



**HAL**  
open science

# Block-diagonal covariance estimation and application to the Shapley effects in sensitivity analysis

Baptiste Broto, François Bachoc, Jean-Marc Martinez

► **To cite this version:**

Baptiste Broto, François Bachoc, Jean-Marc Martinez. Block-diagonal covariance estimation and application to the Shapley effects in sensitivity analysis. 2019. hal-02196583v1

**HAL Id: hal-02196583**

**<https://hal.science/hal-02196583v1>**

Preprint submitted on 29 Jul 2019 (v1), last revised 5 Feb 2020 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Block-diagonal covariance estimation and application to the Shapley effects in sensitivity analysis

Baptiste Broto<sup>1</sup>, François Bachoc<sup>2</sup>, and Jean-Marc Martinez<sup>3</sup>

<sup>1</sup>CEA, LIST, Université Paris-Saclay, F-91120, Palaiseau, France

<sup>2</sup>Institut de Mathématiques de Toulouse, Université Paul Sabatier, F-31062 Toulouse, France

<sup>3</sup>CEA, DEN, Université Paris-Saclay, F-91120, Palaiseau, France

July 29, 2019

## Abstract

In this paper, we aim to estimate block-diagonal covariance matrices for Gaussian data in high dimension and in fixed dimension. We first estimate the block-diagonal structure of the covariance matrix by theoretical and practical estimators which are consistent. We deduce that the suggested estimator of the covariance matrix in high dimension converges with the same rate than if the true decomposition was known. In fixed dimension, we prove that the suggested estimator is asymptotically efficient. Then, we focus on the estimation of sensitivity indices called "Shapley effects", in the high-dimensional Gaussian linear framework. From the estimated covariance matrix, we obtain an estimator of the Shapley effects with a relative error which goes to zero at the parametric rate up to a logarithm factor. Using the block-diagonal structure of the estimated covariance matrix, this estimator is still available for thousands inputs variables, as long as the maximal block is not too large.

## 1 Introduction

Sensitivity analysis has become an important tool in applied sciences. The aim of sensitivity indices is to quantify the impact of the input variables  $X_1, \dots, X_p$  on the output  $Y$  of a model. This information improves the interpretability of the model. In global sensitivity analysis, the input variables are assumed to be random variables. In this framework, the Sobol indices [28] were the first suggested indices to be applicable to general classes of models. Nevertheless, one of the most important limitations of these indices is the assumption of

independence between the input variables. Hence, many variants of the Sobol indices have been suggested for dependent input variables [18, 6, 19].

Recently, Owen defined new sensitivity indices in [21] called "Shapley effects". These sensitivity indices have many advantages over the Sobol indices for dependent inputs [14]. For general models, [29] suggested an estimator of the Shapley effects. However, this estimation requires to be able to generate samples with the conditional distributions of the input variable. A consistent estimator has been suggested in [3], requiring only a sample of the inputs-output. This estimator uses nearest-neighbours methods to mimic the generation of samples with these conditional distributions.

In this paper, we focus on Gaussian linear models in large dimension. Gaussian linear models are widely used as numerical models of physical phenomena (see for example [15], [11], [24]). Indeed, uncertainties are often modelled as Gaussian variables and an unknown function  $Y = f(X_1, \dots, X_p)$  is commonly approximated by its linear approximation around  $E(X)$ . Furthermore, high-dimensional Gaussian linear models are widely studied in statistics [5, 9]. In this particular case of Gaussian linear models, the theoretical values of the Shapley effects can be computed explicitly [22, 14, 4]. These values depend on the covariance matrix  $\Sigma$  of the inputs and on the coefficients  $\beta$  of the linear model.

In this paper, we assume that we observe an i.i.d. sample of the input Gaussian variables in high dimension and that the true covariance matrix is unknown. In this setting, the Shapley effects need to be estimated, replacing the theoretical covariance matrix by an estimated covariance matrix.

There exists a fair amount of work on high-dimensional covariance matrix estimation. Many researchers took an interest in the empirical covariance matrix in high dimension [20, 32, 27, 1]. For particular covariance matrices, different estimators than the empirical covariance can be preferred. For some well-conditioned families of covariance matrices, [2] suggests a banded version of the empirical covariance matrix, and several works address the problem of estimating a sparse covariance matrix [12, 16, 8].

However, in general, given a high-dimensional covariance matrix, the computation cost of the corresponding Shapley effects grows exponentially with the dimension. The only setting where a procedure to compute the Shapley effect with a non-exponential cost is the setting of block-diagonal matrices [4]. Hence, in high dimension, block diagonal covariance matrices are a very favorable setting for the estimation of the Shapley effects. Thus, we address the estimation of high-dimensional block diagonal covariance matrices in this paper. We remark that the above methods are not relevant for the estimation of the Shapley effects, since they do not provide block diagonal matrices.

In our framework, we assume that the true covariance matrix is block diagonal and we want to estimate this matrix with a similar structure to compute

the deduced Shapley effects. Some works address the block diagonal estimation of covariance matrices. [23] gives a numerical procedure to estimate such covariance matrices and [13] suggests a test to verify the independence of the blocks. A block-diagonal estimator of the covariance matrix is proposed in [7]. The authors of [7] choose a more general framework, without assuming that the true covariance matrix is block-diagonal. They obtain the estimated block-diagonal structure by thresholding the empirical correlation matrix. They also give theoretical guaranties by bounding the average of the squared Hellinger distance between the estimated probability density function and the true one. This bound depends on the dimension  $p$  and the sample size  $n$ . When  $p/n$  converges to some constant  $y \in ]0, 1[$ , this bound is larger than 1 and is no more relevant as the Hellinger distance is always smaller than 1.

Here, we focus on the high dimension setting, when  $p/n$  converges to some constant  $y \in ]0, 1[$ , and when the true covariance matrix is assumed to be block-diagonal. We give different estimators of the block-diagonal structure and we show that their complexity is small. Then, we provide new asymptotic results for these estimators. Under mild conditions, we show that the estimators of the block structure are equal to the true block structure, with probability converging to one. Furthermore, the square Frobenius distance between the estimated covariance matrices and the true one, normalized by  $p$ , converge to zero at rate  $1/n$ . Thus, our work complete the one of [7]. We also study the fixed-dimensional setting, where we show that our suggested estimator are asymptotically efficient.

From the estimated block-diagonal covariance matrices, we deduce estimators of the Shapley effects in the high dimensional linear Gaussian framework, with reduced computational cost. We recall that in high dimension, the computation of the Shapley effects requires that the corresponding covariance matrix be block diagonal. We show that the relative estimation error of these estimators goes to zero at the parametric rate  $1/n^{1/2}$ , up to a logarithm factor.

The rest of the paper is organized as follows. In Section 2, we focus on the block-diagonal estimation of the block-diagonal covariance matrix. In Section 3, we apply this block-diagonal estimation of the covariance matrix to deduce Shapley effects estimators. The conclusion is given in Section 4. All the proofs are postponed to the appendix.

## 2 Estimation of block-diagonal covariance matrices

### 2.1 Problem and notations

We assume that we observe  $(X^{(l)})_{l \leq n}$ , an i.i.d. sample with distribution  $\mathcal{N}(\mu, \Sigma)$ , where  $\mu \in \mathbb{R}^p$  and  $\Sigma$  are not known. We assume that  $\Sigma = (\sigma_{ij})_{i,j \leq p} \in S_p^{++}(\mathbb{R})$

(the set of the symmetric positive definite matrices) and has a block diagonal decomposition. To be more precise on this block diagonal decomposition, we need to introduce some notations.

Let us write  $\mathcal{P}_p$  the set of all the partitions of  $[1 : p]$ . We endow the set  $\mathcal{P}_p$  with the following partial order. If  $B, B' \in \mathcal{P}_p$ , we say that  $B$  is finer than  $B'$ , and we write  $B \leq B'$ , if for all  $A \in B'$ , there exists  $A_1, \dots, A_i \in B$  such that  $A = \bigsqcup_{j=1}^i A_j$ . We also compare the elements of a partition  $B \in \mathcal{P}_p$  with their smallest element; that enables us to talk about "the  $k$ -th element" of  $B$ . If  $B \in \mathcal{P}_p$  and  $a_1, \dots, a_i \in [1 : p]$ , we write  $(a_1, \dots, a_i) \in B$  if there exists  $A \in B$  such that  $\{a_1, \dots, a_i\} \subseteq A$  (in other words, if  $a_1, \dots, a_i$  are in the same group of  $B$ ). If  $\Gamma \in S_p^{++}(\mathbb{R})$  with  $\Gamma = (\gamma_{ij})_{i,j \leq p}$  and if  $B \in \mathcal{P}_p$ , we define  $\Gamma_B$  by

$$(\Gamma_B)_{i,j} = \begin{cases} \gamma_{ij} & \text{if } (i, j) \in B \\ 0 & \text{otherwise.} \end{cases}$$

Let us define

$$S_p^{++}(\mathbb{R}, B) := \{\Gamma \in S_p^{++} \mid \Gamma = \Gamma_B, \text{ and } \forall B' < B, \Gamma \neq \Gamma_{B'}\}.$$

Thus  $S_p^{++}(\mathbb{R}) = \bigsqcup_{B \in \mathcal{P}_p} S_p^{++}(\mathbb{R}, B)$  and for all  $\Gamma \in S_p^{++}(\mathbb{R})$ , we can define an unique  $B(\Gamma) \in \mathcal{P}_p$  such that  $\Gamma \in S_p^{++}(\mathbb{R}, B(\Gamma))$ . Here, we assume that  $\Sigma \in S_p^{++}(\mathbb{R}, B^*)$ , i.e.  $B^*$  is the finest decomposition of  $\Sigma$ , i.e.  $B(\Sigma) = B^*$ . We say that  $\Sigma$  has a block diagonal decomposition  $B^*$ .

We also write

$$\bar{X}_n := \frac{1}{n} \sum_{l=1}^n X^{(l)},$$

and

$$S_n := \frac{1}{n} \sum_{l=1}^n (X^{(l)} - \bar{X}_n)(X^{(l)} - \bar{X}_n)^T,$$

which are the empirical estimators of  $\mu$  and  $\Sigma$ . To simplify notations, we write  $\bar{X}$  for  $\bar{X}_n$  and  $S$  for  $S_n$  (the dependency on  $n$  is implicit). We know that, for all  $\Gamma \in S_p^{++}(\mathbb{R})$ ,  $\bar{X}$  maximizes the likelihood  $L_{\Gamma, m}(X^{(1)}, \dots, X^{(n)})$  over the mean parameter  $m$ , where

$$L_{\Gamma, m}(X^{(1)}, \dots, X^{(n)}) := \frac{1}{(2\pi)^{\frac{n}{2}} |\Gamma|} \exp\left(-\frac{1}{2} \sum_{l=1}^n (X^{(l)} - m)^T \Gamma^{-1} (X^{(l)} - m)\right),$$

and  $|\Gamma|$  is the determinant of  $\Gamma$ . Thus, for all  $\Gamma \in S_p^{++}(\mathbb{R})$ , we define

$$l_{\Gamma} := -\frac{2}{np} \log\left(L_{\Gamma, \bar{X}}(X^{(1)}, \dots, X^{(n)})\right) - \frac{n}{2} \log(2\pi) = \frac{1}{p} (\log |\Gamma| + \text{Tr}(\Gamma^{-1} S)).$$

As we assume that the true covariance matrix is block-diagonal, we consider a block-diagonal promoting penalization of the form

$$\text{pen}(\Gamma) := \text{pen}(B(\Gamma)) := \sum_{k=1}^K p_k^2,$$

if  $B(\Gamma) = \{B_1, \dots, B_K\}$  and  $|B_k| = p_k$  for all  $k \in [1 : K]$ . We consider the penalized likelihood criterion

$$\Phi : \begin{array}{ccc} S_p^{++}(\mathbb{R}) & \longrightarrow & \mathbb{R} \\ \Gamma & \longmapsto & l_\Gamma + \kappa \text{pen}(\Gamma), \end{array}$$

where  $\kappa \geq 0$ . In this work, we suggest to estimate  $\Sigma$  by the minimizer of  $\Phi$ , for some choice of penalisation  $\kappa$ . First, we show in Proposition 1 that a minimizer of  $\Phi$  can only be a block diagonal decomposition of  $S$ .

**Proposition 1.** *If  $\Gamma$  is a minimizer of  $\Phi$ , then, there exists  $B \in \mathcal{P}_p$  such that  $\Gamma = S_B$ .*

Hence, the minimization problem on  $S_p^{++}(\mathbb{R})$  becomes a minimization problem on the finite set  $\{S_B, B \in \mathcal{P}_p\}$ . So, we define  $\Psi(B) := \Phi(S_B)$  and we suggest to estimate  $B^*$  by

$$\widehat{B}_{tot} := \arg \min_{B \in \mathcal{P}_p} \Psi(B), \quad (1)$$

as the minimum structure of the penalized log-likelihood. In this paper, we study theoretically this estimator of  $B^*$ . However, it is unimplementable for in high dimension since the number of partitions  $B \in \mathcal{P}_p$  is too large. Hence, we will have to define other estimators less costly.

## 2.2 Convergence in high dimension

### 2.2.1 Assumptions

In Section 2.2, we assume that  $p$  and  $n$  go to infinity. The true covariance matrix  $\Sigma$  is not constant and depend on  $n$  (or  $p$ ). Nevertheless, to simplify notations, we do not write the dependency on  $n$ . In all Section 2.2, we choose a penalisation coefficient  $\kappa = \frac{1}{pn^\delta}$  for a fixed  $\delta \in ]1/2, 1[$ .

We also add the following assumptions on  $\Sigma$  along Section 2.2.

**Condition 1.**  $p/n \rightarrow y \in ]0, 1[$ .

**Condition 2.** *There exist  $\lambda_{\text{inf}} > 0$  and  $\lambda_{\text{sup}} < +\infty$  such that, for all  $n$ , the eigenvalues of  $\Sigma$  are in  $[\lambda_{\text{inf}}, \lambda_{\text{sup}}]$ .*

**Condition 3.** *There exists  $m \in \mathbb{N}^*$  such that for all  $n$ , all the blocks of  $\Sigma$  are smaller than  $m$ , i.e.  $\forall A \in B^*$ , we have  $|A| \leq m$ .*

**Condition 4.** *There exists  $a > 0$  such that for all  $n$  and for all  $B < B^*$ , we have  $\|\Sigma_B - \Sigma\|_{\text{max}} \geq an^{-1/4}$ .*

These four mild assumptions are discussed in Section 2.2.4. However, we also focus on the case when Condition 4 does not hold. We will provide similar results when only Conditions 1, 2 and 3 hold.

### 2.2.2 Convergence of $\widehat{B}$ and reduction of the cost

Now we defined our estimator  $\widehat{B}_{tot}$  of the true decomposition  $B^*$  in Equation 1 and we added assumptions in Section 2.2.1, we give the convergence of  $\widehat{B}$  in Proposition 2. Although  $\widehat{B}_{tot}$  is not computable in practice, its convergence remains interesting to strengthen the choice of the penalized likelihood criterion and will be useful to prove the convergence of more convenient estimators.

**Proposition 2.** *Under Conditions 1 to 4 and for a fixed  $\delta \in ]1/2, 1[$ , we have*

$$\mathbb{P}\left(\widehat{B}_{tot} = B^*\right) \rightarrow 1.$$

Hence, under Conditions 1 to 4, the estimator  $\widehat{B}_{tot}$  is equal to the true decomposition  $B^*$  with probability which goes to one. When Condition 4 does not hold, we can not state such a convergence result but we get a weaker result in Proposition 3. In this case, we need to define  $B(\alpha)$  as the partition given by thresholding  $\Sigma$  by  $n^{-\alpha}$ . In other words,  $B(\alpha)$  is the smallest (or finest) partition  $B$  such that  $\|\Sigma_B - \Sigma\|_{\max} \leq n^{-\alpha}$ .

**Proposition 3.** *Under Conditions 1, 2 and 3, for all  $\alpha_1 < \delta/2$  and  $\alpha_2 > \delta/2$ , we have*

$$\mathbb{P}\left(B(\alpha_1) \not\leq \widehat{B}_{tot} \leq B(\alpha_2)\right) \rightarrow 1.$$

Thus, when Conditions 1 to 4 hold, we defined a consistent estimator  $B^*$  that theoretically solves our problem of the lack of knowledge of the true decomposition  $B^*$ . However, computing  $\widehat{B}_{tot}$  is very costly in practice. Indeed, the number of partitions of  $[1 : p]$  (the Bell number) is exponential in  $p$ . As in [7], we suggest to restrict the research of  $B^*$  to the partitions given by thresholding the empirical correlation matrix  $\widehat{C} := (\widehat{C}_{ij})_{i,j \leq p}$  where  $\widehat{C}_{ij} := \widehat{\sigma}_{ij} / \sqrt{\widehat{\sigma}_{ii}\widehat{\sigma}_{jj}}$ . If  $\lambda \in [0, 1]$ , let  $B_\lambda$  the finest partition such that the thresholded empirical correlation matrix  $\widehat{C}_\lambda := (\widehat{C}_{ij} \mathbb{1}_{|\widehat{C}_{i,j}| > \lambda})_{i,j \leq p}$  has the decomposition  $B_\lambda$ . In other words,  $B_\lambda := B(\widehat{C}_\lambda)$ . For some value  $\lambda \in [0, 1]$ ,  $B_\lambda$  can be found by "Breath-First-Search" (BFS) [17]. Furthermore, we do not need to compute  $B_\lambda$  for all  $\lambda \in [0, 1]$  and we suggest in the following three different choice of grids for  $\lambda$ .

First, we suggest the grid  $A_{\widehat{C}} := \{|\widehat{C}_{ij}| \mid 1 \leq i < j \leq p\}$  and we define the estimator  $\widehat{B}_{\widehat{C}} := \arg \min_{B_\lambda \mid \lambda \in A_{\widehat{C}}} \Psi(B)$ . This grid is the finest one because, because that gives all the partitions  $\{B_\lambda \mid \lambda \in ]0, 1[ \}$ . Almost surely, the coefficients  $(\widehat{C}_{ij})_{i < j}$  are all different. Thus, when we increase the threshold to the next value of  $A_{\widehat{C}}$ , we only remove two symmetric coefficients from the correlation matrix.

**Proposition 4.** *The complexity of  $\widehat{B}_{\widehat{C}}$  is  $O(p^4)$ .*

As the cardinal of  $A_{\widehat{C}}$  is  $p(p-1)/2$  almost surely, we can choose a grid with a smaller cardinal.

Using the rate of convergence of the estimated covariances and by Condition 4, we suggest the estimator  $\widehat{B}_\lambda := B_{n^{-1/3}}$ , the partition of the empirical correlation matrix thresholded by  $n^{-1/3}$ . With this threshold, we can not find all the partitions given by thresholded correlation matrix, but we have to threshold by only one value.

**Proposition 5.** *The complexity of  $\widehat{B}_\lambda$  is  $O(p^2)$ .*

One can see that reducing the grid of thresholds to one value reduce the complexity of the estimator of  $B^*$ . Finally, we suggest a third grid when the maximal size of the groups  $m$  is known.

Let  $A_s := \{s/p, (s+1)/p, \dots, (p-1)/p, 1\}$ , where  $s$  is the smallest integer such that all the groups of  $B_{s/p}$  have a cardinal smaller than  $m$ . The deduced estimator is  $\widehat{B}_s := \arg \min_{B_\lambda \mid \lambda \in A_{s_n}} \Psi(B)$ . So, this grid is the set  $\{l/p \mid l \in [1 : p]\}$  restricted to the thresholds that give fine enough partition (with groups of size smaller than  $m$ ).

**Proposition 6.** *The complexity of  $\widehat{B}_s$  is  $O(p^2)$ .*

One can see that the complexity of this estimator is as small as the complexity of the previous estimator  $\widehat{B}_\lambda$ . Furthermore, it ensures that the estimated blocs are not too large, which was not the case with the previous estimator. However, the computation of  $\widehat{B}_s$  requires the knowledge of  $m$  while the other estimators do not.

Now we have defined new estimators of  $B^*$ , we give their convergence in the following proposition.

**Proposition 7.** *Let  $\widehat{B}$  be either  $\widehat{B}_{tot}$ ,  $\widehat{B}_{\widehat{C}}$ ,  $\widehat{B}_\lambda$  or  $\widehat{B}_s$  indifferently. Under Conditions 1 to 4 and for a fixed  $\delta \in ]1/2, 1[$ , we have*

$$\mathbb{P}(\widehat{B} = B^*) \longrightarrow 1.$$

When Condition 4 is not satisfied, we do not study the convergence of the previous estimators. In this case, we suggest to estimate  $B^*$  by  $B_{n^{-\delta/2}}$ , which is the partition given by the empirical correlation matrix thresholded by  $n^{-\delta/2}$ . The complexity of this estimator is  $O(p^2)$ , as for the previous estimator  $\widehat{B}_\lambda = B_{n^{-1/3}}$ . We show the convergence of this estimator in Proposition 8.

**Proposition 8.** *Under Conditions 1, 2 and 3, if  $\alpha_1 < \delta/2$  and  $\alpha_2 > \delta/2$ ,*

$$\mathbb{P}(B(\alpha_1) \leq B_{n^{-\delta/2}} \leq B(\alpha_2)) \longrightarrow 1.$$

As Condition 4 is not satisfied, the true partition  $B^*$  is again not reached by this estimator. Nevertheless, we get stronger results for the practical estimator  $B_{n^{-\delta/2}}$  than for the theoretical estimator  $\widehat{B}_{tot}$  when Condition 4 is not verified. Actually, the condition "to be larger or equal than" is stronger than "not to be smaller than".



### 2.2.3 Convergence of the estimator of the covariance matrix

We have seen in Propositions 7 and 8 how to estimate the decomposition  $B^*$  by  $\widehat{B}$ . Now to estimate the covariance matrix  $\Sigma$ , it suffices to impose the block-diagonal decomposition  $\widehat{B}$  to the empirical covariance matrix  $S_{\widehat{B}}^{(\prime)}$ . We show in Proposition 9 that the resulting block-diagonal matrix estimator  $S_{\widehat{B}}^{(\prime)}$  reaches the best rate of convergence under Conditions 1 to 4.

**Proposition 9.** *Let  $\|\cdot\|_F$  be the Frobenius norm defined by  $\|\Gamma\|_F^2 := \sum_{i,j=1}^p \gamma_{ij}^2$ . Under Conditions 1 to 4 and for a fixed  $\delta \in ]1/2, 1[$ , we have*

$$\frac{1}{p} \|S_{B^*} - \Sigma\|_F^2 = O_p(1/n),$$

and

$$\frac{1}{p} \|S_{\widehat{B}} - \Sigma\|_F^2 = O_p(1/n),$$

Moreover, it is the best rate that we can have because

$$\frac{1}{p} \|S_{B^*} - \Sigma\|_F^2 \neq o_p(1/n).$$

Thus, we see that the quantity  $\frac{1}{p} \|S_{\widehat{B}} - \Sigma\|_F^2$  decreases to 0 in probability with rate  $1/n$ , which is the same rate that  $S_{B^*}$  if we know the true decomposition  $B^*$ . Thus, the lack of knowledge of  $B^*$  does not deteriorate the convergence of our estimator.

Now we gave the rate of convergence of our estimator  $S_{\widehat{B}}$ , we compare it with that of the empirical estimator  $S$  in Proposition 10.

**Proposition 10.** *The rate of the empirical covariance is*

$$\frac{1}{p} \|S - \Sigma\|_F^2 = O_p(p/n).$$

and we have

$$\mathbb{E} \left( \frac{1}{p} \|S - \Sigma\|_F^2 \right) \geq \frac{\lambda_{\inf}^2 p}{2n}.$$

So, we know that  $\frac{1}{p} \|S - \Sigma\|_F^2$  is lower-bounded in average and is bounded in probability. Thus, the rate of convergence of our suggested estimator  $S_{\widehat{B}}$  is better than the empirical covariance matrix  $S$ .

If Condition 4 does not hold, the rate of convergence is given in the following proposition.

