Dynamics of high-dimensional covariance matrices

Dissertation

zur Erlangung des akademischen Grades
Dr. Rer. Nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Humboldt-Universität zu Berlin von

M.Sc. Valeriy Avanesov

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst
Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Vladimir Spokoiny
2. Dr. Alexandra Carpentier
3. Prof. Dr. Denis Chetverikov

Tag der mündlichen Prüfung: 30. Januar 2018

To my wife

Abstract

We consider the detection and localization of an abrupt break in the covariance structure of high-dimensional random data. The study proposes two novel approaches for this problem. The approaches are essentially hypothesis testing procedures which requires a proper choice of a critical level. In that regard calibration schemes, which are in turn different non-standard bootstrap procedures, are proposed. One of the approaches relies on techniques of inverse covariance matrix estimation, which is motivated by applications in neuroimaging. A limitation of the approach is a sparsity assumption crucial for precision matrix estimation which the second approach does not rely on. The description of the approaches are followed by a formal theoretical study justifying the proposed calibration schemes under mild assumptions and providing the guaranties for the break detection. Theoretical results for the first approach rely on the guaranties for inference of precision matrix procedures. Therefore, we rigorously justify adaptive inference procedures for precision matrices. All the results are obtained in a truly high-dimensional (dimensionality $p \gg n$) finite-sample setting. The theoretical results are supported by simulation studies, most of which are inspired by either real-world neuroimaging or financial data.

Zusammenfassung

Wir betrachten die Detektion und Lokalisation von plötzlichen Änderungen in der Kovarianzstruktur hochdimensionaler zufälliger Daten. Diese Arbeit schlägt zwei neuartige Ansätze für dieses Problem vor. Die Vorgehensweise beinhaltet im Wesentlichen Verfahren zum Test von Hypothesen, welche ihrerseits die Wahl geeigneter kritischer Werte erfordern. Dafür werden Kalibrierungsschemata vorgeschlagen, die auf unterschiedlichen Nichtstandard-Bootstrap-Verfahren beruhen. Der eine der beiden Ansätze verwendet Techniken zum Schätzen inverser Kovarianzmatrizen und ist durch Anwendungen in der neurowissenschaftlichen Bildgebung motiviert. Eine Beschränkung dieses Ansatzes besteht in der für die Schätzung der „Precision matrix“ wesentlichen Voraussetzung ihrer schwachen Besetztheit. Diese Bedingung ist im zweiten Ansatz nicht erforderlich. Die Beschreibung beider Ansätze wird gefolgt durch ihre theoretische Untersuchung, welche unter schwachen Voraussetzungen die vorgeschlagenen Kalibrierungsschemata rechtfertigt und die Detektion von Änderungen der Kovarianzstruktur gewährleistet. Die theoretischen Resultate für den ersten Ansatz basieren auf den Eigenschaften der Verfahren zum Schätzen der Präzisionsmatrix. Wir können daher die adaptiven Schätzverfahren für die Präzisionsmatrix streng rechtfertigen. Alle Resultate beziehen sich auf eine echt hochdimensionale Situation (Dimensionalität $p \gg n$) mit endlichem Stichprobenumfang. Die theoretischen Ergebnisse werden durch Simulationsstudien untermauert, die durch reale Daten aus den Neurowissenschaften oder dem Finanzwesen inspiriert sind.

Acknowledgement

First and foremost I want to thank my advisor Prof. Spokoiny for the immense support, guidance and countless fruitful discussions not to mention that it was a great honour to work under his supervision.

Besides my advisor, I would like to thank the researchers I have worked side by side at the Weierstrass Institute. My special thanks goes to Andzhey Koziuk, Roland Hildebrand, Peter Mathé, Nazar Buzun, Karsten Tabelow, Jörg Polzehl.

The excellent equipment provided by the Weierstrass Institute, especially the large screen and the comfortable chair, has made the struggle physically feasible. For that I express gratitude to the institute's directorate and administration.

Special thanks go to Christine Schneider, the secretary of our group, for her immense help with the paperwork.

Finally, I thank my wife whom the thesis is dedicated to. I thank her for numerous discussions, proofreading and I praise her most of all for giving me the reason.

Contents

1 Introduction **6**

2 Precision matrix inference **9**
2.1 Introduction 9
2.1.1 Contribution 9
2.1.2 Chapter organization 10
2.2 Consistency results 11
2.2.1 Adaptive graphical lasso 11
2.2.2 SCAD graphical lasso 12
2.3 Inference result 14
2.4 Simulation experiments 15
2.4.1 Functional connectivity network from experimental data 15
2.4.2 Software 15
2.4.3 Simulation study 16
2.5 Discussion 25

3 Change point detection based on precision matrix **27**
3.1 Introduction 27
3.2 Proposed approach 28
3.2.1 Definition of the test statistic 28
3.2.2 Bootstrap calibration 29
3.2.3 Change-point localization 30
3.3 Bootstrap validity 31
3.4 Sensitivity result 32
3.5 Simulation study 33
3.5.1 Design 33

4 Change point detection based on covariance matrix **36**
4.1 Introduction 36
4.2 Proposed approach 37
4.2.1 Definition of the test statistics 37
4.2.2 Decision rule and bootstrap calibration scheme 38
4.2.3 Change-point localization 38
4.3 Bootstrap validity 39
4.4 Sensitivity result 41
4.5 Simulation study 42
4.5.1 Real-world covariance matrices 42
4.5.2 Design of the simulation study, results and discussion 42

A Proofs for Chapter 2 **44**
A.1 Proofs of Consistency results . . . 44
A.1.1 Existence and uniqueness of solutions of problems (2.1) and (A.1) . 44
A.1.2 Proof of adaptive lasso consistency result . . . 45
A.1.3 Proof of SCAD graphical lasso consistency result . . . 47
A.2 Proof of the inference result . . . 49

B Proofs for Chapter 3 **52**
B.1 Proof of bootstrap validity result . . . 52
B.2 Proof of sensitivity result . . . 53
B.3 Sandwiching lemma . . . 56
B.4 Similarity of joint distributions of $\{A_n\}_{n\in\mathfrak{N}}$ and $\{A_n^\flat\}_{n\in\mathfrak{N}}$ . . . 58
B.5 Gaussian approximation result for A_n . . . 58
B.6 Gaussian approximation result for $A_n^\flat$ . . . 62
B.7 $\hat{\Sigma}_Y \approx \Sigma_Y^*$ . . . 64

C Proofs for Chapter 4 **70**
C.1 Proof of the sensitivity result . . . 70
C.2 Proof of bootstrap validity result . . . 71
C.3 Similarity of joint distributions of $\{B_n\}_{n\in\mathfrak{N}}$ and $\{B_n^\flat\}_{n\in\mathfrak{N}}$ . . . 71
C.4 Gaussian approximation result for B_n . . . 72
C.5 Gaussian approximation result for $B_n^\flat$ . . . 74
C.6 $\Sigma_Y^* \approx \hat{\Sigma}_Y$ . . . 77

D Known results **79**
D.1 Consistency result for the ℓ_1-penalized estimator by [42] . . . 79
D.2 The bound for $R(\Delta)$ by [42] . . . 79
D.3 The estimation $\hat{\sigma}_{ij}^2$ for σ_{ij}^2 . . . 79
D.4 Probability of the set $\mathcal{T}$ . . . 80
D.5 Gaussian approximation result . . . 81
D.6 Anti-concentration result . . . 81
D.7 Gaussian comparison result . . . 82
D.8 Tail inequality for quadratic forms . . . 82

Chapter 1

Introduction

The analysis of high-dimensional time series is crucial for many fields including neuroimaging and financial engineering. There, one often has to deal with processes involving abrupt structural changes which necessitates a corresponding adaptation of a model and/or a strategy. Structural break analysis comprises determining if an abrupt change is present in the given sample and if so, estimating the change-point, namely the moment in time when it takes place. In literature both problems may be referred to as *change-point* or *break detection*. In this study we will be using terms *break detection* and *change-point localization* respectively in order to distinguish between them. The majority of approaches to the problem consider only a univariate process [17] [1]. However, in recent years the interest for multi-dimensional approaches has increased. Most of them cover the case of fixed dimension [35] [33] [2] [51] [52]. Some approaches [13, 29, 14] feature *high-dimensional* theoretical guaranties but only the case of dimensionality polynomially growing in sample size is covered. The case of exponential growth has not been considered so far.

In order to detect a break, a test statistic is usually computed for each point t (e.g. [35]). The break is detected if the maximum of these values exceeds a certain *threshold*. A proper choice of the latter may be a tricky issue. Consider a pair of plots (Figure 1.1) of the statistic $A(t)$ defined in Section 3.2. It is rather difficult to see how many breaks are there, if any. The classic approach to the problem is based on the asymptotic behaviour of the statistic [17] [1] [2] [29] [8] [52]. As an alternative, permutation [29] [35] or parametric bootstrap may be used [29]. Clearly, it seems attractive to choose the threshold in a solely data-driven way as it is suggested in the recent paper [13], but a careful bootstrap validation is still an open question.

In the current study we are interested in a particular kind of a break – an abrupt transformation in the covariance matrix – which is motivated by applications to neuroimaging and finance. The covariance structure of data in functional Magnetic Resonance Imaging has recently drawn a lot of interest, as it encodes so-called functional connectivity networks [46] which refer to the explicit influence among neural systems [23]. The analysis of the dynamics of these networks is particularly important for the research on neural diseases and also in the context of brain development with emphasis on characterizing the re-configuration of the brain during learning [6].

Analogously, in finance the dynamics of the covariance structure of a high-dimensional process modelling exchange rates and market indexes is crucial for a proper asset allocation in a portfolio [16, 7, 18, 37].

One approach allowing for the change-point localization is developed in [33], the corresponding significance testing problem is considered in [2]. However, neither of these papers address the high-dimensional case.

A widely used break detection approach (named CUSUM) [14, 2, 29] suggests to compute a statistic at a point t as a distance of estimators of some parameter of the underlying distributions obtained using all the data before and after that point. This technique requires the whole sample to be known in advance, which prevents it from being used in *online* setting. In order to overcome this drawback we propose the following augmentation: choose a window size $n \in \mathbb{N}$ and compute parameter estimators using only n points before and n points after the *central point* t (see Section 3.2 for formal definition). Window size n is an important parameter and its choice is case-specific (see Section 3.4 for theoretical treatment of this issue). Using small window results in high variability and low sensitivity, while large window implies higher uncertainty in change-point localization yielding the issue of a proper choice of window size. The *multiscale* nature of the proposed methods enables us to incorporate the advantages of narrower and wider windows by considering multiple window sizes at once in order for wider windows to provide higher sensitivity while narrower ones improve change-point localization.

The contribution of our study is the development of a pair of novel break detection approaches which are

- high-dimensional, allowing for up to exponential growth of the dimensionality with the window size
- suitable for on-line setting
- multiscale, attaining trade-off between break detection sensitivity and change-point localization accuracy
- using a fully data-driven threshold selection algorithm rigorously justified under mild assumptions
- featuring formal sensitivity guaranties in high-dimensional setting

The thesis is comprised of three Chapters. Chapter 3 (based on [5]) establishes theoretical results for pre-existing approaches for estimation and inference of sparse high-dimensional precision matrices. These results are crucial for the construction and theoretical justification of a novel break detection and change-point localization approach which Chapter 2, based on the paper [4], is devoted to. This approach is specifically designed for break detection in a functional connectivity network. A modification of the approach, not relying on precision matrix inference, featuring a wider application range is introduced and analyzed in Chapter 4, which, in turn, is based on the paper [3]. Proofs for Chapters 2, 3 and 4 are collected in Appendices A, B and C. Appendix D lists known results essential for our theoretical study.

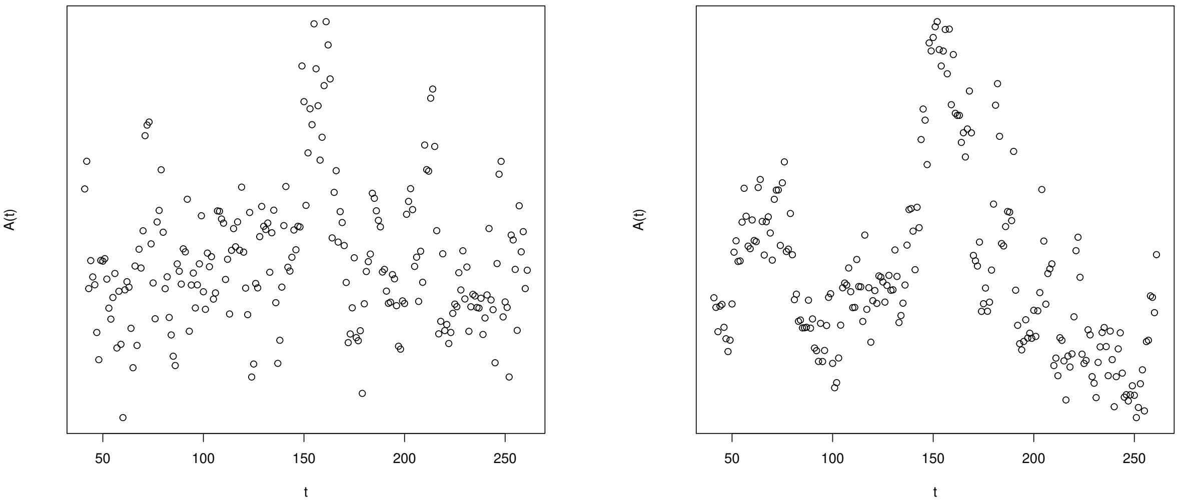


Figure 1.1: Plots of test statstics $A(t)$ computed on synthetically generated data without (left) and with a single change-point at $t = 150$ (right). Clearly, the choice of a threshold is not obvious.

Chapter 2

Precision matrix inference

2.1 Introduction

This chapter is devoted to analysis of adaptive approaches for high-dimensional precision matrix estimation and inference. Specifically, in this chapter we consider an i.i.d. sample $X_1, \ \dots \ X_n \in \mathbb{R}^p$ with zero mean, where n is the length of the fMRI time series. Let X be the $n \times p$ matrix of samples. The FC network is then characterized by the covariance matrix Σ or the precision matrix $\Theta = \Sigma^{-1}$. An estimate of the precision matrix is obtained by minimizing over the cone S^p_{++} of positive-definite $p \times p$ matrices:

$$\arg\min_{\Theta \in S^p_{++}} \left[tr(\Theta\hat{\Sigma}) - \log\det\Theta + p_\lambda(\Theta) \right]$$

with some suitable penalization $p_\lambda(\Theta)$ and the empirical covariance $\hat{\Sigma} = \frac{1}{n} X^T X$.

In order to address the problem ℓ_1-penalization approaches which were initially suggested by [47] may be used imposing the required sparsity on the estimate:

$$\hat{\Theta} = \arg\min_{\Theta \in S^p_{++}} \left[tr(\Theta\hat{\Sigma}) - \log\det\Theta + \|\Lambda * \Theta\|_1 \right] \tag{2.1}$$

where Λ is a $p \times p$ matrix of non-negative off-diagonal elements and zero diagonal elements and $\cdot * \cdot$ denotes matrix element-wise product.

There are consistency results of such estimators for samples of finite size [42] along with asymptotic confidence intervals for the elements of the true precision matrix for the case of equal amount of penalization applied to each element of the precision matrix [27]. On the other hand, there is some experimental evidence that adaptive penalization approaches may perform better [19].

2.1.1 Contribution

In this chapter, we provide consistency results for adaptive ℓ_1-penalized estimators of precision matrices (SCAD graphical lasso [54] [19] [20] and classical adaptive graphical lasso [53] [19]) using the approach by [42]. We also construct confidence intervals based on these estimators for the elements of the true precision matrix following the technique from [27]. We show, that the bias introduced by the penalization and the non-normality of the constructed confidence intervals depends only on the largest amount of penalization applied to non-zero elements of the true precision matrix. In particular, denoting $\theta_{\min}$ the minimal absolute value of non-zero elements of the true precision matrix:

- All the results are obtained in a finite-sample-size setting
- We show improved rates of convergence and non-normality of confidence intervals based on classical adaptive graphical lasso in comparison with graphical lasso for a suitable lower bound for $\theta_{\min}$.
- We demonstrate that the $SCAD$ penalty may not be applicable if $p > n$ since the corresponding optimization problem might be not well-defined
- Nevertheless, we also improve the mentioned rates for $SCAD$ and its one-step approximation in the case $p \leq n$ for a suitable lower bound for $\theta_{\min}$
- We introduce a new $SCAD_\epsilon$ penalty in order to overcome the aforementioned limitation of $SCAD$ for $p > n$. For this penalty as well as for its one-step approximation we provide improved theoretical results in comparison to those for graphical lasso for a suitable lower bound for $\theta_{\min}$
- The obtained rates are never worse than the known rates for non-adaptive graphical lasso for $SCAD$ (if $p \leq n$), $SCAD_\epsilon$ and their one-step approximations
- The aforementioned results are supported by an extensive application-driven simulation study of a functional connectivity network of the human brain based on experimental functional Magnetic Resonance Imaging data

2.1.2 Chapter organization

The chapter is organized as follows. Section 2.1.2 introduces the notation used throughout the chapter. Section 2.2 gives the definition and consistency results for adaptive approaches such as the classical adaptive graphical lasso and SCAD lasso respectively while Section 2.3 comes up with the definition of a de-sparsified estimator and provides the results estimating its distribution which gives rise to confidence intervals construction along with hypotheses testing. Finally, Section 2.4 describes our experimental study. In Section A we provide the proofs of the claimed results.

Notation

We denote the empirical covariance matrix as $\hat{\Sigma} = \frac{1}{n}X^TX$, the true covariance matrix as Σ^* and their difference as $W = \hat{\Sigma} - \Sigma^*$. Throughout the chapter we assume that the true precision matrix ${\Sigma^*}^{-1}$ exists and we denote it as Θ^*.

Also define the set of non-zero entries of Θ^* as $S = \{(i,j) : \Theta^*_{ij} \neq 0\}$ and its complement as $S^c = \{1..p\}^2 \setminus S$.

We use the following notations for matrix norms: $\|A\|_1 = \sum_{i,j} |A_{ij}|$, $\|A\|_\infty = \max_{i,j} |A_{ij}|$ and $|||A|||_\infty = |||A^T|||_1 = \max_j \|A_{\cdot j}\|_1$.

For a matrix A its vectorization is denoted as $\bar{A}$ or, equivalently as $vecA$.

Let $\Gamma^* = \Sigma^* \otimes \Sigma^*$ where $\cdot \otimes \cdot$ stands for Kronecker product, $\kappa_{\Sigma^*} = |||\Sigma^*|||_\infty$, $\kappa_{\Theta^*} = |||\Theta^*|||_1$ and also denote a norm of sub-matrix Γ^*_{SS} as $\kappa_{\Gamma^*} = |||(\Gamma^*_{SS})^{-1}|||_\infty$.

Our main results assume lower bounds on the smallest absolute value of non-zero elements of the true precision matrix which is denoted as $\theta_{min} = \min_{i,j:\Theta_{ij}\neq 0} \left|\Theta^*_{ij}\right|$.

Other values we keep track of are the maximum number of non-zero elements in a row of the true precision matrix $d = \max_i |\{j : \Theta^*_{ij} \neq 0\}|$ and the minimal penalization parameter corresponding to zero elements of the true precision matrix $\rho = \min_{(i,j)\in S^c} \Lambda_{ij}$.

2.2 Consistency results

Assumption 1 (Irrepresentability condition). *Denote an active set*

$$\mathcal{S} \coloneqq \left\{(i, j) \in 1..p \times 1..p : \Theta^*_{ij} \neq 0\right\}$$

and define a $p^2 \times p^2$ matrix $\Gamma^* \coloneqq \Theta^* \otimes \Theta^*$ where $\otimes$ denotes Kronecker product. Irrepresentability condition holds if there exists $\psi \in (0, 1]$ such that

$$\max_{e\notin\mathcal{S}} \left\|\Gamma^*_{e\mathcal{S}}(\Gamma^*_{\mathcal{S}\mathcal{S}})^{-1}\right\|_1 \leq 1 - \psi.$$

The irrepresentability condition is usually interpreted under normality as follows [27] [42]. Define a centered random variable for each edge $(i, j) \in \{1..p\}^2$

$$Y_{(i,j)} = X_{1i}X_{1j} - \mathbb{E}[X_{1i}X_{1j}]$$

then covariances of these variables may be expressed in terms of matrix Γ^* as

$$\mathbf{cov}(Y_{(i,j)}, Y_{(k,l)}) = \Gamma^*_{(i,j),(k,l)} + \Gamma^*_{(j,i),(k,l)}.$$

Assumption 1 requires low correlation between edges from active set S and its complement S_c. The higher the constant ψ is, the stricter upper bound is assumed.

2.2.1 Adaptive graphical lasso

Definition

Let $\hat{\Theta}^{init}$ be a solution of optimization problem (2.1) with penalization parameters $\Lambda_{ij} = \lambda_{init}$ for $i \neq j$.

Then, the adaptive graphical lasso estimator $\hat{\Theta}^{ada}$ is defined as the solution of the optimization problem (2.1) with tuning parameters $\Lambda^{ada}_{ij} = \lambda_{init} \frac{1}{|\hat{\Theta}^{init}_{ij}|^{\gamma}}$ for $i \neq j$ where $\gamma \in (0, 1]$ ($\gamma = 0.5$ is usually used). If $\hat{\Theta}^{init}_{ij} = 0$, we define $\Lambda_{ij} = +\infty$, thereby excluding the corresponding variable from optimization and forcing it to equal zero.

Consistency result

Theorem 1. *Consider a distribution satisfying Assumption 1 with some $\psi \in (0, 1]$. Furthermore, suppose the following sparsity assumption holds for some δ_n:*

$$d \leq \frac{\delta_n}{6\left(\delta_n + \frac{\lambda_n}{(\theta_{min}-r_{init})^{\gamma}}\right)^2 \max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa^2_{\Gamma^*}\kappa^3_{\Sigma^*}\}} \tag{2.2}$$

where $r_{init} \coloneqq 2\kappa_{\Gamma^*}(\delta_n + \lambda_{init})$ and $\lambda_{init} = \frac{8}{\psi}\delta_n$. Also assume that

$$\theta_{min} > r \coloneqq 2\kappa_{\Gamma^*}\left(\delta_n + \frac{\lambda_n}{(\theta_{min} - r_{\lambda})^{\gamma}}\right) \tag{2.3}$$

Then on the set $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$ *the following holds:*
$\left\| \hat{\Theta}^{ada} - \Theta^* \right\|_\infty \leq r$ *and* $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{ada}_{ij} = 0$.

Remark 1. *The main results in the chapter are conditioned on the set* $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$. *The lower bound for the probability of the set* $\mathcal{T}$ *under sub-Gaussianity Assumption D.4.3 is provided by Lemma 41.*

2.2.2 SCAD graphical lasso

Definition

SCAD was suggested in [20] and was applied for sparse precision matrix estimation in [32] as an alternative adaptive penalization approach.

Consider the following optimization problem:

$$\hat{\Theta} = \arg \min_{\Theta \in S^p_{++}} \left[tr(\Theta \hat{\Sigma}) - \log \det \Theta + \sum_{i \neq j} SCAD_{\lambda,a}(|\Theta_{ij}|) \right]$$

for some positive λ and a (usually $a = 3.7$ is used) with the first derivative of $SCAD_{\lambda,a}(\cdot)$ defined as

$$SCAD'_{\lambda,a}(x) = \lambda \left\{ I(x \leq \lambda) + \frac{(a\lambda - x)_+}{(a-1)\lambda} I(x \geq \lambda) \right\}$$

where $(\cdot)_+$ denotes a positive cut: $(x)_+ = \max\{0, x\}$. In order to solve this non-convex optimization problem, the following approximate recurrent algorithm was suggested in [19]

$$\hat{\Theta}^{(k)} = \arg \max_{\Theta \in S^p_{++}} tr(\Theta \hat{\Sigma}) - \log \det \Theta + \sum_{i,j} SCAD'_{\lambda,a}(|\hat{\Theta}^{(k-1)}_{ij}|)|\Theta_{ij}| \tag{2.4}$$

where $\hat{\Theta}^{(0)}$ is obtained as a solution of (2.1) with $\Lambda_{ij} = \lambda \ \forall i \neq j$. Denote the limiting point of the algorithm as $\hat{\Theta}^{SCAD} = \lim_{k \to \infty} \hat{\Theta}^{(k)}$. On the other hand the paper [54] provides asymptotic properties of one-step estimate $\hat{\Theta}^{OSSCAD} = \hat{\Theta}^{(1)}$.

As one can see, $SCAD'_{\lambda,a}(x) = 0$ for x large enough, so the problem (2.4) may have no optimum in case if $\hat{\Sigma}$ is singular. Therefore, in order to establish consistency results for $SCAD$ penalty we have to assume that $\hat{\Sigma}$ is non-singular. However, this rather restrictive assumption may be dropped if we replace $SCAD'_{\lambda,a}(\cdot)$ with

$$SCAD'_{\epsilon,\lambda,a} := \max\{SCAD'_{\lambda,a}(x), \epsilon\}$$

for some positive ϵ. In the same way denote the limiting point as $\hat{\Theta}^{SCAD_\epsilon}$ and one-step estimate as $\hat{\Theta}^{OSSCAD_\epsilon}$.

SCAD graphical lasso consistency results

Theorem 2. *Consider a distribution satisfying Assumption 1 with some* $\psi \in (0, 1]$. *Furthermore, suppose the following sparsity assumption holds for some* δ_n*:*

$$d \leq \frac{\delta_n}{6(\delta_n + \lambda_n)^2 \max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa^2_{\Gamma^*}\kappa^3_{\Sigma^*}\}}$$

and assume that

$$\theta_{min} > r := 2\kappa_{\Gamma^*}(\delta_n + \lambda_n).$$

Also suppose that the matrix $\hat{\Sigma}$ *is non-singular. Then on the set* $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$ *the following holds:*

$$\left\| \hat{\Theta}^{OSSCAD} - \Theta^* \right\|_\infty \leq 2\kappa_{\Gamma^*} \left(\delta_n + SCAD'_{\lambda,a}(\theta_{min} - r) \right)$$

and $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{OSSCAD}_{ij} = 0$.

Theorem 3. *Assume the conditions of Theorem 2. Then on the set* $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$ *the following holds:*

$$\left\| \hat{\Theta}^{SCAD} - \Theta^* \right\|_\infty \leq 2\kappa_{\Gamma^*} \left(\delta_n + \left(\frac{a\lambda_n - \theta_{min} + 2\kappa_{\Gamma^*}\delta_n}{2\kappa_{\Gamma^*} + a - 1} \right)_+ \right)$$

and $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{SCAD}_{ij} = 0$.

The counterparts of these two theorems for $SCAD_\epsilon$ penalty may be proven in a rather similar way.

Theorem 4. *Consider a distribution satisfying Assumption 1 with some* $\psi \in (0, 1]$. *Furthermore, suppose the following sparsity assumption holds for some* δ_n*:*

$$d \leq \frac{\delta_n}{6(\delta_n + \lambda_n)^2 \max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa^2_{\Gamma^*}\kappa^3_{\Sigma^*}\}}$$

Also assume that

$$\theta_{min} > r := 2\kappa_{\Gamma^*}(\delta_n + \lambda_n)$$

Then on the set $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$ *for some* $\epsilon > 0$ *the following holds:*

$$\left\| \hat{\Theta}^{OSSCAD_\epsilon} - \Theta^* \right\|_\infty \leq 2\kappa_{\Gamma^*} \left(\delta_n + SCAD_{\epsilon,\lambda,a}(\theta_{min} - r) \right)$$

and $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{OSSCAD_\epsilon}_{ij} = 0$.

Theorem 5. *Assume the conditions of Theorem 4 holds. Then on the set* $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$ *the following holds:*

$$\left\| \hat{\Theta}^{SCAD_\epsilon} - \Theta^* \right\|_\infty \leq 2\kappa_{\Gamma^*} \left(\delta_n + \left(\frac{a\lambda_n - \theta_{min} + 2\kappa_{\Gamma^*}\delta_n}{2\kappa_{\Gamma^*} + a - 1} \right)_{+,\epsilon} \right)$$

and $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{SCAD_\epsilon}_{ij} = 0$ *where* $(x)_{+,\epsilon} = \max\{x, \epsilon\}$.

2.3 Inference result

In this section we aim to construct confidence intervals for true values of the precision matrix Θ^*_{ij}. In order to do so we mostly follow the approach suggested in [48] and apply it to the problem of estimation of high-dimensional precision matrix introduced in [27]. Consider the stationarity condition corresponding to the problem (2.1):

$$-\hat{\Theta}^{-1} + \hat{\Sigma} + \Lambda * Z = 0$$

where $Z \in \partial \|\Theta\|_1$. Multiply on both sides by $\hat{\Theta}$:

$$\hat{\Theta}\hat{\Sigma}\hat{\Theta} - \hat{\Theta} + \hat{\Theta}(\Lambda * Z)\hat{\Theta} = 0.$$

By rearranging obtain

$$\hat{\Theta} + \hat{\Theta}(\Lambda * Z)\hat{\Theta} = \Theta^* - \Theta^* W \Theta^* + r \tag{2.5}$$

where

$$r = -(\hat{\Theta} - \Theta^*)W\Theta^* - (\hat{\Theta}\hat{\Sigma} - I_p)(\hat{\Theta} - \Theta^*).$$

Finally, we define a de-sparsified estimator as

$$\begin{aligned} \hat{T} &:= 2\hat{\Theta} - \hat{\Theta}\hat{\Sigma}\hat{\Theta} \\ &= \hat{\Theta} + \hat{\Theta}(\Lambda * Z)\hat{\Theta} \\ &= \Theta^* - \Theta^* W \Theta^* + r. \end{aligned} \tag{2.6}$$

Theorem 6. *Consider a distribution satisfying Assumption 1 with some $\alpha \in (0, 1]$, let $\hat{\Theta}$ be the solution of optimization problem (2.1). Suppose also the following restrictions on the penalization parameters Λ hold:*

$$\|\Lambda_S\|_\infty \le \frac{8}{\alpha}\delta_n \text{ and } \rho \ge \frac{8}{\alpha}\delta_n.$$

Furthermore, suppose the following sparsity assumption holds:

$$d \le \frac{\delta_n}{6(\delta_n + \|\Lambda_S\|_\infty)^2 \max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa^2_{\Gamma^*}\kappa^3_{\Sigma^*}\}}.$$

Moreover, suppose, $p_{\mathcal{T}} := \mathbb{P}\{\mathcal{T}\} > 0$. Then, for all (i, j) the following upper and lower bounds hold:

$$\sup_c \left| \mathbb{P}\left\{ \sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij})/\sigma_{ij} \le c \mid \mathcal{T} \right\} - \Phi(c) \right| \le \left(\Phi\left(\frac{R\sqrt{n}}{\sigma_{ij}}\right) - \frac{1}{2} \right) + \frac{A\mu_{ij3}}{\sigma^3_{ij}\sqrt{n}} + 2(1 - p_{\mathcal{T}})$$

where $A < 0.4748$, $\Phi(\cdot)$ is the c.d.f. of a standard normal distribution, $\sigma^2_{ij} = Var[\Theta^*_i X_k \Theta^*_j X_k - \Theta^*_{ij}]$ and μ_{ij3} is the third moment of $|Z_{ijk}|$ (see (A.17)) and R is defined by (A.15).

