Variable selection in multivariate linear models with high-dimensional covariance matrix estimation

Marie Perrot-Dockès, Celine Levy Leduc, Laure Sansonnet, Julien Chiquet

VARIABLE SELECTION IN MULTIVARIATE LINEAR MODELS WITH HIGH-DIMENSIONAL COVARIANCE MATRIX ESTIMATION

M. PERROT-DOCKÈS, C. LÉVY-LEDUC, L. SANSONNET, AND J. CHIQUET

ABSTRACT. In this paper, we propose a novel variable selection approach in the framework of multivariate linear models taking into account the dependence that may exist between the responses. It consists in estimating beforehand the covariance matrix Σ of the responses and to plug this estimator in a Lasso criterion, in order to obtain a sparse estimator of the coefficient matrix. The properties of our approach are investigated both from a theoretical and a numerical point of view. More precisely, we give general conditions that the estimators of the covariance matrix and its inverse have to satisfy in order to recover the positions of the null and non null entries of the coefficient matrix when the size of Σ is not fixed and can tend to infinity. We prove that these conditions are satisfied in the particular case of some Toeplitz matrices. Our approach is implemented in the R package MultiVarSel available from the Comprehensive R Archive Network (CRAN) and is very attractive since it benefits from a low computational load. We also assess the performance of our methodology using synthetic data and compare it with alternative approaches. Our numerical experiments show that including the estimation of the covariance matrix in the Lasso criterion dramatically improves the variable selection performance in many cases.

1. INTRODUCTION

The multivariate linear model consists in generalizing the classical linear model, in which a single response is explained by p variables, to the case where the number q of responses is larger than 1. Such a general modeling can be used in a wide variety of applications ranging from econometrics (Lütkepohl (2005)) to bioinformatics (Meng et al. (2014)). In the latter field, for instance, multivariate models have been used to gain insight into complex biological mechanisms like metabolism or gene regulation. This has been made possible thanks to recently developed sequencing technologies. For further details, we refer the reader to Mehmood et al. (2012). However, the downside of such a technological expansion is to include irrelevant variables in the statistical models. To circumvent this, devising efficient variable selection approaches in the multivariate setting has become a growing concern.

A first naive approach to deal with the variable selection issue in the multivariate setting consists in applying classical univariate variable selection strategies to each response separately. Some well-known variable selection methods include the least absolute shrinkage and selection operator (LASSO) proposed by Tibshirani (1996) and the smoothly clipped absolute deviation (SCAD) approach devised by Fan and Li (2001). However, such a strategy does not take into account the dependence that may exist between the different responses.

In this paper, we shall consider the following multivariate linear model:

$$Y = XB + E, \tag{1}$$

Date: July 14, 2017.

where $Y = (Y_{i,j})_{1\leq i\leq n, 1\leq j\leq q}$ denotes the $n \times q$ random response matrix, X denotes the $n \times p$ design matrix, B denotes a $p \times q$ coefficient matrix and $E = (E_{i,j})_{1\leq i\leq n, 1\leq j\leq q}$ denotes the $n \times q$ random error matrix, where n is the sample size. In order to model the potential dependence that may exist between the columns of E, we shall assume that for each i in $\{1, \ldots, n\}$,

$$(E_{i,1}, \ldots, E_{i,q}) \sim \mathcal{N}(0, \Sigma), \tag{2}$$

where Σ denotes the covariance matrix of the ith row of the error matrix E. We shall moreover assume that the different rows of E are independent. With such assumptions, there is some dependence between the columns of E but not between the rows. Our goal is here to design a variable selection approach which is able to identify the positions of the null and non null entries in the sparse matrix B by taking into account the dependence between the columns of E.

This issue has recently been considered by Lee and Liu (2012) who extended the approach of Rothman et al. (2010). More precisely, Lee and Liu (2012) proposed three approaches for dealing with this issue based on penalized maximum likelihood with a weighted ℓ_1 regularization. In their first approach B is estimated by using a plug-in estimator of Σ^{-1}, in the second one, Σ^{-1} is estimated by using a plug-in estimator of B and in the third one, Σ^{-1} and B are estimated simultaneously. Lee and Liu (2012) also investigate the asymptotic properties of their methods when the sample size n tends to infinity and the number of rows and columns q of Σ is fixed.

In this paper, we propose to estimate Σ beforehand and to plug this estimator in a Lasso criterion, in order to obtain a sparse estimator of B. Hence, our methodology is close to the first approach of Lee and Liu (2012). However, there are two main differences: The first one is the asymptotic framework in which our theoretical results are established and the second one is the strategy that we use for estimating Σ. More precisely, in our asymptotic framework, q is allowed to depend on n and thus to tend to infinity as n tends to infinity at a polynomial rate. Moreover, in Lee and Liu (2012), Σ^{-1} is estimated by using an adaptation of the Graphical Lasso (GLASSO) proposed by Friedman et al. (2008). This technique has also been considered by Yuan and Lin (2007), Banerjee et al. (2008) and Rothman et al. (2008). In this paper, we give general conditions that the estimators of Σ and Σ^{-1} have to satisfy in order to be able to recover the support of B that is to find the positions of the null and non null entries of the matrix B. We prove that when Σ is a particular Toeplitz matrix, namely the covariance matrix of an AR(1) process, the assumptions of the theorem are satisfied.

Let us now describe more precisely our methodology. We start by "whitening" the observations Y by applying the following transformation to Model (1):

$$Y\,\Sigma^{-1/2} = XB\,\Sigma^{-1/2} + E\,\Sigma^{-1/2}. \tag{3}$$

The goal of such a transformation is to remove the dependence between the columns of Y. Then, for estimating B, we proceed as follows. Let us observe that (3) can be rewritten as:

$$\mathcal{Y} = \mathcal{X}\mathcal{B} + \mathcal{E}, \tag{4}$$

with

$$\mathcal{Y} = vec(Y\,\Sigma^{-1/2}),\ \mathcal{X} = (\Sigma^{-1/2})' \otimes X,\ \mathcal{B} = vec(B) \text{ and } \mathcal{E} = vec(E\,\Sigma^{-1/2}), \tag{5}$$

where vec denotes the vectorization operator and $\otimes$ the Kronecker product.

With Model (4), estimating B is equivalent to estimate $\mathcal{B}$ since $\mathcal{B} = vec(B)$. Then, for estimating $\mathcal{B}$, we use the classical LASSO criterion defined as follows for a nonnegative λ:

$$\widehat{\mathcal{B}}(\lambda) = \text{Argmin}_{\mathcal{B}} \left\{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda\|\mathcal{B}\|_1 \right\}, \tag{6}$$

where $\|.\|_1$ and $\|.\|_2$ denote the classical ℓ_1-norm and ℓ_2-norm, respectively. Inspired by Zhao and Yu (2006), Theorem 1 established some conditions under which the positions of the null and non null entries of $\mathcal{B}$ can be recovered by using $\widehat{\mathcal{B}}$.

In practical situations, the covariance matrix Σ is generally unknown and has thus to be estimated. Let $\widehat{\Sigma}$ denote an estimator of Σ. Then, the estimator $\widehat{\Sigma}^{-1/2}$ of $\Sigma^{-1/2}$ is such that

$$\widehat{\Sigma}^{-1} = \widehat{\Sigma}^{-1/2}(\widehat{\Sigma}^{-1/2})'.$$

When $\Sigma^{-1/2}$ is replaced by $\widehat{\Sigma}^{-1/2}$, (3) becomes

$$Y\,\widehat{\Sigma}^{-1/2} = XB\,\widehat{\Sigma}^{-1/2} + E\,\widehat{\Sigma}^{-1/2}, \tag{7}$$

which can be rewritten as follows:

$$\widetilde{\mathcal{Y}} = \widetilde{\mathcal{X}}\mathcal{B} + \widetilde{\mathcal{E}}, \tag{8}$$

where

$$\widetilde{\mathcal{Y}} = vec(Y\,\widehat{\Sigma}^{-1/2}),\ \widetilde{\mathcal{X}} = (\widehat{\Sigma}^{-1/2})' \otimes X,\ \mathcal{B} = vec(B) \text{ and } \widetilde{\mathcal{E}} = vec(E\,\widehat{\Sigma}^{-1/2}). \tag{9}$$

In Model (8), $\mathcal{B}$ is estimated by

$$\widetilde{\mathcal{B}}(\lambda) = \text{Argmin}_{\mathcal{B}} \left\{ \|\widetilde{\mathcal{Y}} - \widetilde{\mathcal{X}}\mathcal{B}\|_2^2 + \lambda\|\mathcal{B}\|_1 \right\}. \tag{10}$$

By extending Theorem 1, Theorem 5 gives some conditions on the eigenvalues of Σ^{-1} and on the convergence rate of $\widehat{\Sigma}$ and its inverse to Σ and Σ^{-1}, respectively, under which the positions of the null and non null entries of $\mathcal{B}$ can be recovered by using $\widetilde{\mathcal{B}}$.

We prove in Section 2.3 that when Σ is a particular Toeplitz matrix, namely the covariance matrix of an AR(1) process, the assumptions of Theorem 5 are satisfied. This strategy has been implemented in the R package MultiVarSel, which is available on the Comprehensive R Archive Network (CRAN), for more general Toeplitz matrices Σ such as the covariance matrix of ARMA processes or general stationary processes. For a successful application of this methodology to particular "-omic" data, namely metabolomic data, we refer the reader to Perrot-Dockès et al. (2017). For a review of the most recent methods for estimating high-dimensional covariance matrices, we refer the reader to Pourahmadi (2013).

The paper is organized as follows. Section 2 is devoted to the theoretical results of the paper. The assumptions under which the positions of the non null and null entries of $\mathcal{B}$ can be recovered are established in Theorem 1 when Σ is known and in Theorem 5 when Σ is unknown. Section 2.3 studies the specific case of the AR(1) model. We present in Section 3 some numerical experiments in order to support our theoretical results. The proofs of our main theoretical results are given in Section 4.

2. Theoretical results

2.1. **Case where Σ is known.** Let us first introduce some notations. Let

$$C = \frac{1}{nq}\mathcal{X}'\mathcal{X} \text{ and } J = \{1 \le j \le pq, \mathcal{B}_j \neq 0\}, \tag{11}$$

where $\mathcal{X}$ is defined in (5) and where $\mathcal{B}_j$ denotes the jth component of the vector $\mathcal{B}$ defined in (5).

