binom{n-1}{k-1}$, one has

$$\begin{aligned}
&\sum_{k=0}^{j-\ell}\binom{j-\ell}{k}(-1)^k\underline{\odot}(d-i-k+1-\ell;(s+1)(d-d_0)-\ell)\\
&= \underline{\odot}((s+1)(d-d_0)-\ell;1)\\
&\qquad\times\sum_{k=0}^{j-\ell-1}\binom{j-\ell-1}{k}(-1)^k\underline{\odot}(d-i-k-\ell;(s+1)(d-d_0)-1-\ell)\\
&\qquad\dots\\
&= \underline{\odot}((s+1)(d-d_0)-\ell;j-\ell-1)\\
&\qquad\times\sum_{k=0}^{1}\binom{1}{k}(-1)^k\underline{\odot}(d-i-k+2-j;(s+1)(d-d_0)+1-j)\\
&= \underline{\odot}((s+1)(d-d_0)-\ell;j-\ell)\underline{\odot}(d-i+1-j;(s+1)(d-d_0)-j).
\end{aligned}$$

Lastly, by direct calculation, we have

$$\begin{aligned}A_\ell &= \frac{C(i,j,\ell)}{\overline{\odot}(i_0+1;(s+1)(d-d_0))}\cdot\underline{\odot}((s+1)(d-d_0)-\ell;j-\ell)\\&\qquad\times\underline{\odot}(d-i+1-j;(s+1)(d-d_0)-j)\\&= D(i,j)\binom{(s+1)(d-d_0)-\ell}{j-\ell}.\end{aligned}$$

The proof is complete. □

Proof of Lemma 41. Define for $i \in [1;(s+1)d_0-sd+1]$

$$M_i \equiv \sum_{k=i}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}}T_k^2.$$

Inequality (B.38) entails that $1 \gtrsim c\sum_{i=1}^{(s+1)d_0-sd+1} M_i$ for some $c = c(d)$. We have for $i \in [1;(s+1)d_0-sd+1]$,

$$\begin{aligned}M_i &= \frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}T_i^2+\sum_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}}\frac{\left(\overline{v}^{s+1}_{\bar{i},k-i}\cdot T_k\right)^2}{(\overline{v}^{s+1}_{\bar{i},k-i})^2}\\&\geq\left(\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\right)\left(T_i^2+\sum_{k=i+1}^{(s+1)d_0-sd+1}(\overline{v}^{s+1}_{\bar{i},k-i}\cdot T_k)^2\right)\\&\gtrsim\left(\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\right)\left(T_i+\sum_{k=i+1}^{(s+1)d_0-sd+1}\overline{v}^{s+1}_{\bar{i},k-i}\cdot T_k\right)^2\\&=\left(\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\right)\left(\sum_{k=0}^{(d-d_0)(s+1)}\overline{\beta}^{s+1}_{i,k}a^{k_0-2-s}_{i+k}\right)^2\\&\gtrsim\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\cdot\left(\sum_{k=0}^{(d-d_0)(s+1)}\overline{\beta}^{s+1}_{i,k}a^{k_0-2-s}_{i+k}\right)^2.\end{aligned}$$

Here, the second identity follows from Lemma 42, and the last inequality follows, by definition of $\{\overline{v}_{\cdot,\cdot}^{\cdot}\}$ and the condition $n_{k_0;k_0-1} \geq n_{k_0-1-s;k_0-2-s}$, from the calculation:

$$\begin{aligned}&\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}} \wedge \bigwedge_{k=i+1}^{(s+1)d_0-sd+1} \frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}_{\bar{i},k-i}^{s+1})^2}\\ &\asymp \bigwedge_{k=i}^{(s+1)d_0-sd+1} \frac{(n-n_{k_0-1})^{2k-1}(n_{k_0-1-s}-n_{k_0-2-s})^{-2(k-i)}}{n^{2(k-1)}} \asymp \frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}.\end{aligned}$$

Putting together the lower bounds for M_i, $i \in [1;(s+1)d_0 - sd + 1]$ yields the result. □

Lemma 43. *Fix any $1 \leq s \leq \lfloor (d_0+1)/(d-d_0) \rfloor$ and $1 \leq i \leq sd_0 - (s-1)d + 1$. For any $0 \leq j_1 \leq j_2 \leq s(d-d_0)$, define the following two quantities:*

$$\underline{S}(j_1) \equiv \underline{S}(j_1; d, d_0, s) \equiv \prod_{\ell=1}^{\lfloor j_1/(d-d_0) \rfloor} n_{k_0-\ell;k_0-1-s}^{d-d_0} \times n_{k_0-1-\lfloor j_1/(d-d_0)\rfloor;k_0-1-s}^{Mod(j_1;d-d_0)},$$

$$\overline{S}(j_2) \equiv \overline{S}(j_2; d, d_0, s) \equiv \prod_{\ell=-\lfloor -j_2/(d-d_0)\rfloor+1}^{s} n_{k_0-\ell;k_0-1-s}^{d-d_0} \times n_{k_0-(-\lfloor -j_2/(d-d_0)\rfloor);k_0-1-s}^{Mod(-j_2;d-d_0)}.$$

Then, there exists some positive constant $c = c(d)$ such that

$$\frac{\overline{\beta}_{i,j_2}^{s}}{\overline{\beta}_{i,j_1}^{s}} \geq c \frac{\prod_{\ell=1}^{s} n_{k_0-\ell;k_0-1-s}^{d-d_0}}{\underline{S}(j_1)\overline{S}(j_2)}.$$

When $j_1 = j_2$, the product on the right hand side is to be understood as 1.

Proof. We only prove the special case $d_0 = d-1$ (the proof for the general case is completely analogous). Then $k_0 = d+2$, and

$$\underline{S}(j_1) = \prod_{\ell=1}^{j_1} n_{d+2-\ell;d+1-s}, \quad \overline{S}(j_2) = \prod_{\ell=j_2+1}^{s} n_{d+2-\ell;d+1-s},$$

so we only need to prove for $s \in [1;d]$, $i \in [1;d+1-s]$, and $0 \leq j_1 \leq j_2 \leq s$,

$$\frac{\overline{\beta}_{i,j_2}^{s}}{\overline{\beta}_{i,j_1}^{s}} \geq c \prod_{k=j_1+1}^{j_2} n_{d+2-k;d+1-s}.$$

We prove this by induction on s.

First consider $s = 1$. Then $\overline{\beta}_j^1 = n_{d+1;d}^j$, and $\overline{\beta}_{i,j}^1 = D(i,j)\overline{\beta}_j^1 \asymp n_{d+1;d}^j$. The only non-trivial case is $j_1 = 0, j_2 = 1$, so the claim follows.

Suppose the claim holds up to $s-1$. Fix any $1 \leq j_1 \leq j_2 \leq s$. The claim clearly holds for $j_1 = j_2 = s$. If $j_2 = s$ and $j_1 \leq s-1$, then it holds by the recursion formula of $\{\overline{\beta}_j^s\}_{j=0}^s$ in (B.10) that $\overline{\beta}_{i,s}^s/\overline{\beta}_{i,j_1}^s \asymp n_{d+2-s;d+1-s}\overline{\beta}_{s-1}^s/\overline{\beta}_{j_1}^s$, and we can reduce to the following case with $1 \leq j_1 \leq j_2 \leq s-1$. For this case, note that

$$\begin{aligned}
&\prod_{k=j_1+1}^{j_2} n_{d+2-k;d+1-s} = \prod_{k=j_1+1}^{j_2} \left(n_{d+2-k;d+2-s} + n_{d+2-s;d+1-s}\right)\\
&= \sum_{k=0}^{j_2-j_1} n_{d+2-s;d+1-s}^k \sum_{j_1+1 \leq m_1 \neq \ldots \neq m_{j_2-j_1-k} \leq j_2} n_{d+2-m_1;d+2-s} \ldots n_{d+2-m_k;d+2-s}\\
&\asymp \bigvee_{k=0}^{j_2-j_1} \left\{ n_{d+2-s;d+1-s}^k \prod_{m=1}^{j_2-j_1-k} n_{d+2-j_1+m;d+2-s} \right\}.
\end{aligned}$$

Treating the above display as a polynomial of $n_{d+2-s;d+1-s}$, it suffices to match the corresponding coefficients of $n_{d+2-s;d+1-s}^k$ for $k \in [0; j_2 - j_1]$ in $\overline{\beta}_{i,j_2}^s/\overline{\beta}_{i,j_1}^s$. To this end, we have

$$\begin{aligned}
\frac{\overline{\beta}_{i,j_2}^s}{\overline{\beta}_{i,j_1}^s} &\asymp \frac{\bigvee_{\ell=0}^{j_2} n_{d+2-s;d+1-s}^{j_2-\ell}\overline{\beta}_\ell^{s-1}}{\bigvee_{\ell=0}^{j_1} n_{d+2-s;d+1-s}^{j_1}\overline{\beta}_\ell^{s-1}} \asymp \bigvee_{k=0}^{j_2-j_1} \frac{\bigvee_{\ell=0}^{j_1} n_{d+2-s;d+1-s}^{j_1+k-\ell}\overline{\beta}_{j_2-j_1-k+\ell}^{s-1}}{\bigvee_{\ell=0}^{j_1} n_{d+2-s;d+1-s}^{j_1-\ell}\overline{\beta}_\ell^{s-1}}\\
&\geq \bigvee_{k=0}^{j_2-j_1} \left\{ n_{d+2-s;d+1-s}^k \bigwedge_{\ell=0}^{j_1} \frac{\overline{\beta}_{j_2-j_1-k+\ell}^{s-1}}{\overline{\beta}_\ell^{s-1}} \right\} \quad \text{(by Lemma 38)}\\
&\gtrsim \bigvee_{k=0}^{j_2-j_1} \left\{ n_{d+2-s;d+1-s}^k \bigwedge_{\ell=0}^{j_1} \prod_{m=\ell+1}^{j_2-j_1-k+\ell} n_{d+2-m;d+2-s} \right\} \quad \text{(by induction)}\\
&= \bigvee_{k=0}^{j_2-j_1} \left\{ n_{d+2-s;d+1-s}^k \prod_{m=j_1+1}^{j_2-k} n_{d+2-m;d+2-s} \right\} \quad \text{(minimum at } \ell = j_1\text{)},
\end{aligned}$$

matching the calculation in the previous display, completing the proof. $\square$

Appendix C

APPENDIX OF CHAPTER 4

C.1 Proofs of results in Section 4.3

C.1.1 Proof of Theorem 4

We need the following proposition, which can be proved using techniques similar to (Goldstein et al., 2017, Theorem 2.1). We provide the details of its proof in Appendix C.3.2 for the convenience of the reader.

Proposition 29. *Suppose K_0, K are two non-empty closed convex sets in $\mathbb{R}^n$. Let*

$$T_{K_0,K}(y) \equiv \|y - \Pi_{K_0}(y)\|^2 - \|y - \Pi_K(y)\|^2.$$

Then for any $\mu \in \mathbb{R}^n$, under the model (4.1),

$$d_{\mathrm{TV}}\bigg(\frac{T_{K_0,K}(Y) - \mathbb{E}_\mu T_{K_0,K}(Y)}{\sqrt{Var_\mu(T_{K_0,K}(Y))}}, \mathcal{N}(0,1)\bigg) \leq \frac{16\sqrt{\mathbb{E}_\mu \|\widehat{\mu}_K - \widehat{\mu}_{K_0}\|^2}}{Var_\mu(T_{K_0,K}(Y))}.$$

The next lemma provides a lower bound for the variance of $F(\xi)$, where the absolute continuity of $F : \mathbb{R}^n \to \mathbb{R}$ is valid up to its first derivatives. The proof is based on Fourier analysis in the Gaussian space in the spirit of (Nourdin and Peccati, 2012, Proposition 1.5.1).

Lemma 44. *Let $F : \mathbb{R}^n \to \mathbb{R}$ be such that $\{\partial_{\boldsymbol{k}} F : |\boldsymbol{k}| \leq 1\}$ are absolutely continuous and $\{\partial_{\boldsymbol{k}} F : |\boldsymbol{k}| \leq 2\}$ have sub-exponential growth at ∞. Then*

$$Var(F(\xi)) \geq \sum_i \big(\mathbb{E}\partial_i F(\xi)\big)^2 + \sum_{i \neq j} \big(\mathbb{E}\partial_{ij} F(\xi)\big)^2 + \frac{1}{2}\sum_i \big(\mathbb{E}\partial_{ii} F(\xi)\big)^2.$$

Proof. We only need to verify the above claimed inequality for $\mathbb{E}F(\xi) = 0$. Let $H_k(x) = (-1)^k e^{x^2/2} \frac{\mathrm{d}^k}{\mathrm{d}x^k} e^{-x^2/2}$ be the Hermite polynomial of order k. For a multi-index $\boldsymbol{k} = (k_1, \ldots, k_n)$

and $y \in \mathbb{R}^n$, let $H_{\boldsymbol{k}}(y) \equiv \prod_{i=1}^n H_{k_i}(y_i)$. Then $\{H_{\boldsymbol{k}} : \boldsymbol{k} \in \mathbb{Z}_{\geq 0}^n\}$ is a complete orthogonal basis of $L_2(\gamma_n)$, where γ_n is the standard Gaussian measure on $\mathbb{R}^n$. On the other hand, the absolute continuity and growth condition on F ensures the validity of the following Gaussian integration-by-parts: For all multi-indices $\boldsymbol{k}$ such that $|\boldsymbol{k}| \leq 2$,

$$\mathbb{E}\big[F(\xi)H_{\boldsymbol{k}}(\xi)\big] = \mathbb{E}\partial_{\boldsymbol{k}} F(\xi).$$

As $\mathbb{E}|H_{\boldsymbol{k}}(\xi)|^2 = \boldsymbol{k}!$, it follows by Plancherel's theorem that

$$\mathrm{Var}(F(\xi)) = \mathbb{E}F^2(\xi) \geq \sum_{\boldsymbol{k}:|\boldsymbol{k}|\leq 2} \frac{\big(\mathbb{E}F(\xi)H_{\boldsymbol{k}}(\xi)\big)^2}{\mathbb{E}|H_{\boldsymbol{k}}(\xi)|^2},$$

which equals the right hand side of the claimed inequality. □

Proof of Theorem 4. Let

$$F(\xi) \equiv T(\mu_0 + \xi) = \|\mu_0 + \xi - \mu_0\|^2 - \|\mu_0 + \xi - \Pi_K(\mu_0 + \xi)\|^2.$$

By Lemma 1-(1),

$$\nabla F(\xi) = \nabla T(\mu_0 + \xi) = 2\big(\Pi_K(\mu_0 + \xi) - \mu_0\big).$$

Hence

$$\partial_{ij} F(\xi) = \partial_{ij} T(\mu_0 + \xi) = 2(J_{\Pi_K}(\mu_0 + \xi))_{ji}.$$

We verify that F satisfies the condition of Lemma 44. By the above closed-form expression of F and ∇F, the absolute continuity for $\{\partial_{\boldsymbol{k}} F : |\boldsymbol{k}| \leq 1\}$ holds by noting that ∇F is 2-Lipschitz. On the other hand, as

$$|F(\xi)| = \big|\|\xi\|^2 - \|\mu_0 + \xi - \Pi_K(\mu_0 + \xi)\|^2\big| \leq C \cdot \big(\|\xi\|^2 \vee \|\mu_0\|^2\big),$$
$$\|\nabla F(\xi)\| \leq C \cdot \big(\|\mu_0\| \vee \|\xi\|\big), \qquad \|\nabla^2 F(\xi)\| = 2\|J_{\Pi_K}(\mu_0 + \xi)^\top\| \leq 2,$$

it follows that $\{\partial_{\boldsymbol{k}} F : |\boldsymbol{k}| \leq 2\}$ have sub-exponential growth at ∞. Now we may apply Lemma 44 to see that

$$\begin{aligned}\sigma_{\mu_0}^2 = \mathrm{Var}(T(Y)) &\geq \sum_i \big(\mathbb{E}\partial_i F(\xi)\big)^2 + \frac{1}{2}\sum_{i,j}\big(\mathbb{E}\partial_{ij} F(\xi)\big)^2 \\ &= 4\|\mathbb{E}\Pi_K(\mu_0+\xi) - \mu_0\|^2 + 2\sum_{i,j}\big(\mathbb{E}J_{\Pi_K}(\mu_0+\xi)\big)_{ij}^2,\end{aligned}$$

as desired. The claim of the theorem now follows from Proposition 29. □

C.1.2 Proof of Theorem 5

A simple but important observation in the proof of Theorem 5 is the following.

Proposition 30. *Let*

$$\begin{aligned}Z(\mu,\mu_0) &\equiv \Delta T_{\mu,\mu_0}(\xi) - \mathbb{E}(\Delta T_{\mu,\mu_0}) \\ &= T(\mu+\xi) - T(\mu_0+\xi) - (m_\mu - m_{\mu_0}).\end{aligned}$$

Then for any $t \geq 0$,

$$\mathbb{P}\big(Z(\mu,\mu_0) > t\big) \vee \mathbb{P}\big(Z(\mu,\mu_0) < -t\big) \leq \exp\bigg(-\frac{t^2}{8\|\mu-\mu_0\|^2}\bigg).$$

Proof. As

$$\nabla T(y) = \nabla\big(\|y-\mu_0\|^2 - \|y-\Pi_K(y)\|^2\big) = 2(\Pi_K(y) - \mu_0)$$

by Lemma 1-(1), it follows that

$$\|\nabla_\xi Z(\mu,\mu_0)\| = 2\|\Pi_K(\mu+\xi) - \Pi_K(\mu_0+\xi)\| \leq 2\|\mu-\mu_0\|.$$

The claim now follows by Gaussian concentration inequality for Lipschitz functions, cf. (Bousquet, 2003, Theorem 5.6). □

Lemma 45. *For any $t \in \mathbb{R}$, there exists some $C_t > 0$ such that for all $u \in \mathbb{R}, \eta \in [-1/2, 1/2]$,*

$$\big|\mathbb{P}\big(\mathcal{N}(u,1) \leq t\big) - \mathbb{P}\big(\mathcal{N}((1+\eta)u,1) \leq t\big)\big| \leq C_t \cdot |\eta|.$$

Furthermore, $\sup_{t\in M} C_t < \infty$ for any compact subset M of $\mathbb{R}$.

Proof. We assume $\eta \geq 0$ without loss of generality. Note that with φ denoting the d.f. for standard normal,

$$\begin{aligned}\mathbb{P}\big(\mathcal{N}(0,1) \leq t - (1+\eta)u\big) &\leq \mathbb{P}\big(\mathcal{N}(0,1) \leq t - u\big) + \eta \cdot \sup_{v\in[(t-u)\pm\eta|u|]} \varphi(v)|u| \\ &\leq \mathbb{P}\big(\mathcal{N}(0,1) \leq t - u\big) + \eta \cdot C_t,\end{aligned}$$

where $C_t \equiv \sup_u \sup_{v\in[(t-u)\pm(|u|/2)]} \varphi(v)|u| < \infty$ depends on t only. □

Proof of Theorem 5. First note that under the model (4.1), the normalized LRS $T(Y)$ satisfies the decomposition

$$\frac{T(Y) - m_{\mu_0}}{\sigma_{\mu_0}} = \frac{T(\mu+\xi) - T(\mu_0+\xi)}{\sigma_{\mu_0}} + \frac{T(\mu_0+\xi) - m_{\mu_0}}{\sigma_{\mu_0}}. \tag{C.1}$$

Using Proposition 30, on an event E_u with $\mathbb{P}(E_u) \geq 1-2e^{-u^2}$, we have $|Z(\mu,\mu_0)| \leq 3u\|\mu-\mu_0\|$ with $Z(\mu,\mu_0)$ defined therein. Then for any $t \in \mathbb{R}$,

$$\begin{aligned}&\mathbb{P}\bigg(\frac{T(\mu+\xi) - m_{\mu_0}}{\sigma_{\mu_0}} \leq t\bigg) \\ &= \mathbb{P}\bigg(\frac{m_\mu - m_{\mu_0} + Z(\mu,\mu_0)}{\sigma_{\mu_0}} + \frac{T(\mu_0+\xi) - m_{\mu_0}}{\sigma_{\mu_0}} \leq t\bigg) \\ &\leq \mathbb{P}\bigg(\frac{m_\mu - m_{\mu_0} - 3u\|\mu-\mu_0\|}{\sigma_{\mu_0}} + \mathcal{N}(0,1) \leq t\bigg) + 2e^{-u^2} + \mathrm{err}_{\mu_0} \qquad \text{(C.2)} \\ &= \mathbb{P}\bigg(\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}}\big(1+\eta(u)\big) + \mathcal{N}(0,1) \leq t\bigg) + 2e^{-u^2} + \mathrm{err}_{\mu_0},\end{aligned}$$

where

$$\eta(u) \equiv -3u \cdot \frac{\|\mu-\mu_0\|}{m_\mu - m_{\mu_0}}.$$

By choosing $u \le |m_\mu - m_{\mu_0}|/\big(6\|\mu-\mu_0\|\big)$, we have $|\eta(u)| \le 1/2$, so we may apply Lemma 45 to see that,

$$\begin{aligned}\Delta^* &\equiv \mathbb{P}\bigg(\frac{T(\mu+\xi)-m_{\mu_0}}{\sigma_{\mu_0}} \le t\bigg) - \mathbb{P}\bigg(\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}} + \mathcal{N}(0,1) \le t\bigg)\\ &\le 2e^{-u^2} + C_t u \cdot \frac{\|\mu-\mu_0\|}{|m_\mu - m_{\mu_0}|} + \mathrm{err}_{\mu_0}.\end{aligned}$$

Optimizing $u \le |m_\mu - m_{\mu_0}|/\big(6\|\mu-\mu_0\|\big)$, the first two terms in the error bound above can be bounded, up to an absolute constant, by

$$(1\vee C_t)\cdot \mathcal{L}\bigg(1 \bigwedge \frac{\|\mu-\mu_0\|}{|m_\mu - m_{\mu_0}|}\bigg).$$

Next we will obtain a similar upper bound for Δ^*, but replacing $|m_\mu - m_{\mu_0}|$ in the above display by σ_{μ_0}. To see this, (C.2) along with

$$\begin{aligned}&\mathbb{P}\bigg(\frac{m_\mu - m_{\mu_0} - 3u\|\mu-\mu_0\|}{\sigma_{\mu_0}} + \mathcal{N}(0,1) \le t\bigg)\\ &\quad\le \mathbb{P}\bigg(\frac{m_\mu - m_{\mu_0}}{\sigma_{\mu_0}} + \mathcal{N}(0,1) \le t\bigg) + \|\varphi\|_\infty \cdot \frac{3u\|\mu-\mu_0\|}{\sigma_{\mu_0}}\end{aligned}$$

yields that

$$\Delta^* \le \inf_{u>0}\bigg\{2e^{-u^2} + 3u\cdot\frac{\|\mu-\mu_0\|}{\sigma_{\mu_0}}\bigg\} + \mathrm{err}_{\mu_0} \le C\cdot\mathcal{L}\bigg(1\bigwedge\frac{\|\mu-\mu_0\|}{\sigma_{\mu_0}}\bigg) + \mathrm{err}_{\mu_0}.$$

Similar lower bounds can be derived. Applying the above arguments to the (at most 2) end point(s) of $\mathcal{A}_\alpha$ proves the inequality (4.18). Now (1) is a direct consequence of (4.18), while (2) follows by further noting $\Delta_{A_\alpha}(w_n) \to \beta$ if and only if all limit points of the sequence $\{w_n\}$ are contained in $\Delta_{A_\alpha}^{-1}(\beta)$. □

C.1.3 Proof of Theorem 6

By Lemma 3, we only need to consider $\mu = 0$. Note that: (i) $\|\xi - \Pi_{K'}(\xi)\|^2 = \|\xi\|^2 - \|\Pi_{K'}(\xi)\|^2$ for $K' \in \{K_0, K\}$, (ii) (K_0, K) is a non-oblique pair of closed convex cones in that $\Pi_{K_0} =$

$\Pi_{K_0} \circ \Pi_K$, so $\Pi_K(\xi) = \Pi_{K_0}(\xi) + \Pi_{K\cap K_0^*}(\xi)$ with $\langle \Pi_{K_0}(\xi), \Pi_{K\cap K_0^*}(\xi)\rangle = 0$ (cf. (Wei et al., 2019, Equation (25))), and hence $\|\Pi_K(\xi)\|^2 = \|\Pi_{K_0}(\xi)\|^2 + \|\Pi_{K\cap K_0^*}(\xi)\|^2$. Thus,

$$\begin{aligned}\mathbb{E}\|\Pi_K(\xi) - \Pi_{K_0}(\xi)\|^2 &= \mathbb{E}\|\Pi_{K\cap K_0^*}(\xi)\|^2 \\ &= \mathbb{E}\big[\|\Pi_K(\xi)\|^2 - \|\Pi_{K_0}(\xi)\|^2\big] = \delta_K - \delta_{K_0},\end{aligned}$$

and

$$\begin{aligned}\sigma_0^2 &= \mathrm{Var}\big(\|\Pi_{K_0}(\xi)\|^2 - \|\Pi_K(\xi)\|^2\big) \\ &= \mathrm{Var}\big(\|\Pi_{K\cap K_0^*}(\xi)\|^2\big) \overset{(*)}{\geq} 2\delta_{K\cap K_0^*} = 2\big(\delta_K - \delta_{K_0}\big).\end{aligned}$$

Here the inequality $(*)$ follows by Lemma 2-(2). The claim now follows from Proposition 29. □

C.1.4 Proof of Theorem 7

First note that we have the decomposition

$$\frac{T(\mu+\xi) - m_0}{\sigma_0} = \frac{T(\mu+\xi) - T(\xi)}{\sigma_0} + \frac{T(\xi) - m_0}{\sigma_0}. \tag{C.3}$$

As

$$\begin{aligned}m_\mu &= \mathbb{E}\big[\|\mu+\xi-\Pi_{K_0}(\mu+\xi)\|^2 - \|\mu+\xi-\Pi_K(\mu+\xi)\|^2\big] \\ &= \mathbb{E}\Bigg[\|\Pi_{K_0}(\mu+\xi)-\mu\|^2 - 2\langle\xi, \Pi_{K_0}(\mu+\xi)\rangle + \|\xi\|^2 \\ &\qquad - \Big(\|\Pi_K(\mu+\xi)-\mu\|^2 - 2\langle\xi, \Pi_K(\mu+\xi)\rangle + \|\xi\|^2\Big)\Bigg] \\ &= \Big\{\|\mu - \Pi_{K_0}(\mu)\|^2 + 2\mathbb{E}\langle\xi, \Pi_K(\mu+\xi)\rangle - \mathbb{E}\|\Pi_K(\mu+\xi)-\mu\|^2\Big\} - \delta_{K_0},\end{aligned}$$

we have (as $\delta_K = \mathbb{E}\|\Pi_K(\xi)\|^2 = \mathbb{E}\langle \xi, \Pi_K(\xi)\rangle$)

$$
\begin{aligned}
m_\mu - m_0 &= \mathbb{E}\bigg[2\langle \xi, \Pi_K(\mu+\xi)\rangle - \|\Pi_K(\mu+\xi)-\mu\|^2 - \langle \xi, \Pi_K(\xi)\rangle\bigg] + \|\mu - \Pi_{K_0}(\mu)\|^2 \\
&= \mathbb{E}\bigg[2\langle \xi, \Pi_K(\mu - \Pi_{K_0}(\mu)+\xi)\rangle \\
&\qquad - \|\Pi_K(\mu - \Pi_{K_0}(\mu)+\xi) - \big(\mu - \Pi_{K_0}(\mu)\big)\|^2 - \langle \xi, \Pi_K(\xi)\rangle\bigg] \\
&\qquad + \|\mu - \Pi_{K_0}(\mu)\|^2 \qquad\qquad \text{(by Lemma 3)} \\
&= \mathbb{E}\bigg[2\langle \mu - \Pi_{K_0}(\mu)+\xi, \Pi_K(\mu - \Pi_{K_0}(\mu)+\xi)\rangle \\
&\qquad - \|\Pi_K\big(\mu - \Pi_{K_0}(\mu)+\xi\big)\|^2 - \|\mu - \Pi_{K_0}(\mu)\|^2 - \langle \xi, \Pi_K(\xi)\rangle\bigg] \\
&\qquad + \|\mu - \Pi_{K_0}(\mu)\|^2 \\
&= \mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu)+\xi\big)\|^2 - \mathbb{E}\|\Pi_K(\xi)\|^2 = \Gamma_{K,2}(\mu - \Pi_{K_0}(\mu)).
\end{aligned}
$$

Here in the last line of the above display we used that

$$
\begin{aligned}
\mathbb{E}\langle \mu - \Pi_{K_0}(\mu)+\xi, \Pi_K(\mu - \Pi_{K_0}(\mu)+\xi)\rangle &= \mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu)+\xi\big)\|^2, \\
\mathbb{E}\langle \xi, \Pi_K(\xi)\rangle &= \mathbb{E}\|\Pi_K(\xi)\|^2.
\end{aligned}
$$

Let

$$
Z_0(\mu) \equiv T(\mu+\xi) - T(\xi) - (m_\mu - m_0).
$$

As $\nabla T(y) = \nabla\big(\|y - \Pi_{K_0}(y)\|^2 - \|y - \Pi_K(y)\|^2\big) = 2(\Pi_K(y) - \Pi_{K_0}(y))$ by Lemma 1-(1),

$$
\begin{aligned}
\nabla_\xi Z_0(\mu) &= 2\big(\Pi_K(\mu+\xi) - \Pi_K(\xi)\big) - 2\big(\Pi_{K_0}(\mu+\xi) - \Pi_{K_0}(\xi)\big) \\
&= 2\big(\Pi_K(\mu - \Pi_{K_0}(\mu)+\xi) - \Pi_K(\xi)\big) \\
&\qquad - 2\big(\Pi_{K_0}(\mu - \Pi_{K_0}(\mu)+\xi) - \Pi_{K_0}(\xi)\big), \qquad \text{(by Lemma 3)}
\end{aligned}
$$

and hence

$$
\|\nabla_\xi Z_0(\mu)\| \le 4\|\mu - \Pi_{K_0}(\mu)\|.
$$

Now using the Gaussian concentration inequality for Lipschitz functions, cf. (Boucheron et al., 2013, Theorem 5.6), it holds for any $t > 0$ that

$$\mathbb{P}\big(Z_0(\mu) > t\big) \vee \mathbb{P}\big(Z_0(\mu) < -t\big) \leq \exp\bigg(-\frac{t^2}{32\|\mu - \Pi_{K_0}(\mu)\|^2}\bigg).$$

From here we may conclude (4.28) by using similar arguments as in the proof of Theorem 5. Furthermore, by the proof of (Wei et al., 2019, Lemma E.1), $\Gamma_{K,2}(\nu) \geq \|\nu\|^2 \geq 0$ for any $\nu \in K$, so

$$\frac{\|\mu - \Pi_{K_0}(\mu)\|}{\big|\Gamma_{K,2}\big(\mu - \Pi_{K_0}(\mu)\big)\big| \vee \sigma_0} \overset{(*)}{\leq} \sup_{\nu \in K} \frac{\|\nu\|}{\Gamma_{K,2}(\nu) \vee \sigma_0} \leq \sup_{\nu \in K} \frac{1}{\|\nu\| \vee (\sigma_0/\|\nu\|)} \overset{(**)}{\leq} \frac{1}{\sigma_0^{1/2}}.$$

The inequality $(*)$ follows as $\mu - \Pi_{K_0}(\mu) \in K$ for $\mu \in K$, and $(**)$ follows as $\inf_{\nu \in K}\big\{\|\nu\| \vee (\sigma_0/\|\nu\|)\big\} \geq \inf_{t \geq 0}\big\{t \vee (\sigma_0/t)\big\} = \sigma_0^{1/2}$. As $\sigma_0^2 \asymp \delta_K - \delta_{K_0}$, the second inequality (4.29) follows by the bound $\mathrm{err}_0 \leq 8\big(\delta_K - \delta_{K_0}\big)^{-1/2}$ via Theorem 6.

Note that (1) is a direct consequence of (4.28) (as err_0 can be bounded above by Theorem 6 and $\mathcal{L}(0) = 0$) so we prove (2) below. To see the claimed power characterization, note that

$$\begin{aligned}
&\frac{\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\|^2 - \mathbb{E}\|\Pi_K(\xi)\|^2}{\sigma_0} \\
&= \frac{\big(\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\|\big)^2 - \big(\mathbb{E}\|\Pi_K(\xi)\|\big)^2}{\sigma_0} + \mathcal{O}(\sigma_0^{-1}) \\
&= \big(\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\| - \mathbb{E}\|\Pi_K(\xi)\|\big) \\
&\qquad \times \bigg[\frac{2\mathbb{E}\|\Pi_K(\xi)\|}{\sigma_0} + \frac{\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\| - \mathbb{E}\|\Pi_K(\xi)\|}{\sigma_0}\bigg] + \mathcal{O}(\sigma_0^{-1}) \\
&= \Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)\bigg[2\sqrt{\frac{\delta_K + \mathcal{O}(1)}{2\big(\delta_K - \delta_{K_0}\big) + \mathrm{Var}\big(V_{K \cap K_0^*}\big)}} \\
&\qquad + \frac{\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)}{\sigma_0}\bigg] + \mathcal{O}(\sigma_0^{-1}) \\
&= \Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)\bigg[2\sqrt{\frac{1 + \mathcal{O}(\delta_K^{-1})}{\big(2 + \mathrm{Var}\big(V_{K \cap K_0^*}\big)/\delta_{K \cap K_0^*}\big) \cdot \big(1 - \delta_{K_0}/\delta_K\big)}} \\
&\qquad + \frac{\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)}{\sigma_0}\bigg] + \mathcal{O}(\sigma_0^{-1}).
\end{aligned}$$

Under the growth condition $\sigma_0 \to \infty$, direct calculation now entails that

$$\begin{aligned} & \frac{\mathbb{E}\|\Pi_K\big(\mu - \Pi_{K_0}(\mu) + \xi\big)\|^2 - \mathbb{E}\|\Pi_K(\xi)\|^2}{\sigma_0} \to w^* \in [0, +\infty] \\ \Leftrightarrow \quad & \frac{2\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big)}{\sqrt{2 + \mathrm{Var}(V_{K\cap K_0^*})/\delta_{K\cap K_0^*}}\sqrt{1 - \delta_{K_0}/\delta_K}} \to w^* \in [0, +\infty]. \end{aligned}$$

The proof is now complete. □

C.1.5 Proof of Corollary 1

We will prove a slightly stronger (than (4.31)) claim that condition (4.33) implies

$$\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big) \to \infty. \tag{C.4}$$

Suppose $\|\mu - \Pi_{K_0}(\mu)\|$ is greater or equal than L_n times the right hand side of (4.33) for some slowly growing sequence $L_n \uparrow \infty$. Then either (i) $\|\mu - \Pi_{K_0}(\mu)\| \geq L_n\delta_K^{1/4}$, or (ii) $\|\mu - \Pi_{K_0}(\mu)\| < L_n\delta_K^{1/4}$ and $\langle \mu - \Pi_{K_0}(\mu), \mathbb{E}\Pi_K(\xi)\rangle \geq L_n\delta_K^{1/2}$. In both cases, we have $\|\mu - \Pi_{K_0}(\mu)\| \to \infty$ as there exists some universal constant $c_0 > 0$ such that the right hand side of (4.33) is bounded below by c_0. In case (i), using (Wei et al., 2019, (74a)),

$$\begin{aligned} \Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big) &\geq \frac{\|\mu - \Pi_{K_0}(\mu)\|^2}{2\|\mu - \Pi_{K_0}(\mu)\| + 8\mathbb{E}\|\Pi_K(\xi)\|} - 2/\sqrt{e} \\ &\geq (1/16)\|\mu - \Pi_{K_0}(\mu)\| \bigwedge \Big(\|\mu - \Pi_{K_0}(\mu)\|^2/\delta_K^{1/2}\Big) - 2/\sqrt{e} \\ &\geq (1/16)\|\mu - \Pi_{K_0}(\mu)\| \bigwedge L_n^2 - 2/\sqrt{e} \to \infty \end{aligned} \tag{C.5}$$

as $n \to \infty$, so (C.4) is verified. In case (ii), we may assume without loss of generality that $\|\mu - \Pi_{K_0}(\mu)\| \leq L_n^{1/4}\delta_K^{1/4}$ because otherwise we can follow the same arguments as in the previous case. Then using (Wei et al., 2019, (74b)) with

$$\begin{aligned} \alpha \equiv \alpha\big(\mu - \Pi_{K_0}(\mu)\big) &= 1 - e^{-\langle \mu - \Pi_{K_0}(\mu), \mathbb{E}\Pi_K(\xi)\rangle^2/8\|\mu - \Pi_{K_0}(\mu)\|^2} \\ &\geq 1 - e^{-\delta_K^{1/2}/8} \to 1, \end{aligned}$$

we have

$$\Gamma_K\big(\mu - \Pi_{K_0}(\mu)\big) \geq \alpha \cdot \frac{\langle \mu - \Pi_{K_0}(\mu), \mathbb{E}\Pi_K \xi\rangle - \|\mu - \Pi_{K_0}(\mu)\|^2}{\alpha\|\mu - \Pi_{K_0}(\mu)\| + 2\mathbb{E}\|\Pi_K(\xi)\|_2} - \frac{2}{\sqrt{e}}$$
$$\gtrsim \frac{(L_n - L_n^{1/2})\delta_K^{1/2}}{L_n^{1/4}\delta_K^{1/4} + \delta_K^{1/2}} - \mathcal{O}(1) \to \infty$$

as $n \to \infty$, so (C.4) is verified. The proof is complete. □

C.2 Proofs of results in Section 4.4

C.2.1 Proof of Theorem 8

Lemma 46. *Let ξ_1 be a standard normal random variable. Then for $x > 0$,*

$$\mathbb{E}[\xi_1 \mathbf{1}_{\xi_1 \geq x}] = \varphi(x), \quad \mathbb{E}[\xi_1^2 \mathbf{1}_{\xi_1 \geq x}] = x\varphi(x) + \int_x^\infty \varphi(y)\,\mathrm{d}y.$$

Proof. The first equality follows as $\mathbb{E}[\xi_1 \mathbf{1}_{\xi_1 \geq x}] = \int_x^\infty y\varphi(y)\,\mathrm{d}y = \varphi(x)$. The second equality follows as $\mathbb{E}[\xi_1^2 \mathbf{1}_{\xi_1 \geq x}] = \int_x^\infty y^2\varphi(y)\,\mathrm{d}y = -\int_x^\infty y\varphi'(y)\,\mathrm{d}y = x\varphi(x) + \int_x^\infty \varphi(y)\,\mathrm{d}y$. □

Proof of Theorem 8. Note that $\widehat{\mu}_{K_+} = \big((\mu_i + \xi_i)_+\big)$. For $\mu_0 \in K_+$, so

$$\mathbb{E}_{\mu_0}\|\widehat{\mu}_{K_+} - \mu_0\|^2 = \sum_{i=1}^n \mathbb{E}\Big[\big((\mu_0)_i + \xi_i\big)_+ - (\mu_0)_i\Big]^2$$
$$= \sum_{i=1}^n \Big[\mathbb{E}\xi_i^2 \mathbf{1}_{\xi_i \geq -(\mu_0)_i} + (\mu_0)_i^2 \mathbb{P}\big(\xi_i < -(\mu_0)_i\big)\Big].$$

As $(\mu_0)_i \geq 0$ for $1 \leq i \leq n$, and $\sup_{x>0} x^2 \mathbb{P}(\xi < -x) < \infty$, it follows that

$$\mathbb{E}_{\mu_0}\|\widehat{\mu}_{K_+} - \mu_0\|^2 \asymp n.$$

On the other hand, as under the null $J_{\widehat{\mu}_{K_+}} = \big(\mathbf{1}_{i=j}\mathbf{1}_{\xi_i \geq -(\mu_0)_i}\big)_{ij}$,

$$\|\mathbb{E}_{\mu_0} J_{\widehat{\mu}_{K_+}}\|_F^2 = \sum_{i,j} \big(\mathbb{E}_{\mu_0} J_{\widehat{\mu}_{K_+}}\big)_{ij}^2 = \sum_{i=1}^n \big(\mathbb{P}(\xi_i \geq -(\mu_0)_i)\big)^2 \asymp n. \tag{C.6}$$

The claim (1) now follows from Theorem 4.

