COMBINATORIAL INFEERENCE FOR LARGE-SCALE DATA ANALYSIS

Junwei Lu

A DISSERTATION
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Operations Research and Financial Engineering
Advisers: Professor Han Liu
Professor Jianqing Fan

June 2018
Abstract

Problems of inferring the combinatorial structures of networks arise in many real applications ranging from genomic regulatory networks, brain networks to social networks. This poses new and challenging problems on the uncertainty assessment and replicability analysis for statistical methods inferring these network topological structures. This thesis develops combinatorial inference, a new inferential framework for networks, to conduct hypothesis tests and variable selection for combinatorial structures in large-scale graphs.

In the first part of the thesis, we propose a unified inferential method to test hypotheses on the global combinatorial properties of graphical models. We showed that my method works for general monotone graph properties that can be preserved under edge deletion including bipartite, planar, k-colorable, etc. We develop a new combinatorial minimax theory to justify the optimality of the skip-down algorithm. We introduce a new notion of graph packing entropy to sharply characterize the complexity of combinatorial inference problems for graphical models. It can be viewed as the combinatorial counterpart of the famous metric entropy theory on parametric minimax lower bounds.

In the second part of the thesis, we generalize the combinatorial inference for larger family of graphical models. We propose a novel class of dynamic nonparanormal graphical models, which allows us to model high dimensional heavy-tailed systems and the evolution of their latent network structures. Under this model we develop statistical tests for presence of edges both locally at a fixed index value and globally over a range of values. The tests are developed for a high-dimensional regime, are robust to model selection mistakes and do not require commonly assumed minimum signal strength. The testing procedures are based on a high dimensional, debiasing-free moment estimator, which uses a novel kernel smoothed Kendall’s tau correlation matrix as an input statistic.
Acknowledgements

First, I would like thank my advisor, Professor Han Liu. When I was wandering outside of the grand gate of statistics and helplessly facing the dazzlingly forking paths behind, it was Han who lead me to the picturesque palace of statistics and machine learning. From then on, I am always exposed to his radiation of passionate enthusiasm for solving the theoretically charming problems and motivated by his steady steps towards the foggy territory of new research problems. I will never forget the days and nights in front of whiteboard, fighting with him for revealing certainty from clouds of randomness; being enlightened by him to disclose correlation in seemingly unrelated areas; and being saved by him from getting drowned in the boundless ocean of data.

I feel especially privileged to have Professor Jianqing Fan to sit in my dissertation committee. It is my greatest honor to work with one of the greatest statisticians of our time, who nurtured my interest and taste of statistics. I learned all the attributes of a mature researcher through his suggestions, encouragements, criticism and insights. Collaborating with Professor Fan is the most exciting adventure in my life.

I also owe special thanks to Professor Guang Cheng and Professor Mladen Kolar. I really appreciate the discussions with them about the interesting problems we worked together. They are always so generous to share their deep thoughts and marvelous ideas. I benefited a lot from their invaluable advices for my research, writing and presentation. I would to express my great attitude to Professor Mengdi Wang for serving on my dissertation committee. I also learned a lot about modern reinforcement learning from her, which is very inspiring.

I wish to thank my other collaborators Professor Fang Han, Professor Tuo Zhao, Professor Zhaoran Wang, Professor Kean-Ming Tan, Professor Wei Sun, Professor Matey Neykov, Dr. Heather Battey, Dr. Ziwei Zhu, Dr. Xinguo Li, Mr. Cong Ma and Mr. Yichen Chen. I am so proud of having an excellent group of collaborators, such that all of you have or will become faculty members in top-tier universities.
I was so fortunate to spend my graduate student life in the Department of Operations Research and Financial Engineering. The vibrant and diverse environment exposed me to various interesting research areas. I want to thank Professors Amir Ali Ahmadi, Sebastien Bubeck, Marc Hallin, Alain Kornhauser, Samory Kpotufe, Warren Powell, Philippe Rigollet, Ramon van Handel, Robert Vanderbei. Thank you for enriching my knowledge and skills in both research and teaching. I also want to express my gratitude to all the staffs at ORFE, Michael, Kim, Carol, Tara, Melissa, Connie, Tara, Lisa and Tabitha, who always help me to solve problems perfectly.

I would like to thank all my friends in ORFE. Life would not have been the same without you: Xingyuan Fang, Yang Ning, Tracy Ke, Xin Tong, Wei Dai, Lucy Xia, Juan Sagredo, Xiaofeng Shi, Jiawei Yao, Weichen Wang, Che-Yu Liu, Xiuneng Zhu, Zhao Chen, Dong Wang, Yuan Ke, Yan Li, Changle Lin, Lingzhou Xue, Dan Wang, Yuan Cao, Boyang Song, Tianqi Zhao, Quefeng Li, Yuyan Wang, Georgina Hall, Nana Aboagye, Donghwa Shin, Peiqi Wang, Ziwei Zhu, Koushiki Bose, Yiqiao Zhong, Quanquan Gu, Qiang Sun, Wenxin Zhou, Huanran Lu, Junchi Li, Wenyan Gong, Yongyi Guo, Suqi Liu, Han Hao, Jason Ge, Nongchao Li, Zongxi Li, Sinem Uysal, Levon Avanesyan, Hao Lu, Kaizheng Wang and Zhuoran Yang.

Finally, my deepest gratitude to my parents for their unconditional love and support. Thank you for every moment we spent together.
To my beloved parents
# Contents

Abstract .................................................................................................................. iii
Acknowledgements ................................................................................................... iv
List of Tables ............................................................................................................. xi
List of Figures ........................................................................................................... xii

1 Introduction ........................................................................................................... 1
   1.1 Graphical Models .......................................................................................... 1
   1.2 Local Statistical Inference .......................................................................... 3
   1.3 Outline of Combinatorial Inference .......................................................... 4
   1.4 Bibliographical Remarks ........................................................................... 6

2 Inference Methods and Theory ........................................................................... 10
   2.1 Preliminaries on Graph Theory ................................................................ 10
   2.2 Inferential Methods for Graph Invariants ................................................. 14
      2.2.1 A Generic Skip-Down Algorithmic Framework ............................. 16
      2.2.2 Case Study for Skip-Down Algorithm ........................................ 21
   2.3 Theory of Skip-down Method .................................................................. 26
      2.3.1 Validity of Tests and Confidence Intervals ................................... 26
   2.4 Power Analysis of the Tests .................................................................... 29
      2.4.1 Optimality and Adaptivity of the Confidence Intervals ............... 32
   2.5 Numerical Results ..................................................................................... 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.3</td>
<td>Proof of Theorem 4.3.6</td>
<td>155</td>
</tr>
<tr>
<td>C.3.1</td>
<td>Proof of Maximum Norm in (4.3.12)</td>
<td>156</td>
</tr>
<tr>
<td>C.3.2</td>
<td>Proof of $\ell_1$ Norm in (4.3.13)</td>
<td>159</td>
</tr>
<tr>
<td>C.4</td>
<td>Convergence Rate of Kendall’s Tau Estimator</td>
<td>161</td>
</tr>
<tr>
<td>C.4.1</td>
<td>Proof of Lemma C.1.1</td>
<td>163</td>
</tr>
<tr>
<td>C.4.2</td>
<td>Proof of Lemma C.1.2</td>
<td>164</td>
</tr>
<tr>
<td>C.5</td>
<td>Concentration of $U$-statistics</td>
<td>165</td>
</tr>
<tr>
<td>C.5.1</td>
<td>Proof of Lemma C.2.1</td>
<td>165</td>
</tr>
<tr>
<td>C.5.2</td>
<td>Proof of Lemma C.2.2</td>
<td>174</td>
</tr>
<tr>
<td>C.6</td>
<td>Proof of Theorem 4.3.3</td>
<td>176</td>
</tr>
<tr>
<td>C.7</td>
<td>Auxiliary Lemmas for Score Statistics</td>
<td>181</td>
</tr>
<tr>
<td>C.7.1</td>
<td>Approximation Error for Score Statistics</td>
<td>181</td>
</tr>
<tr>
<td>C.7.2</td>
<td>First Step Approximation Results</td>
<td>187</td>
</tr>
<tr>
<td>C.7.3</td>
<td>Properties of Bootstrap Score Statistics</td>
<td>191</td>
</tr>
<tr>
<td>C.7.4</td>
<td>Proof of Lemma 4.3.1</td>
<td>197</td>
</tr>
<tr>
<td>C.8</td>
<td>Results on Covering Number</td>
<td>199</td>
</tr>
<tr>
<td>C.9</td>
<td>Some Useful Results</td>
<td>211</td>
</tr>
</tbody>
</table>
List of Tables

2.1 The estimated coverage probability (the column Prob.) and averaged confidence interval length (the column Length) for $I_{\text{Conn}}$. We set the dimension $d = 100$, the sample size $n \in \{400, 600\}$, the values of the invariant $k = -25$, $k_\mu \in \{-25, -26, -27, -28\}$ and the signal strength $\mu \in \{0.2, 0.4, 0.6, 0.8\}$. The results are calculated based on 500 repetitions. ........................................ 38

2.2 The estimated coverage probability (the column Prob.) and averaged confidence interval length (the column Length) for $I_{\text{Deg}}$. We set the dimension $d = 100$, the sample size $n \in \{400, 600\}$, the values of the invariant $k = 5$, $k_\mu \in \{5, 4, 3, 2\}$ and the signal strength $\mu \in \{0.2, 0.4, 0.6, 0.8\}$. The results are calculated based on 500 repetitions. ........................................ 39

2.3 The estimated coverage probability (the column Prob.) and averaged confidence interval length (the column Length) for $I_{\text{Iso}}$. We set the dimension $d = 100$, the sample size $n \in \{400, 600\}$, the values of the invariant $k = -3$, $k_\mu \in \{-3, -5, -7, -9\}$ and the signal strength $\mu \in \{0.2, 0.4, 0.6, 0.8\}$. The results are calculated based on 500 repetitions. ........................................ 39
List of Figures

2.1 Examples of graph invariants and properties. .............................................. 12

2.2 Two examples of the critical edge sets $C_I(E_0)$ (in red edges) for (a) $I$ is the negative number of connected subgraphs, (b) $I$ is the maximum degree, (c) $I$ is the negative number of isolated nodes, and (d) $I = 0$ when the graph is a forest and $I = 1$ otherwise. The first row is for $E_0 = E_0^{(1)}$ and the second row is for the second example when $E_0 = E_0^{(2)}$. .......................................................... 17

2.3 An example of Algorithm 3 for $|\mathcal{I}_{\text{conn}}|$ from the 1st to 4th iteration. The blue dashed circles are the connected node sets $\mathcal{V}$, the red edges are the critical edge set $C_t$ and the black edges are $E_{t-1}$. In the 4th iteration, since $|\mathcal{V}| = 1$, we stop the algorithm. .......................................................... 24

2.4 The connected subgraphs in brain networks changing with the filtration level $\mu$. The upper panel illustrates the connected regions of interest by same colors. The lower panel shows how the number of connected subgraphs change with the filtration level $\mu$ for both intact story (in red) and word scrambled (in blue). 40

2.5 The degree of region of interest in brain networks changing with filtration level $\mu$. The upper panel illustrates the degree of each region of interest. The lower panel shows how the maximum degree changes with the filtration level $\mu$ for both intact story (in red) and word scrambled (in blue). 42
2.6 Structural differences in cerebral cortices between the intact story and word scrambled settings. (a) shows the connected regions of interest in brain networks under the filtration level \( \mu = 0.4 \). (b) shows the degree of regions of interest in brain networks under the filtration level \( \mu = 0 \).

3.1 The graph \( G_0 \) with two edges \( e, e' \in C : d_{G_0}(e, e') = 2, d = 10 \).

3.2 Null base graph \( G_0 \) with \( d - m - 1 \) edges, divider \( C \) (dashed), \( d_{G_0}(e, e') = \infty \), \( d = 7 \).

3.3 The graph \( G_0 \) with two (dashed) edges \( e, e' \in C \) such that \( d_{G_0}(e, e') = 2, d = 7 \).

3.4 The graph \( G_1 \) with \( d - m \) edges and \( m - 1 \) isolated nodes, \( d = 7 \).

3.5 The graph \( G_1 \) with \( m + 1 \) edges and two edges \( e, e' \in C \) with \( d_{G_1}(e, e') = 1 \).

3.6 Test for the maximum degree \( \mathcal{G}_0 := \{ G \mid d_{\text{max}}(G) \leq s_0 \} \) vs \( \mathcal{G}_1 = \{ G \mid d_{\text{max}}(G) \geq s_1 \} \) with \( s_0 = 3, s_1 = 5 \) and \( d = 18 \). The solid edges represent \( G_0 \in \mathcal{G}_0 \) with maximum degree \( s_0 = 3 \). We construct the divider \( C = \{(1, 5), (1, 6)\}, \{(7, 11), (7, 12)\}, \{(13, 17), (13, 18)\}\).

3.7 Visualization of the vertex buffer in \( V_{S, S'} \). Here \( S, S' \) are plotted with dashed and dotted edges respectively and \( G_0 \) is in solid edges. The vertices in the buffer are marked in the dashed squares.

3.8 Test for maximum degree \( \mathcal{G}_0 = \{ G \mid d_{\text{max}}(G) \leq s_0 \} \) and \( \mathcal{G}_1 = \{ G \mid d_{\text{max}}(G) \geq s_1 \} \) with \( s_0 = 3 \) and \( s_1 = 6 \).

3.9 Sparse clique and cycle detection with \( s = 5 \).

3.10 A visualization of a construction satisfying the conditions of Theorem 3.3.1 for the invariants \( \mathcal{I} \): (a) the maximum degree, (b) the negative number of connected subgraphs and (c) the negative number of isolated nodes.

A.1 Visualization of the proof of Proposition 2.2.1.
B.1 The null base $G_0$ and the divider $C$ (dashed) with $d_{G_0}(e,e') = \infty$, $d = 15,$ $m = 3$. \\
B.2 Test for the maximum degree $G_0 := \{ G \mid d_{\text{max}}(G) \leq s_0 \}$ vs $G_1 = \{ G \mid d_{\text{max}}(G) \geq s_1 \}$ with $s_0 = 3$, $s_1 = 5$ and $d = 18$. The solid edges represent $G_0 \in G_0$ with maximum degree $s_0 = 3$. We construct the divider $C = \{ ((1,5),(1,6)) , (7,11),(7,12)\},\{ (13,17),(13,18)\}$. \\
B.3 A closed walk $C$ from the set $C \in C_k(v,o,i,j)$. 
Chapter 1

Introduction

In this thesis, we are interested in developing statistical methods and theory for the hypothesis testing problems for large-scale graphical models. Graphical models are statistical models characterizing the conditional independency structures for random variables, which are widely used for modeling complex networks such as gene regulatory networks and brain connectivity networks.

Unlike classical inference which aims at testing a set of Euclidean parameters in the graphical models, this thesis aims to test some global structural properties. We aim to conduct hypothesis tests and construct confidence intervals for a large family of global graph properties including connectivity, degree and other topological structures of graphs.

1.1 Graphical Models

Given an undirected graph $G = (V, E)$ to represent the conditional dependency structure of a $d$-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$, where each vertex in $V = \{1, \ldots, d\}$ corresponds to a component of the random vector $\mathbf{X}$. Specifically, two nodes $j$ and $k$ are connected if and only if $X_j$ and $X_k$ are conditionally dependent given the other variables. We
say $X$ is Markov to the graph $G = (V, E)$, if its joint density $p_G(x)$ bears the factorization

$$p_G(x) = \exp \left( \sum_{c \in C} \theta_c(x_c) - A(\theta) \right), \quad (1.1.1)$$

where $C$ is the set of cliques in $G$, $x_c$ is a vector indexed by the clique $c$, and $A(\theta)$ is a probability density normalizer. The factorization in (1.1.1) reveals the topological structure of the graphical model. For example, the Gaussian Markov random field is a special family of graphical model with the density factorization

$$p_G(x) = \exp \left( - \sum_{(i,j) \in E} \Omega_{ij}(x_i - \mu_i)(x_j - \mu_j) - \frac{1}{2} \sum_{k=1}^d \Omega_{kk}(x_k - \mu_k)^2 - \frac{1}{2} \log[(2\pi)^d \det(\Sigma)] \right),$$

where $\det(\cdot)$ is the determinant, $\Sigma$ is the covariance matrix and $\Omega = \Sigma^{-1}$. The factorization in (1.1) implies that $(i, j) \in E$ if and only if $\Omega_{ij} \neq 0$. The Ising model characterizing binary values $\{+1, -1\}^d$ follows the probability mass function

$$P_G(x) = \exp \left( \sum_{(i,j) \in E} \theta_{ij}x_i x_j + \sum_{k=1}^d \phi_k x_k - A(\theta) \right),$$

for all $x \in \{+1, -1\}^d$. The graphical structure can be recovered by the support of the interaction parameters $\{\theta_{ij}\}_{i,j=1}^d$. A forest model is a nonparametric graphical model whose graph is a forest $F = (V, E)$ and its density $p_F(\cdot)$ can be written as

$$p_F(x) = \prod_{(i,j) \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{k=1}^d p(x_k).$$

Here $p(x_i, x_j)$ is the bivariate marginal density of $X_i$ and $X_j$, and $p(x_k)$ is the univariate marginal density of $X_k$.  

2
1.2 Local Statistical Inference

We use Gaussian graphical model as an example to show the local statistical estimation and inference for graph structures. Let \( X_1, \ldots, X_n \) be the i.i.d. observations of the random vector \( X \sim N(0, \Sigma) \). The precision matrix \( \Theta = \Sigma^{-1} \) encodes the underlying conditional independence graph \( G = (V, E) \). The local statistical inference for the Gaussian graphical model aims to estimate the entries of \( \Theta \) and conduct a hypothesis on a single edge \( H_0 : (j, k) \not\in E \).

Let the sample covariance be \( \hat{\Sigma} = n^{-1} \sum_{i=1}^{n} X_i X_i^T \). The CLIME estimator for the \( j \)-th column of the precision matrix \( \Theta \) for any \( j = 1, \ldots, d \) is defined as

\[
\hat{\Theta}_j = \arg\min_{\beta \in \mathbb{R}^d} \|\beta\|_1 \\
s.t. \|\hat{\Sigma}\beta - e_j\|_\infty \leq \lambda.
\]

(1.2.1)

where \( e_j \) is the \( j \)-th canonical basis in \( \mathbb{R}^d \). We assemble the columns and have the precision matrix estimator \( \hat{\Theta} = (\hat{\Theta}_1, \ldots, \hat{\Theta}_d) \). Cai et al. (2011) show that if we choose the tuning parameter \( \lambda \) in (4.1.1) as \( \lambda \geq CM\sqrt{\log d/n} \) for some sufficiently large \( C \), the CLIME estimator has the rate \( \|\hat{\Theta} - \Theta\|_1 = O_P(\sqrt{(s_2 \log d)/n}) \). We refer a detailed review of the literature on the graphical model estimation in Section 1.4.

For the single edge hypothesis test \( H_0 : \Theta_{jk} = \theta \), Neykov et al. (2015) propose the score function

\[
\hat{S}_{jk}(\beta) = \hat{\Theta}_j^T (\hat{\Sigma}\beta - e_k),
\]

(1.2.2)
given any vector \( \beta \), where \( e_k \) is the \( k \)-th canonical basis in \( \mathbb{R}^d \). We choose \( \beta \) by replacing \( \hat{\Theta}_{jk} \) by the tested value \( \Theta_{jk} = \theta \) and define the vector

\[
\hat{\Theta}_{k\setminus j}(\theta) = (\hat{\Theta}_{1k}, \ldots, \hat{\Theta}_{(j-1)k}, \theta, \hat{\Theta}_{(j+1)k}, \ldots, \hat{\Theta}_{dk})^T \in \mathbb{R}^d.
\]
Under $H_0 : \Theta_{jk} = \theta$, the score statistic

$$
\sqrt{n} S_{jk}(\hat{\Theta}_{k|j}(\theta)) \approx \sqrt{n} \Theta_j^T (\hat{\Sigma} - \Sigma) \Theta_k = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \Theta_j^T X_i X_i^T \Theta_k - \mathbb{E}[\Theta_j^T X_i X_i^T \Theta_k] \right),
$$

is shown to have asymptotic normality in Neykov et al. (2015). Therefore, they solve $\theta$ from the score equation $\sqrt{n} S_{jk}(\hat{\Theta}_{k|j}(\theta)) = 0$ and obtain the following de-biased CLIME estimator

$$
\hat{\Theta}^d_{jk} = \hat{\Theta}_{jk} - \frac{\hat{\Theta}_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k)}{\hat{\Theta}_j^T \hat{\Sigma}_j}.
$$

The debiased estimator $\hat{\Theta}^d_{jk}$ is also a kind of one-step estimator (Van der Vaart, 2000) by applying Newton-Rhapson’s method to (4.2.3) and it can be shown that that

$$
\sqrt{n}(\hat{\Theta}^d_{jk} - \Theta_{jk}) \overset{d}{\rightarrow} N(0, \text{Var}(\Theta_j^T X X^T \Theta_k)).
$$

### 1.3 Outline of Combinatorial Inference

This thesis aims to develop a new framework, *combinatorial inference*, to test global topological structures for graphical models. Comparing to the local statistical inference discussed above, the combinatorial inference consider the hypothesis in the form of

$$
H_0 : G \text{ does not satisfy certain property} \quad \text{versus} \quad H_1 : G \text{ satisfies certain property},
$$

where the property could be related connectivity, degree, clique size, etc. The topological structure of a graph is an important feature in many real applications. For instance, in a gene regulatory network, it is scientifically interesting to infer the number of metabolic cycles where cells move through in the genomic network (Luscombe et al., 2004). In neuroscience applications, one critical problem is to infer the degrees of various cerebral areas in the brain network during certain cognitive processes (Hagmann et al., 2008).
This thesis is organized as the following contributions to the combinatorial inference for graphical models.

• **Methodology.** In Chapter 2, we propose a fully data-driven method for the hypotheses on a wide family of graph properties, which also derives confidence intervals for quantities related to graph structures. We propose a skip-down algorithm for the nested hypotheses on combinatorial quantities, which iteratively screens critical edge sets for the graph invariant of interest. We also derive confidence intervals from the skip-down algorithm. Constructing confidence intervals for combinatorial quantities is challenging as the standard confidence interval theory on continuous and smooth parameters does not directly apply. We show that our proposed confidence interval is asymptotically honest for all monotone graph invariants. We provide the theoretical analysis of the method we proposed. We prove that our testing method can control the type I error and is optimal in power analysis. We also demonstrate that the length of our proposed confidence interval is adaptive to the signal strength of the precision matrix $\Theta$.

• **Fundamental Limits.** In Chapter 3, we develop a novel strategy for obtaining minimax lower bounds on the signal strength required to distinguish combinatorial graph structures which are separable via a single-edge divider. In particular, we relate the information-theoretic lower bounds to the packing number of the divider, which is an intuitive combinatorial quantity. To obtain this connection, we relate the chi-square divergence between two probability measures to the number of “closed walks” on their corresponding Markov graphs. Our analysis hinges on several technical tools including Le Cam’s Lemma, matrix perturbation inequalities and spectral graph theory. The usefulness of the approach is demonstrated by obtaining generic and interpretable lower bounds in numerous examples such as testing connectivity, connected components, self-avoiding paths, and cycles.

The lower bound of the confidence interval length for graph invariants is also studied. We provide a sufficient condition under which the confidence interval length is optimal for a
large family of invariants. Our sufficient condition is uniquely characterized by the geometry of the graph invariant. This makes it easier to show the lower bound of confidence interval length for many graph invariants.

- **Generalization to Complex Models.** Combinatorial inference can be generalized to complex graphical models. In Chapter 4, we propose a novel class of dynamic nonparanormal graphical models, which allows us to model high dimensional heavy-tailed systems and the evolution of their latent network structures. Under this model we develop statistical tests for presence of edges both locally at a fixed index value and globally over a range of values. The tests are developed for a high-dimensional regime, are robust to model selection mistakes and do not require commonly assumed minimum signal strength.

1.4 Bibliographical Remarks

Learning the structure of the graph in a graphical model has been widely studied in the literature. Many theoretical studies focus on achieving a perfect recovery of the true graph, i.e., they aim to find a consistent graph estimator \( \hat{G} \) satisfying \( P(\hat{G} = G) \to 1 \) as the sample size goes to infinity. For the Gaussian graphical model, many works estimate the graph through the precision matrix \( \Theta \) satisfying \( \Theta_{jk} \neq 0 \) if and only if the edge \((j, k) \in E\) (Meinshausen and Bühlmann, 2006; Yuan and Lin, 2007; Friedman et al., 2008; Peng et al., 2009; Lam and Fan, 2009; Ravikumar et al., 2011; Cai et al., 2011; Shen et al., 2012). For the Ising model, the true graph is determined by the sparsity pattern of the interaction parameters and we can estimate the parameters by sparse logistic regression (Ravikumar et al., 2010). A number of authors relax the Gaussian or Ising assumption by allowing the node-conditional distribution to belong to a univariate exponential family (see, e.g., Lee et al., 2006; Höfling and Tibshirani, 2009; Jalali et al., 2011; Anandkumar et al., 2012; Yang et al., 2012; Allen and Liu, 2012; Yang et al., 2013). This line of research extents to the mixed graphical model framework in which the conditional distributions of two nodes can belong
to two different distributions from a univariate exponential family (see, e.g., Fellinghauer et al., 2013; Lee and Hastie, 2015; Yang et al., 2014; Chen et al., 2015b). Semiparametric extensions using copulas are also developed in Liu and Wasserman (2009), Liu et al. (2012a), Xue and Zou (2012), and Liu et al. (2012b), and extended for mixed data in Fan et al. (2014). Danaher et al. (2014), Qiu et al. (2013), Mohan et al. (2014) consider joint estimation of multiple graphical models. However, in order to achieve the perfect graph recovery in these works, some strong conditions especially the minimal signal strength assumption are usually needed for the CLIME (Cai et al., 2011), the neighborhood selection (Meinshausen and Bühlmann, 2006; Zhou et al., 2009), the graphical Lasso (Lam and Fan, 2009) and the transelliptical graphical model (Liu et al., 2012b). The minimal signal strength assumption imposes that every edge has strong enough signal which is unlikely to hold in reality. All of the above mentioned work assumes that the graphical structure is static. However, in analysis of many complex systems, such an assumption is not valid. There are two major types of time-varying graphical model: directed and undirected. The directed time-varying graphical models are mainly studied in the context of autoregressive models with time-varying parameters (Punskaya et al., 2002; Fujita et al., 2007; Rao et al., 2007; Grzegorczyk and Husmeier, 2011; Kolar and Xing, 2009; Robinson and Hartemink, 2010; Jia and Huan, 2010; Lébre et al., 2010; Husmeier et al., 2010; Wang et al., 2011; Grzegorczyk and Husmeier, 2012; Dondelinger et al., 2012; Lébre et al., 2010). For the time-varying undirected graphical models, Zhou and Wasserman (2010), Kolar and Xing (2011), Yin et al. (2010), Kolar et al. (2010b) and Kolar et al. (2010a) consider the kernel-smoothed type estimator for graphical models. Kolar and Xing (2012) assume the graphical model evolves in a piecewise-constant fashion and estimate it by the temporally smoothed $\ell_1$ penalized regression. Talih and Hengartner (2005) and Xuan and Murphy (2007) consider a Bayesian framework to model the time-varying of graphs and estimate the graph by Markov chain Monte Carlo methods.
Hypothesis testing and confidence intervals for the high dimensional M-estimators are studied in Zhang and Zhang (2013), van de Geer et al. (2014); Javanmard and Montanari (2014), Belloni et al. (2013a), Belloni et al. (2013b), Javanmard and Montanari (2014) and Meinshausen (2013). Lu et al. (2015) considered the confidence bands for the high dimensional nonparametric models. Neykov et al. (2015) proposed the inference for high dimensional method of moments estimators. Lee et al. (2013) and Taylor et al. (2014) consider the conditional inference based on post-selection methods. Various inferential methods for high-dimensional graphical models were suggested (Liu, 2013; Jankova and van de Geer, 2015; Chen et al., 2015a; Ren et al., 2015; Neykov et al., 2015; Gu et al., 2015, e.g.), most of which focus on testing the presence of a single edge (except Liu (2013) who took the FDR approach (Benjamini and Hochberg, 1995) to conduct multiple tests and Gu et al. (2015) who developed procedures of edge testing in Gaussian copula models). None of the aforementioned works address the problem of combinatorial structure testing.

In addition to estimation and model selection procedures, efforts have been made to understand the fundamental limits of these problems. Lower bounds on estimation were obtained by Ren et al. (2015), where the authors show that the parametric estimation rate $n^{-1/2}$ is unattainable unless $s \log d/\sqrt{n} = o(1)$. Lower bounds on the minimal sample size required for model selection in Ising models were established by Santhanam and Wainwright (2012), where it is shown that support recovery is unattainable when $n \ll s^2 \log d$. In a follow up work, Wang et al. (2010) studied model selection limits on the sample size in Gaussian graphical models. The latter two works are remotely related to ours, in that both works exploit graph properties to obtain information-theoretic lower bounds. However, our problem differs significantly from theirs since we focus on developing lower bounds for testing graph structure, which is a fundamentally different problem from estimating the whole graph.

Our problem is most closely related to those in (Arias-Castro et al., 2012, 2015, 2011a; Addario-Berry et al., 2010) , which are inspired by the large body of research on minimax
hypothesis testing (Ingster, 1982; Ingster et al., 2010; Arias-Castro et al., 2015, 2011b, e.g.) among many others. Addario-Berry et al. (2010) quantify the signal strength as the mean parameter of a standard Gaussian distribution, while Arias-Castro et al. (2012, 2015) impose models on the covariance matrix of a multivariate Gaussian distribution. In our setup the parameter spaces of interest are designed to reflect the graphical model structure, and hence the signal strength is naturally imposed on the precision matrix. Arias-Castro et al. (2011b) provide detection bounds for the linear model. This is related to our work since one can view a linear model with Gaussian design as a Gaussian graphical model. Arias-Castro et al. (2015) address testing on a lattice based Gaussian Markov random field. For specific problems they establish lower bounds on the signal strength required to test the empty graph versus an alternative hypothesis. This is different from the setting of our problems, where the null hypothesis is usually not the empty graph.


Chapter 2

Inference Methods and Theory

In this chapter, we introduce the general framework for testing hypotheses for graph properties and then derive confidence intervals. Theoretical results on the validity of the tests and confidence intervals are also provided.

2.1 Preliminaries on Graph Theory

Let \( V = \{1, \ldots, d\} \) be the set of nodes and the edge set \( E \) be a subset of \( V \times V \). \( G = (V, E) \) denotes an undirected graph with vertices in \( V \) and edges in \( E \). We say \( G = (V, E) \) is isomorphic to \( G' = (V', E') \) if there exists a one-to-one mapping \( \sigma : V \to V' \), such that \((j, k) \in E\) if and only if \((\sigma(j), \sigma(k)) \in E'\). We also write \( G \preceq G' \) if \( V \subseteq V' \) and \( E \subseteq E' \). Let \( \mathcal{G}_d \) be the set of all graphs with \( d \) vertices. A graph invariant is a function \( \mathcal{I} : \mathcal{G}_d \to \mathbb{Z} \) such that \( \mathcal{I}(G) = \mathcal{I}(G') \) if \( G \) is isomorphic to \( G' \). In other words, a graph invariant is a geometric characterization of the graph, which is invariant to vertex permutations. We are interested in a special family of graph invariants called monotone invariants defined as follows.

Definition 2.1.1. A graph invariant \( \mathcal{I} \) is monotone whenever \( \mathcal{I}(G) \leq \mathcal{I}(G') \) for all \( G \preceq G' \).

Specifically, if a graph invariant takes binary values, i.e., \( \mathcal{P} : \mathcal{G}_d \to \{0, 1\} \) such that \( \mathcal{P}(G) = \mathcal{P}(G') \) if \( G \) is isomorphic to \( G' \), we call this invariant \( \mathcal{P} \) a graph property. We also
say a graph $G$ satisfies a property $\mathcal{P}$ if $\mathcal{P}(G) = 1$. Similarly to Definition 2.1.1, we define the monotone graph property as follows.

**Definition 2.1.2.** A graph property $\mathcal{P}$ is **monotone** whenever $\mathcal{P}(G) \leq \mathcal{P}(G')$ for all $G \preceq G'$.

In this thesis, we will always take the vertex set $V = \{1, \ldots, d\}$. Therefore, for simplicity of notation, given a graph $G = (V, E)$, we also write an invariant or a property as a function of edge set, i.e., $\mathcal{I}(E) = \mathcal{I}(G)$ and $\mathcal{P}(E) = \mathcal{P}(G)$. From the definitions above, we can see that a monotone graph invariant is non-decreasing under addition of edges. Similarly, if a graph $G$ satisfies a monotone property, the property is preserved under addition of edges to $G$.

Many extensively used graph invariants are monotone. For example, we can easily check that the following invariants are monotone:

- The maximum degree of $G$, which is the largest number of vertices connected to a single vertex in the graph (see Figure 2.1(a) for an example of a graph with maximum degree 5);

- The size of the longest chain in $G$, where a chain is a set of edges connecting a sequence of distinct vertices consecutively, and the size of a chain is the number of edges it contains (see Figure 2.1(b) for an example of a graph with longest chain of size 5);

- The size of the largest clique in $G$, where a clique is a subgraph such that any pair of its vertices are connected and the size of a clique is the number of vertices it contains (see Figure 2.1(c) for an example of a graph with largest clique size 5);

- The chromatic number of $G$, which is the smallest number of colors needed to color all vertices, so that no vertices sharing the same color are adjacent (see Figure 2.1(d) for an example of a graph with chromatic number 3);
The negative number of isolated nodes of $G$, where a isolated node is a vertex with no neighbor. Here we take the negative number because the number of isolated nodes is non-decreasing under the addition of edges (see Figure 2.1 (e) for an example of a graph with 5 isolated nodes);

The negative girth of $G$, where the girth of a graph is the length of the shortest cycle and the girth equals infinity when $G$ has no cycle. We take the negative girth also because the girth itself is non-increasing when adding edges (see Figure 2.1 (f) for an example of a graph with girth 5);

The negative number of connected subgraphs of $G$, where a subgraph is connected if any pair of its vertices are connected by a chain (see Figure 2.1 (g) for an example of a graph with connected subgraphs). Notice that the number of connected subgraphs is again non-increasing under the addition of edges and thus the negative number of connected subgraphs is monotone.

In the last three examples of monotone invariants listed above, we consider the negative values as more natural quantities. The negative sign is introduced since the invariants – number of isolated nodes, girth and number of connected subgraphs are non-increasing under the addition of edges, i.e., they are “monotone decreasing” instead of monotone increasing.
In order to unify our methodology, we focus only on monotone increasing invariants. Any monotone decreasing invariant, can be simply converted to a monotone increasing one by taking its negative value.

Any graph invariant $I$, induces a graph property $P_{I,k}$ defined as $P_{I,k}(G) = 0$ if $I(G) \leq k$ and $P_{I,k}(G) = 1$ if $I(G) > k$. It is easy to see that $P_{I,k}$ is a monotone graph property if $I$ is monotone. For example, when $I$ is the chromatic number, the induced property $P_{I,k}(G) = 0$ if and only if $G$ is $k$-colorable, which means we can assign each vertex a color from $k$ colors such that no two adjacent vertices share the same color, and in particular, $P_{I,2}(G) = 0$ if and only if $G$ is bipartite\(^1\). See Figure 2.1(e) for an example of a 3-colorable graph and Figure 2.1(j) for an example of a bipartite graph. Another example is when $I$ is the negative girth. The induced property $P_{I,-\infty}(G) = 0$ if and only if $G$ is a forest and the induced property $P_{I,-4}(G) = 0$ if and only if $G$ is triangle-free.

Therefore, the above examples of monotone invariants naturally imply corresponding examples of monotone properties. In addition to these natural examples, we also have the following examples of monotone properties:

- $G$ has a perfect matching, i.e. $G$ has a subset of edges without common vertices such that each vertex of $G$ is an endpoint of one of these edges (see Figure 2.1(h) for an example of a graph having perfect matching);

- $G$ is not planar, where a graph is planar if it can be drawn on the plane in such a way so that its edges only intersect at the vertices of the graph (see Figure 2.1(i) for an example of a planar graph);

- $G$ has a subgraph which is isomorphic to a given graph $H$.

The last property can actually derive a family of monotone graph properties given different subgraph pattern $H$. If the given graph $H$ is a $k$-star, the last property becomes the property

\(^1\)Recall that a bipartite graph is a graph whose vertices can be separated into two disjoint sets so that every edge connects vertices from one set to another.
that $G$ contains a star of size equal to or larger than $k$, which is equivalent to the induced property $\mathcal{P}_{I,k}$ when $I$ is the maximum degree. Similarly, we can also set $H$ to be a chain of size $k$ or a clique of size $k$ which corresponds to the induced property $\mathcal{P}_{I,k}$ for $I$ to be the size of the longest chain or the size of the largest clique. Another example is when $H$ is a graph with $d/2$ disjoint edges, which is equivalent to the existence of perfect matching. We visualize the aforementioned examples of $H$’s with the red edges in Figures 2.1(a) - 2.1(d) and 2.1(h).

2.2 Inferential Methods for Graph Invariants

In this section, we introduce the general framework for testing nested hypotheses for monotone graph invariants and then derive confidence intervals from the tests. Theoretical results on the validity of the tests and confidence intervals are also provided.

We begin with formulating our inferential problems on graph invariants. In order to illustrate the main idea of our testing procedure, we focus on the Gaussian graphical model first and discuss further extensions to other models in Chapter 4. Let $X = (X_1, \ldots, X_n)$ be i.i.d. observations of the random vector $X \sim N(0, \Sigma)$. The precision matrix $\Theta = \Sigma^{-1}$ encodes the underlying conditional independence graph $G = (V, E)$. In particular, the edge $(j, k) \in E$ if and only if $\Theta_{jk} \neq 0$. We consider the parameter space of precision matrices

$$\mathcal{U}_s = \left\{ \Theta \in \mathbb{R}^{d \times d} \mid \lambda_{\min}(\Theta) \geq 1/C, \max_{j \in [d]} \|\Theta_j\|_0 \leq s, \|\Theta\|_1 \leq L \right\}. \quad (2.2.1)$$

The above class requires the precision matrix to have at most $s$ nonzero entries for each column and therefore the graphs induced by such matrices have maximum degrees at most $s$.

We denote the range of a graph invariant $I$ as $[I_L^*, I_U^*]$. Throughout the thesis, $[I_L^*, I_U^*]$ usually represents the default range of a graph invariant. For example, for any graph property
\(\mathcal{P}\), the default range is \([0, 1]\). If \(\mathcal{I}\) is the negative number of isolated nodes, the default range is \([-d, 0]\) and for \(\mathcal{I}\) being the negative number of connected subgraphs, the range is \([-d, -1]\). Sometimes we may have prior knowledge on the precision matrix which can make the default range smaller. For example, if we know that \(\Theta \in \mathcal{U}_s\), the range of maximum degree can be set to \([0, s]\). In this thesis, we mainly focus on two inferential problems on graphical models. Let \(G\) be the graph which the graphical model is Markov with respect to. The first problem is testing multiple hypotheses with a nested structure

\[
H_{0k} : \mathcal{I}(G) \leq k \text{ versus } H_{1k} : \mathcal{I}(G) > k, \text{ for } k \in [I_L^*, I_U^*].
\]  

If \(\mathcal{I}(G) = k^*\), we have \(H_{0k}\) is true for \(k \geq k^*\) and \(H_{1k}\) is true for \(k < k^*\). For each \(k \in [I_L^*, I_U^*]\), we aim to propose a test \(\psi_k \in \{0, 1\}\) for \(H_{0k}\), such that the family-wise type-I error is controlled, i.e.,

\[
\lim_{n \to \infty} \mathbb{P}_{\mathcal{I}(G) = k^*} (\text{There exists a type-I error}) = \lim_{n \to \infty} \mathbb{P}_{\mathcal{I}(G) = k^*} (\exists k \geq k^* \text{ s.t. } \psi_k = 1) \leq \alpha.
\]

When an invariant is a property, since it only takes binary values, we have a special single hypothesis on graph property

\[
H_0 : \mathcal{P}(G) = 0 \text{ versus } H_1 : \mathcal{P}(G) = 1,
\]

where \(\mathcal{P}\) is a monotone property of interest. We aim to propose a test \(\psi \in \{0, 1\}\) at significance level \(\alpha\) such that

\[
\lim_{n \to \infty} \mathbb{P}_{H_0} (\psi = 1) \leq \alpha.
\]
The second problem is constructing a confidence interval of a monotone invariant $I$. We aim to construct a confidence interval $[\hat{I}_L, \hat{I}_U]$ at significance level $\alpha$ such that

$$\lim_{n \to \infty} \mathbb{P}(I(G) \in [\hat{I}_L, \hat{I}_U]) \geq 1 - \alpha.$$ 

### 2.2.1 A Generic Skip-Down Algorithmic Framework

In this subsection, we first describe the proposed skip-down algorithm (See Algorithm 1) for testing the hypothesis in (2.2.2). We then show that the same algorithm can also deliver a valid confidence interval for the graph invariant being tested.

To describe the algorithm, we first introduce a concept called "the critical edge set". Intuitively, for any graph invariant $I$ to be tested, we need to find a set of critical edges which may potentially change the value of $I$. A formal definition is as follows.

**Definition 2.2.1 (Critical edge set).** For any monotone graph invariant $I$ and an edge set $E_0 \subseteq V \times V$, we define the critical edge set of $I$ with respect to $E_0$ as

$$C_I(E_0) := \{ e \in E_0^c | \exists E' \supseteq E_0 \text{ s.t. } I(E') > I(E' \{e\}) \}.$$  

(2.2.3)

The definition of $C_I(E_0)$ is quite abstract. To understand its intuition, we first consider a special case when $I$ is a monotone graph property (denoted by $P$). As $P$ only takes binary values, $C_P(E_0)$ can be equivalently written as

$$C_P(E_0) := \{ e \in E_0^c | \exists E' \supseteq E_0 \text{ s.t. } P(E') = 1 \text{ and } P(E' \{e\}) = 0 \}.$$  

(2.2.4)

An edge $e$ is critical for $P$ with respect to $E_0$, if there exits a set of edges $\{e_1, \ldots, e_k\}$ that do not belong to $E_0 \cup \{e\}$ such that adding them to $E_0$ makes $(V, E_0 \cup \{e_1, \ldots, e_k, e\})$ satisfies $P$ but $(V, \{e_1, \ldots, e_k\})$ does not. This implies that $e$ has the "potential" to change the property of $E_0$ through edge addition. In Figure 2.2, we provide examples of critical
Examples of $E_0$ (a) Connectivity (b) Max degree (c) Singletons (d) Acyclic

Figure 2.2: Two examples of the critical edge sets $C_I(E_0)$ (in red edges) for (a) $I$ is the negative number of connected subgraphs, (b) $I$ is the maximum degree, (c) $I$ is the negative number of isolated nodes, and (d) $I = 0$ when the graph is a forest and $I = 1$ otherwise. The first row is for $E_0 = E_0^{(1)}$ and the second row is for the second example when $E_0 = E_0^{(2)}$.

edge sets for various invariants. From Figure 2.2, we see that $C_I(E_0)$ can be strictly smaller than $E_0^c$ for some invariants. This implies that we can ignore some edges and leads to a more powerful test. For a more detailed discussion of critical edges, see Section 2.2.2.

We begin by introducing the high level idea of testing the nested hypotheses in (2.2.2). At first, we do not have any information on the true graph. Therefore, we test the existence of edges in $C_I(\emptyset)$. Suppose we have rejected edges in some edge set $E_0 \subseteq C_I(\emptyset)$, then we can update our knowledge that $E_0$ is a subgraph of the true graph. Since $I$ is a monotone invariant, we also know that $I(G) \geq I(E_0)$. In the next step, we can further update our knowledge of the true graph by testing the existence of edges in $C_I(E_0)$. We can repeat this procedure for multiple times and iteratively refine our knowledge on $I(G)$.

An essential component step of the above idea is to test the existence of a certain edge set $E$. This can be achieved by a generic precision matrix estimator $\hat{\Theta}^d$ and a $(1 - \alpha)$ quantile estimator $c(\alpha, E)$ for the statistic $T_E = \max_{e \in E} \sqrt{n}(\hat{\Theta}_e^d - \Theta_e)$ given any edge set $E \subseteq V \times V$ satisfying

$$\lim_{n \to \infty} \sup_{\Theta \in \mathcal{U}_e} \left| \mathbb{P} \left( \max_{e \in E} \sqrt{n}(\hat{\Theta}_e^d - \Theta_e) > c(\alpha, E) \right) - \alpha \right| = 0.$$

(2.2.5)
In Section 1.2, we have introduced a specific debiased estimator $\hat{\Theta}^d$ in (1.2.4). For any edge set $E \subseteq V \times V$, we aim to estimate the quantile of the statistic

$$T_E := \max_{(j,k) \in E} \sqrt{n} (\hat{\Theta}_{jk}^d - \Theta_{jk}).$$

We consider the Gaussian multiplier bootstrap (Chernozhukov et al., 2013) by constructing the random variable

$$T^B := \sup_{(j,k) \in E} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\Theta}^T_j X_i X_i^T \hat{\Theta}^T_k - e_k \xi_i,$$

and we estimate the conditional quantile by $c(\alpha, E) = \inf \{ t \in \mathbb{R} \mid \mathbb{P}(|T^B_{jk}| > t \mid \{X_i\}_{i=1}^{n}) \leq \alpha \}$. We have the following lemma on controlling the tail-probability uniformly.

**Lemma 2.2.1.** Suppose that $\Omega \in \mathcal{U}(M,s)$. If $(\log(dn))^7/n + s^2(\log dn)^4/n = o(1)$, for any edge set $E \subseteq V \times V$, we have

$$\lim_{n \to \infty} \sup_{\Theta \in \mathcal{U}(M,s)} \mathbb{P} \left( \max_{e \in E} \sqrt{n} (\hat{\Theta}_{e}^d - \Theta_e) > c(\alpha, E) \right) \leq \alpha. \quad (2.2.6)$$

**Remark 1.** This lemma shows that $c(\alpha, E)$ obtained from the Gaussian multiplier bootstrap is a good quantile estimator of $\max_{e \in E} \sqrt{n} (\hat{\Theta}_{e}^d - \Theta_e)$. Similar result is also considered in Eq. (23) in Chernozhukov et al. (2013) for the multiple hypothesis testing. The first scaling condition $(\log(dn))^7/n = o(1)$ comes from the Berry-Essen bound of the Gaussian approximation for the maximum of a sum of high dimensional random vectors. The second condition $s^2(\log dn)^4/n = o(1)$ comes from the approximating the statistic $\sqrt{n} (\hat{\Theta}_{e}^d - \Theta_e)$ by a sum of i.i.d. random vectors. Similar condition can be found in Condition (M) of Chernozhukov et al. (2013).

Lemma 2.2.1 actually provides a generic condition for the property test introduced in Section 2.2. Given any precision matrix estimator $\hat{\Theta}^d$ and the quantile estimator $c(\alpha, E)$
satisfying (2.2.6), we can use them as the input of our property test and it can proved to be valid.

Algorithm 1 Skip-Down Method for Inferring Graph Invariant $I$

**Input:** $\{\hat{\Theta}^d_e\}_{e \in V \times V}$ and the range $[I_L^*, I_U^*]$.  
Initialize $t = 0$, $E_0 = \emptyset$ and the interval $[\hat{I}_L, \hat{I}_U] = [I_L^*, I_U^*]$.

repeat
    $t \leftarrow t + 1$;
    Select the screening set: $A \leftarrow C_I(E_{t-1})$;
    Update the rejected set: $E_t \leftarrow E_{t-1} \cup \{e \in A \mid \sqrt{n} \cdot |\hat{\Theta}^d_e| > c(\alpha, A)\}$;
    Update the interval $[\hat{I}_L, \hat{I}_U] \leftarrow [I(E_t), I_U^*]$;
until $I(E_t) \geq I_U^*$ or $E_t = E_{t-1}$

**Output:**
- **Nested hypotheses:** Let $\hat{I}_L = \min(\max(I(E_t), I_L^*), I_U^*)$. At the significance level $\alpha$, we reject $H_{0k}$ for $k < \hat{I}_L$ and not reject $H_{0k}$ for $k \in [\hat{I}_L, \hat{I}_U^*]$.
- **Confidence interval:** The $1 - \alpha$ confidence interval $[\hat{I}_L, \hat{I}_U^*]$.

Algorithm 1 implements the high level idea above to test hypotheses in (2.2.2) and construct a confidence interval for the monotone invariant $I$. We call this algorithm the skip-down method. The skip-down method is motivated by the step-down method (Romano and Wolf, 2005), which is designed for controlling the family-wise error of general multiple hypothesis tests. Our skip-down method explicitly exploits the nested structure of multiple hypotheses and has stronger power. Particularly, in each step, the step-down method tests all hypotheses which have not been previously rejected, while the skip-down method “skips” edges not belonging to the critical edge set, and hence improves the power of the test.

To construct confidence intervals, we choose the lower side as $\hat{I}_L$ in Algorithm 1 while the upper side is simply chosen to be $I_U^*$. This asymmetry is due to the monotonicity of graph invariants. In fact, we cannot construct a better upper bound for maximum degree according to the following theorem.
Theorem 2.2.1. Denote $\mathcal{I}_{\text{Deg}}(G)$ as the maximum degree of $G$. Define the family of confidence upper bound of $\mathcal{I}_{\text{Deg}}$ as

$$U(\mathcal{I}_{\text{Deg}}, \alpha) = \{ \hat{U}(\cdot) : \mathbb{R}^{d \times n} \to [0, s] \mid \inf_{\Theta \in \mathcal{U}} \mathbb{P}_{\Theta}(\mathcal{I}(\Theta) \leq \hat{U}(X)) \geq 1 - \alpha \}.$$  \hspace{1cm} (2.2.7)

Let $\mathcal{U}_s(\theta) = \{ \Theta \in \mathcal{U} \mid \min_{e \in E(\Theta)} |\Theta_e| \geq \theta \}$. If $s = o(d^{1/2})$ and $\theta \leq C \sqrt{\log d/n}$ for some sufficiently small positive constant $C$, we have

$$\liminf_{n \to \infty} \inf_{\hat{U} \in U(\mathcal{I}_{\text{Deg}}, \alpha)} \sup_{\Theta \in \mathcal{U}_s(\theta)} \mathbb{E}_{\Theta}[\hat{U} - \mathcal{I}_{\text{Deg}}(\Theta)] \geq s(1 - \alpha).$$  \hspace{1cm} (2.2.8)

We can achieve the lower bound (2.2.8) by a naive upper bound $\hat{U} = s$ with probability $1 - \alpha$ and $\hat{U} = 0$ with probability $\alpha$. Theorem 2.2.1 explains why we simply choose the confidence upper bound to be $I_U^*$ for maximum degree. Similar results hold for many other invariant, we refer to Theorems 2.4.2 for details.

Since graph properties are a special type of graph invariants with range $[0, 1]$, Algorithm 1 induces a test for $H_0 : \mathcal{P}(G) = 0$ versus $H_1 : \mathcal{P}(G) = 1$. Using the input $\{ \hat{\Theta}_e \}_{e \in V \times V}$ and the range $[0, 1]$ for Algorithm 1, we construct the test at a significance level $\alpha$ as $\psi_\alpha = 0$ if the output $\hat{I}_L = 0$ and $\psi_\alpha = 1$ otherwise. Following this idea, Algorithm 2 summarizes the procedure for testing graph properties.

An important step in the implementation of Algorithms 1 and 2 is scanning through the critical edge set $\mathcal{C}_T(E_0)$ for any given $E_0$. One general procedure for doing this is to exhaustively search all edges in $\mathcal{C}_T(E_0)$. Specifically, for all edge set $E' \supseteq E_0$, we search for every edge $e \in E' \setminus E_0$ and if $\mathcal{I}(E') > \mathcal{I}(E' \setminus \{e\})$, we add the edge $e$ into $\mathcal{C}_T(E_0)$. Although this is a valid procedure for any monotone invariant or property, the computational complexity is exponential to the dimension. On the other hand, for many specific invariants and properties, the structure of $\mathcal{C}_T(E_0)$ is simple and fast algorithms finding $\mathcal{C}_T(E_0)$ exist. In Section 2.2.2, we will give specific algorithms for finding the critical edge sets for three invariants and one
property: the maximum degree, the negative number of isolated nodes, the negative number of connected subgraphs and the property that a graph is not a forest.

**Algorithm 2** Skip-Down Method for Testing a Graph Property $\mathcal{P}$

**Input:** $\{\hat{\Theta}^d_e\}_{e \in V \times V}$

Initialize $t = 0, E_0 = \emptyset$.

repeat
\hspace{1em} $t \leftarrow t + 1$;
\hspace{1em} Select the screening set: $\mathcal{A} \leftarrow \mathcal{C}(E_{t-1})$;
\hspace{1em} Update the rejected set: $E_t \leftarrow E_{t-1} \cup \{e \in \mathcal{A} | \sqrt{n} \cdot |\hat{\Theta}^d_e| > c(\alpha, \mathcal{A})\}$;
until $\mathcal{P}(E_t) = 1$ or $E_t = E_{t-1}$

**Output:** $\psi_\alpha = 0$ if $\mathcal{P}(E_t) = 0$ and $\psi_\alpha = 1$ otherwise.

### 2.2.2 Case Study for Skip-Down Algorithm

In this section, we provide examples of graph invariants and show how to implement the skip-down algorithm for testing nested hypotheses and constructing confidence intervals of the following graph invariants:

- $\mathcal{I}_{\text{Conn}}(G) =$ the negative number of connected subgraphs of $G$;
- $\mathcal{I}_{\text{Deg}}(G) =$ the maximum degree of $G$;
- $\mathcal{I}_{\text{Iso}}(G) =$ the negative number of isolated nodes of $G$.

Recall that the induced graph property $\mathcal{P}_{\mathcal{I},k}(G) = 0$ if $\mathcal{I}(G) \leq k$ and $\mathcal{P}_{\mathcal{I},k}(G) = 1$ if $\mathcal{I}(G) > k$.

The above four invariants therefore naturally induce the following properties:

- $\mathcal{P}_{\text{Conn},-k}(G) = 1$ if and only if $G$ has less than $k$ connected subgraphs;
- $\mathcal{P}_{\text{Deg},k}(G) = 1$ if and only if $G$ has maximum degree larger than $k$;
- $\mathcal{P}_{\text{Sig},-k}(G) = 1$ if and only if $G$ has less than $k$ isolated nodes.

We also consider a fourth example of graph property:
• \( P_{\text{Cycle}}(G) = 0 \) if and only if \( G \) is a forest.

In order to apply Algorithms 1 and 2 for the above invariants and properties, there are two important steps; we need to first specify critical edge set \( C_I(E_0) \) given any edge set \( E_0 \), and second to calculate the invariant for any given graph. The second step can directly utilize existing algorithms for deterministic graphs. As mentioned previously, the critical edge set can be found via exhaustive search but the computation is not efficient in general. However, for the examples we give above, the critical edge sets can be selected explicitly. In the following proposition, we show the explicit forms of the critical edge sets for the above four properties.

**Proposition 2.2.1** (Critical edge sets for induced properties). Given any graph \( G_0 = (V, E_0) \), we have the following concrete forms of the critical edge sets.

- **Connected subgraphs.** Denote all connected subgraphs of \( G_0 \) by \( \{G_{0\ell} = (V_{0\ell}, E_{0\ell})\}^{k'}_{\ell=1} \). If \( P_{\text{Conn},-k}(G_0) = 1 \), i.e., \( k' < k \), we have \( C_{P_{\text{Conn},-k}}(E_0) = \emptyset \), otherwise

  \[
  C_{P_{\text{Conn},-k}}(E_0) = \{(u, v) \in E_0^c \mid u \in V_{0\ell}, v \in V_{0\ell'}, \ell \neq \ell'\}. \tag{2.2.9}
  \]

  Thus the critical edge set consists of all edges that link the connected subgraphs of \( G_0 \).

- **Maximum degree.** If \( P_{\text{Deg},k}(G_0) = 1 \), i.e., the maximum degree of \( G_0 \) is larger than \( k \), we have \( C_{P_{\text{Deg},k}}(E_0) = \emptyset \), otherwise \( C_{P_{\text{Deg},k}}(E_0) = E_0^c \).

- **Singletons.** Denote the set of all isolated nodes of \( G_0 \) by \( V_{\text{Sig}} \). If \( P_{\text{Sig},-k}(G_0) = 1 \), i.e., \( |V_{\text{Sig}}| < k \), we have \( C_{P_{\text{Sig},-k}}(E_0) = \emptyset \), otherwise

  \[
  C_{P_{\text{Sig},-k}}(E_0) = \{(u, v) \in E_0^c \mid u \in V_{\text{Sig}} \text{ or } v \in V_{\text{Sig}}\}. \tag{2.2.10}
  \]

  Therefore, the critical edge set contains the edges that connect the isolated nodes.
• **Acyclic.** If \( \mathcal{P}_{\text{Cycle}}(G) = 1 \), we have \( \mathcal{C}_{\mathcal{P}_{\text{Cycle}}}(E_0) = \emptyset \), otherwise when \( G_0 \) is a forest, \( \mathcal{C}_{\mathcal{P}_{\text{Cycle}}}(E_0) = E_0^c \).

The critical edge set for an invariant can be obtained from the one of its induced property. In fact, there exists a direct connection between the critical edge set of an invariant and its induced property:

\[
\mathcal{C}_I(E_0) = \bigcup_{k = I_L^U}^{I_U} \mathcal{C}_{\mathcal{P}_{k}}(E_0). \tag{2.2.11}
\]

This can be directly derived from Definition 2.2.1. On the one hand, for any \( e \in \mathcal{C}_I(E_0) \), there exists \( E' \supseteq E_0 \) such that \( I(E') > I(E' \setminus \{e\}) \), therefore \( e \in \mathcal{C}_{\mathcal{P}_{k}}(E_0) \) by (2.2.4). On the other hand, if \( e \in \mathcal{C}_{\mathcal{P}_{k}}(E_0) \) for some \( k_0 \in [I_L^*, I_U^*] \), by (2.2.4) and the definition of induced property, there exists \( E' \supseteq E_0 \) such that \( I(E') > k_0 \geq I(E' \setminus \{e\}) \), which implies that \( e \in \mathcal{C}_I(E_0) \). Therefore, by (2.2.11), we have a corollary of Proposition 2.2.1 on the critical edge set of invariants.

**Corollary 2.2.2 (Critical edge sets for invariants).** Given a graph \( G_0 = (V, E_0) \), as in Proposition 2.2.1, we denote all the connected subgraphs of \( G_0 \) as \( \{G_0^\ell = (V_0^\ell, E_0^\ell)\}_{\ell=1}^{k_0'} \) and all the set of isolated nodes of \( G_0 \) as \( V_{\text{Sig}} \). The critical edge sets for the four invariants are as follows.

- **Connected subgraphs:** \( \mathcal{C}_{I_{\text{Conn}}}(E_0) = \{(u, v) \in E_0^c \mid u \in V_0^\ell, v \in V_0^\ell', \ell \neq \ell'\} \);

- **Maximum degree:** \( \mathcal{C}_{I_{\text{Deg}}}(E_0) = E_0^c \);

- **Singletons:** \( \mathcal{C}_{I_{\text{Iso}}}(E_0) = \{(u, v) \in E_0^c \mid u \in V_{\text{Sig}} \text{ or } v \in V_{\text{Sig}}\} \).

The examples of the critical edges for invariants are visualized in Figure 2.2. Having the explicit forms of critical edge sets, we are now ready to implement Algorithms 1 and 2 for the above examples.

**Example 2.2.1 (Number of connected subgraphs).** By Corollary 2.2.2, we need to partition the node set \( V \) into disjoint connected node sets \( V = \{V_1, V_2, \ldots, V_{k'}\} \) representing connected
Figure 2.3: An example of Algorithm 3 for $I_{\text{Conn}}$ from the 1st to 4th iteration. The blue dashed circles are the connected node sets $V$, the red edges are the critical edge set $C_t$ and the black edges are $E_{t-1}$. In the 4th iteration, since $|V| = 1$, we stop the algorithm.

**Algorithm 3** Skip-Down Method for $I_{\text{Conn}}$

**Input:** $\{\hat{\Theta}_e^d\}_{e \in V \times V}$.

Initialize $t = 0, E_0 = \emptyset$ and connected node sets $V = \{\{1\}, \ldots, \{d\}\}$.

**repeat**

$t \leftarrow t + 1$;

Find the critical edge set $C_t = \{(u, v) \in E_{t-1} \mid u \in S, v \in T, S \neq T$ and $S, T \in V\}$.

Update the rejected set: $R_t = \{e \in C_t \mid \sqrt{n} \cdot |\hat{\Theta}_e^d| > c(\alpha, C_t)\}$;

$E_t \leftarrow E_{t-1} \cup R_t$;

Update the connected node sets $V$:

**for** $(u, v) \in R_t$ **do**

if $u, v$ belong to different node sets $S, T$ in $V$, i.e., $u \in S, v \in T$ and $S \neq T$ **then**

$V \leftarrow (V \backslash \{S, T\}) \cup \{S \cup T\}$

**end if**

**end for**

until $|V| = 1$ or $R_t = \emptyset$

**Output:**

- **Nested hypotheses:** We reject $H_0k$ for $k < -|V|$ and not reject $H_0k$ for $k \in [-|V|, -1]$.

- **Confidence interval:** The $1 - 2\alpha$ confidence interval $[\hat{I}_L, \hat{I}_U]$, where $\hat{I}_L = -|V|$ and $\hat{I}_U = -1$ with probability $1 - \alpha$ and $\hat{I}_U = -|V|$ with probability $\alpha$.

subgraphs in each iteration. We update the connected node sets $V_1, V_2, \ldots, V_k$, in each iteration of skip-down algorithm as follows. If an edge $(u, v)$ is rejected and $u \in V_1$ and $v \in V_2$, we take the union of $V_1$ and $V_2$ and update the connected node sets as $V = \{V_1 \cup V_2, V_3, \ldots, V_k\}$. A detailed description of the algorithm for the confidence interval of $I_{\text{Conn}}$ with a default range $[-d, -1]$ is shown in Algorithm 3. We visualize the algorithm in Figure 2.3. Similarly, we can also test $P_{\text{Conn,–}k}$ by modifying Algorithm 3.
Algorithm 4 Skip-Down Method for $I_{\text{iso}}$

**Input:** $\{\hat{\Theta}_e^d\}_{e \in V \times V}$

Initialize $t = 0$, $E_0 = \emptyset$ and isolated node set $V_{\text{Sig}} = \{1, \ldots, d\}$.

repeat

\begin{align*}
& t \leftarrow t + 1; \\
& \text{Find the critical edge set } C_t = \{(u, v) \in E_{t-1}^c \mid u \in V_{\text{Sig}} \text{ or } v \in V_{\text{Sig}}\}. \\
& \text{Update the rejected set: } R_t = \{e \in C_t \mid \sqrt{n} \cdot |\hat{\Theta}_e^d| > c(\alpha, C_t)\}; \\
& E_t \leftarrow E_{t-1} \cup R_t; \\
& \text{Update the connected node sets } V_{\text{Sig}}: \\
& \quad \text{for } (u, v) \in R_t \text{ do} \\
& \quad \quad V_{\text{Sig}} \leftarrow V_{\text{Sig}} \backslash \{u, v\}; \\
& \quad \text{end for}
\end{align*}

until $|V_{\text{Sig}}| = 0$ or $R_t = \emptyset$

**Output:**

- **Nested hypotheses:** We reject $H_0k$ for $k < -|V_{\text{Sig}}|$ and not reject $H_0k$ for $k \in [-|V|, 0]$.
- **Confidence interval:** The $1 - 2\alpha$ confidence interval $[\hat{I}_L, \hat{I}_U]$, where $\hat{I}_L = -|V_{\text{Sig}}|$ and $\hat{I}_U = 0$ with probability $1 - \alpha$ and $\hat{I}_U = -|V_{\text{Sig}}|$ with probability $\alpha$.

**Example 2.2.2** (Maximum degree). According to Corollary 2.2.2, the critical edge set is simply the complement of the rejected edge set. Therefore, for $I_{\text{Deg}}$, in the $t$-th iteration of Algorithm 1, we select the screening set $A = E_{t-1}^c$. In order to obtain $I(E_t)$, we can directly count the maximal number of neighbors for each node. Therefore, the algorithm for maximum degree confidence interval can directly implement Algorithm 1 by plugging in the explicit forms of $A = E_{t-1}^c$ and $I(E_t)$. Similar methods can be applied to test $P_{\text{Deg},k}$ by plugging in the set $C_{P_{\text{Deg},k}}(E_0)$ defined in Proposition 2.2.1 into Algorithm 2.

**Example 2.2.3** (Singletons). The critical edge set for $C_{I_{\text{Sig},-k}}(E_0)$ is shown in (2.2.10). It is apparent from this explicit form that we need to keep the track of the set of isolated nodes for each iteration. If an edge $(u, v)$ is rejected, we simply delete $u$ and $v$ from the set of isolated nodes if they belong to the set. A detailed description of step-down algorithm for $I_{\text{Sig},-k}$ with a default range $[-d, 0]$ is shown in Algorithm 4.

**Example 2.2.4** (Acyclic graph). The critical edge set is also the complement of the rejected edge set for the cyclicity property. Similarly to the maximum degree test, in the $t$-th iteration,
we also select the screening set $\mathcal{A} = E_i^{c-1}$. The procedure for checking whether $E_t \in \mathcal{P}_{\text{Cycle}}$, i.e., $E_t$ contains a cycle is similar to the one of detecting connectivity in Example 2.2.1. A detailed implementation of the test is shown in Algorithm 5.

**Algorithm 5 Skip-Down Method for $\mathcal{P}_{\text{Cycle}}$**

**Input:** $\{\hat{\Theta}^d_e\}_{e \in V \times V}$.

Initialize $t = 0$, $E_0 = \emptyset$ and path node sets $\mathcal{V}_{\text{Paths}} = \{\{1\}, \ldots, \{d\}\}$.

repeat
    $t \leftarrow t + 1$;
    Update the rejected set: $\mathcal{R}_t = \{e \in E^c_i | \sqrt{n} \cdot |\hat{\Theta}^d_e| > c(\alpha, E^c_i)\}$;
    $E_t \leftarrow E_{t-1} \cup \mathcal{R}_t$;
    Detect the cycles in the graph:
    for $(u, v) \in \mathcal{R}_t$ do
        if $u, v$ belong to different node sets $S, T$ in $\mathcal{V}_{\text{Paths}}$, i.e., $u \in S, v \in T$ and $S \neq T$ then
            $\mathcal{V}_{\text{Paths}} \leftarrow (\mathcal{V}_{\text{Paths}}\backslash\{S, T\}) \cup \{S \cup T\}$
        else
            $\mathcal{V}_{\text{Paths}} \leftarrow \emptyset$, Break;
        end if
    end for
until $\mathcal{R}_t = \emptyset$

Output: $\psi_\alpha = 0$ if $\mathcal{V}_{\text{Paths}} \neq \emptyset$ and $\psi_\alpha = 1$ otherwise.