Let also define

$$C_{J,J} = \frac{1}{nq}(\mathcal{X}_{\bullet,J})'\mathcal{X}_{\bullet,J} \text{ and } C_{J^c,J} = \frac{1}{nq}(\mathcal{X}_{\bullet,J^c})'\mathcal{X}_{\bullet,J}, \tag{12}$$

where $\mathcal{X}_{\bullet,J}$ and $\mathcal{X}_{\bullet,J^c}$ denote the columns of $\mathcal{X}$ belonging to the set J defined in (11) and to its complement J^c, respectively.

More generally, for any matrix A, $A_{I,J}$ denotes the partitioned matrix extracted from A by considering the rows of A belonging to the set I and the columns of A belonging to the set J, with $\bullet$ indicating all the rows or all the columns.

The following theorem gives some conditions under which the estimator $\widehat{\mathcal{B}}$ defined in (6) is sign-consistent as defined by Zhao and Yu (2006), namely,

$$\mathbb{P}\left(\text{sign}(\widehat{\mathcal{B}}) = \text{sign}(\mathcal{B})\right) \to 1, \text{ as } n \to \infty,$$

where the sign function maps positive entries to 1, negative entries to -1 and zero to 0.

Theorem 1. *Assume that $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_{nq})'$ satisfies Model (4). Assume also that there exist some positive constants M_1, M_2, M_3 and positive numbers c_1, c_2 such that $0 < c_1 + c_2 < 1/2$ satisfying:*

- (A1) *for all $n \geq 1$, for all $j \in \{1, \dots, pq\}$, $\frac{1}{n}(\mathcal{X}_{\bullet,j})'\mathcal{X}_{\bullet,j} \leq M_1$, where $\mathcal{X}_{\bullet,j}$ is the jth column of $\mathcal{X}$ defined in (5),*
- (A2) *for all $n \geq 1$, $\frac{1}{n}\lambda_{\min}\left((\mathcal{X}'\mathcal{X})_{J,J}\right) \geq M_2$, where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A,*
- (A3) *$|J| = O(q^{c_1})$, where J is defined in (11) and $|J|$ is the cardinality of the set J,*
- (A4) *$q^{c_2} \min_{j \in J} |\mathcal{B}_j| \geq M_3$.*

Assume also that the following strong Irrepresentable Condition holds:

(IC) *There exists a positive constant vector η such that*

$$\left|(\mathcal{X}'\mathcal{X})_{J^c,J}((\mathcal{X}'\mathcal{X})_{J,J})^{-1} \text{sign}(\mathcal{B}_{\mathrm{J}})\right| \leq \mathbf{1} - \eta,$$

where $\mathbf{1}$ is a $(pq - |J|)$ vector of 1 and the inequality holds element-wise.

Then, for all λ that satisfies

$$\text{(L)} \qquad q = q_n = o\left(n^{\frac{1}{2(c_1+c_2)}}\right), \quad \frac{\lambda}{\sqrt{n}} \to \infty \text{ and } \frac{\lambda}{n} = o\left(q^{-(c_1+c_2)}\right), \text{ as } n \to \infty,$$

we have

$$\mathbb{P}\left(\text{sign}(\widehat{\mathcal{B}}(\lambda)) = \text{sign}(\mathcal{B})\right) \to 1, \text{ as } n \to \infty,$$

where $\widehat{\mathcal{B}}(\lambda)$ is defined by (6).

Remark 1. Observe that if $c_1 + c_2 < (2k)^{-1}$, for some positive k, then the first condition of (L) becomes $q = o(n^k)$. Hence for large values of k, the size q of Σ is much larger than n.

The proof of Theorem 1 is given in Section 4. It is based on Proposition 2 which is an adaptation to the multivariate case of Proposition 1 in Zhao and Yu (2006).

Proposition 2. *Let $\widehat{\mathcal{B}}(\lambda)$ be defined by (6). Then*

$$\mathbb{P}\left(\text{sign}(\widehat{\mathcal{B}}(\lambda)) = \text{sign}(\mathcal{B})\right) \geq \mathbb{P}(A_n \cap B_n),$$

where

$$A_n = \left\{ \left|(C_{J,J})^{-1}W_J\right| < \sqrt{nq}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}|(C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_\mathrm{J})| \right) \right\} \tag{13}$$

and

$$B_n = \left\{ \left|C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c}\right| \leq \frac{\lambda}{2\sqrt{nq}}\left(\mathbf{1} - \left|C_{J^c,J}(C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_\mathrm{J})\right|\right) \right\}, \tag{14}$$

with $W = \mathcal{X}'\mathcal{E}/\sqrt{nq}$. *In (13) and (14),* $C_{J,J}$ *and* $C_{J^c,J}$ *are defined in (12) and* W_J *and* W_{J^c} *denote the components of* W *being in* J *and* J^c*, respectively. Note that the previous inequalities hold element-wise.*

The proof of Proposition 2 is given in Section 4.

We give in the following proposition which is proved in Section 4 some conditions on X and Σ under which Assumptions (A1) and (A2) of Theorem 1 hold.

Proposition 3. *If there exist some positive constants* M_1', M_2', m_1, m_2 *such that, for all* $n \geq 1$,

- (C1) *for all* $j \in \{1, \ldots, p\}$, $\frac{1}{n}(X_{\bullet,j})'X_{\bullet,j} \leq M_1'$,
- (C2) $\frac{1}{n}\lambda_{\min}(X'X) \geq M_2'$,
- (C3) $\lambda_{\max}(\Sigma^{-1}) \leq m_1$,
- (C4) $\lambda_{\min}(\Sigma^{-1}) \geq m_2$,

then Assumptions (A1) *and* (A2) *of Theorem 1 are satisfied.*

Remark 2. Observe that (C1) and (C2) hold in the case where the columns of the matrix X are orthogonal.

We give in Proposition 6 in Section 2.3 some conditions under which Condition (IC) holds in the specific case where Σ is the covariance matrix of an AR(1) process.

2.2. Case where Σ is unknown.

Similarly as in (11) and (12), we introduce the following notations:

$$\widetilde{C} = \frac{1}{nq}\widetilde{\mathcal{X}}'\widetilde{\mathcal{X}} \tag{15}$$

and

$$\widetilde{C}_{J,J} = \frac{1}{nq}(\widetilde{\mathcal{X}}_{\bullet,J})'\widetilde{\mathcal{X}}_{\bullet,J} \text{ and } \widetilde{C}_{J^c,J} = \frac{1}{nq}(\widetilde{\mathcal{X}}_{\bullet,J^c})'\widetilde{\mathcal{X}}_{\bullet,J}, \tag{16}$$

where $\widetilde{\mathcal{X}}_{\bullet,J}$ and $\widetilde{\mathcal{X}}_{\bullet,J^c}$ denote the columns of $\widetilde{\mathcal{X}}$ belonging to the set J defined in (11) and to its complement J^c, respectively.

A straightforward extension of Proposition 2 leads to the following proposition for Model (8).

Proposition 4. *Let* $\widetilde{\mathcal{B}}(\lambda)$ *be defined by (10). Then*

$$\mathbb{P}\left(\mathrm{sign}(\widetilde{\mathcal{B}}(\lambda)) = \mathrm{sign}(\mathcal{B})\right) \geq \mathbb{P}(\widetilde{A_n} \cap \widetilde{B_n}),$$

where

$$\widetilde{A_n} = \left\{ \left|(\widetilde{C}_{J,J})^{-1}\widetilde{W}_J\right| < \sqrt{nq}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}|(\widetilde{C}_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_\mathrm{J})| \right) \right\} \tag{17}$$

and

$$\widetilde{B_n} = \left\{ \left| \widetilde{C}_{J^c,J} (\widetilde{C}_{J,J})^{-1} \widetilde{W}_J - \widetilde{W}_{J^c} \right| \leq \frac{\lambda}{2\sqrt{nq}} \left(\mathbf{1} - \left| \widetilde{C}_{J^c,J} (\widetilde{C}_{J,J})^{-1} \mathrm{sign}(\mathcal{B}_\mathrm{J}) \right| \right) \right\}, \tag{18}$$

with $W = \mathcal{X}'\widetilde{\mathcal{E}}/\sqrt{nq}$. *In (17) and (18),* $\widetilde{C}_{J,J}$ *and* $\widetilde{C}_{J^c,J}$ *are defined in (16) and* $\widetilde{W}_J$ *and* $\widetilde{W}_{J^c}$ *denote the components of* $\widetilde{W}$ *being in* J *and* J^c*, respectively. Note that the previous inequalities hold element-wise.*

The following theorem extends Theorem 1 to the case where Σ is unknown and gives some conditions under which the estimator $\widetilde{\mathcal{B}}$ defined in (10) is sign-consistent. The proof of Theorem 5 is given in Section 4 and is based on Proposition 4.

Theorem 5. *Assume that Assumptions* (A1)*,* (A2)*,* (A3)*,* (A4)*,* (IC) *and* (L) *of Theorem 1 hold. Assume also that, there exist some positive constants* M_4*,* M_5*,* M_6 *and* M_7*, such that for all* $n \geq 1$*,*

(A5) $\|(X'X)/n\|_\infty \leq M_4$,
(A6) $\lambda_{\min}((X'X)/n) \geq M_5$,
(A7) $\lambda_{\max}(\Sigma^{-1}) \leq M_6$,
(A8) $\lambda_{\min}(\Sigma^{-1}) \geq M_7$.

Suppose also that

(A9) $\|\Sigma^{-1} - \widehat{\Sigma}^{-1}\|_\infty = O_P((nq)^{-1/2})$*, as* n *tends to infinity,*
(A10) $\rho(\Sigma - \widehat{\Sigma}) = O_P((nq)^{-1/2})$*, as* n *tends to infinity.*

Let $\widetilde{\mathcal{B}}(\lambda)$ *be defined by (10), then*

$$\mathbb{P}\left(\mathit{sign}(\widetilde{\mathcal{B}}(\lambda)) = \mathit{sign}(\mathcal{B}) \right) \to 1, \text{ as } n \to \infty.$$

In the previous assumptions, $\lambda_{\max}(A)$*,* $\lambda_{\min}(A)$*,* $\rho(A)$ *and* $\|A\|_\infty$ *denote the largest eigenvalue, the smallest eigenvalue, the spectral radius and the infinite norm (induced by the associated vector norm) of the matrix* A*.*

Remark 3. Observe that Assumptions (A5) and (A6) hold in the case where the columns of the matrix X are orthogonal. Note also that (A7) and (A8) are the same as (C3) and (C4) in Proposition 3.