For (2), let for $x \geq 0$

$$
\begin{aligned}
Q(x) &\equiv \mathbb{E}\xi_1^2 \mathbf{1}_{\xi_1 \geq -x} + 2x\mathbb{E}\xi_1 \mathbf{1}_{\xi_1 \geq -x} - x^2 \mathbb{P}\big(\xi_1 < -x\big) \\
&= 1 - \mathbb{E}\xi_1^2 \mathbf{1}_{\xi_1 \geq x} + 2x\mathbb{E}\xi_1 \mathbf{1}_{\xi_1 \geq x} - x^2 \mathbb{P}\big(\xi_1 \geq x\big) \\
&= \int_{-\infty}^{x} \varphi(y)\,\mathrm{d}y + x\varphi(x) - x^2\big(1 - \Phi(x)\big).
\end{aligned}
$$

The last equality follows from Lemma 46. Hence for all $x \geq 0$,

$$
\begin{aligned}
Q'(x) &= 2\varphi(x) + x\varphi'(x) - \big[2x(1-\Phi(x)) - x^2\varphi(x)\big] \\
&= 2\varphi(x) - 2x\big(1 - \Phi(x)\big), \\
Q''(x) &= 2\Big[\varphi'(x) - 1 + \Phi(x) + x\varphi(x)\Big] = 2\Big(-1 + \Phi(x)\Big) < 0.
\end{aligned}
$$

This means that Q' is nonnegative, decreasing with $Q'(0) = 2\varphi(0) = 2/\sqrt{2\pi}$ and $Q'(\infty) = 0$, and Q is strictly increasing, concave and bounded on $[0,\infty)$ with $Q(0) = 1/2$.

Now note that for any $\mu \in K_+$,

$$
\begin{aligned}
& m_\mu - \|\mu - \mu_0\|^2 \\
&= \mathbb{E}\big[2\langle \xi, \Pi_{K_+}(\mu+\xi) - \mu\rangle - \|\Pi_{K_+}(\mu+\xi) - \mu\|^2\big] \\
&= \sum_{i=1}^n \left[2\mathbb{E}\xi_i^2 \mathbf{1}_{\xi_i \geq -\mu_i} + 2\mu_i \mathbb{E}\xi_i \mathbf{1}_{\xi_i \geq -\mu_i} - \Big(\mathbb{E}\xi_i^2 \mathbf{1}_{\xi_i \geq -\mu_i} + \mu_i^2 \mathbb{P}\big(\xi_i < -\mu_i)\big)\Big)\right] \\
&= \sum_{i=1}^n \left[\mathbb{E}\xi_i^2 \mathbf{1}_{\xi_i \geq -\mu_i} + 2\mu_i \mathbb{E}\xi_i \mathbf{1}_{\xi_i \geq -\mu_i} - \mu_i^2 \mathbb{P}\big(\xi_i < -\mu_i)\big)\right] = \sum_{i=1}^n Q(\mu_i).
\end{aligned}
$$

Using the lower bound (C.6) for $\sigma_{\mu_0}^2$, and an easy matching upper bound (by e.g. triangle inequality), we have $\sigma_{\mu_0}^2 \asymp n$. The condition (4.19) reduces to

$$
\|\mu - \mu_0\| \ll \left|\sum_{i=1}^n \big\{\overline{S}_+(\mu_i) - \overline{S}_+((\mu_0)_i)\big\} + \|\mu - \mu_0\|^2\right| \bigvee n^{1/2}. \tag{C.7}
$$

(C.7) clearly holds for $\|\mu - \mu_0\| \ll n^{1/2}$. For $\|\mu - \mu_0\| \gg n^{1/2}$, as

$$
\left|\sum_{i=1}^n \big\{\overline{S}_+(\mu_i) - \overline{S}_+((\mu_0)_i)\big\}\right| \leq (2/\sqrt{2\pi})\|\mu - \mu_0\|_1 \lesssim \sqrt{n}\|\mu - \mu_0\|,
$$

the right hand side of (C.7) is bounded from below by

$$\big(\|\mu-\mu_0\|^2 - C\sqrt{n}\|\mu-\mu_0\|\big)_+ \vee n^{1/2} \asymp \|\mu-\mu_0\|^2 \gg \|\mu-\mu_0\|,$$

so (C.7) holds. Hence in these two regimes, the claim follows from Theorem 5-(2). For $\|\mu-\mu_0\| \asymp n^{1/2}$, by the decomposition (C.1), the LRT is powerful if and only if $|m_\mu - m_{\mu_0}|/\sigma_{\mu_0} \to \infty$, i.e., $\big|\sum_{i=1}^n \big\{\overline{S}_+(\mu_i) - \overline{S}_+((\mu_0)_i)\big\} + \|\mu-\mu_0\|^2\big| \gg n^{1/2}$. The proof is now complete. □

C.2.2 Proof of Theorem 9

We write $K_{\times,\alpha}$ for $K_\times$ in the proof for notational convenience.

(1). This claim follows from the fact that $\sigma_0 \asymp \delta_{K_\times}^{1/2} \asymp \delta_{K_\alpha}^{1/2} \asymp n^{1/2}$ (cf. (McCoy and Tropp, 2014, Section 6.3)) and Theorem 7-(1).

(2)(a). We only need to prove that the LRT is not powerful for $\mu \in K_\alpha$ such that $\|\mu\| = \mathcal{O}(1)$. Using the decomposition (C.3), it suffices to show $T(\mu+\xi) - T(\xi) = \mathcal{O}_{\mathbf{P}}(n^{1/2})$. This follows as

$$\begin{aligned}
&T(\mu+\xi) - T(\xi)\\
&= \|\mu+\xi\|^2 - \|\mu+\xi-\Pi_{K_\alpha}(\mu+\xi)\|^2 - \big(\|\xi\|^2 - \|\xi - \Pi_{K_\alpha}(\xi)\|^2\big)\\
&= \|\mu\|^2 + 2\langle \mu,\xi\rangle - \|\Pi_{K_\alpha}(\xi) - \Pi_{K_\alpha}(\mu+\xi) + \mu\|^2\\
&\qquad - 2\langle \xi - \Pi_{K_\alpha}(\xi), \Pi_{K_\alpha}(\xi) - \Pi_{K_\alpha}(\mu+\xi) + \mu\rangle\\
&= \mathcal{O}_{\mathbf{P}}\bigg(\|\mu\|^2 + \|\mu\| + \|\Pi_{K_\alpha}(\xi) - \Pi_{K_\alpha}(\mu+\xi)\|^2 + \|\mu\|^2\\
&\qquad + \|\xi - \Pi_{K_\alpha}(\xi)\|\big[\|\Pi_{K_\alpha}(\xi) - \Pi_{K_\alpha}(\mu+\xi)\| \vee \|\mu\|\big]\bigg)\\
&= \mathcal{O}_{\mathbf{P}}(n^{1/2}).
\end{aligned}$$

(2)(b). By (2)(a) and Theorem 7-(2), we have $\|\mu^1\| \gg 1$ if and only if

$$\frac{\mathbb{E}\|\Pi_{K_\alpha}(\mu^1+\xi^1)\|^2 - \mathbb{E}\|\Pi_{K_\alpha}(\xi^1)\|^2}{n^{1/2}} \to \infty.$$

Now using Theorem 7-(2) again for $K_\times$ to conclude by noting that

$$
\begin{aligned}
&\mathbb{E}\|\Pi_{K_\times}(\mu+\xi)\|^2 - \mathbb{E}\|\Pi_{K_\times}(\xi)\|^2 \\
&= \mathbb{E}\|\Pi_{K_\alpha}(\mu^1+\xi^1)\|^2 - \mathbb{E}\|\Pi_{K_\alpha}(\xi^1)\|^2 + (\mu^2)^2.
\end{aligned}
$$

This completes the proof. □

C.2.3 Proof of Theorem 10

We first prove Proposition 19. The following lemma will be used. We present its proof at the end of this subsection.

Lemma 47. *Fix $0.1n \le i \le 0.9n$. Let $u^* \le i$ and $h_1^* \ge 0$ be defined through the following max-min formula for the isotonic LSE:*

$$
\widehat{\mu}_i = \max_{u\le i}\min_{v\ge i} \overline{Y}|_{[u,v]} \equiv \min_{v\ge i} \overline{Y}|_{[u^*,v]} \equiv \min_{h_2\ge 0} \overline{Y}|_{[i-h_1^* n^{2/3}, i+h_2 n^{2/3}]}. \tag{C.8}
$$

Then there exists some $C = C(L) > 0$ such that for any $t > 0$

$$
\mathbb{P}\big(|\widehat{\mu}_i - \mu_i| > n^{-1/3}t\big) \vee \mathbb{P}(h_1^* > t) \le C\exp(-t^2/C).
$$

Proof of Proposition 19. We write in the proof $\widehat{\mu} = \widehat{\mu}_{K_\uparrow}$ and $\mu = \mu_0$ for simplicity of notation. Note that $(J_{\widehat{\mu}})_{ij} = \mathbf{1}_{\widehat{\mu}_i=\widehat{\mu}_j}(1/|\{k:\widehat{\mu}_k=\widehat{\mu}_i\}|)$. Note that

$$
\begin{aligned}
\mathbb{P}(\widehat{\mu}_i = \widehat{\mu}_j) &= \mathbb{E}\bigg[\mathbf{1}_{\widehat{\mu}_i=\widehat{\mu}_j}\cdot \frac{1}{|\{k:\widehat{\mu}_k=\widehat{\mu}_i\}|^{1/2}} \cdot |\{k:\widehat{\mu}_k=\widehat{\mu}_i\}|^{1/2}\bigg] \\
&\le \sqrt{(\mathbb{E}J_{\widehat{\mu}})_{ij}}\cdot\sqrt{\mathbb{E}|\{k:\widehat{\mu}_k=\widehat{\mu}_i\}|}.
\end{aligned}
$$

This implies that

$$
(\mathbb{E}J_{\widehat{\mu}})_{ij} \ge \frac{\mathbb{P}^2(\widehat{\mu}_i=\widehat{\mu}_j)}{\mathbb{E}|\{k:\widehat{\mu}_k=\widehat{\mu}_i\}|}. \tag{C.9}
$$

We will bound the denominator from above and the numerator from below in the above display separately in the regime $\{(i,j): |i-j| \le \kappa n^{2/3}, 0.1n \le i,j \le 0.9n\}$, where $\kappa = \kappa(L) > 0$ is a constant to be specified below.

Fix $0.1n \le i \le 0.9n$. First we provide an upper bound for the denominator in (C.9). By Lemma 47 and using the notation defined therein, there exists some large $c = c(L, \varepsilon) > 1$ such that on an event E_0 with probability $1 - \varepsilon$,

$$|\widehat{\mu}_i - \mu_i| \le c n^{-1/3}, \tag{C.10}$$

and

$$\mathbb{P}\big(E_1 \equiv \big\{h_1^* \ge c\big\}\big) \le C e^{-c^2/C},$$

where $C = C(L) > 0$ is a constant depending on L only. Hence integrating the tail leads to the following: for some constant $C' = C'(L) > 0$,

$$\mathbb{E}|\{k \le i : \widehat{\mu}_k = \widehat{\mu}_i\}| \le C' n^{2/3}.$$

Similarly we can handle the case $k \ge i$, so we arrive at

$$\mathbb{E}|\{k : \widehat{\mu}_k = \widehat{\mu}_i\}| \le C'' n^{2/3} \tag{C.11}$$

for some constant $C'' = C''(L) > 0$.

Next we provide a lower bound for the numerator of (C.9). On the event $E_2 = \{1 \vee (i - c^{-100} n^{2/3}) \le u^* \le i\}$ (there is nothing special about the constant 100—a large enough value suffices), we have

$$\begin{aligned}
&n^{1/3}\big(\widehat{\mu}_i - \mu_i\big) = \min_{v \ge i} n^{1/3}\big(\overline{\mu}|_{[u^*,v]} - \mu_i + \overline{\xi}|_{[u^*,v]}\big)\\
&\le \min_{i \le v \le n \wedge (i + c^{-10} n^{2/3})} n^{1/3}\big(\overline{\mu}|_{[u^*,v]} - \mu_i + \overline{\xi}|_{[u^*,v]}\big)\\
&\le \min_{i \le v \le n \wedge (i + c^{-10} n^{2/3})} n^{1/3} \overline{\xi}|_{[u^*,v]} + \max_{i \le v \le n \wedge (i + c^{-10} n^{2/3})} n^{1/3}\big(\overline{\mu}|_{[u^*,v]} - \mu_i\big)\\
&\le \min_{0 \le h_2 \le c^{-10}} \frac{W(-h_1^*) + W(h_2) + R_n}{h_1^* + h_2 + |\mathcal{O}(n^{-2/3})|} + \mathcal{O}(c),
\end{aligned}$$

where $R_n = \mathcal{O}_{a.s.}(\log n/n^{1/3})$ comes from Kolmós-Major-Tusnády strong embedding, and W denotes a standard two-sided Brownian motion starting from 0. The bound $\mathcal{O}(c)$ for the bias term follows as

$$\begin{aligned}
&\max_{i\le v\le n\wedge(i+c^{-10}n^{2/3})} n^{1/3}\big(\overline{\mu}|_{[u^*,v]}-\mu_i\big) \le \max_{i\le v\le n\wedge(i+cn^{2/3})} n^{1/3}\big(\overline{\mu}|_{[i,v]}-\mu_i\big)\\
&\le \max_{n^{-2/3}\le h_2\le c} \frac{n^{1/3}\sum_{\ell\in[i,i+h_2n^{2/3}]\cap\mathbb{Z}}(\mu_\ell-\mu_i)}{\lfloor h_2 n^{2/3}\rfloor+1}\\
&\le \max_{n^{-2/3}\le h_2\le c} \frac{n^{1/3}(L/n)\sum_{\ell\in[0,h_2n^{2/3}]\cap\mathbb{Z}}\ell}{\lfloor h_2 n^{2/3}\rfloor+1} = \mathcal{O}(c). \qquad \text{(C.12)}
\end{aligned}$$

Now on the event E_2,

$$W(-h_1^*) \le \sup_{0\le h_1\le c^{-100}} W(-h_1) \stackrel{d}{=} c^{-50}\sup_{0\le t\le 1} W(t) \equiv c^{-50}Z.$$

By reflection principle for Brownian motion, for any $u>0$,

$$\mathbb{P}\big(\{W(-h_1^*)>c^{-50}u\}\cap E_2\big) \le \mathbb{P}\big(Z>u\big) = 2\mathbb{P}\big(W(1)>u\big) \le 2e^{-u^2/2}.$$

Let h_2° be such that

$$W(h_2^\circ) \equiv \inf_{0\le h_2\le c^{-10}} W(h_2) \stackrel{d}{=} c^{-5}\inf_{0\le t\le 1} W(t) = -c^{-5}Z.$$

So for $u>0$,

$$\mathbb{P}(W(h_2^\circ)<-c^{-5}u) = \mathbb{P}(Z>u) = 1-\mathbb{P}\big(|\mathcal{N}(0,1)|\le u\big) \ge 1-2u.$$

Hence on the event E_2 intersected with an event with probability at least $1-4\varepsilon$,

$$\begin{aligned}
&\min_{0\le h_2\le c^{-10}} \frac{W(-h_1^*)+W(h_2)+R_n}{h_1^*+h_2+|\mathcal{O}(n^{-2/3})|} \le \frac{W(-h_1^*)+W(h_2^\circ)+R_n}{h_1^*+h_2^\circ+|\mathcal{O}(n^{-2/3})|}\\
&\le \frac{c^{-50}\sqrt{2\log(1/\varepsilon)}-c^{-5}\varepsilon+R_n}{h_1^*+h_2^\circ+|\mathcal{O}(n^{-2/3})|} \le -C_\varepsilon\cdot c^5.
\end{aligned}$$

where the last inequality follows by choosing for $c=c(\varepsilon)$ large enough followed by n large enough, and $h_1^*+h_2^\circ \le c^{-100}+c^{-10} \le 2c^{-10}$ on E_2. Combining the above estimates, we see that

$$n^{1/3}\big(\widehat{\mu}_i-\mu_i\big)\leq -C'_\varepsilon\cdot c^5$$

on the event E_2 intersected with an event with probability at least $1-4\varepsilon$, when c and n are chosen large enough, depending on L,ε. This event must occur with small enough probability for c large in view of (C.10), so we have proved that $\mathbb{P}(E_2)\leq 5\varepsilon$ for large enough $c=c(L,\varepsilon)>1$ and $n=n(L,\varepsilon)\in\mathbb{N}$. This means that $\mathbb{P}(\widehat{\mu}_i=\widehat{\mu}_j)\geq 1-5\varepsilon$ for $1\vee(i-c^{-100}n^{2/3})\leq j\leq i$ for large enough $c=c(L,\varepsilon)>1$ and $n=n(L,\varepsilon)\in\mathbb{N}$. Similarly one can handle the regime $i\leq j\leq (i+c^{-100}n^{2/3})\vee n$. In summary, we have proved there exists some $\kappa=\kappa(L)>0$ such that

$$\mathbb{P}(\widehat{\mu}_i=\widehat{\mu}_j)\geq 1/2 \tag{C.13}$$

holds for $\{(i,j):|i-j|\leq \kappa n^{2/3}, 0.1n\leq i,j\leq 0.9n\}$ for n large enough. The claim of the proposition now follows by plugging (C.11) and (C.13) into (C.9). □

Now we are in position to prove Theorem 10.

Proof of Theorem 10-(1). We write in the proof $\widehat{\mu}=\widehat{\mu}_K$ and $\mu=\mu_0$ for simplicity of notation. $\mathbb{P}_\mu,\mathbb{E}_\mu$ are shorthanded to $\mathbb{P},\mathbb{E}$ if no confusion could arise. For $\kappa>0$, let $I_\ell\equiv I_\ell(\kappa)\equiv\{i:0.1n+(\ell-1)\cdot\kappa n^{2/3}\leq i\leq \big(0.1n+\ell\cdot\kappa n^{2/3}\big)\wedge 0.9n\}$ and ℓ_0 be the maximum integer for which $I_{\ell_0}\subset[0.1n,0.9n]$. Clearly $|I_\ell|\asymp\kappa n^{2/3}$ for all $1\leq\ell\leq\ell_0$ and $\ell_0\asymp n^{1/3}/\kappa$. Using the κ specified in Proposition 19, we have

$$\begin{aligned}\|\mathbb{E}J_{\widehat{\mu}}\|_F^2=\sum_{i,j}\big(\mathbb{E}J_{\widehat{\mu}}\big)_{ij}^2&\geq\sum_{\ell=1}^{\ell_0}\sum_{(i,j)\in I_\ell\times I_\ell}\big(\mathbb{E}(J_{\widehat{\mu}})_{ij}\big)^2\\&\gtrsim\sum_{\ell=1}^{\ell_0}\sum_{(i,j)\in I_\ell\times I_\ell}n^{-4/3}\asymp n^{1/3}.\end{aligned} \tag{C.14}$$

On the other hand, by e.g., Zhang (2002), under the condition of Theorem 10, we have

$$\mathbb{E}\|\widehat{\mu}-\mu\|^2\lesssim n^{1/3}.$$

The claim now follows by applying Theorem 4 (by ignoring the bias term in the denominator) with the above two displays. □

Proof of Theorem 10-(2). Following the notation used in Meyer and Woodroofe (2000), let $\widetilde{W}$ be the greatest convex minorant of $t \mapsto W(t) + t^2/2, t \in \mathbb{R}$, and $a = -\mathbb{E}[\widetilde{W}(0)] > 0, b = \mathbb{E}[\widetilde{W}'(0)^2] > 0$. Using the same techniques as in (Meyer and Woodroofe, 2000, Theorem 2, Corollary 4) but by performing Taylor expansion to the second order, it can be shown that for all C^2 monotone functions $f : [0,1] \to \mathbb{R}$ with bounded first derivative f' away from 0 and ∞, and bounded second derivative f'' away from ∞,

$$\begin{aligned}\mathbb{E}_{\mu_f} \operatorname{div} \widehat{\mu}_{K_\uparrow} &= (a+b) \cdot n^{1/3} \int_0^1 (f'(t))^{2/3} \,\mathrm{d}t + \mathcal{O}(1), \\ \mathbb{E}_{\mu_f} \|\widehat{\mu}_{K_\uparrow} - \mu_f\|^2 &= b \cdot n^{1/3} \int_0^1 (f'(t))^{2/3} \,\mathrm{d}t + \mathcal{O}(1). \end{aligned} \tag{C.15}$$

Here $\mu_f = (f(i/n))_{i=1}^n$ for a generic $f : [0,1] \to \mathbb{R}$, and the $\mathcal{O}(1)$ term in the above display depends only on the upper and lower bounds for f' and the upper bound of f''. Hence for the prescribed f,

$$\begin{aligned} m_{\mu_f} - \|\mu_f - \mu_{f_0}\|^2 &= 2\mathbb{E}_{\mu_f} \operatorname{div} \widehat{\mu}_{K_\uparrow} - \mathbb{E}_{\mu_f} \|\widehat{\mu}_{K_\uparrow} - \mu_f\|^2 \\ &= (2a+b) \cdot n^{1/3} \int_0^1 (f'(t))^{2/3} \,\mathrm{d}t + \mathcal{O}(1). \end{aligned}$$

On the other hand, for the prescribed f, (C.14) provides a lower bound for $\sigma^2_{\mu_f}$, while the Gaussian-Poincaré inequality yields a matching upper bound:

$$n^{1/3} \lesssim \sigma^2_{\mu_f} \leq 4\mathbb{E}_{\mu_f} \|\widehat{\mu}_{K_\uparrow} - \mu_f\|^2 \lesssim n^{1/3}.$$

Now with $\|\delta\|_{[1]} \equiv \int \delta'$, condition (4.19) reduces to

$$\begin{aligned}
& \|\mu_f - \mu_{f_0}\| \ll \left| m_{\mu_f} - m_{\mu_{f_0}} \right| \vee \sigma_{\mu_f} \\
\Leftrightarrow \quad & \sqrt{n \int (f - f_0)^2 + \mathcal{O}(1)} \\
& \quad \ll \left| (2a+b) n^{1/3} \int \left\{ (f')^{2/3} - (f_0')^{2/3} \right\} + \mathcal{O}(1) + n \int (f - f_0)^2 \right| \bigvee n^{1/6} \\
\Leftrightarrow \quad & \sqrt{n\rho_n^2 + \mathcal{O}(1)} \ll \left| -|\mathcal{O}(1)| n^{1/3} \rho_n \|\delta\|_{[1]} + \mathcal{O}(1) + n\rho_n^2 \right| \bigvee n^{1/6}, \qquad \text{(C.16)}
\end{aligned}$$

where in the last equivalence we used that

$$\int \left\{ (f')^{2/3} - (f_0')^{2/3} \right\} = \int \left\{ (f')^{2/3} - (f' + \rho_n \delta')^{2/3} \right\} = -\mathcal{O}(\rho_n) \int \delta'.$$

By Theorem 5-(2), under (C.16) the LRT is power consistent if and only if

$$\frac{\left| -|\mathcal{O}(1)| n^{1/3} \rho_n \|\delta\|_{[1]} + n\rho_n^2 \right|}{n^{1/6}} \to \infty. \qquad \text{(C.17)}$$

We have two cases:

1. If $n\rho_n^2 \gg n^{1/3}\rho_n |\|\delta\|_{[1]}| \Leftrightarrow \rho_n \gg n^{-2/3} |\|\delta\|_{[1]}|$, then (C.17) requires $\rho_n \gg n^{-5/12}$.

2. If $n\rho_n^2 \ll n^{1/3}\rho_n |\|\delta\|_{[1]}| \Leftrightarrow \rho_n \ll n^{-2/3} |\|\delta\|_{[1]}|$, then (C.17) requires $\rho_n \gg n^{-1/6}/|\|\delta\|_{[1]}|$. This is not feasible as $|\|\delta\|_{[1]}| = |\int \delta'| = \mathcal{O}(1)$.

To summarize, (C.17) is equivalent to requiring $\rho_n \gg n^{-5/12}$. In this regime (C.16) also holds. The proof is complete. □

Proof of Lemma 47. By the monotonicity of μ, we have

$$\begin{aligned}
\widehat{\mu}_i - \mu_i &= \min_{h_2 \geq 0} \overline{Y}|_{[i - h_1^* n^{2/3}, i + h_2 n^{2/3}]} - \mu_i \leq \overline{Y}|_{[i - h_1^* n^{2/3}, i + n^{2/3}]} - \mu_i \\
&= \left(\overline{\mu}|_{[i - h_1^* n^{2/3}, i + n^{2/3}]} - \mu_i \right) + \overline{\xi}|_{[i - h_1^* n^{2/3}, i + n^{2/3}]} \\
&\leq \left(\mu_{\lceil i + n^{2/3} \rceil} - \mu_i \right) + \max_{h_1 \geq 0} |\overline{\xi}|_{[i - h_1 n^{2/3}, i + n^{2/3}]}| \\
&\leq C_1 n^{-1/3} + \max_{h_1 \geq 0} |\overline{\xi}|_{[i - h_1 n^{2/3}, i + n^{2/3}]}|,
\end{aligned}$$

where the last inequality follows by (4.35). Note for any $t > 0$, a standard blocking argument (cf. (Han and Zhang, 2019, Lemma 2)) yields

$$\mathbb{P}\bigg(\max_{h_1\geq 0} |\overline{\xi}|_{[i-h_1 n^{2/3}, i+n^{2/3}]}| > t n^{-1/3}\bigg) \leq C e^{-t^2/C}. \tag{C.18}$$

This concludes the one-sided estimate $\mathbb{P}(\widehat{\mu}_i - \mu_i > n^{-1/3} t)$. The other side is similar.

Now consider $\mathbb{P}(h_1^* > t)$. On the event $\{h_1^* > t\}$, we have

$$\begin{aligned}
\widehat{\mu}_i - \mu_i &= \min_{h_2\geq 0} \overline{Y}|_{[i-h_1^* n^{2/3}, i+h_2 n^{2/3}]}\\
&\leq \big(\overline{\mu}|_{[i-h_1^* n^{2/3}, i+n^{2/3}]} - \mu_i\big) + \overline{\xi}|_{[i-h_1^* n^{2/3}, i+n^{2/3}]}\\
&\leq \big(\overline{\mu}|_{[i-t n^{2/3}, i+n^{2/3}]} - \mu_i\big) + \max_{h_1\geq 0} \big|\overline{\xi}|_{[i-h_1 n^{2/3}, i+n^{2/3}]}\big|\\
&\leq -C_2 \cdot t n^{-1/3} + C_3 \cdot n^{-1/3} + \max_{h_1\geq 0} \big|\overline{\xi}|_{[i-h_1 n^{2/3}, i+n^{2/3}]}\big|,
\end{aligned}$$

where the last inequality follows from calculations similar to (C.12), but now using both the upper and lower bound parts of (4.35). Choosing $t \geq 2C_3/C_2$, and replacing t in (C.18) by $C_2 t/4$, we see that

$$\mathbb{P}\big(h_1^* > t\big) \leq \mathbb{P}\big(\widehat{\mu}_i - \mu_i \leq -(C_2/4) t n^{-1/3}\big) + C_4 e^{-t^2/C_4}.$$

The claim follows by adjusting constants. □

C.2.4 Proof of Theorem 11

We first prove Proposition 20. The following lemma will be useful to control the term $\mathfrak{p}_{\lambda,\mu_0}$ therein. We present its proof at the end of this subsection.

Lemma 48. *For any $\theta \in \mathbb{R}^p$ with $\|\theta\|_1 \leq \lambda$, let $\mu \equiv X\theta \in K_{X,\lambda}$. There exists some universal constant $C > 0$ such that for $t \geq 1$,*

$$\mathbb{P}_\mu\bigg(\|\widehat{\theta}^0\|_1 \geq \|\theta\|_1 + t\sqrt{\frac{p}{n\lambda_{\min}(\Sigma)}}\bigg) \leq e^{-t^2/C}.$$

Proof of Proposition 20. (1). We will derive an explicit formula for $J_{\widehat{\mu}}$ using the results of Kato (2009). First note by the chain rule that

$$J_{\widehat{\mu}}(\xi) = \frac{\partial \widehat{\mu}}{\partial \xi} = \frac{\partial \widehat{\mu}}{\partial \widehat{\theta}} \frac{\partial \widehat{\theta}}{\partial \widehat{\theta}^0} \frac{\partial \widehat{\theta}^0}{\partial \xi} = X \frac{\partial \widehat{\theta}}{\partial \widehat{\theta}^0} (X^\top X)^{-1} X^\top.$$

Let $\widetilde{K} \equiv \widetilde{K}_\lambda \equiv \{\theta \in \mathbb{R}^p : \|\theta\|_1 \le \lambda\}$. For each $m \in \{1, \ldots, p\}$, suppose there are N_m faces of $\widetilde{K}_\lambda$ of dimension m, denoted as $\{F_{m,\ell}\}_{\ell=1}^{N_m}$. Then we can partition $\partial \widetilde{K}_\lambda$ as $\{F_{m,\ell}\}_{m,\ell}$. Let $\big\{E_0, \{E_{m,\ell}\}_{m,\ell}\big\}$ be a partition of $\mathbb{R}^p$ defined as $E_0 \equiv \widetilde{K}_\lambda$, $E_{m,\ell} \equiv \{y \in \mathbb{R}^p : \Pi_{\widetilde{K}_\lambda}(y) \in F_{m,\ell}\}$. Let $E_0^\circ, E_{m,\ell}^\circ$ be the interiors of $E_0, E_{m,\ell}$, respectively. Since $\widetilde{K}_\lambda$ is a polyhedron, it follows by (Kato, 2009, Equation (3.6) and Remark 3.3) that when $\widehat{\theta}^0 \in E_{m,\ell}^\circ$,

$$\frac{\partial \widehat{\theta}}{\partial \widehat{\theta}^0} = B_{m,\ell}(B_{m,\ell}^\top X^\top X B_{m,\ell})^{-1} B_{m,\ell}^\top X^\top X,$$

where $B_{m,\ell} = [b_{1,\ell}, \ldots, b_{p-m,\ell}] \in \mathbb{R}^{p \times (p-m)}$ whose columns are linearly independent and span the tangent space at any point in $F_{m,\ell}$. Hence on the event $\{\widehat{\theta}^0 \in E_{m,\ell}^\circ\}$,

$$J_{\widehat{\mu}}(\xi) = X B_{m,\ell}(B_{m,\ell}^\top X^\top X B_{m,\ell})^{-1}(B_{m,\ell}^\top X^\top),$$

which is a projection matrix onto the column space of $XB_{m,\ell}$. In other words, a.e. on $\mathbb{R}^n$,

$$J_{\widehat{\mu}}(\xi) = \mathbf{1}_{\widehat{\theta}^0 \in E_0^\circ} \cdot Z_0 + \sum_{m=1}^{p} \sum_{\ell=1}^{N_m} \mathbf{1}_{\widehat{\theta}^0 \in E_{m,\ell}^\circ} \cdot Z_{m,\ell},$$

where

$$Z_0 \equiv X(X^\top X)^{-1} X^\top, \quad Z_{m,\ell} \equiv X B_{m,\ell}(B_{m,\ell}^\top X^\top X B_{m,\ell})^{-1}(B_{m,\ell}^\top X^\top).$$

Hence,

$$\mathbb{E}_{\mu_0} J_{\widehat{\mu}}(\xi) = Z_0 \cdot \mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \in E_0^\circ\big) + \sum_{m=1}^{p} \sum_{\ell=1}^{N_m} \mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \in E_{m,\ell}^\circ\big) \cdot Z_{m,\ell},$$

and therefore

$$
\begin{aligned}
\|\mathbb{E}_{\mu_0} J_{\widehat{\mu}}\|_F^2 &= \sum_{i,j=1}^{n} \left(\mathbb{E}_{\mu_0} J_{\widehat{\mu}}\right)_{ij}^2 \\
&= \sum_{i,j=1}^{n} \left[(Z_0)_{ij}\mathbb{P}\big(\widehat{\theta}^0 \in E_0^\circ\big) + \sum_{m,\ell} \mathbb{P}\big(\widehat{\theta}^0 \in E_{m,\ell}^\circ\big)(Z_{m,\ell})_{ij}\right]^2 \\
&\geq \sum_{i,j=1}^{n} \left[(Z_0)_{ij}\mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \in E_0^\circ\big) - \mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \notin E_0^\circ\big) \max_{m,\ell}|(Z_{m,\ell})_{ij}|\right]_+^2 \\
&\overset{(*)}{\geq} \sum_{i,j=1}^{n} \left[(Z_0)_{ij}\mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \in E_0^\circ\big) - \mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \notin E_0^\circ\big)\right]_+^2 \\
&\overset{(*)}{\geq} \sum_{i,j=1}^{n} \left[(Z_0)_{ij} - 2\mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \notin E_0^\circ\big)\right]_+^2 \\
&\overset{(**)}{\geq} \sum_{i,j=1}^{n} (Z_0)_{ij}^2/2 - 4n^2\mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \notin E_0^\circ\big)^2 \\
&= p/2 - 4n^2\mathbb{P}_{\mu_0}\big(\widehat{\theta}^0 \notin E_0^\circ\big)^2.
\end{aligned}
$$

Here we have used the following:

- In $(*)$, we apply the estimate

$$\pm Z_{ij} = \pm e_i^\top Z e_j \leq \sup_{u,v:\|u\|=\|v\|=1} u^\top Z v = \|Z\| \leq 1, \quad Z \in \{Z_0, Z_{m,\ell}\}.$$

  This means $\max_{i,j}|Z_{ij}| \leq 1$ for $Z \in \{Z_0, Z_{m,\ell}\}$.

- In $(**)$ we use the estimate $(a-b)_+^2 \geq a^2/2 - b^2$.

- In the last equality we use $\sum_{i,j}(Z_0)_{ij}^2 = \mathrm{Tr}(Z_0 Z_0^\top) = p$.

Thus, claim (1) follows.

(2). Note that

$$\begin{aligned}
\mathbb{E}_\mu \operatorname{div} \widehat{\mu} &= \mathbb{E}_\mu \mathrm{Tr}(J_{\widehat{\mu}}) \\
&= \mathbb{P}_\mu\big(\widehat{\theta}^0 \in E_0^\circ\big)\mathrm{Tr}(Z_0) + \sum_{m=1}^{p}\sum_{\ell=1}^{N_m} \mathbb{P}_\mu\big(\widehat{\theta}^0 \in E_{m,\ell}^\circ\big)\mathrm{Tr}(Z_{m,\ell}) \\
&= p - p\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big) + \sum_{m=1}^{p}\sum_{\ell=1}^{N_m} \mathbb{P}_\mu\big(\widehat{\theta}^0 \in E_{m,\ell}^\circ\big)(p-m).
\end{aligned}$$

Hence

$$\big|\mathbb{E}_\mu \operatorname{div} \widehat{\mu} - p\big| \le 2p\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big),$$

proving claim (2).

(3). When $\widehat{\theta}^0 \in E_0^\circ$, $\widehat{\theta} = \widehat{\theta}^0$ as $\widehat{\theta}$ is the projection of $\widehat{\theta}^0$ onto $\widetilde{K}$ with respect to $\|\cdot\|_X \equiv \big(\cdot^\top X^\top X\,\cdot\big)^{1/2}$, cf. (Kato, 2009, Equation (1.6)). This means

$$\begin{aligned}
\mathbb{E}_\mu\|\widehat{\mu} - \mu\|^2 &= \mathbb{E}_\mu\big(\|\widehat{\mu} - \mu\|^2 \mathbf{1}_{\widehat{\theta}^0 \in E_0^\circ}\big) + \mathbb{E}_\mu\big(\|\widehat{\mu} - \mu\|^2 \mathbf{1}_{\widehat{\theta}^0 \notin E_0^\circ}\big) \\
&= \mathbb{E}_\mu\big(\|X\widehat{\theta}^0 - X\theta\|^2 \mathbf{1}_{\widehat{\theta}^0 \in E_0^\circ}\big) + \mathbb{E}_\mu\big(\|\widehat{\mu} - \mu\|^2 \mathbf{1}_{\widehat{\theta}^0 \notin E_0^\circ}\big) \\
&= \mathbb{E}_\mu\|X\widehat{\theta}^0 - X\theta\|^2 + R_{n,\mu} = p + R_{n,\mu},
\end{aligned}$$

where

$$R_{n,\mu} \equiv \mathbb{E}_\mu\big(\|\widehat{\mu} - \mu\|^2 \mathbf{1}_{\widehat{\theta}^0 \notin E_0^\circ}\big) - \mathbb{E}_\mu\big(\|Z_0\xi\|^2 \mathbf{1}_{\widehat{\theta}^0 \notin E_0^\circ}\big).$$

As $\|\theta\|_1 \le \lambda$, $\|\widehat{\mu} - \mu\|^2 \le 2\|Y - \widehat{\mu}\|^2 + 2\|\xi\|^2 \le 4\|\xi\|^2$ by using $\|Y - \widehat{\mu}\|^2 \le \|Y - \mu\|^2 = \|\xi\|^2$ via the definition of projection, we have

$$\mathbb{E}_\mu\big(\|\widehat{\mu} - \mu\|^2 \mathbf{1}_{\widehat{\theta}^0 \notin E_0^\circ}\big) \le 4\sqrt{\mathbb{E}\|\xi\|^4}\sqrt{\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big)} \le Cn\sqrt{\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big)}.$$

On the other hand, a similar estimate yields

$$\mathbb{E}_\mu\big(\|Z_0\xi\|^2 \mathbf{1}_{\widehat{\theta}^0 \notin E_0^\circ}\big) \le Cp\sqrt{\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big)} \le Cn\sqrt{\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big)}.$$

Hence

$$\big|\mathbb{E}_\mu\|\widehat{\mu}-\mu\|^2 - p\big| \leq Cn\sqrt{\mathbb{P}_\mu\big(\widehat{\theta}^0 \notin E_0^\circ\big)}.$$

This completes the proof of claim (3). □

Proof of Theorem 11. The first claim follows from Proposition 20-(1)(3) and Theorem 4 (by ignoring the bias term in the denominator). For the second claim, by Proposition 20-(2)(3),

$$m_\mu - \big(\|\mu-\mu_0\|^2 + p\big) = 2\mathbb{E}_\mu \operatorname{div} \widehat{\mu}_{K_{X,\lambda}} - \mathbb{E}_\mu\|\widehat{\mu}_{K_{X,\lambda}} - \mu\|^2 - p = \mathcal{O}(n\cdot \mathfrak{p}_{\lambda,\mu}^{1/2}).$$

This entails that $m_\mu - m_{\mu_0} = \|\mu-\mu_0\|^2 + n\cdot\mathcal{O}(\mathfrak{p}_{\lambda,\mu}^{1/2}\vee \mathfrak{p}_{\lambda,\mu_0}^{1/2})$. Furthermore, using Gaussian-Poincaré inequality along with Proposition 20-(1)(3), we have

$$\sigma_{\mu_0}^2 \leq 4\mathbb{E}_{\mu_0}\|\widehat{\mu}-\mu_0\|^2 \leq 4p + C(n\mathfrak{p}_{\lambda,\mu_0}^{1/2}) = \mathcal{O}(p),$$

where the last inequality follows from the condition $n\mathfrak{p}_{\lambda,\mu_0}^{1/2} = \mathfrak{o}(1)$. This, along with lower bound for $\sigma_{\mu_0}^2$ derived in Proposition 20-(1), yields that $\sigma_{\mu_0}^2 \asymp p$. Therefore, under the condition $n\cdot(\mathfrak{p}_{\lambda,\mu}^{1/2}\vee\mathfrak{p}_{\lambda,\mu_0}^{1/2}) = \mathfrak{o}(1)$, (4.19) is satisfied automatically, and (4.21) is equivalent to

$$\left|\frac{n\cdot\mathcal{O}\big(\mathfrak{p}_{\lambda,\mu}^{1/2}\vee\mathfrak{p}_{\lambda,\mu_0}^{1/2}\big) + \|\mu-\mu_0\|^2}{p^{1/2}}\right| \to\infty \Leftrightarrow \|\mu-\mu_0\| \gg p^{1/4}.$$

The proof is complete. □

Proof of Lemma 48. Recall that $\Sigma = X^\top X/n$. Note that

$$\widehat{\theta}^0 = (X^\top X)^{-1}X^\top Y = \theta + (X^\top X)^{-1}X^\top\xi \stackrel{d}{=} \theta + (X^\top X)^{-1/2}Z$$

with $Z\sim\mathcal{N}(0,I_p)$. For any $b\in\mathbb{R}^p$, let $f_b:\mathbb{R}^p\to\mathbb{R}$ be defined as $f_b(y)\equiv f_b(y;X)\equiv \langle (X^\top X)^{-1/2}y, b\rangle$. Then $\|\widehat{\theta}^0\|_1 = \sup_{b:\|b\|_\infty\leq 1}\big[f_b(Z)+b^\top\theta\big]$. Hence by Gaussian concentration (cf. (Boucheron et al., 2013, Theorem 5.8)), for any $t>0$,

$$\mathbb{P}\Big(\|\widehat{\theta}^0\|_1 - \mathbb{E}\|\widehat{\theta}^0\|_1 > t\Big) \leq \exp\big(-t^2/2\sigma^2\big), \tag{C.19}$$

where $\sigma^2 = \sup_{b:\|b\|_\infty \le 1} \mathrm{Var}\big(f_b(Z)\big)$. Next we bound $\mathbb{E}\|\widehat{\theta}^0\|_1$ and σ^2. For σ^2, note that

$$\begin{aligned}\sigma^2 &= \sup_{b:\|b\|_\infty \le 1} \mathbb{E}\langle (X^\top X)^{-1/2} Z, b\rangle^2 = n^{-1} \sup_{b:\|b\|_\infty \le 1} b^\top \Sigma^{-1} b \\ &\le (p/n) \cdot \sup_{b:\|b\|_2 \le 1} b^\top \Sigma^{-1} b = p/\big(n\lambda_{\min}(\Sigma)\big).\end{aligned}$$

For the mean term, we have $\mathbb{E}\|\widehat{\theta}^0\|_1 \le \|\theta\|_1 + \mathbb{E}\|(X^\top X)^{-1/2}\xi\|_1 = \|\theta\|_1 + \mathbb{E}\sup_{b:\|b\|_\infty \le 1} f_b(Z)$. Note that the natural metric d induced by the Gaussian process $\big(f_b(Z) : b \in \mathbb{R}^p\big)$ takes the form

$$\begin{aligned}d^2(b_1, b_2) &\equiv \mathbb{E}\big(f_{b_1}(Z) - f_{b_2}(Z)\big)^2 \\ &= (b_1 - b_2)^\top (n\Sigma)^{-1}(b_1 - b_2) \le n^{-1}\lambda_{\min}^{-1}(\Sigma)\|b_1 - b_2\|^2,\end{aligned}$$

and a simple volume estimate yields that

$$\mathcal{N}(\varepsilon, \{b : \|b\|_\infty \le 1\}, d) \lesssim \big[(n\lambda_{\min}(\Sigma))^{-1/2}/\varepsilon\big]^p.$$

Hence by Dudley's entropy integral (cf. (Giné and Nickl, 2016, Theorem 2.3.6)),

$$\begin{aligned}\mathbb{E}\|(X^\top X)^{-1/2}\xi\|_1 &\lesssim \int_0^\infty \sqrt{\log\big(1 \vee \mathcal{N}(\varepsilon, \{b : \|b\|_\infty \le 1\}, d)\big)}\, \mathrm{d}\varepsilon \\ &\lesssim \sqrt{p/\big(n\lambda_{\min}(\Sigma)\big)}.\end{aligned}$$

The claim now follows from (C.19). □

C.2.5 Proof of Theorem 12

By definition of $K_{0,k}$, we have $\delta_{K_{0,k}} = \dim(K_{0,k}) = k + 1$. We will now show that

$$L_k^{-1}\big(\mathbf{1}_{k\ge 1}\log\log(16n) + \mathbf{1}_{k=0}\log(en)\big) \le \delta_{K_{\uparrow,k}} \le L_k \log(en), \tag{C.20}$$

where $L_k > 0$ only depends on k.