Remark 2. *The residual term R involved in the statement of Theorem 6 is quite complicated. Here we note that under sub-Gaussianity Assumption D.4.3 for $n, d(n) \to +\infty$ it holds that*

$$R = O\left(\frac{d^2}{n^{3/2}}\right)$$

and hence, R is small ($R = o(1)$), if

$$d = o\left(n^{3/4}\right).$$

Remark 3. *Theorem 6 applies to solutions of* (2.1) *which all of $\hat{\Theta}^{ada}$, $\hat{\Theta}^{SCAD}$ and $\hat{\Theta}^{SCAD_\epsilon}$ are. Moreover, the residual term R involved in the statement of Theorem 6 is smaller for smaller r_Λ which can be reduced using either $\hat{\Theta}^{ada}$, $\hat{\Theta}^{SCAD}$ or $\hat{\Theta}^{SCAD_\epsilon}$ (see results in Section 2.2).*

2.4 Simulation experiments

2.4.1 Functional connectivity network from experimental data

For our experiments we rely on a functional connectivity network that we determined from an experimental fMRI dataset in a recent study [40] that examined learning-dependent plasticity in the human auditory cortex. There, fMRI data with a total of 1680 EPI volumes were acquired with a 3 T Siemens MAGNETOM Trio MRI scanner (Siemens AG, Erlangen, Germany) with an eight-channel head array. We randomly selected a dataset from a single subject. The subject performed a learning experiment with auditory stimuli, number comparison task and reward. The details of the experiment and data acquisition can be found in [40]. We do not repeat them here, because we used the fMRI data only to obtain a realistic network with a natural sparsity pattern for the simulation experiments.

In order to define suitable nodes for the functional connectivity network we used the parcellation atlas defined in a recent study [21] which is available online at the BioImage Suite NITRC page [1]. We normalized the atlas to the motion-corrected functional dataset using SPM12 [2] with standard parameters. Mean time courses of the $p = 256$ regions-of-interest were determined to estimate a functional connectivity network. In order to exclude changes in the network due to the learning effect in the experiment, only the last $n = 300$ time points were used. The network analysis was conducted on the residuals of linear modeling common in fMRI experiments [39]. This way a matrix X^* of size 256×300 of real data was acquired.

2.4.2 Software

All simulations were performed with the R language and environment for statistical computing and graphics [41]. The data and the R script running the simulations are available at `http://www.wias-berlin.de/preprint/2229/codeANDdata_2229.zip`. The following R packages were used: `oro.nifti` [50] was used in order to work with the format the data were stored in, as an implementation of graphical lasso the package `glasso` [22] was used, `igraph` package [15] was used in order to manipulate and visualize graphs, sampling

[1] `https://www.nitrc.org/frs/?group_id=51`
[2] `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`

from multivariate normal distribution was conducted by `MASS` package [49], and an implementation of partial correlation matrix estimator by Pearson's method was borrowed from `ppcor` package [30].

2.4.3 Simulation study

Data generation

In the first step we estimate the precision matrix from the observed mean time courses X^* using thresholded graphical lasso with penalization parameter λ_1 and threshold 0.1 for absolute values of non-diagonal elements of the resulting partial correlation matrix. Let C be the largest connected subgraph in the graph V defined by this partial correlation matrix and X^*_C be the matrix containing the columns of X^* that correspond to nodes within C. We define our ground truth as the network obtained from X^*_C by thresholded graphical lasso with penalization parameter λ_2 and threshold 0.1. It is easy to see that λ_1 controls the size of the ground truth network and λ_2 controls its sparsity.

Simulated data were drawn independently from a Gaussian distribution $\mathcal{N}(0, {\Theta^*}^{-1})$ varying n from 50 to 4500.

In all the experiments involving either adaptive or non-adaptive graphical lasso the penalization parameter was chosen as $\lambda = \sqrt{\frac{\log p}{n}}$ which is an asymptotically optimal choice [27]. In all the experiments one-step SCAD graphical lasso was used as an adaptive approach.

Hypotheses testing

For each non-diagonal element of the precision matrix the null-hypothesis $\mathbb{H}_0^{ij} = \{\Theta^*_{ij} = 0\}$ can be tested against an alternative $\mathbb{H}_1^{ij} = \{\Theta^*_{ij} \neq 0\}$. In order to do so a de-sparsified estimator $\hat{T}_{ij} \rightsquigarrow \mathcal{N}(\Theta^*_{ij}, \sigma^2_{ij})$ was used with σ^2_{ij} replaced by the suitable estimator

$$\hat{\sigma}^2_{ij} := \hat{\Theta}_{ii}\hat{\Theta}_{jj} + \hat{\Theta}^2_{ij}$$

(see also Lemma 40). Finally, Bonferroni-Hochberg multiplicity correction was applied and the power of the test was computed.

In our experiments we compared tests based on the de-sparsified estimator produced by the ℓ_1-penalized estimator, the adaptive estimator and on the classical approach employing Fisher z-transform $z(\cdot)$ on the elements of the partial correlation matrix. The classical approach can be summarized as follows: the partial correlation matrix was estimated with the Pearson method. Fisher z-transform was applied afterwards producing approximately normally distributed values $z_{ij} \rightsquigarrow \mathcal{N}(z(\rho^*_{ij}), n-p-1)$ where ρ^* is the true partial correlation matrix. Clearly, $\rho^*_{ij} = 0$ iff. $\Theta^*_{ij} = 0$, so one can use values z_{ij} as a test statistic.

The powers (the fraction of null-hypotheses $\mathbb{H}_0^{ij}$ rejected for non-zero elements of Θ^*) of these tests were compared.

Confidence intervals

Using the de-sparsified estimator $\hat{T}_{ij}$ approximate $(1-\alpha)100\%$ confidence intervals for the individual values of precision matrix were constructed as

$$I^{\alpha,n}_{ij}(\hat{\Theta}) = [\hat{T}_{ij} - \Phi^{-1}(1-\alpha/2)\sigma_{ij}/\sqrt{n}, \hat{T}_{ij} + \Phi^{-1}(1-\alpha/2)\sigma_{ij}/\sqrt{n}]$$

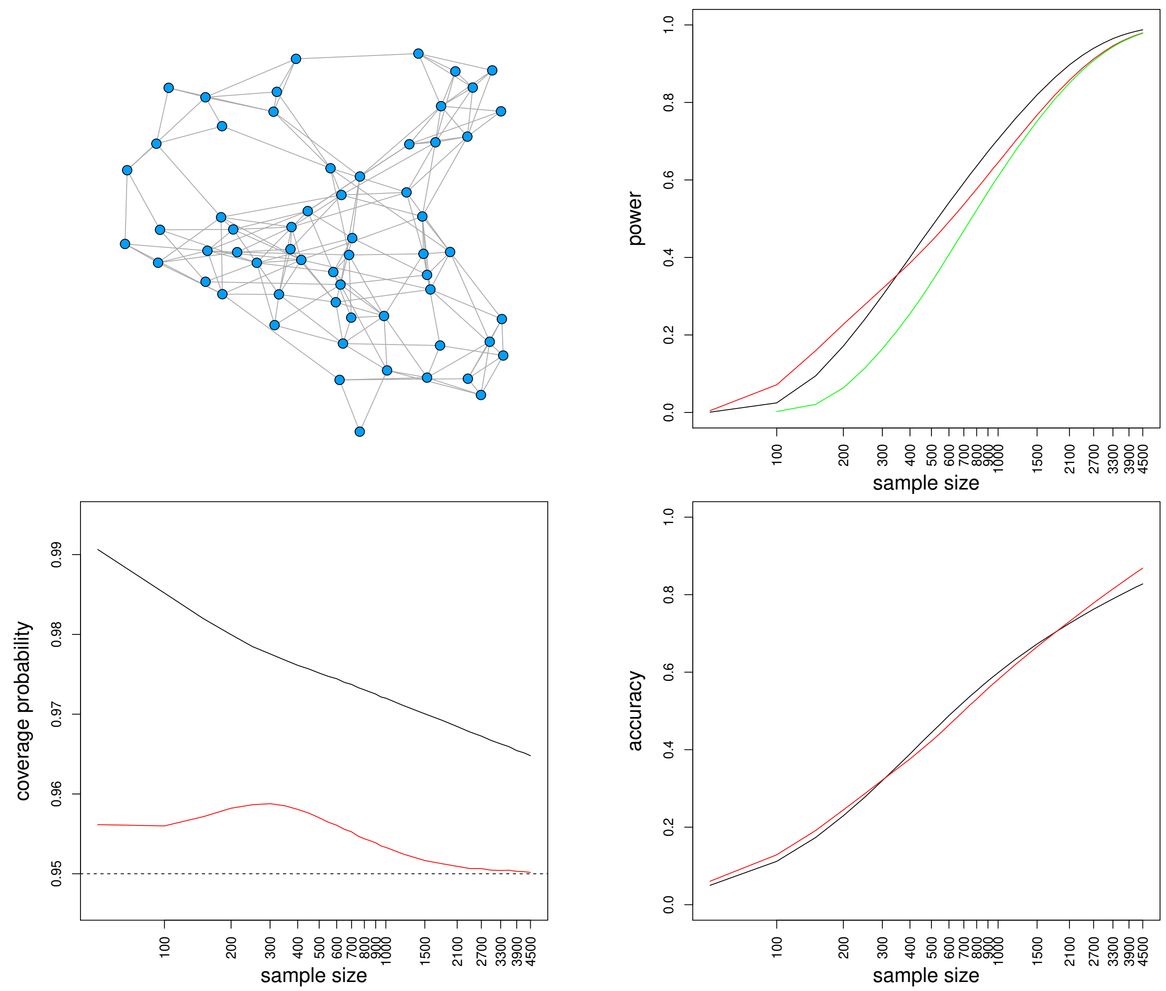


Figure 2.1: Upper-left: Graph obtained with $\lambda_1 = 0.6$, $\lambda_2 = 0.3$, resulting in $p = 60$ and sparsity $= 0.100$. Upper-right: power of hypotheses testing for adaptive (red), non-adaptive (black) and the classical approach (green). Lower-left: coverage probability for adaptive (red) and non-adaptive (black) approach. Lower-right: accuracy of classification between zero and non-zero parameters using adaptive (red) and non-adaptive approach (black)

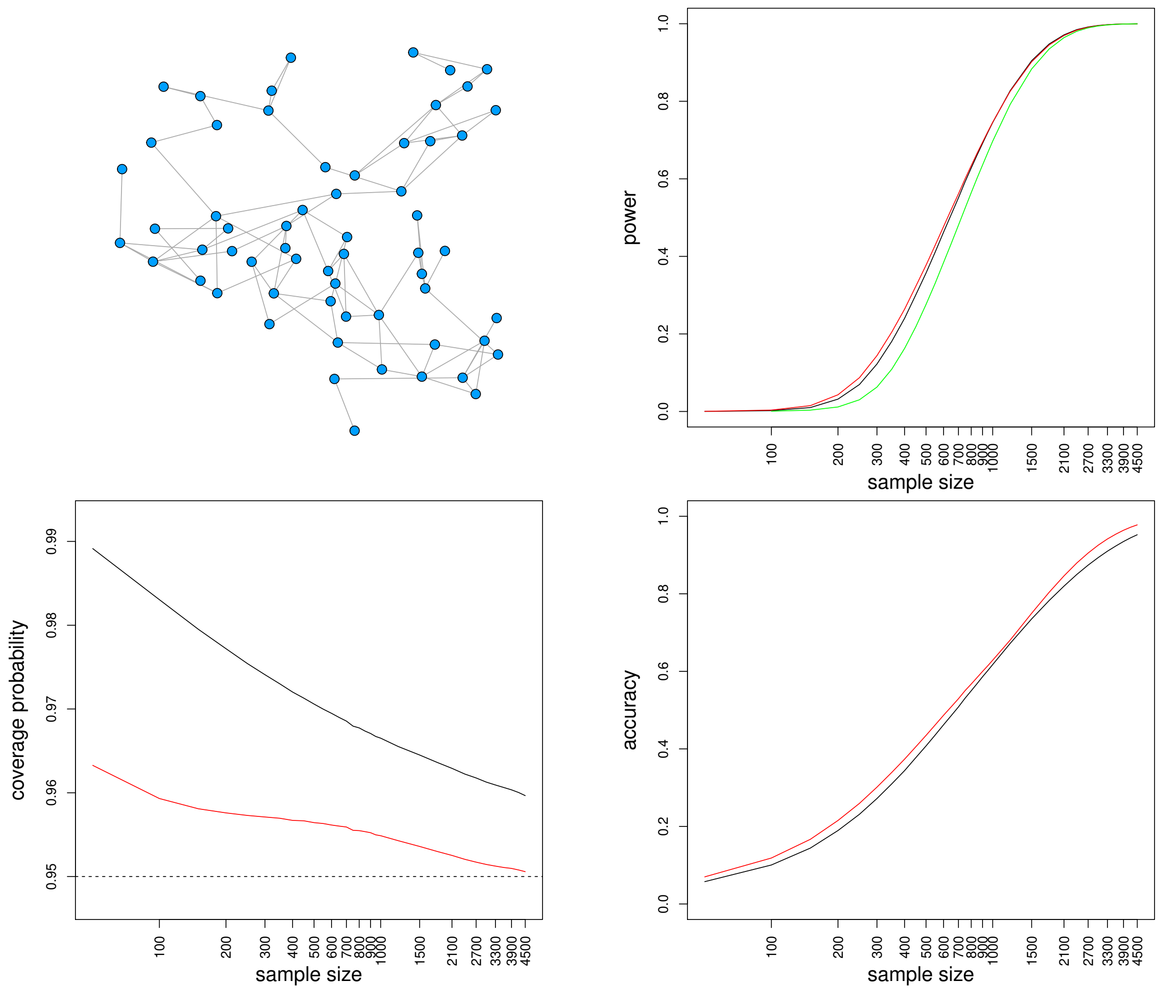


Figure 2.2: As Fig 2.1 but with $\lambda_1 = 0.6$, $\lambda_2 = 0.6$, $p = 60$, $sparsity = 0.050$

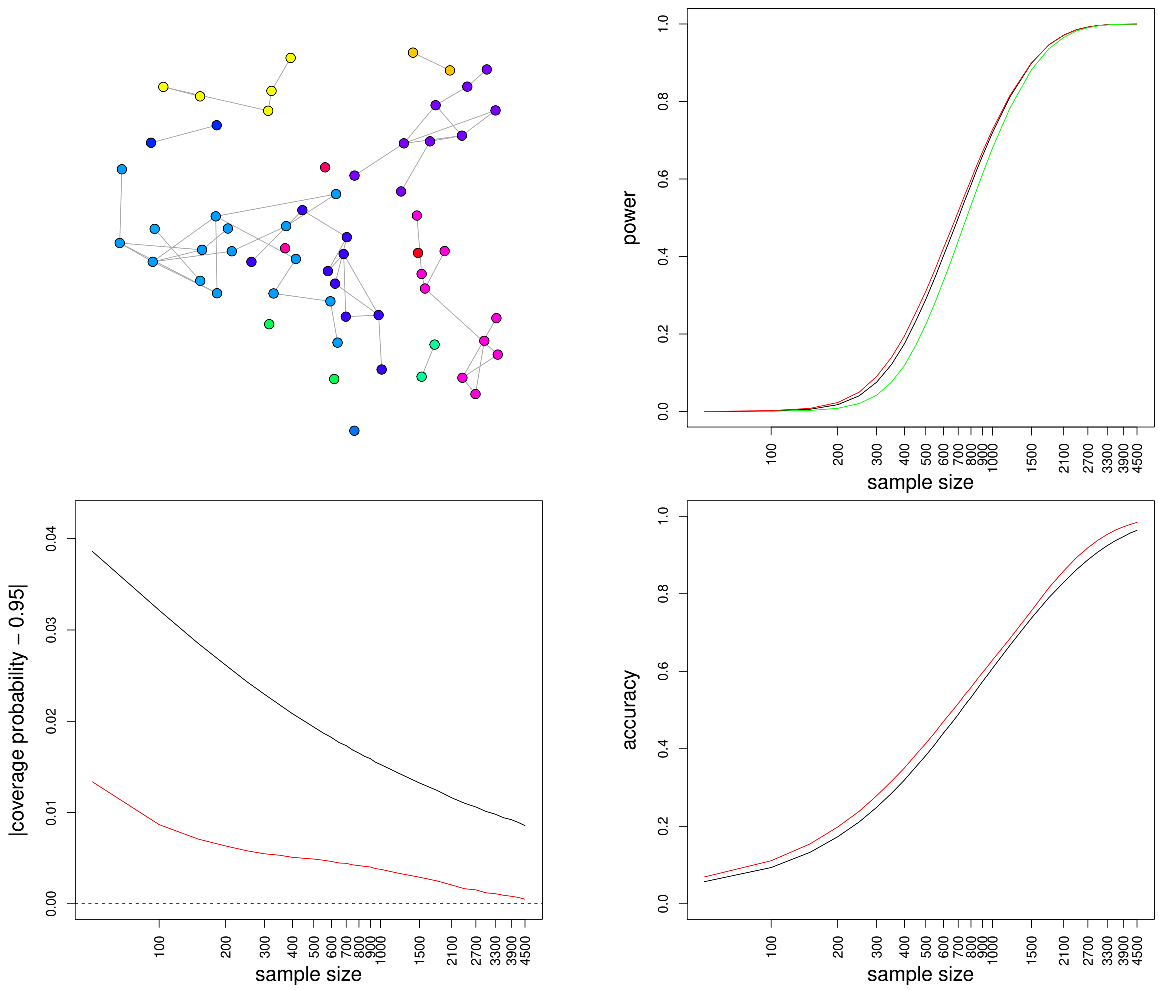


Figure 2.3: As Fig 2.1 but $\lambda_1 = 0.6$, $\lambda_2 = 0.65$, $p = 60$, $sparsity = 0.034$

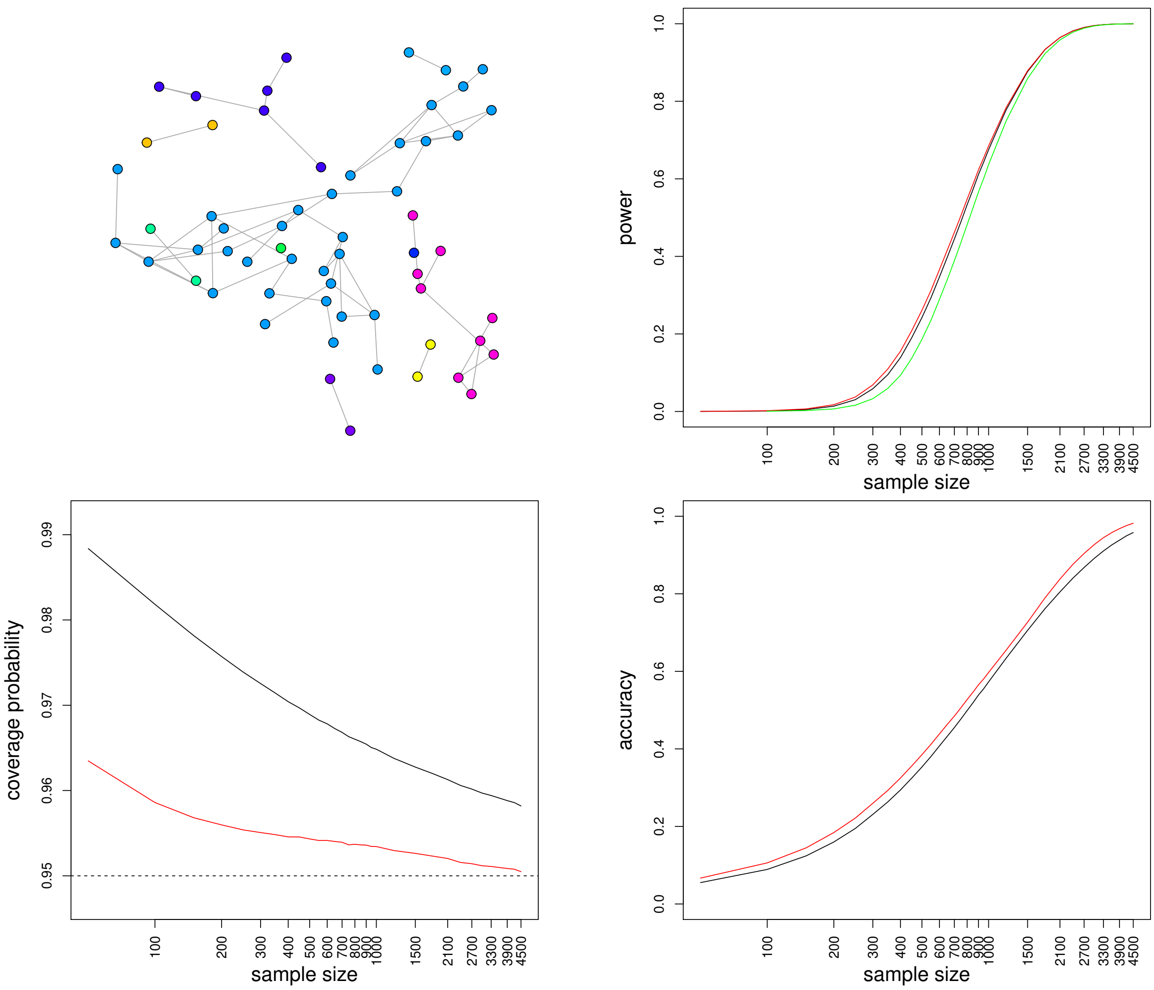


Figure 2.4: As Fig 2.1 but $\lambda_1 = 0.6$, $\lambda_2 = 0.67$, $p = 60$, $sparsity = 0.030$

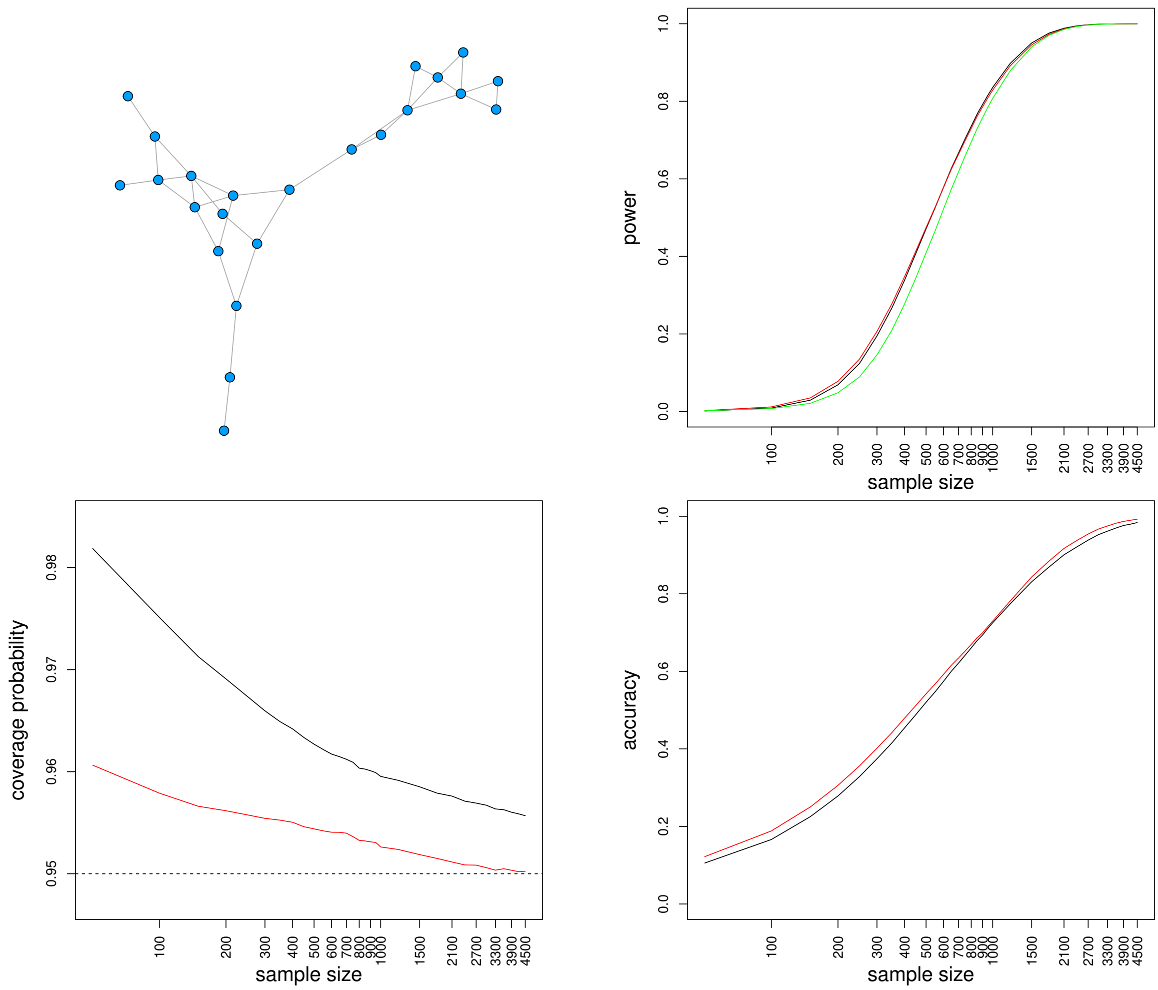


Figure 2.5: As Fig 2.1 but $\lambda_1 = 0.65$, $\lambda_2 = 0.65$, $p = 23$, $sparsity = 0.130$

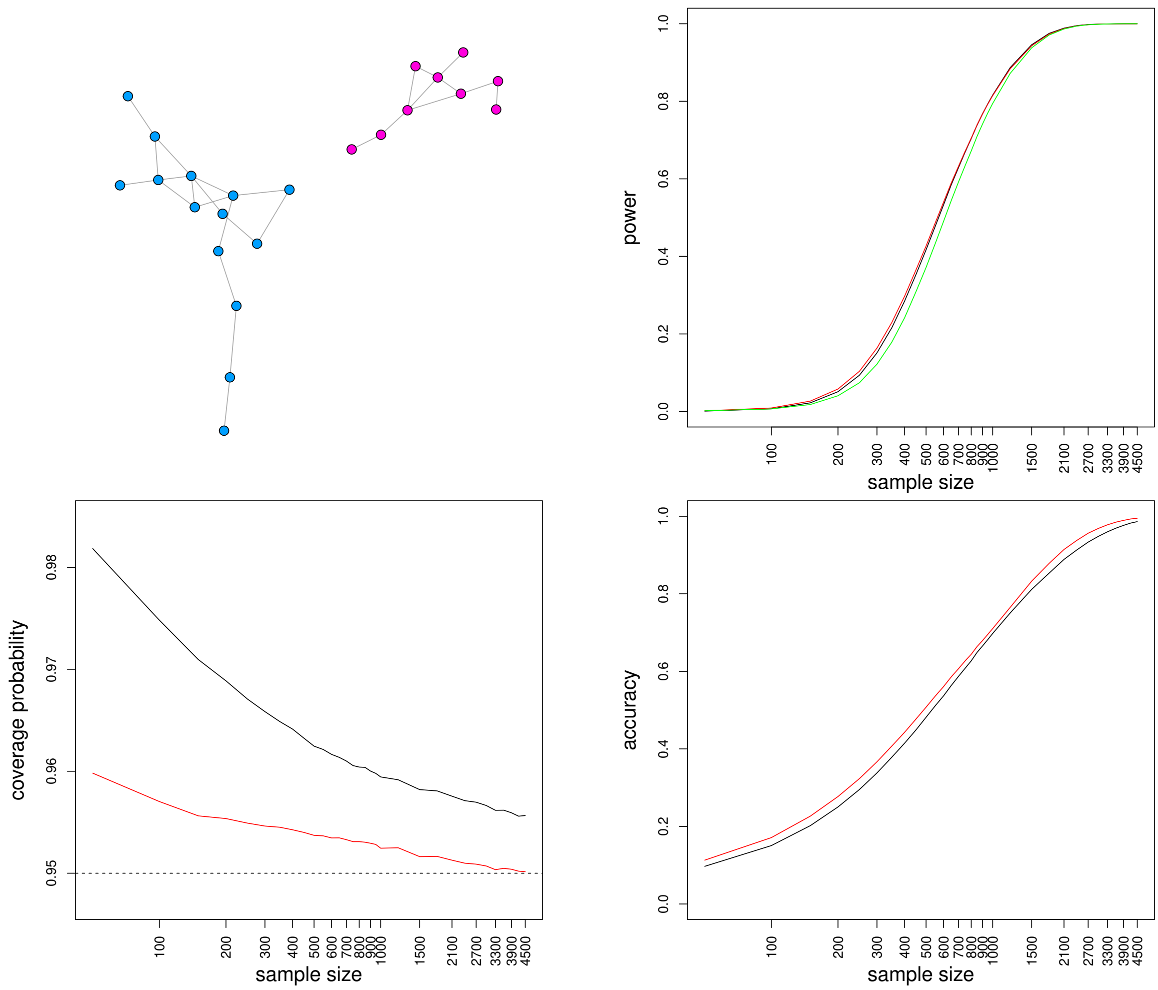


Figure 2.6: As Fig 2.1 but $\lambda_1 = 0.65$, $\lambda_2 = 0.67$, $p = 23$, $sparsity = 0.107$