In order to estimate Σ, we propose the following strategy:

- Fitting a classical linear model to each column of the matrix Y in order to have access to an estimation $\widehat{E}$ of the random error matrix E. It is possible since p is assumed to be fixed and smaller than n.
- Estimating Σ from $\widehat{E}$ by assuming that Σ has a particular structure, Toeplitz for instance.

More precisely, $\widehat{E}$ defined in the first step is such that:

$$\widehat{E} = \left(\mathrm{Id}_{\mathbb{R}^n} - X(X'X)^{-1}X' \right) E =: \Pi E, \tag{19}$$

which implies that

$$\widehat{\mathcal{E}} = vec(\widehat{E}) = [\mathrm{Id}_{\mathbb{R}^q} \otimes \Pi] \, \mathcal{E}, \tag{20}$$

where $\mathcal{E}$ is defined in (5).

We prove in Proposition 7 below that our strategy for estimating Σ provides an estimator satisfying the assumptions of Theorem 5 in the case where $(E_{1,t})_t$, $(E_{2,t})_t$, ..., $(E_{n,t})_t$ are assumed to be independent AR(1) processes.

2.3. **The AR(1) case.**

2.3.1. *Sufficient conditions for Assumption* (IC) *of Theorem 1.* The following proposition gives some conditions under which the strong Irrepresentable Condition (IC) of Theorem 1 holds.

Proposition 6. *Assume that $(E_{1,t})_t$, $(E_{2,t})_t$, ..., $(E_{n,t})_t$ in Model (1) are independent AR(1) processes satisfying:*

$$E_{i,t} - \phi_1 E_{i,t-1} = Z_{i,t},\ \forall i \in \{1,\dots,n\},$$

where the $Z_{i,t}$'s are zero-mean i.i.d. Gaussian random variables with variance σ^2 and $|\phi_1| < 1$. Assume also that X defined in (1) is such that $X'X = \nu Id_{\mathbb{R}^p}$, where ν is a positive constant. Moreover, suppose that if $j \in J$, then $j > p$ and $j < pq - p$. Suppose also that for all j, $j-p$ or $j+p$ is not in J. Then, the strong Irrepresentable Condition (IC) *of Theorem 1 holds.*

The proof of Proposition 6 is given in Section 4.

2.3.2. *Sufficient conditions for Assumptions* (A7), (A8), (A9) *and* (A10) *of Theorem 5.* The following proposition establishes that in the particular case where the $(E_{1,t})_t$, $(E_{2,t})_t$, ..., $(E_{n,t})_t$ are independent AR(1) processes, our strategy for estimating Σ provides an estimator satisfying the assumptions of Theorem 5.

Proposition 7. *Assume that $(E_{1,t})_t$, $(E_{2,t})_t$, ..., $(E_{n,t})_t$ in Model (1) are independent AR(1) processes satisfying:*

$$E_{i,t} - \phi_1 E_{i,t-1} = Z_{i,t},\ \forall i \in \{1,\dots,n\},$$

where the $Z_{i,t}$'s are zero-mean i.i.d. Gaussian random variables with variance σ^2 and $|\phi_1| < 1$. Let

$$\widehat{\Sigma} = \frac{1}{1-\widehat{\phi}_1^2}\begin{pmatrix} 1 & \widehat{\phi}_1 & \widehat{\phi}_1^2 & \dots & \widehat{\phi}_1^{q-1} \\ \widehat{\phi}_1 & 1 & \widehat{\phi}_1 & \dots & \widehat{\phi}_1^{q-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \widehat{\phi}_1^{q-1} & \dots & \dots & \dots & 1 \end{pmatrix},$$

where

$$\widehat{\phi}_1 = \frac{\sum_{i=1}^n \sum_{\ell=2}^q \widehat{E}_{i,\ell}\widehat{E}_{i,\ell-1}}{\sum_{i=1}^n \sum_{\ell=1}^{q-1} \widehat{E}_{i,\ell}^2}, \tag{21}$$

where $\widehat{E} = (\widehat{E}_{i,\ell})_{1\le i\le n, 1\le \ell\le q}$ is defined in (19). Then, Assumptions (A7), (A8), (A9) *and* (A10) *of Theorem 5 are valid.*

The proof of Proposition 7 is given in Section 4. It is based on the following lemma.

Lemma 8. *Assume that $(E_{1,t})_t$, $(E_{2,t})_t$, ..., $(E_{n,t})_t$ in Model (1) are independent AR(1) processes satisfying:*

$$E_{i,t} - \phi_1 E_{i,t-1} = Z_{i,t},\ \forall i \in \{1,\dots,n\},$$

where the $Z_{i,t}$'s are zero-mean i.i.d. Gaussian random variables with variance σ^2 and $|\phi_1| < 1$. Let

$$\widehat{\phi}_1 = \frac{\sum_{i=1}^{n} \sum_{\ell=2}^{q} \widehat{E}_{i,\ell} \widehat{E}_{i,\ell-1}}{\sum_{i=1}^{n} \sum_{\ell=1}^{q-1} \widehat{E}_{i,\ell}^2},$$

where $\widehat{E} = (\widehat{E}_{i,\ell})_{1 \leq i \leq n, 1 \leq \ell \leq q}$ is defined in (19). Then,

$$\sqrt{nq_n}(\widehat{\phi}_1 - \phi_1) = O_p(1), \text{ as } n \to \infty.$$

Lemma 8 is proved in Section 4. Its proof is based on Lemma 10 in Section 5.

3. Numerical experiments

The goal of this section is twofold: *i*) to provide sanity checks for our theoretical results in a well-controlled framework; and *ii*) to investigate the robustness of our estimator to some violations of the assumptions of our theoretical results. The latter may reveal a broader scope of applicability for our method than the one guaranteed by the theoretical results.

We investigate *i*) in the AR(1) framework presented in Section 2.3. Indeed, all assumptions made in Theorems 1 and 5 can be specified with well-controllable simulation parameters in the AR(1) case with balanced design matrix X.

Point *ii*) aims to explore the limitations of our theoretical framework and assess its robustness. To this end, we propose two numerical studies relaxing some of the assumptions of our theorems: first, we study the effect of an unbalanced design – which violates the sufficient condition of the irrepresentability condition (IC) given in Proposition 6 – on the sign-consistency; and second, we study the effect of other types of dependence than an AR(1).

In all experiments, the performance are assessed in terms of sign-consistency. In other words, we evaluate the probability for the sign of various estimators to be equal to sign($\mathcal{B}$). We compare the performance of three different estimators:

- $\widehat{\mathcal{B}}$ defined in (6), which corresponds to the LASSO criterion applied to the data whitened with the true covariance matrix Σ; we call this estimator `oracle`. Its theoretical properties are established in Theorem 1.
- $\widetilde{\mathcal{B}}$ defined in (10), which corresponds to the LASSO criterion applied to the data whitened with an estimator of the covariance matrix $\widehat{\Sigma}$; we refer to this estimator as `whitened-lasso`. Its theoretical properties are established in Theorem 5.
- the LASSO criterion applied to the raw data, which we call `raw-lasso` hereafter. Its theoretical properties are established only in the univariate case in Alquier and Doukhan (2011).

3.1. AR(1) dependence structure with balanced one-way ANOVA.

In this section, we consider Model (1) where X is the design matrix of a one-way ANOVA with two balanced groups. Each row of the random error matrix E is distributed as a centered Gaussian random vector as in Equation (2) where the matrix Σ is the covariance matrix of an AR(1) process defined in Section 2.3.

In this setting, Assumptions (A1), (A2) and Condition (IC) of Theorem 1 are satisfied, see Propositions 3 and 6. The three remaining assumptions (A3), (A4) and (L) are related to more practical quantities: (A3) controls the sparsity level of the problem, involving c_1; (A4) basically controls the signal-to-noise ratio, involving c_2 and (L) links the sample size n, q and the two constants c_1, c_2, so that an appropriate range of penalty λ exists for having a large probability of support recovery. This latter assumption is used in our experiments to tune the difficulty of the support recovery as follows: we consider different values of n, q, c_1, c_2 and we choose a sparsity level $|J|$ and a minimal magnitude in $\mathcal{B}$ such that Assumptions (A3) and (A4) are fulfilled. Hence, the problem difficulty is essentially driven by the validity of Assumption (L) where $q = o(n^k)$ with $c_1 + c_2 = 1/2k$, and so by the relationship between n, q and k.

We consider a large range of sample sizes n varying from 10 to 1000 and three different values for q in $\{10, 50, 1000\}$. The constants c_1, c_2 are chosen such that $c_1 + c_2 = 1/2k$ with $c_1 = c_2$ and k in $\{1, 2, 4\}$. Additional values of c_1 and c_2 have also been considered and the corresponding results are available upon request. Finally, we consider two values for the parameter ϕ_1 appearing in the definition of the AR(1) process: $\phi_1 \in \{0.5, 0.95\}$.

Note that in this AR(1) setting with the estimator $\widehat{\phi}_1$ of ϕ_1 defined in (21) , all the assumptions of Theorem 5 are fulfilled, see Proposition 7.

The frequencies of support recovery for the three estimators averaged over 1000 replications is displayed in Figure 1.