We first prove the upper bound in (C.20) by induction. The baseline case $k = 0$ follows by (Amelunxen et al., 2014, Equation (D.12)). Suppose the claim holds for some $k \in \mathbb{Z}_{\ge 0}$.

For $k+1$, note that $K_{\uparrow,k+1} = \cup_{\ell=1}^n K_{\uparrow,k+1;\ell}$ where $K_{\uparrow,k+1;\ell}$ contains all $\nu \in K_{\uparrow,k+1}$ such that $-\nu|_{[1:\ell]}$ is k-monotone, and $\nu|_{(\ell:n]}$ is k-monotone. Hence for any $\ell \in [1:n]$, it follows by (Amelunxen et al., 2014, Proposition 3.1) that

$$\delta_{K_{\uparrow,k+1;\ell}} \leq L_k\big(\log(e\ell) + \log(e(n-\ell))\big) \leq 2L_k \log(en),$$

where the second inequality follows by induction. On the other hand, let $Z_k \equiv \sup_{\nu \in K_{\uparrow,k} \cap B(1)} \langle \nu, \xi \rangle = \|\Pi_{K_{\uparrow,k}}(\xi)\|$, then Gaussian concentration (cf. (Boucheron et al., 2013, Theorem 5.8)) entails that for any $t > 0$,

$$\mathbb{P}(Z_k \geq \mathbb{E}Z_k + t) \leq \exp(-t^2/2).$$

Hence, using the induction hypothesis $\mathbb{E}Z_k \leq \big(\mathbb{E}Z_k^2\big)^{1/2} \leq L_k^{1/2}\sqrt{\log(en)}$ and the union bound, it holds w.p. at least $1 - \exp(-t)$ that

$$\begin{aligned} Z_{k+1} \equiv \sup_{\nu \in K_{\uparrow,k+1} \cap B(1)} \langle \nu, \xi \rangle &\leq \max_{1 \leq \ell \leq n} \delta^{1/2}_{K_{\uparrow,k+1;\ell}} + \sqrt{2(t + \log(en))} \\ &\leq \big(2L_k \log(en)\big)^{1/2} + \sqrt{2(t + \log(en))}. \end{aligned}$$

Now the bound for $\delta_{K_{\uparrow,k+1}} = \mathbb{E}Z_{k+1}^2$ follows by integrating the tail.

Next we prove the lower bound in (C.20) for $k \geq 1$. By Sudakov's minorization (cf. (Giné and Nickl, 2016, Theorem 2.4.12)), we have

$$\begin{aligned} \delta^{1/2}_{K_{\uparrow,k+1}} \geq \mathbb{E}Z_{k+1} &\gtrsim \sup_{\varepsilon > 0} \varepsilon \sqrt{\log \mathcal{N}(\varepsilon, K_{\uparrow,k+1} \cap B(1), \|\cdot\|)} \\ &\geq \sup_{\varepsilon > 0} \varepsilon \sqrt{\log \mathcal{D}(2\varepsilon, K_{\uparrow,k+1} \cap B(1), \|\cdot\|)}, \end{aligned}$$

where $\mathcal{D}(\varepsilon, T, d)$ is the maximal ε-packing number of set T with respect to the metric d. By taking ε to be small enough, the construction in (Shen et al., 2020, Theorem 3.4) yields an (2ε)-packing set of cardinality of the order $\log(en)$. This completes the lower bound proof.

Now the claim (1) follows from Theorem 7-(1) and the lower bound in (C.20). (2) follows from the upper bound in (C.20) and Theorem 7-(2). $\square$

C.3 Additional proofs

C.3.1 Proof of Lemma 2

We provide the proof for (1)-(2) assuming K is a polyhedral cone. The claim for a general convex cone K follows from polyhedral approximation (McCoy and Tropp, 2014, Section 7.3).

(1) As $V_K \stackrel{d}{=} \operatorname{div}\Pi_K(\xi) = \operatorname{Tr}\big(J_{\Pi_K}(\xi)\big)$, we have $\mathbb{E}V_K = \mathbb{E}\operatorname{div}\Pi_K(\xi) = \mathbb{E}\langle\xi,\Pi_K(\xi)\rangle = \mathbb{E}\|\Pi_K(\xi)\|^2 = \delta_K$.

(2) The claim is proved in (McCoy and Tropp, 2014, Proposition 4.4) using the 'Master Steiner formula', cf. (McCoy and Tropp, 2014, Theorem 3.1), which is a restatement of the chi-bar squared distribution. Below we provide a simple alternative proof of this claim using Gaussian integration-by-parts only.

By expanding $V_K \stackrel{d}{=} \langle\xi,\Pi_K(\xi)\rangle - \big(\langle\xi,\Pi_K(\xi)\rangle - \operatorname{div}\Pi_K(\xi)\big)$ and noting that $\mathbb{E}\langle\xi,\Pi_K(\xi)\rangle = \mathbb{E}\operatorname{div}\Pi_K(\xi)$,

$$
\begin{aligned}
\operatorname{Var}(V_K) &= \operatorname{Var}\big(\langle\xi,\Pi_K(\xi)\rangle - \operatorname{div}\Pi_K(\xi)\big) + \operatorname{Var}\big(\langle\xi,\Pi_K(\xi)\rangle\big)\\
&\qquad - 2\mathbb{E}\big[\big(\langle\xi,\Pi_K(\xi)\rangle - \operatorname{div}\Pi_K(\xi)\big)\langle\xi,\Pi_K(\xi)\rangle\big]\\
&= \mathbb{E}\operatorname{Tr}J^2_{\Pi_K}(\xi) + \mathbb{E}\|\Pi_K(\xi)\|^2 + \operatorname{Var}\big(\langle\xi,\Pi_K(\xi)\rangle\big)\\
&\qquad - 2\mathbb{E}\big[\langle\Pi_K(\xi),\nabla\langle\xi,\Pi_K(\xi)\rangle\rangle\big].
\end{aligned}
$$

The last equality follows from Gaussian integration-by-parts: (i) $\operatorname{Var}(\langle\xi,f(\xi)\rangle - \operatorname{div} f(\xi)) = \mathbb{E}\operatorname{Tr}J_f^2(\xi) + \mathbb{E}\|f(\xi)\|^2$ (see e.g., (Stein, 1981, Theorem 3), or (Bellec and Zhang, 2018, Theorem 2.1)) and (ii) $\mathbb{E}[(\langle\xi,f(\xi)\rangle - \operatorname{div} f(\xi))g(\xi)] = \mathbb{E}[\langle f(\xi),\nabla g(\xi)\rangle]$ (Bellec and Zhang, 2018, Equation (2.4)). Note that (i) $\nabla\langle\xi,\Pi_K(\xi)\rangle = \nabla\|\Pi_K(\xi)\|^2 = \nabla\|\xi-\Pi_{K^*}(\xi)\|^2 = 2(\xi-\Pi_{K^*}(\xi)) = 2\Pi_K(\xi)$ using the fact K is a cone and Lemma 1-(1), and (ii) $\mathbb{E}\operatorname{Tr}J^2_{\Pi_K}(\xi) = \mathbb{E}\operatorname{Tr}J_{\Pi_K}(\xi) = \mathbb{E}\operatorname{div}\Pi_K(\xi)$ using the fact that when K is polyhedral, J_{Π_K} is a.e. a projection matrix (cf. (Kato, 2009, Remark 3.3)). Finally using that $\mathbb{E}\operatorname{div}\Pi_K(\xi) = \mathbb{E}\langle\xi,\Pi_K(\xi)\rangle = \mathbb{E}\|\Pi_K(\xi)\|^2$ to conclude.

(3) The right inequality follows by an application of the improved Gaussian-Poincaré inequality stated in (Goldstein et al., 2017, Theorem A.2) as follows: By Lemma 1-(1) again, $\nabla\|\Pi_K(\xi)\|^2 = \nabla\|\xi - \Pi_{K^*}(\xi)\|^2 = 2(\xi - \Pi_{K^*}(\xi)) = 2\Pi_K(\xi)$, so (Goldstein et al., 2017, Theorem A.2) yields that

$$\begin{aligned}\mathrm{Var}(\|\Pi_K(\xi)\|^2) &\leq \frac{1}{2}\mathbb{E}\big\|\nabla\|\Pi_K(\xi)\|^2\big\|^2 + \frac{1}{2}\big\|\mathbb{E}\nabla\|\Pi_K(\xi)\|^2\big\|^2\\ &= 2\mathbb{E}\|\Pi_K(\xi)\|^2 + 2\|\mathbb{E}\Pi_K(\xi)\|^2.\end{aligned}$$

The left inequality is an immediately consequence of (2). $\square$

C.3.2 Proof of Proposition 29

Recall the following second-order Poincaré inequality due to Chatterjee (2009).

Lemma 49 (Second-order Poincaré inequality)**.** *Let ξ be an n-dimensional standard normal random vector. Let $F : \mathbb{R}^n \to \mathbb{R}$ be absolute continuous such that F and its derivatives have sub-exponential growth at ∞. Let ξ' be an independent copy of ξ. Define $T : \mathbb{R}^n \to \mathbb{R}$ by*

$$T(y) \equiv \int_0^1 \frac{1}{2\sqrt{t}}\langle \nabla F(y), \mathbb{E}_{\xi'}\nabla F(\sqrt{t}y + \sqrt{1-t}\xi')\rangle\,\mathrm{d}t.$$

Then with $W \equiv F(\xi)$,

$$d_{\mathrm{TV}}\bigg(\frac{W - \mathbb{E}W}{\sqrt{\mathrm{Var}(W)}}, \mathcal{N}(0,1)\bigg) \leq \frac{2\sqrt{\mathrm{Var}(T(\xi))}}{\mathrm{Var}(W)}.$$

Proof. Let $W' \equiv \big(F(\xi) - \mathbb{E}F(\xi)\big)/\sqrt{\mathrm{Var}(F(\xi))}$, and $T' \equiv T/\mathrm{Var}(F(\xi))$. Then (Chatterjee, 2009, Lemma 5.3) says that

$$d_{\mathrm{TV}}(W', \mathcal{N}(0,1)) \leq 2\sqrt{\mathrm{Var}(T'(\xi))} = 2\sqrt{\mathrm{Var}(T(\xi))}/\mathrm{Var}(F(\xi)).$$

The claim follows by the invariance of the total variation metric by translation and scaling.

$\square$

Proof of Proposition 29. For any fixed $\mu \in \mathbb{R}^n$, let $F(\xi) \equiv F_\mu(\xi) \equiv \|\mu + \xi - \Pi_{K_0}(\mu+\xi)\|^2 - \|\mu + \xi - \Pi_K(\mu+\xi)\|^2$. By Lemma 1-(1), we have

$$
\begin{aligned}
\nabla F(\xi) &= 2(\mu + \xi - \Pi_{K_0}(\mu+\xi)) - 2(\mu + \xi - \Pi_K(\mu+\xi)) \\
&= 2\big(\Pi_K(\mu+\xi) - \Pi_{K_0}(\mu+\xi)\big).
\end{aligned}
$$

To use the second-order Poincaré inequality, let ξ' be an independent copy of ξ and $\xi_t \equiv \sqrt{t}\xi + \sqrt{1-t}\xi'$, and let

$$
\begin{aligned}
T(\xi) &\equiv \int_0^1 \frac{1}{2\sqrt{t}} \langle \nabla F(\xi), \mathbb{E}_{\xi'} \nabla F(\xi_t)\rangle \, \mathrm{d}t \\
&= 4\mathbb{E}_{\xi'} \int_0^1 \frac{1}{2\sqrt{t}} \langle \Pi_K(\mu+\xi) - \Pi_{K_0}(\mu+\xi), \Pi_K(\mu+\xi_t) - \Pi_{K_0}(\mu+\xi_t)\rangle \, \mathrm{d}t.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\nabla T(\xi) = 4\mathbb{E}_{\xi'} \int \frac{1}{2\sqrt{t}} \Big[&(J_{\Pi_K} - J_{\Pi_{K_0}})(\mu+\xi)^\top \big(\Pi_K(\mu+\xi_t) - \Pi_{K_0}(\mu+\xi_t)\big) \\
&+ \sqrt{t}(J_{\Pi_K} - J_{\Pi_{K_0}})(\mu+\xi_t)^\top \big(\Pi_K(\mu+\xi) - \Pi_{K_0}(\mu+\xi)\big)\Big] \, \mathrm{d}t.
\end{aligned}
$$

The terms involved in the integral in T are all absolute continuous, so we may continue to use Gaussian-Poincaré inequality:

$$
\begin{aligned}
&\mathrm{Var}(T(\xi)) \leq \mathbb{E}\|\nabla T(\xi)\|^2 \\
&\leq 16 \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E}\Big\| (J_{\Pi_K} - J_{\Pi_{K_0}})(\mu+\xi)^\top \big(\Pi_K(\mu+\xi_t) - \Pi_{K_0}(\mu+\xi_t)\big) \\
&\qquad + \sqrt{t}(J_{\Pi_K} - J_{\Pi_{K_0}})(\mu+\xi_t)^\top \big(\Pi_K(\mu+\xi) - \Pi_{K_0}(\mu+\xi)\big)\Big\|^2 \, \mathrm{d}t \\
&\qquad\qquad \text{(by Jensen's inequality applied to the measure } \mathrm{d}t/2\sqrt{t}) \\
&\leq 16 \times 8 \int_0^1 \frac{1}{2\sqrt{t}} \Big(\mathbb{E}\|\Pi_K(\mu+\xi_t) - \Pi_{K_0}(\mu+\xi_t)\|^2 \\
&\qquad + \mathbb{E}\|\Pi_K(\mu+\xi) - \Pi_{K_0}(\mu+\xi)\|^2 \Big) \, \mathrm{d}t.
\end{aligned}
$$

Here in the last inequality we used that $\|J_{\Pi_K} - J_{\Pi_{K_0}}\| \leq \|J_{\Pi_K}\| + \|J_{\Pi_{K_0}}\| \leq 2$. Now using that ξ_t has the same distribution as ξ for each $t \in [0,1]$, we arrive at

$$\mathrm{Var}(T(\xi)) \leq 16^2 \mathbb{E}\|\widehat{\mu}_K - \widehat{\mu}_{K_0}\|^2.$$

The claim now follows from the second-order Poincaré inequality in Lemma 58. $\square$

Appendix D

APPENDIX OF CHAPTER 5

D.1 Some spectral estimates

First we introduce some convention on notation: Let $\mathcal{I}_1$, $\mathcal{I}_2$ be finite index sets. For $A = (A_{\iota_1,\iota_2})_{\iota_1\in\mathcal{I}_1,\iota_2\in\mathcal{I}_2} \in \mathbb{R}^{\mathcal{I}_1\times\mathcal{I}_2}$, its operator norm is defined as

$$\|A\|_{\mathrm{op}} \equiv \sup_{v\in B_{\mathcal{I}_2}(1)} \|Av\|_{\ell_2(\mathbb{R}^{\mathcal{I}_2})}. \tag{D.1}$$

It can be readily verified that $\|A\|_{\mathrm{op}} = \sup_{u\in B_{\mathcal{I}_1},v\in B_{\mathcal{I}_2}} \langle u, Av\rangle_{\mathcal{I}_1}$, and for a symmetric matrix $A \in \mathbb{R}^{\mathcal{I}_1\times\mathcal{I}_1}$, $\|A\|_{\mathrm{op}} = \sup_{u\in B_{\mathcal{I}_1}} |\langle u, Au\rangle_{\mathcal{I}_1}|$. Here $\langle\cdot,\cdot\rangle_{\mathcal{I}_1}$ is the standard inner product on $\mathbb{R}^{\mathcal{I}_1}$. Clearly, the definition of the operator norm does not depend on the choice of the ordering of the index sets.

Under this notational convention, with the index set $\Lambda \equiv \{(ij) : i \in [N], j \in [p]\}$, we present below two results on the spectral norm of some special $\Lambda\times\Lambda$ matrices that are crucial to the proof of the quantitative central limit theorems. We do not specify a particular ordering on Λ as we will be only interested in the operator norm as defined above. In the following we use $\mathbb{N}$ to denote the set of natural numbers. Recall the data matrix $X = [X_1,\ldots,X_N]^\top \in \mathbb{R}^{N\times p}$ and the definition of S in (5.16).

Proposition 31. *1. Suppose $p/N \le 1-\varepsilon$ for some $\varepsilon > 0$. For $\ell, m \in \mathbb{N}$ such that $\ell + m \ge 1$, let $U_{\ell,m} \in \mathbb{R}^{\Lambda\times\Lambda}$ be defined by*

$$(U_{\ell,m})_{(ij),(i'j')} \equiv N^{-1} X_i^\top S^{-\ell} X_{i'} (S^{-m})_{jj'}. \tag{D.2}$$

Then for any $q \in \mathbb{N}$, there exists some $C = C(\varepsilon,\ell,m,q) > 0$ such that $\mathbb{E}\|U_{\ell,m}\|_{\mathrm{op}}^q \le C$ for $p \ge C$.

2. When X_i, S and N is replaced by $X_i - \overline{X}$, S_* and n, the conclusion of (1) still holds.

When the inverse S^{-1} in (D.2) is replaced by S, the condition on p/N can be substantially relaxed.

Proposition 32. *Let $y \equiv p/N$. For $\ell, m \in \mathbb{N}$, let $U_{\ell,m;+} \in \mathbb{R}^{\Lambda\times\Lambda}$ be defined by*

$$(U_{\ell,m;+})_{(ij),(i'j')} \equiv N^{-1} X_i^\top S^\ell X_{i'} (S^m)_{jj'}. \tag{D.3}$$

Then for any $q \in \mathbb{N}$, there exists some $C = C(\ell, m, q) > 0$ such that $\mathbb{E}\|U_{\ell,m;+}\|_{\mathrm{op}}^q \leq C(\sqrt{y} \vee y)^{q(\ell+m+1)}$.

The proof of Proposition 31 relies crucially on the following stable moment estimate for $\|S^{-1}\|_{\mathrm{op}}$. Its proof utilizes two main technical tools: (i) rigidity estimates on the eigenvalues of the sample covariance matrix (cf. Pillai and Yin (2014)); (ii) closed form distributional formula of sample eigenvalues via zonal polynomials (Muirhead, 1982, Chapter 9.7). Details are presented in Appendix D.10.

Lemma 50. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0, I)$ in $\mathbb{R}^p$. Suppose $p/N \leq 1-\varepsilon$ for some fixed $\varepsilon > 0$ and every $N, p \geq 2$. Then for any positive integer $q \leq (N-p-1)/8$, we have $\mathbb{E}\|S_Z^{-1}\|_{\mathrm{op}}^q \leq C$ for some positive $C = C(\varepsilon, q)$.*

The following corollary of the Koltchinskii-Lounici theorem Koltchinskii and Lounici (2017) will also be repeatedly used.

Lemma 51. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0, I)$ in $\mathbb{R}^p$. Then for any positive integer q, there exists some positive $C = C(q)$ such that*

$$\mathbb{E}\|S_Z - I\|_{\mathrm{op}}^q \leq C \cdot \Big(\sqrt{\frac{p}{N}} \bigvee \frac{p}{N}\Big)^q.$$

Proof. This is a direct consequence of (Koltchinskii and Lounici, 2017, Corollary 2). $\square$

Now we prove Propositions 31 and 32.

Proof of Proposition 31. We only prove (1); claim (2) follows from completely same arguments by noting that (D.4) and (D.6) below still hold with the prescribed substitution. Note that $U_{\ell,m}$ is symmetric in that $(U_{\ell,m})_{(ij)(i'j')} = (U_{\ell,m})_{(i'j')(ij)}$, and satisfies that for any non-negative integers (ℓ_1, m_1) and (ℓ_2, m_2) such that $(\ell_1 + m_1) \wedge (\ell_2 + m_2) \geq 1$,

$$\begin{aligned}
&(U_{\ell_1,m_1} \cdot U_{\ell_2,m_2})_{(ij),(i'j')} \\
&= N^{-2} \sum_{(\bar{i}\bar{j})} X_i^\top S^{-\ell_1} X_{\bar{i}} (S^{-m_1})_{j\bar{j}} X_{\bar{i}}^\top S^{-\ell_2} X_{i'} (S^{-m_2})_{\bar{j}j'} \\
&= N^{-2} X_i^\top S^{-\ell_1} \Big(\sum_{\bar{i}} X_{\bar{i}} X_{\bar{i}}^\top \Big) S^{-\ell_2} X_{i'} \cdot \Big(\sum_{\bar{j}} (S^{-m_1})_{j\bar{j}} (S^{-m_2})_{\bar{j}j'} \Big) \\
&= N^{-1} X_i^\top S^{-(\ell_1+\ell_2-1)} X_{i'} (S^{-(m_1+m_2)})_{jj'} \\
&= (U_{\ell_1+\ell_2-1,m_1+m_2})_{(ij),(i'j')}.
\end{aligned} \tag{D.4}$$

Consequently the above argument entails that for any $q \in \mathbb{N}$, $(U_{\ell,m})^q = U_{\ell',m'}$ with

$$\ell' \equiv \ell'(q) \equiv q(\ell - 1) + 1, \quad m' \equiv m'(q) \equiv qm. \tag{D.5}$$

Using that $\|U_{\ell,m}\|_{\mathrm{op}} = \sup_{u \in B_{N\times p}(1)} \big|\sum_{(ij)(i'j')} u_{ij} (U_{\ell,m})_{(ij),(i'j')} u_{i'j'}\big|$, we have

$$\begin{aligned}
\|U_{\ell,m}\|_{\mathrm{op}} &= \sup_{u \in B_{N\times p}(1)} N^{-1} \Big| \sum_{i,i',j,j'} X_i^\top S^{-\ell} X_{i'} (S^{-m})_{jj'} u_{ij} u_{i'j'} \Big| \\
&= N^{-1} \sup_{u \in B_{N\times p}(1)} \Big| \sum_{i,i'} X_i^\top S^{-\ell} X_{i'} \cdot \sum_{jj'} u_{ij} \cdot (S^{-m})_{jj'} u_{i'j'} \Big| \\
&= N^{-1} \sup_{u \in B_{N\times p}(1)} \Big| \sum_{i,i'} X_i^\top S^{-\ell} X_{i'} \cdot \Big[u \cdot S^{-m} \cdot u^\top \Big]_{ii'} \Big|.
\end{aligned}$$

As the (i,i')-th entry of $X^\top S^{-\ell} X$ is $X_i^\top S^{-\ell} X_{i'}$ and that $\mathrm{Tr}(AB) = \sum_{i,i'=1}^N A_{ii'} B_{ii'}$ for two symmetric matrices in $\mathbb{R}^{N\times N}$, we have

$$\begin{aligned}
\|U_{\ell,m}\|_{\mathrm{op}} &= N^{-1} \sup_{u \in B_{N\times p}(1)} \big| \mathrm{Tr}\big[X S^{-\ell} X^\top \cdot u S^{-m} u^\top \big] \big| \\
&= N^{-1} \sup_{u \in B_{N\times p}(1)} \big| \mathrm{Tr}\big[(u^\top X) S^{-\ell} (X^\top u) \cdot S^{-m} \big] \big|.
\end{aligned}$$

Further using twice the fact that $\mathrm{Tr}(AB) \le \mathrm{Tr}(A)\|B\|_{\mathrm{op}}$ for any two p.s.d. and symmetric matrices A, B, we arrive at

$$
\begin{aligned}
\|U_{\ell,m}\|_{\mathrm{op}} &\le N^{-1}\|S^{-m}\|_{\mathrm{op}}\|S^{-\ell}\|_{\mathrm{op}} \cdot \sup_{u \in B_{N\times p}(1)} \big|\mathrm{Tr}\big(XX^\top uu^\top\big)\big| \\
&\le \|S^{-(\ell+m)}\|_{\mathrm{op}} \cdot N^{-1}\|XX^\top\|_{\mathrm{op}} \cdot \sup_{u \in B_{N\times p}(1)} \mathrm{Tr}(uu^\top) \\
&= \|S^{-(\ell+m)}\|_{\mathrm{op}} \cdot \|S\|_{\mathrm{op}} = \|S^{-1}\|_{\mathrm{op}}^{\ell+m-1}. \qquad \text{(D.6)}
\end{aligned}
$$

Hence for any $q \in \mathbb{N}$, equations (D.4)-(D.6) and Lemma 50 entail that

$$
\begin{aligned}
\mathbb{E}\|U_{\ell,m}\|_{\mathrm{op}}^q &= \mathbb{E}\|U_{\ell,m}^q\|_{\mathrm{op}} = \mathbb{E}\|U_{\ell',m'}\|_{\mathrm{op}} \le \mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{\ell'+m'-1} \\
&= \mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q(\ell+m-1)} \le C_{\ell,m,q},
\end{aligned}
$$

completing the proof. □

Proof of Proposition 32. The proof largely follows that of Proposition 31 with modifications. We sketch the difference below. Using the same calculations as in (D.4), we have

$$
\begin{aligned}
&(U_{\ell_1,m_1;+} \cdot U_{\ell_2,m_2;+})_{(ij),(i'j')} \\
&= N^{-2} \sum_{(\bar{i}\bar{j})} X_i^\top S^{\ell_1} X_{\bar{i}} (S^{m_1})_{j\bar{j}} X_{\bar{i}}^\top S^{\ell_2} X_{i'} (S^{m_2})_{\bar{j}j'} \\
&= N^{-2} X_i^\top S^{\ell_1} \bigg(\sum_{\bar{i}} X_{\bar{i}} X_{\bar{i}}^\top\bigg) S^{\ell_2} X_{i'} \cdot \bigg(\sum_{\bar{j}} (S^{m_1})_{j\bar{j}} (S^{m_2})_{\bar{j}j'}\bigg) \\
&= N^{-1} X_i^\top S^{\ell_1+\ell_2+1} X_{i'} (S^{m_1+m_2})_{jj'} = (U_{\ell_1+\ell_2+1,m_1+m_2;+})_{(ij),(i'j')}.
\end{aligned}
$$

Hence for any $q \in \mathbb{N}$, $(U_{\ell,m;+})^q = U_{\ell',m';+}$ with ℓ' now defined by $\ell' \equiv \ell'(q) \equiv \ell + (q-1)(\ell+1)$ and $m' = qm$ remains the same as in (D.5).

Then using the same arguments as in (D.6), we have $\|U_{\ell,m;+}\|_{\mathrm{op}} \le \|S\|_{\mathrm{op}}^{\ell+m+1}$, hence for any $q \in \mathbb{N}$,

$$
\mathbb{E}\|U_{\ell,m;+}\|_{\mathrm{op}}^q \le \mathbb{E}\|S\|_{\mathrm{op}}^{\ell'+m'+1} \lesssim_{\ell,m,q} (\sqrt{y} \vee y)^{q(\ell+m+1)},
$$

where the last inequality follows by Lemma 51 and the fact that $\ell'+m'+1=\ell+(q-1)(\ell+1)+qm+1=q(\ell+m+1)$. □

D.2 Proofs for Section 5.3.1

Recall $\Lambda=\{(ij):i\in[N],j\in[p]\}$. In the following sections, for a sufficiently smooth function $T:\mathbb{R}^\Lambda\to\mathbb{R}$, its gradient $\nabla T:\mathbb{R}^\Lambda\to\mathbb{R}^\Lambda$ and Hessian $\nabla^2 T:\mathbb{R}^\Lambda\to\mathbb{R}^{\Lambda\times\Lambda}$ are defined respectively by

$$\big(\nabla T(x)\big)_{(ij)}\equiv\frac{\partial T}{\partial x_{(ij)}}(x)\quad\text{and}\quad\big(\nabla^2 T(x)\big)_{(ij),(i'j')}\equiv\frac{\partial^2 T}{\partial x_{(ij)}\partial x_{(i'j')}}(x),$$

with $x=(x_{(ij)})\in\mathbb{R}^\Lambda$. Slightly abusing notation, we using $\|\nabla T(x)\|_F\equiv\|\nabla T(x)\|_{\ell_2(\mathbb{R}^\Lambda)}$. The operator norm $\|\nabla^2 T(x)\|_{\mathrm{op}}$ is defined in (D.1).

D.2.1 Evaluation of derivatives

In the following, we use $\{e_j\}_{j=1}^p$ to represent the canonical basis in $\mathbb{R}^p$. Let δ_{ij} be the Kronecker delta.

Lemma 52. *Recall the form of $T_{\mathrm{LRT}}(X)$ in (5.19). We assume without loss of generality that $\mu=0$. Then for any $(i,j),(i',j')\in[N]\times[p]$,*

1. $\big(\nabla T_{\mathrm{LRT}}(X)\big)_{(ij)}=\big(X(I-S^{-1})\big)_{(ij)}=e_j^\top(I-S^{-1})X_i.$

2. $\big(\nabla^2 T_{\mathrm{LRT}}(X)\big)_{(ij),(i'j')}=N^{-1}X_i^\top S^{-1}(e_{j'}X_{i'}^\top+X_{i'}e_{j'}^\top)S^{-1}e_j+\delta_{ii'}e_j^\top(I-S^{-1})e_{j'}.$

Proof. We use T as a shorthand for T_{LRT}.
(1). By definition, we have

$$\partial_{(ij)}T(X)=\frac{N}{2}\big(\partial_{(ij)}\mathrm{Tr}(S)-\partial_{(ij)}\log\det S\big).$$

For the first partial derivative, using $\partial_{(ij)} X_k = \delta_{ik} e_j$, we have

$$
\begin{aligned}
\partial_{(ij)} \mathrm{Tr}(S(X)) &= N^{-1} \sum_k \frac{\partial}{\partial X_{ij}} \mathrm{Tr}(X_k X_k^\top) \\
&= N^{-1} \sum_k \delta_{ik} \mathrm{Tr}\big[e_j X_k^\top + X_k e_j^\top\big] \\
&= N^{-1} \sum_k \delta_{ik} \cdot 2X_{kj} = 2N^{-1} X_{ij}. \qquad \text{(D.7)}
\end{aligned}
$$

For the second partial derivative, using the well-known fact that $\nabla \log\det A = A^{-1}$ for any invertible and symmetric matrix A (see e.g., (Boyd and Vandenberghe, 2004, Section A.4.1)), we have

$$
\begin{aligned}
\partial_{(ij)} \log\det S &= \sum_{k,\ell} \frac{\partial \log\det S}{\partial S_{k\ell}} \frac{\partial S_{k\ell}}{\partial X_{ij}} \\
&\overset{(*)}{=} \sum_{k,\ell} (S^{-1})_{k\ell} \cdot \frac{1}{N} (\delta_{jk} X_{i\ell} + \delta_{j\ell} X_{ik}) \\
&= \frac{1}{N} \sum_{\ell} (S^{-1})_{j\ell} X_{i\ell} + \sum_k (S^{-1})_{kj} X_{ik} = \frac{2}{N} (XS^{-1})_{ij}, \qquad \text{(D.8)}
\end{aligned}
$$

where in $(*)$, we use

$$
\frac{\partial S_{k\ell}}{\partial X_{ij}} = \frac{1}{N} \frac{\partial}{\partial X_{ij}} \langle X e_k, X e_\ell \rangle = \frac{1}{N} (\delta_{jk} X_{i\ell} + \delta_{j\ell} X_{ik}). \qquad \text{(D.9)}
$$

Combining (D.7) and (D.8) yields the first claim.