### 2.3 Theory of Skip-down Method

In this section, we first prove the validity of obtained tests and confidence intervals described in Algorithm 1. We then show the optimality of the length of the confidence intervals for a family of monotone invariants. We show that the length of the confidence interval is adaptive to the signal strength.

#### 2.3.1 Validity of Tests and Confidence Intervals

Given the precision matrix $\Theta$, recall that $G(\Theta)$ is the graph induced by the support of $\Theta$. We also use the shorthand $\mathcal{I}(\Theta) = \mathcal{I}(G(\Theta))$ for the corresponding invariant for $G(\Theta)$.
and similarly put $\mathcal{P}(\Theta) = \mathcal{P}(G(\Theta))$ for the corresponding property. Given a graph invariant $\mathcal{I}$, we define the parameter space

$$
\mathcal{U}_s(I_L^*, I_U^*) = \{ \Theta \in \mathcal{U}_s | \mathcal{I}(\Theta) \in [I_L^*, I_U^*] \}.
$$

Here we assume $I_L^*$ and $I_U^*$ are known.

The following theorem shows the family-wise error of nested hypotheses and the asymptotic coverage probability of the confidence interval given by the skip-down method in Algorithm 1.

**Theorem 2.3.1 (Asymptotic coverage probability).** Suppose $\Theta \in \mathcal{U}_s$ and (2.2.6) is satisfied. Given any monotone invariant $\mathcal{I}$, the multiple test given by Algorithm 1 has

$$
\limsup_{n \to \infty} \sup_{\Theta \in \mathcal{U}_s(I_L^*, I_U^*)} \mathbb{P}_\Theta(\exists k \geq \mathcal{I}(\Theta) \text{ such that } H_0k \text{ is rejected}) \leq \alpha,
$$

and the confidence interval $[\hat{I}_L, I_U^*]$ satisfies

$$
\lim inf_{n \to \infty} \inf_{\Theta \in \mathcal{U}_s(I_L^*, I_U^*)} \mathbb{P}_\Theta(\mathcal{I}(\Theta) \in [\hat{I}_L, I_U^*]) \geq 1 - \alpha.
$$

Since graph properties are a special types of graph invariants, Theorem 2.3.1 actually implies the validity of the test $\psi_\alpha$ in Algorithm 2.

**Corollary 2.3.1 (Uniform asymptotic validity).** Given any monotone property $\mathcal{P}$, we define the parameter space

$$
\mathcal{G}_0 := \{ \Theta \in \mathcal{U}_s | \mathcal{P}(\Theta) = 0 \}.
$$

Under the same conditions of Theorem 2.3.1, the test $\psi_\alpha$ given by Algorithm 2 has

$$
\liminf_{n \to \infty} \inf_{\Theta \in \mathcal{G}_0} \mathbb{P}_\Theta(\psi_\alpha = 0) \geq 1 - \alpha.
$$
Both results of Theorem 2.3.1 and Corollary 2.3.1 are uniform over the parameter space \( U_s(I^*_L, I^*_U) \). Besides, the parameter space \( U_s(I^*_L, I^*_U) \) does not impose any restriction on the signal strength. The signal strength assumption is usually required to obtain a consistent graph estimator (Ravikumar et al., 2011; Cai et al., 2011; Liu et al., 2012b), such that \( \hat{G} \) satisfies \( \mathbb{P}(\hat{G} = G) \to 1 \) as \( n \to \infty \). The consistency of graph recovery has been shown for many graphical model estimators including CLIME (Cai et al., 2011), neighborhood selection (Meinshausen and Bühlmann, 2006; Zhou et al., 2009), graphical Lasso (Lam and Fan, 2009) and transelliptical graphical model (Liu et al., 2012b) and all these results need the minimal signal strength condition. Given any consistent graph estimator \( \hat{G} \), there is a plug-in invariant estimator \( \mathcal{I}(\hat{G}) \) such that \( \mathbb{P}(\mathcal{I}(\hat{G}) = \mathcal{I}(G)) \to 1 \). Similarly, to test the properties, one can also construct a plug-in test \( \psi := \mathcal{P}(\hat{G}) \). In comparison with the plug-in methods, the confidence intervals and tests obtained by Algorithms 1 and 2 have two major advantages: (1) they do not require any signal strength conditions and (2) one can choose different significance levels and thus have controllable uncertainty assessment for our inferential procedures.

From Theorems 2.3.1 and 2.4.1, we show that the test \( \psi_\alpha \) obtained from Algorithm 1 is optimal for the four examples we discuss in Section 2.2.2: \( \mathcal{P}_{\text{Conn},-k} \), \( \mathcal{P}_{\text{Deg},k} \) and \( \mathcal{P}_{\text{Sig},-k} \). To achieve this goal, we define the risk of a test \( \psi \) as

\[
R_\theta(\psi, \mathcal{P}) = \sup_{\Theta \in \mathcal{G}_0} \mathbb{P}_\Theta(\psi = 1) + \sup_{\Theta \in \mathcal{G}_1(\theta; \mathcal{P})} \mathbb{P}_\Theta(\psi = 0), \tag{2.3.5}
\]

where the parameter spaces \( \mathcal{G}_0 \) and \( \mathcal{G}_1(\theta; \mathcal{P}) \) are defined in (2.3.3) and (2.4.1) respectively. We say that the hypotheses \( H_0 : \Theta \in \mathcal{G}_0 \) v.s. \( H_{1,\theta} : \Theta \in \mathcal{G}_1(\theta; \mathcal{P}) \) are asymptotically separated by a test \( \psi \) if \( \lim_{n \to \infty} R_\theta(\psi, \mathcal{P}) = 0 \). On the other hand, \( H_0 \) and \( H_{1,\theta} \) are asymptotically inseparable if

\[
\liminf_{n \to \infty} \inf_{\psi} R_\theta(\psi, \mathcal{P}) = 1.
\]
From Theorems 2.3.1 and 2.4.1, the test $\psi_\alpha$ obtained from Algorithm 1 can asymptotically separate $H_0$ and $H_{1,\theta}$ for any monotone property $\mathcal{P}$ if the signal strength is strong enough. In specific, we have the following corollary directly from (2.3.2) and (2.4.2).

**Corollary 2.3.2.** Under the same conditions as Theorem 2.4.1, for any monotone property $\mathcal{P}$, we choose the level of significance $\alpha_n \in (0, 1)$ such that $1/(\alpha_n(d\sqrt{n})) = o(1)$ and $\alpha_n = o(1)$. The hypotheses $H_0 : \Theta \in G_0$ v.s. $H_{1,\theta} : \Theta \in G_1(\theta; \mathcal{P})$ can be asymptotically separated by the test $\psi_\alpha$ when $\theta \geq C_1 \sqrt{\log d/n}$ for some universal constant $C_1 > 0$.

On the other hand, we will show the lower bound on the signal strength in Chapter 3 and show that the signal strength in Corollary 2.4.2 is optimal for the properties including connectivity, maximum degrees and cyclic graph. In specific, we will show that there exists a constant $C_2 > 0$ such that if $\theta \leq C_2 \sqrt{\log d/n}$, and $\log d/n = o(1)$, for the following properties: (1) $\mathcal{P}_{\text{Conn},-k}$ for $k \in [1, d)$; (2) $\mathcal{P}_{\text{Cycle}}$; (3) $\mathcal{P}_{\text{Deg}, k}$ for $k \in [1, \sqrt{d})$ and $k \log d/n = o(1)$, the hypotheses $H_0 : \Theta \in G_0$ v.s. $H_{1,\theta} : \Theta \in G_1(\theta; \mathcal{P})$ are asymptotically inseparable. For $\mathcal{P}_{\text{Sig},-k}$ for $k \in [0, d)$, we can establish the same lower bound.

### 2.4 Power Analysis of the Tests

In this section, we discuss the power analysis for the test $\psi_\alpha$ obtained from Algorithm 2. Under the alternative $H_1 : \mathcal{P}(\Theta) = 1$, we define the parameter space

$$
G_1(\theta; \mathcal{P}) = \left\{ \Theta \in U_s \left| \mathcal{P}(G(\Theta)) = 1 \text{ and } \max_{E' \subseteq E(\Theta), \mathcal{P}(E') = 1} \min_{e \in E'} |\Theta_e| \geq \theta \right. \right\}.
$$

(2.4.1)

If $\Theta \in G_1(\theta; \mathcal{P})$, by (2.4.1), its induced graph $G(\Theta) = (V, E(\Theta))$ must has a sub-edge set $E_0 \subseteq E(\Theta)$ such that $\mathcal{P}(E_0) = 1$ and the minimal signal strength on $E_0$ is larger than $\theta$.

**Theorem 2.4.1** (Power analysis). Suppose $\Theta \in U_s$ and (2.2.6) is satisfied. Under the alternative hypothesis $H_1 : \mathcal{P}(G) = 1$, there exists a positive constant $C$ such that for any
\[ \alpha \in (0, 1) \text{ with } 1/(\alpha (d \lor n)) = o(1), \]

\[ \lim_{n \to \infty} \inf_{\Theta \in G_1(C\tau_n; \mathcal{P})} \mathbb{P}_\Theta(\psi_\alpha = 1) = 1, \quad (2.4.2) \]

where \( \tau_n = \sqrt{\log d/n} \).

Notice that the parameter space \( G_1(\theta; \mathcal{P}) \) defined in (2.4.1) is larger than the parameter space such that \( \Theta \) has the minimal signal strengths \( \theta \) on its support \( E(\Theta) \). Namely, we have

\[ G_1(\theta; \mathcal{P}) \supseteq G'_1(\theta; \mathcal{P}) = \left\{ \Theta \in \mathcal{U}_s \mid \mathcal{P}(G(\Theta)) = 1 \text{ and } \min_{e \in E(\Theta)} |\Theta_e| \geq \theta \right\}. \quad (2.4.3) \]

The parameter space \( G'_1(\theta) \) is usually considered for the power analysis of the global hypothesis tests (Han and Liu, 2014) or high dimensional two sample tests (Cai et al., 2013, 2014). For example, Han and Liu (2014) consider the null hypothesis \( H_0 : \Sigma = I_d \) and show their test is powerful if \( \min_{j \neq k} |\Sigma_{jk}| > C \sqrt{\log d/n} \) for some constant \( C > 0 \). Theorem 2.4.1 demonstrates that for the power analysis of graph property test, it suffices to impose the minimal signal strength condition on a subgraph \( E' \) of the true support \( E(\Theta) \) if \( \mathcal{P}(E') = 1 \).

The following proposition gives a concrete characterization of \( G_1(\theta) \).

**Proposition 2.4.1.** Given the graph \( G(\Theta) = (V, E(\Theta)) \), we set the weights on the edge \( (j, k) \) as \( |\Theta_{jk}| \) for any \( j, k \in V \). We order the weights \( |\Theta_e| \) for \( e \in V \times V \) as \( |\Theta_{e[1]}| \geq \ldots \geq |\Theta_{e[\lfloor d(d-1)/2 \rfloor]}| \). Let \( t^* = \arg\min \{ t \mid \mathcal{P}(E_{\overline{t}}) = \mathcal{P}(\{e[1], \ldots, e[t]\}) = 1 \} \). If the property \( \mathcal{P} \) is monotone, we have

\[ G_1(\theta; \mathcal{P}) = \left\{ \Theta \in \mathcal{U}_s \mid |\Theta_{e[t^*]}| \geq \theta \right\}. \]

**Proof.** By the definition in (2.4.1), since \( \mathcal{P}(E_{\overline{t^*}}) = 1 \), we have

\[ G_1(\theta; \mathcal{P}) \supseteq \left\{ \Theta \in \mathcal{U}_s \mid |\Theta_{e[t^*]}| \geq \theta \right\}. \]

30
It suffices to prove the other direction. If $\Theta \in G_1(\theta; \mathcal{P})$, there exists $E' \subseteq E(\Theta)$ with $\mathcal{P}(E') = 1$ such that $\min_{e \in E'} |\Theta_e| \geq \theta$. Since $\mathcal{P}$ is monotone, we have $\mathcal{P}(E_{t^*} \cup E') = 1$. As $\min_{e \in E'} |\Theta_e| \geq \theta$ and $E_{t^*}$ contains the top $t^*$ largest weight edges, we have $|\Theta_{e_{t^*}}| \geq \theta$. This completes the proof of the proposition.

**Remark 2.** Proposition 2.4.1 implies that we can add the edges in $E(\Theta)$ from the largest to the smallest until we stop at the $t^*$-th step when $\mathcal{P}(E_{t^*}) = 1$. For example, for the connectivity property $\mathcal{P}_{\text{Conn}(1)}$, we greedily add edges in order until the graph is connected. Comparing to the procedure of Kruskal’s algorithm (Kruskal, 1956), let $E_{\text{MSF}}$ be the edge set of the maximum spanning tree of $G(\Theta)$ with $\Theta$ as the weights, we can have

$$G_1(\theta; \mathcal{P}_{\text{Conn},-1}) = \left\{ \Theta \in \mathcal{U}_s \mid \min_{e \in E_{\text{MSF}}} |\Theta_e| \geq \theta \right\}.$$

From Theorems 2.3.1 and 2.4.1, we show that the test $\psi_\alpha$ obtained from Algorithm 1 is optimal for the four examples we discuss in Section 2.2.2: $\mathcal{P}_{\text{Conn},-k}$, $\mathcal{P}_{\text{Deg},k}$ and $\mathcal{P}_{\text{Sig},-k}$. To achieve this goal, we define the risk of a test $\psi$ as

$$R_\theta(\psi, \mathcal{P}) = \sup_{\Theta \in G_0} \mathbb{P}_\Theta(\psi = 1) + \sup_{\Theta \in G_1(\theta; \mathcal{P})} \mathbb{P}_\Theta(\psi = 0),$$

(2.4.4)

where the parameter spaces $G_0$ and $G_1(\theta; \mathcal{P})$ are defined in (2.3.3) and (2.4.1) respectively. We say that the hypotheses $H_0 : \Theta \in G_0$ v.s. $H_{1,\theta} : \Theta \in G_1(\theta; \mathcal{P})$ are asymptotically separated by a test $\psi$ if $\lim_{n \to \infty} R_\theta(\psi, \mathcal{P}) = 0$. On the other hand, $H_0$ and $H_{1,\theta}$ are asymptotically inseparable if

$$\liminf_{n \to \infty} \inf_{\psi} R_\theta(\psi, \mathcal{P}) = 1.$$

From Theorems 2.3.1 and 2.4.1, the test $\psi_\alpha$ obtained from Algorithm 1 can asymptotically separate $H_0$ and $H_{1,\theta}$ for any monotone property $\mathcal{P}$ if the signal strength is strong enough. In specific, we have the following corollary directly from (2.3.2) and (2.4.2).
Corollary 2.4.2. Under the same conditions as Theorem 2.4.1, for any monotone property $\mathcal{P}$, we choose the level of significance $\alpha_n \in (0, 1)$ such that $1/(\alpha_n(d\vee n)) = o(1)$ and $\alpha_n = o(1)$. The hypotheses $H_0 : \Theta \in \mathcal{G}_0$ v.s. $H_{1,\theta} : \Theta \in \mathcal{G}_1(\theta; \mathcal{P})$ can be asymptotically separated by the test $\psi_{\alpha_n}$ when $\theta \geq C_1 \sqrt{\log d/n}$ for some universal constant $C_1 > 0$.

On the other hand, the lower bound on the signal strength (see Chapter 3) shows that the signal strength in Corollary 2.4.2 is optimal for the properties including connectivity, maximum degrees and cyclic graph. In specific, we show that there exists a constant $C_2 > 0$ such that if $\theta \leq C_2 \sqrt{\log d/n}$ and $\log d/n = o(1)$, for the following properties: (1) $\mathcal{P}_{\text{Conn},-k}$ for $k \in [1, d)$; (2) $\mathcal{P}_{\text{Cycle}}$; (3) $\mathcal{P}_{\text{Deg},k}$ for $k \in [1, \sqrt{d})$ and $k \log d/n = o(1)$, the hypotheses $H_0 : \Theta \in \mathcal{G}_0$ v.s. $H_{1,\theta} : \Theta \in \mathcal{G}_1(\theta; \mathcal{P})$ are asymptotically inseparable. For $\mathcal{P}_{\text{Sig},-k}$ for $k \in [0, d)$, we can establish the same lower bound.

2.4.1 Optimality and Adaptivity of the Confidence Intervals

Theorem 2.3.1 proves that the confidence intervals $[\hat{I}_L, I_U^*]$ constructed by Algorithm 1 is asymptotically honest uniformly over the parameter space $\mathcal{U}_{s}(I_L^*, I_U^*)$. In this subsection, we show that the averaged length of confidence interval obtained in Algorithm 1 is optimal in the minimax sense.

Define a parameter space with minimal signal strength $\theta$ as

$$
\mathcal{U}_{\zeta}(I_L^*, I_U^*; \theta) = \left\{ \Theta \in \mathcal{U}_{s} \left| I(\Theta) \in [I_L^*, I_U^*], \min_{e \in E(\Theta)} |\Theta_e| \geq \theta \right. \right\}. 
$$

(2.4.5)

Let the family of confidence upper bound of $I$ be

$$
U(I, \alpha) = \left\{ \hat{U}(\cdot) : \mathbb{R}^{d \times n} \rightarrow [I_L^*, I_U^*] \left| \inf_{\Theta \in \mathcal{U}_{\zeta}(I_L^*, I_U^*)} \mathbb{P}_{\Theta}(I(\Theta) \leq \hat{U}(X)) \geq 1 - \alpha \right. \right\}.
$$

Theorem 2.2.1 shows we cannot construct non-trivial confidence upper bound for maximum degree. The following theorem shows that a similar result holds for $\mathcal{I}_{\text{Conn}}$ and $\mathcal{I}_{\text{Iso}}$ as well.
Theorem 2.4.2 (Lower bound of confidence upper bounds). Suppose $I^*_L \geq d/2$ and $I^*_U - I^*_L = o(d^{1/2})$. If $\theta \leq C \sqrt{\log d/n}$ for some sufficiently small positive constant $C$, we have

$$\liminf_{n \to \infty} \inf_{\tilde{U} \in U(\mathcal{I}, \alpha)} \sup_{\Theta \in \mathcal{U}(I^*_L, I^*_U; \theta)} \mathbb{E}_\Theta[\tilde{U} - \mathcal{I}(\Theta)] \geq (I^*_U - I^*_L)(1 - \alpha), \quad (2.4.6)$$

for $\mathcal{I} = \mathcal{I}_{\text{Conn}}$ or $\mathcal{I} = \mathcal{I}_{\text{Iso}}$.

Similar to Theorem 2.2.1, the lower bound in (2.4.6) can be achieved by a naive upper side $\hat{U} = I^*_U$ with probability $1 - \alpha$ and $\hat{U} = I^*_L$ with probability $\alpha$. This explains why in Algorithm 1, we only simply choose the upper bound as $I^*_U$. For the result of generic invariant confidence upper bound, we refer to Theorem 3.3.2.

Although we cannot construct a non-trivial confidence upper bound for many invariants, we can still construct good enough confidence intervals. Define the family of honest confidence intervals as

$$I(\mathcal{I}, \alpha) = \left\{ [\hat{L}, \hat{U}] \mid \hat{L}(X) \leq \hat{U}(X) \text{ a.s., } \inf_{\Theta \in \mathcal{U}(I^*_L, I^*_U)} \mathbb{P}_\Theta(I(\Theta) \in [\hat{L}(X), \hat{U}(X)]) \geq 1 - \alpha \right\}. \quad (2.4.7)$$

The following theorem gives the lower bounds of confidence interval length for $\mathcal{I} = \mathcal{I}_{\text{Deg}}, \mathcal{I}_{\text{Conn}}$ and $\mathcal{I}_{\text{Iso}}$.

Theorem 2.4.3 (Lower bound of confidence intervals). Given a precision matrix $\Theta \in \mathcal{U}_I(I^*_L, I^*_U; \theta)$, we define its significant edge set as

$$E_{\text{Sig}}(\Theta) := \{(j, k) \mid |\Theta_{jk}| \geq C \sqrt{\log d/n}\},$$

where $C$ is some sufficiently large constant. We then define the oracle length as

$$\text{Oracle Length}(\Theta) = I^*_U - \mathcal{I}(E_{\text{Sig}}(\Theta)). \quad (2.4.8)$$
For $\mathcal{I} = \mathcal{I}_{\text{Deg}}, \mathcal{I}_{\text{Conn}}$ or $\mathcal{I}_{\text{Iso}}$, if $\theta \leq C' \sqrt{\log d/n}$ for some sufficiently small positive constant $C'$, we have

$$\liminf_{n \to \infty} \inf_{[\hat{L}, \hat{U}] \in \mathcal{I}(\mathcal{I}, \alpha)} \sup_{\Theta \in \mathcal{U}(\mathcal{I}, \mathcal{I}^*_L, \mathcal{I}^*_U; \theta)} \frac{\mathbb{E}_{\Theta}[\hat{U} - \hat{L}]}{\text{Oracle Length}(\Theta)} \geq 1 - 2\alpha. \quad (2.4.9)$$

By the definition in (2.4.8), Oracle Length$(\Theta)$ becomes smaller, if there are more entries in $\Theta$ with signal strength larger than $C \sqrt{\log d/n}$. Namely, the oracle length is adaptive to the number of edges to strong signal strength. Therefore, (2.4.9) implies that the lower bound of the confidence interval length is adaptive to the significant edge set $E_{\text{Sig}}(\Theta)$. For the lower bound of generic invariants, we refer to Theorem 3.3.1 in Section 3.3.

From (2.4.9), it is straightforward to derive that

$$\liminf_{n \to \infty} \inf_{[\hat{L}, \hat{U}] \in \mathcal{I}(\mathcal{I}, \alpha)} \sup_{\Theta \in \mathcal{U}(\mathcal{I}^*_L, \mathcal{I}^*_U; \theta)} \mathbb{E}_{\Theta}[\hat{U} - \hat{L}] \geq (1 - 2\alpha)(I^*_U - I^*_L).$$

This implies that if there is no edge with strong enough signal strength, we only have a trivial rate $O(I^*_U - I^*_L)$ for the confidence interval length.

Now we discuss the upper bound the confidence interval length from Algorithm 1. The following theorem shows that it achieves the lower bound in Theorem 2.4.3.

**Theorem 2.4.4** (Size of confidence interval). Suppose $\Theta \in \mathcal{U}_a$ and (2.2.6) is satisfied. For any monotone invariant $\mathcal{I}$ with range $[I^*_L, I^*_U]$, if $I^*_U - I^*_L = O(d^2)$, for any $\alpha \in (0, 1)$ and $\theta > 0$,

$$\lim_{n \to \infty} \sup_{\Theta \in \mathcal{U}(\mathcal{I}_L, \mathcal{I}_U; \theta)} \frac{\mathbb{E}_{\Theta}[\hat{I}_U - \hat{I}_L]}{\text{Oracle Length}(\Theta) + 1} \leq 1. \quad (2.4.10)$$

We add one in the denominator Oracle Length$(\Theta) + 1$ of (2.4.10) just to avoid singularity when $I^*_U = \mathcal{I}(E_{\text{Sig}}(\Theta))$. We can see that the length is adaptive to the value $\mathcal{I}(E_{\text{Sig}}(\Theta))$. As we argued above, the oracle length in (2.4.10) shows the first level of adaptivity for the skip-down algorithm: the length of our confidence interval is smaller if there are more edges with strong enough signal strength. The assumption that $I^*_U - I^*_L = O(d^2)$ is satisfied for all examples.
in Section 2.1. In fact, this assumption is mild in the sense that for monotone \( \mathcal{I} \), there are at most \( d(d - 1)/2 \) possible values and we can easily rescale \( \mathcal{I} \) such that \( I^*_U - I^*_L = O(d^2) \).

Theorem 2.4.2 shows that it is impossible to construct an adaptive upper side of confidence interval. In fact, the following theorem shows that the oracle length in (2.4.10) mainly comes from the lower side.

**Theorem 2.4.5** (Size of confidence lower bound). Suppose \( \Theta \in \mathcal{U}_s \). If (2.2.6) is satisfied, for any monotone invariant \( \mathcal{I} \) with range \([I^*_L, I^*_U]\), if \( I^*_U - I^*_L = O(d^2) \), for any \( \alpha \in (0, 1) \) and \( \theta > 0 \), we have

\[
\lim_{n \to \infty} \sup_{\Theta \in \mathcal{U}(I^*_L, I^*_U; \theta)} \frac{\mathbb{E}_{\Theta}[\mathcal{I}(\Theta) - \tilde{I}_L]}{\mathcal{I}(\Theta) - \mathcal{I}(E_{\text{Sig}}(\Theta)) + 1} \leq 1. \tag{2.4.11}
\]

Similar to (2.4.10), we add one in the denominator \( I^*_U - I(E_{\text{Sig}}(\Theta)) + 1 \) of (2.4.11) just to avoid singularity. We remark that (2.4.11) also gives the type II error analysis for nested hypotheses in (2.2.2). In Algorithm 1, we do not reject \( H_{0k} \) for \( k \in [\hat{I}_L, I^*_U] \), thus the number of type II errors is \( \max\{\mathcal{I}(\Theta) - \hat{I}_L, 0\} \). In fact, the proof of Theorem 2.4.5 shows that the expected number of type II errors has

\[
\lim_{n \to \infty} \mathbb{E}_{\Theta}[\mathcal{I}(\Theta) - \tilde{I}_L] \leq \mathcal{I}(\Theta) - \mathcal{I}(E_{\text{Sig}}(\Theta)).
\]

Therefore, when the minimal signal strength satisfies \( \min_{(j,k) \in E} |\Theta_{jk}| \geq C \sqrt{\log d/n} \) for sufficiently large \( C \), there is asymptotically no type II error.

### 2.5 Numerical Results

In this section we show the numerical performance for the proposed confidence interval to three graph invariants: the negative number of connected subgraphs, the maximum
degree and the negative number of isolated nodes. In addition we apply our method to a neuroimaging dataset.

2.5.1 Synthetic Data

We first implement Algorithm 1 for synthetic simulations. Here we use the confidence interval to illustrate the performance of the skip-down algorithm. We consider three invariants: $I_{\text{Conn}}$, $I_{\text{Deg}}$, and $I_{\text{Iso}}$. Let $X_1, \ldots, X_n$ be i.i.d. $n$ samples generated from $N(0, \Theta^{-1})$, where the precision matrix $\Theta \in \mathbb{R}^{d \times d}$. In order to illustrate Theorem 2.4.4 and show how the confidence interval length changes adaptively with the value of $I(E_{\text{Sig}}(\Theta))$, we choose two parameters $k \geq k_\mu$ and generate the precision matrix $\Theta$ satisfying $I(\Theta) = k$ and $I(E_{\text{Sig}}(\Theta)) = k_\mu$. To generate such a precision matrix, we set the minimal signal strength $\theta = 0.01\sqrt{\log d/n}$ and we set part of the non-zero entries of $\Theta$ as $\mu > \theta$. To calculate the average coverage probability and average confidence interval length, we repeat the simulation 500 times. We allow the precision matrix to change in different repetitions to show that our confidence interval is honest.

Specifically, we construct the precision matrices following different scenarios for three different invariants as follows.

- $I_{\text{Conn}}$: Denote the adjacency matrix of a $d$-chain with $A_{\text{chain}(d)} \in \mathbb{R}^{d \times d}$. Formally, $[A_{\text{chain}(d)}]_{s,t} = 1$ for any $1 \leq s \neq t \leq d$ satisfying $|s - t| = 1$ and $[A_{\text{chain}(d)}]_{s,t} = 0$ otherwise. Recall that $I_{\text{Conn}}$ denotes the negative number of connected subgraphs, and hence the parameter $k$ takes negative values. We construct our graph with $-k$ connected subgraphs by removing $-k - 1$ edges from the $d$-chain. Given any $k \in [-d, -1]$, let the adjacency matrix for the edges to cut be $A_{\text{cut}(k)}$ such that $[A_{\text{cut}(k)}]_{s,t} = 1$ if and only if $|s - t| = 1$ and $s = j[-d/k]$ for $j = 1, 2, \ldots, -k - 1$ and otherwise $[A_{\text{cut}(k)}]_{s,t} = 0$. We construct the precision matrix as $\Theta_{\mu,k} = I_d + \mu(A_{\text{chain}(d)} - A_{\text{cut}(k)})$. We construct the precision matrix as $\Theta_{\mu,k} = I_d + \mu(A_{\text{chain}(d)} - A_{\text{cut}(k)})$. Given $k$ and $k_\mu$ to be the values of $I_{\text{Conn}}(\Theta)$ and $I_{\text{Conn}}(E_{\text{Sig}}(\Theta))$ respectively, for each
repetition, we uniformly sample $k - k_{\mu}$ edges from $E(\Theta_{\mu,k})$ and change the values of entries on $\Theta_{\mu,k}$ corresponding to these edges from $\mu$ to $\theta$. Denote this new precision matrix as $\tilde{\Theta}$ and we can see that $\mathcal{I}_{\text{Conn}}(\tilde{\Theta}) = k$ and $\mathcal{I}_{\text{Conn}}(E_{\text{Sig}}(\tilde{\Theta})) = k_{\mu}$. We generate $X_1, \ldots, X_n$ i.i.d. from $N(0, \tilde{\Theta}^{-1})$ and construct $[\tilde{I}_L, I^*_U]$ from Algorithm 3. For $\mathcal{I}_{\text{Conn}}$, we consider $k = -25$, $k_{\mu} = -25, -26, -27$ and $-28$ and set $[I^*_L, I^*_U] = [-d, -1]$.

- $\mathcal{I}_{\text{Deg}}$: Let the adjacency matrix of a $k$-star graph be $A_{\text{star}(k)} \in \mathbb{R}^{(k+1) \times (k+1)}$ such that $[A_{\text{star}(k)}]_{tt} = 1$ for $1 \leq t \leq k + 1$ and $[A_{\text{star}(k)}]_{st} = 0$ if $s \neq 1$. We construct the precision matrix with signal strength $\mu$ for the graph assembling $\lfloor d/(k + 1) \rfloor$ number of $k$-stars as $\Theta_{\mu,k} = I_{d} + \mu \cdot \text{diag}(A_{\text{star}(k)}, \ldots, A_{\text{star}(k), I_{d-k[d/(k+1)]}})$. Given $k$ and $k_{\mu}$, for each repetition, denote $\widetilde{A} = \text{diag}(A_{\text{star}(k-k_{\mu}), I_{k-k_{\mu}}})$ and we construct the precision matrix as

\[
\tilde{\Theta} = I_{d} + \mu \text{diag}(A_{\text{star}(k), \ldots, A_{\text{star}(k), I_{d-k[d/(k+1)]}}}) + (\theta - \mu) \text{diag}(\widetilde{A}, \ldots, \widetilde{A}, I_{d-k[d/(k+1)]}).
\]

We can see that $\mathcal{I}_{\text{Deg}}(\tilde{\Theta}) = k$ and $\mathcal{I}_{\text{Deg}}(E_{\text{Sig}}(\tilde{\Theta})) = k_{\mu}$. We generate $X_1, \ldots, X_n$ i.i.d. from $N(0, \tilde{\Theta}^{-1})$ and construct $[\tilde{I}_L, I^*_U]$ from Algorithm 1 for $\mathcal{I} = \mathcal{I}_{\text{Deg}}$. For $\mathcal{I}_{\text{Deg}}$, we consider $k = 5$, $k_{\mu} = 5, 4, 3$ and $2$ and set $[I^*_L, I^*_U] = [0, 20]$.

- $\mathcal{I}_{\text{Iso}}$: Since $\mathcal{I}_{\text{Iso}}$ is the negative number of isolated nodes, the parameters $k$ and $k_{\mu}$ are negative. If $d + k$ is even, we construct the precision matrix with $-k$ isolated nodes as $\Theta_{\mu,k} = \mu \text{diag}(I_{-k}, A_{\text{chain}(1), \ldots, A_{\text{chain}(1)}})$. The precision matrix represents the graph containing $-k$ isolated nodes and $(d + k)/2$ disconnected single edges. If $d + k$ is even we construct the precision matrix as $\Theta_{\mu,k} = I_{d} + \mu \text{diag}(I_{-k}, A_{\text{chain}(1), \ldots, A_{\text{chain}(1), A_{\text{chain}(2)}}})$. Since $d + k$ is odd, we let the last chain in the graph contain 2 edges. Given $k$ and $k_{\mu}$ such that $k - k_{\mu}$ is even, for each repetition, we uniformly sample $(k - k_{\mu})/2$ edges from the single edge chain in $E(\Theta_{\mu,k})$ and change the weights on these edges from $\mu$ to $\theta$. Denote this new
Table 2.1: The estimated coverage probability (the column Prob.) and averaged confidence interval length (the column Length) for $I_{\text{Conn}}$. We set the dimension $d = 100$, the sample size $n \in \{400, 600\}$, the values of the invariant $k = -25, k_\mu \in \{-25, -26, -27, -28\}$ and the signal strength $\mu \in \{0.2, 0.4, 0.6, 0.8\}$. The results are calculated based on 500 repetitions.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k_\mu$</th>
<th>$\mu = 0.2$</th>
<th>Length</th>
<th>$\mu = 0.4$</th>
<th>Length</th>
<th>$\mu = 0.6$</th>
<th>Length</th>
<th>$\mu = 0.8$</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>-25</td>
<td>0.978</td>
<td>43.81</td>
<td>0.848</td>
<td>23.86</td>
<td>0.802</td>
<td>23.26</td>
<td>0.920</td>
<td>23.39</td>
</tr>
<tr>
<td></td>
<td>-26</td>
<td>0.984</td>
<td>44.63</td>
<td>0.962</td>
<td>24.97</td>
<td>0.970</td>
<td>24.39</td>
<td>0.976</td>
<td>24.47</td>
</tr>
<tr>
<td></td>
<td>-27</td>
<td>0.970</td>
<td>45.13</td>
<td>0.970</td>
<td>25.65</td>
<td>0.968</td>
<td>25.01</td>
<td>0.970</td>
<td>25.08</td>
</tr>
<tr>
<td></td>
<td>-28</td>
<td>0.976</td>
<td>46.25</td>
<td>0.976</td>
<td>26.74</td>
<td>0.976</td>
<td>26.09</td>
<td>0.976</td>
<td>26.25</td>
</tr>
<tr>
<td>600</td>
<td>-25</td>
<td>0.976</td>
<td>35.91</td>
<td>0.756</td>
<td>23.16</td>
<td>0.826</td>
<td>23.25</td>
<td>0.896</td>
<td>23.33</td>
</tr>
<tr>
<td></td>
<td>-26</td>
<td>0.974</td>
<td>36.72</td>
<td>0.940</td>
<td>24.08</td>
<td>0.972</td>
<td>24.15</td>
<td>0.966</td>
<td>24.25</td>
</tr>
<tr>
<td></td>
<td>-27</td>
<td>0.976</td>
<td>37.69</td>
<td>0.970</td>
<td>25.13</td>
<td>0.976</td>
<td>25.20</td>
<td>0.976</td>
<td>25.25</td>
</tr>
<tr>
<td></td>
<td>-28</td>
<td>0.986</td>
<td>38.88</td>
<td>0.984</td>
<td>26.93</td>
<td>0.986</td>
<td>26.41</td>
<td>0.986</td>
<td>26.50</td>
</tr>
</tbody>
</table>

We generate $X_1, \ldots, X_n$ i.i.d. from $N(0, \tilde{\Theta}^{-1})$ and construct $[\tilde{I}_L, \tilde{I}_U]$ from Algorithm 4. For $I_{\text{iso}}$, we consider $k = -3, k_\mu = -3, -5, -7$ and $-9$ and set $[I^*_L, I^*_U] = [-d, 0]$.

Given the data $X_1, \ldots, X_n$, we estimate the precision matrix by the CLIME estimator

$$\tilde{\Theta}_j = \arg\min_{\beta \in \mathbb{R}^d} \|\beta\|_1 \text{ s.t. } \|\tilde{\Sigma}\beta - e_j\|_\infty \leq \lambda$$ (2.5.1)

where $e_j$ is the $j$-th canonical basis in $\mathbb{R}^d$ for any $j = 1, \ldots, d$. The tuning parameter $\lambda$ in (4.1.1) is chosen through minimizing a 3-fold cross validation

$$\text{CV}(\lambda) = \sum_{k=1}^{3} \|\tilde{\Sigma}^{(k)}\tilde{\Theta}_\lambda^{(-k)} - I_d\|_F^2,$$ (2.5.2)

where $\tilde{\Sigma}^{(k)}$ is the sample covariance matrix only using the $k$-th fold of the dataset and $\tilde{\Theta}_\lambda^{(-k)}$ is the CLIME estimator using the remaining data. In the simulations for all three invariants, we set the dimension $d = 100$ and sample size $n = 400$ and 600. We set $\mu = 0.2, 0.4, 0.6$ and 0.8. We choose the significance level of confidence intervals as 5%.
Table 2.2: The estimated coverage probability (the column Prob.) and averaged confidence interval length (the column Length) for $I_{\text{Deg}}$. We set the dimension $d = 100$, the sample size $n \in \{400, 600\}$, the values of the invariant $k = 5$, $k_{\mu} \in \{5, 4, 3, 2\}$ and the signal strength $\mu \in \{0.2, 0.4, 0.6, 0.8\}$. The results are calculated based on 500 repetitions.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k_{\mu}$</th>
<th>$\mu = 0.2$</th>
<th></th>
<th>$\mu = 0.4$</th>
<th></th>
<th>$\mu = 0.6$</th>
<th></th>
<th>$\mu = 0.8$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>5</td>
<td>0.962</td>
<td>15.68</td>
<td>0.964</td>
<td>15.38</td>
<td>0.962</td>
<td>15.39</td>
<td>0.962</td>
<td>15.39</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.962</td>
<td>15.68</td>
<td>0.964</td>
<td>15.38</td>
<td>0.962</td>
<td>15.39</td>
<td>0.962</td>
<td>15.39</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.962</td>
<td>16.44</td>
<td>0.962</td>
<td>16.31</td>
<td>0.962</td>
<td>16.34</td>
<td>0.962</td>
<td>16.35</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.962</td>
<td>17.31</td>
<td>0.962</td>
<td>17.28</td>
<td>0.962</td>
<td>17.29</td>
<td>0.962</td>
<td>17.31</td>
</tr>
<tr>
<td>600</td>
<td>5</td>
<td>0.978</td>
<td>15.63</td>
<td>0.978</td>
<td>15.63</td>
<td>0.978</td>
<td>15.65</td>
<td>0.978</td>
<td>15.65</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.980</td>
<td>15.63</td>
<td>0.978</td>
<td>15.64</td>
<td>0.978</td>
<td>15.65</td>
<td>0.978</td>
<td>15.65</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.978</td>
<td>16.60</td>
<td>0.978</td>
<td>16.59</td>
<td>0.978</td>
<td>16.62</td>
<td>0.978</td>
<td>16.62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.978</td>
<td>17.57</td>
<td>0.978</td>
<td>17.55</td>
<td>0.978</td>
<td>17.56</td>
<td>0.978</td>
<td>17.59</td>
</tr>
</tbody>
</table>

The estimated coverage probability and the averaged confidence interval length calculated through 500 repetitions are reported in Tables 2.1, 2.2 and 2.3 for $I_{\text{Conn}}$, $I_{\text{Deg}}$ and $I_{\text{Iso}}$ respectively. From these results, we can see that when the value $\mu$ is relatively small, the confidence interval lengths are larger in order to guarantee the confidence interval cover the true invariant under small signal strength. When $\mu$ becomes larger, the confidence interval lengths converge to the optimal rate $O(I_{U}^{c} - I(F_{\text{Sig}}(\Theta)))$ shown in (2.4.10). This illustrates that the proposed confidence interval is adaptive to $k_{\mu}$.

Table 2.3: The estimated coverage probability (the column Prob.) and averaged confidence interval length (the column Length) for $I_{\text{Iso}}$. We set the dimension $d = 100$, the sample size $n \in \{400, 600\}$, the values of the invariant $k = -3$, $k_{\mu} \in \{-3, -5, -7, -9\}$ and the signal strength $\mu \in \{0.2, 0.4, 0.6, 0.8\}$. The results are calculated based on 500 repetitions.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k_{\mu}$</th>
<th>$\mu = 0.2$</th>
<th></th>
<th>$\mu = 0.4$</th>
<th></th>
<th>$\mu = 0.6$</th>
<th></th>
<th>$\mu = 0.8$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>-3</td>
<td>0.962</td>
<td>32.97</td>
<td>0.964</td>
<td>4.32</td>
<td>0.958</td>
<td>2.84</td>
<td>0.938</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td>-5</td>
<td>0.972</td>
<td>33.49</td>
<td>0.972</td>
<td>6.25</td>
<td>0.972</td>
<td>4.78</td>
<td>0.972</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td>-7</td>
<td>0.970</td>
<td>51.42</td>
<td>0.970</td>
<td>8.28</td>
<td>0.970</td>
<td>6.70</td>
<td>0.970</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td>-9</td>
<td>0.964</td>
<td>36.63</td>
<td>0.964</td>
<td>10.15</td>
<td>0.964</td>
<td>8.58</td>
<td>0.964</td>
<td>8.51</td>
</tr>
<tr>
<td>600</td>
<td>-3</td>
<td>0.978</td>
<td>21.16</td>
<td>0.976</td>
<td>2.93</td>
<td>0.972</td>
<td>2.91</td>
<td>0.962</td>
<td>2.89</td>
</tr>
<tr>
<td></td>
<td>-5</td>
<td>0.974</td>
<td>22.63</td>
<td>0.974</td>
<td>4.87</td>
<td>0.974</td>
<td>4.83</td>
<td>0.974</td>
<td>4.81</td>
</tr>
<tr>
<td></td>
<td>-7</td>
<td>0.968</td>
<td>24.15</td>
<td>0.968</td>
<td>6.74</td>
<td>0.968</td>
<td>6.71</td>
<td>0.968</td>
<td>6.69</td>
</tr>
<tr>
<td></td>
<td>-9</td>
<td>0.974</td>
<td>26.19</td>
<td>0.974</td>
<td>8.70</td>
<td>0.974</td>
<td>8.68</td>
<td>0.974</td>
<td>8.66</td>
</tr>
</tbody>
</table>
Figure 2.4: The connected subgraphs in brain networks changing with the filtration level $\mu$. The upper panel illustrates the connected regions of interest by same colors. The lower panel shows how the number of connected subgraphs change with the filtration level $\mu$ for both intact story (in red) and word scrambled (in blue).

### 2.5.2 Neuroscience Application

We apply our inferential method to the brain imaging dataset studied in Simony et al. (2016). The dataset contains fMRI scans from 36 subjects taken while the subjects were listening to the stimuli generated from a seven-minute story *Pieman* (told by Jim O’Grady at the “The Moth” storytelling event). The 36 subjects also listened to a word-scrambled version of the story. In particular, the story was segmented into 608 short words and their order was scrambled randomly. The raw functional data was preprocessed to correct head motion, time slicing, spatial smoothing and temporal filtering in Simony et al. (2016). For both the intact story and the word scrambled settings, each subject had 300 fMRI measurements and the measurements were taken every 1.4 seconds.
The original fMRI dataset has the dimensional 271,633 representing 271,633 3-mm isotropic voxels. We reduce the dimension to 172 regions of interest (ROIs) introduced by Baldassano et al. (2015) through averaging the voxels in the same ROI. Therefore, for each subject, we have the data with dimension \( d = 172 \) and sample size \( n = 300 \). We average the data across the 36 subject to obtain a single \( 300 \times 172 \) dataset and standardize each ROI such that they have mean zero and standard deviation one. We apply the Gaussian graphical model to the dataset so that the brain network is induced by the precision matrix \( \Theta \) and each ROI corresponds to a node in the network. Our goal is to infer two combinatorial quantities of the brain network: the number of connected subgraphs and the maximum degree. In fact, to grasp more detailed structural information of the network, we aim to infer the above two invariants for the precision matrix at different filtration levels. In specific, given a precision matrix \( \Theta \) and a filtration level \( \mu > 0 \), we define the thresholded matrix \( [T_\mu(\Theta)]_{jk} = \Theta_{jk} \mathbb{I}\{|\Theta_{jk}| \geq \mu\} \) for all \( 1 \leq j, k \leq d \). We want to construct confidence intervals for invariants \( I_{\text{Conn}}(T_\mu(\Theta)) \) and \( I_{\text{Deg}}(T_\mu(\Theta)) \) for different levels of \( \mu \) under both the intact story and word scrambled settings. It is easy to check that \( I_{\text{Conn}}(T_\mu(\Theta)) \) and \( I_{\text{Deg}}(T_\mu(\Theta)) \) are also monotone invariants, therefore the skip-down algorithm can be applied.

As before, the tuning parameter of the CLIME estimator is chosen via 3-fold cross validation using the risk in (2.5.2). The significance level of the confidence interval is set to \( \alpha = 5\% \) and since there is no prior information on how large the maximum degree of the network is, or how many connected subgraphs the network contains, we consider the largest possible range for both invariants.

We implement Algorithm 1 to construct the confidence intervals. Figures 2.4 and 2.5 visualize how the lower endpoints of the confidence intervals, of the number of connected subgraphs and the maximum degree respectively, change with the filtration level \( \mu \). In addition, we visualize the structural information of the output edge set \( E_t^* \) generated from the skip-down algorithm, where \( t^* \) is the number of iterations needed for Algorithm 1 to
Figure 2.5: The degree of region of interest in brain networks changing with filtration level $\mu$. The upper panel illustrates the degree of each region of interest. The lower panel shows how the maximum degree changes with the filtration level $\mu$ for both intact story (in red) and word scrambled (in blue).

Figure 2.6: Structural differences in cerebral cortices between the intact story and word scrambled settings. (a) shows the connected regions of interest in brain networks under the filtration level $\mu = 0.4$. (b) shows the degree of regions of interest in brain networks under the filtration level $\mu = 0$. 

42
conclude. Figures 2.4 illustrates the different connected ROIs in $G_{t^*} = (V, E_{t^*})$ by different colors and Figure 2.5 illustrates the degree of each ROI in $G_{t^*}$. For the number of connected subgraphs, Figure 2.4 shows that the brain network has fewer connected subgraphs when the subject is listening to the intact story compared to when the subject is listening to the scrambled word version. In particular, we can see in Figure 2.6(a) that compared to the word scrambled setting, the dorsolateral prefrontal cortex (DL-PFC) is connected to the inferior frontal gyrus (IFG) under the intact story setting. The inferior frontal gyrus is the brain area responsible for language processing and comprehension (Grewe et al., 2005; Caplan, 2006) and the dorsolateral prefrontal cortex is responsible for working memory tasks (Barbey et al., 2013). The fact that these two areas are connected in the intact story setting suggests that both language processing and memory are working together in the procedure of understanding the intact Pieman story.

Regarding maximum degree, Figure 2.5 shows that the brain network has higher maximum degree when the subject is listening to the intact story compared to when the subject is listening to scrambled words. In Figure 2.6(b) we also observe that the precuneus area, which is known to be involved with understanding high-level concepts in stories (Lerner et al., 2011; Ames et al., 2015), has a higher degree under the intact story setting compared to the word scrambled setting.
Chapter 3

Theoretical Lower Bound

In this chapter, we provide a unified theoretical framework for the minimaxity of hypothesis tests and confidence interval length.

3.1 Single-Edge Null-Alternative Dividers

In this section we derive a novel and generic lower bound strategy, applicable to null and alternative hypotheses which differ in one single edge: i.e., under the Gaussian model, there exist two matrices $\Theta_0 \in S_0$ and $\Theta_1 \in S_1$ whose induced graphs $G_0 := G(\Theta_0)$ and $G_1 := G(\Theta_1)$ differ in a single edge. We aim to characterize necessary conditions on the pair $S_0, S_1$ under which the combinatorial inference problem

\[ H_0 : \Theta^* \in S_0 \text{ vs } H_1 : \Theta^* \in S_1. \]  

(3.1.1)

is testable. Specifically, recall that a test is any measurable function $\psi : \{X_i\}_{i=1}^n \mapsto \{0, 1\}$. Define the minimax risk of testing $S_0$ against $S_1$ as:

\[ \gamma(S_0, S_1) = \inf_{\psi} \left[ \max_{\Theta \in S_0} \mathbb{P}_\Theta(\psi = 1) + \max_{\Theta \in S_1} \mathbb{P}_\Theta(\psi = 0) \right]. \]  

(3.1.2)
We formalize this concept of single-edge divider in the definition below.

**Definition 3.1.1 (Single-Edge Null-Alternative Divider).** For a sub-decomposition \((\mathcal{G}_0, \mathcal{G}_1)\) of \(\mathcal{G}\), let \(G_0 = (\overline{V}, E_0) \in \mathcal{G}_0\) be a graph under the null. We refer to an edge set \(C = \{e_1, \ldots, e_m\}\), as a (single-edge) null-alternative divider with the null base \(G_0\) if for any \(e \in C\) the graphs \(G_e := (\overline{V}, E_0 \cup \{e\}) \in \mathcal{G}_1\).

If a large divider exists, it is expected that differentiating \(G_0\) from an alternative graph \(G_e \in \mathcal{G}_1\) is more challenging. Indeed, our main result of this section confirms this intuition. We proceed to define a predistance for a graph \(G\) and two edges \(e, e'\) (which need not belong to \(E(G)\)) which plays a key role in our lower bound result.

**Definition 3.1.2 (Edge Geodesic Predistance).** Let \(G = (\overline{V}, E)\) and \(\{e, e'\}\) be a pair of edges (\(e\) and \(e'\) may or may not belong to \(E\)). We define

\[
d_G(e, e') := \min_{u \in e, v \in e'} d_G(u, v),
\]

where \(d_G(u, v)\) denotes the geodesic distance between vertices \(u\) and \(v\) on the augmented graph \(G\). If such a path does not exist \(d_G(e, e') = \infty\).

By definition \(d_G(e, e')\) is a predistance, i.e., \(d_G(e, e) = 0\) and \(d_G(e, e') \geq 0\). Moreover, \(d_G(e, e')\) has the same value regardless of whether \(e, e' \in E(G)\). See Figure 3.1 for an illustration of \(d_G(e, e')\). Inspired by the classical concept of packing entropy on metric spaces (e.g., Yang and Barron, 1999) we propose the structural packing entropy for graphs in an attempt to characterize information-theoretic lower bounds for combinatorial inference.

**Definition 3.1.3 (Structural Packing Entropy).** Let \(\mathcal{C}\) be a non-empty edge set and \(G\) be a graph. For any \(r \geq 0\) we call the edge set \(N_r \subset \mathcal{C}\) an \(r\)-packing of \(\mathcal{C}\) if for any \(e, e' \in N_r\) we have \(d_G(e, e') \geq r\). Define the structural \(r\)-packing entropy as:

\[
M(\mathcal{C}, d_G, r) := \log \max \{|N_r| \mid N_r \subset \mathcal{C}, N_r \text{ is a } r\text{-packing of } \mathcal{C}\}. \quad (3.1.3)
\]
The packing entropy in Definition 3.1.3 is an analog to the classical packing entropy on metric spaces in the sense that it is defined over an edge set $\mathcal{C}$ equipped with a predistance $d_G(e, e')$ based on the graph $G$.

To study minimax lower bounds, we only need to focus on the Gaussian graphical model whose structural properties are completely characterized by the precision matrices. We now formally define the sets of precision matrices $\mathcal{S}_0$ and $\mathcal{S}_1$ used in this section. Let:

$$
\mathcal{S}_0(\theta, s) := \left\{ \Theta \in \mathcal{U}_s \mid G(\Theta) \in \mathcal{G}_0, \min_{e \in E(G(\Theta))} |\Theta_e| \geq \theta \right\} \text{ and,}
$$

$$
\mathcal{S}_1(\theta, s) := \left\{ \Theta \in \mathcal{U}_s \mid G(\Theta) \in \mathcal{G}_1, \min_{e \in E(G(\Theta))} |\Theta_e| \geq \theta \right\},
$$

where $\mathcal{U}_s$ is defined in (4.3.2). The parameter $\theta$ in the definitions of $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$ denotes the signal strength, and as we show below, its magnitude plays an important role in determining whether one can distinguish between graphical models in $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$.

**Theorem 3.1.1** (Necessary Signal Strength). Let $D$ be a fixed integer. Suppose that

$$
\theta \leq \max_{G_0 \in \mathcal{G}_0 : C \text{ divider with null base } G_0} C_{\text{max}}(G_0) \leq D, \quad \mathcal{C} \in \mathcal{G}_0 \quad \kappa \sqrt{M(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|)} \frac{n}{\sqrt{2(D+2)}} \quad \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)),
$$

where $C$ is defined in (4.3.2). Then if $M(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|) \to \infty$ as $n \to \infty$, there exists a sufficiently small constant $\kappa$ in (3.1.6) (depending on $D, C, L$) such that

$$
\liminf_{n \to \infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1.
$$

Theorem 3.1.1 allows us to quantify the signal strength necessary for combinatorial inference via combinatorial constructions. The radius $\log |\mathcal{C}|$ of the packing entropy in (3.1.6) ensures that the pairs of distinct edges are sufficiently far apart. The constant term $(1 - C^{-1}) \wedge e^{-\frac{1}{2}} \sqrt{2(D+2)}$ in (3.1.6) ensures that precision matrices with signal strength $\theta$ indeed belong to $\mathcal{U}_s$.

**Proof Sketch.** The proof of Theorem 3.1.1 can roughly be divided into four steps. Full details of the proof will be provided in Section B.1 in appendix.
Step 1 (Connect the structural parameters to geometric parameters). Given the adjacency matrices of the null and alternative graphs $G_0$ and $\{G_e\}_{e \in \mathcal{E}}$, we construct the corresponding precision matrices and make sure that they belong to $\mathcal{S}_0(\theta,s)$ and $\mathcal{S}_1(\theta,s)$.

Step 2 (Construct minimax risk lower bound via Le Cam’s method). The second step uses Le Cam’s method to lower bound $\gamma(\mathcal{S}_0(\theta,s), \mathcal{S}_1(\theta,s))$. This requires us to evaluate the chi-square divergence between a normal and a mixture normal distribution. The chi-square divergence can be expressed via ratios of determinants. In particular, we show that the log chi-square divergence can be equivalently re-expressed via an infinite sum of differences among trace operators of adjacency matrix powers.

Step 3 (Represent the lower bound by the number of shortest closed walks in the graph). In this step we control the deviations of the differences of the trace operators. Since the trace of the power of an adjacency matrix equals the number of closed walks within the corresponding graph, we eliminate the trace powers which are smaller than the shortest closed walks. The traces of the higher powers are handled via matrix perturbation bounds.

Step 4 (Characterize the smallest magnitude of the geometric parameter using the packing entropy). Lastly, we show that condition (3.1.6) ensures that the closed walks on the packing of the divider are sufficiently lengthy, which implies that the chi-square divergence vanishes when the signal strength $\theta$ is small.

A typical application of Theorem 3.1.1 proceeds by constructing a graph $G_0$ under the null hypothesis, which is one edge apart from the alternative. Next, one builds a divider $\mathcal{C}$ with as large as possible packing number, so that adding any edge from $\mathcal{C}$ to $G_0$ results in an alternative graph. Clearly choosing the graph $G_0$ is crucial for this strategy to work. Below we give several examples of explicit constructions of $G_0$ and divider. At the end of the section we also provide somewhat general guidance how to select $G_0$. 

47
3.1.1 Some Applications

In this section we give several examples of combinatorial testing, which readily fall into the framework developed in Section 3.1. Although some of the examples We show one more additional example on self-avoiding paths in Section B.2 of the appendix.

Example 3.1.1 (Connectivity Testing). Consider the sub-decomposition $G_0 = \{ G \in \mathcal{G} | G$ disconnected $\} \text{ vs } G_1 = \{ G \in \mathcal{G} | G$ connected $\}$. We construct a base graph $G_0 := (\mathcal{V}, E_0)$ where

$$E_0 := \{(j, j + 1)_{j=1}^{\lfloor d/2 \rfloor -1}, ([d/2], 1), (j, j + 1)_{j=\lfloor d/2 \rfloor +1}^{d}, ([d/2] + 1, d)\},$$

and let $\mathcal{C} := \{(j, \lfloor d/2 \rfloor + j)_{j=1}^{\lfloor d/2 \rfloor} \}$ (see Figure 3.1). Clearly adding any edge from $\mathcal{C}$ to $G_0$ connects the graph, so $\mathcal{C}$ is a single edge divider with a null base $G_0$. Furthermore, the maximum degree of $G_0$ equals 2 by construction. To construct a packing set of $\mathcal{C}$, we collect all edges $(j, \lfloor d/2 \rfloor + j)$ satisfying $\lceil \log |\mathcal{C}| \rceil$ divides $j$ except if $j > \lfloor d/2 \rfloor - \lceil \log |\mathcal{C}| \rceil$. This procedure results in a packing set with radius at least $\lceil \log |\mathcal{C}| \rceil$ which has cardinality of at least $\left\lfloor \frac{|\mathcal{C}|}{\log |\mathcal{C}|} \right\rfloor - 1$. Therefore

$$M(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|) \geq \log \left[ \left\lfloor \frac{|\mathcal{C}|}{\log |\mathcal{C}|} \right\rfloor - 1 \right] \asymp \log |\mathcal{C}| \asymp \log d.$$

Theorem 3.1.1 implies that the asymptotic minimax risk is 1 if $\theta < \kappa \sqrt{\log d/n \wedge (1-\mathcal{C}^{-1})^\frac{1}{\Theta^2}}$. 

Figure 3.1: The graph $G_0$ with two edges $e, e' \in \mathcal{C} : d_{G_0}(e, e') = 2, d = 10.$
Example 3.1.2 ($m + 1$ vs $m$ Connected Components, $m \geq \sqrt{d}$). Let $m \geq \sqrt{d}$ be an integer. In this example we are interested in testing whether the graph contains $m + 1$ connected components vs $m$ connected components. The reason to assume $m \geq \sqrt{d}$ is to make sure there are sufficiently many edges for constructing a single edge divider in order to obtain sharp bounds. The case when $m < \sqrt{d}$ is treated in Example 3.1.6 via a different divider construction (In fact, the case $m < \sqrt{d}$ requires deleting edges from the alternative rather than adding edges to the null base. See Section B.2 of the appendix for more details). Formally we have the sub-decomposition $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ has } \geq m + 1 \text{ connected components}\}$ vs $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ has } \leq m \text{ connected components}\}$. Construct the null base graph $G_0 = (V, E_0)$, where $E_0 := \{(j, j + 1)^{d-m-1}_{j=1}\}$, and we let $C := \{(j, j + 1)^{d-1}_{j=d-m}\}$ (see Figure 3.2). Adding an edge $e \in C$ to $G_0$ converts the base graph $G_0$ into a graph with $m$ connected components and therefore $C$ is a single edge divider with a null base $G_0$. Additionally, the maximum degree of $G_0$ is 2 by construction. Note that the distance between any two edges in $C$ is 0 if and only if they share a common vertex, and $\infty$ in all other cases. This implies that we can construct a packing set by taking every other edge in the set $C$. We conclude that $M(C, d_{G_0}, \log |C|) \asymp \log(|C|/2) \asymp \log d$. Hence, by Theorem 3.1.1 the minimax risk goes to 1 when $\theta < \kappa \sqrt{\log d/n} \wedge \frac{(1-C^{-1})e^{-\frac{1}{4}}}{4\sqrt{2}}$.

Example 3.1.3 (Cycle Testing). Consider testing whether the graph is a forest vs the graph contains a cycle. Let $\mathcal{G}_0 = \{G \in \mathcal{G} \mid G \text{ is cycle-free}\}$ and $\mathcal{G}_1 = \{G \in \mathcal{G} \mid G \text{ contains a cycle}\}$. Define the null base graph $G_0 = (V, E_0)$, where $E_0 := \{(j, j + 1)^{d-1}_{j=1}\}$. Let the edge set $C := \{(j, j + 2)^d_{j=1}\}$, where the addition is taken modulo $d$ (refer to Figure 3.3 for a visualization). By construction we have $G_0 \in \mathcal{G}_0$ and $|C| = d$. Adding any edge from $C$ to $G_0$ results in a graph with a cycle, and hence the edge set $C$ is a single edge divider with a null base $G_0$. The maximum degree of $G_0$ equals 2, and is thus bounded. Moreover, there exists a
(log |C|)-packing set of C of cardinality at least \( \frac{|C|-2}{\log |C| + 2} \) which can be produced by collecting the edges \((j, j + 2)\) for \( j = k(\log |C| + 2) + 1 \) for \( k = 0, 1, \ldots \) and \( j \leq d - 2 \). The last observation implies that \( M(C, d, \log |C|) \asymp \log \frac{|C|-2}{\log |C| + 2} \asymp \log d \). Hence by Theorem 3.1.1 we conclude that the minimax risk goes to 1 when \( \theta < \kappa \sqrt{\log d / n} \wedge (1 - C^{-1})/e^{-1/2} \).

Figure 3.3: The graph \( G_0 \) with two (dashed) edges \( e, e' \in C \) such that \( d_{G_0}(e, e') = 2, d = 7 \).

**Example 3.1.4** (Tree vs Connected Graph with Cycles). The construction in Example 3.1.3 also shows that we have the same signal strength limitation to test for cycles, even if we restrict to the subclass of connected graphs, i.e., the class of graphs under the null hypothesis is the class of all trees \( \mathcal{G}_0 = \{ G \in \mathcal{G} \mid G \text{ is a tree} \} \), and the alternative is the class of all connected graphs which contain a cycle — \( \mathcal{G}_1 \in \{ G \in \mathcal{G} \mid G \text{ is connected but is not a tree} \} \).

**Example 3.1.5** (Triangle-Free Graph). Consider testing whether the graph contains a triangle (i.e., 3-clique). More formally let the decomposition of \( \mathcal{G} \) be \( \mathcal{G}_0 = \{ G \in \mathcal{G} \mid G \text{ is triangle-free} \} \) and \( \mathcal{G}_1 = \{ G \in \mathcal{G} \mid \exists 3\text{-clique in } G \} \). It is clear that in this case we can reuse the set \( C \) and its null base \( \tilde{G}_0 = (V, E_0 \cup \{(1, d)\}) \), where \( E_0 \) and \( C \) are taken as in Example 3.1.3.

Though Theorem 3.1.1 delivers valid lower bounds, the obtained bounds may not always be tight. For example, recall the construction of Example B.2.1. When \( m = d - 2 \) the divider contains only a single edge, and therefore (3.1.6) is not sharp. Sometimes tighter bounds can be obtained via constructions which delete edges from the alternative graph instead of adding edges to the graph under null. In this subsection we extend the divider definition to allow for deleting edges from the alternative hypothesis, which is sometimes more convenient and gives more informative lower bounds as we demonstrate below.
Definition 3.1.4 (Single-Edge Deletion Null-Alternative Divider). Let $G_1 = (V, E_1)$ be an alternative graph. An edge set $C$ is called a (single edge) deletion divider with alternative base $G_1$, if for all $e \in C$ the graphs $G \setminus e = (V, E_1 \setminus \{e\}) \in G_0$.

In parallel to Theorem 3.1.1 we have the following result using the deletion divider.

**Theorem 3.1.2.** Let $G_1 \in G_1$ be a graph whose maximum degree bounded by a fixed integer $D$, and let $C$ be a single-edge deletion divider with alternative base $G_1$. Suppose that $n \geq (\log |C|)/c_0$ for some $c_0 > 0$ and that:

$$\theta \leq \kappa \sqrt{\frac{M(C, d_{G_1}, \log |C|)}{n}} \wedge \frac{1 - C^{-1}}{\sqrt{2D}}. \tag{3.1.7}$$

Then, if $M(C, d_{G_1}, \log |C|) \to \infty$ as $n \to \infty$, for a sufficiently small $\kappa$ (depending on $D, C, L, c_0$) we have $\liminf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1$.

To illustrate the usefulness of Theorem 3.1.2 for edge deletion divider, we consider the examples below.

**Example 3.1.6 (m+1 vs m Connected Components, m < \sqrt{d}).** In this example we consider a complementary setting to Example 3.1.2, and test whether the graph contains $m+1$ connected components vs $m$ connected components for $m < \sqrt{d}$. Recall the decomposition $G_0 = \{G \in G \mid G$ has $\geq m + 1$ conn. components$\}$ vs $G_1 = \{G \in G \mid G$ has $\leq m$ conn. components$\}$. Construct an alternative base graph $G_1 = (\overline{V}, E_1)$: $E_1 := \{(j, j + 1)_{j=1}^{d-m}\}$, and set $C := E_1 = \{(j, j + 1)_{j=1}^{d-m}\}$ (see Fig 3.4 for a visualization). Since removing any edge from $C$ results in increasing the number of connected components by 1, the set $C$ is a single edge deletion divider with an alternative base graph $G_1$. Moreover, notice that the maximum degree of $G_1$ is 2.

The entropy $M(C, d_{G_1}, \log |C|)$ satisfies $M(C, d_{G_1}, \log |C|) \asymp \log \frac{|C|}{\log |C|} \asymp \log |C| \asymp \log d$. Hence by Theorem 3.1.2 we conclude that testing connectivity is impossible when $\theta < \kappa \sqrt{\log d/n} \wedge \frac{1 - C^{-1}}{2\sqrt{2}}$. 51
Example 3.1.7 (Self-Avoiding Path (SAP) of Length $m$ vs $m + 1$, $m \geq \sqrt{d}$). In parallel to Example B.2.1 we now consider testing that SAP $\leq m$ vs SAP of length $\geq m + 1$ when $m \geq \sqrt{d}$. Define $G_0 = \{ G \in \mathcal{G} \mid \forall$ SAP have length $\leq m \}$ vs $G_1 = \{ G \in \mathcal{G} \mid \exists$ SAP of length $m + 1 \}$. Construct the alternative graph $G_1 = (\mathcal{V}, E_1)$, where $E_1 := \{(j, j + 1)_{j=1}^{m+1}\}$. Define the set $\mathcal{C} := E_1 = \{(j, j + 1)_{j=1}^{m+1}\}$, and note that removing any edge from $\mathcal{C}$ results in a graph from the null. Hence $\mathcal{C}$ is a deletion divider with base $G_1$. Moreover, the maximum degree of $G_1$ is 2. Similarly to Example 3.1.6 it follows that $M(\mathcal{C}, d_{G_1}, \log |\mathcal{C}|) \asymp \log |\mathcal{C}| \asymp \log d$. An application of Theorem 3.1.2 implies that for values of $\theta < \kappa \sqrt{\log d/n} \wedge \frac{1-C^{-1}}{2\sqrt{2}}$ testing the length of SAP is impossible.

3.1.2 General remarks on choosing a null base $G_0$

The following result sheds some light on reasonable choices of $G_0$.