**Proposition 11.** *Under Conditions 1,2 and 3, for all penalization coefficient  $\delta \in ]0, 1[$  and for all  $\varepsilon > 0$ , we have*

$$\frac{1}{p} \|S_{B_{n^{-\delta/2}}} - \Sigma\|_F^2 = o_p \left( \frac{1}{n^{\delta-\varepsilon}} \right).$$

We remark that this rate of convergence almost reaches the best rate of  $S_{B^*}$ , whereas the partition estimator  $B_{n-\delta/2}$  does not reach the true decomposition  $B^*$ . That comes from the fact that the elements  $\sigma_{ij}$  of  $\Sigma$  such that the indices  $(i, j)$  are not in the estimated partition  $B_{n-\delta/2}$  are small (with high probability). Hence, estimating these values by 0 does not increase so much the error  $\frac{1}{p} \|S_{B_{n-\delta/2}} - \Sigma\|_F^2$ .

Theoretical guaranties for a block-diagonal estimator of the covariance matrix are also provided in [7]. Their framework is more general, with a true covariance matrix which is not necessarily block-diagonal. They bound the average of the square Hellinger distance between the true normal density and the density with the block-diagonal estimated covariance matrix. However, when  $p/n$  goes to  $+\infty$ , their theoretical results do not hold. Indeed, they give an upper-bound which goes to  $+\infty$ , while the square Hellinger distance remains always smaller than 1.

#### 2.2.4 Discussion about the assumptions

For the previous results, we needed to make four assumptions on  $\Sigma$  (Conditions 1 to 4, given in Section 2.2.1).

Condition 1 provides a standard setting for high-dimensional problems, in particular for estimation of covariance matrices [20, 27]. Studying an higher dimensional setting where  $p/n \rightarrow +\infty$  would be interesting in future work.

Condition 2 is needed to bound the operator norm of  $\Sigma$  and  $\Sigma^{-1}$  and the eigenvalues of the empirical covariance matrix (with high probability). It also enables to bound the diagonal term of  $\Sigma$  and so the rate of convergence of each component of the empirical covariance (using in particular Bernstein's inequality, see the proofs for more details).

Condition 3 states that the blocks of the true decomposition have a maximal size. It implies that the number of non-zero term of  $\Sigma$  is  $O(p)$ .

Condition 4 requires that a finer block decomposition  $\Sigma_B$  is not too close to the true  $\Sigma$ . This condition is needed to not confuse  $B^*$  with a finer decomposition. However, Condition 4 seems to be less mild than the others. That is why we also focus on the case when Condition 4 is not satisfied.

Nevertheless, even Condition 4 is not so restrictive. Indeed, we suggest in Proposition 12 a reasonable way to generate  $\Sigma$  which satisfies a slightly different Condition 4.

**Proposition 12.** *Let  $L \in \mathbb{N}$ . Assume that for all  $p$ ,  $\Sigma$  is generated in the following way:*

- *Let  $B^*$  be a partition of  $[1 : p]$  such that all its elements have a cardinal between 10 and  $m \geq 10$ . Let  $K$  be the number of groups (the cardinal of  $B^*$ ). For all  $k \in [1 : K]$ , let  $p_k$  be the cardinal of the " $k$ -th element" of  $B^*$ .*

- For all  $k \in [1 : K]$ , let  $(U_i^{(l)})_{i \leq p_k, l \leq L}$  be i.i.d. with distribution  $\mathcal{U}([-1, 1])$ . Let  $U \in \mathcal{M}_{L, p_k}(\mathbb{R})$  such that the coefficient  $(l, i)$  is  $U_i^{(l)}$ . Let  $\Sigma_{B_k^*} = U^T U + \varepsilon I_{p_k}$ , where  $\Sigma_{B_k^*}$  is the sub-matrix of  $\Sigma$  indexed by the elements of  $B_k^*$ .
- Let  $\sigma_{ij} = 0$  for all  $(i, j) \notin B^*$ .

Then, Conditions 2 and 3 are verified and the following slightly modified version of Condition 4 is satisfied for all  $a > 0$ :

$$\mathbb{P} \left( \exists B < B^*, \|\Sigma_B - \Sigma\|_{\max} < an^{-\frac{1}{4}} \right) \rightarrow 0.$$

Thus, if  $p/n \rightarrow y \in ]0, 1[$ , the conclusions of Propositions 2, 7 and 9 remain true when the probabilities are defined with respect to  $\Sigma$  and on  $X$  which distribution conditionally to  $\Sigma$  is  $\mathcal{N}(\mu, \Sigma)$ .

### 2.2.5 Numerical applications

We present here numerical applications of the previous results with simulated data. We generate a covariance matrix  $\Sigma$  as in Proposition 12 with blocks of random size distributed uniformly on  $[10 : 15]$ , with  $L = 5$  and  $\varepsilon = 0.2$ . We assume here that we know that the maximal size of the block is  $m = 15$ , so we can use the estimator  $\hat{B} = \hat{B}_s$  given in Proposition 7 to reduce the complexity to  $O(p^2)$  and to prevent the blocks from being too large.

We plot in Figure 1 the Frobenius norm of the error of the empirical covariance matrix  $S$  and the Frobenius norm of the error of the suggested estimator  $S_{\hat{B}}$ , with  $n = Np$  for different values of  $N$ . We can remark that the error of  $S$  is in  $\sqrt{K}$  (where  $K$  is the number of groups) whereas the error of  $S_{\hat{B}}$  stays bounded as in Proposition 9. For  $K = 100$ , the Frobenius error of  $S_{\hat{B}}$  on Figure 1 is about 10 times smaller than the one of  $S$ .

## 2.3 Convergence and efficiency in fixed dimension

In this section,  $p$  and  $\Sigma$  are fixed and  $n$  go to  $+\infty$ . We choose a different penalisation  $\kappa = \frac{1}{pn^\delta}$  with  $\delta \in ]0, 1/2[$  (instead of  $\delta \in ]1/2, 1[$  in the previous setting). This framework enables to study the efficiency of estimators of  $\Sigma$ . Contrary to the high-dimensional setting of Section 2.2, we do not assume particular condition in addition to the ones given in Section 2.1.

We first give the convergence of  $\hat{B}_{tot}$  defined in Equation (1) in the next proposition.

**Proposition 13.** *We have*

$$\mathbb{P} \left( \hat{B}_{tot} = B^* \right) \rightarrow 1.$$

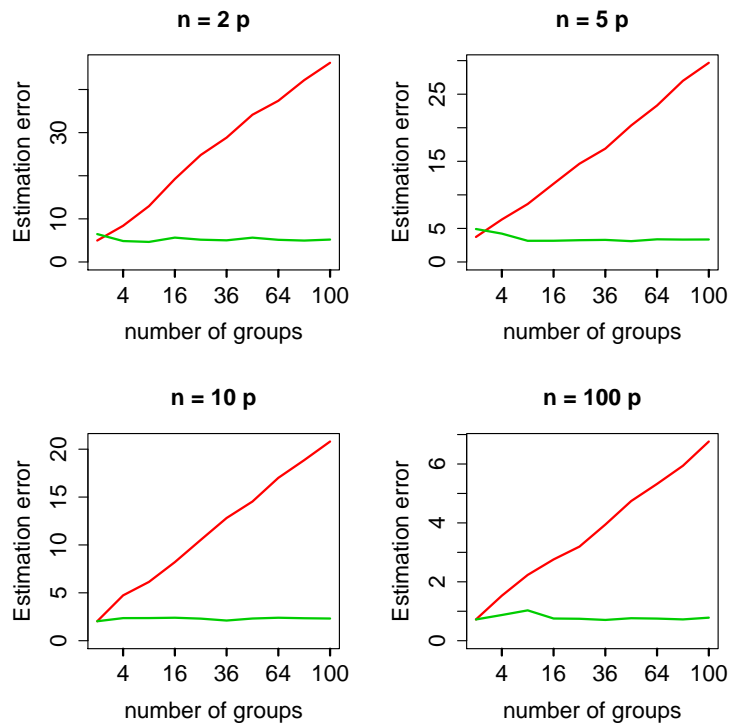


Figure 1: Frobenius error of the empirical covariance matrix  $S$  in red and the suggested estimator  $S_{\hat{B}}$  in green, in function of the number of groups  $K$ . The scale of the x-axis is in  $\sqrt{K}$ .

**Corollary 1.** Let  $\widehat{B}_{\widehat{C}} := \arg \min_{B_\lambda \mid \lambda \in A_{\widehat{C}}} \Psi(B)$ , where  $A_{\widehat{C}} := \{|\widehat{C}_{ij}| \mid 1 \leq i < j \leq p\}$  as in Proposition 7. Then

$$\mathbb{P} \left( \widehat{B}_{\widehat{C}} = B^* \right) \rightarrow 1.$$

In the rest of Section 2.3, we write  $\widehat{B}$  for  $\widehat{B}_{tot}$  or  $\widehat{B}_{\widehat{C}}$ . The aim of this framework is to show that the suggested estimator  $S_{\widehat{B}}$  is asymptotically efficient as if the true decomposition  $B^*$  were known.

As the parameter  $\Sigma$  is in the set  $S_p^{++}(\mathbb{R})$  or even  $S_p^{++}(\mathbb{R}, B^*)$ , which are not open subsets of  $\mathbb{R}^{p^2}$ , the classical Cramér-Rao bound is no longer a lower-bound for the estimation error. Furthermore, as  $B^*$  is not known, the number of parameters of  $S_{\widehat{B}}$  is not constant. That is why the classical Cramér-Rao bound is not relevant in our setting. We remark that applying this classical Cramér-Rao bound to a subset of the matrix estimator does not solve this problem.

A specific Cramér-Rao bound is suggested in [30] for parameters and estimators which satisfy continuously differentiable constraints. We shall consider linear constraints here. We let  $\theta \in \mathbb{R}^d$  be the parameter, that is assumed to be restricted to a linear subspace  $V$  of dimension  $q$  in  $\mathbb{R}^d$ . In this case, if  $U \in \mathcal{M}_{d,q}(\mathbb{R})$  is a matrix whose columns are the elements of an orthonormal basis of  $V$  and if  $J$  is the Fisher Information Matrix (FIM) of  $\theta$  in the non-constraint case, [30] states that for unbiased estimator  $\widehat{\theta} \in V$ , we have

$$\mathbb{E} \left[ (\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^T \right] \leq U(U^T J U)^{-1} U^T, \quad (2)$$

where  $\leq$  is the partial order on the symmetric positive semi-definite matrices.

In our setting, remark that  $S_p^{++}(\mathbb{R})$  is an open subset of the linear subspace  $S_p(\mathbb{R})$  of symmetric matrices and  $S_p^{++}(\mathbb{R}, B^*)$  is an open subset of the linear subspace  $\overline{S_p(\mathbb{R}, B^*)} := \{\Gamma \in S_p(\mathbb{R}), \Gamma_{B^*} = \Gamma\}$ . We let  $\text{vec}(\Sigma)$  be the column vectorization of  $\Sigma$ . Hence, the parameter is  $\text{vec}(\Sigma)$  and there are  $p(p-1)/2$  linear constraints arising from the symmetry and  $p(p-1)/2 - \sum_{k=1}^K p_k(p_k-1)/2$  linear constraints arising from the block structure  $B^*$ .

So, the Cramér-Rao bound of Equation (2) is adapted to our framework, by considering the parameter  $\text{vec}(\Sigma) \in \mathbb{R}^{p^2}$ , and we say that an estimator is efficient if it reaches the Cramér-Rao bound (2) (meaning that there is an equality in this equation), where the constraints (symmetry only or symmetry and block structure) will be stated explicitly.

Proposition 14 states that, in general, the empirical covariance matrix is efficient with this Cramér-Rao bound. This supports this choice of Cramér-Rao Bound, since in fixed dimension, one would expect that the empirical matrix is the most appropriate estimator.

If the empirical covariance matrix did not reach the Cramér-Rao Bound, we could not hope that  $S_{\widehat{B}}$  would be efficient in the model where  $B^*$  was known, and this Cramér-Rao bound would be not well tuned to our problem.

**Proposition 14.** *If  $\mu$  is known, the empirical estimator  $S$  is an efficient estimator of  $\Sigma$  in the model  $\{\mathcal{N}(\mu, \Sigma), \Sigma \in S_p^{++}(\mathbb{R})\}$ .*

**Remark 1.** *In Proposition 14, we assume that  $\mu$  is known to reach the Cramér-Rao bound for fixed  $n$  (and not only asymptotically). This will be the same in Proposition 15.*

Now, we deduce the efficiency of  $S_{B^*}$  when  $B^*$  is known.

**Proposition 15.** *If  $\mu$  and  $B^*$  are known,  $S_{B^*}$  is an efficient estimator of  $\Sigma$  in the model  $\{\mathcal{N}(0, \Sigma), \Sigma \in S_p^{++}(\mathbb{R}, B^*)\}$ .*

Finally, Proposition 16 states the asymptotic efficiency of our estimator  $S_{\hat{B}}$  (even for unknown  $\mu$ )

**Proposition 16.**

$$\sqrt{n}(\text{vec}(S_{\hat{B}}) - \text{vec}(\Sigma)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \text{CR}(\Sigma, B^*)),$$

where  $\text{CR}(\Sigma, B^*)$  is the Cramér-Rao bound of  $\text{vec}(\Sigma)$  in the model  $\{\mathcal{N}(0, \Sigma), \Sigma \in S_p^{++}(\mathbb{R}, B^*)\}$ .

The explicit expression of the  $p^2 \times p^2$  matrix  $\text{CR}(\Sigma, B^*)$  can be found in the appendix where Propositions 14, 15 and 16 are proved.

### 3 Application to the estimation of the Shapley effects

In this section, we apply the block-diagonal estimation of the covariance matrix  $\Sigma$  to estimate the Shapley effects in high dimension and for Gaussian linear models. In Section 3.1, we recall the definition of the Shapley effects with their particular expression in the Gaussian linear framework with a block-diagonal covariance matrix. In Section 3.2, we address the problem of estimating the Shapley effects when the covariance matrix  $\Sigma$  is estimated. We derive the convergence the estimators of Shapley effects from the results of Section 2.

#### 3.1 The Shapley effects

Let  $(X_i)_{i \in [1:p]}$  be random inputs variables on  $\mathbb{R}^p$  and let  $Y = f(X)$  be the real random output variable in  $L^2$ . We assume that  $\text{Var}(Y) \neq 0$ . Here,  $f$  can be a numerical simulation model [26].

If  $u \subseteq [1 : p]$  and  $x = (x_i)_{i \in [1:p]} \in \mathbb{R}^p$ , we write  $x_u := (x_i)_{i \in u}$ . We can define the Shapley effects as in [21] for the input variable  $X_i$  as:

$$\eta_i := \frac{1}{p \text{Var}(Y)} \sum_{u \subseteq -i} \binom{p-1}{|u|}^{-1} (\text{Var}(\mathbb{E}(Y|X_{u \cup \{i\}})) - \text{Var}(\mathbb{E}(Y|X_u))) \quad (3)$$

where  $-i$  is the set  $[1 : p] \setminus \{i\}$ . One can see in Equation (3) that adding a  $X_i$  to  $X_u$  changes the conditional expectation of  $Y$ , and increases the variability of this conditional expectation. The Shapley effect  $\eta_i$  is large when, on average, the variance of this conditional expectation increases significantly when  $X_i$  is observed. Thus, a large Shapley effect  $\eta_i$  corresponds to an important input variable  $X_i$ .

The Shapley effects have interesting properties for global sensitivity analysis. Indeed, there is only one Shapley effect for each variable (contrary to the Sobol indices). Moreover, the sum of all the Shapley effects is equal to 1 (see [21]) and all these values lie in  $[0, 1]$  even with dependent inputs. This is very convenient for the interpretation of these sensitivity indices.

Here, we assume that  $X \sim \mathcal{N}(\mu, \Sigma)$ , that  $\Sigma \in S_p^{++}(\mathbb{R})$  and that the model is linear, that is  $f : x \mapsto \beta_0 + \beta^T x$ , for a fixed  $\beta_0 \in \mathbb{R}$  and a fixed vector  $\beta$ . This framework is widely used to model physical phenomena (see for example [15], [11], [24]). Indeed, uncertainties are often modelled as Gaussian variables and an unknown function is commonly estimated by its linear approximation. Furthermore, the main focus on this paper is on the high-dimensional case, where  $p$  is large. In high dimension, linear models are often considered, as more complex models are not necessarily more relevant. In this framework, the sensitivity indices can be calculated explicitly [22]:

$$\eta_i := \frac{1}{p \text{Var}(Y)} \sum_{u \subseteq -i} \binom{p-1}{|u|}^{-1} (\text{Var}(Y|X_u) - \text{Var}(Y|X_{u \cup \{i\}})). \quad (4)$$

with

$$\text{Var}(Y|X_u) = \text{Var}(\beta_{-u}^T X_{-u} | X_u) = \beta_{-u}^T (\Sigma_{-u, -u} - \Sigma_{-u, u} \Sigma_{u, u}^{-1} \Sigma_{u, -u}) \beta_{-u} \quad (5)$$

where  $\beta_u := (\beta_i)_{i \in u}$  and  $\Gamma_{u, v} := (\Gamma_{i, j})_{i \in u, j \in v}$ . Thus, in the Gaussian linear framework, the Shapley effects are functions of the parameters  $\beta$  and  $\Sigma$ .

In this paper, we assume that  $n > p$ , so  $\beta$  is known. Indeed,  $(\beta_0 \ \beta^T)^T$  is given by  $(A^T A)^{-1} A^T Y$ , where  $A \in \mathcal{M}_{n, p+1}(\mathbb{R})$  is defined by  $A_{i, i+1} := X_i^{(l)}$  and  $A_{i, 1} = 1$ , and  $Y \in \mathcal{M}_{n, 1}(\mathbb{R})$  is defined by  $Y_l := Y^{(l)}$ .

Despite the analytical formula (5), even in the case where  $\Sigma$  is known, the computational cost of the Shapley effects remains an issue when the number of input variables  $p$  is too large ( $p \geq 30$ ), as it is highlighted in [4]. Indeed, the Shapley effects depend on  $2^p$  values, namely the  $(\text{Var}(Y|X_u))_{u \subseteq [1: p]}$ . However, when the covariance matrix is block-diagonal, [4] showed that this high dimensional computational problem boils down to a collection of lower dimensional problems.

Indeed, assume that  $\Sigma \in S_p^{++}(\mathbb{R}, B^*)$  with  $B^* = \{B_1^*, B_2^*, \dots, B_K^*\}$ . If  $i \in [1 : p]$ , let  $[i]$  denotes the group of  $i$ , that is  $i \in B_{[i]}$ . Using Corollary 2 of [4], we have for all  $i \in [1 : p]$ ,

$$\eta_i = \frac{1}{\beta^T \Sigma \beta} \frac{1}{|B_{[i]}|} \sum_{u \subseteq B_{[i]}^* - i} \binom{|B_{[i]}| - 1}{|u|}^{-1} (V_u^{B_{[i]}^*} - V_{u \cup \{i\}}^{B_{[i]}^*}), \quad (6)$$

where for all  $v \subseteq B_{[i]}^*$ ,

$$V_v^{B_{[i]}^*} := \text{Var} \left( \beta_{B_{[i]}^*}^T X_{B_{[i]}^*} | X_v \right) = \beta_{B_{[i]}^* - v}^T \left( \Sigma_{B_{[i]}^* - v, B_{[i]}^* - v} - \Sigma_{B_{[i]}^* - v, v} \Sigma_{v, v}^{-1} \Sigma_{v, B_{[i]}^* - v} \right) \beta_{B_{[i]}^* - v}. \quad (7)$$

Thus, when  $\Sigma$  is known, to compute all the Shapley effects  $(\eta_i)_{i \in [1:p]}$ , we only have to compute the  $\sum_{k=1}^K 2^{|B_k^*|}$  values  $\{\text{Var}(Y|X_u), u \subseteq B_k^*, k \in [1:K]\}$  instead of all the  $2^p$  values  $\{\text{Var}(Y|X_u), u \subseteq [1:p]\}$ . Some numerical experiments highlighting this gain are given in [4]. Because of the matrix inversion in Equation 6, the complexity of the computation of the Shapley effects is  $O(K2^m m^3)$ , where  $m$  denotes the size of the maximal group in  $B^*$ .

If  $\Sigma$  is known, but the decomposition  $B^*$  is unknown, we can compute  $B^*$  from  $\Sigma$ . We can for example use "Breath-First-Search" (BFS). The complexity of this algorithm is in  $O(pm^2)$ .

To conclude, when the parameters  $\beta$  and  $\Sigma$  are known with  $\Sigma \in S_p^{++}(\mathbb{R}, B^*)$ , the computation of all the Shapley effects has a complexity  $O(K2^m m^3)$ .

### 3.2 Estimation of the Shapley effects in high dimension

We now address the problem when the parameters  $\mu$ ,  $\Sigma$  and thus  $B^*$  are unknown, and we just observe a sample  $(X^{(l)}, Y^{(l)})_{l \leq n}$  (or  $(X^{(l)})_{l \leq n}$  and  $\beta$ ).

At first glance, we could estimate  $\Sigma$  by the empirical covariance matrix  $S$  and replace it in the computation of the Shapley effects given by Equations (4) and (5) or (6) and (7). However,  $B^*$  is not known and we can not find it using BFS with the empirical covariance matrix  $S$  (which usually has the simple structure  $\{[1:p]\}$  with probability one). Thus, we can not use the formula (6) of the Shapley effects with independent groups. So, the only way to estimate the Shapley effects is using Equations (4) and (5), replacing  $\Sigma$  by the empirical covariance matrix  $S$ . However, as we have seen, the complexity of this computation would be exponential in  $p$  and it would be no more tractable for  $p \geq 30$ . Furthermore, in high dimension, the Frobenius error between  $S$  and  $\Sigma$  does not go to 0 (see Proposition 9). Thus, using the empirical covariance matrix could yield to estimators of the Shapley effects that do not converge.

For that reason, to estimate  $\eta = (\eta_i)_{i \leq p}$ , when Conditions 1 to 4 hold, we suggest to estimate  $B^*$  by  $\widehat{B}$  (defined in Section 2.2.2) and  $\Sigma$  by  $S_{\widehat{B}}$  and to replace them in the analytical formula (6). We write  $\widehat{\eta} = (\widehat{\eta}_i)_{i \leq p}$  this estimator. We use our previous results on the estimation of the covariance matrix to obtain the convergence rate of  $\widehat{\eta}$ .

We focus on the high dimensional case, when  $p$  and  $n$  go to  $+\infty$ . In this case,  $\beta$  and  $\Sigma$  are not fixed but depend on  $n$  (or  $p$ ). As in Section 2.2, we choose  $\kappa = \frac{1}{pn^\delta}$  with  $\delta \in ]1/2, 1[$  to compute  $\widehat{B}$ . To prevent problematic cases, we also add an assumption on the vector  $\beta$ .

**Condition 5.** *There exist  $\beta_{\text{inf}} > 0$  and  $\beta_{\text{sup}} < +\infty$  such that for all  $n$  and for all  $j \leq p$ , we have  $\beta_{\text{inf}} \leq |\beta_j| \leq \beta_{\text{sup}}$ .*



**Proposition 17.** *Under Conditions 1 to 5 and if  $\delta \in ]1/2, 1[$ , then for all  $\gamma > 1/2$ , we have*

$$\sum_{i=1}^p |\hat{\eta}_i - \eta_i| = o_p \left( \frac{\log(n)^\gamma}{\sqrt{n}} \right).$$

Recall that  $\sum_{i=1}^p \eta_i = 1$ . Thus, to quantify the error estimation, the value of  $\sum_{i=1}^p |\hat{\eta}_i - \eta_i|$  is a relative error. Proposition 17 states that this relative error goes to zero at the parametric rate  $1/n^{1/2}$ , up to a logarithm factor.

We have seen in Section 3.1 that, once we have the block-diagonal covariance matrix, the computation of the Shapley effects has the complexity  $O(K2^m m^3)$  which is equal to  $O(n)$  under Condition 4. In section 2.2, we gave four different choices of  $\hat{B}$ , with four different complexities, all larger than  $O(n)$ . Thus, the complexity of the whole estimation of the Shapley effects (including the estimation of  $\Sigma$ ) is the same that the complexity of  $\hat{B}$  (see Section 2.2.2).

