Change-point detection in high-dimensional covariance structure

Valeriy Avanesov and Nazar Buzun

WIAS
Mohrenstr. 39
10117 Berlin
Germany
e-mail: avanesov@wias-berlin.de; buzun@wias-berlin.de

Abstract: In this paper we introduce a novel approach for an important problem of break detection. Specifically, we are interested in detection of an abrupt change in the covariance structure of a high-dimensional random process – a problem, which has applications in many areas e.g., neuroimaging and finance. The developed approach is essentially a testing procedure involving a choice of a critical level. To that end a non-standard bootstrap scheme is proposed and theoretically justified under mild assumptions. Theoretical study features a result providing guaranties for break detection. All the theoretical results are established in a high-dimensional setting (dimensionality $p \gg n$). Multiscale nature of the approach allows for a trade-off between sensitivity of break detection and localization. The approach can be naturally employed in an on-line setting. Simulation study demonstrates that the approach matches the nominal level of false alarm probability and exhibits high power, outperforming a recent approach.

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 necessitate 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 [14] [2]. However, in recent years the interest for multi-dimensional approaches has increased. Most of them cover the case of fixed dimensionality [31] [29] [1] [41] [42]. Some approaches [11, 28, 12] 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. [31]). 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) of the statistic $A(t)$ defined in Section 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 [14] [2] [1] [28] [7] [42]. As an alternative, permutation [28] [31] or parametric bootstrap may be used [28]. Clearly, it seems attractive to choose the threshold in a solely data-driven way as it is suggested in the recent paper [11], 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 inverse covariance matrix – which is motivated by applications to neuroimaging. 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 [40] which refer to the explicit influence among neural systems [22]. A rather popular approach to inferencing these networks is based on estimating inverse covariance or precision matrices [21]. The technique generally makes use of the observation that functional connectivity networks are of small-world type [40], which makes sparsity assumptions feasible. Analysing 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 [4].

A similar problem is found 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 [13, 6, 16, 33].

One approach to the change-point localization is developed in [29], the corresponding significance testing problem is considered in [1]. However, neither of these papers address the high-dimensional case.

A widely used break detection approach (named CUSUM) [12, 1, 28] 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 a method ideologically similar to MOSUM [5] [15]: 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 2 for formal definition). Window size n is an important parameter and its choice is case-specific (see Section 4 for theoretical treatment of this issue). Using a small window results in high variability and low sensitivity, while a 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 method 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. Moreover, the local nature
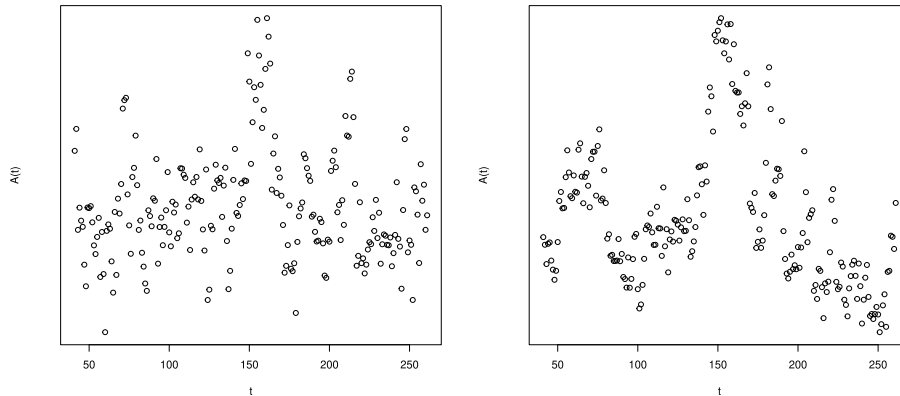
FIGURE 1. *Plots of test statstics $A(t)$ computed on synthetically generated data without (left) and with a single change-point at $t = 150$ (right). The ticks and their corresponding values are intentionally hidden for the vertical axis, since we aim to draw a conclusion about a brake based only on a single time series, not a collection of them. Clearly, the choice of a threshold is not obvious.*

of the proposed statistic allows for detection of multiple change points if the change-points are not too close to each other (see Section 4).

The contribution of our study is the development of a novel break detection approach which is

- high-dimensional, allowing for up to exponential growth of the dimensionality with the window size
- suitable for on-line setting
- suitable for detection of multiple change-points
- 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

We consider the following setup. Let $X_1, \dots X_N \in \mathbb{R}^p$ denote sample of independent zero-mean vectors (the on-line setting is discussed in Section 3) and we want to test a hypothesis

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

versus an alternative suggesting the existence of a break:

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

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

The approach proposed in the paper focuses on applications in neuroimaging. Independence of the vectors X_i is fulfilled only approximately in practice but functional connectivity network analysis, which assumes temporal independence, has been proven to be very successful and validated [35].

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 2.1 which is followed by Section 2.2 describing a data-driven (bootstrap) calibration scheme and Section 2.3 describing change-point localization procedure. The theoretical part of the paper justifies the proposed procedure in a high-dimensional setting. The result justifying the validity of the proposed calibration scheme is stated in Section 3. Section 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 5) demonstrating conservativeness of the proposed test and higher sensitivity compared to the other algorithms and by analysis of real-world datasets (Section 6). Appendix A contains a finite-sample version of sensitivity result along with the proofs. Appendix B provides a a finite-sample version of bootstrap sensitivity result which is followed by the proofs. Finally, Appendix H lists results which were essential for our theoretical study.

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.

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 [26] [27] as

$$\hat{T}_n^{\mathfrak{S}}(t) := \hat{\Theta}_n^{\mathfrak{S}}(t) + \hat{\Theta}_n^{\mathfrak{S}}(t)^T - \hat{\Theta}_n^{\mathfrak{S}}(t)^T \hat{\Sigma}_n^{\mathfrak{S}}(t) \hat{\Theta}_n^{\mathfrak{S}}(t) \qquad (2.1)$$

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 [37] or node-wise procedure [27] (see Definition 3.1 and Appendix H.5 for details). Note, the symmetricity of $\hat{\Theta}_n^{\mathfrak{S}}(t)$ is not required, yet $\hat{T}_n^{\mathfrak{S}}(t)$ is symmetric by construction.

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

$$Z_{i,uv} := \Theta_u^* X_i \Theta_v^* X_i - \Theta_{uv}^* \qquad (2.2)$$

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 3.1 for details) of the precision matrix Θ^* obtained based on the sub-sample $\{X_i\}_{i \in \mathcal{I}_s}$ as $\hat{\Theta}$. 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 [26] [3]. If the node-wise procedure is employed, the uniform consistency of an empirical estimate of σ_{uv}^2 has been shown under some mild assumptions (not including normality) [27].

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

$$A_n(t) := \left|\left|\sqrt{\frac{n}{2}} S^{-1} \overline{(\hat{T}_n^l(t) - \hat{T}_n^r(t))}\right|\right|_\infty \qquad (2.3)$$

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 $\mathbb{H}_0$

$$\sqrt{n}(\hat{T}_n^{\mathfrak{S}}(t) - \Theta^*) = \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}_n^{\mathfrak{S}}(t)} Z_i + r_n^{\mathfrak{S}}(t)\sqrt{n}, \qquad (2.4)$$

where the residual term

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

can be controlled under mild assumptions [26] [27] [3] as

$$\max_{\substack{\mathfrak{S} \in \{l,r\} \\ n \in \mathfrak{N}, t \in \mathbb{T}_n}} \left\| r_n^{\mathfrak{S}}(t) \right\|_\infty = O_P \left(\frac{d \log p}{n} \right).$$

The main reason why we prefer to use de-sparsified estimators $\hat{T}_n^{\mathfrak{S}}(t)$ over using ℓ_1-penalized estimators $\hat{\Theta}_n^{\mathfrak{S}}(t)$ is that the former allows for the expansion (2.4) which pitches an idea behind the bootstrap procedure we suggest and makes the theoretical analysis possible.