Proposition 3.1.1. Let the graph $G_0 = (\mathcal{V}, E_0) \in \mathcal{G}_0$ have bounded maximum degree. Suppose there exist constants $0 < c, \gamma \leq 1$ so that for each vertex $v \in \mathcal{V}$, one can find a set of vertices $W_v$ satisfying $|W_v| \geq cd^\gamma$ and for all $w \in W_v$, we have $(\mathcal{V}, E_0 \cup \{(v, w)\}) \in \mathcal{G}_1$. Then there exists a divider with null base $G_0$ satisfying $M(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|) \asymp \log d$.

Of note, for any edge set $\mathcal{C}$ one has $M(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|) \leq \log \binom{d}{2} \asymp \log d$, which implies that graphs $G_0 \in \mathcal{G}_0$ as in Proposition 3.1.1 give scalar optimal bounds. The existence of such graphs is dependent on the sub-decomposition $(\mathcal{G}_0, \mathcal{G}_1)$. Notably, all examples in Section 3.1.1 fall under the framework of Proposition 3.1.1. Its proof can be found in the appendix.
Remark 3. When $\Theta \in \mathcal{U}$, the results in Theorem 3.1.1 and Theorem 3.1.2 suggest that a signal strength of order $\sqrt{\log d/n}$ is necessary for controlling the minimax risk (3.1.2). In fact, Theorem 7 of Cai et al. (2011) shows that under such signal strength condition, support recovery of $\Theta$ is indeed achievable, which further implies that controlling the minimax risk (3.1.2) is possible. A naive procedure for matching the lower bound is to first perfectly recover the graph structure. Then construct a test based on examining whether the graph has the desired combinatorial structure. Though such an approach is theoretically feasible, it is not practical. First, such an approach is overly conservative and does not allow us to tightly control the type I error at a desired level. Second, such an approach crucially depends on having a suitable thresholding parameter to estimate the graph, which is in general not realistic.

3.2 Multi-Edge Dividers

The results of Section 3.1 (and Section B.2 of the appendix) have two major limitations. First, the null base $G_0$ is assumed to be of bounded degree. Second, our results cover only tests for which there exists a single-edge divider. In this section we relax both of these conditions. The following motivating example illustrates a relevant testing problem which does not fall into the framework of Section 3.1.

Maximum Degree Testing. Consider testing whether the maximum degree of the graph $d_{\text{max}}$ satisfies $d_{\text{max}} \leq s_0$ vs $d_{\text{max}} \geq s_1$, where $s_0 < s_1 \leq s$ are integers which are allowed to scale with $n$. In this case it is impossible to simultaneously construct a null base graph $G_0$ of bounded degree and a single-edge divider $C$.

To handle multiple edge dividers, we first extend Definitions 3.1.1 and 3.1.2 to allow for the above examples.
Definition 3.2.1 (Null-Alternative Divider). Let \( G_0 = (V, E_0) \in \mathcal{G}_0 \) be a fixed graph under the null with adjacency matrix \( A_0 \). We call a collection of edge sets \( C \) a (multi-edge) divider with null base \( G_0 \), if for all edge sets \( S \in C \) we have \( S \cap E_0 = \emptyset \) and \((V, E_0 \cup S) \in \mathcal{G}_1\). For any edge set \( S \in C \), we denote the adjacency matrix of the graph \((V, S)\) with \( A_S \).

Definition 3.2.2 (Edge Set Geodesic Predistance). For two edge sets \( S \) and \( S' \) and a given graph \( G \) let \( d_G(S, S') = \min_{e \in S, e' \in S'} d_G(e, e') \).

We provide two generic strategies for obtaining combinatorial inference lower bounds on the signal strength. The first strategy, described in Section 3.2.1, assumes that all \( S \in C \) satisfy \(|S| \leq U \) for some fixed constant \( U \). The second strategy, presented in Section 3.2.2, does not require bounded cardinality of the edge sets \( S \), but requires that the null bases and dividers have some special combinatorial properties.

### 3.2.1 Bounded Edge Sets

Below we consider an extension of Theorem 3.1.1 for multi-edge dividers, where the number of edges in each set \( S \in C \) satisfy \(|S| \leq U \) for some fixed integer \( U \in \mathbb{N} \). In contrast to Section 3.1, here the graph \( G_0 \) is allowed to have unbounded degree.

**Theorem 3.2.1.** Let \( G_0 \in \mathcal{G}_0 \) be a graph under the null, and let \( C \) be a multi-edge divider with null base \( G_0 \). Suppose that for some sufficiently small absolute constant \( \kappa > 0 \):

\[
\theta \leq \kappa \sqrt{\frac{M(C, d_{G_0}, \log |C|)}{nU}} \wedge \frac{\kappa}{U(\|A_0\|_2 + 2U)} \wedge \frac{1 - C^{-1}}{4(\|A_0\|_1 + 2U)}. \tag{3.2.1}
\]

If \( M(C, d_{G_0}, \log |C|) \to \infty \) we have \( \liminf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1 \).

Theorem 3.2.1 is an extension of Theorem 3.1.1. Specifically, Theorem 3.1.1 corresponds the setting where \( U = 1 \), and \( \|A_0\|_2 \leq \|A_0\|_1 \leq D \) (recall that \( D \) is an upper bound of the graph degree). Even though by assumption \( U = \max_{S \in C} |S| \) is bounded, we explicitly keep
the dependency on $U$ in (3.2.1) to reflect how the bound changes if $U$ is allowed to scale. The first term on the right hand side of (3.2.1) is the structural packing entropy, while the remaining two terms ensure the parameter $\theta$ is small enough to construct a valid packing set (More details are provided in the proof).

We illustrate the usefulness of Theorem 3.2.1 by an example similar to the ones in Section 3.1.1. Consider testing whether the maximum degree of the graph $G(\Theta^*)$ is at most $s_0$ vs it is at least $s_1$, where $s_0 < s_1 \leq s$ can increase with $n$ but the null-alternative gap $s_1 - s_0$ remains bounded. Therefore we cannot apply Theorem 3.1.1 but should use Theorem 3.2.1 instead.

Define the sub-decomposition $G_0 = \{G \mid d_{\max}(G) \leq s_0\}$ and $G_1 = \{G \mid d_{\max}(G) \geq s_1\}$ respectively.

**Example 3.2.1** (Maximum Degree Test with Bounded Null-Alternative Gap). Let $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$ be defined in (3.1.4) and (3.1.5). Assume that $s \log d/n = o(1), s \sqrt{\log d/n} = O(1)$ and $s = O(d^\gamma)$ for some $\gamma < 1$. Then if $\kappa$ is small enough and $\theta < \kappa \sqrt{\log d/n}$, we have $\liminf_{n \to \infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

**Proof.** In order to show the above result, we start by building a graph $G_0$ by constructing $\left\lfloor \frac{d}{s_1 + 1} \right\rfloor$ non-intersecting $s_0$-star graphs first (see Fig B.2). Define the star graph centers $C_j = (s_1 + 1)j + 1$ for $j = 0, \ldots, \left\lfloor \frac{d}{s_1 + 1} \right\rfloor - 1$. Next, define $G_0$ by

$$G_0 := \{ \mathcal{V}, \bigcup_{j=0}^{\left\lfloor \frac{d}{s_1 + 1} \right\rfloor - 1} s_0 \bigcup_{k=1}^{s_0} \{(C_j, C_j + k)\} \}.$$

We define the divider $C$ as follows:

$$C := \left\{ \bigcup_{k=s_0+1}^{s_1} \{(C_j, C_j + k)\} \mid j = 0, \ldots, \left\lfloor \frac{d}{s_1 + 1} \right\rfloor - 1 \right\},$$

where we simply connect each of the vertices $C_j$ to the remaining $s_1 - s_0$ vertices in the block.
Figure 3.6: Test for the maximum degree $G_0 := \{ G \mid d_{\text{max}}(G) \leq s_0 \}$ vs $G_1 = \{ G \mid d_{\text{max}}(G) \geq s_1 \}$ with $s_0 = 3$, $s_1 = 5$ and $d = 18$. The solid edges represent $G_0 \in G_0$ with maximum degree $s_0 = 3$. We construct the divider $C = \{ ((1,5), (1,6)), ((7,11), (7,12)), ((13,17), (13,18)) \}$.

Since the predistance between any two different edge sets $S, S' \in C$, is $d_{G_0}(S, S') = \infty$, the set $C$ itself is a $(\log |C|)$-packing set. The latter implies that $M(C, d_{G_0}, \log |C|) = \log |C| \asymp \log(d/(s_1 + 1)) \asymp \log d$. In addition, one can easily check that $\|A_0\|_2 = \sqrt{s_0}$ and $\|A_0\|_1 = s_0$. By Theorem 3.2.1 we have that under the required scaling $\liminf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1$, when $\theta < \kappa \sqrt{\log d/n}$ for a sufficiently small $\kappa$. 

### 3.2.2 Scaling Edge Sets

Theorem 3.2.1 requires the cardinalities of the edge sets in the divider $C$ to be bounded. In this section, we consider multi-edge dividers $C$ allowing the sizes of $S \in C$ to increase with $n$. For this case, the previous notion of packing entropy based on geodesic predistence is no longer effective. Instead, we introduce a new mechanism called buffer entropy to quantify the lower bound under scaling multi-edge dividers.

We first intuitively explain why the structural entropy in Theorem 3.2.1 may not be sufficient for handling dividers with scaling edge sets. Recall that Theorem 3.2.1 uses the structural entropy $M(C, d_{G_0}, \log |C|)$ to characterize the lower bound. In turn, the structural entropy is calculated based on the edge set geodesic predistance $d_{G_0}$ in Definition 3.2.2. One difference between fixed and scaling edge sets sizes is that, one can only pack a limited number of edge sets or large size which are sufficiently far apart (and hence do not overlap). A less wasteful strategy would be to allow for the edge sets to overlap. However, in general,
different edge sets $S, S' \in C$ may have multiple overlapping vertices and the notion of geodesic predistance is no longer precise enough to reflect the closeness between $S$ and $S'$.

Below we introduce a concept called vertex buffer, which helps to measure the closeness between edge sets $S$ and $S'$ more precisely than the geodesic predistance.

**Definition 3.2.3 (Vertex Buffer).** Let $G_0 = (\bar{V}, E_0)$ be a given graph and $S, S'$ be two edge sets. The vertex buffer of $S, S'$ under $G_0$ is defined as

$$
\mathcal{V}_{S,S'} := \{V(E_0 \cup S) \cap V(S')\} \cup \{V(E_0 \cup S') \cap V(S)\}.
$$

An important property of the set $\mathcal{V}_{S,S'}$ is that all paths passing through at least one edge in both $S$ and $S'$ must contain at least one vertex in $\mathcal{V}_{S,S'}$. In that sense, a large buffer size $|\mathcal{V}_{S,S'}|$ indicates that the edge sets $S$ and $S'$ are close to each other. We visualize an example of a vertex buffer in Figure 3.7.

![Figure 3.7: Visualization of the vertex buffer in $\mathcal{V}_{S,S'}$. Here $S, S'$ are plotted with dashed and dotted edges respectively and $G_0$ is in solid edges. The vertices in the buffer are marked in the dashed squares.](image)

In contrast to the bounded edge sets case, when the edge sets in $C$ are allowed to scale in size, it is not effective to build packing sets based on the predistance, since this strategy limits the number of edge sets we can build. One way to increase the cardinality of $C$ is to consider a larger number of potentially overlapping structures, and use the buffer size as a more precise closeness measure between these structures. Below we formalize the concept of buffer entropy which quantifies this intuition.

\footnote{We suppress the dependence of $\mathcal{V}_{S,S'}$ on $G_0$ to ease the notation.}
**Definition 3.2.4** (Buffer Entropy). Let \( C \) be a multi-edge divider with a base graph \( G_0 \). The buffer entropy is defined as:

\[
M_B(C, G_0) := \log \left( \max_{S \in C} \mathbb{E}_{S'} |\mathcal{V}_{S,S'}|^{-1} \right),
\]

where the expectation \( \mathbb{E}_{S'} \) is taken from uniformly sampling \( S' \) from \( C \).

We want the buffer entropy to be as large as possible to achieve sharp lower bounds. Note the following trivial bound on the size \( |\mathcal{V}_{S,S'}| \)

\[
|\mathcal{V}_{S,S'}| \leq \sum_{v \in V(S)} \mathbb{1}(v \in \mathcal{V}_{S,S'}) + \sum_{v \in V(S')} \mathbb{1}(v \in \mathcal{V}_{S,S'}). 
\]

An important condition allowing us to relate the signal strength lower bounds to buffer entropy requires that the divider is such that the variables \( \{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v \in V(S)} \) are negatively associated.

**Definition 3.2.5** (Incoherent Divider). The collection of edge sets \( C \) is called an incoherent divider with a null base \( G_0 \), if for any fixed \( S \in C \), the random variables \( \{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v \in V(S)} \) with respect to a uniformly sampled \( S' \) from \( C \) are negatively associated. In other words, for any pair of disjoint sets \( I, J \subseteq V(S) \) and any pair of coordinate-wise nondecreasing functions \( f, g \) we have:

\[
\text{Cov} \left( f(\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v \in I}), g(\{\mathbb{1}(v \in \mathcal{V}_{S,S'})\}_{v \in J}) \right) \leq 0.
\]

We show concrete constructions of incoherent dividers in Examples 3.2.2, 3.2.3 and 3.2.4. As a remark, negative association is satisfied by a variety of classical discrete distributions such as the multinomial and hypergeometric, and even more generally by the class of permutation distributions (Joag-Dev and Proschan, 1983; Dubhashi and Ranjan, 1996, e.g.). It is a standard assumption that has been exploited in other works (Addario-Berry et al., 2010, e.g.) for obtaining lower bounds.
Besides the packing entropy, the lower bound in Theorem 3.2.1 involves the maximum degree $\|A_0\|_1$ and the spectral norm $\|A_0\|_2$. We define similar quantities for the scaling edge sets case. For a divider $C$ with null base $G_0$ and any two edge sets $S, S' \in C$ define the notation:

$$A_{S,S'} := A_0 + A_S + A_{S'}.$$  \hfill (3.2.3)

As the sizes of $S, S' \in C$ are no longer ignorable, we need to consider the matrix $A_{S,S'}$ (3.2.3) instead. Denote the uniform maximum degree as $\Gamma := \max_{S, S' \in C} \|A_{S,S'}\|_1$ and uniform spectral norm as $\Lambda := \max_{S, S' \in C} \|A_{S,S'}\|_2$. We define

$$R := \max_{S, S' \in C} \frac{|S \cap S'|}{|\mathcal{V}_{S,S'}|}, \quad B := \Lambda^4 \wedge \max_{S, S' \in C} (\Gamma^2 |\mathcal{V}_{S,S'}|).$$

$R$ is an edge-node ratio measuring how dense the edge set $S \cap S'$ is compared to the vertex buffers. The quantity $B$ is an auxiliary quantity which assembles maximum degrees, spectral norms and buffer sizes and helps to obtain a compact lower bound formulation.

Below we connect the structural features we defined above to the lower bound. Recall definitions (3.1.4) and (3.1.5) on $S_0(\theta, s)$ and $S_1(\theta, s)$. We have the following theorem.

**Theorem 3.2.2.** Let $C$ be an incoherent divider with a null base $G_0$. Then if $M_B(C, G_0) \to \infty$ and

$$\theta \leq \sqrt{\frac{M_B(C, G_0)}{4nR}} \wedge \sqrt{\frac{R}{B}} \wedge \frac{1 - C^{-1}}{2\sqrt{2\Gamma}},$$  \hfill (3.2.4)

the minimax risk satisfies $\liminf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1$.

When the sample size $n$ is sufficiently large, the buffer entropy term on the right hand side of (3.2.4) is the smallest term and drives the bound which bares similarity to Theorem 3.2.1.
Figure 3.8: Test for maximum degree $\mathcal{G}_0 = \{ G | d_{\text{max}}(G) \leq s_0 \}$ and $\mathcal{G}_1 = \{ G | d_{\text{max}}(G) \geq s_1 \}$ with $s_0 = 3$ and $s_1 = 6$.

To better illustrate the usage of Theorem 3.2.2 we consider three examples. First we focus on the problem of testing whether the maximum degree in the graph is $\leq s_0$ vs $\geq s_1$. When $s_0 = 0$, this problem is related to the problem of detecting a set of $s_1$ signals in the normal means model (Ingster, 1982; Baraud, 2002; Donoho and Jin, 2004; Addario-Berry et al., 2010; Verzelen and Villers, 2010; Arias-Castro et al., 2011b, e.g.). However the two problems are distinct, since we are studying structural testing in the graphical model setting. Given $s_0 < s_1 \leq s$, we let the sub-decomposition be $\mathcal{G}_0 = \{ G | d_{\text{max}}(G) \leq s_0 \}$ and $\mathcal{G}_1 = \{ G | d_{\text{max}}(G) \geq s_1 \}$. We summarize our results in the following

Example 3.2.2 (Maximum Degree Test with Scaling Divider). Let $\mathcal{S}_0(\theta, s)$ and $\mathcal{S}_1(\theta, s)$ be defined in (3.1.4) and (3.1.5). As is shown in Figure 3.8, we split the vertices into two parts $\{1, \ldots, \lceil \sqrt{d} \rceil \}$ and $\{ \lfloor \sqrt{d} \rfloor + 1, \ldots, d \}$. We use the first part of vertices to construct $s_0$-star graphs as $G_0$ (visualized with solid edges). To construct the divider $\mathcal{C}$, we select any $s_1 - s_0$ vertices (e.g., vertices 13, 14, 16) from the second vertices part $\{ \lfloor \sqrt{d} \rfloor + 1, \ldots, d \}$ and connect them to any center of the $s_0$-star graphs in $G_0$ (e.g., vertex 1). This gives us one of the edge sets $S \in \mathcal{C}$ (e.g., $S = \{(1, 13), (1, 14), (1, 16)\}$ depicted in dashed edges in the figure). $\mathcal{C}$ is comprised by all such edge sets. We depicted the vertex buffer $\mathcal{V}_{S, S'} = \{14, 16\}$ for $S = \{(1, 13), (1, 14), (1, 16)\}$ and $S' = \{(5, 14), (5, 16), (5, 18)\}$ and $S \cap S' = \emptyset$. Assume that $s \sqrt{\log d / n} = O(1)$ and $s = O(d^\gamma)$ for some $\gamma < 1/2$. Then for a small enough absolute constant $\kappa$ if $\theta < \kappa \sqrt{\log d / n}$ we have $\liminf_{n \to \infty} \gamma(\mathcal{S}_0(\theta, s), \mathcal{S}_1(\theta, s)) = 1$.

Due to space limitations, we show how this example follows from Theorem 3.2.2 in Section B.3 of the appendix. Here, we simply sketch the construction of the divider in Figure 3.8.
On an important note, the negative association of the random variables \( \{1(v \in \mathcal{V}_{S,S'})\}_{v \in \mathcal{V}(S)} \) can be easily deduced by a result of Joag-Dev and Proschan (1983). Our second example further illustrates the usage of Theorem 3.2.2 with a clique detection problem. Define the null and alternative parameter spaces: \( S_0 := \{ \mathbf{I}_d \} \) and

\[
S_1(\theta, s) := \{ \mathbf{I}_d + \theta (\mathbf{v} \mathbf{v}^T - \mathbf{I}_d) \mid \theta \in (0, 1), \forall j : \mathbf{v}_j \in \{ \pm 1, 0 \}, \|\mathbf{v}\|^2_2 = s \}.
\]

This setup is related to that in Berthet and Rigollet (2013); Johnstone and Lu (2009). Our case is different from previous works because we parametrize the precision matrix rather than the covariance matrix, and the parametrization is distinct. Under our parametrization, the graph in the alternative hypothesis consists of a single \( s \)-clique.

**Example 3.2.3** (Sparse Clique Detection). Suppose \( s = O(d^\gamma) \) for a \( \gamma < 1/2 \). In Figure 3.9a, we set \( G_0 = (\mathcal{V}, \emptyset) \) and visualize two intersecting \( S, S' \in \mathcal{C} \): \( S \) is the 5-clique \( K_{\{1,3,5,7,9\}} \), \( S' \) is the 5-clique \( K_{\{1,2,3,9,10\}} \) and \( \mathcal{V}_{S,S'} = \{1, 3, 9\} \). For values of \( \theta < \frac{1}{4\sqrt{2}s} \wedge \sqrt{\frac{\log(d/s^2)}{4ns}} \) we have

\[
\liminf_{n \to \infty} \gamma(S_0, S_1(\theta, s)) = 1.
\]

We show how Example 3.2.3 follows from Theorem 3.2.2 in Section B.3 of the appendix. The divider construction we use is simply drawing \( s \) vertices and connecting them to form a \( s \)-clique. Figure 3.9a illustrates two sets from the divider along with their vertex buffer. Algorithm matching the lower bound is discussed in Section B.5 of the appendix.
We conclude this Section by a final example on cycle detection. In this example the sub-decomposition is $G_0 = \{(V, \emptyset)\}$, and $G_1 = \{(V, C) \mid |C| = |(v_1, v_2), \ldots, (v_{s-1}, v_s), (v_s, v_1)|, v_i \in V \}$ for an integer $s \in \mathbb{N}$. We have the following example, whose proof can be found in Section B.3 of the appendix. We show two sets from the divider on Figure 3.9b.

**Example 3.2.4** (Sparse Cycle Detection). Suppose $s = O(d^\gamma)$ for a $\gamma < 1/2$. In Figure 3.9b $S$ is the 5-cycle $C_{\{1,3,5,7,9\}}$, $S'$ is the 5-cycle $C_{\{1,2,3,9,10\}}$ and $\mathcal{V}_{S,S'} = \{1, 3, 9\}$. Then for a small enough absolute constant $\kappa$ if $\theta < \kappa \sqrt{\log d / n}$ we have

$$\liminf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1.$$  

Therefore, the skip-down algorithm matches the lower bounds. See Section 2.4 for the power analysis in the upper bound.

### 3.3 Lower Bounds of Confidence Interval Length

In this section, we show the lower bound of the confidence interval length for a generic family of invariants characterized by a concept called “hollow graphs”, whose formal definition is given below.

**Definition 3.3.1** (Hollow graph). A graph $G = (V, E)$ is called $R$-hollow, if

$$\max_{\emptyset \neq F \subseteq E} \frac{|F|}{|V(F)| - 1} \leq R. \quad (3.3.1)$$

The quantity $R$ in (3.3.1) measures the maximal “density” of edges. A $k$-clique is $k$-hollow as the edges are fully connected and a $k$-chain is $(1 - 1/k)$-hollow as it is relatively sparse. Therefore, we can see that any graph with maximal degree $s$ is at most $s$-hollow. Definition 3.3.1 is a classical definition proposed by Nash-Williams (1964). If a graph is $R$-hollow for some constant $R$ independent of the graph size, we say the graph is hollow.
Before presenting the theorem on the lower bound, we introduce a few notations on the graph. The maximum degree of a graph \( G = (V, E) \) is denoted by \( d_{\text{max}}(G) \). The union of two graphs \( G_1 = (V, E_1) \) and \( G_2 = (V, E_2) \) is \( G_1 \cup G_2 := (V, E_1 \cup E_2) \). We say \( G' = (V, E') \) is an isomorphic copy of \( G = (V, E) \) if \( G' \) is isomorphic to \( G \) and \( V(E') \cap V(E) = \emptyset \). Furthermore, we say \( G_1, \ldots, G_N \) are different isomorphic copies of \( G = (V, E) \) if each \( G_j \) is an isomorphic copy of \( G \) and \( V(E_j) \cap V(E_{j'}) = \emptyset \) for any \( 1 \leq j \neq j' \leq N \).

Now we briefly explain the intuition behind the proof of the lower bound of confidence interval length. Given an invariant \( \mathcal{I} \), the idea of the proof is to reduce the problem to a lower bound in a certain hypothesis test (Cai and Guo, 2015). In order to obtain sharp bounds, we compare a graph with invariant equal to \( I^*_L \) with multiple graphs with invariant equal to \( I^*_U \). The construction of the alternative graphs with invariant \( I^*_U \) relies on the existence of isomorphic copies of a certain graph.

**Theorem 3.3.1** (Lower bound of confidence interval length). Given any monotone invariant \( \mathcal{I} \) in the range \( [I^*_L, I^*_U] \), suppose there exist two graphs \( G_L = (V, E_L) \) and \( G_U = (V, E_U) \) with \( G_L \preceq G_U \) and they satisfy \( |V(E_L)| = O(1) \), \( |V(E_U)| = o(d^{1/2}) \) and \( G_U \) is hollow. Given some \( N \) satisfying \( d^{1/2} \leq N \leq d/(2|V(E_L)|) \), we assume there exist \( N \) different isomorphic copies of \( G_L \) denoted as \( G_L, 1, \ldots, G_L, N \) such that

\[
\mathcal{I}(\bigcup_{j=1}^N G_{L,j} \cup G_L) = I^*_L \quad \text{and} \quad \mathcal{I}(\bigcup_{j=1}^N G_{L,j} \cup G_U) = I^*_U. \tag{3.3.2}
\]

If there exist constants \( C_1 \) and \( C_2 \) such that

\[
\theta \leq C_1 \sqrt{\log d/n} \quad \text{and} \quad d_{\text{max}}(G_U) \sqrt{\log d/n} \leq C_2, \tag{3.3.3}
\]

we have the following lower bound on the confidence interval length

\[
\liminf_{n \to \infty} \inf_{[\hat{L}, \hat{U}] \in \mathcal{I}(I, \alpha)} \sup_{\Theta \in \mathcal{U}_n(I^*_L, I^*_U; \theta)} \frac{\mathbb{E}_\Theta[\hat{U} - \hat{L}]}{\text{Oracle Length}(\Theta)} \geq 1 - 2\alpha, \tag{3.3.4}
\]
where Oracle Length($\Theta$) is defined in (2.4.8).

We now sketch the high level idea behind the proof of Theorem 3.3.1. The first step reduces the minimax result in (3.3.4) to a lower bound of testing $H_0 : \mathcal{I}(\Theta) = I^*_L$ versus $H_1 : \mathcal{I}(\Theta) = I^*_U$. Next we further reduce the above test to the test $H_0 : \Theta = \Theta_0$ versus $H_1 : \Theta \in \{\Theta_1, \ldots, \Theta_M\}$, where $\mathcal{I}(\Theta_0) = I^*_L$ and $\mathcal{I}(\Theta_j) = I^*_L$ for $1 \leq j \leq M$. In order to obtain sharp bounds one needs to construct a maximally challenging set of matrices \{\Theta_0, \Theta_1, \ldots, \Theta_M\} for hypothesis test. Condition (3.3.2) enables our construction. We choose a $\Theta_0$ with $G(\Theta_0) = \bigcup_{j=1}^N G_{L,j} \cup G_L$ and select $\Theta_1, \ldots, \Theta_M$ so that each $G(\Theta_j)$ for $1 \leq j \leq M$ is isomorphic to $\bigcup_{j=1}^N G_{L,j} \cup G_U$. The reason the graph $\bigcup_{j=1}^N G_{L,j} \cup G_U$ is used is to reproduce multiple isomorphic $G(\Theta_j)$'s in the alternative which makes it challenging to tell them from the graph $G(\Theta_0)$ under the null. Assumptions on the sizes of $G_L$ and $G_U$, as well as the range of $N$ are imposed to ensure the existence of sufficiently many graphs in the alternative. The second condition of (3.3.3) guarantees the positive definiteness of the precision matrix. It is satisfied when the condition $s \sqrt{\log d/n} = o(1)$ holds.

The following theorem gives a generic lower bound of confidence upper bound, i.e., the right endpoint of the confidence interval.

**Theorem 3.3.2** (Negative result of confidence upper bound). Given any monotone invariant $\mathcal{I}$ in the range $[I^*_L, I^*_U]$. Under the same conditions as Theorem 3.3.1, if (3.3.3) is satisfied, we have

$$\liminf_{n \to \infty} \inf_{\hat{U} \in \mathcal{U}(\mathcal{I}, \alpha)} \sup_{\Theta \in \mathcal{U}(I^*_L; I^*_U; \theta)} \mathbb{E}_\Theta [\hat{U} - \mathcal{I}(\Theta)] \geq (I^*_U - I^*_L)(1 - \alpha). \quad (3.3.5)$$

In the final part of this section, we give concrete examples of invariants satisfying the conditions of Theorem 3.3.1. Specifically, the following examples show how to derive Theorems 2.4.2 and 2.4.3 from Theorems 3.3.1 and 3.3.2.

**Example 3.3.1.** The maximum degree $I_{\text{Deg}}$ satisfies the conditions of Theorem 3.3.1 for $I^*_L = O(1)$ and $I^*_U = o(d^{1/2})$. We visualize the construction in Figure 3.10(a). We let $G_L$ be
be an $I_L^*$-star and $G_U$ be an $I_U^*$-star. By definition, we immediately have that $G_U$ is hollow. Since $I_U^* = o(d^{1/2})$ and $I_L^* = O(1)$, we have $|V(E_L)| = O(1)$ and $|V(E_U)| = o(d^{1/2})$. We let $N = d/(2(I_L^* + 1)) = \Omega(d^{1/2})$ and reproduce $N$ different isomorphic copies of the $I_L^*$-star, as $G_{L,1}, \ldots, G_{L,N}$ shown in Figure 3.10(a). One can easily verify that this construction satisfies (3.3.2).

**Example 3.3.2.** The negative number of connected subgraphs $\mathcal{I}_{\text{conn}}$ satisfies the conditions of Theorem 3.3.1 for any pair of $I_L^*$ and $I_U^*$ such that $I_L^* = -\lfloor \gamma d \rfloor$ for some $\gamma \in (1/2, 1]$ and $I_U^* - I_L^* = o(d^{1/2})$. We let $G_L$ be a $1/(2\gamma - 1)$-loop$^2$ and $G_U$ be a graph connecting an $(I_U^* - I_L^*)$-chain to $G_L$, as is shown in Figure 3.10(b). It is easy to see that $G_U$ is hollow, $|V(E_L)| = O(1)$ and $|V(E_U)| = o(d^{1/2})$. Choose $N = (2k - 1)d/2$ and reproduce $N$ different isomorphic copies of the $1/(2\gamma - 1)$-loop. It can be checked that (3.3.2) is satisfied.

**Example 3.3.3.** The negative number of isolated nodes $\mathcal{I}_{\text{iso}}$ satisfies the conditions of Theorem 3.3.1 for any pair of $I_L^*$ and $I_U^*$ satisfying $I_L^* = -\lfloor \gamma d \rfloor$ for some $\gamma \in (0, 1)$ and $I_U^* - I_L^* = o(d^{1/2})$. We let $G_L$ be a 2-chain and $G_U$ be a graph connecting an $(I_U^* - I_L^*)$-chain to $G_L$, as is shown in Figure 3.10(c). By construction $G_U$ is hollow, $|V(E_L)| = O(1)$ and $|V(E_U)| = o(d^{1/2})$. Choose $N = (d - I_L^*)/2$ and reproduce $N$ different isomorphic copies of the 2-chain. One can easily check that (3.3.2) is satisfied.

---

$^2$Without loss of generality, we assume $1/(2\gamma - 1)$ is an integer and same for $N$. 

65
Chapter 4

Time-Varying Nonparanormal Graphical Models

We consider the problem of inferring time-varying undirected graphical models from high dimensional non-Gaussian distributions. Undirected graphical models have been widely used as a powerful tool for exploring the dependency relationships between variables. We are interested in graphical models which have non-static graphical structures and can handle heavy-tail distributions as well as data contaminated with outliers. To that end, we develop a class of time-varying nonparanormal models, which can be used to explore Markov dependencies of a random vector $X$ given the index variable $Z$. Specifically, we assume the random variables $(X, Z)$ follow the following joint distribution: the conditional distribution of $X | Z = z$ follows a nonparanormal distribution

$$X | Z = z \sim \text{NPN}_d(0, \Sigma(z), f)$$  \hspace{1cm} (4.0.1)

where $f = \{f_1, \ldots, f_d\}$ is a set of $d$ univariate, strictly increasing functions and $Z$ is a random variable with a continuous density. A variable follows a nonparanormal distribution $Y \sim \text{NPN}_d(\mu, \Sigma, f)$ if $f(Y) \sim N(\mu, \Sigma)$ (Liu and Wasserman, 2009).
4.1 Preliminaries

We start by providing background on the nonparanormal distribution and discuss how it relates to the time-varying nonparanormal graphical model in (4.0.1). The nonparanormal distribution was introduced in Liu and Wasserman (2009). A random variable $X = (X_1, \ldots, X_d)^T$ is said to follow a nonparanormal distribution if there exists a set of monotone univariate functions $f = \{f_1, \ldots, f_d\}$ such that $f(X) := (f_1(X_1), \ldots, f_d(X_d))^T \sim N(0, \Sigma)$, where $\Sigma$ is a latent correlation matrix satisfying $\text{diag}(\Sigma) = 1$. We denote $X \sim \text{NPN}_d(0, \Sigma, f)$.

Given $n$ independent copies of $X \sim \text{NPN}_d(0, \Sigma, f)$, Liu et al. (2012a) study how to estimate the latent correlation matrix $\Sigma$. The key idea lies in relating the Kendall’s tau correlation matrix with the Pearson correlation. The Kendall’s tau correlation between $X_j$ and $X_k$, two coordinates of $X$, is defined as

$$
\tau_{jk} = \mathbb{E} \left[ \text{sign} \left( (X_j - \tilde{X}_j)(X_k - \tilde{X}_k) \right) \right],
$$

where $(\tilde{X}_j, \tilde{X}_k)$ is an independent copy of $(X_j, X_k)$. It can be related to the latent correlation matrix using the fact that $\tau_{jk} = (2/\pi) \arcsin(\Sigma_{jk})$ when $X$ follows a nonparanormal distribution (Fang et al., 1990). The inverse covariance matrix $\Omega = \Sigma^{-1}$ encodes the graph structure of a nonparanormal distribution (Liu and Wasserman, 2009). Specifically $\Omega_{jk} = 0$ if and only if $X_j$ is independent of $X_k$ conditionally on $X_{\{j,k\}}$.

The above observations lead naturally to the following estimation procedure for $\Omega$. We estimate the Kendall’s tau correlation matrix $\hat{T} = [\hat{\tau}_{jk}] \in \mathbb{R}^{d \times d}$ elementwise using the following $U$-statistic

$$
\hat{\tau}_{jk} = \frac{2}{n(n-1)} \sum_{1 \leq i < i' \leq n} \text{sign}(X_{ij} - X_{i'j}) \text{sign}(X_{ik} - X_{i'k}).
$$
An estimate of the latent correlation matrix is given as \( \widehat{\Sigma} = \sin\left(\pi \widehat{T}/2\right) \), where \( \sin(\cdot) \) is applied elementwise. Finally, the estimate of the latent correlation matrix \( \widehat{\Sigma} \) is used as a plug-in statistic in the CLIME estimator (Cai et al., 2011), or calibrated CLIME estimator (Zhao and Liu, 2014), to obtain the inverse covariance estimator \( \widehat{\Omega} \).

The CLIME estimator solves the following optimization program

\[
\widehat{\Omega}^{\text{CLIME}}_j = \arg\min_{\beta \in \mathbb{R}^d} \|\beta\|_1 \quad \text{subject to} \quad \|\widehat{\Sigma}\beta - e_j\|_{\infty} \leq \lambda,
\]

(4.1.1)

where \( e_j \) is the \( j \)-th canonical basis in \( \mathbb{R}^d \) and the penalty parameter \( \lambda \) that controls the sparsity of the resulting estimator is commonly chosen as \( \lambda \approx \|\Omega\|_1 \sqrt{\log d/n} \) (Cai et al., 2011). Note that the tuning parameter depends on the unknown \( \Omega \) through \( \|\Omega\|_1 \), which makes practical selection of \( \lambda \) difficult. The calibrated CLIME is a tuning-insensitive estimator, which alleviates this problem. The calibrated CLIME estimator solves

\[
(\widehat{\Omega}^{\text{CCLIME}}_j, \widehat{\kappa}_j) = \arg\min_{\beta \in \mathbb{R}^d, \kappa \in \mathbb{R}} \|\beta\|_1 + \gamma \kappa \quad \text{subject to} \quad \|\widehat{\Sigma}\beta - e_j\|_{\infty} \leq \lambda \kappa, \|\beta\|_1 \leq \kappa,
\]

(4.1.2)

where \( \gamma \) is any constant in \((0, 1)\) and the tuning parameter can be chosen as \( \lambda = C \sqrt{\log d/n} \) with \( C \) being a universal constant independent of the problem parameters. In what follows, we will adapt the calibrated clime to estimation of the parameters of the model in (4.0.1).

### 4.1.1 Model Definition

The time-varying nonparanormal graphical model in (4.0.1) is an extension of the nonparanormal distribution. For every fixed value of the index variable \( Z = z \), we have a static nonparanormal distribution \( X \mid Z = z \sim \text{NPN}_d(0, \Sigma(z), f) \) that can be easily interpreted. However, as the index variable changes, the conditional distribution of \( X \mid Z \) can change in an unspecified way. In this sense, time-varying nonparanormal graphical models extend
nonparanormal graphical models in the same way varying coefficient models extend linear regression models.

Let \( Y = (X, Z) \) denote a random pair distributed according to the time-varying nonparanormal distribution. Specifically \( Z \sim f_Z(z) \) with \( f_Z(\cdot) \) being a continuous density function supported on \([0, 1]\) and \( X \mid Z = z \sim \text{NPN}_d(0, \Sigma(z), f) \) for all \( z \in [0, 1] \). For any fixed \( z \in [0, 1] \), we denote the inverse of the correlation matrix as \( \Omega(z) = \Sigma^{-1}(z) \). Both \( f_Z(z) \) and each entry of \( \Omega(z) \) are second-order differentiable (we will formalize assumptions in Section 4.3). We denote the undirected graph encoding the conditional independence of \( X \mid Z = z \) as \( G^*(z) = (V,E^*(z)) \), with \( (j,k) \in E^*(z) \) when \( \Omega_{jk}(z) \neq 0 \). As in the static case, we relate the Kendall’s tau correlation matrix with the latent correlation matrix. Let \( T(z) = [\tau_{jk}(z)]_{jk} \) be the Kendall’s tau correlation matrix corresponding to \( X \mid Z = z \) with elements defined as

\[
\tau_{jk}(z) = \mathbb{E} \left[ \text{sign} \left( (X_j - \tilde{X}_j)(X_k - \tilde{X}_k) \right) \mid Z = z \right],
\]

where \( \tilde{X} \) is an independent copy of \( X \) conditionally on \( Z = z \). Given \( n \) independent copies of \( Y = (X, Z), \{Y_i = (X_i, Z_i)\}_{i \in [n]} \), we estimate an element of the Kendall’s tau correlation matrix using the following kernel estimator

\[
\hat{\tau}_{jk}(z) = \frac{n^{-2} \sum_{i<i'} \omega_z(Z_i, Z_i') \text{sign}(X_{ij} - X_{i'j}) \text{sign}(X_{jk} - X_{i'k})}{n^{-2} \sum_{i<i'} \omega_z(Z_i, Z_i')}, \quad \text{where} \quad \omega_z(Z_i, Z_i') = K_h(Z_i - z) K_h(Z_i' - z)
\]

with the kernel function \( K(\cdot) \) being a symmetric density function, \( K_h(\cdot) = h^{-1}K(\cdot/h) \) and \( h > 0 \) is the bandwidth parameter. We can choose the kernel function as long as it satisfies some regularity conditions, which will be specified in Assumption 4.3.3. The kernel \( U \)-statistic in (4.1.3) is a generalization of classical kernel regression (Opsomer and Ruppert, 1997; Fan and Jiang, 2005). For example, given i.i.d. samples \( \{Y_i, Z_i\}_{i=1}^n \) from the model
\[ Y = f(Z) + \epsilon, \]

the Nadaraya-Watson estimator (Bierens, 1988) is

\[ \hat{f}(z) = \frac{n^{-1} \sum_{i=1}^{n} K_h(Z_i - z)Y_i}{n^{-1} \sum_{i=1}^{n} K_h(Z_i - z)}, \] (4.1.5)

where we take the weighted average of \( Y_i \)'s and the weight \( K_h(Z_i - z) \) is related to the distance between \( Z_i \) and \( z \). In order to normalize the weights, we add the denominator in (4.1.5), which is the kernel density estimator for the density of \( Z \). Comparing (4.1.3) with the Nadaraya-Watson estimator, since the kernel \( U \)-statistic involves both \( X_i \) and \( X_{i'} \), we need to multiply the weights as (4.1.4) to ensure that both \( Z_i \) and \( Z_{i'} \) are in the neighborhood of \( z \). We also normalize the weights by dividing the denominator \( n^{-2} \sum_{i<i'} \omega_z(Z_i, Z_{i'}) \) which is the estimator of \( f_Z^2(z) \). The denominator \( n^{-2} \sum_{i<i'} \omega_z(Z_i, Z_{i'}) \) is also related to the density \( f_Z(z) \). In fact, it is the estimator of \( f_Z^2(z) \). Intuitively, we can see it from

\[ \mathbb{E} \left[ \frac{1}{n^2} \sum_{i<i'} \omega_z(Z_i, Z_{i'}) \right] = \int \int K(t_1)K(t_2)f_Z(z + t_1 h)f_Z(z + t_2 h)dt_1 dt_2 = f_Z^2(z) + O(h^2). \]

Lemma 14 provides the details and is given in the supplementary material.

Based on the above estimator, we obtain the corresponding latent correlation matrix for any index value \( z \in (0, 1) \) as

\[ \hat{\Sigma}(z) = \sin \left( \frac{\pi}{2} \hat{T}(z) \right), \] (4.1.6)

where \( \hat{T}(z) = [\hat{T}_{jk}(z)] \in \mathbb{R}^{d \times d} \).

Finally, similar to the static case, we can plug \( \hat{\Sigma}(z) \) into a procedure that gives an estimate of the inverse correlation matrix, such as the CLIME in (4.1.1) or calibrated CLIME in (4.1.2). Due to practical advantages of the calibrated CLIME, we will use it for our simulations. However, at this point we note that the inferential framework only requires the estimator of the inverse correlation matrix to converge at a fast enough rate. Therefore, in
what follows, we denote $\hat{\Omega}(z)$ a generic estimator of $\Omega(z)$. Concrete statistical properties required of the calibrated CLIME will be discussed in details in Section 4.3.2.

### 4.2 Inferential Methods

In this section, we develop a framework for statistical inference about the parameters in a time-varying nonparanormal graphical model. We focus on the following three testing problems:

- **Edge presence test**: $H_0 : \Omega_{jk}(z_0) = 0$ for a fixed $z_0 \in (0, 1)$ and $j, k \in [d]$;

- **Super-graph test**: $H_0 : G^*(z_0) \subset G$ for a fixed $z_0 \in (0, 1)$ and a fixed graph $G$;

- **Uniform edge presence test**: $H_0 : G^*(z) \subset G$ for all $z \in [z_L, z_U] \subset (0, 1)$ and a fixed graph $G$.

For all the testing problems, the alternative hypotheses is the negation of the null. The edge presence test is concerned with a local hypothesis that $X_j$ and $X_k$ are conditionally independent given $X_{\{j,k\}}$ for a particular value of the index $z_0$. Equivalently, under the null hypothesis of the edge presence test, the nodes $j$ and $k$ are not connected at a particular index value $z_0$. The null hypothesis under the super-graph test postulates that the true graph is a subgraph of a given graph $G$ for $Z = z_0$. It can also be interpreted as multiple-edge presence tests, since

$$H_0 : \Omega_{jk}(z_0) = 0, \quad \text{for all } (j, k) \in E^c,$$

where $E$ is the edge set of the graph $G$. The null hypothesis under the uniform edge presence test postulates that the true graph is a subgraph of $G$ for all index values in the range $[z_L, z_U]$. It is a generalization of the first two local tests to a global test over a range of index values.
Similar to the super-graph test, this hypothesis is equivalent to the following

\[ H_0 : \sup_{z \in [z_L, z_U]} |\Omega_{jk}(z)| = 0, \quad \text{for all } (j, k) \in E^c. \]

If the graph \( G \) consists of the edge set \( E = \{(a, b) \in V \times V \mid (a, b) \neq (j, k)\} \), the uniform edge presence test becomes a uniform single-edge test \( H_0 : \sup_{z \in [z_L, z_U]} |\Omega_{jk}(z)| = 0. \)

Next, we provide details on how to construct tests for the above three hypothesis.

### 4.2.1 Edge Presence Testing

We consider the hypothesis \( H_0 : \Omega_{jk}(z_0) = 0 \), for a fixed \( z_0 \in (0, 1) \) and \( j, k \in [d] \). In order to construct a test for this hypothesis, we introduce the score function

\[ \hat{S}_{z\sim(j,k)}(\beta) = \hat{\Omega}^T_j(z)(\hat{\Sigma}(z)\beta - e_k). \] (4.2.2)

The argument \( \beta \) of the score function corresponds to the \( k \)-th column of \( \Omega(z) \). Our test is based on the score function evaluated at \( \hat{\Omega}_{k\backslash j}(z) \) which is an estimator of \( \Omega_k(z) \) under the null hypothesis, defined as

\[ \hat{\Omega}_{k\backslash j}(z) = (\hat{\Omega}_{1k}(z), \ldots, \hat{\Omega}_{(j-1)k}(z), 0, \hat{\Omega}_{(j+1)k}(z), \ldots, \hat{\Omega}_{dk}(z))^T \in \mathbb{R}^d, \]

where \( \hat{\Omega}(z) \) is an estimator of \( \Omega(z) \). That is, we use \( \hat{S}_{z\sim(j,k)}(\hat{\Omega}_{k\backslash j}(z)) \) as the score statistic. We establish statistical properties of this statistic later. We first develop intuition for why \( \hat{S}_{z\sim(j,k)}(\hat{\Omega}_{k\backslash j}(z)) \) is a good testing statistic for \( H_0 : \Omega_{jk}(z_0) = 0 \). If we replace \( \hat{\Omega} \) and \( \hat{\Sigma} \) by the truth \( \Omega \) and \( \Sigma \) in (4.2.2), we can find the score statistic \( \hat{S}_{z\sim(j,k)}(\hat{\Omega}_{k\backslash j}(z)) \approx \Omega_j^T(z)(\Sigma(z)\Omega_{k\backslash j}(z) - e_k) = \Omega_{jk}(z) \). Therefore, the score statistic is close to zero under the null, and \( \hat{S}_{z\sim(j,k)}(\hat{\Omega}_{k\backslash j}(z_0)) \approx \hat{\Omega}_{jk}(z_0) \) under the alternative. In specific, under \( H_0 \), the testing
The statistic is close to
\[
\hat{S}_{z|j,k}(\hat{\Omega}_{k|j}) \approx \Omega_j^T(z)(\hat{\Sigma}(z) - \Sigma(z))\Omega_k(z) \approx \Omega_j^T(z)[\hat{\Sigma}(z) \circ (\hat{T}(z) - T(z))]\Omega_k(z),
\]
where \(\hat{\Sigma}_{jk}(z) = (\pi/2) \cos((\pi/2)\tau_{jk}(z))\) is the derivative of \(\Sigma_{jk}(\cdot)\) and "\(\approx\)" denotes equality up to a smaller order term. The first approximation in (4.2.3) is due to replacing \(\hat{\Omega}\) with the truth and the second approximation is due to the Taylor expansion. A rigorous derivation of this argument can be found in Appendix C.2.1. We note that the right-hand side of (4.2.3) is a linear function of \(\hat{T}(z)\), which is a \(U\)-statistic. By applying the central limit theorem for \(U\)-statistics, we will show the asymptotic normality of \(\hat{S}_{z|j,k}(\hat{\Omega}_{k|j})\). See Theorem 4.3.1 for details.

Under the null, we have that
\[
\sqrt{nh} \cdot \sigma^{-1}_{jk}(z_0)\hat{S}_{z_0|j,k}(\hat{\Omega}_{k|j}(z_0)) \rightsquigarrow N(0,1),
\]
where \(\sigma^2_{jk}(z_0) = f^{-1}_z(z_0) \text{Var}(\Omega_j^S(z_0)\Theta_{z_0}\Omega_k(z_0))\), and \(\Theta_z\) is a random matrix with elements
\[
(\Theta_z)_{jk} = \pi \cos((\pi/2)\tau_{jk}(z))\tau^{(1)}_{jk}(Y), \quad \text{where}\]
\[
\tau^{(1)}_{jk}(x,z) = \sqrt{h} \cdot \mathbb{E}[K_h(z - z_0)K_h(Z-z_0)(\text{sign}(X_j - x_j)\text{sign}(X_k - x_k) - \tau_{jk}(z_0))],
\]
where the expectation is taken for \(Z\) and \(X\). For simplicity, we denote \(y = (x^T, z)^T\) and write \(\tau^{(1)}_{jk}(y) := \tau^{(1)}_{jk}(x,z)\). The form of the asymptotic variance comes from the Hoeffding decomposition of the \(U\)-statistics in (4.1.3), with \(\tau^{(1)}_{jk}\) being the leading term of the decomposition. Technical details will be provided in Section 4.3.1 and Section C.2.1. The score statistic is a generalization of Rao’s score tests in fixed dimensional parametric models. We can apply a one-step debiased estimator of \(\Omega_{jk}(z)\) from the score statistic \(\hat{S}_{z|j,k}(\hat{\Omega}_{k|j}(z))\) following the procedure similar to (Ning and Liu, 2014). They show that the one-step es-
The estimation procedure is asymptotically equivalent to the score statistic and we choose to use the score statistic in this chapter. Janková and van de Geer (2016) considered a similar debiasing procedure specific for the nodewise regression estimator. On the other hand, we will show in Theorem 4.3.1 that the score statistic $\hat{S}_{z|j,k}(\hat{\boldsymbol{\Omega}}_{k\setminus j}(z))$ can be applied to any estimator $\hat{\boldsymbol{\Omega}}$ has sharp enough statistical rate.

In order to use the score function as a test statistic, we need to estimate its asymptotic variance $\sigma_{jk}^2(z_0)$. For any $1 \leq s \leq n$ and $1 \leq j, k \leq d$, let

$$q_{s,jk}(z) = \frac{\sqrt{n}}{n-1} \sum_{s' \neq s} \omega_z(Z_s, Z_{s'}) \left( \text{sign} \left( (X_{sj} - X_{s'j})(X_{sk} - X_{s'k}) \right) - \hat{\tau}_{jk}(z) \right), \quad (4.2.6)$$

$$\hat{\Theta}^{(s)}(z) = \pi \cos \left( (\pi/2)\hat{\tau}_{jk}(z) \right) q_{s,jk}(z).$$

With this notation, the leave-one-out Jackknife estimator for $\sigma_{jk}^2(z_0)$ is given as

$$\hat{\sigma}_{jk}^2(z_0) = \left[ U_n(\omega_{z_0}) \right]^{-2} \cdot \frac{1}{n} \sum_{s=1}^{n} \left( \hat{\Omega}_j^S(z_0) \hat{\Theta}^{(s)}(z_0) \hat{\Omega}_k(z_0) \right)^2, \quad (4.2.7)$$

where the matrix $\hat{\Theta}^{(s)}(z) = [\hat{\Theta}_{j,k}^{(s)}(z)]$. As we have remarked in Section 4.1.1, $\hat{\sigma}_{jk}^2(z_0)$ is an estimator for $f_Z^2(z)$. We divide $[U_n(\omega_{z_0})]^{-2}$ in (4.2.7) in order to normalized the weights $\omega_z(Z_s, Z_{s'})$ in the $U$-statistics $q_{s,jk}(z)$ defined in (4.2.6). The Jackknife estimator is widely used when estimating the variance of a U-statistics, which is not an average of independent random variables. The leave-one-out statistic $q_{s,jk}(z)$ in (4.2.6) is estimates the expectation in (4.2.5) by leaving $Y_s$ out of the summation in $q_{s,jk}(z)$.

Finally, a level $\alpha$ test for $H_0 : \Omega_{jk}(z_0) = 0$ is given as

$$\psi_{z_0|j,k}(\alpha) = \begin{cases} 1 \text{ if } \sqrt{n}\hat{h} \cdot |\hat{S}_{z_0|j,k}(\hat{\Omega}_{k\setminus j}(z_0))/\hat{\sigma}_{jk}(z_0)| > \Phi^{-1}(1 - \alpha/2); \\ 0 \text{ if } \sqrt{n}\hat{h} \cdot |\hat{S}_{z_0|j,k}(\hat{\Omega}_{k\setminus j}(z_0))/\hat{\sigma}_{jk}(z_0)| \leq \Phi^{-1}(1 - \alpha/2), \end{cases}$$

74
where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. The single-edge presence test is a cornerstone of more general hypothesis tests described in the next two sections. The properties of the test are given in Theorem 4.3.1.

### 4.2.2 Super-Graph Testing

In this section, we discuss super-graph testing. Recall that for a fixed $z_0$ and a predetermined graph $G = (V, E)$, the null hypothesis is

$$H_0 : G^*(z_0) \subset G.$$  \hspace{1cm} (4.2.8)

From (4.2.1), we have that the super-graph test can be seen as a multiple test for presence of several edges. Therefore, we propose the following testing statistic based on the score function in (4.2.2):

$$S(z_0) = \sqrt{nh} \cdot \mathbb{U}_n(\omega_{z_0}) \max_{(j,k) \in E^c} \hat{S}_{z_0\mid(j,k)}(\hat{\Omega}_{k\setminus j}(z_0)).$$  \hspace{1cm} (4.2.9)

In order to estimate the quantile of $S(z_0)$, we develop a novel Gaussian multiplier bootstrap for $U$-statistics. Let $\{\xi_i\}_{i \in [n]}$ be $n$ independent copies of $N(0, 1)$. Let $\hat{T}^B(z) = [\hat{\tau}_{jk}^B(z)]$ where

$$\hat{\tau}_{jk}^B(z) = \frac{\sum_{i \neq i'} K_h(Z_i - z) K_h(Z_{i'} - z) \text{sign}((X_{ij} - X_{i'j})(X_{ik} - X_{i'k})) \left(\xi_i + \xi_{i'}\right)}{\sum_{i \neq i'} K_h(Z_i - z) K_h(Z_{i'} - z) \left(\xi_i + \xi_{i'}\right)},$$  \hspace{1cm} (4.2.10)

$$\hat{\Sigma}^B(z) = \sin \left(\frac{\pi}{2} \hat{T}^B(z)\right).$$  \hspace{1cm} (4.2.11)

The Gaussian multiplier bootstrap statistic in (4.2.12) is motivated by the method developed in Chernozhukov et al. (2013), who proposed a bootstrap procedure to estimate a quantile of the supremum of high dimensional empirical processes. Their method, however, cannot be directly applied to our kernel Kendall’s tau estimator in (4.1.3), which is a ratio of two
U-statistics. If we compare (4.2.10) with (4.1.3), we add $\xi_i + \xi_i'$ into the bootstrap estimator in order to simulate the distribution of $\hat{\tau}_{jk}(z)$ in (4.1.3).

The bootstrap estimator of the test statistic $S(z_0)$ in (4.2.9) is

$$S^B(z_0) = \sqrt{n h} \cdot U_n[\omega^B_{z_0}] \max_{(j,k) \in E^c} \hat{\Omega}_j^S(z_0) \left( \hat{\Sigma}^B(z_0) \hat{\Omega}_{k\setminus j} - e_k^T \right),$$

(4.2.12)

$$U_n[\omega^B_{z_0}] = \frac{2}{n(n-1)} \sum_{i \neq i'} K_h(Z_i - z_0) K_h(Z_{i'} - z_0) (\xi_i + \xi_{i'}).$$

(4.2.13)

Here we multiply $U_n[\omega^B_{z_0}]$ in (4.2.12) in order to eliminate the denominator in (4.2.10) such that the leading term of the bootstrap statistic $S^B(z_0)$ is a Gaussian multiplier bootstrap $U$-statistic. The correlation estimator in (4.1.6) has an additional sin transform. Therefore, a new nonlinear type of multiplier bootstraps in (4.2.10) and (4.2.12) are introduced to overcome these problems and novel technical tools are then developed to study statistical properties. See Theorem 4.3.2 for the statement of statistical properties.

Denote the conditional $(1 - \alpha)$-quantile of $S^B(z_0)$ given $\{Y_i\}_{i=1}^n$ as $\hat{c}_T(1 - \alpha, \{Y_i\}_{i=1}^n)$. The level-$\alpha$ super-graph test is constructed as

$$\psi_{z_0|G}(\alpha) = \begin{cases} 1 & \text{if } S(z_0) > \hat{c}_T(1 - \alpha, \{Y_i\}_{i=1}^n); \\ 0 & \text{if } S(z_0) \leq \hat{c}_T(1 - \alpha, \{Y_i\}_{i=1}^n). \end{cases}$$

(4.2.14)

Note that the quantile $\hat{c}_T(1 - \alpha, \{Y_i\}_{i=1}^n)$ can be estimated by a Monte-Carlo method.

An alternative approach for the super-graph test in (4.2.8) is the multiple hypothesis testing. We can apply the Holm’s multiple testing procedure (Holm, 1979) to control the family-wise error for the hypotheses set $\{H_{0,(jk)}\}_{(j,k) \in E^c}$ where $H_{0,(jk)} : \Omega_{jk}(z_0) = 0$ and $G^*(z_0) \subset G = (V, E)$. However, it is not straightforward to obtain the nominal probability for the family-wise error in Holm’s method. Moreover, we will show in Theorem 4.3.2 that our testing procedure is nominal.
4.2.3 Uniform Edge Presence Testing

In this section, we develop the uniform presence test for which the null hypothesis is given as

\[ H_0 : G^*(z) \subset G \text{ for all } z \in [z_L, z_U]. \]

This test is a generalization of the edge presence test to the uniform version over both edges and index. We again use the score function in (4.2.2) to construct the test statistic

\[ W_G = \sqrt{n h} \sup_{z \in [z_L, z_U]} \max_{(j, k) \in E^c} \mathbb{U}_n[\omega_z] \hat{S}_z | (j, k)(\hat{\Omega}_{k \setminus j}(z)) \] (4.2.15)

and estimate a quantile of \( W_G \) by developing a Gaussian multiplier bootstrap. Let

\[ W_G^B = \sqrt{n h} \sup_{z \in [z_L, z_U]} \max_{(j, k) \in E^c} \mathbb{U}_n[\omega^B_z] \cdot \hat{\Omega}_j^T(z)(\hat{\Sigma}^B(z) \hat{\Omega}_{k \setminus j}(z) - e_k^T), \] (4.2.16)

where \( \hat{\Sigma}^B(z) \) is defined in (4.2.11). Let \( \hat{\omega}_W(1 - \alpha, \{Y_i\}_{i=1}^n) \) denote the conditional \((1 - \alpha)\)-quantile of \( W_G^B \) given \( \{Y_i\}_{i=1}^n \). Similar to (4.2.14), the level \( \alpha \) uniform edge presence test is constructed as

\[ \psi_G(\alpha) = \begin{cases} 
1 & \text{if } W_G > \hat{\omega}_W(1 - \alpha, \{Y_i\}_{i=1}^n); \\
0 & \text{if } W_G \leq \hat{\omega}_W(1 - \alpha, \{Y_i\}_{i=1}^n). 
\end{cases} \] (4.2.17)

Theorem 4.3.3 provides statistical properties of the test. The statistics \( W_G \) in (4.2.15) and \( W_G^B \) in (4.2.16) involve taking supreme over \( z \in [z_L, z_U] \). In practice, we approximate the suprema by evenly dividing \([z_L, z_U]\) into discrete grids and taking the maximum of the statistic over these discrete values in \([z_L, z_U]\).
4.3 Theoretical Properties

In this section, we establish the validity of tests proposed in the previous sections. Validity of tests rely on existence of estimators for the latent inverse correlation matrix with fast enough convergence. We show that the calibrated CLIME satisfy the testing requirements and, in addition, show that it achieves the minimax rate of convergence for a large class of models.

To facilitate the argument, we need the regularity and smoothness of the density function of $Z$ and the time-varying correlation matrix $\Sigma(z)$. Let us first introduce the Hölder class $\mathcal{H}(\gamma, L)$ of smooth functions. The Hölder class $\mathcal{H}(\gamma, L)$ on $(0,1)$ is the set of $\ell = \lfloor \gamma \rfloor$ times differentiable functions $g : \mathcal{X} \mapsto \mathbb{R}$ whose derivative $g^{(\ell)}$ satisfies

$$|g^{(\ell)}(x) - g^{(\ell)}(y)| \leq L|x - y|^\gamma - \ell, \text{ for any } x, y \in \mathcal{X}$$

and $\lfloor \gamma \rfloor$ denotes the largest integer smaller than $\gamma$. In this chapter, we need some regularity conditions for the functions in our model.

**Assumption 4.3.1 (Density function of $Z$).** There exist constants $0 < \underline{f}_Z < \bar{f}_Z < \infty$ such that the marginal density $f_Z$ of the index variable $Z$ has its image in $[\underline{f}_Z, \bar{f}_Z]$ and $f_Z \in \mathcal{H}(2, \bar{f}_Z)$.

**Assumption 4.3.2 (Regularization of $\Sigma_{jk}(\cdot)$).** The correlations $\Sigma_{jk}(\cdot) \in \mathcal{H}(2, M_\sigma)$ for some constant $M_\sigma < \infty$ given any $1 \leq j, k \leq d$.

The above two assumptions are standard assumptions on the marginal distribution of $Z$ (Pagan and Ullah, 1999) and time-varying graphical models (see, for example, Kolar et al., 2010a).

**Assumption 4.3.3 (Kernel function).** Through this chapter, we assume the kernel function $K$, used in (4.1.3), is a symmetric density function supported on $[-1,1]$ with bounded
variation, i.e., $\|K\|_{\infty} \vee \text{TV}(K) < \infty$,

$$\int_{-1}^{1} K(u)du = 1 \text{ and } \int_{-1}^{1} uK(u)du = 0.$$  

These properties are also required in Zhou and Wasserman (2010). Many widely used kernels, including the uniform kernel $K(u) = 0.5 \mathbb{I}(|u| < 1)$, the triangular kernel $K(u) = (1 - |u|) \mathbb{I}(|u| < 1)$, and the Epanechnikov kernel $K(u) = 0.75(1 - u^2) \mathbb{I}(|u| < 1)$, satisfy this assumption.

Finally, we list a generic assumption on the properties of $\hat{\Sigma}(z)$ in (4.1.6) and the an inverse correlation matrix estimator $\hat{\Omega}(z)$.

**Assumption 4.3.4 (Statistical rates).** There are sequences $r_{1n}, r_{2n}, r_{3n} = o(1)$ such that

$$\sup_{z \in (0,1)} \|\hat{\Sigma}(z) - \Sigma(z)\|_{\max} \leq r_{1n}, \quad \sup_{z \in (0,1)} \|\hat{\Omega}(z) - \Omega(z)\|_{1} \leq r_{2n}, \quad \text{and}$$

$$\sup_{z \in (0,1)} \max_{j \in [d]} \|\hat{\Sigma}(z)\hat{\Omega}_j(z) - e_j\|_{\infty} \leq r_{3n},$$

with probability at least $1 - 1/d$.

Assumption 4.3.4 is a generic condition on the consistency of $\hat{\Sigma}(z)$ and $\hat{\Omega}(z)$. We aim to show that our testing methods are independent to any specific procedure to estimate $\Omega(\cdot)$. The three rates in Assumption 4.3.4 are sufficient for the validity of our tests. Our inferential framework can thus be easily generalized to other inverse correlation matrix estimators as long as their rates satisfy Assumption 4.3.4. Under this assumption, the score statistic used for testing can be approximated by an asymptotically normal leading term. For our estimators $\hat{\Sigma}(\cdot)$ in (4.1.6) and $\hat{\Omega}(\cdot)$ in (4.1.1) or (4.1.2), we will show in Theorems 4.3.4 and
4.3.5 that

\[ r_{1n} = O(\sqrt{\log(d/h)/(nh)}), \quad r_{2n} = O(s\sqrt{\log(d/h)/(nh)}) \quad {\text{and}} \quad r_{3n} = O(\sqrt{\log(d/h)/(nh)}). \]  

(4.3.1)

See Section 4.3.2 for more details.

### 4.3.1 Validity of Tests

In this section, we state theorems on asymptotic validity of the tests considered in Section 4.2. We first define the parameter space

\[ \mathcal{U}_s(M, \rho) = \left\{ \Omega \in \mathbb{R}^{d \times d} \mid \Omega > 1/\rho, \| \Omega \|_2 \leq \rho, \max_{j \in [d]} \| \Omega_j \|_0 \leq s, \| \Omega \|_1 \leq M \right\}. \]  

(4.3.2)

This matrix class was considered in the literature on inverse covariance matrix estimation (Cai et al., 2012) and time-varying covariance estimation (Chen and Leng, 2016).

The following theorem gives us the limiting distribution of the score function defined in (4.2.2).

**Theorem 4.3.1** (Edge presence test). For a fixed \( z_0 \in (0, 1) \), suppose \( \Omega(z_0) \in \mathcal{U}_s(M, \rho) \), Assumption 4.3.4 holds with \( \sqrt{nh} \cdot (r_{2n}(r_{1n} + r_{3n})) = o(1) \) and the bandwidth \( h \) satisfies

\[ \sqrt{nh} \left( \log(dn)/(nh) + h^2 \right) + s^3/\sqrt{nh} = o(1). \]  

(4.3.3)

Furthermore, for a fixed and \( j, k \in [d] \), suppose there exists \( \theta_{\min} > 0 \) such that

\[ \mathbb{E}(\Omega_j^T(z_0) \Theta_{z_0} \Omega_k(z_0))^2 \geq \theta_{\min} \| \Omega_j(z_0) \|_2^2 \| \Omega_k(z_0) \|_2^2, \]
then under $H_0 : \Omega_{jk}(z_0) = 0$, we have that

$$\sqrt{nh} \cdot \sigma^{-1}_{jk}(z_0) \tilde{S}_{z_0|[j,k]}(\tilde{\Omega}_{k|j}(z_0)) \sim N(0, 1),$$

where $\sigma^2_{jk}(z_0) = f^{-4}_Z(z_0) \text{Var}(\Omega^T_j(z_0) \Theta_{z_0} \Omega_k(z_0))$ and $\Theta_{z_0}$ is defined in (4.2.4).

We have two sets of scaling conditions in the above theorem. Under the condition $\sqrt{nh}(r_2 n(r_1 n + r_3 n)) = o(1)$, the first heuristic approximation “≈” in (4.2.3) is valid. The condition in (4.3.3) guarantees that the leading term on the right hand side of (4.2.3) is asymptotically normal. In particular, the first term of (4.3.3) makes the second heuristic approximation in (4.2.3) valid and allows for control of the higher order term of the Hoeffding decomposition of the $U$-statistics in (4.1.3). If $r_{1n}, r_{2n}$ and $r_{3n}$ have the rates as in (4.3.1) (see Theorems 4.3.4 and 4.3.5 for more details), this condition becomes $\sqrt{nh} \cdot s(h^2 + \sqrt{\log(\hat{d}n)/(nh)^2}) = o(1)$. We choose $h \asymp n^{-\nu}$, where $\nu > 1/5$ in order to remove the bias. The two scaling conditions in Theorem 4.3.1 can be replaced by $(s^3 + s \log(\hat{d}n)) / n^{1-\nu/2} = o(1)$. This is similar to the condition $s^2 \log d / \sqrt{n}$ in the inference for the Lasso estimator (Zhang and Zhang, 2013; van de Geer et al., 2014; Javanmard and Montanari, 2014). Here, the slower $n^{-(1-\nu)/2}$ term originates from the nonparametric relationship between the index and correlation matrix. The additional $s^2$ term comes from the matrix structure and is ignorable if $s = o\left(\sqrt{\log(\hat{d}n)}\right)$.