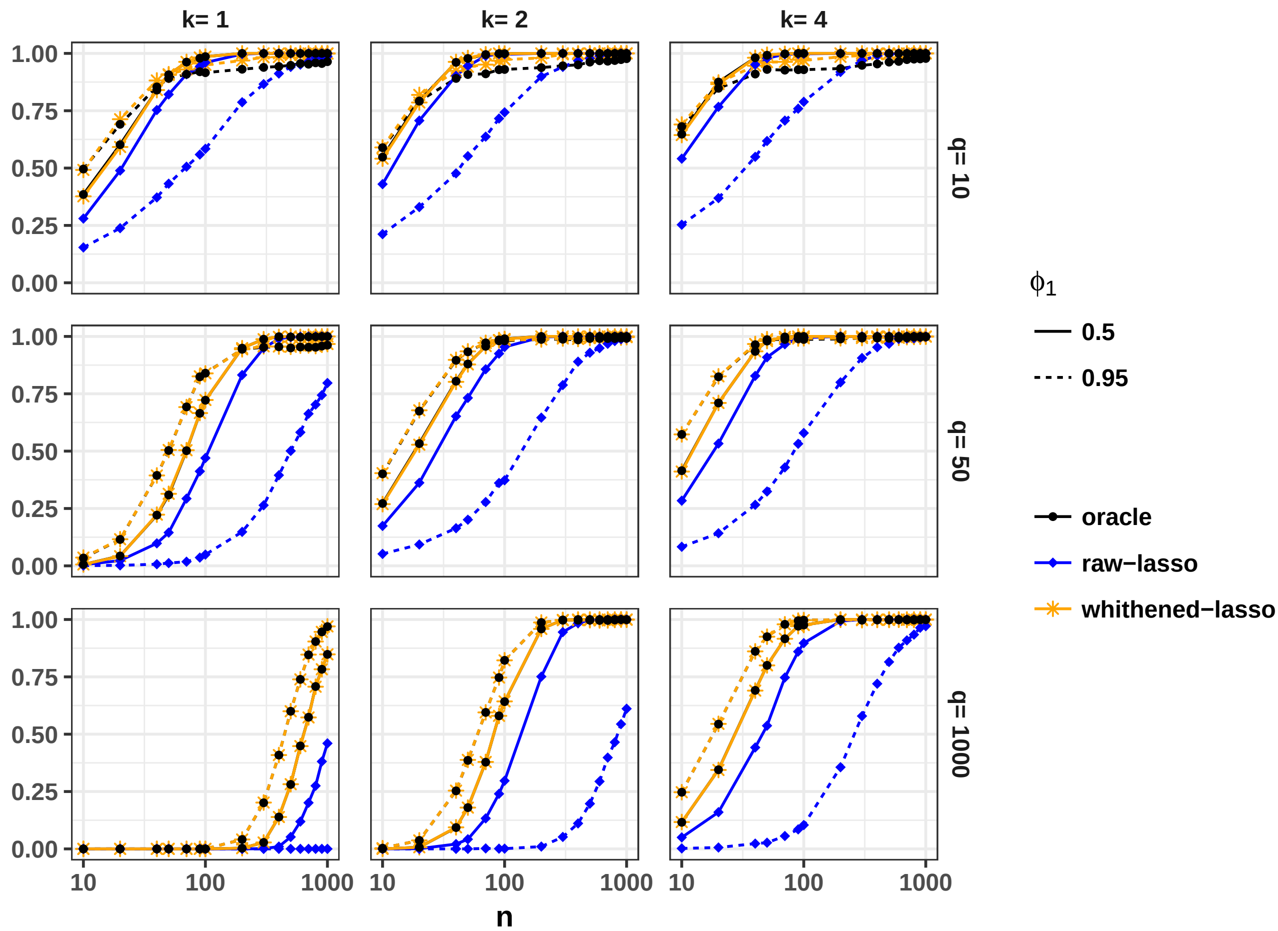

FIGURE 1. Frequencies of support recovery in a multivariate one-way ANOVA model with two balanced groups and an AR(1) dependence.

We observe from Figure 1 that `whitened-lasso` and `oracle` have similar performance since ϕ_1 is well estimated. These two approaches always exhibit better performance than `raw-lasso`, especially when $\phi_1 = 0.95$. In this case, the sample size n required to reach the same performance is indeed ten time larger for `raw-lasso` than for `oracle` and `whitened-lasso`.

Finally, the performance of all estimators are altered when n is too small, especially in situations where the signal to noise ratio (SNR) is small and the signal is not sparse enough, these two characteristics corresponding to small values of k.

3.2. **Robustness to unbalanced designs and correlated features.** The goal of this section is to study some particular design matrices X in Model (1) that may lead to violation of the Irrepresentability Condition (IC).

To this end, we consider the multivariate linear model (1) with the same AR(1) dependence as the one considered in Section 3.1. Then, two different matrices X are considered: First, an one-way ANOVA model with two unbalanced groups with respective sizes n_1 and n_2 such that $n_1 + n_2 = n$; and second, a multiple regression model with p correlated Gaussian predictors such that the rows of X are i.i.d. $\mathcal{N}(0, \Sigma^X)$.

For the one-way ANOVA, violation of (IC) may occur when $r = n_1/n$ is too different from $1/2$, as stated in Proposition 6. For the regression model, we choose for Σ^X a 9×9 matrix ($p = 9$) such that $\Sigma^X_{i,i} = 1$, $\Sigma^X_{i,j} = \rho$, when $i \neq j$. The other simulation parameters are fixed as in Section 3.1.

We report in Figure 2 the results for the case where $q = 1000$ and $k = 2$ both for unbalanced one-way ANOVA (top panels) and regression with correlated predictors (bottom panels). For the one-way ANOVA, r varies in $\{0.4, 0.2, 0.1\}$. For the regression case, ρ varies in $\{0.2, 0.6, 0.9\}$. In both cases, the gray lines correspond to the ideal situation (that is, either unbalanced or uncorrelated) denoted `Ideal` in the legend of Figure 2. The probability of support recovery is estimated over 1000 runs.

From this figure, we note that correlated features or unbalanced designs deteriorate the support recovery of all estimators. This was expected for these LASSO-based methods which all suffer from the violation of the irrepresentability condition (IC). However, we also note that `whitened-lasso` and `oracle` have similar performance, which means that the estimation of Σ is not altered, and that whitening always improves the support recovery.

3.3. **Robustness to more general autoregressive processes.** In this section, we consider the case where X is the design matrix of a one-way ANOVA with two balanced groups and where Σ is the covariance matrix of an AR(m) process with m in $\{5, 10\}$. Figure 3 displays the performance of the different estimators when $q = 500$. Here, for computing $\widehat{\Sigma}$ in `whitened-lasso`, the parameters $\phi_1, \dots, \phi_m$ of the AR(m) process are estimated as follows. They are obtained by averaging over the n rows of $\widehat{E}$ defined in (19) the estimations $\widehat{\phi}_1^{(i)}, \dots, \widehat{\phi}_m^{(i)}$ obtained for the ith row of $\widehat{E}$ by using standard estimation approaches for AR processes described in Brockwell and Davis (1990). As previously, we observe from this figure that `whitened-lasso` and `oracle` have better performance than `raw-lasso`.

4. Proofs

Proof of Proposition 2. For a fixed nonnegative λ, by (6),

$$\widehat{\mathcal{B}} = \widehat{\mathcal{B}}(\lambda) = \text{Argmin}_{\mathcal{B}} \left\{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda \|\mathcal{B}\|_1 \right\}.$$

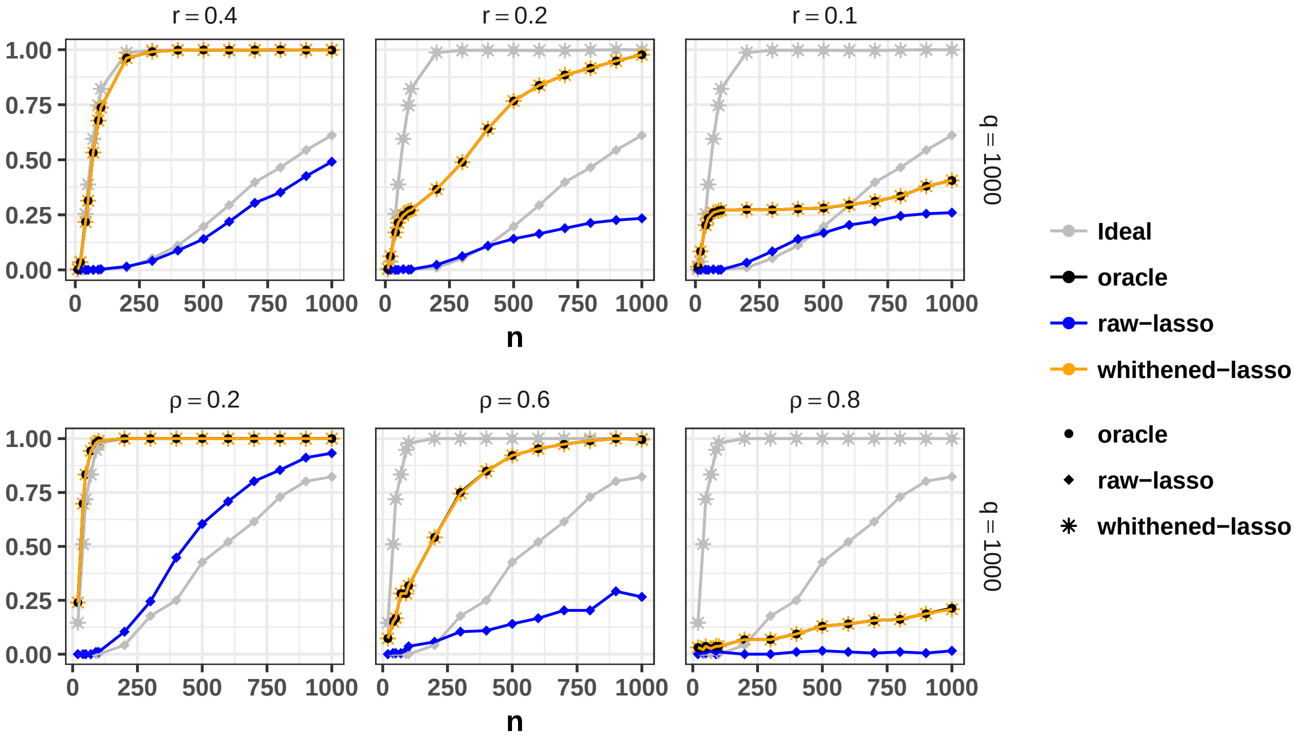

FIGURE 2. Frequencies of support recovery in general linear models with unbalanced designs: one-way ANOVA and regression.