(2). Again by definition, we have

$$
\partial_{(ij)(i'j')} T(X) = \frac{N}{2} \big(\partial_{(ij)(i'j')} \mathrm{Tr}(S) - \partial_{(ij)(i'j')} \log\det S\big).
$$

For the first derivative, it follows from (D.7) that

$$
\partial_{(ij)(i'j')} \mathrm{Tr}(S) = 2N^{-1} \partial_{(i'j')} X_{ij} = 2N^{-1} \delta_{ii'} \delta_{jj'}. \qquad \text{(D.10)}
$$

For the second derivative, it follows from (D.8) that

$$\begin{aligned}
&\partial_{(ij)(i'j')} \log\det S \\
&= \frac{2}{N}\frac{\partial}{\partial X_{i'j'}} X_i^\top S^{-1} e_j = \frac{2}{N}\Big(\frac{\partial X_i}{\partial X_{i'j'}}^\top S^{-1} e_j + X_i^\top \frac{\partial S^{-1}}{\partial X_{i'j'}} e_j\Big) \\
&\overset{(**)}{=} \frac{2}{N}\Big(\delta_{ii'} e_{j'}^\top S^{-1} e_j - N^{-1} X_i^\top S^{-1}\big(e_{j'} X_{i'}^\top + X_{i'} e_{j'}^\top\big) S^{-1} e_j\Big), \qquad \text{(D.11)}
\end{aligned}$$

where in $(**)$ we use the following calculation with the help of (D.9):

$$\begin{aligned}
\frac{\partial S^{-1}}{\partial X_{i'j'}} &= -S^{-1}\frac{\partial S}{\partial X_{i'j'}} S^{-1} = -S^{-1}\Big(\sum_{k\ell} e_k e_\ell^\top \frac{\partial S_{k\ell}}{\partial X_{i'j'}}\Big) S^{-1} \\
&= -\frac{1}{N} S^{-1} \cdot \Big(\sum_{k\ell} e_k e_\ell^\top (\delta_{j'k} X_{i'\ell} + \delta_{j'\ell} X_{i'k})\Big) S^{-1} \\
&= -\frac{1}{N} S^{-1}\big(e_{j'} X_{i'}^\top + X_{i'} e_{j'}^\top\big) S^{-1}. \qquad \text{(D.12)}
\end{aligned}$$

We obtain the second claim by combining (D.10) and (D.11). □

D.2.2 Normal approximation

Proof of Theorem 14. We again shorthand $T_{\mathrm{LRT};\Sigma}$ by T. By Lemma 52,

$$\begin{aligned}
\|\nabla T(X)\|_F^2 = \sum_{i,j}\big(\partial_{(ij)} T(X)\big)^2 &= \sum_i \|(I - S^{-1}) X_i\|^2 \\
&\le \|S^{-1}\|_{\mathrm{op}}^2 \|I - S\|_{\mathrm{op}}^2 \sum_i \|X_i\|^2.
\end{aligned}$$

Using Lemma 50 and Lemma 51,

$$\begin{aligned}
\mathbb{E}\|\nabla T(X)\|_F^4 &\lesssim \mathbb{E}\Big(\|S^{-1}\|_{\mathrm{op}}^2 \|I - S\|_{\mathrm{op}}^2 \sum_i \|X_i\|^2\Big)^2 \\
&= \sum_{i,i'} \mathbb{E}\Big[\|S^{-1}\|_{\mathrm{op}}^4 \|I - S\|_{\mathrm{op}}^4 \|X_i\|^2 \|X_{i'}\|^2\Big] \\
&\le \sum_{i,i'} \mathbb{E}^{1/4}\|S^{-1}\|_{\mathrm{op}}^{16} \cdot \mathbb{E}^{1/4}\|I - S\|_{\mathrm{op}}^{16} \cdot \mathbb{E}^{1/4}\|X_i\|^8 \cdot \mathbb{E}^{1/4}\|X_{i'}\|^8 \\
&\lesssim N^2 \cdot \Big(\sqrt{\frac{p}{N}}\Big)^4 \cdot p \cdot p = p^4. \qquad \text{(D.13)}
\end{aligned}$$

Again by Lemma 52, the second derivatives are

$$\begin{aligned}\partial_{(ij),(i'j')}T(X) &= N^{-1}X_i^\top S^{-1}e_{j'}X_{i'}^\top S^{-1}e_j\\ &\qquad + N^{-1}X_i^\top S^{-1}X_{i'}(S^{-1})_{jj'} + \delta_{ii'}\big(I-S^{-1}\big)_{jj'}\\ &\equiv \big(T_1+T_2+T_3\big)_{(ij),(i'j')}.\end{aligned}$$

Recall the definition of $U_{\ell,m}$ in Proposition 31-(1). Then

$$\begin{aligned}(T_1^2)_{(ij),(i'j')} &= N^{-2}\sum_{(\bar{i}\bar{j})} X_i^\top S^{-1}e_{\bar{j}}\cdot X_{\bar{i}}^\top S^{-1}e_j\cdot X_{\bar{i}}^\top S^{-1}e_{j'}\cdot X_{i'}^\top S^{-1}e_{\bar{j}}\\ &= N^{-2}\bigg(\sum_{\bar{i}} e_j^\top S^{-1}X_{\bar{i}}X_{\bar{i}}^\top S^{-1}e_{j'}\bigg)\cdot\bigg(\sum_{\bar{j}} X_i^\top S^{-1}e_{\bar{j}}e_{\bar{j}}^\top S^{-1}X_{i'}\bigg)\\ &= N^{-1}(S^{-1})_{jj'}\cdot X_i^\top S^{-2}X_{i'} = (U_{2,1})_{(ij),(i'j')},\end{aligned}\tag{D.14}$$

and $T_2 = U_{1,1}$. Proposition 31-(1) entails that

$$\mathbb{E}\|T_1\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_2\|_{\mathrm{op}}^4 = \mathcal{O}(1).\tag{D.15}$$

On the other hand, T_3 has a block diagonal structure with respect to the index (i,i'), so its spectral norm equals that of $(I-S^{-1})\in\mathbb{R}^{p\times p}$, and hence

$$\mathbb{E}\|T_3\|_{\mathrm{op}}^4 = \mathbb{E}\|I-S^{-1}\|_{\mathrm{op}}^4 \le \mathbb{E}\big[\|S^{-1}\|_{\mathrm{op}}^4\|I-S\|_{\mathrm{op}}^4\big] = \mathcal{O}(1).\tag{D.16}$$

Combining all the estimates above, we find that

$$\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \lesssim \mathbb{E}\|T_1\|_{\mathrm{op}}^4 + \mathbb{E}\|T_2\|_{\mathrm{op}}^4 + \mathbb{E}\|T_3\|_{\mathrm{op}}^4 = \mathcal{O}(1).\tag{D.17}$$

Let X' be an independent copy of X and let $X_t' \equiv \sqrt{t}X + \sqrt{1-t}X' \in \mathbb{R}^{N\times p}$. Let $\mathbb{E}'$ denote expectation only with respect to X' and

$$\overline{T}(X) \equiv \int_0^1 \frac{1}{2\sqrt{t}}\langle \nabla T(X), \mathbb{E}'\nabla T(X_t')\rangle\,\mathrm{d}t.$$

Then by the Gaussian-Poincaré inequality

$$\mathrm{Var}\big(\overline{T}(X)\big) \leq \mathbb{E}\|\nabla \overline{T}(X)\|_F^2 \lesssim \sqrt{\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4}\sqrt{\mathbb{E}\|\nabla T(X)\|_F^4} \lesssim p^2.$$

The claim now follows from the second-order Poincaré inequality in Lemma 58 and Proposition 21-(4). $\square$

D.2.3 Ratio control

Proof of Proposition 21. We shorthand $(T_{\mathrm{LRT}}, m_{\Sigma;\mathrm{LRT}}, \sigma_{\Sigma;\mathrm{LRT}}, V_{\Sigma;\mathrm{LRT}})$ by $(T, m_\Sigma, \sigma_\Sigma, V_\Sigma)$.
(1). Recall that $Z_1, \ldots, Z_n$ are i.i.d. samples from $\mathcal{N}(0, I_p)$. By Lemma 52, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$, we have

$$\mathcal{T}_\Sigma(Z) = Z\Sigma^{1/2}\big(I - \Sigma^{-1/2}S_Z^{-1}\Sigma^{-1/2}\big)\Sigma^{1/2} = Z(\Sigma - S_Z^{-1}).$$

Hence with $\{\lambda_j\}_{j=1}^p$ denoting the eigenvalues of Σ, we have

$$\begin{aligned} V_\Sigma^2 &= \mathbb{E}\|\mathcal{T}_\Sigma(Z) - \mathcal{T}_I(Z)\|_F^2 = \mathbb{E}\|Z(\Sigma - I)\|_F^2 \\ &= \mathbb{E}\mathrm{Tr}\big((\Sigma - I)Z^\top Z(\Sigma - I)\big) = \mathrm{Tr}\big(\mathbb{E}Z^\top Z(\Sigma - I)^2\big) \\ &= N\|\Sigma - I\|_F^2 = N\sum_j (\lambda_j - 1)^2. \end{aligned}$$

(2). Note that

$$\begin{aligned} m_\Sigma &= (N/2)\mathbb{E}\Big[\mathrm{Tr}\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big) - \log\det(\Sigma^{1/2}S_Z\Sigma^{1/2}) - p\Big] \\ &= (N/2)\big[\mathrm{Tr}(\Sigma) - \log\det\Sigma - p\big] - (N/2)\mathbb{E}\log\det S_Z, \end{aligned}$$

so

$$\begin{aligned} m_\Sigma - m_I &= (N/2)d_S(\Sigma, I) \\ &= (N/2)\sum_{j=1}^p \big(\lambda_j - \log\lambda_j - 1\big) \gtrsim N\sum_{j=1}^p \big[|\lambda_j - 1| \wedge (\lambda_j - 1)^2\big]. \end{aligned}$$

(3). It is shown by the proof of (Chen and Jiang, 2018, Theorem 1) that with $\mu_{n,0}, \sigma_{n,0}^2$ defined in (Chen and Jiang, 2018, Corollary 1), and $Y_n \equiv \big(T(X) - \mu_{n,0}\big)/(n\sigma_{n,0})$, for $s \in (-s_0, s_0)$ for some $s_0 > 0$,

$$\lim_{n\wedge p\to\infty, n\geq p+2} M_{Y_n}(s) = M_{\mathcal{N}(0,1)}(s) = e^{s^2/2},$$

where $M_Y(s) \equiv \mathbb{E}e^{sY}$ denotes the moment generating function of a generic random variable Y. Now using that for any $s \in (0, s_0)$,

$$\mathbb{E}Y_n^4 = 4\int_0^\infty t^3 \mathbb{P}\big(Y_n > t\big)\,\mathrm{d}t \leq 4\int_0^\infty t^3 e^{-st} M_{Y_n}(s)\,\mathrm{d}t = (6/s^3) M_{Y_n}(s),$$

it follows that $\sup_n \mathbb{E}Y_n^4 < \infty$, and hence convergence of moments yields that $\mathbb{E}Y_n \to 0, \mathbb{E}Y_n^2 \to 1$. This implies $\sigma_I^2/(n^2\sigma_{n,0}^2) = \mathrm{Var}(Y_n) = \mathbb{E}Y_n^2 - (\mathbb{E}Y_n)^2 \to 1$. Hence the asymptotic formula for σ_I^2 holds. In particular, this means that there exists some sufficiently large M such that for $n \wedge p \geq M, n \geq p+2$,

$$\sigma_I^2 \geq M^{-1}\cdot n^2\sigma_{n,0}^2 = \frac{1}{2M}\cdot n^2\bigg[-\frac{p}{N} - \log\bigg(1 - \frac{p}{N}\bigg)\bigg] \overset{(*)}{\geq} \frac{1}{4M}n^2\cdot\frac{p^2}{N^2} \geq \frac{p^2}{4M},$$

where in $(*)$ we used the inequality $-x - \log(1-x) \geq x^2/2$ that holds for $x \in (0,1)$.

(4). Recall that $\{\lambda_j\}_{j=1}^p$ are eigenvalues of Σ. By (1)-(3), we only need to show that for some universal $C > 0$,

$$\frac{\sqrt{N\sum_j(\lambda_j - 1)^2}}{\big(N\sum_j\big(|\lambda_j - 1| \wedge (\lambda_j - 1)^2\big)\big)\bigvee\sigma_I} \leq \frac{C}{\big(\sigma_I \wedge N\big)^{1/2}}. \tag{D.18}$$

To see this, let $\nu_j \equiv |\lambda_j - 1|$, and $J \equiv \{j \in [p] : \nu_j \leq 1\}$, it suffices to prove

$$\frac{\sqrt{N\sum_{j\in J}\nu_j^2}\bigvee\sqrt{N\sum_{j\in J^c}\nu_j^2}}{\big(N\sum_{j\in J}\nu_j^2\big)\bigvee\big(N\sum_{j\in J^c}\nu_j\big)\bigvee\sigma_I} \leq \frac{C}{\big(\sigma_I\wedge N\big)^{1/2}}. \tag{D.19}$$

This follows as

$$\begin{aligned}\text{LHS of (D.19)} &\leq \frac{\sqrt{N\sum_{j\in J}\nu_j^2}}{\big(N\sum_{j\in J}\nu_j^2\big)\bigvee\sigma_I} + \frac{\sqrt{N}\sum_{j\in J^c}\nu_j}{\big(N\sum_{j\in J^c}\nu_j\big)\bigvee\sigma_I}\\ &\leq \frac{1}{\inf_{x\geq 0}\big(x\vee\frac{\sigma_I}{x}\big)} + N^{-1/2} \lesssim \big(\sigma_I\wedge N\big)^{-1/2}.\end{aligned}$$

The proof is complete. □

D.3 Proofs for Section 5.3.2

D.3.1 Evaluation of derivatives

Lemma 53. *Recall the form of $T_{(\mu,\Sigma);\mathrm{LRT}}(X)$ in (5.23). Then for any $(i,j),(i',j') \in [n]\times[p]$,*

1. $\big(\nabla T_{(\mu,\Sigma);\mathrm{LRT}}(X)\big)_{(ij)} = \big(X(I-S_*^{-1}) + \mathbf{1}_n\overline{X}^\top S_*^{-1}\big)_{(ij)} = e_j^\top\big((I-S_*^{-1})X_i + S_*^{-1}\overline{X}\big)$.

2. $\big(\nabla^2 T_{(\mu,\Sigma);\mathrm{LRT}}(X)\big)_{(ij)(i'j')} = n^{-1}(X_i-\overline{X})^\top S_*^{-1}\big(e_{j'}(X_{i'}-\overline{X})^\top + (X_{i'}-\overline{X})e_{j'}^\top\big)S_*^{-1}e_j + \delta_{ii'}e_j^\top(I-S_*^{-1})e_{j'} + n^{-1}e_{j'}^\top S_*^{-1}e_j$.

Proof. We shorthand $T_{(\mu,\Sigma);\mathrm{LRT}}$ as T.

(1). By definition, we have

$$\partial_{(ij)}T(X) = \frac{n}{2}\big[\partial_{(ij)}\mathrm{Tr}(S_*) - \partial_{(ij)}\log\det S_* + \partial_{(ij)}\overline{X}^\top\overline{X}\big].$$

For the first term, using $\partial_{(ij)}X_k = \delta_{ik}e_j$ and $\partial_{(ij)}\overline{X} = n^{-1}\sum_k \partial_{ij}X_k = n^{-1}e_j$, we have

$$\begin{aligned}
\partial_{(ij)}\mathrm{Tr}(S_*(X)) &= n^{-1}\sum_k \frac{\partial}{\partial X_{ij}}\mathrm{Tr}\big((X_k-\overline{X})(X_k-\overline{X})^\top\big)\\
&= n^{-1}\sum_k \mathrm{Tr}\big[(\delta_{ik}e_j - n^{-1}e_j)(X_k-\overline{X})^\top + (X_k-\overline{X})(\delta_{ik}e_j-n^{-1}e_j)^\top\big]\\
&= n^{-1}\sum_k \delta_{ik}\cdot 2(X_k-\overline{X})_j = 2n^{-1}(X_{ij}-\overline{X}_j).
\end{aligned} \tag{D.20}$$

For the second term, using the well-known fact that $\nabla\log\det A = A^{-1}$ for any invertible and symmetric matrix A (see e.g., (Boyd and Vandenberghe, 2004, Section A.4.1)), we have

$$\begin{aligned}
\partial_{(ij)}\log\det S_* &= \sum_{k,\ell}\frac{\partial\log\det S_*}{\partial(S_*)_{k\ell}}\frac{\partial(S_*)_{k\ell}}{\partial X_{ij}}\\
&\overset{(*)}{=} \sum_{k,\ell}(S_*^{-1})_{k\ell}\cdot\frac{1}{n}\big(\delta_{jk}(X_{i\ell}-\overline{X}_\ell) + \delta_{j\ell}(X_{ik}-\overline{X}_k)\big)\\
&= \frac{2}{n}\sum_\ell (S_*^{-1})_{j\ell}(X_{i\ell}-\overline{X}_\ell) = \frac{2}{n}\big((X-\mathbf{1}_n\overline{X}^\top)S_*^{-1}\big)_{ij},
\end{aligned} \tag{D.21}$$

where in $(*)$, we use

$$\frac{\partial (S_*)_{k\ell}}{\partial X_{ij}} = \frac{1}{n}\frac{\partial}{\partial X_{ij}}\sum_{m=1}^{n}\big((X_m-\overline{X})^\top e_k\big)\big((X_m-\overline{X})^\top e_\ell\big)$$
$$= \frac{1}{n}\big[\delta_{jk}(X_{i\ell}-\overline{X}_\ell)+\delta_{j\ell}(X_{ik}-\overline{X}_k)\big]. \tag{D.22}$$

Lastly, for the third term, we have

$$\partial_{(ij)}\overline{X}^\top\overline{X} = 2(\partial_{(ij)}\overline{X})^\top\overline{X} = \frac{2}{n}\overline{X}_j. \tag{D.23}$$

We combine (D.20)-(D.23) to obtain the first claim.

(2). Again by definition, we have

$$\partial_{(ij)(i'j')}T(X) = \frac{n}{2}\big[\partial_{(ij)(i'j')}\mathrm{Tr}(S_*) - \partial_{(ij)(i'j')}\log\det S_* + \partial_{(ij)(i'j')}\overline{X}^\top\overline{X}\big].$$

For the first term, it follows from (D.20) that

$$\partial_{(ij)(i'j')}\mathrm{Tr}(S_*) = 2n^{-1}\partial_{(i'j')}(X_{ij}-\overline{X}_j) = 2n^{-1}\big(\delta_{ii'}-n^{-1}\big)\delta_{jj'}. \tag{D.24}$$

For the second term, it follows from (D.21) that

$$\partial_{(ij)(i'j')}\log\det S_* = \frac{2}{n}\frac{\partial}{\partial X_{i'j'}}(X_i-\overline{X})^\top S_*^{-1}e_j$$
$$= \frac{2}{n}\Big[\frac{\partial (X_i-\overline{X})}{\partial X_{i'j'}}^\top S_*^{-1}e_j + (X_i-\overline{X})^\top\frac{\partial S_*^{-1}}{\partial X_{i'j'}}e_j\Big]$$
$$\overset{(**)}{=} \frac{2}{n}\Big[(\delta_{ii'}-n^{-1})e_{j'}^\top S_*^{-1}e_j$$
$$- n^{-1}(X_i-\overline{X})^\top S_*^{-1}\big(e_{j'}(X_{i'}-\overline{X})^\top + (X_{i'}-\overline{X})e_{j'}^\top\big)S_*^{-1}e_j\Big], \tag{D.25}$$

where in $(**)$ we use the following calculation with the help of (D.22):

$$
\begin{aligned}
\frac{\partial S_*^{-1}}{\partial X_{i'j'}} &= -S_*^{-1}\frac{\partial S_*}{\partial X_{i'j'}}S_*^{-1}\\
&= -S_*^{-1}\Big[\sum_{k,\ell} e_k e_\ell^\top \frac{\partial (S_*)_{k\ell}}{\partial X_{i'j'}}\Big]S_*^{-1}\\
&= -\frac{1}{n}S_*^{-1}\cdot\Big[\sum_{k,\ell} e_k e_\ell^\top\big(\delta_{j'k}(X_{i'\ell}-\overline{X}_\ell)+\delta_{j'\ell}(X_{i'k}-\overline{X}_k)\big)\Big]S_*^{-1}\\
&= -\frac{1}{n}S_*^{-1}\big[e_{j'}(X_{i'}-\overline{X})^\top+(X_{i'}-\overline{X})e_{j'}^\top\big]S_*^{-1}.
\end{aligned}
$$

Lastly

$$
\partial_{(ij),(i'j')}\overline{X}^\top\overline{X} = \frac{2}{n}\partial_{(i'j')}\overline{X}_j = \frac{2}{n^2}\delta_{jj'}. \tag{D.26}
$$

We combine (D.24)-(D.26) to obtain the second claim. □

D.3.2 Normal approximation

Proof of Theorem 16. We again shorthand $T_{(\mu,\Sigma);\mathrm{LRT}}$ as T. First we bound the norm for the gradient: by Lemma 53-(1),

$$
\begin{aligned}
\|\nabla T(X)\|_F^2 &= \sum_{i,j}\big(X(I-S_*^{-1})+\mathbf{1}_n\overline{X}^\top S_*^{-1}\big)_{ij}^2\\
&\le 2\sum_i \|(I-S_*^{-1})X_i\|^2 + 2\sum_{i,j}\big(\overline{X}^\top S_*^{-1}e_j\big)^2\\
&= 2\sum_i \|(I-S_*^{-1})X_i\|^2 + 2n\overline{X}^\top S_*^{-2}\overline{X} \equiv (I)+(II).
\end{aligned}
$$

By essentially the same arguments as in (D.13) in the proof of Theorem 14, we have $\mathbb{E}(I)^2 \lesssim p^4$, so we only need to handle (II):

$$
\begin{aligned}
\mathbb{E}(II)^2 &= 4n^2\cdot\mathbb{E}\big(\overline{X}^\top S_*^{-2}\overline{X}\,\overline{X}^\top S_*^{-2}\overline{X}\big) \le 4n^2\cdot\mathbb{E}\big(\|S_*^{-1}\|_{\mathrm{op}}^4\|\overline{X}\|^4\big)\\
&\overset{(*)}{=} 4n^2\mathbb{E}\|S_*^{-1}\|_{\mathrm{op}}^4\cdot\mathbb{E}\|\overline{X}\|^4 \overset{(**)}{\lesssim} n^2\cdot\frac{p^2}{n^2} = p^2.
\end{aligned}
$$

Here $(*)$ follows from the fact that S_* is independent of $\overline{X}$, and $(**)$ follows from Lemma 50. Combining the bounds we have $\mathbb{E}\|\nabla T(X)\|_F^4 \lesssim p^4$.

Next we bound the spectral norm for the Hessian. By Lemma 53-(2),

$$\begin{aligned}
\partial_{(ij),(i'j')} T(X) &= n^{-1}(X_i - \overline{X})^\top S_*^{-1} e_{j'} (X_{i'} - \overline{X})^\top S_*^{-1} e_j \\
&\quad + n^{-1}(X_i - \overline{X})^\top S_*^{-1}(X_{i'} - \overline{X})(S_*^{-1})_{jj'} \\
&\quad + \delta_{ii'}\big(I - S_*^{-1}\big)_{jj'} + n^{-1}(S_*^{-1})_{jj'} \\
&\equiv \big(T_1 + T_2 + T_3 + T_4\big)_{(ij),(i'j')}.
\end{aligned}$$

Using the same calculation as in (D.14), we have

$$(T_1^2)_{(ij),(i'j')} = n^{-1}(X_i - \overline{X})^\top S_*^{-2}(X_{i'} - \overline{X})(S_*^{-1})_{jj'}.$$

Proposition 31-(2) entails that $\mathbb{E}\|T_1\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_2\|_{\mathrm{op}}^4 = \mathcal{O}(1)$. Similar to (D.16), $\mathbb{E}\|T_3\|_{\mathrm{op}}^4 = \mathcal{O}(1)$. For T_4, note that

$$\begin{aligned}
\|T_4\|_{\mathrm{op}} &= n^{-1} \sup_{\substack{u=(u_1,\ldots,u_n)\in\mathbb{R}^{n\times p} \\ \sum_{k=1}^n \|u_k\|^2 \le 1}} \Big| \sum_{1\le k,\ell\le n} u_k^\top S_*^{-1} u_\ell \Big| \\
&\le n^{-1} \sup_{\substack{u=(u_1,\ldots,u_n)\in\mathbb{R}^{n\times p} \\ \sum_{k=1}^n \|u_k\|^2 \le 1}} \sum_{1\le k,\ell\le n} \|u_k\|\|u_\ell\| \cdot \|S_*^{-1}\|_{\mathrm{op}} \le \|S_*^{-1}\|_{\mathrm{op}},
\end{aligned}$$

where in the last inequality we use the Cauchy-Schwarz inequality that $\sum_{1\le k,\ell\le n}\|u_k\|\|u_\ell\| = \big(\sum_{k=1}^n \|u_k\|\big)^2 \le n\sum_k \|u_k\|^2 \le n$. Hence $\mathbb{E}\|T_4\|_{\mathrm{op}}^4 = \mathcal{O}(1)$. Combining the bounds we arrive at $\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 = \mathcal{O}(1)$. The rest of the proof proceeds along the lines in the proof of Theorem 14, with the help of the variance formula in Proposition 22-(3). □

D.3.3 Ratio control

Proof of Proposition 22. We shorthand $(T_{(\mu,\Sigma);\mathrm{LRT}}, m_{(\mu,\Sigma);\mathrm{LRT}}, \sigma_{(\mu,\Sigma);\mathrm{LRT}}, V_{(\mu,\Sigma);\mathrm{LRT}})$ by $(T, m_{(\mu,\Sigma)}, \sigma_{(\mu,\Sigma)}, V_{(\mu,\Sigma)}$
(1). Recall that $Z_1, \ldots, Z_n$ are i.i.d. samples from $\mathcal{N}(0, I_p)$. Let

$$S_{*,Z} \equiv n^{-1}\sum_{i=1}^n (Z_i - \overline{Z})(Z_i - \overline{Z})^\top.$$

By Lemma 53,

$$
\begin{aligned}
\mathcal{T}_{(\mu,\Sigma)} &= \Big(\big(Z\Sigma^{1/2} + \mathbf{1}_n\mu^\top\big)\big(I - \Sigma^{-1/2}S_{*,Z}\Sigma^{-1/2}\big) \\
&\qquad + \mathbf{1}_n\big(\overline{Z}^\top\Sigma^{1/2} + \mu^\top\big)\cdot\Sigma^{-1/2}S_{*,Z}\Sigma^{-1/2}\Big)\Sigma^{1/2} \\
&= Z(\Sigma - S_{*,Z}) + \mathbf{1}_n\overline{Z}^\top S_{*,Z} + \mathbf{1}_n\mu^\top\Sigma^{1/2}.
\end{aligned}
$$

Hence with $\|\mu\|_\Sigma^2 \equiv \mu^\top\Sigma\mu$ and $\{\lambda_{j=1}^p\}_{j=1}^p$ denoting the eigenvalues of Σ,

$$
\begin{aligned}
V_{(\mu,\Sigma)}^2 &= \mathbb{E}\big\|\mathcal{T}_{(\mu,\Sigma)}(Z) - \mathcal{T}_{(0,I)}(Z)\big\|_F^2 = \mathbb{E}\|Z(\Sigma - I) + \mathbf{1}_n\mu^\top\Sigma^{1/2}\|_F^2 \\
&= n\big(\|\Sigma - I\|_F^2 + \mu^\top\Sigma\mu\big) = n\Big[\sum_j \big(\lambda_j - 1\big)^2 + \|\mu\|_\Sigma^2\Big].
\end{aligned}
$$

(2). Note that

$$
\begin{aligned}
m_{(\mu,\Sigma)} &= \frac{n}{2}\mathbb{E}\big[\mathrm{Tr}\big(\Sigma^{1/2}S_{*,Z}\Sigma^{1/2}\big) - \log\det(\Sigma^{1/2}S_{*,Z}\Sigma^{1/2}) - p + \overline{X}^\top\overline{X}\big] \\
&= \frac{n}{2}\big[\mathrm{Tr}(\Sigma) - \log\det\Sigma - p + \|\mu\|^2\big] - \frac{n}{2}\cdot\mathbb{E}\log\det S_{*,Z},
\end{aligned}
$$

where the second equality follows as

$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}(S_{*,Z}) &= \frac{N}{n}\mathbb{E}\mathrm{Tr}\Big(N^{-1}\sum_{k=1}^n (Z_k - \overline{Z})(Z_k - \overline{Z})^\top\Big) \\
&= \frac{n-1}{n}\mathbb{E}\mathrm{Tr}\Big(N^{-1}\sum_{i=1}^N Z_k Z_k^\top\Big) = \frac{n-1}{n}\mathrm{Tr}(\Sigma),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\overline{X}^\top\overline{X} &= n^{-2}\mathbb{E}\Big(\sum_k X_k\Big)^\top\Big(\sum_\ell X_\ell\Big) = n^{-2}\Big[\Big(\sum_{k\neq\ell} + \sum_{k=\ell}\Big)\mathbb{E}X_k^\top X_\ell\Big] \\
&= n^{-2}\big[n(n-1)\|\mu\|^2 + n\big(\|\mu\|^2 + \mathrm{Tr}(\Sigma)\big)\big] = \|\mu\|^2 + n^{-1}\mathrm{Tr}(\Sigma).
\end{aligned}
$$

Hence

$$
m_{(\mu,\Sigma)} - m_{(0,I)} = \frac{n}{2}\big(d_S(\Sigma, I) + \|\mu\|^2\big).
$$

(3). The proof is the same as Proposition 21-(3) by invoking (Chen and Jiang, 2018, Theorem 2).

(4). By (1)-(3), we only need to show that for some universal constant $C > 0$,

$$\frac{\sqrt{n\big(\sum_j(\lambda_j-1)^2\vee\|\mu\|_\Sigma^2\big)}}{\big(n\sum_j\big(|\lambda_j-1|\wedge(\lambda_j-1)^2\big)\big)\bigvee\big(n\|\mu\|^2\big)\bigvee\sigma_{(0,I)}}\leq\frac{C}{\big(\sigma_{(0,I)}\wedge n\big)^{1/2}}\tag{D.27}$$

holds. In view of (D.18), we only need to prove that

$$\frac{\sqrt{n\|\mu\|_\Sigma^2}}{\big(n\sum_j\big(|\lambda_j-1|\wedge(\lambda_j-1)^2\big)\big)\bigvee\big(n\|\mu\|^2\big)\bigvee\sigma_{(0,I)}}\leq\frac{C}{\big(\sigma_{(0,I)}\wedge n\big)^{1/2}}.$$

As $\|\mu\|_\Sigma^2 = \|\mu\|^2+\mu^\top(\Sigma-I)\mu \leq \|\mu\|^2+\|\mu\|^2\cdot\max_j|\lambda_j-1|$, with $\nu_j = |\lambda_j-1|$ and $J=\{j\in[p]:\nu_j\leq 1\}$, we only need to prove

$$\frac{\sqrt{n\|\mu\|^2}}{\big(n\sum_{j\in J}\nu_j^2\big)\bigvee\big(n\sum_{j\in J^c}\nu_j\big)\bigvee\big(n\|\mu\|^2\big)\bigvee\sigma_{(0,I)}}\leq\frac{1}{\sigma_{(0,I)}^{1/2}},\tag{D.28}$$

$$\frac{\sqrt{n\|\mu\|^2\cdot\max_j\nu_j}}{\big(n\sum_{j\in J}\nu_j^2\big)\bigvee\big(n\sum_{j\in J^c}\nu_j\big)\bigvee\big(n\|\mu\|^2\big)\bigvee\sigma_{(0,I)}}\leq\frac{C}{\big(\sigma_{(0,I)}\wedge n\big)^{1/2}}.\tag{D.29}$$

To see (D.28), note that

$$\text{LHS of (D.28)}\leq\frac{\sqrt{n\|\mu\|^2}}{\big(n\|\mu\|^2\big)\bigvee\sigma_{(0,I)}}\leq\frac{1}{\inf_{x\geq 0}\big(x\vee\frac{\sigma_{0,I}}{x}\big)}\leq\frac{1}{\sigma_{0,I}^{1/2}}.$$

To see (D.29), using that $ab\leq(a^2+b^2)/2$, we have

$$\begin{aligned}\text{LHS of (D.29)}&\lesssim\frac{n^{1/2}\|\mu\|^2}{\big(n\|\mu\|^2\big)}+\frac{n^{1/2}\max_j\nu_j}{\big(n\sum_{j\in J}\nu_j^2\big)\bigvee\big(n\sum_{j\in J^c}\nu_j\big)\bigvee\sigma_{(0,I)}}\\&\leq n^{-1/2}+\frac{n^{1/2}(\max_j\nu_j)\cdot\mathbf{1}_{\max_j\nu_j>1}}{\big(n\sum_{j\in J^c}\nu_j\big)}+\frac{n^{1/2}(\max_j\nu_j)\cdot\mathbf{1}_{\max_j\nu_j\leq 1}}{\big(n\max_j\nu_j^2\big)\bigvee\sigma_{(0,I)}}\\&\leq 2n^{-1/2}+\frac{n^{1/2}}{\inf_{x\geq 0}\big(nx\vee\frac{\sigma_{(0,I)}}{x}\big)}\lesssim n^{-1/2}\bigvee\sigma_{(0,I)}^{-1/2},\end{aligned}$$

as desired. $\square$

D.4 Proofs for Section 5.3.3

D.4.1 Evaluation of derivatives

Lemma 54. *Recall the form of $T_{\mathrm{LNW}}(X)$ in (5.24). We assume without loss of generality that $\mu = 0$. Then for any $(i,j),(i',j') \in [N]\times[p]$,*

1. $\big(\nabla T_{\mathrm{LNW}}(X)\big)_{(ij)} = \big(X(S-I) - (Tr(S)/N)X\big)_{ij} = e_j^\top(S-I)X_i - (Tr(S)/N)X_{ij}$.

2. $\big(\nabla^2 T_{\mathrm{LNW}}(X)\big)_{(ij)(i'j')} = N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}(S-I)_{jj'} - (2/N^2)X_{ij}X_{i'j'} - (Tr(S)/N)\delta_{ii'}\delta_{jj'}$.

Furthermore, for any $(i_\ell, j_\ell) \in [N]\times[p], \ell = 1,2,3,4$,

$$\begin{aligned}
&\partial_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)}T_{\mathrm{LNW}}(X)\\
&= N^{-1}\big(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_2}\delta_{j_3j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_2}\delta_{j_3j_4}\\
&\qquad + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_4}\delta_{j_2j_3}\\
&\qquad + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_4}\delta_{j_2j_3}\big)\\
&\qquad - 2N^{-2}\big(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_4}\delta_{j_2j_3} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_2}\delta_{j_3j_4}\big).
\end{aligned}$$

Proof. We shorthand T_{LNW} as T. As $\partial_{ij}S(X) = N^{-1}(e_jX_i^\top + X_ie_j^\top)$, for the first-order derivatives we have

$$\begin{aligned}
\partial_{(ij)}T(X) &= \frac{N}{4}\left(\mathrm{Tr}\big[\partial_{(ij)}(S-I)^2\big] - \frac{1}{N}\cdot 2\mathrm{Tr}(S)\mathrm{Tr}\big[\partial_{(ij)}S\big]\right)\\
&= \frac{1}{2}\mathrm{Tr}\big[(S-I)(e_jX_i^\top + X_ie_j^\top)\big] - \frac{\mathrm{Tr}(S)X_{ij}}{N}\\
&= \big(X(S-I)\big)_{ij} - \frac{\mathrm{Tr}(S)}{N}X_{ij}\\
&= e_j^\top(S-I)X_i - \frac{\mathrm{Tr}(S)}{N}X_{ij}.
\end{aligned}$$

For the second-order derivatives we have

$$
\begin{aligned}
\partial_{(ij),(i'j')}T(X) &= \partial_{(i'j')}\big(e_j^\top(S-I)X_i\big) - N^{-1}\partial_{(i'j')}\big(\mathrm{Tr}(S)X_{ij}\big)\\
&= N^{-1}e_j^\top(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top)\cdot X_i + \delta_{ii'}e_j^\top(S-I)e_{j'}\\
&\quad - N^{-1}\big((2/N)X_{ij}X_{i'j'} + \delta_{ii'}\delta_{jj'}\mathrm{Tr}(S)\big)\\
&= N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}(S-I)_{jj'}\\
&\quad - 2N^{-2}X_{ij}X_{i'j'} - N^{-1}\mathrm{Tr}(S)\delta_{ii'}\delta_{jj'}.
\end{aligned}
$$

For the third-order derivatives we have

$$
\begin{aligned}
&\partial_{(i_1j_1)(i_2j_2)(i_3j_3)}T(X)\\
&= N^{-1}\delta_{j_1j_2}\partial_{(i_3j_3)}(X_{i_1}^\top X_{i_2}) + N^{-1}\partial_{(i_3j_3)}(X_{i_2j_1}X_{i_1j_2})\\
&\quad + N^{-1}\delta_{i_1i_2}e_{j_1}^\top\big(e_{j_3}X_{i_3}^\top + X_{i_3}e_{j_3}^\top\big)e_{j_2}\\
&\quad - 2N^{-2}\partial_{(i_3j_3)}\big(X_{i_1j_1}X_{i_2j_2}\big) - 2N^{-2}\delta_{i_1i_2}\delta_{j_1j_2}X_{i_3j_3}\\
&= N^{-1}\big(\delta_{i_1i_3}\delta_{j_1j_2}X_{i_2j_3} + \delta_{i_2i_3}\delta_{j_1j_2}X_{i_1j_3}\big)\\
&\quad + N^{-1}\big(\delta_{i_2i_3}\delta_{j_1j_3}X_{i_1j_2} + \delta_{i_1i_3}\delta_{j_2j_3}X_{i_2j_1}\big)\\
&\quad + N^{-1}\big(\delta_{i_1i_2}\delta_{j_1j_3}X_{i_3j_2} + \delta_{i_1i_2}\delta_{j_2j_3}X_{i_3j_1}\big)\\
&\quad - 2N^{-2}\big(\delta_{i_1i_3}\delta_{j_1j_3}X_{i_2j_2} + \delta_{i_2i_3}\delta_{j_2j_3}X_{i_1j_1} + \delta_{i_1i_2}\delta_{j_1j_2}X_{i_3j_3}\big).
\end{aligned}
$$

For the fourth-order derivatives we have

$$
\begin{aligned}
&\partial_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)}T(X)\\
&= N^{-1}\big(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_2}\delta_{j_3j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_2}\delta_{j_3j_4}\\
&\quad + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_4}\delta_{j_2j_3}\\
&\quad + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_4}\delta_{j_2j_3}\big)\\
&\quad - 2N^{-2}\big(\delta_{i_1i_3}\delta_{i_2i_4}\delta_{j_1j_3}\delta_{j_2j_4} + \delta_{i_1i_4}\delta_{i_2i_3}\delta_{j_1j_4}\delta_{j_2j_3} + \delta_{i_1i_2}\delta_{i_3i_4}\delta_{j_1j_2}\delta_{j_3j_4}\big).
\end{aligned}
$$

The proof is complete. $\square$

D.4.2 Normal approximation

Proof of Theorem 18. Let $y \equiv p/N$. We start by showing that

$$\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \leq C(1\vee y)^4 \tag{D.30}$$

for some absolute constant $C>0$. Reorganizing the terms in Lemma 54, we have

$$\begin{aligned}\big(\nabla^2 T(X)\big)_{(ij),(i'j')} &= N^{-1}X_i^\top X_{i'}\delta_{jj'} + N^{-1}X_{ij'}X_{i'j} - 2N^{-2}X_{i'j'}X_{ij}\\ &\quad + \delta_{ii'}e_{j'}^\top(S - I - N^{-1}\mathrm{Tr}(S)I)e_j\\ &\equiv (T_{2,1}+T_{2,2}-T_{2,3}+T_{2,4})_{(ij),(i'j')}.\end{aligned}$$

Recall the definition of $U_{\ell,m;+}$ from Proposition 32. As $T_{2,1}=U_{0,0;+}$ and

$$\begin{aligned}(T_{2,2}^2)_{(ij),(i'j')} &= N^{-2}\sum_{(\bar{i}\bar{j})} X_{\bar{i}j}X_{i\bar{j}}X_{i'\bar{j}}X_{\bar{i}j'} = N^{-2}\bigg(\sum_{\bar{i}} X_{\bar{i}j}X_{\bar{i}j'}\bigg)\bigg(\sum_{\bar{j}} X_{i\bar{j}}X_{i'\bar{j}}\bigg)\\ &= N^{-1}S_{jj'}X_i^\top X_{i'} = (U_{0,1;+})_{(ij),(i'j')},\end{aligned}$$

Proposition 32 entails that $\mathbb{E}\|T_{2,1}\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_{2,2}\|_{\mathrm{op}}^4 = \mathcal{O}((1\vee y)^4)$. For $T_{2,3}$, as

$$\|T_{2,3}\|_{\mathrm{op}} = (2/N^2)\sup_{u,v\in B_{N\times p}}\bigg|\sum_{(ij),(i'j')} u_{ij}X_{ij}X_{i'j'}v_{i'j'}\bigg| = (2/N^2)\|X\|_F^2.$$

Hence $\mathbb{E}\|T_{2,3}\|_{\mathrm{op}}^4 = \mathcal{O}(y^4) = \mathcal{O}((1\vee y)^4)$. For $T_{2,4}$, it holds by the block diagonal structure that

$$\|T_{2,4}\|_{\mathrm{op}} = \|S - I - N^{-1}\mathrm{Tr}(S)I\|_{\mathrm{op}} \leq \|S-I\|_{\mathrm{op}} + N^{-1}\mathrm{Tr}(S).$$

Hence it holds by Lemma 51 that

$$\mathbb{E}\|T_{2,4}\|_{\mathrm{op}}^4 \lesssim \big(y\vee\sqrt{y}\big)^4 + N^{-4}\cdot N^{-4}\mathbb{E}\|X\|_F^8 \lesssim (1\vee y)^4.$$

By collecting the estimates of $T_{2,1}$-$T_{2,4}$, we complete the proof of (D.30).

Next we show that $\mathbb{E}\|\nabla T(X)\|_F^4 \lesssim p^4$. This will be done by two estimates below.