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 2.1. *A different test statistic can be defined as the maximum distance between elements of empirical covariance matrices $\hat{\Sigma}_n^l(t)$ and $\hat{\Sigma}_n^r(t)$. Such a method would be computationally less burdensome, since it does not involve precision matrix estimation. However, there are extremely effective implementations of the latter which makes the computational costs of calibration procedure dominate. In turn, the calibration complexity is the same for both of the approaches, since matrix inversion is not involved therein (see Section 2.2). Therefore, the computational gain would be negligible. Furthermore, 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 (2.3) increases the sensitivity to this kind of alternative (see Remark 4.1 for more details).*

Remark 2.2. *The estimator $\hat{T}_n^{\mathfrak{S}}(t)$ is indeed a de-biased estimator. We can easily rearrange the definition (2.1) as*

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

in order to represent it as a difference of ℓ_1-penalized estimator and bias-correcting term.

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 $x_n^\flat(\alpha)$ or formally if $\exists n \in \mathfrak{N} : A_n > x_n^\flat(\alpha)$.

In order to properly choose the thresholds, we define bootstrap statistics $A_n^\flat$ 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 ($|\mathcal{I}_s| \le p$) 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

$$X_i^\flat = X_{\varkappa_i} - \mathbb{E}_s\left[X_j\right]$$

where $\mathbb{E}_s\left[\cdot\right]$ stands for averaging over values of index belonging to $\mathcal{I}_s$ e.g., $\mathbb{E}_s\left[X_j\right] = \frac{1}{|\mathcal{I}_s|}\sum_{j\in\mathcal{I}_s}X_j$. Denote the measure $X_i^\flat$ are distributed with respect to as $\mathbb{P}^\flat$. In accordance with (2.2) define

$$Z_{i,uv}^\flat := \hat{\Theta}_u X_i^\flat \hat{\Theta}_v X_i^\flat - \hat{\Theta}_{uv}$$

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 point t define a bootstrap counterpart of $A_n(t)$

$$A_n^\flat(t) := \left\|\frac{1}{\sqrt{2n}}S^{-1}\overline{\left(\sum_{i\in\mathcal{I}_n^l(t)}Z_i^\flat - \sum_{i\in\mathcal{I}_n^r(t)}Z_i^\flat\right)}\right\|_\infty \tag{2.5}$$

which is intuitively reasonable due to expansion (2.4). And finally we define the bootstrap counterpart of A_n as

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

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

$$z_n^\flat(\mathrm{x}) := \inf\left\{z : \mathbb{P}^\flat\left\{A_n^\flat > z\right\} \le \mathrm{x}\right\}. \tag{2.6}$$

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

$$\alpha^* := \sup\left\{\mathrm{x} : \mathbb{P}^\flat\left\{\exists n\in\mathfrak{N} : A_n^\flat > z_n^\flat(\mathrm{x})\right\} \le \alpha\right\} \tag{2.7}$$

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

Remark 2.3. *In order to detect multiple breaks we suggest to repeat the calibration procedure after each detected break using a portion of data acquired after the break.*

Remark 2.4. *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 3.1 and Theorem 4.1, since it effectively means $s = N$. However, in practise such a decision might lead to reduction of sensitivity due to overestimation of the thresholds.*

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 $\hat{n}$:

$$\hat{n} := \min \left\{ n \in \mathfrak{N} : A_n > x_n^\flat(\alpha) \right\} \qquad (2.8)$$

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

$$\hat{\tau} := \min \left\{ t \in \mathbb{T}_{\hat{n}} : A_{\hat{n}}(t) > x_{\hat{n}}^\flat(\alpha) \right\}. \qquad (2.9)$$

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. 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 3.1 (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 [37], adaptive graphical lasso [43] and thresholded de-sparsified estimator based on node-wise procedure [27]. 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 [37, 3, 27] or formally

Definition 3.1. *Consider an i.i.d. sample $x_1, x_2, ...x_n \in \mathbb{R}^p$. Denote their precision matrix as $\Theta^* = \mathbb{E}\left[X_1 X_1^T\right]^{-1}$. 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*

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

and

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

Graphical lasso and its adaptive versions impose an assumption, common for ℓ_1-penalized approaches.

Assumption 3.2 (Irrepresentability condition)**.** *Denote an active set*

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

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

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

The interpretation of irrepresentability condition under normality assumption is given in [26] [37]. Particularly, Assumption 3.2 requires low correlation between the elements of empirical covariance matrix from the active set $\mathcal{S}$ and from its complement. The higher the constant ψ is, the stricter upper bound is assumed.

While the Assumption 3.2 allows for the recovery of the active set, the test statistic does not explicitly require it, since it is based on the de-sparsified estimators. But this assumption is still essential for demonstrating consistency of estimation of the non-zero elements $\Theta_{\mathcal{S}}^*$ by either graphical lasso or its adaptive versions. Alternatively, one can use thresholded de-sparsified estimator, for which the theoretical guaranties can be established in the absence of Assumption 3.2.

These observations give rise to the two following assumptions.

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 3.2.*

Assumption 3.3.B. *Suppose, thresholded de-sparsified estimator based on nodewise 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 3.1. *Assume $\mathbb{H}_0$ holds and furthermore, let $X_1, X_2, ...X_N \in \mathbb{R}^p$ be i.i.d. Let Assumption 3.1 and either Assumption 3.3.A or Assumption 3.3.B hold. Also assume, the spectrum of Θ^* is bounded. Allow the maximal number*

d of non-zero elements in a row of the matrix Θ^*, *the size s of the set* $\mathcal{I}_s$, *the dimensionality p, the number* $|\mathfrak{N}|$ *of window sizes being considered, the maximum and minimum window sizes* n_+ *and* n_- *grow with the sample size* N. *Further let* $N > 2n_+$, $n_+ \geq n_-$ *and also impose the sparsity assumption*

$$d = o\left(\frac{\sqrt[4]{\min\{s, n_-^2\}}}{|\mathfrak{N}|^3 \log^{10}(pN)}\right).$$

Then

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

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

Bootstrap validity result discussion Theorem 3.1 guarantees under mild assumptions (Assumption 3.2 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 [38]). 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 3.1 guarantees the capability of the proposed bootstrap scheme to control the aforementioned probability to raise a false alarm.

Proof discussion The proof of the bootstrap validity result, presented in Appendix B, mostly relies on the high-dimensional central limit theorems obtained in [10], [9]. 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 C.1), initially used in [39]. The high-level structure of the proof can be summarized as follows:

1. Approximate statistics A_n by norms of sub-vectors η_n of a high-dimensional Gaussian vector η up to the residual R_A using the high dimensional central limit theorem by [10].

2. Similarly, approximate bootstrap counterparts $A_n^\flat$ of the statistics by norms of sub-vectors ζ_n of a high-dimensional Gaussian vector ζ up to the residual $R_{A^\flat}$.
3. Prove that the covariance matrix $\mathrm{Var}\,[\zeta]$ is concentrated in the ball of radius Δ_Y centered at its real-world counterpart $\mathrm{Var}\,[\eta]$.
4. By employing the Gaussian comparison result provided by [10] and [9] obtain similarity of joint distributions of the norms of ζ and η, which in combination with the steps 1 and 2 yields similarity of joint distributions of A_n and $A_n^\flat$.
5. Finally, obtain the bootstrap validity result using the sandwiching Lemma C.1.

The rigorous treatment of steps 1 through 4 is presented in Sections E, F, G and D respectively, while step 5 is formalized in Sections B and C.

4. Sensitivity and consistency results

Consider the following setting. Let there be index τ, such that $\{X_i\}_{i \leq \tau}$ 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

$$\Delta := \|\Theta_1 - \Theta_2\|_\infty. \tag{4.1}$$

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 4.1. *Let Assumption 3.1 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 maximal number d of non-zero elements in a row of the matrix Θ^*, the size s of the set $\mathcal{I}_s$, the dimensionality p, the number $|\mathfrak{N}|$ of window sizes being considered, the maximum and minimum window sizes n_+ and n_- grow with the sample size N and let the break extent Δ decay with N. Further let $N > 2n_+$, $n_+ \geq n_-$,*

$$d = o\left(\frac{\sqrt{\max\{s, n_-\}}}{|\mathfrak{N}| \log^7(pN)}\right) \tag{4.2}$$

and there exists $n^* \in \mathfrak{N}$ such that

$$\frac{\log^2(pN)}{n^* \Delta} = o(1). \tag{4.3}$$

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 A.