The following lemma shows that the asymptotic variance of the score function can be consistently estimated.

**Lemma 4.3.1.** Suppose the conditions of Theorem 4.3.1 hold. If $r_{2n}/h = o(1)$ and $\log(\hat{d}n)/(nh^3) = o(1)$, then the variance estimator $\hat{\sigma}^2_{jk}(z_0)$ in (4.2.7) has $\tilde{\sigma}^2_{jk}(z_0) \rightarrow_{P} \sigma^2_{jk}(z_0)$.

The proofs of Theorem 4.3.1 and Lemma 4.3.1 are deferred to Appendix C.2.1 and C.7.4 respectively. Stronger scaling conditions are needed for consistent estimation of variance as its estimator in (4.2.7) relies on controlling higher moments. Under the rates in (4.3.1) with
$h \asymp n^{-\nu}$, for some $\nu > 1/5$, the scaling $(s^3 + s^2 \log(dn))/n^{(1-\nu)/2} = o(1)$ suffices for the estimator to consistently estimate the variance.

We also have the following theorems on the asymptotic validity of the super-graph test.

**Theorem 4.3.2** (Super-graph test). Let $z_0 \in (0,1)$ and $j, k \in [d]$ be fixed. Assume the conditions of Theorem 4.3.1 hold. Suppose there exists $\theta_{\min} > 0$ such that for all $j \neq k \in [d]$,

$$
\mathbb{E}(\Omega_j^T(z_0) \Theta_{z_0} \Omega_k(z_0))^2 \geq \theta_{\min} \|\Omega_j(z_0)\|_2^2 \|\Omega_k(z_0)\|_2^2,
$$

and there exists a constant $\epsilon > 0$ such that

$$
\sqrt{nh}(r_3n(r_1n + r_3n)) = O(n^{-\epsilon}) \quad \text{and}
$$

$$
\sqrt{nh}(\log(dn)/(nh) + h^2) + \log d/(nh^2) + (\log(dn))^7/(nh) = O(n^{-\epsilon}).
\quad (4.3.4)
$$

Let $G = (V,E)$ be any fixed graph and $G^*(z_0)$ is the Markov graph corresponding to the index value $z_0$. Under the null hypothesis $H_0 : G^*(z_0) \subset G$, the test $\psi_{z_0,G}(\alpha)$ defined in (4.2.14) satisfies

$$
\sup_{\alpha \in (0,1)} \left| \mathbb{P}_{H_0}(\psi_{z_0,G}(\alpha) = 1) - \alpha \right| = O(n^{-c}) \quad (4.3.5)
$$

for some universal constant $c$.

The next theorem shows the asymptotic validity of the uniform edge presence test.

**Theorem 4.3.3** (Uniform edge presence test). Assume that $\Omega(z) \in \mathcal{U}(M, \rho)$ for any $z \in (0,1)$ and Assumption 4.3.4 is true. Suppose there exists $\theta_{\min} > 0$ such that for all $j \neq k \in [d]$ and $z \in (0,1)$, $\mathbb{E}(\Omega_j^T(z) \Theta_{z} \Omega_k(z))^2 \geq \theta_{\min} \|\Omega_j(z)\|_2^2 \|\Omega_k(z)\|_2^2$, and there exists a constant $\epsilon > 0$ such that

$$
\sqrt{nh}(r_2n(r_1n + r_3n)) = O(n^{-\epsilon}) \quad \text{and}
$$

$$
\sqrt{nh}(\log(dn)/(nh) + h^2) + \log d/(nh^2) + (\log(dn))^7/(nh) = O(n^{-\epsilon}).
\quad (4.3.6)
$$

Under the null hypothesis $H_0 : G^*(z) \subset G$ for all $z \in [z_L, z_U]$, the test $\psi_G(\alpha)$ defined in (4.2.17) satisfies

$$
\sup_{\alpha \in (0,1)} \left| \mathbb{P}_{H_0}(\psi_G(\alpha) = 1) - \alpha \right| = O(n^{-c}) \quad (4.3.7)
$$
for some universal constant $c > 0$.

We defer the proof of Theorem 4.3.2 to Appendix C.2.2 and the proof of Theorem 4.3.3 to Appendix C.6.

Theorems 4.3.2 and 4.3.3 only depend on the estimation rates of $\hat{\Sigma}(z)$ and $\hat{\Omega}(z)$ through Assumption 4.3.4. This implies that our inferential framework does not rely on exact model selection. We have $O(n^{-\epsilon})$ in (4.3.4) and (4.3.6) instead of $o(1)$ in (4.3.3) to achieve the polynomial convergence rate for type I error in (4.3.5) and (4.3.7). Comparing the scaling condition in (4.3.4) with the one in (4.3.3), the second term in (4.3.4) is dominated by the first term under a mild bandwidth rate $h = o(n^{-1/3})$. The third term $(\log(dn))/n(h)$ in (4.3.4) comes from a Berry-Essen bound on the suprema of increasing dimensional $U$-processes. Such a scaling condition is similar to the one in Chernozhukov et al. (2013). They showed that for the empirical process $W = (W_1, \ldots, W_d)^T$ having the same covariance as the centered Gaussian vector $U = (U_1, \ldots, U_d)^T$, the following Berry-Essen bound holds

$$
\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left( \max_j W_j \leq t \right) - \mathbb{P}\left( \max_j U_j \leq t \right) \right| = O\left( \left( (\log(dn))/n \right)^{1/6} \right).
$$

Comparing with our condition that $(\log(dn))/n(h) = O(n^{-\epsilon})$, the additional term $nh$ in the denominator comes from the nonparametric part of our estimator. Furthermore, Theorem 4.3.3 requires a stronger scaling condition in (4.3.6), where the term $(\log(dn))/n(h) = O(n^{-\epsilon})$ arises from the additional supremum over $z \in [z_L, z_U]$ in the uniform edge presence test.

### 4.3.2 Consistency of Estimation

In this section, we show that the Assumption 4.3.4 holds under mild conditions on the data generating process. We give explicit rates for $r_{1n}, r_{2n}$ and $r_{3n}$ under concrete estimation
procedures. We first show the estimation rate of \( \hat{\Sigma} \) given in (4.1.6). Next, we give the rate of convergence for \( \Omega(z) \) when using the (calibrated) CLIME estimator.

We establish rates of convergence that are uniform in bandwidth \( h \). Uniform in bandwidth results are important as they ensure consistency of our estimators even when the bandwidth is chosen in a data-driven way, which is the case in practice, including the cross-validation over integrated squared error Hall (1992) and other risks (Muller and Stadtmuller, 1987; Ruppert et al., 1995; Fan and Gijbels, 1995). See Jones et al. (1996) for a survey of other methods. Existing literature on uniform in bandwidth consistency focuses on low dimensional problems (see, for example, Einmahl and Mason, 2005). High dimensional statistical methods usually have more tuning parameters and it is hard to guarantee that the selected bandwidth satisfies an optimal scaling condition. To the best of our knowledge this is the first result established in a high-dimensional regime that shows uniform consistency for a wide range of possible bandwidths.

The following theorem shows the rate of the covariance matrix estimator.

**Theorem 4.3.4.** Assume \( \log d/n = o(1) \) and the bandwidths \( 0 < h_l < h_u < 1 \) satisfy

\[
h_l n / \log(dn) \to \infty \quad \text{and} \quad h_n = o(1).
\]

There exists a universal constant \( C_\Sigma > 0 \) such that for any \( \delta \in (0, 1) \), we have

\[
\sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \frac{\| \hat{\Sigma}(z) - \Sigma(z) \|_{\max}}{h^2 + \sqrt{(nh)^{-1} \left[ \log(d/h) \vee \log \left( \delta^{-1} \log(h_u h_l^{-1}) \right) \right]}} \leq C_\Sigma \tag{4.3.8}
\]

with probability \( 1 - \delta \).

The proof of this theorem is deferred to Appendix C.1.1. Using (4.3.8), we can determine the rate of \( r_n \) in Assumption 4.3.4. The supremum over the bandwidth \( h \) in (4.3.8) implies that if a data-driven bandwidth \( \hat{h}_n \) satisfies \( P(h_l \leq \hat{h}_n \leq h_u) \to 1 \), then with high
probability,

$$\sup_{z \in (0,1)} \| \hat{\Sigma}(z) - \Sigma(z) \|_{\max} \leq C_\Sigma \left( \frac{\tilde{h}_n^2}{n} + \sqrt{\frac{\log(d/\tilde{h}_n)}{nh_n}} \right). \quad (4.3.9)$$

The first term in the rate is the bias and the second is the variance. Our result is sharper than the rate $O\left( h^2 \sqrt{\log d} + \sqrt{\log d/(nh)} \right)$ established in Lemma 9 of Chen and Leng (2016). The uniform consistency result $\sup_{h \leq h \leq h_u} \sup_{z \in (0,1)} \| \hat{\Sigma}(z) - \Sigma(z) \|_{\max} = o_P(1)$ holds for a wide range of bandwidths satisfying $h_t n / \log(dn) \to \infty$ and $h_u = o(1)$, which allows flexibility for data-driven methods. In fact, such $h_t$ is the smallest to make the variance (4.3.9) converge and $h_u$ is the largest for the convergence of bias. When $d = 1$, the range $[h_t, h_u]$ is the same as for the kernel-type function estimators (Einhahl and Mason, 2005).

Due to the large capacity of the estimator $\hat{\Sigma}(z)$ in (4.3.8), which varies with both the bandwidth $h$ and the index $z$, the routine proof based on uniform entropy numbers does not easily apply here. We use the peeling method (Van de Geer, 2000) by slicing the range of $h$ into smaller intervals, for which the uniform entropy number is controllable. Finally, we assemble the interval specific bounds to obtain (4.3.8). See Section C.5.1 for more details.

Next, we give a result on the estimation consistency of the inverse correlation matrix. Let $\hat{\Omega}(z) = (\hat{\Omega}_1(z), \ldots, \hat{\Omega}_d(z))$ where each column $\hat{\Omega}_j(z)$ is constructed either by using the CLIME in (4.1.1) or calibrated CLIME in (4.1.2). We recommend using the calibrated CLIME in practice due to the tuning issues discussed in Section 4.1.1.

**Theorem 4.3.5.** Suppose $\Omega(z) \in U_\mathcal{A}(M, \rho)$ for all $z \in (0,1)$. Assume $\log d/n = o(1)$, the bandwidths $0 < h_t < h_u < 1$ satisfy $h_t n / \log(dn) \to \infty$ and $h_u = o(1)$. The regularization parameter $\lambda$ is chosen to satisfy $\lambda \geq \lambda_{n,h} := C_\Sigma (h^2 + \sqrt{\log(d/h)/(nh)})$, where $C_\Sigma$ is the constant in (4.3.8), for the calibrated CLIME and $\lambda \geq M\lambda_{n,h}$ for the CLIME estimator. Then there exists a universal constant $C > 0$ such that

$$\sup_{h \in [h_t, h_u]} \sup_{z \in (0,1)} \frac{1}{\lambda M^2} \| \hat{\Omega}(z) - \Omega(z) \|_{\max} \leq C; \quad (4.3.10)$$
sup \sup \frac{1}{\lambda s M} \left\| \hat{\Omega}(z) - \Omega(z) \right\|_1 \leq C; \\
\sup \max \frac{1}{\lambda M} \left\| \hat{\Omega}_j \hat{\Sigma} - \mathbf{e}_j \right\|_\infty \leq C, \quad (4.3.11)

with probability $1 - 1/d$.

The proof is deferred to Appendix C.1.2. From this theorem, we can see that (4.3.10) determines $r_{2n}$ and (4.3.11) determines $r_{3n}$ in Assumption 4.3.4. We can plug the $r_{1n}$, $r_{2n}$ and $r_{3n}$ given in (4.3.8), (4.3.10) and (4.3.11) into the condition \( \sqrt{n h (r_{2n} (r_{1n} + r_{3n}))} = O(n^{-\epsilon}) \) stated in Theorems 4.3.1, 4.3.2 and 4.3.3 to get an explicit condition for $n, h, s, d$.

**Corollary 4.3.5.** Let \( \hat{\Omega}(z) \) be the calibrated CLIME estimator with $\lambda \geq \lambda_{n,h} = C_\Sigma(h^2 + \sqrt{\log(d/h)/(nh)})$ or the CLIME estimator with $\lambda \geq M \lambda_{n,h}$. Then the Assumption 4.3.4 and all conditions on $r_{1n}$, $r_{2n}$ and $r_{3n}$ in Theorems 4.3.1, 4.3.2 and 4.3.3 can be replaced with \( \sqrt{n h \cdot s \left( h^2 + \sqrt{\log(dn)/(nh)} \right)^2} = o(1) \).

Theorem 4.3.5 implies that if the bandwidth satisfies $h \asymp (\log(dn)/n)^{1/5}$, then

\[
\sup_{z \in (0,1)} \left\| \hat{\Omega}(z) - \Omega(z) \right\|_{\max} \leq C M^2 \left( \frac{\log d + \log n}{n} \right)^{2/5} ; \\
\sup_{z \in (0,1)} \left\| \hat{\Omega}(z) - \Omega(z) \right\|_1 \leq C M s \left( \frac{\log d + \log n}{n} \right)^{2/5}
\]

with probability $1 - 1/d$. When $\log d \gg \log n$, the optimal bandwidth for selection is larger than the standard scaling $h \asymp (\log n/n)^{1/5}$ for univariate nonparametric regression (Tsybakov, 2009). This is because we need to over-regularize each entry of \( \hat{\Sigma}(z) \) to reduce the variance of entire matrix. The optimal bandwidth is also larger than the scaling for inference $h \asymp n^{-\nu}$ for some $\nu > 1/5$ in (4.3.3), since we also need to over-regularize to remove bias for inference.
Optimality of Estimation Rate

The following theorem shows that the rate in (4.3.10) is minimax optimal.

**Theorem 4.3.6.** Consider the following class of the inverse correlation matrices

\[
\mathcal{U}_s(M, \rho, L) := \{ \Omega(\cdot) \mid \Omega(z) \in \mathcal{U}_s(M, \rho) \text{ for any } z \in (0, 1), \text{ and } \Omega_{jk}(\cdot) \in \mathcal{H}(2, L) \text{ for } j, k \in [d] \},
\]

where \( \mathcal{U}_s(M, \rho) \) is defined in (4.3.2). We have the following two results on the minimax risk:

1. If \( \log(dn)/n = o(1) \), then

\[
\inf_{\Omega(z)} \sup_{\Omega(\cdot) \in \mathcal{U}_s(M, \rho, L)} \mathbb{E} \left[ \sup_{z \in (0, 1)} \| \hat{\Omega}(z) - \Omega(z) \|_{\max} \right] \geq c \left( \frac{\log d + \log n}{n} \right)^{2/5}. \tag{4.3.12}
\]

2. If \( s^2 \log(dn)/n = o(1) \) and \( s^{-v}d \leq 1 \) for some \( v > 2 \), then

\[
\inf_{\Omega(z)} \sup_{\Omega(\cdot) \in \mathcal{U}_s(M, \rho, L)} \mathbb{E} \left[ \sup_{z \in (0, 1)} \| \hat{\Omega}(z) - \Omega(z) \|_1 \right] \geq cs \left( \frac{\log d + \log n}{n} \right)^{2/5}. \tag{4.3.13}
\]

The proof is deferred to Appendix C.3. We prove it by applying Le Cam’s lemma (LeCam, 1973) and constructing a finite collection \( \Omega(\cdot) \) from the function value matrices space \( \mathcal{U}_s(M, \rho, L) \). If we take the dimension \( d = 1 \), our problem degenerates to the univariate twice differentiable function estimation. The risk on the left hand side of (4.3.12) becomes to the supreme norm between the estimated function and truth. The right hand side of (4.3.12) degenerates to \( O((\log n/n)^{2/5}) \) which matches the typical minimax rate for nonparametric regression in \( \| \cdot \|_\infty \) risk (Tsybakov, 2009). This indicates the reason why we have the power 2/5 in the rates.
Appendices
Appendix A

Proofs of Chapter 2

We prove the theorems in Chapter 2 on the properties of skip-down method.

A.1 Proofs of Results on the Skip-Down Method

In this section, we prove theoretical results on skip-down algorithm. We will prove Theorems 2.3.1, 2.4.4 and 2.4.5 and Proposition 2.2.1.

A.1.1 Proof of Theorem 2.3.1

Under (2.2.6), it suffices to prove (2.3.1) and (2.3.2) by showing that

$$\lim\inf_{n \to \infty} \inf_{\Theta \in \mathcal{U}(I^*_L, I^*_U)} \mathbb{P}_\Theta(I(\Theta) \leq \hat{I}_L) \geq 1 - \alpha.$$  \hspace{1cm} (A.1.1)

We denote the true edge set as $E^* = E(\Theta)$. We aim to bound the probability of the event $\{I(\Theta) < \hat{I}_L\}$. Assume that the output of Algorithm 1 satisfies $I(\Theta) < \hat{I}_L$ and that the algorithm has stopped at the $t = K$-th step. Under these assumptions the final rejected edge set $E_K$ satisfies $I(E_K) > I(\Theta)$. Our high-level outline to prove this theorem is to show the following two facts:
1. At least one rejected edge belongs to $C_T(E^*)$;

2. If the $\ell$-th iteration is the step before which no edge in $C_T(E^*)$ is rejected, then $C_T(E^*) \subseteq C_T(E_{\ell-1})$.

Suppose the above two claims have been proved. We denote the first edge rejected in $C_T(E^*)$ as $\bar{e}$. According to the steps in Algorithm 1, we have

$$\max_{e \in C_T(E^*)} \sqrt{n} \hat{\Theta}_e^d \geq \sqrt{n} \hat{\Theta}_{\bar{e}}^d \geq c(\alpha, C_T(E_{\ell-1})) \geq c(\alpha, C_T(E^*)) \quad (A.1.2)$$

where the first inequality is by $\bar{e} \in C_T(E^*)$, the second inequality is by the mechanism of updating the rejected set and the third inequality is by $C_T(E^*) \subseteq C_T(E_{\ell-1})$. Therefore, by (2.2.6) and the fact that $\Theta_e = 0$ for any $e \in C_T(E^*) \subseteq (E^*)^c$, we have

$$\sup_{\Theta \in U(I_L, I_U)} P_{\Theta}(I(\Theta) < \hat{I}_L) \leq \sup_{\Theta \in U(I_L, I_U)} P_{\Theta}\left( \max_{e \in C_T(E^*)} \sqrt{n} \hat{\Theta}_e^d \geq c(\alpha, C_T(E^*)) \right) \leq \alpha + o(1).$$

So in the rest part of the proof, we only need to show the above two facts.

We first prove that at least one rejected edge belongs to $C_T(E^*)$. To begin with, there must exist at least one edge in $(E^*)^c$ which is rejected, otherwise $E_K \subseteq E^*$ which implies a contradictory result that $\hat{I}_L = I(E_K) \leq I(E^*) = I(\Theta)$. To show the stronger result that $E_K \cap C_T(E^*) \neq \emptyset$, we consider the edge set $E_K \cup E^*$. Since $I(E_K) > I(E^*)$ and $I$ is monotone, we have $I(E_K \cup E^*) > I(E^*)$ as well. As we have shown that $E_K \setminus E^* \neq \emptyset$, we denote the edges in this set as $E_K \setminus E^* = \{e_1, \ldots, e_m\}$. Here the order of the edges’ indices is arbitrary. Consider the nested sequence

$$E^* \subseteq E^* \cup \{e_1\} \subseteq E^* \cup \{e_1, e_2\} \subseteq \cdots \subseteq E^* \cup \{e_1, \ldots, e_m\} = E_K \cup E^*,$$

where $I(E_K \cup E^*) > I(E^*)$. 

90
Notice that the invariant of the first edge set in the above sequence is strictly smaller than the one of the last one. The fact that $I$ is monotone implies that there exists an integer $m_0 \in [0, m]$ such that $I(E^* \cup \{e_1, \ldots, e_{m_0}\}) < I(E^* \cup \{e_1, \ldots, e_{m_0+1}\})$, where for $m_0 = 0$, we denote $E^* \cup \{e_1, \ldots, e_{m_0}\} = E^*$. In particular, we have the sequence

$$E^* \subseteq \cdots \subseteq E^* \cup \{e_1, \ldots, e_{m_0}\} \subseteq E^* \cup \{e_1, \ldots, e_{m_0+1}\} \subseteq \cdots \subseteq E_K \cup E^*. \quad \text{(A.1.3)}$$

Comparing to (2.2.3), we have $e_{m_0+1} \in C_I(E^*)$ and by the construction, $e_{m_0+1} \in E_K$. Therefore, we prove that $E_K \cap C_I(E^*) \neq \emptyset$.

Recall $\ell$ denotes the first iteration of Algorithm 1 in which edge from $C_I(E^*)$ has been rejected. Now, we are going to show that $C_I(E^*) \subseteq C_I(E_{\ell-1})$. We have $E_{\ell-1} \cap C_I(E^*) = \emptyset$ and $I(E_{\ell-1}) \leq I(E^*)$. For any $e' \in C_I(E^*)$, denote $E'$ is the set satisfying (2.2.3) for $C_I(E^*)$ such that $E' \supseteq E^*$ and $I(E') > I(E^\setminus \{e'\})$. The monotone invariant implies that $I(E' \cup E_{\ell-1}) \geq I(E')$. To prove $e' \in C_I(E_{\ell-1})$, it suffices to show that

$$I(E' \cup E_{\ell-1}) > I((E' \cup E_{\ell-1}) \setminus \{e'\}).$$

Combining $E_{\ell-1} \cap C_I(E^*) = \emptyset$ with $e' \in C_I(E^*)$, it is equivalent to show

$$I(E' \cup E_{\ell-1}) > I((E' \setminus \{e'\}) \cup (E_{\ell-1} \setminus E')).$$

This is obviously true if $E_{\ell-1} \setminus E' = \emptyset$ as $I(E') > I(E' \setminus \{e'\})$. For the case $E_{\ell-1} \setminus E' \neq \emptyset$, we prove by contradiction. If $I((E' \setminus \{e'\}) \cup (E_{\ell-1} \setminus E')) \geq I(E' \cup E_{\ell-1})$, we have

$$E' \setminus \{e'\} \subseteq (E' \setminus \{e'\}) \cup E_{\ell-1} \text{ and } I(E' \setminus \{e'\}) < I(E' \cup E_{\ell-1}) \leq I((E' \setminus \{e'\}) \cup E_{\ell-1}).$$
Similar to (A.1.3), we denote \( E_{\ell-1} \backslash E' = \{ e'_1, \ldots, e'_m \} \) and there exists an integer \( m'_0 \in [0, m'] \) such that
\[
(E' \backslash \{ e' \}) \cup \{ e'_1, \ldots, e'_{m'_0} \} \subseteq (E' \backslash \{ e' \}) \cup \{ e'_1, \ldots, e'_{m'_0+1} \}.
\]
(A.1.4)

Since \( e'_{m'_0+1} \in E_{\ell-1} \backslash E' \) and \( E^* \subseteq E' \backslash \{ e' \} \), we have \( e'_{m'_0+1} \notin E^* \). Combining with (A.1.4), the definition in (2.2.3) is satisfied for \( e'_{m'_0+1} \) and thus \( e'_{m'_0+1} \notin C_I(E^*) \). This contradicts the fact that \( E_{\ell-1} \cap C_I(E^*) = \emptyset \). In summary, we show that \( I(T\mu(\Theta)) = I(E_{\Sigma}(\Theta)) \) when \( \mu = C\sqrt{\log d/n} \).

A.1.2 Proof of Theorems 2.4.4 and 2.4.5

We first define a few notations before presenting the proof. For a positive number \( \mu > 0 \), recall that the thresholded matrix \( [T\mu(\Theta)]_{jk} = \Theta_{jk} \mathbb{1}\{|\Theta_{jk}| \geq \mu\} \) for all \( 1 \leq j, k \leq d \).

We define \( T\mu(\Theta) \) in order to show the adaptivity of our confidence interval. We can see \( I(T\mu(\Theta)) = I(E_{\Sigma}(\Theta)) \) when \( \mu = C\sqrt{\log d/n} \).

We also define the parameter space
\[
U_s(I^*_L, I^*_U; \theta, \mu) = \left\{ \Theta \in U_s \left| I(\Theta) \leq I^*_U, \min_{e \in E(\Theta)} |\Theta_e| \geq \theta, I(T\mu(\Theta)) \geq I^*_L \right. \right\}.
\]
(A.1.5)

The proof of both Theorems 2.4.4 and 2.4.1 can be directly derived from the following lemma.

**Lemma A.1.1.** Suppose \( \Theta \in U_s \) and \( (\log(dn))^6/n + s^2(\log dn)^4/n = o(1) \). For any monotone invariant \( I \) with range \([I^*_L, I^*_U]\), there exists a positive constant \( C \) such that if \( \mu \geq C\sqrt{\log d/n} \), for any \( \alpha \in (0, 1) \) and \( \theta > 0 \),
\[
\sup_{\Theta \in U_s(I^*_L, I^*_U; \theta)} \mathbb{P}_{\Theta}(I(T\mu(\Theta)) \leq \hat{I}_L) = 1 - O(1/d^3).
\]
(A.1.6)
We defer the proof of Lemma A.1.1 to Section A.1.3 in the appendix. We first give the proof of Theorem 2.4.4. By the construction of \([\hat{I}_L, \hat{I}_U]\) in Algorithm 1, we have

\[
\mathbb{E}_\Theta [I_U^* - \hat{I}_L] \leq \mathbb{E}_\Theta [I_U^* - \mathcal{I}(\mathcal{T}_\mu(\Theta)) \lor I_L^* \mid \mathcal{I}(\mathcal{T}_\mu(\Theta)) \leq \hat{I}_L] \mathbb{P}(\mathcal{I}(\mathcal{T}_\mu(\Theta)) \leq \hat{I}_L) + (I_U^* - I_L^*) \mathbb{P}(\mathcal{I}(\mathcal{T}_\mu(\Theta)) > \hat{I}_L)
\]

\[
\leq [I_U^* - \mathcal{I}(\mathcal{T}_\mu(\Theta)) \lor I_L^* + 1] \mathbb{P}(\mathcal{I}(\mathcal{T}_\mu(\Theta)) \leq \hat{I}_L) + (I_U^* - I_L^*) \mathbb{P}(\mathcal{I}(\mathcal{T}_\mu(\Theta)) > \hat{I}_L).
\]

Therefore, by (A.1.6), we have

\[
\sup_{\Theta \in \mathcal{U}(I_L, I_U; \theta)} \mathbb{E}_\Theta [\hat{I}_U - \hat{I}_L] \leq 1 + o(1) + \frac{I_U^* - I_L^*}{I_U^* - \mathcal{I}(\mathcal{T}_\mu(\Theta)) + 1} \cdot O(1/d^3).
\]

Since \(I_U^* - I_L^* = O(d^2)\), we have the second term on the right hand side above to be \(o(1)\).

Therefore, we finish the proof of (2.4.10).

Similarly, we can also have

\[
\mathbb{E}_\Theta [\mathcal{I}(\Theta) - \hat{I}_L] \leq [\mathcal{I}(\Theta) - \mathcal{I}(\mathcal{T}_\mu(\Theta)) \lor I_L^*] \mathbb{P}(\mathcal{I}(\mathcal{T}_\mu(\Theta)) \leq \hat{I}_L) + [\mathcal{I}(\Theta) - I_L^*] \mathbb{P}(\hat{I}_L > \mathcal{I}(\mathcal{T}_\mu(\Theta))).
\]

By (A.1.6), following the same argument as the proof of (2.4.10) above, we prove (2.4.11).

Theorem 2.4.1 is a direct corollary of (A.1.6). For a monotone property \(\mathcal{P}\), we have the range \(I_L^* = 0\) and \(I_U^* = 1\). Moreover, if \(\Theta \in \mathcal{G}_1(\mu; \mathcal{P})\) defined in (2.4.1), we have \(\mathcal{T}_\mu(\Theta) = \Theta\)

Since Algorithm 2 is derived from Algorithm 1 in the sense that the test \(\psi_\alpha = 1\) is equivalent to a \(1 - 2\alpha\) confidence lower side \(\hat{I}_L = 1\). Thus, under the alternative that \(\mathcal{P}(\Theta) = 1\), we have

\[
\{\psi_\alpha = 1\} = \{\mathcal{P}(\Theta) \leq \hat{I}_L\} = \{\mathcal{P}(\mathcal{T}_\mu(\Theta)) \leq \hat{I}_L\}.
\]

Therefore, (2.4.2) can be directly derived from (A.1.6).
A.1.3 Proof of Lemma A.1.1

We first consider a special case that $I(T_\mu(\Theta)) = I(\emptyset)$. Since $I$ is monotone, according to Algorithm 1, $\hat{I}_L \geq I(T_\mu(\Theta)) = I(\emptyset)$ almost surely and (A.1.6) is trivially true.

Therefore, in the following of the proof, we consider the case $I(T_\mu(\Theta)) > I(\emptyset)$. This implies that there exists a non-empty edge set $E_0'$ such that

$$E_0' \subseteq E(T_\mu(\Theta)), I(E_0') = I(T_\mu(\Theta))$$

and

$$\min_{e \in E_0'} |\Theta_e| > C \sqrt{\log d/n},$$

(A.1.7)

where the constant $C$ is determined later. By $I(E_0') > I(\emptyset)$, similar to (A.1.4), we claim that $I(E_0' \cap C_\mu(\emptyset)) = I(E_0')$. To prove this claim, we find a subgraph $E_0'' \subset E_0'$ such that $I(E_0'') = I(E_0')$ and for any $\tilde{E} \subset E_0''$, we have $I(\tilde{E}) < I(E_0'')$. Such graph $E_0''$ can be constructed by deleting edges from $E_0'$ and making the invariant equal to $I(E_0')$ until it is impossible to further delete edges without reducing the value of the invariant. By Definition 2.2.1, we have $E_0'' \subseteq C_\mu(\emptyset)$ which implies $E_0'' \subseteq E_0' \cap C_\mu(\emptyset)$. As $I(E_0'') = I(E_0')$, by monotone property, we prove the claim that $I(E_0' \cap C_\mu(\emptyset)) = I(E_0')$. Consider the following event

$$\mathcal{E}_1 = \left\{ \min_{e \in E_0' \cap C_\mu(\emptyset)} \sqrt{n} |\hat{\Theta}_e^d| > c(\alpha, C_I(\emptyset)) \right\}.$$

According to Algorithm 1, the rejected set in the first iteration is

$$E_1 = \left\{ e \in C_\mu(\emptyset) : \sqrt{n} |\hat{\Theta}_e^d| > c(\alpha, C_I(\emptyset)) \right\}.$$

Under the event $\mathcal{E}_1$, we have $E_0' \cap C_\mu(\emptyset) \subseteq E_1$ and since $I(E_0' \cap C_\mu(\emptyset)) = I(E_0')$, we have $I(E_1) = I(E_0')$, which makes $\hat{I}_L \geq I(E_0') = I(T_\mu(\Theta))$. Therefore, we have

$$\mathbb{P}(I(T_\mu(\Theta)) \leq \hat{I}_L) \geq \mathbb{P}(\mathcal{E}_1).$$

(A.1.8)
We will bound $P(\mathcal{E}_1)$ next. We have
\[
P(\mathcal{E}_1) \geq P\left( \min_{e \in E_0} |\Theta_e| > 2c(\alpha, C_I(\emptyset)) \quad \text{and} \quad \max_{e \in V \times V} |\hat{\Theta}_e^d - \Theta_e| \leq c(\alpha, C_I(\emptyset)) \right). \tag{A.1.9}
\]

We next estimate the rate of $c(\alpha, C_I(\emptyset))$. We have the following lemma.

**Lemma A.1.2.** For large enough $n$ and $d$ we have:
\[
\sup_{\Theta \in \mathcal{M}(s)} P\left( \max_{j,k \in [d]} \sqrt{n}|(\hat{\Theta}_j^d - \Theta_{jk}) + \Theta_j^s(T_\Sigma - \Sigma )_k| > \xi_1 \right) < \xi_2, \tag{A.1.10}
\]
where $\xi_1 = \Xi_1 s \log d/\sqrt{n}$, $\xi_2 = 2/d$, where $\Xi_1$ is an absolute constant depending solely on $K, L$ and $C$.

We leave the proof of this lemma to the end of this section and continue with the main proof. Since $\Theta \in \mathcal{U}_s$ and $\|\Theta\|_2 \leq \rho$, we have $\|\Theta_j^T X X^T \Theta_k\|_{\psi_1} \leq C_2 \rho^2$. By the maximal inequality (see Lemma 2.2.2 in van der Vaart and Wellner (1996)), we have for some constant $C_3 > 0$
\[
P\left( \max_{j,k \in [d]} |\Theta_j^T (\hat{\Sigma} \Theta_k - e_k)| > C_3 \rho^2 \sqrt{\log d / n} \right)
= P\left( \max_{j,k \in [d]} \frac{1}{n} \sum_{i=1}^n (\Theta_j^T \hat{X}_i X_i^T \Theta_k - E[\Theta_j^T \hat{X}_i X_i^T \Theta_k]) \right) > C_3 \rho^2 \sqrt{\log d / n} \leq 1/d^3.
\]

Let $C_0 = C_1 + C_3$ and we have
\[
P\left( \max_{j,k \in [d]} |\hat{\Theta}_j^d - \Theta_{jk}| > C_0 \sqrt{\log d / n} \right) < \frac{2}{d^3}. \tag{A.1.11}
\]

Applying (2.2.6) and (A.2.14), for any fixed $\alpha \in (0,1)$ and sufficiently large $d, n$ such that $P(T_{V \times V} > c(\alpha, V \times V)) > \alpha$ and $2/d^3 \leq \alpha$, we have
\[
c(\alpha, C_I(\emptyset)) \leq c(\alpha, V \times V) \leq C_0 \sqrt{\log d / n}.
\]
We set the constant \( C = 2C_0 \) in \( \mu \geq C \sqrt{\log d/n} \) and it follows from (A.1.7) and (A.2.14) that

\[
\min_{e \in E_0^* \cap C_2(\emptyset)} |\Theta_e| > 2C_0 \sqrt{\log d/n} > 2c(\alpha, C_I(\emptyset)) \quad \text{and} \quad \mathbb{P}\left( \max_{e \in V \times V} |\hat{\Theta}_e^d - \Theta_e| \leq c(\alpha, C_I(\emptyset)) \right) > 1 - 2/d^3.
\]

Combining the above inequalities with (A.1.8) and (A.1.9), we have

\[
\mathbb{P}(T(\mu(\Theta)) \leq \tilde{T}_L) \geq \mathbb{P}(E_1) > 1 - 2/d^3.
\]

Since the right hand side of the above inequality is universal for any \( \Theta \in U_T(I_L^*, I_U^*; \theta) \), we complete the proof of the lemma.

**Proof of Lemma A.1.2.** By elementary algebra we obtain the following representation:

\[
(\Theta_{jk}^d - \Theta_{jk}) + \Theta_{j}^T (\hat{\Sigma} - \Sigma) \Theta_k
= - \Theta_{j}^T \hat{\Sigma}_{\backslash j} \left( \hat{\Theta}_{\backslash j,k} - \Theta_{\backslash j,k} \right) + \left( \hat{\Theta}_j - \Theta_j^* \right)^T (\hat{\Sigma} - \Sigma) \Theta_k
\]

\[
= \left[ 1 - \Theta_{j}^T \hat{\Sigma}_{\backslash j} \right] \Theta_k \left( \frac{1}{\Theta_{j}^T \hat{\Sigma}_{\backslash j}} - 1 \right),
\]

where indexing with \( \backslash j \) means dropping the corresponding column or element from the matrix or vector respectively. We deal with the terms \( I_{1}^{jk} \) first. By the triangle inequality followed by Hölder’s inequality we have:

\[
\max_{j,k \in [d]} |I_{1}^{jk}| \leq \max_{j,k \in [d]} \{ |\hat{\Theta}_j^T \hat{\Sigma}_{\backslash j}|^{-1} (\|\hat{\Theta}_j^T \hat{\Sigma}_{\backslash j}\|_{\infty} \|\hat{\Theta}_{\backslash j,k} - \Theta_{\backslash j,k}\|_1
\]

\[
+ \|\hat{\Theta}_j - \Theta_j^*\|_1 \|\hat{\Sigma} - \Sigma\|_{\max} \|\Theta_k\|_1)\}
\]

96
Since it is assumed that \( \log d/n = o(1) \), take \( d \) and \( n \) large enough so that \( \log d/n \leq 1/(4K^2) \).

Let \( E \) be the event where (A.2.1) hold. Then on \( E \) we have:

\[
\max_{j \in [d]} |\hat{\Theta}_j^T \hat{\Sigma}_j - 1| \leq K \sqrt{\log d/n}.
\]

This implies that on the same event \( \sup_{j \in [d]} |\hat{\Theta}_j^T \hat{\Sigma}_j|^{-1} \leq 2 \). Continuing our bounds on the event \( E \) we have:

\[
\|\hat{\Theta}_j^T \hat{\Sigma}_*\|_\infty \|\hat{\Theta}_j - \Theta_{j,k}\|_1 \leq K^2 s \log d/n = o(n^{-1/2}).
\]

Furthermore on the intersection of \( E \), which holds with probability no less than \( 1 - 2/d \), we have:

\[
\|\hat{\Theta}_j - \Theta_j\|_1 \|\hat{\Sigma} - \Sigma\|_{\text{max}} \|\Theta_k\|_1 \leq RC^2 KL s \log d/n = o(n^{-1/2}).
\]

Combining the last four bounds we conclude that \( \max_{j,k \in [d]} |I_{1j}^k| = o_k(n^{-1/2}) \). Next we handle the terms \( |I_{2j}^k| \). We have:

\[
\max_{j,k \in [d]} |I_{2j}^k| \leq \max_{j,k \in [d]} \left\{ |\Theta_j^* (\hat{\Sigma} - \Sigma ) \Theta_k| \left\| \frac{1}{\hat{\Theta}_j^T \hat{\Sigma}_j} - 1 \right\| \right\}.
\]

Clearly on the event \( E \) under the assumption \( \log d/n \leq 1/(4K^2) \), we have that:

\[
\max_{j \in [d]} \left| \frac{1}{\hat{\Theta}_j^T \hat{\Sigma}_j} - 1 \right| \leq 2K \sqrt{\log d/n}. \tag{A.1.13}
\]

Next we consider the random variables \( \Theta_j^* X^{\otimes 2} \Theta_k^* \) for all \( j, k \in [d] \). By Theorem A.2.1 we have:

\[
\|\Theta_j^* X^{\otimes 2} \Theta_k^*\|_{\psi_1} \leq 2\|\Theta_j^* X\|_{\psi_2} \|X^T \Theta_k^*\|_{\psi_2} \leq 2 \sup_{j \in [d]} \|\Theta_j^*\|_{\psi_2}^2 C^2 \leq 2C^4,
\]

with the last inequality following by the fact that \( \sup_{j \in [d]} \|\Theta_j^*\|_{\psi_2}^2 \leq \|\Theta\|_2^2 \leq C^2 \) since \( \Theta \in \mathcal{M}(s) \). Clearly then \( \|\Theta_j^* X^{\otimes 2} \Theta_k^* - \Theta_{jk}\|_{\psi_1} \leq 4C^4 \). Finally by the union bound
and Proposition 5.16 (Vershynin, 2012), one concludes that there exists an absolute constant \( \tilde{C} \), such that when \( \sqrt{\log d/n} \leq 4C^4 \), we have:

\[
\max_{j,k \in [d]} |\Theta_j^T (\tilde{\Sigma} - \Sigma) \Theta_k| \leq 4\tilde{C}C^4 \sqrt{\log d/n},
\]

with probability at least \( 1 - 2/d \). Combining this bound with (A.1.13) and our conditions we conclude that \( \max_{j,k \in [d]} |I_{jk}^2| = o_p(n^{-1/2}) \). This completes the proof.

A.1.4 Proof of Proposition 2.2.1

We first prove a general result on when the critical edge set is empty. We claim that given any monotone property \( \mathcal{P} \), if \( \mathcal{P}(G_0) = 1 \) then \( \mathcal{C}_\mathcal{P}(E_0) = \emptyset \). This is straightforward from Definition 2.2.1. If \( \mathcal{P}(G_0) = 1 \), by the monotonicity of \( \mathcal{P} \), for any \( E' \supseteq E_0 \), we have \( \mathcal{P}(E') = 1 \). Therefore, there is no edge \( e \in E_0^{\emptyset} \) such that \( \mathcal{P}(E' \setminus \{e\}) = 0 \). This implies that \( \mathcal{C}_\mathcal{P}(E_0) = \emptyset \). Therefore, we prove the four results in Proposition 2.2.1 when the critical edge sets are empty. In the following part of the proof, we will discuss four properties case by case. An illustration of the proof are shown in Figure A.1.

- Connected subgraphs. Now we prove the results on the connected subgraphs. We can consider the case \( \mathcal{P}_{\text{Conn},-k}(G_0) = 0 \). We choose the \( k' \) nodes, each of which from one of the connected subgraphs \( \{G_{0\ell} = (V_{0\ell}, E_{0\ell})\}_{\ell=1}^{k'} \). Namely, we choose arbitrary \( v_{\ell} \in V_{0\ell} \) for \( 1 \leq \ell \leq k' \). For any edge \( (u, v) \in E_0 \) such that \( j \in V_{0\ell}, k \in V_{0\ell'}, \ell \neq \ell' \). We arbitrarily select

![Figure A.1: Visualization of the proof of Proposition 2.2.1.](image)
$k' - k - 1$ edges from the set $\{(v_s, v_{s'})|(s, s') \neq (\ell, \ell') \text{ or } (s, s') \neq (\ell', \ell)\}$ and denote this set $\widetilde{E}$. We construct the set

$$E' = E_0 \cup \{e\} \cup \widetilde{E}.$$ 

We can find an illustration of the construction in Figure A.1(a). Notice that $G' = (V, E')$ has $k - 1$ connected subgraphs and thus $\mathcal{P}_{\text{Conn}, -k}(E') = 1$. We can also check that $E' \setminus \{e\} = E_0 \cup \widetilde{E}$ with has $k$ connected subgraphs. By (2.2.3), we have $(u, v) \in \mathcal{C}_{\mathcal{P}_{\text{Conn}, -k}}(E_0)$ and thus

$$\mathcal{C}_{\mathcal{P}_{\text{Conn}, -k}}(E_0) \supseteq \{(u, v) \in E_0 \mid j \in V_0\ell, k \in V_0\ell', \ell \neq \ell'\}.$$ 

To prove the other direction, if $e \not\in \{(u, v) \in E_0 \mid u \in V_0\ell, v \in V_0\ell'\}$, we must have $e \in E_0\ell$ for some $\ell$. For any $E' \supseteq E$ such that $\mathcal{P}_{\text{Conn}, -k}(E') = 1$, since $e \in E_0\ell$, $E' \setminus \{e\}$ has the same number of connected subgraphs, therefore $e \not\in \mathcal{C}_{\mathcal{P}_{\text{Conn}, -k}}(E_0)$. In summary, we prove (2.2.9).

- Maximum degree. We next move to the critical edge set of maximum degree. By (2.2.3), we have $\mathcal{C}_{\mathcal{P}_{\text{Deg}, k}}(E_0) \subseteq E_0^c$. It suffices to prove another direction. Denote the maximum degree of $G_0$ as $k'$. For any $(u, v) \in E_0^c$, since the maximum degree $k' \leq k$, we can select arbitrary $k' - k$ edges from $\{(u, u') \in E_0^c | u' \neq v\}$ and denote the set containing these edges as $\widetilde{E}$. See an illustration of the proof in Figure A.1(b). It is easy to check that $E' = E_0 \cup \{(u, v)\} \cup \widetilde{E}$ has $\mathcal{P}_{\text{Deg}, k}(E') = 1$ and $\mathcal{P}_{\text{Deg}, k}(E' \setminus \{(u, v)\}) = 0$. Therefore, we have $\mathcal{C}_{\mathcal{P}_{\text{Deg}, k}}(E_0) = E_0^c$.

- Acyclic. Now we turn to show the critical edge set for the cyclic property when $G_0 = (V, E_0)$ is a forest. Similar to the preceding part, it suffices to show $E_0^c \subseteq \mathcal{C}_{\mathcal{P}_{\text{Cycle}}}(E_0)$. For any $(u, v) \in E_0^c$, if $E_0 \cup \{(u, v)\}$ forms a cycle, then $(u, v) \in \mathcal{C}_{\mathcal{P}_{\text{Cycle}}}(E_0)$ by the definition in (2.2.3). If $E_0 \cup \{(u, v)\}$ is still a forest, we apply the Kruskal’s algorithm (Kruskal, 1956) with the initial input $E_0 \cup \{(u, v)\}$ and the weights of all edges are 1. Specifically, the procedure in described in Algorithm 6 and in the algorithm we can choose arbitrary order of adding edges in $(E_0 \cup \{(u, v)\})^c$ in the for loop since the edge weights are the same. We denote the output edge set of Algorithm 6 as $\widetilde{E}$ and $(V, \widetilde{E})$ is a tree by the property of Kruskal’s
algorithm. We illustrate the graph and the following of the construction in Figure A.1(c). We start to construct $E'$ to satisfy (2.2.3). Since $(V, E')$ is a tree, there are only two cases: (1) neither of $u, v$ is a leaf; and (2) only one of $u, v$ is a leaf. We first consider the case that neither of $u, v$ is a leaf in $\tilde{E}$. Then, in the graph $(V, \tilde{E})$, we can choose $u'$ as any neighbor of $u$ but $v$ and choose $v'$ as any neighbor of $v$ but $u$. This is feasible since neither of $u, v$ is a leaf. We can see that $u' \neq v'$, otherwise the triangle $u \rightarrow v \rightarrow u' \rightarrow u$ forms a loop. Let $E' = \tilde{E} \cup \{(u', v')\}$ and it has a loop $u \rightarrow v \rightarrow v' \rightarrow u$ and has only this loop (as $\tilde{E}$ is a tree and adding one edge to it will only form one loop). Therefore, if we delete the edge $(u, v)$ from the loop, the graph $(V, E' \setminus \{(u, v)\})$ has no loop. Therefore, we check (2.2.3) that $E' = \tilde{E} \cup \{(u, u')\}$ has $P_{\text{Cycle}}(E') = 1$ and $P_{\text{Cycle}}(E' \setminus \{(u, u')\}) = 0$. For the second case that only one of $u, v$ is a leaf (we assume the leaf is $u$), the proof is similar to the first one. In the graph $(V, \tilde{E})$, we find $v'$ as any neighbor of $v$ but $u$. We can check as the first case that $E' = \tilde{E} \cup \{(u, v')\}$ has $P_{\text{Cycle}}(E') = 1$ and $P_{\text{Cycle}}(E' \setminus \{(u, v')\}) = 0$. We can now conclude that $E_0^* = C_{P_{\text{Cycle}}}(E_0)$.

**Algorithm 6** Kruskal’s algorithm for the proof of Proposition 2.2.1

Input: $E^{(0)} = E_0 \cup \{(u, v)\}, t = 0.$

for $e \in (E_0 \cup \{(u, v)\})^t$ do

$t \leftarrow t + 1$;

if $\{e\} \cup E^{(t)}$ does not contain a cycle then 

$E^{(t)} \leftarrow E^{(t-1)} \cup \{e\}$

end if

end for

Output: $E^{(t)}$.

- Singletons. We finally discuss the critical edge set for isolated nodes. Denote the number of isolated nodes $|V_{\text{Sig}}| = k' \geq k$. For any $(u, v) \in E_0^*$ such that $u \in V_{\text{Sig}}$ or $v \in V_{\text{Sig}}$, we can check that at least one of $u$ and $v$ is an isolated node. If both $u, v$ are isolated nodes, this implies that $k' > k + 1$ and we can select arbitrary $k' - k - 1$ nodes from $V_{\text{Sig}} \setminus \{u, v\}$ if only one of $u, v$ is a isolated node, we select arbitrary $k' - k - 2$ nodes from $V_{\text{Sig}} \setminus \{u, v\}$. Denote the set of nodes selected as $\tilde{V}$ and define $\tilde{E} = \{(u', v') \in E_0^*|u', v' \in \tilde{V}\}$. We can check
that $E' = E_0 \cup \{(u, v)\} \cup \tilde{E}$ has $P_{\text{Sig}, -k}(E') = 1$ and $P_{\text{Sig}, -k}(E' \setminus \{(u,v)\}) = 0$. Therefore, we have $C_{P_{\text{Sig}, -k}}(E_0) \supseteq \{(u, v) \in E_0^c \mid u \in V_{\text{Sig}} \text{ or } v \in V_{\text{Sig}}\}$. The construction is illustrated in Figure A.1(d). On the other hand, let the edge $(u,v)$ satisfy $u \not\in V_{\text{Sig}}$ and $v \not\in V_{\text{Sig}}$. For any $E' \supseteq E_0$ and $P_{\text{Sig}, -k}(E') = 1$, as $E' \setminus \{(u,v)\}$ does not include new isolated nodes, $P_{\text{Sig}, -k}(E' \setminus \{(u,v)\}) = 1$. This implies that $C_{P_{\text{Sig}, -k}}(E_0) = \{(u, v) \in E_0^c \mid u \in V_{\text{Sig}} \text{ or } v \in V_{\text{Sig}}\}.

### A.2 Proof of Lemma 2.2.1

In order to prove Lemma 2.2.1, we need a preliminary results on the estimation rates of CLIME estimator in (4.1.1). Cai et al. (2011) show the following theorem.

**Theorem A.2.1.** Suppose $\Omega \in \mathcal{U}(M, s)$ and we choose $\lambda \geq CM \sqrt{\log d/n}$ in (4.1.1). With probability no smaller than $1 - c/d$, we have

$$
\|\hat{\Sigma} - \Sigma\|_{\max} \leq C \sqrt{\frac{\log d}{n}}, \|\hat{\Theta} \hat{\Sigma} - I\|_{\max} \leq CM \sqrt{\frac{\log d}{n}} \text{ and } \|\hat{\Theta} - \Theta\|_1 \leq CM \sqrt{\frac{s^2 \log d}{n}},
$$

(A.2.1)

where the constant $C$ is a universal constant only depends on $C$ in (4.3.2).

**Remark 4.** The first inequality in (A.2.1) follows from Eq. (26), Cai et al. (2011), the second inequality follows from the constraint in (4.1.1) and the third follows from Theorem 6 of Cai et al. (2011).

Given a random variable $Z$, we define the $\psi_\ell$-norm for $\ell \geq 1$ as $\|Z\|_{\psi_\ell} = \sup_{p \geq 1} p^{-1/\ell} (\mathbb{E}|Z|^p)^{1/p}$. We also need to control the $\psi_\ell$-norm of $X$ and the lower bound of the variance of the debiased estimator in the following lemma.
Lemma A.2.1. There exist universal constants $c, C$ only depending on $C$ in in (4.3.2) such that

$$
\sup_{\|v\|_2 = 1} \|v^T \Sigma^{-1/2} X\|_{\psi_2} \leq C \quad \text{and} \quad \min_{j,k \in [d]} \mathbb{E}[\Theta_j^T (XX^T - \Sigma) \Theta_k]^2 \geq c.
$$

(A.2.2)

Proof. Since $v^T \Sigma^{-1/2} X \sim N(0, 1)$ for any $\|v\|_2 = 1$, the first inequality in (A.2.2) is straightforward.

We now study the second inequality of (A.2.2). Since $\mathbb{E}[\Theta_j^T (XX^T - \Sigma) \Theta_k]^2 = \text{Var}(\Theta_j^T XX^T \Theta_k)$, we calculate a more general form as follows. We apply Isserlis’ theorem (Isserlis, 1918) to calculate the moment of Gaussian. For any vectors $u, v \in \mathbb{R}^d$, Isserlis’ theorem gives us

$$
\text{Var}(u^T XX^T v) = \mathbb{E}[(u^T X)^2 (v^T X)^2] - (\mathbb{E}[u^T X v^T X])^2
$$

$$
= \mathbb{E}[(u^T X)^2] \mathbb{E}[(v^T X)^2] + (\mathbb{E}[u^T X v^T X])^2
$$

$$
= (u^T \Sigma u^T)(v^T \Sigma v^T) + (u^T \Sigma v^T)^2.
$$

(A.2.3)

Therefore, we have

$$
\mathbb{E}[\Theta_j^T (XX^T - \Sigma) \Theta_k]^2 = (\Theta_j^T \Sigma \Theta_j^T)(\Theta_k^T \Sigma \Theta_k^T) + (\Theta_j^T \Sigma \Theta_k^T)^2 = \Theta_{jj} \Theta_{kk} + \Theta_{jk}^2 \geq 1/C^2.
$$

where the last inequality is due to $\lambda_{\text{min}}(\Theta) \geq 1/C$. 

Now we start the main proof of Lemma 2.2.1.

In order to approximate the statistic $T_E := \max_{(j,k) \in E} \sqrt{n}(\hat{\Theta}_{jk} - \Theta_{jk})$ by the multiplier bootstrap process

$$
T^B := \max_{(j,k) \in E} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\Theta}_j^T X_i X_i^T \hat{\Theta}_k - e_k) \xi_i.
$$

102
we define two intermediate processes

\[
T_0 = \max_{(j,k) \in E} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Theta_j^T (X_i X_i^T \Theta_k - e_k) \quad \text{and} \quad T_0^B = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Theta_j^T (X_i X_i^T \Theta_k - e_k) \xi_i.
\]

The strategy of proving this lemma is to check the three conditions in Corollary 3.1 of Chernozhukov et al. (2013) as follows.

1. \( \min_{j,k} \mathbb{E}[(\Theta_j^T (XX^T \Theta_k - e_k))^2] > c \) and \( \max_{j,k \in [d]} \|\Theta_j^T (XX^T \Theta_k - e_k)\|_{\psi} \leq C \) for some positive constants \( c \) and \( C \);

2. \( \mathbb{P}(\{|T_E - T_0| > \zeta_1\}) < \zeta_2 \);

3. And \( \mathbb{P}(\mathbb{P}(\{|T^B - T_0^B| > \zeta_1 \mid \{X_i\}_{i=1}^{n}\} > \zeta_2) < \zeta_2 \) hold for \( \zeta_1 \sqrt{\log d} + \zeta_2 = o(1) \).

Notice in Chernozhukov et al. (2013), the original conditions require the last scaling to be \( \zeta_1 \sqrt{\log d} + \zeta_2 = o(n^{-c_1}) \). This is because they have stronger result that

\[
\|\mathbb{P}(T_E > c(\alpha, E)) - \alpha\| = O(n^{-c_1}).
\]

Since we do not emphasize on the polynomial decaying in our result, we only need \( \zeta_1 \sqrt{\log d} + \zeta_2 = o(1) \) and same for the scaling condition \( (\log (dn))^7/n = o(1) \).

First, we check the first condition. Lemma A.2.1 directly implies the first part. By the second condition in (A.2.2), we have \( \|X_jX_k - \mathbb{E}[X_jX_k]\|_{\psi} \leq C \). By the definition of the \( \psi \)-norms, we have

\[
\max_{j,k \in [d]} \|\Theta_j^T (X_i X_i^T \Theta_k - e_k)\|_{\psi} \leq C^2 \|\Theta_j^T (X_jX_k - \mathbb{E}[X_jX_k])\|_{\psi} \leq C^2 \sup_{\|v\|_2=1} \|v^T XX^T v - \mathbb{E}[v^T XX^T v]\|_{\psi} = O(1). \quad (A.2.4)
\]

103
Second, we check the next condition by bounding the difference $|T_0 - T_E|$. We denote
\[ \tilde{\Theta}_k = (\tilde{\Theta}_{k1}, \ldots, \tilde{\Theta}_{k(j-1)}, \Theta_{kj}, \tilde{\Theta}_{k(j+1)}, \ldots, \tilde{\Theta}_{kd})^T \in \mathbb{R}^d. \]

The high-level strategy is to decompose the following quantity.
\[ \sqrt{n}(\hat{\Theta}_j^d - \Theta_{jk}) = -\sqrt{n} \cdot \frac{\hat{\Theta}_j^T(\tilde{\Sigma}\tilde{\Theta}_k - e_k)}{\hat{\Theta}_j^T \tilde{\Sigma}_j}. \]

In what follows, we quantify $\sqrt{n}\hat{\Theta}_j^T(\tilde{\Sigma}\tilde{\Theta}_k - e_k^T)$. We first have
\[ \sqrt{n} \cdot \hat{\Theta}_j^T(\tilde{\Sigma}\tilde{\Theta}_k - e_k^T) = \sqrt{n} \cdot \hat{\Theta}_j^T(\tilde{\Sigma}\tilde{\Theta}_k - e_k^T) + \sqrt{n} \cdot \hat{\Theta}_j^T \tilde{\Sigma}(\tilde{\Theta}_k - \Theta_k). \quad (A.2.5) \]

We can bound $I_1$ by
\[ I_1 = \sqrt{n} \cdot \hat{\Theta}_j^T(\tilde{\Sigma}\tilde{\Theta}_k - e_k^T) + \sqrt{n} \cdot (\hat{\Theta}_j^T - \Theta_j^T)(\tilde{\Sigma}\tilde{\Theta}_k - e_k). \quad (A.2.6) \]

Here $I_{11} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Theta_j^T(X_iX_i^T(\tilde{\Theta}_k - e_k))$ which is same as what is inside of the maximum of $T_0$.

We next bound $I_{12}$ by Hölder inequality,
\[ |I_{12}| = \sqrt{n} \cdot (\hat{\Theta}_j - \Theta_j)^T(\tilde{\Sigma} - \Sigma)\Theta_k \leq \sqrt{n} \cdot \|\hat{\Theta}_j - \Theta_j\|_1\|\tilde{\Sigma} - \Sigma\|_{\text{max}}\|\Theta_k\|_1. \quad (A.2.7) \]

According to Theorem A.2.1, (A.2.7) yields that with probability $1 - 1/d$,
\[ \max_{j, k \in [d]} |I_{12}| \lesssim M^2 s \log d \frac{\log d}{\sqrt{n}}. \quad (A.2.8) \]

104
We finally bound $I_2$ by Hölder inequality and Theorem A.2.1 such that with probability $1 - 1/d,$

$$\max_{j,k \in [d]} |I_2| \leq \sqrt{n} \cdot \max_{j,k \in [d]} \|\hat{\Theta}_j^T \hat{\Sigma} - \Theta_k\|_1 \lesssim M^2 s \log d \sqrt{n}.$$  (A.2.9)

Therefore, we conclude that by (C.7.12) and (C.7.13), we have with probability $1 - 1/d,$

$$\sup_{j,k \in [d]} \sqrt{n} \cdot |\hat{\Theta}_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k^T) - \Theta_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k)| \lesssim M^2 s \log d \sqrt{n}.$$  (A.2.10)

Theorem A.2.1 also gives

$$\max_{j \in [d]} |\hat{\Theta}_j^T \hat{\Sigma}_j - 1| \leq \max_{j \in [d]} \|\hat{\Theta}_j^T \hat{\Sigma} - e_j\|_\infty \lesssim M \sqrt{\frac{\log d}{n}}.$$  (A.2.11)

Combining (A.2.5), (A.2.6) with (A.2.10) and (A.2.11), for sufficiently large $d, n$, we have with probability $1 - 1/d,$

$$|T_E - T_0| \leq \max_{(j,k) \in E} \sqrt{n} \left|\hat{\Theta}_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k) - \Theta_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k)\right| \leq \max_{(j,k) \in E} \sqrt{n} \left|\hat{\Theta}_j^T (\hat{\Sigma} - \Sigma) \Theta_k\right| + 2 \max_{(j,k) \in E} \left|\hat{\Theta}_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k) - \Theta_j^T (\hat{\Sigma} \hat{\Theta}_k - e_k)\right| \leq 2M \sqrt{n} \max_{j \in [d]} \left|\hat{\Theta}_j^T \hat{\Sigma}_j - 1\right| \lesssim M^2 s \log d \sqrt{n},$$  (A.2.12)

where the second inequality uses $|x/(1 + \delta) - y| \leq 2|y\delta| + 2|x - y|$ for any $|\delta| < 1/2$. Therefore, by (A.2.12), we have $\zeta_1 = s \log d / \sqrt{n}$ and $\zeta_2 = 1/d$ which derives the condition $s(\log d)^{3/2} / \sqrt{n} = o(1)$.

Third, we bound the difference between $T_B$ and $T_0^B$ as

$$|T_B - T_0^B| \leq \max_{(j,k) \in E} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\Theta}_j^T (X_i X_i^T \hat{\Theta}_k - e_k) - \Theta_j^T (X_i X_i^T \Theta_k - e_k)\right) \xi_i\right|$$
Conditioning on the data \( \{X_i\}_{i=1}^n \), the right hand side of the above inequality is a suprema of a Gaussian process. Therefore, we need to bound the following conditional variance

\[
\max_{(j,k) \in E} \frac{1}{n} \sum_{i=1}^n \left[ \hat{\Theta}_j^T (X_i X_i^T \hat{\Theta}_k - e_k) - \Theta_j^T (X_i X_i^T \Theta_k - e_k) \right]^2 
\leq \max_{(j,k) \in E} \left| \hat{\Theta}_j^T (X_i X_i^T \hat{\Theta}_k - e_k) - \Theta_j^T (X_i X_i^T \Theta_k - e_k) \right|^2 
\lesssim \left[ 2M \| \hat{\Theta} - \Theta \|_1 \max_i \| X_i X_i^T - \Sigma \|_{\max} \right]^2.
\]

According to Lemma A.2.1, we have with probability \(1 - 1/d\), \( \max_i \| X_i X_i^T - \Sigma \|_{\max} \leq C \sqrt{\log(dn)} \). Therefore, the event

\[
\mathcal{E} = \left\{ \max_{(j,k) \in E} \frac{1}{n} \sum_{i=1}^n \left[ \hat{\Theta}_j^T (X_i X_i^T \hat{\Theta}_k - e_k) - \Theta_j^T (X_i X_i^T \Theta_k - e_k) \right]^2 \leq CM^2 \frac{(s \log(dn))^2}{n} \right\}
\]

has \( \mathbb{P}(\mathcal{E}^c) < 1/d \). Therefore, by maximal inequality, under the event \( \mathcal{E} \), we have

\[
\mathbb{E} \left[ \max_{(j,k) \in E} \frac{1}{n} \sum_{i=1}^n \left( \hat{\Theta}_j^T (X_i X_i^T \hat{\Theta}_k - e_k) - \Theta_j^T (X_i X_i^T \Theta_k - e_k) \right) \xi_i | \{X_i\}_{i=1}^n \right] \lesssim M^2 \frac{(s \log dn) \sqrt{\log d}}{\sqrt{n}}.
\]

Applying Borell’s inequality, we have with probability \(1 - 1/d\),

\[
\mathbb{P} \left( \max_{(j,k) \in E} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \hat{\Theta}_j^T (X_i X_i^T \hat{\Theta}_k - e_k) - \Theta_j^T (X_i X_i^T \Theta_k - e_k) \right) \xi_i > C \sqrt{\frac{s^2 \log^4 dn}{n}} | \{X_i\}_{i=1}^n \right) \leq 1/d.
\]

This implies that

\[
\mathbb{P}(|T^B - T_0^B| > \sqrt{(s^2 \log^4 dn)/n}) > 1/d < 1/d.
\]

Therefore, by Corollary 3.1 of Chernozhukov et al. (2013),

\[
\lim_{n \to \infty} |\mathbb{P}(T_E > c(\alpha, E)) - \alpha| = 0. \tag{A.2.13}
\]
Since the tail probability is independent to the edge set $E$, the proof is complete.

**Lemma A.2.2.** Under the same conditions as Lemma 2.2.1, we have

$$
P\left( \max_{j,k \in [d]} |\hat{\Theta}_{j,k}^{d} - \Theta_{j,k}| > C_0 \sqrt{\frac{\log d}{n}} \right) < \frac{2}{d}, \tag{A.2.14} \]

for some constant $C_0 > 0$.

**Proof.** By (A.2.12), we have with probability $1 - 1/d$,

$$\max_{j,k \in [d]} |\hat{\Theta}_{j,k}^{d} - \Theta_{j,k} + \Theta_j^T (\hat{\Sigma} \Theta_k - e_k)| \leq C_1 \frac{s \log d}{n}.$$

By Lemma A.2.1 and $\|\Theta\|_2 \leq C$, we have $\|\Theta_j^T X X^T \Theta_k\|_{\psi_1} \leq C_2 C^2$. By the maximal inequality (see Lemma 2.2.2 in van der Vaart and Wellner (1996)), we have for some constant $C_3 > 0$

$$P\left( \max_{j,k \in [d]} |\Theta_j^T (\hat{\Sigma} \Theta_k - e_k)| > C_3 C^2 \sqrt{\frac{\log d}{n}} \right)$$

$$\leq P\left( \max_{j,k \in [d]} \left| \frac{1}{n} \sum_{i=1}^{n} (\Theta_j^T X_i X_i^T \Theta_k - E[\Theta_j^T X_i X_i^T \Theta_k]) \right| > C_3 C^2 \sqrt{\frac{\log d}{n}} \right) \leq 1/d.$$

Let $C_0 = C_1 + C_3$ and the lemma is proved. \qed
Appendix B

Proofs in Chapter 3

B.1 Proof of Theorem 3.1.1

In this section we prove the main result of Section 3.1. To begin with, we give a high level picture of our proof. The argument consists of four major steps. Our first three steps will show that, there exists a constant $R$ such that if $\theta \leq \frac{1-C^{-1}}{\sqrt{2}\|A_0\|_1+2}$ we have:

$$\gamma(S_0(\theta, s), S_1(\theta, s)) \geq 1 - \frac{1}{2} \sqrt{\frac{1}{|C|^2} \sum_{e, e' \in C} \exp \left( n\frac{(R\theta)^2d_{G_0}(e, e') + 2}{d_{G_0}(e) + 1} \right) - 1}. \quad (B.1.1)$$

To establish this result, in the first step, we select one precision matrix from the null $S_0(\theta, s)$ and a set of precision matrices from the alternative $S_1(\theta, s)$. In the second step, we apply Le Cam’s Lemma to the precision matrices constructed above to get a lower bound of $\gamma(S_0(\theta, s), S_1(\theta, s))$. In the third step, we establish trace perturbation inequalities to further connect the lower bound achieved in the second step to the geometric quantities of the graphs. In the fourth step, we prove the theorem by showing that the right hand side of (B.1.1) goes to 1 if (3.1.6) is satisfied.
Step 1 (Matrix Construction).