(**Estimate 1**) By Lemma 54-(1),

$$\begin{aligned}\|\nabla T(X)\|_F^2 &\lesssim \sum_i \|(S-I)X_i\|^2 + N^{-2}\mathrm{Tr}^2(S)\|X\|_F^2\\ &\leq \big(\|S-I\|_{\mathrm{op}}^2 + N^{-2}\mathrm{Tr}^2(S)\big)\sum_i \|X_i\|^2,\end{aligned}$$

so by Lemma 51 and Proposition 32,

$$\begin{aligned}\mathbb{E}\|\nabla T(X)\|_F^4 &\lesssim \mathbb{E}\Big[\big(\|S-I\|_{\mathrm{op}}^2 + N^{-2}\mathrm{Tr}^2(S)\big)\sum_i \|X_i\|^2\Big]^2\\ &\lesssim \sum_{i,i'} \mathbb{E}\big[\big(\|S-I\|_{\mathrm{op}}^4 + N^{-4}\mathrm{Tr}^4(S)\big)\|X_i\|^2\|X_{i'}\|^2\big]\\ &\leq \sum_{i,i'} \Big(\mathbb{E}^{1/2}\|S-I\|_{\mathrm{op}}^8 + N^{-4}\mathbb{E}^{1/2}\mathrm{Tr}^8(S)\Big)\cdot \mathbb{E}^{1/4}\|X_i\|^8 \cdot \mathbb{E}^{1/4}\|X_{i'}\|^8\\ &\lesssim N^2 \cdot \Big[\Big(\frac{p}{N}\Big)^2 + \Big(\frac{p}{N}\Big)^4 \cdot \mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^8\Big]\cdot p \cdot p \lesssim p^4(1+y^6).\end{aligned}$$

(**Estimate 2**) Note that

$$\nabla T(X) = X\big(S - N^{-1}\mathrm{Tr}(S)I\big) - X \equiv T_{1,1} + T_{1,2}.$$

It is clear that $\mathbb{E}\|T_{1,2}\|_F^2 \lesssim Np$. To handle $T_{1,1}$, note that

$$\begin{aligned}\|T_{1,1}\|_F^2 &= N\mathrm{Tr}\Big(\big(S - N^{-1}\mathrm{Tr}(S)I\big)^2 S\Big)\\ &= N\mathrm{Tr}\Big(S^3 + N^{-2}\mathrm{Tr}^2(S)S - 2N^{-1}\mathrm{Tr}(S)S^2\Big)\\ &= N\Big[\mathrm{Tr}(S^3) + N^{-2}\mathrm{Tr}^3(S) - 2N^{-1}\mathrm{Tr}(S)\mathrm{Tr}(S^2)\Big].\end{aligned}$$

Then using Lemma 64, we have under the prescribed asymptotics that

$$\begin{aligned}\mathbb{E}\|T_{1,1}\|_F^2 &= N\Big[py^2 + 3py + p + 3y^2 + 3y + 4N^{-1}y + N^{-2}\big(p^3 + 6py + 8N^{-1}y\big)\\ &\qquad - 2N^{-1}\big(p^2y + p^2 + py + 4(y^2+y) + 4N^{-1}y\big)\Big]\\ &= p^2\Big(1 + \frac{N}{p} + \frac{1}{N} + \frac{3}{p} - \frac{4}{Np} - \frac{2}{N^2}\Big)\\ &= p^2\big[1 + \mathcal{O}((N\wedge p)^{-1})\big] + pN.\end{aligned}$$

Hence we have

$$\begin{aligned}\mathbb{E}\|\nabla T(X)\|_F^4 &= \big(\mathbb{E}\|\nabla T(X)\|_F^2\big)^2 + \mathrm{Var}\big(\|\nabla T(X)\|_F^2\big)\\ &= \mathcal{O}(p^4(1+y^{-2})) + \mathrm{Var}\big(\|\nabla T(X)\|_F^2\big).\end{aligned} \tag{D.31}$$

By the Gaussian-Poincaré inequality, we have

$$\begin{aligned}\mathrm{Var}\big(\|\nabla T(X)\|_F^2\big) &\leq \mathbb{E}\|\nabla\|\nabla T(X)\|_F^2\|_F^2 = 4\mathbb{E}\|\big(\nabla^2 T(X)\big)^\top \nabla T(X)\|_F^2\\ &\leq 4\mathbb{E}^{1/2}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \cdot \mathbb{E}^{1/2}\|\nabla T(X)\|_F^4.\end{aligned}$$

Combining the above display with (D.31) yields that

$$\mathbb{E}\|\nabla T(X)\|_F^4 \leq \mathcal{O}(p^4(1+y^{-2})) + 4\mathbb{E}^{1/2}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \cdot \mathbb{E}^{1/2}\|\nabla T(X)\|_F^4.$$

Solving the quadratic inequality above and using (D.30), we arrive at

$$\mathbb{E}\|\nabla T(X)\|_F^4 = \mathcal{O}\big(p^4(1+y^{-2}) \vee \mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4\big) = \mathcal{O}(p^4(1+y^{-2})).$$

Combining the above two estimates, we have

$$\mathbb{E}\|\nabla T(X)\|_F^4 \lesssim p^4 \max_{y\geq 0} \min\big\{(1+y^6),(1+y^{-2})\big\} \asymp p^4.$$

The rest of the proof proceeds along the lines in the proof of Theorem 14, with the help of the variance formula in Proposition 23-(3). $\square$

D.4.3 Ratio control

Proof of Proposition 23. (1). Recall that $Z_1,\ldots,Z_n$ are i.i.d. samples from $\mathcal{N}(0,I_p)$. By Lemma 54, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$,

$$\begin{aligned}\mathcal{T}_{\Sigma;\mathrm{LNW}}(Z) &= \Big[Z\Sigma^{1/2}\big(\Sigma^{1/2}S_Z\Sigma^{1/2} - I\big) - N^{-1}\mathrm{Tr}(\Sigma S_Z) Z\Sigma^{1/2}\Big]\Sigma^{1/2}\\ &= Z\Sigma S_Z\Sigma - Z\Sigma - N^{-1}\mathrm{Tr}(\Sigma S_Z) Z\Sigma,\end{aligned}$$

so

$$
\begin{aligned}
&\mathcal{T}_{\Sigma;\mathrm{LNW}}(Z) - \mathcal{T}_{I;\mathrm{LNW}}(Z)\\
&= \Big[Z\Sigma(S_Z\Sigma - I) - Z(S_Z - I)\Big] - \frac{1}{N}\Big[\mathrm{Tr}(\Sigma S_Z)Z\Sigma - \mathrm{Tr}(S_Z)Z\Big]\\
&= \Big[Z\Sigma(S_Z\Sigma - I) - Z(S_Z\Sigma - I) + Z(S_Z\Sigma - I) - Z(S_Z - I)\Big]\\
&\qquad - \frac{1}{N}\Big[\mathrm{Tr}(\Sigma S_Z)Z\Sigma - \mathrm{Tr}(\Sigma S_Z)Z + \mathrm{Tr}(\Sigma S_Z)Z - \mathrm{Tr}(S_Z)Z\Big]\\
&= Z(\Sigma - I)(S_Z\Sigma - I) + ZS_Z(\Sigma - I)\\
&\qquad - \frac{1}{N}\mathrm{Tr}(\Sigma S_Z)Z(\Sigma - I) - \frac{1}{N}\mathrm{Tr}\big((\Sigma - I)S_Z\big)Z\\
&\equiv V_1(Z) + V_2(Z) + V_3(Z) + V_4(Z).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\|V_1(Z)\|_F^2 &\le \|S_Z\Sigma - I\|_{\mathrm{op}}^2\|Z(\Sigma - I)\|_F^2 \le \|S_Z\Sigma - I\|_{\mathrm{op}}^2\|Z\|_{\mathrm{op}}^2\|\Sigma - I\|_F^2,\\
\|V_2(Z)\|_F^2 &\le \|ZS_Z\|_{\mathrm{op}}^2\|\Sigma - I\|_F^2 \le \|Z\|_{\mathrm{op}}^2\|S_Z\|_{\mathrm{op}}^2\|\Sigma - I\|_F^2,\\
\|V_3(Z)\|_F^2 &\le N^{-2}\mathrm{Tr}^2(\Sigma S_Z)\|Z(\Sigma - I)\|_F^2\\
&\le p^2N^{-2}\|\Sigma\|_{\mathrm{op}}^2\|S_Z\|_{\mathrm{op}}^2\|Z(\Sigma - I)\|_F^2,\\
\|V_4(Z)\|_F^2 &\le N^{-2}\mathrm{Tr}^2\big((\Sigma - I)S_Z\big)\|Z\|_F^2\\
&\le N^{-2}\|S_Z\|_F^2\|Z\|_F^2\|\Sigma - I\|_F^2 \le pN^{-2}\|S_Z\|_{\mathrm{op}}^2\|Z\|_F^2\|\Sigma - I\|_F^2.
\end{aligned}
$$

Under $p/N \le M$, we have

$$
V_{\Sigma;\mathrm{LNW}}^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.
$$

(2). By Lemma 63, with $\delta_N \equiv N^{-1} - 2N^{-2}$,

$$\begin{aligned}
m_\Sigma &= \frac{N}{4}\left[\mathbb{E}\mathrm{Tr}(S-I)^2 - \frac{1}{N}\mathbb{E}\mathrm{Tr}^2(S)\right] \\
&= \frac{N}{4}\left[\mathbb{E}\mathrm{Tr}(S^2) - 2\mathbb{E}\mathrm{Tr}(S) + p - \frac{1}{N}\mathbb{E}\mathrm{Tr}^2(S)\right] \\
&= \frac{N}{4}\left[\big(1+N^{-1}\big)\mathrm{Tr}(\Sigma^2) + N^{-1}\mathrm{Tr}^2(\Sigma) - 2\mathrm{Tr}(\Sigma) + p - N^{-1}\mathrm{Tr}^2(\Sigma) - 2N^{-2}\mathrm{Tr}(\Sigma^2)\right] \\
&= \frac{N}{4}\big[(1+\delta_N)\mathrm{Tr}(\Sigma^2) - 2\mathrm{Tr}(\Sigma) + p\big].
\end{aligned}$$

Hence

$$\begin{aligned}
m_\Sigma - m_I &= \frac{N}{4}\Big[(1+\delta_N)\mathrm{Tr}(\Sigma^2 - I) - 2\mathrm{Tr}(\Sigma - I)\Big] \\
&= \frac{N}{4}\big[\|\Sigma - I\|_F^2 + \delta_N \mathrm{Tr}(\Sigma^2 - I)\big].
\end{aligned}$$

(3). By the Plancherel's theorem (i.e., (Chatterjee, 2014b, formula (6.2))), we have

$$\begin{aligned}
\sigma^2_{I;\mathrm{LNW}} &= \sum_{(ij)} \big[\mathbb{E}\partial_{(ij)}T(X)\big]^2 + \frac{1}{2!}\sum_{(i_1j_1)(i_2j_2)} \big[\mathbb{E}\partial_{(i_1j_1)(i_2j_2)}T(X)\big]^2 \\
&\quad + \frac{1}{3!}\sum_{(i_1j_1)(i_2j_2)(i_3j_3)} \big[\mathbb{E}\partial_{(i_1j_1)(i_2j_2)(i_3j_3)}T(X)\big]^2 \\
&\quad + \frac{1}{4!}\sum_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)} \big[\mathbb{E}\partial_{(i_1j_1)(i_2j_2)(i_3j_3)(i_4j_4)}T(X)\big]^2 \\
&\equiv (I) + (II) + (III) + (IV).
\end{aligned}$$

Terms (I) - (IV) are handled as follows:

- To handle (I), note that

$$\mathbb{E}\partial_{(ij)}T(X) = \mathbb{E}e_j^\top(S-I)X_i - \mathbb{E}\big[(\mathrm{Tr}(S)/N)X_{ij}\big].$$

  The first term satisfies

$$\begin{aligned}
\mathbb{E}e_j^\top(S-I)X_i &= \mathbb{E}e_j^\top\left(\frac{1}{N}\sum_{k=1}^N X_kX_k^\top\right)X_i = N^{-1}e_j^\top\mathbb{E}(X_i\cdot\|X_i\|^2) \\
&= N^{-1}e_j^\top\mathbb{E}\left(\frac{X_i}{\|X_i\|}\cdot\|X_i\|^3\right) = N^{-1}e_j^\top\mathbb{E}\left(\frac{X_i}{\|X_i\|}\right)\cdot\mathbb{E}\|X_i\|^3 = 0.
\end{aligned}$$

A similar identity holds for the second term, so $(I) = 0$.

- $(II) \lesssim p/N = \mathfrak{o}(p^2)$ by noting that $\mathbb{E}\partial_{(i_1 j_1)(i_2 j_2)} T(X) = (N^{-1} - 2N^{-2}) \cdot \delta_{i_1 i_2}\delta_{j_1 j_2}$.

- $(III) = 0$ by direct calculation.

- $(IV) = 6p^2\big(1 + \mathfrak{o}(1)\big)$ by direct calculation.

The proof is now complete by collecting all of the estimates.

(4). By (1)-(3), $\|\Sigma\|_{\mathrm{op}} \le \|\Sigma - I\|_F + 1$ and the condition $p/N \le M$, we only need to show that

$$\frac{\sqrt{N}\|\Sigma - I\|_F^2 \bigvee \sqrt{N\|\Sigma - I\|_F^2}}{\big(N\|\Sigma - I\|_F^2 - N\delta_N |\mathrm{Tr}(\Sigma^2 - I)|\big)_+ \bigvee \sigma_{I;\mathrm{LNW}}} \le \frac{C_M}{(\sigma_{I;\mathrm{LNW}} \wedge N)^{1/2}}. \tag{D.32}$$

Note that with $\{\lambda_j\}_{j=1}^p$ denoting the eigenvalues of Σ,

$$\begin{aligned} |\mathrm{Tr}(\Sigma^2 - I)| &= \left|\sum_{j=1}^p (\lambda_j^2 - 1)\right| \le \max_j(\lambda_j + 1) \cdot \sum_{j=1}^p |\lambda_j - 1| \\ &\le \sqrt{p}(\|\Sigma\|_{\mathrm{op}} + 1)\|\Sigma - I\|_F \lesssim_M \sqrt{N}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F, \end{aligned} \tag{D.33}$$

so for N large enough, (D.32) is satisfied provided that

$$\frac{\sqrt{N}\|\Sigma - I\|_F^2 \bigvee \sqrt{N\|\Sigma - I\|_F^2}}{\big(N\|\Sigma - I\|_F^2 - C_M'\sqrt{N}\|\Sigma - I\|_F\big)_+ \bigvee \sigma_{I;\mathrm{LNW}}} \le \frac{C_M}{(\sigma_{I;\mathrm{LNW}} \wedge N)^{1/2}}. \tag{D.34}$$

To see this, note that the left hand side of the above display is bounded, up to a constant that may depend on M, by

$$\begin{aligned} &\mathbf{1}_{\sqrt{N}\|\Sigma - I\|_F \le 2C_M'} \frac{1}{\sigma_{I;\mathrm{LNW}}} + \mathbf{1}_{\sqrt{N}\|\Sigma - I\|_F > 2C_M'} \frac{\sqrt{N}\|\Sigma - I\|_F^2 \bigvee \sqrt{N\|\Sigma - I\|_F^2}}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LNW}}} \\ &\lesssim \frac{1}{\sigma_{I;\mathrm{LNW}}} + \frac{\sqrt{N}\|\Sigma - I\|_F^2}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LNW}}} + \frac{\sqrt{N\|\Sigma - I\|_F^2}}{N\|\Sigma - I\|_F^2 \vee \sigma_{I;\mathrm{LNW}}} \\ &\le \frac{1}{\sigma_{I;\mathrm{LNW}}} + \frac{1}{N^{1/2}} + \frac{1}{\inf_{x \ge 0}\left(x \vee \frac{\sigma_{I;\mathrm{LNW}}}{x}\right)} \le \text{RHS of (D.34)}. \end{aligned}$$

This completes the proof. $\square$

D.4.4 Completing the proof for power expansion

Proof of Theorem 19. Abbreviate Ψ_{LNW} by Ψ. By Theorem 18 and Proposition 23, we have

$$\left|\mathbb{E}_\Sigma \Psi(X) - \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \left(\|\Sigma - I\|_F^2 + Q_{\mathrm{LNW}}(\Sigma)\right)}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right)\right| \leq C \cdot p^{-1/3}.$$

We only need to remove the residual term $Q_{\mathrm{LNW}}(\Sigma)$. To see this, note that by (D.33),

$$|Q_{\mathrm{LNW}}(\Sigma)| \leq C_M N^{-1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F.$$

So using Lemma 45 we have

$$\Delta P \leq \frac{C_{\alpha,M}(\|\Sigma - I\|_F \vee 1)}{N^{1/2}\|\Sigma - I\|_F},$$

where

$$\begin{aligned}\Delta P \equiv \mathbb{P}&\left(\mathcal{N}\left(\frac{N \cdot \left(\|\Sigma - I\|_F^2 + Q_{\mathrm{LNW}}(\Sigma)\right)}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right) \\ &- \mathbb{P}\left(\mathcal{N}\left(\frac{N \cdot \|\Sigma - I\|_F^2}{4\sigma_{I;\mathrm{LNW}}}, 1\right) > z_\alpha\right).\end{aligned}$$

On the other hand, by anti-concentration of normal random variable,

$$\Delta P \leq C_M \frac{N^{1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F}{\sigma_{I;\mathrm{LNW}}}.$$

Hence

$$\begin{aligned}\Delta P &\lesssim_{\alpha,M} \frac{(\|\Sigma - I\|_F \vee 1)}{N^{1/2}\|\Sigma - I\|_F} \bigwedge \frac{N^{1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F}{\sigma_{I;\mathrm{LNW}}} \\ &\leq \mathbf{1}_{\|\Sigma - I\|_F > 1} \frac{1}{N^{1/2}} + \mathbf{1}_{\|\Sigma - I\|_F \leq 1}\left[\frac{1}{N^{1/2}\|\Sigma - I\|_F} \bigwedge \frac{N^{1/2}\|\Sigma - I\|_F}{\sigma_{I;\mathrm{LNW}}}\right] \\ &\leq \frac{1}{N^{1/2}} + \frac{1}{\inf_{\alpha \geq 0}\left(x \vee \frac{\sigma_{I;\mathrm{LNW}}}{x}\right)} \asymp \frac{1}{(\sigma_{I;\mathrm{LNW}} \wedge N)^{1/2}}.\end{aligned}$$

Similarly we may get a lower bound for ΔP. The proof is complete. □

D.5 Proofs for Section 5.3.4

In the proof of this subsection, we write $S_n \equiv n^{-1}\sum_{k=1}^n X_k X_k^\top$, where X_i's are i.i.d. $\mathcal{N}(0, \Sigma)$.

D.5.1 Evaluation of derivatives

Lemma 55. *Recall the form of $T_{\mathrm{CM}}(X)$ in (5.25). Then for any $(i,j),(i',j') \in [n]\times[p]$,*

1. $\partial_{(ij)} T_{\mathrm{CM}}(X) = \frac{2n}{n-1} X_i^\top (S_n - I) e_j - \frac{2}{n-1}(\|X_i\|^2 - 1) X_{ij}$.

2. $\partial_{(ij),(i'j')} T_{\mathrm{CM}}(X) = \frac{2n}{n-1}\big[\delta_{ii'}(S_n - I)_{jj'} + n^{-1} X_i^\top X_{i'} \delta_{jj'} + n^{-1} X_{ij'} X_{i'j}\big]$
$- \frac{2}{n-1}\big[2\delta_{ii'} X_{ij'} X_{ij} + (\|X_i\|^2 - 1)\delta_{ii'}\delta_{jj'}\big]$.

Proof. (1). Note that for any $1 \le k < \ell \le n$, we have

$$\begin{aligned}
\partial_{(ij)} h(X_k, X_\ell) &= \partial_{(ij)} (X_k^\top X_\ell)^2 - \partial_{(ij)} (X_k^\top X_k + X_\ell^\top X_\ell) \\
&= 2(X_k^\top X_\ell)(\delta_{ki} X_{\ell j} + \delta_{\ell i} X_{kj}) - \big(2\delta_{ik} X_{ij} + 2\delta_{\ell i} X_{ij}\big) \\
&= 2\delta_{ki}\big[(X_k^\top X_\ell) X_{\ell j} - X_{ij}\big] + 2\delta_{\ell i}\big[(X_k^\top X_\ell) X_{kj} - X_{ij}\big].
\end{aligned}$$

The above display entails that

$$\begin{aligned}
\partial_{(ij)} T_{\mathrm{CM}}(X) &= \frac{n}{2}\binom{n}{2}^{-1} \sum_{k<\ell} \partial_{(ij)} h(X_k, X_\ell) \\
&= \frac{n}{2}\binom{n}{2}^{-1} \sum_{k<\ell} 2\Big[\delta_{ki}\big((X_k^\top X_\ell) X_{\ell j} - X_{ij}\big) + \delta_{\ell i}\big((X_k^\top X_\ell) X_{kj} - X_{ij}\big)\Big] \\
&= \frac{2}{n-1} \sum_{k\in[n]:k\neq i} \big[(X_i^\top X_k) X_{kj} - X_{ij}\big] \\
&= \frac{2}{n-1} X_i^\top \Big[\sum_{k\in[n]:k\neq i} (X_k X_k^\top - I)\Big] e_j \\
&= \frac{2n}{n-1} X_i^\top (S_n - I) e_j - \frac{2}{n-1}\big(\|X_i\|^2 - 1\big) X_{ij}.
\end{aligned}$$

(2). By (1),

$$\begin{aligned}
\partial_{(ij),(i'j')} T_{\mathrm{CM}}(X) &= \frac{2n}{n-1} \partial_{(i'j')} \big[X_i^\top (S_n - I) e_j\big] - \frac{2}{n-1} \partial_{(i'j')} \Big[(\|X_i\|^2 - 1) X_{ij}\Big] \\
&= \frac{2n}{n-1}\Big[\delta_{ii'}(S_n - I)_{jj'} + n^{-1} X_i^\top X_{i'} \delta_{jj'} + n^{-1} X_{ij'} X_{i'j}\Big] \\
&\quad - \frac{2}{n-1}\Big[2\delta_{ii'} X_{ij'} X_{ij} + (\|X_i\|^2 - 1)\delta_{ii'}\delta_{jj'}\Big].
\end{aligned}$$

The proof is then completed. $\square$

D.5.2 Normal approximation

Proof of Theorem 20. We abbreviate T_{CM} as T. We first bound the operator norm of the Hessian. By Lemma 55-(2), we have

$$
\begin{aligned}
\partial_{(ij),(i'j')}T(X) = \frac{2n}{n-1}\delta_{ii'}(S_n - I)_{jj'} + \frac{2}{n-1}X_i^\top X_{i'}\delta_{jj'} + \frac{2}{n-1}X_{ij'}X_{i'j} \\
- \frac{4}{n-1}\delta_{ii'}X_{ij}X_{ij'} - \frac{2}{n-1}(\|X_i\|^2 - 1)\delta_{ii'}\delta_{jj'} \\
\equiv (T_{2,1} + T_{2,2} + T_{2,3} - T_{2,4} - T_{2,5})_{(ij),(i'j')}.
\end{aligned}
$$

In view of the proof of Theorem 18, we have $\mathbb{E}\|T_{2,1}\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_{2,2}\|_{\mathrm{op}}^4 \vee \mathbb{E}\|T_{2,3}\|_{\mathrm{op}}^4 = \mathcal{O}((1\vee y)^4)$. To handle $T_{2,4}$, note that

$$
\begin{aligned}
\|T_{2,4}\|_{\mathrm{op}} &\lesssim n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \Big| \sum_{(ij),(i'j')} u_{ij}u_{i'j'}\delta_{ii'}X_{ij}X_{ij'} \Big| \\
&= n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \Big| \sum_{i,j,j'} u_{ij}u_{ij'}X_{ij}X_{ij'} \Big| \\
&= n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \Big| \sum_i \Big(\sum_j u_i^\top e_j e_j^\top X_i \Big) \cdot \Big(\sum_{j'} u_i^\top e_{j'} e_{j'}^\top X_i \Big) \Big| \\
&= n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \sum_i \big(u_i^\top X_i\big)^2 = n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \sum_i \big(e_i^\top U X^\top e_i\big)^2 \\
&= n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \mathrm{Tr}(UX^\top X U^\top) \\
&= n^{-1} \cdot \sup_{u\in B_{n\times p}(1)} \|XU^\top\|_F^2 \le n^{-1}\|X\|_{\mathrm{op}}^2 = \|S_n\|_{\mathrm{op}}.
\end{aligned}
$$

Hence by Lemma 51, we have $\mathbb{E}\|T_{2,4}\|_{\mathrm{op}}^4 \lesssim (y \vee \sqrt{y})^4 = \mathcal{O}((1 \vee y)^4)$. To handle $T_{2,5}$, note that

$$\begin{aligned}
\|T_{2,5}\|_{\mathrm{op}} &\lesssim n^{-1} \cdot \sup_{u \in B_{n\times p}(1)} \Big| \sum_{(ij),(i'j')} u_{ij} u_{i'j'} (\|X_i\|^2 - 1)\delta_{ii'}\delta_{jj'} \Big| \\
&= n^{-1} \cdot \sup_{u \in B_{n\times p}(1)} \Big| \sum_{ij} u_{ij}^2 (\|X_i\|^2 - 1) \Big| \\
&= n^{-1} \cdot \sup_{u \in B_{n\times p}(1)} \Big| \sum_{i} \|u_i\|^2 (\|X_i\|^2 - 1) \Big| \\
&\le n^{-1} \cdot \sup_{u \in B_{n\times p}(1)} \sum_{i} \|u_i\|^2 \|X_i\|^2 + n^{-1} \\
&= n^{-1} \cdot \max_{i \in [n]} \|X_i\|^2 + n^{-1}.
\end{aligned}$$

This entails that, with $\|X_i\|^2$ following $\chi^2(p)$,

$$\mathbb{E}\|T_{2,5}\|_{\mathrm{op}}^4 \lesssim n^{-4}\mathbb{E}\big(\max_{i\in[n]} \|X_i\|^8\big) + n^{-4} \lesssim (y^4 \log^4 n) \vee 1.$$

By putting together the estimates for $T_{2,1}$-$T_{2,5}$, we have that

$$\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \lesssim y^4 \log^4 n \vee 1. \tag{D.35}$$

Next we bound the norm of the gradient. By Lemma 55-(1), we have

$$\begin{aligned}
\partial_{(ij)} T(X) &= \frac{2n}{n-1} X_i^\top (S_n - I) e_j - \frac{2}{n-1} (\|X_i\|^2 - 1) X_{ij} \\
&\equiv (T_{1,1} - T_{1,2})_{(ij)}.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}\|\nabla T(X)\|_F^2 &= \mathbb{E}\|T_{1,1}\|_F^2 + \mathbb{E}\|T_{1,2}\|_F^2 - 2\mathbb{E}\mathrm{Tr}(T_{1,1}^\top T_{1,2}) \\
&\equiv (I) + (II) - 2(III).
\end{aligned}$$

For (I), we have by Lemmas 63 and 64,

$$\begin{aligned}
(I) &= \Big(\frac{2n}{n-1}\Big)^2 \cdot n \cdot \mathbb{E}\mathrm{Tr}\Big[\big(S_n - I\big)^2 S_n\Big] \\
&= \Big(\frac{2n}{n-1}\Big)^2 \cdot n \cdot \big[\mathbb{E}\mathrm{Tr}(S_n^3) - 2\mathbb{E}\mathrm{Tr}(S_n^2) + \mathbb{E}\mathrm{Tr}(S_n)\big] \\
&= \Big(\frac{2n}{n-1}\Big)^2 \cdot n \cdot (py^2 + py + 3y^2 + y + 4yn^{-1}) \\
&= \Big(\frac{2n}{n-1}\Big)^2 \cdot n\big[n^{-2}p^3 + n^{-1}p^2(1 + \mathcal{O}(n \wedge p)^{-1})\big] \\
&= \frac{4n}{(n-1)^2}p^3 + 4p^2\big[1 + \mathcal{O}(n \wedge p)^{-1}\big].
\end{aligned}$$

For (II), we have

$$\begin{aligned}
(II) &= \Big(\frac{2}{n-1}\Big)^2 \mathbb{E}\sum_{i,j} \big(\|X_i\|^2 - 1\big)^2 X_{i,j}^2 \\
&= \Big(\frac{2}{n-1}\Big)^2 \cdot n \cdot \mathbb{E}\big(\|X_1\|^6 - 2\|X_1\|^4 + \|X_1\|^2\big) \\
&= \frac{4n}{(n-1)^2}p^3 + \mathcal{O}(n^{-1}p^2).
\end{aligned}$$

For (III), we have

$$\begin{aligned}
(III) &= \frac{2n}{n-1} \cdot \frac{2}{n-1} \cdot n \cdot \mathbb{E}X_1^\top(S_n - I)X_1(\|X_1\|^2 - 1) \\
&= \frac{4n^2}{(n-1)^2}\left[n^{-1}\mathbb{E}\bigg(X_1^\top \sum_j X_j X_j^\top X_1(\|X_1\|^2 - 1)\bigg) - \mathbb{E}\|X_1\|^4 + \mathbb{E}\|X_1\|^2\right] \\
&= \frac{4n^2}{(n-1)^2}\bigg[n^{-1}\Big(\mathbb{E}\|X_1\|^6 - \mathbb{E}\|X_1\|^4 + (n-1)\mathbb{E}\|X_1\|^4 - (n-1)\mathbb{E}\|X_1\|^2\Big) \\
&\qquad\qquad - \mathbb{E}\|X_1\|^4 + \mathbb{E}\|X_1\|^2\bigg] \\
&= \frac{4n}{(n-1)^2}\big(\mathbb{E}\|X_1\|^6 - 2\mathbb{E}\|X_1\|^4 + \mathbb{E}\|X_1\|^2\big) \\
&= \frac{4n}{(n-1)^2}p^3 + \mathcal{O}(n^{-1}p^2).
\end{aligned}$$

Combine the estimates for (I)-(III) to yield that $\mathbb{E}\|\nabla T(X)\|_F^2 = 4p^2(1 + \mathcal{O}(n \wedge p)^{-1})$. Then by following the same argument as in the proof of Theorem 18 and using (D.35), we again arrive at

$$\begin{aligned}\mathbb{E}\|\nabla T(X)\|_F^4 &= \big(\mathbb{E}\|\nabla T(X)\|_F^2\big)^2 + \mathcal{O}\big(\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4\big)\\ &= 16p^4\big(1+\mathcal{O}(n\wedge p)^{-1}\big).\end{aligned}$$

The rest of the proof follows from the same lines in the proof of Theorem 14, with the help of the variance formula in Proposition 24-(3). The normal approximation error bound then becomes a constant multiple of

$$\frac{(y\log n \vee 1)\cdot p}{p^2} = \frac{\frac{p\log n}{n}\vee 1}{p} = \frac{\log n}{n}\bigvee\frac{1}{p},$$

as desired. □

D.5.3 Ratio control

Proof of Proposition 24. (1). Recall that $Z_1,\ldots,Z_n$ are i.i.d. samples from $\mathcal{N}(0,I_p)$. Let $S_X \equiv S_n = n^{-1}\sum_{i=1}^n X_iX_i^\top$ and $S_Z \equiv n^{-1}\sum_{i=1}^n Z_iZ_i^\top$. Then for any $(i,j)\in[n]\times[p]$, Lemma 55-(1) implies that

$$\begin{aligned}&\big(\mathcal{T}_\Sigma(Z)\big)_{(ij)} = \sum_{\bar{j}}\big(\nabla T(X)\big)_{(i\bar{j})}(\Sigma^{1/2})_{\bar{j}j}\\ &= \frac{2n}{n-1}\sum_{\bar{j}} X_i^\top(S_X-I)e_{\bar{j}}e_{\bar{j}}^\top\Sigma^{1/2}e_j - \frac{2}{n-1}(\|X_i\|^2-1)\sum_{\bar{j}}X_i^\top e_{\bar{j}}e_{\bar{j}}^\top\Sigma^{1/2}e_j\\ &= \frac{2n}{n-1}X_i^\top(S_X-I)\Sigma^{1/2}e_j - \frac{2}{n-1}(\|X_i\|^2-1)X_i^\top\Sigma^{1/2}e_j\\ &= \frac{2n}{n-1}Z_i^\top(\Sigma S_Z\Sigma-\Sigma)e_j - \frac{2}{n-1}(\|\Sigma^{1/2}Z_i\|^2-1)Z_i^\top\Sigma e_j.\end{aligned}$$

This entails that

$$\begin{aligned}\|\mathcal{T}_\Sigma(Z)-\mathcal{T}_I(Z)\|_F^2 &\lesssim \sum_{i,j}\Big[Z_i^\top(\Sigma S_Z\Sigma-\Sigma-S_Z+I)e_j\Big]^2\\ &\qquad + n^{-2}\sum_{i,j}\Big[(\|\Sigma^{1/2}Z_i\|^2-1)Z_i^\top\Sigma e_j - (\|Z_i\|^2-1)Z_i^\top e_j\Big]^2\\ &\equiv V_1+V_2.\end{aligned}$$

To handle V_1, note that

$$\begin{aligned}
V_1 &= n\mathrm{Tr}\Big[(\Sigma S_Z \Sigma - \Sigma - S_Z + I)^2 S_Z\Big] \\
&\leq n\|S_Z\|_{\mathrm{op}} \cdot \|(\Sigma - I)S_Z\Sigma + S_Z(\Sigma - I) - (\Sigma - I)\|_F^2 \\
&\lesssim n\|\Sigma - I\|_F^2 \cdot \|S_Z\|_{\mathrm{op}}\big(\|S_Z\|_{\mathrm{op}}^2 \vee 1\big) \cdot \big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big).
\end{aligned}$$

Hence by Lemma 51, we have $\mathbb{E}V_1 \lesssim n(1 \vee y^3)\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2$.

To handle V_2, note that

$$\begin{aligned}
n^2 V_2 \lesssim &\sum_{i,j}\Big[(\|\Sigma^{1/2}Z_i\|^2 - 1)Z_i^\top \Sigma e_j - (\|Z_i\|^2 - 1)Z_i^\top \Sigma e_j\Big]^2 \\
&+ \sum_{i,j}\Big[(\|Z_i\|^2 - 1)(Z_i^\top \Sigma e_j - Z_i^\top e_j)\Big]^2 \equiv V_{2,1} + V_{2,2}.
\end{aligned}$$

To handle $V_{2,1}$, we have

$$\begin{aligned}
V_{2,1} &= \sum_{i,j}\big(Z_i^\top(\Sigma - I)Z_i\big)^2 \cdot \big(Z_i^\top \Sigma e_j\big)^2 \\
&= \sum_i \big(Z_i^\top(\Sigma - I)Z_i\big)^2 \cdot Z_i^\top \Sigma^2 Z_i \\
&= \sum_i \mathrm{Tr}\big((\Sigma - I)Z_i Z_i^\top(\Sigma - I)Z_i Z_i^\top \Sigma^2 Z_i Z_i^\top\big) \\
&\leq \sum_i \mathrm{Tr}\big((\Sigma - I)Z_i Z_i^\top(\Sigma - I)\big) \cdot \max_{i\in[n]}\|Z_i Z_i^\top \Sigma^2 Z_i Z_i^\top\|_{\mathrm{op}} \\
&\leq n \cdot \|\Sigma - I\|_F^2 \cdot \|S_Z\|_{\mathrm{op}} \cdot \|\Sigma\|_{\mathrm{op}}^2 \cdot \max_{i\in[n]}\|Z_i\|^4,
\end{aligned}$$

where in the above we repeatedly use the fact that $\mathrm{Tr}(AB) \leq \mathrm{Tr}(A)\|B\|_{\mathrm{op}}$ for any p.s.d. matrices A, B. Hence by Lemma 51, we have

$$\begin{aligned}
\mathbb{E}V_{2,1} &\leq n \cdot \|\Sigma - I\|_F^2 \cdot \|\Sigma\|_{\mathrm{op}}^2 \cdot \mathbb{E}^{1/2}\|S_Z\|_{\mathrm{op}}^2 \cdot \mathbb{E}^{1/2}\max_{i\in[n]}\|Z_i\|^8 \\
&\lesssim np^2(1 \vee y)(\log n)^2 \cdot \|\Sigma\|_{\mathrm{op}}^2\|\Sigma - I\|_F^2.
\end{aligned}$$

To handle $V_{2,2}$, we have

$$\begin{aligned}V_{2,2} &= \sum_i (\|Z_i\|^2-1)^2 Z_i^\top(\Sigma-I)^2 Z_i\\ &\lesssim n\cdot\|\Sigma-I\|_F^2\cdot\|S_Z\|_{\mathrm{op}}\cdot\big(\max_{i\in[n]}\|Z_i\|^4\vee 1\big).\end{aligned}$$

Hence the same bound as above implies that $\mathbb{E}V_{2,2}\lesssim np^2(1\vee y)(\log n)^2\big(\|\Sigma\|_{\mathrm{op}}^2\vee 1\big)\|\Sigma-I\|_F^2$. Combining the estimates of $\mathbb{E}V_{2,1}$ and $\mathbb{E}V_{2,2}$ completes the proof of claim (1).

(2) and (3). These follow directly from the mean and variance formula (5.26).

(4). By (1)-(3), it remains to prove that

$$\frac{\sqrt{n\big(\|\Sigma\|_{\mathrm{op}}^2\vee 1\big)\|\Sigma-I\|_F^2}}{n\|\Sigma-I\|_F^2\vee p}\leq\frac{C}{(n\wedge p)^{1/2}}$$

holds for some universal constant $C>0$. Using $\|\Sigma\|_{\mathrm{op}}\leq\|\Sigma-I\|_{\mathrm{op}}+1\leq\|\Sigma-I\|_F+1$, it suffices to prove

$$\frac{\sqrt{n}\|\Sigma-I\|_F^2\bigvee\sqrt{n\|\Sigma-I\|_F^2}}{n\|\Sigma-I\|_F^2\vee p}\leq\frac{C}{(n\wedge p)^{1/2}}.$$

This is weaker than the proven inequality (D.34). $\square$

D.5.4 Completing the proof for power expansion

Proof of Theorem 21. By Corollary 3 and Proposition 24, the error of power expansion is bounded by a constant multiple of

$$\begin{aligned}&\Big(\frac{\log n}{n}\vee\frac{1}{p}\Big)+\Big(\frac{1\vee y}{(n\wedge p)^{1/3}}\Big)\bigvee\Big(\frac{y^{2/3}(1\vee y)^{1/3}(\log n)^{2/3}}{(n\wedge p)^{1/3}}\Big)\\ &\equiv\Big(\frac{\log n}{n}\vee\frac{1}{p}\Big)+(I)\vee(II).\end{aligned}$$

Using the condition $p/n\leq M$, we have $(I)\lesssim_M p^{-1/3}$. For (II), we have $(II)\leq p^{1/3}n^{-2/3}(\log n)^{2/3}\leq n^{-1/3}(\log n)^{2/3}$ when $p\leq n$ and $(II)\lesssim_M n^{-1/3}(\log n)^{2/3}$ otherwise. The proof is complete. $\square$

D.6 Proofs for Section 5.4.1

D.6.1 Evaluation of derivatives

Lemma 56. *Recall the form of $T_{\mathrm{LRT},s}(X)$ in (5.28) and the definition of $b(S)$ in (5.29). Then for any $(i,j),(i',j') \in [N]\times[p]$,*

1. $\partial_{(ij)} T_{\mathrm{LRT},s}(X) = \big(X(I-S^{-1})\big)_{(ij)} + \big(1/b(S)-1\big)X_{ij} = e_j^\top\big[(I-S^{-1})X_i + \big(1/b(S)-1\big)X_i\big]$.

2. $\partial_{(ij),(i'j')} T_{\mathrm{LRT},s}(X) = N^{-1}X_i^\top S^{-1}(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top)S^{-1}e_j + \delta_{ii'}e_j^\top(I-S^{-1})e_{j'} + \big(1/b(S) - 1\big)\delta_{ii'}\delta_{jj'} - (2/Np)X_{ij}X_{i'j'}/b^2(S)$.