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

Next we formulate a trivial corollary of Theorem 4.1 establishing consistency of the change-point estimator $\hat{\tau}$ defined by (2.9).

Corollary 4.1. *Let assumptions of [Theorem 4.1](#) hold. Moreover, assume a suitable window is being considered*

$$n^* := \left\lceil \left(\frac{\log^2(pN)}{\Delta} \right)^\eta \right\rceil \in \mathfrak{N}, \ where \ \eta > 1. \tag{4.4}$$

then

$$\mathbb{P}\left\{ |\tau - \hat{\tau}| \le n^* \right\} \gtrsim 1 - \alpha.$$

Sensitivity and consistency results discussion   Assumptions (4.2) and (4.3) are essentially a sparsity bound and a bound for the sufficient windows size n^*. Clearly, they do not yield a particular value n^* necessary to detect a break, since it depends on the underlying distributions. A more restrictive assumption (4.4) suggests a particular choice of n^* which is suitable in asymptotic setting. Note, the results include dimensionality p only under the sign of logarithm, which guarantees high sensitivity of the test and proper localization of the change-point in high-dimensional setting.

Online setting   Theorem 4.1 and [Corollary 4.1](#) are established in off-line setting. In on-line setting they guarantee that the proposed approach can reliably detect a break of an extent not less than Δ with a delay at most n^*.

Multiple change-point detection   Consider a setting which allows for multiple breaks: let there be n_b change-points $\{\tau_j\}_{j=1}^{n_b} \subset N$ and let $\{X_i\}_{i=\tau_j+1}^{\tau_{j+1}}$ be i.i.d for all $1 \le j \le n_b$. Also define $n_b + 1$ precision matrices $\{\Theta_j\}_{j=0}^{n_b}$ – one for each region between a pair of consecutive change-points. Define the minimal extent of the breaks as

$$\Delta_{\min} = \min_j \|\Theta_j - \Theta_{j+1}\|_\infty.$$

Then under assumption that two change-points are not too close to each other

$$\min_j |\tau_j - \tau_{j+1}| \ge n^* + s$$

we can apply [Theorem 4.1](#) and conclude that the method detects all the change points with probability approaching 1.

Remark 4.1. *The sensitivity and consistency results heavily depend on the break extent Δ defined by (4.1). As [Remark 2.1](#) suggests, a test statistic based on the ℓ_∞ norms of the covariance matrices, not of their inverses, may also be considered. For such a statistic a similar result could be established, but the break extent would have to be correspondingly redefined as a ℓ_∞ distance between the covariance matrices instead of Θ_1 and Θ_2 which would unnecessarily restrict the application of this approach in the field of neuroimaging where the precision matrix is the main object of interest.*

5. Simulation study

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 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 and 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 assess the performance of the proposed approach in both on-line and off-line settings. Multiple break detection is left out of the scope because the suggested method attacks the problem repetitively detecting the breaks one-by-one in on-line fashion. In on-line setting the method is calibrated in order to raise a false alarm with probability $\alpha = 0.05$ on N data points using for calibration the set $\{X_i\}_{i \in \mathcal{I}_s}$ which is known in advance (unlike the rest of the dataset).

5.2. Experiment results

We have also come up with an approach to the same problem not involving bootstrap. The paper [30] 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 [23] and eventually compare them against α. This example is considered in order to demonstrate inapplicability of a straightforward application of a two-sample test for change-point detection.

The paper [31] 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 [25] was used.

In [12] a high-dimensional approach aiming to detect a change-point in second-order structure of the time series is suggested. In order to investigate its performance in our setting we use the implementation kindly provided by the authors of the paper.

Our approach is implemented as an R-package covcp, which is available on GitHub.[1]

Below we report and discuss results for a particular alternative matrix Σ_1 generated in such a way suggested in Section 5.1. All the methods being considered exhibit comparable performance for other matrices Σ_1 drawn from the same distributions.

The first type error rate and power for our approach are reported in Table 1. As one can see, our approach allows to properly control first type error rate in both off-line and on-line setting. In fact, the test is conservative and we believe this is caused by the $\leq$ signs entering the definitions of $z_n^\flat(\cdot)$ (2.6) and the multiplicity correction (2.7). As expected, the power of the test 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. The power of the test is rather similar in on-line and off-line settings which is due to its local nature.

In our study the approaches proposed in [31], [12] and the one based on the two sample test [30] turned out to be conservative, but neither of them exhibited power above 0.1.

In order to justify application of multiscale approach (i. e. $|\mathfrak{N}| > 1$) for the sake of better change-point localization in off-line setting we report the distribution of the narrowest detecting window $\hat{n}$ (defined by (2.8)) over $\mathfrak{N}$ in Figure 2. The Table 1 represents average precision of change-point localization for various choices of set of window sizes $\mathfrak{N}$. One can see, that multiscale approach significantly improves the precision of localization.

Similarly, in on-line setting we analyze *delay of detection* – the average number of data points after the break which have to be considered before the method detects a break. As Table 1 shows, the delay of detection is significantly decreased by the use of multiscale approach which justifies its use in on-line setting as well.

6. Analysis of real-world datasets

The paper [36] presents a functional magnetic resonance imaging study on human subjects, who had to learn the relation between different auditory stimuli and a monetary reward while being scanned. Based on their recorded performance (e.g. success rate) those of them who have managed to learn the relation within the time course of the experiment are considered *learners*. The authors

---

[1]https://github.com/akopich/covcp.

TABLE 1
First type error rate and power of change-point localization of the proposed approach for various sets of window sizes $\mathfrak{N}$. Localization precision and delay of detection are also reported for off-line and on-line settings respectively.

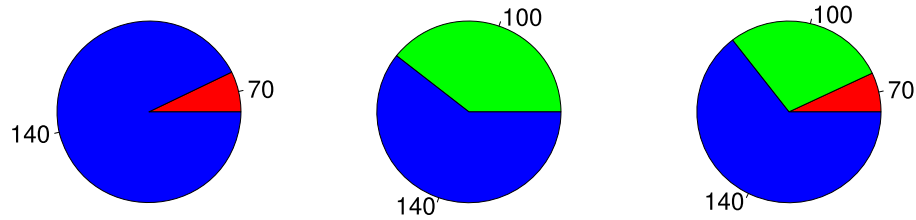| $\mathfrak{N}$	Off-line			On-line		
	I type error rate	Power	Localization precision	I type error rate	Power	Delay of detection
{70}	0.02	0.09	70	0.02	0.08	46
{100}	0.00	0.37	100	0.00	0.37	97
{140}	0.01	0.81	140	0.01	0.81	127
{200}	0.00	0.99	200	0.00	0.99	160
{140, 200}	0.00	0.99	153	0.00	0.99	140
{100, 140, 200}	0.00	0.98	146	0.00	0.99	134
{70, 140}	0.01	0.76	135	0.01	0.76	124
{100, 140}	0.01	0.75	124	0.01	0.75	118
{70, 100, 140}	0.01	0.74	123	0.01	0.72	118

FIGURE 2. *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.*

have investigated the activity of 4 brain regions-of-interest (ROI) which are believed to be involved in solving problems like the suggested one. Among the learners 3 of the ROIs exhibit a statistically significant change in the Blood Oxygenation Level Dependent (BOLD) response between the first and the last quarters of the experiment.

In this paper we analyze the BOLD responses for 18 subjects which were classified as learners. We have considered a finer-grained brain atlas with $p = 256$ ROIs [19]. For each subject and for each ROI a time-course of length $N = 1680$ was acquired. Datasets for each subject were analyzed separately. We have applied the proposed approach to the residuals of linear modeling usual in fMRI experiments [35]. The residuals are publicly available.[2] We have used graphical lasso with the penalty parameter $\lambda_n = \sqrt{\frac{\log p}{n}}$, a single window size $n = 50$ was considered and the first 200 data points were used for bootstrap simulations ($\mathcal{I}_s = \{1...200\}$). The proposed approach has detected a change-point in the covariance structure of residuals at significance level $\alpha = 0.05$ for each of the subjects.