In this step we construct a class of precision matrices based on the null base graph $G_0$ and the divider set $C$ and verify that these matrices indeed belong to the sets $S_0(\theta, s)$ and $S_1(\theta, s)$. We begin with giving the upper bound of matrix norms of adjacent inequalities. Let $A_0$ be the adjacency matrix of the graph $G_0$. Observe that since $A_0$ is symmetric, by Hölder’s inequality $\|A_0\|_2 \leq \sqrt{\|A_0\|_1 \cdot \|A_0\|_\infty} = \|A_0\|_1 \leq D$.

Similarly, denote with $A_e$ the adjacency matrix of the graph $(V, \{e\})$ for $e \in C$. Under our assumptions it follows that $A_0 + A_e$ is the adjacency matrix of the graph $G_e = (V, E_0 \cup \{e\})$. For brevity, for any two edges $e, e' \in C$ we define the shorthand notation $A_{e,e'} := A_0 + A_e + A_{e'}$. Take $\Theta_0 = I + \theta A_0$, $\Theta_e = I + \theta (A_0 + A_e)$, $\Theta_{e,e'} = I + \theta A_{e,e'}$, for $e, e' \in C$, $\theta > 0$. By the triangle inequality for any $e, e' \in C$ we have:

$$\max(\|A_0\|_2, \|A_0 + A_e\|_2, \|A_{e,e'}\|_2) \leq \|A_0\|_2 + 2,$$
$$\max(\|A_0\|_1, \|A_0 + A_e\|_1, \|A_{e,e'}\|_1) \leq \|A_0\|_1 + 2.$$

Recall definition (4.3.2) of the set $U_s$. Next we make sure that the matrices $\Theta_0$ and $\Theta_e$ fall into $U_s$ and in addition the matrix $\Theta_{e,e'} > 0$. For the upper bounds, it suffices to choose $\eta$ satisfying:

$$\max(\|\Theta_0\|_2, \|\Theta_e\|_2) \leq 1 + (\|A_0\|_2 + 2)\theta \leq C,$$
$$\max(\|\Theta_0\|_1, \|\Theta_e\|_1) \leq 1 + (\|A_0\|_1 + 2)\theta \leq L.$$

Recall that $\|A_0\|_2 \leq \|A_0\|_1$, and $C \leq L$ hence both inequalities are implied if $1 + (\|A_0\|_1 + 2)\theta \leq C$. This inequality holds since

$$\theta < \frac{(1 - C^{-1}) \wedge e^{-1/2}}{\sqrt{2}(D+2)} \leq \frac{C - 1}{(\|A_0\|_1 + 2)},$$
where the last inequality is true since $D = \|A_0\|_1$, and $C \geq 1$ and therefore $C - 1 \geq 1 - C^{-1}$.

Furthermore, by Weyl’s inequality:

$$\lambda_d(\Theta_0), \lambda_d(\Theta_e), \lambda_d(\Theta_{e,e'}) \geq 1 - \theta(\|A_0\|_2 + 2) \geq 1 - \theta(\|A_0\|_1 + 2), \quad (B.1.2)$$

where $\lambda_d$ denotes the smallest eigenvalue of the corresponding matrix. We want to ensure that the last term is at least $C^{-1}$. Since by assumption $\theta < \frac{1 - C^{-1}}{\sqrt{2(\|A_0\|_1 + 2)}}$ the above inequalities are satisfied. Furthermore, we have $G_0 \in G_0, G_e \in G_1$ for all $e \in C$ and hence the induced graphs $G(\Theta_0) \in G_0$ and $G(\Theta_e) \in G_1$ for all $e \in C$. This shows that $\Theta_0 \in S_0(\theta, s)$ and $\Theta_e \in S_1(\theta, s)$. We also obtain as a by-product that $\Theta_{e,e'} \geq 0$.

**Step 2 (Minimax Risk Lower Bound).**

In this step we obtain a lower bound on the minimax risk driven by Le Cam’s Lemma (LeCam, 1973). Using a determinant identity we control the chi-square divergence by the traces of adjacency matrices’ powers. Put the uniform prior on $C$ and consider the models generated by $N(0, (\Theta_e)^{-1})$ where $e \in C$. Define:

$$\overline{P} = \frac{1}{|C|} \sum_{e \in C} P_{\Theta_e},$$

where $P_{\Theta_e}$ we define the probability measure when the data is i.i.d. $X_i \sim N(0, (\Theta_e)^{-1})$, and let $P_{\Theta_0}$ be the probability measure when the data is i.i.d. $X_i \sim N(0, (\Theta_0)^{-1})$. Next, by Neyman-Pearson’s lemma we have:

$$\gamma(S_0, S_1) \geq \inf_{\psi} \left[ P_{\Theta_0}(\psi = 1) + P(\psi = 0) \right] = 1 - TV(\overline{P}, P_{\Theta_0}), \quad (B.1.3)$$

110
where for two probability measures $P, Q \ll \lambda$ on a measurable space $(\Omega, \mathcal{A})$, TV stands for total variation distance, and is defined as

$$\text{TV}(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \frac{1}{2} \int \left| \frac{dP}{d\lambda}(\omega) - \frac{dQ}{d\lambda}(\omega) \right| d\lambda(\omega).$$

By Cauchy-Schwartz one has:

$$1 - \text{TV}(\mathbb{P}_\pi, \mathbb{P}_{\Theta_0}) \geq 1 - \frac{1}{2} \sqrt{D_{\chi^2}(\mathbb{P}_\pi, \mathbb{P}_{\Theta_0})}, \quad (B.1.4)$$

where $D_{\chi^2}(P, Q)$ is the chi-square divergence between the measures $P, Q$ and is defined as:

$$D_{\chi^2}(P, Q) = \sum \left( \left( \frac{dP}{dQ}(\omega) - 1 \right)^2 dQ(\omega) \right) = \int \left( \frac{dP}{dQ}(\omega) \right)^2 dQ(\omega) - 1,$$

assuming that $P \ll Q$. Observe that $D_{\chi^2}(\mathbb{P}, \mathbb{P}_{\Theta_0})$ can be equivalently expressed as:

$$D_{\chi^2}(\mathbb{P}, \mathbb{P}_{\Theta_0}) = \mathbb{E}_{\Theta_0} L_{\Theta_0}^2 - 1, \quad (B.1.5)$$

where $L_{\Theta_0} = \frac{1}{|C|} \sum_{e \in C} \frac{d\mathbb{P}_{\Theta_0}}{d\mathbb{P}_e}$ is the integrated likelihood ratio, and $\mathbb{E}_{\Theta_0}$ denotes the expectation under $X_i \sim N(0, (\Theta_0)^{-1})$. Hence by (B.1.3) and (B.1.4), it suffices to obtain upper bounds on the integrated likelihood ratio in order to lower bound the minimax risk (3.1.2). Writing out the likelihood ratio comparing the normal distribution with precision matrix $\Theta_0$ to the uniform mixture of normal distribution with precision matrix $\Theta_e$ for $e \in C$ we get:

$$L_{\Theta_0} = \frac{1}{|C|} \sum_{e \in C} \left( \frac{\det(\Theta_e)}{\det(\Theta_0)} \right)^{n/2} \prod_{i=1}^n \exp(-X_i^T \Theta A_i X_i / 2).$$
To calculate the chi-square distance in (B.1.5), next we square this expression and take its expectation under \( \mathbb{P}_{\Theta_0} \) to obtain:

\[
\mathbb{E}_{\Theta_0} L^2_{\Theta_0} = \frac{1}{|C|^2} \sum_{e,e' \in C} \left( \frac{\text{det}(\Theta_e)}{\text{det}(\Theta_0)} \right)^{n/2} \frac{\text{det}(\Theta_0)}{\text{det}(\Theta_0)} \exp \left( - \frac{\sum_{i=1}^{n} X_i^T \theta(A_e + A_{e'}) X_i}{2} \right)
\]

\[
= \frac{1}{|C|^2} \sum_{e,e' \in C} \left( \frac{\text{det}(\Theta_e)}{\text{det}(\Theta_0)} \right)^{n/2} \left( \frac{\text{det}(\Theta_{e'})}{\text{det}(\Theta_{e,e'})} \right)^{n/2}.
\]

(B.1.6)

Next, we will expand the determinants above. Recall that we have ensured that \( 1 - \theta(\|A_0\|_2 + 2) > 0 \) (see B.1.2). This implies

\[
\theta \max(\|A_0\|_2, \|A_0 + A_e\|_2, \|A_0 + A_{e'}\|_2, \|A_{e,e'}\|_2) \leq 1.
\]

For what follows for a symmetric matrix \( A_{d \times d} \) we denote its ordered eigenvalues with \( \lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_d(A) \). Let \( A \in \mathbb{R}^{d \times d} \) be a symmetric matrix such that \( \|A\|_2 \leq 1 \). Then we have:

\[
\log \det(I + A) = \sum_{j=1}^{d} \log \lambda_j(I + A) = \sum_{j=1}^{d} \log(1 + \lambda_j(A))
\]

\[
= \sum_{j=1}^{d} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\lambda_j^k(A)}{k} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\text{Tr}(A^k)}{k}.
\]

Using the form \( \det(I + A) = \exp(\log \det(I + A)) \) and plugging the above equation into (B.1.6), we conclude that:

\[
\mathbb{E}_{\Theta_0} L^2_{\Theta_0} = \frac{1}{|C|^2} \sum_{e,e' \in C} \exp \left( \frac{n}{2} \sum_{k=1}^{\infty} (-\theta)^k \frac{\text{Tr}(A^k)}{k} \left( T_1^k + T_2^k \right) \right),
\]

where
$T_1^k := \text{Tr} [A_0^k - (A_0 + A_e)^k]$ \hspace{1em} $T_2^k := \text{Tr} [(A_{e,e'})^k - (A_0 + A_e)^k]$.

**Step 3 (Trace Perturbation Inequalities).**

In this step, we control $T_1^k + T_2^k$ in terms of $k$ and link it with the geometric quantities of the graph. We view $T_1^k$ as the perturbation difference between $\text{Tr} [A_0^k]$ and $\text{Tr} [(A_0 + A_e)^k]$ and we treat $T_2^k$ similarly. In the following step, we aim to develop the perturbation inequalities for the trace of matrix powers.

First we will argue that $T_1^k + T_2^k \geq 0$ for all $k \in \mathbb{N}$. To see this recall that the trace operator of an adjacency matrix $M$ satisfies

$$\text{Tr}(M^k) = \text{number of all closed walks of length } k.$$  

First we consider case $e \neq e'$. Notice that all closed walks in $G(A_0 + A_e)$ that do not belong to $G(A_0)$ have to pass through the edge $e$ at least once. Similarly all closed walks in $G(A_0 + A_e')$ that do not belong to $G(A_0)$ have to pass through the edge $e'$ at least once. Furthermore, all closed walks of length $k$ passing through either $e$ or $e'$ belong to $G(A_{e,e'})$. In addition $G(A_{e,e'})$ might contain extra closed walks passing through both $e$ and $e'$. This shows:

$$T_1^k + T_2^k \geq 0,$$

for all $k$. This shows that when $k$ is odd we have $(-\theta)^k (T_1^k + T_2^k) \leq 0$, and thus to control $\mathbb{E}_{\Theta_0} L_\Theta^2$ it suffices to focus only on even $k$.

Next we prove that for $k < 2d_{\mathcal{G}_0}(e,e') + 2$, we have $T_1^k + T_2^k \equiv 0$. To see this, first consider the case $e \neq e'$. Notice that the graph $G(A_{e,e'})$ cannot contain paths passing through both $e$ and $e'$ unless $k \geq 2d_{\mathcal{G}_0}(e,e') + 2$. To see this, notice that no even length closed walk between $e$ and $e'$ can exist if the length of this walk is smaller than $2d_{\mathcal{G}_0}(e,e')$ plus the two edges $e$
and $e'$. This proves our claim in the case $e \neq e'$. In the special case $e = e'$, the length of the path trivially needs to be at least of length 2 to pass through both $e$ and $e'$.

We will now argue that for even $k \in \mathbb{N}$ we have $T_1^k + T_2^k \leq 2(\|A_0\|_2 + 2)^k$. In fact we will prove that $T_1^k \leq 0$ for all $k$ and $T_2^k \leq 2(\|A_0\|_2 + 2)^k$ for all even $k$. To see that $T_1^k \leq 0$, note that $G(A_0)$ contains less closed walks than $G(A_0 + A_e)$.

Recall that for a symmetric matrix $A_{d \times d}$ we denote its ordered eigenvalues with $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_d(A)$. To this end we state a helpful result whose proof is deferred to the supplement.

**Lemma B.1.1.** For two symmetric $m \times m$ matrices $A$ and $B$, and any constants $c_1 \geq c_2 \geq \ldots \geq c_m$, and a permutation $\sigma$ on $\{1, \ldots, m\}$ we have:

$$\sum_{j=1}^{m} c_{\sigma(j)} \lambda_j(A + B) \leq \sum_{j=1}^{m} c_{\sigma(j)} \lambda_j(A) + \sum_{j=1}^{m} c_j \lambda_j(B).$$

Using Lemma B.1.1 for the matrices $A = A_e, B = A_0 + A_{e'}$ with constants

$$c_{\sigma(j)} = \text{sign}(\lambda_j(A_{e,e'}) - \lambda_j(A_e))|\lambda_j(A_{e,e'}) - \lambda_j(A_e)|^{k-1},$$

we obtain:

$$\sum_{j=1}^{d} |\lambda_j(A_{e,e'}) - \lambda_j(A_e)|^k \leq \sum_{j=1}^{d} c_j \lambda_j(A_0 + A_{e'})$$

$$\leq \left[ \sum_{j=1}^{d} |c_j|^{\frac{k}{k-1}} \right]^{\frac{k-1}{k}} \left[ \sum_{j=1}^{d} |\lambda_j(A_0 + A_{e'})|^k \right]^{\frac{1}{k}},$$

where the last inequality follows by Hölder’s inequality. We conclude that:

$$\sum_{j=1}^{d} |\lambda_j(A_{e,e'}) - \lambda_j(A_e)|^k \leq \sum_{j=1}^{d} |\lambda_j(A_0 + A_{e'})|^k. \quad (B.1.7)$$
Next, observe that the negative adjacency matrix $-A_e$ of the single edge graph $(\Omega, \{e\})$ has very simple eigenvalue structure: $1$, $-1$ and $d-2$ zeros. Hence we conclude that for even $k$:

$$T_2^k = \text{Tr}((A_{e,e'})^k) - \text{Tr}((A_0 + A_{e'})^k) = \sum_{j=1}^{d} |\lambda_j(A_{e,e'})|^k - \sum_{j=1}^{d} |\lambda_j(A_0 + A_{e'})|^k$$

$$\leq |\lambda_1(A_{e,e'})|^k - |\lambda_1(A_{e,e'}) - 1|^k + |\lambda_d(A_{e,e'})|^k - |\lambda_d(A_{e,e'}) + 1|^k$$

$$\leq |\lambda_1(A_{e,e'})|^k + |\lambda_d(A_{e,e'})|^k \leq 2\|A_{e,e'}\|_2^k \leq 2(h_0 + 2)^k.$$

The last shows that indeed $T_1^k + T_2^k \leq 2(h_0 + 2)^k$ as claimed. Putting everything together we obtain

$$\sum_{k=1}^{\infty} \frac{(-\theta)^k}{k} [T_1^k + T_2^k] \leq \sum_{2|k, k \geq 2dG_0(e,e')} \frac{\theta^k}{k} [T_1^k + T_2^k]$$

$$\leq \sum_{2|k, k \geq 2dG_0(e,e')} \frac{2((h_0 + 2)\theta)^k}{k}$$

$$\leq \frac{2((h_0 + 2)\theta)^{2dG_0(e,e')} + 2}{(2dG_0(e,e') + 2)(1 - (\theta(h_0 + 2))^2)^2} \leq \frac{2((h_0 + 2)\theta)^{2dG_0(e,e')} + 2}{dG_0(e,e') + 1},$$

where in the last inequality we used the fact that $\theta \leq \frac{1}{\sqrt{2(\|A_0\|_2 + 2)}}$ which follows by the requirements on $\theta$. This completes the proof of (B.1.1) where $R = \sqrt{2(\|A_0\|_2 + 2)}$.

**Step 4 (Rate Control).**

The goal in this final step is to show that if (3.1.6) holds, the minimax risk

$$\lim \inf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1.$$

The proof is technical, but the high-level idea is to clip the first $\log |C|$ degrees in (B.1.1) and deal with two separate summations. It turns out that the scaling assumed on $\theta$ in (3.1.6) is precisely enough to control both the summation of all degrees below $\log |C|$ and the
summation of all degrees above \( \log |C| \). Define the following quantities:

\[
K_r := |\{(e, e') \mid e, e' \in C, d_{G_0}(e, e') = r\}|
\]

where \((e, e')\) are unordered edge pairs, and observe that \( \sum_{r \geq 0} K_r = \binom{|C|}{2} + |C| \) by definition. We will in fact, first show that if \( \theta \leq \kappa \sqrt{\frac{\log |C|}{n}} \) for some small \( \kappa \), and

\[
\sum_{r=0}^{\log |C|} K_r = O(|C|^{2-\gamma}),
\]

(B.1.8)

for some \( 1/2 < \gamma \leq 1 \), then \( \liminf \gamma(S_0(\theta, s), S_1(\theta, s)) = 1 \) provided that \( |C| \to \infty \). We will then derive the Theorem as a corollary to this observation.

By (B.1.1) it suffices to control:

\[
\frac{2}{|C|^2} \sum_{(e, e') : d_{G_0}(e, e') \geq 1} \exp \left( n \frac{\bar{\theta}^{2d_{G_0}(e, e')} + 2}{d_{G_0}(e, e') + 1} \right) + \frac{2K_0 - |C|}{|C|^2} \exp(n\bar{\theta}^2),
\]

(B.1.9)

where we will write \( \bar{\theta} \) for \( R\theta \) for brevity.

First, observe that since \( \theta < \frac{1-C^{-1}}{\sqrt{2(D+2)}} \) then we have

\[
\bar{\theta} < R \frac{(1-C^{-1}) \wedge e^{-1/2}}{\sqrt{2(D+2)}} \leq (1-C^{-1}) \wedge e^{-1/2} < e^{-1/2} < 1.
\]

We will show that when \( \bar{\theta} \) is small, (B.1.9) is bounded by 1 asymptotically, which in turn suffices to show that \( \liminf \gamma(S_0(\theta, s), S_1(\theta, s)) = 1 \). Notice that \( \bar{\theta} \) and \( \theta \) are the same quantity up to the constant \( R \) and hence \( \theta < \kappa \sqrt{\frac{\log |C|}{n}} \) is equivalent to \( \bar{\theta} < \bar{\kappa} \sqrt{\frac{\log |C|}{n}} \) for some sufficiently small \( \bar{\kappa} \). We will require:

\[
\bar{\kappa} < \sqrt{\gamma \wedge \sqrt{2\gamma/c_0 \wedge (ec_0)^{-1/2}}}.
\]
Observe that since $K_0 = O(|C|^{2-\gamma})$, $\bar{\kappa}^2 < \gamma$, and $\bar{\theta} < \bar{\kappa} \sqrt{\frac{\log |C|}{n}}$ we have:

$$I_2 \leq \frac{(2K_0 - |C|)|C|\bar{\kappa}^2}{|C|^2} \to 0.$$ 

Next we tackle the term $I_1$ in (B.1.9). We will show that since $\bar{\theta} < 1$ by assumption, this term goes to 1.

$$I_1 = \frac{2}{|C|^2} \sum_{r \geq 1} K_r \exp(n\bar{\theta}^{2r+2}/(r+1)) = \frac{2}{|C|^2} \sum_{r=1}^{d-1} K_r |C|\bar{\kappa}^{2\bar{\theta}^{2r}}/(r+1) + \frac{2K_\infty}{|C|^2}$$

$$< \frac{2}{|C|^2} \sum_{r=1}^{d-1} K_r |C|\bar{\theta}^{2r}/(r+1) + \frac{2K_\infty}{|C|^2},$$

where the last inequality follows by the fact that $\bar{\kappa}^2 < \gamma < 1$. Splitting out the first $\lfloor \log |C| \rfloor$ terms out of this summation yields:

$$I_1 \leq \frac{2}{|C|^2} \sum_{r=1}^{\lfloor \log |C| \rfloor} K_r |C|\bar{\theta}^{2r}/(r+1) + \frac{2}{|C|^2} \sum_{r=\lfloor \log |C| \rfloor+1}^{d-1} K_r |C|\bar{\theta}^{2r}/(r+1) + \frac{2K_\infty}{|C|^2}.$$ 

The first term is bounded by

$$I_{11} \leq 2 \left( \sum_{r=1}^{\lfloor \log |C| \rfloor} K_r \right) \frac{|C|\bar{\theta}^{2r}/(r+1)}{|C|^2} = o(1),$$

where we used

$$\left( \sum_{r=2}^{\lfloor \log |C| \rfloor} K_r \right) = O(|C|^{2-\gamma})$$

and the fact that $\bar{\theta}^2/2 < 1/2 < \gamma$. Next, we will argue that $|C|\bar{\theta}^{2r}/(r+1) \leq 1 + 3\bar{\theta}^{2r}(|C| - 1)/(r + 1)$. This follows by:

$$\exp(\log(|C|)\bar{\theta}^{2r}/(r + 1)) \leq 1 + 3 \log(|C|)\bar{\theta}^{2r}/(r + 1) \leq 1 + 3(|C| - 1)\bar{\theta}^{2r}/(r + 1),$$
with the first inequality holding when \( \log(|C|) \tilde{\theta}^{2r}/(r + 1) < 1 \), which is true since \( \tilde{\theta} < 1 \), and \( r \geq \lceil \log |C| \rceil + 1 \). Hence we have:

\[
I_{12} \leq \frac{2}{|C|^2} \sum_{r=\lceil \log |C| \rceil + 1}^{d-1} K_r (1 + 3(|C| - 1) \tilde{\theta}^{2r}/(r + 1)) + \frac{2K_\infty}{|C|^2}
\]

\[
\leq \left( 1 - \frac{O(|C|^{2-\gamma})}{|C|^2} \right) + \frac{6(|C| - 1)}{|C|^2} \sum_{r=\lceil \log |C| \rceil + 1}^{d-1} K_r \tilde{\theta}^{2r} \frac{r}{r}.
\]

\[
\leq \left( 1 - \frac{O(|C|^{2-\gamma})}{|C|^2} \right) + \frac{6\tilde{\theta}^{2\lceil \log |C| \rceil + 2}}{1 - \tilde{\theta}^2} |C|^{-1} \max_{\lceil \log |C| \rceil + 1 \leq r} K_r \frac{r}{r}.
\]

Paying closer attention to the second term we have:

\[
\frac{6\tilde{\theta}^{2\lceil \log |C| \rceil + 2}}{1 - \tilde{\theta}^2} |C|^{-1} \max_{\lceil \log |C| \rceil + 1 \leq r} K_r \frac{r}{r} \leq \frac{6}{1 - e^{-1}} \tilde{\theta}^{2\lceil \log |C| \rceil + 2} \left( \frac{|C|}{2} \right) + |C| \frac{\lceil \log |C| \rceil + 2}{|C|} |C| = o(1),
\]

with the last equality holds since \( \tilde{\theta} < \exp(-1/2) \), as we required. This combined with (B.1.1) concludes the proof of \( \liminf_n \gamma(S_0(\theta,s), S_1(\theta,s)) = 1 \), when \( \theta \leq \kappa \sqrt{\frac{\log |C|}{n}} \).

Finally, notice that any subset of a divider \( C \) is trivially a divider. Hence we can apply what we just showed to the set \( N_{\log |C|} \subset C \) — the maximal \( \log |C| \)-packing of \( C \). Evaluating the constants \( K_r \) on \( N_{\log |C|} \) gives:

\[ K_0 = |N_{\log |C|}|, K_r = 0 \text{ for all } r \leq \lceil \log |C| \rceil, \]

and since \( |N_{\log |C|}| \leq \lceil \log |C| \rceil \) we conclude that \( \sum_{r=0}^{\lceil N_{\log |C|} \rceil} K_r = |N_{\log |C|}| = O(|N_{\log |C|}|^{2-\gamma}) \) for any \( 0 < \gamma \leq 1 \).
B.1.1 Proof of Theorem 3.1.2

The proof of this result follows the same line of argument as the proof of Theorem 3.1.1. Here we just sketch the differences. The first important observation is that the risk $\gamma(S_0, S_1)$ is symmetric in the sense that $\gamma(S_0, S_1) = \gamma(S_1, S_0)$. This implies that if controlling the eigenvalues of precision matrices made in the proof of Theorem 3.1.1 hold for edge deletion divider, the proof will continue to hold. If we denote by $A_e$ the adjacency matrix of $(\overline{V}, \{e\})$ for any $e \in \mathcal{C}$, then the adjacency matrices of the graphs under the null $G_{\setminus e}$ are given by $A_1 - A_e$. It is now a simple exercise to check that:

$$\text{Tr}(A_1^k) - \text{Tr}((A_1 - A_e)^k) \leq 2\|A_1\|_2^k,$$

$$\text{Tr}((A_1 - A_{e'})^k) - \text{Tr}((A_1 - A_e - A_{e'})^k) \leq 0,$$

and the proof is completed as in Theorem 3.1.1.

B.1.2 Auxiliary Results

Recall that for a symmetric matrix $A \in \mathbb{R}^{d \times d}$ we denote its eigenvalues in decreasing with $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_d(A)$.

**Theorem B.1.1** (Lidskii’s Inequality (Helmke and Rosenthal, 1995)). For two symmetric $m \times m$ matrices $A$ and $B$ we have:

$$\sum_{j=1}^{k} \lambda_{i_j}(A + B) \leq \sum_{j=1}^{k} \lambda_{i_j}(A) + \sum_{j=1}^{k} \lambda_{j}(B),$$

for any $1 \leq i_1 < \ldots < i_k \leq m$.

We now derive Lemma B.1.1 as a Corollary to Theorem B.1.1. Lemma B.1.1 can also be viewed as a more general formulation of the latter result.
Proof of Lemma B.1.1. First notice that since
\[
\sum_{j=1}^{m} \lambda_j (A + B) = \sum_{j=1}^{m} \lambda_j (A) + \sum_{j=1}^{m} \lambda_j (B),
\]
it suffices to show the bound when for all \( j \) we have \( c_j \geq 0 \). Let \( C = \{ j \mid c_j \neq 0 \} \). We will induct on the cardinality \( |C| \), which we denote with \( k \). When \( k = 1 \), the inequality immediately follows by Theorem B.1.1. Suppose the inequality holds for \( k = l \). Notice that the inequality can equivalently be expressed as
\[
\sum_{j=1}^{m} c_j \lambda_{\sigma-1(j)} (A + B) \leq \sum_{j=1}^{m} c_j \lambda_{\sigma-1(j)} (A) + \sum_{j=1}^{m} c_j \lambda_j (B), \tag{B.1.10}
\]
To see why it holds for \( k = l + 1 \) let \( c^* = \min_{j \in C} c_j \), and notice that by Theorem B.1.1 we have:
\[
\sum_{j \in C} c^* \lambda_{\sigma-1(j)} (A + B) \leq \sum_{j \in C} c^* \lambda_{\sigma-1(j)} (A) + \sum_{j \in C} c^* \lambda_j (B).
\]
This implies that inequality (B.1.10) follows by an inequality in which we subtract \( c^* \) from all \( c_j, j \in C \). This completes the proof by the induction hypothesis. \( \square \)

B.2 Proof of Proposition 3.1.1 and SAP example

In this section, we prove Proposition 3.1.1 and we provide the example of self-avoiding path.

Proof of Proposition 3.1.1. To prove this result we construct the set \( C \) explicitly in an iterative manner. For a vertex set \( V \subset \overline{V} \) define
\[
\mathcal{N}_l (V) := \{ v \in \overline{V} \mid d_{G_0}(v, w) \leq l \ \forall w \in V \}.
\]
Start with $C = \emptyset$. Take any vertex $v_1 \in V$, find $w_1 \in W_v$, and set $C \leftarrow C \cup \{(v_1, w_1)\}$. We now bound the cardinality of $\mathcal{N}_l({v_1, w_1})$. Suppose that the maximum degree of $G_0$ is at most $D$. We have $|\mathcal{N}_l({v_1, w_1})| \leq 2 + 2D^l$. Update the graph $G_0$ by deleting all vertices and edges associated with the set $|\mathcal{N}_l({v_1, w_1})|$. The new graph has at least $d - (2 + 2D^l)$ vertices. If $cd^\gamma - k(2 + 2D^l) > 0$ then there exists a pair of vertices $v_2, w_2$ so that $(V, E_0 \cup \{(v_2, w_2)\}) \in \mathcal{G}_1$. Set $C \leftarrow C \cup \{(v_2, w_2)\}$. Iterating this procedure $k$ times while maintaining $cd^\gamma - k(2 + 2D^l) > 0$ gives a set $C$ of cardinality $k + 1$, each two edges of which are at least $l$ apart. Since $2 + 2D^l \leq 4D^l$ the latter inequality is implied if $\log(cd^\gamma) - \log 4 \geq \log k + \log D$. Selecting $l = \lceil \log k \rceil + 1 \geq \lceil \log(k+1) \rceil$ yields that we can select $k$ as large as $\log k \leq \log(cd^\gamma) - \log 4 - 2 \log D$, which completes the proof.

**Example B.2.1** (Self-Avoiding Path (SAP) of Length $m$ vs $m + 1$, $m < \sqrt{d}$). In this example we test the existence of a SAP of length $\leq m$ vs SAP of length $\geq m + 1$, i.e., $\mathcal{G}_0 = \{G \in \mathcal{G} \mid \text{All SAPs have length } \leq m\}$ vs $\mathcal{G}_1 = \{G \in \mathcal{G} \mid \exists \text{ SAP of length } m + 1\}$. We assume that $m < \sqrt{d}$ in order to be able to construct sufficiently large divider set which yields tight bounds. The case $m \geq \sqrt{d}$ is treated in Example 3.1.7. Without loss of generality we further suppose that $m + 2$ divides $d$ (at the sake of assigning some isolated vertices in the graph). Construct the following graph in $\mathcal{G}_0$: $G_0 = (V, E_0)$, where

$$E_0 := \{(j, j + 1) \mid \text{unless } (m + 2) \text{ divides } j \text{ or } (m + 2) \text{ divides } (j + 1)\},$$

and hence $|E_0| = dm/(m + 2)$ (see Fig B.1). Consider the class $\mathcal{C} := \{(j(m + 2) - 1, j(m + 2))_{j \in [d/(m+2)]}\}$. The set $\mathcal{C}$ is a divider since adding any edge from $\mathcal{C}$ to $G_0$ results in a graph with a longer self-avoiding path. Furthermore the maximum degree of $G_0$ is 2. Notice that the set $\mathcal{C}$ is its own a $(\log |\mathcal{C}|)$-packing, since each two edges $e, e' \in \mathcal{C}$ satisfy $d_{G_0}(e, e') = \infty$. Hence $M(\mathcal{C}, d_{G_0}, \log |\mathcal{C}|) = \log |\mathcal{C}| \propto \log d$. We conclude that if $\theta \leq \kappa \sqrt{\log d/n} \wedge \frac{1-C^{-1}}{4\sqrt{2}}$ we cannot differentiate the null from the alternative hypothesis.
Figure B.1: The null base $G_0$ and the divider $C$ (dashed) with $d_{G_0}(e, e') = \infty$, $d = 15$, $m = 3$.

Figure B.2: Test for the maximum degree $G_0 := \{G \mid d_{\text{max}}(G) \leq s_0\}$ vs $G_1 = \{G \mid d_{\text{max}}(G) \geq s_1\}$ with $s_0 = 3$, $s_1 = 5$ and $d = 18$. The solid edges represent $G_0 \in G_0$ with maximum degree $s_0 = 3$. We construct the divider $C = \{(1, 5), (1, 6), (7, 11), (7, 12), (13, 17), (13, 18)\}$.

B.3 Multi-Edge Divider Examples

B.3.1 Bounded Edge Sets

Proof of Example 3.2.1. In order to show the above result, we start by building a graph $G_0$ by constructing $\lfloor \frac{d}{s_1 + 1} \rfloor$ non-intersecting $s_0$-star graphs first (see Fig B.2). Define the star graph centers $C_j = (s_1 + 1)j + 1$ for $j = 0, \ldots, \lfloor \frac{d}{s_1 + 1} \rfloor - 1$. Next, define $G_0$ by

$$G_0 := \left( V, \bigcup_{j=0}^{\lfloor \frac{d}{s_1 + 1} \rfloor - 1} \bigcup_{k=1}^{s_0} \{(C_j, C_j + k)\} \right).$$

We define the divider $C$ as follows:

$$C := \left\{ \bigcup_{k=s_0+1}^{s_1} \{(C_j, C_j + k)\} \mid j = 0, \ldots, \lfloor \frac{d}{s_1 + 1} \rfloor - 1 \right\},$$

where we simply connect each of the vertices $C_j$ to the remaining $s_1 - s_0$ vertices in the block.
Since the predistance between any two different edge sets $S, S' \in C$, is $d_{G_0}(S, S') = \infty$, the set $C$ itself is a $\log |C|$-packing set. The latter implies that $M(C, d_{G_0}, \log |C|) = \log |C| \asymp \log(d/(s_1 + 1)) \asymp \log d$. In addition, one can easily check that $\|A_0\|_2 = \sqrt{s_0}$ and $\|A_0\|_1 = s_0$. By Theorem 3.2.1 we have that under the required scaling $\liminf_{n \to \infty} \gamma(S_0(\theta, s), S_1(\theta, s)) = 1$, when $\theta < \kappa \sqrt{\log d/n}$ for a sufficiently small $\kappa$.

**B.3.2 Unbounded Edge Sets**

*Proof of Example 3.2.2.* Before we show how this example follows from Theorem 3.2.2 we would like to highlight the difference between Examples 3.2.1 and 3.2.2. First, note that Example 3.2.2 is more flexible compared to Example 3.2.1 since the number of edges is allowed to scale with $n$. However Example 3.2.2 is also more restrictive in that we require $s = O(d^\gamma)$ for some $\gamma < 1/2$, which is not required by Example 3.2.1.

We now first explain the construction of $(G_0, C)$ and then we formalize it. We start by splitting the vertices into two parts $\{1, \ldots, \lfloor \sqrt{d} \rfloor\}$ and $\{\lfloor \sqrt{d} \rfloor + 1, \ldots, d\}$. The graph $G_0$ is constructed based only on the first part of vertices, and consists of non-intersecting $s_0$-star graphs. To build the divider $C$, we select any $s_1 - s_0$ vertices from the second set vertices $\{\lfloor \sqrt{d} \rfloor + 1, \ldots, d\}$ and connect them to any center of the $s_0$-star graphs in $G_0$ (e.g., vertex 1). The set $C$ contains all such edge sets (see Fig 3.8 in the main text). Formally, to construct a graph $G_0$ we split the vertex set $\{1, \ldots, \lfloor \sqrt{d} \rfloor\}$ into $[\frac{\lfloor \sqrt{d} \rfloor}{s_0+1}]$ non-intersecting $s_0$-star graphs. We take centers $C_j = (s_0 + 1)j + 1$ for $j = 0, \ldots, [\frac{\lfloor \sqrt{d} \rfloor}{s_0+1}] - 1$ and connect them to the remaining vertices as follows:

$$G_0 := \left( V, \bigcup_{j=0}^{[\frac{\lfloor \sqrt{d} \rfloor}{s_0+1}] - 1} \bigcup_{k=1}^{s_0} (C_j, C_j + k) \right).$$

123
Since \( s_0 < s_1 \leq s = O(d) \), this creates at least \( \lfloor \sqrt{d} \rfloor \simeq d^{1/2-\gamma} \) graphs. Denote the set of centers of the \( s_0 \)-star graphs with

\[
J := \bigcup_{j=0}^{\lfloor \sqrt{d/s_0} \rfloor - 1} \{C_j\}.
\]

To construct the divider \( C \) from the remaining \( d - \lfloor \sqrt{d} \rfloor \) isolated vertices we choose a set \( I \) of \( s_1 - s_0 \) vertices and connect them with any of the vertex \( C \) from the center set \( J \). Formally we let

\[
S_{C,I} = \{(C,i)\}_{i \in I} \quad \text{and} \quad \mathcal{I} = \left\{ I \in \{M \in 2^d : |M| = s_1 - s_0\} \mid \min_{i \in I} i > \lfloor \sqrt{d} \rfloor \right\}.
\]

Consider the collection of vertex buffers \( V_{S,S'} = V(S) \cap V(S') \). A visualization of the key quantities \( G_0, C, V_{S,S'} \) for two specific \( S, S' \in C \) is provided on Fig 3.8 in the main text. Now we will argue that the set \( \{1(v \in V_{S,S'})\}_{v \in V(S)} \) is negatively associated, as required by Assumption 3.2.5. Note that uniformly selecting a set \( S' \in C \) is equivalent to uniformly selecting a center \( C \in J \) and a vertex set \( I \in \mathcal{I} \) and construct \( S' \) by connecting \( C \) to every vertex in \( I \). By construction the selection of a center \( C \) is independent of the selection of the vertex set \( I \). By a result in Section 3.1 (c) of Joag-Dev and Proschan (1983) the random variables \( \{1(v \in V_{S,S'})\}_{v \in V(S)} \) are negatively associated. Next we calculate the remaining quantities from Theorem 3.2.2.

By the triangle inequality \( \|A_{S,S'}\|_2 \leq \sqrt{s_1} + \sqrt{s_1 - s_0} \leq 2\sqrt{s_1} \) and \( \|A_{S,S'}\|_1 \leq 2s_1 \) and therefore \( \Lambda \leq 2\sqrt{s_1}, \Gamma \leq 2s_1, \text{ and } B \leq \Lambda^4 \leq 16s_1^2 \leq 16s^2 \). To calculate upper and lower bounds of \( \mathcal{R} \) we only need to consider \( |S \cap S'| \neq 0 \) and in this case \( S \) and \( S' \) must share the same center. In this situation by definition we have \( |S \cap S'| = |V_{S,S'}| - 1 \) and therefore \( 1/2 \leq \mathcal{R} \leq 1 \).
Finally we calculate \( \max_{S \in C} \mathbb{E}_{S'} |V_{S,S'}| \). Recall that \( V_{S,S'} = V(S) \cap V(S') \) and that a set \( S' \) can be constructed by first selecting a center \( C \) uniformly from the set \( J \) and then selecting a set \( I \) uniformly from \( I \). Hence for all \( S \in C \)

\[
\mathbb{E}_{S'} |V_{S,S'}| = \mathbb{E}[\mathbf{1}(C \in V(S))|S] + \mathbb{E}[|I \cap V(S)| \mid S] = \frac{1}{|J|} + \frac{(s_1 - s_0)^2}{d - (s_0 + 1)|J|}.
\]

The last equality implies that \( M_B(C, G_0) \approx \log d \). An application of Theorem 3.2.2, and taking into account the required scaling completes the result.

**Proof of Example 3.2.3.** Consider \( G_0 = (\overline{V}, \emptyset) \), and take the divider:

\[
C = \{ \text{all s cliques with vertices in } \overline{V} \}.
\]

Take two sets \( S, S' \in C \). Define the vertex buffer as \( V_{S,S'} = V(S) \cap V(S') \). Similarly to Example 3.2.2, conditioning on \( S \) the random variables \( \{\mathbf{1}(v \in V_{S,S'})\}_{v \in V(S)} \) are negatively associated (see Section 3.1 (c) of Joag-Dev and Proschan (1983)). By the definition of vertex buffer sets we have that \( |V_{S,S'}| \leq s \), \( |S \cap S'| = \left( \frac{|V_{S,S'}|}{2} \right)^2 \leq s|V_{S,S'}| \) and hence \( s - 1/2 \leq \mathcal{R} \leq s \). Furthermore \( \Gamma \leq 2s \), and therefore \( \Lambda \leq 2s \). Therefore \( B \leq 4s^3 \). For all \( v \in V(S) \) we have \( \mathbb{E}_{S'}[\mathbf{1}(v \in V_{S,S'})] = s/d \) which implies \( \mathbb{E}_{S'}[|V_{S,S'}|] = s^2/d \) for all \( S \in C \), and therefore \( M_B(C, G_0) = \log(d/s^2) \). An application of Theorem 3.2.2 shows the statement.

**Proof of Example 3.2.4.** Consider \( G_0 = (\overline{V}, \emptyset) \), and take the divider:

\[
C = \{ \text{all cycles joining } s\text{-vertices from } \overline{V} \text{ in an anti-clockwise increasing order} \}.
\]

Two cycles from the set \( C \) are visualized on Fig 3.9b. Take two sets \( S, S' \in C \). Define the vertex buffer as \( V_{S,S'} = V(S) \cap V(S') \). Conditioning on \( S \) the random variables \( \{\mathbf{1}(v \in V_{S,S'})\}_{v \in V(S)} \) are negatively associated (see Section 3.1 (c) of Joag-Dev and Proschan (1983)). By the definition of vertex buffer sets we have that \( |V_{S,S'}| \leq s \), \( |V_{S,S'}| - 1 \leq |S \cap S'| \leq |V_{S,S'}| \leq 125 \).
s and hence \( R = 1 \). Furthermore \( \Gamma \leq 2 \), and therefore \( \Lambda \leq 2 \). Therefore \( B \leq 16 \). For all \( v \in V(S) \) we have \( \mathbb{E}_{\mathcal{S}'}[\mathbb{I}(v \in \mathcal{V}_{S,S'})] = s/d \) which implies \( \mathbb{E}_{\mathcal{S}'}[|\mathcal{V}_{S,S'}|] = s^2/d \) for all \( S \in \mathcal{C} \), and therefore \( M_B(C, G_0) = \log(d/s^2) \). An application of Theorem 3.2.2 shows the statement. \( \square \)

### B.4 Multi-Edge Divider Proofs

Before proceeding with the proofs we remind the reader definition (3.2.3) of \( A_{S,S'} = A_0 + A_S + A_{S'} \). In what follows we will use the shorthand notation:

\[
\mathcal{V}_{S,S'|S} = \mathcal{V}_{S',S} \cap V(S).
\]

**Proof of Theorem 3.2.1.** We denote by \( \mathcal{L} := \exp(M(C, d_{G_0}, \log|\mathcal{C}|)) \) the cardinality of the \((\log|\mathcal{C}|)\)-packing. Using Proposition B.4.1 with the constants from Setting 1, and the fact that we have a \((\log|\mathcal{C}|)\)-packing it suffices to control:

\[
\frac{2}{\mathcal{L}^2} \sum_{d_{G_0}(S,S') \geq \log|\mathcal{C}|} \exp \left( \frac{n |\mathcal{V}_{S,S'|S}||A_S||A_{S'}||2(2\|A_{S,S'}\|_2\theta)^2d_{G_0}(S,S')\theta^2}{2d_{G_0}(S,S') + 2} \right)
\]

\[
+ \frac{2}{\mathcal{L}^2} \sum_{d_{G_0}(S,S') = 0} \exp \left( n|S \cap S'|\theta^2 + \frac{n |\mathcal{V}_{S,S'|S}||A_S||A_{S'}||2(2\|A_{S,S'}\|_2\theta)^2\theta^2}{4} \right),
\]

When the cardinality of each \(|S| \leq U\) the above expression can further be controlled by:

\[
\frac{2}{\mathcal{L}^2} \sum_{(S,S'):d_{G_0}(S,S') \geq \log|\mathcal{C}|} \exp \left( \frac{2nU^3(2\Lambda_0\theta)^2d_{G_0}(S,S')\theta^2}{2d_{G_0}(S,S') + 2} \right)
\]

\[
+ \frac{2}{\mathcal{L}} \exp \left( nU \theta^2 + nU^3(\Lambda_0 \theta)^2 \theta^2 \right),
\]

where \( \Lambda_0 := \|A_0\|_2 + 2U \), and we used the facts that \(|S \cap S'|, \|A_S\|_2, \|A_{S'}\|_2 \leq U, |\mathcal{V}_{S',S'|S}| \leq |V(S)| \leq 2U\) and that \( d_{G_0}(S,S') = 0 \) is only possible when \( S \equiv S' \). We first deal with the
term $I_2$. For values of $\theta < \frac{1}{2\Lambda_0}$, we have $I_2 \leq \frac{2}{\mathcal{L}} \exp(2nU\theta^2)$. Hence when $\theta \leq \kappa \frac{\log \mathcal{L}}{nU}$ for some sufficiently small $0 < \kappa$ we have $I_2 = o(1)$. Next, we proceed similarly to the proof of Step 4 in Theorem 3.1.1. Observe the identity

$$I_1 \leq \frac{2}{\mathcal{L}^2} \sum_{(S,S'): d_{G_0}(S,S') \geq \log |C|} \exp \left( \log \mathcal{L} \frac{\kappa^2 U (2\Lambda_0 \theta)^{2d_{G_0}(S,S')}}{d_{G_0}(S,S') + 1} \right),$$

which follows since $\theta \leq \kappa \frac{\log \mathcal{L}}{nU}$. Using $\theta < \frac{1}{2\Lambda_0}$, $\mathcal{L} < |C|$ for $r \geq \log |C|$ we have

$$\log \mathcal{L} \frac{\kappa^2 U (2\Lambda_0 \theta)^{2r}}{r + 1} < \log \mathcal{L} \frac{\kappa^2 U - 2r + 1}{r + 1} < 1,$$

provided that $\kappa$ is small enough (e.g. $\kappa < 1$). By the inequities $e^x \leq 1 + 3x$ for $x \leq 1$ and $\log x \leq x - 1$ we have

$$I_1 \leq \frac{2}{\mathcal{L}^2} \sum_{(S,S'): d_{G_0}(S,S') \geq \log |C|} 1 + \frac{3(\mathcal{L} - 1)(2\Lambda_0 \theta)^{2d_{G_0}(S,S') - 1}}{d_{G_0}(S,S') + 1}$$

Let $K_r := \{|(S, S')| S, S' \in C, d_{G_0}(S, S') = r\}$. We have

$$I_1 < \left(1 - \frac{1}{\mathcal{L}}\right) + \frac{12(2\Lambda_0 \theta)^{2[\log \mathcal{L}] - 1}}{1 - (2\Lambda_0 \theta)^2} \mathcal{L}^{-1} \max_{[\log \mathcal{L}] + 1 \leq r} \frac{K_r}{r}.$$

Paying closer attention to the second term we have:

$$\frac{12(2\Lambda_0 \theta)^{2[\log \mathcal{L}] - 1}}{1 - (2\Lambda_0 \theta)^2} \mathcal{L}^{-1} \max_{[\log \mathcal{L}] + 1 \leq r} \frac{K_r}{r} \leq 16(2\Lambda_0 \theta)^{2[\log \mathcal{L}] - 1} \frac{\mathcal{L}}{r} \leq 16(2\Lambda_0 \theta)^{2[\log \mathcal{L}] - 1} \frac{\mathcal{L}}{2} \leq 16(2\Lambda_0 \theta)^{2[\log \mathcal{L}] - 1} = o(1),$$

with the last equalities hold for $2\Lambda_0 \theta \leq \frac{1}{2} < \exp(-1/2)$. This completes the proof. \qed
Proof of Theorem 3.2.2. Note that by the definition of $B$, the fact that $|V_{S,S'|S}| \lor |V_{S,S'|S'}| \leq |V_{S,S'}|$, we have:

$$\max_{S,S' \in \mathcal{C}} \left( (\|A_S\|_2\|A_{S'}\|_2\|A_{S,S'}\|_2^2) \land (|V_{S,S'|S'}|\|A_{S,S'}\|_1^2) \right) \leq B.$$  

Therefore using Proposition B.4.1 it suffices to control:

$$\overline{D}_{\chi^2} := \frac{1}{|\mathcal{C}|^2} \sum_{S,S' \in \mathcal{C}} \exp \left[ |V_{S,S'|S}|n \theta^2 \left( R + B \theta^2 \right) \right].$$

First note that by $|V_{S,S'|S}| = \sum_{v \in V(S)} 1(v \in V_{S,S'})$, we have:

$$\overline{D}_{\chi^2} \leq \frac{1}{|\mathcal{C}|^2} \sum_{S,S' \in \mathcal{C}} \exp \left[ n \theta^2 \left( R + B \theta^2 \right) \sum_{v \in V(S)} 1(v \in V_{S,S'}) \right].$$

Denote by $\mathbb{P}_{S'}$ the measure induced by drawing $S'$ uniformly from $\mathcal{C}$. Under the assumption: $\theta < \sqrt{R/B}$, and using the fact that the random variables $\{1(v \in V_{S,S'}) \mid v \in V(S)\}$ are negatively associated for every fixed $S \in \mathcal{C}$, we obtain:

$$\log \overline{D}_{\chi^2} \leq \max_{S \in \mathcal{C}} \left[ \sum_{v \in V(S)} \log \left[ \exp(2Rn \theta^2)\mathbb{P}_{S'}(v \in V_{S,S'}) + (1 - \mathbb{P}_{S'}(v \in V_{S,S'})) \right] \right]$$

$$\leq \max_{S \in \mathcal{C}} \left[ (\exp(2Rn \theta^2) - 1) \sum_{v \in V(S')} \mathbb{P}_{S'}(v \in V_{S,S'}) \right]$$

$$\leq \exp(2Rn \theta^2) \max_{S \in \mathcal{C}} \mathbb{E}_{S'} |V_{S,S'}|,$$

where the expectation $\mathbb{E}_{S'}$ is taken with respect to a uniform draw of $S' \in \mathcal{C}$. The first inequality above is due to negative association, the second inequality is due to $\log(1+x) \leq x$. Hence for values of $\theta$

$$\theta \leq \sqrt{\frac{\log \left[ \max_{S \in \mathcal{C}} \mathbb{E}_{S'} |V_{S,S'}|^{-1} \right]}{4nR}},$$

128
we have $D_{\chi^2} \leq \exp(\max_{S \in C} E_S|\mathcal{V}_{S,S'}|^{1/2})$, and therefore:

$$\liminf_n \gamma(S_0(\theta, s), S_1(\theta, s)) \geq 1 - \frac{1}{2} \sqrt{\exp(\max_{S \in C} E_S|\mathcal{V}_{S,S'}|^{1/2}) - 1} = 1,$$

where the last equality holds since $M_B(C, G_0) \to \infty$ implies

$$\max_{S \in C} E_S|\mathcal{V}_{S,S'}| = o(1).$$

\[\square\]

**Proposition B.4.1.** Let $G_0 \in \mathcal{G}_0$ and $C$ be a divider with null base $G_0$. Then for any collection of vertex buffers $V = \{\mathcal{V}_{S,S'}\}_{S,S' \in C}$ and any of the following two settings:

- **S1:** $H_{S,S'} := \frac{(|\mathcal{V}_{S,S'}| \wedge |\mathcal{V}_{S',S'}|)\|A_s\|_2\|A_{s'}\|_2}{\|A_{S,S'}\|_2^2}$ and $K_{S,S'} := 2\|A_{S,S'}\|_2$;
- **S2:** $H_{S,S'} := \frac{|\mathcal{V}_{S,S'}|\|\mathcal{V}_{S',S'}|}{\|A_{S,S'}\|_1^2}$ and $K_{S,S'} := 2\|A_{S,S'}\|_1$.

when the signal strength satisfies

$$\theta < \min_{S,S' \in C} \frac{1 - C^{-1}}{2\sqrt{2}\|A_{S,S'}\|_1}, \quad (B.4.1)$$

we have:

$$\gamma(S_0(\theta, s), S_1(\theta, s)) \geq 1 - \frac{1}{2} \sqrt{\frac{1}{|C|^2} \sum_{S,S' \in C} \exp \left[ n \left( |S \cap S'| \theta^2 + \frac{H_{S,S'}(K_{S,S'}\theta)^{2(d_{G_0}(S,S') \vee 1 + 1)}}{2(d_{G_0}(S,S') \vee 1 + 1)} \right) \right] - 1},$$

**Remark 5.** Proposition B.4.1 continues to hold for parameter classes $\mathcal{U}_s$ not imposing bounds on the $\ell_1$ norm of the precision matrix $\Theta$, if we substitute $\|A_{S,S'}\|_1$ with $\|A_{S,S'}\|_2$ in (B.4.1). In fact tracking the proof, it is easy to see that the theorem also remains valid for
parameter spaces such that $I + \theta A_0 \in \mathcal{S}_0(\theta, s)$ and $I + \theta (A_0 + A_S) \in \mathcal{S}_1(\theta, s)$ for all $S \in C$, with (B.4.1) replaced by $\theta < \min_{S, S' \in C} \frac{1 - C^{-1}}{\|A_{s,s'}\|_2 \sqrt{2\mathcal{K}_{S,S'}}}$.

**Definition B.4.1.** Let $C$ be a divider with null base $G_0 \in \mathcal{G}_0$ whose adjacency matrix is $A_0$. We call a set of constants $\{H_{S,S'}, K_{S,S'} \mid S, S' \in C\}$ **admissible** with respect to the pair $(G_0, C)$ if for all even integers $k \geq 4$ the following holds:

$$\text{Tr}(A_{S,S'}^k + A_0^k - (A_0 + A_S)^k - (A_0 + A_{S'})^k) \leq H_{S,S'} K_{S,S'}^k, \quad (B.4.2)$$

for all $S, S' \in C$.

We will see that any admissible set of constants $\{H_{S,S'}, K_{S,S'} \mid S, S' \in C\}$ yields the lower bound on the minimax risk claimed by Proposition B.4.1.

**Proof of Proposition B.4.1.** We will in fact show a slightly stronger result than presented, involving the following additional combinatorial quantity:

$$\Delta_{S,S'} := |\{\text{triangles in } G(A_{S,S'}) \text{ with } \geq 1 \text{ edges in } S \text{ and } \geq 1 \text{ edges in } S'\}|.$$

**Step 1 (Matrix Construction).**

In this step we construct a set of precision matrices and argue that they fall into the parameter set $\mathcal{U}_\theta$. Take $\Theta_0 = I + \theta A_0$, $\Theta_S = I + \theta (A_0 + A_S)$, $\Theta_{S,S'} = I + \theta A_{S,S'}$, for $S, S' \in C$ and some $\theta > 0$. For any $S, S' \in C$ we have:

$$\max(\|A_0\|_2, \|A_0 + A_S\|_2, \|A_{S,S'}\|_2) \leq \|A_{S,S'}\|_2 \leq \|A_{S,S'}\|_1$$

$$\max(\|A_0\|_1, \|A_0 + A_S\|_1, \|A_{S,S'}\|_1) \leq \|A_{S,S'}\|_1,$$

where these inequalities hold since all matrices $A_0, A_0 + A_S$ and $A_{S,S'}$ consist only of non-negative entries.
Similarly to the proof of Theorem 3.1.1 we can make sure that the matrices $\Theta_0$ and $\Theta_S$ fall into the set $\mathcal{U}_s$ and in addition the matrix $\Theta_{S,S'}$ is strictly positive definite if $\theta < \frac{1-C^{-1}}{\Gamma}$. Thus by assumption the graphs $G(\Theta_0) \in \mathcal{G}_0$ and $G(\Theta_S) \in \mathcal{G}_1$ for all $S \in \mathcal{C}$, and hence $\Theta_0 \in S_0(\theta,s)$ and $\Theta_S \in S_1(\theta,s)$ for all $S \in \mathcal{C}$. In addition we also have that matrices $\Theta_0$, $\Theta_S$, $\Theta_{S,S'}$ are strictly positive definite for any $S, S' \in \mathcal{C}$.

**Step 2 (Risk and Trace Bounds).**

In this step we will lower bound the risk, and will further derive some combinatorial bounds on the traces of powers of adjacency matrices. These bounds are more detail tracking compared to bounds discussed in Theorem 3.1.1. Similarly to Step 2 of the proof of Theorem 3.1.1 it suffices to bound:

$$
\frac{\left( \det(I + \theta(A_0 + A_S)) \right)^{n/2}}{\det(I + \theta A_0)} \left( \frac{\left( \det(I + \theta(A_0 + A_{S'})) \right)^{n/2}}{\det(I + \theta A_{S,S'})} \right) = \exp \left( \frac{n}{2} \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k} \text{Tr} \left( A_{S,S'}^k + A_0^k - (A_0 + A_S)^k - (A_0 + A_{S'})^k \right) \right)
$$

Similarly to Step 3 of Theorem 3.1.1 it is easy to argue that for any $k \in \mathbb{N}$ we have:

$$
\text{Tr}(A_{S,S'}^k + A_0^k - (A_0 + A_S)^k - (A_0 + A_{S'})^k) \geq 0.
$$

We will consider three cases: (1) $k < 2(d_{G_0}(S, S') + 1)$, (2) $k < 4$ and $k \geq 2(d_{G_0}(S, S') + 1)$ and (3) $k \geq 4$ and $k \geq 2(d_{G_0}(S, S') + 1)$. For $k < 2(d_{G_0}(S, S') + 1)$, similarly to the argument in the Step 3 of the proof of Theorem 3.1.1, the above is in fact an equality. However, for the case $k < 4$ and $k \geq 2(d_{G_0}(S, S') + 1)$, we will instead show the following two more precise bounds:

$$
\text{Tr}(A_0^2 + A_{S,S'}^2 - (A_0 + A_S)^2 - (A_0 + A_{S'})^2) \leq 4|S \cap S'|, \quad (B.4.3)
$$

$$
\text{Tr}(A_0^3 + A_{S,S'}^3 - (A_0 + A_S)^3 - (A_0 + A_{S'})^3) \geq 6\Delta_{S,S'}. \quad (B.4.4)
$$
The left hand side of (B.4.3) contains edges lying only in the intersection \( S \cap S' \) since all closed walks containing at least one edge in \( G_0 \) cancel out. By definition the number of such edges is \(|S \cap S'|\). In addition, each closed walk of length 2 in (B.4.3), has precisely one edge in \( S \) and one edge in \( S' \). Fixing the first edge to be in \( S \) and the second edge to be in \( S' \) we notice that each edge appears twice in the total count — once for each of its two vertices. Further multiplying by 2 to adjust for the ordering of the edges we obtain \( 4|S \cap S'| \). Next, notice that expression (B.4.4) contains only closed walks which are triangles. In addition, similarly to the logic above, only walks containing one edge from \( S \) and \( S' \) survive in (B.4.4). Further, each triangle is contained at least 6 times — once per each of its vertices and once per its 2 orientations. This completes the proof of (B.4.4).

We will check the third case in the next step by checking the admissibility in Definition B.4.1.

**Step 3 (Verifying Admissibility).**

In this step we show that both settings below are admissible for any pair \((G_0, C)\):

\[
\begin{align*}
S1: \quad & \mathcal{H}_{S,S'} := \frac{(|V_{S,S'}| |V_{S,S'}|)}{\|A_{S,S'}\|_2^2} \text{ and } K_{S,S'} := 2\|A_{S,S'}\|_2; \\
S2: \quad & \mathcal{H}_{S,S'} := \frac{|V_{S,S'}|}{\|A_{S,S'}\|_1} \text{ and } K_{S,S'} := 2\|A_{S,S'}\|_1.
\end{align*}
\]

Before we prove that the constants in Settings 1 and 2 are admissible, we will show a simple and general bound on closed walks over a sequence of graphs. Let \( E_1, \ldots, E_j \subset \bar{E} \) be fixed edge sets and \( G_1 = (\bar{V}, E_1), G_2 = (\bar{V}, E_2), \ldots G_j = (\bar{V}, E_j) \) be graphs with vertex set \( \bar{V} \), adjacency matrices \( A_1, \ldots, A_j \). Denote \( w_{ii} \) as the number of closed walks of length \( j \) starting and ending at vertex \( i \) such that its \( \ell \)-th edge locates on \( G_\ell \) for all \( \ell \in [j] \). Note that \( w_{ii} \), precisely equals to the \((i, i)\)-th entry of the matrix \( A \), where \( A = \prod_{\ell \in [j]} A_\ell \), i.e.

\[
w_{ii} = A_{ii} = \left( \prod_{\ell \in [j]} A_k \right)_{ii} \leq \|A\|_2 \leq \prod_{\ell \in [j]} \|A_\ell\|_2. \tag{B.4.5}
\]
We conclude that for any fixed vertex \( i \in \mathcal{V} \), we have that the number of closed walks starting and ending at vertex \( i \) walking on the edges of \( G_\ell \) for \( \ell \in [j] \) is at most: \( \prod_{\ell \in [j]} \|A_\ell\|_2 \).

Following the above argument, we will prove below the admissibility of the constants in Setting 1. More precisely we will prove the following

\[
\text{Tr}(A_0^k + A_{S,S'}^k - (A_0 + A_S)^k - (A_0 + A_{S'})^k) \\
\leq \left[ 2^{(k)} \left( |\mathcal{V}_S \cup S'| \wedge |\mathcal{V}_{S'}| \right) \|A_S\|_2 \|A_{S'}\|_2 \right] \|A_{S,S'}\|_2^k, 1 \tag{B.4.6}
\]

which implies the admissibility of Setting 1, by the trivial bound \( 2^{(k)} \leq 2^k \).

To prove (B.4.6), we remind the reader that the trace of an adjacency matrix, counts the number of closed walks of length \( k \) in the graph. A walk will only be counted in the LHS of (B.4.6), if it contains an edge from the set \( S \) and another edge from the set \( S' \). In the remainder of this step of the proof, we will bound the number of closed walks containing edges from both \( S \) and \( S' \).

Denote \( C_{S,S'}^{(k)} = \{ \text{closed walks } C \text{ of length } k \text{ on } G(A_{S,S'}) \text{, with edges } e, e' \in C, e \neq e', e \in S, e' \in S' \} \). We will denote closed walks of length \( k \) by \( C = v_0^C \to v_1^C \to \ldots \to v_{k-1}^C \to v_0^C \), where \( v_j^C \) is the \( j \)th vertex of \( C \) and \((v_j^C, v_{j+1}^C)\) is its \( j \)th edge, and indexation is taken modulo \( k \). All indexation below will also be taken modulo \( k \) where applicable.

For any \( 0 \leq i, j < k, i \neq j \), we first count the number of closed walks in the set \( C_{k}(v, o, i, j) = \{ C \in C_{S,S'}^{(k)} | v_{i+1}^C = v, (v_i^C, v_{i+1}^C) \in S, (v_j^C, v_{j+1}^C) \in S', |i + 1 - j| \wedge (k - |i + 1 - j|) \} \).
mod $k$) is min from any edge in $S$ to any edge in $S'$ on $C$. Following (B.4.5), we have

$$|C_k(v, i, j)| \leq \|A_S\|_2 \|A_{S'}\|_2 \|A_{S, S'}\|_2^{k-2}. \quad (B.4.7)$$

Note that by the definition of $V_{S, S'|S}$ we have $v^C_{i+1} \in V_{S, S'|S}$. To see this first consider the path segment subset of $C$, connecting $v^C_{i+1}$ with $v^C_j$ (see Fig B.3 for a visualization). All edges on the path between $v^C_{i+1}$ and $v^C_j$ belong to the set $E_0$, or else it will not hold that it is the shortest possible path from an edge in $S$ to and edge in $S'$ on $C$. Hence vertex $v^C_{i+1} \in V(E_0 \cup S')$ (the $S'$ comes since possibly $i + 1 = j$). On the other hand since $(v^C_i, v^C_{i+1}) \in S$ we also have $v^C_{i+1} \in V(E_0 \cup S') \cap V(S)$ and thus $v^C_{i+1} \in V_{S, S'|S}$.

$$C_k(i, j) := \bigcup_{v \in V} C_k(v, i, j) \subseteq \bigcup_{v \in V_{S, S'|S}} C_k(v, i, j).$$

Hence:

$$|C_k(i, j)| \leq |V_{S, S'|S}| \|A_S\|_2 \|A_{S'}\|_2 \|A_{S, S'}\|_2^{k-2},$$

and therefore by symmetry:

$$|C_k(i, j)| \leq (|V_{S, S'|S}| \land |V_{S, S'|S'}|) \|A_S\|_2 \|A_{S'}\|_2 \|A_{S, S'}\|_2^{k-2}, \quad (B.4.8)$$

We now observe that:

$$C^{(k)}_{S, S'} \subseteq \bigcup_{0 \leq i, j < k} C_k(i, j)$$

Hence: applying (B.4.8) we obtain:

$$|C^{(k)}_{S, S'}| \leq 2 \binom{k}{2} (|V_{S, S'|S}| \land |V_{S, S'|S'}|) \|A_S\|_2 \|A_{S'}\|_2 \|A_{S, S'}\|_2^{k-2}$$

which completes the proof of (B.4.6).
To check the admissibility for Setting 2, just as before we must have two edges in \( S \) and \( S' \) respectively within a closed walk of length \( k \) which is counted in the LHS of (B.4.2). Then, by the definition of vertex buffer set \( \mathcal{V}_{S,S'} \), we certainly have two vertices in the set \( \mathcal{V}_{S,S'} \) (one in \( \mathcal{V}_{S,S'|S} \) and one in \( \mathcal{V}_{S,S'|S'} \)). Notice that each vertex on the path is of degree at most \( \| A_{S,S'} \|_1 \), and hence can give rise to at most \( \| A_{S,S'} \|_1 \) continuations of the path. Therefore, having fixed two vertices from the sets \( \mathcal{V}_{S,S'|S} \) and \( \mathcal{V}_{S,S'|S'} \), we are left with at most \( \| A_{S,S'} \|_1^{k-2} \) paths. Taking into account that we can position the two vertices on at most \( k(k-1) \leq 2^k \) spots completes the proof.