Proof. (1). We shorthand $T_{\mathrm{LRT},s}(X)$ as T. By definition, (D.7) and (D.9), we have

$$
\begin{aligned}
\partial_{(ij)} T(X) &= \frac{N}{2}\big(p\cdot\partial_{(ij)}\log\mathrm{Tr}(S) - \partial_{ij}\log\det S\big)\\
&= \frac{N}{2}\left(p\frac{\partial_{(ij)}\mathrm{Tr}(S)}{\mathrm{Tr}(S)} - \sum_{k,\ell=1}^{p}\frac{\partial\log\det S}{\partial S_{k\ell}}\frac{\partial S_{k\ell}}{\partial X_{ij}}\right)\\
&= \frac{N}{2}\left[\frac{2p}{N}\frac{X_{ij}}{\mathrm{Tr}(S)} - \sum_{k,\ell=1}^{p}(S^{-1})_{k\ell}\cdot\frac{1}{N}\big(\delta_{kj}X_{i\ell} + \delta_{\ell j}X_{ik}\big)\right]\\
&= \frac{p}{\mathrm{Tr}(S)}X_{ij} - \sum_{k=1}^{p}(S^{-1})_{kj}X_{ik}\\
&= \big(X(I-S^{-1})\big)_{ij} + \left(\frac{p}{\mathrm{Tr}(S)} - 1\right)X_{ij}.
\end{aligned}
$$

(2). By the previous part, we have

$$
\partial_{(ij),(i'j')} T(X) = \partial_{(i'j')}\big(X(I-S^{-1})\big)_{ij} + \partial_{(i'j')}\left(\frac{p}{\mathrm{Tr}(S)} - 1\right)X_{ij} \equiv (I) + (II).
$$

The first term above is already calculated in Lemma 52-(2):

$$
(I) = N^{-1}X_i^\top S^{-1}(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top)S^{-1}e_j + \delta_{ii'}e_j^\top(I-S^{-1})e_{j'}.
$$

So we only need to evaluate the second term:

$$\begin{aligned}(II) &= p\cdot\partial_{(i'j')}\mathrm{Tr}^{-1}(S)\cdot X_{ij} + \Big(\frac{p}{\mathrm{Tr}(S)}-1\Big)\partial_{(i'j')}X_{ij}\\ &= -p\cdot\partial_{(i'j')}\mathrm{Tr}(S)\cdot X_{ij}\cdot\mathrm{Tr}^{-2}(S) + \Big(\frac{p}{\mathrm{Tr}(S)}-1\Big)\delta_{ii'}\delta_{jj'}\\ &= -\frac{2p}{N}X_{ij}X_{i'j'}\cdot\mathrm{Tr}^{-2}(S) + \Big(\frac{p}{\mathrm{Tr}(S)}-1\Big)\delta_{ii'}\delta_{jj'}.\end{aligned}$$

The proof is complete. □

D.6.2 Normal approximation

Proof of Theorem 22. We abbreviate $T_{\mathrm{LRT},s}(X)$ as T. First we bound the norm for the gradient. Comparing Lemmas 52-(1) and 56-(1), we only need to control

$$\begin{aligned}\mathbb{E}\|\big(b^{-1}(S)-1\big)X\|_F^4 &= \mathbb{E}\Big(N\big(b^{-1}(S)-1\big)^2\mathrm{Tr}(S)\Big)^2\\ &\leq N^2p^2\cdot\mathbb{E}^{1/2}b^4(S)\cdot\mathbb{E}^{1/2}\big(b^{-1}(S)-1\big)^8\\ &\lesssim N^2p^2\cdot\Big(\frac{p}{N}\Big)^2 = p^4.\end{aligned}$$

The inequality in the final line of the above display follows as

$$\mathbb{E}b^4(S)\leq\mathbb{E}\|S\|_{\mathrm{op}}^4\lesssim 1, \tag{D.36}$$

$$\mathbb{E}\big(b^{-1}(S)-1\big)^8 = \mathbb{E}^{1/2}b^{-16}(S)\cdot\mathbb{E}^{1/2}\big(b(S)-1\big)^{16}\overset{(*)}{\lesssim}(pN^{-1})^4. \tag{D.37}$$

Here $(*)$ follows from Lemma 63-(3). Now by combining with (D.13) derived in the proof of Theorem 14, we see that $\mathbb{E}\|\nabla T(X)\|_F^4\lesssim p^4$.

Next we bound the spectral norm of the Hessian. Comparing Lemmas 52-(1) and 56-(1), we only need to control the spectral norms of T_4 and T_5, where

$$\begin{aligned}(T_4)_{(ij),(i'j')} &\equiv \big(b^{-1}(S)-1\big)\delta_{ii'}\delta_{jj'},\\ (T_5)_{(ij),(i'j')} &\equiv -\frac{2}{Np}X_{ij}X_{i'j'}\cdot b^{-2}(S).\end{aligned}$$

For T_4, clearly $\|T_4\|_{\mathrm{op}} = |1/b(S)-1|$, so $\mathbb{E}\|T_4\|_{\mathrm{op}}^4 = \mathbb{E}\big(1/b(S)-1\big)^4 \lesssim (p/N)^2$ by (D.36). For T_5, note that

$$\begin{aligned}\|T_5\|_{\mathrm{op}} &= \frac{2}{Np\cdot b^2(S)} \sup_{u,v\in B_{N\times p}(1)} \bigg|\sum_{(ij),(i'j')} u_{ij}X_{ij}X_{i'j'}v_{i'j'}\bigg|\\ &= \frac{2}{Np\cdot b^2(S)}\|X\|_F^2 = \frac{2}{b(S)}.\end{aligned}$$

So $\mathbb{E}\|T_5\|_{\mathrm{op}}^4 \lesssim \mathbb{E}b^{-4}(S) = \mathcal{O}(1)$ by Lemma 63-(3). By combining with (D.17) derived in the proof of Theorem 14, we see that $\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 = \mathcal{O}(1)$. The rest of the proof proceeds along the lines in the proof of Theorem 14, with the help of the variance formula in Proposition 25-(3). □

D.6.3 Ratio control

Proof of Proposition 25. We abbreviate $(T_{\mathrm{LRT},s}, m_{\Sigma;\mathrm{LRT},s}, \sigma_{\Sigma;\mathrm{LRT},s}, V_{\Sigma;\mathrm{LRT},s}$ as $(T, m_{\Sigma;s}, \sigma_{\Sigma;s}, V_{\Sigma;s})$, and assume without loss of generality that $b(\Sigma) = \mathrm{Tr}(\Sigma)/p = 1$ (otherwise we may replace Σ by $\Sigma\cdot b^{-1}(\Sigma)$).

(1). By Lemma 56, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_iZ_i^\top$, we have

$$\begin{aligned}\mathcal{T}_{\Sigma;s} &= \bigg[Z\Sigma^{1/2}(I-\Sigma^{-1/2}S_Z^{-1}\Sigma^{-1/2}) + \bigg(\frac{1}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-1\bigg)Z\Sigma^{1/2}\bigg]\Sigma^{1/2}\\ &= Z(\Sigma - S_Z^{-1}) + \bigg(\frac{1}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-1\bigg)Z\Sigma\\ &= \frac{Z\Sigma}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - ZS_Z^{-1}.\end{aligned}$$

Hence

$$\begin{aligned}V_{\Sigma;s}^2 = \mathbb{E}\|\mathcal{T}_{\Sigma;s} - \mathcal{T}_{I;s}\|_F^2 &= \mathbb{E}\bigg\|\frac{Z\Sigma}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{Z}{b(S_Z)}\bigg\|_F^2\\ &\leq 2\bigg\{\mathbb{E}\bigg[\bigg(\frac{1}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{b(S_Z)}\bigg)^2\|Z\Sigma\|_F^2\bigg]\\ &\qquad + \mathbb{E}\big[b^{-2}(S_Z)\|Z(\Sigma-I)\|_F^2\big]\bigg\}\\ &\equiv 2\mathbb{E}\big((I)+(II)\big).\end{aligned}$$

We bound (I) and (II) separately:

$$\begin{aligned}(I) &= b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-2}(S_Z)b^2\big((\Sigma-I)S_Z\big)\|Z\Sigma\|_F^2\\ &\le b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-2}(S_Z)\|S_Z\|_{\mathrm{op}}^2\cdot\big(\|\Sigma\|_F^2/p\big)\|Z\|_{\mathrm{op}}^2\|\Sigma-I\|_F^2;\\ (II) &\le b^{-2}(S_Z)\cdot\|Z\|_{\mathrm{op}}^2\|\Sigma-I\|_F^2.\end{aligned}$$

Using Lemmas 61 and 51, we have

$$V_{\Sigma;s}^2 \lesssim (p^{-1}\|\Sigma-I\|_F^2+1)N\|(\Sigma-I)\|_F^2.$$

On the other hand, a trivial bound for $V_{\Sigma;s}^2$ is

$$\begin{aligned}V_{\Sigma;s}^2 &= \mathbb{E}\bigg\|\frac{Z\Sigma}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})}-\frac{Z}{b(S_Z)}\bigg\|_F^2\\ &\lesssim \mathbb{E}b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})\|Z\Sigma\|_F^2+\mathbb{E}b^{-2}(S_Z)\|Z\|_F^2\\ &\lesssim N\big(\|\Sigma-I\|_F^2\vee p\big).\end{aligned}$$

Collecting the two bounds, we have

$$\begin{aligned}V_{\Sigma;s}^2 &\lesssim \big[\big(p^{-1}\|\Sigma-I\|_F^2+1\big)N\|(\Sigma-I)\|_F^2\big]\bigwedge N\big(\|\Sigma-I\|_F^2\vee p\big)\\ &\asymp N\|(\Sigma-I)\|_F^2.\end{aligned}$$

(2). As

$$m_{\Sigma;s}=\frac{N}{2}\Big[p\cdot\mathbb{E}\log\mathrm{Tr}(\Sigma S_Z)-\log\det(\Sigma)-p\log p-\mathbb{E}\log\det(S_Z)\Big],$$

by Lemma 62 we have

$$m_{\Sigma;s}-m_{I;s}=\frac{N}{2}\big[-\log\det(\Sigma)+Q_s(\Sigma)\big],$$

where

$$\begin{aligned}|Q_s(\Sigma)| &\equiv |p\big(\mathbb{E}\log\mathrm{Tr}(\Sigma S_Z)-\mathbb{E}\log\mathrm{Tr}(S_Z)\big)|\\ &\lesssim N^{-1}\Big\{1+b(\Sigma^2)+e^{-cN}\big[1+b^{1/2}(\Sigma^2)\big]\Big\}\\ &\lesssim N^{-1}\Big[1+b(\Sigma^2)\Big]\lesssim N^{-1}b(\Sigma^2),\end{aligned}$$

where the last inequality follows as $b(\Sigma^2) = p^{-1}\sum_{j=1}^p \lambda_j^2 \geq p^{-2}(\sum_{j=1}^p \lambda_j)^2 = 1$.

(3). Recall T_{LRT} defined in (5.19). Define

$$\Delta(X) \equiv T_{\mathrm{LRT}}(X) - T_{\mathrm{LRT},s}(X).$$

Then for any $\varepsilon > 0$, there exists some $C_\varepsilon > 0$ such that under the null (i.e., $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(0, I_p)$),

$$\big[(1-\varepsilon)\sigma^2_{I;\mathrm{LRT}} - C_\varepsilon \mathrm{Var}_I(\Delta)\big]_+ \leq \sigma^2_{I;\mathrm{LRT},s} \leq (1+\varepsilon)\sigma^2_{I;\mathrm{LRT}} + C_\varepsilon \mathrm{Var}_I(\Delta). \tag{D.38}$$

We will now bound $\mathrm{Var}_I(\Delta)$. By Lemmas 52-(1) and 56-(1), we have for any $i, j \in [N] \times [p]$

$$\partial_{(ij)}\Delta(X) = \partial_{(ij)}T_{\mathrm{LRT}}(X) - \partial_{(ij)}T_{\mathrm{LRT},s}(X) = X_{ij}\big[b^{-1}(S) - 1\big].$$

By the Gaussian-Poincaré inequality (Boucheron et al., 2013, Theorem 3.20),

$$\begin{aligned}
\mathrm{Var}_I \Delta(X) &\leq \mathbb{E}\big[b^{-1}(S) - 1\big]^2 \|X\|_F^2 = Np\mathbb{E}\big[b(S) - 1\big]^2 b^{-1}(S) \\
&\leq Np \cdot \mathbb{E}^{1/2}\big(b(S) - 1\big)^4 \cdot \mathbb{E}^{1/2} b^{-2}(S) \\
&\overset{(*)}{\lesssim} Np \cdot (Np)^{-1} = 1.
\end{aligned}$$

Here $(*)$ follows from Lemma 63-(3). Hence by choosing ε in (D.38) to be decaying to 0 slowly enough, $\sigma^2_{I;\mathrm{LRT}}$ and $\sigma^2_{I;\mathrm{LRT},s}$ share the same asymptotic formula in Proposition 21-(3).

(4). By (1)-(2), and using that $b(\Sigma^2) = \|\Sigma\|_F^2/p$, we only need to prove that for a given constant $C_0 > 0$, there exists some constant $C = C(C_0) > 0$ such that

$$\begin{aligned}
&\frac{\sqrt{N\|\Sigma - I\|_F^2}}{\Big(-N\log\det(\Sigma) - C_0(1 + \frac{\|\Sigma\|_F^2}{p}) - C_0 e^{-cN}\big(\frac{\|\Sigma\|_F}{p^{1/2}} + 1\big)\Big)_+ \bigvee \sigma_{I;s}} \\
&\qquad \leq \frac{C}{\big(\sigma_{I;s} \wedge N\big)^{1/2}}.
\end{aligned}$$

Equivalently, with $\lambda = (\lambda_1, \ldots, \lambda_p) \in (0, \infty)^p$ and $\overline{\lambda} \equiv p^{-1}\sum_j \lambda_j = 1$, we only need to show

$$\frac{\sqrt{N\sum_j (\lambda_j - 1)^2}}{\Big(N\sum_j - \log(1 + (\lambda_j - 1)) - C_0 - C_0\big(\frac{\sum_j \lambda_j^2}{p}\big) - C_0 e^{-cN}\frac{(\sum_j \lambda_j^2)^{1/2}}{p^{1/2}}\Big)_+ \bigvee \sigma_{I;s}}$$

is at most a multiple of $\big(\sigma_{I;s} \wedge N\big)^{-1/2}$. Let $J \equiv \{j : |\lambda_j - 1| \le 1\}$ and $J^c \equiv \{j : |\lambda_j - 1| > 1\}$. As $|\lambda_j - 1| \lesssim p$, so the first term in the denominator becomes

$$
\begin{aligned}
& N \sum_j \big[-\log(1 + (\lambda_j - 1)) + (\lambda_j - 1)\big] - C_0 - C_0 \frac{\sum_j \lambda_j^2}{p} - C_0 e^{-cN} \frac{(\sum_j \lambda_j^2)^{1/2}}{p^{1/2}} \\
& \gtrsim N \sum_j (\lambda_j - 1)^2 \wedge |\lambda_j - 1| - C_1 p^{-1} \sum_j (\lambda_j - 1)^2 - C_2.
\end{aligned}
$$

Next, by breaking the summation in $\sum_j (\lambda_j - 1)^2$ into J and J^c, the above display equals

$$
\begin{aligned}
& N \sum_{j \in J} (\lambda_j - 1)^2 + N \sum_{j \in J} |\lambda_j - 1| - C_1 \frac{\sum_{j \in J} (\lambda_j - 1)^2 + \sum_{j \in J^c} (\lambda_j - 1)^2}{p} - C_2 \\
& \ge (N - C_1 p^{-1}) \sum_{j \in J} (\lambda_j - 1)^2 + (N - \mathcal{O}(1)) \sum_{j \in J^c} |\lambda_j - 1| - C_2 \\
& \ge \frac{N}{2} \sum_j (\lambda_j - 1)^2 \wedge |\lambda_j - 1| - C_2
\end{aligned}
$$

for N and p large enough. Hence with $\nu_j \equiv |\lambda_j - 1|$, we only need to show that for given $C_0 > 0$,

$$
\frac{\sqrt{N \sum_{j \in J} \nu_j^2} \bigvee \sqrt{N \sum_{j \in J^c} \nu_j^2}}{\Big(N \sum_{j \in J} \nu_j^2 + N \sum_{j \in J^c} \nu_j - C_0\Big)_+ \bigvee \sigma_{I;s}} \le \frac{C}{\big(\sigma_{I;s} \wedge N\big)^{1/2}}.
$$

Equivalently, we only need to show

$$
\frac{\sqrt{N \sum_{j \in J} \nu_j^2}}{\big(N \sum_{j \in J} \nu_j^2 - C_0\big)_+ \bigvee \sigma_{I;s}} \le \frac{C}{\sigma_{I;s}^{1/2}}, \tag{D.39}
$$

$$
\frac{\sqrt{N \sum_{j \in J^c} \nu_j^2}}{\big(N \sum_{j \in J^c} \nu_j - C_0\big)_+ \bigvee \sigma_{I;s}} \le \frac{C}{N^{1/2}}. \tag{D.40}
$$

To see these inequalities, note that the left side of (D.39) is bounded by

$$
\begin{aligned}
& \mathbf{1}_{N \sum_{j \in J} \nu_j^2 \le 2C_0} \frac{(2C_0)^{1/2}}{\sigma_{I;s}} + \mathbf{1}_{N \sum_{j \in J} \nu_j^2 > 2C_0} \frac{\sqrt{N \sum_{j \in J} \nu_j^2}}{(N/2) \sum_{j \in J} \nu_j^2 \bigvee \sigma_{I;s}} \\
& \lesssim \frac{1}{\sigma_{I;s}} + \frac{1}{\inf_{x \ge 0} \big(x \vee \frac{\sigma_{I;s}}{x}\big)} \lesssim \sigma_{I;s}^{-1/2}.
\end{aligned}
$$

Also, the left side of (D.40) is bounded by

$$\begin{aligned}
&\frac{\sqrt{N}\sum_{j\in J^c}\nu_j}{\big(N\sum_{j\in J^c}\nu_j - C_0\big)_+ \bigvee \sigma_{I;s}}\\
&\leq \mathbf{1}_{N\sum_{j\in J^c}\nu_j\leq 2C_0}\frac{(2C_0)^{1/2}}{\sqrt{N}\sigma_{I;s}} + \mathbf{1}_{N\sum_{j\in J^c}\nu_j> 2C_0}\frac{\sqrt{N}\sum_{j\in J^c}\nu_j}{N\sum_{j\in J^c}\nu_j\bigvee\sigma_{I;s}}\\
&\lesssim \frac{1}{\sqrt{N}\sigma_{I;s}} + \frac{1}{\sqrt{N}} \lesssim \frac{1}{N^{1/2}},
\end{aligned}$$

proving the claim. $\square$

D.6.4 Completing the proof for power expansion

Proof of Theorem 23. The proof is similar to that of Theorem 19, we provide some details for the convenience of the reader. Without loss of generality we assume $b(\Sigma) = 1$. Abbreviate $\Psi_{\mathrm{LRT},s}$ by Ψ and $Q_{\mathrm{LRT},s}(\Sigma)$ by $Q(\Sigma)$. By Theorem 22 and Proposition 25, we have

$$\bigg|\mathbb{E}_\Sigma\Psi(X) - \mathbb{P}\bigg(\mathcal{N}\bigg(\frac{N\cdot\big(-\log\det\big(\Sigma\big)+Q(\Sigma)\big)}{2\sigma_{I;s}},1\bigg)>z_\alpha\bigg)\bigg| \leq C\cdot p^{-1/3}.$$

We only need to remove the residual term $Q(\Sigma)$. To this end, we claim that

$$|\Delta P| \leq C_\alpha\Big[\frac{NQ(\Sigma)}{\sigma_{I;s}}\bigwedge\frac{Q(\Sigma)}{|\log\det\big(\Sigma\big)|}\Big]. \tag{D.41}$$

where

$$\begin{aligned}
\Delta P \equiv\ & \mathbb{P}\bigg(\mathcal{N}\bigg(\frac{N\cdot\big(-\log\det\big(\Sigma\big)+Q(\Sigma)\big)}{2\sigma_{I;s}},1\bigg)>z_\alpha\bigg)\\
&- \mathbb{P}\bigg(\mathcal{N}\bigg(\frac{-N\log\det\big(\Sigma\big)}{2\sigma_{I;s}},1\bigg)>z_\alpha\bigg).
\end{aligned}$$

Here the first bound in (D.41) is by anti-concentration of the normal distribution, and the second bound in (D.41) follows from Lemma 45.

Let $\{\lambda_j\}_{j=1}^p$ be the eigenvalues of Σ so that $\sum_{j=1}^p\lambda_j = p$. Then by (5.30), $Q(\Sigma)\lesssim 2(Np)^{-1}\sum_{j=1}^p\lambda_j^2$. Hence using the bound $\sigma_{I;s}\geq cp$, (D.41) entails that

$$|\Delta P| \leq C'_\alpha\cdot\Big[\frac{p^{-1}\sum_{j=1}^p\lambda_j^2}{p}\bigwedge\frac{(Np)^{-1}\sum_{j=1}^p\lambda_j^2}{\sum_{j=1}^p\lambda_j-\log\lambda_j-1}\Big]. \tag{D.42}$$

If $\max_j \lambda_j \le 10$, we use the first bound in (D.42) to conclude that $\Delta P \lesssim_\alpha p^{-1}$. Otherwise, by writing $J \equiv \{j \in [p] : |\lambda_j - 1| \ge 1\}$ and $J^c \equiv [p]\backslash J$, the second bound in (D.42) yields that

$$\begin{aligned}
|\Delta P| &\lesssim_\alpha \frac{(Np)^{-1}\sum_{j=1}^p \lambda_j^2}{\sum_{j=1}^p |\lambda_j - 1| \wedge (\lambda_j - 1)^2} \\
&\lesssim \frac{(Np)^{-1}\sum_{j\in J}(\lambda_j - 1)^2 + (Np)^{-1}(|J| + |J^c|)}{\sum_{j\in J}|\lambda_j - 1|} \\
&\le \frac{(Np)^{-1}\sum_{j\in J}(\lambda_j - 1)^2}{\sum_{j\in J}|\lambda_j - 1|} + \frac{N^{-1}}{\sum_{j\in J}|\lambda_j - 1|} \\
&\equiv (I) + (II).
\end{aligned}$$

Now $(II) \lesssim N^{-1}$ as $\max_j \lambda_j > 10$, and (I) satisfies

$$(I) \le (Np)^{-1} \max_{j\in J} |\lambda_j - 1| \lesssim N^{-1}$$

by using the trivial bound that $\max_j \lambda_j \le p$ due to the normalization $b(\Sigma) = 1$. The proof is complete. $\square$

D.7 Proofs for Section 5.4.2

D.7.1 Evaluation of derivatives

Lemma 57. *Recall the form of $T_J(X)$ in (5.31) and the definition of $b_\ell(S)$ in (5.29). Then the following hold:*

1. *For the first-order partial derivatives: for any $(i, j) \in [N] \times [p]$,*

$$\partial_{(ij)} T_J(X) = \Big(\frac{XS}{b^2(S)} - X \cdot \frac{b_2(S)}{b^3(S)}\Big)_{ij} = \frac{X_i^\top S e_j}{b^2(S)} - X_{ij}\frac{b_2(S)}{b^3(S)}.$$

2. For the second-order partial derivatives: for any $(i,j),(i',j') \in [N] \times [p]$,

$$\begin{aligned}
&\partial_{(ij),(i'j')} T_J(X) \\
&= b(S)^{-2}\big(N^{-1}\delta_{jj'} X_i^\top X_{i'} + N^{-1} X_{i'j} X_{ij'} + \delta_{ii'} S_{jj'}\big) - \delta_{ii'}\delta_{jj'} \frac{b_2(S)}{b^3(S)} \\
&\quad + X_{ij} X_{i'j'} \frac{6 b_2(S)}{b^4(S) N p} - \frac{4}{b^3(S) N p}\Big[X_i^\top S e_j \cdot X_{i'j'} + X_{i'}^\top S e_{j'} \cdot X_{ij}\Big].
\end{aligned}$$

Proof. We abbreviate $T_J(X)$ by $T(X)$ and write $b = b(S)$ in the proof if no confusion could arise.

(1). Note that $\partial_{ij} S(X) = N^{-1}(e_j X_i^\top + X_i e_j^\top)$, $\partial_{(ij)} \mathrm{Tr}(S) = 2N^{-1} X_{ij}$ and

$$\begin{aligned}
\partial_{(ij)} b(S) &= \frac{2}{Np} X_{ij}, \\
\partial_{(ij)} b_2(S) &= \frac{2}{Np} \mathrm{Tr}\big(S(e_j X_i^\top + X_i e_j^\top)\big) = \frac{4}{Np} X_i^\top S e_j.
\end{aligned} \tag{D.43}$$

For the first-order derivatives we have

$$\begin{aligned}
\partial_{(ij)} T(X) &= \frac{N}{4} \mathrm{Tr}\left[2\left(\frac{S}{b} - I\right) \partial_{(ij)}\left(\frac{S}{b}\right)\right] \\
&= \frac{N}{2} \mathrm{Tr}\left[\left(\frac{S}{b} - I\right) \cdot \frac{N^{-1}(e_j X_i^\top + X_i e_j^\top) b - 2(Np)^{-1} S X_{ij}}{b^2}\right] \\
&= \frac{1}{2b^2} \mathrm{Tr}\big[(S - bI)(e_j X_i^\top + X_i e_j^\top)\big] - \frac{X_{ij}}{b^3 p} \mathrm{Tr}\big[(S - bI) S\big] \\
&= \frac{(XS)_{ij}}{b^2} - \frac{X_{ij}}{b} - \left[\frac{X_{ij} b_2}{b^3} - \frac{X_{ij}}{b}\right] \\
&= \frac{(XS)_{ij}}{b^2} - X_{ij} \cdot \frac{b_2}{b^3} \equiv T_{1,(ij)}(X) - T_{2,(ij)}(X).
\end{aligned}$$

(2). For the second-order derivatives,

$$
\begin{aligned}
\partial_{(i'j')}T_{1,(ij)}(X) &= \frac{\partial_{(i'j')}(X_i^\top Se_j)b^2 - (X_i^\top Se_j)\partial_{(i'j')}b^2}{b^4} \\
&= \frac{\delta_{ii'}e_{j'}^\top Se_j + N^{-1}X_i^\top(e_{j'}X_{i'}^\top + X_{i'}e_{j'}^\top)e_j}{b^2} - \frac{4}{pN}\frac{X_i^\top Se_j \cdot X_{i'j'}}{b^3} \\
&= \frac{N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}S_{jj'}}{b^2} - \frac{4}{pN}\frac{X_i^\top Se_j \cdot X_{i'j'}}{b^3}, \\
\partial_{(i'j')}T_{2,(ij)}(X) &= \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3} + X_{ij}\cdot\partial_{(i'j')}\left[\frac{b_2}{b^3}\right] \\
&= \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3} + X_{ij}\cdot\left[\frac{4X_{i'}^\top Se_{j'}}{b^3Np} - \frac{6b_2X_{i'j'}}{b^4Np}\right] \\
&= \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3} - X_{ij}X_{i'j'}\frac{6b_2}{b^4Np} + 4X_{i'}^\top Se_{j'}\cdot X_{ij}\frac{1}{b^3Np}.
\end{aligned}
$$

Combining the above two displays, we have

$$
\begin{aligned}
\partial_{(ij),(i'j')}T(X) = b^{-2}\big(N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}S_{jj'}\big) - \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3} \\
+ X_{ij}X_{i'j'}\frac{6b_2}{b^4Np} - \frac{4}{b^3Np}\Big[X_i^\top Se_j\cdot X_{i'j'} + X_{i'}^\top Se_{j'}\cdot X_{ij}\Big].
\end{aligned}
$$

The proof is complete. □

D.7.2 Normal approximation

Proof of Theorem 24. We abbreviate T_J by T and write $b = b(S)$ in the proof if no confusion could arise. First we bound the operator norm of the Hessian. By Lemma 57-(2),

$$
\begin{aligned}
&\partial_{(ij),(i'j')}T(X) = b^{-2}\big(N^{-1}\delta_{jj'}X_i^\top X_{i'} + N^{-1}X_{i'j}X_{ij'} + \delta_{ii'}S_{jj'}\big) \\
&- \delta_{ii'}\delta_{jj'}\frac{b_2}{b^3} + X_{ij}X_{i'j'}\frac{6b_2}{b^4Np} - \frac{4}{b^3Np}\Big[X_i^\top Se_j\cdot X_{i'j'} + X_{i'}^\top Se_{j'}\cdot X_{ij}\Big] \\
&\equiv (T_1 - T_2 + T_3 - T_4)_{(ij),(i'j')}.
\end{aligned}
$$

Following the proof of Theorem 18 along with Lemma 61, we have $\mathbb{E}\|T_1\|_{\mathrm{op}}^4 \lesssim (1\vee y)^4$. Next for T_2, we have by Lemma 61 and Lemma 51 that

$$
\mathbb{E}\|T_2\|_{\mathrm{op}}^4 \lesssim \mathbb{E}(b_2^4\cdot b^{-12}) \le \mathbb{E}^{1/2}b_2^8\cdot\mathbb{E}^{1/2}b^{-24} \lesssim \mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^8 \lesssim (1\vee y)^4.
$$

The operator norm of T_3 can be similarly bounded by

$$\mathbb{E}\|T_3\|_{\mathrm{op}}^4 = \frac{6^4}{(Np)^4}\mathbb{E}\Big[\Big(\frac{b_2}{b^4}\Big)^4\|X\|_F^8\Big] \lesssim (Np)^{-4}\mathbb{E}^{1/2}b_2^8 \cdot \mathbb{E}^{1/4}b^{-64}\mathbb{E}^{1/4}\|X\|_F^{32}$$
$$\lesssim (Np)^{-4}\cdot \mathbb{E}^{1/2}b_2^8 \cdot (Np)^4 \lesssim (1\vee y)^4.$$

Lastly,

$$\|T_4\|_{\mathrm{op}} \lesssim \frac{1}{b^3Np}\cdot \sup_{u,v\in B_{N\times p}(1)}\Big|\sum_{(ij),(i'j')} X_i^\top Se_j\cdot X_{i'j'}u_{ij}v_{i'j'}\Big|$$
$$= \frac{1}{b^3Np}\cdot \sup_{u,v\in B_{N\times p}(1)}\Big|\Big(\sum_{i,j} X_i^\top Se_j u_{ij}\Big)\Big(\sum_{i'j'} X_{i'j'}v_{i'j'}\Big)\Big|$$
$$\le \frac{1}{b^3Np}\cdot \|XS\|_F\cdot\|X\|_F \le \frac{1}{b^3Np}\cdot\|S\|_{\mathrm{op}}\|X\|_F^2.$$

Hence by Lemma 51 and Lemma 61, $\mathbb{E}\|T_4(X)\|_{\mathrm{op}}^4 \lesssim (1\vee y)^4$. Putting together the bounds for T_1 - T_4 yields that $\mathbb{E}\|\nabla^2 T(X)\|_{\mathrm{op}}^4 \lesssim (1\vee y)^4$.

Next we bound the norm of the gradient. We will show that $\mathbb{E}\|\nabla T(X)\|_F^2 \lesssim p^2$ by considering the two cases $p/N\le 1$ and $p/N>1$ separately.

(**Case** $p/N\le 1$) By Lemma 57-(1), we may write

$$\nabla T(X) = b^{-1}X\big(b^{-1}S - I\big) - b^{-1}X\cdot b\big(b^{-1}S-I\big)^2,$$

so

$$\|\nabla T(X)\|_F^4 \lesssim b^{-8}\|X\|_F^4\|S-bI\|_{\mathrm{op}}^4 + b^{-12}\|X\|_F^4\|S-bI\|_{\mathrm{op}}^8$$
$$\lesssim \|X\|_F^4\big(b^{-8}\|S-I\|_{\mathrm{op}}^4 + b^{-8}|b-1|^4 + b^{-12}\|S-I\|_{\mathrm{op}}^8 + b^{-12}|b-1|^8\big).$$

By Lemma 51 and Lemma 63, it holds under the condition $p/N\le 1$ that

$$\mathbb{E}\|T(X)\|_F^4 \lesssim (Np)^2\big((N^{-1}p)^2 + (N^{-1}p)^4\big) \lesssim p^4.$$

(**Case** $p/N>1$) By Lemma 57-(1), we have

$$\mathbb{E}\|\nabla T(X)\|_F^2 = \mathbb{E}\Big[\Big\|\frac{XS}{b^2}\Big\|_F^2 + \Big\|\frac{Xb_2}{b^3}\Big\|_F^2 - 2\langle\frac{XS}{b^2},\frac{Xb_2}{b^3}\rangle\Big] = Np\cdot\mathbb{E}\Big[\frac{bb_3-b_2^2}{b^5}\Big]$$
$$= Np\cdot\mathbb{E}(bb_3-b_2^2) + Np\cdot\mathbb{E}(bb_3-b_2^2)(b^{-5}-1) \equiv (I)+(II).$$

To handle (I), it holds by Lemma 64-(4)(5) that under $p > N$,

$$(I) = \frac{N}{p}\mathbb{E}\big[\mathrm{Tr}(S)\mathrm{Tr}(S^3) - \mathrm{Tr}^2(S^2)\big] = \frac{N}{p}N^{-4}\mathcal{O}(N^3p^3) = \mathcal{O}(p^2).$$

To handle (II), it holds by Lemmas 61, 63-(3), and 64-(6) that under $p > N$,

$$\begin{aligned}(II) &= Np\cdot\mathbb{E}(bb_3 - b_2^2)b^{-5}(1-b^5) \leq Np\cdot\mathbb{E}^{1/2}(bb_3 - b_2^2)^2\mathbb{E}^{1/4}b^{-20}\mathbb{E}^{1/4}(b^5-1)^4\\ &\leq Np\cdot\Big(|\mathbb{E}(bb_3 - b_2^2)| + \mathrm{Var}^{1/2}(bb_3 - b_2^2)\Big)\cdot\mathbb{E}^{1/4}b^{-20}\mathbb{E}^{1/4}(b^5-1)^4\\ &= Np\cdot\mathcal{O}(N^{-1}p)\cdot\mathcal{O}(1)\cdot\mathcal{O}((Np)^{-1/2}) = \mathfrak{o}(p^2).\end{aligned}$$

Putting together the estimates for (I) and (II) yield that $\mathbb{E}\|\nabla T(X)\|_F^2 = \mathcal{O}(p^2)$ under the considered case $p > N$. The rest of the proof proceeds along the lines in the proof of Theorem 18, with the help of the variance formula in Proposition 26-(3). The normal approximation error bound becomes a constant multiple of

$$\frac{(1\vee y)\cdot p}{p^2} = \frac{\frac{p}{n}\vee 1}{p} = \frac{1}{n\wedge p},$$

as desired. $\square$

D.7.3 Ratio control

Proof of Proposition 26. We assume without loss of generality that $b(\Sigma) = \mathrm{Tr}(\Sigma)/p = 1$ (otherwise we replace Σ by $\Sigma\cdot b^{-1}(\Sigma)$).

(1). By Lemma 57, with $S_Z \equiv N^{-1}\sum_{i=1}^N Z_iZ_i^\top$, we have

$$\begin{aligned}\mathcal{T}_{\Sigma;\mathrm{J}} &= \bigg\{\frac{Z\Sigma^{1/2}\Sigma^{1/2}S_Z\Sigma^{1/2}}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - Z\Sigma^{1/2}\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg\}\Sigma^{1/2}\\ &= \frac{Z\Sigma S_Z\Sigma}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - Z\Sigma\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})},\end{aligned}$$

so

$$
\begin{aligned}
&\mathcal{T}_{\Sigma;\mathrm{J}} - \mathcal{T}_{I;\mathrm{J}} \\
&= \bigg\{\frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(S_Z)}\bigg\} - \bigg\{Z\Sigma \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} - Z\frac{b_2(S_Z)}{b^3(S_Z)}\bigg\} \\
&= \bigg\{\frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(S_Z)}\bigg\} - Z\bigg\{\frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(S_Z)}\bigg\} \\
&\qquad - (Z\Sigma - Z)\cdot \frac{b_2(\Sigma^{1/2} S_Z \Sigma^{1/2})}{b^3(\Sigma^{1/2} S_Z \Sigma^{1/2})} \equiv V_1(Z) + V_2(Z) + V_3(Z).
\end{aligned}
$$

We will handle the Frobenius norms of $V_1(Z), V_2(Z), V_3(Z)$ separately below. For $V_1(Z)$,

$$
\begin{aligned}
\|V_1(Z)\|_F^2 &\lesssim \bigg\|\frac{Z\Sigma S_Z \Sigma}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})}\bigg\|_F^2 + \bigg\|\frac{Z S_Z}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2})} - \frac{Z S_Z}{b^2(S_Z)}\bigg\|_F^2 \\
&= \|Z\Sigma S_Z \Sigma - Z S_Z\|_F^2 \cdot \frac{1}{b^4(\Sigma^{1/2} S_Z \Sigma^{1/2})} \\
&\qquad + \|Z S_Z\|_F^2 \cdot \bigg[\frac{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2}) - b^2(S_Z)}{b^2(\Sigma^{1/2} S_Z \Sigma^{1/2}) b^2(S_Z)}\bigg]^2 \equiv V_{1,1} + V_{1,2}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
V_{1,1} &\lesssim b^{-4}(\Sigma^{1/2} S_Z \Sigma^{1/2})\big(\|Z\Sigma S_Z(\Sigma - I)\|_F^2 + \|Z(\Sigma - I)S_Z\|_F^2\big) \\
&\lesssim \Big[b^{-4}(\Sigma^{1/2} S_Z \Sigma^{1/2}) \cdot \|S_Z\|_{\mathrm{op}}^2\Big] \cdot \big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big) \cdot \|Z\|_{\mathrm{op}}^2 \|\Sigma - I\|_F^2, \\
V_{1,2} &\leq \|S_Z\|_{\mathrm{op}}^2 \|Z\|_F^2 b^{-4}(\Sigma^{1/2} S_Z \Sigma^{1/2}) b^{-4}(S_Z) \\
&\qquad \times \big(\mathrm{Tr}((\Sigma - I)S_Z)/p\big)^2 \big(b^2(\Sigma^{1/2} S_Z \Sigma^{1/2}) \vee b^2(S_Z)\big) \\
&\lesssim \Big[\|S_Z\|_{\mathrm{op}}^4 b^{-4}(\Sigma^{1/2} S_Z \Sigma^{1/2}) b^{-4}(S_Z)\big(b^2(\Sigma^{1/2} S_Z \Sigma^{1/2}) \vee b^2(S_Z)\big)\Big] \\
&\qquad \times p^{-1}\|Z\|_F^2 \|\Sigma - I\|_F^2.
\end{aligned}
$$

So under $p/N \leq M$, by Lemma 61 and Lemma 51, we have

$$
\mathbb{E}\|V_1(Z)\|_F^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.
$$

For $V_2(Z)$,

$$
\begin{aligned}
\|V_2(Z)\|_F^2 &= \|Z\|_F^2\Big(\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(S_Z)}\Big)^2\\
&\lesssim \|Z\|_F^2\Big\{\Big(\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})}\Big)^2\\
&\qquad + \Big(\frac{b_2(S_Z)}{b^3(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{b_2(S_Z)}{b^3(S_Z)}\Big)^2\Big\} \equiv V_{2,1} + V_{2,2}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\big(b_2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b_2(S_Z)\big)^2 = p^{-2}\mathrm{Tr}^2\big(S_Z\Sigma S_Z\Sigma - S_Z^2\big)\\
&\lesssim p^{-2}\Big\{\mathrm{Tr}^2\big(S_Z(\Sigma - I)S_Z\Sigma\big) + \mathrm{Tr}^2\big(S_Z^2(\Sigma - I)\big)\Big\}\\
&\lesssim p^{-1}\|S_Z\|_{\mathrm{op}}^4\big(\|\Sigma\|_F^2/p + 1\big)\|\Sigma - I\|_F^2 \qquad \text{(D.44)}\\
&\lesssim p^{-1}\|S_Z\|_{\mathrm{op}}^4\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2,
\end{aligned}
$$

so

$$
\begin{aligned}
V_{2,1} &\lesssim \Big[b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})\|S_Z\|_{\mathrm{op}}^4\Big]\cdot\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\cdot\big(p^{-1}\|Z\|_F^2\big)\cdot\|\Sigma - I\|_F^2,\\
V_{2,2} &\le \|Z\|_F^2 b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-6}(S_Z)b_2^2(S_Z)\big(b(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b(S_Z)\big)^2\\
&\qquad\times\Big(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) + b(\Sigma^{1/2}S_Z\Sigma^{1/2})b(S_Z) + b^2(S_Z)\Big)^2\\
&\lesssim \Big[b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-6}(S_Z)\big(b^4(\Sigma^{1/2}S_Z\Sigma^{1/2}) \vee b^4(S_Z)\big)\|S_Z\|_{\mathrm{op}}^6\Big]\\
&\qquad\times p^{-1}\|Z\|_F^2\|\Sigma - I\|_F^2.
\end{aligned}
$$

Hence under $p/N \le M$, by Lemma 61 and Lemma 51, we have

$$
\mathbb{E}\|V_2(Z)\|_F^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.
$$

Lastly, recall that $\mathrm{Tr}(\Sigma) = p$ so using trace Hölder inequality we have $\mathrm{Tr}(S_Z\Sigma S_Z\Sigma) \le \mathrm{Tr}(\Sigma)\|S_Z\Sigma S_Z\|_{\mathrm{op}} \le p\|S_Z\|_{\mathrm{op}}^2\|\Sigma\|_{\mathrm{op}}$, so $V_3(Z)$ satisfies

$$
\begin{aligned}
\|V_3(Z)\|_F^2 &\le p^{-2}\|\Sigma - I\|_F^2\cdot\|Z\|_{\mathrm{op}}^2\cdot b^{-6}(\Sigma S_Z)\cdot\mathrm{Tr}^2(S_Z\Sigma S_Z\Sigma)\\
&\le \Big[b^{-6}(\Sigma^{1/2}S_Z\Sigma^{1/2})\|S_Z\|_{\mathrm{op}}^4\Big]\cdot\|\Sigma\|_{\mathrm{op}}^2\cdot\|Z\|_{\mathrm{op}}^2\|\Sigma - I\|_F^2
\end{aligned}
$$

Hence under $p/N \le M$, by Lemma 61 and Lemma 51, we have

$$\mathbb{E}\|V_3(Z)\|_F^2 \lesssim_M N\big(\|\Sigma\|_{\mathrm{op}}^2 \vee 1\big)\|\Sigma - I\|_F^2.$$

Combining the estimates proves the claim.