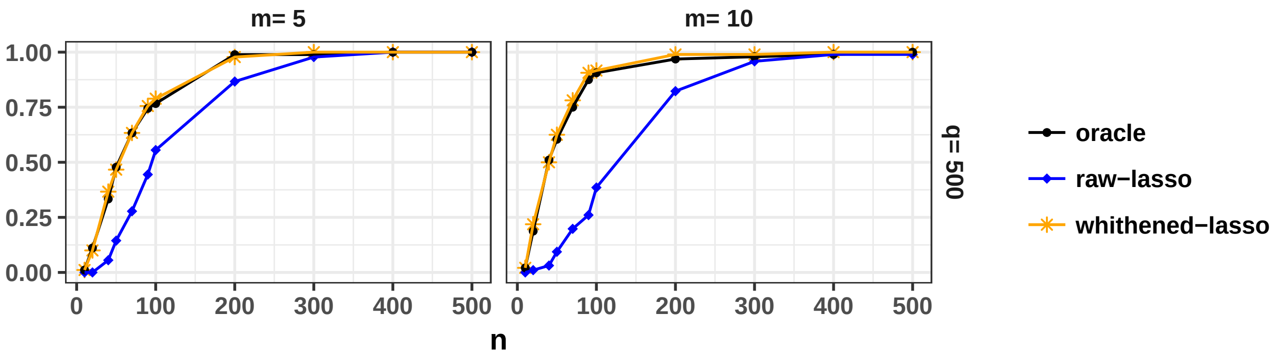

FIGURE 3. Frequencies of support recovery in one-way ANOVA with AR(m) covariance matrix.

Denoting $\widehat{u} = \widehat{\mathcal{B}} - \mathcal{B}$, we get

$$\|\mathcal{Y} - \mathcal{X}\widehat{\mathcal{B}}\|_2^2 + \lambda\|\widehat{\mathcal{B}}\|_1 = \|\mathcal{X}\mathcal{B} + \mathcal{E} - \mathcal{X}\widehat{\mathcal{B}}\|_2^2 + \lambda\|\widehat{u} + \mathcal{B}\|_1 = \|\mathcal{E} - \mathcal{X}\widehat{u}\|_2^2 + \lambda\|\widehat{u} + \mathcal{B}\|_1$$
$$= \|\mathcal{E}\|_2^2 - 2\widehat{u}'\mathcal{X}'\mathcal{E} + \widehat{u}'\mathcal{X}'\mathcal{X}\widehat{u} + \lambda\|\widehat{u} + \mathcal{B}\|_1.$$

Thus,

$$\widehat{u} = \text{Argmin}_u\, V(u),$$

where

$$V(u) = -2(\sqrt{nq}u)'W + (\sqrt{nq}u)'C(\sqrt{nq}u) + \lambda\|u + \mathcal{B}\|_1.$$

Since the first derivative of V with respect to u is equal to

$$2\sqrt{nq}\big(C(\sqrt{nq}u) - W\big) + \lambda\,\text{sign}(\mathrm{u} + \mathcal{B}),$$

$\widehat{u}$ satisfies

$$C_{J,J}(\sqrt{nq}\widehat{u}_J) - W_J = -\frac{\lambda}{2\sqrt{nq}}\text{sign}(\widehat{\mathrm{u}}_{\mathrm{J}} + \mathcal{B}_{\mathrm{J}}) = -\frac{\lambda}{2\sqrt{\mathrm{nq}}}\text{sign}(\widehat{\mathcal{B}}_{\mathrm{J}}),\ \text{ if } \widehat{\mathrm{u}}_{\mathrm{J}} + \mathcal{B}_{\mathrm{J}} = \widehat{\mathcal{B}}_{\mathrm{J}} \neq 0$$

and

$$\left|C_{J^c,J}(\sqrt{nq}\widehat{u}_J) - W_{J^c}\right| \leq \frac{\lambda}{2\sqrt{nq}}.$$

Note that, if $|\widehat{u}_J| < |\mathcal{B}_J|$, then $\widehat{\mathcal{B}}_J \neq 0$ and $\text{sign}(\widehat{\mathcal{B}}_J) = \text{sign}(\mathcal{B}_J)$.

Let us now prove that when A_n and B_n, defined in (13) and (14), are satisfied then there exists $\widehat{u}$ satisfying:

$$C_{J,J}(\sqrt{nq}\widehat{u}_J) - W_J = -\frac{\lambda}{2\sqrt{nq}}\text{sign}(\mathcal{B}_J), \tag{22}$$

$$|\widehat{u}_J| < |\mathcal{B}_J|, \tag{23}$$

$$\left|C_{J^c,J}(\sqrt{nq}\widehat{u}_J) - W_{J^c}\right| \leq \frac{\lambda}{2\sqrt{nq}}. \tag{24}$$

Note that A_n implies:

$$\sqrt{nq}\left(-|\mathcal{B}_J| + \frac{\lambda}{2nq}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right) < (C_{J,J})^{-1}W_J < \sqrt{nq}\left(|\mathcal{B}_J| + \frac{\lambda}{2nq}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right). \tag{25}$$

By denoting

$$\widehat{u}_J = \frac{1}{\sqrt{nq}}(C_{J,J})^{-1}W_J - \frac{\lambda}{2nq}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J), \tag{26}$$

we obtain from (25) that (22) and (23) hold. Note that B_n implies:

$$\begin{aligned} &-\frac{\lambda}{2\sqrt{nq}}\left(\mathbf{1} - C_{J^c,J}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right) \\ &\qquad \leq C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c} \leq \frac{\lambda}{2\sqrt{nq}}\left(\mathbf{1} + C_{J^c,J}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right). \end{aligned}$$

Hence,

$$\left|C_{J^c,J}\left((C_{J,J})^{-1}W_J - \frac{\lambda}{2\sqrt{nq}}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right) - W_{J^c}\right| \leq \frac{\lambda}{2\sqrt{nq}},$$

which is (24) by (26). This concludes the proof. □

Proof of Theorem 1. By Proposition 2,

$$\mathbb{P}\left(\text{sign}(\widehat{\mathcal{B}}(\lambda)) = \text{sign}(\mathcal{B})\right) \geq \mathbb{P}(A_n \cap B_n) = 1 - \mathbb{P}(A_n^c \cup B_n^c) \geq 1 - \mathbb{P}(A_n^c) - \mathbb{P}(B_n^c),$$

where A_n and B_n are defined in (13) and (14). It is thus enough to prove that $\mathbb{P}(A_n^c)$ and $\mathbb{P}(B_n^c)$ tend to zero as n tends to infinity.

By definition of A_n,

$$\mathbb{P}(A_n^c) = \mathbb{P}\left(\left|(C_{J,J})^{-1}W_J\right| \geq \sqrt{nq}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}|(C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J)|\right)\right)$$

$$\leq \sup_{j\in J}\mathbb{P}\left(|\xi_j| \geq \sqrt{nq}\left(|\mathcal{B}_j| - \frac{\lambda}{2nq}|b_j|\right)\right), \tag{27}$$

where

$$\xi = (\xi_j)_{j\in J} = (C_{J,J})^{-1}W_J = \frac{1}{\sqrt{nq}}(C_{J,J})^{-1}(\mathcal{X}_{\bullet,J})'\mathcal{E} =: H_A\,\mathcal{E},$$

and

$$b = (b_j)_{j\in J} = (C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J).$$

By definition of B_n and (IC),

$$\mathbb{P}(B_n^c) = \mathbb{P}\left(\left|C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c}\right| > \frac{\lambda}{2\sqrt{nq}}\left(\mathbf{1} - \left|C_{J^c,J}(C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J)\right|\right)\right)$$

$$\leq \mathbb{P}\left(\left|C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c}\right| > \frac{\lambda}{2\sqrt{nq}}\eta\right)$$

$$\leq \sup_{j\in J^c}\mathbb{P}\left(|\zeta_j| > \frac{\lambda}{2\sqrt{nq}}\eta\right), \tag{28}$$

where

$$\zeta = (\zeta_j)_{j\in J^c} = C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c} = \frac{1}{\sqrt{nq}}\left(C_{J^c,J}(C_{J,J})^{-1}(\mathcal{X}_{\bullet,J})' - (\mathcal{X}_{\bullet,J^c})'\right)\mathcal{E} =: H_B\,\mathcal{E}.$$

Note that, for all j in J,

$$|b_j| \leq \sum_{j\in J}|b_j| \leq \sqrt{|J|}\left(\sum_{j\in J}b_j^2\right)^{1/2} = \sqrt{|J|}\|b\|_2.$$

Moreover,

$$\|b\|_2 = \|(C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J)\|_2 \leq \|(C_{J,J})^{-1}\|_2\sqrt{|J|} := \lambda_{\max}((C_{J,J})^{-1})\sqrt{|J|},$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of the matrix A. Observe that

$$\lambda_{\max}((C_{J,J})^{-1}) = \frac{1}{\lambda_{\min}(C_{J,J})} = \frac{q}{\lambda_{\min}((\mathcal{X}'\mathcal{X})_{J,J})/n} \leq \frac{q}{M_2}, \tag{29}$$

by Assumption (A2) of Theorem 1. Thus, for all j in J,

$$|b_j| \leq \frac{q|J|}{M_2}. \tag{30}$$

By Assumption (A4) of Theorem 1, we get thus that for all j in J,

$$\sqrt{nq}\left(|\mathcal{B}_j| - \frac{\lambda}{2nq}\left|\left((C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J)\right)_j\right|\right) = \sqrt{nq}\left(|\mathcal{B}_j| - \frac{\lambda}{2nq}|b_j|\right) \geq \sqrt{nq}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right). \tag{31}$$

Thus,

$$\mathbb{P}(A_n^c) \leq \sup_{j\in J} \mathbb{P}\left(|\xi_j| \geq \sqrt{nq}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right)\right). \tag{32}$$

Since $\mathcal{E}$ is a centered Gaussian random vector having a covariance matrix equal to identity, $\xi = H_A\,\mathcal{E}$ is a centered Gaussian random vector with a covariance matrix equal to:

$$H_A H_A' = \frac{1}{nq}(C_{J,J})^{-1}(\mathcal{X}_{\bullet,J})'\mathcal{X}_{\bullet,J}(C_{J,J})^{-1} = (C_{J,J})^{-1}.$$