---

[2]http://www.wias-berlin.de/preprint/2404/wias_preprints_2404.zip.

Appendix A: Proof of sensitivity result

Proof of Theorem 4.1. Proof consists in applying of the finite-sample Theorem A.1. Its applicability is guaranteed by the consistency results given in papers [37, 3, 27] and by the results from [26, 27, 3] bounding the term $R_{\hat{T}}$. High probability of set $\mathcal{T}_{\leftrightarrow}$ is ensured by Lemma G.1. $\qquad\square$

Theorem A.1. *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 3.1 holds and there exists $R_{\hat{T}}$ such that $\left\|r_{n_+}^{\mathfrak{S}}(t)\right\|_\infty \leq R_{\hat{T}}$ for all $\mathfrak{S} \in \{l, r\}$ and $t \in \mathbb{T}_{n_+}$ on some set

$$\mathcal{T}_{\leftrightarrow} := \left\{\forall t \leq \tau - n_+ : \left\|\hat{\Sigma}_n^{\mathfrak{S}}(t) - \Sigma_1^*\right\|_\infty \leq \delta_{n_+}\right\}$$
$$\bigcap \left\{\forall t \geq \tau + n_+ : \left\|\hat{\Sigma}_n^{\mathfrak{S}}(t) - \Sigma_2^*\right\|_\infty \leq \delta_{n_+}\right\}.$$

Moreover, let the residual $R_{A^\flat}$ defined in Lemma F.2 be bounded:

$$R_{A^\flat} \leq \frac{\alpha}{6\,|\mathfrak{M}|}.$$

Also let

$$\sqrt{\frac{n_+}{2}}\,\|S\|_\infty\left(\Delta - 2R_{\hat{T}}\right) \geq \mathrm{q}, \tag{A.1}$$

where

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

and Δ_Y is defined in Lemma G.2. Then on set $\mathcal{T}_{\leftrightarrow}$ with probability at least

$$1 - p_s^{\Sigma_Y}(x, \mathrm{q}),$$

where $p_s^{\Sigma_Y}(x, \mathrm{q})$ is defined in Lemma G.2, $\mathbb{H}_0$ will be rejected.

Discussion of finite-sample sensitivity result The assumption (A.1) is rather complicated. Here we note that if either graphical lasso [37], adaptive graphical lasso [43] or thresholded de-sparsified estimator based on node-wise procedure [27] 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

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

for some positive constant D_6 independent of $N, \mathfrak{M}, p, d, S$ 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 G.1). At the same time the remainder $R_{A^\flat}$ 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 3.1, the bound increases with the sample size N (only logarithmically) since we use only $2n$ data points.

Proof of Theorem A.1. The strategy of the proof is straightforward.

1. Bound the probability of a large deviation of the Gaussian approximation $\left\|\zeta_{n_+}\right\|_\infty$ of $A_{n_+}^\flat$ using Chernoff bound.
2. Bound the corresponding critical level $x_{n_+}^\flat(\alpha)$ up to the remainder term $R_{A^\flat}$, using the approximation guaranties provided by Lemma F.2.
3. Conclude that $A_{n_+} > x_{n_+}^\flat(\alpha)$ and therefore the $\mathbb{H}_0$ is rejected, by construction of the test statistic A_{n_+}.

First, denote all the window sizes being considered as $n_1, n_2, ... n_{|\mathfrak{N}|}$:

$$\mathfrak{N} = \{n_1, n_2, ..., n_{|\mathfrak{N}|}\},$$

$$n_- = n_1 < n_2 < ... < n_{|\mathfrak{N}|} = n_+.$$

Consider a pair of centered normal vectors

$$\eta := \begin{pmatrix} \eta_{n_1} & \eta_{n_2} & ... & \eta_{n_{|\mathfrak{N}|}} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_Y^*),$$

$$\zeta := \begin{pmatrix} \zeta_{n_1} & \zeta_{n_2} & ... & \zeta_{n_{|\mathfrak{N}|}} \end{pmatrix} \sim \mathcal{N}(0, \hat{\Sigma}_Y),$$

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

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

where vectors $Y_{\cdot j}^n$ and $Y_{\cdot j}^{n\flat}$ are defined in proofs of Lemma E.2 and Lemma F.1 respectively.

Lemma A.2 applies here and yields for all positive q

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

where $\hat{\Sigma}_Y = \mathrm{Var}\left[\zeta\right]$ and $\left|\mathbb{T}_{n_+}\right|$ is the number of central points for window of size n_+. Applying Lemma G.2 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,

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

Due to Lemma F.2 and continuity of Gaussian c.d.f.

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

and due to Lemma F.2 along with the fact that $\left|\mathbb{T}_{n_+}\right| < N$, choosing q as proposed by equation (A.2) we ensure that $x_{n_+}^\flat(\alpha) \leq \mathsf{q}$.

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 (A.1) $A_{n_+} > \mathsf{q}$ and therefore, $\mathbb{H}_0$ will be rejected. □

Remark A.1. *The assumption $\mathcal{I}_s \subseteq 1..\tau$ is only technical. A similar result may be proven without relying on it by methodologically the same argument. Really, if we drop the assumption, the matrix S (depending only on the distribution before the break) will fail to normalize the vectors $Y_{\cdot j}^{n\flat}$ and therefore $\left\|\hat{\Sigma}_Y\right\|_\infty$ will significantly deviate from 1. Yet a bound (omitted for brevity) of sort $\left\|\hat{\Sigma}_Y\right\|_\infty < C$ where C would depend on Δ_Y, distributions before and after the break as well as on the portions of the data points before and after the break included in the $\mathcal{I}_s$ can be established. The term C will have to enter the definition (A.2) of q instead of $(1 + \Delta_Y)$.*

Lemma A.1. *Consider a centered random Gaussian vector $\xi \in \mathbb{R}^p$ with arbitrary covariance matrix Σ. For any positive q it holds that*

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

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

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

Chernoff bound yields for any t

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

Finally, optimization over t yields the claim. □

As a trivial corollary, one obtains

Lemma A.2. *Consider a centered random Gaussian vector $\xi \in \mathbb{R}^p$ with arbitrary covariance matrix Σ. For any positive* q *it holds that*

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

Appendix B: Proof of bootstrap validity result

Proof of Theorem 3.1. Proof consists in applying of the finite-sample Theorem B.1. Its applicability is guaranteed by the consistency results given in papers [37, 3, 27] and by the results from [26, 27, 3] bounding the term $R_{\hat{T}}$. High probability of set $\mathcal{T}_T$ is ensured by Lemma G.1. □

Theorem B.1. *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 3.1 holds and there exists $R_{\hat{T}}$ such that $\sqrt{n}\left\| r_n^{\mathfrak{S}}(t) \right\|_\infty \leq R_{\hat{T}}$ for all $\mathfrak{S} \in \{l, r\}$, $n \in \mathfrak{N}$ and $t \in \mathbb{T}_n$ on set

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

Moreover, let

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

where the remainders R_A, $R_{A^\flat}$, $R_\Sigma^\pm$ are defined in Lemma E.1, Lemma F.2 and Lemma C.1 respectively and the mis-tie Δ_Y involved in the definition of $R_\Sigma^\pm$ comes from Lemma G.2. Then on set $\mathcal{T}_T$ it holds that

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

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 G.2, Lemma G.1 and Lemma F.2 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 B.1 are rather complicated. The exact expressions for them are provided by Lemma G.2, Lemma E.1, Lemma F.2 and Lemma C.1 respectively, 3rd and 4th moments M_3^3 and M_4^4 involved therein are bounded by Lemma E.4 and Lemma G.3 while asymptotic bounds for $R_{\hat{T}}$ are provided in [27] (for node-wise procedure) and [26] (for graphical lasso). For the case of graphical lasso an explicit form of $R_{\hat{T}}$ is given in [3].