**Step 4 (Proof Completion).**

In this step we complete the proof by arguing that if we are given any set of admissible constants \( \{ \mathcal{H}_{S,S'}, \mathcal{K}_{S,S'} \} \) for each \( (S,S') \in \mathcal{C} \), the bound on the minimax risk from the Theorem statement follows.

\[
\sum_{k=1}^{\infty} \theta^k \text{Tr} \left( A_{S,S'}^k + A^k_0 - (A_0 + A_S)^k - (A_0 + A_{S'})^k \right) / k \\
\leq 2|S \cap S'| \theta^2 - 2 \Delta_{S,S'}^2 \theta^3 + \mathcal{H}_{S,S'} \sum_{2|k, \ k \geq 2(d_{G_0}(S,S') \lor 1 + 1)} \frac{(\mathcal{K}_{S,S'} \theta)^k}{k} \\
\leq 2|S \cap S'| \theta^2 - 2 \Delta_{S,S'}^2 \theta^3 + \frac{2 \mathcal{H}_{S,S'}(\mathcal{K}_{S,S'} \theta)^2(d_{G_0}(S,S') \lor 1 + 1)}{2(d_{G_0}(S, S') \lor 1 + 1)},
\]

where in the last inequality we used that \( (\mathcal{K}_{S,S'} \theta)^2 \leq \frac{1}{2} \) by (B.4.1). In step 3 we verified that the constants \( \{ \mathcal{H}_{S,S'}, \mathcal{K}_{S,S'} \} \) given in the Theorem statement are admissible for \( (G_0, \mathcal{C}) \) in the sense of Definition B.4.1. This completes the proof. \( \square \)

**B.5 Clique Detection Test**

Here we devise a test to match the lower bound of Example 3.2.3. Unlike previous tests, the clique detection test is not computationally feasible. Recall that \( \hat{\Sigma} = n^{-1} \sum_{i=1}^{n} X_i \otimes X_i \).
Define $\hat{\lambda}_{\min} = \min_{|C| = s} \lambda_d(\hat{\Sigma}_{CC})$, where $\hat{\Sigma}_{CC}$ is a sub-matrix of $\hat{\Sigma}$ with both column and row indices in $C$, and $\lambda_d(A)$ is the smallest eigenvalue of the matrix $A$. Consider the test $\psi = 1(\hat{\lambda}_{\min} < \nu)$, where $\nu$ is

$$\nu := \left(1 - (\sqrt{2} + 1)\sqrt{(s \log(ed/s) + \log(2\alpha^{-1}))/n}\right)^2,$$

for some small constant $\alpha \geq 0$. We have

**Proposition B.5.1.** Suppose that $(s \log(ed/s) + \log(2\alpha^{-1}))/n = o(1)$ for a constant $\alpha \in (0,1)$. Then for values of $\theta \in (0,1)$ satisfying $\theta > \kappa \sqrt{\frac{\log(ed/s)}{sn}}$, for an absolute constant $\kappa$, we have

$$\limsup_{n \to \infty} \sup_{\Theta \in S_0} \mathbb{P}_\Theta(\text{reject } H_0) \leq \alpha \liminf_{n \to \infty} \inf_{\Theta \in S_1(\theta,s)} \mathbb{P}_\Theta(\text{reject } H_0) = 1.$$

**Proof.** The proof relies on an implication of Gordon’s comparison theorem for Gaussian processes, see (Vershynin, 2012, Corollary 5.35 e.g.). By this result we have that:

$$\sqrt{\lambda_{\min}(\hat{\Sigma}_{CC})} \geq 1 - \frac{s}{n} - t,$$

with probability at least $1 - 2\exp(-nt^2/2)$ for any $t \geq 0$. Using the union bound in conjunction to the standard inequality $(\frac{d}{s}) \leq (\frac{ed}{s})^s$, by setting $t = \sqrt{2^{\frac{s \log(ed/s) + \log\alpha^{-1}}{n}}}$, we can ensure that (B.5.2) holds. \hfill \square

**Lemma B.5.1.** Under $H_1$ we have:

$$\sqrt{\hat{\lambda}_{\min}} \leq \frac{1 + \sqrt{\frac{2}{n}} + \sqrt{\frac{2 \log \alpha^{-1}}{n}}}{\sqrt{1 + (s-1)\theta}} \leq \frac{1 + (\sqrt{2} + 1)\sqrt{\frac{s \log(ed/s) + \log \alpha^{-1}}{n}}}{\sqrt{1 + (s-1)\theta}}, \quad \text{(B.5.1)}$$

with probability at least $1 - 2\alpha$.

**Proof of Lemma B.5.1.** Taking in mind that $\Sigma = \Theta^{-1}$ by a simple calculation one can verify that for the set $C^* = \text{supp}(v)$, $\lambda_{\min}(\Sigma_{C^*C^*}) = [\lambda_{\max}(\Theta_{C^*C^*})]^{-1} = \frac{1}{1+(s-1)\theta}$ with a
corresponding eigenvector \( \frac{\mathbf{w}}{\sqrt{s}} \). Again using Corollary 5.35 of Vershynin (2012), and the fact that for two symmetric psd matrices \( \mathbf{A}, \mathbf{B} \) we have \( \lambda_{\text{min}}(\mathbf{ABA}) \leq \lambda_{\text{min}}(\mathbf{A})^2 \lambda_{\text{max}}(\mathbf{B}) \), we have:

\[
\sqrt{\lambda_d(\tilde{\Sigma}_{C^*C^*})} \leq \frac{1 + \sqrt{\frac{2}{n}} + \sqrt{\frac{2\log \alpha^{-1}}{n}}}{\sqrt{1 + (s - 1)\theta}},
\]

with probability at least \( 1 - 2\alpha \). \( \Box \)

**Proof of Proposition B.5.1.** Combining the results of Lemma B.5.2 and setting \( \alpha = d^{-1} \) in Lemma B.5.1 it suffices to show there will be a gap between the bounds in (B.5.2) and (B.5.1). By simple algebra when

\[
\theta > \kappa \sqrt{\frac{\log(ed/s)}{sn}},
\]

for a sufficiently large \( \kappa \) the gap between (B.5.2) and (B.5.1) is implied, which completes the proof. \( \Box \)

**Lemma B.5.2.** Under \( H_0 \) we have:

\[
\sqrt{\lambda_{\text{min}}} \geq 1 - \sqrt{\frac{s}{n}} - \sqrt{\frac{2\log(ed/s) + \log \alpha^{-1}}{n}} \geq 1 - (\sqrt{2} + 1)\sqrt{\frac{s\log(ed/s) + \log \alpha^{-1}}{n}}, \tag{B.5.2}
\]

with probability no less than \( 1 - 2\alpha \).

### B.6 Proofs for Lower Bounds of Confidence Interval Length

In this section, we give a general framework of the lower bound of confidence interval length.
B.6.1 Proof of Theorem 3.3.1

We define the pre-distance on a graph $G = (V, E)$ between two vertex sets $V_1$ and $V_2$ as $d_G(V_1, V_2) = \text{the length of the shortest path on } G \text{ connecting one of } v_1 \in V_1 \text{ and one of } v_2 \in V_2$, and if there is no such path, we let $d_G(V_1, V_2) = \infty$. In order to prove Theorem 3.3.1, we need the following lemma.

Lemma B.6.1. Given an invariant, if there exists a construction as follows:

1. There exists $G_L = (V, E_L)$ satisfying $\overline{V}_L := V(E_L) \subset V = \{1, \ldots, d\}$ and $\mathcal{I}(G_L) = I_L^*$.
2. There exist $N$ disjoint anchor vertex sets $\{A_j\}_{j=1}^N$ such that $\bigcup_{j=1}^N A_j \subseteq \overline{V}_L$.
3. There exists $\overline{V}_U \subseteq V \setminus \overline{V}_L$ such that for each vertex set $A_j$, we can find an edge set $E_j$ with $V(E_j) \subseteq A_j \cup \overline{V}_U$ and the graph $G_j = (V, E_L \cup E_j)$ has $\mathcal{I}(G_j) = I_U^*$.
4. There exists a universal constant $R$ independent to the such that for any $1 \leq j \leq N$, $G_j$ is $R$-hollow and $|A_j| \leq R$. Moreover, the number of anchor vertex sets has $N \geq d^\gamma$ with some $\gamma > 0$.

Suppose $\forall 1 \leq j \neq k \leq N$, $d_{G_L}(A_j, A_k) = \infty$, $|\overline{V}_U| = o(d^{1/2})$ and $|\overline{V}_L|/d < 1$. there exist constants $C_1$ and $C_2$ such that if

$$\theta \leq C_1 \sqrt{\log d/n} \text{ and } \max_j d_{\max}(G_j) \sqrt{\log d/n} \leq C_2,$$

we have the following lower bound on the confidence interval length

$$\liminf_{n \to \infty} \inf_{[\hat{L}, \hat{U}] \in I(I, \alpha)} \sup_{\Theta \in \mathcal{I}(I^*_L, I^*_U, \theta)} \frac{\mathbb{E}_\Theta[\hat{U} - \hat{L}]}{\text{Oracle Length}(\Theta)} \geq 1 - 2\alpha.$$

We now prove Theorem 3.3.1 using Lemma B.6.1. Let

$$G_L = \bigcup_{j=1}^N G_{L,j} \cup G_L \text{ and } G_U = \bigcup_{j=1}^N G_{L,j} \cup G_U.$$
We choose anchor vertex sets $A_j = V(G_{L,j})$ for $1 \leq j \leq N$. Since $|V(E_L)| = O(1)$ and $G_{L,j}$'s are isomorphic copies of $G_L$, we have $|A_j| = O(1)$ as well. It is easy to check $\bigcup_{j=1}^{N} A_j \subseteq V_L$.

We now construct $G_j$’s as isomorphic copies of $\overline{G_U}$. In order to obtain these copies, we denote the isomorphic map between $G_L$ and $G_{L,j}$ as $\sigma_j : V(G_L) \rightarrow V(G_{L,j})$. We extend the domain of $\sigma_j$ to $V$ by letting $\bar{\sigma}_j(u) = \sigma_j(u)$ if $u \in V(G_L)$ and $\bar{\sigma}_j(u) = u$ otherwise. We construct $E_j := \{(\sigma_j(u), \sigma_j(v)) | (u, v) \in (E_{\overline{U}} \setminus E_L))\}$ and we can see the graph $G_j = (V, E_L \cup E_j)$ is isomorphic to $\overline{G_U}$. Therefore, $\mathcal{I}(G_j) = I_U^*$ for any $1 \leq j \leq N$. Therefore, by Lemma B.6.1, we complete the proof.

**B.6.2 Proof of Lemma B.6.1**

The high level idea of the proof is to reduce the length of confidence interval to the bound of $\chi^2$-divergence of two distributions.

In fact, we will prove a stronger lower bound in a smaller parameter space $\mathcal{U}_s(I_L^*, I_U^*; \theta, \mu)$ defined in (A.1.5). We consider this parameter space because it contains precision matrices $\Theta$ such that $\mathcal{I}(T_\mu(\Theta)) \geq I_L^*$, for $\mu = C \sqrt{\log d/n}$. Namely, part of entries of precision matrix has strong enough signal strength.

We can lower bound the length of confidence interval as

$$
\inf_{[\hat{L}, \hat{U}] \in \mathcal{I}(I, \alpha)} \sup_{\Theta \in \mathcal{U}_s(I_L^*, I_U^*; \theta, \mu)} \frac{\mathbb{E}_\Theta[\hat{U} - \hat{L}]}{I_U^* - I_L^*} \geq \inf_{[\hat{L}, \hat{U}] \in \mathcal{I}(I, \alpha)} \sup_{\Theta \in \mathcal{U}_s(I_L^*, I_U^*; \theta, \mu)} \frac{\mathbb{E}_\Theta[\hat{U} - \hat{L}]}{I_U^* - I_L^*}. \quad (B.6.2)
$$

The following lemma proved in Cai and Guo (2015) reduces the length of confidence interval to the $\chi^2$-divergence. Given two distributions $\mathbb{P}$ and $\mathbb{Q}$, the $\chi^2$-divergence between $\mathbb{P}$ and $\mathbb{Q}$ is defined as

$$
D_{\chi^2}(\mathbb{P}, \mathbb{Q}) = \int \left( \frac{d\mathbb{P}}{d\mathbb{Q}} \right)^2 d\mathbb{Q} - 1.
$$

139
Lemma B.6.2 (Lemma 1, Cai and Guo (2015)). Given any monotone invariant $I$, suppose there exist $\Theta_0, \Theta_1, \ldots, \Theta_N \in U_s(I_L, I_U; \theta, \mu)$ satisfying $I(\Theta_0) = I_L$ and $I(\Theta_j) = I_U$ for all $1 \leq j \leq N$. If $P_L = \sum_{j=1}^{N} P_{\Theta_j}$, we have

$$\inf_{[L, U] \in I(I, I)} \mathbb{E}_{\Theta} [\hat{U} - \hat{L}] \geq (I_U^* - I_L^*) \left( 1 - 2\alpha - \sqrt{D\chi^2(PL, PU)} \right).$$

Given two edge sets $E_1, E_2$, we denote the pre-distance on the graph $G$ between $E_1$ and $E_2$ as $d_G(V_1, V_2) = d_G(V(E_1), V(E_2))$. Combining (B.6.2) with Lemma B.6.2, the proof of Lemma B.6.1 can be deduced from the following theorem.

Theorem B.6.1. Given graphs $G_0 = (V, E_0)$ and $G_1, \ldots, G_N$, we denote the difference edge set as $C = \{E(G_j) \setminus E(G_0)\}_{j=1}^{N}$. Given an edge set $S \in C$, we denote $A_S$ as the adjacency matrix of $S$. We also define $A_0$ as the adjacency matrix of $G_0$ and $A_{S,S'} = A_0 + A_S + A_{S'}$. Let $V_{S,S'} = \{V(E_0 \cup S) \cap V(S')\} \cup \{V(E_0 \cup S) \cap V(S)\}$. Suppose the following assumption holds:

A1: Denote the uniform maximum degree as $\Gamma = \max_{S,S' \in C} \|A_{S,S'}\|_1$ and uniform spectral norm as $\Lambda = \max_{S,S' \in C} \|A_{S,S'}\|_2$. We further define

$$R = \max_{S,S' \in C} \frac{|S \cap S'|}{|V_{S,S'}|} \quad \text{and} \quad B = \max_{S,S' \in C} \left( (\Gamma^2 |V_{S,S'}|) \wedge \Lambda^4 \right).$$

Suppose $S'$ is uniformly sampled from $C$. If for any fixed $S \in C$, $V(S)$ can be split into $\ell$ groups: $V(S) = \bigcup_{j=1}^{\ell} V_j(S)$, so that the random variables $\{ \mathbb{I}(V_{S,S'} \cap V_j(S) \neq \emptyset) \}_{j \in [\ell]}$ with respect to a uniformly sampled $S'$ from $C$ are negatively associated. In other words, for any pair of disjoint sets $I, J \subseteq [\ell]$ and any pair of coordinate-wise nondecreasing functions $f, g$ we have:

$$\text{Cov} \left( f\left( \{ \mathbb{I}(V_{S,S'} \cap V_j(S) \neq \emptyset) \}_{j \in I} \right), g\left( \{ \mathbb{I}(V_{S,S'} \cap V_j(S) \neq \emptyset) \}_{j \in J} \right) \right) \leq 0.$$
Denote the largest cardinality of vertices as $V_{\text{max}} = \max_{S \in C} \max_j |V_j(S)|$. Assume the following holds: $\left[ \max_{S \in C} E_{S'} |V_{S,S'}| \right]^{-1} \to \infty$,

$$\mu \leq \sqrt{\frac{R}{B}} \wedge \frac{1 - C^{-1}}{2\sqrt{2} \Gamma} \text{ and } \theta \leq \sqrt{\frac{\log \left( \left[ \max_{S \in C} E_{S'} |V_{S,S'}| \right]^{-1} \right)}{4nV_{\text{max}} R \wedge \mu}}$$

Under either Setting A or B, for the parameters $\Theta_0 = \mu A_0$, $\Theta_S = \mu A_0 + \theta A_S$ for all $S \in C$, the $\chi^2$-divergence between $P_L = P_{\Theta_0}$ and $P_U = \frac{1}{N} \sum_{S \in C} P_{\Theta_S}$ satisfies

$$\lim_{n \to \infty} D_{\chi^2}(P_U, P_L) = 0.$$ 

The proof follows the same strategy of the proof of Proposition B.4.1. The only difference is that our adjacency matrices $\tilde{A}_0 = (\mu/\theta)A_0$, and $\tilde{A}_{S,S'} = \tilde{A}_0 + A_S + A_{S'}$ have weights on edges while Proposition B.4.1 handles the adjacency matrices all have the same weight one. Therefore, we skip the proof.

### B.6.3 Proofs of Theorem 3.3.2

The proof of Theorem 3.3.2 is same as the one of Lemma B.6.1. We just set $\theta$ in the proof of Lemma B.6.1 as $\mu = \theta$. We change the parameters in Theorem B.6.1 into $\Theta_0 = \theta A_0$, $\Theta_S = \theta A_0 + \theta A_S$ for all $S \in C$. Considering two distributions $P_L = P_{\Theta_0}$ and $P_U = \frac{1}{N} \sum_{j=1}^{N} P_{\Theta_j}$, Theorem B.6.1 gives us $\lim_{n \to \infty} D_{\chi^2}(P_U, P_L) = 0$. Applying Lemma B.6.2, we have

$$\inf_{[\hat{L}, \hat{U}] \in I(\hat{L}, \alpha)} \sup_{\Theta \in [\hat{L}, \hat{U}; \theta]} \mathbb{E}_\Theta [\hat{U} - \hat{L}] \geq (I_U^* - I_L^*)(1 - 2\alpha - \sqrt{D_{\chi^2}(P_U, P_L)}) \to (I_U^* - I_L^*)(1 - 2\alpha).$$

141
We now start to prove Theorem 3.3.2.

\[
\inf_{\hat{U} \in U(I, \alpha)} \sup_{\Theta \in U(I^*_L, I^*_U; \theta, \mu)} E_\Theta[\hat{U} - I(\Theta)] \geq \inf_{\hat{U} \in U(I, \alpha)} \sup_{I(\Theta) = I^*_L} E_\Theta[\hat{U}] - I^*_L
\]

\[
\geq \inf_{\hat{U} \in U(I, \alpha)} \sup_{I(\Theta) = I^*_L} (I^*_U \mathbb{P}_\Theta[\hat{U} = I^*_U] + I^*_L \mathbb{P}_\Theta[\hat{U} < I^*_U]) - I^*_L
\]

\[
\geq \inf_{\hat{U} \in U(I, \alpha)} \sup_{I(\Theta) = I^*_L} (I^*_U - I^*_L) \mathbb{P}_\Theta[\hat{U} = I^*_U].
\]  

(B.6.3)

We aim to bound \( \mathbb{P}_\Theta[\hat{U} = I^*_U] \) by Theorem B.6.1. Since \( I \) is hollow, similarly to the proof of Lemma B.6.1, we can construct \( \Theta_0, \Theta_1, \ldots, \Theta_N \in U(I^*_U, I^*_U; \theta, \mu) \) satisfying \( I(\Theta_0) = I^*_L \) and \( I(\Theta_j) = I^*_U \) for all \( 1 \leq j \leq N \) and the \( \chi^2 \)-divergence between \( \mathbb{P}_L = \mathbb{P}_{\Theta_0} \) and \( \mathbb{P}_U = \frac{1}{N} \sum_{S \in C} \mathbb{P}_{\Theta_S} \) satisfies

\[
\lim_{n \to \infty} D_{\chi^2}(\mathbb{P}_U, \mathbb{P}_L) = 0. \tag{B.6.4}
\]

Next notice that

\[
|\mathbb{P}_L(\hat{U} < I^*_U) - \mathbb{P}_U(\hat{U} < I^*_U)| \leq \text{TV}(\mathbb{P}_L, \mathbb{P}_U) \leq \sqrt{D_{\chi^2}(\mathbb{P}_L, \mathbb{P}_U)},
\]

where the next to last inequality simply follows by the definition of total variation norm and the last inequality is due to Cauchy-Schwartz. Hence we have:

\[
\sup_{I(\Theta) = I^*_L} \mathbb{P}_\Theta[\hat{U} = I^*_U] \geq 1 - \mathbb{P}_L(\hat{U} < I^*_U) 
\]

\[
\geq 1 - \sqrt{D_{\chi^2}(\mathbb{P}_L, \mathbb{P}_U)}, \mathbb{P}_U(\hat{U} < I^*_U) 
\]

\[
\geq 1 - \sup_{I(\Theta) = I^*_L} \mathbb{P}_\Theta(\hat{U} < I^*_U) - \sqrt{D_{\chi^2}(\mathbb{P}_L, \mathbb{P}_U)} \geq 1 - \alpha - \sqrt{D_{\chi^2}(\mathbb{P}_L, \mathbb{P}_U)}.
\]

Combining the above inequality with (B.6.4) and (B.6.3), we prove Theorem 3.3.2.
Appendix C

Proofs in Chapter 4

In this chapter, we provide the proofs of results in Chapter 4.

C.1 Proof of Estimation Consistency

In the appendix, we use $c, C, c_1, C_1, c_2, C_2, \ldots$ to denote universal constants, independent of $n$ and $d$, whose values may change from line to line.

In this section, we prove uniform rates of convergence for the covariance matrix estimator $\hat{\Sigma}(z)$ and the inverse covariance estimator $\hat{\Omega}(z)$. These rates are uniformly valid over both the index $z$ and the kernel bandwidth $h$ used for the estimator $\hat{\tau}_{jk}(z)$ in (4.1.3).

C.1.1 Proof of Theorem 4.3.4

We apply the bias-variance decomposition for the kernel smoothed Kendall’s tau statistic in the following two lemmas. The first lemma controls the variance term $|\hat{\tau}_{jk}(z) - E[\hat{\tau}_{jk}(z)]|$ uniformly in $j, k \in [d], z \in (0, 1)$, and $h \in [h_l, h_u]$. The second lemma controls the bias term $|E[\hat{\tau}_{jk}(z)] - \tau_{jk}(z)|$.

Lemma C.1.1 (Variance of Kendall’s tau estimator). Assume that $n^{-1} \log d = o(1)$ and the bandwidths $0 < h_l < h_u < 1$ satisfy $h_l n / \log (dn) \to \infty$ and $h_u = o(1)$. There exists a
universal constant $C > 0$ such that, with probability $1 - \delta$, for sufficiently large $n$,

$$
\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \frac{\sqrt{nh}}{\log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))} |\hat{\tau}_{jk}(z) - \mathbb{E} [\hat{\tau}_{jk}(z)]| \leq C.
$$

**Lemma C.1.2** (Bias of Kendall’s tau estimator). Assume that the bandwidths $0 < h_l < h_u < 1$ satisfy $h_u = o(1)$. There exists a constant $C > 0$ such that

$$
\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \frac{|\mathbb{E}[\hat{\tau}_{jk}(z)] - \tau_{jk}(z)|}{h^2 + 1/(nh)} \leq C.
$$

We defer the proof of these two lemmas to Appendix C.4.

By the definition of $\hat{\Sigma}(z)$ in (4.1.6), for any $j, k \in [d]$ and $z \in (0, 1)$, we have

$$
|\hat{\Sigma}_{jk}(z) - \Sigma_{jk}(z)| = |\sin(\pi \hat{\tau}_{jk}(z)/2) - \sin(\pi \tau_{jk}(z)/2)| \leq \frac{\pi}{2} |\hat{\tau}_{jk}(z) - \tau_{jk}(z)|, \quad (C.1.1)
$$

where the last inequality is due to $|\sin(x) - \sin(y)| \leq |x - y|$ for any $x, y \in [-\pi/2, \pi/2]$. Therefore, the rate of $\hat{\Sigma}(z)$ can be bounded by the rate of $\hat{\tau}_{jk}(z)$ up to a constant. Recall that $T = [\tau_{jk}]_{jk}$ and $\hat{T} = [\hat{\tau}_{jk}]_{jk}$. We have

$$
\|\hat{T}(z) - T(z)\|_{\max} \leq \sup_{j,k \in [d]} |\hat{\tau}_{jk}(z) - \mathbb{E} \hat{\tau}_{jk}(z)| + \sup_{j,k \in [d]} |\mathbb{E} \hat{\tau}_{jk}(z) - \tau_{jk}(z)|.
$$

Lemma C.1.1 and Lemma C.1.2 together with (C.1.1) give us

$$
\sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \frac{\|\hat{\Sigma}(z) - \Sigma(z)\|_{\max}}{h^2 + \sqrt{(nh)^{-1} \log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))}} \leq \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \frac{\|\hat{T}(z) - T(z)\|_{\max}}{h^2 + \sqrt{(nh)^{-1} \log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))}} \leq 2\pi(C_1 + C_2),
$$

with probability $1 - \delta$, since $1/(nh) = o(1)$. We complete the proof of the theorem by setting $C_\Sigma = 2\pi(C_1 + C_2)$. 144
C.1.2 Proof of Theorem 4.3.5

Using the uniform rate of convergence of $\hat{\Sigma}(z)$ established in Theorem 4.3.4, we establish the corresponding rate for $\hat{\Omega}(z)$ when estimated using the CLIME (Cai et al., 2011) or the calibrated CLIME (Zhao and Liu, 2014). Modifying the proofs in the above two theses, we can establish a bound on $\|\hat{\Omega} - \Omega\|_{\text{max}}$, if $\|\hat{\Sigma} - \Sigma\|_{\text{max}}$ is controlled. For simplicity, we recall the results for the calibrated CLIME estimator. Similar results for the CLIME estimator can be found in the proof of Theorem 6 in Cai et al. (2011).

**Theorem C.1.1** (Adapted from Zhao and Liu 2014). Suppose $\Omega \in \mathcal{U}_s(M, \rho)$ and the tuning parameter satisfies $s\lambda = o(1)$. On the event $\{\|\hat{\Sigma} - \Sigma\|_{\text{max}} \leq \lambda\}$, there exist universal constants $C_1, C_2, C_3$ such that the output of the calibrated CLIME satisfies

$$\|\hat{\Omega} - \Omega\|_{\text{max}} \leq C_1 M^2 \lambda, \quad \|\hat{\Omega} - \Omega\|_1 \leq C_2 sM \lambda \quad \text{and} \quad \max_{j \in [d]} \|\hat{\Sigma}\hat{\Omega}_j - e_j\|_{\infty} \leq C_3 \lambda M.$$  

The formal statements of Theorem IV.1 and Theorem IV.2 in Zhao and Liu (2014) are not the same as the statement above. The result of Theorem C.1.1 is more general and directly follows from the proofs of Theorem IV.1 and Theorem IV.2 in Zhao and Liu (2014). For example, the last inequality in Theorem C.1.1 follows from Equation (E.12) of Zhao and Liu (2014).

Let $\lambda_{n,h} = C\Sigma\left(h^2 + \sqrt{\log(dn)/(nh)}\right)$ and define the event

$$\mathcal{E} = \left\{ \sup_{h \in [h_l, h_u]} \sup_{z \in (0,1)} \lambda_{n,h}^{-1/2} \|\hat{\Sigma}(z) - \Sigma(z)\|_{\text{max}} \leq 1 \right\}.$$
Since the constants $C_1, C_2$ and $C_3$ in Theorem C.1.1 are universal and the penalty parameter $\lambda \geq \lambda_{n,h}$, it follows that

$$\sup_{h \in [h_l, h_u]} \sup_{z \in (0,1)} \lambda^{-1} \left\| \mathbf{\hat{\Omega}}(z) - \mathbf{\Omega}(z) \right\|_{\max} \leq C_1 M^2;$$

$$\sup_{h \in [h_l, h_u]} \sup_{z \in (0,1)} \lambda^{-1} \left\| \mathbf{\hat{\Omega}}(z) - \mathbf{\Omega}(z) \right\|_{1} \leq C_2 s M;$$

$$\sup_{z \in (0,1)} \max_{j \in [d]} \lambda^{-1} \cdot \left\| \mathbf{\hat{\Omega}}^T_j \mathbf{\hat{\Sigma}} - \mathbf{e}_j \right\|_{\infty} \leq C_3 M,$$

on the event $\mathcal{E}$. Theorem 4.3.4 gives us $\mathbb{P}(\mathcal{E}) \geq 1 - 1/d$, which completes the proof.

### C.2 Asymptotic Properties of Testing Statistics

In this section, we prove asymptotic properties of the testing statistics for three kinds of hypothesis tests: (1) edge presence test in Theorem 4.3.1, (2) super-graph test in Theorem 4.3.2 and (3) uniform edge presence test in Theorem 4.3.3.

Let $\mathcal{S}_{j,z} = \{k \in [d] \mid \Omega_{kj}(z) \neq 0\}$. For any index sets $\mathcal{S}, \mathcal{S}' \subset [d]$, we define $\mathbf{\Omega}_{SS'} \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}'|}$ to be the submatrix of $\mathbf{\Omega}$ obtained from rows indexed by $\mathcal{S}$ and columns indexed by $\mathcal{S}'$. For any function $f(x)$, we define

$$G_n^\xi[f] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i) \cdot \xi_i,$$  \hspace{1cm} (C.2.1)

where $\xi_1, \ldots, \xi_n \sim N(0, 1)$. We also define $\mathbb{P}_\xi(\cdot) := \mathbb{P}(\cdot\mid \{Y_i\}_{i \in [n]})$ and $\mathbb{E}_\xi[\cdot] := \mathbb{E}[\cdot\mid \{Y_i\}_{i \in [n]}]$.

At a high-level, we establish the asymptotic results by carefully studying the Hoeffding decomposition of the kernel smoothed Kendall’s tau estimator. From (4.1.3), we observe that $\hat{\tau}_{jk}(z)$ is a quotient of two $U$-statistics. To study this quotient, we introduce some additional notation. For a fixed $j, k \in [d]$, we define the following bivariate function

$$g_{z|j,k}(y_i, y_i') = \omega_z(z_i, z_i') \text{sign}(x_{ij} - x_{i'j}) \text{sign}(x_{ik} - x_{i'k}),$$  \hspace{1cm} (C.2.2)

146
for \( y_i = (z_i, x_i) \) and \( y_{i'} = (z_{i'}, x_{i'}) \). Recalling the definition of \( \omega_z(z_i, z_{i'}) \) in (4.1.4), we have that

\[
\hat{\tau}_{jk}(z) = \frac{U_n[g_{z|j,k}]}{U_n[\omega_z]}.
\]

Let us define the Hoeffding decomposition of \( g_{z|j,k} \) as

\[
g^{(1)}_{z|j,k}(y) = \mathbb{E}[g_{z|j,k}(y, Y)] - \mathbb{E}[U_n[g_{z|j,k}]], \tag{C.2.3}
\]

\[
g^{(2)}_{z|j,k}(y_1, y_2) = g_{z|j,k}(y_1, y_2) - g^{(1)}_{z|j,k}(y_1) - g^{(1)}_{z|j,k}(y_2) - \mathbb{E}[U_n[g_{z|j,k}]]. \tag{C.2.4}
\]

Then we can reformulate the centered \( U \)-statistic as

\[
U_n[g_{z|j,k}] - \mathbb{E}[U_n[g_{z|j,k}]] = 2\mathbb{E}_n[g^{(1)}_{z|j,k}(Y_i)] + U_n[g^{(2)}_{z|j,k}]. \tag{C.2.5}
\]

In the above display, we decomposed the centered \( U \)-statistic into an empirical process and a higher order \( U \)-statistic. Our proof strategy is to study the asymptotic property of the leading empirical process term \( 2\mathbb{E}_n[g^{(1)}_{z|j,k}(Y_i)] \) and show that the higher order term can be ignored. Similarly, let

\[
\omega^{(1)}_z(s) = \mathbb{E}[\omega_z(s, Z)] - \mathbb{E}[U_n[\omega_z]], \tag{C.2.6}
\]

\[
\omega^{(2)}_z(s, t) = \omega_z(s, t) - \omega^{(1)}_z(s) - \omega^{(1)}_z(t) - \mathbb{E}[U_n[\omega_z]], \tag{C.2.7}
\]

which leads to the following Hoeffding decomposition

\[
U_n[\omega_z] - \mathbb{E}[U_n[\omega_z]] = 2\mathbb{E}_n[\omega^{(1)}_z] + U_n[\omega^{(2)}_z]. \tag{C.2.8}
\]
According to the heuristic approximation of \( \hat{S}_{z|j,k}(\hat{\Omega}_{k\setminus j}(z_0)) \) in (4.2.3), we have

\[
\hat{S}_{z|j,k}(\hat{\Omega}_{k\setminus j}(z)) \approx \Omega^T_j(z) \left[ \hat{\Sigma}(z) \odot (\hat{T}(z) - T(z)) \right] \Omega_k(z)
\approx \left[ U_n[\omega_z] \right]^{-1} \sum_{u,v \in [d]} \Omega_{ju}(z) \Omega_{kv}(z) \pi \cos \left( \tau_{uv}(z) \frac{\pi}{2} \right) \cdot \left[ U_n[g_{z|u,v}] - \mathbb{E}[U_n[g_{z|u,v}]] : U_n[\omega_z] \right],
\]

(C.2.9)

where the last “\( \approx \)” comes from (4.1.3) and Lemma C.1.2. Combining (C.2.5) and (C.2.8) with (C.2.9),

\[
\sqrt{n}h \cdot \hat{S}_{z|j,k} \approx \left[ U_n[\omega_z] \right]^{-1} \mathbb{G}_n \left[ J_{z|j,k} \right],
\]

where the leading term of the Hoeffding decomposition is defined as

\[
J_{z|j,k}(y') := \sum_{u,v \in [d]} \Omega_{ju}(z) \Omega_{kv}(z) \pi \cos \left( \tau_{uv}(z) \frac{\pi}{2} \right) \sqrt{n} \cdot \left[ g_{z|u,v}^{(1)}(y') - \tau_{uv}(z) \omega_z^{(1)}(z') \right],
\]

(C.2.10)

for any \( y' = (z', x') \). We find the asymptotic distribution of \( \left[ U_n[\omega_z] \right]^{-1} \mathbb{G}_n \left[ J_{z|j,k} \right] \) in the next part.

C.2.1 Proof of Theorem 4.3.1

Let the operator \( u_n[\cdot] \) be defined as

\[
u_n[H] = \sqrt{n} \cdot (U_n[H] - \mathbb{E}[U_n[H]])
\]

(C.2.11)

for any bivariate function \( H(x, x') \). In order to prove Theorem 4.3.1, we need the convergence rate of terms related to the Kendall’s tau estimator, especially the two lemmas below.

Lemma C.2.1. Suppose that \( n^{-1} \log d = o(1) \) and the bandwidths \( 0 < h_l < h_u < 1 \) satisfy \( h_l n / \log(dn) \to \infty \) and \( h_u = o(1) \). There exists a universal constant \( C > 0 \) such that with
probability $1 - \delta$,

$$
\sup_{j,k \in [d]} \sup_{h \in [h_l,h_u]} \sup_{z \in (0,1)} \left| \frac{\sqrt{h}}{\sqrt{\log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))}} \left( u_n [\omega_z] \lor u_n [g_{z(j,k)}] \right) \right| \leq C,
$$

(C.2.12)

for large enough $n$.

**Lemma C.2.2.** Suppose the bandwidths $0 < h_l < h_u < 1$ satisfy $h_u = o(1)$. Then

$$
\sup_{z \in (0,1)} \left| \mathbb{E} \left[ U_n [g_{z(j,k)}] \right] - f^2_Z(z) \tau_{jk}(z) \right| = O(h^2),
$$

(C.2.13)

$$
\sup_{z \in (0,1)} \left| \mathbb{E} \left[ U_n [\omega_z] \right] - f^2_Z(z) \right| = O(h^2),
$$

(C.2.14)

$$
\sup_{z \in (0,1)} n^{-1} \mathbb{E} \left[ u_n \left[ g_{z(j,k)} \right] \cdot u_n \left[ \omega_z \right] \right] = O((nh)^{-1}),
$$

(C.2.15)

$$
\sup_{z \in (0,1)} n^{-1} \mathbb{E} \left[ (u_n \left[ \omega_z \right])^2 \right] = O((nh)^{-1}).
$$

(C.2.16)

We defer the proof of the above two lemmas to Appendix C.5 in the supplementary material.

Using Lemma C.2.1 and Lemma C.2.2, we have

$$
\inf_{z \in (0,1)} U_n[\omega_z] \geq \inf_{z \in (0,1)} \mathbb{E}[U_n[\omega_z]] - \sup_{z \in (0,1)} n^{-1/2} |u_n[\omega_z]| \geq f^2_Z/2,
$$

(C.2.17)

$$
\sup_{z \in (0,1)} U_n[\omega_z] \leq \sup_{z \in (0,1)} \mathbb{E}[U_n[\omega_z]] + \sup_{z \in (0,1)} n^{-1/2} |u_n[\omega_z]| \leq 2f^2_Z,
$$

with probability $1 - 1/d$ for sufficiently large $n$. The last inequality is due to the fact that $f_Z$ is bounded from above and below, $h = o(1)$ and $\log(1/h)/nh = o(1)$. 

149
Combining the above display with Lemma C.7.1, we have

\[ \left| \sqrt{n}h \cdot S_{z_{(j,k)}}(\hat{\Omega}_{k\backslash j}(z)) - [U_n[z]]^{-1} G_n \left[ J_{z_{(j,k)}} \right] \right| \leq \left| \sqrt{n}h \cdot U_n[z] S_{z_{(j,k)}}(\hat{\Omega}_{k\backslash j}(z)) - G_n \left[ J_{z_{(j,k)}} \right] \right| \leq 2\int_{\text{Z}}^2 n^{-c}. \]  

(C.2.18)

Therefore, it suffices to derive the limiting distribution of \( G_n \left[ J_{z_{(j,k)}} \right] \).

In order to apply central limit theorem to \( G_n \left[ J_{z_{(j,k)}} \right] \), we check Lyapunov’s condition. By the definition of \( g_{z_{(j,k)}}^{(1)} \) in (C.2.3) and \( \omega_{z_{(j,k)}}^{(1)} \) in (C.2.6), we have

\[ E \left[ J_{z_{(j,k)}}(Y) \right] = 0 \text{ for all } i \in [n]. \]

The matrix \( \Theta_z \) defined in (4.2.4) can be rewritten as

\[ (\Theta_z)_{jk} = \pi \cos \left( \frac{\tau_{jk}(z)^{\pi}}{2} \right) \sqrt{h} \cdot \left[ g_{z_{(j,k)}}^{(1)}(Y) - \tau_{jk}(z)\omega_{z_{(j,k)}}^{(1)}(Z) \right]. \]  

(C.2.19)

In order to apply the Lyapunov condition to show the asymptotic normality, we begin to control the third moments of \( \omega_{z_{(j,k)}}^{(1)} \) and \( g_{z_{(j,k)}}^{(1)} \). We bound the third moment of \( \omega_{z_{(j,k)}}^{(1)} \) by

\[ \sup_z E \left[ \left| \omega_{z_{(j,k)}}^{(1)}(Z) \right|^3 \right] = \sup_z \left| E[K_h(z - Z)] \right|^3 E \left[ |K_h(z - Z) - E[K_h(z - Z)]|^3 \right] \leq \sup_z \left| E[K_h(z - Z)] \right|^3 E \left[ |K_h(z - Z)|^3 \right] = \sup_z \left| f_Z(z) + O(h^2) \right|^3 \cdot h^{-2} \int |K^3(t)|f_Z(z + th)dt \leq h^{-2} \cdot \bar{f}_Z^4 \int |K^3(t)|dt, \]  

(C.2.20)

where we used that \((1 + x)^3 \leq 4(1 + x^3)\) for \( x > 0 \) and \(|E[K_h(z - Z)]|^3 \leq E[|K_h(z - Z)|^3] \).

By the definition of \( g_{z_{(j,k)}}^{(1)} \), we also have

\[ E \left[ \left| g_{z_{(j,k)}}^{(1)}(Y) \right|^3 \right] \leq 4E \left[ \left| E[g_{z_{(j,k)}}(Y', Y) | Y'] \right|^3 \right] + 4 \left| E[g_{z_{(j,k)}}(Y', Y)] \right|^3, \]  

(C.2.21)
where \( Y' \) is an independent copy of \( Y \). Using (C.5.19),

\[
\sup_z \left| \mathbb{E}[g_{z(j,k)}(Y', Y)] \right|^3 = \sup_z \left| f_Z^2(z) \tau_{jk}(z) + O(h^2) \right|^3 \lesssim \bar{r}_2^3.
\]  
(C.2.22)

We now bound the conditional expectation in (C.2.21). From (A.3) of Mitra and Zhang (2014), denoting \( \mathbf{x}' = (x_1', \ldots, x_d')^T \) and \( y' = (z', \mathbf{x}') \), we have

\[
\mathbb{E} \left[ g_{z(j,k)}(y', Y) \bigg| Z = s \right] = K_h(z' - z) K_h(s - z) \varphi \left( x'_j, x'_k, \Sigma_{jk}(s) \right), \text{ where}
\]

\[
\varphi(u, v, \rho) = 2 \int \text{sign}(u - x) \phi(x) \cdot \Phi \left( \frac{v - \rho x}{\sqrt{1 - \rho^2}} \right) dx,
\]  
(C.2.23)

with \( \phi(\cdot) \) and \( \Phi(\cdot) \) being the probability density and cumulative distribution function of a standard normal variable, respectively. From (C.2.3), we have

\[
\mathbb{E} \left[ g_{z(j,k)}(y', Y) \bigg| Z = s \right] = K_h(z' - z) \int K_h(s - z) \varphi \left( x'_j, x'_k, \Sigma_{jk}(s) \right) f_Z(s) ds.
\]  
(C.2.24)

Let \( \phi_{\rho}(x, y) \) be the density function of bivariate normal distribution with mean zero, variance one and correlation \( \rho \). Notice that \( \sup_{x, y, \rho} |\varphi(x, y, \rho)| \leq 2 \). Since the minimum eigenvalue of \( \Sigma(z) \) is strictly positive for any \( z \), there exists a \( \gamma_\sigma < 1 \) such that \( \sup_z |\Sigma_{jk}(z)| \leq \gamma_\sigma < 1 \) for any \( j \neq k \). We also have \( \sup_{x, y, \rho} |\phi_{\rho}(x, y)| \leq (2\pi \sqrt{1 - \gamma_\sigma^2})^{-1} \). By (C.2.24), for any \( z \in (0, 1) \)

\[
\mathbb{E} \left[ \left| \mathbb{E}[g_{z(j,k)}(Y', Y) \big| g'] \right|^3 \right] = 2 \int h^{-2} K^3(t_1) f_Z(z + t_1 h) \times \int K(t_2) f_Z(z + t_2 h) \varphi(u, v, \Sigma_{jk}(z + t_2 h)) dt_2 \left| \phi_{\Sigma_{jk}(z+t_1 h)}(u, v) dt_1 \right|^3 \lesssim 1 \int h^{-2} \pi^2 \frac{K^3}{\sqrt{1 - \gamma_\sigma^2}} dt_1 \leq h^{-2} \cdot \frac{1}{\pi} \frac{\bar{r}_2^3}{\sqrt{1 - \gamma_\sigma^2}}.
\]  
(C.2.25)
Combining (C.2.22), (C.2.25) with (C.2.21), we have

\[ E\left[ |g_{z_0,(j,k)}^{(1)}(Y)|^3 \right] \lesssim h^{-2} \left( \frac{\bar{r}_f^4 \| K \|^3}{\pi \sqrt{1 - \gamma_2^2}} \right). \]  

(C.2.26)

By the assumption of Theorem 4.3.3, there exists a \( \theta_{\text{min}} > 0 \) such that

\[ \text{Var}(J_{z_0,(j,k)}(Y)) = E(\Omega_j^T(z_0) \Theta_{z_0} \Omega_k(z_0))^2 \geq \theta_{\text{min}} \| \Omega_j(z_0) \|^2 \| \Omega_k(z_0) \|^2. \]  

(C.2.27)

We are now ready to check the Lyapunov’s condition. We have

\[ \sum_{i=1}^n E|J_{z_0,(j,k)}(Y_i)|^3 \leq (\theta_{\text{min}} n)^{-3/2} \sum_{i=1}^n E|J_{z_0,(j,k)}(Y_i)|^3 \lesssim \frac{1}{n^{3/2}} \sum_{i=1}^n E\| \text{Vec}((\Theta_{z_0})_{S_{j,z_0},S_{k,z_0}}) \|^3. \]  

(C.2.28)

Since \( |\Theta_{jk}^{(1)}| \leq \pi \sqrt{h}|g_{z_0,(j,k)}^{(1)}(Y)| + \pi \sqrt{h}|\tau_{jk}(z_0)|\omega_{z_0}^{(1)}(Z)| \), we have

\[ E\| \text{Vec}((\Theta_{z_0})_{S_{j,z_0},S_{k,z_0}}) \|^3 \leq |S_{j,z_0}|^{3/2}|S_{k,z_0}|^{3/2} \pi^3 h^{3/2} \left( E\left[ |g_{z_0,(j,k)}^{(1)}(Y)|^3 \right] + |\tau_{jk}(z_0)|^3 E\left[ |\omega_{z_0}^{(1)}(Z)|^3 \right] \right). \]

Using (C.2.20) and (C.2.26), together with \( s^3/\sqrt{nh} = o(1) \),

\[ \sum_{i=1}^n E|J_{z_0,(j,k)}(Y_i)|^3 \leq \frac{s^3}{\sqrt{nh}} = o(1), \]

which implies that the Lyapunov’s condition is satisfied. Moreover, by Lemma C.2.1, for any \( z_0 \in (0, 1) \), \( U_n[\omega_{z_0}] - E[U_n[\omega_{z_0}]] \) converges to 0 in probability. Combining this with (C.2.14) and \( h = o(1) \), we have that \( U_n[\omega_{z_0}] \) converges to \( f_Z^2(z_0) \) in probability. Therefore, by the
central limit theorem and Slutsky’s theorem, for any \( j, k \in [d], \)

\[
\frac{[\mathbb{U}_n[\omega_{z_0}]]^{-1} \mathbb{G}_n [J_{z_0}(j,k)]}{f_Z^{-2}(z_0) \{ \mathbb{E}[(\Omega_j^T \Theta_{z_0} \Omega_k)^2] \}^{1/2}} \xrightarrow{d} N(0, 1).
\]

Combining with (C.2.18), the proof is complete.

### C.2.2 Proof of Theorem 4.3.2

The strategy is to apply the theory for multiplier bootstrap developed in Chernozhukov et al. (2013) to the score function in (4.2.9). A similar strategy is applied to prove Theorem 4.3.3, whose proof is deferred to Section C.6.

Let \( T_0(z) := \max_{(j,k) \in E^c} \mathbb{G}_n [J_{z_0}(j,k)] \) and

\[
S_0^B(z_0) = \max_{(j,k) \in E^c} \mathbb{G}_n^\xi [J_{z_0}(j,k)] = \max_{(j,k) \in E^c} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_{z_0}(j,k)(Y_i) \cdot \xi_i
\]

be the bootstrap counterpart to \( T_0(z_0) \). Recall that \( S^B(z) \) is defined in (4.2.12). We denote

\[
\Delta_z := \max_{(j,k) \in E^c} \left| \frac{1}{n} \sum_{i=1}^{n} \left( J_{z_0}(j,k)(Y_i) J_{z_0}(j',k')(Y_i) - \mathbb{E}[J_{z_0}(j,k)(Y_i) J_{z_0}(j',k')(Y_i)] \right) \right|.
\]

In order to Use Theorem 3.2 in Chernozhukov et al. (2013), we check four conditions.

1. With probability \( 1 - 1/d, \)

\[
|S(z_0) - T_0(z_0)| \leq \sup_{j,k \in [d]} \left| \sqrt{n} h \cdot \mathbb{U}_n[\omega_{z_0}] \hat{S}_{z_0}(j,k) \left( \hat{\Theta}_{k \setminus j}(z_0) \right) - \mathbb{G}_n [J_{z_0}(j,k)] \right| \leq n^{-c}.
\]

2. With probability \( 1 - 1/d, \)

\[
\mathbb{P}_\xi(|S^B(z_0) - S_0^B(z_0)| \leq n^{-c}) \geq 1 - 1/d.
\]

3. There exists a constant \( c > 0, \) such that \( \text{Var}(\mathbb{G}_n [J_{z_0}(j,k)]) > c. \)

4. There exists a constant \( c > 0, \) such that \( \mathbb{P}(\Delta_{z_0} > n^{-c}) \leq n^{-c}. \)
The first condition is proven in Lemma C.7.1. We defer the proof of the second condition in Lemma C.7.2. The third condition is due to (C.2.27). The following of the proof verifies the last condition.

Define

\[
\gamma_{z\lfloor(j,k,j',k')}(Y_i) = J_{z\lfloor(j,k)}(Y_i)J_{z0\lfloor(j',k')}(Y_i) - \mathbb{E}[J_{z\lfloor(j,k)}(Y_i)J_{z\lfloor(j',k')}(Y_i)].
\]  

(C.2.29)

We will apply Lemma A.1 in van de Geer (2008) on the concentration of empirical processes. For the self-consistence, we have restated the lemma in Lemma C.9.2. In order to apply Lemma C.9.2, we need to bound \(\|\gamma_{z\lfloor(j,k,j',k')}\|_{\infty}\) and \(n^{-1} \sum_{i=1}^{n} \mathbb{E}[\gamma_{z\lfloor(j,k,j',k')}(Z_i)]\). By the definition of \(J_{z0\lfloor(j,k)}\) in (C.2.10), we have for any \(z_0 \in (0, 1)\),

\[
\max_{(j,k),(j',k') \in E^c} \|\gamma_{z_0\lfloor(j,k,j',k')}\|_{\infty} \\
\leq \max_{i \in [n]} \max_{(j,k) \in E^c} 2|J_{z_0\lfloor(j,k)}(Y_i)|^2 \\
\leq \max_{(j,k) \in E^c} 2\|\Omega_j(z_0)\|_1^2 \|\Omega_k(z_0)\|_1^2 \pi^2 h \cdot \left(\|g_{z_0\lfloor(j,k)}(Y_i)\|_{\infty} + \|\omega_{z_0}^{(1)}(Z_i)\|_{\infty}\right)^2 \\
\leq CM^2 h^{-1},
\]  

(C.2.30)

where the second inequality follows from Hölder’s inequality, similar to (C.2.28), and the final inequality is due to (C.5.4) and (C.5.14). Since the right hand size of (C.2.30) does not depend on \(z_0\), we also have

\[
\max_{z \in (0,1)} \max_{(j,k),(j',k') \in E^c} \|\gamma_{z\lfloor(j,k,j',k')}\|_{\infty} \leq CM^2 h^{-1} \quad \text{and} \quad \max_{z \in (0,1)} \max_{(j,k),(j',k') \in E^c} \mathbb{E}[\gamma_{z\lfloor(j,k,j',k')(Z_i)}] \leq \max_{z \in (0,1)} \max_{(j,k),(j',k') \in E^c} \|\gamma_{z\lfloor(j,k,j',k')}\|_{\infty}^2 \leq CM^4 h^{-2}.
\]  

(C.2.31)
According to Lemma C.9.2, the expectation of $\Delta z_0$ is bounded by
\[
E[\Delta z_0] \lesssim \sqrt{\frac{2M^4 \log(2d)}{nh^2} + \frac{M^2 \log(2d)}{nh}}.
\]
Since $\log d/(nh^2) = o(n^{-\epsilon})$, there exists $c_1 > 0$ such that $E[\Delta z_0] \leq n^{-2c_1}$ for sufficiently large $n$. By Markov’s inequality, $P(\Delta z_0 > n^{-c_1}) \leq n^{c_1} E[\Delta z_0] \leq n^{-c_1}$ for sufficiently large $n$, which verifies the last condition.

By Theorem 3.2 in Chernozhukov et al. (2013),
\[
\sup_{\alpha \in (0,1)} \left| P_{H_0}(\psi_{z_0} | G(\alpha) = 1) - \alpha \right| \lesssim n^{-c},
\]
for some constant $c > 0$, which completes the proof.

### C.3 Proof of Theorem 4.3.6

In this section, we prove the minimax rate of convergence for estimating time-varying inverse covariance matrices. Section C.3.1 proves the minimax rate in terms of $\| \cdot \|_{\text{max}}$ norm, while Section C.3.2 establishes the minimax rate for the $\| \cdot \|_1$ norm.

At a high-level, both results will use Le Cam’s lemma applied to a finite collection of time-varying inverse covariance matrices. Given a time-varying inverse covariance matrix $\Omega(\cdot)$, let $P_\Omega$ be the joint distribution of $(X_1, Z_1), \ldots, (X_n, Z_n)$ where $(X_i, Z_i)$ are independent copies of $(X, Z)$ with $Z \sim \text{Unif}((0,1))$ and $X \mid Z \sim N(0, \Omega(z)^{-1})$. Let $U_0 = \{\Omega_0(\cdot), \Omega_1(\cdot), \ldots, \Omega_m(\cdot)\}$ be a collection of time-varying inverse covariance matrices, which are going to be defined later. With these we define the mixture distribution
\[
P = m^{-1} \sum_{\ell=1}^m P_{\Omega_\ell}.
\]For two measures $P$ and $Q$, the total variation is given as $\|P \wedge Q\| := \int (dP/d\mu) \wedge (dQ/d\mu)d\mu$, where $d\mu$ is the Lebesgue measure. Now, Le Cam’s lemma (LeCam, 1973) gives us the following lower bound.
Lemma C.3.1. Let $\hat{\Omega}(\cdot)$ be any estimator of $\Omega(\cdot)$ based on the data generated from the distribution family $\{P_{\Omega} | \Omega \in U_0\}$. Then

$$\max_{1 \leq \ell \leq m} \mathbb{E} \left[ \sup_{z \in (0,1)} \| \hat{\Omega}(z) - \Omega_{\ell}(z) \|_{\text{max}} \right] \geq r_{\min} \| \mathbb{P} \wedge \mathbb{P}_{\Omega_0} \|,$$

where $r_{\min} = \min_{1 \leq \ell \leq m} \sup_{z \in (0,1)} \| \Omega_0(z) - \Omega_{\ell}(z) \|_{\text{max}}$.

We will use the above lemma in the following two subsections.

C.3.1 Proof of Maximum Norm in (4.3.12)

We start by constructing the collection of inverse covariance matrices $U_0$. Let $\Omega_0(\cdot) \equiv I$.

Let $M_0 = \lceil c_0(n/\log(dn))^{1/5} \rceil$ where $c_0$ is some constant to be determined. Then

$$U_0 = \left\{ \Omega_{(j,m)}(\cdot) | \Omega_{(j,m)}^{-1}(z) = \Sigma_{(j,m)}(z) = I + \tau_m(z) E_{jj}, z \in (0,1), j \in [d-1], m \in [M_0] \right\},$$

where $E_{jj} = e_j e_j^T + e_{j+1} e_{j+1}^T$, $e_j$ is the $j$-th canonical basis in $\mathbb{R}^d$ and for any $m \in [M_0],$

$$\tau_m(z) = L h^2 K_0 \left( \frac{z - z_m}{h} \right), \quad z_m = \frac{m - 1/2}{M_0}, \quad h = 1/M_0. \quad \text{(C.3.1)}$$

Here, $K_0(\cdot)$ is any function supported on $(-1/2,1/2)$ and satisfying $\|K_0\|_{\text{sup}} \leq K_{\text{max}}$. For example, consider $K_0(\cdot) = a \psi(2z)$, where $\psi(z) = \exp(-1/(1 - z^2)) 1_{|z| \leq 1}$, for some sufficiently small $a$ (Tsybakov, 2009). It is easy to check that $U_0 \subset \bar{U}_s(M, \rho, L)$ if $n^{-1} \log(dn) = o(1)$.

For any $j \in [d-1], m \in [M_0]$, we have

$$\sup_{z \in (0,1)} \| \Omega_{(j,m)}(z) - \Omega_0(z) \|_{\text{max}} \geq \left\| \frac{\tau_m(\cdot)}{1 - \tau^2_m(\cdot)} \right\|_{\text{sup}} \geq \| \tau_m \|_{\text{sup}} \geq L h^2 K_0(0)$$

by direct calculation, which gives us $r_{\min} \geq L h^2 K_0(0) \asymp (\log(dn)/n)^{2/5}$.
In the remainder of the proof we show that for \( \bar{P} = ((d - 1)M_0)^{-1} \sum_{j \in [d], m \in [M_0]} P_{\Omega(j,m)} \), we have \( \|P_{\Omega} \land \bar{P}\| \geq 1/2 \). Let \( f_{jm} \) be the density of \( P_{\Omega(j,m)} \) for \( j \in [d - 1], m \in [M_0] \) and \( f_0 \) the density of \( P_0 \). Under our setting,

\[
\|P_{\Omega} \land P_0\| \geq 1/2.
\]

Let \( f_{jm} \) be the density of \( P_{\Omega(j,m)} \) for \( j \in [d - 1], m \in [M_0] \) and \( f_0 \) the density of \( P_0 \). Under our setting,

\[
f_{jm}((x_i, z_i)_{i=1}^n) = \prod_{i=1}^n g_{jm}(x_i) \mathbb{1}\{z_i \in (0, 1)\},
\]

where \( g_{jm} \) is the density function of \( N(0, \Sigma(j,m)) \). Note that for any two densities \( f \) and \( f' \)

\[
\int (f \land f')d\mu = 1 - \frac{1}{2} \int |f - f'|d\mu \geq 1 - \frac{1}{2} \left( \int \frac{f^2}{f'}d\mu - 1 \right)^{1/2}.
\]

Therefore, it suffices to show that

\[
\Delta := \int \left( ((d - 1)M_0)^{-1} \sum_{j,m} f_{jm} \right)^2 / f_0 d\mu - 1 \to 0. \quad \text{(C.3.2)}
\]

Expanding the square of the mixture in (C.3.2), we have

\[
\frac{1}{((d - 1)M_0)^2} \left\{ \sum_{j=1}^{d-1} \sum_{m=1}^{M_0} \int \frac{f_{jm}^2}{f_0} d\mu + \sum_{j=1}^{d-1} \sum_{m_1 \neq m_2} \int \frac{f_{jm_1}f_{jm_2}}{f_0} d\mu 
\right.
\]

\[
+ \sum_{j_1 \neq j_2} \sum_{m_1, m_2 \in [M_0]} \int \frac{f_{jm_1}f_{jm_2}}{f_0} d\mu \right\} - 1.
\]

To proceed, we recall the following result of Ren et al. (2015) given in their equation (94). Let \( g_i \) be the density function of \( N(0, \Sigma_i) \) for \( i = 0, 1, 2 \). Then

\[
\int \frac{g_1 g_2}{g_0} = \left[ \det \left( I - \Sigma_0^{-1}(\Sigma_1 - \Sigma_0)\Sigma_0^{-1}(\Sigma_2 - \Sigma_0) \right) \right]^{-1/2}. \quad \text{(C.3.3)}
\]

Using the above display, for \( j_1 \neq j_2 \) and \( m_1, m_2 \in [M_0] \), since \( (\Sigma_{j_1} - I)(\Sigma_{j_2} - I) = 0 \), we have

\[
\int f_{j_1 m_1} f_{j_2 m_2} / f_0 d\mu = 1.
\]

For \( |j_1 - j_2| = 1 \), we have

\[
\int f_{j_1 m_1} f_{j_2 m_2} / f_0 d\mu = [\det (I_3 - M_3)]^{-n/2} = 1,
\]

where \( M_3 \in \mathbb{R}^{3 \times 3} \) with \((M_3)_{13} = \tau_{m_1}(z)\tau_{m_2}(z)\) and the other entries are zero. For any
\(m_1, m_2 \in [M_0]\) and \(j \in [d - 1]\), we have
\[
\int \frac{f_{jm_1} f_{jm_2}}{f_0} \, d\mu = \prod_{1 \leq i \leq n} \int_{0}^{1} \left( 1 - \tau_{m_1}(z_i) \tau_{m_2}(z_i) \right)^{-1} dz_i = \left[ \int_{0}^{1} \left( 1 - \tau_{m_1}(z) \tau_{m_2}(z) \right)^{-1} dz \right]^n.
\]
\[(C.3.4)\]

If \(m_1 \neq m_2\), since the supports of \(\tau_{m_1}(\cdot)\) and \(\tau_{m_2}(\cdot)\) are disjoint, (C.3.4) implies
\[
\int f_{jm_1} f_{jm_2} / f_0 \, d\mu = 1.
\]

Finally, if \(m_1 = m_2 = m\), we have
\[
\int \frac{f_{jm}^2}{f_0} \, d\mu \leq \left[ \int_{0}^{1} \left( 1 - \tau_m^2(z) \log^{4/5}(dn) \right)^{-1} dz \right]^n \leq \left[ \frac{M_0 - 1}{M_0} + \frac{1}{M_0} (1 - L^2 h^4 K^2_{max})^{-1} \right]^n = \left[ 1 + \frac{L^2 h^4 K^2_{max}}{M_0(1 - L^2 h^4 K^2_{max})} \right]^n.
\]
\[(C.3.5)\]

In summary,
\[
\Delta = \frac{1}{((d - 1)M_0)^2} \sum_{j=1}^{d-1} \sum_{m=1}^{M} \left( \int \frac{f_{jm}^2}{f_0} \, d\mu - 1 \right)
\leq \exp \left[ - \log((d - 1)M_0) + n \log \left( 1 + \frac{L^2 h^4 K^2_{max}}{M_0(1 - L^2 h^4 K^2_{max})} \right) \right] - \frac{1}{(d - 1)M_0}.
\]

Recall that \(M_0 = \lceil c_0(n/\log(dn))^{1/5} \rceil\) and \(h = 1/M_0\). We choose \(c_0\) sufficiently large such that \(c_0^5 > 1 - L^2 K^2_{max}\). Using \(\log(1 + x) \leq x\) and \(x/(1 - x) \leq 2x\) for \(x \in (0, 1/2)\), we have
\[
\Delta \leq \exp \left[ - \log(d(c_0n/\log(dn))^{1/5}) + c_0^{-5} L^2 K^2_{max} \log(dn) \right] - \frac{1}{c_0dn^{1/5}} \to 0.
\]

Since \(r_{min} \geq Lh^2 K_0(0) \asymp (\log(dn)/n)^{2/5}\), the proof of (4.3.12) is complete by Lemma C.3.1.
C.3.2 Proof of $\ell_1$ Norm in (4.3.13)

In order to show the lower bound for $\| \cdot \|_1$, we need to construct a different $U_0$. We still choose $\Omega_0(\cdot) \equiv \text{I}$. Let $B$ be the set of matrices defined as

$$B = \left\{ B \in \mathbb{R}^{d \times d} \mid B = \begin{pmatrix} 0 & v \\ v & 0 \end{pmatrix} \text{ s.t. } v \in \{0, 1\}^{d-1}, \|v\|_0 = s, v_2 = 0 \right\}.$$ 

With this, we define

$$U'_0 = \left\{ \Omega_{(j,m)}(\cdot) \mid \Omega_{(j,m)}^{-1}(z) = \Sigma_{(j,m)}(z) = \text{I} + \tau_m(z)B_j, \text{ for } z \in (0,1), B_j \in B, m \in [M_0] \right\},$$

where the index $j$ corresponds to an element in the set $B$ and the function $\tau_m(\cdot)$ is defined in (C.3.1). Let $D := |B| = (d^2 - 2)/s$. We still choose $M_0 = \lceil c_0(n/\log(dn))^{1/5} \rceil$ for some constant $c_0$. It can be easily shown that $U_0 \subset \overline{U}_s(M, \rho, L)$ if $s^2 \log(dn)/n^{4/5} = o(1)$.

The rest of the proof is similar to the proof in Section C.3.1. For any $j \in [D], m \in [M_0],

$$\sup_{z \in (0,1)} \|\Omega_{(j,m)}(z) - \Omega_0(z)\|_1 \geq s\|\tau_m\|_{\sup} \geq sLh^2K_0(0),$$

giving $r_{\min} \geq sLh^2K_0(0) \asymp s(\log(dn)/n)^{2/5}$. We proceed to show $\|P \Omega_0 \wedge \mathbb{P}\| \geq 1/2$, where $\mathbb{P} = (DM_0)^{-1} \sum_{j \in [D], m \in [M_0]} \mathbb{P}_{\Omega_{(j,m)}},$ by proving

$$\Delta := \int \left( \frac{1}{DM_0} \sum_{j,m} f_{jm} \right)^2 / f_0 \, d\mu - 1 \to 0.$$ 

We establish the above display by modifying the proof of Lemma 2 in Ren et al. (2015).
Let \( J(j_1, j_2) = \text{Vec}(B_{j_1})^T \text{Vec}(B_{j_2})/2 \), where \( \text{Vec}(M) = (M_1^T, \ldots, M_d^T)^T \) for any matrix \( M = [M_1, \ldots, M_d] \). From (C.3.3) and the definition of \( B \), we have

\[
\int \frac{f_{j_1m_1}f_{j_2m_2}}{f_0} d\mu = \left[ \int_0^1 \left[ \text{det } \left( I - \tau_{m_1}(z)\tau_{m_2}(z)B_{j_1}B_{j_2} \right) \right]^{-1/2} dz \right]^n
\]

(C.3.6)

Similar to (C.3.4), when \( m_1 \neq m_2 \), (C.3.6) yields that \( \int f_{j_1m_1}f_{j_2m_2}/f_0 d\mu = 1 \). Similar to (C.3.5), we also have

\[
\int \frac{f_{j_1m}f_{j_2m}}{f_0} d\mu \leq \left[ 1 + \frac{2J(j_1, j_2)L^2h^4K_{\max}^2}{M_0} \right]^n.
\]

Since

\[
\{|(j_1, j_2) \in [D]^2 \mid J(j_1, j_2) = j\} = \binom{d-2}{s} \binom{s}{j} \binom{d-2-s}{s-j},
\]

we have

\[
\Delta = \frac{1}{D^2M_0^2} \sum_{m=1}^{M_0} \sum_{0 \leq j \leq s} \sum_{J(j_1, j_2) = j} \left( \int \frac{f_{j_1m}f_{j_2m}}{f_0} d\mu - 1 \right)
\]

\[
\leq \frac{1}{D^2M_0^2} \sum_{m=1}^{M_0} \sum_{0 \leq j \leq s} \sum_{J(j_1, j_2) = j} \left[ 1 + \frac{2J(j_1, j_2)L^2h^4K_{\max}^2}{M_0} \right]^n - 1
\]

\[
\leq \frac{1}{D^2M_0^2} \sum_{m=1}^{M_0} \sum_{0 \leq j \leq s} \binom{d-2}{s} \binom{s}{j} \binom{d-2-s}{s-j} \left[ 1 + \frac{2J(j_1, j_2)L^2h^4K_{\max}^2}{M_0} \right]^n.
\]

Let \( a_0 = 2c_0^{-5}L^2K_{\max}^2 \). Similar to the proof of Lemma 6 in Ren et al. (2015), we have

\[
\left[ 1 + \frac{2J(j_1, j_2)L^2h^4K_{\max}^2\log(dn)}{M_0} \right]^n \leq \exp \left( 2nM^{-1}jL^2h^4K_{\max}^2 \right) \leq (nd)^{a_0j}.
\]

160
This further gives us

\[ \Delta \leq \frac{1}{M^2_0} \sum_{m=1}^{M_0} \sum_{1 \leq j \leq s} \frac{(s) (d - 2 - s) (nd)^{a_0j}}{s - j} \frac{(d - 2)}{s} \]

\[ = \frac{1}{M^2_0} \sum_{m=1}^{M_0} \sum_{0 \leq j \leq s} \frac{1}{j!} \left( \frac{s!(d - s)!}{(s - j)!} \right)^2 \frac{(nd)^{a_0j}}{(d - 2)!(d - 2s + j - 2)!} \]

\[ \leq \frac{1}{M_0} \sum_{1 \leq j \leq s} \left( \frac{s^2 (nd)^{a_0j}}{d - s} \right)^j , \]

where the last inequality is due to \( \frac{s!}{(s-j)!} \leq s^j \) and \( \frac{(d-2)!(d-2s+j-2)!}{(d-s)^2} \geq (d-s)^j \). By assumption \( s^{-v}d \geq 1 \) for some \( v > 2 \), and therefore

\[ \Delta \leq \frac{1}{c_0(n/\log n)^{1/5}} \sum_{0 \leq j \leq s} (2d^{v-1}(dn)^{a_0j}j \leq \frac{2c_0(n/\log (dn))^{-1/5}d^{2/v-1+a_0}a_0}{1 - 2d^{2/v-1+2a_0}}, \]

for sufficiently large \( d, d - s \geq d/2, d \geq n \). By choosing \( c_0 \) in \( a_0 = 2c_0^{-5}L^2K_{\max}^2 \) sufficiently small, so that \( a_0 < \min(1/5, 1/v - 1/2) \), we have \( \Delta \to 0 \). This completes the proof.