When Condition 4 is not satisfied, we still have the convergence of the relative error, with almost the same rate.

**Proposition 18.** *Under Conditions 1, 2, 3 and 5, for all penalization coefficient  $\delta \in ]1/2, 1[$ , choosing the partition  $B_{n-\delta/2}$  and for all  $\varepsilon > 0$ , we have*

$$\sum_{i=1}^p |\hat{\eta}_i - \eta_i| = o_p \left( \frac{1}{n^{-(\delta-\varepsilon)/2}} \right).$$

**Remark 2.** *When the dimension  $p$  is fixed, the rate of convergence is  $O_p(1/\sqrt{n})$ , as if we estimated  $\Sigma$  by the empirical covariance matrix. Moreover, we have seen in Proposition 16 that the computation of  $S_{\hat{B}}$  enables to reach asymptotically the Cramér-Rao bound of [30] as if  $B^*$  were known. We then deduce the asymptotic efficiency of  $\hat{\eta}$ . If we define  $g : \Sigma \mapsto \eta$ , let  $\text{CR}(\eta, B^*) := \text{Dg}(\Sigma)\text{CR}(\Sigma, B^*)\text{Dg}(\Sigma)$  be the Cramér-Rao bound of  $\eta$  in the model  $\{\mathcal{N}(\mu, \Sigma), \Sigma \in S_p^{++}(\mathbb{R}, B^*)\}$ . Thus,*

$$\sqrt{n}(\hat{\eta} - \eta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \text{CR}(\eta, B^*)).$$

### 3.3 Numerical application

We have seen in Proposition 12 a way to generate  $\Sigma$  which verifies Conditions 1 to 3 and some slightly modified version of Condition 4. So, with this choice of  $\Sigma$ , we derive in Proposition 19 the convergence of the Shapley effects estimation.

**Proposition 19.** *Under Condition 5, if  $\Sigma$  is generated as in Proposition 12, then*

$$\sum_{i=1}^p |\hat{\eta}_i - \eta_i| = o_p \left( \frac{\log(n)^\gamma}{\sqrt{n}} \right).$$

*when the probabilities are defined with respect to  $\Sigma$  and  $X$ , which distribution conditionally to  $\Sigma$  is  $\mathcal{N}(\mu, \Sigma)$ .*

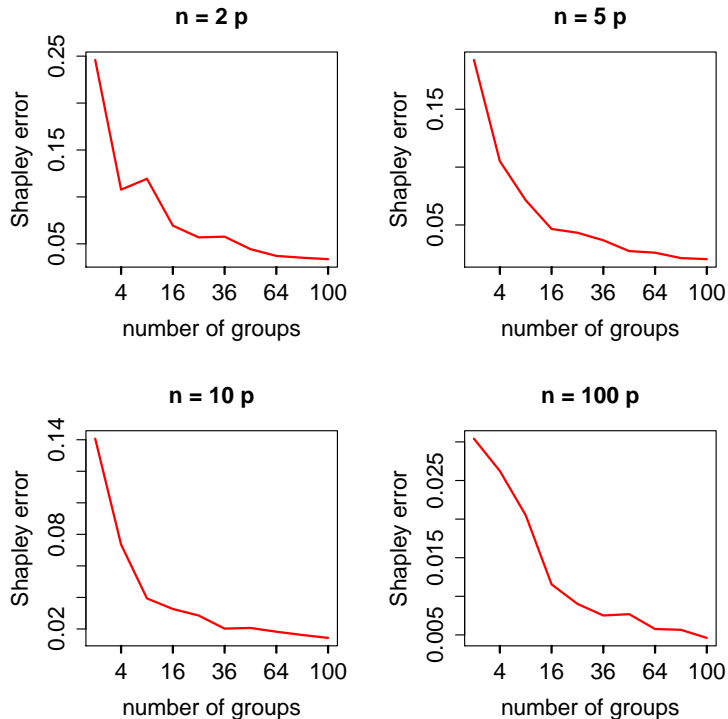


Figure 2: Sum of errors of the Shapley effects estimations in function of the number of groups  $K$ . The scale of the x-axis is in  $\sqrt{K}$ .

We now present a numerical application of Proposition 19. The matrix  $\Sigma$  is generated by Proposition 12 as in Section 2.2.5, with blocks of random size distributed uniformly on  $[10, 15]$ ,  $L = 5$  and  $\varepsilon = 0.2$ . For all  $p$ , the vector  $\beta$  is generated with distribution  $\mathcal{U}([1, 2]^p)$ , so satisfies Condition 5. As in Section 2.2.5, we assume that we know that the maximal size of the block is  $m = 15$ , so we can use the estimator  $\hat{B} = \hat{B}_s$  given in Proposition 7. As the computation of the Shapley effects is exponential in the maximal block size, the estimator  $\hat{B}_s$  is preferred. The complexity of the estimation of the Shapley effects is then in  $O(p^2)$ .

We plot in Figure 2 the sum of the Shapley effects estimation error  $\sum_{i=1}^p |\hat{\eta}_i - \eta_i|$ , with  $n = Np$  for different values of  $N$ . We can remark that the sum of the errors seems to be in  $1/\sqrt{K}$ , which is confirmed by Proposition 19.

## 4 Conclusion

In this work, we suggest an estimator of a block-diagonal covariance matrix for Gaussian data. We prove that in high dimension, this estimator converges to

the same block-diagonal structure with complexity in  $O(p^2)$ . For fixed dimension, we also prove the asymptotic efficiency of this estimator, that performs asymptotically as well as as if the true block-diagonal structure were known. Then, we deduce a convergent estimator of the Shapley effects in high dimension for Gaussian linear models. This estimator is still available for thousands input variables, as long as the maximal block is not too large.

In future works, it would be interesting to generalize the convergence of the Shapley effects estimators when the observations of the output are noisy and so the parameter  $\beta$  is estimated. Finally, it remains to treat the higher dimension setting when  $p/n$  goes to  $+\infty$ .

## Acknowledgements

We acknowledge the financial support of the Cross-Disciplinary Program on Numerical Simulation of CEA, the French Alternative Energies and Atomic Energy Commission. We would like to thank BPI France for co-financing this work, as part of the PIA (Programme d'Investissements d'Avenir) - Grand Défi du Numérique 2, supporting the PROBANT project. We thank Vincent Prost for his help.

## References

- [1] BAI, Z., AND SILVERSTEIN, J. W. *Spectral Analysis of Large Dimensional Random Matrices*, 2 ed. Springer Series in Statistics. Springer-Verlag, New York, 2010.
- [2] BICKEL, P. J., AND LEVINA, E. Regularized estimation of large covariance matrices. *The Annals of Statistics* 36, 1 (Feb. 2008), 199–227.
- [3] BROTO, B., BACHOC, F., AND DEPECKER, M. Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. Dec. 2018.
- [4] BROTO, B., BACHOC, F., DEPECKER, M., AND MARTINEZ, J.-M. Sensitivity indices for independent groups of variables. *Mathematics and Computers in Simulation* 163 (Sept. 2019), 19–31.
- [5] BÜHLMANN, P., AND VAN DE GEER, S. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.
- [6] CHASTAING, G. *Indices de Sobol généralisés pour variables dépendantes*. phdthesis, Université de Grenoble, Sept. 2013.
- [7] DEVIJVER, E., AND GALLOPIN, M. Block-diagonal covariance selection for high-dimensional gaussian graphical models. *Journal of the American Statistical Association* 113, 521 (2018), 306–314.

- [8] EL KAROUI, N. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics* 36, 6 (2008), 2717–2756.
- [9] GIRAUD, C. *Introduction to high-dimensional statistics*. Chapman and Hall/CRC, 2014.
- [10] GUPTA, A., AND NAGAR, D. *Matrix Variate Distributions*, vol. 104. CRC Press, 1999.
- [11] HAMMER, H., AND TJELMELAND, H. Approximate forward-backward algorithm for a switching linear Gaussian model. *Computational Statistics & Data Analysis* 55, 1 (Jan. 2011), 154–167.
- [12] HUANG, J. Z., LIU, N., POURAHMADI, M., AND LIU, L. Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* 93, 1 (2006), 85–98.
- [13] HYODO, M., SHUTOH, N., NISHIYAMA, T., AND PAVLENKO, T. Testing block-diagonal covariance structure for high-dimensional data. *Statistica Neerlandica* 69, 4 (2015), 460–482.
- [14] IOOSS, B., AND PRIEUR, C. Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol’ indices, numerical estimation and applications. *arXiv:1707.01334 [math, stat]* (July 2017).
- [15] KAWANO, T., HANSON, K. M., FRANKLE, S., TALOU, P., CHADWICK, M. B., AND LITTLE, R. C. Evaluation and Propagation of the  $^{239}\text{Pu}$  Fission Cross-Section Uncertainties Using a Monte Carlo Technique. *Nuclear Science and Engineering* 153, 1 (May 2006), 1–7.
- [16] LAM, C., AND FAN, J. Sparsistency and rates of convergence in large covariance matrix estimation. *Annals of statistics* 37, 6B (2009), 4254.
- [17] LEE, C. Y. An Algorithm for Path Connections and Its Applications. *IRE Transactions on Electronic Computers EC-10*, 3 (Sept. 1961), 346–365.
- [18] MARA, T. A., AND TARANTOLA, S. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety* 107 (Nov. 2012), 115–121.
- [19] MARA, T. A., TARANTOLA, S., AND ANNONI, P. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental Modelling & Software* 72 (Oct. 2015), 173–183.
- [20] MARČENKO, V. A., AND PASTUR, L. A. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* 1, 4 (apr 1967), 457–483.
- [21] OWEN, A. B. Sobol’ Indices and Shapley Value. *SIAM/ASA Journal on Uncertainty Quantification* 2, 1 (Jan. 2014), 245–251.

- [22] OWEN, A. B., AND PRIEUR, C. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification* 5, 1 (2017), 986–1002.
- [23] PERROT-DOCKÈS, M., AND LÉVY-LEDUC, C. Estimation of large block covariance matrices: Application to the analysis of gene expression data. *arXiv preprint arXiv:1806.10093* (2018).
- [24] ROSTI, A.-V. I. *Linear Gaussian models for speech recognition*. PhD thesis, University of Cambridge, 2004.
- [25] ROTHMAN, A. J., BICKEL, P. J., LEVINA, E., AND ZHU, J. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* 2 (2008), 494–515.
- [26] SANTNER, T. J., WILLIAMS, B. J., NOTZ, W., AND WILLIAMS, B. J. *The design and analysis of computer experiments*, vol. 1. Springer, 2003.
- [27] SILVERSTEIN, J. The smallest eigenvalue of a large dimensional wishart matrix. *The annals of probability* 13, 4 (1985), 1364–1368.
- [28] SOBOL, I. M. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments* 1, 4 (1993), 407–414.
- [29] SONG, E., NELSON, B. L., AND STAUM, J. Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *SIAM/ASA Journal on Uncertainty Quantification* 4, 1 (Jan. 2016), 1060–1083.
- [30] STOICA, P., AND NG, B. C. On the Cramer-Rao bound under parametric constraints. *IEEE Signal Processing Letters* 5, 7 (1998), 177–179.
- [31] VERSHYNIN, R. *High-dimensional probability: An introduction with applications in data science*, vol. 47. Cambridge University Press, 2018.
- [32] WACHTER, K. W. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* 6, 1 (02 1978), 1–18.

## Appendix

### Notations

We will write  $C_{\text{sup}}$  for a generic non-negative finite constant (depending only on  $\lambda_{\text{inf}}$ ,  $\lambda_{\text{sup}}$  and  $m$  in Conditions 2 and 3). The actual value of  $C_{\text{sup}}$  is of no interest and can change in the same sequence of equations. Similarly, we will write  $C_{\text{inf}}$  for a generic strictly positive constant.

If  $B, B' \in \mathcal{P}_p$ , and  $(i, j) \in [1 : p]^2$ , we will write  $(i, j) \in B \setminus B'$  if  $(i, j) \in B$  and  $(i, j) \notin B'$ , that is, if  $i$  and  $j$  are in the same group with the partition  $B$  and are in different groups with the partition  $B'$ .

If  $B, B' \in \mathcal{P}_p$ , we define  $B \cap B'$  as the maximal partition  $B''$  such that  $B'' \leq B$  and  $B'' \leq B'$ .

If  $\Gamma \in \mathcal{M}_p(\mathbb{R})$  (the set of the matrices of dimension  $p \times p$ ), and if  $u, v \subset [1 : p]$ , we define  $\Gamma_{u,v} := (\Gamma_{i,j})_{i \in u, j \in v}$  and  $\Gamma_u := \Gamma_{u,u}$ .

Recall that  $\text{vec} : \mathcal{M}_p(\mathbb{R}) \rightarrow \mathbb{R}^{p^2}$  is defined by  $(\text{vec}(M))_{p(j-1)+i} := M_{i,j}$ .

If  $\Gamma \in S_p(\mathbb{R})$  (the set of the symmetric positive definite matrices) and  $i \in [1 : p]$ , let  $\phi_i(M)$  be the  $i$ -th largest eigenvalue of  $M$ . We also write  $\lambda_{\max}(M)$  (resp.  $\lambda_{\min}(M)$ ) for the largest (resp. smallest) eigenvalue of  $M$ .

We define  $\widehat{\Sigma} := \frac{1}{n-1} \sum_{l=1}^l (X^{(l)} - \bar{X})(X^{(l)} - \bar{X})^T = \frac{n}{n-1} S$ , the unbiased empirical estimator of  $\Sigma$ . Let  $(\widehat{\sigma}_{ij})_{i,j \leq p}$  be the coefficients of  $\widehat{\Sigma}$  and  $(s_{ij})_{i,j \leq p}$  be the coefficients of  $S$ .

Recall that when Condition 4 does not hold, we need to define  $B(\alpha)$  as the partition given by thresholding  $\Sigma$  by  $n^{-\alpha}$ . We also define  $K(\alpha) := |B(\alpha)|$  and write  $B(\alpha) = \{B_1(\alpha), B_2(\alpha), \dots, B_{K(\alpha)}(\alpha)\}$ .

### Proof of Proposition 1

*Proof.* Let us write

$$\overline{S_p^{++}(\mathbb{R}, B)} := \bigsqcup_{B' \leq B} S_p^{++}(\mathbb{R}, B') = \{\Gamma \in S_p^{++}, \Gamma = \Gamma_B\},$$

which is the closure of  $S_p^{++}(\mathbb{R}, B)$  in  $S_p^{++}(\mathbb{R})$ .

First, let us show that, for all  $B$ ,  $S_B$  is the minimum of  $\Gamma \mapsto l_\Gamma$  on  $\overline{S_p^{++}(\mathbb{R}, B)}$ . If  $\Gamma_B \in \overline{S_p^{++}(\mathbb{R}, B)}$ , we have

$$\begin{aligned} p(l_{\Gamma_B} - l_{S_B}) &= -\log(|\Gamma_B^{-1}|) + \text{Tr}(\Gamma_B^{-1}S) + \log(|S_B^{-1}|) - \text{Tr}(S_B^{-1}S) \\ &= -\log(|\Gamma_B^{-1}S_B|) + \text{Tr}(\Gamma_B^{-1}S_B) - p \\ &= \sum_{i=1}^p \left\{ -\log\left(\phi_i\left[\Gamma_B^{-1/2}S_B\Gamma_B^{-1/2}\right]\right) + \phi_i\left(\Gamma_B^{-1/2}S_B\Gamma_B^{-1/2}\right) - 1 \right\}. \end{aligned}$$

The function  $f : \mathbb{R}_+^* \rightarrow \mathbb{R}$  defined by  $f(t) := -\log(t) + t - 1$  has an unique minimum at 1. Thus, the function  $g : S_p^{++} \rightarrow \mathbb{R}$  defined by  $g(M) := \sum_{i=1}^p -\log(\phi_i[M]) + \phi_i(M) - 1$  has an unique minimum at  $I_p$ . Thus  $\Gamma_B \in \overline{S_p^{++}(\mathbb{R}, B)} \mapsto l_{\Gamma_B} - l_{S_B}$  has an unique minimum at  $\Gamma_B = S_B$ .

Now, the penalisation term is constant on each  $S_p^{++}(\mathbb{R}, B)$ . Thus  $\Phi$  has a global minimum (not necessary unique) at  $S_B$ , for some  $B \in \mathcal{P}_p$ .  $\square$

### Notations for Section 2.2

Here and in all the proofs of Section 2.2, we assume Conditions 1 to 3 of Section 2.2.1.

In the following, we introduce some notations.

We know that

$$\sum_{k=1}^n (X^{(k)} - \bar{X})(X^{(k)} - \bar{X})^T \sim \mathcal{W}(n-1, \Sigma),$$

where  $\mathcal{W}(n-1, \Sigma)$  is the Wishart distribution with parameter  $n-1$  and  $\Sigma$  [10]. Thus, if we write  $(M^{(k)})_k$  i.i.d. with distribution  $\mathcal{N}(0, \Sigma)$ , we have

$$\hat{\Sigma} := \frac{1}{n-1} \sum_{k=1}^n (X^{(k)} - \bar{X})(X^{(k)} - \bar{X})^T \sim \frac{1}{n-1} \sum_{k=1}^{n-1} M^{(k)} M^{(k)T}.$$

**Lemma 1.** For all  $C_{\text{inf}} > 0$ ,

$$\mathbb{P}(\lambda_{\max}(S) > \lambda_{\text{sup}}(1 + \sqrt{y})^2 + C_{\text{inf}}) \longrightarrow 0,$$

$$\mathbb{P}(\lambda_{\max}(S) < \lambda_{\text{inf}}(1 + \sqrt{y})^2 - C_{\text{inf}}) \longrightarrow 0,$$

$$\mathbb{P}(\lambda_{\min}(S) < \lambda_{\text{inf}}(1 - \sqrt{y})^2 - C_{\text{inf}}) \longrightarrow 0,$$

and

$$\mathbb{P}(\lambda_{\min}(S) > \lambda_{\text{sup}}(1 - \sqrt{y})^2 + C_{\text{inf}}) \longrightarrow 0.$$

Moreover, that holds also for  $\hat{\Sigma}$  instead of  $S$ .

*Proof.* Let  $(A^{(k)})_k$  i.i.d. with distribution  $\mathcal{N}(0, I_p)$ . Using the result in [27] which states that

$$\lambda_{\max} \left( \frac{1}{n-1} \sum_{k=1}^{n-1} A^{(k)} A^{(k)T} \right) \xrightarrow[n \rightarrow +\infty]{a.s.} (1 + \sqrt{y})^2,$$

we have,

$$\begin{aligned} \lambda_{\max}(S) &= \frac{n}{n-1} \lambda_{\max} \left( \frac{1}{n-1} \sum_{k=1}^{n-1} M^{(k)} M^{(k)T} \right) \\ &\leq \frac{n}{n-1} \lambda_{\text{sup}} \lambda_{\max} \left( \frac{1}{n-1} \sum_{k=1}^{n-1} A^{(k)} A^{(k)T} \right) \\ &= \lambda_{\text{sup}}(1 + \sqrt{y})^2 + o_p(1), \end{aligned}$$

and

$$\lambda_{\max}(S) \geq \frac{n}{n-1} \lambda_{\text{inf}} \lambda_{\max} \left( \frac{1}{n-1} \sum_{k=1}^{n-1} A^{(k)} A^{(k)T} \right) = \lambda_{\text{inf}}(1 + \sqrt{y})^2 + o_p(1).$$

Thus,

$$\lambda_{\text{inf}}(1 + \sqrt{y})^2 + o_p(1) \leq \lambda_{\max}(S) \leq \lambda_{\text{sup}}(1 + \sqrt{y})^2 + o_p(1).$$

The proof is the same for  $\lambda_{\min}$ . □

We also verify the assumptions of Bernstein's inequality (see for example Theorem 2.8.1 in [31]). For all  $i, j, k$ , let

$$Z_{ij}^{(k)} := M_i^{(k)} M_j^{(k)} - \sigma_{ij}. \quad (8)$$

The random-variables  $(Z_{ij}^{(k)})_k$  are independent, mean zero, sub-exponential and we have  $\|Z_{ij}^{(k)}\|_{\psi_1} \leq \|M_i\|_{\psi_2} \|M_j\|_{\psi_2} \leq C_{\text{sup}} \sqrt{\sigma_{ii} \sigma_{jj}} \leq C_{\text{sup}}$ , where  $\|\cdot\|_{\psi_1}$  is the sub-exponential norm (for example, see Definition 2.7.5 in [31]). So, we can use Bernstein's inequality with  $(Z_{ij}^{(k)})_k$ : there exists  $C_{\text{inf}}$  such that, for all  $\varepsilon > 0$  and  $n \in \mathbb{N}$ ,

$$\max_{i,j \in [1:p]} \mathbb{P} \left( \left| \sum_{k=1}^n Z_{ij}^{(k)} \right| \geq n\varepsilon \right) \leq 2 \exp(-C_{\text{inf}} n \min(\varepsilon, \varepsilon^2)).$$

## Proof of Proposition 2

In this proof, we assume that Conditions 1 to 4 are satisfied. We first show several Lemmas.

**Lemma 2.** *For all symmetric positive definite  $\Gamma$  and for all  $B \in \mathcal{P}_p$ , if we write  $\Delta = \Gamma - \Gamma_B$ , we have:*

- $v \mapsto \lambda_{\min}(\Gamma_B + v\Delta)$  decreases and so  $\min_{v \in [0,1]} \lambda_{\min}(\Gamma_B + v\Delta) = \lambda_{\min}(\Gamma)$ .
- $v \mapsto \lambda_{\max}(\Gamma_B + v\Delta)$  increases and so  $\max_{v \in [0,1]} \lambda_{\max}(\Gamma_B + v\Delta) = \lambda_{\max}(\Gamma)$ .

*Proof.* Let us show that  $v \mapsto \lambda_{\max}(\Gamma_B + v\Delta)$  increases (the proof if the same for  $\lambda_{\min}$ ).

For all  $v \in [0, 1]$ , let  $\Gamma_v = \Gamma_B + v\Delta$ ,  $\lambda_v = \lambda_{\max}(\Gamma_v)$  and  $e_v$  an unit eigenvector of  $\Gamma_v$  associated to  $\lambda_v$ . Let  $v, v' \in [0, 1]$ ,  $v < v'$ . Thus

$$\lambda_{v'} = \max_{u, \|u\|=1} u^T (\Gamma_B + v'\Delta) u \geq e_v^T (\Gamma_B + v'\Delta) e_v = \lambda_v + (v - v') e_v^T \Delta e_v.$$

If we show that  $e_v^T \Delta e_v \geq 0$ , we proved that  $v \mapsto \lambda_v$  increases. First, assume that  $v = 0$ . If we write  $B_k$  the group of the largest eigenvalue of  $\Gamma_B$ , then  $(e_0)_i$  is equal to zero for all  $i \notin B_k$ , so  $(e_0^T \Delta)_j$  is equal to zero for all  $j \in B_k$ , and so  $e_0^T \Delta e_0$  is equal to zero.

Assume now that  $v > 0$  and let us show that  $e_v^T \Delta e_v \geq 0$  by contradiction. Assume that  $e_v^T \Delta e_v < 0$ . Then

$$e_v^T (\Gamma_B + v\Delta) e_v < e_v^T \Gamma_B e_v \leq e_0^T \Gamma_B e_0.$$

Furthermore, we have seen that  $e_0^T \Delta e_0 = 0$ . Thus, we have

$$e_v^T (\Gamma_B + v\Delta) e_v < e_0^T (\Gamma_B + v\Delta) e_0,$$

that is in contradiction with  $e_v \in \arg \max_{u, \|u\|=1} u^T (\Gamma_B + v\Delta) u$ .  $\square$



In the following, let  $\Delta_{B,B'} := S_B - S_{B'}$  for all  $B, B' \in \mathcal{P}_p$ .