Hence, by (29), we get that for all j in J,

$$\mathrm{Var}(\xi_j) = \left((C_{J,J})^{-1}\right)_{jj} \leq \lambda_{\max}(C_{J,J}^{-1}) \leq \frac{q}{M_2}.$$

Thus,

$$\mathbb{P}\left(|\xi_j| \geq \sqrt{nq}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right)\right) \leq \mathbb{P}\left(|Z| \geq \frac{\sqrt{M_2}}{\sqrt{q}}\left(M_3 q^{-c_2}\sqrt{nq} - \frac{\lambda q|J|}{2\sqrt{nq}M_2}\right)\right),$$

where Z is a standard Gaussian random variable. By Chernoff inequality, we thus obtain that for all j in J,

$$\mathbb{P}\left(|\xi_j| \geq \sqrt{nq}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right)\right) \leq 2\exp\left(-\frac{M_2}{2q}\left\{M_3 q^{-c_2}\sqrt{nq} - \frac{\lambda q|J|}{2\sqrt{nq}M_2}\right\}^2\right).$$

By Assumption (A3) of Theorem 1, we get that under the last condition of (L),

$$\frac{\lambda q|J|}{\sqrt{nq}} = o\left(q^{-c_2}\sqrt{nq}\right), \text{ as } n \to \infty. \tag{33}$$

Thus,

$$\mathbb{P}(A_n^c) \to 0, \quad \text{as } n \to \infty. \tag{34}$$

Let us now bound $\mathbb{P}(B_n^c)$. Observe that $\zeta = H_B\,\mathcal{E}$ is a centered Gaussian random vector with a covariance matrix equal to:

$$\begin{aligned}
H_B H_B' &= \frac{1}{nq}(C_{J^c,J}(C_{J,J})^{-1}(\mathcal{X}_{\bullet,J})' - \mathcal{X}_{\bullet,J^c}')(\mathcal{X}_{\bullet,J}(C_{J,J})^{-1}C_{J,J^c} - \mathcal{X}_{\bullet,J^c})\\
&= C_{J^c,J^c} - C_{J^c,J}(C_{J,J})^{-1}C_{J,J^c} = \frac{1}{nq}(\mathcal{X}_{\bullet,J^c})'\left(\mathrm{Id}_{\mathbb{R}^{nq}} - \mathcal{X}_{\bullet,J}((\mathcal{X}_{\bullet,J})'\mathcal{X}_{\bullet,J})^{-1}(\mathcal{X}_{\bullet,J})'\right)\mathcal{X}_{\bullet,J^c}\\
&= \frac{1}{nq}(\mathcal{X}_{\bullet,J^c})'\left(\mathrm{Id}_{\mathbb{R}^{nq}} - \Pi_{\mathrm{Im}(\mathcal{X}_{\bullet,J})}\right)\mathcal{X}_{\bullet,J^c},
\end{aligned}$$

where $\Pi_{\mathrm{Im}(\mathcal{X}_{\bullet,J})}$ denotes the orthogonal projection onto the column space of $\mathcal{X}_{\bullet,J}$. Note that, for all j in J^c,

$$\begin{aligned}
\mathrm{Var}(\zeta_j) &= \frac{1}{nq}\left((\mathcal{X}_{\bullet,J^c})'\left(\mathrm{Id}_{\mathbb{R}^{nq}} - \Pi_{\mathrm{Im}(\mathcal{X}_{\bullet,J})}\right)\mathcal{X}_{\bullet,J^c}\right)_{jj}\\
&= \frac{1}{nq}\left((\mathcal{X}_{\bullet,J^c})'\mathcal{X}_{\bullet,J^c}\right)_{jj} - \frac{1}{nq}\left((\mathcal{X}_{\bullet,J^c})'\,\Pi_{\mathrm{Im}(\mathcal{X}_{\bullet,J})}\mathcal{X}_{\bullet,J^c}\right)_{jj}\\
&\leq \frac{1}{nq}\left((\mathcal{X}_{\bullet,J^c})'\mathcal{X}_{\bullet,J^c}\right)_{jj} \leq \frac{M_1}{q},
\end{aligned}$$

where the inequalities come from Lemma 9 and Assumption (A1) of Theorem 1. Thus, for all j in J^c,

$$\mathbb{P}\left(|\zeta_j| > \frac{\lambda}{2\sqrt{nq}}\eta\right) \leq \mathbb{P}\left(|Z| > \frac{\lambda\sqrt{q}}{2\sqrt{M_1}\sqrt{nq}}\eta\right),$$

where Z is a standard Gaussian random variable. By Chernoff inequality, for all j in J^c,

$$\mathbb{P}\left(|\zeta_j| > \frac{\lambda}{2\sqrt{nq}}\eta\right) \leq 2\exp\left\{-\frac{1}{2}\left(\frac{\lambda}{2\sqrt{M_1}\sqrt{n}}\eta\right)^2\right\}.$$

Hence, under the following assumption

$$\frac{\lambda}{\sqrt{n}} \to \infty,$$

which is the second condition of (L),

$$\mathbb{P}(B_n^c) \to 0, \quad \text{as } n \to \infty. \tag{35}$$

□

Proof of Proposition 3. Let us first prove that (C1) and (C3) imply (A1). For $j \in \{1, \ldots, pq\}$, by considering the Euclidian division of $j-1$ by p given by $(j-1) = pk_j + r_j$, we observe that

$$\begin{aligned}
(\mathcal{X}_{\bullet,j})'\mathcal{X}_{\bullet,j} &= (((\Sigma^{-1/2})' \otimes X)_{\bullet,j})'((\Sigma^{-1/2})' \otimes X)_{\bullet,j} \\
&= ((\Sigma^{-1/2}) \otimes X')_{j,\bullet})((\Sigma^{-1/2})' \otimes X)_{\bullet,j} \\
&= ((\Sigma^{-1/2})_{k_j+1,\bullet} \otimes (X_{\bullet,r_j+1})')(((\Sigma^{-1/2})_{\bullet,k_j+1})' \otimes X_{\bullet,r_j+1}) \\
&= (\Sigma^{-1/2})_{k_j+1,\bullet}((\Sigma^{-1/2})_{\bullet,k_j+1})' \otimes (X_{\bullet,r_j+1})'X_{\bullet,r_j+1} \\
&= (\Sigma^{-1})_{k_j+1,k_j+1} \otimes (X_{\bullet,r_j+1})'X_{\bullet,r_j+1} \\
&= (\Sigma^{-1})_{k_j+1,k_j+1}(X_{\bullet,r_j+1})'X_{\bullet,r_j+1}.
\end{aligned}$$

Hence, using (C1), we get that for all j in $\{1, \ldots, pq\}$,

$$\begin{aligned}
\frac{1}{n}(\mathcal{X}_{\bullet,j})'\mathcal{X}_{\bullet,j} &\leq M_1'(\Sigma^{-1})_{k_j+1,k_j+1} \leq M_1' \sup_{k\in\{0,\ldots,q-1\}}((\Sigma^{-1})_{k+1,k+1}) \\
&\leq M_1'\lambda_{\max}(\Sigma^{-1}) \leq M_1' m_1,
\end{aligned}$$

where the last inequality comes from (C3), which gives (A1).

Let us now prove that (C2) and (C4) imply (A2). Note that

$$\begin{aligned}
(\mathcal{X}'\mathcal{X})_{J,J} &= (((\Sigma^{-1/2})' \otimes X)'((\Sigma^{-1/2})' \otimes X))_{J,J} \\
&= (\Sigma^{-1/2}(\Sigma^{-1/2})' \otimes X'X)_{J,J} \\
&= (\Sigma^{-1} \otimes X'X)_{J,J}.
\end{aligned}$$

Then, by Theorem 4.3.15 of Horn and Johnson (1986),

$$\begin{aligned}
\lambda_{\min}((\mathcal{X}'\mathcal{X})_{J,J}) &= \lambda_{\min}((\Sigma^{-1} \otimes X'X)_{J,J}) \\
&\geq \lambda_{\min}(\Sigma^{-1} \otimes X'X) \\
&= \lambda_{\min}(X'X)\lambda_{\min}(\Sigma^{-1}). \qquad (36)
\end{aligned}$$

Finally, by using Conditions (C2) and (C4), we obtain

$$\frac{1}{n}\lambda_{\min}(\mathcal{X}'\mathcal{X})_{J,J} \geq \frac{1}{n}\lambda_{\min}(X'X)\lambda_{\min}(\Sigma^{-1}) \geq M_2' m_2,$$

which gives (A2). □

Proof of Theorem 5. By Proposition 4,

$$\mathbb{P}\left(\text{sign}(\widetilde{\mathcal{B}}(\lambda)) = \text{sign}(\mathcal{B})\right) \geq \mathbb{P}(\widetilde{A}_n \cap \widetilde{B}_n) = 1 - \mathbb{P}(\widetilde{A}_n^c \cup \widetilde{B}_n^c) \geq 1 - \mathbb{P}(\widetilde{A}_n^c) - \mathbb{P}(\widetilde{B}_n^c),$$

where $\widetilde{A}_n$ and $\widetilde{B}_n$ are defined in (17) and (18). By definition of $\widetilde{A}_n$, we get

$$\mathbb{P}(\widetilde{A}_n^c) = \mathbb{P}\left(\left\{\left|(\widetilde{C}_{J,J})^{-1}\widetilde{W}_J\right| \geq \sqrt{nq}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}|(\widetilde{C}_{J,J})^{-1}\text{sign}(\mathcal{B}_J)|\right)\right\}\right).$$