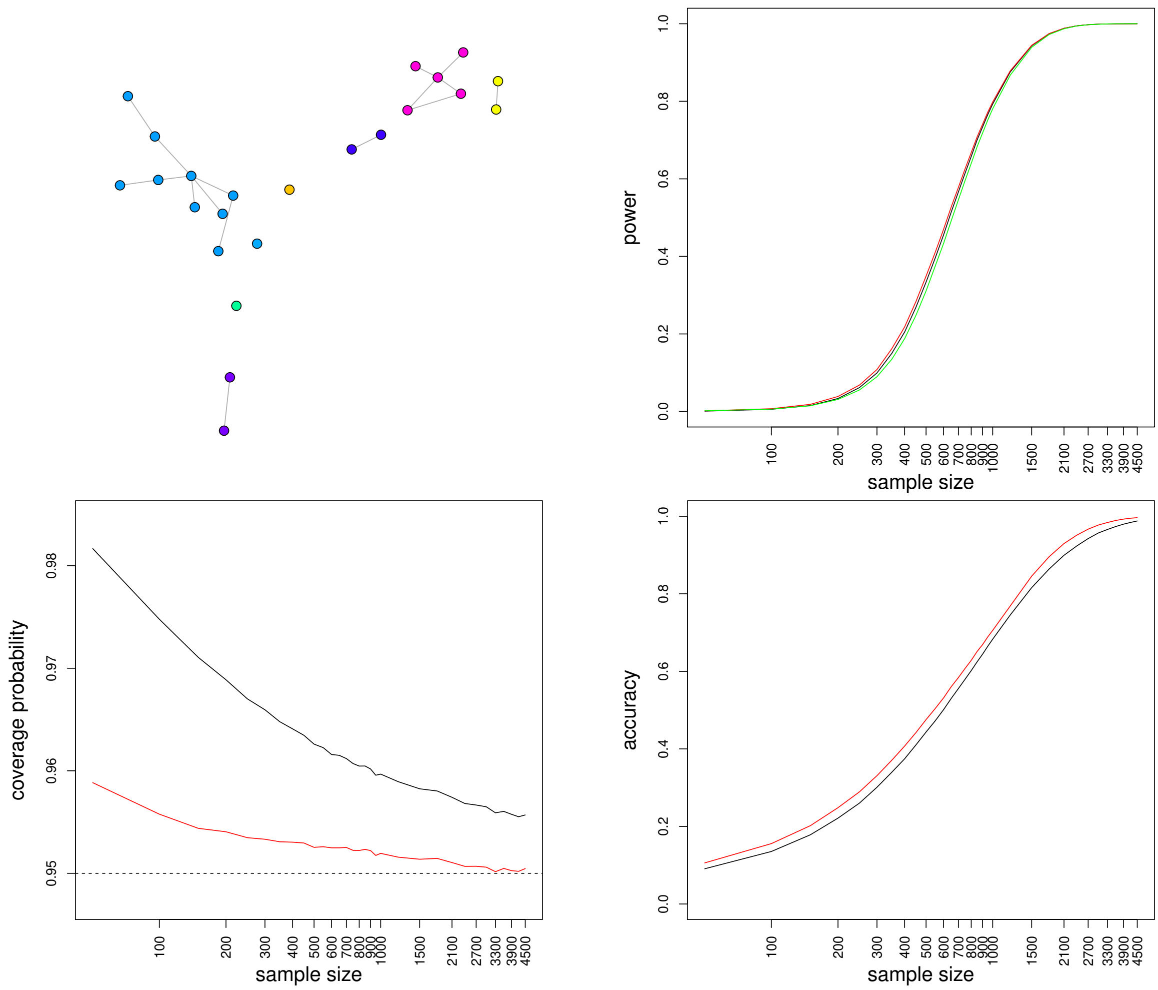


Figure 2.7: As Fig 2.1 but $\lambda_1 = 0.65$, $\lambda_2 = 0.7$, $p = 23$, *sparsity* $= 0.063$

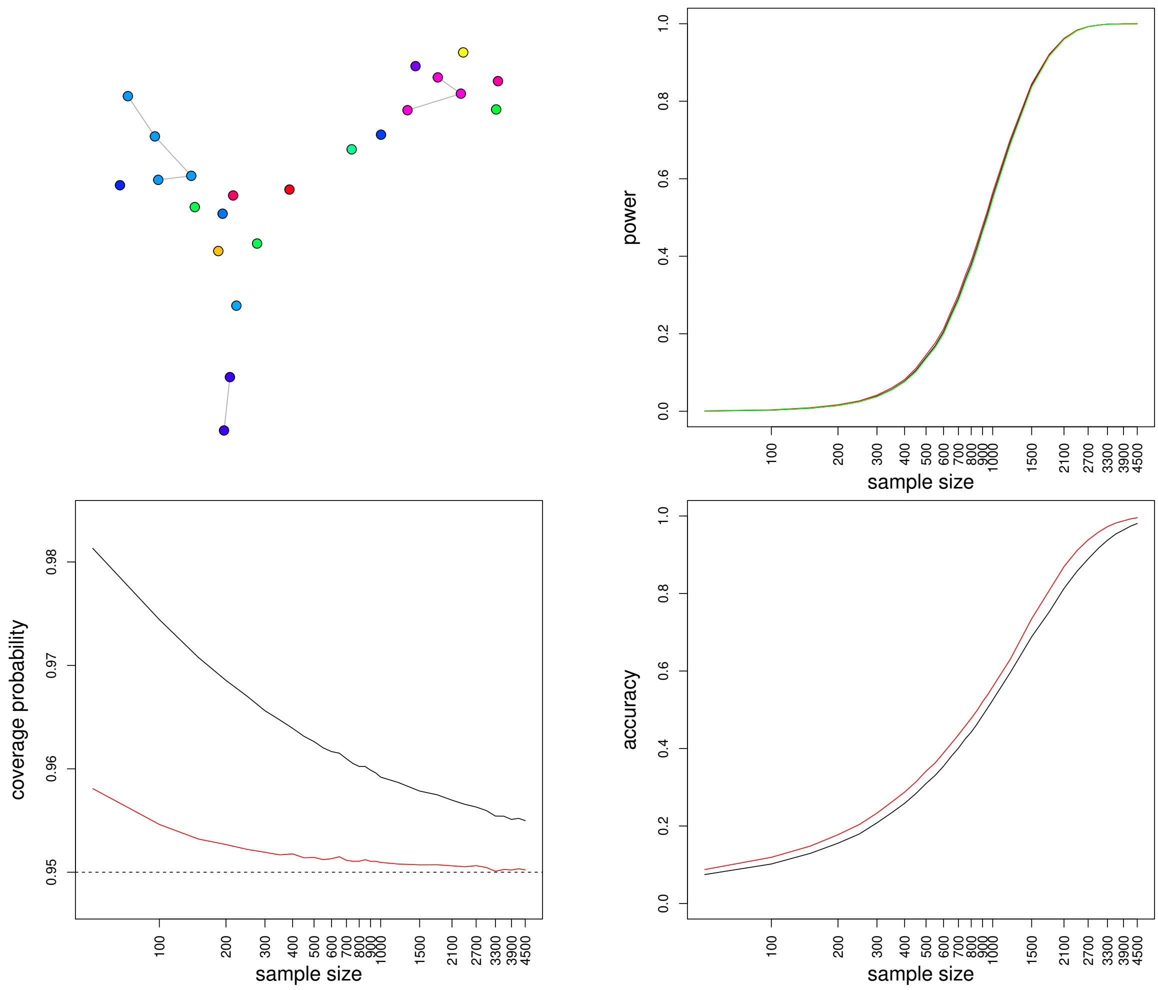


Figure 2.8: As Fig 2.1 but $\lambda_1 = 0.65$, $\lambda_2 = 0.75$, $p = 23$, $sparsity = 0.024$

where $\Phi(\cdot)$ stands for cumulative distribution function of standard normal distribution and $\Phi^{-1}(\cdot)$ denotes its inverse.

In order to compare the approaches we estimated the mean probability for the interval to cover the true value $\frac{1}{p(p-1)}\sum_{i\neq j}\mathbb{P}\{\Theta^*_{ij}\in I^{\alpha,n}_{ij}(\hat{\Theta})\}$ and compared its absolute deviation from $1-\alpha$ for the approaches under comparison.

In our experiments the confidence level was chosen as $\alpha = 0.05$.

Classification between zero and non-zero elements

We compared the ability of the adaptive and non-adaptive approaches to classify between zero and non-zero elements of the precision matrix. The estimated elements of the precision matrix are compared against 0.05 (half of the threshold parameter, see sub-Section 2.4.3) in order to determine whether the correlation coefficient equals zero or not.

The accuracies (probabilities of correct classification between zero and non-zero elements) of such classifiers were compared.

Description of the figures

The graphs obtained in the manner described in Section 2.4.3 along with the results obtained (see Sections 2.4.3–2.4.3 for details) are given in Fig 2.1 – 2.8. The values of the penalization parameters λ_1 and λ_2 used to produce these graphs along with dimensionality p and sparsity (the fraction of non-zero off-diagonal elements of Θ^*) are given in the captions. The upper-left plots represent the ground truth graph. In all these plots each vertex occupies the same spot and disconnected components are shown in different colors. The upper-right plots report the powers of hypotheses testing based on the adaptive, non-adaptive graphical lasso and on a classical approach (see sub-Section 2.4.3). The lower-left plots compare the coverage probabilities of the constructed confidence intervals using the estimator based on the adaptive and non-adaptive estimator (see sub-Section 2.4.3). The lower-right plots represent the comparison of accuracies of classification between zero and non-zero parameters based on the adaptive and non-adaptive approach (see sub-Section 2.4.3). The performance of non-adaptive graphical lasso is shown in black, of the adaptive approach in red and the performance of the classical approach (on the plots reporting the powers of statistical tests) is shown in green.

2.5 Discussion

The experiments showed that the tests based on the classical approach are always outperformed by those based on graphical lasso approaches (apart from the cases where $n \gg p$ where all the approaches perform nearly perfect). At the same time adaptive graphical lasso tends to notably outperform non-adaptive graphical lasso in case of short samples, though sometimes (in case of a denser true precision matrix, see Fig 2.1) non-adaptive approach performs better for sufficiently large samples.

The confidence intervals constructed using the adaptive graphical lasso estimate exhibit coverage probabilities significantly closer to the desired confidence level in comparison to those obtained using non-adaptive graphical lasso estimates.

In experiments on the accuracy of classification between zero and non-zero elements the adaptive approach performs notably better for all sample sizes n apart from the case of a denser precision matrix (see Fig 2.1).

We believe, that superiority of adaptive graphical lasso over non-adaptive graphical lasso is related to the fact that non-adaptive lasso penalizes all the values with the same penalty parameter λ whereas adaptive graphical lasso might reduce penalization of non-zero parameters which leads to the reduction of bias brought in by penalization (compare Theorem 2 and Lemma 38). At the same time, non-normality of the de-sparsified estimator depends on the largest penalization parameter corresponding to a non-zero element $\|\Lambda_S\|_\infty$ (see Theorem 6), which in case of non-adaptive graphical lasso equals λ while in case of an adaptive approach it might be smaller.

Chapter 3

Change point detection based on precision matrix

3.1 Introduction

This chapter presents a novel approach to break detection and change-point localization. The approach is based on precision matrix inference and is designed specifically for break detection in brain function connectivity networks.

Formally, we consider the following setup. Let $X_1, \dots X_N \in \mathbb{R}^p$ denote an independent sample of zero-mean vectors (the on-line setting is discussed in Section 4.3) and we want to test a hypothesis

$$\mathbb{H}_0 := \{\forall i : \mathrm{Var}\left[X_i\right] = \mathrm{Var}\left[X_{i+1}\right]\}$$

versus an alternative suggesting the existence of a break:

$$\mathbb{H}_1 := \{\exists \tau : \mathrm{Var}\left[X_1\right] = \mathrm{Var}\left[X_2\right] = \dots = \mathrm{Var}\left[X_\tau\right] \neq \mathrm{Var}\left[X_{\tau+1}\right] = \dots = \mathrm{Var}\left[X_N\right]\}$$

and localize the change-point τ as precisely as possible or (in online setting) to detect a break as soon as possible.

In the current study it is also assumed that some subset of indices $\mathcal{I}_s \subseteq 1..N$ of size s (possibly, $s = N$) is chosen. The threshold is chosen relying on the sub-sample $\{X_i\}_{i \in \mathcal{I}_s}$ while the test-statistic is computed based on the whole sample.

To this end we define a family of test statistics in Section 3.2.1 which is followed by Section 3.2.2 describing a data-driven (bootstrap) calibration scheme and Section 3.2.3 describing change-point localization procedure. The theoretical part of the chapter justifies the proposed procedure in a high-dimensional setting. The result justifying the validity of the proposed calibration scheme is stated in Section 3.3. Section 3.4 is devoted to the sensitivity result yielding a bound for the window size n necessary to reliably detect a break of a given extent and hence bounding the uncertainty of the change-point localization (or the delay of detection in online setting). The theoretical study is supported by a comparative simulation study (described in Section 3.5) demonstrating conservativeness of the proposed test and higher sensitivity compared to a recent algorithm. Appendix B.1 provides a a finite-sample version of bootstrap sensitivity result which is followed by the proofs. Appendix B.2 contains a finite-sample version of sensitivity result along with the proofs.

3.2 Proposed approach

This section describes the proposed approach along with a data-driven calibration scheme. Informally the proposed statistic can be described as follows. Provided that the break may happen only at moment t, one could estimate some parameter of the distribution using n data-points to the left of t, estimate it again using n data-points to the right and use the norm of their difference as a test-statistic $A_n(t)$. Yet, in practice one does not usually possess such knowledge, therefore we propose to maximize these statistics over all possible locations t yielding A_n. Finally, in order to attain a trade-off between break detection sensitivity and change-point localization accuracy we build a multiscale approach: consider a family of test statistics $\{A_n\}_{n \in \mathfrak{N}}$ for multiple window sizes $n \in \mathfrak{N} \subset \mathbb{N}$ at once.

3.2.1 Definition of the test statistic

Now we present a formal definition of the test statistic. In order to detect a break we consider a set of window sizes $\mathfrak{N} \subset \mathbb{N}$. Denote the size of the widest window as n_+ and of the narrowest as n_-. Given a sample of length N, for each window size $n \in \mathfrak{N}$ define a set of central points $\mathbb{T}_n := \{n+1, ..., N-n+1\}$. Next, for all $n \in \mathfrak{N}$ define a set of indices which belong to the window on the left side of the central point $t \in \mathbb{T}_n$ as $\mathcal{I}_n^l(t) := \{t-n, ..., t-1\}$ and correspondingly for the window on the right side define $\mathcal{I}_n^r(t) := \{t, ..., t+n-1\}$. Denote the sum of numbers of central points for all window sizes $n \in \mathfrak{N}$ as

$$T := \sum_{n \in \mathfrak{N}} |\mathbb{T}_n| .$$

For each window size $n \in \mathfrak{N}$, each central point $t \in \mathbb{T}_n$ and each side $\mathfrak{S} \in \{l, r\}$ we define a de-sparsified estimator of precision matrix [27] [28] as

$$\hat{T}_n^{\mathfrak{S}}(t) := \hat{\Theta}_n^{\mathfrak{S}}(t) + \hat{\Theta}_n^{\mathfrak{S}}(t)^T - \hat{\Theta}_n^{\mathfrak{S}}(t)^T \hat{\Sigma}_n^{\mathfrak{S}}(t) \hat{\Theta}_n^{\mathfrak{S}}(t)$$

where

$$\hat{\Sigma}_n^{\mathfrak{S}}(t) = \frac{1}{n} \sum_{i \in \mathcal{I}_n^{\mathfrak{S}}(t)} X_i X_i^T$$

and $\hat{\Theta}_n^{\mathfrak{S}}(t)$ is a consistent estimator of precision matrix which can be obtained by graphical lasso [43] or node-wise procedure [28] (see Definition 1 for details).

Now define a matrix of size $p \times p$ with elements

$$Z_{i,uv} := \Theta_u^* X_i \Theta_v^* X_i - \Theta_{uv}^* \tag{3.1}$$

where $\Theta^* := \mathbb{E}\left[X_i X_i^T\right]^{-1}$ for $i \leq \tau$, Θ_u^* stands for the u-th row of Θ^*. Denote their variances as $\sigma_{uv}^2 := \mathrm{Var}\left[Z_{1,uv}\right]$ and introduce the diagonal matrix $S = diag(\sigma_{1,1}, \sigma_{1,2} ... \sigma_{p,p-1}, \sigma_{p,p})$. Denote a consistent estimator (see Definition 1 for details) of the precision matrix Θ^* obtained based on the sub-sample $\{X_i\}_{i \in \mathcal{I}_s}$ as $\hat{\Theta}$, where $\mathcal{I}_s \subset 1..N$. In practice, the variances σ_{uv}^2 are unknown, but under normality assumption one can plug in $\hat{\sigma}_{uv}^2 := \hat{\Theta}_{uu}\hat{\Theta}_{vv} + \hat{\Theta}_{uv}^2$ which have been proven to be consistent (uniformly for all u and v) estimators of σ_{uv}^2 [27] [5]. If the node-wise procedure is employed, the uniform consistency of an empirical estimate of σ^2_{uv} has been shown under some mild assumptions (not including normality) [28].

For each window size $n \in \mathfrak{N}$ and a central point $t \in \mathbb{T}_n$ we define a statistic

$$A_n(t) := \left\| \sqrt{\frac{n}{2}} S^{-1} \overline{(\hat{T}^l_n(t) - \hat{T}^r_n(t))} \right\|_\infty \tag{3.2}$$

where we write $\overline{M}$ for a vector composed of stacked columns of matrix M. Finally we define our family of test statistics for all $n \in \mathfrak{N}$ as

$$A_n = \max_{t \in \mathbb{T}_n} A_n(t).$$

Our approach heavily relies on the following expansion under $\mathbb{H}_0$

$$\sqrt{n}(\hat{T}^{\mathfrak{S}}_n(t) - \Theta^*) = \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}^{\mathfrak{S}}_n(t)} Z_i + r^{\mathfrak{S}}_n(t)\sqrt{n}, \tag{3.3}$$

where the residual term

$$r^{\mathfrak{S}}_n(t) := \hat{T}^{\mathfrak{S}}_n(t) - \left(\Theta^* - \Theta^* \left(\hat{\Sigma}^{\mathfrak{S}}_n(t) - \Sigma^*\right) \Theta^*\right)$$

can be controlled under mild assumptions [27] [28] [5].

This expansion might have been used in order to investigate the asymptotic properties of A_n and obtain the threshold, however we propose a data-driven scheme.

Remark 4. *A different test statistic $A_n(t)$ can be defined as the maximum distance between elements of empirical covariance matrices $\hat{\Sigma}(t)^l_n$ and $\hat{\Sigma}(t)^r_n$. However, application to neuroimaging motivates the search for a structural change in a functional connectivity network which is encoded by the structure of the corresponding precision matrix. Clearly, a change in the precision matrix also means a change in the covariance matrix, though we believe that the definition* (4.2) *increases the sensitivity to this kind of alternative – compare the definitions of break extent in sensitivity results Theorem 8 and Theorem 10.*

3.2.2 Bootstrap calibration

Our approach rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ if at least one of statistics A_n exceeds the corresponding threshold $x^\flat_n(\alpha)$ or formally if $\exists n \in \mathfrak{N} : A_n > x^\flat_n(\alpha)$.

In order to properly choose the thresholds, we define bootstrap statistics $A^\flat_n$ in the following non-standard way. Note, that we cannot use an ordinary scheme with replacement or weighted bootstrap since in a high-dimensional case ($|\mathcal{I}_s| \leq p$) the covariance matrix of bootstrap distribution would be singular which would make inverse covariance matrix estimation procedures meaningless.

First, draw with replacement a sequence $\{\varkappa_i\}_{i=1}^N$ of indices from $\mathcal{I}_s$ and denote

$$X^\flat_i = X_{\varkappa_i} - \mathbb{E}_{\mathcal{I}_s}[X_j]$$

where $\mathbb{E}_{\mathcal{I}_s}[\cdot]$ stands for averaging over values for indexes belonging to $\mathcal{I}_s$ e.g., $\mathbb{E}_{\mathcal{I}_s}[X_j] = \frac{1}{|\mathcal{I}_s|}\sum_{j \in \mathcal{I}_s} X_j$. Denote the measure $X^\flat_i$ are distributed with respect to as $\mathbb{P}^\flat$. In accordance with (A.17) define

$$Z^{\flat}_{i,uv} := \hat{\Theta}_u X^{\flat}_i \hat{\Theta}_v X^{\flat}_i - \hat{\Theta}_{uv}$$

and for technical purposes define

$$\hat{Z}_{i,uv} := \hat{\Theta}_u X_i \hat{\Theta}_v X_i - \hat{\Theta}_{uv}.$$

Now for all central points t define a bootstrap counterpart of $A_n(t)$

$$A^{\flat}_n(t) := \left\| \frac{1}{\sqrt{2n}} S^{-1} \overline{\left(\sum_{i \in \mathcal{I}^l_n(t)} Z^{\flat}_i - \sum_{i \in \mathcal{I}^r_n(t)} Z^{\flat}_i \right)} \right\|_{\infty} \tag{3.4}$$

which is intuitively reasonable due to expansion (3.3). And finally we define the bootstrap counterpart of A_n as

$$A^{\flat}_n = \max_{t \in \mathbb{T}_n} A^{\flat}_n(t).$$

Now for each given $\mathrm{x} \in [0, 1]$ we can define quantile functions $z^{\flat}_n(\mathrm{x})$ such that

$$z^{\flat}_n(\mathrm{x}) := \inf \left\{ z : \mathbb{P}^{\flat} \left\{ A^{\flat}_n > z \right\} \le \mathrm{x} \right\}.$$

Next for a given significance level α we apply multiplicity correction choosing α^* as

$$\alpha^* := \sup \left\{ \mathrm{x} : \mathbb{P}^{\flat} \left\{ \exists n \in \mathfrak{N} : A^{\flat}_n > z^{\flat}_n(\mathrm{x}) \right\} \le \alpha \right\}$$

and finally choose thresholds as $x^{\flat}_n(\alpha) := z^{\flat}_n(\alpha^*)$.

Remark 5. *One can choose $\mathcal{I}_s = 1, 2, ..., N$ and use the whole given sample for calibration as well as for detection. In fact, it would improve the bounds in Theorem 7 and Theorem 8, since it effectively means $s = N$. However, in practise such a decision might lead to reduction of sensitivity due to overestimation of the thresholds.*

3.2.3 Change-point localization

In order to localize a change-point we have to assume that $\mathcal{I}_s \subseteq 1..\tau$. Consider the narrowest window detecting a change-point:

$$\hat{n} := \min \left\{ n \in \mathfrak{N} : A_n > x^{\flat}_n(\alpha) \right\} \tag{3.5}$$

and denote the central point where this window detects a break for the first time as

$$\hat{\tau} := \min \left\{ t \in \mathbb{T}_{\hat{n}} : A_{\hat{n}}(t) > x^{\flat}_{\hat{n}}(\alpha) \right\}.$$

By construction of the family of test statistics we conclude (up to the confidence level α) that the change-point τ is localized in the interval

$$[\hat{\tau} - \hat{n}; \hat{\tau} + \hat{n} - 1].$$

Clearly, if a non-multiscale version of the approach is employed, i.e. $|\mathfrak{N}| = \{n\}$, $n = \hat{n}$ and the precision of localization (delay of the detection in online setting) equals n.

3.3 Bootstrap validity

This section states and discusses the theoretical result demonstrating the validity of the proposed bootstrap scheme i.e.

$$\mathbb{P}\left\{\forall n \in \mathfrak{N} : A_n \leq x_n^{\flat}(\alpha)\right\} \approx 1 - \alpha.$$

Our theoretical results require the tails of the underlying distributions to be light. Specifically, we impose Sub-Gaussianity vector condition.

Assumption 2 (Sub-Gaussianity vector condition)**.**

$$\exists L : \forall i \in 1..N \sup_{\substack{a \in \mathbb{R}^p \\ ||a||_2 \leq 1}} \mathbb{E}\left[\exp\left(\left(\frac{a^T X_i}{L}\right)^2\right)\right] \leq 2.$$

Naturally, in order to establish a theoretical result we have to assume that a method featuring theoretical guaranties was used for estimating the precision matrices. Such methods include graphical lasso [43], adaptive graphical lasso [53] and thresholded de-sparsified estimator based on node-wise procedure [28]. These approaches overcome the high dimensionality of the problem by imposing a sparsity assumption, specifically bounding the maximum number of non-zero elements in a row: $d := \max_i \left|\{j|\Theta^*_{ij} \neq 0\}\right|$. These approaches are guaranteed to yield a root-n consistent estimate revealing the sparsity pattern of the precision matrix [43, 5, 28] or formally

Definition 1. *Consider an i.i.d. sample $x_1, x_2, ...x_n \in \mathbb{R}^p$. Denote their precision matrix as $\Theta^* = \mathbb{E}\left[x_1\right]^{-1}$. Let p and d grow with n. A positive-definite matrix $\hat{\Theta}^n$ is a consistent estimator of the high-dimensional precision matrix if*

$$\left\|\Theta^* - \hat{\Theta}^n\right\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right)$$

and

$$\forall i, j \in 1..p \text{ and } \Theta^*_{ij} = 0 \Rightarrow \hat{\Theta}^n_{ij} = 0.$$

Assumption 3.3.A. *Suppose, either graphical lasso or its adaptive version was used with regularization parameter $\lambda_n \asymp \sqrt{\log p/n}$ and also impose Assumption 1.*

Assumption 3.3.B. *Suppose, thresholded de-sparsified estimator based on node-wise procedure was used with regularization parameter $\lambda_n \asymp \sqrt{\log p/n}$.*

Now we are ready to establish a result which guarantees that the suggested bootstrap procedure yields proper thresholds.

Theorem 7. *Assume $\mathbb{H}_0$ holds and furthermore, let $X_1, X_2, ...X_N \in \mathbb{R}^p$ be i.i.d. Let Assumption 2 and Assumption 3.3.A hold. Also assume, the spectrum of Θ^* is bounded. Allow the parameters $d, s, p, |\mathfrak{N}|, n_-, n_+$ grow with N. Further let $N > 2n_+$, $n_+ \geq n_-$ and also impose the sparsity assumption*

$$d = o\left(\frac{\sqrt{\max\{s, n_-\}}}{\log p}\right).$$

Then

$$\left|\mathbb{P}\left\{\forall n \in \mathfrak{N} : A_n \le x_n^\flat(\alpha)\right\} - (1-\alpha)\right|$$
$$= O\Bigg(|\mathfrak{N}| \left\{ \left(\frac{d \log^7(pN)}{n_-}\right)^{1/6} \left(\log^2(ps) + d^{1/6}\log p\right) + \left(\frac{d^4}{s}\right)^{1/6} \log^{2/3}(pN) \right\}\Bigg).$$

The finite-sample version of this result, namely, Theorem 11, is given in Appendix B.1 along with the proofs.

Bootstrap validity result discussion Theorem 7 guarantees under mild assumptions (Assumption 1 seems to be the most restrictive one, yet it may be dropped if the node-wise procedure is employed) that the first-type error rate meets the nominal level α if the narrowest window size n_- and the set $\mathcal{I}_s$ are large enough. Clearly, the dependence on dimensionality p is logarithmic which establishes applicability of the approach in a high-dimensional setting. It is worth noticing that, unusually, the sparsity bound gets stricter with N but the dependence is only logarithmic. Indeed, we gain nothing from longer samples, since we use only $2n$ data points each time.

On-line setting As one can easily see, the theoretical result is stated in off-line setting, when the whole sample of size N is acquired in advance. In on-line setting we suggest to control the probability α to raise a false alarm for at least one central point t among N data points (which differs from the classical techniques controlling the mean distance between false alarms [44]). Having α and N chosen one should acquire s data-points (the set $\mathcal{I}_s$), use the proposed bootstrap scheme with bootstrap samples of length N in order to obtain the thresholds. Next the approach can be naturally applied in on-line setting and Theorem 7 guarantees the capability of the proposed bootstrap scheme to control the aforementioned probability to raise a false alarm.

Proofs The proof of the bootstrap validity result, presented in Appendix B.1, mostly relies on the high-dimensional central limit theorems obtained in [12], [11]. These papers also present bootstrap justification results, yet do not include a comprehensive bootstrap validity result. The theoretical treatment is complicated by the randomness of $x_n^\flat(\alpha)$. We overcome it by applying the so-called "sandwiching" proof technique (see Lemma 10), initially used in [45] and extended by [9]. The authors of [45] had to assume normality and low dimensionality of the data, while in [9] only continuous probability measures $\mathbb{P}$ and $\mathbb{P}^\flat$ were considered. Our result is free of such limitations.

3.4 Sensitivity result

Consider the following setting. Let there be index τ, such that $\{X_i\}_{i\le\tau}$ are i.i.d. and $\{X_i\}_{i>\tau}$ are i.i.d. as well. Denote precision matrices $\Theta_1{}^{-1} := \mathbb{E}\left[X_1 X_1^T\right]$ and $\Theta_2{}^{-1} := \mathbb{E}\left[X_{\tau+1} X_{\tau+1}^T\right]$. Define the break extent Δ as

$$\Delta := \left\|\Theta_1 - \Theta_2\right\|_\infty.$$

The question is, how large the window size n should be in order to reliably reject $\mathbb{H}_0$ and how firmly can we localize the change-point.

Theorem 8. *Let Assumption 2 and either Assumption 3.3.A or Assumption 3.3.B hold. Also assume, the spectrums of* Θ_1 *and* Θ_2 *are bounded. Allow the parameters* $d, s, p, |\mathfrak{N}|, n_-, n_+$ *grow with* N *and let* Δ *decay with* N*. Further let* $N > 2n_+$, $n_+ \geq n_-$,

$$d = o\left(\frac{\sqrt{\max\{s, n_-\}}}{d \log^7(pN)}\right) \tag{3.6}$$

and

$$\frac{\log^2(pN)}{n_+\Delta} = o(1). \tag{3.7}$$

Then $\mathbb{H}_0$ *will be rejected with probability approaching* 1.

This result is a direct corollary of the finite-sample sensitivity result established and discussed in Appendix B.2.

The assumption $\mathcal{I}_s \subseteq 1..\tau$ is only technical. The result may be proven without relying on it by methodologically the same argument.

Sensitivity result discussion Assumptions (3.6) and (3.7) are essentially a sparsity bound and a bound for the largest window size n_+. Clearly, they do not yield a particular value n_+ necessary to detect a break, since it depends on the underlying distributions, however, the result includes dimensionality p only under the sign of logarithm, which guarantees high sensitivity of the test in high-dimensional setting.

Online setting Theorem 8 is established in offline setting as well. In online setting it guarantees that the proposed approach can reliably detect a break of an extent not less than Δ with a delay at most n_+ bounded by (3.7).

Change-point localization guaranties Theorem 8 implies by construction of statistic A_n that the change-point can be localized with precision up to n_+. Hence condition (3.7) provides the bound for change-point localization accuracy.