Here we just note that if $\hat{\Theta}$ is a root-n consistent estimator, recovering sparsity pattern (graphical lasso [37], adaptive graphical lasso [43] or thresholded desparsified estimator based on node-wise procedure [27]), 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

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

If either graphical lasso, adaptive graphical lasso or node-wise procedure [32] 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 G.1 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 expression of $p_s^{\Sigma_Y}(x, q)$, $p^\Sigma(\gamma)$ and $p_s^M(x)$ see Lemma G.2, Lemma G.1 and Lemma F.2.

Proof of Theorem B.1. The proof consists in application of Lemma F.1, Lemma E.2 and Lemma D.1 justifying applicability of Lemma C.1. $\qquad\square$

Appendix C: Sandwiching lemma

Lemma C.1. *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*

$$\sup_{\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}} |\mathbb{P}\{\forall n \in \mathfrak{N} : A_n \leq x_n\} - \mathbb{P}\{\forall n \in \mathfrak{N} : \|\eta_n\|_\infty \leq x_n\}| \leq R_A, \tag{C.1}$$

$$\sup_{\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}} \left| \mathbb{P}^\flat \left\{ \forall n \in \mathfrak{N} : A_n^\flat \leq x_n \right\} - \mathbb{P}^\flat \left\{ \forall n \in \mathfrak{N} : ||\zeta_n||_\infty \leq x_n \right\} \right| \leq R_{A^\flat},$$

(C.2)

$$\sup_{\{x_n\}_{n \in \mathfrak{N}} \subset \mathbb{R}} \left| \mathbb{P} \left\{ \forall n \in \mathfrak{N} : A_n \leq x_n \right\} - \mathbb{P}^\flat \left\{ \forall n \in \mathfrak{N} : A_n^\flat \leq x_n \right\} \right| \leq R. \quad \text{(C.3)}$$

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

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

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

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

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

and their boundaries

$$\partial \mathcal{Z}_+(\delta) := \{ z : \mathcal{N}(z) = 1 - \alpha - \delta \}, \quad \text{(C.4)}$$

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

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

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

here and below comparison of vectors should be understood element-wise. Notice that due to continuity of multivariate normal distribution and assumption (C.2) $\forall z^\flat \in \mathcal{Z}^\flat$

$$\left| P^\flat(z^\flat) - (1 - \alpha) \right| \leq R_{A^\flat}. \quad \text{(C.5)}$$

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

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

where we have consequently used (C.3), (C.1), (C.4) and (C.5). 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 $z^N : [0, 1] \to \mathbb{R}^{|\mathfrak{N}|}$:

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

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 $\mathbf{x} \in [0, 1]$,

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

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 $0 \leq \kappa \leq 2\delta$

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

$$= \mathcal{N}^\flat\left(z^{N^\flat}(\alpha^*)\right) + \sum_{n \in \mathfrak{N}} \partial_{z_n^\flat}\mathcal{N}^\flat(z^{N^\flat}(\alpha^*))\partial_\alpha z_n^{N^\flat}(\alpha^*)\kappa + \delta$$

$$\leq 1 - \alpha + \sum_{n \in \mathfrak{N}} \partial_{z_n^\flat}\mathcal{N}^\flat(z^{N^\flat}(\alpha^*))\partial_\alpha z_n^{N^\flat}(\alpha^*)\kappa + 3\delta.$$

Next successively using [Lemma C.2](#) and the fact that the quantile function is an inverse function of 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

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

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

In the same way one obtains

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

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

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

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

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

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

(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

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

Lemma C.2. *Consider a random variable ξ and an event A defined on the same probability space. Let c.d.f. $\mathbb{P}\left\{\xi \leq x\right\}$ and $\mathbb{P}\left\{\xi \leq x \& A\right\}$ be differentiable. Then*

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

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

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

Using the fact that derivative of c.d.f. is non-negative we finalize the proof. $\quad\square$

Appendix D: Similarity of joint distributions of $\{A_n\}_{n \in \mathfrak{N}}$ and $\{A_n^{\flat}\}_{n \in \mathfrak{N}}$

Lemma D.1. *Under assumptions of Theorem 3.1 it holds that on set $\mathcal{T}$ with probability at least*

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

that

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

Proof. The proof consists in applying Lemma F.1, Lemma E.2, Lemma G.2 and Lemma H.3. $\qquad\square$

Appendix E: Gaussian approximation result for A_n

Lemma E.1. *Suppose there exists $R_{\hat{T}}$ such that $\sqrt{n}\left|\left|r^{\mathfrak{S}}(t)\right|\right|_{\infty} \leq R_{\hat{T}}$ for all $\mathfrak{S}$ and t on some set $\mathcal{T}$. Then on set $\mathcal{T}$ it holds that*

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

where F is defined by (E.2) *and η^n by* (E.1).

Proof. Substituting (2.4) to (2.3) yields

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

Now denote stacked $S_Z^n(t)$ for all $n \in \mathfrak{N}$ and as S_Z^n and for all n as S_Z. Lemma E.2 bounds the c.d.f. of $||S_Z||_{\infty}$ as

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

But clearly on set $\mathcal{T}$

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

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

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

Now notice that $\forall i : (\Sigma_Y^*)_{ii} = 1$ and bound the latter two terms by means of Lemma H.2:

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

Lemma E.2. *Let [Assumption 3.1](#) hold. Then*

$$\sup_x \left| \mathbb{P}\left\{ \forall n \in \mathfrak{N} : ||S_Z^n||_\infty \le x_n \right\} - \mathbb{P}\left\{ \forall n \in \mathfrak{N} : ||\eta^n||_\infty \le x_n \right\} \right|$$

$$\le C_A \left(F \log^7(2p^2 T n_+) \right)^{1/6}$$

Where

$$\left(\begin{array}{cccc} \eta^1 & \eta^2 & ... & \eta^{|\mathfrak{N}|} \end{array} \right) \sim \mathcal{N}(0, \Sigma_Y^*), \tag{E.1}$$

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

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

with γ defined by [(E.5)](#), β by [(E.6)](#) and Y by [(E.3)](#) and an independent constant C_A .

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

$$Y_n^T := \sqrt{\frac{n_+}{n}} \times$$

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

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:

$$Y := \begin{pmatrix} \dfrac{Y_1}{\dfrac{Y_2}{\dfrac{...}{Y_{|\mathfrak{N}|}}}} \end{pmatrix} \tag{E.3}$$

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 H.1. Denote

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

By means of Lemma G.3 one shows that the assumptions of Lemma E.3 hold for components of Z_i^S with

$$\gamma := 12 L^2 \sqrt{d} \Lambda \left(\Theta^* \right) \left\| \Theta^* \right\|_\infty \left\| S^{-1} \right\|_\infty \qquad \text{(E.5)}$$

$$\beta := \left(\frac{9}{2} L^2 \sqrt{d} \Lambda \left(\Theta^* \right) + 1 \right) \left\| \Theta^* \right\|_\infty \left\| S^{-1} \right\|_\infty \qquad \text{(E.6)}$$

where $\Lambda \left(\Theta^* \right)$ denotes the maximal eigen value of Θ^*. Therefore condition (H.1) holds with B_n defined by equation (E.4).