Observing that

$$\begin{aligned}(\widetilde{C}_{J,J})^{-1}\widetilde{W}_J &= (C_{J,J})^{-1}W_J + (C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right)\\ &+ \left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)W_J\\ &+ \left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\left(\widetilde{W}_J - W_J\right),\end{aligned}$$

$$(\widetilde{C}_{J,J})^{-1}\text{sign}(\mathcal{B}_J) = (C_{J,J})^{-1}\text{sign}(\mathcal{B}_J) + \left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J),$$

and using the triangle inequality, we obtain that

$$\begin{aligned}&\mathbb{P}(\widetilde{A}_n^c) \leq \mathbb{P}\left(\left\{\left|(C_{J,J})^{-1}W_J\right| \geq \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}\left|(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right|\right)\right\}\right)\\ &+ \mathbb{P}\left(\left\{\left|(C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right)\right| \geq \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}\left|(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right|\right)\right\}\right)\\ &+ \mathbb{P}\left(\left\{\left|\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)W_J\right| \geq \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}\left|(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right|\right)\right\}\right)\\ &+ \mathbb{P}\left(\left\{\left|\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\left(\widetilde{W}_J - W_J\right)\right| \geq \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}\left|(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right|\right)\right\}\right)\\ &+ \mathbb{P}\left(\left\{\frac{\lambda}{2\sqrt{nq}}\left|\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J)\right| \geq \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}\left|(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right|\right)\right\}\right).\end{aligned} \tag{37}$$

The first term in the r.h.s of (37) tends to 0 by the definition of A_n^c and (34). By (31), the last term of (37) satisfies, for all $j \in J$:

$$\begin{aligned}&\mathbb{P}\left(\left|\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J)\right| \geq \frac{2nq}{5\lambda}\left(|\mathcal{B}_J| - \frac{\lambda}{2nq}\left|(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right|\right)\right)\\ &\leq \mathbb{P}\left(\left|\left(\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J)\right)_j\right| \geq \frac{2nq}{5\lambda}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right)\right).\end{aligned}$$

Let $U = (\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}$ and $s = \mathrm{sign}(\mathcal{B}_J)$ then for all j in J:

$$|(Us)_j| = \left|\sum_{k\in J} U_{jk}s_k\right| \leq \sqrt{|J|}\|U\|_2. \tag{38}$$

We focus on

$$\|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2 = \|(\widetilde{C}_{J,J})^{-1}(C_{J,J} - \widetilde{C}_{J,J})(C_{J,J})^{-1}\|_2 \leq \|(\widetilde{C}_{J,J})^{-1}\|_2 \; \|C_{J,J} - \widetilde{C}_{J,J}\|_2 \; \|(C_{J,J})^{-1}\|_2$$
$$\leq \frac{\rho(C_{J,J} - \widetilde{C}_{J,J})}{\lambda_{\min}(\widetilde{C}_{J,J})\lambda_{\min}(C_{JJ})} \leq \frac{\rho(C_{J,J} - \widetilde{C}_{J,J})}{\lambda_{\min}(\widetilde{C}_{J,J})(M_2/q)},$$

where the last inequality comes from Assumption (A2) of Theorem 1, which gives that

$$\|(C_{J,J})^{-1}\|_2 \leq \frac{q}{M_2}. \tag{39}$$

Using Theorem 4.3.15 of Horn and Johnson (1986), we get

$$\|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2 \leq \frac{q\rho(C - \widetilde{C})}{\lambda_{\min}(\widetilde{C})M_2}.$$

By definition of C and $\widetilde{C}$ given in (11) and (15), respectively, we get

$$C = \frac{\Sigma^{-1} \otimes (X'X)}{nq} \text{ and } \widetilde{C} = \frac{\widehat{\Sigma}^{-1} \otimes (X'X)}{nq}. \tag{40}$$

By using that the eigenvalues of the Kronecker product of two matrices is equal to the product of the eigenvalues of the two matrices, we obtain

$$\|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2 \leq \frac{\rho(\Sigma^{-1} - \widehat{\Sigma}^{-1})\lambda_{\max}((X'X)/n)q}{\lambda_{\min}(\widehat{\Sigma}^{-1})\lambda_{\min}((X'X)/n)M_2} \leq \frac{\rho(\Sigma^{-1} - \widehat{\Sigma}^{-1})\lambda_{\max}(\widehat{\Sigma})\lambda_{\max}((X'X)/n)q}{\lambda_{\min}((X'X)/n)M_2}$$
$$\leq \frac{\rho(\Sigma^{-1} - \widehat{\Sigma}^{-1})\left(\rho(\widehat{\Sigma} - \Sigma) + \lambda_{\max}(\Sigma)\right)\lambda_{\max}((X'X)/n)q}{\lambda_{\min}((X'X)/n)M_2}, \tag{41}$$

where the last inequality follows from Theorem 4.3.1 of Horn and Johnson (1986). Thus, by Assumptions (A5), (A6), (A8), (A9) and (A10), we get that

$$\|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2 = O_P(q(nq)^{-1/2}), \text{ as } n \to \infty. \tag{42}$$

Hence, by (38), we get for all j in J that

$$\mathbb{P}\left(\left|\left(\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\mathrm{sign}(\mathcal{B}_J)\right)_j\right| \geq \frac{2nq}{5\lambda}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right)\right)$$
$$\leq \mathbb{P}\left(\sqrt{|J|}\;\|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2 \geq \frac{2\sqrt{nq}}{5\lambda}\left(M_3 q^{-c_2}\sqrt{nq} - \frac{\lambda q|J|}{2\sqrt{nq}M_2}\right)\right).$$

By (33), (42) and (A3), it is enough to prove that

$$\mathbb{P}\left(q^{c_1/2}q(nq)^{-1/2} \geq \frac{nq}{\lambda}q^{-c_2}\right) \to 0, \text{ as } n \to \infty.$$

By the last condition of (L),

$$\frac{\frac{nq}{\lambda}q^{-c_2}}{q^{1+c_1}} \to \infty, \text{ as } n \to \infty,$$

and the result follows since n tends to infinity. Hence, the last term of (37) tends to zero as n tends to infinity.

Let us now study the second term in the r.h.s of (37).

$$\widetilde{W}_J - W_J = \frac{1}{\sqrt{nq}}\left(\left(\widetilde{\mathcal{X}}'\widetilde{\mathcal{E}}\right)_J - (\mathcal{X}'\mathcal{E})_J\right) = \frac{1}{\sqrt{nq}}\left(\widetilde{\mathcal{X}}'\widetilde{\mathcal{E}} - \mathcal{X}'\mathcal{E}\right)_J$$

$$= \frac{1}{\sqrt{nq}}\left[\left(\widehat{\Sigma}^{-1/2}\otimes X'\right)\left((\widehat{\Sigma}^{-1/2})'\otimes \mathrm{Id}_{\mathbb{R}^n}\right)\mathrm{Vec}(E) - \left(\Sigma^{-1/2}\otimes X'\right)\left((\Sigma^{-1/2})'\otimes \mathrm{Id}_{\mathbb{R}^n}\right)\mathrm{Vec}(E)\right]_J$$

(43)

$$= \frac{1}{\sqrt{nq}}\left[\left\{\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\otimes X'\right\}\mathrm{Vec}(E)\right]_J \stackrel{d}{=} AZ,$$

where Z is a centered Gaussian random vector having a covariance matrix equal to identity and

$$A = \frac{1}{\sqrt{nq}}\left[\left\{\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\otimes X'\right\}\left\{(\Sigma^{1/2})'\otimes \mathrm{Id}_{\mathbb{R}^n}\right\}\right]_{J,\bullet}. \tag{44}$$

By Cauchy-Schwarz inequality, we get for all $K\times nq$ matrix B, and all $nq\times 1$ vector U that for all k in $\{1,\dots,K\}$,

$$|(BU)_k| = \left|\sum_{\ell=1}^{nq} B_{k,\ell}U_\ell\right| \le \|B\|_2\,\|U\|_2. \tag{45}$$

Thus, for all j in J, for all γ in $\mathbb{R}$ and all $|J|\times|J|$ matrix D,

$$\mathbb{P}\left(\left|\left(D\left(\widetilde{W}_J - W_J\right)\right)_j\right| \ge \gamma\right) = \mathbb{P}\left(\left|(DAZ)_j\right|\ge\gamma\right) \le \mathbb{P}\left(\|D\|_2\,\|A\|_2\|Z\|_2 \ge \gamma\right), \tag{46}$$

where A is defined in (44) and Z is a centered Gaussian random vector having a covariance matrix equal to identity. Hence, for all j in J,

$$\mathbb{P}\left(\left|\left((C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right)\right)_j\right| \ge \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_j| - \frac{\lambda}{2nq}\left|\left((C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J)\right)_j\right|\right)\right)$$
$$\le \mathbb{P}\left(\|(C_{J,J})^{-1}\|_2\|A\|_2\|Z\|_2 \ge \frac{\sqrt{nq}}{5}\left(M_3 q^{-c_2} - \frac{\lambda q|J|}{2nqM_2}\right)\right).$$

Let us bound $\|A\|_2$. Observe that

$$\left\|\left[\left\{\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\otimes X'\right\}\left\{(\Sigma^{1/2})'\otimes \mathrm{Id}\right\}\right]_{J,\bullet}\right\|_2$$
$$= \rho\left(\left[\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\Sigma\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\otimes(X'X)\right]_{J,J}\right)^{1/2}$$
$$\le \rho\left(\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\Sigma\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\right)^{1/2}\lambda_{\max}\left(X'X\right)^{1/2}$$
$$\le \rho\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)\lambda_{\max}(\Sigma)^{1/2}\lambda_{\max}\left(X'X\right)^{1/2},$$

(47)

where the first inequality comes from Theorem 4.3.15 of Horn and Johnson (1986). Hence, by (A5), (A8) and (A9)

$$\|A\|_2 = \frac{1}{\sqrt{nq}} \left\| \left[\left\{ \left(\widehat{\Sigma}^{-1} - \Sigma^{-1} \right) \otimes X' \right\} \left\{ (\Sigma^{1/2})' \otimes \mathrm{Id} \right\} \right]_{J,\bullet} \right\|_2 = O_P(q^{-1/2}(nq)^{-1/2}). \tag{48}$$

By (33), (39) and (48), it is enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq nq\, n\, q^{-2c_2} \right) \to 0, \text{ as } n \to \infty.$$

The result follows from the Markov inequality and the first condition of (L).