3.5 Simulation study

3.5.1 Design

In our simulation we test

$$\mathbb{H}_0 = \left\{\{X_i\}_{i=1}^N \sim \mathcal{N}(0, I)\right\}$$

versus an alternative

$$\mathbb{H}_1 = \left\{\exists \tau : \{X_i\}_{i=1}^{\tau} \sim \mathcal{N}(0, I) \text{ and } \{X_i\}_{i=\tau+1}^N \sim \mathcal{N}(0, \Sigma_1)\right\}$$

The alternative covariance matrix Σ_1 was generated in the following way. First we draw $k \sim Poiss(3)$. The matrix Σ_1 is composed as a block-diagonal matrix of k matrices of size 2×2 with ones on their diagonals and their off-diagonal element drawn uniformly from $[-0.6; -0.3] \cup [0.3; 0.6]$ and an identity matrix of size $(p-2k) \times (p-2k)$. The dimensionality

Table 3.1: First type error rate, power and precision of change-point localization of the proposed approach for various sets of window sizes $\mathfrak{N}$

$\mathfrak{N}$	I type error rate	Power	Localization precision
{70}	0.02	0.09	70
{100}	0.00	0.37	100
{140}	0.01	0.81	140
{70, 140}	0.01	0.76	135
{100, 140}	0.01	0.75	124
{70, 100, 140}	0.01	0.74	123

of the problem is chosen as $p = 50$, the length of the sample $N = 1000$ and we choose the set $\mathcal{I}_s = [1, 2, ..100]$. The absence of positive effect of large sample size N is discussed in Sections 3.3 and 3.4. Moreover, in all the simulations under alternative the sample was generated with the change point in the middle: $\tau = N/2$ but the algorithm was oblivious about this as well as about either of the covariance matrices. The significance level $\alpha = 0.05$ was chosen. In all the experiments graphical lasso with penalization parameter $\lambda_n = \sqrt{\frac{\log p}{n}}$ was used in order to obtain $\hat{\Theta}_n^{\mathfrak{S}}(t)$. In the same way, graphical lasso with penalization parameter λ_s was used in order to obtain $\hat{\Theta}$.

We have also come up with an approach to the same problem not involving bootstrap. The paper [34] defines a high-dimensional two-sample test for equality of matrices. Moreover, the authors prove asymptotic normality of their statistic which makes computing p-value possible. We suggest to run this test for every $t \in \mathbb{T}_n$ and every $n \in \mathfrak{N}$, adjust the obtained p-values using Holm method [24] and eventually compare them against α.

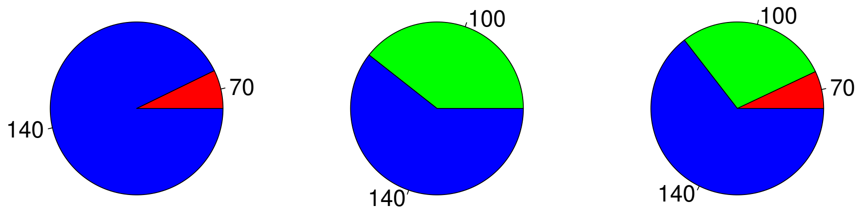
The paper [35] suggests an approach based on comparing characteristic functions of random variables. The critical values were chosen with permutation test as proposed by the authors. In our experiments the method was allowed to consider all the sample at once. The R-package ecp [26] was used.

The first type error rate and power for our approach are reported in Table 3.1. As one can see, our approach allows to properly control first type error rate. As expected, its power is higher for larger windows and it is decreased by adding narrower windows into consideration which is the price to be paid for better localization of a change point.

In our study the approach proposed in [35] and the one based on the two sample test [34] turned out to be conservative, but neither of them exhibited power above 0.1.

Also, in order to justify application of the multiscale approach (i. e. $|\mathfrak{N}| > 1$) for the sake of better change-point localization we report the distribution of the narrowest detecting window $\hat{n}$ (defined by (3.5)) over $\mathfrak{N}$ in Figure 3.1. The Table 3.1 represents average precision of change-point localization for various choices of set of window sizes $\mathfrak{N}$. One can see, that the multiscale approach significantly improves the precision of localization.

Figure 3.1: Pie charts representing distribution of narrowest detecting window $\hat{n}$ *and the precision of localization in cases of* $|\mathfrak{N}| = \{70, 140\}$, $|\mathfrak{N}| = \{100, 140\}$ *and* $|\mathfrak{N}| = \{70, 100, 140\}$ *respectively*



Chapter 4

Change point detection based on covariance matrix

4.1 Introduction

This chapter presents a novel approach to break detection and change-point localization. In contrast to the approach suggested in Chapter 2 it does not impose any sparsity assumption, which makes it naturally applicable in the field of finance, yet it does not focus on precision matrices and hence might be less suitable for applications in neuroimaging.

Formally, we consider the following setup. Let $X_1, ...X_N \in \mathbb{R}^p$ denote an independent sample of zero-mean vectors (the on-line setting is discussed in Section 4.3) and we want to test a hypothesis

$$\mathbb{H}_0 \coloneqq \{\forall i : \mathrm{Var}\,[X_i] = \mathrm{Var}\,[X_{i+1}]\}$$

versus an alternative suggesting the existence of a break:

$$\mathbb{H}_1 \coloneqq \{\exists \tau : \mathrm{Var}\,[X_1] = \mathrm{Var}\,[X_2] = ... = \mathrm{Var}\,[X_\tau] \neq \mathrm{Var}\,[X_{\tau+1}] = ... = \mathrm{Var}\,[X_N]\}$$

and localize the change-point τ as precisely as possible or (in online setting) to detect a break as soon as possible.

In the current study it is also assumed that some subset of indices $\mathcal{I}_s \subseteq 1..N$ of size s (possibly, $s = N$) is chosen. The threshold is chosen relying on the sub-sample $\{X_i\}_{i \in \mathcal{I}_s}$ while the test-statistic is computed based on the whole sample.

To this end we define a family of test statistics in Section 4.2.1 which is followed by Section 4.2.2 describing a data-driven (bootstrap) calibration scheme and Section 4.2.3 proposing a change-point localization procedure. Section 4.3 presents and discusses a theoretical result justifying the bootstrap scheme while Section 4.4 presents a sensitivity result providing a lower bound for a window size n necessary to detect a break of a given extent and hence bounding the uncertainty of the change-point localization (or the delay of detection in online setting). Finally, Section 4.5 presents a simulation study inspired by real-world financial data supporting the theoretical findings and demonstrating superiority of our approach to a recent one.

4.2 Proposed approach

The first part of this Section formally defines the test statistics while the second part concentrates on the calibration scheme. Informally, the test statistics may be defined as follows. Provided that the break may happen only at point t, one could estimate the covariance matrix using n data-points to the left of t, estimate it again using n data-points to the right of it and use the norm of their difference as a test-statistic $B_n(t)$. Yet, in practice one does not usually possess such knowledge, therefore we propose to maximize these statistics over all possible locations t yielding B_n. Finally, in order to attain a trade-off between break detection sensitivity and change-point localization we propose a multiscale approach considering multiple window sizes $n \in \mathfrak{N}$ and multiple respective test statistics $\{B_n\}_{n \in \mathfrak{N}}$ at once.

4.2.1 Definition of the test statistics

Now we present a formal definition of the test statistics. In order to detect a break we consider a set of window sizes $\mathfrak{N} \subset \mathbb{N}$. Denote the size of the widest window as n_+ and of the narrowest as n_-. Given a sample of length N for each window size $n \in \mathfrak{N}$ define a set of central points $\mathbb{T}_n := \{n+1, n+2, ..., N-n+1\}$. Next, for all $n \in \mathfrak{N}$ define a set of indices which belong to the window on the left side from the central point $t \in \mathbb{T}_n$ as $\mathcal{I}_n^l(t) := \{t-n, t-n+1, ..., t-1\}$ and correspondingly $\mathcal{I}_n^r(t) := \{t, t+1, ..., t+n-1\}$. Denote the sum of numbers of central points for all window sizes $n \in \mathfrak{N}$ as

$$T := \sum_{n \in \mathfrak{N}} |\mathbb{T}_n| . \tag{4.1}$$

For each window size $n \in \mathfrak{N}$ and each central point $t \in \mathbb{T}_n$ define a pair of estimators of covariance matrix as

$$\hat{\Sigma}_n^l(t) := \frac{1}{n} \sum_{i \in \mathcal{I}_n^l(t)} X_i X_i^T \text{ and } \hat{\Sigma}_n^r(t) := \frac{1}{n} \sum_{i \in \mathcal{I}_n^r(t)} X_i X_i^T .$$

Let some subset of indices $\mathcal{I}_s \subseteq 1..N$ of size s (possibly, $s = N$) be chosen. Define a scaling diagonal matrix

$$S = diag(\sigma_{1,1}, \sigma_{1,2} ... \sigma_{p,p-1}, \sigma_{p,p})$$

where the elements $\sigma_{j,k}$ are standard deviations of corresponding elements of $X_i X_i^T$ averaged over $\mathcal{I}_s$:

$$\sigma_{j,k}^2 := \frac{1}{s} \sum_{i \in \mathcal{I}_s} \text{Var}\left[(X_i X_i^T)_{jk}\right] .$$

In practice the matrix S is usually unknown, hence we propose to plug-in empirical estimators $\hat{\sigma}_{j,k}$.

For each window size $n \in \mathfrak{N}$ and central point $t \in \mathbb{T}_n$ we define a test statistic $B_n(t)$

$$B_n(t) := \left\| \sqrt{\frac{n}{2}} S^{-1} \overline{(\hat{\Sigma}_n^l(t) - \hat{\Sigma}_n^r(t))} \right\|_\infty . \tag{4.2}$$

Here and below we write $\overline{A}$ for a vector composed of stacked columns of matrix A and use $\|\cdot\|_\infty$ to denote sup norm. Finally the family of test statistics $\{B_n\}_{n \in \mathfrak{N}}$ is obtained via maximization over the central points:

$$B_n := \max_{t \in \mathbb{T}_n} B_n(t).$$

4.2.2 Decision rule and bootstrap calibration scheme

Our approach rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ if at least one of statistics B_n exceeds a corresponding threshold $x_n^\flat(\alpha)$ or formally if $\exists n \in \mathfrak{N} : B_n > x_n^\flat(\alpha)$.

In order to choose thresholds $x_n^\flat(\alpha)$ the following bootstrap scheme is proposed. Define vectors $\hat{Z}_i$ for $i \in \mathcal{I}_s$ as

$$\hat{Z}_i := \overline{X_i X_i^T - \frac{1}{s} \sum_{i \in \mathcal{I}_s} X_i X_i^T}.$$

Elements $Z_i^\flat$ for $i \in 1..N$ of bootstrap sample are proposed to be drawn with replacements from the set $\bigcup_{i \in \mathcal{I}_s} \{\hat{Z}_i, -\hat{Z}_i\}$. Denote the measure which $Z_i^\flat$ are distributed with respect to as $\mathbb{P}^\flat$.

Now we are ready to define a bootstrap counterpart $B_n^\flat(t)$ of $B_n(t)$ for all $n \in \mathfrak{N}$ and $t \in \mathbb{T}_n$ as

$$B_n^\flat(t) := \left\| \frac{1}{\sqrt{2n}} S^{-1} \left(\sum_{i \in \mathcal{I}_n^l(t)} Z_i^\flat - \sum_{i \in \mathcal{I}_n^r(t)} Z_i^\flat \right) \right\|_\infty . \tag{4.3}$$

The counterparts $B_n^\flat$ of B_n for all $n \in \mathfrak{N}$ are naturally defined as

$$B_n^\flat := \max_{t \in \mathbb{T}_n} B_n^\flat(t).$$

Now for each given $\mathrm{x} \in (0, 1)$ we can define quantile functions $z_n^\flat(\mathrm{x})$ such that

$$z_n^\flat(\mathrm{x}) := \inf \left\{ z : \mathbb{P}^\flat \left\{ B_n^\flat > z \right\} \le \mathrm{x} \right\}.$$

Next for a given significance level α we apply multiplicity correction choosing α^* as

$$\alpha^* := \sup \left\{ \mathrm{x} : \mathbb{P}^\flat \left\{ \exists n \in \mathfrak{N} : B_n^\flat > z_n^\flat(\mathrm{x}) \right\} \le \alpha \right\}$$

and finally choose thresholds as $x_n^\flat(\alpha) := z_n^\flat(\alpha^*)$.

Remark 6. *In most of the cases one may simply choose $\mathcal{I}_s = 1...N$ but at the same time it seems appealing to use some sub-sample which a priori does not include a break, if such information is available. On the other hand, the bootstrap justification result (Theorem 9) and sensitivity result (Theorem 10) benefit from larger set $\mathcal{I}_s$. The experimental comparison of these options is given in Section 4.5.*

4.2.3 Change-point localization

In order to localize a change-point we have to assume that $\mathcal{I}_s \subseteq 1..\tau$. Consider the narrowest window detecting a change-point as $\hat{n}$:

$$\hat{n} := \min \left\{ n \in \mathfrak{N} : B_n > x_n^\flat(\alpha) \right\}$$

and the central point where this window detects a break for the first time as

$$\hat{\tau} := \min \left\{ t \in \mathbb{T}_{\hat{n}} : B_{\hat{n}}(t) > x_{\hat{n}}^\flat(\alpha) \right\}.$$

By construction of the family of the test statistics we conclude (up to the confidence level α) that the change-point τ is localized in the interval

$$[\hat{\tau} - \hat{n}; \hat{\tau} + \hat{n} - 1].$$

Clearly, if a non-multiscale version of the approach is employed, i.e. $|\mathfrak{N}| = \{n\}$, $n = \hat{n}$ and precision of localization (delay of the detection in online setting) equals n.

4.3 Bootstrap validity

This section states and discusses the theoretical result demonstrating validity of the proposed bootstrap scheme i.e.

$$\mathbb{P}\left\{\forall n \in \mathfrak{N} : B_n \leq x_n^\flat(\alpha)\right\} \approx 1 - \alpha. \tag{4.4}$$

Our theoretical results require the tails of the underlying distributions to be light. Specifically, we impose Sub-Gaussianity vector condition Assumption 2.

Theorem 9. *Let Assumption 2 hold and let $X_1, X_2, ..., X_N$ be i.i.d. Moreover, assume that the residual $R < \alpha/2$ where*

$$R := (3 + 2|\mathfrak{N}|)\left(R_B + R_{B^\flat} + R_\Sigma^\pm\right),$$

$$R_\Sigma^\pm := C\Delta_Y^{1/3} \log^{2/3}\left(Tp^2\right),$$

Δ_Y, R_B and $R_{B^\flat}$ are defined in Lemma 36, Lemma 28 and Lemma 32 respectively and C is an independent positive constant. Then for all positive x, t and χ it holds that

$$\left|\mathbb{P}\left\{\forall n \in \mathfrak{N} : B_n \leq x_n^\flat(\alpha)\right\} - (1 - \alpha)\right| \leq R + 2(1 - q),$$

where

$$q := 1 - p_{\mathcal{Z}_s}(\kappa) - p_s^\Omega(t, \mathrm{x}) - p_s^{\mathcal{W}}(\mathrm{x}) - p^\Sigma(\chi), \tag{4.5}$$

probabilities $p_{\mathcal{Z}_s}(\kappa)$, $p_s^\Omega(t, \mathrm{x})$, $p_s^{\mathcal{W}}(\mathrm{x})$ and $p^\Sigma(\chi)$ come from Lemma 31, Lemma 35, Lemma 33 and Lemma 20 respectively and quantiles $\{x_n^\flat(\alpha)\}_{n \in \mathfrak{N}}$ are yielded by bootstrap procedure described in Section 4.2.2.

Proof sketch The proof consists of four straightforward steps.

1. Approximate statistics B_n by norms of a high-dimensional Gaussian vector up to the residual R_B using the high dimensional central limits theorem by [12].

2. Similarly, approximate bootstrap counterparts $B_n^\flat$ of the statistics up to the residual $R_{B^\flat}$.

3. Prove that the covariance matrix of Gaussian vector used to approximate $B_n^\flat$ in step 2 is concentrated in the ball of radius Δ_Y centered at its real-world counterpart involved in the step 1 and employ the Gaussian comparison result provided by [12] and [11].

4. Finally, obtain the bootstrap validity result combining the results of steps 1-3.

The formal treatment for each of these steps is given in Appendices C.4, C.5, C.6 and in Lemma 10 respectively.

Proof discussion The proof of the bootstrap validity result mostly relies on the high-dimensional central limit theorems obtained by [12]. That paper also presents bootstrap justification results, yet does not include a comprehensive bootstrap validity statement. The theoretical treatment is complicated by the randomness of $x_n^\flat(\alpha)$. Indeed, consider Lemma 27 which is a straightforward combination of steps 1-3. One cannot trivially obtain result of sort (4.4) substituting $\{x_n^\flat(\alpha)\}_{n\in\mathfrak{N}}$ in (C.1) due to randomness of $x_n^\flat(\alpha)$ and dependence between $x_n^\flat(\alpha)$ and B_n. We overcome this by means of so-called "sandwiching" proof technique (see Lemma 10), initially used by [45]. The authors had to assume normality and low dimensionality of the data. Our result is free of such a limitations.

Bootstrap validity result discussion The remainder terms R_B, $R_{B^\flat}$ and $R_\Sigma^\pm$ involved in the statement of Theorem 9 are rather complicated. Here we just note that for p, s, N, n_-, $n_+ \to +\infty$, $N > 2n_+$, $n_+ \geq n_-$

$$R_B \leq C_1 \left(\frac{L^4 \log^7 (p^2 T n_+)}{n_-} \right)^{1/6},$$

$$R_{B^\flat} \leq C_2 \left(\frac{L^4 \log^7 (p^2 T n_+)}{n_-} \right)^{1/6} \log^2(ps), \tag{4.6}$$

$$R_\Sigma^\pm \leq C_3 \left(\frac{L^4 \log^4 (ps)}{s} \right)^{1/6} \log^{2/3} \left(p^2 T\right),$$

while the parameters $\kappa, \mathrm{x}, \chi, t$ are chosen in order to ensure the probability q defined by (4.5) to be above 0.995, e.g.

$$\mathrm{x} = 7.61 + \log(ps), \tag{4.7}$$

$$\kappa = 6.91 + \log s, \tag{4.8}$$

$$t = 7.61 + 2\log p, \tag{4.9}$$

$$\chi = 6.91.$$

Here C_1, C_2, C_3 are some positive constants independent of $N, \mathfrak{N}, p, s, L$. In fact, probability q can be made arbitrarily close to 1 at the cost of worse constants.

It is worth noticing that, unusually, remainder terms R_B, $R_{B^\flat}$ and $R_\Sigma^\pm$ grow with T defined by (4.1) and hence with the sample size N but the dependence is logarithmic. Really, we gain nothing from longer samples since we use only $2n$ data points each time.

Online setting As one can easily see, the theoretical result is stated in off-line setting, when the whole sample of size N is acquired in advance. In online setting we suggest to control the probability α to raise a false alarm for at least one central point t among N data points (which differs from classical techniques controlling the mean distance between false alarms [44]). Having α and N chosen, one should acquire s data-points (set $\{X_i\}_{i\in\mathcal{I}_s}$), employ the proposed bootstrap scheme with the bootstrap samples of length N in order to obtain the critical values. Next the approach can be naturally applied in online setting and Theorem 9 guarantees the capability of the proposed bootstrap scheme to control the aforementioned probability to raise a false alarm.

4.4 Sensitivity result

Consider the following setting. Let there be index τ, such that $\{X_i\}_{i\le\tau}$ are i.i.d. and $\{X_i\}_{i>\tau}$ are i.i.d. as well. Denote covariance matrices $\Sigma_1 := \mathbb{E}\left[X_1 X_1^T\right]$ and $\Sigma_2 := \mathbb{E}\left[X_{\tau+1} X_{\tau+1}^T\right]$. Define the break extent Δ as

$$\Delta := \left\|\Sigma_1 - \Sigma_2\right\|_\infty .$$

The question is, how large the window size n_+ should be in order to reliably reject $\mathbb{H}_0$.

Theorem 10. *Let Assumption 2 hold. Also let* $\Delta_Y < 1/2$ *and*

$$R_{B^\flat} < \frac{\alpha}{6\left|\mathfrak{N}\right|},$$

where Δ_Y *and* $R_{B^\flat}$ *come from Lemma 36 and Lemma 32. Moreover, assume* $\mathcal{I}_s \subseteq 1..\tau$ *and* $\tau \ge n_{suff}$*, where*

$$n_{suff} := \left(\frac{\mathrm{q}\left\|S^{-1}\right\|_\infty - 2\rho + \sqrt{\left(2\rho - \mathrm{q}\left\|S^{-1}\right\|_\infty\right)^2 - 4\Delta\rho^2}}{\sqrt{2}\Delta}\right)^2, \tag{4.10}$$

$$\mathrm{q} = \sqrt{2\left(1+\Delta_Y\right)\log\left(\frac{2N\left|\mathfrak{N}\right|p^2}{\alpha - 3\left|\mathfrak{N}\right|R_{B^\flat}}\right)}, \tag{4.11}$$

$$\rho = \sqrt{2\log p + \chi}.$$

Let it hold for the widest window that $n_+ > n_{suff}$*. Then with probability at least*

$$1 - p_{\mathcal{Z}_s}(\kappa) - 3p^\Sigma(\chi) - p_s^\Omega(t, \mathrm{x}) - p_s^{\mathcal{W}}(\mathrm{x}) \tag{4.12}$$

where $p_{\mathcal{Z}_s}(\kappa)$, $p_s^\Omega(t,\mathrm{x})$, $p_s^{\mathcal{W}}(\mathrm{x})$ *and* $p^\Sigma(\chi)$, *come from Lemma 31, Lemma 35, Lemma 33 and Lemma 41 respectively, the hypothesis* $\mathbb{H}_0$ *will be rejected by the proposed approach at confidence level* α.

The formal proof is given in Appendix C.1.

Discussion of the sensitivity result The expression (4.10) and the residual $R_{B^\flat}$ involved in the statement of Theorem 10 are rather complicated. Here we note that for N, s and $p \to +\infty$, for some positive constant C_4 independent of N, s, p and Δ it holds that

$$n_{suff} \le C_4\left(1 + \frac{\log^2(ps)}{\sqrt{s}}\right)\frac{\log\left(\left|\mathfrak{N}\right|Np^2\right)}{\Delta^2}$$

while the bound (4.6) for $R_{B^\flat}$ holds as well, and the parameters x, t and κ may be chosen as specified by (4.7), (4.9) and (4.8) respectively and χ may be chosen as $\chi = 7.32$ in order to ensure the probability (4.12) to be at least 0.99.

As expected, the bound for sufficient window size decreases with the growth of the break extent Δ and the size of the set $\mathcal{I}_s$, but increases with dimensionality p. It is worth noticing, that the latter dependence is only logarithmic. And again, in the same way as with Theorem 9, the bound increases with the sample size N (only logarithmically) since we use only $2n$ data points.

The assumption $\mathcal{I}_s \subseteq 1..\tau$ is only technical. The result may be proven without relying on it by methodologically the same argument.

Obviously, we still cannot exactly compute n_{suff}, since it depends on the underlying distributions. However this result guarantees that the sensitivity of the test does not vanish in high-dimensional setting.

Online setting Theorem 10 is established in offline setting as well. In online setting it guarantees that the proposed approach can reliably detect a break of an extent not less than Δ with a delay at most n_{suff}.

4.5 Simulation study

4.5.1 Real-world covariance matrices

We have downloaded stock market quotes for $p = 87$ companies included in S&P 100 with 1-minute intervals for approximately a week ($N = 2211$) using the API provided by Google Finance[1]. A sample of interest was composed of 1-minute log returns for each of the companies. Our approach with window size $\mathfrak{N} = \{30\}$ has detected a break at confidence level $\alpha = 0.05$, while the approach proposed by [35] (referred to as ecp below) has detected nothing. The change-point was localized at the morning of Monday 19 December 2016 (the day when the Electoral College had voted).

Discarding the portion of the data around the estimated change-point we have acquired a pair of data samples which both approaches fail to detect a break in. Denote the realistic covariance matrices estimated on each of these samples as Σ_1 and Σ_2.

4.5.2 Design of the simulation study, results and discussion

The goal of the current simulation study is to verify that the bootstrap procedure controls first type error rate and evaluate the power of the test and compare it to the power of ecp. Hence we need to generate two types of realistic datasets – with and without a break for power and first type error rate estimation respectively. In order to generate a dataset without a break we independently draw 520 vectors from normal distribution $\mathcal{N}(0, \Sigma_1)$. As for the datasets including a break, they are generated by binding 400 vectors independently drawn from $\mathcal{N}(0, \Sigma_1)$ and 120 vectors independently drawn from $\mathcal{N}(0, \Sigma_2)$.

The results obtained in the simulation study are given in Table 4.1. One can easily see that the proposed test exhibits proper control of the first type error rate. Being tested in the same setting, ecp has demonstrated proper first type error rate as well, but the power did not exceed 0.1. So, our approach outperforms ecp in all cases apart from $\mathfrak{N} = \{7\}$ and $\mathcal{I}_s = 1..100$.

As expected, the power is higher for larger windows and it may be decreased by adding narrower windows into consideration which is the price to be paid for better change-point

[1] `https://www.google.com/finance`

Table 4.1: First type error rate and power exhibited by the proposed approach for various choice of set of window sizes $\mathfrak{N}$ and sub-set used for bootstrap $\mathcal{I}_s$ at significance level $\alpha = 0.05$. For the case $\mathcal{I}_s \subset 1..\tau$ mean precision of change-point localization is reported as well.

	$\mathcal{I}_s = 1..520$		$\mathcal{I}_s = 1..100$		
$\mathfrak{N}$	I type error rate	power	I type error rate	power	localization
$\{60\}$	.02	1.00	.00	.90	60
$\{30\}$	.01	.90	.00	.52	30
$\{15\}$	.00	.76	.00	.38	15
$\{7\}$	.00	.34	.00	.03	7
$\{60, 30\}$	.01	.99	.00	.84	47.1
$\{60, 30, 15\}$	.01	.99	.00	.82	41.1
$\{60, 30, 15, 7\}$	.01	.99	.00	.78	42.0
$\{30, 15\}$	.01	.90	.00	.49	21.8
$\{30, 15, 7\}$	.01	.84	.00	.34	19.9

localization.

It should be noted that contrary to the intuition expressed in Remark 6 using only a data sub-sample which a priori does not include a break does not necessarily improve the power of the test.

For the case of $\mathcal{I}_s = 1..100 \subset 1..\tau$ Table 4.1 also provides mean precision of change-point localization. One can see, that multiscale approach significantly improves it.

Appendix A

Proofs for Chapter 2

A.1 Proofs of Consistency results

In order to prove the claimed consistency results we employ the primal-dual witness technique which suggest to consider the following optimization problem:

$$\tilde{\Theta} = \arg \min_{\substack{\Theta \in S^p_{++} \\ \Theta_{S^c}=0}} \left[tr(\Theta\hat{\Sigma}) - \log\det\Theta + \|\Lambda * \Theta\|_1 \right]. \tag{A.1}$$

The only difference between the problems (2.1) and (A.1) is that the latter one forces all zero elements to be estimated as zero, e.g. $\tilde{\Theta}_{S^c} = 0$. The main idea of the technique is to show that $\tilde{\Theta} = \hat{\Theta}$ on some set of high probability.

We use $\Delta = \tilde{\Theta} - \Theta^*$ to denote the mis-tie between the true precision matrix and the solution of the problem (A.1).

In our derivations we also make use of properties of the residuals of the first-order Taylor expansion of the gradient of the log-det functional which takes form:

$$R(\Delta) = \tilde{\Theta}^{-1} - \Theta^{*-1} + \Theta^{*-1}\Delta\Theta^{*-1}.$$

A.1.1 Existence and uniqueness of solutions of problems (2.1) and (A.1)

Since we are about to investigate the properties of solutions of the problems (2.1) and (A.1), we first need to give sufficient conditions for their existence and uniqueness. The lemma below is a slightly generalized version of Lemma 3 given in [42] and can be proven by exactly the same argument.

Lemma 1. *Let $\forall i \neq j$ $\Lambda_{ij} > 0$, $\Lambda_{ii} = 0$ and $\Sigma_{ii} > 0$ $\forall i$, then the problems* (2.1) *and* (A.1) *have unique solutions.*

We also give sufficient conditions which do not include positiveness of all non-diagonal elements of Λ but in turn rely on non-singularity of the sample covariance matrix $\hat{\Sigma}$.

Lemma 2. *Suppose, $\hat{\Sigma}$ is non-singular. Then the problems* (2.1) *and* (A.1) *have unique solutions.*

Proof. We give the proof for the problem (2.1). The uniqueness of the solution for the problem (A.1), as well as for a problem with any set of non-diagonal values of Θ restricted to zero (in case it does not violate symmetry) can be established by the same argument.

By Lagrange duality we can rewrite the problem (2.1) in form

$$\hat{\Theta} = \min_{\substack{\Theta \in S^p_{++} \\ \|C(\Lambda) * \Theta\|_1 \le 1}} \left[tr(\Theta \hat{\Sigma}) - \log\det\Theta \right]$$

for some $C_{ij}(\Lambda) < +\infty$ for $\Lambda_{ij} > 0$ and $C_{ij}(\Lambda) = 0$ for $\Lambda_{ij} = 0$. Now, since $\hat{\Sigma}$ is non-singular and it is a covariance matrix, it is positive-definite. Thus, there exists and orthogonal transform S such that $S^T \hat{\Sigma} S = D = diag(d_1 ... d_p)$ and $\forall i \ d_i > 0$.

Then, by using the fact that $tr \Theta \hat{\Sigma} = tr S^T \Theta \hat{\Sigma} S$ and by noting that $\det S = 1$, we further rewrite the problem as

$$\hat{\Theta} = \min_{\substack{\Theta' \in S^p_{++} \\ \|C(\Lambda) * (S \Theta' S^T)\|_1 \le 1}} \left[tr(\Theta' D) - \log\det\Theta' \right],$$

where $\Theta' = S^T \Theta S$. Here we have also used the fact that $\Theta' \in S_{++}$ iff. $\Theta \in S_{++}$.

Now we just substitute the definition of the trace:

$$\hat{\Theta} = \min_{\substack{\Theta' \in S^p_{++} \\ \|C(\Lambda) * (S \Theta' S^T)\|_1 \le 1}} \left[\sum_i d_i \Theta'_{ii} - \log\det\Theta' \right].$$

But, due to the fact that $d_i > 0$ by Lagrange duality we finally obtain

$$\hat{\Theta} = \min_{\substack{\Theta' \in S^p_{++} \\ \|C(\Lambda) * (S \Theta' S^T)\|_1 \le 1 \\ \forall i \ |\Theta'_{ii}| \le C_i(d_i)}} - \log\det\Theta' \tag{A.2}$$

for some $C_i(d_i) < +\infty$.