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

Hence, Assumption H.1 is fulfilled with $b = 1$. Next notice that for some k-th component of Z_i^S and 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 H.2 holds with B_{n_+} so Lemma H.1 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 depends on nothing. $\square$

Lemma E.3. *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

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

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

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

$$\leq 2 \qquad\qquad\qquad\qquad \square$$

By the same technique the following lemma can be proven

Lemma E.4. *Under assumptions of Lemma E.3*

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

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

Appendix F: Gaussian approximation result for $A_n^\flat$

Lemma F.1.

$$\sup_{\{x_n\}_{n \in \mathfrak{N} \subset \mathbb{R}}}\left|\mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : A^\flat \leq x_n\right\} - \mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : ||\zeta^n||_\infty \leq x_n\right\}\right|$$

$$\leq \hat{C}_{A^\flat}\left(F^\flat \log^7(2p^2Tn_+)\right)^{1/6}.$$

Where

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

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

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

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

$Y_n^\flat$ are defined by (F.1), 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 (2.5) as $S_Z^{n\flat}$

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

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

Consider a matrix

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

$$\begin{pmatrix}
Z_1^{S\flat} & O & \dots & O & -Z_{2n_++1}^{S\flat} & \dots \\
Z_2^{S\flat} & Z_2^{S\flat} & \dots & \dots & \dots & \dots \\
\dots & Z_3^{S\flat} & \dots & \dots & \dots & \dots \\
Z_n^{S\flat} & \dots & \dots & \dots & \dots & \dots \\
-Z_{n+1}^{S\flat} & Z_{n+1}^{S\flat} & \dots & \dots & \dots & \dots \\
-Z_{n+2}^{S\flat} & -Z_{n+2}^{S\flat} & \dots & \dots & \dots & \dots \\
\dots & -Z_{n+3}^{S\flat} & \dots & O & \dots & \dots \\
-Z_{2n}^{S\flat} & \dots & \dots & Z_{2n_+-2n+1}^{S\flat} & O & \dots \\
O & -Z_{2n+1}^{S\flat} & \dots & Z_{2n_+-2n+2}^{S\flat} & Z_{2n_+-2n+2}^{S\flat} & \dots \\
O & O & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & -Z_{2n_+-1}^{S\flat} & -Z_{2n_+-1}^{S\flat} & \dots \\
O & O & \dots & -Z_{2n_+}^{S\flat} & -Z_{2n_+}^{S\flat} & \dots
\end{pmatrix} \quad (F.1)$$

which is a bootstrap counterpart of Y_n from the proof of Lemma E.2 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

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

V. Avanesov and N. Buzun

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

And finally apply Lemma H.1. $\qquad\square$

Lemma F.2. *Let $\hat{\Theta}$ denote an estimator of Θ^* s.t. for some positive r*

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

and $\Theta_{ij}^* = 0 \Rightarrow \hat{\Theta}_{ij} = 0$, furthermore, let $\Delta_Y(r) < 1/2$, also suppose Assumption 3.1 holds. Then at least with probability $1 - p_s^M(x) - p_s^{\Sigma_Y}(x, q)$

$$\sup_{\{x_n\}_{n \in \mathfrak{N} \subset \mathbb{R}}} \left|\mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : A^\flat \leq x_n\right\} - \mathbb{P}^\flat\left\{\forall n \in \mathfrak{N} : ||\zeta^n||_\infty \leq x_n\right\}\right|$$

$$\leq R_{A^\flat} := C_{A^\flat} \left(\hat{F} \log^7(2p^2 T n_+)\right)^{1/6}$$

where

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

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

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

Proof. The proof consists in subsequently applying Lemma F.1 and Lemma F.3 ensuring $C^\flat \geq M^\flat = \max_{i \in \mathcal{I}_s} \left|\left|\hat{Z}_i\right|\right|_\infty$ with probability at least $1 - p_s^M(x)$ and applying Lemma G.2 providing that $\left|\left|\Sigma_Y^* - \hat{\Sigma}_Y\right|\right|_\infty \leq \Delta_Y \leq 1 = \min_{1 \leq k \leq p}(\Sigma_Y^*)_{kk}$ with probability at least $1 - p_s^{\Sigma_Y}(x, q)$ which implies the existence of a deterministic constant $C_{A^\flat} > \hat{C}_{A^\flat}$. $\qquad\square$

Lemma F.3. *Let $\hat{\Theta}$ denote an estimator of Θ^* s.t. for some positive r*

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

and $\Theta_{ij}^* = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Also let Assumption 3.1 hold. Then with probability at least $1 - p_s^M(x)$

$$M^\flat \leq \mathcal{Z}_s(x) + \Delta_Z(x) \tag{F.2}$$

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

Proof. Direct application of Lemma G.4 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 G.6) that $\left|\left|\hat{Z}_i - Z_i\right|\right|_\infty \leq \Delta_Z(x)$ implies (F.2). $\qquad\square$

Appendix G: $\hat{\Sigma}_Y \approx \Sigma_Y^*$

First of all, if $\Sigma_Z^* := \mathrm{Var}\left[\overline{Z_i}\right] \approx \mathrm{Var}_\flat\left[\overline{Z_i^\flat}\right]$, then $\Sigma_Y^* \approx \hat{\Sigma}_Y$ as well (Lemma G.2). The idea is to notice that

$$\mathrm{Var}\left[\overline{Z_i^\flat}\right] = \hat{\Sigma}_{\hat{Z}} := \mathbb{E}_s\left[\left(\overline{\hat{Z}_i} - \mathbb{E}_s\left[\overline{\hat{Z}_i}\right]\right)\left(\overline{\hat{Z}_i} - \mathbb{E}_s\left[\overline{\hat{Z}_i}\right]\right)^T\right]$$

due to the choice of the bootstrap scheme. Next we show that $\Sigma_Z^* \approx \hat{\Sigma}_Z := \mathbb{E}_s\left[\left(\overline{Z_i} - \mathbb{E}_s\left[\overline{Z_i}\right]\right)\left(\overline{Z_i} - \mathbb{E}_s\left[\overline{Z_i}\right]\right)^T\right]$ (Lemma G.5) and finalize the proof by proving that $\hat{\Sigma}_Z \approx \hat{\Sigma}_{\hat{Z}}$ (Lemma G.7).

The results of this section rely on a lemma which is a trivial corollary of Lemma 6 by [26] providing the concentration result for the empirical covariance matrix

Lemma G.1. *Let Assumption 3.1 hold for some $L > 0$. Then for any positive γ*

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

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

Lemma G.2. *Assume, Assumption 3.1 holds. Moreover, let*

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

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 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

$$p_s^{\Sigma_Y}(x, q) := p_s^{\Sigma_{Z_1}}(x, q) + p_s^{\Sigma_{Z_2}}(x)$$

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

and $\Delta_{\Sigma_Z}^{(1)}$ and $\Delta_{\Sigma_Z}^{(2)}$ along with the probabilities $p_s^{\Sigma_{Z_1}}(x, q)$ and $p_s^{\Sigma_{Z_2}}(x)$ are defined in Lemma G.5 and Lemma G.7 respectively.

Proof. Notice that

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

V. Avanesov and N. Buzun

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 G.7 and Lemma G.5

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

and hence

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

with probability at least

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

Lemma G.3. *Under Assumption 3.1 it holds for arbitrary $1 \leq u, v \leq p$ and positive x that*

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

Proof. Re-write the definition (2.2) of an element $Z_{i,uv}$ for arbitrary $1 \leq u, v \leq p$

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

The first term is clearly a value of a quadratic form defined by the matrix $B = \Theta_u^*(\Theta_v^*)^T$. Note that $rank B = 1$ which implies that it is either positive semi-definite or negative semi-definite. Next we apply Lemma H.4 and obtain for all positive x

$$\mathbb{P}\left\{\left|X_i^T B X_i\right| \geq 3L^2\left(|tr B| + 2\sqrt{tr(B^2)x} + 2\left|\Lambda\left(B\right)\right|x\right)\right\} \leq e^{-x}. \qquad (G.1)$$

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

$$tr B = \Lambda\left(B\right) = \sqrt{tr B^2} \qquad (G.2)$$

and by construction of matrix B

$$\begin{aligned} |tr B| &= \left|\Theta_u^*(\Theta_v^*)^T\right| \\ &\leq ||\Theta_u^*||_1 ||\Theta^*||_\infty \\ &\leq \sqrt{d}||\Theta_u^*||_2 ||\Theta^*||_\infty \\ &\leq \sqrt{d}\Lambda\left(\Theta^*\right)||\Theta^*||_\infty. \end{aligned} \qquad (G.3)$$

Substitution of (G.2) and (G.3) to (G.1) yields

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

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

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

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

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

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

Lemma G.4. *Consider an i.i.d. sample X_i of length n. Under Assumption 3.1 for positive x it holds that*