### C.4 Convergence Rate of Kendall’s Tau Estimator

In this section, we study the rate of convergence of kernel Kendall’s tau estimator \( \hat{\tau}_{jk}(z) \) in (4.1.3) to \( \tau_{jk}(z) \) uniformly in the bandwidth \( h \in [h_l, h_u] \), the index \( z \in (0, 1) \) and the dimension \( j, k \in [d] \). We start by establishing the bias-variance decomposition for \( \hat{\tau}_{jk}(z) \) and then show how to bound the bias and variance separately.

Recall that in (C.2.2), we define

\[ g_{z|j,k}(y_i, y'_i) = \omega_{z}(z_i, z'_i) \text{sign}(x_{ia} - x'_{ia}) \text{sign}(x_{ib} - x'_{ib}), \]
and \( \omega(z_i, z_i') \) is defined in (4.1.4). The estimator \( \hat{\tau}_{jk}(z) \) can be written as a quotient of two \( U \)-statistics

\[
\hat{\tau}_{jk}(z) = \frac{U_n[g_{z(i,j,k)}]}{U_n[\omega_z]}.
\]

Recall that the operator \( u_n[\cdot] \) is defined as

\[
u_n[H] = \sqrt{n} \cdot (U_n[H] - E[U_n[H]])
\]

for any bivariate function \( H(x, x') \). With this, following an argument similar to that in Equation (3.45) of Pagan and Ullah (1999), we have the following decomposition

\[
\hat{\tau}_{jk}(z) = E[U_n[g_{z(i,j,k)}| (j,k)] + \frac{u_n[g_{z(i,j,k)}]}{\sqrt{n} \cdot E[U_n[\omega_z]]} - \frac{E[U_n[g_{z(i,j,k)}] \cdot u_n[\omega_z]}{\sqrt{n} \cdot (E[U_n[\omega_z]])^2} + n^{-1}O\left(u_n[g_{z(i,j,k)}] \cdot u_n[\omega_z] + (u_n[\omega_z])^2 \right),
\]

under the condition that

\[
\left| \frac{u_n[\omega_z]}{\sqrt{n} \cdot E[U_n[\omega_z]]} \right| < 1 \text{ and } E[U_n[\omega_z]] \neq 0.
\]

Taking expectation on both sides of (C.4.1), we obtain

\[
E[\hat{\tau}_{jk}(z)] = E[U_n[g_{z(i,j,k)}]] + n^{-1}O\left(E[u_n[g_{z(i,j,k)}] \cdot u_n[\omega_z] + (u_n[\omega_z])^2 \right).
\]

With this, we are ready to prove Lemma C.1.1 and Lemma C.1.2.
C.4.1 Proof of Lemma C.1.1

We first check that the condition in (C.4.2) is satisfied under the assumptions. Using (C.2.12) and (C.2.14), for sufficiently large $n$,

\[
\sup_{j,k \in [d]} \sup_{z \in (0,1)} \left| \frac{u_n[z]}{\sqrt{n} E[U_n[z]]} \right| \lesssim \sqrt{\frac{\log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))}{nh}} \cdot \frac{1}{f^2_Z(z) + O(h^2)} < 1.
\]

Since $f_Z$ is bounded from below, $E[U_n[z]] = f^2_Z(z) + O(h^2) \geq f^2_Z(z)/2 > 0$ for large enough $n$. Therefore, condition in (C.4.2) holds and the expansion in (C.4.1) is valid.

From (C.4.1) and (C.4.3), we have

\[
\hat{\tau}_{jk}(z,h) - E[\hat{\tau}_{jk}(z,h)] = \frac{u_n[g_z(j,k)]}{\sqrt{n E[U_n[z]]}} - \frac{\mathbb{E}[U_n[g_z(j,k)]] \cdot u_n[z]}{\sqrt{n \mathbb{E}[U_n[z]]^2}} + I_3,
\]

where

\[
I_3 = n^{-1}O\left( u_n[g_z(j,k)] \cdot u_n[z] + (u_n[z])^2 + \mathbb{E}[u_n[g_z(j,k)] \cdot u_n[z]] + \mathbb{E}\left((u_n[z])^2\right) \right).
\]

We bound $I_1$, $I_2$ and $I_3$ separately. Using (C.2.12) and (C.2.14), we have that

\[
\sup_{j,k \in [d]} \sup_{h \in [h_l,h_u]} \sup_{z \in (0,1)} \left| \frac{\sqrt{nh} \cdot |I_1|}{\sqrt{\log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))}} \right| \leq \sup_{j,k \in [d]} \sup_{h \in [h_l,h_u]} \sup_{z \in (0,1)} \left| \frac{\sqrt{nh} \cdot |u_n[g_z(j,k)]|}{\sqrt{\log(d/h) \lor \log(\delta^{-1} \log(h_u h_l^{-1}))}} \right| \cdot \left| \frac{f^2_Z(z) + O(h^2)}{f^2_Z(z)} \right| \leq C;
\]
with probability $1 - \delta$ for large enough $n$. Similarly, using (C.2.12), (C.2.13) and (C.2.14), we also have

$$
\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{\sqrt{nh} \cdot |I_2|}{\log(d/h) \lor \log (\delta^{-1} \log (h_i h^{-1}))} \right| \leq \sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{\sqrt{h}}{\log(d/h) \lor \log (\delta^{-1} \log (h_i h^{-1}))} \left| u_n [\omega_z] \left| \frac{(f_2^2(z) \tau_{jk}(z) + O(h^2))}{(f_2^2(z) + O(h^2))^2} \right| \right| \right| \leq C,
$$

with probability $1 - \delta$. Finally, using (C.2.12), (C.2.15) and (C.2.16), we have

$$
\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{nh \cdot |I_3|}{\log(d/h) \lor \log (\delta^{-1} \log (h_i h^{-1}))} \right| \leq \sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{2h}{\log(d/h) \lor \log (\delta^{-1} \log (h_i h^{-1}))} \left( u_n [\omega_z] \lor u_n [g_{z(j,k)}] \right)^2 \right| \leq C,
$$

with probability $1 - \delta$.

Combining the above three displays with (C.4.4) completes the proof.

### C.4.2 Proof of Lemma C.1.2

It follows from Lemma C.2.2 and the decomposition in (C.4.3) that

$$
\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{\mathbb{E}[\tau_{jk}(z)] - \tau_{jk}(z)}{h^2 + 1/(nh)} \right| \leq \sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{f_2^2(z) \tau_{jk}(z) + O(h^2)}{f_2^2(z) + O(h^2)} - \tau_{jk}(z) + O((nh)^{-1}) \right| \frac{1}{h^2 + (nh)^{-1}} \leq C.
$$
C.5 Concentration of \(U\)-statistics

In this section, we study certain properties of the \(U\)-statistics used in this thesis. We state and prove Lemma C.2.1 and Lemma C.2.2, which were used to establish results on the rate of convergence of Kendall’s tau estimator in Appendix C.4. In particular, we will prove the following two results in this section, with the notations on \(U\)-statistics and empirical processes defined in Appendix C.4.

C.5.1 Proof of Lemma C.2.1

At a high-level, we will use the Hoeffding decomposition to represent the \(U\)-statistics \(U_n[\omega z]\) and \(U_n[g_z(j,k)]\) defined in (C.2.5) and (C.2.8). Next, we will use concentration inequalities for suprema of empirical processes and \(U\)-statistics to bound individual terms in the decomposition.

Recall that (C.2.5) and (C.2.8) give

\[
n^{-1/2} \cdot u_n [g_z(j,k)] = U_n [g_z(j,k)] - \mathbb{E} [U_n [g_z(j,k)]] = 2\mathbb{E}_n [g_z^{(1)}(j,k)(Y_i)] + U_n [g_z^{(2)}(j,k)]
\]

and

\[
n^{-1/2} \cdot u_n [\omega z] = U_n [\omega z] - \mathbb{E} [U_n [\omega z]] = 2\mathbb{E}_n [\omega_z^{(1)}] + U_n [\omega_z^{(2)}].
\]

In order to bound \(u_n [\omega z]\) and \(u_n [g_z(j,k)]\) it is sufficient to bound \(g_z^{(1)}(j,k), g_z^{(2)}(j,k), \omega_z^{(1)}\) and \(\omega_z^{(2)}\). We do so in the following lemmas.
Lemma C.5.1. We assume $n^{-1} \log d = o(1)$ and the bandwidths $0 < h_l < h_u < 1$ satisfy $h_l n / \log(d n) \to \infty$ and $h_u = o(1)$. There exist a universal constant $C > 0$ such that

$$\sup_{j,k} \sup_{z \in (0,1)} \left| \mathbb{G}_n \left[ \frac{\sqrt{h g_{z|j,k}}}{\sqrt{\log(d/h) \vee \log \left( \frac{\delta^{-1} \log(h_u h_l^{-1})}{h} \right)}} \right] \right| \leq C$$

(C.5.1)

with probability $1 - \delta$.

Proof. The general strategy to show (C.5.1) can be separated into two steps.

**Step 1.** Splitting the supreme over $h \in [h_l, h_u]$ into smaller intervals such that the empirical process (C.5.1) is easier to bound for each interval. Let $S$ be the smallest integer such that $2^S h_l \geq h_u$. Note that $S \leq \log_2(h_u/h_l)$ and $[h_l, h_u] \subseteq \bigcup_{\ell=1}^S [2^{\ell-1} h_l, 2^\ell h_l] =: \mathcal{H}_\ell$. Then

$$\mathbb{P} \left[ \sup_{z,j,k} \sup_{h \in [h_l, h_u]} \left| \mathbb{G}_n \left[ \frac{\sqrt{h g_{z|j,k}}}{\sqrt{\log(d/h) \vee \log \left( \frac{\delta^{-1} \log(h_u h_l^{-1})}{h} \right)}} \right] \right| \geq t \right] \leq \sum_{\ell=1}^S \mathbb{P} \left[ \sup_{j,k} \sup_{h \in \mathcal{H}_\ell} \sup_{z \in (0,1)} \left| \mathbb{G}_n \left[ \frac{\sqrt{h g_{z|j,k}}}{\sqrt{\log(d/h) \vee \log \left( \frac{\delta^{-1} \log(h_u h_l^{-1})}{h} \right)}} \right] \right| \geq t \right].$$

(C.5.2)

**Step 2.** We apply Talagrand’s inequality (Bousquet, 2002) to each term in the summation on the right hand side of (C.5.2).

Consider the class of functions

$$\mathcal{F}_\ell = \left\{ \sqrt{h g_{z|j,k}} : h \in \mathcal{H}_\ell, z \in (0,1), j, k \in [d] \right\}.$$

In order to apply Talagrand’s inequality to functions in the class $\mathcal{F}_\ell$, we have to check three conditions:

- The class $\mathcal{F}_\ell$ is uniformly bounded;
- Bounding the variance of $g_{z|j,k}^{(1)}$;
- Bounding the covering number of $\mathcal{F}_\ell$. 

166
We verify the above three conditions next. First, we show that the class $\mathcal{F}_\ell$ is uniformly bounded.

Recall that in (C.2.24), we show that

$$
E \left[ g_{z(j,k)}(y', Y) \right] = K_h(z' - z) \int K_h(s - z) \varphi \left( x'_j, x'_k, \Sigma_{jk}(s) \right) f_Z(s) ds,
$$

where $\varphi(\cdot)$ is defined in (C.2.23). Since $|\varphi(u, v, \rho)| \leq 2$ for any $u$, $v$, and $\rho$, and $h \in \mathcal{H}_\ell$, the above display gives us

$$
\sup_{z \in (0, 1)} \max_{j, k \in \mathcal{D}} \left\| g^{(1)}_{z(j,k)} \right\|_{\infty} \leq \frac{2\|K\|_{\infty}}{h} \sup_{z \in (0, 1)} \int K_h(s - z) f_Z(s) ds + \sup_{z \in (0, 1)} \max_{j, k \in \mathcal{D}} \mathbb{E} \left[ U_n \left[ g_{z(j,k)} \right] \right]
$$

$$
\leq \frac{2\|K\|_{\infty}}{h} \sup_{z \in (0, 1)} (f_Z(z) + O(h^2)) + \mathcal{T}_Z^2 + O(h^2) \leq \frac{3\|K\|_{\infty} \mathcal{T}_Z}{h},
$$

where the second inequality is due to (C.2.13). This results shows that the class $\mathcal{F}_\ell$ is uniformly bounded by $(2^\ell h_l)^{-1/3}\|K\|_{\infty} \mathcal{T}_Z$ and has the envelope $F_\ell = 3(2^{\ell-1} h_l)^{-1/2}\|K\|_{\infty} \mathcal{T}_Z$.

Next, we bound the variance of $g^{(1)}_{z(j,k)}$. Let $\bar{\varphi}_{u,v}(s) = \varphi(u, v, \Sigma_{jk}(s)) f_Z(s)$. We can rewrite (C.5.3) as

$$
E \left[ g_{z(j,k)}(y', Y) \right] = K_h(z' - z) \cdot (K_h \ast \bar{\varphi}_{u,v})(z),
$$

where "\ast" denotes the convolution operator. Then we bound the convolution by

$$
\sup_{u, v} \|K_h \ast \bar{\varphi}_{u,v}\|_{\infty} \leq \sup_{u, v} \|K_h\|_1 \|\bar{\varphi}_{u,v}\|_{\infty} \leq 2\mathcal{T}_Z,
$$

(C.5.5)
where the first inequality follows using the Young’s inequality for convolution. Now, for any fixed \( h \in H_\ell \) and \( z \in (0, 1) \), we have

\[
\sup_{z,j,k} \mathbb{E} \left[ \left( \sqrt{h} g_{z(j,k)}(1) \right)^2 \right] \leq \sup_{z,j,k} \mathbb{E} \left[ \max_{u,v} \left( \sqrt{h} g_{z(j,k)}(1) \right)^2 \right]
\]

\[
\leq \sup_z 2 \mathbb{E} \left[ h \left( K_h(Z - z) \right)^2 \right] \sup_{u,v} \| K_h * \tilde{\varphi}_{u,v} \|_\infty^2 + \sup_{z,j,k} \left\{ \mathbb{E} \left[ U_n \left[ g_{z(j,k)} \right] \right] \right\}^2
\]

\[
\leq C f_Z^2 \cdot \sup_z \left( \int \frac{1}{h} K^2 \left( \frac{x - z}{h} \right) f_Z(x) dx \right) + Chf_Z^2 + O(h^3)
\]

\[
\leq C f_Z^2 \sup_z (f_Z(z)\|K\|_2^2 + O(h^2)) \leq C f_Z^3 \|K\|_2^2.
\]

The reason why we show a stronger result above on the expectation of maximal is because we need (C.5.6) in the proof of Section C.6.

Using the above display, we can bound the variance for any \( \ell \in [S] \),

\[
\sigma^2 := \sup_{f \in F_\ell} \mathbb{E} \left[ f^2 \right] \leq C f_Z^3 \|K\|_2^2.
\]

(C.5.7)

Finally, we need a bound on the covering number of \( F_\ell \). Using Lemma C.8.4 we have that

\[
\sup_Q N(F_\ell, L_2(Q), \epsilon) \leq \frac{Cd^2}{(2^{\ell-1}h_\ell)^{v+9\epsilon v+6}}.
\]

Theorem 3.12 of Koltchinskii (2011) then gives us that for some universal constant \( C \),

\[
\mathbb{E} \left[ \sup_{f \in F_\ell} |n \mathbb{E}_n[f]| \right] \leq C \sqrt{n \log \left( \frac{Cd}{2^{\ell-1}h_\ell} \right) + \frac{C}{(2^{\ell-1}h_\ell)^{1/2}} \log \left( \frac{Cd}{2^{\ell-1}h_\ell} \right)}.
\]

Furthermore, since \( h_\ell n / \log(d n) \to \infty \), the above display simplifies to

\[
\mathbb{E} \left[ \sup_{f \in F_\ell} |n \mathbb{E}_n[f]| \right] \leq C \sqrt{n \log \left( \frac{Cd}{2^{\ell-1}h_\ell} \right)}.
\]

(C.5.8)
Using Theorem 2.3 of Bousquet (2002) together with (C.5.7), (C.5.8), and \( n^{-1} \log (dh^{-1}) = o(1) \), we obtain

\[
\sup_{j,k} \sup_{h \in \mathcal{H}_\ell} \sup_{z \in (0,1)} \left| G_n \left[ \sqrt{n} g_{z(j,k)}^{(1)} \right] \right| \leq C \left( \sqrt{\log \left( \frac{Cd}{2^{\ell-1}h_l} \right)} + \sqrt{\log(1/\delta)} \right) \tag{C.5.9}
\]

with probability \( 1 - \delta \). Observe that \( 2^{\ell-1}h_l \geq h/2 \) for any \( h \in \mathcal{H}_\ell \). Therefore, combining (C.5.9), (C.5.2), we obtain that for some constant \( C > 0 \)

\[
\sup_{j,k} \sup_{h \in \mathcal{H}_\ell} \sup_{z \in (0,1)} \left| G_n \left[ \frac{\sqrt{n} g_{z(j,k)}^{(1)}}{\log(d/h) \vee \log(\delta^{-1} \log(\delta h^{-1})))} \right] \right| \leq C,
\]

with probability \( 1 - \delta \).

\[\square\]

**Lemma C.5.2.** We assume \( n^{-1} \log d = o(1) \) and the bandwidths \( 0 < h_l < h_u < 1 \) satisfy \( h_l n / \log(dn) \to \infty \) and \( h_u = o(1) \). There exists a universal constant \( C > 0 \) such that

\[
\sup_{j,k} \sup_{h \in \mathcal{H}_\ell} \sup_{z \in (0,1)} \left| h U_n \left[ g_{z(j,k)}^{(2)} \right] \right| \leq C
\]

with probability \( 1 - \delta \).

**Proof.** The method to prove this lemma is similar to the proof of Lemma C.5.1. The only difference is that instead of bounding the empirical process, we bound a suprema of a \( U \)-statistic process. Therefore, we will use Theorem C.9.1 instead of Talagrand’s inequality.

We apply the trick of splitting \([h_l, h_u]\) again. Let \( S \) be the smallest integer such that \( 2^S h_l \geq h_u \). We have that \( S \leq \log_2(h_u/h_l) \) and \([h_l, h_u] \subseteq \bigcup_{\ell=1}^{S} [2^{\ell-1}h_l, 2^\ell h_l] \). For simplicity, we define \( \mathcal{H}_\ell = [2^{\ell-1}h_l, 2^\ell h_l] \). Therefore,

\[
P \left[ \sup_{j,k} \sup_{h \in [h_l, h_u]} \left| h U_n \left[ g_{z(j,k)}^{(2)} \right] \right| \geq t \right] \leq \sum_{\ell=1}^{S} P \left[ \sup_{j,k} \sup_{h \in \mathcal{H}_\ell} \left| h U_n \left[ g_{z(j,k)}^{(2)} \right] \right| \geq t \right]. \tag{C.5.10}
\]

169
As \( g_{z|j,k}^{(2)} \) is a degenerate kernel, we can use Theorem C.9.1 to bound the right hand side of (C.5.10). Consider the class of functions

\[
\mathcal{F}_\ell^{(2)}(j,k) = \left\{ (2^\ell - 1 \eta_l) h g_{z|j,k}^{(2)} \mid h \in \mathcal{H}_\ell, z \in (0,1) \right\} \quad \text{and} \quad \mathcal{F}_\ell^{(2)} = \left\{ f \in \mathcal{F}_\ell^{(2)}(j,k) \mid j, k \in [d] \right\}.
\]

where \( L^{-1} := 7 \| K \|_\infty \bar{f}^2_Z \). From the expansion in (C.2.4), with (C.5.4) and (C.2.13), we have

\[
\sup_{z \in (0,1)} \max_{j,k} \| g_{z|j,k}^{(2)} \|_\infty \leq C \| K \|_\infty \bar{f}^2_Z h^{-2}, \quad \text{for } h \in [\eta_l, \eta_u].
\]

Therefore, the class \( \mathcal{F}_\ell^{(2)} \) is uniformly bounded by 1 and has the envelope \( F_\ell^{(2)} = 1 \), which verifies the condition in (C.9.1). Furthermore, by (C.5.6) and (C.2.13), we can bound the variance as

\[
\sup_{z,j,k} \mathbb{E} \left[ (h g_{z|j,k}^{(2)})^2 \right] \leq \sup_{z,j,k} \left\{ 4\bar{h}^2 \mathbb{E} \left[ (g_{z|j,k}(Y_i, Y_i'))^2 \right] + \mathbb{E} \left[ \left( \sqrt{h} g_{z|j,k}^{(1)} \right)^2 \right] + 4 \mathbb{E} \left[ U_n[g_{z|j,k}] \right]^2 \right\} \\
\leq \sup_z h^{-2} \mathbb{E} \left[ K ((Z_t - z)/h)^2 K ((Z_t - z)/h)^2 + \bar{f}^2_Z h + \bar{f}^2_Z h^2 \right] \\
\leq \sup_z (h^{-1} \int K ((x - z)/h)^2 f_Z(x) dx)^2 + \bar{f}^2_Z h + \bar{f}^2_Z h^2 \\
\leq \bar{f}^2_Z \mathbb{E} \left[ \| K \|_2^4 + \bar{f}^2_Z h + \bar{f}^2_Z h^2 \right] \lesssim \bar{f}^2_Z \mathbb{E} \left[ \| K \|_2^4 \right].
\]

Therefore, for any \( \ell \in [S] \),

\[
\sup_{f \in \mathcal{F}_\ell^{(2)}} \mathbb{E} \left[ f^2 \right] \lesssim 2 \bar{f}^2_Z \| K \|_2^4 (2^\ell - 1 \eta_l L)^2 := \sigma_\ell^2 < 1.
\]

According to Lemma C.8.4, we have that

\[
\sup_{Q} N(\mathcal{F}_\ell^{(2)}(j,k), L_2(Q), \epsilon) \leq C \frac{(2^\ell - 1 \eta_l L)^{-4v-15}}{(2^\ell - 1 \eta_l)^{2v+18} \epsilon^{4v+15}}.
\]
By setting $t = C(2^{\ell-1}h_l)\log (C/(2^{\ell-1}h_l)) \lor \log(C/\delta)$ in (C.9.2) for a sufficiently large constant $C$, as $h_l n / \log (dn) \to \infty$, we have

$$n \sigma_t^2 = n(2^{\ell-1}h_l)^2 \geq \log \left(1/(2^{\ell-1}h_l)\right) \lor \log(1/\delta) = \frac{t}{\sigma_t} \geq C' \left( \frac{\log(1/(2^{\ell-1}h_l))}{\log n} \right)^{3/2} \log \left( \frac{2}{2^{\ell-1}h_l} \right),$$

for some constant $C' > 0$. This verifies the condition (C.9.3). Now (C.9.2) gives us that

$$n \sigma_t^2 \sigma_t \geq C' \left( \frac{\log(1/(2^{\ell-1}h_l))}{\log n} \right)^{3/2} \log \left( 2^{2\ell-1}h_l \right),$$

for some constant $C' > 0$. This verifies the condition (C.9.3). Now (C.9.2) gives us that

$$\sup_{h \in H_{\ell}} \sup_{z \in (0, 1)} (2^{\ell-1}h_l L) h \cup_n \left[ g_{z(j,k)}^{(2)} \right] \leq C n^{-1}(2^{\ell-1}h_l) \log \left( 1/(2^{\ell-1}h_l) \right) \lor \log(1/\delta). \quad (C.5.11)$$

Since $2^{\ell-1}h_l \geq h/2$ for any $h \in H_{\ell}$, applying (C.5.10) and (C.5.11) with the union bound over $j, k \in [d]$, we obtain

$$\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{nh}{\log(d/h) \lor \log (\delta^{-1} \log (h_u h_l^{-1}))} \cup_n \left[ g_{z(j,k)}^{(2)} \right] \right| \leq C, \quad (C.5.12)$$

with probability $1 - (h_u \lor \delta)$. As $h_u = o(1)$, (C.5.12) follows with probability larger than $1 - \delta$.

**Lemma C.5.3.** We suppose $n^{-1} \log d = o(1)$ and the bandwidths $0 < h_l < h_u < 1$ satisfy $h_l n / \log (dn) \to \infty$ and $h_u = o(1)$. There exists a universal constant $C > 0$ such that

$$\sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| G_n \left[ \frac{\sqrt{h_l} \omega_z^{(1)}}{\sqrt{\log(1/h) \lor \log (\delta^{-1} \log (h_u h_l^{-1}))}} \right] \right| \leq C$$

with probability $1 - \delta$.

**Proof.** The proof is similar to that of Lemma C.5.1. We again decompose $[h_l, h_u] \subseteq \cup_{\ell=1}^S H_{\ell}$, where $H_{\ell} = [2^{\ell-1}h_l, 2^\ell h_l]$ and $S$ is the smallest integer such that $2^S h_l \geq h_u$. By the union

171
bound,

\[
\Pr \left[ \sup_{h \in [h, h_u]} \sup_{z \in (0,1)} |G_n [h^2 \omega_z^{(l)}]| \geq t \right] \leq \sum_{j=1}^{S} \Pr \left[ \sup_{h \in H_\ell} \sup_{z \in (0,1)} |G_n [h^2 \omega_z^{(l)}]| \geq t \right].
\]

(C.5.13)

Talagrand’s inequality (Bousquet, 2002) is applied once again to control (C.5.13). Consider the class of functions

\[
\mathcal{K}_\ell = \left\{ \sqrt{h} \omega_z^{(1)} \mid h \in H_\ell, z \in (0,1) \right\}.
\]

According to the definition of \( \omega_z^{(1)} \) in (C.2.6), we have

\[
\omega_z^{(1)}(s) = K_h(s - z) \mathbb{E}[K_h(z - Z)] - (\mathbb{E}[K_h(z - Z)])^2.
\]

We can bound the supremum norm by

\[
\sup_{z \in (0,1)} \|\omega_z^{(1)}\|_\infty \leq 2\bar{f}_Z \|K\|_\infty h^{-1},
\]

which implies that the envelope function is

\[F_\ell = 2\bar{f}_Z \|K\|_\infty (2^{\ell-1}h_\ell)^{-1}/2.\]

Similar to (C.5.6), we can also bound the variance by

\[
\sup_z \mathbb{E} \left[ \left( \sqrt{h} \omega_z^{(1)} \right)^2 \right] \leq \sup_z 2\mathbb{E} \left[ h (K_h(Z - z))^2 \right] \cdot (\mathbb{E}[K_h(z - Z)])^2 + \sup_z 2h \cdot \mathbb{E} \left[ \mathbb{E} \left[ U_n \left[ \omega_z \right] \right] \right]^2
\]

\[
\leq 2(\bar{f}_Z^3 \|K\|_2^2 + O(h^2)) + 2h \bar{f}_Z^2 + O(h^3) \leq 2\bar{f}_Z^3 \|K\|_2^2 := \sigma_\ell^2.
\]

(C.5.15)

Using Lemma C.8.5 we have

\[
\sup_Q N(\mathcal{K}_\ell, L_2(Q), \epsilon) \leq \frac{C}{(2^{\ell-1}h_\ell)^{4\epsilon^{2\nu+1}}} \cdot \frac{\|F_\ell\|_{L_2}}{2\|K\|_\infty (2^{\ell-1}h_\ell)^{-1/2}}.
\]

172
which combined with Theorem 3.12 of Koltchinskii (2011) with \( h_t n/ \log(dn) \to \infty \) implies that

\[
\mathbb{E} \left[ \sup_{f \in K} |n \mathbb{E}_n[f]| \right] \leq C \sqrt{n \log \left( \frac{C}{2^{\ell-1} h_t} \right)}.
\]

Theorem 2.3 of Bousquet (2002) derives that for some constant \( C > 0 \)

\[
\sup_{h \in H_t} \sup_{z \in (0, 1)} |G_n \left[ \sqrt{\omega_z^{(1)}} \right] | \leq C \left( \sqrt{\log \left( \frac{C}{2^{\ell-1} h_t} \right)} + \sqrt{\log(1/\delta)} \right) \quad \text{(C.5.16)}
\]

with probability \( 1 - \delta \). Using the union bound in (C.5.16) with (C.5.13), we have

\[
\sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| G_n \left[ \sqrt{\omega_z^{(1)}} \right] \right| \leq C,
\]

for some constant \( C > 0 \) with probability \( 1 - \delta \).

**Lemma C.5.4.** We assume \( n^{-1} \log d = o(1) \) and the bandwidths \( 0 < h_l < h_u < 1 \) satisfy \( h_t n/ \log(dn) \to \infty \) and \( h_u = o(1) \). There exists a universal constant \( C > 0 \) such that for sufficiently large \( n \),

\[
\sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| G_n \left[ \omega_z^{(2)} \right] \right| \leq C
\]

with probability \( 1 - \delta \).

**Proof.** The proof is similar to that of Lemma C.5.2. Instead of applying Lemma C.8.5, we use Lemma C.8.4. We omit the details of the proof.

Let

\[
r(d, n, h, \delta, h_u, h_l) = \frac{\log(d/h) \vee \log(\delta^{-1} \log(h_u h_l^{-1}))}{nh}.
\]
Applying Lemma C.5.1, Lemma C.5.2, Lemma C.5.3 and Lemma C.5.4 to (C.2.5) and (C.2.8), we obtain that with probability $1 - \delta$, there exists a constant $C > 0$ such that

$$\sup_{j,k \in [d]} \sup_{h \in [h_l, h_u]} \sup_{z \in (0, 1)} \left| \frac{n^{-1/2} \cdot u_n [\omega_z] \vee u_n [g_{z| (j,k)}]}{r^{1/2}(d, n, h, \delta, h_u, h_l)} + r(d, n, h, \delta, h_u, h_l) \right| \leq C/2. \quad (C.5.17)$$

Since $\log(dn) / (nh_l) = o(1)$ and $h_u = o(1)$, we have $r(d, n, h, \delta, h_u, h_l) \leq r^{1/2}(d, n, h, \delta, h_u, h_l)$ for sufficiently large $n$. Using this in (C.5.17), we have Lemma C.2.1 proved.

### C.5.2 Proof of Lemma C.2.2

The high-level idea for proving this lemma is to write the expectations as the integrals and apply Taylor expansions to the density functions and other nonparametric functions. Afterwards, we bound the remainder terms of the Taylor expansions.

We compute $\mathbb{E} \left[ U_n \left[ g_{z| (j,k)} \right] \right]$ first. Using Corollary C.9.1, we have

$$\mathbb{E} \left[ U_n \left[ g_{z| (j,k)} \right] \right] = \frac{2}{\pi} \mathbb{E} \left[ K_h(Z_1 - z)K_h(Z_2 - z) \arcsin \left( \frac{\Sigma_{jk}(Z_1) + \Sigma_{jk}(Z_2)}{2} \right) \right], \quad (C.5.18)$$

where $Z_1, Z_2$ are independent and equal to $Z$ in distribution. After a change of variables, we can further expand the right hand side of (C.5.18) as

$$\frac{2}{\pi} \int \int K(t_1)K(t_2)f_Z(z + t_1 h)f_Z(z + t_2 h) \arcsin \left( \frac{\Sigma_{jk}(z + t_1 h) + \Sigma_{jk}(z + t_2 h)}{2} \right) dt_1 dt_2.$$

Using the Taylor series expansion of $\arcsin(\cdot)$, we have

$$\arcsin \left( \frac{\Sigma_{jk}(z + t_1 h) + \Sigma_{jk}(z + t_2 h)}{2} \right) = \arcsin (\Sigma_{jk}(z)) + \left( t_1 h \Sigma_{jk}(z) + t_2 h \Sigma_{jk}(z) \right) + \frac{(t_1 h)^2 \Sigma_{jk}(z) + (t_2 h)^2 \Sigma_{jk}(z)}{4} \frac{1}{\sqrt{1 - \rho^2}},$$

174
where \( \tilde{z} \) is between \( z \) and \( z + t_1h \), and \( \tilde{\rho} \) is between \( \Sigma_{jk}(z) \) and \((\Sigma_{jk}(z + t_1h) + \Sigma_{jk}(z + t_2h))/2 \). Since the minimum eigenvalue of \( \Sigma(z) \) is strictly positive for any \( z \), there exists a \( \gamma_\sigma < 1 \) such that \( \sup_z \left| \Sigma_{jk}(z) \right| \leq \gamma_\sigma < 1 \) for any \( j \neq k \). We thus have \((1 - \tilde{\rho}^2)^{-1/2} \leq (1 - \gamma_\sigma^2)^{-1/2} < \infty \). Similarly, we can expand \( f_Z(z + th) = f_Z(z) + th \tilde{f}_Z(z) + (th)^2 \tilde{f}_Z(\tilde{z})/2 \), where \( \tilde{z} \in (z, z + th) \). Therefore, using regularity conditions on \( f_Z \) and \( \Sigma(\cdot) \), we obtain

\[
\max_{j,k} \sup_{d \in [d]} \E \left[ \left| \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] - f_Z^2(z) \tau_{jk}(z) \right| \right] \leq \frac{M_\sigma \tilde{f}_Z}{\sqrt{1 - \gamma_\sigma^2}} h^2 \tag{C.5.19}
\]

as desired. Proof of (C.2.14) follows in the same way since

\[
\E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \right] = \int \int K(t_1)K(t_2)f_Z(z + t_1h)f_Z(z + t_2h) dt_1 dt_2
\]

\[
= f_Z^2(z) + O(h^2).
\]

Similarly, we can also bound the expectation of the cross product term

\[
n^{-1} \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \times \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \right]
\]

\[
= \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \right] - \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \right] \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \right]
\]

\[
= \frac{4}{n^2(n-1)^2} \sum_{i \neq j, i \neq k} \E \left[ K_h(z_i - z)K_h(z_j - z)g_{z,(j,k)}(y_i, y_j) \right] - (f_Z^2(z) \tau_{jk}(z) + O(h^2))^2
\]

\[
= \frac{24}{n^2(n-1)^2} \left\{ \sum_{i \neq j, i \neq k} \E \left[ K_h(z_i - z) \right] \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)} \right] \right] \right. 
\]

\[
+ \left( \frac{n}{4} \right) \E \left[ K_h(z_i - z) \right] \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)}(y_i, y_j)K_h(z_i - z) \right] \right] 
\]

\[
+ \left( \frac{n}{3} \right) \E \left[ \sum_{i \in (0,1)} \left[ g_{z,(j,k)}(y_i, y_j)K_h(z_i - z)K_h(z_j - z) \right] \right] \left\} - (f_Z^2(z) + O(h^2))^2 = O \left( \frac{1}{nh} \right).
\]
and the expectation of the square term

\[
    n^{-1} \mathbb{E} \left[ (u_n [K_h(z_i - z)K_h(z_i - z)])^2 \right] \\
    = \mathbb{E}[\mathbb{U}_n^2 [K_h(z_i - z)K_h(z_i - z)]] - \mathbb{E}^2[\mathbb{U}_n [K_h(z_i - z)K_h(z_i - z)]] \\
    = \frac{4}{n^2(n-1)^2} \sum_{i \neq j, s \neq t} \mathbb{E} [K_h(z_i - z)K_h(z_j - z)K_h(z_s - z)K_h(z_t - z)] - (f^2_Z(z) + O(h^2))^2 \\
    = \frac{24}{n^2(n-1)^2} \left\{ \left( \frac{n}{4} \right) \mathbb{E}^4 [K_h(z_i - z)] + \left( \frac{n}{3} \right) \mathbb{E}^2 [K_h(z_i - z)] \mathbb{E} [K_h(z_i - z)]^2 \\
    + \left( \frac{n}{3} \right) \mathbb{E} [K_h(z_i - z)]^4 \right\} - (f^2_Z(z) + O(h^2))^2 = O \left( \frac{1}{nh} \right).
\]

This completes the proof.

### C.6 Proof of Theorem 4.3.3

The proof is similar to the proof of Theorem 4.3.2. Instead of only bounding the maximum over \((j, k) \in E^c\), we also need to control the supremum over \(z \in [z_L, z_U]\). Without loss of generality, we consider the case \(E = \emptyset\) and \([z_L, z_U] = (0, 1)\). We will apply the multiplier bootstrap theory for continuous suprema developed in Chernozhukov et al. (2014a). Let \(W_0\) and its bootstrap counterpart \(W_0^B\) be defined as

\[
    W_0 = \max_{j, k \in [d]} \sup_{z \in (0, 1)} \frac{1}{\sqrt{n}} \sum_{i=1}^n J_{z,(j,k)}(Y_i) \quad \text{and} \quad W_0^B = \max_{j, k \in [d]} \sup_{z \in (0, 1)} \frac{1}{\sqrt{n}} \sum_{i=1}^n J_{z,(j,k)}(Y_i) \cdot \xi_i.
\]

**Step 1.** We aim to approximate \(W_0\) by a Gaussian process. In order to apply Theorem A.1 of Chernozhukov et al. (2014a), we need to study the variance of \(J_{z,(j,k)}\) and covering number of the function class \(\mathcal{J} = \{ J_{z,(j,k)} \mid z \in (0, 1), j, k \in [d] \} \).
By the definition of $J_{z|j,k}$ in (C.2.10), we have

$$
\sup_{z,j,k} \mathbb{E}[J_{z|j,k}^2(Y)] \leq \sup_{z,j,k} 4\|\Omega_j(z)\|^2 \|\Omega_k(z)\|^2 \cdot \pi^2 h \cdot \left( \mathbb{E}[\max_{u,v} g_{z(u,v)}^{(1)}(Y_i)]^2 + \mathbb{E}[\omega_z^{(1)}(Z_i)]^2 \right)
$$

$$
\leq CM^2 := \sigma_j^2, \quad (C.6.1)
$$

where the last inequality is due to (C.5.6) and (C.5.15). Similar to (C.2.30), we also have for some constant $C > 0$,

$$
\sup_{z,j,k} \|J_{z|j,k}^2(Y)\|_\infty \leq C/h := b_j.
$$

Furthermore, by Lemma C.8.6, the covering number of $\mathcal{J}$ satisfies

$$
\sup_Q N(\mathcal{J}, L_2(Q), b_j \epsilon) \leq \left( \frac{Cd}{\epsilon h} \right)^c.
$$

Therefore $\mathcal{J}$ is a VC($b_j, C(d/h)^c, c$) type class (see, for example, Definition 3.1 Chernozhukov et al., 2014a) and we can apply Theorem A.1 in Chernozhukov et al. (2014a). Let $K_n = C (\log n \lor \log(d/h))$ for some sufficiently large $C > 0$. Using Theorem A.1 of Chernozhukov et al. (2014a), there exists a random variable $W^0$ such that for any $\gamma \in (0, 1)$,

$$
\mathbb{P}\left( |W_0 - W^0| \geq \frac{K_n/\sqrt{h}}{(\gamma n)^{1/2}} + \frac{(\sigma_j/\sqrt{h})^{1/2} K_n^{3/4}}{\gamma^{1/2} n^{1/4}} + \frac{h^{-1/6} \sigma_j^{2/3} K_n^{2/3}}{\gamma^{1/3} n^{1/6}} \right) \leq C \left( \gamma + \frac{\log n}{n} \right).
$$

Choosing $\gamma = (\log^4(dn)/(nh))^{1/8}$, we have

$$
\mathbb{P}\left( |W_0 - W^0| > C(\log^4(dn)/(nh))^{1/8} \right) \leq C(\log^4(dn)/(nh))^{1/8}. \quad (C.6.2)
$$

**Step 2.** We next bound the difference between $W^B_0$ and $W^0$. Define

$$
\psi_n = \sqrt{\frac{\sigma_j^2 K_n}{n}} + \left( \frac{\sigma_j^2 K_n^3}{nh} \right)^{1/4} \quad \text{and} \quad \gamma_n(\delta) = \frac{1}{\delta} \left( \frac{\sigma_j^2 K_n^3}{nh} \right)^{1/4} + \frac{1}{n}.
$$

177
Since $K_n/h \lesssim (\log(dn))/h \lesssim n\sigma_j^2$, Theorem A.2 of Chernozhukov et al. (2014a) gives us

$$
P\left( |W_0^B - W_0| > \psi_n + \delta \mid \{Y_i\}_{i \in [n]} \right) \leq C\gamma_n(\delta),$$

with probability $1 - 3/n$. Choosing $\delta = (\log(dn))^3/(nh))^{1/8}$, with probability $1 - 3/n,$

$$
P\left( |W_0^B - W_0| > C((\log(dn))^3/(nh))^{1/8} \mid \{Y_i\}_{i \in [n]} \right) \leq C((\log d)^3/(nh))^{1/8}. \quad (C.6.3)$$

**Step 3.** The last step is to assemble the above results to quantify the difference between $W$ and $W^B$. According to Lemma C.7.1 and Lemma C.7.2, there exists a constant $c > 0$ such that

$$
P(\pi - \pi_0 > n^{-c}) \leq n^{-c} \quad \text{and} \quad P\left( P_\pi(\pi_G - \pi_0 > n^{-c}) > n^{-c} \right) \leq n^{-c}. \quad (C.6.4)$$

Let $q_n := [(\log(dn))^3/(nh)]^{1/8}$. From (C.6.4) and (C.6.2), we have

$$
P(\pi - \pi_0 > q_n) \leq q_n. \quad \text{(C.6.5)}$$

Define the event

$$
W = \left\{ P\left( |W_0^B - W_0| > q_n \mid \{Y_i\}_{i \in [n]} \right) \leq q_n \right\}. \quad \text{(C.6.6)}
$$

Using (C.6.4) and (C.6.3), $P(W) \geq 1 - n^{-c}$. Recall that $\hat{c}_W(1 - \alpha, \{Y_i\}_{i=1}^n)$ is $(1 - \alpha)$-quantile of $W_0^B$ conditionally on $\{Y_i\}_{i=1}^n$. Let

$$W_0^B(\{Y_i\}_{i=1}^n) = \max_{j, k \in [d], z \in (0, 1)} n^{-1/2} \sum_{i=1}^n J_{z((j,k)}(Y_i) \cdot \xi_i.$$
and define \( W^B(\{Y_i\}_{i=1}^n) \) similarly. Let \( \hat{\sigma}_{z,j,k}^2 = n^{-1} \sum_{i=1}^n J^2_{z,(j,k)}(Y_i) \), \( \sigma_f = \inf_{z,j,k} \hat{\sigma}_{z,j,k} \) and \( \bar{\sigma}_f = \sup_{z,j,k} \hat{\sigma}_{z,j,k} \). Using the triangle inequality, we have

\[
P(W \leq \hat{\sigma}_W(1 - \alpha, \{Y_i\}_{i=1}^n)) \geq P(W^0 \leq \hat{\sigma}_W(1 - \alpha, \{Y_i\}_{i=1}^n) - q_n) - q_n\]

\[
\geq P(W^0(\{Y_i\}_{i=1}^n) \leq \hat{\sigma}_W(1 - \alpha, \{Y_i\}_{i=1}^n) - 2q_n) - 2q_n
\]

\[
\geq P(W^B(\{Y_i\}_{i=1}^n) \leq \hat{\sigma}_W(1 - \alpha, \{Y_i\}_{i=1}^n))
\]

\[
- C(\sigma_f, \bar{\sigma}_f) q_n (E_\xi[W^0_B] + \sqrt{1 \lor \log(n\sigma_f)}) - Cq_n,
\]

where the last inequality follows from the anti-concentration for suprema of Gaussian processes, given in Lemma A.1 of Chernozhukov et al. (2014b), and \( C(\sigma_f, \bar{\sigma}_f) \) is a constant that only depends on \( \sigma_f \) and \( \bar{\sigma}_f \).

In the following of the proof, we will bound the right hand side of (C.6.7). We first show the constant \( C(\sigma_f, \bar{\sigma}_f) \) is independent to \( h, n \) or \( d \) and then bound \( E_\xi[W^B] \). We bound \( E_\xi[W^B] \) by bounding \( E_\xi[W^0_B] \) and \( E_\xi[[W^B - W^0_B]] \). Since \( W^0_B \) is a supreme of a Gaussian process given data, we can bound its expectation by quantifying its conditional variance and related covering number. We first bound its conditional variance \( n^{-1} \sum_{i=1}^n J^2_{z,(j,k)}(Y_i) \). Using the notation defined in (C.2.29), we have

\[
\sup_{z,j,k} \frac{1}{n} \sum_{i=1}^n (J^2_{z,(j,k)}(Y_i) - E[J^2_{z,(j,k)}(Y_i)]) = \sup_{z,j,k} \frac{1}{n} \sum_{i=1}^n \gamma_{z,(j,k),(j,k)}(Y_i).
\]

Therefore, (C.2.31) and (C.2.32) give an upper bound and variance of \( (J^2_{z,(j,k)}(Y_i) - E[J^2_{z,(j,k)}(Y_i)]) \). Define the function class \( \mathcal{J}_{(2)} = \left\{ J^2_{z,(j,k)} \mid z \in (0, 1), j, k \in [d] \right\} \). Using Lemma C.8.2 and Lemma C.8.6,

\[
\sup_{Q} N(\mathcal{J}_{(2)}, \| \cdot \|_{L^2(Q), \epsilon / \sqrt{h}}) \leq \left( \frac{C(d)}{h\epsilon} \right)^c.
\]
As the upper bound, variance and covering number are quantified above, similar to (C.5.8), we apply Theorem 3.12 of Koltchinskii (2011) to get

\[
\mathbb{E} \left[ \sup_{z,j,k} \frac{1}{n} \sum_{i=1}^{n} (J^2_{z_{j,k}}(Y_i) - \mathbb{E}[J^2_{z_{j,k}}(Y_i)]) \right] \lesssim \sqrt{\frac{\log(2d/h)}{nh^2}} + \frac{\log(2d/h)}{nh}.
\]

Under the assumptions of the theorem, \(\sqrt{\log(2d/h)/(nh^2)} + \log(2d/h)/(nh) = O(n^{-2c})\) and the Markov’s inequality give

\[
P \left( \sup_{z,j,k} \frac{1}{n} \sum_{i=1}^{n} (J^2_{z_{j,k}}(Y_i) - \mathbb{E}[J^2_{z_{j,k}}(Y_i)]) > n^{-c} \right) \leq Cn^{-c}.
\]

Combining with (C.6.1),

\[
\sigma^2_J = \sup_{z,j,k} \frac{1}{n} \sum_{i=1}^{n} J^2_{z_{j,k}}(Y_i) \leq \sigma^2_J + n^{-c} \leq 2\sigma^2_J,
\]

with probability \(1 - Cn^{-c}\). By the assumption in Theorem 4.3.3,

\[
\inf_{z,j,k} \mathbb{E}[J^2_{z_{j,k}}(Y)] = \inf_{z,j,k} \text{Var}(\Omega_j(z)\Theta_k(z)) > c > 0.
\]

Therefore, we have

\[
\sigma_J = \inf_{z,j,k} \frac{1}{n} \sum_{i=1}^{n} J^2_{z_{j,k}}(Y_i)
\geq \inf_{z,j,k} \mathbb{E}[J^2_{z_{j,k}}(Y)] - \sup_{z,j,k} \frac{1}{n} \sum_{i=1}^{n} (J^2_{z_{j,k}}(Y_i) - \mathbb{E}[J^2_{z_{j,k}}(Y_i)]) \geq c/2 > 0
\]

with probability \(1 - Cn^{-c}\). The constant \(C(\sigma_J, \sigma_J)\) does not depend on \(n, d\) and \(h\).

Combining Lemma 2.2.8 in van der Vaart and Wellner (1996) and Lemma C.8.6, we have

\[
\mathbb{E}_\xi [W^B_0] \leq C\sigma_J \sqrt{\log \left( C(d\sigma_J^{-1}/h) \right)} \leq C\sqrt{\log(d/h)}.
\]
From Lemma C.7.2, we have \( \mathbb{E}_\xi[W^B] \leq \mathbb{E}_\xi[W_0^B] + \mathbb{E}_\xi[W^B - W_0^B] \leq C \sqrt{\log(d/h)} \), since \( q_n \sqrt{\log(d/h)} = (\log^8(d/h)/(nh))^{1/8} = O(n^{-c}) \). Due to (C.6.7) and the fact that \( \mathbb{P}(W) \geq 1 - n^{-c} \), we have \( \mathbb{P}(W \leq c_W(1 - \alpha)) \geq 1 - \alpha - 3n^{-c} \). Similarly, we can also show that \( \mathbb{P}(W \geq c_W(1 - \alpha, \{Y_i^n\}_{i=1}^n)) \geq \alpha - 3n^{-c} \), which completes the proof.

\section*{C.7 Auxiliary Lemmas for Score Statistics}

In this section, we provide the auxiliary results for proving auxiliary lemmas on the asymptotic properties on the score statistics.

\subsection*{C.7.1 Approximation Error for Score Statistics}

In this section, we prove Lemma C.7.1 and Lemma C.7.2, which approximate the score statistics by a leading linear term.

\textbf{Lemma C.7.1.} Under the same conditions as Theorem 4.3.3, there exists a universal constant \( c > 0 \) such that

\begin{equation}
\sup_{j,k \in [d]} \sup_{z \in (0,1)} \left| \sqrt{nh} \cdot U_n[\omega_z] \hat{S}_{z|(j,k)}(\hat{\Omega}_{k \setminus j}(z)) - \mathbb{G}_n[J_{z|(j,k)}] \right| \leq n^{-c},
\end{equation}

with probability \( 1 - c/d \).

\textit{Proof.} We have

\begin{align*}
&\left| \sqrt{nh} \cdot U_n[\omega_z] \hat{S}_{z|(j,k)}(\hat{\Omega}_{k \setminus j}(z)) - \mathbb{G}_n[J_{z|(j,k)}] \right| \\
&\leq \sqrt{nh} \cdot U_n[\omega_z] \left| \hat{S}_{z|(j,k)}(\hat{\Omega}_{k \setminus j}(z)) - \Omega_j^T(z)(\hat{\Sigma}(z)\hat{\Omega}_k(z) - e_k^T) \right| \\
&\quad + \left| \sqrt{nh} \cdot U_n[\omega_z] \Omega_j^T(z)(\bar{\Sigma}(z)\Omega_k(z) - e_k^T) - \mathbb{G}_n[J_{z|(j,k)}] \right|
\end{align*}

181
A bound on $I$ is obtained in Lemma C.7.3. Here, we proceed to obtain a bound on $II$.

To simplify the notation, we sometimes omit the argument $z_0$, that is, we write $\Omega(z_0)$ as $\Omega$ and similarly for other parameters indexed by $z$. Applying the Taylor expansion to $\sin(\cdot)$ we obtain

$$\sqrt{n\hbar} \cdot \Omega_j^T(z) \left( \tilde{\Sigma}(z) \Omega_k(z) - e_k \right) = \sqrt{n\hbar} \cdot \Omega_j^T(z) \left( \tilde{\Sigma}(z) - \Sigma(z) \right) \Omega_k(z)$$

$$= \sqrt{n\hbar} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \left( \sin \left( \tau_{ab} \frac{\pi}{2} \right) - \sin \left( \tilde{\tau}_{ab} \frac{\pi}{2} \right) \right)$$

$$= \sqrt{n\hbar} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \cos \left( \tau_{ab} \frac{\pi}{2} \right) \left( \frac{\pi}{2} \left( \tilde{\tau}_{ab} - \tau_{ab} \right) \right)$$

$$- \sqrt{n\hbar} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \sin \left( \tau_{ab} \frac{\pi}{2} \right) \left( \frac{\pi}{2} \left( \tau_{ab} - \tilde{\tau}_{ab} \right) \right)^2,$$

where $\tilde{\tau}_{ij} = (1 - \alpha) \tilde{\tau}_{ij} + \alpha \tau_{ij}$ for some $\alpha \in (0, 1)$.

In order to study the properties of $T_1$ and $T_2$, we first analyze the $\sqrt{n\hbar}(\tilde{\tau}_{ab} - \tau_{ab})$ shared by both terms. Let

$$r_{z_{i(a,b)}}(Y_i, Y_{i'}) := g_{z_{i(a,b)}}(Y_i, Y_{i'}) - \tau_{ab}(z) \omega_z(Z_i, Z_{i'}), \quad \text{(C.7.3)}$$

where $g_{z_{i(a,b)}}$ is defined in (C.2.2). From (4.1.3), we have

$$\sqrt{n\hbar}(\tilde{\tau}_{ab}(z) - \tau_{ab}(z)) = \sqrt{n\hbar} \cdot \sum_{i \neq i'} r_{z_{i(a,b)}}(Y_i, Y_{i'}) \omega_z(Z_i, Z_{i'}).$$

We divide the $U$-statistic in the numerator into two parts

$$\frac{1}{\sqrt{n\hbar}(n-1)} \sum_{i \neq i'} r_{z_{i(a,b)}}(Y_i, Y_{i'}) = \sqrt{\hbar} \cdot u_n[r_{z_{i(a,b)}}] + \sqrt{n\hbar} \cdot \mathbb{E}[r_{z_{i(a,b)}}(Y_i, Y_{i'})].$$
For any \( z \in (0, 1) \) and \( a, b \in [d] \), using Lemma C.2.2 we obtain

\[
\sqrt{nh} \cdot \mathbb{E}[r_{z|(a,b)}(Y_i, Y_{i'})] = \sqrt{nh} \cdot \mathbb{E}\left[\sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \frac{\pi}{2} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \right]
\]

\[
= \sqrt{nh} \cdot \mathbb{E}\left[\sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \frac{\pi}{2} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \right] + \mathbb{E}[r_{z|(a,b)}(Y_i, Y_{i'})] - \tau_{ab}(z) \mathbb{E}[r_{z|(a,b)}(Y_i, Y_{i'})]
\]

\[
= O(\sqrt{nh^5}).
\]

For the leading term \( \sqrt{h} \cdot u_n[r_{z|(a,b)}] \), we combine (C.7.3) with (C.2.5) and (C.2.8), to obtain

\[
\sqrt{h} \cdot u_n[r_{z|(a,b)}] = 2\sqrt{nh} \cdot \mathbb{E}_{n}\left[ g_{z|(a,b)}^{(1)} - \tau_{ab}(z) \omega_{z}^{(1)} \right] + \sqrt{nh} \cdot \mathbb{E}_{n}\left[ g_{z|(a,b)}^{(2)} - \tau_{ab}(z) \omega_{z}^{(2)} \right],
\]

where \( g_{z|(a,b)}^{(1)}, g_{z|(a,b)}^{(2)}, \omega_{z}^{(1)}, \omega_{z}^{(2)} \) are defined in (C.2.3), (C.2.4), (C.2.6), (C.2.7) respectively.

With this,

\[
T_1 = [\mathbb{U}_n[\omega_z]]^{-1} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \frac{\pi}{2} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \left( \sqrt{h} \cdot u_n[r_{z|(a,b)}] + \sqrt{nh} \cdot \mathbb{E}[r_{z|(a,b)}(Y_i, Y_{i'})] \right)
\]

\[
= [\mathbb{U}_n[\omega_z]]^{-1} (T_{11} + T_{12} + T_{13}),
\]

where

\[
T_{11} = \sqrt{nh} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \frac{\pi}{2} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \mathbb{E}_{n}\left[ g_{z|(a,b)}^{(1)} - \tau_{ab}(z) \omega_{z}^{(1)} \right],
\]

\[
T_{12} = \sqrt{nh} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \frac{\pi}{2} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \mathbb{U}_{n}\left[ g_{z|(a,b)}^{(2)} - \tau_{ab}(z) \omega_{z}^{(2)} \right]
\]

\[
T_{13} = \sqrt{nh} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \frac{\pi}{2} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \mathbb{E}[r_{z|(a,b)}(Y_i, Y_{i'})].
\]
From (C.2.10), we have that $T_{11} = G_n(J_{z|\{(j,k)\}})$. We proceed to bound the other terms. Using Lemma C.5.2 and Lemma C.5.4, we have

$$
\sup_{z,j,k} |T_{12}| \leq \sup_{z,j,k} \sqrt{nh} \cdot \|\Omega_j\|_1 \|\Omega_k\|_1 \max_{a,b \in [d]} \left( \left| \mathbb{E}_n \left[ g_{z|\{(a,b)\}}^{(2)} \right] \right| + \left| \mathbb{E}_n \left[ \omega_z^{(2)} \right] \right| \right) \lesssim \frac{\log(d/h)}{\sqrt{nh}},
$$

(C.7.5)

with probability $1 - 1/d$. Using (C.7.4), we can bound $T_{13}$ as

$$
\sup_{z,j,k} |T_{13}| \lesssim \frac{\pi}{2} \|\Omega_j\|_1 \|\Omega_k\|_1 \cdot \sqrt{nh} \lesssim \pi M^2 \sqrt{nh^5}.
$$

(C.7.6)

The final step is to bound $T_2$. Using Lemma C.1.1 and Lemma C.1.2,

$$
\sup_{z,j,k} |T_2| \leq \frac{\pi^2}{8} \|\Omega_j\|_1 \|\Omega_k\|_1 \max_{z,a,b} \sqrt{nh} \cdot |\tau_{ab}(z) - \hat{\tau}_{ab}(z)|^2 \leq CM^2 \sqrt{nh} \cdot \left( \frac{\log(d/h)}{nh} + h^4 + \frac{1}{n^2h^2} \right),
$$

(C.7.7)

with probability $1 - 1/d$.

Combining (C.2.17), (C.7.5), (C.7.6), and (C.7.7) with (C.7.2), we finally have

$$
\sup_{j,k \in [d]} \sup_{z \in (0,1)} \left| \sqrt{nh} \cdot \mathbb{E}_n[\omega_z] (\hat{\Omega}_k^{\{(j,k)\}}(z)) - G_n(J_{z|\{(j,k)\}}) \right| \lesssim \frac{\log(d/h)}{\sqrt{nh}} + \sqrt{nh^5}.
$$

with probability $1 - 1/d$. Under the assumptions of the lemma, there exists a constant $c > 0$ such that $\log(d/h)/\sqrt{nh} = o(n^{-c})$ and $\sqrt{nh^5} = o(n^{-c})$, which completes the proof.

\[\square\]

The following lemma states a result analogous to Lemma C.7.1 for the bootstrap test statistic.

184
Lemma C.7.2. Under the same conditions as Theorem 4.3.3, there exists a universal constant $c > 0$ such that for sufficiently large $n$,

$$
\mathbb{P}_\xi \left( \sup_{j, k \in [d]} \sup_{z \in (0,1)} \left| \sqrt{n h} \cdot U_n [\omega^B_z] \hat{S}^B_{z \setminus (j,k)}(\hat{\Omega}_{k \setminus j}(z)) - G^\xi_n [J_{z \setminus (j,k)}] \leq n^{-c} \right| \right) \geq 1 - c/d.
$$

Proof. We have

$$
\left| \sqrt{n h} \cdot U_n [\omega^B_z] \hat{S}^B_{z \setminus (j,k)}(\hat{\Omega}_{k \setminus j}(z)) - G^\xi_n (J_{z \setminus (j,k)}) \right|
\leq \sqrt{n h} \cdot U_n [\omega^B_z] \left( \hat{S}^B_{z \setminus (j,k)}(\hat{\Omega}^B_{k \setminus j}(z)) - \Omega^T_j(z) \left( \hat{\Sigma}(z) \Omega_k(z) - e^T_k \right) \right)
+ \left( \sqrt{n h} \cdot U_n [\omega^B_z] \Omega^T_j(z) \left( \hat{\Sigma}^B(z) \Omega_k(z) - e^T_k \right) - G^\xi_n [J_{z \setminus (j,k)}] \right).
$$

Lemma C.7.4 gives a bound on $I$. Here, we focus on obtaining a bound on $II$. That is, we show that $\sqrt{n h} \cdot U_n [\omega^B_z] \Omega^T_j(z) \left( \hat{\Sigma}^B(z) \Omega_k(z) - e^T_k \right)$ is close to the linear leading term

$$
T^B_0 := G^\xi_n [J_{z \setminus (j,k)}] = \frac{1}{\sqrt{n}} \sum_{i=1}^n J_{z \setminus (j,k)}(Y_i) \cdot \xi_i.
$$

Similar to (C.7.2), we have

$$
\max_{j, k \in [d]} \sup_{z \in (0,1)} \sqrt{n h} \cdot U_n [\omega^B_z] \Omega^T_j(\hat{\Sigma}^B(z) \Omega_k(z) - e^T_k) = T^B_1 + T^B_2,
$$

where

$$
T^B_1 = \sqrt{n h} \sum_{a, b \in [d]} \Omega_{ja} \Omega_{bk} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \frac{\pi}{2} U_n [\omega^B_z] (\hat{\tau}^B_{ab} - \tau_{ab})
$$
and
\[ T_2^B = -\frac{\sqrt{n}h}{2} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \sin \left( \tau_{ab} \frac{\pi}{2} \right) \mathbb{U}_n[\omega_z] \left( \frac{\pi}{2} (\tilde{\tau}_{ab} - \tau_{ab}) \right)^2. \]

We bound \( T_2^B \) first. Note that
\[ f_Z^2 - \sup_{z \in (0,1)} |\mathbb{U}_n[\omega_z^B] - f_Z^2(z)| \leq \inf_{z \in (0,1)} \mathbb{U}_n[\omega_z^B] \leq \sup_{z \in (0,1)} \mathbb{U}_n[\omega_z^B] \leq f_Z^2 + \sup_{z \in (0,1)} |\mathbb{U}_n[\omega_z^B] - f_Z^2(z)|. \]

Using Lemma C.7.6 and \( \log(h^{-1})/(nh^2) = o(1) \), we have
\[ f_Z^2/2 \leq \inf_{z \in (0,1)} \mathbb{U}_n[\omega_z] \leq \sup_{z \in (0,1)} \mathbb{U}_n[\omega_z] \leq 2f_Z^2, \quad (C.7.8) \]
with probability at least \( 1 - 1/n \). Similar to (C.7.7), (C.7.8) and the Hölder’s inequality give us
\[ |T_2^B| \leq \frac{\pi^2}{4f_Z^2} \|\Omega_j\| \|\Omega_k\| \max_{a,b \in [d]} \sqrt{n}h \cdot |\mathbb{U}_n[\omega_z^B](\tilde{\tau}_{ab} - \tau_{ab})|^2. \]

Under the assumptions, \( h \asymp n^{-\delta} \) for \( \delta \in (1/5, 1/4) \), and Lemma C.7.5 gives us
\[ \mathbb{P}_\epsilon \left( \max_{z,j,k} |T_2^B| \leq C \log(d/h)/\sqrt{nh^3} \right) \geq 1 - 1/d \quad (C.7.9) \]
for some constant \( C > 0 \) with probability \( 1 - c/d \).

Next, we handle the difference between \( T_1^B \) and \( T_0^B \). We denote
\[ \Delta W_z|(a,b) = \mathbb{U}_n[\omega_z^B] (\tilde{\tau}_{ab}(z) - \tau_{ab}(z)) - \frac{2}{n} \sum_{i=1}^n \left[ g_{z|(a,b)}(Y_i) - \tau_{ab}(z) \omega_z^{(1)}(Z_i) \right] \xi_i. \quad (C.7.10) \]
Using the Hölder’s inequality, we have

\[ |T^B_1 - T^B_0| \leq \sqrt{nh} \sum_{a,b \in [d]} \Omega_{ja} \Omega_{bk} \cos \left( \frac{\tau_{ab} \pi}{2} \right) \frac{\pi}{2} \cdot \Delta W_{z|a,b} \leq CM^2 \sqrt{nh} \max_{a,b} |\Delta W_{z|a,b}|. \]

Combining with (C.7.17), there exists a constant \( c > 0 \) such that with probability \( 1 - c/d \)

\[ P_\xi \left( \sup_{z \in (0,1)} \max_{j,k \in [d]} |T^B_1 - T^B_0| \geq C \sqrt{\log(d/h)/(nh^2)} \right) \leq 1/d. \]  

(C.7.11)

If \( \log(d/h)/(nh^2) \approx n^{-c} \), (C.7.14), (C.7.9) and (C.7.11) give us with probability \( 1 - c/d \)

\[ P_\xi \left( \sup_{z \in (0,1)} \max_{j,k \in [d]} |\hat{S}^B_{z|j,k} (\hat{\Omega}_{k\backslash j}(z)) - T^B_0| \leq n^{-c} \right) \geq 1 - 1/d, \]

which completes the proof of the lemma.

\( \square \)

### C.7.2 First Step Approximation Results

Here we establish results needed for the first step in the proofs of Lemma C.7.1 and Lemma C.7.2.

**Lemma C.7.3.** Under the same conditions as Theorem 4.3.3, there exists a universal constant \( c > 0 \) such that

\[ \sup_{j,k \in [d]} \sup_{z \in (0,1)} \sqrt{nh} \cdot \left| \hat{S}_{z|j,k} (\hat{\Omega}_{k\backslash j}(z)) - \Omega^T_{j}(z) \left( \hat{\Sigma}(z) \Omega_k(z) - \mathbf{e}_k^T \right) \right| \leq n^{-c} \]

with probability \( 1 - c/d \).

**Proof.** For any matrix \( \mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_d) \in \mathbb{R}^{d \times d} \), we define

\[ \mathbf{A}_{-j} := (\mathbf{A}_1, \ldots, \mathbf{A}_{j-1}, \mathbf{A}_{j+1}, \ldots, \mathbf{A}_d) \in \mathbb{R}^{d \times (d-1)}, \]

187
which is a submatrix of $A$ with column $j$ removed. We also denote
\[
\gamma^*(z) = (\Omega_{k1}(z), \ldots, \Omega_{k(j-1)}(z), \Omega_{k(j+1)}(z) \ldots \Omega_{kd}(z))^T \in \mathbb{R}^{d-1}
\]
and
\[
\hat{\gamma}(z) = (\hat{\Omega}_{k1}(z), \ldots, \hat{\Omega}_{k(j-1)}(z), \hat{\Omega}_{k(j+1)}(z), \ldots, \hat{\Omega}_{kd}(z))^T \in \mathbb{R}^{d-1}.
\]

To simplify the notation, we sometimes omit the varying variable $z$ in the proof.