So, the diagonal elements of Θ' are bounded. Therefore, its trace is bounded, thus the sum of its eigenvalues is bounded, so the feasible set is compact. Thus (recalling the convexity of the log-det functional) the optimum exists and is unique.

Using the fact of equivalence of the problems (A.2) and (2.1) we obtain the claimed statement.

□

A.1.2 Proof of adaptive lasso consistency result

Lemma 3 (generalization of Lemma 6, [42]). *Suppose that*

$$r := 2\kappa_{\Gamma^*}(\|W\|_\infty + \|\Lambda_S\|_\infty) \le \min\left\{ \frac{1}{3\kappa_{\Sigma^*} d}, \frac{1}{3\kappa^3_{\Sigma^*}\kappa_{\Gamma^*} d} \right\}, \tag{A.3}$$

then

$$\left\| \Theta^* - \tilde{\Theta} \right\|_\infty \le r.$$

Proof (adaptation of the one given in [42]). The problem (A.1) has a unique solution, thus the gradient condition holds:

$$G(\Theta_S) := -[\Theta^{-1}]_S + \hat{\Sigma}_S + \Lambda_S * Z_S = 0$$

where Z_S denotes an element of the sub-gradient: $Z_S \in \partial_S \left\|\tilde{\Theta}\right\|_1$.

Now we define a continuous function $F\colon B(r) \to \mathbb{R}^{|S|}$ (where $B(r)$ stands for a zero-centered $|S|$-dimensional l_∞ ball of radius r)

$$F(\Delta_S) \coloneqq -(\Gamma^*_{SS})^{-1}\overline{G}(\Theta^* + \Delta_S) + \overline{\Delta}_S.$$

We now claim that $F(B(r)) \subseteq B(r)$. First, rewrite the expression for $G(\tilde{\Theta}_S)$ as

$$G(\Theta^*_S + \Delta_S) = [-[(\Theta^* + \Delta)^{-1}]_S + [\Theta^{*-1}]_S] + W_S + \Lambda_S * Z_S. \tag{A.4}$$

By Lemma 39 (which applies due to assumption (A.3) and the choice of Δ) we have

$$\overline{R}(\Delta_S)_S = vec((\Theta^* + \Delta)^{-1} - \Theta^{*-1})_S + \Gamma^*_{SS}\overline{\Delta}_S = vec(\Theta^{*-1}\Delta\Theta^{*-1}\Delta J\Theta^{*-1})_S. \tag{A.5}$$

Using (A.4) and (A.5) obtain

$$F(\overline{\Delta}_S) = \underbrace{(\Gamma^*_{SS})^{-1}vec(\Theta^{*-1}\Delta\Theta^{*-1}\Delta J\Theta^{*-1})_S}_{T_1} - \underbrace{(\Gamma^*_{SS})^{-1}(\overline{W}_S + \overline{\Lambda}_S * \overline{Z}_S)}_{T_2}.$$

Clearly, $\|T_2\|_\infty \le \kappa_{\Gamma^*}(\|W\|_\infty + \|\Lambda\|_\infty) = r/2$. As for T_1, by Lemma 39 we have

$$\|T_1\|_\infty \le \frac{3}{2}d\kappa^3_{\Sigma^*}\kappa_{\Gamma^*}\|\Delta\|^2_\infty \le \frac{3}{2}d\kappa^3_{\Sigma^*}\kappa_{\Gamma^*}r^2$$

and again, by assumption (A.3), we obtain $\|T_1\|_\infty \le r/2$.

Now, we have shown that the continuous function $F(\cdot)$ maps a ball $B(r)$ into itself. Thus, we can apply the fixed-point theorem. Obviously, this function has a fixed point iff. $\exists \Delta_S \in B(r)\colon G(\Theta^*_S + \Delta_S) = 0$ which is a sufficient and necessary condition for $\Theta^* + \Delta$ to be a solution of optimization problem (A.1) and thus $\left\|\Theta^* - \tilde{\Theta}\right\|_\infty \le r$.

□

Proof of Theorem 1. First, we note that $\hat{\Theta}^{init}_{ij} = 0$ iff. $\Theta^*_{ij} = 0$ (by Lemma 38). Thus, by the choice of Λ^{ada}, $\hat{\Theta}^{ada} = \tilde{\Theta}^{ada}$, so we can analyze the problem (A.1). Note, that this also implies the fact that $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{ada}_{ij} = 0$. Also, by Lemma 38 $\left\|\hat{\Theta}^{ada} - \Theta^*\right\|_\infty \le r$. Thus,

$$\left\|\Lambda^{ada}_S\right\|_\infty \le \frac{\lambda_n}{\left(\min\limits_{i,j\colon \hat{\Theta}^{init}_{ij} \neq 0} \hat{\Theta}^{init}_{ij}\right)^\gamma} \le \frac{\lambda_n}{(\theta_{min} - r)^\gamma}. \tag{A.6}$$

Lemma 3 applies to the problem (A.1) with tuning parameters Λ^{ada} due to the sparsity bound (2.2) and the bound we have just obtained. Thus, $\left\|\Theta^* - \tilde{\Theta}^{ada}\right\|_\infty \le 2\kappa_{\Gamma^*}(\|W\|_\infty + \left\|\Lambda^{ada}_S\right\|_\infty)$. Substituting the bound (A.6), recalling that we are considering the set $\mathcal{T}$ and that $\hat{\Theta}^{ada} = \tilde{\Theta}^{ada}$ we obtain the claimed bound.

□

A.1.3 Proof of SCAD graphical lasso consistency result

Lemma 4 (generalization of Lemma 4, [42])**.** *Let*

$$\max\{\|W\|_\infty, \|R(\Delta)\|_\infty\} \leq \frac{\alpha}{8}\rho$$

and

$$\frac{\|\Lambda_S\|_\infty}{\rho} \leq 1. \tag{A.7}$$

Also, suppose Assumption 1 holds for some $\alpha \in (0, 1]$*. Then* $\hat{\Theta} = \tilde{\Theta}$*.*

Proof (adaptation of the one given in [42]). First, rewrite the stationarity condition for the problem (2.1) as

$$\Theta^{*-1}\Delta\Theta^{*-1} + W - R(\Delta) + \Lambda * Z = 0.$$

By vectorizing obtain:

$$\Gamma^*\overline{\Delta} + \overline{W} - \overline{R} + \overline{\Lambda * Z} = 0.$$

Now, using the fact that $\Delta_{S^c} = 0$ rewrite it in terms of disjoint decomposition:

$$\Gamma^*_{SS}\overline{\Delta_S} + \overline{W_S} - \overline{R}_S + \overline{\Lambda_S * Z_S} = 0, \tag{A.8}$$

$$\Gamma^*_{S^cS}\overline{\Delta_S} + \overline{W_{S^c}} - \overline{R}_{S^c} + \overline{\Lambda_{S^c} * Z_{S^c}} = 0. \tag{A.9}$$

Solving (A.8) we obtain

$$\overline{\Delta_S} = -(\Gamma^*_{SS})^{-1}[\overline{W_{S^c}} - \overline{R}_{S^c} + \overline{\Lambda_{S^c} * Z_{S^c}}].$$

Now, by solving (A.9) for $\overline{Z}_{S^c}$ and by substituting $\overline{\Delta_S}$:

$$\begin{aligned}
\overline{Z}_{S^c} &= -[\Gamma^*_{S^cS}\overline{\Delta_S} + \overline{W_{S^c}} - \overline{R}_{S^c}] \oslash \overline{\Lambda}_{S^c} \\
&= \left[(I - \Gamma^*_{S^cS}{\Gamma^*_{SS}}^{-1})(\overline{W}_S + \overline{R}_S) - \Gamma^*_{S^cS}{\Gamma^*_{SS}}^{-1}\overline{\Lambda}_S * \overline{Z}_S\right],
\end{aligned}$$

where $\cdot \oslash \cdot$ denotes matrix element-wise division. Now we take the ℓ_∞ norm of both sides and recall Assumption 1

$$\begin{aligned}
\left\|\overline{Z}_{S^c}\right\|_\infty &\leq \frac{2-\alpha}{\rho}(\|W\|_\infty + \|R\|_\infty) + (1-\alpha)\frac{\|\Lambda_S\|_\infty}{\rho} \\
&\leq \frac{2}{\rho}(\|W\|_\infty + \|R\|_\infty) + (1-\alpha) \\
&\leq \frac{2}{\rho}\left(\frac{2\alpha}{8}\rho\right) + (1-\alpha) \\
&= 1 - \frac{\alpha}{2} \\
&< 1.
\end{aligned}$$

The strict dual feasibility condition holds. Therefore, we have $\hat{\Theta} = \tilde{\Theta}$.

□

Lemma 5 (generalization of Theorem 1, [42]). *Consider a distribution satisfying Assumption 1 with some $\alpha \in (0,1]$, let $\hat{\Theta}$ be the solution of the optimization problem* (2.1). *Suppose also the following restrictions on the penalization parameters Λ hold*

$$\|\Lambda_S\|_\infty \leq \frac{8}{\alpha}\delta_n$$

and

$$\rho \geq \frac{8}{\alpha}\delta_n.$$

Furthermore, suppose the following sparsity assumption holds:

$$d \leq \frac{\delta_n}{6(\delta_n + \|\Lambda_S\|_\infty)^2 \max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa_{\Gamma^*}^2\kappa_{\Sigma^*}^3\}}. \tag{A.10}$$

Then on the set $\mathcal{T} = \left\{\left\|\hat{\Sigma} - \Sigma^*\right\|_\infty < \delta_n\right\}$ the following hold:

$$\left\|\hat{\Theta} - \Theta^*\right\|_\infty \leq r_\Lambda := 2\kappa_{\Gamma^*}(\delta_n + \|\Lambda_S\|_\infty)$$

and

$$\Theta^*_{ij} = 0 \Rightarrow \hat{\Theta}_{ij} = 0. \tag{A.11}$$

Proof. First, we show that Lemma 3 applies. The inequality

$$2\kappa_{\Gamma^*}(\|W\|_\infty + \|\Lambda_S\|_\infty) \leq \min\left\{\frac{1}{3\kappa_{\Sigma^*}d}, \frac{1}{3\kappa_{\Sigma^*}^3\kappa_{\Gamma^*}d}\right\} \tag{A.12}$$

holds due to assumption (A.10). Therefore, we have a bound

$$\left\|\tilde{\Theta} - \Theta^*\right\|_\infty \leq 2\kappa_{\Gamma^*}(\|W\|_\infty + \|\Lambda_S\|_\infty). \tag{A.13}$$

Now, we show the applicability of Lemma 4. First, observe that

$$\|W\|_\infty \leq \delta_n \leq \frac{\alpha}{8}\rho.$$

In order to bound $R(\Delta)$ we use Lemma 39 which applies due to bounds (A.13) and (A.12) and make use of the sparsity bound (A.10):

$$\|R(\Delta)\|_\infty \leq \frac{3}{2}d\,\|\Delta\|_\infty^2\,\kappa_{\Sigma^*}^3 \leq \delta_n \leq \frac{\alpha}{8}\rho.$$

The assumption (A.7) of Lemma 4 clearly holds as well.

Thus, $\tilde{\Theta} = \hat{\Theta}$ which combined with (A.13) gives the claimed bound along with the sparsistency property (A.11).

□

In the next two proofs we denote the penalization matrix used at the k-th iteration as $\Lambda_{ij}^{(k)} = SCAD'_{\lambda,a}(|\Theta_{ij}^{(k-1)}|)$ and its minimal value corresponding to zero elements of the true precision matrix as $\rho^{(k)} = \min_{(i,j)\in S^c} \Lambda_{ij}^{(k)}$.

Proof of Theorem 2. Since, the conditions of the Lemma 38 hold, $\left\|\hat{\Theta}^{(0)} - \Theta^*\right\|_\infty \leq r$ and $\Theta^*_{ij} = 0 \Leftrightarrow \Theta^{(0)}_{ij} = 0$.

Therefore, $\left\|\Lambda^{(1)}_S\right\|_\infty \leq \lambda_n \leq \frac{8}{\alpha}\delta_n$ and $\rho^{(1)} = \lambda_n \geq \frac{8}{\alpha}\delta_n$ and, due to $\hat{\Sigma}$ being non-singular, the problem (2.4) has a unique solution. Thus, Lemma 5 applies here giving the bound for $\hat{\Theta}^{OSSCAD}$. Moreover, due to the bound (2.3) we have $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{OSSCAD}_{ij} = 0$ (since the bound for $\hat{\Theta}^{(1)}_{ij}$ is not less strict that the one for $\hat{\Theta}^{(0)}_{ij}$). □

Proof of Theorem 3. Theorem 2 provides the bound for $\hat{\Theta}^{(1)}$ along with the sparsistency property: $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}^{(1)}_{ij} = 0$.

Following the same argument we prove the following bound for every $\hat{\Theta}^{(k)}$:

$$\left\|\hat{\Theta}^{(k)} - \Theta^*\right\|_\infty \leq 2\kappa_{\Gamma^*}\left(\delta_n + \left\|\Lambda^{(k)}_S\right\|_\infty\right) \tag{A.14}$$

and we have the following recurrent expression for $\Lambda^{(k)}_S$

$$\left\|\Lambda^{(k)}_S\right\|_\infty \leq \left(\frac{a\lambda_n - |\theta_{min} - 2\kappa_{\Gamma^*}(\delta_n + \Lambda^{(k-1)}_S)|}{a-1}\right)_+ .$$

Some algebra yields

$$\left\|\Lambda^{(k)}_S\right\|_\infty \xrightarrow{k\to\infty} \left\|\Lambda^{(\infty)}_S\right\|_\infty \leq \left(\frac{a\lambda_n - \theta_{min} + 2\kappa_{\Gamma^*}\delta_n}{2\kappa_{\Gamma^*} + a - 1}\right)_+ .$$

And the passage to the limit in inequality (A.14) yields the claimed bound. The second statement of the theorem follows from (2.3).

□

Theorem 4 and Theorem 5 can be proved in the same way as Theorem 2 and Theorem 3 but Lemma 1 should be used instead of Lemma 5 in order to show the existence and uniqueness of the underlying optimization problems.

A.2 Proof of the inference result

The next lemma bounds the remainder r on the set $\mathcal{T} = \left\{\left\|\hat{\Sigma} - \Sigma^*\right\|_\infty < \delta_n\right\}$.

Lemma 6. *Suppose, assumptions of Lemma 5 hold. Then, on the set* $\mathcal{T} = \left\{\left\|\hat{\Sigma} - \Sigma^*\right\|_\infty < \delta_n\right\}$ *it holds that*

$$\|r\|_\infty \leq R := dr_\Lambda(dr_\Lambda(\nu_{\Sigma^*} + \delta_n) + 2\kappa_{\Theta^*}\delta_n) \tag{A.15}$$

where $\nu_{\Sigma^*} := \|\Theta^*\|_\infty$.

Proof.

$$\begin{aligned}
\|r\|_\infty &\leq \left\|(\hat{\Theta} - \Theta^*)W\Theta^*\right\|_\infty + \left\|(\hat{\Theta}\hat{\Sigma} - I_p)(\hat{\Theta} - \Theta^*)\right\|_\infty \\
&\leq |||(\hat{\Theta} - \Theta^*)|||_1(\|W\Theta^*\|_\infty + \left\|(\hat{\Theta}\hat{\Sigma} - I_p)\right\|_\infty) \\
&\leq dr_\Lambda(\|W\Theta^*\|_\infty + \left\|\hat{\Theta}\hat{\Sigma} - I_p\right\|_\infty)
\end{aligned}$$

$$\left\|(\hat{\Theta}\hat{\Sigma} - I_p)\right\|_\infty = \left\|(\hat{\Theta} - \Theta^*)\hat{\Sigma} + \Theta^*(\hat{\Sigma} - \Sigma^*)\right\|_\infty$$
$$\leq dr_\Lambda(\nu_{\Sigma^*} + \delta_n) + \kappa_{\Theta^*}\delta_n$$
$$\|r\|_\infty \leq dr_\Lambda(dr_\Lambda(\nu_{\Sigma^*} + \delta_n) + 2\kappa_{\Theta^*}\delta_n).$$

□

The next lemma shows that conditioning on a set of high probability does not significantly change the measure of the set.

Lemma 7. *Consider a measure $\mathbb{P}$ and a pair of sets A and B. Then denoting $p := \mathbb{P}\{B\}$*

$$|\mathbb{P}\{A\} - \mathbb{P}\{A|B\}| \leq 2(1-p).$$

Proof.

$$\begin{aligned}|\mathbb{P}\{A\} - \mathbb{P}\{A|B\}| &= \left|\mathbb{P}\{A|B\}\mathbb{P}\{B\} + \mathbb{P}\left\{A|\overline{B}\right\}\mathbb{P}\left\{\overline{B}\right\} - \mathbb{P}\{A|B\}\right| \\ &= \left|\mathbb{P}\{A|B\}(p-1) + \mathbb{P}\left\{A|\overline{B}\right\}(1-p)\right| \\ &\leq |\mathbb{P}\{A|B\}(p-1)| + \left|\mathbb{P}\left\{A|\overline{B}\right\}(1-p)\right| \\ &\leq 2(1-p).\end{aligned}$$

□

Proof of Theorem 6. Using (2.5), and the definition of $\hat{T}$ (4.9) we obtain for all (i,j)

$$\sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij}) = \frac{1}{\sqrt{n}}\sum_k Z_{ijk} + \frac{r}{\sqrt{n}}, \tag{A.16}$$

where

$$Z_{ijk} := \Theta^*_i X_k \Theta^*_j X_k - \Theta^*_{ij}. \tag{A.17}$$

Observe that Z_{ijk} are i.i.d. (for (i,j) fixed) and $\mathbb{E}[Z_{ijk}] = 0$.

Now we divide both sides of (A.16) by $\sigma_{ij} := \sqrt{Var[Z_{ijk}]}$

$$\sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij})/\sigma_{ij} = \underbrace{\frac{1}{\sigma_{ij}\sqrt{n}}\sum_k Z_{ijk}}_{S} + \frac{r\sqrt{n}}{\sigma_{ij}}.$$

The cumulative distribution function of S can be estimated by Berry-Esseen inequality [31]

$$|\mathbb{P}\{S < c\} - \Phi(c)| \leq \frac{A\mu_{ij3}}{\sigma^3_{ij}\sqrt{n}}$$

with $A < 0.4748$.

Now from Lemma 7 we have

$$|\mathbb{P}\{S < c\} - \mathbb{P}\{S < c|\,\mathcal{T}\}| \leq 2(1 - p_{\mathcal{T}}).$$

Combining the latter two inequalities yields

$$|\mathbb{P}\{S < c|\,\mathcal{T}\} - \Phi(c)| \leq \frac{A\mu_{ij3}}{\sigma^3_{ij}\sqrt{n}} + 2(1 - p_{\mathcal{T}}). \tag{A.18}$$

Next we make use of the bound for the residual r provided by Lemma 6.

$$\mathbb{P}\left\{\sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij})/\sigma_{ij} \leq c \,|\, \mathcal{T}\right\} \leq \mathbb{P}\left\{S - \frac{R\sqrt{n}}{\sigma_{ij}} < c \,|\, \mathcal{T}\right\}.$$

And making use of (A.18)

$$\mathbb{P}\left\{\sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij})/\sigma_{ij} \leq c \,|\, \mathcal{T}\right\} \leq \Phi\left(c + \frac{R\sqrt{n}}{\sigma_{ij}}\right) + \frac{A\mu_{ij3}}{\sigma^3_{ij}\sqrt{n}} + 2(1 - p_{\mathcal{T}}). \tag{A.19}$$

In the same manner one obtains

$$\Phi\left(c - \frac{R\sqrt{n}}{\sigma_{ij}}\right) - \frac{A\mu_{ij3}}{\sigma^3_{ij}\sqrt{n}} - 2(1 - p_{\mathcal{T}}) \leq \mathbb{P}\left\{\sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij})/\sigma_{ij} \leq c \,|\, \mathcal{T}\right\}. \tag{A.20}$$

Also notice that

$$\forall a : \arg\max_c |\Phi(c) - \Phi(c + a)| = 0. \tag{A.21}$$

Combining (A.19), (A.20) and (A.21) yields

$$\sup_c \left|\mathbb{P}\left\{\sqrt{n}(\hat{T}_{ij} - \Theta^*_{ij})/\sigma_{ij} \leq c \,|\, \mathcal{T}\right\} - \Phi(c)\right| \leq \left(\Phi\left(\frac{R\sqrt{n}}{\sigma_{ij}}\right) - \frac{1}{2}\right) + \frac{A\mu_{ij3}}{\sigma^3_{ij}\sqrt{n}} + 2(1 - p_{\mathcal{T}}).$$

□

Appendix B

Proofs for Chapter 3

B.1 Proof of bootstrap validity result

Proof of Theorem 7. Proof consists in applying the finite-sample Theorem 11. Its applicability is guaranteed by the consistency results given in papers [43, 5, 28] and by the results from [27, 28, 5] bounding the term $R_{\hat{T}}$. High probability of set $\mathcal{T}_T$ is ensured by Lemma 41. □

Theorem 11. *Assume $\mathbb{H}_0$ holds and furthermore, let $X_1, X_2, ...X_N$ be i.i.d. Let $\hat{\Theta}$ denote a symmetric estimator of Θ^* s.t. for some positive r*

$$\left\| \Theta^* - \hat{\Theta} \right\|_\infty < r$$

and $\Theta^*_{ij} = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Suppose Assumption 2 holds and there exists $R_{\hat{T}}$ such that $\sqrt{n} \left\| r_n^{\mathfrak{S}}(t) \right\|_\infty \le R_{\hat{T}}$ for all $\mathfrak{S} \in \{l, r\}$, $n \in \mathfrak{N}$ and $t \in \mathbb{T}_n$ on set

$$\mathcal{T}_T := \left\{ \forall \mathfrak{S} \in \{l, r\}, n \in \mathfrak{N}, t \in \mathbb{T}_n : \left\| \hat{\Sigma}_n^{\mathfrak{S}}(t) - \mathbb{E}\left[X_1 X_1^T \right] \right\|_\infty \le \delta_n \right\}.$$

Moreover, let

$$R := (3 + 2\,|\mathfrak{N}|) \left(2R_A(R_{\hat{T}}) + 2R_{A^\flat} + R_\Sigma^\pm(r) \right) \le \frac{\alpha}{2},$$

where the remainders R_A, $R_{A^\flat}$, $R_\Sigma^\pm$ are defined in Lemma 13, Lemma 18 and Lemma 10 respectively and the mis-tie Δ_Y involved in the definition of $R_\Sigma^\pm$ comes from Lemma 21. Then on set $\mathcal{T}_T$ it holds that

$$\left| \mathbb{P}\left\{ \forall n \in \mathfrak{N} : A_n \le x_n^\flat(\alpha) \right\} - (1 - \alpha) \right| \le R + 2(1 - q),$$

where

$$q = 1 - p_s^{\Sigma_Y}(x, q) - p^\Sigma(\gamma) - p_s^M(x) \tag{B.1}$$

and the terms $p_s^{\Sigma_Y}(x, q)$, $p^\Sigma(\gamma)$ and $p_s^M(x)$ are defined in Lemma 21, Lemma 41 and Lemma 18 respectively.

Discussion of finite-sample bootstrap validity result The terms Δ_Y, R_A, $R_{A^\flat}$ and $R_\Sigma^\pm$ involved in the statement of Theorem 11 are rather complicated. The exact expressions for them are provided by Lemma 21, Lemma 13, Lemma 18 and Lemma 10 respectively, 3rd and 4th moments M_3^3 and M_4^4 involved therein are bounded by Lemma 16 and Lemma 22 while asymptotic bounds for $R_{\hat{T}}$ are provided in [28] (for node-wise procedure) and [27] (for graphical lasso). For the case of graphical lasso an explicit form of $R_{\hat{T}}$ is given in [5].

Here we just note that if $\hat{\Theta}$ is a root-n consistent estimator, recovering sparsity pattern (graphical lasso [43], adaptive graphical lasso [53] or thresholded de-sparsified estimator based on node-wise procedure [28]), then for $d, s, p, N, n_-, n_+ \to \infty$, $N > 2n_+$, $n_+ \geq n_-$, $s \geq n_-$ and $\frac{d^2}{n_-} = o(1)$ given the spectrum of Θ^* is bounded

$$R_{A^b} \leq D_1 \left(\frac{L^4 d \log^7(2p^2 T n_+)}{n_-} \right)^{1/6} \log^2(ps). \tag{B.2}$$

If either graphical lasso, adaptive graphical lasso or node-wise procedure [36] is used with $\lambda_n \asymp \sqrt{\frac{\log p}{n}}$ in order to obtain $\hat{\Theta}_n^{\mathfrak{S}}(t)$, then on set $\mathcal{T}_T$ it holds that

$$R_A \leq D_2 \left(\frac{L^4 d \log^7(2p^2 T n_+)}{n_-} \right)^{1/6} + D_3 \sqrt{\frac{\log 2p^2 T}{n_-}} d \log p.$$

The high probability of $\mathcal{T}_T$ may be ensured by means of Lemma 41 e.g., choosing $\gamma = \log(500T)$ for $\mathbb{P}\{\mathcal{T}_T\} \geq 0.99$. Further

$$\Delta_Y \leq D_4 \frac{L^4 d^2}{\sqrt{s}},$$

$$R_\Sigma^{\pm} \leq D_5 \left(\frac{L^4 d^2}{\sqrt{s}} \right)^{1/3} \log^{2/3}(2p^2 T).$$

Here $D_1, ..., D_5$ are positive constants independent of N, $\mathfrak{N}$, d, p and s. We also note that the proper choice of x, γ and q in (B.1) is

$$x = 6, \tag{B.3}$$

$$\gamma = \log(500T), \tag{B.4}$$

$$q = 7 + 4\log(p) \tag{B.5}$$

which ensures the probability defined by (B.1) to be above 0.99. For exact expressions for $p_s^{\Sigma_Y}(x, q)$, $p^{\Sigma}(\gamma)$ and $p_s^M(x)$ see Lemma 21, Lemma 20 and Lemma 18.

Proof of Theorem 11. The proof consists in application of Lemma 17, Lemma 14 and Lemma 12 justifying applicability of Lemma 10. □

B.2 Proof of sensitivity result

Proof of Theorem 8. Proof consists in applying the finite-sample Theorem 12. Its applicability is guaranteed by the consistency results given in papers [43, 5, 28] and by the results from [27, 28, 5] bounding the term $R_{\hat{T}}$. High probability of set $\mathcal{T}_{\leftrightarrow}$ is ensured by Lemma 20. □

Theorem 12. *Let $\mathcal{I}_s \subseteq 1..\tau$. Let $\hat{\Theta}$ denote a symmetric estimator of Θ_1 s.t. for some $r \in \mathbb{R}$ it holds that*

$$\left\| \Theta_1 - \hat{\Theta} \right\|_\infty < r$$

and $(\Theta_1)_{ij} = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Suppose Assumption 2 holds and there exists $R_{\hat{T}}$ such that $\left\| r^{\mathfrak{S}}_{n_+}(t) \right\|_\infty \leq R_{\hat{T}}$ for all $\mathfrak{S} \in \{l, r\}$ and $t \in \mathbb{T}_{n_+}$ on some set

$$\mathcal{T}_{\leftrightarrow} := \left\{ \forall t \leq \tau - n_+ : \left\| \hat{\Sigma}^{\mathfrak{S}}_n(t) - \Sigma^*_1 \right\|_\infty \leq \delta_{n_+} \right\}$$
$$\bigcap \left\{ \forall t \geq \tau + n_+ : \left\| \hat{\Sigma}^{\mathfrak{S}}_n(t) - \Sigma^*_2 \right\|_\infty \leq \delta_{n_+} \right\}.$$

Moreover, let the residual R_{A^b} defined in Lemma 18 be bounded:

$$R_{A^b} \leq \frac{\alpha}{6\,|\mathfrak{N}|}.$$

Also let

$$\sqrt{\frac{n_+}{2}} \left\| S \right\|_\infty (\Delta - 2R_{\hat{T}}) \geq \mathrm{q}, \tag{B.6}$$

where

$$\mathrm{q} := \sqrt{2\,(1 + \Delta_Y(r)) \log \left(\frac{2N\,|\mathfrak{N}|\,p^2}{\alpha - 3\,|\mathfrak{N}|\,R_{A^b}} \right)} \tag{B.7}$$

and Δ_Y is defined in Lemma 21. Then on set $\mathcal{T}_{\leftrightarrow}$ with probability at least

$$1 - p^{\Sigma_Y}_s(x, q),$$

where $p^{\Sigma_Y}_s(x, q)$ is defined in Lemma 21, $\mathbb{H}_0$ will be rejected.

Discussion of finite-sample sensitivity result The assumption (B.6) is rather complicated. Here we note that if either graphical lasso [43], adaptive graphical lasso [53] or thresholded de-sparsified estimator based on node-wise procedure [28] with penalization parameter chosen as $\lambda_s \asymp o(\sqrt{\log p/n})$ was used, given $d, s, p, N, n_-, n_+ \to \infty$, $N > 2n_+$, $n_+ \geq n_-$, $s \geq n_-$ and $d = o(\sqrt{n_+})$ it boils down to

$$n_+ \geq D_6 \frac{1}{\Delta} \left(\left\| S^{-1} \right\|_\infty \log(N\,|\mathfrak{N}|\,p^2) \right)^2$$

for some positive constant D_6 independent of $N, \mathfrak{N}, p, d, S$ while the parameters q, γ and x may be chosen as in (B.5), (B.4), (B.3) (high probability of $\mathcal{T}_{\leftrightarrow}$ is ensured by Lemma 41). At the same time the remainder R_{A^b} can be bounded by (B.2).

As expected, the bound for sufficient window size decreases with growth of the break extent Δ and the size of the set $\mathcal{I}_s$, but increases with dimensionality p. It is worth noticing, that the latter dependence is only logarithmic. And again, in the same way as with Theorem 7, the bound increases with the sample size N (only logarithmically) since we use only $2n$ data points.