**Lemma 3.** *For all  $B \in \mathcal{P}_p$ , we have*

$$l_{S_{B \cap B^*}} - l_{S_B} \leq \frac{1}{2\lambda_{\min}(S)} \frac{1}{p} \|\Delta_{B,B \cap B^*}\|_F^2. \quad (9)$$

Moreover, for all  $B < B^*$ , we have

$$l_{S_B} - l_{S_{B^*}} \geq \frac{1}{2\lambda_{\max}(S_{B^*})} \frac{1}{p} \|\Delta_{B^*,B}\|_F^2. \quad (10)$$

*Proof.* First, we prove Equation (11). Doing the Taylor expansion of  $t \mapsto \log \circ \det(S_{B \cap B^*} + t\Delta_{B,B \cap B^*})$  and using the integral form of the remainder (as Equation (9) of [25] or in [16]), we have

$$\begin{aligned} & p(l_{S_{B \cap B^*}} - l_{S_B}) \\ &= \log(|S_{B \cap B^*}|) - \log(|S_B|) \\ &= -\text{Tr}(S_{B \cap B^*} \Delta_{B,B \cap B^*}) + \text{vec}(\Delta_{B,B \cap B^*})^T \left[ \int_0^1 (S_{B \cap B^*} + v\Delta_{B,B \cap B^*})^{-1} \right. \\ & \quad \left. \otimes (S_{B \cap B^*} + v\Delta_{B,B \cap B^*})^{-1} (1-v) dv \right] \text{vec}(\Delta_{B,B \cap B^*}), \end{aligned}$$

where  $\otimes$  is the Kronecker product. The trace is equal to zero. Now,

$$\begin{aligned} p(l_{S_{B \cap B^*}} - l_{S_B}) &\leq 1/2 \max_{0 \leq v \leq 1} \lambda_{\max}^2[(S_{B \cap B^*} + v\Delta_{B,B \cap B^*})^{-1}] \|\text{vec}(\Delta_{B,B \cap B^*})\|^2 \\ &= 1/2 \max_{0 \leq v \leq 1} \lambda_{\min}^{-2}(S_{B \cap B^*} + v\Delta_{B,B \cap B^*}) \|\text{vec}(\Delta_{B,B \cap B^*})\|^2 \\ &= \frac{1}{2 \min_v (\lambda_{\min}(S_{B \cap B^*} + v\Delta_{B,B \cap B^*}))^2} \|\text{vec}(\Delta_{B,B \cap B^*})\|^2 \\ &= \frac{1}{2\lambda_{\min}(S)^2} \|\text{vec}(\Delta_{B,B \cap B^*})\|^2 \\ &\leq \frac{1}{2\lambda_{\min}(S)^2} \|\text{vec}(\Delta_{B,B \cap B^*})\|^2, \end{aligned}$$

using Lemma 2 for the two last steps.

Now, we prove Equation (12) similarly. We have, using Lemma 2,

$$\begin{aligned}
p(l_{S_B} - l_{S_{B^*}}) &= -\text{Tr}(S_B \Delta_{B^*, B}) + \text{vec}(\Delta_{B^*, B})^T \left[ \int_0^1 (S_B + v \Delta_{B^*, B})^{-1} \right. \\
&\quad \left. \otimes (S_B + v \Delta_{B^*, B})^{-1} (1 - v) dv \right] \text{vec}(\Delta_{B^*, B}) \\
&\geq 1/2 \min_{0 \leq v \leq 1} \lambda_{\min}^2[(S_B + v \Delta_{B^*, B})^{-1}] \|\text{vec}(\Delta_{B^*, B})\|^2 \\
&= 1/2 \min_{0 \leq v \leq 1} \lambda_{\max}^{-2}(S_B + v \Delta_{B^*, B}) \|\text{vec}(\Delta_{B^*, B})\|^2 \\
&= \frac{1}{2 \max_v (\lambda_{\max}(S_B + v \Delta_{B^*, B}))^2} \|\text{vec}(\Delta_{B^*, B})\|^2 \\
&= \frac{1}{2 \lambda_{\max}(S_B)^2} \|\text{vec}(\Delta_{B^*, B})\|^2 \\
&\geq \frac{1}{2 \lambda_{\max}(S)^2} \|\text{vec}(\Delta_{B^*, B})\|^2.
\end{aligned}$$

□

**Lemma 4.**

$$\mathbb{P} \left( \max_{B \not\leq B^*} \Phi(B \cap B^*) - \Phi(B) \geq 0 \right) \rightarrow 0.$$

*Proof.* Using Lemma 3, we have

$$\begin{aligned}
&\mathbb{P} \left( \max_{B \not\leq B^*} \Phi(B \cap B^*) - \Phi(B) \geq 0 \right) \\
&\leq \mathbb{P} \left( \max_{B \not\leq B^*} \left[ l_{S_{B \cap B^*}} - l_{S_B} - \frac{1}{pn^\delta} (\text{pen}(B) - \text{pen}(B \cap B^*)) \right] \geq 0 \right) \\
&\leq \mathbb{P} \left( \max_{B \not\leq B^*} \left[ \frac{1}{2 \lambda_{\min}(S)^2} \|\Delta_{B, B \cap B^*}\|_F^2 - \frac{1}{n^\delta} (\text{pen}(B) - \text{pen}(B \cap B^*)) \right] \geq 0 \right) \\
&\leq \mathbb{P} \left( \lambda_{\min}(S) \leq \frac{1}{2} \lambda_{\inf} (1 - \sqrt{y})^2 \right) \\
&\quad + \mathbb{P} \left( \max_{B \not\leq B^*} \left[ \frac{1}{(1 - \sqrt{y})^4 \lambda_{\inf}^2} \|\Delta_{B, B \cap B^*}\|_F^2 - \frac{1}{n^\delta} (\text{pen}(B) - \text{pen}(B \cap B^*)) \right] \geq 0 \right).
\end{aligned}$$

We show that the two terms go to 0. The first term goes to 0 with Lemma 1.

For the second term, we have

$$\begin{aligned}
& \mathbb{P} \left( \max_{B \not\subseteq B^*} \left[ \frac{1}{(1-\sqrt{y})^4 \lambda_{\inf}^2} \|\Delta_{B, B \cap B^*}\|_F^2 - \frac{1}{n^\delta} (\text{pen}(B) - \text{pen}(B \cap B^*)) \right] \geq 0 \right) \\
&= \mathbb{P} \left( \max_{B \not\subseteq B^*} \left[ \sum_{(i,j) \in B \setminus B^*} \left\{ \frac{1}{(1-\sqrt{y})^4 \lambda_{\inf}^2} s_{ij}^2 - \frac{1}{n^\delta} \right\} \right] \geq 0 \right) \\
&\leq \mathbb{P} \left( \exists (i,j) \notin B^*, \frac{1}{(1-\sqrt{y})^4 \lambda_{\inf}^2} s_{ij}^2 - \frac{1}{n^\delta} \geq 0 \right) \\
&\leq \mathbb{P} \left( \exists (i,j) \notin B^*, \frac{2}{(1-\sqrt{y})^4 \lambda_{\inf}^2} \widehat{\sigma}_{ij}^2 - \frac{1}{n^\delta} \geq 0 \right) \\
&\leq p^2 \max_{(i,j) \notin B^*} \mathbb{P} \left( \frac{2}{(1-\sqrt{y})^4 \lambda_{\inf}^2} \widehat{\sigma}_{ij}^2 - \frac{1}{n^\delta} \geq 0 \right) \\
&\leq p^2 \max_{(i,j) \notin B^*} \mathbb{P} \left( \frac{\sqrt{2}}{(1-\sqrt{y})^2 \lambda_{\inf}} |\widehat{\sigma}_{ij}| \geq \frac{1}{n^{\delta/2}} \right) \\
&\leq p^2 \max_{(i,j) \notin B^*} \mathbb{P} \left( \left| \sum_{k=1}^{n-1} Z_{ij}^{(k)} \right| \geq \frac{(1-\sqrt{y})^2}{\sqrt{2}} \lambda_{\inf} n^{1-\delta/2} \right) \\
&\leq 2p^2 \exp(-C_{\inf} n^{1-\delta}) \rightarrow 0,
\end{aligned}$$

using Bernstein's inequality, where  $Z_{ij}^{(k)}$  is defined in Equation (8). That concludes the proof.  $\square$

**Lemma 5.**

$$\mathbb{P} \left( \max_{B < B^*} \Phi(B^*) - \Phi(B) \geq 0 \right) \rightarrow 0.$$

*Proof.* Using Lemma 3, we have

$$\begin{aligned}
& \mathbb{P} \left( \max_{B < B^*} \Phi(B^*) - \Phi(B) \geq 0 \right) \\
&\leq \mathbb{P} \left( \min_{B < B^*} \left[ l_{S_B} - l_{S_{B^*}} - \frac{1}{pn^\delta} (\text{pen}(B^*) - \text{pen}(B)) \right] \leq 0 \right) \\
&\leq \mathbb{P} \left( \min_{B < B^*} \left[ \frac{1}{2\lambda_{\max}(S_{B^*})^2} \|\Delta_{B^*, B}\|_F^2 - \frac{1}{n^\delta} (\text{pen}(B^*) - \text{pen}(B)) \right] \leq 0 \right) \\
&\leq \mathbb{P} \left( \lambda_{\max}(S_B) \leq \frac{\lambda_{\inf}(1+\sqrt{y})^2}{2} \right) \\
&\quad + \mathbb{P} \left( \min_{B < B^*} \left[ \frac{1}{(1+\sqrt{y})^4 \lambda_{\inf}^2} \|\Delta_{B^*, B}\|_F^2 - \frac{1}{n^\delta} (\text{pen}(B^*) - \text{pen}(B)) \right] \leq 0 \right).
\end{aligned}$$

The first term goes to 0 with Lemma 1. The second term is

$$\begin{aligned}
& \mathbb{P} \left( \exists B < B^*, \sum_{(i,j) \in B^* \setminus B} \left[ \frac{1}{(1 + \sqrt{y})^4 \lambda_{\inf}^2} s_{i,j}^2 - n^{-\delta} \right] \leq 0 \right) \\
& \leq \mathbb{P} \left( \exists B < B^*, \sum_{(i,j) \in B^* \setminus B} \left[ \frac{1}{\lambda_{\inf}^2} s_{i,j}^2 - n^{-\delta} \right] \leq 0 \right) \\
& \leq \mathbb{P} \left( \exists k \in [1 : K], \emptyset \subsetneq B_1 \subsetneq B_k^*, \sum_{i \in B_1, j \in B_k^* \setminus B_1} \left[ \frac{1}{\lambda_{\inf}^2} s_{ij}^2 - n^{-\delta} \right] \leq 0 \right) \\
& \leq p2^m \max_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \mathbb{P} \left( \sum_{i \in B_1, j \in B_k^* \setminus B_1} \left[ \frac{1}{\lambda_{\inf}^2} s_{ij}^2 - n^{-\delta} \right] \leq 0 \right) \\
& \leq p2^m \max_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \mathbb{P} \left( \sum_{i \in B_1, j \in B_k^* \setminus B_1} \left[ \frac{1}{2\lambda_{\inf}^2} \widehat{\sigma}_{ij}^2 - n^{-\delta} \right] \leq 0 \right).
\end{aligned}$$

Now, for all  $k \in [1 : K]$  and for all  $\emptyset \subsetneq B_1 \subsetneq B_k^*$ , let  $(i^*, j^*) \in \arg \max_{i \in B_1, j \in B_k^* \setminus B_1} |\sigma_{ij}|$  (with an implicit dependence on  $k$  and  $B_1$ ). Remark that

$$\frac{1}{2\lambda_{\inf}^2} \widehat{\sigma}_{i^*j^*}^2 \geq m^2 n^{-\delta} \implies \sum_{i \in B_1, j \in B_k^* \setminus B_1} \left( \frac{1}{2\lambda_{\inf}^2} \widehat{\sigma}_{ij}^2 - n^{-\delta} \right) \geq 0.$$

Thus,

$$\begin{aligned}
& \mathbb{P} \left( \exists B < B^*, \sum_{(i,j) \in B^* \setminus B} \left[ \frac{1}{\lambda_{\inf}^2} s_{i,j}^2 - n^{-\delta} \right] \leq 0 \right) \\
& \leq p2^m \max_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \mathbb{P} \left( \frac{1}{2\lambda_{\inf}^2} \widehat{\sigma}_{i^*j^*}^2 \leq m^2 n^{-\delta} \right) \\
& = p2^m \max_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \mathbb{P} \left( |\widehat{\sigma}_{i^*j^*}| \leq \sqrt{2} \lambda_{\inf} m n^{-\delta/2} \right) \\
& \leq p2^m \max_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \mathbb{P} \left( |\widehat{\sigma}_{i^*j^*} - \sigma_{i^*j^*}| \geq |\sigma_{i^*j^*}| - \sqrt{2} \lambda_{\inf} m n^{-\delta/2} \right) \\
& = p2^m \max_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \mathbb{P} \left( \left| \sum_{k=1}^{n-1} Z_{i^*j^*}^{(k)} \right| \geq n \left[ |\sigma_{i^*j^*}| - \sqrt{2} \lambda_{\inf} m n^{-\delta/2} \right] \right) \\
& = p2^m \mathbb{P} \left( \left| \sum_{k=1}^{n-1} Z_{i^*j^*}^{(k)} \right| \geq n \min_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \left[ |\sigma_{i^*j^*}| - \sqrt{2} \lambda_{\inf} m n^{-\delta/2} \right] \right).
\end{aligned}$$

Now, by Condition 4, we know that  $\min_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} |\sigma_{i^*j^*}| \geq an^{-1/4}$ , so, for  $n$  large enough,

$$\min_{\substack{k \in [1:K], \\ \emptyset \subsetneq B_1 \subsetneq B_k^*}} \left[ |\sigma_{i^*j^*}| - \sqrt{2}\lambda_{\text{inf}}mn^{-\delta/2} \right] \geq C_{\text{inf}}(n^{-1/4} - n^{-\delta/2}) \geq C_{\text{inf}}(\delta)n^{-1/4}.$$

Thus, by Bernstein's inequality, for  $n$  large enough,

$$\begin{aligned} & \mathbb{P} \left( \exists B < B^*, \sum_{(i,j) \in B^* \setminus B} \left[ \frac{1}{(1-\sqrt{y})^2 \lambda_{\text{inf}}^2} s_{i,j}^2 - n^{-\delta} \right] \leq 0 \right) \\ & \leq p2^{m+1} \exp \left( -C_{\text{inf}}(\delta)n^{1/2} \right) \longrightarrow 0. \end{aligned}$$

□

Now, we can prove Proposition 2.

*Proof.* We have

$$\begin{aligned} \mathbb{P} \left( \max_{B \neq B^*} \Phi(B^*) - \Phi(B) \geq 0 \right) & \leq \mathbb{P} \left( \max_{B < B^*} \Phi(B^*) - \Phi(B) \geq 0 \right) \\ & \quad + \mathbb{P} \left( \max_{B > B^*} \Phi(B^*) - \Phi(B) \geq 0 \right) \\ & \quad + \mathbb{P} \left( \max_{B \not\leq B^*, B \not\geq B^*} \Phi(B^*) - \Phi(B) \geq 0 \right). \end{aligned}$$

The two first terms go to 0 thanks to Lemmas 4 and 5. For the last term, we have

$$\begin{aligned} & \mathbb{P} \left( \max_{B \not\leq B^*, B \not\geq B^*} \Phi(B^*) - \Phi(B) \geq 0 \right) \\ & = \mathbb{P} \left( \max_{B \not\leq B^*, B \not\geq B^*} \Phi(B^*) - \Phi(B \cap B^*) + \Phi(B \cap B^*) - \Phi(B) \geq 0 \right) \\ & \leq \mathbb{P} \left( \max_{B \not\leq B^*, B \not\geq B^*} \Phi(B^*) - \Phi(B \cap B^*) \geq 0 \right) \\ & \quad + \mathbb{P} \left( \max_{B \not\leq B^*, B \not\geq B^*} \Phi(B \cap B^*) - \Phi(B) \geq 0 \right) \\ & \leq \mathbb{P} \left( \max_{B' < B^*} \Phi(B^*) - \Phi(B') \geq 0 \right) \\ & \quad + \mathbb{P} \left( \max_{B \not\leq B^*} \Phi(B \cap B^*) - \Phi(B) \geq 0 \right). \end{aligned}$$

These two last terms go to 0 thanks to Lemmas 4 and 5.

□

**Proofs of Proposition 3**

In this proof, we assume that Conditions 1 to 3 hold.

**Lemma 6.** For all  $B \in \mathcal{P}_p$ , we have

$$l_{S_{B \cap B(\alpha_2)}} - l_{S_B} \leq \frac{1}{2\lambda_{\min}(S)} \frac{1}{p} \|\Delta_{B, B \cap B(\alpha_2)}\|_F^2. \quad (11)$$

Moreover, for all  $B < B(\alpha_2)$ , we have

$$l_{S_B} - l_{S_{B(\alpha_2)}} \geq \frac{1}{2\lambda_{\max}(S_{B(\alpha_2)})} \frac{1}{p} \|\Delta_{B(\alpha_2), B}\|_F^2. \quad (12)$$

*Proof.* Same proof as Lemma 3 replacing  $B^*$  by  $B(\alpha_2)$ .  $\square$

**Lemma 7.** If  $\alpha_2 > \delta/2$ , then,

$$\mathbb{P} \left( \max_{B \not\leq B(\alpha_2)} \Phi(B \cap B(\alpha_2)) - \Phi(B) \geq 0 \right) \longrightarrow 0.$$

*Proof.* Following the proof of Lemma 4 (and using Lemma 6), it is enough to prove that the following term goes to 0:

$$\begin{aligned} & \mathbb{P} \left( \max_{B \not\leq B(\alpha_2)} \left[ \frac{1}{(1-\sqrt{y})^4 \lambda_{\inf}^2} \|\Delta_{B, B \cap B(\alpha_2)}\|_F^2 - \frac{1}{n^\delta} (\text{pen}(B) - \text{pen}(B \cap B(\alpha_2))) \right] \geq 0 \right) \\ &= \mathbb{P} \left( \max_{B \not\leq B(\alpha_2)} \left[ \sum_{(i,j) \in B \setminus B(\alpha_2)} \left\{ \frac{1}{(1-\sqrt{y})^4 \lambda_{\inf}^2} s_{ij}^2 - \frac{1}{n^\delta} \right\} \right] \geq 0 \right) \\ &\leq \mathbb{P} \left( \exists (i,j) \notin B(\alpha_2), \frac{1}{(1-\sqrt{y})^4 \lambda_{\inf}^2} s_{ij}^2 - \frac{1}{n^\delta} \geq 0 \right) \\ &\leq \mathbb{P} \left( \exists (i,j) \notin B(\alpha_2), \frac{2}{(1-\sqrt{y})^4 \lambda_{\inf}^2} \hat{\sigma}_{ij}^2 - \frac{1}{n^\delta} \geq 0 \right) \\ &\leq p^2 \max_{(i,j) \notin B(\alpha_2)} \mathbb{P} \left( \frac{2}{(1-\sqrt{y})^4 \lambda_{\inf}^2} \hat{\sigma}_{ij}^2 - \frac{1}{n^\delta} \geq 0 \right) \\ &\leq p^2 \max_{(i,j) \notin B(\alpha_2)} \mathbb{P} \left( \frac{\sqrt{2}}{(1-\sqrt{y})^2 \lambda_{\inf}} |\hat{\sigma}_{ij}| \geq \frac{1}{n^{\delta/2}} \right) \\ &\leq p^2 \max_{(i,j) \notin B(\alpha_2)} \mathbb{P} \left( \frac{\sqrt{2}}{(1-\sqrt{y})^2 \lambda_{\inf}} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq \frac{1}{n^{\delta/2}} - \frac{\sqrt{2}}{(1-\sqrt{y})^2 \lambda_{\inf}} n^{-\alpha_2} \right) \\ &\leq 2p^2 \exp(-C_{\inf} n^{1-\delta}) \longrightarrow 0, \end{aligned}$$

using again Bernstein's inequality. That concludes the proof.  $\square$

**Lemma 8.** If  $\alpha_1 < \delta/2$ , then,

$$\mathbb{P} \left( \max_{B < B(\alpha_1)} \Phi(B(\alpha_1)) - \Phi(B) \geq 0 \right) \longrightarrow 0.$$

*Proof.* Following the proof of Lemma 5, it suffices to prove that

$$p2^m \mathbb{P} \left( \left| \sum_{k=1}^{n-1} Z_{i^* j^*}^{(k)} \right| \geq n \min_{\substack{k \in [1:K(\alpha_1)], \\ \emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1)}} \left[ |\sigma_{i^* j^*}| - \sqrt{2} \lambda_{\text{inf}} m n^{-\delta/2} \right] \right) \longrightarrow 0.$$

Now, by definition of  $B(\alpha_1)$ , we know that  $\min_{\substack{k \in [1:K(\alpha_1)], \\ \emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1)}} |\sigma_{i^* j^*}| \geq n^{-\alpha_1}$ , so, for  $n$  large enough,

$$\min_{\substack{k \in [1:K(\alpha_1)], \\ \emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1)}} \left[ |\sigma_{i^* j^*}| - \sqrt{2} \lambda_{\text{inf}} m n^{-\delta/2} \right] \geq C_{\text{inf}}(n^{-\alpha_1} - n^{-\delta/2}) \geq C_{\text{inf}}(\alpha_1, \delta) n^{-\alpha_1}.$$

Thus, by Bernstein's inequality, for  $n$  large enough,

$$\begin{aligned} & \mathbb{P} \left( \exists B < B(\alpha_1), \sum_{(i,j) \in B(\alpha_1) \setminus B} \left[ \frac{1}{(1-\sqrt{y})^2 \lambda_{\text{inf}}^2} s_{i,j}^2 - n^{-\delta} \right] \leq 0 \right) \\ & \leq p2^{m+1} \exp(-C_{\text{inf}}(\alpha_1, \delta) n^{1-2\alpha_1}) \longrightarrow 0. \end{aligned}$$

□

We can now prove Proposition 3.

*Proof.* We have

$$\begin{aligned} & \mathbb{P} \left( \left\{ B(\alpha_1) \not\prec \widehat{B}_{\text{tot}} \leq B(\alpha_2) \right\}^c \right) \\ & = \mathbb{P} \left( \min_{B < B(\alpha_1) \text{ or } B \not\prec B(\alpha_2)} \Phi(B) \leq \min_{B \not\prec B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right) \\ & \leq \mathbb{P} \left( \min_{B < B(\alpha_1)} \Phi(B) \leq \min_{B \not\prec B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right) \\ & \quad + \mathbb{P} \left( \min_{B \not\prec B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) \not\prec B(\alpha_1)} \Phi(B) \leq \min_{B \not\prec B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right) \\ & \quad + \mathbb{P} \left( \min_{B \not\prec B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) < B(\alpha_1)} \Phi(B) \leq \min_{B \not\prec B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right). \end{aligned}$$

First,

$$\begin{aligned} & \mathbb{P} \left( \min_{B < B(\alpha_1)} \Phi(B) \leq \min_{B \not\prec B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right) \\ & \leq \mathbb{P} \left( \min_{B < B(\alpha_1)} \Phi(B) - \Phi(B(\alpha_1)) \leq 0 \right) \longrightarrow 0, \end{aligned}$$

from Lemma 8. Secondly,

$$\begin{aligned} & \mathbb{P} \left( \min_{B \not\prec B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) \not\prec B(\alpha_1)} \Phi(B) \leq \min_{B \not\prec B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right) \\ & \leq \mathbb{P} \left( \min_{B \not\prec B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) \not\prec B(\alpha_1)} \Phi(B) - \Phi(B \cap B(\alpha_2)) \leq 0 \right) \longrightarrow 0, \end{aligned}$$

from Lemma 7. Finally,

$$\begin{aligned}
& \mathbb{P} \left( \min_{B \not\leq B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) < B(\alpha_1)} \Phi(B) \leq \min_{B \not\leq B(\alpha_1) \text{ and } B \leq B(\alpha_2)} \Phi(B) \right) \\
\leq & \mathbb{P} \left( \min_{B \not\leq B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) < B(\alpha_1)} \Phi(B) \leq \Phi(B(\alpha_1)) \right) \\
\leq & \mathbb{P} \left( \min_{B \not\leq B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) < B(\alpha_1)} \Phi(B) - \Phi(B \cap B(\alpha_2)) \right. \\
& \left. + \Phi(B \cap B(\alpha_2)) - \Phi(B(\alpha_1)) \leq 0 \right) \\
\leq & \mathbb{P} \left( \min_{B \not\leq B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) < B(\alpha_1)} \Phi(B) - \Phi(B \cap B(\alpha_2)) \leq 0 \right) \\
& + \mathbb{P} \left( \min_{B \not\leq B(\alpha_2) \text{ s.t. } B \cap B(\alpha_2) < B(\alpha_1)} \Phi(B \cap B(\alpha_2)) - \Phi(B(\alpha_1)) \leq 0 \right) \\
\leq & \mathbb{P} \left( \min_{B \not\leq B(\alpha_2)} \Phi(B) - \Phi(B \cap B(\alpha_2)) \leq 0 \right) \\
& + \mathbb{P} \left( \min_{B < B(\alpha_1)} \Phi(B) - \Phi(B(\alpha_1)) \leq 0 \right) \\
\rightarrow & 0,
\end{aligned}$$

from Lemmas 7 and 8.  $\square$

### Proofs of Propositions 4, 5 and 6

*Proof.* In the three cases, the computation of  $\widehat{B}$  requires carrying out the BFS algorithm for  $B_\lambda$  and the computation of a determinant for  $\Psi(B_\lambda)$ . Recall that if  $G = (V, E)$  is a graph (where  $V$  is the set of vertices and  $E$  the set of edges), the complexity of the BFS algorithm is  $O(|V| + |E|)$ . Recall that, if  $M$  is a squared matrix of size  $p$ , the complexity of  $\det(M)$  is  $O(p^3)$  using the LU decomposition.