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

where

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

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

Lemma G.5. *Under Assumption 3.1 for positive x and q*

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

where

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

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

Proof. Denote

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

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

$$\mathbb{P}\left\{\mathbb{E}_s\left[W_{kl}^{(i)}\right]\right.$$

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

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

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

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

Lemma G.6. *Let [Assumption 3.1](#) holds and let $\hat{\Theta}$ be a symmetric estimator of Θ^* s.t.*
$$
\left|\left|\Theta^* - \hat{\Theta}\right|\right|_\infty < r
$$
and $\Theta_{ij}^* = 0 \Rightarrow \hat{\Theta}_{ij} = 0$. Then for positive x
$$
\mathbb{P}\left\{\forall i \in \mathcal{I}_s : \left|\left|Z_i - \hat{Z}_i\right|\right|_\infty \leq \Delta_Z(x)\right\} \geq 1 - p_s^X(x)
$$
where
$$
\Delta_Z(x) := 2rd^{3/2}x^2\,||\Theta^*||_\infty + (rdx)^2
$$
$$
p_s^X(x) := se^{-x^2/L^2}
$$

Proof. Due to sub-Gaussianity,
$$
\forall \alpha \in \mathbb{R}^p : \mathbb{P}\left\{\left|\alpha^T X_i\right| \leq x\right\} \geq 1 - se^{-x^2/L^2} \tag{G.4}
$$

Now consider the mis-tie of arbitrary elements $Z_{i,uv}$ and $\hat{Z}_{i,uv}$:
$$
\left|Z_{i,uv} - \hat{Z}_{i,uv}\right| = \left|\Theta_u^* X_i \Theta_v^* X_i + \Theta_{uv}^* - \hat{\Theta}_u X_i \hat{\Theta}_v X_i - \hat{\Theta}_{uv}\right|
$$
$$
\leq \left|(\Theta_u^* - \hat{\Theta}_u)X_i\Theta_v^* X_i\right| + \left|(\Theta_u^* - \hat{\Theta}_u)X_i\hat{\Theta}_v X_i\right| + r
$$

Now note that due to [(G.4)](#) and assumptions imposed on Θ^*
$$
|\Theta_v^* X_i| \leq \sqrt{d}\,||\Theta^*||_\infty\, x
$$
$$
\left|(\Theta_v^* - \hat{\Theta}_v)X_i\right| \leq rdx
$$
$$
\left|\hat{\Theta}_v X_i\right| \leq |\Theta_v^* X_i| + \left|(\Theta_v^* - \hat{\Theta}_v)X_i\right| \leq \sqrt{d}\,||\Theta^*||_\infty\, x + rdx
$$

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

Lemma G.7. *Assume [Assumption 3.1](#) 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

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

where

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

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

Proof. By [Lemma G.4](#) 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 G.6](#) 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

$$\xi_i := \overline{Z_i} - \mathbb{E}_s\left[\overline{Z_i}\right] \text{ and } \hat{\xi}_i := \overline{\hat{Z}_i} - \mathbb{E}_s\left[\overline{\hat{Z}_i}\right]$$

And deliver the bound

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

$\square$

Appendix H: Known results

H.1. *Gaussian approximation result*

In this section we briefly describe the result obtained in [10].

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 H.1. *We call a set A of the form $A = \{w \in \mathbb{R}^p : a_i \le w_i \le b_i \;\forall i \in \{1..p\}\}$ a hyperrectangle. The family of all hyperrectangles is denoted as A^{re}.*

Assumption H.1. $\exists b > 0$ *such that*

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[x_{ij}^2\right] \geq b \text{ for all } j \in 1..p$$

Assumption H.2. $\exists G_n \geq 1$ *such that*

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[|x_{ij}|^{2+k}\right] \leq G_n^{2+k} \text{ for all } j \in 1..p \text{ and } k \in \{1, 2\}$$

$$\mathbb{E}\left[\exp\left(\frac{|x_{ij}|}{G_n}\right)\right] \leq 2 \text{ for all } j \in 1..p \text{ and } i \in 1..n \qquad \text{(H.1)}$$

Lemma H.1 (Proposition 2.1 by [10]). *Let Assumption H.1 hold for some b and Assumption H.2 hold for some G_n. Then*

$$\sup_{A \in A^{re}} \left|\mathbb{P}\left\{S_n^X \in A\right\} - \mathbb{P}\left\{S_n^Y \in A\right\}\right| \leq C \left(\frac{G_n^2 \log^7(pn)}{n}\right)^{1/6}$$

and the constant C depends only on b.

H.2. Anti-concentration result

Lemma H.2 (Nazarov's inequality [34]). *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*

$$\mathbb{P}\left\{X \leq y + a\right\} - \mathbb{P}\left\{X \leq y\right\} \leq C a \sqrt{\log p},$$

where C is an independent constant.

H.3. Gaussian comparison result

By the technique given in the proof of Theorem 4.1 by [10] one obtains the following generalization of the result given in [8]

Lemma H.3. *Consider a pair of covariance matrices Σ_1 and Σ_2 of size $p \times p$ such that*

$$\|\Sigma_1 - \Sigma_2\|_\infty \leq \Delta$$

and $\forall k : C_1 \geq \Sigma_{1,kk} \geq c_1 > 0$. Then for random vectors $\eta \sim \mathcal{N}(0, \Sigma_1)$ and $\zeta \sim \mathcal{N}(0, \Sigma_2)$ it holds that

$$\sup_{A \in A^{re}} |\mathbb{P}\left\{\eta \in A\right\} - \mathbb{P}\left\{\zeta \in A\right\}| \leq C \Delta^{1/3} \log^{2/3} p,$$

where C is a positive constant which depends only on C_1 and c_1.

H.4. Tail inequality for quadratic forms

The following result is a direct corollary of Theorem 1 in [24]

Lemma H.4. *Consider a positive semi-definite or negative semi-definite matrix B and suppose Assumption 3.1 holds. Then for all $t > 0$*

$$\mathbb{P}\left\{\left|X_1^T B X_1\right| \geq 3L^2\left(|trB| + 2\sqrt{tr(B^2)t} + 2\left|\Lambda\left(B\right)\right|t\right)\right\} \leq e^{-t}$$

H.5. High-dimensional precision matrix estimation

In order to address the problem of high-dimensional precision matrix estimation one has to assume its sparsity. Below we describe two approaches exploiting this assumption. In both of them we assume that an i.i.d. sample $X_1, ... X_n \in \mathbb{R}^p$ is supplied.

H.5.1. Graphical lasso

In [20] the graphical lasso approach was suggested. An estimate may be obtained as the solution of the following optimization problem over a positive-definite cone S_{++}^p of $p \times p$ dimensional matrices.

$$\hat{\Theta}^{GL} := \arg\min_{\Theta \in S_{++}^p}\left[tr(\Theta\hat{\Sigma}) - \log det\Theta + \lambda\left|\left|\Theta\right|\right|_1\right] \tag{H.2}$$

where $\hat{\Sigma}$ stands for the empirical covariance matrix

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^T.$$

The theoretical treatment of the approach keeps track on the following Schatten norms: $\kappa_{\Sigma^*} = \left|\left|\left|\Sigma^*\right|\right|\right|_\infty$ and $\kappa_{\Gamma^*} = \left|\left|\left|(\Gamma_{SS}^*)^{-1}\right|\right|\right|_\infty$. The following result establishes consistency of the estimator in the sense of Definition 3.1.

Lemma   H.5 (Theorem   1,   [37])**.** *Consider   a   distribution   satisfying Assumption 3.2 with some $\phi \in (0, 1]$, let $\hat{\Theta}$ be a solution of the optimization problem (H.2) with tuning parameter $\lambda_n = \frac{8}{\psi}\delta_n$. Furthermore, impose the following sparsity assumption:*

$$d \leq \frac{1}{6(\delta_n + \lambda_n)\max\{\kappa_{\Gamma^*}\kappa_{\Sigma^*}, \kappa_{\Gamma^*}^2\kappa_{\Sigma^*}^3\}}.$$

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

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

and

$$\Theta_{ij}^* = 0 \Rightarrow \hat{\Theta}_{ij} = 0.$$

A rather similar result is provided in paper [3] for adaptive versions of graphical lasso suggested and studied in [44] [17] [18] [43] .

H.5.2. Node-wise lasso

This section describes the node-wise lasso approach which was suggested in [32].