Proof of Theorem 12. Consider a pair of centered normal vectors

$$\eta := \begin{pmatrix} \eta^1 & \eta^2 & \ldots & \eta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \Sigma^*_Y),$$

$$\zeta := \begin{pmatrix} \zeta^1 & \zeta^2 & \dots & \zeta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \hat{\Sigma}_Y),$$
$$\Sigma_Y^* := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^n\right],$$
$$\hat{\Sigma}_Y := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^{n\flat}\right],$$

where vectors $Y_{\cdot j}^n$ and $Y_{\cdot j}^{n\flat}$ are defined in proofs of Lemma 14 and Lemma 17 respectively. Lemma 9 applies here and yields for all positive q

$$\mathbb{P}\left\{\|\zeta^{n_+}\|_\infty \geq \mathrm{q}\right\} \leq 2\left|\mathbb{T}_{n_+}\right| p^2 \exp\left(-\frac{\mathrm{q}^2}{2\left\|\hat{\Sigma}_Y\right\|_\infty}\right),$$

where $\hat{\Sigma}_Y = \mathrm{Var}\left[\zeta\right]$ and $\left|\mathbb{T}_{n_+}\right|$ is the number of central points for the window of size n_+. Applying Lemma 21 on a set of probability at least $1 - p_s^{\Sigma_Y}(x, q)$ yields $\left\|\Sigma_Y^* - \hat{\Sigma}_Y\right\|_\infty \leq \Delta_Y$, and hence, due to the fact that $\|\Sigma_Y^*\|_\infty = 1$ by construction,

$$\mathbb{P}\left\{\|\zeta^{n_+}\|_\infty \geq \mathrm{q}\right\} \leq 2\left|\mathbb{T}_{n_+}\right| p^2 \exp\left(-\frac{\mathrm{q}^2}{2\left(1 + \Delta_Y\right)}\right).$$

Due to Lemma 18 and continuity of Gaussian c.d.f.

$$\mathbb{P}^\flat\left\{A_{n_+}^\flat \geq x_{n_+}^\flat(\alpha)\right\} \geq \alpha / |\mathfrak{N}| - 2R_{A^\flat}$$

and due to Lemma 18 along with the fact that $\left|\mathbb{T}_{n_+}\right| < N$, choosing q as proposed by equation (B.7) we ensure that $x_{n_+}^\flat(\alpha) \leq \mathrm{q}$.

Now by assumption of the theorem and by construction of the test statistics A_n

$$A_{n_+} \geq \sqrt{\frac{n_+}{2}} \|S\|_\infty \left(\Delta - 2R_{\hat{T}}\right).$$

Finally, we notice that due to assumption (B.6) $A_{n_+} > \mathrm{q}$ and therefore, $\mathbb{H}_0$ will be rejected. □

Lemma 8. *Consider a centered random Gaussian vector $\xi \in \mathbb{R}^p$ with arbitrary covariance matrix Σ. For any positive* q *it holds that*

$$\mathbb{P}\left\{\max_i \xi_i \geq \mathrm{q}\right\} \leq p \exp\left(-\frac{\mathrm{q}^2}{2\|\Sigma\|_\infty}\right).$$

Proof. By convexity we obtain the following chain of inequalities for any t

$$e^{t\mathbb{E}[t \max_i \xi_i]} \leq \mathbb{E}\left[e^{t \max_i \xi_i}\right] \leq \mathbb{E}\left[e^{t \sum_i \xi_i}\right] \leq p e^{t^2 \|\Sigma\|_\infty / 2}.$$

Chernoff bound yields for any t

$$\mathbb{P}\left\{\max_i \xi_i \geq \mathrm{q}\right\} \leq \frac{p e^{t^2 \|\Sigma\|_\infty / 2}}{e^{t\mathrm{q}}}.$$

Finally, optimization over t yields the claim. □

As a trivial corollary, one obtains

Lemma 9. *Consider a centered random Gaussian vector $\xi \in \mathbb{R}^p$ with arbitrary covariance matrix Σ. For any positive* q *it holds that*

$$\mathbb{P}\left\{\|\xi\|_\infty \geq \mathrm{q}\right\} \leq 2p \exp\left(-\frac{\mathrm{q}^2}{2\|\Sigma\|_\infty}\right).$$

B.3 Sandwiching lemma

The following lemma is a generalization covering the case of non-continuous probability measures of Lemma 21 of [9].

Lemma 10. *Consider a normal multivariate vector* η *with a deterministic covariance matrix and a normal multivariate vector* ζ *with a possibly random covariance matrix such that*

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}\left\{\forall n\in\mathfrak{N}: A_n\le x_n\right\} - \mathbb{P}\left\{\forall n\in\mathfrak{N}: \|\eta_n\|_\infty \le x_n\right\}\right| \le R_A, \tag{B.8}$$

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}^\flat\left\{\forall n\in\mathfrak{N}: A^\flat_n\le x_n\right\} - \mathbb{P}^\flat\left\{\forall n\in\mathfrak{N}: \|\zeta_n\|_\infty \le x_n\right\}\right| \le R_{A^\flat}, \tag{B.9}$$

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}\left\{\forall n\in\mathfrak{N}: A_n\le x_n\right\} - \mathbb{P}^\flat\left\{\forall n\in\mathfrak{N}: A^\flat_n \le x_n\right\}\right| \le R. \tag{B.10}$$

where η_n *and* ζ_n *are sub-vectors of* η *and* ζ *respectively. Then*

$$\left|\mathbb{P}\left\{\forall n\in\mathfrak{N}: A_n\le x^\flat_n(\alpha)\right\} - (1-\alpha)\right| \le (3+2|\mathfrak{N}|)\left(R+R_A+R_{A^\flat}\right).$$

Proof. Let us introduce some notation. Denote multivariate cumulative distribution functions of $A_n, A^\flat_n, \|\eta_n\|_\infty, \|\zeta_n\|_\infty$ as $P, P^\flat, \mathcal{N}, \mathcal{N}^\flat : \mathbb{R}^{|\mathfrak{N}|} \to [0,1]$ respectively. Define the following sets for all $\delta\in[0,\alpha]$

$$\mathcal{Z}_+(\delta) := \left\{z : \mathcal{N}(z) \ge 1-\alpha-\delta\right\},$$

$$\mathcal{Z}_-(\delta) := \left\{z : \mathcal{N}(z) \le 1-\alpha+\delta\right\}$$

and their boundaries

$$\partial\mathcal{Z}_+(\delta) := \left\{z : \mathcal{N}(z) = 1-\alpha-\delta\right\}, \tag{B.11}$$

$$\partial\mathcal{Z}_-(\delta) := \left\{z : \mathcal{N}(z) = 1-\alpha+\delta\right\}.$$

Consider $\delta = R+R_A+R_{A^\flat}$ and denote sets $\mathcal{Z}_+ = \mathcal{Z}_+(\delta)$, $\mathcal{Z}_- = \mathcal{Z}_-(\delta)$, $\partial\mathcal{Z}_- = \partial\mathcal{Z}_-(\delta)$, $\partial\mathcal{Z}_+ = \partial\mathcal{Z}_+(\delta)$. Define a set of thresholds satisfying the confidence level

$$\mathcal{Z}^\flat := \left\{z : P^\flat(z) \ge 1-\alpha \;\&\; \forall z_1 < z : P^\flat(z_1) < 1-\alpha\right\}$$

here and below comparison of vectors should be understood element-wise. Notice that due to continuity of multivariate normal distribution and assumption (B.9) $\forall z^\flat \in \mathcal{Z}^\flat$

$$\left|P^\flat(z^\flat) - (1-\alpha)\right| \le R_{A^\flat}. \tag{B.12}$$

Now for all $z_- \in \partial\mathcal{Z}_-$ and for all $z^\flat\in\mathcal{Z}^\flat$ it holds that

$$\begin{aligned} P^\flat(z_-) &\le P(z_-) + R \\ &\le \mathcal{N}(z_-) + R + R_A \\ &\le 1-\alpha-R_{A^\flat} \\ &\le P^\flat(z^\flat) \end{aligned}$$

where we have consequently used (B.10), (B.8), (B.11) and (B.12). In the same way one obtains for all $z_+ \in \partial\mathcal{Z}_+$ and for all $z^\flat \in \mathcal{Z}^\flat$

$$P^\flat(z_+) \geq P^\flat(z^\flat)$$

which implies that $\mathcal{Z}^\flat \subset \mathcal{Z}_- \cap \mathcal{Z}_+$.

Now denote quantile functions of $\|\eta_n\|_\infty$ as $z^N : [0,1] \to \mathbb{R}^{|\mathfrak{N}|}$:

$$\forall n \in \mathfrak{N} : \mathbb{P}\left\{\|\eta_n\|_\infty \geq z_n^N(\mathrm{x})\right\} = \mathrm{x}.$$

In exactly the same way define quantile functions $z^{N^\flat} : [0,1] \to \mathbb{R}^{|\mathfrak{N}|}$ of $\|\zeta_n\|_\infty$. Clearly for all $\mathrm{x} \in [0,1]$,

$$z^N(\mathrm{x}+\delta) \leq z^\flat(\mathrm{x}) \leq z^N(\mathrm{x}-\delta)$$

and hence

$$z^\flat(\alpha^*) \leq z^N(\alpha^* - \delta) \leq z^\flat(\alpha^* - 2\delta),$$
$$1 - \alpha \leq P^\flat(z^N(\alpha^* - \delta)) \leq P^\flat(z^\flat(\alpha^* - 2\delta)).$$

Using Taylor expansion with Lagrange remainder term we obtain for some $0 \leq \kappa \leq 2\delta$

$$\begin{aligned}
\mathcal{N}^\flat\left(z^\flat(\alpha^* - 2\delta)\right) &\leq \mathcal{N}^\flat\left(z^{N^\flat}(\alpha^* - 2\delta)\right) + \delta \\
&= \mathcal{N}^\flat\left(z^{N^\flat}(\alpha^*)\right) + \sum_{n\in\mathfrak{N}} \partial_{z_n^\flat}\mathcal{N}^\flat(z^{N^\flat}(\alpha^*))\partial_\alpha z_n^{N^\flat}(\alpha^*)\kappa + \delta \\
&\leq 1 - \alpha + \sum_{n\in\mathfrak{N}} \partial_{z_n^\flat}\mathcal{N}^\flat(z^{N^\flat}(\alpha^*))\partial_\alpha z_n^{N^\flat}(\alpha^*)\kappa + 3\delta.
\end{aligned}$$

Next successively using Lemma 11 and the fact that the quantile function is the inverse function of the c.d.f. we obtain

$$\mathcal{N}^\flat\left(z^\flat(\alpha^* - 2\delta)\right) \leq 1 - \alpha + 3\delta + 2\delta\,|\mathfrak{N}|$$

and therefore

$$1 - \alpha \leq P^\flat\left(z^\flat(\alpha^* - 2\delta)\right) \leq 1 - \alpha + \delta\,(3 + 2\,|\mathfrak{N}|)\,,$$
$$1 - \alpha \leq P^\flat\left(z^N(\alpha^* - \delta)\right) \leq 1 - \alpha + \delta\,(3 + 2\,|\mathfrak{N}|)\,.$$

In the same way one obtains

$$1 - \alpha - \delta\,(3 + 2\,|\mathfrak{N}|) \leq P^\flat\left(z^N(\alpha^* + \delta)\right) \leq 1 - \alpha.$$

Next, by the argument used in the beginning of the proof we obtain

$$z^N(\alpha^* + \delta), z^N(\alpha^* - \delta) \in \mathcal{Z}_-(\delta\,(3 + 2\,|\mathfrak{N}|)) \cap \mathcal{Z}_+\left(\delta\,(3 + 2\,|\mathfrak{N}|)\right).$$

As the final ingredient, we need to choose deterministic α^+ and α^- such that

$$N(z^N(\alpha^- + \delta)) = 1 - \alpha - \delta\,(3 + 2\,|\mathfrak{N}|)\,,$$

$$N(z^N(\alpha^+ - \delta)) = 1 - \alpha + \delta\,(3 + 2\,|\mathfrak{N}|)$$

(which is possible due to continuity), so $\alpha^- \leq \alpha^* \leq \alpha^+$ and hence by monotonicity

$$z^N(\alpha^- + \delta) \leq z^N(\alpha^* + \delta) \leq z^\flat(\alpha^*) \leq z^N(\alpha^* - \delta) \leq z^N(\alpha^+ - \delta)$$

and finally

$$
\begin{aligned}
1-\alpha-\delta\left(3+2\left|\mathfrak{N}\right|\right) &\le P(z^{N}(\alpha^{-}+\delta)) \\
&\le P(z^{\flat}(\alpha^{*})) \\
&\le P(z^{N}(\alpha^{+}-\delta)) \\
&\le 1-\alpha+\delta\left(3+2\left|\mathfrak{N}\right|\right).
\end{aligned}
$$

□

Lemma 11. *Consider a random variable ξ and an event A defined on the same probability space. Let c.d.f. $\mathbb{P}\left\{\xi \le x\right\}$ and $\mathbb{P}\left\{\xi \le x \& A\right\}$ be differentiable. Then*

$$
\frac{\partial_x \mathbb{P}\left\{\xi \le x \& A\right\}}{\partial_x \mathbb{P}\left\{\xi \le x\right\}} \le 1
$$

Proof. Indeed, denoting the complement of set A as $\overline{A}$ we obtain,

$$
\begin{aligned}
\frac{\partial_x \mathbb{P}\left\{\xi \le x \& A\right\}}{\partial_x \mathbb{P}\left\{\xi \le x\right\}} &= \frac{\partial_x \mathbb{P}\left\{\xi \le x \& A\right\}}{\partial_x\left(\mathbb{P}\left\{\xi \le x \& A\right\}+\mathbb{P}\left\{\xi \le x \& \overline{A}\right\}\right)} \\
&= \frac{\partial_x \mathbb{P}\left\{\xi \le x \& A\right\}}{\partial_x \mathbb{P}\left\{\xi \le x \& A\right\}+\partial_x \mathbb{P}\left\{\xi \le x \& \overline{A}\right\}} \\
&= \frac{1}{1+\frac{\partial_x \mathbb{P}\left\{\xi \le x \& \overline{A}\right\}}{\partial_x \mathbb{P}\left\{\xi \le x \& A\right\}}}
\end{aligned}
$$

Using the fact that derivative of c.d.f. is non-negative we finalize the proof. □

B.4 Similarity of joint distributions of $\{A_n\}_{n\in\mathfrak{N}}$ and $\{A_n^{\flat}\}_{n\in\mathfrak{N}}$

Lemma 12. *Under assumptions of Theorem 7 it holds on set $\mathcal{T}$ with probability at least*

$$
1-p_s^{\Sigma_Y}(x,q)-p^{\Sigma}(\gamma)-p_s^{M}(x)
$$

that

$$
\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}\left\{\forall n\in\mathfrak{N}: A_n \le x_n\right\}-\mathbb{P}^{\flat}\left\{\forall n\in\mathfrak{N}: A_n^{\flat} \le x_n\right\}\right| \le R_A+R_{A^{\flat}}+R_{\Sigma}^{\pm}.
$$

Proof. The proof consists in applying Lemma 17, Lemma 14, Lemma 21 and Lemma 44. □

B.5 Gaussian approximation result for A_n

Lemma 13. *Suppose there exists $R_{\hat{T}}$ such that $\sqrt{n}\left\|r^{\mathfrak{S}}(t)\right\|_{\infty} \le R_{\hat{T}}$ for all $\mathfrak{S}\in\{l,r\}$ and t on some set $\mathcal{T}$. Then on set $\mathcal{T}$ it holds that*

$$
\begin{aligned}
&\sup_x \left|\mathbb{P}\left\{\forall n\in\mathfrak{N}: A_n \le x_n\right\}-\mathbb{P}\left\{\forall n\in\mathfrak{N}: \left\|\eta^n\right\|_{\infty} \le x_n\right\}\right| \le R_A \\
&\quad := C_A\left(\left(F\log^7(p^2 T n_+)\right)^{1/6}+4R_{\hat{T}}\sqrt{\log(2p^2T)}\right).
\end{aligned}
$$

where F is defined by (B.14) *and η^n by* (B.13).

Proof. Substituting (3.3) to (4.2) yields

$$A_n(t) = \left\| \underbrace{\frac{1}{\sqrt{2n}} S^{-1} \left(\sum_{i \in \mathcal{I}_n^l(t)} \overline{Z}_i - \sum_{i \in \mathcal{I}_n^r(t)} \overline{Z}_i \right)}_{S_Z^n(t)} + \frac{1}{\sqrt{2}} (\overline{r_n^l} - \overline{r_n^r}) \right\|_\infty .$$

Now denote stacked $S_Z^n(t)$ for all t as S_Z^n and for all n and t as S_Z. Lemma 14 bounds the c.d.f. of $\|S_Z\|_\infty$ as

$$\sup_x |\mathbb{P}\{\forall n \in \mathfrak{N} : \|S_Z^n\|_\infty \le x_n\} - \mathbb{P}\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \le x_n\}| \le C_A \left(F \log^7(p^2 T n_+) \right)^{1/6} .$$

But clearly on set $\mathcal{T}$

$$|A_n - \|S_Z^n\|_\infty| \le \sqrt{2} R_{\hat{T}}$$

and hence for all $\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}$

$$\begin{aligned} |\mathbb{P}\{\forall n \in \mathfrak{N} : A_n < x_n | \mathcal{T}\} - \mathbb{P}\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \le x_n\}| \le C_A \left(F \log^7(p^2 T n_+) \right)^{1/6} \\ + \mathbb{P}\left\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \le x_n + \sqrt{2} R_{\hat{T}}\right\} \\ - \mathbb{P}\left\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \le x_n - \sqrt{2} R_{\hat{T}}\right\} . \end{aligned}$$

Now notice that $\forall i : (\Sigma_Y^*)_{ii} = 1$ and bound the latter two terms by means of Lemma 43:

$$\begin{aligned} \sup_{\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}^{|\mathfrak{N}|}} |\mathbb{P}\{\forall n \in \mathfrak{N} : A_n < x_n | \mathcal{T}\} - \mathbb{P}\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \le x_n\}| \le C_A \left(F \log^7(p^2 T n_+) \right)^{1/6} \\ + 4 R_{\hat{T}} (\sqrt{\log(2p^2 T)}) \end{aligned}$$

□

Lemma 14. *Let Assumption 2 hold. Then*

$$\sup_x |\mathbb{P}\{\forall n \in \mathfrak{N} : \|S_Z^n\|_\infty \le x_n\} - \mathbb{P}\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \le x_n\}| \le C_A \left(F \log^7(2p^2 T n_+) \right)^{1/6} ,$$

where

$$\begin{pmatrix} \eta^1 & \eta^2 & \dots & \eta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_Y^*), \tag{B.13}$$

$$\Sigma_Y^* = \frac{1}{N} \sum_{i=1}^N \mathrm{Var}[Y_{\cdot i}] ,$$

$$F = \frac{1}{2n_-} \left(\beta \log 2 \vee \frac{\sqrt{2}}{\sqrt{2} - 1} \gamma \right)^2 \vee \frac{1}{2n_+} \left(\frac{n_+}{n_-} \right)^{1/3} M_3^2 \vee \sqrt{\frac{1}{2n_+ n_-}} M_4^2 \tag{B.14}$$

with γ defined by (B.17), *β by* (B.18) *and Y by* (B.15) *and an independent constant C_A .*

Proof. Consider a matrix Y_n with $2n_+$ columns

$$
Y_n^T := \sqrt{\frac{n_+}{n}} \times
\begin{pmatrix}
Z_1^S & O & \dots & O & -Z_{2n_+ +1}^S & \dots \\
Z_2^S & Z_2^S & \dots & \dots & \dots & \dots \\
\dots & Z_3^S & \dots & \dots & \dots & \dots \\
Z_n^S & \dots & \dots & \dots & \dots & \dots \\
-Z_{n+1}^S & Z_{n+1}^S & \dots & \dots & \dots & \dots \\
-Z_{n+2}^S & -Z_{n+2}^S & \dots & \dots & \dots & \dots \\
\dots & -Z_{n+3}^S & \dots & O & \dots & \dots \\
-Z_{2n}^S & \dots & \dots & Z_{2n_+-2n+1}^S & O & \dots \\
O & -Z_{2n+1}^S & \dots & Z_{2n_+-2n+2}^S & Z_{2n_+-2n+2}^S & \dots \\
O & O & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & -Z_{2n_+-1}^S & -Z_{2n_+-1}^S & \dots \\
O & O & \dots & -Z_{2n_+}^S & -Z_{2n_+}^S & \dots
\end{pmatrix}
$$

where $Z_i^S := (S^{-1}\overline{Z}_i)^T$. Clearly, columns of the matrix are independent and

$$
S_Z^n = \frac{1}{\sqrt{2n_+}} \sum_{l=0}^{2n_+} (Y_n)_{\cdot l}
$$

Next define a block matrix composed of Y_n matrices:

$$
Y := \begin{pmatrix} Y_1 \\ \hline Y_2 \\ \hline \dots \\ \hline Y_{|\mathfrak{N}|} \end{pmatrix} \tag{B.15}
$$

Clearly vectors $Y_{\cdot l}$ are independent and

$$
S_Z = \frac{1}{\sqrt{2n_+}} \sum_{l=0}^{2n_+} Y_{\cdot l}
$$

In order to complete the proof we make use of Lemma 42. Denote

$$
B_{n_+} = \sqrt{\frac{n_+}{n_-}} \left(\beta \log 2 \vee \frac{\sqrt{2}}{\sqrt{2}-1} \gamma \right) \vee \left(\frac{n_+}{n_-} \right)^{1/6} M_3 \vee \left(\frac{n_+}{n_-} \right)^{1/4} M_4 \tag{B.16}
$$

By means of Lemma 22 one shows that the assumptions of Lemma 15 hold for components of Z_i^S with

$$
\gamma := 12 L^2 \sqrt{d} \Lambda(\Theta^*) \|\Theta^*\|_\infty \left\|S^{-1}\right\|_\infty \tag{B.17}
$$

$$
\beta := \left(\frac{9}{2} L^2 \sqrt{d} \Lambda(\Theta^*) + 1 \right) \|\Theta^*\|_\infty \left\|S^{-1}\right\|_\infty \tag{B.18}
$$

where $\Lambda(\Theta^*)$ denotes the maximal eigen value of Θ^*. Therefore condition (D.1) holds with B_n defined by equation (B.16). Further,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[(Y_{ij}^n)^2\right] \geq \min_j \operatorname{Var}\left[Z_{1j}^S\right] = 1.$$

Hence, Assumption D.5.4 is fulfilled with $b = 1$. Next notice that for some k-th component of Z_i^S and a central point t (both defined by j):

$$\begin{aligned}
\frac{1}{2n_+}\sum_{i=1}^{2n_+}\mathbb{E}\left[\left|Y_{ij}^n\right|^3\right] &= \frac{1}{2n_+}\sum_{i\in\mathcal{I}_n^l(t)\cup\mathcal{I}_n^r(t)}\mathbb{E}\left[\left(\sqrt{\frac{n_+}{n}}\left|Z_{ik}^S\right|\right)^3\right] \\
&= \frac{1}{2n_+}\sum_{i\in\mathcal{I}_n^l(t)\cup\mathcal{I}_n^r(t)}\left(\frac{n_+}{n}\right)^{3/2}\mathbb{E}\left[\left|Z_{ik}^S\right|^3\right] \\
&= \frac{2n}{2n_+}\left(\frac{n_+}{n}\right)^{3/2}\mathbb{E}\left[\left|Z_{ik}^S\right|^3\right] \\
&= \sqrt{\frac{n_+}{n}}\mathbb{E}\left[\left|Z_{ik}^S\right|^3\right] \\
&\leq \sqrt{\frac{n_+}{n_-}}M_3^3
\end{aligned}$$

and in the same way:

$$\frac{1}{2n_+}\sum_{i=1}^{N}\mathbb{E}\left[\left|Y_{ij}^n\right|^4\right] \leq \frac{n_+}{n_-}M_4^4.$$

Therefore Assumption D.5.5 holds with B_{n_+} so Lemma 42 applies here and provides us with the claimed bound. Moreover, C_A depends only on b which equals one which implies that the constant C_A is independent.

□

Lemma 15. *Consider a random variable ξ. Suppose the following bound holds $\forall x \geq 0$:*

$$\mathbb{P}\left\{|\xi| \geq \gamma x + \beta\right\} \leq e^{-x}.$$

Then

$$\mathbb{E}\left[\exp\left(\frac{|\xi|}{B}\right)\right] \leq 2$$

for

$$B = \beta\log 2 \vee \frac{\sqrt{2}}{\sqrt{2}-1}\gamma$$

Proof. Integration by parts yields

$$\mathbb{E}\left[\exp\left(\frac{|\xi|}{B}\right)\right] \leq \exp\left(\frac{\beta}{B}\right) + \frac{\gamma}{B}\int_0^{+\infty}\exp\left(\frac{\gamma x+\beta}{B}\right)e^{-x}dx$$

$$\int_0^{+\infty} \exp\left(\frac{\gamma x + \beta}{B}\right) e^{-x} dx = \frac{B}{B-\gamma} \exp\left(\frac{\beta}{B}\right)$$

$$\mathbb{E}\left[\exp\left(\frac{|\xi|}{B}\right)\right] \le \frac{B}{B-\gamma} \exp\left(\frac{\beta}{B}\right) \\ \le 2$$

□

By the same technique the following lemma can be proven

Lemma 16. *Under assumptions of Lemma 15*

$$\mathbb{E}\left[|\xi|^3\right] \le \beta^3 + 3\gamma\beta^2 + 6\beta\gamma^2 + 2\gamma^3,$$

$$\mathbb{E}\left[\xi^4\right] \le \beta^4 + 4\gamma\beta^3 + 12\beta^2\gamma^2 6\beta\gamma^3 + 24\gamma^4.$$

B.6 Gaussian approximation result for $A_n^\flat$

Lemma 17.

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : A^\flat \le x_n\right\} - \mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : \|\zeta^n\|_\infty \le x_n\right\}\right| \le \hat{C}_{A^\flat}\left(F^\flat \log^7(2p^2Tn_+)\right)^{1/6},$$

where

$$\begin{pmatrix} \zeta^1 & \zeta^2 & \dots & \zeta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \hat{\Sigma}_Y),$$

$$\hat{\Sigma}_Y = \frac{1}{N}\sum_{i=1}^{N} \operatorname{Var}\left[Y_{\cdot i}^\flat\right],$$

$$F^\flat = \left(\frac{1}{2n_- \log^2 2} \vee \frac{1}{2n_+}\left(\frac{n_+}{n_-}\right)^{1/3} \vee \sqrt{\frac{1}{2n_+n_-}}\right) \left\|S^{-1}\right\|_\infty^2 (M^\flat)^2$$

$$M^\flat = \max_{i\in\mathcal{I}_s} \left\|\hat{Z}_i\right\|_\infty$$

$Y_n^\flat$ *are defined by* (B.19)*, and* $\hat{C}_{A^\flat}$ *depends only on* $\min_{1\le k\le p}(\hat{\Sigma}_Y)_{kk}$.

Proof. Denote the term under the sign of $\|\cdot\|_\infty$ in (4.3) as $S_Z^{n\flat}(t)$

$$S_Z^{n\flat}(t) := \frac{1}{\sqrt{2n}}\left(\sum_{i\in\mathcal{I}_n^l(t)} Z_i^{S\flat} - \sum_{i\in\mathcal{I}_n^r(t)} Z_i^{S\flat}\right)^T$$

where $Z_i^{S\flat} := (S^{-1}\overline{Z_i^\flat})^T$ and let $S_Z^\flat$ be a vector composed of stacked vectors $S_Z^{n\flat}(t)$ for all $n \in \mathfrak{N}$ and central points t.

Consider a matrix

$$
(Y_n^\flat)^T := \sqrt{\frac{n_+}{n}} \times
\begin{pmatrix}
Z_1^{S\flat} & O & \dots & O & -Z_{2n_++1}^{S\flat} & \dots \\
Z_2^{S\flat} & Z_2^{S\flat} & \dots & \dots & \dots & \dots \\
\dots & Z_3^{S\flat} & \dots & \dots & \dots & \dots \\
Z_n^{S\flat} & \dots & \dots & \dots & \dots & \dots \\
-Z_{n+1}^{S\flat} & Z_{n+1}^{S\flat} & \dots & \dots & \dots & \dots \\
-Z_{n+2}^{S\flat} & -Z_{n+2}^{S\flat} & \dots & \dots & \dots & \dots \\
\dots & -Z_{n+3}^{S\flat} & \dots & O & \dots & \dots \\
-Z_{2n}^{S\flat} & \dots & \dots & Z_{2n_+-2n+1}^{S\flat} & O & \dots \\
O & -Z_{2n+1}^{S\flat} & \dots & Z_{2n_+-2n+2}^{S\flat} & Z_{2n_+-2n+2}^{S\flat} & \dots \\
O & O & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & -Z_{2n_+-1}^{S\flat} & -Z_{2n_+-1}^{S\flat} & \dots \\
O & O & \dots & -Z_{2n_+}^{S\flat} & -Z_{2n_+}^{S\flat} & \dots
\end{pmatrix}
\tag{B.19}
$$

which is a bootstrap counterpart of Y_n from the proof of Lemma 14 and construct a block matrix $Y^\flat$:

$$
Y^\flat = \begin{pmatrix} Y_1^\flat \\ \hline Y_2^\flat \\ \hline \dots \\ \hline Y_{|\mathfrak{N}|}^\flat \end{pmatrix}
$$

Clearly vectors $Y_{\cdot l}^\flat$ are independent and

$$
S_Z^\flat = \frac{1}{\sqrt{2n_+}} \sum_{l=0}^{N} Y_{\cdot l}^\flat
$$

Now notice

$$
\frac{1}{2n_+} \sum_{i=1}^{N} \mathbb{E}\left[|Y_{ij}|^3\right] \le \sqrt{\frac{n_+}{n_-}} \max_{i \in \mathcal{I}_s} \left\|\hat{Z}_i\right\|_\infty^3 \left\|S^{-1}\right\|_\infty^3
$$

$$
\frac{1}{2n_+} \sum_{i=1}^{N} \mathbb{E}\left[|Y_{ij}|^4\right] \le \frac{n_+}{n_-} \max_{i \in \mathcal{I}_s} \left\|\hat{Z}_i\right\|_\infty^4 \left\|S^{-1}\right\|_\infty^4
$$

And finally apply Lemma 42.