Now, we compute the complexity of the three estimators  $\widehat{B}_{\widehat{C}}$ ,  $\widehat{B}_A$  and  $\widehat{B}_s$ .

- For all  $\lambda \in A_{\widehat{C}}$ , the complexity of  $B_\lambda$  is  $O(p^2)$ , and the cardinal of  $A_{\widehat{C}}$  is  $O(p^2)$ . Thus, the complexity of the computation of  $\{B_\lambda \mid \lambda \in A_{\widehat{C}}\}$  is  $O(p^4)$ .

Now, for all  $\lambda \in A_{\widehat{C}}$ , the complexity of  $\Psi(B_\lambda)$  is  $O(p^3)$  and the cardinal of  $\{B_\lambda \mid \lambda \in A_{\widehat{C}}\}$  is  $O(p)$  (because the function  $\lambda \mapsto B_\lambda$  decreases). Thus, the complexity of the evaluations  $\{\Psi(B), B \in \{B_\lambda \mid \lambda \in A_{\widehat{C}}\}\}$  is  $O(p^4)$ .

So the complexity of  $\widehat{B}_{\widehat{C}}$  is  $O(p^4)$ .

- For the threshold  $n^{-1/3}$ , the complexity of  $B_{n^{-1/3}}$  is  $O(p^2)$ .  
So the complexity of  $\widehat{B}_\lambda$  is  $O(p^2)$ .



- One can divide the computation of  $\widehat{B}_s$  into two steps.

For the first step, as we do not know the value of  $s$ , we have to compute  $B_{l/p}$  from  $l = p$  to  $l = s - 1$ , verifying each time if the maximal size of group is smaller than  $m$  or not. First, for each value of  $l$  from  $p$  decreasing to  $s$ , the complexity of the BFS algorithm to  $B_{l/p}$  is  $O(p \times m^2) = O(p)$ , thus, the complexity of all these partitions is  $O(p^2)$ . Then, for  $l = s - 1$ , the complexity of  $B_{(s-1)/p}$  is  $O(p^2)$ . So, the complexity of this first step is  $O(p^2)$ .

In the second step, we have to evaluate  $\Psi(B_{l/p})$  for all  $l \in [s : p]$ . The complexity of each evaluation is  $O(pm^3) = O(p)$ , and the number of evaluations is  $O(p)$ . Thus, the complexity of this second step is  $O(p^2)$ .

□

### Proof of Proposition 7

To prove the convergence of  $\widehat{B}$  in the three cases, we need the three following Lemmas.

**Lemma 9.** *For all sequence  $(\lambda_n)_n$  such that for all  $n$ ,  $\lambda_n \in [n^{-1/3}, an^{-1/4}/3\lambda_{\text{sup}}(1 + \sqrt{y})^2]$  (we assume that  $n$  is large enough and that subset is not empty), we have*

$$\mathbb{P}(B_\lambda = B^*) \longrightarrow 1.$$

*Proof.* Step 1:  $B_\lambda \leq B^*$  with probability which goes to 1.

$$\begin{aligned} & \mathbb{P}(B_\lambda \not\leq B^*) \\ &= \mathbb{P}\left(\exists(i, j) \notin B^*, |\widehat{C}_{ij}| \geq \lambda\right) \\ &\leq \mathbb{P}\left(\exists(i, j) \notin B^*, |\widehat{\sigma}_{ij}| \geq \lambda \frac{\lambda_{\text{inf}}(1 - \sqrt{y})^2}{2}\right) + \mathbb{P}\left(\exists i \leq p, \widehat{\sigma}_{ii} < \lambda_{\text{inf}} \frac{(1 - \sqrt{y})^2}{2}\right) \\ &\leq p^2 \max_{(i, j) \notin B^*} \mathbb{P}\left(|\widehat{\sigma}_{ij}| \geq \lambda \frac{\lambda_{\text{inf}}(1 - \sqrt{y})^2}{2}\right) + \mathbb{P}\left(\lambda_{\text{min}}(\widehat{\Sigma}) < \lambda_{\text{inf}} \frac{(1 - \sqrt{y})^2}{2}\right) \\ &\leq p^2 \max_{(i, j) \notin B^*} \mathbb{P}\left(|\widehat{\sigma}_{ij}| \geq \frac{\lambda_{\text{inf}}(1 - \sqrt{y})^2}{2} n^{-1/3}\right) + o(1) \\ &\leq 2p^2 \exp\left(-C_{\text{inf}} n^{1/3}\right) + o(1) \longrightarrow 0, \end{aligned}$$

using Lemma 1 and Bernstein's inequality.

Step 2:  $B_\lambda \geq B^*$  with probability which goes to 1.

For all  $k \in [1 : K]$ , and all  $\emptyset \subsetneq B_1 \subsetneq B_k^*$ , let  $B_2 := B_k^* \setminus B_1$  and  $(i^*, j^*) := \arg \max_{(i, j) \in B_1 \times B_2} |\sigma_{ij}|$ , where the dependency on  $k$  and  $B_1$  is implicit. Thanks

to Condition 4, we have  $|\sigma_{i^*j^*}| \geq an^{-1/4}$ . Then, using Lemma 1,

$$\begin{aligned}
& \mathbb{P}(B_\lambda \not\subseteq B^*) \\
&= \mathbb{P}\left(\exists k \in [1 : K], \exists \emptyset \subsetneq B_1 \subsetneq B_k^*, \max_{(i,j) \in B_1 \times B_2} |\widehat{C}_{ij}| < \lambda\right) \\
&\leq \mathbb{P}\left(\exists k \in [1 : K], \exists \emptyset \subsetneq B_1 \subsetneq B_k^*, \max_{(i,j) \in B_1 \times B_2} |\widehat{\sigma}_{ij}| < 2\lambda\lambda_{\text{sup}}(1 + \sqrt{y})^2\right) \\
&\quad + \mathbb{P}(\exists i \leq p, \widehat{\sigma}_{ii} \geq 2\lambda_{\text{sup}}(1 + \sqrt{y})^2) \\
&\leq \mathbb{P}\left(\exists k \in [1 : K], \exists \emptyset \subsetneq B_1 \subsetneq B_k^*, |\widehat{\sigma}_{i^*j^*}| < \frac{2}{3}an^{-1/4}\right) \\
&\quad + \mathbb{P}\left(\lambda_{\max}(\widehat{\Sigma}) \geq 2\lambda_{\text{sup}}(1 + \sqrt{y})^2\right) \\
&\leq \mathbb{P}\left(\exists k \in [1 : K], \exists \emptyset \subsetneq B_1 \subsetneq B_k^*, |\widehat{\sigma}_{i^*j^*} - \sigma_{i^*j^*}| > \frac{1}{3}an^{-1/4}\right) + o(1) \\
&\leq \mathbb{P}\left(\exists (i, j) \in [1 : p]^2, |\widehat{\sigma}_{ij} - \sigma_{ij}| > \frac{1}{3}an^{-1/4}\right) + o(1) \\
&\leq p^2 \max_{(i,j)} \mathbb{P}\left(|\widehat{\sigma}_{ij} - \sigma_{ij}| > \frac{1}{3}an^{-1/4}\right) + o(1) \\
&\leq 2p^2 \exp\left(-C_{\text{inf}}n^{1/2}\right) + o(1) \longrightarrow 0,
\end{aligned}$$

by Bernstein's inequality.  $\square$

**Lemma 10.** *Let  $c > 0$ . Let  $\tilde{A} := \{a_0, a_1, \dots, a_L\}$  such that  $a_0 = 0$ ,  $a_L = 1$ ,  $0 < a_{l+1} - a_l < c/\sqrt{p}$  for all  $l \in [0 : L - 1]$ . Then,*

$$\mathbb{P}\left(B^* \in \left\{B_\lambda, \lambda \in \tilde{A}\right\}\right) \longrightarrow 1.$$

*Proof.* Thanks to Lemma 9, it suffices to show that, for  $n$  large enough, there exists  $l \in [0 : L]$  such that  $a_l \in [n^{-1/3}, an^{-1/4}/3\lambda_{\text{sup}}(1 + \sqrt{y})^2]$ . By contradiction, let us assume that there does not exist such  $l$ . Let  $j \in [0 : L]$  such that  $a_j < n^{-1/3}$  and  $a_{j+1} > an^{-1/4}/3\lambda_{\text{sup}}$ . Thus, we have

$$\begin{aligned}
\sqrt{p}(a_{j+1} - a_j) &> \sqrt{p}\left(\frac{an^{-1/4}}{3\lambda_{\text{sup}}(1 + \sqrt{y})^2} - n^{-1/3}\right) \\
&\geq C_{\text{inf}}n^{1/4} \longrightarrow +\infty,
\end{aligned}$$

which is in contradiction with the definition of  $\tilde{A}$ .  $\square$

**Lemma 11.** *We have,*

$$\mathbb{P}(B^* \in \{B_\lambda, \lambda \in A_s\}) \longrightarrow 1.$$

*Proof.* Let  $\mathcal{P}_p(m)$  be the set of the partitions of  $[1 : p]$  such that all their elements have cardinal smaller than  $m$ . By assumption (Condition 3),  $B^* \in \mathcal{P}_p(m)$ . Let  $G := \{l/p \mid l \in [0, p]\}$ . Thus  $G$  verifies the assumption of  $\tilde{A}$  in Lemma 10, so

$$\mathbb{P}(B^* \in \{B_\lambda, \lambda \in G\}) \longrightarrow 1.$$

Thus

$$\mathbb{P}(B^* \in \{B_\lambda, \lambda \in G\} \cap \mathcal{P}_p(m)) \longrightarrow 1.$$

To conclude, it suffices to prove that  $\{B_\lambda, \lambda \in G_s\} \cap \mathcal{P}_p(m) = \{B_\lambda, \lambda \in A_s\}$ .

We have immediately  $\{B_\lambda, \lambda \in A_s\} \subset \{B_\lambda, \lambda \in G\} \cap \mathcal{P}_p(m)$ . We have to prove the other inclusion. Assume that  $B \in \{B_\lambda, \lambda \in G\} \cap \mathcal{P}_p(m)$ . We know that there exists  $\lambda = l/p \in G$  such that  $B = B_\lambda$ . As  $B_{l/p} \in \mathcal{P}_p(m)$ , we know by definition of  $s$  that  $l \geq s$  and thus  $\lambda \in A$ .  $\square$

Now, we prove Proposition 7.

*Proof.* • Using Lemma 9, Proposition 2, and the fact that  $\{B_\lambda \mid \lambda \in A_{\hat{C}}\} = \{B_\lambda \mid \lambda \in [0, 1]\}$ , we have  $\mathbb{P}(\hat{B}_{\hat{C}} = B^*) \longrightarrow 1$ .

• Using Lemma 9 and Proposition 2, we have  $\mathbb{P}(\hat{B}_\lambda = B^*) \longrightarrow 1$ .

• Using Lemma 11 and Proposition 2, we have  $\mathbb{P}(\hat{B}_s = B^*) \longrightarrow 1$ .  $\square$

### Proof of Proposition 8

*Proof.* We follow the proof of Lemma 9.

Step 1:  $B_{n^{-\delta/2}} \leq B(\alpha_2)$  with probability which goes to 1.

$$\begin{aligned} & \mathbb{P}(B_{n^{-\delta/2}} \not\leq B(\alpha_2)) \\ = & \mathbb{P}\left(\exists(i, j) \notin B(\alpha_2), |\hat{C}_{ij}| \geq n^{-\delta/2}\right) \\ \leq & \mathbb{P}\left(\exists(i, j) \notin B(\alpha_2), |\hat{\sigma}_{ij}| \geq n^{-\delta/2} \frac{\lambda_{\inf}(1 - \sqrt{y})^2}{2}\right) + \mathbb{P}\left(\exists i \leq p, \hat{\sigma}_{ii} < \lambda_{\inf} \frac{(1 - \sqrt{y})^2}{2}\right) \\ \leq & p^2 \max_{(i, j) \notin B(\alpha_2)} \mathbb{P}\left(|\hat{\sigma}_{ij}| \geq n^{-\delta/2} \frac{\lambda_{\inf}(1 - \sqrt{y})^2}{2}\right) + \mathbb{P}\left(\lambda_{\min}(\hat{\Sigma}) < \lambda_{\inf} \frac{(1 - \sqrt{y})^2}{2}\right) \\ \leq & p^2 \max_{(i, j) \notin B(\alpha_2)} \mathbb{P}\left(|\hat{\sigma}_{ij} - \sigma_{ij}| \geq n^{-\delta/2} \frac{\lambda_{\inf}(1 - \sqrt{y})^2}{2} - n^{-\alpha_2}\right) + o(1) \\ \leq & 2p^2 \exp(-C_{\inf}(\delta, \alpha_2)n^{1-\delta}) + o(1) \longrightarrow 0, \end{aligned}$$

using Lemma 1 and Bernstein's inequality.

Step 2:  $B_{n^{-\delta/2}} \geq \mathcal{B}(\alpha_1)$  with probability which goes to 1.

For all  $k \in [1 : K(\alpha_1)]$ , and all  $\emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1)$ , let  $B_2 := B_k(\alpha_1) \setminus B_1$  and  $(i^*, j^*) := \arg \max_{(i,j) \in B_1 \times B_2} |\sigma_{ij}|$ , where the dependency on  $k$  and  $B_1$  is implicit. Then, using Lemma 1,

$$\begin{aligned}
& \mathbb{P}(B_{n^{-\delta/2}} \not\geq \mathcal{B}(\alpha_1)) \\
&= \mathbb{P}\left(\exists k \in [1 : K(\alpha_1)], \exists \emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1), \max_{(i,j) \in B_1 \times B_2} |\widehat{C}_{ij}| < n^{-\alpha_1}\right) \\
&\leq \mathbb{P}(\exists k \in [1 : K(\alpha_1)], \exists \emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1), |\widehat{\sigma}_{i^*j^*}| < 2\lambda_{\text{sup}}(1 + \sqrt{y})^2 n^{-\alpha_1}) \\
&\quad + \mathbb{P}(\exists i \leq p, \widehat{\sigma}_{ii} \geq 2\lambda_{\text{sup}}(1 + \sqrt{y})^2) \\
&\leq \mathbb{P}(\exists k \in [1 : K(\alpha_1)], \exists \emptyset \subsetneq B_1 \subsetneq B_k(\alpha_1), |\widehat{\sigma}_{i^*j^*}| < 2\lambda_{\text{sup}}(1 + \sqrt{y})^2 n^{-\alpha_1}) \\
&\quad + \mathbb{P}\left(\lambda_{\max}(\widehat{\Sigma}) \geq 2\lambda_{\text{sup}}(1 + \sqrt{y})^2\right) \\
&\leq \mathbb{P}\left(\exists k \in [1 : K], \exists \emptyset \subsetneq B_1 \subsetneq B_k^*, |\widehat{\sigma}_{i^*j^*} - \sigma_{i^*j^*}| > n^{-\alpha_1} - 2\lambda_{\text{sup}}(1 + \sqrt{y})^2 n^{-\delta/2}\right) + o(1) \\
&\leq \mathbb{P}\left(\exists (i, j) \in [1 : p]^2, |\widehat{\sigma}_{ij} - \sigma_{ij}| > n^{-\alpha_1} - 2\lambda_{\text{sup}}(1 + \sqrt{y})^2 n^{-\delta/2}\right) + o(1) \\
&\leq p^2 \max_{(i,j)} \mathbb{P}\left(|\widehat{\sigma}_{ij} - \sigma_{ij}| > n^{-\alpha_1} - 2\lambda_{\text{sup}}(1 + \sqrt{y})^2 n^{-\delta/2}\right) + o(1) \\
&\leq 2p^2 \exp(-C_{\text{inf}}(\delta, \alpha_1) n^{1-2\alpha_1}) + o(1) \longrightarrow 0,
\end{aligned}$$

by Bernstein's inequality. □

### Proof of Proposition 9

*Proof.* First, we prove the results for  $\widehat{\Sigma}_{B^*}$ . We have, using again the notation  $M \sim \mathcal{N}(0, \Sigma)$ ,

$$\begin{aligned}
\mathbb{E}\left(\frac{n}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2\right) &\leq n m^2 \max_{(i,j) \in B^*} \mathbb{E}[(\widehat{\sigma}_{ij} - \sigma_{ij})^2] \\
&= n m^2 \max_{(i,j) \in B^*} \text{Var}(\widehat{\sigma}_{ij}) \\
&\leq m^2 \frac{n}{n-1} \max_{(i,j) \in B^*} \text{Var}(M_i M_j) \\
&\leq 2m^2 \max_{(i,j) \in B^*} (\sigma_{ii} \sigma_{jj} + 2\sigma_{ij}^2) \\
&\leq 6m^2 \lambda_{\text{sup}}^2.
\end{aligned}$$

By Markov's inequality, that proves

$$\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 = O_p(1/n).$$

Now, we want to prove that

$$\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 \neq o_p(1/n).$$

First, we have

$$\begin{aligned} \mathbb{E} \left( \frac{n}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 \right) &\geq n \min_{i \in [1:p]} \mathbb{E} [(\widehat{\sigma}_{ii} - \sigma_{ii})^2] \\ &= n \min_{i \in [1:p]} \text{Var}(\widehat{\sigma}_{ii}) \\ &\geq \frac{n}{n-1} \text{Var}(M_{ii}^2) \\ &\geq \min_{i \in [1:p]} 2\sigma_{ii}^2 \\ &\geq 2\lambda_{\inf}^2. \end{aligned}$$

Now, the variance is

$$\text{Var} \left( \frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 \right) = \sum_{k=1}^K \text{Var} \left( \frac{1}{p} \|\widehat{\Sigma}_{B_k^*} - \Sigma_{B_k^*}\|_F^2 \right) \leq p \max_{k \in [1:K]} \text{Var} \left( \frac{1}{p} \|\widehat{\Sigma}_{B_k^*} - \Sigma_{B_k^*}\|_F^2 \right).$$

Now,

$$\text{Var} \left( \frac{1}{p} \|\widehat{\Sigma}_{B_k^*} - \Sigma_{B_k^*}\|_F^2 \right) = \frac{1}{p^2} \text{Var} \left( \sum_{i,j \in B_k^*} (\widehat{\sigma}_{ij} - \sigma_{ij})^2 \right).$$

Remark that if  $A_1, \dots, A_d$  are random variables, we have

$$\text{Var} \left( \sum_{i=1}^d A_i \right) = \sum_{i,j=1}^d \text{cov}(A_i, A_j) \leq \sum_{i,j=1}^d \sqrt{\text{Var}(A_i)} \sqrt{\text{Var}(A_j)} = \left( \sum_{i=1}^d \sqrt{\text{Var}(A_i)} \right)^2 \leq d \sum_{i=1}^d \text{Var}(A_i).$$

Thus

$$\text{Var} \left( \frac{1}{p} \|\widehat{\Sigma}_{B_k^*} - \Sigma_{B_k^*}\|_F^2 \right) \leq \frac{m^4}{p^2} \max_{i,j \in B_k^*} \text{Var}((\widehat{\sigma}_{ij} - \sigma_{ij})^2).$$

Let  $i, j \in B_k^*$  for some  $k$ . We want to upper-bound  $\text{Var}((\widehat{\sigma}_{ij} - \sigma_{ij})^2)$ . Let us define  $a_k := X_i^{(k)} X_j^{(k)} - \sigma_{ij}$ . We know that

$$(\widehat{\sigma}_{ij} - \sigma_{ij})^2 = \left( \frac{1}{n} \sum_{k=1}^n a_k \right)^2 = \frac{1}{n^2} \sum_{k=1}^n a_k^2 + \frac{1}{n^2} \sum_{k \neq k'} a_k a_{k'}.$$

So, using the independence of  $a_1, \dots, a_n$ , we obtain

$$\begin{aligned}
\text{Var}((\widehat{\sigma}_{ij} - \sigma_{ij})^2) &= \frac{1}{n^4} \sum_{k_1, k_2=1}^n \text{cov}(a_{k_1}^2, a_{k_2}^2) + 2 \frac{1}{n^4} \sum_{\substack{k_1, k_2, k'_2=1, \\ k_2 \neq k'_2}}^n \text{cov}(a_{k_1}^2, a_{k_2} a_{k'_2}) \\
&\quad + \frac{1}{n^4} \sum_{\substack{k_1, k'_1, k_2, k'_2=1, \\ k_1 \neq k'_1, k_2 \neq k'_2}}^n \text{cov}(a_{k_1} a_{k'_1}, a_{k_2} a_{k'_2}) \\
&= \frac{1}{n^3} \text{cov}(a_1^2, a_1^2) + 4 \frac{n-1}{n^3} \text{cov}(a_1^2, a_1 a_2) \\
&\quad + 2 \frac{n-1}{n^3} \text{cov}(a_1 a_2, a_1 a_2),
\end{aligned}$$

where we observed that  $\text{cov}(a_1 a_2, a_1, a_3) = 0$ . Now, by Isserlis' theorem and using the fact that  $\sigma_{ij}$  is upper-bounded by  $\lambda_{\text{sup}}$ , we have  $\text{cov}(a_1^2, a_1^2) \leq C_{\text{sup}}$ ,  $\text{cov}(a_1^2, a_1 a_2) \leq C_{\text{sup}}$  and  $\text{cov}(a_1 a_2, a_1 a_2) \leq C_{\text{sup}}$  (and these bounds do not depend on  $k, i, j$ ). So

$$\text{Var}(\widehat{\sigma}_{ij}^2) \leq \frac{C_{\text{sup}}}{n^2}.$$

Thus,

$$\text{Var}\left(\frac{1}{p} \|\widehat{\Sigma}_{B_k^*} - \Sigma_{B_k^*}\|_F^2\right) \leq \frac{C_{\text{sup}}}{p^2 n^2},$$

and

$$\text{Var}\left(\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2\right) \leq \frac{C_{\text{sup}}}{p n^2}.$$

Thus, by Chebyshev's inequality

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 < \frac{\lambda_{\text{inf}}^2}{n}\right) \\
&\leq \mathbb{P}\left(\left|\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 - \mathbb{E}\left[\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2\right]\right| > \frac{\lambda_{\text{inf}}^2}{n}\right) \\
&\leq \frac{\text{Var}\left(\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2\right) n^2}{\lambda_{\text{inf}}^4} \\
&\leq \frac{C_{\text{sup}}}{p} \rightarrow 0.
\end{aligned}$$

So, we proved that  $\frac{1}{p} \|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2$  is not an  $o_p(1/n)$ .