We start by decomposing the score function into two parts. We will then identify the leading term and bound the remainder. With the above introduced notation, we have
\[
\sqrt{nh} \cdot \tilde{S}_{z|j,k}(\hat{\Omega}_{k\setminus j}) = \sqrt{nh} \cdot \hat{\Omega}_j^T (\hat{\Sigma}_{-j} \hat{\gamma} - e_k^T) = \sqrt{nh} \cdot \Omega_j^T (\Sigma_{-j} \gamma^* - e_k^T) + \sum_{I_1} \sqrt{nh} \cdot \hat{\Omega}_j^T \Sigma_{-j} (\hat{\gamma} - \gamma^*). \]

The first term in the display above is the leading term on the right hand side of (C.7.1) as desired. The remaining part of the proof is to bound $I_1$ and $I_2$. By the Hölder’s inequality, we have
\[
|I_1| = \sqrt{nh} \cdot (\hat{\Omega}_j - \Omega_j)^T (\hat{\Sigma}_{-j} - \Sigma_{-j}) \gamma^* \leq \sqrt{nh} \cdot \|\hat{\Omega}_j - \Omega_j\|_1 \|\hat{\Sigma} - \Sigma\|_{\max} \|\Omega_k\|_1.
\]

Assumption 4.3.4 together with the display above gives
\[
\sup_{z \in (0,1)} \max_{j,k \in [d]} |I_1| \leq M \sqrt{nh} \cdot r_{1n} r_{2n} \quad \text{(C.7.12)}
\]
with probability $1 - 1/d$. Next, using the Hölder’s inequality we have
\[
|I_2| \leq \sqrt{nh} \cdot \|\hat{\Omega}_j^T \Sigma_{-j}\|_\infty \|\hat{\gamma} - \gamma^*\|_1.
\]
From Assumption 4.3.4, we have \( \|\hat{\Omega}_j^T \hat{\Sigma}_{-j}\|_\infty \leq \|\hat{\Omega}_j^T \hat{\Sigma}\|_\infty \leq r_{3n} \) with probability 1 \(- 1/d \)
and, therefore,

\[
\sup_{z \in (0,1)} \max_{j,k \in [d]} |I_2| \leq \sqrt{nh} \cdot r_{3n} r_{2n} \tag{C.7.13}
\]

with probability 1 \(- 1/d \). Combining (C.7.12) and (C.7.13), we have

\[
\sup_{j,k \in [d]} \sup_{z \in (0,1)} (|I_1| + |I_2|) \leq \sqrt{nh} \cdot r_{2n} (r_{1n} + r_{3n}) \leq n^{-c}
\]

with probability 1 \(- 2/\overline{d} \), which completes the proof.

\[\square\]

**Lemma C.7.4.** Under the same conditions as Theorem 4.3.3, there exists a universal constant \( c > 0 \) such that with probability 1 \(- 1/d \),

\[
P_\xi \left( \max_{j,k \in [d]} \sup_{z \in (0,1)} \sqrt{nh} \cdot \left| \hat{\Sigma}_{z,(j,k)}^B (\hat{\Omega}_{k\setminus j}(z)) - \hat{\Omega}_j^T(z) \left( \hat{\Sigma}_j(z) \Omega_k(z) - \mathbf{e}_k^T \right) \right| \leq n^{-c} \right) \geq 1 - 1/d. \tag{C.7.14}
\]

**Proof.** The proof is similar to the proof of Lemma C.7.3. Compared to the proof in Lemma C.7.3, there are two differences: (1) we need to bound \( \bar{U}_n[\omega^B_z]\|\hat{\Sigma}^B - \Sigma\|_{\text{max}} \) instead of \( \|\hat{\Sigma} - \Sigma\|_{\text{max}} \) and (2) obtain a rate for \( \bar{U}_n[\omega^B_z]\|\hat{\Omega}_j^T \hat{\Sigma}^B\|_\infty \) instead of \( \|\hat{\Omega}_j^T \hat{\Sigma}\|_\infty \).

According to Lemma C.7.5 and (C.7.8), we have

\[
P_\xi \left( \max_{z \in (0,1)} \sup_{j \neq k \in [d]} \sqrt{nh} \cdot \left| \hat{\Sigma}_{z,(j,k)}^B (\hat{\Omega}_{k\setminus j}(z)) - \hat{\Omega}_j^T(z) \left( \hat{\Sigma}_j(z) \Omega_k(z) - \mathbf{e}_k^T \right) \right| \right) \leq 1/d. \tag{C.7.15}
\]

Next, (C.7.8) and the Hölder’s inequality, give us

\[
\bar{U}_n[\omega^B_z]\|\hat{\Omega}_j^T \hat{\Sigma}^B\|_\infty \leq 2\overline{f}_Z \left( \|\hat{\Omega}_j^T \hat{\Sigma}\|_\infty + \left( \|\hat{\Omega}_j - \Omega_j\|_1 + \|\Omega_j\|_1 \right) \bar{U}_n[\omega^B_z]\|\hat{\Sigma}^B - \hat{\Sigma}\|_{\text{max}} \right).
\]

189
Therefore, by Assumption 4.3.4, (C.7.15) and Theorem 4.3.5, with probability $1 - 1/d$,

$$
\mathbb{P}_\xi \left( \| U_n [\omega^B \varphi] \| \hat{\Omega}_j \hat{\Sigma}^B \|_\infty > C \left( M \lambda + M \sqrt{\log(d/h)/(nh^2)} \right) \right) \leq 1/d. 
$$

(C.7.16)

Compared to the rate on $I_1$ in (C.7.12), we have

$$
I_{12}^B := \sqrt{nh} \cdot (\hat{\Omega}_j - \Omega_j)^T (\hat{\Sigma}^B_j - \Sigma_j) \gamma^* \leq \sqrt{nh} \cdot \| \hat{\Omega}_j - \Omega_j \|_1 \| \hat{\Sigma}^B_j - \Sigma_j \|_\max \| \gamma^* \|_1.
$$

For $\lambda = \kappa \sqrt{\log(dn)} \cdot (h^2 + 1/\sqrt{nh})$ and $h = n^{-\delta}$, for $1/5 < \delta < 1/4$, by (C.7.15), we have with probability $1 - 1/d$, there exists a constant $c$ such that

$$
\mathbb{P}_\xi \left( \sup_{z,j,k} |I_{12}^B| > sM^2 \lambda \sqrt{\log(d/h)/h} \right) \leq 1/d.
$$

Instead of $I_2$ in (C.7.13), we define $I_2^B := \sqrt{nh} \cdot \| U_n^B (\omega \varphi) \| \hat{\Omega}_j \hat{\Sigma}^B \|_\infty \| \hat{\gamma} - \gamma^* \|_1$. By (C.7.16) and Theorem 4.3.5, we have with probability $1 - 1/d$,

$$
\mathbb{P}_\xi \left( \sup_{z,j,k} |I_2^B| \leq C \sqrt{nh} \cdot sM^2 \lambda (\lambda + \sqrt{\log(d/h)(nh^2)}) \right) \leq 1/d.
$$

For $\lambda = C \Sigma \left( h^2 + \sqrt{\log(d/h)/(nh)} \right)$, as $s \log d/\sqrt{nh^3} = o(n^{-c})$ and $\sqrt{nh^5} = o(n^{-c})$, we have with probability $1 - 1/d$,

$$
\mathbb{P}_\xi \left( \sup_{j,k} \sup_{z \in (0,1)} \sqrt{nh} \cdot \left| \hat{S}_{\varphi j,k} (\varphi \varphi_k) (\hat{\Omega}_j \hat{\Sigma}^B_k (\varphi \varphi_k) - \Omega_j^T (\varphi \varphi_k) \Omega_k (\varphi \varphi_k) - \varepsilon_k^T) \right| > n^{-c} \right)
\leq \mathbb{P}_\xi \left( \sup_{j,k} \sup_{z \in (0,1)} \left( |I_{12}^B| + |I_2^B| \right) > n^{-c} \right) \leq 1/d,
$$

following the same proof of Lemma C.7.3. The proof is therefore complete.

\[ \square \]
C.7.3 Properties of Bootstrap Score Statistics

In this section, we focus on establishing certain properties of the Gaussian multiplier bootstrap statistics introduced in this thesis. The main goal is to prove Lemma C.7.2, which states the approximation rate of a leading linear term to the bootstrap score statistic. To that end, we establish a rate of convergence for the bootstrap Kendall’s tau estimator \( \hat{\tau}_{jk}(z) \) parallel to the results for \( \hat{\tau}_{jk}^B(z) \) in Lemma C.2.1.

Recall from (4.2.10) and (4.2.13) that
\[
\hat{\tau}_{jk}^B(z) = \frac{\sum_{i \neq i'} K_h(Z_i - z) K_h(Z_{i'} - z) \text{sign}(X_{ij} - X_{i'j}) \text{sign}(X_{ik} - X_{i'k})(\xi_i + \xi_{i'})}{\sum_{i \neq i'} K_h(Z_i - z) K_h(Z_{i'} - z) (\xi_i + \xi_{i'})},
\]
and
\[
U_n[\omega^B_z] = \frac{2}{n(n-1)} \sum_{i \neq i'} K_h(Z_i - z_0) K_h(Z_{i'} - z_0) (\xi_i + \xi_{i'}).
\]

The following lemma presents a convergence rate of the bootstrap Kendall’s tau estimator \( \hat{\tau}_{jk}^B(z) \).

Lemma C.7.5. Under the conditions of Lemma C.7.2, with probability \( 1 - c/d \),
\[
P_{\xi} \left( \max_{j, k \in [d]} \sup_{z \in (0, 1)} \left| U_n[\omega^B_z] (\hat{\tau}_{jk}^B(z) - \tau_{jk}(z)) \right| > C \frac{\sqrt{\log(d/h)}}{(nh^2)} \right) \leq \frac{1}{d}
\]
and
\[
P_{\xi} \left( \max_{j, k \in [d]} \sup_{z \in (0, 1)} \sqrt{nh} |\Delta W_{z|(j,k)}| > C \frac{\sqrt{\log(d/h)}}{(nh^2)} \right) \leq \frac{1}{d},
\]
(C.7.17)
with \( \Delta W_{z|(j,k)} \) defined in (C.7.10).
Proof. We first introduce some notation to simplify the proof. Let

$$W_{z(j,k)}(Y_i) = \frac{2}{n-1} \sum_{i' \neq i} \omega_z(Z_i, Z_{i'}) \left( \text{sign}(X_{ij} - X_{ij'}) \text{sign}(X_{ik} - X_{ik'}) - \tau_{jk}(z) \right). \quad (C.7.18)$$

From the definition of $\hat{\tau}_{jk}^B(z)$ in (4.2.10), conditionally on \{\{Y_i\}_{i \in [n]}\}, we have

$$\sqrt{n} \cdot U_n[\omega_{\tau_{jk}}^B] \left( \hat{\tau}_{jk}^B(z) - \tau_{jk}(z) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{z(j,k)}(Y_i) \xi_i \sim N \left( 0, \frac{1}{n} \sum_{i=1}^n W_{z(j,k)}^2(Y_i) \right). \quad (C.7.19)$$

Since the bootstrap process in (C.7.18) is a Gaussian process, we bound its supreme using the Borell’s inequality (see Proposition A.2.1, van der Vaart and Wellner (1996)). The Borell’s inequality requires us bound the following three quantities:

1. the variance of $n^{-1} \sum_{i=1}^n W_{z(j,k)}^2(Y_i)$;
2. the supremum norm of $n^{-1} \sum_{i=1}^n W_{z(j,k)}^2(Y_i)$;
3. the $L^2$ norm covering number of the function class

$$\mathcal{F}_W = \left\{ \omega^{(i)}_{z(j,k)} | z \in (0, 1), j, k \in [d] \right\}, \quad (C.7.20)$$

under the empirical measure $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{Y_i}$, where

$$\omega^{(i)}_{z(j,k)} := \frac{1}{n-1} \sum_{i' \neq i} \omega_z(Z_i, Z_{i'}) \text{sign}(X_{ij} - X_{ij'}) \text{sign}(X_{ik} - X_{ik'}).$$

To bound the variance, we first study $W_{z(j,k)}(Y_i)$ for each single $i \in [n]$. For any bivariate function $f(y_1, y_2)$, define the operator

$$\mathbb{G}^{(i)}_{n-1}[f] = \frac{1}{\sqrt{n-1}} \sum_{i' \neq i}^n (f(Y_{i'}, Y_i) - \mathbb{E}[f(Y_{i'}, Y_i) | Y_i]).$$
Now, can be written as

\[
W_{z_{(j,k)}}(Y_i) = \mathbb{E}[W_{z_{(j,k)}}(Y_i)] + \frac{2}{\sqrt{n-1}} \left( G_{n-1}^{(i)}(g_{z_{(j,k)}}) - \tau_{jk}(z) G_{n-1}^{(i)}(\omega_z) \right)_{j^{(1)}(Y_i)}
\]

\[
+ 2 \left( g_{z_{(j,k)}}^{(1)}(Y_i) - \tau_{jk}(z) \omega_z^{(1)}(Z_i) \right)_{j^{(2)}(Y_i)}. \tag{C.7.21}
\]

From (C.5.4) and (C.5.14), we have that almost surely

\[
\max_{i \in [n]} \sup_{j,k} \sup_{z \in (0,1)} J^{(2)}(Y_i) \leq C h^{-1}. \tag{C.7.22}
\]

Using Lemma C.2.2, we have

\[
\sup_{z,j,k} \mathbb{E}[W_{z_{(j,k)}}(Y_i)] \leq \sup_{z,j,k} 2 \left( |\mathbb{E} \left[ U_n \left[ g_{z_{(j,k)}} \right] \right] - f_Z^2(z) \tau_{jk}(z) | + |\mathbb{E} \left[ U_n \left[ \omega_z \right] \right] - f_Z^2(z) | \right) \leq C h^2. \tag{C.7.23}
\]

Similar to the proof of Lemma C.5.1 and Lemma C.5.3, with probability \(1 - \delta\), we have

\[
\max_{i \in [n]} \sup_{j,k} \sup_{z \in (0,1)} \left| G_{n-1}^{(i)} \left[ h^{3/2} \left( g_{z_{(j,k)}} - \tau_{jk}(z) \omega_z \right) \right] / \sqrt{\log(d/h) \vee \log(n/\delta)} \right| \leq C. \tag{C.7.24}
\]

Plugging (C.7.22), (C.7.23) and (C.7.24) into (C.7.21), with probability 1 − 1/d, we have

\[
\max_{j,k} \frac{1}{n} \sum_{i=1}^{n} W_{z_{(j,k)}}^2(Y_i) \leq \max_{i \in [n]} \max_{j,k} \sup_{z \in (0,1)} W_{z_{(j,k)}}^2(Y_i) \leq C h^{-2}, \tag{C.7.25}
\]

as \(\log(d/h)/(nh) = o(1)\) and \(h = o(1)\).

Next, we bound the covering number of the function class \(\mathcal{F}_W\) defined in (C.7.20). For some \(M_0\) to be determined later, let \(\{K_h(z_{\ell} - \cdot)\}_{\ell \in [M_0]}\) be the \(\epsilon\)-net of

\[
\mathcal{K} = \{ K \left( (s - \cdot)/h \right) \mid s \in (0,1) \}. \]

193
That is, for any \( z \in (0, 1) \), there exists a \( z_\ell \) such that \( \| K_h(z_\ell - \cdot) - K_h(z - \cdot) \|_{L^2(P_n)}^2 \leq \epsilon \). For this \( z_\ell \), we also have

\[
\| \omega^{(i)}_{z(i,j,k)} - \omega^{(i)}_{z_\ell(i,j,k)} \|_{L^2(P_n)}^2 \leq \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{n-1} \sum_{i' \neq i} |\omega_{z_i, z_{i'}} - \omega_{z_\ell, z_{i'}}| \right)^2 \\
\leq \frac{1}{n} \sum_{i=1}^n (K_h(z - Z_i))^2 \left( \frac{1}{n-1} \sum_{i' \neq i} |K_h(z - Z_{i'}) - K_h(z_\ell - Z_{i'})|^2 \right) \\
+ \frac{1}{n} \sum_{i=1}^n (K_h(z - Z_i) - K_h(z_\ell - Z_i))^2 \left( \frac{1}{n-1} \sum_{i' \neq i} |K_h(z_\ell - Z_{i'})|^2 \right) \\
\leq C \epsilon^2 h^{-2}.
\]

Therefore, as \( M_0 \leq (C/\epsilon)^{3w} \), we have \( N(F_W, \| \cdot \|_{L^2(P_n)}, h^{-1}\epsilon) \leq d^2(C/\epsilon)^{3w} \).

Similarly, for the function class

\[
F'_W = \{ W_{z(i,j,k)}(Y_i) \ | \ z \in (0, 1), j, d \in [d] \}
\]

we have

\[
N(F'_W, \| \cdot \|_{L^2(P_n)}, h^{-1}\epsilon) \leq N(F_W, \| \cdot \|_{L^2(P_n)}, h^{-1}\epsilon) \leq d^2(C/\epsilon)^{6w}.
\]

This bound follows by combining the the fact that \( \tau_{jk}(\cdot) \) is Lipschitz (since \( \Sigma_{jk}(\cdot) \in \mathcal{H}(2, L) \)) with Lemma C.8.2 and Lemma C.8.3.

Now, the Dudley’s inequality (see Lemma 2.2.8 in van der Vaart and Wellner, 1996), together with the fact that \( \cup_n [\omega^{B}_{z_0}](\widehat{\tau}^{B}_{jk}(z) - \tau_{jk}(z)) \) is normally distributed conditionally on data (see (C.7.19)), the upper bound and variance bound in (C.7.25) and the covering number on \( F'_W \) above, gives us

\[
E \left[ \sup_{z \in (0, 1), j,k \in [d]} \max_{i} \cup_n [\omega^{B}_{z_0}](\widehat{\tau}^{B}_{jk}(z) - \tau_{jk}(z)) \right] \leq C \sqrt{\frac{\log(d/h)}{nh^2}}.
\]

194
Using the Borell’s inequality, on the event that (C.7.25) is true, we have

\[ P_{\xi} \left( \sup_{z \in (0,1)} \max_{j,k \in [d]} U_n [\omega_{\tau_{z0}}] \left( \hat{\tau}_{jk}^B (z) - \tau_{jk} (z) \right) \geq C \frac{\log (d/h)}{nh^2} \right) \geq \frac{1}{d}. \]

Since (C.7.25) is true with probability \( 1 - \frac{1}{d} \), the first part of the lemma is proved.

Similarly, we can bound \( \Delta W_{z|(j,k)} \). By (C.7.21), we have

\[ \Delta W_{z|(j,k)} = \bar{\xi} \cdot E[W_{z|(j,k)}(Y)] + \frac{1}{n} \sum_{i=1}^{n} \frac{2J^{(1)}(Y_i)}{\sqrt{n-1}} \xi_i, \]  

where \( \bar{\xi} = n^{-1} \sum_{i=1}^{n} \xi_i \). From the concentration of sub-Gaussian random variables, we have

\[ P (|\bar{\xi}| < C \sqrt{\log d/n}) \geq 1 - \frac{1}{d}. \]  

Combining with (C.7.23), we have

\[ P_{\xi} \left( \sup_{z \in (0,1)} \max_{j,k \in [d]} \bar{\xi} \cdot E[W_{z|(j,k)}(Y)] \geq \sqrt{\frac{Ch^4 \log d}{n}} \right) \leq \frac{1}{d}. \]  

According to (C.7.24), with probability \( 1 - \frac{1}{d} \), we have

\[ \sup_{z \in (0,1)} \max_{j,k \in [d]} \frac{1}{n} \sum_{i=1}^{n} \left( \frac{2J^{(1)}(Y_i)}{\sqrt{n-1}} \right)^2 \leq \max_{i \in [n]} \sup_{z \in (0,1)} \max_{j,k \in [d]} \left( \frac{2J^{(1)}(Y_i)}{\sqrt{n-1}} \right)^2 \leq \frac{C \log (d/h)}{nh^3}. \]  

Define the function class \( \tilde{F}_W = \{ J^{(1)}(\cdot) \mid z \in (0,1), j, k \in [d] \} \). By the definition of \( J^{(1)} \) in (C.7.21), we apply Lemma C.8.2 to the covering number of function classes consisting of \( W_{z|(j,k)} \), \( g_{z|(j,k)}^{(1)} \) and \( \omega_{z}^{(1)} \) in (C.7.26), (C.8.2) and (C.8.11) to obtain

\[ N(\tilde{F}_W, \| \cdot \|_{L^2(P_n)}, h^{-1} \epsilon) \leq d^2 (C/\epsilon)^c. \]

The Borell’s inequality, on the event that (C.7.29) is true, gives us

\[ P_{\xi} \left( \max_{j,k \in [d]} \sup_{z \in (0,1)} \frac{1}{n} \sum_{i=1}^{n} J^{(1)}(Y_i) \xi_i \geq C \sqrt{\frac{\log (d/h)}{n^2 h^3}} \right) \leq \frac{1}{d}. \]  

195
Plugging (C.7.28) and (C.7.30) into (C.7.27), the proof of the second part is complete.

The following lemma presents a convergence rate of \( \mathbb{U}_n^B(\omega_z) \) to \( f_{Z}^2(z) \).

**Lemma C.7.6.** Under the conditions of Lemma C.7.2, with probability \( 1 - 1/n \),

\[
\mathbb{P}_\xi \left( \sup_{z \in (0,1)} \left| \mathbb{U}_n[\omega_z^B] - f_{Z}^2(z) \right| > C \sqrt{\log(1/h)/(nh^2)} \right) \leq 1/n.
\]

**Proof.** The proof is similar to that of Lemma C.7.5. We define

\[
\mathbb{W}_{z|(j,k)}(Z_i) = \frac{2}{n-1} \sum_{i' \neq i} \left( \omega_z(Z_i, Z_{i'}) - f_{Z}^2(z) \right).
\]

Conditionally on the data \( \{Y_i\}_{i \in [n]} \),

\[
\sqrt{n} \cdot (\mathbb{U}_n[\omega_z^B] - f_{Z}^2(z)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{W}_{z|(j,k)}(Z_i) \xi_i \sim N \left( 0, \frac{1}{n} \sum_{i=1}^{n} \mathbb{W}_{z|(j,k)}^2(Z_i) \right).
\]

Note that

\[
\mathbb{W}_{z|(j,k)}(Y_i) = \mathbb{E}[\mathbb{W}_{z|(j,k)}(Y_i)] + 2(n - 1)^{-1} \mathcal{C}_{n-1}[\omega_z] + 2 \omega_z^{(1)}(Z_i).
\]

Similar to the proof of Lemma C.5.3 and (C.7.24), with probability \( 1 - \delta \),

\[
\max_{j,k \in [d]} \sup_{z \in (0,1)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{W}_{z|(j,k)}^2(Y_i) \leq \max_{i \in [n]} \max_{j,k \in [d]} \sup_{z \in (0,1)} \mathbb{W}_{z|(j,k)}^2(Y_i) \leq C h^{-2}. \quad (C.7.31)
\]

Using Lemmas C.8.5, C.8.2 and C.8.3, we can bound the covering number of the function class

\[
\mathcal{F}'_W = \{ \mathbb{W}_{z|(j,k)}(Y_i) \mid z \in (0,1) \}
\]

by

\[
N(\mathcal{F}'_W, \| \cdot \|_{L^2(\mathbb{P}_n)}, h^{-1} \epsilon) \leq (C/\epsilon)^{6v}.
\]
The remained of the proof follows the proof of Lemma C.7.5. Using the Dudley’s and Borell’s inequality (see Lemma 2.2.8 and Proposition A.2.1 van der Vaart and Wellner, 1996)), on the event that (C.7.31) is true, we have

\[ \mathbb{P}_\xi \left( \sup_{z \in (0,1)} \left( U_n \omega_z^B - f_Z^2(z) \right) \geq C \sqrt{\frac{\log(n/h)}{nh^2}} \right) \geq \frac{1}{n}. \]

The lemma follows since (C.7.31) holds with probability \( 1 - 1/n \).

### C.7.4 Proof of Lemma 4.3.1

Recall that we defined the matrix \( \Theta^{(i)} \) in (C.2.19) with elements

\[ \Theta_{jk}^{(i)}(z) = \pi \cos \left( \frac{\pi}{2} \tau_{jk}(z) \right) \cdot \frac{1}{n-1} \sum_{i' \neq i} \tau_{jk}^{(1)}(Y_{i'}), \]

and \( \tau_{jk}^{(1)} \) defined in (4.2.5). The strategy of the proof is to establish

\[ \frac{1}{n} \sum_{i=1}^n \left( \Omega^T_j(z_0) \Theta_j(z_0) \Omega_k(z_0) \right)^2 \xrightarrow{P} \operatorname{Var}(\Omega^T_j(z_0) \Theta z_0 \Omega_k(z_0)) \quad \text{and} \quad (C.7.32) \]

\[ [U_n[\omega_{z_0}]]^2 \xrightarrow{P} f_Z^4(z_0). \quad (C.7.33) \]

The lemma then follows from the Slutsky’s theorem.

We first establish (C.7.32). Let

\[ \Delta_1 = \frac{1}{n} \sum_{i=1}^n \left( \Omega^T_j(z_0) \Theta_j(z_0) \Omega_k(z_0) \right)^2 - \frac{1}{n} \sum_{i=1}^n \left( \Omega^T_j(z_0) \Theta_j(z_0) \Omega_k(z_0) \right)^2 \quad \text{and} \]

\[ \Delta_2 = \frac{1}{n} \sum_{i=1}^n \left( \Omega^T_j(z_0) \Theta_j(z_0) \Omega_k(z_0) \right)^2 - \frac{1}{n} \sum_{i=1}^n \left( \Omega^T_j(z_0) \Theta_j(z_0) \Omega_k(z_0) \right)^2. \]
We can bound $\Delta_1$ as

$$|\Delta_1| = \left| (\hat{\Omega}_j(z_0) - \Omega_j(z_0))^T \cdot \frac{1}{n} \sum_{i=1}^{n} (\hat{\Theta}^{(i)}(z_0)\hat{\Omega}_k(z_0)\hat{\Theta}^{(i)}(z_0) \cdot (\hat{\Omega}_j(z_0) + \Omega_j(z_0))^T \right|$$

$$\leq \|\hat{\Omega}_j(z_0) - \Omega_j(z_0)\|_1 \|\hat{\Omega}_j(z_0) + \Omega_j(z_0)\|_1 \|\hat{\Theta}^{(i)}(z_0)\|^2_{\max}$$

$$= O_P(M^4r_{2n}(2\pi)^2 h^{-1}) = o_P(1), \quad (C.7.34)$$

where the second equality is by Assumption 4.3.4 and $\|\hat{\Theta}^{(i)}(z_0)\|_{\max} \leq 2\pi h^{-1/2}$ with probability $1 - c/d$. Similarly,

$$|\Delta_2| = O_P(M^4r_{2n}(2\pi)^2 h^{-1}) = o_P(1). \quad (C.7.35)$$

Next, $\max_{i\in[n]} h^{1/2}\|\hat{\Theta}^{(i)} - \Theta^{(i)}\|_{\max} \leq \Delta_31 + \Delta_32$, where

$$\Delta_31 = \max_{i\in[n], j,k\in[d]} \pi |\cos((\pi/2)\hat{\tau}_{jk}(z_0)) - \cos((\pi/2)\tau_{jk}(z_0))| \cdot \left| \frac{h^{1/2}}{n-1} \sum_{i'\neq i} \tau^{(1)}_{jk}(Y_{i'}) \right|,$$

$$\Delta_32 = \max_{i\in[n], j,k\in[d]} \pi |\cos((\pi/2)\tau_{jk}(z_0))| \cdot h^{1/2} \left| q_{i,jk}(z_0) - \frac{1}{n-1} \sum_{i'\neq i} \tau^{(1)}_{jk}(Y_{i'}) \right|,$$

where $q_{i,jk}$ is defined in (4.2.6). Using Lemma C.1.1 and Lemma C.1.2, with probability $1 - c/d$

$$\Delta_31 \leq \pi^2 \|K\|_{\infty} \cdot \max_{j,k} |\hat{\tau}_{jk}(z_0) - \tau_{jk}(z_0)| \lesssim \left( h^2 + \sqrt{\log(dn)/(nh)} \right).$$

Let $\bar{q}_{i,jk} = q_{i,jk} + \hat{\tau}_{jk}(z_0)$. By the Hoeffding’s inequality and union bound, we have

$$P \left( \max_{i,j,k} h^{3/2} |\bar{q}_{i,jk} - E[\bar{q}_{i,jk}]| > t \right| Y_i \right) \leq 2nd^2 \exp \left( -(n-1)t^2/2 \right).$$

Setting $t = \sqrt{\log(dn)/n}$ and taking the expectation above, with probability $1 - c/d$,\n
$$\max_{i,j,k} h^{1/2} |q_{i,jk} - E[q_{i,jk}]| \leq \sqrt{\log(dn)/n}.$$
Similarly, with probability $1 - c/d$,

$$
\Delta_{32} \leq \pi h^{1/2} \max_{j,k} |\hat{\tau}_{jk}(z_0) - \tilde{\tau}_{jk}(z_0)| + \max_{i,j,k} h^{1/2}|\tilde{q}_{i,j,k} - \mathbb{E}[\tilde{q}_{i,j,k}]| \lesssim \sqrt{\log(dn)/n} + h^{3/2}.
$$

Therefore,

$$
\Delta_3 := \frac{1}{n} \sum_{i=1}^{n} \left( \Omega_j^T(z_0) \hat{\Theta}^{(i)}(z_0) \Omega_k(z_0) \right)^2 - \frac{1}{n} \sum_{i=1}^{n} \left( \Omega_j^T(z_0) \Theta^{(i)}(z_0) \Omega_k(z_0) \right)^2
\leq 4M^4 \pi h^{-1/2} \max_{i \in [n]} \|\hat{\Theta}^{(i)} - \Theta^{(i)}\|_{\text{max}} = O_P\left(h + \sqrt{\log(dn)/(nh^3)}\right) = o_P(1).
\tag{C.7.36}
$$

Finally, by the law of large numbers,

$$
\frac{1}{n} \sum_{i=1}^{n} \left( \Omega_j^T(z_0) \Theta^{(i)}(z_0) \Omega_k(z_0) \right)^2 \xrightarrow{P} \mathbb{E}[\Omega_j^T(z_0) \Theta^{(i)}(z_0) \Omega_k(z_0)]^2 = \text{Var}(\Omega_j^T(z_0) \Theta_{z_0} \Omega_k(z_0)).
$$

Combining (C.7.34), (C.7.35) and (C.7.36), we prove (C.7.32).

Using Lemma C.7.6 and the continuous mapping theorem, we also prove (C.7.33). By the Slutsky’s theorem, we have \( \hat{\sigma}_{jk}(z_0) \xrightarrow{P} \sigma_{jk}(z_0) \).

**C.8 Results on Covering Number**

In this section, we present several results on the covering number of certain function classes. The first two lemmas, Lemmas C.8.2 and C.8.3, are preliminary technical lemmas that will be used to prove Lemmas C.8.4, C.8.5 and C.8.6. Lemma C.8.2 provides bounds on the covering numbers for function classes generated from products and additions of two function classes. Lemma C.8.3 provides a bound on the covering number of a class of constant functions.

Before presenting these lemmas, we first state a result on the covering number of kernel functions.
Lemma C.8.1. (Lemma 22 of Nolan and Pollard, 1987). Let $K : \mathbb{R} \mapsto \mathbb{R}$ be a bounded variation function. The following function class

$$
\mathcal{K} = \left\{ K \left( \frac{s - h}{h} \right) \mid h > 0, s \in \mathbb{R} \right\}
$$

indexed by the kernel bandwidth satisfies the uniform entropy condition

$$
\sup_Q N(\mathcal{K}, L_2(Q), \epsilon) \leq C \epsilon^{-v}, \quad \text{for all } \epsilon \in (0, 1), \quad (C.8.1)
$$

for some $C > 0$ and $v > 0$.

The following lemma is about the covering number of summation and product of functions.

Lemma C.8.2. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two function classes satisfying

$$
N(\mathcal{F}_1, \| \cdot \|_{L_2(Q)}, a_1 \epsilon) \leq C_1 \epsilon^{-v_1} \quad \text{and} \quad N(\mathcal{F}_2, \| \cdot \|_{L_2(Q)}, a_2 \epsilon) \leq C_2 \epsilon^{-v_2}
$$

for some $C_1, C_2, a_1, a_2, v_1, v_2 > 0$ and any $0 < \epsilon < 1$. Define $\| \mathcal{F}_\ell \|_\infty = \sup\{\|f\|_\infty, f \in \mathcal{F}_\ell\}$ for $\ell = 1, 2$ and $U = \| \mathcal{F}_1 \|_\infty \vee \| \mathcal{F}_2 \|_\infty$. For the function classes $\mathcal{F}_x = \{f_1 f_2 \mid f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$ and $\mathcal{F}_+ = \{f_1 + f_2 \mid f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$, we have for any $\epsilon \in (0, 1)$,

$$
N(\mathcal{F}_x, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C_1 C_2 \left( \frac{2a_1 U}{\epsilon} \right)^{v_1} \left( \frac{2a_2 U}{\epsilon} \right)^{v_2},
$$

$$
N(\mathcal{F}_+, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C_1 C_2 \left( \frac{2a_1}{\epsilon} \right)^{v_1} \left( \frac{2a_2}{\epsilon} \right)^{v_2}.
$$

Proof. For any $\epsilon \in (0, 1)$, let $\mathcal{N}_1 = \{f_{11}, \ldots, f_{1N_1}\}$ and $\mathcal{N}_2 = \{f_{21}, \ldots, f_{2N_2}\}$ be the $\epsilon/(2U)$-net of $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively with

$$
N_1 \leq C_1 \left( \frac{2a_1 U}{\epsilon} \right)^{v_1} \quad \text{and} \quad N_2 \leq C_2 \left( \frac{2a_2 U}{\epsilon} \right)^{v_2}.
$$
Define the set \( N = \{ f_{1j}f_{2k} \mid f_{1j} \in N_1, f_{2k} \in N_2 \} \). We now show that \( N \) is an \( \epsilon \)-net for \( F \). For any \( f_{1j}f_{2k} \in F \), there exist two functions \( f_{1j} \in N_1 \) and \( f_{2k} \in N_2 \) such that \( \| f_1 - f_{1j} \|_{L^2(Q)} \leq \epsilon/(2U) \) and \( \| f_2 - f_{2k} \|_{L^2(Q)} \leq \epsilon/(2U) \). Moreover, we have \( f_{1j}f_{2k} \in N \) and

\[
\| f_{1j}f_{2k} - f_{1j}f_{2k} \|_{L^2(Q)} \leq \| f_1 - f_{1j} \|_{L^2(Q)} + \| f_2 - f_{2k} \|_{L^2(Q)} \leq \epsilon.
\]

Therefore \( N \) is the \( \epsilon \)-net for \( F \). Similarly, we also have

\[
\| (f_1 + f_2) - (f_{1j} + f_{2k}) \|_{L^2(Q)} \leq \| f_1 - f_{1j} \|_{L^2(Q)} + \| f_2 - f_{2k} \|_{L^2(Q)} \leq \epsilon/2U.
\]

So \( N' = \{ f_{1j} + f_{2k} \mid f_{1j} \in N_1, f_{2k} \in N_2 \} \) is the \( \epsilon/2U \)-net of \( F \). We finally complete the proof by showing that

\[
|N'| = |N| = N_1N_2 \leq C_1C_2 \left( \frac{2a_1U}{\epsilon} \right)^{v_1} \left( \frac{2a_2U}{\epsilon} \right)^{v_2}.
\]

\[\square\]

**Lemma C.8.3.** Let \( f(s) \) be a Lipschitz function defined on \( [a, b] \) such that \( |f(s) - f(s')| \leq L_f|s - s'| \) for any \( s, s' \in [a, b] \). We define the constant function class \( F_c = \{ g_s(\cdot) \equiv f(s) \mid s \in [a, b] \} \). For any probability measure \( Q \), the covering number of \( F_c \) satisfies for any \( \epsilon \in (0, 1) \),

\[
N(F_c, \|\cdot\|_{L^2(Q)}, \epsilon) \leq L_f \cdot \frac{|b - a|}{\epsilon}.
\]

**Proof.** Let \( N = \{ a + i\epsilon/L_f \mid i = 0, \ldots, \lfloor L_f|b - a|/\epsilon \rfloor \} \). For any \( g_{s_0} \in F_c \), there exists a \( s \in N \) such that \( |s_0 - s| \leq \epsilon/L_f \) and we have

\[
\| g_{s_0} - g_s \|_{L^2(Q)} = |f(s_0) - f(s)| \leq L_f|s_0 - s| \leq \epsilon.
\]
Therefore $\{g_s \mid s \in \mathcal{N}\}$ is the $\epsilon$-net of $\mathcal{F}_c$. As $|\mathcal{N}| \leq L_f|b - a|/\epsilon$, the lemma is proved. 

The following lemma presents the covering number of function classes consisting of $g_{z|_{(j,k)}}^{(1)}(\cdot)$ or $g_{z|_{(j,k)}}^{(2)}(\cdot)$ defined in (C.2.3) and (C.2.4).

**Lemma C.8.4.** For some $0 < \underline{h} < \bar{h} < 1$, we consider the class of functions

$$
\mathcal{F}^{(1)} = \left\{ \sqrt{h} \cdot g_{z|_{(j,k)}}^{(1)} \mid h \in [\underline{h}, \bar{h}], z \in (0, 1), j, k \in [d] \right\}; \\
\mathcal{F}^{(2)} = \left\{ h \cdot g_{z|_{(j,k)}}^{(2)} \mid h \in [\underline{h}, \bar{h}], z \in (0, 1), j, k \in [d] \right\},
$$

where $g_{z|_{(j,k)}}^{(1)}$ and $g_{z|_{(j,k)}}^{(2)}$ are defined in (C.2.3) and (C.2.4). There exist constants $C_{(1)}$ and $C_{(2)}$ such that for any $\epsilon \in (0, 1)$

$$
\sup_Q N(\mathcal{F}^{(1)}, \|\cdot\|_{L_2(Q)}, \epsilon) \leq \frac{d^2 C_{(1)}}{\bar{h}^{v+9} \epsilon^{v+6}} \text{ and } \sup_Q N(\mathcal{F}^{(2)}, \|\cdot\|_{L_2(Q)}, \epsilon) \leq \frac{d^2 C_{(2)}}{\bar{h}^{2v+18} \epsilon^{4v+15}}.
$$

**Proof.** Recall that

$$
g_{z|_{(j,k)}}^{(1)}(y) = \mathbb{E}[g_{z|_{(j,k)}}(y, Y)] - \mathbb{E}\left[\mathbb{U}_n(g_{z|_{(j,k)}})\right] \quad \text{and} \quad g_{z|_{(j,k)}}^{(2)}(y_1, y_2) = g_{z|_{(j,k)}}(y_1, y_2) - g_{z|_{(j,k)}}^{(1)}(y_1) - g_{z|_{(j,k)}}^{(1)}(y_2) - \mathbb{E}\left[\mathbb{U}_n(g_{z|_{(j,k)}})\right].
$$

Our proof strategy for bounding the covering numbers of $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ is to decompose them into the following three auxiliary function classes:

$$
\mathcal{F}_{1,jk}^{(1)} = \left\{ \mathbb{E}[g_{z|_{(j,k)}}(y, Y)] \mid h \in [\underline{h}, \bar{h}], z \in (0, 1) \right\}; \\
\mathcal{F}_{2,jk}^{(1)} = \left\{ \mathbb{E}[\mathbb{U}_n(g_{h,z})] \mid h \in [\underline{h}, \bar{h}], z \in (0, 1) \right\}; \\
\mathcal{F}_{jk}^{(2)} = \left\{ g_{z|_{(j,k)}}(y_1, y_2) \mid h \in [\underline{h}, \bar{h}], z \in (0, 1) \right\}.
$$
Observe that we can write

\[ \mathcal{F}^{(1)} = \{ \sqrt{h} \cdot (f_1 - f_2) \mid h \in [h, \bar{h}], f_1 \in \mathcal{F}^{(1)}_{1,jk}, f_2 \in \mathcal{F}^{(1)}_{2,j,k} \}; \]

\[ \mathcal{F}^{(2)} = \{ h \cdot (f_1 - f_2 - f_3 - f_4) \mid h \in [h, \bar{h}], f_1 \in \mathcal{F}^{(2)}_{jk}, f_2, f_3 \in \mathcal{F}^{(1)}_{1,jk}, f_4 \in \mathcal{F}^{(1)}_{2,j,k} \}. \]

Therefore, we can apply Lemma C.8.2 on the addition and product of functions classes and Lemma C.8.3 on the constant functions to bound the covering numbers of \( \mathcal{F}^{(1)} \) and \( \mathcal{F}^{(2)} \).

**Covering number of** \( \mathcal{F}^{(1)}_{1,jk} \). We bound the covering number of \( \mathcal{F}^{(1)}_{1,jk} \) first. Recall from (C.2.24) that for \( y' = (z', x') \)

\[ \mathbb{E} \left[ g_{z,(j,k)}(y', Y) \right] = K_h(z' - z) \int K_h(s - z) \varphi \left( x'_j, x'_k, \Sigma_{jk}(s) \right) f_Z(s)ds. \]

We have \( \mathcal{F}^{(1)}_{1,jk} = \{ f_1 \cdot f_2 \mid f_1 \in \{ K_h(\cdot - z) \}, f_2 \in \mathcal{F}^{(1)}_{3,jk} \} \) where

\[ \mathcal{F}^{(1)}_{3,jk} = \left\{ q_{z,h}(x, y) = \int K_h(s - z) \varphi (x, y, \Sigma_{jk}(s)) f_Z(s)ds, h \in [h, \bar{h}], z \in (0, 1) \right\}. \tag{C.8.3} \]

Let \( \varphi_{x,y}(s) = \varphi (x, y, \Sigma_{jk}(s)) f_Z(s) \). Then \( \mathcal{F}^{(1)}_{3,jk} \) is the class of functions generated by the convolution \( q_{z,h}(x, y) = (K_h * \varphi_{x,y})(z) \). The \( L_1 \) norm of the derivative of \( K_h \) can be bounded by

\[ \| K'_h \|_1 = \int \frac{1}{h^2} \left| K' \left( \frac{t}{h} \right) \right| dt = h^{-1} \int |K'(t)|dt = h^{-1} \text{TV}(K), \tag{C.8.4} \]

where \( \text{TV}(K) \) is the total variation of the kernel \( K \). Similarly we have for any \( h \in [h, \bar{h}] \),

\[ \left\| \frac{\partial}{\partial h} K_h \right\|_1 \leq \int h^{-2}|K(t/h)|dt + \int h^{-3}|K'(t/h)|dt = h^{-1} \| K \|_1 + h^{-2} \text{TV}(K). \]
We can apply a similar argument as in (C.5.5) and derive that

\[
\sup_{z_0, h, x, y} \left| \frac{\partial}{\partial z} q_{z, h}(x, y) \right|_{z=z_0} = \sup_{h, x, y} \| K'_h * \tilde{\varphi}_{x,y} \|_\infty \leq \sup_{h, x, y} \| K'_h \|_1 \| \tilde{\varphi}_{x,y} \|_\infty \leq 2h^{-1}\text{TV}(K)\bar{f}_Z, \tag{C.8.5}
\]

where the first equality is due to the property of the derivative of a convolution, the first inequality is because of Young's inequality and the last inequality is by (C.8.4). Similarly, we have

\[
\sup_{z_0, h, x, y} \left| \frac{\partial}{\partial h} q_{z, h}(x, y) \right|_{h=h_0} = \sup_{h_0, x, y} \| \nabla_{h=h_0} K_h * \tilde{\varphi}_{x,y} \|_\infty \leq \sup_{h_0, x, y} \| \nabla_{h=h_0} K_h \|_1 \| \tilde{\varphi}_{x,y} \|_\infty \leq 2\bar{f}_Z(h^{-1}\|K\|_1 + h^{-2}\text{TV}(K)). \tag{C.8.6}
\]

Therefore for any \( z_1, z_2 \in (0, 1), \) \( h_1, h_2 \in [h, \bar{h}] \), denoting \( C_h := 2\bar{f}_Z[(h^{-1} + h^{-2})\text{TV}(K) + h^{-1}\|K\|_1] \), we have

\[
\sup_{x, y} |q_{z_1, h_1}(x, y) - q_{z_2, h_2}(x, y)| \leq C_h \max(|z_1 - z_2|, |h_1 - h_2|).
\]

Given any measure \( Q \) on \( \mathbb{R}^2 \), let \( \mathcal{Z} \) be the \( \epsilon/C_h \)-net of \( (0, 1) \times [h, \bar{h}] \) under \( \| \cdot \|_\infty \). For any \( z \in (0, 1), h \in [h, \bar{h}] \), choose \( (z_0, h_0) \in \mathcal{Z} \) such that \( \max(|z - z_0|, |h - h_0|) \leq \epsilon/C_h \) and we have

\[
\|q_{z, h} - q_{z_0, h_0}\|_{\| \cdot \|_{L^2(Q)}} \leq \|q_{z, h} - q_{z_0, h_0}\|_\infty \leq \epsilon.
\]

This shows that \( \{q_{z, h} \mid (z, h) \in \mathcal{Z}\} \) is the \( \epsilon \)-net of \( \mathcal{F}^{(1)}_{\mathcal{A}, jk} \) and

\[
\sup_Q N(\mathcal{F}^{(1)}_{\mathcal{A}, jk}, \| \cdot \|_{L^2(Q), \epsilon}) \leq |\mathcal{Z}| \leq \left( \frac{C_h}{\epsilon} \right)^2. \tag{C.8.7}
\]
According to the formulation in (C.2.24) and Lemma C.8.2 and (C.8.1), we have

$$
\sup_Q N(F_{1,jk}^{(1)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C \left( \frac{1}{h\epsilon} \right)^v \left( \frac{h - h}{h^2 \epsilon} \right) \left( \frac{C_h}{h\epsilon} \right)^2 \leq \frac{CC_h^2}{h^{v+4}\epsilon^{v+3}}. \quad (C.8.8)
$$

**Covering number of $\mathcal{F}_{2}^{(1)}$.** According to (C.2.13) and Assumptions (T) and (D), we have that for any $z_1, z_2 \in (0, 1), h_1, h_2 \in [h, \bar{h}]

$$
\left| \mathbb{E} \left[ \mathcal{U}_n(g^{(1)}_{z_1\{j,k\}}) \right] - \mathbb{E} \left[ \mathcal{U}_n(g^{(1)}_{z_2\{j,k\}}) \right] \right| \leq |f_Z^2(z_1)\tau_{jk}(z_1) - f_Z^2(z_2)\tau_{jk}(z_2)| + C|h_1^2 - h_2^2|
$$

$$
\leq \bar{f}_Z^2 C_\tau |z_1 - z_2| + 2C\bar{h}|h_1 - h_2|.
$$

Using Lemma C.8.3, we have

$$
\sup_Q N(\mathcal{F}_{2}^{(1)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq \left( \frac{\bar{f}_Z^2 C_\tau + 2C\bar{h}}{\epsilon} \right)^2. \quad (C.8.9)
$$

**Covering number of $\mathcal{F}^{(1)}$.** Observe that the function $g(h) = \sqrt{h}$ is Lipschitz on $[h, \bar{h}]$ with Lipschitz constant $(h)^{-1/2}/2$. Combining (C.8.8) and (C.8.9) with Lemma C.8.2, we have

$$
\sup_Q N(\mathcal{F}^{(1)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq d^2 \cdot \frac{2CC_h^2(\bar{f}_Z^2 C_\tau + 2C\bar{h})^2}{h^{v+5}\epsilon^{v+6}}. \quad (C.8.10)
$$

Here the additional $d^2$ on the right hand side of (C.8.10) is because we also take supreme over $j, k \in [d]$ in the definition of $\mathcal{F}^{(1)}$. As $C_h \leq 2\bar{f}_Z(2TV(K) + \| K \|_1) \cdot \bar{h}^{-2}$, defining

$$
C_{(1)} := 4\bar{f}_Z^2(2TV(K) + \| K \|_1)(\bar{f}_Z^2 C_\tau + 2C)^2
$$

and the first part of lemma in (C.8.2) is proved.

**Covering numbers of $\mathcal{F}_{jk}^{(2)}$ and $\mathcal{F}^{(2)}$.** Let $\mathcal{N}_K = \{ K_h(x-z)K_h(y-z) \mid z \in \mathcal{Z}_K, h \in \mathcal{H}_K \}$ be the $\epsilon$-net of the function class $\mathcal{K}^2 = \{ K_h(x-z)K_h(y-z) \mid h \in [h, \bar{h}], z \in (0, 1) \}$. According to Lemma C.8.2 and (C.8.1), we have $|\mathcal{N}_K| \leq C^2(2\| K \|_\infty/\epsilon)^{2n}$. Given any $g_{h,a,z_0} \in \mathcal{F}_{jk}^{(2)}$, there
exist \( h_1 \in \mathcal{Z}_K, z_1 \in \mathcal{H}_K \) such that

\[
\|g_{h_0,z_0} - g_{h_1,z_1}\|_{L_2(Q)} \leq \|K_{h_0}(\cdot - z_0)K_{h_0}(\cdot - z_0) - K_{h_1}(\cdot - z_1)K_{h_1}(\cdot - z_1)\|_{L_2(Q)} \leq \epsilon.
\]

Therefore \( \mathcal{N}_g = \{g_{z}(j,k)(y_1, y_2) \mid z \in \mathcal{Z}_K, h \in \mathcal{H}_K\} \) is an \( \epsilon \)-net of \( \mathcal{H} \) and \( |\mathcal{N}_g| = |\mathcal{N}_K| \leq C^2(2\|K\|_{\infty}/\epsilon)^{2v} \). Applying Lemma C.8.2 again with (C.8.9) and (C.8.10), we have

\[
\sup_Q N(\mathcal{F}^{(2)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq d^2C^2 \left( \frac{\bar{h} - h}{\epsilon} \right) \left( \frac{2\|K\|_{\infty}}{\epsilon} \right)^{2v} \left( \frac{C(1)}{h^{v+9}\epsilon^{v+6}} \right)^2 \left( \frac{\bar{f}_2C + 2C\bar{h}}{\epsilon} \right)^2 
\]

\[
\leq \frac{d^2C(2)}{h^{2v+18}\epsilon^{4v+15}},
\]

where \( C(2) := C^24^v\|K\|_{\infty}^{2v}C(1)^2(\bar{f}_2C + 2C)^2 \). Therefore, we complete the proof of the lemma.

\[\square\]

Similar to Lemma C.8.4, we can also establish the covering number for function classes consisting of \( \omega_z^{(1)} \) or \( \omega_z^{(2)} \) in the following lemma.

**Lemma C.8.5.** For some \( 0 < h < \bar{h} < 1 \), we consider the class of functions

\[
\mathcal{K}^{(1)} = \left\{ \sqrt{\bar{h}} \cdot \omega_z^{(1)} \mid h \in [h, \bar{h}], z \in (0,1) \right\}; \quad (C.8.11)
\]

\[
\mathcal{K}^{(2)} = \left\{ h \cdot \omega_z^{(2)} \mid h \in [h, \bar{h}], z \in (0,1) \right\},
\]

where \( \omega_z^{(1)} \) and \( \omega_z^{(2)} \) are defined in (C.2.6) and (C.2.7). There exist constants \( C'_1 \) and \( C'_2 \) such that for any \( \epsilon \in (0, 1) \)

\[
\sup_Q N(\mathcal{K}^{(1)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq \frac{C'_1}{h^{2v+7}\epsilon^{v+3}} \quad \text{and} \quad \sup_Q N(\mathcal{K}^{(2)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq \frac{C'_2}{h^{4v+14}\epsilon^{4v+8}}.
\]
Proof. The proof is similar to Lemma C.8.4. We first bound \( \sup_Q N(\mathcal{K}^{(1)}, \| \cdot \|_{L_2(Q)}, \epsilon) \). We have

\[
\omega_1(s) = \mathbb{E}[K_h(s - z)K_h(Z - z)] - \mathbb{E} \left[ \mathbb{E} \left[ \mathcal{K} \left( Z_i - z \right) \mathcal{K} \left( Z_{i'} - z \right) \right] \right]
\]

\[
= K_h(s - z)\mathbb{E}[K_h(z - Z)] - \left\{ \mathbb{E}[K_h(z - Z)] \right\}^2.
\]

We first study the covering number of the function class

\[
\mathcal{C}_K = \{ \mathbb{E}[K_h(z - Z)] = (K_h \ast f_Z)(z) | h \in [\underline{h}, \overline{h}], z \in (0, 1) \},
\]

where “\( \ast \)” denotes the convolution. Just as (C.8.3), \( \mathcal{C}_K \) is also generated by convolutions.

Similar to (C.8.5) and (C.8.6), we have

\[
\sup_{z_0, h} \left\| \nabla_z \mathbb{E}[K_h(z - Z)] \right\|_{z = z_0} \leq 2h^{-1} \text{TV}(K)\bar{f}_Z \quad \text{and}
\]

\[
\sup_{z, h_0} \left\| \nabla_h \mathbb{E}[K_h(z - Z)] \right\|_{h = h_0} \leq 2\bar{f}_Z (h^{-1} \| K \|_1 + h^{-2} \text{TV}(K)).
\]

Therefore, following the derivation of (C.8.7), for any \( z_1, z_2 \in (0, 1) \), \( h_1, h_2 \in [\underline{h}, \overline{h}] \), we have

\[
\left| \mathbb{E}[K_{h_1}(z_1 - Z)] - \mathbb{E}[K_{h_2}(z_2 - Z)] \right| \leq C_h \max(|z_1 - z_2|, |h_1 - h_2|),
\]

where \( C_h := 2\bar{f}_Z ([h^{-1} + h^{-2}] \text{TV}(K) + h^{-1} \| K \|_1) \). From Lemma C.8.3, we have

\[
\sup_Q N(\mathcal{C}_K, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C_h / \epsilon \quad \text{and}
\]

\[
\sup_Q N(\{ \mathbb{E}[K_h(z - Z)] \}^2 | z, h, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C_h / \epsilon.
\]

Using the fact that the function \( q(h) = 1/\sqrt{h} \) is Lipschitz on \([\underline{h}, \overline{h}]\) with Lipschitz constant \((h)^{-3/2} / 2\) together with Lemma C.8.2, Lemma C.8.3 and (C.8.1), we have

\[
\sup_Q N(\mathcal{K}^{(1)}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C \left( \frac{1}{h^{2/\epsilon}} \right)^v \left( \frac{C_h}{h\epsilon} \right) \left( \frac{\overline{h} - h}{h^{3/2} \epsilon} \right) \leq \frac{C'_{(1)}}{h^{2v+7} \epsilon^{v+3}},
\]

207
where $C'_1 := [2f_Z(2TV(K) + \|K\|_1)C]^2$. Function class $\mathcal{K}^{(2)}$ contains functions in the form

$$\omega^{(2)}_z(s, t) = K_h(s - z)K_h(t - z) - \omega^{(1)}_z(s) - \omega^{(1)}_z(t) - \mathbb{E} \left[ \mathbb{U}_n(K_h(Z_i - z)K_h(Z_i - z)) \right].$$

By Lemma C.8.2, it suffices to study the covering number of

$$C'_K = \{K_h(s - z)K_h(t - z) \mid h \in [h, \overline{h}], z \in (0, 1)\}.$$  

Using Lemma C.8.2 and (C.8.1) again, we have

$$\sup_Q N(C'_K, \|\cdot\|_{L_2(Q)}, \epsilon) \leq C^2 \left( \frac{2\|K\|_{\infty}}{\epsilon} \right)^{2v},$$

and therefore combining with the covering number in (C.8.11) and (C.8.9)

$$\sup_Q N(\mathcal{K}^{(2)}, \|\cdot\|_{L_2(Q)}, \epsilon) \leq C^2 \left( \frac{2\|K\|_{\infty}}{\epsilon} \right)^{2v} \left( \frac{C'_1}{\epsilon} \right)^2 \left( \frac{C_h}{\epsilon} \right) \left( \frac{\overline{h} - h}{\epsilon} \right) \leq \frac{C'_2}{\epsilon^{4v+8}}.$$

where $C'_2 := C^2 4^v\|K\|_{\infty}^2(C'_1)^2(\bar{Z}_\tau C + 2C)^2$. This completes the proof.  

**Lemma C.8.6.** Suppose $\Omega(z) \in \mathcal{U}(c, M, \rho)$ for all $z \in (0, 1)$. Consider the class of functions $\mathcal{J} = \{J_{z(j, k)} \mid z \in (0, 1), j, k \in [d]\}$, where $J_{z(j, k)}$ is defined in (C.2.10). There exists positive constants $C$ and $c$ such that

$$\sup_Q N(\mathcal{J}, \|\cdot\|_{L_2(Q)}, \epsilon/\sqrt{h}) \leq \left( \frac{Cd}{ht\epsilon} \right)^c.$$

**Proof.** We denote for any $u, v \in [d]$ and $z \in (0, 1)$ that

$$\Phi_{uv}(z; Y) = \pi \cos \left( \tau_{uv}(z) \frac{\pi}{2} \right) \sqrt{h} \cdot \left[ g^{(1)}_{z(u,v)}(Y) - \tau_{uv}(z)\omega^{(1)}(Z) \right] \quad (C.8.12)$$
and the matrix $\Phi(z; Y) = [\Phi_{uv}(z; Y)]_{u,v\in[d]}$. In order to bound the covering number of $\mathcal{J}$, we define a larger function class

$$
\mathcal{J}' = \{ \Omega_j^T(z)\Phi(x; \cdot)\Omega_k(z) \mid z, x \in (0, 1), j, k \in [d] \}.
$$

Given any measure $Q$, $j, k \in [d]$ and $x_1, x_2, z_1, z_2 \in (0, 1)$, we first bound the difference

$$
\| \Omega_j^T(z_1)\Phi(x_1; Y)\Omega_k(z_1) - \Omega_j^T(z_2)\Phi(x_2; Y)\Omega_k(z_2) \|_{L_2(Q)}^2
\leq 3 \| \Omega_j(z_1) - \Omega_j(z_2) \|_1^2 \max_{u,v} \| \Phi_{uv}(x_1; Y) \|_{L_2(Q)}^2 \| \Omega_k(z_1) \|_1^2
\leq 3 \| \Phi_{uv}(x_1; Y) - \Phi_{uv}(z_1; Y) \|_{L_2(Q)}^2 \| \Omega_k(z_1) \|_1^2
\leq 3 \| \Phi_{uv}(x_2; Y) - \Phi_{uv}(z_2; Y) \|_{L_2(Q)}^2 \| \Omega_k(z_1) - \Omega_k(z_2) \|_1^2.
\tag{C.8.13}
$$

Since $\Omega(z) \in \mathcal{U}(c, M, \rho)$ for any $z \in (0, 1)$, we have $\sup_z \| \Omega_j(z) \|_1 \leq M$. Next, using Theorem 2.5 of Stewart et al. (1990), for any $z_1, z_2 \in (0, 1),$

$$
\| \Omega(z_1) - \Omega(z_2) \|_2 \leq \| \Omega(z_1) \|_2 \| \Omega(z_2) - \Sigma(z_1) - \Sigma(z_2) \|_2.
$$

Since $\Omega(z) \in \mathcal{U}(c, M, \rho)$ for any $z \in (0, 1)$, we further have

$$
\| \Omega(z_1) - \Omega(z_2) \|_1 \leq \sqrt{d} \| \Omega(z_1) - \Omega(z_2) \|_2 \leq \rho^2 \| \Sigma(z_1) - \Sigma(z_2) \|_2 \leq \rho^2 d^{3/2} \| \Sigma(z_1) - \Sigma(z_2) \|_{\max}.
$$

Since $\Sigma_{jk}(\cdot) \in \mathcal{H}(2, M_\sigma)$, we have

$$
\| \Omega(z_1) - \Omega(z_2) \|_1 \leq \rho^2 d^{3/2} \| \Sigma(z_1) - \Sigma(z_2) \|_{\max}
\leq \rho^2 d^{3/2} \| T(z_1) - T(z_2) \|_{\max} \leq \rho^2 M_\sigma d^{3/2} |z_1 - z_2|.
\tag{C.8.14}
$$
We next study the covering number of the function class \( \mathcal{J}_{uv} = \{ \Phi_{uv}(z; \cdot) \mid z \in (0,1) \} \).

By (C.5.4) and (C.5.14), we have

\[
\max_{u,v,x} \| \Phi_{uv}(x; Y) \|_{L_2(Q)}^2 \leq \max_{u,v,x} \| \Phi_{uv}(x; Y) \|_\infty^2 \leq Ch^{-1}. \tag{C.8.15}
\]

According to the definition in (C.8.12), \( \Phi_{uv}(z; \cdot) \) is obtained from products and summations of functions with known covering numbers, quantified in Lemmas C.8.4 and C.8.5. By Lemmas C.8.3 and C.8.2 and fixing the bandwidth \( h = \overline{h} = \overline{h} \), \( \sup_Q N(\mathcal{J}_{uv}, \| \cdot \|_{L_2(Q)}, \epsilon) \leq C/(h\epsilon)^{n_1} \) for any \( u, v \in [d] \). Notice that the construction of covering sets in the proofs of Lemmas C.8.4, and C.8.5 is independent to the indices \( j, k \). Therefore, we can construct a set \( \mathcal{N}(2) \subset (0, 1) \) with \( |\mathcal{N}(2)| \leq C/(h\epsilon)^c \) such that for any \( x \in (0,1) \), there exists a \( x_\ell \in \mathcal{N}(2) \) with

\[
\max_{u,v} \| \Phi_{uv}(x; Y) - \Phi_{uv}(x_\ell; Y) \|_{L_2(Q)} \leq \epsilon. \tag{C.8.16}
\]

With this, we construct the covering set for \( \mathcal{J}' \) as

\[
\mathcal{N}(3) = \mathcal{N}(2) \times \{ \ell \epsilon \sqrt{\overline{h}} | \ell = 0, \ldots, [1/(\epsilon \sqrt{\overline{h}})] \}.
\]

For any \((x, z) \in (0,1)^2\), we select \((x_\ell, z_\ell) \in \mathcal{N}(3)\) such that (C.8.16) holds and \(|z - z_\ell| \leq \epsilon \sqrt{\overline{h}}\).

Therefore, by (C.8.13), (C.8.14) and (C.8.15), we have

\[
\| \Omega_j^T(z) \Phi(x; Y) \Omega_k(z) - \Omega_j^T(z_\ell) \Phi(x_\ell; Y) \Omega_k(z_\ell) \|_{L_2(Q)}^2 \leq Cd^3 M^4 \epsilon^2
\]

and \( \sup_Q N(\mathcal{J}', \| \cdot \|_{L_2(Q)}, d^{3/2} M^2 \epsilon) \leq d^2 |\mathcal{N}(3)| = (Cd/(h\epsilon))^c \).

\( \square \)
C.9 Some Useful Results

**Lemma C.9.1.** Let \((Y_1, Y_2, Y_3, Y_4)^T \sim N_4(0, K)\) with \(K = [K_{ab}]_{ab}\). We then have

\[
E [\text{sign}(Y_1 - Y_2) \text{sign}(Y_3 - Y_4)] = \frac{2}{\pi} \arcsin \left( \frac{K_{13} + K_{24} - K_{23} - K_{14}}{\sqrt{(K_{11} + K_{22} - 2K_{12})(K_{33} + K_{44} - 2K_{34})}} \right).
\]

**Proof.** Observe that \((Y_1 - Y_2, Y_3 - Y_4)^T\) is distributed according to a bivariate Gaussian distribution with mean zero and Pearson correlation coefficient

\[
\text{Corr} [Y_1 - Y_2, Y_3 - Y_4] = \frac{K_{13} + K_{24} - K_{23} - K_{14}}{\sqrt{(K_{11} + K_{22} - 2K_{12})(K_{33} + K_{44} - 2K_{34})}}.
\]

The result follows directly from the correspondence between Pearson correlation and Kendall's tau (Fang et al., 1990). \(\square\)

**Corollary C.9.1.** Let \((X_1, Z_1), (X_2, Z_2)\) be independently distributed according to the model in (4.0.1). Then we have

\[
E [\text{sign}(X_{1a} - x_{2a}) \text{sign}(X_{1b} - x_{2b}) \mid Z_1 = z_1, Z_2 = z_2] = \frac{2}{\pi} \arcsin \left( \frac{\Sigma_{ab}(z_1) + \Sigma_{ab}(z_2)}{2} \right).
\]

**Proof.** Follows directly from Lemma C.9.1 by observing that

\[
\text{sign}(X_{1a} - X_{2a}) = \text{sign}(f(X_{1a}) - f(X_{2a})),
\]

since \(f\) is monotone, and using the fact that \(f(X_i)\) follows a Gaussian distribution. \(\square\)

Let \(H : S^2 \mapsto \mathbb{R}\) be a symmetric kernel function. In the setting of our thesis, we have \(S^2 = \mathbb{R}^2 \times (0, 1)\). A kernel is completely degenerate if

\[
E[H(Y_1, Y_2) \mid Y_2] = 0.
\]

211
$U$-statistic based on the kernel $H$ is called degenerate of order 1. See, for example, Serfling (2001).

**Theorem C.9.1** (Theorem 2, Major (2006)). Let $\{Y_i\}_{i \in [n]}$ be independent and identically distributed random variables on a probability space $(S, \mathcal{S}, \mu)$. Let $\mathcal{F}$ be a separable space (with respect to $\mu$) of $S$-measurable $\mu$-degenerate kernel functions that satisfies

$$
N(\epsilon, \mathcal{F}, L_2(\mu)) \leq A\epsilon^{-v}, \quad \text{for all } 1 \geq \epsilon > 0,
$$

where $A$ and $v$ are some fixed constants. Furthermore, we assume that the envelope function is bounded by 1, that is,

$$
\sup_{y_1, y_2} |H(y_1, y_2)| \leq 1, \quad \text{for all } H \in \mathcal{F} \quad \text{(C.9.1)}
$$

$$
\sup_{H \in \mathcal{F}} \mathbb{E} \left[ H^2(Y_1, Y_2) \right] \leq \sigma^2
$$

for some $0 < \sigma \leq 1$. Then there exist constants $C_1, C_2, C_3$ that depend only on $A$ and $v$, such that

$$
\mu \left[ \sup_{H \in \mathcal{F}} \left| \frac{1}{n} \sum_{i \neq i'} H(Y_i, Y_{i'}) \right| \geq t \right] \leq C_1 \exp \left( -C_2 \frac{t}{\sigma} \right) \quad \text{(C.9.2)}
$$

for all $t$ such that

$$
n\sigma^2 \geq \frac{t}{\sigma} \geq C_3 \left( v + \frac{\log A}{\log n} \right)^{3/2} \log \left( \frac{2}{\sigma} \right). \quad \text{(C.9.3)}
$$

**Lemma C.9.2** (Lemma A.1, van de Geer 2008). Let $X_1, \ldots, X_n$ be independent random variables. The $\gamma_1, \ldots, \gamma_d$ be real-valued bounded functions satisfying

$$
\mathbb{E}[\gamma_j(Z_i)] = 0, \text{ for any } i \in [n], j \in [d]; \quad \|\gamma_k\|_\infty \leq \eta_n \quad \text{and} \quad \max_{j \in [d]} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \gamma_j^2(Z_i) \leq \tau_n^2.
$$
We have

\[ \mathbb{E} \left[ \max_{j \in [d]} \left| \frac{1}{n} \sum_{i=1}^{n} \gamma_j(Z_i) \right| \right] \leq \sqrt{\frac{2\tau_n^2 \log(2d)}{n}} + \frac{\eta_n \log(2m)}{n}. \]
References


Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal
frequency distribution in any number of variables. *Biometrika*, pages 134–139.

models using group-sparse regularization. In *International Conference on Artificial
Intelligence and Statistics*, pages 378–387.

Janková, J. and van de Geer, S. (2016). Honest confidence regions and optimality in high-


Jia, Y. and Huan, J. (2010). Constructing non-stationary dynamic bayesian networks with

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applica-


Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, *Proc. of
NIPS*, pages 1732–1740.


Mach. Learn.*, Haifa, Israel.


Lectures from the 38th Probability Summer School held in Saint-Flour, 2008, École d’Été
de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

218