□

Lemma 18. *Let $\hat{\Theta}$ denote an estimator of Θ^* s.t. for some positive r*

$$
\left\|\Theta^* - \hat{\Theta}\right\|_\infty < r
$$

and $\Theta_{ij}^* = 0 \Rightarrow \hat{\Theta}_{ij} = 0$, furthermore, let $\Delta_Y(r) < 1/2$, also suppose Assumption 2 holds. Then at least with probability $1 - p_s^M(x) - p_s^{\Sigma_Y}(x, q)$

$$
\sup_{\{x_n\}_{n\in\mathfrak{N}} \subset \mathbb{R}} \left|\mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : A^\flat \le x_n\right\} - \mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : \|\zeta^n\|_\infty \le x_n\right\}\right| \le R_{A^\flat} := C_{A^\flat}\left(\hat{F}\log^7(2p^2Tn_+)\right)^{1/6}
$$

where

$$\hat{F} = \left(\frac{1}{2n_- \log^2 2} \vee \frac{1}{2n_+} \left(\frac{n_+}{n_-} \right)^{1/3} \vee \sqrt{\frac{1}{2n_+ n_-}} \right) \left\| S^{-1} \right\|_\infty^2 (C^\flat)^2$$

$$C^\flat \coloneqq \mathcal{Z}_s(x) + (3(dx)^2 + 1)r$$

and constant $C_{A^\flat}$ depends only on Δ_Y.

Proof. The proof consists in subsequently applying Lemma 17 and Lemma 19 ensuring $C^\flat \geq M^\flat = \max_{i \in \mathcal{I}_s} \left\| \hat{Z}_i \right\|_\infty$ with probability at least $1 - p_s^M(x)$ and applying Lemma 21 providing that $\left\| \Sigma_Y^* - \hat{\Sigma}_Y \right\|_\infty \leq \Delta_Y \leq 1 = \min_{1 \leq k \leq p} (\Sigma_Y^*)_{kk}$ with probability at least $1 - p_s^{\Sigma_Y}(x, q)$ which implies the existence of a deterministic constant $C_{A^\flat} > \hat{C}_{A^\flat}$. ☐

Lemma 19. *Let $\hat{\Theta}$ denote an estimator of Θ^* s.t. for some positive r*

$$\left\| \Theta^* - \hat{\Theta} \right\|_\infty < r$$

and $\Theta^*_{ij} = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Also let Assumption 2 hold. Then with probability at least $1 - p_s^M(x)$

$$M^\flat \leq \mathcal{Z}_s(x) + \Delta_Z(x) \tag{B.20}$$

where $p_s^M(x) \coloneqq p_{\mathcal{Z}_s}(x) + p_s^X(x)$.

Proof. Direct application of Lemma 23 yields

$$\mathbb{P}\left\{ \forall i \in \mathcal{I}_s : \|Z_i\|_\infty \leq \mathcal{Z}_s(x) \right\} \geq 1 - p_{\mathcal{Z}_s}(x)$$

which in combination with the fact (provided by Lemma 25) that $\left\| \hat{Z}_i - Z_i \right\|_\infty \leq \Delta_Z(x)$ implies (B.20).

☐

B.7 $\hat{\Sigma}_Y \approx \Sigma_Y^*$

First of all, if $\Sigma_Z^* \coloneqq \operatorname{Var}\left[\overline{Z_i}\right] \approx \operatorname{Var}_\flat\left[\overline{Z_i^\flat}\right]$, then $\Sigma_Y^* \approx \hat{\Sigma}_Y$ as well (Lemma 21). The idea is to notice that

$$\operatorname{Cov}\left[\overline{Z_i^\flat}\right] = \hat{\Sigma}_{\hat{Z}} \coloneqq \mathbb{E}_{\mathcal{I}_s}\left[\left(\overline{\hat{Z}_i} - \mathbb{E}_{\mathcal{I}_s}\left[\overline{\hat{Z}_i}\right] \right) \left(\overline{\hat{Z}_i} - \mathbb{E}_{\mathcal{I}_s}\left[\overline{\hat{Z}_i}\right] \right)^T \right]$$

due to the choice of the bootstrap scheme. Next we show that $\Sigma_Z^* \approx \hat{\Sigma}_Z \coloneqq \mathbb{E}_{\mathcal{I}_s}\left[\left(\overline{Z_i} - \mathbb{E}_{\mathcal{I}_s}\left[\overline{Z_i}\right] \right) \left(\overline{Z_i} - \mathbb{E}_{\mathcal{I}_s} \right. \right.$ (Lemma 24) and finalize the proof by proving that $\hat{\Sigma}_Z \approx \hat{\Sigma}_{\hat{Z}}$ (Lemma 26).

The results of this section rely on a lemma which is a trivial corollary of Lemma 6 by [27] providing the concentration result for the empirical covariance matrix

Lemma 20. *Let Assumption 2 hold for some $L > 0$. Then for any positive γ*

$$\mathbb{P}\left\{\left\|\hat{\Sigma} - \Sigma^*\right\|_\infty \geq \delta_n(\gamma)\right\} \leq p^{\Sigma}(\gamma) := 2e^{-\gamma},$$

where

$$\delta_n(\gamma) := 2L^2\left(\frac{2\log p + \gamma}{n} + \sqrt{\frac{4\log p + 2\gamma}{n}}\right).$$

Lemma 21. *Assume, Assumption 2 holds. Moreover, let*

$$\left\|\mathbb{E}_{\mathcal{I}_s}\left[X_i X_i^T\right] - \Sigma^*\right\|_\infty \leq \delta_s$$

and let $\hat{\Theta}$ denote a symmetric estimator of Θ^* s.t.

$$\left\|\Theta^* - \hat{\Theta}\right\|_\infty < r$$

and $\Theta^*_{ij} = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Then for any positive x and q

$$\mathbb{P}\left\{\left\|\hat{\Sigma}_Y - \Sigma^*_Y\right\|_\infty \geq \Delta_Y\right\} \leq p_s^{\Sigma_Y}(x, q)$$

where

$$p_s^{\Sigma_Y}(x, q) := p_s^{\Sigma_{Z1}}(x, q) + p_s^{\Sigma_{Z2}}(x)$$

$$\Delta_Y := \left\|S^{-1}\right\|_\infty^2\left(\Delta^{(1)}_{\Sigma_Z} + \Delta^{(2)}_{\Sigma_Z}\right)$$

and $\Delta^{(1)}_{\Sigma_Z}$ and $\Delta^{(2)}_{\Sigma_Z}$ along with the probabilities $p_s^{\Sigma_{Z1}}(x, q)$ and $p_s^{\Sigma_{Z2}}(x)$ are defined in Lemma 24 and Lemma 26 respectively.

Proof. Notice that

$$\left\|\hat{\Sigma}_Y - \Sigma^*_Y\right\|_\infty = \left\|S^{-1}\hat{\Sigma}_{\hat{Z}}S^{-1} - S^{-1}\Sigma^*_Z S^{-1}\right\|_\infty \leq \left\|S^{-1}\right\|_\infty^2\left\|\hat{\Sigma}_{\hat{Z}} - \Sigma^*_Z\right\|_\infty$$

because the matrices $\hat{\Sigma}_Y$ and Σ^*_Y are composed of blocks $S^{-1}\hat{\Sigma}_Z S^{-1}$ and $S^{-1}\Sigma^*_Z S^{-1}$ respectively, each block multiplied by some positive value not greater than 1 (which can be verified by simple algebra).

By Lemma 26 and Lemma 24

$$\left\|\hat{\Sigma}_{\hat{Z}} - \Sigma^*_Z\right\|_\infty \leq \Delta^{(1)}_{\Sigma_Z} + \Delta^{(2)}_{\Sigma_Z}$$

and hence

$$\left\|\hat{\Sigma}_Y - \Sigma^*_Y\right\|_\infty \leq \left\|S^{-1}\right\|_\infty^2 (\Delta^{(1)}_{\Sigma_Z} + \Delta^{(2)}_{\Sigma_Z})$$

with probability at least

$$1 - p_s^{\Sigma_{Z1}}(x, q) - p_s^{\Sigma_{Z2}}(x)$$

□

Lemma 22. *Under Assumption 2 it holds for arbitrary $u, v \in 1..p$ and positive x that*

$$\mathbb{P}\left\{|Z_{1,uv}| \le \left(3L^2\sqrt{d}\Lambda\left(\Theta^*\right)\left(\frac{3}{2}+4x\right)+1\right)\|\Theta^*\|_\infty\right\} \ge 1 - e^{-x}$$

Proof. Re-write the definition (A.17) of an element $Z_{i,uv}$ for arbitrary $u, v \in 1..p$

$$\begin{aligned} Z_{i,uv} &= \Theta_u^* X_i \Theta_v^* X_i - \Theta_{uv}^* \\ &= X_i^T\left[\Theta_u^*(\Theta_v^*)^T\right]X_i - \Theta_{uv}^*. \end{aligned}$$

The first term is clearly a value of a quadratic form defined by the matrix $B = \Theta_u^*(\Theta_v^*)^T$. Note that $rankB = 1$ which implies that it is either positive semi-definite or negative semi-definite. Next we apply Lemma 45 and obtain for all positive x

$$\mathbb{P}\left\{\left|X_i^T B X_i\right| \ge 3L^2\left(|trB| + 2\sqrt{tr(B^2)x} + 2\left|\Lambda\left(B\right)\right|x\right)\right\} \le e^{-x}. \tag{B.21}$$

Again, due to the fact that B is a rank-1 matrix

$$trB = \Lambda\left(B\right) = \sqrt{trB^2} \tag{B.22}$$

and by construction of matrix B

$$\begin{aligned} |trB| &= \left|\Theta_u^*(\Theta_v^*)^T\right| \\ &\le \|\Theta_u^*\|_1\|\Theta^*\|_\infty \\ &\le \sqrt{d}||\Theta_u^*||_2\|\Theta^*\|_\infty \\ &\le \sqrt{d}\Lambda\left(\Theta^*\right)\|\Theta^*\|_\infty. \end{aligned} \tag{B.23}$$

Substitution of (B.22) and (B.23) to (B.21) yields

$$\mathbb{P}\left\{\left|X_i^T B X_i\right| \ge 3L^2\sqrt{d}\Lambda\left(\Theta^*\right)\|\Theta^*\|_\infty\left(1+2\sqrt{x}+2x\right)\right\} \le e^{-x}.$$

And since $\sqrt{x} \le x + \frac{1}{4}$

$$\mathbb{P}\left\{\left|X_i^T B X_i\right| \ge 3L^2\sqrt{d}\Lambda\left(\Theta^*\right)\|\Theta^*\|_\infty\left(\frac{3}{2}+4x\right)\right\} \le e^{-x}.$$

Finally, we obtain a bound for $Z_{i,uv}$ as

$$\mathbb{P}\left\{|Z_{i,uv}| \ge \left(3L^2\sqrt{d}\Lambda\left(\Theta^*\right)\left(\frac{3}{2}+4x\right)+1\right)\|\Theta^*\|_\infty\right\} \le e^{-x}.$$

□

Correction for all i, u and v establishes the following result

Lemma 23. *Consider an i.i.d. sample $X_1, .., X_n$. Under Assumption 2 for positive x it holds that*

$$\mathbb{P}\left\{\forall i \in \{1..n\} : \|Z_i\|_\infty \le \mathcal{Z}_n(x)\right\} \ge 1 - p_{\mathcal{Z}_n}(x)$$

where

$$\mathcal{Z}_n(x) := \left(3L^2\sqrt{d}\Lambda\left(\Theta^*\right)\left(\frac{3}{2}+4\log p^2 n + 4x\right)+1\right)\|\Theta^*\|_\infty,$$

$$p_{\mathcal{Z}_n}(x) := e^{-x}.$$

Lemma 24. *Under Assumption 2 for positive x and q*

$$\mathbb{P}\left\{\left\|\hat{\Sigma}_Z - \Sigma_Z^*\right\|_\infty \geq \Delta_{\Sigma_Z}^{(1)}\right\} \leq p_s^{\Sigma_{Z1}}(x, q)$$

where

$$\Delta_{\Sigma_Z}^{(1)} := \frac{s}{s-1} \frac{\left(4\mathcal{Z}_s^2(x) + \frac{s-1}{s} \left\|\Sigma_Z^*\right\|_\infty\right) q}{3s} \left(1 + \sqrt{1 + \frac{9s\sigma_W^2}{q\left(4\mathcal{Z}_s^2(x) + \frac{s-1}{s} \left\|\Sigma_Z^*\right\|_\infty\right)^2}}\right)$$

$$p_s^{\Sigma_{Z1}}(x, q) := p^4 e^{-q} + p_{\mathcal{Z}_s}(x)$$

Proof. Denote

$$W^{(i)} := (\overline{Z}_i - \mathbb{E}_{\mathcal{I}_s}\left[\overline{Z}_i\right])(\overline{Z}_i - \mathbb{E}_{\mathcal{I}_s}\left[\overline{Z}_i\right])^T - \frac{s-1}{s}\Sigma_Z^*$$

and note that

$$\frac{s-1}{s}\left(\hat{\Sigma}_Z - \Sigma_Z^*\right) = \frac{1}{s}\sum_{i \in \mathcal{I}_s} W^{(i)}.$$

By Lemma 23 we have $\|Z_i\|_\infty \leq \mathcal{Z}_s(x)$ with probability at least $1 - p_{\mathcal{Z}_s}(x)$ which implies $\left\|W^{(i)}\right\|_\infty \leq 4\mathcal{Z}_s^2(x) + \frac{s-1}{s}\left\|\Sigma_Z^*\right\|_\infty$. Since $W_{kl}^{(i)}$ are i.i.d., bounded and centered, Bernstein inequality applies here:

$$\mathbb{P}\left\{\mathbb{E}_{\mathcal{I}_s}\left[W_{kl}^{(i)}\right] \geq \frac{\left(4\mathcal{Z}_s^2(x) + \frac{s-1}{s} \left\|\Sigma_Z^*\right\|_\infty\right) q}{3s} \left(1 + \sqrt{1 + \frac{9s\sigma_W^2}{q\left(4\mathcal{Z}_s^2(x) + \frac{s-1}{s} \left\|\Sigma_Z^*\right\|_\infty\right)^2}}\right)\right\} \leq e^{-q}$$

where σ_W^2 is the smallest variance of components of $W^{(i)}$. Therefore

$$\mathbb{P}\left\{\left\|\mathbb{E}_{\mathcal{I}_s}\left[W^{(i)}\right]\right\|_\infty \geq \frac{\left(4\mathcal{Z}_s^2(x) + \frac{s-1}{s} \left\|\Sigma_Z^*\right\|_\infty\right) q}{3s} \left(1 + \sqrt{1 + \frac{9s\sigma_W^2}{q\left(4\mathcal{Z}_s^2(x) + \frac{s-1}{s} \left\|\Sigma_Z^*\right\|_\infty\right)^2}}\right)\right\} \leq p^4 e^{-q}.$$

□

The following lemma bounds the mis-tie between Z_i and $\hat{Z}_i$.

Lemma 25. *Let Assumption 2 hold and let $\hat{\Theta}$ be a symmetric estimator of Θ^* s.t.*

$$\left\|\Theta^* - \hat{\Theta}\right\|_\infty < r$$

and $\Theta_{ij}^* = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Then for positive x

$$\mathbb{P}\left\{\forall i \in \mathcal{I}_s : \left\|Z_i - \hat{Z}_i\right\|_\infty \leq \Delta_Z(x)\right\} \geq 1 - p_s^X(x)$$

where

$$\Delta_Z(x) := 2rd^{3/2}x^2 \left\|\Theta^*\right\|_\infty + (rdx)^2,$$

$$p_s^X(x) := se^{-x^2/L^2}$$

Proof. Due to sub-Gaussianity,

$$\forall \alpha \in \mathbb{R}^p : \mathbb{P}\left\{\left|\alpha^T X_i\right| \le x\right\} \ge 1 - se^{-x^2/L^2}. \tag{B.24}$$

Now consider the mis-tie of arbitrary elements $Z_{i,uv}$ and $\hat{Z}_{i,uv}$:

$$\begin{aligned}\left|Z_{i,uv} - \hat{Z}_{i,uv}\right| &= \left|\Theta^*_u X_i \Theta^*_v X_i + \Theta^*_{uv} - \hat{\Theta}_u X_i \hat{\Theta}_v X_i - \hat{\Theta}_{uv}\right| \\ &\le \left|(\Theta^*_u - \hat{\Theta}_u) X_i \Theta^*_v X_i\right| + \left|(\Theta^*_u - \hat{\Theta}_u) X_i \hat{\Theta}_v X_i\right| + r.\end{aligned}$$

Now note that due to (B.24) and assumptions imposed on Θ^*

$$|\Theta^*_v X_i| \le \sqrt{d}\,\|\Theta^*\|_\infty x,$$

$$\left|(\Theta^*_v - \hat{\Theta}_v) X_i\right| \le rdx,$$

$$\left|\hat{\Theta}_v X_i\right| \le |\Theta^*_v X_i| + \left|(\Theta^*_v - \hat{\Theta}_v) X_i\right| \le \sqrt{d}\,\|\Theta^*\|_\infty x + rdx.$$

And hence

$$\left|Z_{i,uv} - \hat{Z}_{i,uv}\right| \le 2rd^{3/2}x^2\,\|\Theta^*\|_\infty + (rdx)^2.$$

□

Lemma 26. *Assume Assumption 2 holds. Let $\hat{\Theta}$ be a symmetric estimator of Θ^* s.t.*

$$\left\|\Theta^* - \hat{\Theta}\right\|_\infty < r$$

and $\Theta^*_{ij} = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Then for positive x

$$\mathbb{P}\left\{\left\|\hat{\Sigma}_Z - \hat{\Sigma}_{\hat{Z}}\right\|_\infty \ge \Delta^{(2)}_{\Sigma_Z}\right\} \le p_s^{\Sigma_{Z2}}(x),$$

where

$$p_s^{\Sigma_{Z2}}(x) \coloneqq p_s^X(x) + p_{\mathcal{Z}_s}(x),$$

$$\Delta^{(2)}_{\Sigma_Z} = \Delta_Z(x)(2\mathcal{Z}_s(\mathrm{x}) + \Delta_Z(x)).$$

Proof. By Lemma 23 with probability at least $1 - p_{\mathcal{Z}_s}(\mathrm{x})$ we have $\|Z_i\|_\infty \le \mathcal{Z}_s(\mathrm{x})$ and in combination with Lemma 25 we obtain $\left\|\overline{\hat{Z}_i}\right\|_\infty \le \mathcal{Z}_s(\mathrm{x}) + \Delta_Z(x)$ with probability at least $1 - p_{\mathcal{Z}_s}(x) - p_s^X(x)$. Now denote

$$\xi_i \coloneqq \overline{Z_i} - \mathbb{E}_{\mathcal{I}_s}\left[\overline{Z_i}\right] \text{ and } \hat{\xi}_i \coloneqq \overline{\hat{Z}_i} - \mathbb{E}_{\mathcal{I}_s}\left[\overline{\hat{Z}_i}\right].$$

And deliver the bound

$$\begin{aligned}\left\|\hat{\Sigma}_Z - \hat{\Sigma}_{\hat{Z}}\right\|_\infty &\le \mathbb{E}_{\mathcal{I}_s}\left[\xi_i(\xi_i - \hat{\xi}_i)^T + (\xi_i - \hat{\xi}_i)\hat{\xi}_i^T\right] \\ &\le \left(\left\|\hat{\xi}_i\right\|_\infty + \|\xi_i\|_\infty\right)\left\|\xi_i - \hat{\xi}_i\right\|_\infty \\ &\le \Delta_Z(x)(2\mathcal{Z}_s(\mathrm{x}) + \Delta_Z(x)).\end{aligned}$$

□

Appendix C

Proofs for Chapter 4

C.1 Proof of the sensitivity result

Proof of Theorem 10. Consider a pair of centered normal vectors

$$\eta := \begin{pmatrix} \eta^1 & \eta^2 & \dots & \eta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_Y^*),$$
$$\zeta := \begin{pmatrix} \zeta^1 & \zeta^2 & \dots & \zeta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \hat{\Sigma}_Y),$$

where

$$\Sigma_Y^* := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^n\right],$$

$$\hat{\Sigma}_Y := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^{n\flat}\right],$$

where vectors $Y_{\cdot j}^n$ and $Y_{\cdot j}^{n\flat}$ are defined in proofs of Lemma 28 and Lemma 30 respectively. Lemma 9 applies here and yields for all positive q

$$\mathbb{P}\left\{\|\zeta^{n_+}\|_\infty \geq \mathrm{q}\right\} \leq 2\left|\mathbb{T}_{n_+}\right| p^2 \exp\left(-\frac{\mathrm{q}^2}{2\left\|\hat{\Sigma}_Y\right\|_\infty}\right),$$

where $\hat{\Sigma}_Y = \mathrm{Var}[\zeta]$ and $\left|\mathbb{T}_{n_+}\right|$ is the number of central points for window of size n_+. Applying Lemma 36 on a set of probability at least $1 - p_s^\Omega(t, \mathrm{x}) - p_s^{\mathcal{W}}(\mathrm{x}) - p^\Sigma(\chi)$ yields $\left\|\Sigma_Y^* - \hat{\Sigma}_Y\right\|_\infty \leq \Delta_Y$, and hence, due to the fact that $\|\Sigma_Y^*\|_\infty = 1$ by construction,

$$\mathbb{P}\left\{\|\zeta^{n_+}\|_\infty \geq \mathrm{q}\right\} \leq 2\left|\mathbb{T}_{n_+}\right| p^2 \exp\left(-\frac{\mathrm{q}^2}{2(1+\Delta_Y)}\right).$$

Due to Lemma 32 and continuity of Gaussian c.d.f.

$$\mathbb{P}^\flat\left\{B_{n_+}^\flat \geq x_{n_+}^\flat(\alpha)\right\} \geq \alpha/\left|\mathfrak{N}\right| - 2R_{B^\flat}$$

and due to Lemma 32 along with the fact that $\left|\mathbb{T}_{n_+}\right| < N$, choosing q as proposed by equation (4.11) we ensure that $x_{n_+}^\flat(\alpha) \leq \mathrm{q}$.

Now using Lemma 20 twice for $\hat{\Sigma}_n^l(\tau)$ and $\hat{\Sigma}_n^r(\tau)$ respectively we obtain that with probability at least $1 - 2p^\Sigma(\chi)$

$$B_n \geq \sqrt{\frac{n}{2}} \|S\|_\infty (\Delta - 2\delta_n(\chi)).$$

Finally, we notice that due to definition (4.10) of n_{suff} and since $n_+ > n_{suff}$

$$B_{n_+} > \mathrm{q}$$

and therefore, $\mathbb{H}_0$ will be rejected. □

C.2 Proof of bootstrap validity result

Proof of Theorem 9. The proof consists in applying Lemma 28, Lemma 32 and Lemma 27 justifying applicability of sandwiching Lemma 10 on a set of probability not less than q (defined by (4.5)) which are followed by applying Lemma 7. □

C.3 Similarity of joint distributions of $\{B_n\}_{n\in\mathfrak{N}}$ and $\{B_n^\flat\}_{n\in\mathfrak{N}}$

Lemma 27. *Let Assumption 2 hold and $\Delta_Y < 1/2$ where Δ_Y comes from Lemma 36. Also let $X_1, X_2, ..., X_N$ be i.i.d. Then for all positive* x*, t and χ on a set of probability at least $1 - p_{\mathcal{Z}_s}(\kappa) - p_s^\Omega(t, \mathrm{x}) - p_s^\mathcal{W}(\mathrm{x}) - p^\Sigma(\chi)$*

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}\{\forall n \in \mathfrak{N} : B_n \leq x_n\} - \mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : B_n^\flat \leq x_n\right\}\right| \leq R \tag{C.1}$$

where

$$R := R_B + R_{B^\flat} + R_\Sigma^\pm$$

$$R_\Sigma^\pm := C\Delta_Y^{1/3} \log^{2/3}\left(Tp^2\right)$$

$p_{\mathcal{Z}_s}(\kappa)$, $p_s^\Omega(t, \mathrm{x})$, $p_s^\mathcal{W}(\mathrm{x})$ and $p^\Sigma(\chi)$, come from Lemma 31, Lemma 35, Lemma 33 and Lemma 20 respectively, R_B and $R_{B^\flat}$ are defined in Lemma 28 and Lemma 32 respectively and C is an independent constant.

Proof. Consider a pair of normal vectors η and ζ

$$\eta := \begin{pmatrix} \eta^1 & \eta^2 & ... & \eta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_Y^*),$$

$$\zeta := \begin{pmatrix} \zeta^1 & \zeta^2 & ... & \zeta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \hat{\Sigma}_Y),$$

where

$$\Sigma_Y^* := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^n\right],$$

$$\hat{\Sigma}_Y := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^{n\flat}\right],$$

where vectors $Y_{\cdot j}^n$ and $Y_{\cdot j}^{n\flat}$ are defined in proofs of Lemma 28 and Lemma 32 respectively. Applying Lemma 44 along with Lemma 37 yields

$$\sup_{A \in A^{re}} |\mathbb{P}\{\eta \in A\} - \mathbb{P}\{\zeta \in A\}| \leq C \Delta_Y^{1/3} \log^{2/3}\left(Tp^2\right)$$

and the fact that $\forall k \in 1..p : (\mathrm{Var}[\zeta])_{kk} = 1$ provides independence of the constant C. Here A^{re} denotes a set of hyperrectangles in the sense of Definition 2 and clearly for all $\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}$ the set $\{\forall n \in \mathfrak{N} : B_n < x_n\}$ is a hyperrectangle. Subsequently applying Lemma 28 and Lemma 32 we finalize the proof. □

C.4 Gaussian approximation result for B_n

Lemma 28. *Let Assumption 2 hold. Then*

$$\sup_{\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}} |\mathbb{P}\{\forall n \in \mathfrak{N} : B_n \leq x_n\} - \mathbb{P}\{\forall n \in \mathfrak{N} : \|\eta^n\|_\infty \leq x_n\}| \leq R_B,$$

where

$$\begin{pmatrix} \eta^1 & \eta^2 & \ldots & \eta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_Y^*),$$

$$\Sigma_Y^* := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y_{\cdot j}^n\right],$$

$$R_B := C_B \left(F \log^7(2p^2 T n_+)\right)^{1/6},$$

$$F := \frac{1}{2n_-}\left(\beta \log 2 \vee \frac{\sqrt{2}}{\sqrt{2}-1}\gamma\right)^2 \vee \frac{1}{2n_+}\left(\frac{n_+}{n_-}\right)^{1/3} (\|S^{-1}\|_\infty M_3)^2 \vee \sqrt{\frac{1}{2n_+ n_-}} (\|S^{-1}\|_\infty M_4)^2$$

with γ defined by (C.6)*, β by* (C.7) *and Y along with its sub-matrices Y^n by* (C.4) *and* (C.3)*. Also, M_3^3 and M_4^4 stand for the third and the fourth maximal centered moments of $(X_1)_k (X_1)_l$ and C_B is an independent constant.*

Proof. First, we define for all $i \in 1..N$

$$Z_i := S^{-1}\left(\overline{X_i X_i^T - \Sigma^*}\right) \tag{C.2}$$

and notice that

$$B_n(t) := \left\| \frac{1}{\sqrt{2n}} \left(\sum_{i \in \mathcal{I}_n^l(t)} Z_i - \sum_{i \in \mathcal{I}_n^r(t)} Z_i \right) \right\|_\infty.$$

Next consider a matrix Y^n with $2n_+$ columns

$$(Y^n)^T := \sqrt{\frac{n_+}{n}} \times$$

$$\begin{pmatrix} Z_1 & O & \dots & O & -Z_{2n_+ +1} & \dots \\ Z_2 & Z_2 & \dots & \dots & \dots & \dots \\ \dots & Z_3 & \dots & \dots & \dots & \dots \\ Z_n & \dots & \dots & \dots & \dots & \dots \\ -Z_{n+1} & Z_{n+1} & \dots & \dots & \dots & \dots \\ -Z_{n+2} & -Z_{n+2} & \dots & \dots & \dots & \dots \\ \dots & -Z_{n+3} & \dots & O & \dots & \dots \\ -Z_{2n} & \dots & \dots & Z_{2n_+ -2n+1} & O & \dots \\ O & -Z_{2n+1} & \dots & Z_{2n_+ -2n+2} & Z_{2n_+ -2n+2} & \dots \\ O & O & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & -Z_{2n_+ -1} & -Z_{2n_+ -1} & \dots \\ O & O & \dots & -Z_{2n_+} & -Z_{2n_+} & \dots \end{pmatrix} \tag{C.3}$$

Clearly, columns of the matrix are independent and

$$B_n = \frac{1}{\sqrt{2n_+}} \sum_{l=0}^{2n_+} (Y^n)_{\cdot l}$$

Next we define a block matrix composed of Y^n matrices:

$$Y := \begin{pmatrix} Y^1 \\ \hline Y^2 \\ \hline \dots \\ \hline Y^{|\mathfrak{N}|} \end{pmatrix}. \tag{C.4}$$

Again, vectors $Y_{\cdot l}$ are independent and for all $\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}$ the set

$$\{\forall n \in \mathfrak{N} : B_n \le x_n\}$$

is a hyperrectangle in the sense of Definition 2.