Now, we show that the same results hold for  $S_{B^*}$  proving that  $\frac{1}{p} \|S_{B^*} -$

$\Sigma\|_F^2 - \frac{1}{p}\|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 = o_p(1/n)$ . We have

$$\begin{aligned}
& \left| \frac{1}{p}\|S_{B^*} - \Sigma\|_F^2 - \frac{1}{p}\|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 \right| \\
&= \left| \frac{1}{p} \sum_{(i,j) \in B^*} 2\sigma_{ij}\widehat{\sigma}_{ij} \frac{1}{n} - \widehat{\sigma}_{ij}^2 \frac{2n-1}{n^2} \right| \\
&\leq \frac{m^2}{n} \max_{(i,j) \in B^*} \left| 2\sigma_{ij}\widehat{\sigma}_{ij} - \frac{2n-1}{n}\widehat{\sigma}_{ij}^2 \right| \\
&\leq \frac{m^2}{n} \max_{(i,j) \in B^*} (2|\widehat{\sigma}_{ij}||\widehat{\sigma}_{ij} - \sigma_{ij}| + |\widehat{\sigma}_{ij}^2|/n).
\end{aligned}$$

Yet, by Bernstein's inequality,

$$\max_{(i,j) \in B^*} |\widehat{\sigma}_{ij}| = O_p(1),$$

$$\max_{(i,j) \in B^*} |\widehat{\sigma}_{ij} - \sigma_{ij}| = o_p(1),$$

and

$$\max_{(i,j) \in B^*} \widehat{\sigma}_{ij}^2 = O_p(1).$$

That proves

$$\frac{1}{p}\|S_{B^*} - \Sigma\|_F^2 - \frac{1}{p}\|\widehat{\Sigma}_{B^*} - \Sigma\|_F^2 = o_p(1/n).$$

Now, on the one hand, we have

$$\frac{1}{p}\|S_{B^*} - \Sigma\|_F^2 = O_p(1/n),$$

and by Proposition 7,

$$\frac{1}{p}\|S_{\widehat{B}} - \Sigma\|_F^2 = O_p(1/n).$$

On the other hand,

$$\frac{1}{p}\|S_{B^*} - \Sigma\|_F^2 \neq o_p(1/n).$$

□

### Proof of Proposition 10

*Proof.* It suffices to prove that

$$\frac{\lambda_{\inf}^2}{2} \leq \mathbb{E} \left( \frac{n}{p^2} \|S - \Sigma\|_F^2 \right) \leq C_{\sup}. \quad (13)$$

First,

$$\begin{aligned}
\mathbb{E} \left( \frac{n}{p^2} \|S - \Sigma\|_F^2 \right) &\leq n \max_{(i,j) \in [1:p]^2} \mathbb{E} [(s_{ij} - \sigma_{ij})^2] \\
&= n \max_{(i,j) \in [1:p]^2} \text{Var}(s_{ij}) + \frac{\sigma_{ij}^2}{n} \\
&= \frac{n-1}{n} \max_{(i,j) \in [1:p]^2} \text{Var}(M_i M_j) + \frac{\sigma_{ij}^2}{n} \\
&\leq \max_{(i,j) \in [1:p]^2} \left( \sigma_{ii} \sigma_{jj} + \sigma_{ij}^2 + \frac{\sigma_{ij}^2}{n} \right) \\
&\leq 3\lambda_{\text{sup}}^2.
\end{aligned}$$

Secondly,

$$\begin{aligned}
\mathbb{E} \left( \frac{n}{p^2} \|S - \Sigma\|_F^2 \right) &\geq n \min_{(i,j) \in [1:p]^2} \mathbb{E} [(s_{ij} - \sigma_{ij})^2] \\
&\geq n \min_{(i,j) \in [1:p]^2} \text{Var}(s_{ij}) \\
&= \frac{n-1}{n} \min_{(i,j) \in [1:p]^2} (\sigma_{ii} \sigma_{jj} + \sigma_{ij}^2) \\
&\geq \frac{1}{2} \lambda_{\text{inf}}^2.
\end{aligned}$$

□

### Proof of Proposition 11

*Proof.* We follow the proof of Proposition 9. Let  $\delta \in ]1/2, 1[$ ,  $\varepsilon > 0$ , and  $\alpha_1 := \delta/2 - \varepsilon/4$ .

We have

$$\begin{aligned}
&\max_{B(\alpha_1) \leq B \leq B^*} \mathbb{E} \left( \frac{n^{\delta-\varepsilon}}{p} \|\widehat{\Sigma}_B - \Sigma\|_F^2 \right) \\
&= \max_{B(\alpha_1) \leq B \leq B^*} \frac{n^{\delta-\varepsilon}}{p} \sum_{k=1}^K \left[ \sum_{\substack{i,j \in B_k^* \\ (i,j) \in B}} \mathbb{E} ((\widehat{\sigma}_{ij} - \sigma_{ij})^2) + \sum_{\substack{i,j \in B_k^* \\ (i,j) \in B}} \sigma_{ij}^2 \right] \\
&\leq n^{\delta-\varepsilon} m^2 \left( \max_{(i,j) \in B^*} \mathbb{E} (\widehat{\sigma}_{ij} - \sigma_{ij})^2 \right) + \max_{B(\alpha_1) \leq B \leq B^*} \max_{(i,j) \in B^* \setminus B} \sigma_{ij}^2 \\
&\leq n^{\delta-\varepsilon} m^2 \left( O \left( \frac{1}{n} \right) + n^{-2\alpha_1} \right) \longrightarrow 0.
\end{aligned}$$



Thus,

$$\max_{B(\alpha_1) \leq B \leq B^*} \frac{1}{p} \|\widehat{\Sigma}_B - \Sigma\|_F^2 = o_p \left( \frac{1}{n^{\delta-\varepsilon}} \right),$$

and thus

$$\max_{B(\alpha_1) \leq B \leq B^*} \frac{1}{p} \|S_B - \Sigma\|_F^2 = o_p \left( \frac{1}{n^{\delta-\varepsilon}} \right).$$

We conclude using Proposition 8 and using that  $B(\alpha_2) \leq B^*$ .  $\square$

### Proof of Proposition 12

*Proof.* The eigenvalues of  $\Sigma$  are lower-bounded by  $\varepsilon$  and upper-bounded by  $mL$ , so  $\Sigma$  verifies Condition 2. Condition 3 is verified by construction. It remains to prove the slightly modified Condition 4 given in Proposition 12. Let  $a > 0$ .

$$\begin{aligned} & \mathbb{P} \left( \exists B < B^*, \|\Sigma_B - \Sigma\|_{\max} < an^{-1/4} \right) \\ &= \mathbb{P} \left( \exists k, \max_{i,j \in B_k^*, i \neq j} |\sigma_{ij}| < an^{-1/4} \right) \\ &\leq p \mathbb{P} \left( \max_{i,j \in [1:10], i \neq j} \left| \sum_{l=1}^L U_i^{(l)} U_j^{(l)} \right| \leq an^{-1/4} \right), \end{aligned}$$

using an union bound and the fact that all the blocks have a size larger than 10. Then, by independence of  $\left( \sum_{l=1}^L U_{2k-1}^{(l)} U_{2k}^{(l)} \right)_{k \leq 5}$ , we have

$$\mathbb{P} \left( \exists B < B^*, \|\Sigma_B - \Sigma\|_{\max} < an^{-1/4} \right) \leq p \mathbb{P} \left( \left| \sum_{l=1}^L U_1^{(l)} U_2^{(l)} \right| \leq an^{-1/4} \right)^5$$

Let  $U_i := (U_i^{(l)})_{l \leq L} \in \mathbb{R}^L$  for  $i = 1, 2$ . Then,  $U_1$  and  $U_2$  are independent and uniformly distributed on  $[-1, 1]^L$ . Thus

$$\mathbb{P} \left( \left| \sum_{l=1}^L U_1^{(l)} U_2^{(l)} \right| \leq an^{-1/4} \right) = \mathbb{E} \left[ \mathbb{P} \left( |\langle U_1, U_2 \rangle| \leq an^{-1/4} \mid U_2 \right) \right]$$

Let  $u_2 \in [-1, 1]^L \setminus \{0\}$ . The set  $\{u_1 \in [-1, 1]^L \mid |\langle u_1, u_2 \rangle| \leq an^{-1/4}\}$  is a subset of  $\{\sum_{i=1}^L x_i e_i \mid -an^{-1/4} \|u_2\| \leq x_1 \leq an^{-1/4} \|u_2\|, |x_l| \leq \sqrt{L} \forall l\}$  where  $e_1 = u_2 / \|u_2\|$  and  $(e_1, \dots, e_L)$  is an orthonormal basis of  $\mathbb{R}^L$ . The Lebesgue measure of this subset is  $(2\sqrt{L})^{L-1} 2an^{-1/4} \|u_2\|$ . Furthermore, (conditionally to  $U_2 = u_2$ ) the probability density function of  $U_1$  on this set is either 0 or  $2^{-L}$ . So, for all  $u_2 \in [-1, 1]^L \setminus \{0\}$ ,

$$\mathbb{P} \left( |\langle U_1, U_2 \rangle| \leq an^{-1/4} \mid U_2 = u_2 \right) \leq (2\sqrt{L})^{L-1} 2an^{-1/4} \|u_2\| 2^{-L} \leq \sqrt{L}^{L-1} an^{-1/4}.$$

Thus

$$\mathbb{P}\left(\left|\sum_{l=1}^L U_1^{(l)} U_2^{(l)}\right| \leq an^{-1/4}\right) \leq \sqrt{L}^{L-1} an^{-1/4}.$$

Then

$$\mathbb{P}\left(\exists B < B^*, \|\Sigma_B - \Sigma\|_{\max} < an^{-1/4}\right) \leq p(\sqrt{L}^{L-1} an^{-1/4})^5 \rightarrow 0.$$

Hence, it remains to prove that the conclusion of Proposition 2 holds. That will imply the same for Propositions 7 and 9. Let  $a > 0$  and  $E := \{\Gamma \in S_p^{++}(\mathbb{R}, B^*) \mid \forall B < B^*, \|\Sigma_B - \Sigma\|_{\max} \geq an^{-1/4}\}$ , where the generation of  $B^*$  is defined in Proposition 12. We have

$$\begin{aligned} \mathbb{P}\left(\widehat{B}_{tot} \neq B^*\right) &\leq \mathbb{P}(\Sigma \notin E) + \mathbb{P}\left(\widehat{B}_{tot} \neq B^* \mid \Sigma \in E\right) \\ &\leq o(1) + \int_E \mathbb{P}\left(\widehat{B}_{tot} \neq B^* \mid \Sigma = \Gamma\right) d\mathbb{P}_\Sigma(\Gamma). \end{aligned}$$

Yet, for all  $\Sigma \in E$ ,  $\mathbb{P}\left(\widehat{B}_{tot} \neq B^* \mid \Sigma = \Gamma\right) \rightarrow 0$  thanks to Proposition 2 (even in Condition 4 is not verified, the proof is still valid since the covariance matrix is in  $E$ ). We conclude by dominated convergence theorem.  $\square$

### Notations for the proofs of Section 2.3

For all  $i, j \in [1 : p]$ , let  $e_i \in \mathbb{R}^p$  be such that all coefficients are zero except the  $i$ -th one which is equal to 1, and let  $e_{ij} \in \mathcal{M}_p(\mathbb{R})$  be such that all coefficients are zero except the  $(i, j)$ -th one which is equal to 1. Let  $\gamma_{ij}$  be the  $(i, j)$ -th coefficient of  $\Sigma^{-1}$ . Finally, as we use matrices  $M$  of size  $p^2 \times p^2$ , and vectors  $v$  of size  $p^2$ , we define  $v_{ij} := v_{(j-1)p+i}$  and  $M_{ij,kl} := M_{(j-1)p+i, (l-1)p+k}$ .

### Proof of Proposition 13

We see that, for all  $B \in \mathcal{P}_p$ ,  $l_{S_B} = \log(|S_B|)/p + \frac{n-1}{n}$  converges almost surely to  $\log(|\Sigma_B|)/p + 1$ . The following Lemma gives a central limit theorem for this convergence.

**Lemma 12.** *For all  $B \in \mathcal{P}_p$ , we have*

$$\sqrt{n}(\log |S_B| - \log |\Sigma_B|) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 2 \operatorname{Tr}(\Sigma_B^{-1} \Sigma \Sigma_B^{-1} \Sigma)/p), \quad (14)$$

with  $2 \operatorname{Tr}(\Sigma_B^{-1} \Sigma \Sigma_B^{-1} \Sigma)/p \leq 2p$ . In particular

$$\sqrt{n}(\log |S| - \log |\Sigma|) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 2).$$

*Proof.* Let  $Z^{(k)} = M^{(k)} M^{(k)T}$ , where  $M^{(k)} = (M_i^{(k)})_{i \leq p} \in R^p$ . We know that  $\mathbb{E}(Z) = \Sigma$  and  $\operatorname{cov}(Z_{i,j}, Z_{k,l}) = \mathbb{E}(X_i X_j X_k X_l) - \sigma_{ij} \sigma_{kl} = \sigma_{ij} \sigma_{kl} + \sigma_{ik} \sigma_{jl} +$

$\sigma_{il}\sigma_{jk} - \sigma_{ij}\sigma_{kl} = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$ . Let  $\Gamma \in \mathcal{M}_{p^2, p^2}$ , be such that  $\Gamma_{ij,kl} := \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} = \text{cov}(Z_{i,j}, Z_{k,l})$ . Using the central limit Theorem,

$$\sqrt{n-1} \left( \text{vec}(\widehat{\Sigma}_B) - \text{vec}(\Sigma_B) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_B),$$

and by Slutsky Lemma,

$$\sqrt{n} (\text{vec}(S_B) - \text{vec}(\Sigma_B)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_B), \quad (15)$$

where  $(\Gamma_B)_{ij,kl} = \Gamma_{ij,kl}$  if  $(i, j) \in B$  and  $(k, l) \in B$  and  $(\Gamma_B)_{ij,kl} = 0$  otherwise.

Let us apply the Delta-method to (15) with the function  $\log \circ \det \circ \text{mat}$ , where  $\text{mat} = \mathbb{R}^{p^2} \rightarrow \mathcal{M}_p(\mathbb{R})$  is the inverse function of  $\text{vec}$ . If we write  $L$  the Jacobian matrix of  $\log \circ \det \circ \text{mat}$ , we have:

$$\sqrt{n} (\log |S_B| - \log |\Sigma_B|) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, L(\text{vec}(\Sigma_B)) \Gamma_B L(\text{vec}(\Sigma_B))^T).$$

Let us compute the linear map  $L(\text{vec}(\Sigma_B)) : \mathbb{R}^{p^2} \rightarrow \mathbb{R}$ , that we identify with its matrix. Let us recall that, for the dot product  $\langle A, B \rangle := \text{Tr}(A^T B)$ , the gradient of  $\log \circ \det$  on  $A$  is  $A^{-1}$ . Thus, if  $v \in \mathbb{R}^{p^2}$ , we have

$$\begin{aligned} L(\text{vec}(\Sigma_B))(v) &= D(\log \circ \det)(\text{mat}(\text{vec}(\Sigma_B)) \circ D \text{mat}(\text{vec}(\Sigma_B))(v)) \\ &= \langle \nabla(\log \circ \det)(\Sigma_B), D \text{mat}(\text{vec}(\Sigma))(v) \rangle, \\ &= \langle \Sigma_B^{-1}, \Sigma \rangle \\ &= \text{Tr}(\Sigma_B^{-1} \text{mat}(v)) \\ &= \text{vec}(\Sigma_B^{-1})^T v. \end{aligned}$$

So  $L(\text{vec}(\Sigma_B)) = \text{vec}(\Sigma_B^{-1})^T$ , then

$$\sqrt{n} (\log |S_B| - \log |\Sigma_B|) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \text{vec}(\Sigma_B^{-1})^T \Gamma_B \text{vec}(\Sigma_B^{-1})).$$

Now,

$$\begin{aligned} & \text{vec}(\Sigma_B^{-1})^T \Gamma_B \text{vec}(\Sigma_B^{-1}) \\ &= \sum_{i,j,k,l} (\Sigma_B)_{i,j}^{-1} (\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) (\Sigma_B)_{k,l}^{-1} \\ &= \sum_{i,j,k,l} (\Sigma_B)_{i,j}^{-1} \sigma_{ik}\sigma_{jl} (\Sigma_B)_{k,l}^{-1} + \sum_{i,j,k,l} (\Sigma_B)_{i,j}^{-1} \sigma_{il}\sigma_{jk} (\Sigma_B)_{k,l}^{-1} \\ &= 2 \text{Tr}(\Sigma_B^{-1} \Sigma \Sigma_B^{-1} \Sigma) \\ &= 2 \text{Tr} \left[ \left( \Sigma_B^{-\frac{1}{2}} \Sigma \Sigma_B^{-\frac{1}{2}} \right) \left( \Sigma_B^{-\frac{1}{2}} \Sigma \Sigma_B^{-\frac{1}{2}} \right) \right] \\ &\leq 2 \text{Tr} \left( \Sigma_B^{-\frac{1}{2}} \Sigma \Sigma_B^{-\frac{1}{2}} \right)^2 \\ &= 2 \text{Tr}(\Sigma_B^{-1} \Sigma)^2 \\ &= 2p^2. \end{aligned}$$

Indeed, as  $A := \Sigma_B^{-\frac{1}{2}} \Sigma \Sigma_B^{-\frac{1}{2}}$  is symmetric positive definite, we have  $\text{Tr}(AA) \leq \text{Tr}(A)^2$ .  $\square$

**Lemma 13.** *For all  $\Gamma \in S_p^{++}(\mathbb{R})$  and for all  $B \in \mathcal{P}_p$  such that  $\Gamma \neq \Gamma_B$ , we have  $\det(\Gamma_B) > \det(\Gamma)$ .*

*Proof.* First, let us prove it for  $|B| = K = 2$ . We have  $B = \{I, J\}$ .

$$\det(\Gamma) = \det(\Gamma_{I,I}) \det(\Gamma_{J,J} - \Gamma_{J,I} \Gamma_{I,I}^{-1} \Gamma_{I,J}).$$

Now,  $\det(\Gamma_B) = \det(\Gamma_{I,I}) \det(\Gamma_{J,J})$ . Thus, it suffices to show that  $\det(\Gamma_{J,J}) > \det(\Gamma_{J,J} - \Gamma_{J,I} \Gamma_{I,I}^{-1} \Gamma_{I,J})$ . We then write  $A_1 := \Gamma_{J,J} - \Gamma_{J,I} \Gamma_{I,I}^{-1} \Gamma_{I,J}$  which is symmetric positive definite (Schur's complement), and  $A_2 = \Gamma_{J,I} \Gamma_{I,I}^{-1} \Gamma_{I,J}$  which is also symmetric positive definite. Then, we have

$$\det(A_1 + A_2) = \det(A_1) \det(I_p + A_1^{-\frac{1}{2}} A_2 A_1^{-\frac{1}{2}}) > \det(A_1),$$

because  $\det(I_p + A_1^{-\frac{1}{2}} A_2 A_1^{-\frac{1}{2}}) = \prod_{i=1}^p (1 + \phi_i(A_1^{-\frac{1}{2}} A_2 A_1^{-\frac{1}{2}}))$ .

Now, we prove the lemma for any value of  $|B| = K$ . Let  $\Gamma \in S_p^{++}(\mathbb{R})$  and  $B \in \mathcal{P}_p$  such that  $\Gamma \neq \Gamma_B$ . Let  $B^{(j)} := \{\bigcup_{i=1}^j B_i, \bigcup_{i=j+1}^K B_i\}$  for all  $j \in [1 : K-1]$ . We now define  $(\Gamma^{(j)})_{j \in [1, K]}$  with the recurrence relation  $\Gamma_{(j+1)} = \Gamma_{B^{(j)}}^{(j)}$  and with  $\Gamma^{(1)} = \Gamma$ , we then have  $\Gamma_B = \Gamma^K$ . Thus

$$\det(\Gamma_B) = \det(\Gamma^{(K)}) \geq \det(\Gamma^{(K-1)}) \geq \dots \geq \det(\Gamma^{(1)}) = \det(\Gamma).$$

Furthermore, as  $\Gamma \neq \Gamma_B$ , there exists  $j$  such that  $\Gamma_{B^{(j)}}^{(j)} \neq \Gamma^{(j)}$ . Thus, at least one of the previous inequality is strict, and so  $\det(\Gamma_B) > \det(\Gamma)$ .  $\square$

Using Lemmas 12 and 13, we can prove Proposition 13.

*Proof.* It suffices to show that, for all  $B \neq B^*$ ,

$$\mathbb{P}(\widehat{B}_{tot} = B) \xrightarrow{n \rightarrow +\infty} 0.$$

We split the proof into two steps: for  $B \not\geq B^*$  and for  $B > B^*$ .

Step 1:  $B \not\geq B^*$ .

Let  $h := \min\{\log(|\Sigma_B|) - \log(|\Sigma|) \mid B \not\geq B^*\} = \min\{\log(|\Sigma_B|) - \log(|\Sigma|), B < B^*\}$ , since  $\Sigma_B = \Sigma_{B \cap B^*}$ . Thanks to Lemma 13, we know that  $h > 0$ .

Let  $B \not\geq B^*$ . Using the convergence in probability of  $l_{S'_B}$ , we know that  $\mathbb{P}(l_{S_B} < \log |\Sigma_B| / p + 1 - h/3) \xrightarrow{n \rightarrow +\infty} 0$  and  $\mathbb{P}(l_{S_{B^*}} > \log |\Sigma| / p + 1 + h/3) \xrightarrow{n \rightarrow +\infty} 0$ .

Now, we know that for  $n > (3p/h)^{1/\delta}$ , the term of penalisation satisfies  $\kappa \text{pen}(B^*) < h/3$ . Thus,

$$\mathbb{P}(\widehat{B}_{tot} = B) \xrightarrow{n \rightarrow +\infty} 0.$$

Step 2:  $B > B^*$ .

Let  $B > B^*$ . We know that

$$\begin{aligned} & \sqrt{n}(\Psi(B) - \Psi(B^*)) \\ &= \sqrt{n}(l_{S_B} + \kappa \text{pen}(B) - l_{S_{B^*}} - \kappa \text{pen}(B^*)) \\ &= \sqrt{n}\kappa(\text{pen}(B) - \text{pen}(B^*)) + \sqrt{n}(l_{S_B} - l_{\Sigma_B}) - \sqrt{n}(l_{S_{B^*}} - l_{\Sigma_{B^*}}), \end{aligned}$$

since  $\Sigma_B = \Sigma_{B^*}$  for  $B > B^*$ . Let  $a_n$  be equal to  $\sqrt{n}\kappa(\text{pen}(B) - \text{pen}(B^*))$  (which converges to  $+\infty$ ),  $b_n$  to be equal to  $\sqrt{n}(l_{S_B} - l_{\Sigma_B})$  (which converges to a zero mean normal distribution) and  $c_n$  to be equal to  $\sqrt{n}(l_{S_{B^*}} - l_{\Sigma_{B^*}})$  (which converges to a zero mean normal distribution). We have

$$\begin{aligned} \mathbb{P}[\sqrt{n}(\Psi(B) - \Psi(B^*)) > 0] &= \mathbb{P}(b_n - c_n < -a_n) \\ &\leq \mathbb{P}(b_n \leq -a_n/2 \text{ or } c_n \geq a_n/2) \\ &\leq \mathbb{P}(b_n \leq -a_n/2) + \mathbb{P}(c_n \geq a_n/2) \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Thus,  $\mathbb{P}(\hat{B}_{tot} = B) \xrightarrow{n \rightarrow +\infty} 0$ .  $\square$

#### Proof of Proposition 14

*Proof.* We follow the notations of [30].