(2). Recall the normalization $b(\Sigma) = 1$. Note that

$$\begin{aligned}
&\mathbb{E}\bigg[\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg]\\
&= \frac{\mathbb{E}b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} + \mathbb{E}\bigg[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg)\bigg]\\
&\stackrel{(*)}{=} \frac{(1+N^{-1})\frac{b_2(\Sigma)}{b^2(\Sigma)} + \frac{p}{N}}{1 + 2\mathrm{Tr}(\Sigma^2)/(Np^2)} + \mathbb{E}\bigg[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg)\bigg]\\
&= \frac{b_2(\Sigma)}{b^2(\Sigma)} + \frac{p}{N} + \bigg\{\mathbb{E}\bigg[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg)\bigg]\\
&\qquad + \bigg[(1+N^{-1})\frac{b_2(\Sigma)}{b^2(\Sigma)} + \frac{p}{N}\bigg]\bigg(\frac{1}{1 + 2\mathrm{Tr}(\Sigma^2)\cdot(Np^2)^{-1}} - 1\bigg) + N^{-1}\frac{b_2(\Sigma)}{b^2(\Sigma)}\bigg\}\\
&\equiv \frac{b_2(\Sigma)}{b^2(\Sigma)} + \frac{p}{N} + R(\Sigma).
\end{aligned}$$

Here we use Lemma 63-(1) in $(*)$ and

$$\begin{aligned}
R(\Sigma) &= \mathbb{E}\bigg[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg)\bigg]\\
&\qquad + \bigg[(1+N^{-1})\frac{b_2(\Sigma)}{b^2(\Sigma)} + \frac{p}{N}\bigg]\bigg(\frac{1}{1 + 2\mathrm{Tr}(\Sigma^2)\cdot(Np^2)^{-1}} - 1\bigg) + N^{-1}\frac{b_2(\Sigma)}{b^2(\Sigma)}\\
&\equiv R_1(\Sigma) + R_2(\Sigma) + R_3(\Sigma).
\end{aligned}$$

As

$$\begin{aligned}
m_{\Sigma;\mathrm{J}} &= \frac{N}{4}\mathbb{E}\mathrm{Tr}\Big(\frac{\Sigma^{1/2}S_Z\Sigma^{1/2}}{b(\Sigma^{1/2}S_Z\Sigma^{1/2})} - I\Big)^2\\
&= \frac{Np}{4}\bigg\{\mathbb{E}\bigg[\frac{b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^2(\Sigma S_Z)}\bigg] - 1\bigg\},
\end{aligned}$$

we have

$$m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}} = \frac{Np}{4}\big(p^{-1}\|\Sigma - I\|_F^2 + R(\Sigma) - R(I)\big)$$

Now we handle $R_\ell(\Sigma) - R_\ell(I)$ for $\ell = 1, 2, 3$.

For $\ell = 1$,

$$
\begin{aligned}
&\big|R_1(\Sigma) - R_1(I)\big| \\
&= \bigg|\mathbb{E}\bigg[b_2(\Sigma^{1/2}S_Z\Sigma^{1/2})\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg) \\
&\qquad - b_2(S_Z)\bigg(\frac{1}{b^2(S_Z)} - \frac{1}{\mathbb{E}b^2(S_Z)}\bigg)\bigg]\bigg| \\
&\le \bigg|\mathbb{E}\bigg[\big(b_2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b_2(S_Z)\big)\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}\bigg)\bigg]\bigg| \\
&\qquad + \bigg|\mathbb{E}\bigg[b_2(S_Z)\bigg(\bigg(\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{b^2(S_Z)}\bigg) \\
&\qquad\qquad\qquad - \bigg(\frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})} - \frac{1}{\mathbb{E}b^2(S_Z)}\bigg)\bigg)\bigg]\bigg| \equiv R_{1,1} + R_{1,2}.
\end{aligned}
$$

The term $R_{1,1}$ can be handled as follows: by (D.44) Lemmas 51, 61, and 63, under $p/N \le M$,

$$
\begin{aligned}
R_{1,1} &\lesssim \mathbb{E}^{1/4}\big(b_2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b_2(S_Z)\big)^4 \cdot \mathbb{E}^{1/4}b^{-8}(\Sigma^{1/2}S\Sigma^{1/2}) \\
&\qquad \cdot \mathrm{Var}^{1/2}\big(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big) \cdot \big(\mathbb{E}b^2(\Sigma^{1/2}S\Sigma^{1/2})\big)^{-1} \\
&\lesssim_M p^{-2} \cdot \Big[p^{-1/2}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F\Big] \cdot \mathrm{Var}^{1/2}\big(\mathrm{Tr}^2(\Sigma S_Z)\big) \\
&\lesssim_M (N^{1/2}p)^{-1}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F.
\end{aligned}
$$

For $R_{1,2}$, we have by Lemmas 51 and 63 that, under $p/N \leq M$,

$$
\begin{aligned}
R_{1,2} &= \bigg|\mathbb{E}\bigg[b_2(S_Z)\bigg(\frac{b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})b^2(S_Z)} - \frac{\mathbb{E}b^2(S_Z) - \mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\mathbb{E}b^2(S_Z)}\bigg)\bigg]\bigg| \\
&\leq \mathbb{E}b_2(S_Z)b^{-2}(\Sigma^{1/2}S_Z\Sigma^{1/2})b^{-2}(S_Z) \\
&\qquad \times \big|b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - \mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \\
&\qquad + \big|\mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \cdot \mathbb{E}b_2(S_Z) \\
&\qquad\qquad \times \bigg|\frac{1}{b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})b^2(S_Z)} - \frac{1}{\mathbb{E}b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\mathbb{E}b^2(S_Z)}\bigg| \\
&\lesssim_M \mathrm{Var}^{1/2}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big) \\
&\qquad + \big|\mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \cdot \big(\mathrm{Var}^{1/2}(b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})) \vee \mathrm{Var}^{1/2}(b^2(S_Z))\big) \\
&\overset{(*)}{\lesssim_M} (N^{1/2}p)^{-1}\|\Sigma - I\|_F + p^{-1/2}\|\Sigma - I\|_F \cdot (N^{1/2}p)^{-1}(\|\Sigma\|_F \vee p^{1/2}) \\
&\lesssim (N^{1/2}p)^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F.
\end{aligned}
$$

Here in $(*)$ we use the fact that

$$
\begin{aligned}
&\big|\mathbb{E}\big(b^2(S_Z) - b^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big)\big| \lesssim p^{-1}\mathbb{E}^{1/2}\mathrm{Tr}^2\big((\Sigma - I)S_Z\big) \\
&\leq p^{-1/2}\|\Sigma - I\|_F \cdot \mathbb{E}^{1/2}\|S_Z\|_{\mathrm{op}}^2 \lesssim_M p^{-1/2}\|\Sigma - I\|_F.
\end{aligned}
$$

Hence

$$
|R_1(\Sigma) - R_1(I)| \lesssim_M (N^{1/2}p)^{-1}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F.
$$

For $\ell = 2$, with $\mathfrak{a}(\Sigma) \equiv 1/\big(1 + 2\mathrm{Tr}(\Sigma^2)/(Np^2)\big) - 1$ (then $|\mathfrak{a}(\Sigma)| \leq 2/N$ and $|\mathfrak{a}(I)| \leq 2/(Np)$), we have

$$
R_2(\Sigma) = (1 + N^{-1})b_2(\Sigma)\mathfrak{a}(\Sigma) + N^{-1}p\mathfrak{a}(\Sigma),
$$

so

$$
\big|R_2(\Sigma) - R_2(I)\big| \lesssim_M \big|b_2(\Sigma)\mathfrak{a}(\Sigma) - b_2(I)\mathfrak{a}(I)\big| + |\mathfrak{a}(\Sigma) - \mathfrak{a}(I)| \equiv R_{2,1} + R_{2,2}.
$$

The two terms $R_{2,1}, R_{2,2}$ can be handled as follows: using $\mathrm{Tr}(\Sigma^2) \le p^2$ under $b(\Sigma) = 1$, we have

$$\begin{aligned} R_{2,1} &\lesssim b_2(\Sigma)|\mathfrak{a}(\Sigma) - \mathfrak{a}(I)| + |\mathfrak{a}(I)||b_2(\Sigma) - b_2(I)| \\ &\lesssim p^{-1}\mathrm{Tr}(\Sigma^2)(Np^2)^{-1}|\mathrm{Tr}\big(\Sigma^2 - I\big)| + (Np)^{-1} \cdot p^{-1} \cdot |\mathrm{Tr}(\Sigma^2 - I)| \\ &\lesssim (Np^{1/2})^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F, \\ R_{2,2} &\lesssim (Np^{1/2})^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F, \end{aligned}$$

so

$$\big|R_2(\Sigma) - R_2(I)\big| \lesssim_M (Np^{1/2})^{-1}\big(\|\Sigma\|_F/p^{1/2} + 1\big)\|\Sigma - I\|_F.$$

For $\ell = 3$,

$$\big|R_3(\Sigma) - R_3(I)\big| = N^{-1}\big|b_2(\Sigma) - b_2(I)\big| \lesssim (Np^{1/2})^{-1}\big(p^{-1/2}\|\Sigma\|_F + 1\big)\|\Sigma - I\|_F.$$

Now with $Q_{\mathrm{J}}(\Sigma) \equiv p\big(R(\Sigma) - R(I)\big)$, we have

$$\begin{aligned} |Q_{\mathrm{J}}(\Sigma)| &\lesssim_M p\max\{(N^{1/2}p)^{-1}, (Np^{1/2})^{-1}\}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F \\ &\lesssim_M N^{-1/2}\big(p^{-1}\|\Sigma\|_F^2 + 1\big)\|\Sigma - I\|_F, \end{aligned}$$

and

$$m_{\Sigma;\mathrm{J}} - m_{I;\mathrm{J}} = \frac{N}{4}\big(\|\Sigma - I\|_F^2 + Q_{\mathrm{J}}(\Sigma)\big).$$

(3). Recall T_{LNW} defined in (5.24). Let

$$\Delta(X) \equiv T_{\mathrm{LNW}}(X) - T_{\mathrm{J}}(X).$$

Then for any $\varepsilon > 0$, there exists some $C_\varepsilon > 0$ such that under the null (i.e., $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(0, I_p)$),

$$\big[(1-\varepsilon)\sigma^2_{I;\mathrm{LNW}} - C_\varepsilon \mathrm{Var}_I(\Delta)\big]_+ \le \sigma^2_{I;\mathrm{J}} \le (1+\varepsilon)\sigma^2_{I;\mathrm{LNW}} + C_\varepsilon \mathrm{Var}_I(\Delta). \tag{D.45}$$

We will now bound $\mathrm{Var}_I(\Delta)$. By Lemmas 54-(1) and 57-(1), we have for any $i,j\in[N]\times[p]$

$$\begin{aligned}\partial_{(ij)}\Delta(X) &= \partial_{(ij)}T_{\mathrm{LNW}}(X)-\partial_{(ij)}T_{\mathrm{J}}(X)\\ &= \big(X(S-I)-N^{-1}\mathrm{Tr}(S)X\big)_{ij}-\bigg[\frac{X}{b}\Big(\frac{S}{b}-I\Big)-\frac{X}{b(S)}\cdot b\Big(\Big(\frac{S}{b(S)}-I\Big)^2\Big)\bigg]_{ij}\\ &= \bigg[X(S-I)-\frac{X}{b}\Big(\frac{S}{b}-I\Big)\bigg]_{ij}+\Big[Xb\big((S-b(S)I)^2\big)\big(b^{-3}(S)-1\big)\Big]_{ij}\\ &\quad+\Big[X\Big(b_2(S)-b^2(S)-N^{-1}\mathrm{Tr}(S)\Big)\Big]_{ij}\equiv\big(\Delta_1+\Delta_2+\Delta_3\big)_{ij}.\end{aligned}$$

We now handle Δ_1-Δ_3 separately below. For Δ_1, by Lemmas 51, 61, and 63, we have

$$\begin{aligned}\mathbb{E}\|\Delta_1\|_F^2 &\lesssim \mathbb{E}b^{-2}(1-b)^2\|X(S-I)\|_F^2+\mathbb{E}b^{-4}(1-b)^2\|XS\|_F^2\\ &\le N\mathbb{E}b^{-2}(1-b)^2\|S\|_{\mathrm{op}}\|S-I\|_F^2+N\mathbb{E}b^{-4}(b-1)^2\mathrm{Tr}(S^3)\\ &\lesssim N\cdot(pN)^{-1}\cdot(1\vee y)(N^{-1}p^2)+N\mathbb{E}^{1/4}b^{-16}\mathbb{E}^{1/4}(b-1)^8\mathbb{E}^{1/2}\mathrm{Tr}^2(S^3)\\ &\overset{(*)}{\lesssim} \mathfrak{o}(p^2)+N\cdot\mathcal{O}(1)\cdot(Np)^{-1}\cdot\mathcal{O}(N^{-2}p^3\vee p^2)=\mathfrak{o}(p^2).\end{aligned}$$

Here in $(*)$ the first bound follows by direct calculation and the second bound follows as: by Lemma 64-(7),

$$\begin{aligned}\mathbb{E}\mathrm{Tr}^2(S^3)&\le\mathbb{E}\mathrm{Tr}^2(S^2)\|S\|_{\mathrm{op}}^2\le\mathbb{E}^{1/2}\mathrm{Tr}^4(S^2)\cdot\mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^4\\ &\lesssim p^4\cdot(y^2\vee 1)=\mathcal{O}(N^{-4}p^6\vee p^4).\end{aligned}$$

For Δ_2, using $b\big((S-b(S)I)^2\big)\le\|(S-b(S)I)^2\|_{\mathrm{op}}\lesssim\|S\|_{\mathrm{op}}^2\vee b^2(S)$, we have

$$\begin{aligned}\mathbb{E}\|\Delta_2\|_F^2&\lesssim\mathbb{E}b^{-6}(b^4\vee b^2\vee 1)(b-1)^2\big(\|S\|_{\mathrm{op}}^2\vee b^2\big)\|X\|_F^2\\ &\lesssim(pN)^{-1}\cdot(pN)\cdot\mathbb{E}^{1/2}\big(\|S\|_{\mathrm{op}}^4\vee b^4\big)\asymp(1\vee y)^2=\mathfrak{o}(p^2).\end{aligned}$$

For Δ_3, let $h(S)\equiv b_2(S)-b^2(S)-N^{-1}\mathrm{Tr}(S)$, we have

$$\begin{aligned}\mathbb{E}\|\Delta_3\|_F^2&\lesssim\mathbb{E}\|X\|_F^2h^2(S)\le N\mathbb{E}^{1/2}\mathrm{Tr}^2(S)\cdot\mathbb{E}^{1/2}h^4(S)\\ &\lesssim Np\cdot\Big[\big(\mathbb{E}h(S)\big)^4+\mathrm{Var}^2\big(h(S)\big)+\mathbb{E}\|\nabla h(S)\|_F^4\Big]^{1/2},\end{aligned}$$

where the last inequality follows since

$$\begin{aligned}\mathbb{E}h^4(S) &= [\mathbb{E}h^2(S)]^2 + \mathrm{Var}(h^2(S)) \le 2\big[\mathbb{E}h(S))^4 + 2\mathrm{Var}^2(h(S)) + \mathrm{Var}(h^2(S)\big] \\ &\le 2(\mathbb{E}h(S))^4 + 2\mathrm{Var}^2(h(S)) + 4\mathbb{E}h^2(S)\|\nabla h(S)\|_F^2 \\ &\le 2(\mathbb{E}h(S))^4 + 2\mathrm{Var}^2(h(S)) + 4\tau\mathbb{E}h^4(S) + C_\tau\mathbb{E}\|\nabla h(S)\|_F^4\end{aligned}$$

and choosing, say, $\tau = 1/8$. For $\mathbb{E}h(S)$, Lemma 63 yields the direct evaluation

$$\begin{aligned}\mathbb{E}h(S) &= \frac{(1+N^{-1})p + N^{-1}p^2}{p} - \frac{p^2 + 2N^{-1}p}{p^2} - \frac{p}{N} \\ &= (1+N^{-1}) + \frac{p}{N} - \Big(1 + \frac{2}{Np}\Big) - \frac{p}{N} = \frac{1}{N} - \frac{2}{Np} = \mathcal{O}(N^{-1}).\end{aligned}$$

For $\mathrm{Var}\big(h(S)\big)$, the Gaussian-Poincaré inequality (Boucheron et al., 2013, Theorem 3.20) yields that

$$\begin{aligned}\mathrm{Var}\big(h(S)\big) &\le \mathbb{E}\sum_{ij}\big(\partial_{(ij)}h(S)\big)^2 \overset{(**)}{=} \sum_{ij}\mathbb{E}\bigg(\frac{4X_i^\top Se_j}{Np} - \frac{4X_{ij}b(S)}{Np} - \frac{2X_{ij}}{N^2}\bigg)^2 \\ &\lesssim (Np)^{-2}\big(\mathbb{E}\|XS\|_F^2 + \mathbb{E}b^2(S)\|X\|_F^2\big) + N^{-4}\mathbb{E}\|X\|_F^2 \\ &\lesssim (Np^2)^{-1}\mathbb{E}\mathrm{Tr}(S^2)\|S\|_{\mathrm{op}} + (Np)^{-2}(Np) + N^{-4}(Np) \\ &\lesssim (Np^2)^{-1}\mathbb{E}^{1/2}\mathrm{Tr}^2(S^2)\mathbb{E}^{1/2}\|S\|_{\mathrm{op}}^2 + (Np)^{-1} + pN^{-3} \\ &\lesssim (Np^2)^{-1}\cdot p^2\cdot(1\vee y) + (Np)^{-1} + pN^{-3} = \mathfrak{o}(N^{-1}p).\end{aligned}$$

Here $(**)$ follows from (D.43). Lastly $\mathbb{E}\|\nabla h(S)\|_F^4$ can be bounded similarly:

$$\mathbb{E}\|\nabla h(S)\|_F^4 \lesssim (Np)^{-4}\big(\mathbb{E}\|XS\|_F^4 + \mathbb{E}b^4(S)\|X\|_F^4\big) + N^{-8}\mathbb{E}\|X\|_F^4 = \mathfrak{o}(N^{-2}p^2).$$

Now by the Gaussian-Poincaré inequality (Boucheron et al., 2013, Theorem 3.20),

$$\mathrm{Var}_I(\Delta) \le \mathbb{E}\sum_{ij}\big(\partial_{(ij)}\Delta(X)\big)^2 \lesssim \mathbb{E}\|\Delta_1\|_F^2 + \mathbb{E}\|\Delta_2\|_F^2 + \mathbb{E}\|\Delta_3\|_F^2 = \mathfrak{o}(p^2).$$

As $\sigma^2_{I;\mathrm{LNW}} \sim p^2/4 \to \infty$ whenever $N \wedge p \to \infty$, by taking ε in (D.45) slowly decaying to 0 we conclude $\sigma^2_{I;\mathrm{LNW}} \sim \sigma^2_{I;\mathrm{J}}$.

(4). By (1)-(2), as $\|\Sigma\|_F^2/p \lesssim \|\Sigma - I\|_F^2/p + 1 \lesssim \|\Sigma - I\|_F \vee 1$ [where we use $\|\Sigma - I\|_F \leq \|\Sigma\|_F + \sqrt{p} \leq p + \sqrt{p}$ under $\mathrm{Tr}(\Sigma) = p$], we only need to prove that given C_0, we may find some constant $C_1 > 0$,

$$\frac{\sqrt{N}\|\Sigma - I\|_F\big(1 \vee \|\Sigma - I\|_F\big)}{\big(N\|\Sigma - I\|_F^2 - C_0 N^{1/2}\|\Sigma - I\|_F\big(1 \vee \|\Sigma - I\|_F\big)\big)_+ \bigvee \sigma_{I;\mathrm{J}}} \leq \frac{C_1}{(\sigma_{I;\mathrm{J}} \wedge N)^{1/2}}. \tag{D.46}$$

Write $\alpha = \|\Sigma - I\|_F$, we only need to prove that

$$\frac{\sqrt{N}\alpha \bigvee \sqrt{N}\alpha^2}{\big(N\alpha^2 - C_0 N^{1/2}\alpha\big)_+ \bigvee \sigma_{I;\mathrm{J}}} \leq \frac{C_1}{(\sigma_{I;\mathrm{J}} \wedge N)^{1/2}}. \tag{D.47}$$

This follows as

$$\begin{aligned}
\text{LHS of (D.47)} &\lesssim \mathbf{1}_{\alpha \leq 2C_0 N^{-1/2}} \frac{1}{\sigma_{I;J}} + \mathbf{1}_{\alpha > 2C_0 N^{-1/2}} \frac{\sqrt{N}\alpha \bigvee \sqrt{N}\alpha^2}{N\alpha^2 \bigvee \sigma_{I;\mathrm{J}}} \\
&\lesssim \frac{1}{\sigma_{I;J}} + \frac{1}{N^{1/2}\inf_{\alpha \geq 0}\big(\alpha \vee \frac{\sigma_{I;\mathrm{J}}/N}{\alpha}\big)} + \frac{1}{N^{1/2}} \\
&\lesssim \frac{1}{\sigma_{I;J}} + \frac{1}{\sigma_{I;\mathrm{J}}^{1/2}} + \frac{1}{N^{1/2}} \asymp \frac{1}{(\sigma_{I;\mathrm{J}} \wedge N)^{1/2}}.
\end{aligned}$$

The proof is complete. □

D.7.4 Completing of the proof for power expansion

Proof of Theorem 25. The proof essentially follows that of Theorem 19 by noting that the key property used therein is $|Q_{\mathrm{LNW}}(\Sigma)| \leq C_M N^{-1/2}(\|\Sigma - I\|_F \vee 1)\|\Sigma - I\|_F$, while here we have $|Q_{\mathrm{J}}(\Sigma \cdot b^{-1}(\Sigma))| \leq C_M N^{-1/2}\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F \leq N^{-1/2}(\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F \vee 1)\|\Sigma \cdot b^{-1}(\Sigma) - I\|_F$. □

D.8 Second-order Poincaré inequality

The main tool used for proving normal approximations is the following second-order Poincaré inequality due to Chatterjee (2009). Recall that $W^{1,2}(\gamma_n)$ is the Gaussian Sobolev space defined in (5.11).

Lemma 58 (Second-order Poincaré inequality)**.** *Let ξ be an n-dimensional standard normal random vector. Let $F : \mathbb{R}^n \to \mathbb{R}$ be an element of $W^{1,2}(\gamma_n)$. Let ξ' be an independent copy of ξ. Define $T : \mathbb{R}^n \to \mathbb{R}$ by*

$$T(y) \equiv \int_0^1 \frac{1}{2\sqrt{t}} \langle \nabla F(y), \mathbb{E}_{\xi'} \nabla F(\sqrt{t} y + \sqrt{1-t}\xi') \rangle \, \mathrm{d}t.$$

Then with $W \equiv F(\xi)$,

$$d_{\mathrm{TV}}\left(\frac{W - \mathbb{E}W}{\sqrt{\mathrm{Var}(W)}}, \mathcal{N}(0,1) \right) \leq \frac{2\sqrt{\mathrm{Var}(T(\xi))}}{\mathrm{Var}(W)}.$$

D.9 Sobolev regularity of matrix functionals

Lemma 59. *The following hold:*

1. *Let $f : \mathbb{R}^{N\times p} \to \mathbb{R}$ be defined by $f(X) = \log\det(X^\top X)$. If $N \geq p+1$, then $f \in W^{1,2}(\gamma_{N\times p})$ provided itself and its pointwise first derivatives live in $L_2(\gamma_{N\times p})$, $f \in W^{2,2}(\gamma_{N\times p})$ provided itself, its pointwise first, and second derivatives live in $L_2(\gamma_{N\times p})$. In particular, if N, p are large enough with $p/N \leq 1-\varepsilon$ for some $\varepsilon \in (0,1)$, then $f \in W^{2,2}(\gamma_{N\times p})$.*

2. *Let $g_\ell : \mathbb{R}^{N\times p} \to \mathbb{R}$ be defined by $g_\ell(X) = Tr^{-\ell}(X^\top X)$ for $\ell \in \mathbb{N}$. Then $g_\ell \in W^{1,2}(\gamma_{N\times p})$ provided itself and its pointwise first derivatives live in $L_2(\gamma_{N\times p})$, $g_\ell \in W^{2,2}(\gamma_{N\times p})$ provided itself, its pointwise first, and second derivatives live in $L_2(\gamma_{N\times p})$. In particular, there exists some $N_\ell \in \mathbb{N}$ such that for $N \geq N_\ell$, $g_\ell \in W^{2,2}(\gamma_{N\times p})$.*

Proof. (1). We first prove the claim involving $f \in W^{1,2}(\gamma_{N\times p})$. Let $W^{r,p}(\mathbb{R}^d)$ be the standard Sobolev class on $\mathbb{R}^d$ (cf. (Bogachev, 1998, Chapter 1.5)) and recall that $C_0^\infty(\mathbb{R}^d)$ is the class of smooth functions on $\mathbb{R}^d$ with compact support. By (Bogachev, 1998, Proposition 1.5.2), we only need to verify that $\zeta f \in W^{1,2}(\mathbb{R}^{N\times p})$, i.e., $\zeta f \in L_2(\mathbb{R}^{N\times p})$ and its first partial derivatives (in the sense of distributions) live in $L_2(\mathbb{R}^{N\times p})$ for every $\zeta \in C_0^\infty(\mathbb{R}^{N\times p})$. $\zeta f \in L_2(\mathbb{R}^{N\times p})$

follows from $N \geq p+1 > p$. Using the absolute continuity on line characterization of the space $W^{1,2}(\mathbb{R}^{N\times p})$ (cf. (Maz'ya, 2011, Section 1.1.3)), we only need to show that ζf is absolutely continuous on almost all straight lines that are parallel to coordinate axes and the first pointwise derivatives of ζf belong to $L_2(\mathbb{R}^{N\times p})$. As ζ has compact support, the latter requirement is satisfied by the assumption that f and its first pointwise derivatives live in $L_2(\gamma_{N\times p})$. To show the almost absolute continuity, we only need to do so for f on a compact subset of $\mathbb{R}^{N\times p}$. Identify $X \in \mathbb{R}^{N\times p}$ in the matrix form $X = [X_1 \cdots X_p]$ where $X_j \in \mathbb{R}^N$ for $1 \leq j \leq p$, and in the coordinate form $X = (X_1^\top, \ldots, X_p^\top)$. Let $L_j \equiv \{(X_1^\top, \ldots, X_p^\top) \in \mathbb{R}^{N\times p} : X_j \in \mathrm{lin}(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)\}$, and $\pi_{-(ij)} : \mathbb{R}^{N\times p} \to \mathbb{R}^{N\times p-1}$ be the natural projection that excludes $(X_j)_i$. Then $\pi_{-(ij)}(L_{j'})$ is a subset of $\mathbb{R}^{N\times p-1}$ of Lebesgue measure 0 for each $(i,j) \in [N]\times[p], j' \in [p]$ under the condition $p \leq N-1$, as we may write

$$L_{j'} = \Big\{ \big(X_1^\top, \ldots, X_{j'-1}^\top, \sum_{j\neq j'} \gamma_j X_j^\top, X_{j'+1}^\top, \ldots, X_p^\top\big) : X_j \in \mathbb{R}^N, \gamma_j \in \mathbb{R}, j \neq j' \Big\}.$$

Hence with $L \equiv \cup_j L_j$, $\pi_{-(ij)}(L)$ is a subset of $\mathbb{R}^{N\times p-1}$ of Lebesgue measure 0 for every $(i,j) \in [N]\times[p]$. In particular, this means that for any $X_{-(ij)} \notin \pi_{-(ij)}(L)$, the map f along the line $x_{(ij)} \mapsto (x_{(ij)}, X_{-(ij)})$ does not touch $\{X \in \mathbb{R}^{N\times p} : \det(X^\top X) = 0\}$, and hence is locally Lipschitz as $\nabla f(X) = 2X(X^\top X)^{-1}$. This verifies the almost absolute continuity property, and hence $f \in W^{1,2}(\gamma_{N\times p})$ provided itself and its first pointwise derivatives live in $L_2(\gamma_{N\times p})$.

The verification of $f \in W^{2,2}(\gamma_{N\times p})$ under L_2 integrability of the pointwise derivatives up to the second order is the same, upon noting the derivatives have singularities only at $\{X \in \mathbb{R}^{N\times p} : \det(X^\top X) = 0\}$ (the precise derivative formula is given in Lemma 52).

The last assertion follows from Lemma 50 and (D.17) that establishes the L_2 integrability of the pointwise first and second derivatives, and the straightforward verification of the L_2 integrability of f itself.

(2). The singularity of g_ℓ occurs only at $\mathrm{Tr}(X^\top X) = \sum_j \|X_j\|^2 = 0$, i.e., $X = 0$. The almost absolute continuity on line characterization is therefore easily verified. The L_2 integrability of the derivatives up to the second order follows from Lemma 61. □

D.10 Proof of Lemma 50

Proof of Lemma 50. Write S_Z for S in the proof for simplicity. Let λ be the smallest eigenvalue of S, and $y \equiv (p-1)/N < 1-\varepsilon$. By (Rudelson and Vershynin, 2009, Theorem 1.1), on an event E with probability at least $1-e^{-cN(1-y)}$, $\lambda \geq c(1-\sqrt{y})^2$ for some absolute constant $c > 0$. A similar estimate can be obtained using rigidity estimate for the eigenvalues of the sample covariance matrix, e.g., (Pillai and Yin, 2014, Theorem 3.1(iii)). Hence

$$\begin{aligned}\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q} &= \mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q}\mathbf{1}_{E} + \mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q}\mathbf{1}_{E^c} \\ &\leq c^{-q}\big(1-\sqrt{y}\big)^{-2q} + \mathbb{E}^{1/2}\|S^{-1}\|_{\mathrm{op}}^{2q}\cdot e^{-cN(1-y)/2}. \end{aligned} \tag{D.48}$$

Now we give an upper bound for $\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{2q}$. Let $r \equiv (N-p-1)/2$ assumed to be a positive integer. For any non-negative integer k, we write $\kappa \vdash k$ if $\kappa = (k_1, k_2, \ldots)$, with convention $k_1 \geq k_2 \geq \ldots$, is a partition of k, i.e., $\sum_i k_i = k$. Let C_κ denote the zonal polynomial (cf. (Muirhead, 1982, Chapter 7)) with respect to the partition κ. Then it follows from

(Muirhead, 1982, Corollary 9.7.4) that, for any $x > 0$,

$$
\begin{aligned}
&\mathbb{P}(\|S^{-1}\|_{\mathrm{op}} > x) = 1 - \mathbb{P}(\lambda > 1/x) \\
&= 1 - e^{-\frac{Np}{2x}} \sum_{k=0}^{pr} \sum_{\kappa \vdash k : k_1 \le r} \frac{C_\kappa\big(NI/(2x)\big)}{k!} \\
&= e^{-\frac{Np}{2x}} \Big[\sum_{k=0}^{\infty} \frac{\big(Np/(2x)\big)^k}{k!} - \sum_{k=0}^{pr} \sum_{\kappa \vdash k : k_1 \le r} \frac{C_\kappa\big(NI/(2x)\big)}{k!} \Big] \\
&= e^{-\frac{Np}{2x}} \Big[\sum_{k=pr+1}^{\infty} \frac{\big(Np/(2x)\big)^k}{k!} + \sum_{k=0}^{pr} \frac{1}{k!} \Big\{ \Big(\frac{Np}{2x}\Big)^k - \sum_{\kappa \vdash k : k_1 \le r} C_\kappa\Big(\frac{NI}{2x}\Big) \Big\} \Big] \\
&\overset{(*)}{=} e^{-\frac{Np}{2x}} \Big[\sum_{k=pr+1}^{\infty} \frac{\big(Np/(2x)\big)^k}{k!} + \sum_{k=0}^{pr} \frac{1}{k!} \sum_{\kappa \vdash k : k_1 > r} C_\kappa\Big(\frac{NI}{2x}\Big) \Big] \\
&\overset{(**)}{=} e^{-\frac{Np}{2x}} \Big[\sum_{k=pr+1}^{\infty} \frac{\big(Np/(2x)\big)^k}{k!} + \sum_{k=r+1}^{pr} \frac{\big(N/(2x)\big)^k}{k!} \sum_{\kappa \vdash k : k_1 > r} C_\kappa\big(I\big) \Big) \Big] \\
&\overset{(***)}{\le} e^{-\frac{Np}{2x}} \cdot \sum_{k=r+1}^{\infty} \frac{\big(Np/(2x)\big)^k}{k!}.
\end{aligned}
$$

Here $(*)$ follows from (Muirhead, 1982, Definition 7.2.1, (iii)): for any $k \ge 0$ and $t > 0$,

$$
\sum_{\kappa \vdash k} C_\kappa(t \cdot I) = \big[\mathrm{Tr}(t \cdot I)\big]^k = (tp)^k; \tag{D.49}
$$

$(**)$ follows from the fact that for each k and partition κ of k, C_κ is a homogeneous polynomial of order k; $(***)$ follows from the non-negativity of zonal polynomial for I (cf. (Muirhead, 1982, Corollary 7.2.4)) and an application of (D.49) with $t = 1$:

$$
\sum_{\kappa \vdash k : k_1 > r} C_\kappa\big(I\big) \le \sum_{\kappa \vdash k} C_\kappa\big(I\big) = p^k.
$$

Hence by using the fact that for any $k \ge 2q + 1$,

$$
\begin{aligned}
\int_0^\infty e^{-\frac{Np}{2x}} \Big(\frac{Np}{2x}\Big)^k \cdot x^{2q-1}\, \mathrm{d}x &= \Big(\frac{Np}{2}\Big)^{2q} \int_0^\infty e^{-y} y^{k-2q-1}\, \mathrm{d}y \\
&= \Big(\frac{Np}{2}\Big)^{2q} (k - 2q - 1)!,
\end{aligned}
$$

we have for every $r \geq 4q$

$$
\begin{aligned}
\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{2q} &= 2q \int_0^\infty x^{2q-1}\mathbb{P}(\|S^{-1}\|_{\mathrm{op}} > x)\,\mathrm{d}x \\
&\leq \Big(\frac{Np}{2}\Big)^{2q} \sum_{k=r+1}^{\infty} \frac{1}{k(k-1)\cdots(k-2q)} \\
&= \Big(\frac{Np}{2}\Big)^{2q} \sum_{k=r+1}^{\infty} \frac{1}{2q}\cdot\Big\{\frac{1}{(k-1)\cdots(k-2q)} - \frac{1}{k\cdots(k-2q-1)}\Big\} \\
&= \Big(\frac{Np}{2}\Big)^{2q} \frac{1}{2q}\frac{1}{r(r-1)\cdots(r-2q+1)} \lesssim_q \frac{(Np)^{2q}}{r^{2q}}. \qquad \text{(D.50)}
\end{aligned}
$$

Combining (D.48) and (D.50), as $p/N \leq 1-\varepsilon$,

$$
\mathbb{E}\|S^{-1}\|_{\mathrm{op}}^{q} \leq C_\varepsilon^q + C_\varepsilon N^q e^{-c_\varepsilon N} \lesssim_{q,\varepsilon} 1 \qquad \text{(D.51)}
$$

with $r = (N-p-1)/2$ being a positive integer. If r is not an integer, write $S = \frac{N-1}{N}S' + \frac{1}{N}X_N X_N^\top$, where $S' \equiv \frac{1}{N-1}\sum_{i=1}^{N-1} X_i X_i^\top$. Then using Sherman-Morrison formula,

$$
\begin{aligned}
S^{-1} &= \frac{N}{N-1}(S')^{-1} - \frac{N}{(N-1)^2}\cdot\frac{(S')^{-1}X_N X_N^\top (S')^{-1}}{1+\frac{1}{N-1}X_N^\top (S')^{-1}X_N} \\
&\equiv \frac{N}{N-1}(S')^{-1} - R. \qquad \text{(D.52)}
\end{aligned}
$$

As $X_N^\top (S')^{-1} X_N \geq \|X_N\|^2/\lambda_{\max}(S')$, we have

$$
\begin{aligned}
\mathbb{E}\|R\|_{\mathrm{op}}^q &\leq \Big(\frac{N}{N-1}\Big)^q \cdot \mathbb{E}\Big[\|(S')^{-1}X_N X_N^\top (S')^{-1}\|_{\mathrm{op}}^q \frac{\lambda_{\max}(S')^{2q}}{\|X_N\|^{2q}}\Big] \\
&\lesssim_q \mathbb{E}\Big(\frac{\lambda_{\max}(S')^{2q}}{\lambda_{\min}(S')^{2q}}\Big) \leq \mathbb{E}^{1/2}\|S'\|_{\mathrm{op}}^{4q}\cdot\mathbb{E}^{1/2}\|(S')^{-1}\|_{\mathrm{op}}^{4q} \lesssim_{q,\varepsilon} 1.
\end{aligned}
$$

The claim for r not being an integer follows from the decomposition (D.51) and the estimate above. $\square$

D.11 Moment and concentration (in)equalities for trace functionals

Lemma 60. *Let $Z \in \mathbb{R}^{N\times p}$ be a random matrix whose entries are i.i.d. $\mathcal{N}(0,1)$, and $A \in \mathbb{R}^{p\times p}$. Then*

$$
\mathbb{E}\|ZA\|_F^4 \leq 4N\|A^\top A\|_F^2 + N^2\|A\|_F^4 \leq 5N^2\|A\|_F^4.
$$

Proof. As $\|ZA\|_F^2 = \mathrm{Tr}(ZAA^\top Z^\top) = \mathrm{Tr}(A^\top Z^\top ZA)$, we have

$$\begin{aligned}\mathbb{E}\|ZA\|_F^4 &= \mathbb{E}\mathrm{Tr}^2(A^\top Z^\top ZA) = \mathrm{Var}\big(\mathrm{Tr}(A^\top Z^\top ZA)\big) + \big(\mathbb{E}\mathrm{Tr}(A^\top Z^\top ZA)\big)^2\\ &= \mathrm{Var}\big(\mathrm{Tr}(A^\top Z^\top ZA)\big) + N^2\mathrm{Tr}^2(A^\top A).\end{aligned}$$

Further note that for any $(i,j) \in [N] \times [p]$,

$$\begin{aligned}\partial_{(ij)}\mathrm{Tr}(A^\top Z^\top ZA) &= \mathrm{Tr}\big(A^\top (e_ie_j^\top)^\top ZA\big) + \mathrm{Tr}\big(A^\top Z^\top e_ie_j^\top A\big)\\ &= (ZAA^\top)_{ij} + (AA^\top Z^\top)_{ji} = 2(ZAA^\top)_{ij},\end{aligned}$$

so by Gaussian-Poincaré inequality,

$$\begin{aligned}\mathrm{Var}\big(\mathrm{Tr}(A^\top Z^\top ZA)\big) &\le \mathbb{E}\sum_{i,j}\big(\partial_{(ij)}\mathrm{Tr}(A^\top Z^\top ZA)\big)^2\\ &= 4\mathbb{E}\|ZAA^\top\|_F^2 = 4\mathbb{E}\mathrm{Tr}(ZAA^\top AA^\top Z^\top) = 4N\mathrm{Tr}(AA^\top AA^\top).\end{aligned}$$

Finally note that

$$\mathrm{Tr}(AA^\top AA^\top) = \|A^\top A\|_F^2 = \sum_i \lambda_i^4(A) \le \bigg(\sum_i \lambda_i^2(A)\bigg)^2 = \|A\|_F^4 = \mathrm{Tr}^2(A^\top A).$$

The claim follows. $\square$

Lemma 61. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_iZ_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0,I)$ in $\mathbb{R}^p$. Then there exists some universal constant $C > 0$ such that for any non-negative definite matrix Σ and any $t > 0$,*

$$\mathbb{P}\bigg(N\big|\big(Tr(\Sigma S_Z) - Tr(\Sigma)\big)\big| > t\bigg) \le 2\exp\bigg(-\frac{t^2}{C(N\|\Sigma\|_F^2 + \|\Sigma\|_{\mathrm{op}}t)}\bigg).$$

Consequently, $\mathbb{P}\big(Tr(\Sigma S_Z) < Tr(\Sigma)/2\big) \le e^{-cN}$ for some universal $c > 0$. Furthermore, for any $\ell \in \mathbb{Z}$ such that $\ell \ge -N/2$, there exists some $C_\ell > 0$ such that [recall (5.29)]

$$\mathbb{E}b^\ell(\Sigma^{1/2}S_Z\Sigma^{1/2}) \le C_\ell \cdot b^\ell(\Sigma).$$

Proof. Let $X_i \equiv \Sigma^{1/2} Z_i$. Then $\mathrm{Tr}(\Sigma S_Z) = N^{-1}\sum_{i=1}^N Z_i^\top \Sigma Z_i = N^{-1}\sum_{i=1}^N \|X_i\|^2$, and $\mathbb{E}\mathrm{Tr}(\Sigma S_Z) = \mathbb{E}\|X_i\|^2 = \mathrm{Tr}(\Sigma)$. By Hanson-Wright inequality (cf. (Boucheron et al., 2013, pp.39)),

$$\mathbb{E}\exp\bigg(\lambda \sum_{i=1}^N (\|X_i\|^2 - \mathbb{E}\|X_i\|^2)\bigg) \le \exp\bigg(\frac{\lambda^2 \cdot N\|\Sigma\|_F^2}{1-2\lambda\|\Sigma\|_{\mathrm{op}}}\bigg),$$

so by (Boucheron et al., 2013, Theorem 2.3), we have

$$\mathbb{P}\Big(N\big|\big(\mathrm{Tr}(\Sigma S_Z) - \mathrm{Tr}(\Sigma)\big)\big| > t\Big) \le 2\exp\bigg(-\frac{t^2}{C(N\|\Sigma\|_F^2 + \|\Sigma\|_{\mathrm{op}} t)}\bigg).$$

In particular, with $t \equiv N\mathrm{Tr}(\Sigma)/2$, we have

$$\mathbb{P}\big(\mathrm{Tr}(\Sigma S_Z) < \mathrm{Tr}(\Sigma)/2\big) \le \exp\bigg(-\frac{N^2\mathrm{Tr}^2(\Sigma)}{C\big(N\|\Sigma\|_F^2 + N\|\Sigma\|_{\mathrm{op}}\mathrm{Tr}(\Sigma)\big)}\bigg) \le e^{-cN}.$$

For the expectation bound, let $\{\lambda_j\}_{j=1}^p$ be the eigenvalues of Σ and assume without loss of generality that $\sum_{j=1}^p \lambda_j = 1$. Then

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}^\ell(S_Z\Sigma) &= \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^N Z_i^\top \Sigma Z_i\Big) = \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^N Z_i^\top \mathrm{diag}(\lambda_1,\ldots,\lambda_p) Z_i\Big)^\ell \\
&= N^\ell \mathbb{E}\bigg(\sum_{i=1}^N \sum_{j=1}^p \lambda_j Z_{ij}^2\bigg)^\ell \equiv N^\ell \mathbb{E}\bigg(\sum_{j=1}^p \lambda_j Y_j\bigg)^\ell \\
&\overset{(*)}{\le} N^\ell \mathbb{E}\sum_{j=1}^p \lambda_j Y_j^\ell \overset{(**)}{\lesssim_\ell} N^\ell \cdot \sum_{j=1}^p \lambda_j N^\ell = 1.
\end{aligned}$$

Here $(*)$ follows as the map $x \mapsto x^\ell$ is convex on $(0,\infty)$ for $\ell \in \mathbb{Z}$, and $(**)$ follows from the following calculations:

- If $\ell \in \mathbb{Z}_{\ge 1}$, $\mathbb{E}Y_1^\ell = \mathbb{E}\big(\chi^2(N)\big)^\ell \lesssim_\ell N^\ell$.