For each $1 \leq j \leq n$ define a vector

$$\hat{\Gamma}_j := (\hat{\gamma}_{j\,1}, ..., \hat{\gamma}_{j\,j-1}, 1, \hat{\gamma}_{j\,j+1}, ..., \hat{\gamma}_{j\,p})$$

where $\hat{\gamma}_j$ is defined as a solution of the following lasso regression:

$$\hat{\gamma}_j := \arg \max_{\gamma \in \mathbb{R}^{p-1}} \frac{1}{n} \sum_{1 \leq i \leq n} \left(X_{ij} - X_{i,-j}^T \gamma \right)^2 + 2\lambda \left\| \gamma \right\|_1$$

and

$$\hat{\tau}_j^2 := \frac{1}{n} \sum_{1 \leq i \leq n} \left(X_{ij} - X_{i,-j}^T \hat{\gamma}_j \right)^2 + \lambda \left\| \gamma \right\|_1.$$

Finally the j-th column of the estimator is defined as

$$\hat{\Theta}_j^{MB} := \hat{\Gamma}_j / \hat{\tau}_j^2.$$

Note, that this estimator might not be symmetric, so one cannot use it as an estimator $\hat{\Theta}$ based on the sub-sample $\{X_i\}_{i \in \mathcal{I}_s}$. The paper [27] suggests to construct a de-sparsified estimator $\hat{T}(\hat{\Theta}^{MB})$ where

$$\hat{T}(\hat{\Theta}) := \hat{\Theta} + \hat{\Theta}^T - \hat{\Theta}^T \hat{\Sigma} \hat{\Theta}$$

and threshold elements of $\hat{T}$ obtaining a positive-definite estimate.

Under Assumption 3.1, the sparsity assumption $\frac{d \log p}{n} = o(1)$ and the assumption of the bounded spectrum (Assumption H.3) the paper [27] establishes the root-n consistency of such an estimator (see Definition 3.1).

Assumption H.3.

$$\exists E : \frac{1}{E} \leq \lambda\left(\Theta^*\right) \leq \Lambda\left(\Theta^*\right) \leq E.$$

H.5.3. Bounds for r

While graphical lasso and node-wise estimate are point estimates, de-sparsified estimators have been suggested in order to obtain confidence intervals [26] [27].

The analysis of these estimators relies on the bounds for the residual term r:

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

The next two lemmas bound the remainder r for the case of graphical lasso and node-wise estimator.

Lemma H.6 (by [26]). *Impose* Assumption 3.1, Assumption 3.2 *and* Assumption H.3. *Then under the sparsity assumption*

$$\frac{d \log p}{\sqrt{n}} = o(1) \tag{H.3}$$

it holds that

$$||r||_\infty = O_p\left(\frac{d \log p}{n}\right).$$

A finite sample-size bound for r along with its adaptations for the case of adaptive graphical lasso may be found in [3]

Lemma H.7 (by [27]). *Let $\hat{\Theta}$ be yielded by the node-wise procedure with $\lambda_n \asymp \sqrt{\frac{\log p}{n}}$. Then under* Assumption 3.1, Assumption H.3 *and sparsity assumption* (H.3)

$$||r||_\infty = O_p\left(\frac{d \log p}{n}\right).$$

Acknowledgements

References

[1] Alexander Aue, Siegfried Hörmann, Lajos Horváth, and Matthew Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 12 2009. MR2572452

[2] Alexander Aue and Lajos Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013. MR3008012

[3] Valeriy Avanesov, Jörg Polzehl, and Karsten Tabelow. Consistency results and confidence intervals for adaptive l1-penalized estimators of the high-dimensional sparse precision matrix. Technical Report 2229, WIAS, 2016.

[4] Danielle S. Bassett, Nicholas F. Wymbs, Mason a. Porter, Peter J. Mucha, Jean M. Carlson, and Scott T. Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641, 2010.

[5] Peter Bauer and Peter Hackl. An extension of the mosum-technique for quality control. 22:1–7, 02 1980.

[6] Luc Bauwens, Sébastien Laurent, and Jeroen V K Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, Jan 2006. MR2225523

[7] Gérard Biau, Kevin Bleakley, and David M. Mason. Long signal change-point detection. *Electron. J. Statist.*, 10(2):2097–2123, 2016. MR3522670

[8] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. Dec 2013. MR3262468

[9] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819, 12 2013. MR3161448

[10] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Central limit theorems and bootstrap in high dimensions. Dec 2014.

[11] Haeran Cho. Change-point detection in panel data via double cusum statistic. *Electron. J. Statist.*, 10(2):2000–2038, 2016. MR3522667

[12] Haeran Cho and Piotr 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. MR3310536

[13] Mihaela Şerban, Anthony Brockwell, John Lehoczky, and Sanjay Srivastava. Modelling the dynamic dependence structure in multivariate financial time series. *Journal of Time Series Analysis*, 28(5):763–782, 2007. MR2395913

[14] M. Csörgö and L. Horváth. *Limit theorems in change-point analysis*. Wiley series in probability and statistics. J. Wiley & Sons, Chichester, New York, 1997. MR2743035

[15] Birte Eichinger and Claudia Kirch. A mosum procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526–564, 02 2018. MR3706768

[16] Robert F. Engle, Victor K. Ng, and Michael Rothschild. Asset pricing with a factor-arch covariance structure. Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1–2):213–237, 1990.

[17] Jianqing Fan, Yang Feng, and Yichao Wu. Network exploration via the adaptive lasso and scad penalties. *Ann. Appl. Stat.*, 3(2):521–541, 06 2009. MR2750671

[18] Jianqing Fan and Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360, 2001. MR1946581

[19] Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18:1664–1671, 2015.

[20] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, July 2008.

[21] Jerome Friedman, Trevor Hastie, Robert Tibshirani, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso, pages 1–14, 2007.

[22] Karl J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011.

[23] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979. MR0538597

[24] Daniel Hsu, Sham M. Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(0), January 2012. MR2994877

[25] Nicholas A. James and David S. Matteson. ecp: An R package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software*, 62(7):1–25, 2014. MR3180567

[26] Jana Janková and Sara van de Geer. Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Statist.*, 9(1):1205–1229, 2015. MR3354336

[27] Jana Janková and Sara van de Geer. Honest confidence regions and optimality in high-dimensional precision matrix estimation. *TEST*, 26(1):143–162, 2017. MR3613609

[28] Moritz Jirak. Uniform change point tests in high dimension. *Ann. Statist.*, 43(6):2451–2483, 12 2015. MR3405600

[29] M. Lavielle and G. Teyssière. Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, 46(3):287–306, 2006. MR2285348

[30] Jun Li and Song Xi Chen. Two sample tests for high-dimensional covariance matrices. *Ann. Statist.*, 40(2):908–940, 04 2012. MR2985938

[31] David S. Matteson and Nicholas A. James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014. MR3180567

[32] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 06 2006. MR2278363

[33] Thomas Mikosch, Søren Johansen, and Eric Zivot. Handbook of Financial Time Series. *Time*, 468(1996):671–693, 2009.

[34] Fedor 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. MR2083397

[35] Russell A. Poldrack, Jeanette A. Mumford, and Thomas E. Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011. MR2839490

[36] Sebastian Puschmann, André Brechmann, and Christiane M. Thiel. Learning-dependent plasticity in human auditory cortex during appetitive operant conditioning. *Human Brain Mapping*, 34(11):2841–2851, 2013.

[37] Pradeep Ravikumar, Martin J. Wainwright, Garvesh Raskutti, and Bin Yu. High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence. *Electron. J. Statist.*, 5:935–980, 2011. MR2836766

[38] A.N. Shiryaev. *Optimal Stopping Rules*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2007. MR2374974

[39] V. Spokoiny and N. Willrich. Bootstrap tuning in ordered model selection. *ArXiv e-prints*, July 2015.

[40] O. Sporns. *Networks of the brain.* The MIT Press, 2011.

[41] Yao Xie and David Siegmund. Sequential multi-sensor change-point detection. *Ann. Statist.*, 41(2):670–692, 04 2013. MR3099117

[42] Changliang Zou, Guosheng Yin, Long Feng, and Zhaojun Wang. Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.*, 42(3):970–1002, 06 2014. MR3210993

[43] H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006. MR2279469

[44] Hui Zou and Runze Li. One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.*, 36(4):1509–1533, 08 2008. MR2435443