An orthonormal basis of  $S_p(\mathbb{R})$  is  $\{\frac{1}{\sqrt{2}}(e_{ij} + e_{ji}) \mid i < j\} \cup \{e_{ii} \mid i \leq p\}$  with the following total order on  $\{(i, j) \in [1 : p]^2 \mid i \leq j\}$ : we write  $(i, j) \leq (i', j')$  if  $j < j'$  or if  $j = j'$  and  $i \leq i'$ . We define  $U \in \mathcal{M}_{p^2, p(p+1)/2}(\mathbb{R})$  as the matrix which columns are the vectorizations of the components of this basis of  $\text{vec}(S_p(\mathbb{R}))$ . Thus  $U_{ij,kl} = \frac{1}{\sqrt{2}}(\mathbb{1}_{(i,j)=(k,l)} + \mathbb{1}_{(i,j)=(l,k)})$ , for all  $k < l$  and  $U_{ij,kk} = \mathbb{1}_{(i,j)=(k,k)}$ .

Thus,  $U(U^T J U)^{-1} U^T$  is the Cramér-Rao bound, where  $J$  is the standard Fisher information matrix in the model  $\{\mathcal{N}(\mu, \Sigma), \Sigma \in \mathcal{M}_p(\mathbb{R})\}$ . As the sample is i.i.d, it suffices to prove it with  $n = 1$ . In the rest of the proof, we compute the Cramér-Rao bound, and we show that this bound is equal to  $\mathbb{E}((S - \Sigma)(S - \Sigma)^T)$ . We split the proof into several Lemmas.

**Lemma 14.** *Recall that  $\Sigma^{-1} = (\gamma_{ij})_{i,j \leq p}$ . Let  $A = (A_{mn,m'n'})_{m \leq n, m' \leq n'} \in \mathcal{M}_{p(p+1)/2}(\mathbb{R})$  defined by*

$$A_{mn,m'n'} = \begin{cases} \frac{1}{2}(\gamma_{mm'}\gamma_{nn'} + \gamma_{mn'}\gamma_{nm'}) & \text{if } m < n \text{ and } m' < n' \\ \frac{1}{\sqrt{2}}\gamma_{mm'}\gamma_{nn'} & \text{if either } m = n \text{ or } m' = n' \\ \frac{1}{2}\gamma_{mm'}^2 & \text{if } m = n \text{ and } m' = n', \end{cases}$$

Then,  $A = U^T J U$ .

*Proof.* Deriving twice the log-likelihood with respect to  $\sigma_{ij}$  and  $\sigma_{kl}$  (for  $i, j, k, l \in [1 : p]$ ) and taking the expectation, we get

$$\begin{aligned} J_{ij,kl} &= \frac{1}{2} \text{Tr} (\Sigma^{-1} e_i e_j^T \Sigma^{-1} e_k e_l^T) \\ &= \frac{1}{2} \gamma_{li} \gamma_{jk}. \end{aligned}$$

Thus, for all  $m < n$ ,  $m' < n'$ , we have

$$\begin{aligned} (U^T J U)_{mn,m'n'} &= \sum_{i,j,k,l=1}^p U_{ij,mn} J_{ij,kl} U_{kl,m'n'} \\ &= \sum_{i,j,k,l=1}^p \frac{1}{\sqrt{2}} (\mathbb{1}_{(i,j)=(m,n)} + \mathbb{1}_{(i,j)=(n,m)}) J_{ij,kl} \frac{1}{\sqrt{2}} (\mathbb{1}_{(k,l)=(m',n')} + \mathbb{1}_{(k,l)=(n',m')}) \\ &= \frac{1}{2} (J_{mn,m'n'} + J_{mn,n'm'} + J_{nm,m'n'} + J_{nm,n'm'}) \\ &= \frac{1}{2} (\gamma_{mm'} \gamma_{nn'} + \gamma_{mn'} \gamma_{nm'}). \end{aligned}$$

Now, if  $m' < n'$ , we have

$$\begin{aligned} (U^T J U)_{mm,m'n'} &= \sum_{i,j,k,l=1}^p U_{ij,mm} J_{ij,kl} U_{kl,m'n'} \\ &= \sum_{i,j,k,l=1}^p \mathbb{1}_{(i,j)=(m,m)} J_{ij,kl} \frac{1}{\sqrt{2}} (\mathbb{1}_{(k,l)=(m',n')} + \mathbb{1}_{(k,l)=(n',m')}) \\ &= \frac{1}{\sqrt{2}} (J_{mm,m'n'} + J_{mm,n'm'}) \\ &= \frac{1}{\sqrt{2}} \gamma_{mm'} \gamma_{mn'}. \end{aligned}$$

If  $m < n$ , we have

$$\begin{aligned} (U^T J U)_{mn,m'm'} &= \sum_{i,j,k,l=1}^p U_{ij,mn} J_{ij,kl} U_{kl,m'm'} \\ &= \sum_{i,j,k,l=1}^p \frac{1}{\sqrt{2}} (\mathbb{1}_{(i,j)=(m,n)} + \mathbb{1}_{(i,j)=(n,m)}) J_{ij,kl} \mathbb{1}_{(k,l)=(m',m')} \\ &= \frac{1}{\sqrt{2}} (J_{mn,m'm'} + J_{nm,m'm'}) \\ &= \frac{1}{\sqrt{2}} \gamma_{mm'} \gamma_{nm'}. \end{aligned}$$

Finally,

$$\begin{aligned}
(U^T J U)_{mm, m' m'} &= \sum_{i, j, k, l=1}^p U_{ij, mm} J_{ij, kl} U_{kl, m' m'} \\
&= \sum_{i, j, k, l=1}^p \mathbf{1}_{(i, j)=(m, m)} J_{ij, kl} \mathbf{1}_{(k, l)=(m', m')} \\
&= J_{mm, m' m'} \\
&= \frac{1}{2} \gamma_{mm'}^2.
\end{aligned}$$

□

**Lemma 15.** Let  $B = (B_{mn, m' n'})_{m \leq n, m' \leq n'} \in \mathcal{M}_{p(p+1)/2}(\mathbb{R})$  defined by

$$B_{mn, m' n'} = \begin{cases} 2(\sigma_{mm'} \sigma_{nn'} + \sigma_{mn'} \sigma_{nm'}) & \text{if } m < n \text{ and } m' < n' \\ 2\sqrt{2} \sigma_{mm'} \sigma_{nn'} & \text{if either } m = n \text{ or } m' = n' \\ 2\sigma_{mm'}^2 & \text{if } m = n \text{ and } m' = n', \end{cases}$$

then,  $B = A^{-1}$ . Moreover  $(UBU^T)_{ij, i' j'} = \sigma_{ii'} \sigma_{j j'} + \sigma_{ij'} \sigma_{j i'}$  for all  $i, j, i', j' \in [1 : p]$ .

*Proof.* We compute the product  $AB$ . First of all, let  $m < n$  and  $m' < n'$ . We have

$$\begin{aligned}
(AB)_{mn, m' n'} &= \sum_{i \leq j} A_{mn, ij} B_{ij, m' n'} \\
&= \sum_{i < j} A_{mn, ij} B_{ij, m' n'} + \sum_{i=j} A_{mn, ij} B_{ij, m' n'} \\
&= \sum_{i < j} (\gamma_{mi} \gamma_{nj} + \gamma_{mj} \gamma_{ni}) (\sigma_{im'} \sigma_{jn'} + \sigma_{in'} \sigma_{jm'}) \\
&\quad + 2 \sum_{i=j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jn'} \\
&= I_1 + I_2,
\end{aligned}$$

with

$$I_1 = \sum_{i < j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jn'} + \sum_{i < j} \gamma_{mj} \gamma_{ni} \sigma_{in'} \sigma_{jm'} + \sum_{i=j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jn'},$$

and

$$I_2 = \sum_{i < j} \gamma_{mj} \gamma_{ni} \sigma_{im'} \sigma_{jn'} + \sum_{i < j} \gamma_{mi} \gamma_{nj} \sigma_{in'} \sigma_{jm'} + \sum_{i=j} \gamma_{mj} \gamma_{ni} \sigma_{im'} \sigma_{jn'}.$$

We then have

$$\begin{aligned}
I_1 &= \sum_{i < j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jn'} + \sum_{j < i} \gamma_{mi} \gamma_{nj} \sigma_{jn'} \sigma_{im'} + \sum_{i=j} \gamma_{im} \gamma_{jn} \sigma_{im'} \sigma_{jn'} \\
&= \sum_{i,j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jn'} \\
&= \sum_i \gamma_{mi} \sigma_{im'} \sum_j \gamma_{nj} \sigma_{jn'} \\
&= \mathbf{1}_{(m,n)=(m',n')}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
I_2 &= \sum_{i < j} \gamma_{mj} \gamma_{ni} \sigma_{im'} \sigma_{jn'} + \sum_{j < i} \gamma_{mj} \gamma_{ni} \sigma_{jn'} \sigma_{im'} + \sum_{i=j} \gamma_{mj} \gamma_{ni} \sigma_{im'} \sigma_{jn'} \\
&= \sum_{i,j} \gamma_{mj} \gamma_{ni} \sigma_{im'} \sigma_{jn'} \\
&= \sum_i \gamma_{ni} \sigma_{im'} \sum_j \gamma_{mj} \sigma_{jn'} \\
&= \mathbf{1}_{(n,m)=(m',n')} \\
&= 0.
\end{aligned}$$

Now, if  $m' < n'$

$$\begin{aligned}
(A B)_{mm,m'n'} &= \sum_{i \leq j} A_{mm,ij} B_{ij,m'n'} \\
&= \sum_{i < j} A_{mm,ij} B_{ij,m'n'} + \sum_{i=j} A_{mm,ij} B_{ij,m'n'} \\
&= \sqrt{2} \sum_{i < j} \gamma_{mi} \gamma_{mj} (\sigma_{im'} \sigma_{jn'} + \sigma_{in'} \sigma_{jm'}) \\
&\quad + \sqrt{2} \sum_{i=j} \gamma_{mi} \gamma_{mj} \sigma_{im'} \sigma_{jn'} \\
&= \sqrt{2} \sum_{i,j} \gamma_{mi} \gamma_{mj} \sigma_{im'} \sigma_{jn'} \\
&= \sqrt{2} \sum_i \gamma_{mi} \sigma_{im'} \sum_j \gamma_{mj} \sigma_{jn'} \\
&= \sqrt{2} \mathbf{1}_{(m,m)=(m',n')} \\
&= 0.
\end{aligned}$$



If  $m < n$ , then

$$\begin{aligned}
(A B)_{mn,m'm'} &= \sum_{i \leq j} A_{mn,ij} B_{ij,m'm'} \\
&= \sum_{i < j} A_{mn,ij} B_{ij,m'm'} + \sum_{i=j} A_{mn,ij} B_{ij,m'm'} \\
&= \sqrt{2} \sum_{i < j} (\gamma_{mi} \gamma_{nj} + \gamma_{mj} \gamma_{ni}) \sigma_{im'} \sigma_{jm'} \\
&\quad + \sqrt{2} \sum_{i=j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jm'} \\
&= \sqrt{2} \sum_{i,j} \gamma_{mi} \gamma_{nj} \sigma_{im'} \sigma_{jm'} \\
&= \sqrt{2} \sum_i \gamma_{mi} \sigma_{im'} \sum_j \gamma_{nj} \sigma_{jn'} \\
&= \sqrt{2} \mathbf{1}_{(m,n)=(m',m')} \\
&= 0.
\end{aligned}$$

Finally,

$$\begin{aligned}
(A B)_{mm,m'm'} &= \sum_{i \leq j} A_{mm,ij} B_{ij,m'm'} \\
&= \sum_{i < j} A_{mm,ij} B_{ij,m'm'} + \sum_{i=j} A_{mm,ij} B_{ij,m'm'} \\
&= 2 \sum_{i < j} \gamma_{mi} \gamma_{mj} \sigma_{im'} \sigma_{jm'} \\
&\quad + \sum_{i=j} \gamma_{mi} \gamma_{mj} \sigma_{im'} \sigma_{jm'} \\
&= \sum_{i,j} \gamma_{mi} \gamma_{mj} \sigma_{im'} \sigma_{jm'} \\
&= \mathbf{1}_{m=m'}.
\end{aligned}$$

We proved that  $B = A^{-1}$ . Let us show that  $(U^T B U)_{ij,i'j'} = \sigma_{ii'} \sigma_{jj'} + \sigma_{ij'} \sigma_{ji'}$  for all  $i, j, i', j'$ . First of all, assume  $i \neq j$  and  $i' \neq j'$ . Assume for example  $i < j$  and  $i' < j'$ . Then we have

$$\begin{aligned}
&(U B U^T)_{ij,i'j'} \\
&= \sum_{m \leq n, m' \leq n'} U_{ij,mn} B_{mn,m'n'} U_{i'j',m'n'} \\
&= \sum_{m \leq n, m' \leq n'} \frac{1}{\sqrt{2}} (\mathbf{1}_{(m,n)=(i,j)} + \mathbf{1}_{(m,n)=(j,i)}) B_{mn,m'n'} \frac{1}{\sqrt{2}} (\mathbf{1}_{(m',n')=(i',j')} + \mathbf{1}_{(m',n')=(j',i')}) \\
&= \frac{1}{2} B_{ij,i'j'} \\
&= \sigma_{ii'} \sigma_{jj'} + \sigma_{ij'} \sigma_{ji'}.
\end{aligned}$$

We apply the same method for  $i < j$  and  $i' > j'$ , for  $i > j$  and  $i' < j'$ , and for  $i > j$  and  $i' > j'$ . Then, let  $i = j$  and  $i' \neq j'$ , for example  $i' < j'$ . We have

$$\begin{aligned}
& (UBU^T)_{ii,i'j'} \\
&= \sum_{m \leq n, m' \leq n'} U_{ii, mn} B_{mn, m'n'} U_{i'j', m'n'} \\
&= \sum_{m \leq n, m' \leq n'} \mathbb{1}_{(m,n)=(i,i)} B_{mn, m'n'} \frac{1}{\sqrt{2}} (\mathbb{1}_{(m',n')=(i',j')} + \mathbb{1}_{(m',n')=(j',i')}) \\
&= \frac{1}{\sqrt{2}} B_{ii,i'j'} \\
&= \sigma_{ii'} \sigma_{ij'} + \sigma_{ij'} \sigma_{ii'}.
\end{aligned}$$

The other cases are similar.  $\square$

We thus have the component of the Cramér-Rao bound:

$$[U(U^T J U)^{-1} U^T]_{ij,i'j'} = \sigma_{ii'} \sigma_{jj'} + \sigma_{ij'} \sigma_{ji'}.$$

This matrix is equal to  $\mathbb{E}((\text{vec}(S) - \text{vec}(\Sigma))(\text{vec}(S) - \text{vec}(\Sigma))^T)$  for  $n = 1$  and when the mean  $\mu$  is known.  $\square$

### Proof of Proposition 15

*Proof.* An orthonormal basis of  $S_p(\mathbb{R}, B^*)$  is  $\{\frac{1}{\sqrt{2}}(e_{ij} + e_{ji}) \mid i < j, (i, j) \in B^*\} \cup \{e_{ii} \mid i \leq p\}$  with the following total order on  $\{(i, j) \in [1 : p]^2 \mid i \leq j, (i, j) \in B^*\}$ : we write  $(i, j) \leq (i', j')$  if  $j < j'$  or if  $j = j'$  and  $i \leq i'$ . Thus, we define  $U$  as the matrix which the columns are the vectorizations of the components of this basis of  $S_p(\mathbb{R}, B^*)$ . We have  $U_{ij,kl} = \frac{1}{\sqrt{2}}(\mathbb{1}_{(i,j)=(k,l)} + \mathbb{1}_{(i,j)=(l,k)})$ , for all  $k < l$  with  $(k, l) \in B^*$  and  $U_{ij,kk} = \mathbb{1}_{(i,j)=(kk)}$ .

Thus,  $U(U^T J U)^{-1} U^T$  is the Cramér-Rao bound. As the sample is i.i.d, it suffices to prove the proposition with  $n = 1$ .

**Lemma 16.** Let  $A = (A_{mn,m'n'})_{(m,n), (m',n') \in B^*, m \leq n, m' \leq n'}$  defined by

$$A_{mn,m'n'} = \begin{cases} \frac{1}{2}(\gamma_{mm'} \gamma_{nn'} + \gamma_{mn'} \gamma_{nm'}) & \text{if } m < n \text{ and } m' < n' \\ \frac{1}{\sqrt{2}} \gamma_{mm'} \gamma_{nn'} & \text{if either } m = n \text{ or } m' = n' \\ \frac{1}{2} \gamma_{mm'}^2 & \text{if } m = n \text{ and } m' = n', \end{cases}$$

Then,  $A = U^T J U$ .

*Proof.* The proof is similar to the proof of Lemma 14, except that the values of  $m, n, m'$  and  $n'$  are more constraint. First of all

$$\begin{aligned}
J_{ij,kl} &= \frac{1}{2} \text{Tr}(\Sigma^{-1} e_i e_j^T \Sigma^{-1} e_k e_l^T) \\
&= \frac{1}{2} \gamma_{il} \gamma_{jk}.
\end{aligned}$$

Now, if  $(m, n) \in B^*$ ,  $(m', n') \in B^*$ ,  $m < n$ ,  $m' < n'$ ,

$$\begin{aligned}
(U^T JU)_{mn, m'n'} &= \sum_{i,j,k,l=1}^p U_{ij,mn} J_{ij,kl} U_{kl, m'n'} \\
&= \sum_{i,j,k,l=1}^p \frac{1}{\sqrt{2}} (\mathbb{1}_{(i,j)=(m,n)} + \mathbb{1}_{(i,j)=(n,m)}) J_{ij,kl} \frac{1}{\sqrt{2}} (\mathbb{1}_{(k,l)=(m',n')} + \mathbb{1}_{(k,l)=(n',m')}) \\
&= \frac{1}{2} (J_{mn, m'n'} + J_{mn, n'm'} + J_{nm, m'n'} + J_{nm, n'm'}) \\
&= \frac{1}{2} (\gamma_{mm'} \gamma_{nn'} + \gamma_{mn'} \gamma_{nm'}).
\end{aligned}$$

If  $m' < n'$  and  $(m', n') \in B^*$ ,

$$\begin{aligned}
(U^T JU)_{mm, m'n'} &= \sum_{i,j,k,l=1}^p U_{ij,mm} J_{ij,kl} U_{kl, m'n'} \\
&= \sum_{i,j,k,l=1}^p \mathbb{1}_{(i,j)=(m,m)} J_{ij,kl} \frac{1}{\sqrt{2}} (\mathbb{1}_{(k,l)=(m',n')} + \mathbb{1}_{(k,l)=(n',m')}) \\
&= \frac{1}{\sqrt{2}} (J_{mm, m'n'} + J_{mm, n'm'}) \\
&= \frac{1}{\sqrt{2}} \gamma_{mm'} \gamma_{mn'}.
\end{aligned}$$

If  $m < n$  and  $(m, n) \in B^*$ , we have

$$\begin{aligned}
(U^T JU)_{mn, m'm'} &= \sum_{i,j,k,l=1}^p U_{ij,mn} J_{ij,kl} U_{kl, m'm'} \\
&= \sum_{i,j,k,l=1}^p \frac{1}{\sqrt{2}} (\mathbb{1}_{(i,j)=(m,n)} + \mathbb{1}_{(i,j)=(n,m)}) J_{ij,kl} \mathbb{1}_{(k,l)=(m',m')} \\
&= \frac{1}{\sqrt{2}} (J_{mn, m'm'} + J_{nm, m'm'}) \\
&= \frac{1}{\sqrt{2}} \gamma_{mm'} \gamma_{nm'}.
\end{aligned}$$

Finally,

$$\begin{aligned}
(U^T JU)_{mm, m'm'} &= \sum_{i,j,k,l=1}^p U_{ij,mm} J_{ij,kl} U_{kl, m'm'} \\
&= \sum_{i,j,k,l=1}^p \mathbb{1}_{(i,j)=(m,m)} J_{ij,kl} \mathbb{1}_{(k,l)=(m',m')} \\
&= J_{mm, m'm'} \\
&= \frac{1}{2} \gamma_{mm'}^2.
\end{aligned}$$

□

**Lemma 17.** Let  $B = (B_{mn,m'n'})_{m \leq n, m' \leq n', (m,n) \in B^*, (m',n') \in B^*}$  defined by

$$B_{mn,m'n'} = \begin{cases} 2(\sigma_{mm'}\sigma_{nn'} + \sigma_{mn'}\sigma_{nm'}) & \text{if } m < n \text{ and } m' < n' \\ 2\sqrt{2}\sigma_{mm'}\sigma_{nn'} & \text{if either } m = n \text{ or } m' = n' \\ 2\sigma_{mm'}^2 & \text{if } m = n \text{ and } m' = n', \end{cases}$$

then,  $B = A^{-1}$ . Moreover  $(UBU^T)_{ij,i'j'} = \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{ji'}$  for all  $(i, j, i', j') \in B^*$  and  $(UBU^T)_{ij,i'j'} = 0$  for all  $(i, j, i', j') \notin B^*$ . Recall that we write  $(i, j, i', j') \in B^*$  if there exists  $A \in B^*$  such that  $\{i, j, i', j'\} \subset A$ .

*Proof.* We introduce the following notation: if  $l \in B_k^*$ , let  $[l]_k$  to be the index of  $l$  in  $B_k^*$ .

Step 1: Let us prove that  $B = A^{-1}$ .

We compute the product  $AB$ . Assume that  $m, n \in B_k^*$  with  $m \leq n$  and  $m', n' \in B_{k'}^*$  with  $m' \leq n'$  and  $k \neq k'$ . We then have

$$\begin{aligned} (AB)_{mn,m'n'} &= \sum_{(a,b) \in B^*, a \leq b} A_{mn,ab} B_{ab,m'n'} \\ &= \sum_{a,b \in B_k^*, a \leq b} A_{mn,ab} B_{ab,m'n'} + \sum_{a,b \in B_{k'}^*, a \leq b} A_{mn,ab} B_{ab,m'n'} \\ &= \sum_{a,b \in B_k^*, a \leq b} A_{mn,ab} 0 + \sum_{a,b \in B_{k'}^*, a \leq b} 0 B_{ab,m'n'} \\ &= 0, \end{aligned}$$

using that  $B_{ab,m'n'} = 0$  if  $a, b \in B_k^*$  and  $m', n' \in B_{k'}^*$  because  $\Sigma$  is block-diagonal, and using that  $A_{mn,a,b} = 0$  if  $m, n \in B_k^*$  and  $a, b \in B_{k'}^*$  because  $\Sigma^{-1}$  is block-diagonal. Assume that  $m, n, m', n' \in B_k^*$  with  $m \leq n$  and  $m' \leq n'$ . We have,

$$\begin{aligned} (AB)_{mn,m'n'} &= \sum_{(a,b) \in B^*, a \leq b} A_{mn,ab} B_{ab,m'n'} \\ &= (A_{B_k^*} B_{B_k^*})_{[m]_k [n]_k, [m']_k [n']_k} \\ &= \mathbb{1}_{([m]_k, [n]_k) = ([m']_k, [n']_k)} \\ &= \mathbb{1}_{(m,n) = (m',n')}, \end{aligned}$$

thanks to Lemma 15 applied to the matrix  $\Sigma_{B_k^*}$ . We proved that  $B = A^{-1}$ .

Step 2.A : We show that  $(UBU^T)_{ij,i'j'} = \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{ji'}$  for all  $(i, j, i', j') \in B^*$ .