The rest of the proof consists in applying Lemma 42. Denote

$$G_{n_+} = \sqrt{\frac{n_+}{n_-}} \left(\beta \log 2 \vee \frac{\sqrt{2}}{\sqrt{2}-1} \gamma \right) \vee \left(\frac{n_+}{n_-} \right)^{1/6} M_3 \vee \left(\frac{n_+}{n_-} \right)^{1/4} M_4. \tag{C.5}$$

In the same way as in Lemma 29 one shows that the assumptions of Lemma 15 hold for components of Z_i with

$$\gamma := L^2 \left\| S^{-1} \right\|_\infty, \tag{C.6}$$

$$\beta := L^2 \left\| S^{-1} \right\|_\infty \left\| \Sigma^* \right\|_\infty. \tag{C.7}$$

Therefore condition (D.1) holds with G_{n_+} defined by equation (C.5). In order to see that Assumption D.5.4 is fulfilled with $b = 1$ notice that

$$\frac{1}{2n_+} \sum_{j=1}^{n_+} \mathbb{E}\left[(Y^n_{ij})^2 \right] \ge \min_j \mathrm{Var}\left[(Z_1)_j \right] = 1.$$

Next observe that for any k-th component Z_{ik} of Z_i and a central point t (both determined by j):

$$\begin{aligned}
\frac{1}{2n_+}\sum_{j=1}^{2n_+}\mathbb{E}\left[\left|Y_{ij}^n\right|^3\right] &= \frac{1}{2n_+}\sum_{i\in\mathcal{I}_n^l(t)\cup\mathcal{I}_n^r(t)}\mathbb{E}\left[\left(\sqrt{\frac{n_+}{n}}\left|Z_{ik}\right|\right)^3\right] \\
&= \frac{1}{2n_+}\sum_{i\in\mathcal{I}_n^l(t)\cup\mathcal{I}_n^r(t)}\left(\frac{n_+}{n}\right)^{3/2}\mathbb{E}\left[\left|Z_{ik}\right|^3\right] \\
&= \frac{2n}{2n_+}\left(\frac{n_+}{n}\right)^{3/2}\mathbb{E}\left[\left|Z_{ik}\right|^3\right] \\
&= \sqrt{\frac{n_+}{n}}\mathbb{E}\left[\left|Z_{ik}\right|^3\right] \\
&\le \sqrt{\frac{n_+}{n_-}}\left(\left\|S^{-1}\right\|_\infty M_3\right)^3.
\end{aligned}$$

In the same way:

$$\frac{1}{2n_+}\sum_{i=1}^{N}\mathbb{E}\left[\left|Y_{ij}^n\right|^4\right] \le \frac{n_+}{n_-}\left(\left\|S^{-1}\right\|_\infty M_4\right)^4.$$

Therefore, Assumption D.5.5 holds with B_{n_+}, so Lemma 42 applies here and provides us with the claimed bound. Moreover, C_B depends only on $b = 1$ which implies that the constant C_B is independent. □

Lemma 29. *Under Assumption Assumption 2 it holds for all $i \in 1..N$ and positive κ that*

$$\mathbb{P}\left\{\forall k \in 1..p : |(Z_i)_k| \le \left\|S^{-1}\right\|_\infty L^2\left(\kappa + \log p + \|\Sigma^*\|_\infty\right)\right\} \ge 1 - e^{-\kappa}.$$

Proof. According to the definition (C.2) of Z_i for its arbitrary element $(Z_i)_k$ one obtains for some $l, m \in 1..p$:

$$(Z_i)_k = {S^{-1}}_{kk}\left((X_i)_l(X_i)_m - \Sigma^*_{lm}\right).$$

By sub-Gaussianity Assumption 2 it holds for all positive x that

$$\mathbb{P}\left\{\forall k \in 1..p : |(X_i)_k| \le x\right\} \ge 1 - pe^{-x^2/L^2}.$$

Hence

$$\mathbb{P}\left\{\forall k \in 1..p : |(Z_i)_k| \le \left\|S^{-1}\right\|_\infty\left(x^2 + \|\Sigma^*\|_\infty\right)\right\} \ge 1 - pe^{-x^2/L^2}$$

and finally a change of variables establishes the claim.

□

C.5 Gaussian approximation result for $B_n^\flat$

Denote

$$\Sigma_Y^* := \frac{1}{2n_+}\sum_{j=1}^{2n_+}\mathrm{Var}\left[Y_{\cdot j}\right],$$

$$\hat{\Sigma}_Y := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y^\flat_{\cdot j}\right],$$

where vectors $Y_{\cdot j}$ and $Y^\flat_{\cdot j}$ are defined by (C.3) and (C.8) respectively.

Lemma 30.

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : B^\flat_n \le x_n\right\} - \mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : \|\zeta^n\|_\infty \le x_n\right\}\right| \le \hat{C}_{B^\flat}, \left(F^\flat \log^7(2p^2 T n_+)\right)^{1/6}$$

where $\hat{C}_{B^\flat}$ *depends only on* $\min_{k\in 1..p}(\hat{\Sigma}_Y)_{kk}$*,*

$$\begin{pmatrix} \zeta^1 & \zeta^2 & ... & \zeta^{|\mathfrak{N}|} \end{pmatrix} \sim \mathcal{N}(0, \hat{\Sigma}_Y),$$

$$\hat{\Sigma}_Y := \frac{1}{2n_+} \sum_{j=1}^{2n_+} \mathrm{Var}\left[Y^{n\flat}_{\cdot j}\right],$$

$$F^\flat = \left(\frac{1}{2n_- \log^2 2} \vee \frac{1}{2n_+}\left(\frac{n_+}{n_-}\right)^{1/3} \vee \sqrt{\frac{1}{2n_+ n_-}}\right) \left\|S^{-1}\right\|_\infty^2 (M^\flat)^2$$

$$M^\flat = \max_{i\in\mathcal{I}_s} \left\|\hat{Z}_i\right\|_\infty.$$

Proof. This proof is similar to the proof of Lemma 28.

Consider a matrix which is a bootstrap counterpart of Y^n

$$(Y^{n\flat})^T := \sqrt{\frac{n_+}{n}} \times$$
$$\begin{pmatrix}
Z^\flat_1 & O & ... & O & -Z^\flat_{2n_++1} & ... \\
Z^\flat_2 & Z^\flat_2 & ... & ... & ... & ... \\
... & Z^\flat_3 & ... & ... & ... & ... \\
Z^\flat_n & ... & ... & ... & ... & ... \\
-Z^\flat_{n+1} & Z^\flat_{n+1} & ... & ... & ... & ... \\
-Z^\flat_{n+2} & -Z^\flat_{n+2} & ... & ... & ... & ... \\
... & -Z^\flat_{n+3} & ... & O & ... & ... \\
-Z^\flat_{2n} & ... & ... & Z^\flat_{2n_+-2n+1} & O & ... \\
O & -Z^\flat_{2n+1} & ... & Z^\flat_{2n_+-2n+2} & Z^\flat_{2n_+-2n+2} & ... \\
O & O & ... & ... & ... & ... \\
... & ... & ... & -Z^\flat_{2n_+-1} & -Z^\flat_{2n_+-1} & ... \\
O & O & ... & -Z^\flat_{2n_+} & -Z^\flat_{2n_+} & ...
\end{pmatrix}.$$

Clearly, columns of the matrix are independent and

$$B^\flat_n = \frac{1}{\sqrt{2n_+}} \sum_{l=0}^{2n_+} (Y^{n\flat})_{\cdot l}$$

Next, we define a block matrix composed of $Y^{n\flat}$ matrices:

$$Y^\flat := \begin{pmatrix} Y^{1\flat} \\ \hline Y^{2\flat} \\ \hline ... \\ \hline Y^{|\mathfrak{N}|\flat} \end{pmatrix}. \tag{C.8}$$

Again, vectors $Y_{.l}^{\flat}$ are independent, and for all $\{x_n\}_{n\in\mathfrak{N}} \subset \mathbb{R}$ the set

$$\left\{\forall n \in \mathfrak{N} : B_n^{\flat} < x_n\right\}$$

is a hyperrectangle in the sense of Definition 2. Now notice that

$$\frac{1}{2n_+}\sum_{j=1}^{2n_+} \mathbb{E}\left[|Y_{ij}|^3\right] \le \sqrt{\frac{n_+}{n_-}} \left\|S^{-1}\right\|_\infty^3 \max_{i\in\mathcal{I}_s} \|Z_i\|_\infty ,$$

$$\frac{1}{2n_+}\sum_{j=1}^{2n_+} \mathbb{E}\left[|Y_{ij}|^4\right] \le \frac{n_+}{n_-} \left\|S^{-1}\right\|_\infty^4 \max_{i\in\mathcal{I}_s} \|Z_i\|_\infty .$$

And finally apply Lemma 42. □

Lemma 31. *Under Assumption 2 it holds for all positive κ that*

$$\mathbb{P}\left\{\forall i \in \mathcal{I}_s : \|Z_i\|_\infty \le \mathcal{Z}_s(\kappa)\right\} \ge 1 - p_{\mathcal{Z}_s}(\kappa),$$

where

$$\mathcal{Z}_s(\kappa) := \left\|S^{-1}\right\|_\infty L^2\left(\kappa + \log p + \|\Sigma^*\|_\infty\right),$$

$$p_{\mathcal{Z}_s}(\kappa) := se^{-\kappa}.$$

Proof. The proof consists in applying Lemma 29 and appropriate multiplicity correction. □

Lemma 32. *Let Assumption 2 hold and $\Delta_Y < 1/2$. Then for all positive κ with probability at least $1 - p_{\mathcal{Z}_s}(\kappa)$*

$$\sup_{\{x_n\}_{n\in\mathfrak{N}}\subset\mathbb{R}} \left|\mathbb{P}^{\flat}\left\{\forall n \in \mathfrak{N} : B_n^{\flat} \le x_n\right\} - \mathbb{P}^{\flat}\left\{\forall n \in \mathfrak{N} : \left\|Y^{n\flat}\right\|_\infty \le x_n\right\}\right| \le R_{B^{\flat}},$$

where

$$R_{B^{\flat}} := C_{B^{\flat}}\left(\hat{F}\log^7(2p^2Tn_+)\right)^{1/6},$$

$$\hat{F} := \left(\frac{1}{2n_-\log^2 2} \vee \frac{1}{2n_+}\left(\frac{n_+}{n_-}\right)^{1/3} \vee \sqrt{\frac{1}{2n_+n_-}}\right)\left\|S^{-1}\right\|_\infty^2 \left(\mathcal{Z}_s(\kappa)\right)^2$$

and $\hat{C}_{B^{\flat}}$ is an independent constant.

Proof. The proof consists in subsequently applying Lemma 30 and Lemma 31 ensuring $M^{\flat} \le \mathcal{Z}_s(\kappa)$ with probability at least $1 - p_{\mathcal{Z}_s}(\kappa)$, while assumed bound $\left\|\Sigma_Y^* - \hat{\Sigma}_Y\right\|_\infty \le \Delta_Y < 1/2 = \min_{1\le k\le p}(\Sigma_Y^*)_{kk}$ implies the existence of a deterministic constant $C_{B^{\flat}} > \hat{C}_{B^{\flat}}$. □

C.6 $\Sigma_Y^* \approx \hat{\Sigma}_Y$

Denote

$$W_i := \overline{X_i X_i^T},$$

$$\Omega^* := \mathbb{E}\left[\left(W_1 - \overline{\Sigma^*}\right)\left(W_1 - \overline{\Sigma^*}\right)^T\right],$$

$$\hat{\Omega} := \mathbb{E}_{\mathcal{I}_s}\left[\left(W_i - \overline{\Sigma^*}\right)\left(W_i - \overline{\Sigma^*}\right)^T\right],$$

where notation $\mathbb{E}_{\mathcal{I}_s}[\cdot]$ is used as a shorthand for averaging over $\mathcal{I}_s$, e.g.

$$\mathbb{E}_{\mathcal{I}_s}[\xi_i] = \frac{1}{s}\sum_{i \in \mathcal{I}_s} \xi_i,$$

and similarly $\mathrm{Var}_{\mathcal{I}_S}[\cdot]$ denotes an empirical covariance matrix computed using the same set, e.g.

$$\mathrm{Var}_{\mathcal{I}_S}[\xi_i] = \mathbb{E}_{\mathcal{I}_s}\left[\xi_i \xi_i^T\right].$$

The results of this section rely on the Lemma 20 which is a trivial corollary of Lemma 6 by [27] providing the concentration result for empirical covariance matrix.

Straightforwardly applying Assumption 2 and a proper multiplicity correction yields the following result.

Lemma 33. *Under Assumption 2 it holds for all positive* x *that*

$$\mathbb{P}\left\{\forall i \in \mathcal{I}_s : \left\|W_i - \overline{\Sigma^*}\right\|_\infty \le \mathcal{W}_s(\mathrm{x})\right\} \ge 1 - p_s^{\mathcal{W}}(\mathrm{x}),$$

where

$$\mathcal{W}_s(\mathrm{x}) := \mathrm{x}^2 + \|\Sigma^*\|_\infty,$$

$$p_s^{\mathcal{W}}(\mathrm{x}) := pse^{-\mathrm{x}}.$$

Lemma 34. *Under Assumption 2 with probability at least* $1 - p_{\mathrm{x}}^{\mathcal{W}}(s) - p^{\Sigma}(\chi)$

$$\left\|\mathrm{Var}_{\mathcal{I}_S}\left[W_i - \mathbb{E}_{\mathcal{I}_s}[W_i]\right] - \hat{\Omega}\right\|_\infty \le 2\mathcal{W}_s(\mathrm{x})\delta_s(\chi) + \delta_s(\chi)^2.$$

Proof. By the construction of bootstrap procedure and definition (C.2)

$$\begin{aligned}
\mathrm{Var}_{\mathcal{I}_S}\left[W_i - \mathbb{E}_{\mathcal{I}_s}[W_i]\right] &= \frac{1}{s}\sum_{i \in \mathcal{I}_s}\left(W_i - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)\left(W_i - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)^T \\
&= \frac{1}{s}\sum_{i \in \mathcal{I}_s}\left(W_i - \overline{\Sigma^*} + \overline{\Sigma^*} - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)\left(W_i - \overline{\Sigma^*} + \overline{\Sigma^*} - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)^T \\
&= \hat{\Omega} + \frac{1}{s}\sum_{i \in \mathcal{I}_s}\left(\overline{\Sigma^*} - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)\left(\overline{\Sigma^*} - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)^T \\
&\quad + \frac{2}{s}\left(W_i - \overline{\Sigma^*}\right)\left(\overline{\Sigma^*} - \mathbb{E}_{\mathcal{I}_s}[W_i]\right)^T.
\end{aligned}$$

Applying Lemmas Lemma 33 and Lemma 20 yields the claim. □

Lemma 35. *Let Assumption 2 hold for some $L > 0$. Then for any positive t and* x

$$\mathbb{P}\left\{\left\|\Omega^* - \hat{\Omega}\right\|_\infty \geq \Delta_s^\Omega(t, \mathrm{x})\right\} \leq p_s^\Omega(t, \mathrm{x}),$$

where

$$\Delta_s^\Omega(t, \mathrm{x}) := \frac{(2(\mathcal{W}_s(\mathrm{x}))^2)\, t}{3s}\left(1 + \sqrt{1 + \frac{9s\sigma_\Omega^2}{t\,(2(\mathcal{W}_s(\mathrm{x}))^2)^2}}\right),$$

$$p_s^\Omega(t, \mathrm{x}) := p^2 e^{-t} + p_s^{\mathcal{W}}(\mathrm{x}).$$

Proof. Consider a random variable

$$\zeta_{lm}^i := \left(W_i - \overline{\Sigma^*}\right)_l \left(W_i - \overline{\Sigma^*}\right)_m - \Omega_{lm}^*.$$

By Lemma 33 we can bound it as $|\zeta_{lm}^i| \leq 2(\mathcal{W}_s(\mathrm{x}))^2$ with probability at least $1 - p_s^{\mathcal{W}}(\mathrm{x})$. Due to ζ_{lm}^i being centered Bernstein inequality applies:

$$\mathbb{P}\left\{\mathbb{E}_{\mathcal{I}_s}\left[\zeta_{lm}^i\right] \geq \frac{(2(\mathcal{W}_s(\mathrm{x}))^2)\, t}{3s}\left(1 + \sqrt{1 + \frac{9s\sigma_\Omega^2}{t\,(2(\mathcal{W}_s(\mathrm{x}))^2)^2}}\right)\right\} \leq e^{-t}.$$

□

Lemma 36. *Under Assumption 2 for any positive t,* x *and χ with probability at least $1 - p_s^\Omega(t, \mathrm{x}) - p_s^{\mathcal{W}}(\mathrm{x}) - p^\Sigma(\chi)$ it holds that*

$$\left\|\mathrm{Var}\left[Z_i\right] - \mathrm{Var}\left[Z_i^\flat\right]\right\|_\infty \leq \Delta_Y := \left\|S^{-1}\right\|_\infty^2 \left(\Delta_s^\Omega(t, \mathrm{x}) + 2\mathcal{W}_s(\mathrm{x})\delta_s(\chi) + \delta_s^2(\chi)\right).$$

Proof. Proof consists in applying Lemma 35 and Lemma 34. □

Using the fact that the covariance matrices Σ_Y^* and $\hat{\Sigma}_Y$ are block matrices composed of blocks $\mathrm{Var}\left[Z_i\right]$ and $\mathrm{Var}\left[Z_i^\flat\right]$ respectively, multiplied by some positive values ≤ 1, we trivially obtain the following result.

Lemma 37. *Under Assumption 2 for any positive t,* x *and χ with probability at least $1 - p_s^\Omega(t, \mathrm{x}) - p_s^{\mathcal{W}}(\mathrm{x}) - p^\Sigma(\chi)$ it holds that*

$$\|\Sigma_Y^* - \Sigma_Y^*\|_\infty \leq \Delta_Y,$$

Δ_Y comes from Lemma Lemma 36.

Appendix D

Known results

D.1 Consistency result for the ℓ_1-penalized estimator by [42]

Lemma 38 (Theorem 1, [42]). *Consider a distribution satisfying Assumption 1 with some* $\alpha \in (0, 1]$, *let* $\hat{\Theta}$ *be a solution of the optimization problem* (2.1) *with tuning parameters* $\Lambda_{ij} = \lambda_n = \frac{8}{\alpha}\delta_n$ *for* $i \neq j$. *Furthermore, suppose the following sparsity assumption:*

$$d \leq \frac{\delta_n}{6(\delta_n + \lambda_n)^2 \max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa_{\Gamma^*}^2\kappa_{\Sigma^*}^3\}}.$$

Also assume that

$$\theta_{min} > r := 2\kappa_{\Gamma^*}(\delta_n + \lambda_n)$$

Then on the set $\mathcal{T} = \left\{\left\|\hat{\Sigma} - \Sigma^*\right\|_\infty < \delta_n\right\}$ *the following holds:* $\left\|\hat{\Theta} - \Theta^*\right\|_\infty \leq r$ *and* $\Theta^*_{ij} = 0 \Leftrightarrow \hat{\Theta}_{ij} = 0$.

D.2 The bound for $R(\Delta)$ by [42]

Lemma 39 (Lemma 5, [42]). *Suppose,* $\|\Delta\|_\infty \leq \frac{1}{3\kappa_{\Sigma^*}d}$. *Then the matrix* $J := \sum_{k=0}^{\infty}(-1)^k({\Theta^*}^{-1}\Delta)^k$ *satisfies the bound* $\|J\|_\infty \leq 3/2$ *and the matrix*

$$R(\Delta) = {\Theta^*}^{-1}\Delta{\Theta^*}^{-1}\Delta J{\Theta^*}^{-1}$$

is bounded as

$$\|R(\Delta)\|_\infty \leq \frac{3}{2}d\,\|\Delta\|_\infty^2\,\kappa_{\Sigma^*}^3$$

.

D.3 The estimation $\hat{\sigma}_{ij}^2$ for σ_{ij}^2

Lemma 40 (generalization of Lemma 2 by [27]). *Assume conditions of Lemma 5 hold. Moreover, let* $X_i \sim \mathcal{N}(0, \Sigma^*)$. *Define the estimator* $\hat{\sigma}_{ij}^2$ *as*

$$\hat{\sigma}_{ij}^2 := \hat{\Theta}_{ii}\hat{\Theta}_{jj} + \hat{\Theta}_{ij}^2.$$

Then on set $\mathcal{T} = \left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty < \delta_n \right\}$

$$\left| \hat{\sigma}_{ij}^2 - \sigma_{ij}^2 \right| \le 2 r_\Lambda (2\nu_{\Theta^*} + r_\Lambda),$$

where $\nu_{\Theta^*} = \|\Theta^*\|_\infty$.

Proof. Since $X_i \sim \mathcal{N}(0, \Sigma^*)$, clearly $\Theta^* X \sim \mathcal{N}(0, \Theta^*)$. Some algebra yields

$$\sigma_{ij}^2 = \Theta_{ii}^* \Theta_{jj}^* + \Theta_{ij}^{*2}.$$

Therefore

$$\left| \hat{\sigma}_{ij}^2 - \sigma_{ij}^2 \right| \le \left| \hat{\Theta}_{ii} \hat{\Theta}_{jj} - \Theta_{ii}^* \Theta_{jj}^* \right| + \left| \hat{\Theta}_{ij}^2 - \Theta_{ij}^{*2} \right|.$$

Now using the bound provided by Lemma 5 we can bound the terms on the right hand

$$\left| \hat{\Theta}_{ii} \hat{\Theta}_{jj} - \Theta_{ii}^* \Theta_{jj}^* \right| \le (\Theta_{ii}^* + \Theta_{jj}^*) r_\Lambda + r_\Lambda^2,$$

$$\begin{aligned} \left| \hat{\Theta}_{ij}^2 - \Theta_{ij}^{*2} \right| &= \left| (\hat{\Theta}_{ij} - \Theta_{ij}^*)(\hat{\Theta}_{ij} + \Theta_{ij}^*) \right| \\ &\le r_\Lambda (2\Theta_{ij}^* + r_\Lambda). \end{aligned}$$

And finally

$$\begin{aligned} \left| \hat{\sigma}_{ij}^2 - \sigma_{ij}^2 \right| &\le r_\Lambda (2\nu_{\Theta^*} + r_\Lambda) + 2\nu_{\Theta^*} r_\Lambda + r_\Lambda^2 \\ &= 2 r_\Lambda (2\nu_{\Theta^*} + r_\Lambda). \end{aligned}$$

□

D.4 Probability of the set $\mathcal{T}$

Assumption D.4.3 (Sub-Gaussianity condition)**.** *Denote the normalized components of the vector* X_1 *as* $\xi_i = \frac{X_{1i}}{\sqrt{\Sigma_{ii}^*}}$. *Then, we say that the Sub-Gaussianity condition holds for vector* X_1 *if*

$$\exists K > 0 : \forall i \ \mathbb{E}\left[\exp\left(\frac{\xi_i^2}{K^2} \right) \right] \le 2.$$

Lemma 41 (by [42] in form by [27])**.** *Let Assumption D.4.3 hold for some* $K > 0$. *Then for*

$$\delta(n, r) = 8(1 + 12K^2) \max_i \Sigma_{ii}^* \sqrt{2 \frac{\log(4r)}{n}}$$

and for any $\gamma > 2$ *and for* n *such that* $\delta(n, p^\gamma) < 8(1 + 12K^2) \max_i \Sigma_{ii}^*$ *we have*

$$\mathbb{P}\left\{ \left\| \hat{\Sigma} - \Sigma^* \right\|_\infty \le \delta(n, p^\gamma) \right\} \ge 1 - \frac{1}{p^{\gamma - 2}}.$$

D.5 Gaussian approximation result

In this section we briefly describe the result obtained in [12].

Throughout this section consider an independent sample $x_1, ..., x_n \in \mathbb{R}^p$ of centered random variables. Define their Gaussian counterparts $y_i \sim \mathcal{N}(0, \mathrm{Var}[x_i])$ and denote their scaled sums as

$$S_n^X := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i,$$

$$S_n^Y := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i.$$

Definition 2. *We call a set A of a form $A = \{w \in \mathbb{R}^p : a_i \le w_i \le b_i \; \forall i \in \{1..p\}\}$ a hyperrectangle. A family of all hyperrectangles is denoted as A^{re}.*

Assumption D.5.4. $\exists b > 0$ *such that*

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[x_{ij}^2\right] \ge b \text{ for all } j \in \{1..p\}.$$

Assumption D.5.5. $\exists G_n \ge 1$ *such that*

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[|x_{ij}|^{2+k}\right] \le G_n^{2+k} \text{ for all } j \in \{1..p\} \text{ and } k \in \{1, 2\},$$

$$\mathbb{E}\left[\exp\left(\frac{|x_{ij}|}{G_n}\right)\right] \le 2 \text{ for all } j \in \{1..p\} \text{ and } i \in 1..n. \tag{D.1}$$

Lemma 42 (Proposition 2.1 by [12])**.** *Let Assumption D.5.4 hold for some b and Assumption D.5.5 hold for some G_n. Then*

$$\sup_{A \in A^{re}} \left|\mathbb{P}\left\{S_n^X \in A\right\} - \mathbb{P}\left\{S_n^Y \in A\right\}\right| \le C \left(\frac{G_n^2 \log^7(pn)}{n}\right)^{1/6}$$

and C depends only on b.

D.6 Anti-concentration result

Lemma 43 (Nazarov's inequality [38])**.** *Consider a normal p-dimensional vector $X \sim \mathcal{N}(0, \Sigma)$ and let $\forall i : \Sigma_{ii} = 1$. Then for any $y \in \mathbb{R}^p$ and any positive a*

$$\mathbb{P}\{X \le y + a\} - \mathbb{P}\{X \le y\} \le Ca\sqrt{\log p},$$

where C is an independent constant.

D.7 Gaussian comparison result

By the technique given in the proof of Theorem 4.1 by [12] one obtains the following generalization of the result given in [10]

Lemma 44. *Consider a pair of covariance matrices* Σ_1 *and* Σ_2 *of size* $p \times p$ *such that*

$$\|\Sigma_1 - \Sigma_2\|_\infty \le \Delta$$

and $\forall k : C_1 \ge \Sigma_{1,kk} \ge c_1 > 0$. *Then for random vectors* $\eta \sim \mathcal{N}(0, \Sigma_1)$ *and* $\zeta \sim \mathcal{N}(0, \Sigma_2)$ *it holds that*

$$\sup_{A \in A^{re}} |\mathbb{P}\{\eta \in A\} - \mathbb{P}\{\zeta \in A\}| \le C\Delta^{1/3} \log^{2/3} p,$$

where C *is a positive constant which depends only on* C_1 *and* c_1.

D.8 Tail inequality for quadratic forms

The following result is a direct corollary of Theorem 1 in [25]

Lemma 45. *Consider a positive semi-definite or negative semi-definite matrix* B *and suppose Assumption 2 holds. Then for all* $t > 0$

$$\mathbb{P}\left\{\left|X_1^T B X_1\right| \ge 3L^2 \left(|trB| + 2\sqrt{tr(B^2)t} + 2\left|\Lambda\left(B\right)\right| t\right)\right\} \le e^{-t}.$$

Bibliography

[1] A. Aue and L. Horvath. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013.

[2] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 12 2009.

[3] V. Avanesov. Structural break analysis in high-dimensional covariance structure. *NIPS, Under review*, 2017.

[4] V. Avanesov and N. Buzun. Change-point detection in high-dimensional covariance structure. *ArXiv e-prints*, Oct. 2016.

[5] V. Avanesov, J. Polzehl, and K. Tabelow. Consistency results and confidence intervals for adaptive l1-penalized estimators of the high-dimensional sparse precision matrix. Technical Report 2229, WIAS, 2016.

[6] D. S. Bassett, N. F. Wymbs, M. a. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641, 2010.

[7] L. Bauwens, S. Laurent, and J. V. K. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, jan 2006.

[8] G. Biau, K. Bleakley, and D. M. Mason. Long signal change-point detection. *Electron. J. Statist.*, 10(2):2097–2123, 2016.

[9] N. Buzsun and V. Avanesov. Bootstrap for change point detection. *Manuscript*, 2017.

[10] V. Chernozhukov, D. Chetverikov, and K. Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. Dec 2013.

[11] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819, 12 2013.

[12] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. Dec 2014.

[13] H. Cho. Change-point detection in panel data via double cusum statistic. *Electron. J. Statist.*, 10(2):2000–2038, 2016.

[14] H. Cho and P. Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society Series B*, 77(2):475–507, 2015.

[15] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[16] M. Şerban, A. Brockwell, J. Lehoczky, and S. Srivastava. Modelling the dynamic dependence structure in multivariate financial time series. *Journal of Time Series Analysis*, 28(5):763–782, 2007.

[17] M. Csörgő and L. Horváth. *Limit theorems in change-point analysis*. Wiley series in probability and statistics. J. Wiley & Sons, Chichester, New York, 1997.

[18] R. F. Engle, V. K. Ng, and M. Rothschild. Asset pricing with a factor-arch covariance structure. Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2):213–237, 1990.

[19] J. Fan, Y. Feng, and Y. Wu. Network exploration via the adaptive lasso and scad penalties. *Ann. Appl. Stat.*, 3(2):521–541, 2009.

[20] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties, 2001.

[21] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nat. Neurosci.*, 18:1664–1671, 2015.

[22] J. Friedman, T. Hastie, and R. Tibshirani. *glasso: Graphical lasso- estimation of Gaussian graphical models*, 2014. R package version 1.8.

[23] K. J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011.

[24] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.

[25] D. Hsu, S. M. Kakade, and T. Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(0), Jan. 2012.

[26] N. A. James and D. S. Matteson. ecp: An R package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software*, 62(7):1–25, 2014.

[27] J. Janková and S. van de Geer. Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Statist.*, 9(1):1205–1229, 2015.

[28] J. Janková and S. van de Geer. Honest confidence regions and optimality in high-dimensional precision matrix estimation, 2015.

[29] M. Jirak. Uniform change point tests in high dimension. *Ann. Statist.*, 43(6):2451–2483, 12 2015.

[30] S. Kim. *ppcor: Partial and Semi-partial (Part) correlation*, 2012. R package version 1.0.

[31] Korolev and I. G. Shevtsova. On the upper bound for the absolute constant in the berry-esseen inequality. *Theory Probab. Appl.*, 54(4):638–658, 2010.

[32] C. Lam and J. Fan. Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.*, 37(6B):4254–4278, 12 2009.

[33] M. Lavielle and G. Teyssière. Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, 46(3):287–306, 2006.

[34] J. Li and S. X. Chen. Two sample tests for high-dimensional covariance matrices. *Ann. Statist.*, 40(2):908–940, 04 2012.

[35] D. S. Matteson and N. A. James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

[36] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 06 2006.

[37] T. Mikosch, S. Johansen, and E. Zivot. Handbook of Financial Time Series. *Time*, 468(1996):671–693, 2009.

[38] F. Nazarov. *On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure*, pages 169–187. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[39] R. A. Poldrack, J. A. Mumford, and T. E. Nichols. *Handbook of functional MRI data analysis.* Cambridge University Press, 2011.

[40] S. Puschmann, A. Brechmann, and C. M. Thiel. Learning-dependent plasticity in human auditory cortex during appetitive operant conditioning. *Hum. Brain Mapp.*, 34(11):2841–2851, 2013.

[41] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2016.

[42] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence. *Electron. J. Statist.*, 5:935–980, 2011.

[43] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence. *Electron. J. Statist.*, 5:935–980, 2011.

[44] A. Shiryaev. *Optimal Stopping Rules.* Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2007.

[45] V. Spokoiny and N. Willrich. Bootstrap tuning in ordered model selection. *ArXiv e-prints*, July 2015.

[46] O. Sporns. *Networks of the brain.* The MIT Press, 2011.

[47] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[48] S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.*, 42(3):1166–1202, 2014.

[49] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S.* Springer, New York, 4th edition, 2002.

[50] B. Whitcher, V. J. Schmid, and A. Thornton. Working with the DICOM and NIfTI data standards in R. *J. Stat. Softw.*, 44(6):1–28, 2011.

[51] Y. Xie and D. Siegmund. Sequential multi-sensor change-point detection. *Ann. Statist.*, 41(2):670–692, 04 2013.

[52] C. Zou, G. Yin, L. Feng, and Z. Wang. Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.*, 42(3):970–1002, 06 2014.

[53] H. Zou. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101:1418–1429, 2006.

[54] H. Zou and R. Li. One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.*, 36(4):1509–1533, 2008.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 7.06.2017 Valeriy Avanesov.