- If $\ell \in \mathbb{Z}_{\le -1}$ and $\ell \ge -N/2$, then

$$\begin{aligned}
\mathbb{E}Y_1^\ell = \mathbb{E}\big(\chi^{-2}(N)\big)^{-\ell} &= \int x^{-\ell} \frac{2^{-\frac{N}{2}}}{\Gamma(N/2)} x^{-\frac{N}{2}-1} e^{-\frac{1}{2x}}\,\mathrm{d}x \\
&= 2^\ell \frac{\Gamma(N/2+\ell)}{\Gamma(N/2)} \lesssim_\ell N^\ell.
\end{aligned}$$

The proof is complete. $\square$

Lemma 62. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_iZ_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0,I)$, and $\Sigma \in \mathbb{R}^{p\times p}$ be a non-negative definite matrix. Recall the definition of $b_\ell(\Sigma)$ in (5.29). Then for some universal constants $C,c>0$,*

$$\begin{aligned}&\big|\mathbb{E}\log \mathit{Tr}(\Sigma S_Z) - \log \mathit{Tr}(\Sigma)\big|\\ &\leq \frac{2b\big[(\Sigma\cdot b^{-1}(\Sigma))^2\big]}{Np} + Ce^{-cN}\Big(\frac{b^{1/2}\big[(\Sigma\cdot b^{-1}(\Sigma))^2\big]}{(Np)^{1/2}}\bigvee 1\Big).\end{aligned}$$

Proof. Let $E \equiv \{\mathrm{Tr}(\Sigma(S_Z-I))/\mathrm{Tr}(\Sigma)\geq -1/2\}$. By Lemma 61, $\mathbb{P}(E^c)\leq e^{-cN}$ for some universal constant $c>0$. As $|\log(1+x)-x|\leq 4x^2$ for $x\geq -1/2$,

$$\begin{aligned}&\big|\mathbb{E}\log\mathrm{Tr}(\Sigma S_Z) - \log\mathrm{Tr}(\Sigma)\big| = \bigg|\mathbb{E}\log\bigg(1+\frac{\mathrm{Tr}\big(\Sigma(S_Z-I)\big)}{\mathrm{Tr}(\Sigma)}\bigg)\big(\mathbf{1}_E+\mathbf{1}_{E^c}\big)\bigg|\\ &\leq \bigg|\mathbb{E}\frac{\mathrm{Tr}\big(\Sigma(S_Z-I)\big)}{\mathrm{Tr}(\Sigma)}\mathbf{1}_E\bigg| + 4\mathbb{E}\bigg(\frac{\mathrm{Tr}\big(\Sigma(S_Z-I)\big)}{\mathrm{Tr}(\Sigma)}\bigg)^2 + \mathbb{E}\bigg[\log\Big(\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)}\Big)\mathbf{1}_{E^c}\bigg]\\ &\leq \mathbb{E}^{1/2}\bigg[\frac{\mathrm{Tr}\big(\Sigma(S_Z-I)\big)}{\mathrm{Tr}(\Sigma)}\bigg]^2\mathbb{P}^{1/2}(E^c) + 4\mathbb{E}\bigg(\frac{\mathrm{Tr}\big(\Sigma(S_Z-I)\big)}{\mathrm{Tr}(\Sigma)}\bigg)^2\\ &\qquad + \mathbb{E}^{1/2}\log^2\Big(\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)}\Big)\cdot\mathbb{P}^{1/2}(E^c)\equiv (I)+(II)+(III).\end{aligned}$$

To handle (I), note that by Gaussian-Poincaré inequality (Boucheron et al., 2013, Theorem 3.20),

$$\begin{aligned}&\mathbb{E}\mathrm{Tr}^2\big(\Sigma(S_Z-I)\big)\leq \mathbb{E}\sum_{i,j}\big[\partial_{(ij)}\mathrm{Tr}\big(\Sigma(S_Z-I)\big)\big]^2\\ &= \mathbb{E}\sum_{i,j}\bigg[N^{-1}\mathrm{Tr}\bigg(\Sigma\sum_k\big(\delta_{ik}e_jZ_k^\top+\delta_{ik}Z_ke_j^\top\big)\bigg)\bigg]^2\\ &= \mathbb{E}\sum_{i,j}\big[N^{-1}\mathrm{Tr}\big(\Sigma e_jZ_i^\top+\Sigma Z_ie_j^\top\big)\big]^2 = \frac{4}{N^2}\sum_{i,j}\mathbb{E}Z_i^\top\Sigma e_je_j^\top\Sigma Z_i = \frac{4\mathrm{Tr}(\Sigma^2)}{N},\end{aligned}$$

so

$$(I) = \frac{2e^{-cN/2}\mathrm{Tr}^{1/2}(\Sigma^2)}{N^{1/2}\mathrm{Tr}(\Sigma)} = 2e^{-cN/2}\cdot\frac{b^{1/2}\big[(\Sigma\cdot b^{-1}(\Sigma))^2\big]}{(Np)^{1/2}}.$$

The second term has closed-form expression: by Lemma 63-(1),

$$(II) = \frac{2\mathrm{Tr}(\Sigma^2)}{N\mathrm{Tr}^2(\Sigma)}.$$

To handle (III), by using $0 \le \log x \le x-1$ for $x \ge 1$ and $-x^{-1} \le \log x < 0$ for $x \in (0,1)$, we have

$$\begin{aligned}
\mathbb{E}\log^2\Big(\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)}\Big) &= \mathbb{E}\bigg[\log^2\Big(\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)}\Big)\mathbf{1}\Big\{\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)} \ge 1\Big\}\bigg] \\
&+ \mathbb{E}\bigg[\log^2\Big(\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)}\Big)\mathbf{1}\Big\{\frac{\mathrm{Tr}(\Sigma S_Z)}{\mathrm{Tr}(\Sigma)} < 1\Big\}\bigg] \\
&\le \mathbb{E}\bigg[\frac{\mathrm{Tr}\big(\Sigma(S_Z - I)\big)}{\mathrm{Tr}(\Sigma)}\bigg]^2 + \mathbb{E}\bigg[\frac{\mathrm{Tr}(\Sigma)}{\mathrm{Tr}(\Sigma S_Z)}\bigg]^2 \lesssim \frac{\mathrm{Tr}(\Sigma^2)}{N\mathrm{Tr}^2(\Sigma)} + 1,
\end{aligned}$$

where in the last inequality we apply Lemma 61. Hence

$$(III) \lesssim e^{-\frac{cN}{2}}\Big[\frac{\mathrm{Tr}^{1/2}(\Sigma^2)}{N^{1/2}\mathrm{Tr}(\Sigma)} \bigvee 1\Big].$$

The proof is complete by collecting the bounds. $\square$

Lemma 63. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_i Z_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0,I)$ in $\mathbb{R}^p$, and $\Sigma \in \mathbb{R}^{p\times p}$ be a non-negative definite matrix.*

1. *There exists some absolute $C > 0$ such that*

$$\begin{aligned}
&\mathbb{E}\,Tr\big[\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big)^2\big] = \big(1+N^{-1}\big)\,Tr(\Sigma^2) + N^{-1}\,Tr^2(\Sigma),\\
&\mathbb{E}\,Tr^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) = Tr^2(\Sigma) + 2N^{-1}\,Tr(\Sigma^2),\\
&\mathbb{E}\,Tr^2\big[\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big)^2\big] \le C\Big[N^{-1}(1\vee(p/N))^3\,Tr(\Sigma^4) + Tr^2(\Sigma^2) + N^{-2}\,Tr^4(\Sigma)\Big].
\end{aligned}$$

2. There exists some absolute $C > 0$ such that

$$\begin{aligned}
&\mathit{Var}\big(\mathit{Tr}(\Sigma S_Z)\big) \le 4N^{-1}\|\Sigma\|_F^2,\\
&\mathit{Var}\big(\mathit{Tr}^2(\Sigma^{1/2}S_Z\Sigma^{1/2})\big) \le C(N^{-2}\,\mathit{Tr}^2(\Sigma))\cdot N\|\Sigma\|_F^2,\\
&\mathit{Var}\big(\mathit{Tr}^2(\Sigma^{1/2}S_Z\Sigma^{1/2}) - \mathit{Tr}^2(S_Z)\big)\\
&\qquad \lesssim \big(N^{-2}\,\mathit{Tr}^2(\Sigma - I)\big)\cdot N\|\Sigma\|_F^2 + (N^{-1}p)^2\cdot N\|\Sigma - I\|_F^2,\\
&\mathit{Var}\Big(\mathit{Tr}\big[\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big)^2\big]\Big) \le CN^{-1}\big[1 \vee (N^{-1}p)\big]^3\,\mathit{Tr}(\Sigma^4).
\end{aligned}$$

3. Recall that $b(\Sigma) = \mathit{Tr}(\Sigma)/p$ from (5.29). For any $\ell \in \mathbb{N}$,

$$\mathbb{E}|b(\Sigma^{1/2}S_Z\Sigma^{1/2}) - b(\Sigma)|^\ell \le C_1\big(\|\Sigma\|_F N^{-1/2}p^{-1}\big)^\ell$$

for some constant $C_1 = C_1(\ell)$.

Proof. Let X_i's be i.i.d. $\mathcal{N}(0,\Sigma)$. We write $S \equiv \Sigma^{1/2}S_Z\Sigma^{1/2}$ in the proof for simplicity.

(1). Note that

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}(S^2) &= N^{-2}\mathbb{E}\mathrm{Tr}\Bigg[\sum_{i,j} X_iX_i^\top X_jX_j^\top\Bigg]\\
&= N^{-2}\Bigg[\sum_{i\neq j}\mathbb{E}\mathrm{Tr}\big(X_iX_i^\top X_jX_j^\top\big) + \sum_{i=j}\mathbb{E}(X_i^\top X_j)^2\Bigg]\\
&= N^{-2}\Big[N(N-1)\mathrm{Tr}(\Sigma^2) + N\mathbb{E}(Z_1^\top\Sigma Z_1)^2\Big]\\
&\overset{(*)}{=} N^{-2}\Big[N(N+1)\mathrm{Tr}(\Sigma^2) + N\mathrm{Tr}^2(\Sigma)\Big]\\
&= \big(1+N^{-1}\big)\mathrm{Tr}(\Sigma^2) + N^{-1}\mathrm{Tr}^2(\Sigma),
\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}^2(S) &= \mathbb{E}\bigg(N^{-1}\sum_{i=1}^{N} X_i^\top X_i\bigg)^2 = N^{-2}\sum_{i,j}\mathbb{E}X_i^\top X_i X_j^\top X_j \\
&= N^{-2}\bigg[\sum_{i\neq j}\mathbb{E}\|X_i\|^2\mathbb{E}\|X_j\|^2 + \sum_i \mathbb{E}(X_i^\top X_i)^2\bigg] \\
&= N^{-2}\Big[N(N-1)\big(\mathbb{E}Z_1^\top \Sigma Z_1\big)^2 + N\mathbb{E}\big(Z_1^\top\Sigma Z_1\big)^2\Big] \\
&\overset{(**)}{=} N^{-2}\Big[N(N-1)\mathrm{Tr}^2(\Sigma) + N\big(\mathrm{Tr}^2(\Sigma) + 2\mathrm{Tr}(\Sigma^2)\big)\Big] \\
&= \mathrm{Tr}^2(\Sigma) + 2N^{-1}\mathrm{Tr}(\Sigma^2).
\end{aligned}
$$

Here $(*), (**)$ follow by the following calculations:

$$
\begin{aligned}
\mathbb{E}Z_1^\top\Sigma Z_1 &= \mathbb{E}\bigg(\sum_j \lambda_j Z_{1j}^2\bigg) = \mathrm{Tr}(\Sigma), \\
\mathbb{E}(Z_1^\top\Sigma Z_1)^2 &= \mathbb{E}\bigg(\sum_j \lambda_j Z_{1j}^2\bigg)^2 = 3\sum_j \lambda_j^2 + \sum_{j\neq j'}\lambda_j\lambda_{j'} \\
&= 2\sum_j \lambda_j^2 + \bigg(\sum_j \lambda_j\bigg)^2 = \mathrm{Tr}^2(\Sigma) + 2\mathrm{Tr}(\Sigma^2),
\end{aligned}
$$

where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of Σ. The final one follows as

$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}^2\big[\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big)^2\big] &= \mathrm{Var}\big(\mathrm{Tr}\big[\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big)^2\big]\big) + \Big(\mathbb{E}\mathrm{Tr}\big[\big(\Sigma^{1/2}S_Z\Sigma^{1/2}\big)^2\big]\Big)^2 \\
&\lesssim N^{-1}\big(1\vee(N^{-1}p)\big)^3\mathrm{Tr}(\Sigma^4) + \mathrm{Tr}^2(\Sigma^2) + N^{-2}\mathrm{Tr}^4(\Sigma).
\end{aligned}
$$

The last inequality used (2) to be proved below.

(2). For the first variance bound, note that

$$
\frac{\partial}{\partial Z_{ij}}\mathrm{Tr}(\Sigma S_Z) = N^{-1}\mathrm{Tr}\big(\Sigma(e_j Z_i^\top + Z_i e_j^\top)\big) = 2N^{-1}(Z\Sigma)_{ij},
$$

so Gaussian-Poincaré inequality yields that

$$
\mathrm{Var}\big(\mathrm{Tr}(\Sigma S_Z)\big) \leq \mathbb{E}\sum_{i,j}\bigg[\frac{\partial}{\partial Z_{ij}}\mathrm{Tr}^2(\Sigma S_Z)\bigg]^2 = 4N^{-2}\mathbb{E}\|Z\Sigma\|_F^2 = 4N^{-1}\|\Sigma\|_F^2.
$$

For the second variance bound, note that

$$\begin{aligned}\frac{\partial}{\partial Z_{ij}}\mathrm{Tr}^2(\Sigma S_Z) &= 2\mathrm{Tr}(\Sigma S_Z)\cdot N^{-1}\mathrm{Tr}\big(\Sigma(e_j Z_i^\top + Z_i e_j^\top)\big)\\ &= 4N^{-1}\mathrm{Tr}(\Sigma S_Z)(Z\Sigma)_{ij},\end{aligned}$$

so Gaussian-Poincaré inequality yields that

$$\begin{aligned}\mathrm{Var}\big(\mathrm{Tr}^2(\Sigma S_Z)\big) &\le \mathbb{E}\sum_{i,j}\left[\frac{\partial}{\partial Z_{ij}}\mathrm{Tr}^2(\Sigma S_Z)\right]^2 = 16N^{-2}\mathbb{E}\mathrm{Tr}^2(\Sigma S_Z)\|Z\Sigma\|_F^2\\ &\overset{(*)}{\le} N^{-2}\mathrm{Tr}^2(\Sigma)\cdot\mathbb{E}^{1/2}\|Z\Sigma\|_F^4 \overset{(**)}{\lesssim} (N^{-2}\mathrm{Tr}^2(\Sigma))\cdot N\|\Sigma\|_F^2.\end{aligned}$$

Here in $(*)$ we use Lemma 61, and in $(**)$ we use Lemma 60.

For the third variance bound, note that

$$\frac{\partial}{\partial Z_{ij}}\big(\mathrm{Tr}^2(\Sigma S_Z) - \mathrm{Tr}^2(S_Z)\big) = 4N^{-1}\big(\mathrm{Tr}(\Sigma S_Z)(Z\Sigma)_{ij} - \mathrm{Tr}(S_Z)Z_{ij}\big).$$

Hence

$$\begin{aligned}&\mathrm{Var}\big(\mathrm{Tr}^2(\Sigma S_Z) - \mathrm{Tr}^2(S_Z)\big)\\ &\lesssim N^{-2}\mathbb{E}\mathrm{Tr}^2\big((\Sigma - I)S_Z\big)\|Z\Sigma\|_F^2 + N^{-2}\mathbb{E}\mathrm{Tr}^2(S_Z)\|Z(\Sigma - I)\|_F^2\\ &\lesssim \big[N^{-2}\mathrm{Tr}^2(\Sigma - I)\big]\cdot N\|\Sigma\|_F^2 + (N^{-1}p)^2\cdot N\|\Sigma - I\|_F^2.\end{aligned}$$

For the fourth variance bound, note that

$$\frac{\partial}{\partial Z_{ij}}\mathrm{Tr}(\Sigma S_Z\Sigma S_Z) = 2N^{-1}\mathrm{Tr}\Big[\Sigma S_Z\Sigma(e_j Z_i^\top + Z_i e_j^\top)\Big] = 4N^{-1}\big(Z\Sigma S_Z\Sigma\big)_{ij},$$

so by Gaussian-Poincaré inequality and Lemma 51,

$$\begin{aligned}&\mathrm{Var}\big(\mathrm{Tr}(\Sigma S_Z\Sigma S_Z)\big) \le 16N^{-2}\mathbb{E}\|Z\Sigma S_Z\Sigma\|_F^2\\ &= 16N^{-2}\mathbb{E}\mathrm{Tr}\big(Z\Sigma S_Z\Sigma\Sigma S_Z\Sigma Z^\top\big) = 16N^{-1}\mathbb{E}\mathrm{Tr}\big(\Sigma S_Z\Sigma\Sigma S_Z\Sigma S_Z\big)\\ &\le 16N^{-1}\mathbb{E}\|S_Z\|_{\mathrm{op}}^3\mathrm{Tr}(\Sigma^4) \lesssim N^{-1}\big(1\vee(N^{-1}p)\big)^3\mathrm{Tr}(\Sigma^4).\end{aligned}$$

(3). This follows by integrating the tail of $|b(\Sigma^{1/2}S_Z\Sigma^{1/2})-b(\Sigma)|$ in Lemma 61:

$$\mathbb{E}|b(\Sigma^{1/2}S_Z\Sigma^{1/2})-b(\Sigma)|^\ell = \int_0^\infty \ell t^{\ell-1}\mathbb{P}\big(|b(\Sigma^{1/2}S_Z\Sigma^{1/2})-b(\Sigma)|>t\big)\,\mathrm{d}t$$

$$\lesssim_\ell \int_0^\infty t^{\ell-1}e^{-\frac{Np^2}{\|\Sigma\|_F^2}t^2}\,\mathrm{d}t + \int_0^\infty t^{\ell-1}e^{-\frac{Np}{\|\Sigma\|_{\mathrm{op}}}t}\,\mathrm{d}t$$

$$\lesssim_\ell \big(\|\Sigma\|_F N^{-1/2}p^{-1}\big)^\ell + \big(\|\Sigma\|_{\mathrm{op}}N^{-1}p^{-1}\big)^\ell \asymp \big(\|\Sigma\|_F N^{-1/2}p^{-1}\big)^\ell.$$

The proof is complete. $\square$

Lemma 64. *Let $S_Z \equiv N^{-1}\sum_{i=1}^N Z_iZ_i^\top$ where Z_i's are i.i.d. $\mathcal{N}(0,I)$ in $\mathbb{R}^p$. With $y\equiv p/N$ the following hold:*

1. $\mathbb{E}\,Tr(S_Z^3) = py^2+3py+p+3y^2+3y+4y/N$.

2. $\mathbb{E}\,Tr^3(S_Z) = p^3+6py+8y/N$.

3. $\mathbb{E}\,Tr(S_Z)\,Tr(S_Z^2) = p^2y+p^2+py+4(y^2+y)+4y/N$.

4. $\mathbb{E}\,Tr^2(S_Z^2)$ *equals*

$$\begin{aligned}N^{-4}\big[&Np(p+2)(p+4)(p+6)+N(N-1)\big(p(p+2)\big)^2\\ &+2N(N-1)3p(p+2)+4N(N-1)p(p+2)(p+4)\\ &+4N(N-1)(N-2)p(p+2)+2N(N-1)(N-2)p^2(p+2)\\ &+N(N-1)(N-2)(N-3)p^2\big].\end{aligned}$$

5. $\mathbb{E}\,Tr(S_Z)\,Tr(S_Z^3)$ *equals*

$$\begin{aligned}N^{-4}\big[&Np(p+2)(p+4)(p+6)+N(N-1)p^2(p+2)(p+4)\\ &+3N(N-1)p(p+2)^2+3N(N-1)p(p+2)(p+4)\\ &+3N(N-1)(N-2)p(p+2)+3N(N-1)(N-2)p^2(p+2)\\ &+N(N-1)(N-2)(N-3)p^2\big].\end{aligned}$$

6. *Var*$\big(b(S_Z)b_3(S_Z) - b_2^2(S_Z)\big) = \mathcal{O}(p^2/N^2)$ *in the asymptotic regime* $p > N \to \infty$.

7. *For any* $k, \ell \in \mathbb{N}$, $\mathbb{E}\,Tr^k(S_Z^\ell) \le Cp^{k\ell}$ *for some constant* $C = C(k, \ell) > 0$.

Proof. Write S_Z for S in the proof for simplicity. Recall that if R follows a chi-squared distribution with an integer ν degrees of freedom, then

$$\mathbb{E}R^2 = \nu^2 + 2\nu, \quad \mathbb{E}R^3 = \nu^3 + 6\nu^2 + 8\nu, \quad \mathbb{E}R^4 = \nu(\nu+2)(\nu+4)(\nu+6).$$

Hence (1)-(3) follows from the following calculations: We have

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}(S^3) &= N^{-3} \cdot \mathbb{E} \sum_{i_1,i_2,i_3} (Z_{i_1}^\top Z_{i_2})(Z_{i_2}^\top Z_{i_3})(Z_{i_3}^\top Z_{i_1}) \\
&= N^{-3} \cdot \Big(\sum_{|(i_1,i_2,i_3)|=1} \mathbb{E}\|Z_1\|^6 + \sum_{|(i_1,i_2,i_3)|=2} \mathbb{E}\|Z_1\|^4 + \sum_{|(i_1,i_2,i_3)|=3} \mathbb{E}\|Z_1\|^2 \Big) \\
&= N^{-3} \cdot \big[N \cdot \mathbb{E}(\chi_p^2)^3 + (3N^2 - 3N)\mathbb{E}(\chi_p^2)^2 + N(N-1)(N-2) \cdot p \big] \\
&= N^{-3} \cdot \big[N(p^3 + 6p^2 + 8p) + (3N^2 - 3N)(p^2 + 2p) + N(N-1)(N-2)p \big] \\
&= N^{-3} \cdot \big[(Np^3 + 3N^2p^2 + N^3p) + 3(Np^2 + N^2p) + 4Np \big] \\
&= N^{-2}p^3 + 3N^{-1}p^2 + p + 3(N^{-1}p)^2 + 3N^{-1}p + 4N^{-2}p \\
&= py^2 + 3py + p + 3y^2 + 3y + 4N^{-1}y,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}^3(S) = N^{-3}\mathbb{E}\bigg(\sum_{i=1}^N \|Z_i\|^2 \bigg)^3 &= N^{-3}\mathbb{E}(\chi_{Np}^2)^3 \\
&= N^{-3}(N^3p^3 + 6N^2p^2 + 8Np) = p^3 + 6N^{-1}p^2 + 8N^{-2}p \\
&= p^3 + 6py + 8N^{-1}y,
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E}\mathrm{Tr}(S)\mathrm{Tr}(S^2) = N^{-3}\mathbb{E}\Big(\sum_{i_1=1}^{N}\|Z_{i_1}\|^2\Big)\Big(\sum_{i_2,i_3=1}^{N}(Z_{i_2}^\top Z_{i_3})^2\Big) \\
&= N^{-3}\mathbb{E}\sum_{i_1,i_2,i_3=1}^{N}\|Z_{i_1}\|^2(Z_{i_2}^\top Z_{i_3})^2 \\
&= N^{-3}\Big[\sum_{|(i_1,i_2,i_3)|=3}\mathbb{E}\|Z_1\|^2\cdot\mathbb{E}(Z_2^\top Z_3)^2 + \sum_{(i_2=i_3)\neq i_1}\mathbb{E}\|Z_1\|^2\cdot\mathbb{E}\|Z_1\|^4 \\
&\qquad + \sum_{|(i_2,i_3)|=|(i_1,i_2,i_3)|=2}\mathbb{E}\|Z_1\|^2(Z_1^\top Z_2)^2 + \sum_{|(i_1,i_2,i_3)|=1}\mathbb{E}\|Z_1\|^6\Big] \\
&= N^{-3}\big[N(N-1)(N-2)p^2 + (N^2-N)(p^3+2p^2) \\
&\qquad + 2(N^2-N)\big(p^2+2p\big) + N\big(p^3+6p^2+8p\big)\big] \\
&= N^{-3}\big[(N^3p^2+N^2p^3) + N^2p^2 + 4(N^2p+Np^2) + 4Np\big] \\
&= p^2y + p^2 + py + 4(y^2+y) + 4N^{-1}y.
\end{aligned}
$$

(4). By definition, we have

$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}^2(S^2) &= N^{-4}\mathbb{E}\Big(\sum_{i_1,i_1'=1}^{N}(X_{i_1}^\top X_{i_1'})^2\Big)^2 \\
&= N^{-4}\sum_{i_1,i_1',i_2,i_2'=1}^{N}\mathbb{E}(X_{i_1}^\top X_{i_1'})^2(X_{i_2}^\top X_{i_2'})^2.
\end{aligned} \tag{D.53}
$$

The right hand side of (D.53) breaks into $\sum_{i=1}^{7} A_i$, where A_1, A_2-A_4, A_5-A_6, and A_7 correspond to the cases where (i_1, i_1', i_2, i_2') take $1, 2, 3, 4$ distinct values, respectively:

- (A_1) When (i_1, i_1', i_2, i_2') take 1 value, there are N such summands each of which take the value $\mathbb{E}\|X_1\|^8 = p(p+2)(p+4)(p+6)$.

- (A_2) When (i_1, i_1', i_2, i_2') take 2 values with $(i_1 = i_1') \neq (i_2 = i_2')$, there are $N(N-1)$ such summands each of which takes the value $\mathbb{E}\|X_1\|^4\|X_2\|^4 = \big(\mathbb{E}\|X_1\|^4\big)^2 = p^2(p+2)^2$.

- (A_3) When (i_1, i_1', i_2, i_2') take 2 values with $(i_1 = i_2) \neq (i_1' = i_2')$ or $(i_1 = i_2') \neq (i_2 = i_1')$, there are $2N(N-1)$ such summands each of which takes the value

$$\begin{aligned}
\mathbb{E}(X_1^\top X_2)^2(X_1^\top X_2)^2 &= \mathbb{E}(X_1^\top X_2)^4 = \mathbb{E}\Big(\sum_{j=1}^{p} X_{1,j}X_{2,j}\Big)^4 \\
&= \sum_{j_1,j_2,j_3,j_4=1}^{p} \big(\mathbb{E}X_{1,j_1}X_{1,j_2}X_{1,j_3}X_{1,j_4}\big)^2 \\
&= \sum_{|(j_1,j_2,j_3,j_4)|=1} \big(\mathbb{E}X_{1,j_1}X_{1,j_2}X_{1,j_3}X_{1,j_4}\big)^2 \\
&\quad + \sum_{|(j_1,j_2,j_3,j_4)|=2} \big(\mathbb{E}X_{1,j_1}X_{1,j_2}X_{1,j_3}X_{1,j_4}\big)^2 \\
&= p \cdot 3^2 + 3p(p-1)\cdot 1 = 3p(p+2).
\end{aligned}$$

- (A_4) When (i_1, i_1', i_2, i_2') take 2 values of the form $(i_1 = i_2 = i_1') \neq i_2'$ or its variants, there are $4N(N-1)$ such summands each of which takes the value

$$\mathbb{E}\|X_1\|^4(X_1^\top X_2)^2 = \mathbb{E}\mathrm{Tr}(\|X_1\|^4 X_1X_1^\top X_2X_2^\top) = \mathbb{E}\|X_1\|^6 = p(p+2)(p+4).$$

- (A_5) When (i_1, i_1', i_2, i_2') take 3 values of the form $(i_1 = i_2) \neq i_1' \neq i_2'$ or its variants, there are $4N(N-1)(N-2)$ such summands each of which takes the value $\mathbb{E}(X_1^\top X_2)^2(X_1^\top X_3)^2 = \mathbb{E}\|X_1\|^4 = p(p+2)$.

- (A_6) When (i_1, i_1', i_2, i_2') take 3 values of the form $(i_1 = i_1') \neq i_2 \neq i_2'$ or its variants, there are $2N(N-1)(N-2)$ such summands each of which takes the value $\mathbb{E}(X_1^\top X_1)^2(X_2^\top X_3)^2 = p \cdot \mathbb{E}\|X_1\|^4 = p^2(p+2)$.

- (A_7) When (i_1, i_1', i_2, i_2') take 4 values, there are $N(N-1)(N-2)(N-3)$ such summands each of which takes the value $\mathbb{E}(X_1^\top X_2)^2(X_3^\top X_4)^2 = p^2$.

(5). By definition, we have

$$\begin{aligned}\mathbb{E}\big[\mathrm{Tr}(S)\mathrm{Tr}(S^3)\big] &= N^{-4}\mathbb{E}\Big(\sum_{i=1}^{N}\|X_i\|^2\Big)\Big(\sum_{j_1,j_2,j_3=1}^{N}\big(X_{j_1}^\top X_{j_2}\big)\big(X_{j_2}^\top X_{j_3}\big)\big(X_{j_3}^\top X_{j_1}\big)\Big)\\ &= N^{-4}\sum_{i,j_1,j_2,j_3=1}^{N}\mathbb{E}\Big[\|X_i\|^2\big(X_{j_1}^\top X_{j_2}\big)\big(X_{j_2}^\top X_{j_3}\big)\big(X_{j_3}^\top X_{j_1}\big)\Big]. \end{aligned} \tag{D.54}$$

The right hand side of (D.54) breaks into $\sum_{i=1}^{7} B_i$, where B_1, B_2-B_4, B_5-B_6, and B_7 correspond to the cases where (i, j_1, j_2, j_3) take $1, 2, 3, 4$ distinct values, respectively:

- (B_1) When (i, j_1, j_2, j_3) takes 1 value, there are N such summands in (D.54), each of which takes the value $\mathbb{E}\|X_1\|^8 = p(p+2)(p+4)(p+6)$.

- (B_2) When (i, j_1, j_2, j_3) take 2 values with $i \neq (j_1 = j_2 = j_3)$, there are $N(N-1)$ such summands each of which takes the value $\mathbb{E}\|X_1\|^2\|X_2\|^6 = p^2(p+2)(p+4)$.

- (B_3) When (i, j_1, j_2, j_3) take 2 values of the form $(i = j_1) \neq (j_2 = j_3)$ and its variants, there are $3N(N-1)$ such summands each of which takes the value

  $$\begin{aligned}&\mathbb{E}\|X_1\|^2(X_1^\top X_2)\|X_2\|^2(X_2^\top X_1) = \mathbb{E}\mathrm{Tr}\Big(\|X_1\|^2 X_1 X_1^\top \cdot \|X_2\|^2 X_2 X_2^\top\Big)\\ &= \mathrm{Tr}\Big(\mathbb{E}\|X_1\|^2 X_1 X_1^\top\Big)^2 \overset{(*)}{=} \mathrm{Tr}\big[(p+2)I_p\big]^2 = p(p+2)^2.\end{aligned}$$

  Here in $(*)$ we use the following fact by direct calculation

  $$\Big(\mathbb{E}\|X_1\|^2 X_1 X_1^\top\Big)_{k\ell} = \mathbb{E}\Big[\big(\sum_{m=1}^{p} X_{1,m}^2\big) X_{1,k} X_{1,\ell}\Big] = (p+2)\delta_{k\ell}.$$

- (B_4) When (i, j_1, j_2, j_3) take 2 values of the form $(i = j_1 = j_2) \neq j_2$ or its variants, there are $3N(N-1)$ such summands each of which takes the value $\mathbb{E}\|X_1\|^4(X_1^\top X_2)^2 = \mathbb{E}\|X_1\|^6 = p(p+2)(p+4)$.

- (B_5) When (i, j_1, j_2, j_3) take 3 values of the form $(i = j_1) \neq j_2 \neq j_3$ or its variants, there are $3N(N-1)(N-2)$ such summands each of which takes the value $\mathbb{E}\|X_1\|^2(X_1^\top X_2)(X_2^\top X_3)(X_3^\top X_1) = \mathbb{E}\|X_1\|^4 = p(p+2)$.

- (B_6) When (i, j_1, j_2, j_3) take 3 values of the form $i \neq (j_1 = j_2) \neq j_3$ or its variants, there are $3N(N-1)(N-2)$ such summands each of which takes the value $\mathbb{E}\|X_1\|^2\|X_2\|^2(X_2^\top X_3)^2 = \mathbb{E}\|X_1\|^2 \cdot \mathbb{E}\|X_2\|^4 = p^2(p+2)$.

- (B_7) When (i, j_1, j_2, j_3) take 4 values, there are $N(N-1)(N-2)(N-3)$ such summands each of which takes the value $\mathbb{E}\big[\|X_1\|^2(X_2^\top X_3)(X_3^\top X_4)$
$\cdot(X_4^\top X_2)\big] = p^2$.

(6). Let $F(X) \equiv b(S)b_3(S) - b_2^2(S)$. Then using for any $(i,j) \in [N] \times [p]$

$$\partial_{ij} b = 2(Np)^{-1}X_{ij}, \quad \partial_{ij} b_2 = 4(Np)^{-1}X_i^\top S e_j, \quad \partial_{ij} b_3 = 6(Np)^{-1}X_i^\top S^2 e_j,$$

we have

$$\begin{aligned}
\partial_{ij} F(X) &= 2(Np)^{-1}X_{ij} \cdot b_3 + 6(Np)^{-1}X_i^\top S^2 e_j \cdot b - 8(Np)^{-1}X_i^\top S e_j \cdot b_2 \\
&= (Np)^{-1}\Big(2b_3 X + 6bXS^2 - 8b_2 XS\Big)_{(ij)}.
\end{aligned}$$

Hence by the Gaussian-Poincaré inequality, we have by direct calculation

$$\begin{aligned}
\mathrm{Var}(F(X)) &\leq \mathbb{E}\|\nabla F(X)\|_F^2 \\
&= (Np)^{-1}\mathbb{E}\big(28bb_3^2 + 36b^2b_5 + 32b_2^2b_3 - 96bb_2b_4\big) \\
&= (Np^4)^{-1}\mathbb{E}\Big[28\mathrm{Tr}(S)\mathrm{Tr}^2(S^3) + 36\mathrm{Tr}^2(S^2)\mathrm{Tr}(S^5) + 32\mathrm{Tr}^2(S^2)\mathrm{Tr}(S^3) \\
&\qquad - 96\mathrm{Tr}(S)\mathrm{Tr}(S^2)\mathrm{Tr}(S^4)\Big].
\end{aligned}$$

The rest of the proof follows from similar arguments as in (4) and (5) by explicit calculation and cancellation of higher order terms; we omit the details.

(7). This follows directly from the property of the chi-squared distribution:

$$\begin{aligned}\mathbb{E}\mathrm{Tr}^k(S^\ell) &= N^{-k\ell}\mathbb{E}\Big[\sum_{i_1,\ldots,i_\ell=1}^{N}(X_{i_1}^\top X_{i_2})\cdots(X_{i_{\ell-1}}^\top X_{i_\ell})(X_{i_\ell}^\top X_{i_1})\Big]^k \\ &\le N^{-k\ell}\mathbb{E}\Big(\sum_{i_1,\ldots,i_\ell=1}^{N}\|X_{i_1}\|^2\cdots\|X_{i_\ell}\|^2\Big)^k = N^{-k\ell}\mathbb{E}\Big(\sum_{i=1}^{N}\|X_i\|^2\Big)^{k\ell} \\ &= N^{-k\ell}\mathbb{E}\big(\chi^2(Np)\big)^{k\ell} \lesssim_{k,\ell} p^{k\ell}.\end{aligned}$$

The proof is complete. $\square$