Assume that  $(i, j, i', j') \in B^*$ . First, assume that  $i \neq j$  and  $i' \neq j'$ . Assume for example that  $i < j$  and  $i' < j'$  (the other cases are similar). We then have

$$\begin{aligned}
& (UBU^T)_{ij, i'j'} \\
&= \sum_{\substack{(m,n) \in B^*, m \leq n, \\ (m',n') \in B^*, m' \leq n'}} U_{ij, mn} B_{mn, m'n'} U_{i'j', m'n'} \\
&= \sum_{\substack{(m,n) \in B^*, m \leq n, \\ (m',n') \in B^*, m' \leq n'}} \frac{1}{\sqrt{2}} (\mathbb{1}_{(m,n)=(i,j)} + \mathbb{1}_{(m,n)=(j,i)}) B_{mn, m'n'} \frac{1}{\sqrt{2}} (\mathbb{1}_{(m',n')=(i',j')} + \mathbb{1}_{(m',n')=(j',i')}) \\
&= \frac{1}{2} B_{ij, i'j'} \\
&= \sigma_{ii'} \sigma_{jj'} + \sigma_{ij'} \sigma_{ji'}.
\end{aligned}$$

Let us take the case where  $(i, j, i', j') \in B^*$  with either  $i = j$ , or  $i' = j'$ . For example  $i = j$  and  $i' < j'$ . We then have

$$\begin{aligned}
& (U^T BU)_{ii, i'j'} \\
&= \sum_{\substack{(m,n) \in B^*, m \leq n, \\ (m',n') \in B^*, m' \leq n'}} U_{mn, ii} B_{mn, m'n'} U_{m'n', i'j'} \\
&= \sum_{\substack{(m,n) \in B^*, m \leq n, \\ (m',n') \in B^*, m' \leq n'}} \mathbb{1}_{(m,n)=(i,i)} B_{mn, m'n'} \frac{1}{\sqrt{2}} (\mathbb{1}_{(m',n')=(i',j')} + \mathbb{1}_{(m',n')=(j',i')}) \\
&= \frac{1}{\sqrt{2}} B_{ii, i'j'} \\
&= \sigma_{ii'} \sigma_{ij'} + \sigma_{ij'} \sigma_{ii'}.
\end{aligned}$$

It is the same for  $i = j$  and  $i' > j'$ , then for  $i \neq j$  and  $i' = j'$ . We also can prove the equality similarly when  $i = i'$  and  $j = j'$ .

Step 2.B: Let us prove that  $(UBU^T)_{ij, i'j'} = 0$  for all  $(i, j, i', j') \notin B^*$ .

Assume that  $(i, j, i', j') \notin B^*$ . If  $(i, j) \notin B^*$ , or if  $(i', j') \notin B^*$ , we have

$$\begin{aligned}
& (UBU^T)_{ij, i'j'} \\
&= \sum_{\substack{(m,n) \in B^*, m \leq n, \\ (m',n') \in B^*, m' \leq n'}} U_{ij, mn} B_{mn, m'n'} U_{i'j', m'n'} \\
&= \sum_{\substack{(m,n) \in B^*, m \leq n, \\ (m',n') \in B^*, m' \leq n'}} \frac{1}{\sqrt{2}} (\mathbb{1}_{(m,n)=(i,j)} + \mathbb{1}_{(m,n)=(j,i)}) B_{mn, m'n'} \frac{1}{\sqrt{2}} (\mathbb{1}_{(m',n')=(i',j')} + \mathbb{1}_{(m',n')=(j',i')}) \\
&= 0,
\end{aligned}$$

because if  $(i, j) \notin B^*$ , the term  $(\mathbb{1}_{(m,n)=(i,j)} + \mathbb{1}_{(m,n)=(j,i)})$  is equal to 0. Similarly, if  $(i', j') \notin B^*$ , the term  $(\mathbb{1}_{(m',n')=(i',j')} + \mathbb{1}_{(m',n')=(j',i')})$  is equal to 0.

It remains the case where  $i, j \in B_k^*$  and  $i', j' \in B_{k'}^*$  with  $k \neq k'$ . Then,  $(U^T BU)_{ij, i'j'} = \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{ji'} = 0$ .  $\square$

To conclude the proof, we remark that, if  $(i, j, i', j') \in B^*$ , then

$$\text{cov}((S_{B^*})_{ij}, (S_{B^*})_{i'j'}) = \text{cov}(X_i X_j, X_{i'} X_{j'}) = \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{ji'}.$$

Now, assume that  $(i, j, i', j') \notin B^*$ . If  $(i, j) \notin B^*$  or if  $(i', j') \notin B^*$ , then  $\text{cov}((S_{B^*})_{ij}, (S_{B^*})_{i'j'}) = 0$  because one of the two terms is zero. Assume that  $i, j \in B_k^*$  and  $i', j' \in B_{k'}^*$  with  $k \neq k'$ . Then

$$\text{cov}((S_{B^*})_{ij}, (S_{B^*})_{i'j'}) = \text{cov}(X_i X_j, X_{i'} X_{j'}) = \sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{ji'} = 0.$$

Thus, the covariance matrix of  $\text{vec}(S_{B^*})$  is equal to the Cramér-Rao bound.  $\square$

### Proof of Proposition 16

*Proof.* Using the central limit Theorem and Proposion 15, we have

$$\sqrt{n-1}(\text{vec}(\widehat{\Sigma}_{B^*}) - \text{vec}(\Sigma)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, CR).$$

Then, by Proposition 13, we have

$$\sqrt{n-1}(\text{vec}(\widehat{\Sigma}_{\widehat{B}}) - \text{vec}(\Sigma)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, CR),$$

and by Slutsky,

$$\sqrt{n}(\text{vec}(S_{\widehat{B}}) - \text{vec}(\Sigma)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, CR).$$

$\square$

### Proof of Proposition 17

**Lemma 18.** *Under Conditions 1 to 4, for all  $\gamma > 1/2$*

$$\|S_{B^*} - \Sigma\|_2 = o_p\left(\frac{\log(n)^\gamma}{\sqrt{n}}\right),$$

where  $\|\cdot\|_2$  is the operator norm, and it is equal to  $\lambda_{\max}(\cdot)$  on the set of the symmetric positive semi-definite matrices.

*Proof.*

$$\begin{aligned}
& \mathbb{P} \left( \|S_{B^*} - \Sigma\|_2 > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) \\
&= \mathbb{P} \left( \exists k \in [1 : K], \|S_{B_k^*} - \Sigma_{B_k^*}\|_2 > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) \\
&\leq K \max_{k \in [1:K]} \mathbb{P} \left( \|S_{B_k^*} - \Sigma_{B_k^*}\|_2 > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) \\
&\leq K \max_{k \in [1:K]} \mathbb{P} \left( m \max_{i,j \in B_k^*} |s_{ij} - \sigma_{ij}| > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) \\
&\leq K \max_{k \in [1:K]} \mathbb{P} \left( \max_{i,j \in B_k^*} |s_{ij} - \hat{\sigma}_{ij}| + |\hat{\sigma}_{ij} - \sigma_{ij}| > \frac{\varepsilon \log(n)^\gamma}{m\sqrt{n}} \right) \\
&\leq Km^2 \max_{k \in [1:K]} \max_{i,j \in B_k^*} \mathbb{P} \left( |s_{ij} - \hat{\sigma}_{ij}| + |\hat{\sigma}_{ij} - \sigma_{ij}| > \frac{\varepsilon \log(n)^\gamma}{m\sqrt{n}} \right).
\end{aligned}$$

Now, on the one hand,

$$\begin{aligned}
& Km^2 \max_{k \in [1:K]} \max_{i,j \in B_k^*} \mathbb{P} \left( |\hat{\sigma}_{ij} - \sigma_{ij}| > \frac{\varepsilon \log(n)^\gamma}{2m\sqrt{n}} \right) \\
&\leq 2Km^2 \exp \left( -C_{\inf} n \frac{\varepsilon^2 \log(n)^{2\gamma}}{4m^2 n} \right) \rightarrow 0,
\end{aligned}$$

by Bernstein's inequality. On the other hand,

$$\begin{aligned}
& Km^2 \max_{k \in [1:K]} \max_{i,j \in B_k^*} \mathbb{P} \left( |s_{ij} - \hat{\sigma}_{ij}| > \frac{\varepsilon \log(n)^\gamma}{2m\sqrt{n}} \right) \\
&= Km^2 \max_{k \in [1:K]} \max_{i,j \in B_k^*} \mathbb{P} \left( |\hat{\sigma}_{ij}| > n \frac{\varepsilon \log(n)^\gamma}{2m\sqrt{n}} \right) \rightarrow 0,
\end{aligned}$$

by Bernstein's inequality.  $\square$

**Lemma 19.** *Under Conditions 1 to 5, for all  $\gamma > 1/2$*

$$\left| \frac{p}{\beta^T S_{B^*} \beta} - \frac{p}{\beta^T \Sigma \beta} \right| = o_p \left( \frac{\log(n)^\gamma}{\sqrt{n}} \right).$$

*Proof.* Remark that

$$\left| \frac{\beta^T (S_{B^*} - \Sigma) \beta}{p} \right| \leq \frac{\|\beta\|_2^2}{p} \|S_{B^*} - \Sigma\|_2 = o_p \left( \frac{\log(n)^\gamma}{\sqrt{n}} \right).$$

With probability which goes to one 1, by Lemma 1, we have  $\beta^T S_{B^*} \beta / p \geq \lambda_{\inf} \beta_{\inf}^2 / 2$ . Moreover,  $\beta^T \Sigma \beta / p \geq \lambda_{\inf} \beta_{\inf}^2 \geq \lambda_{\inf} \beta_{\inf}^2 / 2$ . Thus, with probability which goes to one 1, we have

$$\left| \frac{p}{\beta^T S_{B^*} \beta} - \frac{p}{\beta^T \Sigma \beta} \right| \leq \frac{4}{\lambda_{\inf}^2 \beta_{\inf}^4} \left| \frac{\beta^T S_{B^*} \beta}{p} - \frac{\beta^T \Sigma \beta}{p} \right| = o_p \left( \frac{\log(n)^\gamma}{\sqrt{n}} \right).$$

$\square$

We can now prove Proposition 17

*Proof.* Let  $\tilde{\eta}_i$  be the estimator of  $\eta_i$  given by the estimator  $S_{B^*}$  of  $\Sigma$ . For all  $\varepsilon > 0$  and  $\gamma > 1/2$ , we have

$$\begin{aligned} & \mathbb{P} \left( \sum_{i=1}^p |\hat{\eta}_i - \eta_i| > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) \\ & \leq \mathbb{P} \left( \sum_{i=1}^p |\tilde{\eta}_i - \eta_i| > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) + \mathbb{P}(\hat{B} \neq B^*) \\ & \leq \mathbb{P} \left( p \max_{i \in [1:p]} |\tilde{\eta}_i - \eta_i| > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) + \mathbb{P}(\hat{B} \neq B^*). \end{aligned}$$

The term  $\mathbb{P}(\hat{B} = B^*)$  goes to 0 from Proposition 2. It remains to prove that

$$\mathbb{P} \left( p \max_{i \in [1:p]} |\tilde{\eta}_i - \eta_i| > \frac{\varepsilon \log(n)^\gamma}{\sqrt{n}} \right) \rightarrow 0. \quad (16)$$

For all  $k \in [1 : K]$  and all  $u \subset B_k^*$ , let us write

$$\begin{aligned} V_u^k & := \beta_{B_k^*-u}^T (\Sigma_{B_k^*-u, B_k^*-u} - \Sigma_{B_k^*-u, u} \Sigma_{u, u}^{-1} \Sigma_{u, B_k^*-u}) \beta_{B_k^*-u} \\ \tilde{V}_u^k & := \beta_{B_k^*-u}^T (S_{B_k^*-u, B_k^*-u} - S_{B_k^*-u, u} S_{u, u}^{-1} S_{u, B_k^*-u}) \beta_{B_k^*-u} \\ V & := \beta^T \Sigma \beta \\ \tilde{V} & := \beta^T S_{B^*} \beta \\ \alpha_u & := \frac{V_u^k}{V} \\ \tilde{\alpha}_u & := \frac{\tilde{V}_u^k}{V}. \end{aligned}$$

Let, for all  $C \subset [1 : p]$ ,  $C \neq \emptyset$ ,

$$L((a_u)_{u \in C}; C) := \left( \frac{1}{|C|} \sum_{u \subset C-i} \binom{|C|-1}{|u|}^{-1} [a_{u+i} - a_u] \right)_{i \in C}.$$

We then have  $(\eta_i)_{i \in B_k^*} = L((\alpha_u)_{u \in B_k^*}; B_k^*)$  et  $(\tilde{\eta}_i)_{i \in B_k^*} = L((\tilde{\alpha}_u)_{u \in B_k^*}; B_k^*)$ .

As  $L(\cdot; C)$  is linear, it is Lipschitz continue from  $(R^{2^{|C|}}, \|\cdot\|_\infty)$  to  $(R^{|C|}, \|\cdot\|_\infty)$ , with constant  $l_{|C|}$  (we can show that  $l_{|C|} = 2$ ). Let  $l := \max_{j \in [1:m]} l_j < +\infty$  (we have in fact  $l = 2$ ). We then have,

$$p \max_{i \in [1:p]} |\tilde{\eta}_i - \eta_i| \leq p l \max_{k \in [1:K]} \max_{u \subset B_k^*} |\tilde{\alpha}_u - \alpha_u|.$$

It suffices to show that

$$p \max_{k \in [1:K]} \max_{u \subset B_k^*} |\tilde{\alpha}_u - \alpha_u| = o_p \left( \frac{\log(n)^\gamma}{\sqrt{n}} \right).$$



Now,

$$\begin{aligned} p |\tilde{\alpha}_u - \alpha_u| &\leq \left| \frac{p\tilde{V}_u^k}{\tilde{V}} - \frac{pV_u^k}{\tilde{V}} \right| + \left| \frac{pV_u^k}{\tilde{V}} - \frac{pV_u^k}{V} \right| \\ &\leq \frac{p}{\tilde{V}} \left| \tilde{V}_u^k - V_u^k \right| + V_u^k \left| \frac{p}{\tilde{V}} - \frac{p}{V} \right|. \end{aligned}$$

The term  $\max_{k \in [1:K]} \max_{u \subset B_k^*} V_u^k$  is bounded from Conditions 2 and 5 and  $\left| \frac{p}{\tilde{V}} - \frac{p}{V} \right| = o_p(\log(n)^\gamma / \sqrt{n})$  thanks to Lemma 19. The term  $\frac{p}{\tilde{V}}$  is bounded in probability using Lemma 19, Conditions 2 and 5. Thus, it suffices to show that  $\max_{k \in [1:K]} \max_{u \subset B_k^*} \left| \tilde{V}_u^k - V_u^k \right| = o_p(\log(n)^\gamma / \sqrt{n})$ . We will use that the operator norm of a sub-matrix is smaller than the operator norm of the whole matrix.

For all  $k \in [1:K]$  and  $u \subset B_k^*$ , we have

$$\begin{aligned} \left| \tilde{V}_u^k - V_u^k \right| &\leq m\beta_{\text{sup}}^2 \|S_{B_k^* - u} - \Sigma_{B_k^* - u}\|_2 \\ &\quad + m\beta_{\text{sup}}^2 \|S_{B_k^* - u, u} S_{u, u}^{-1} S_{u, B_k^* - u} - \Sigma_{B_k^* - u, u} \Sigma_{u, u}^{-1} \Sigma_{u, B_k^* - u}\|_2. \end{aligned}$$

Now,  $\|S_{B_k^* - u} - \Sigma_{B_k^* - u}\|_2 \leq \|S_{B_k^*} - \Sigma_{B_k^*}\|_2$  so is  $o_p(\log(n)^\gamma / \sqrt{n})$  from Lemma 18. Moreover

$$\begin{aligned} &\|S_{B_k^* - u, u} S_{u, u}^{-1} S_{u, B_k^* - u} - \Sigma_{B_k^* - u, u} \Sigma_{u, u}^{-1} \Sigma_{u, B_k^* - u}\|_2 \\ &\leq \|S_{B_k^* - u, u} S_{u, u}^{-1} (S_{u, B_k^* - u} - \Sigma_{u, B_k^* - u})\|_2 \\ &\quad + \|S_{B_k^* - u, u} (S_{u, u}^{-1} - \Sigma_{u, u}^{-1}) \Sigma_{u, B_k^* - u}\|_2 \\ &\quad + \|(S_{B_k^* - u, u} - \Sigma_{B_k^* - u, u}) \Sigma_{u, u}^{-1} \Sigma_{u, B_k^* - u}\|_2 \\ &\leq \|S_{B_k^*}\|_2 \|S_{B_k^*}^{-1}\|_2 \|S_{B_k^*} - \Sigma_{B_k^*}\|_2 + \|S_{B_k^*}\|_2 \|S_{B_k^*}^{-1} - \Sigma_{B_k^*}^{-1}\|_2 \|\Sigma_{B_k^*}\|_2 + \|S_{B_k^*} - \Sigma_{B_k^*}\|_2 \|\Sigma_{B_k^*}^{-1}\|_2 \|\Sigma_{B_k^*}\|_2 \\ &\leq \|S\|_2 \|S^{-1}\|_2 \|S - \Sigma\|_2 + \|S\|_2 \|S^{-1} - \Sigma^{-1}\|_2 \|\Sigma\|_2 + \|S - \Sigma\|_2 \|\Sigma^{-1}\|_2 \|\Sigma\|_2, \end{aligned}$$

which do not depend on  $k$  and  $u$ . Finally, remark that  $\|\Sigma\|_2$  and  $\|\Sigma^{-1}\|_2$  are bounded from Condition 2, that  $\|S\|_2$  and  $\|S^{-1}\|_2$  are bounded in probability from Lemma 1, that  $\|S - \Sigma\|_2 = o_p(\log(n)^\gamma / \sqrt{n})$  from Lemma 18 and Proposition 2 and that

$$\|S^{-1} - \Sigma^{-1}\|_2 \leq \|\Sigma^{-1}\|_2 \|S^{-1}\|_2 \|S - \Sigma\|_2 = o_p(\log(n)^\gamma / \sqrt{n}).$$

Thus, we proved that

$$p \max_{k \in [1:K]} \max_{u \subset B_k^*} |\tilde{\alpha}_u - \alpha_u| = o_p(\log(n)^\gamma / \sqrt{n}).$$

□

### Proof of Proposition 18

**Lemma 20.** *Under Conditions 1, 2 and 3, for all penalization coefficient  $\delta \in ]0, 1[$  and for all  $\varepsilon > 0$ ,*

$$\max_{B(\alpha_1) \leq B \leq B^*} \|S_B - \Sigma\|_2 = o_p \left( \frac{1}{n^{(\delta-\varepsilon)/2}} \right),$$

where  $\|\cdot\|_2$  is the operator norm, and it is equal to  $\lambda_{max}(\cdot)$  on the set of the symmetric positive semi-definite matrices.

*Proof.* Let  $\alpha_1 := \delta/2 - \varepsilon/4$ .

$$\begin{aligned} & \mathbb{P} \left( \max_{B(\alpha_1) \leq B \leq B^*} \|S_B - \Sigma\|_2 > \frac{\epsilon}{n^{(\delta-\varepsilon)/2}} \right) \\ &= \mathbb{P} \left( \exists k \in [1 : K], \max_{B(\alpha_1) \leq B \leq B^*} \|(S_B)_{B_k^*} - \Sigma_{B_k^*}\|_2 > \frac{\epsilon}{n^{(\delta-\varepsilon)/2}} \right) \\ &\leq K \max_{k \in [1:K]} \mathbb{P} \left( \max_{B(\alpha_1) \leq B \leq B^*} \|(S_B)_{B_k^*} - \Sigma_{B_k^*}\|_2 > \frac{\epsilon}{n^{(\delta-\varepsilon)/2}} \right) \\ &\leq K \max_{k \in [1:K]} \mathbb{P} \left( m \max_{B(\alpha_1) \leq B \leq B^*} \max_{i,j \in B_k^*} |(S_B)_{i,j} - \sigma_{ij}| > \frac{\epsilon}{n^{(\delta-\varepsilon)/2}} \right) \\ &\leq K \max_{k \in [1:K]} \mathbb{P} \left( m \max_{i,j \in B_k^*} |s_{ij} - \sigma_{ij}| + mn^{-\alpha_1} > \frac{\epsilon}{n^{(\delta-\varepsilon)/2}} \right) \\ &\leq K \max_{k \in [1:K]} \mathbb{P} \left( m \max_{i,j \in B_k^*} |s_{ij} - \sigma_{ij}| > C_{\inf}(\varepsilon, \delta, \epsilon) n^{-(\delta-\varepsilon)/2} \right), \end{aligned}$$

that goes to 0 following the proof of 18.  $\square$

**Lemma 21.** *Under Conditions 1, 2, 3 and 5, for all penalization coefficient  $\delta \in ]0, 1[$  and for all  $\varepsilon > 0$ ,*

$$\left| \frac{p}{\beta^T S_{B^*} \beta} - \frac{p}{\beta^T \Sigma \beta} \right| = o_p \left( \frac{1}{n^{(\delta-\varepsilon)/2}} \right).$$

*Proof.* The proof is similar to the proof of Lemme 19.  $\square$

We now can prove Proposition 18.

*Proof.* For all  $B \in \mathcal{P}_p$ , we define  $\tilde{\eta}(B)_i$  as the Shapley effect given by the estimation  $S_B$  of  $\Sigma$ , and we write  $\hat{\eta}_i$  for the Shapley effect given by the estimation  $S_{B_{n-\delta/2}}$ .

$$\begin{aligned} & \mathbb{P} \left( \sum_{i=1}^p |\hat{\eta}_i - \eta_i| > \frac{\epsilon}{n^{-(\delta-\varepsilon)/2}} \right) \\ &\leq \mathbb{P} \left( \max_{B(\alpha_1) \leq B \leq B^*} \sum_{i=1}^p |\tilde{\eta}(B)_i - \eta_i| > \frac{\epsilon}{n^{-(\delta-\varepsilon)/2}} \right) + \mathbb{P}(\{B(\alpha_1) \leq B_{n-\delta/2} \leq B^*\}^c) \\ &\leq \mathbb{P} \left( p \max_{B(\alpha_1) \leq B \leq B^*} \max_{i \in [1:p]} |\tilde{\eta}(B)_i - \eta_i| > \frac{\epsilon}{n^{-(\delta-\varepsilon)/2}} \right) + \mathbb{P}(\{B(\alpha_1) \leq B_{n-\delta/2} \leq B^*\}^c). \end{aligned}$$

By Proposition 8,  $\mathbb{P}(\{B(\alpha_1) \leq B_{n^{-\delta/2}} \leq B^*\}^c) \rightarrow 0$ .

Finally, we prove that

$$\mathbb{P}\left(p \max_{B(\alpha_1) \leq B \leq B^*} \max_{i \in [1:p]} |\check{\eta}(B)_i - \eta_i| > \frac{\epsilon}{n^{-(\delta-\epsilon)/2}}\right) \rightarrow 0,$$

following the proof of Proposition 17. □

### Proof of Proposition 19

*Proof.* Remark that  $\Sigma$  verifies Conditions 1 to 3. Let  $a > 0$ . Let  $\check{\Sigma} := \Sigma$  if  $\forall B < B^*$ ,  $\|\Sigma_B - \Sigma\|_{\max} \geq an^{-1/4}$  and  $\check{\Sigma} = I_p$  otherwise. Let  $\check{\eta}$  and  $\check{\hat{\eta}}$  be defined as  $\eta$  and  $\hat{\eta}$  in Proposition 17 but replacing  $\Sigma$  by  $\check{\Sigma}$ . As  $\check{\Sigma}$  and  $\beta$  verify the Conditions 1 to 4, conditionally to  $\check{\Sigma}$

$$\sum_{i=1}^p |\check{\eta}_i - \check{\hat{\eta}}_i| = o_p\left(\frac{(\log(n))^\gamma}{\sqrt{n}}\right).$$

Thus, for all  $\epsilon > 0$ ,

$$\mathbb{P}\left(\sum_{i=1}^p |\check{\eta}_i - \check{\hat{\eta}}_i| > \frac{\epsilon \log(n)^\gamma}{\sqrt{n}} \middle| \check{\Sigma}\right) \rightarrow 0,$$

so, by dominated convergence theorem,

$$\sum_{i=1}^p |\check{\eta}_i - \check{\hat{\eta}}_i| = o_p\left(\frac{(\log(n))^\gamma}{\sqrt{n}}\right),$$

unconditionally to  $\check{\Sigma}$ .

We conclude saying that  $\check{\Sigma} = \Sigma$  with probability which converges to 1 from Proposition 12, so  $\check{\hat{\eta}} = \hat{\eta}$  and  $\check{\eta} = \eta$  with probability which converges to 1. □