Let us now study the third term in the r.h.s of (37). Observe that

$$\begin{aligned} W_J &= \frac{1}{\sqrt{nq}} \left[\left(\Sigma^{-1/2} \otimes X' \right) \left((\Sigma^{-1/2})' \otimes \mathrm{Id}_{\mathbb{R}^n} \right) \mathrm{Vec}(E) \right]_J \\ &\stackrel{d}{=} \frac{1}{\sqrt{nq}} \left[\left(\Sigma^{-1} \otimes X' \right) \left((\Sigma^{1/2})' \otimes \mathrm{Id}_{\mathbb{R}^n} \right) \right]_{J,\bullet} Z =: A_1 Z, \end{aligned} \tag{49}$$

where Z is a centered Gaussian random vector having a covariance matrix equal to identity and

$$A_1 = \frac{1}{\sqrt{nq}} \left[\left(\Sigma^{-1} \otimes X' \right) \left((\Sigma^{1/2})' \otimes \mathrm{Id}_{\mathbb{R}^n} \right) \right]_{J,\bullet}. \tag{50}$$

Using (45), we get for all j in J, for all γ in $\mathbb{R}$ and all $|J| \times |J|$ matrix D,

$$\mathbb{P}\left(\left| (D\, W_J)_j \right| \geq \gamma \right) = \mathbb{P}\left(\left| (D A_1 Z)_j \right| \geq \gamma \right) \leq \mathbb{P}\left(\|D\|_2\, \|A_1\|_2\, \|Z\|_2 \geq \gamma \right), \tag{51}$$

where A_1 is defined in (50) and Z is a centered Gaussian random vector having a covariance matrix equal to identity. Hence, for all j in J,

$$\begin{aligned} &\mathbb{P}\left(\left| \left(\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1} \right) W_J \right)_j \right| \geq \frac{\sqrt{nq}}{5} \left(|\mathcal{B}_j| - \frac{\lambda}{2nq} \left| \left((C_{J,J})^{-1} \mathrm{sign}(\mathcal{B}_J) \right)_j \right| \right) \right) \\ &\leq \mathbb{P}\left(\left\| (\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1} \right\|_2 \|A_1\|_2\, \|Z\|_2 \geq \frac{\sqrt{nq}}{5} \left(M_3 q^{-c_2} - \frac{\lambda q |J|}{2nqM_2} \right) \right). \end{aligned}$$

Let us now bound $\|A_1\|_2$. Note that

$$\begin{aligned} &\left\| \left[\left(\Sigma^{-1} \otimes X' \right) \left((\Sigma^{1/2})' \otimes \mathrm{Id}_{\mathbb{R}^n} \right) \right]_{J,\bullet} \right\|_2 = \left\| \left[\left(\Sigma^{-1/2} \otimes X' \right) \right]_{J,\bullet} \right\|_2 \\ &= \rho\left(\left[\Sigma^{-1} \otimes (X'X) \right]_{J,J} \right)^{1/2} \leq \rho\left(\left[\Sigma^{-1} \otimes (X'X) \right] \right)^{1/2} \leq \lambda_{\max}(\Sigma^{-1})^{1/2} \lambda_{\max}(X'X)^{1/2}, \end{aligned}$$

where the first inequality comes from Theorem 4.3.15 of Horn and Johnson (1986). Hence, by (A5) and (A7),

$$\|A_1\|_2 \leq \frac{1}{nq} \lambda_{\max}(\Sigma^{-1})^{1/2} \lambda_{\max}(X'X)^{1/2} = O_P(q^{-1/2}). \tag{52}$$

By (33), (42) and (52) it is thus enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq nq\, n\, q^{-2c_2} \right) \to 0, \text{ as } n \to \infty.$$

The result follows from the Markov inequality and the first condition of (L).

Let us now study the fourth term in the r.h.s of (37). By (46), for all j in J,

$$\mathbb{P}\left(\left|\left(\left((\widetilde{C}_{J,J})^{-1}-C_{J,J})^{-1}\right)\left(\widetilde{W}_J-W_J\right)\right)_j\right| \geq \frac{\sqrt{nq}}{5}\left(|\mathcal{B}_j|-\frac{\lambda}{2nq}\left|\left((C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right)_j\right|\right)\right)$$
$$\leq \mathbb{P}\left(\left\|(\widetilde{C}_{J,J})^{-1}-C_{J,J})^{-1}\right\|_2 \|A\|_2 \|Z\|_2 \geq \frac{\sqrt{nq}}{5}\left(M_3 q^{-c_2}-\frac{\lambda q|J|}{2nqM_2}\right)\right),$$

where A is defined in (44).

By (33), (42) and (48), it is thus enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq (nq)\, n^2\, q^{1-2c_2}\right) \to 0, \text{ as } n \to \infty.$$

The result follows from the Markov inequality and the fact that $c_2 < 1/2$.

Let us now study $\mathbb{P}(\widetilde{B}_n)$. By definition of $\widetilde{B}_n$, we get that

$$\mathbb{P}(\widetilde{B}_n^c) = \mathbb{P}\left(\left\{\left|\widetilde{C}_{J^c,J}(\widetilde{C}_{J,J})^{-1}\widetilde{W}_J - \widetilde{W}_{J^c}\right| \geq \frac{\lambda}{2\sqrt{nq}}\left(1-|\widetilde{C}_{J^c,J}(\widetilde{C}_{J,J})^{-1}\text{sign}(\mathcal{B}_J)|\right)\right\}\right).$$

Observe that

$$\begin{aligned}
\widetilde{C}_{J^c,J}(\widetilde{C}_{J,J})^{-1}\widetilde{W}_J - \widetilde{W}_{J^c} &= C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c} \\
&+ C_{J^c,J}(C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right) \\
&+ C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)W_J \\
&+ C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)\left(\widetilde{W}_J - W_J\right) \\
&+ \left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)(C_{J,J})^{-1}W_J \\
&+ \left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)(C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right) \\
&+ \left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)W_J \\
&+ \left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)\left(\widetilde{W}_J - W_J\right) \\
&+ W_{J^c} - \widetilde{W}_{J^c}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\widetilde{C}_{J^c,J}(\widetilde{C}_{J,J})^{-1}\text{sign}(\mathcal{B}_J) &= C_{J^c,J}(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J) \\
&+ C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J) \\
&+ \left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J) \\
&+ \left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J).
\end{aligned}$$

By (IC) and the triangle inequality, we obtain that

$$
\begin{aligned}
&\mathbb{P}(\widetilde{B}_n^c) \leq \mathbb{P}\left(\left|C_{J^c,J}(C_{J,J})^{-1}W_J - W_{J^c}\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right)\\
&+\mathbb{P}\left(\left|C_{J^c,J}(C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right)\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right)\\
&+\mathbb{P}\left(\left\{\left|C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)W_J\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\left(\widetilde{W}_J - W_J\right)\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J} - C_{J^c,J}\right)(C_{J,J})^{-1}W_J\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J} - C_{J^c,J}\right)(C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right)\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J} - C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)W_J\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J} - C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\left(\widetilde{W}_J - W_J\right)\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|W_{J^c} - \widetilde{W}_{J^c}\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\mathrm{sign}(\mathcal{B}_J)\right| \geq \frac{\eta}{12}\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J} - C_{J^c,J}\right)(C_{J,J})^{-1}\mathrm{sign}(\mathcal{B}_J)\right| \geq \frac{\eta}{12}\right\}\right)\\
&+\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J} - C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)\mathrm{sign}(\mathcal{B}_J)\right| \geq \frac{\eta}{12}\right\}\right).
\end{aligned}
\tag{53}
$$

The first term in the r.h.s of (53) tends to 0 by (35).

Let us now study the second term of (53). By (46), we get that for all j in J^c,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\left(C_{J^c,J}(C_{J,J})^{-1}\left(\widetilde{W}_J - W_J\right)\right)_j\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right)\\
&\leq \mathbb{P}\left(\left\|C_{J^c,J}\right\|_2 \|(C_{J,J})^{-1}\|_2\|A\|_2\|Z\|_2 \geq \frac{\lambda}{24\sqrt{nq}}\eta\right).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\left\|C_{J^c,J}\right\|_2 &= \rho\left(\frac{(\mathcal{X}_{\bullet,J^c})'\mathcal{X}_{\bullet,J}}{nq}\frac{(\mathcal{X}_{\bullet,J})'\mathcal{X}_{\bullet,J^c}}{nq}\right)^{1/2} = \frac{1}{nq}\left\|(\mathcal{X}_{\bullet,J^c})'\mathcal{X}_{\bullet,J}\right\|_2 \leq \frac{\|(\mathcal{X}_{\bullet,J^c})'\|_2}{\sqrt{nq}}\frac{\|\mathcal{X}_{\bullet,J}\|_2}{\sqrt{nq}}\\
&\leq \rho\left(\frac{(\mathcal{X}_{\bullet,J^c})'\mathcal{X}_{\bullet,J^c}}{nq}\right)^{1/2}\rho\left(\frac{(\mathcal{X}_{\bullet,J})'\mathcal{X}_{\bullet,J}}{nq}\right)^{1/2}\\
&= \rho(C_{J^cJ^c})^{1/2}\rho(C_{J,J})^{1/2} \leq \rho(C) = \frac{\lambda_{\max}(\Sigma^{-1})}{q}\lambda_{\max}(X'X/n) = O_P(q^{-1}).
\end{aligned}
\tag{54}
$$

In (54) the last inequality and the fourth equality come from Theorem 4.3.15 of Horn and Johnson (1986) and (40), respectively. The last equality comes from (A5) and (A7).

By (39), (48) and (54), it is thus enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq \left((nq)^{1/2}\sqrt{q}\frac{\lambda}{\sqrt{nq}}\right)^2\right) = \mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq (nq)\left(\frac{\lambda}{\sqrt{n}}\right)^2\right) \to 0, \text{ as } n \to \infty,$$

which holds true by the second condition of (L) and Markov inequality. Hence, the second term of (53) tends to zero as n tends to infinity.

Let us now study the third term of (53). By (51), we get that for all j in J^c,

$$\mathbb{P}\left(\left|\left(C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\right)W_J\right)_j\right| \geq \frac{\lambda}{24\sqrt{nq}}\eta\right)$$
$$\leq \mathbb{P}\left(\|C_{J^c,J}\|_2 \, \|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2\|A_1\|_2\|Z\|_2 \geq \frac{\lambda}{24\sqrt{nq}}\eta\right).$$

By (42), (52) and (54), it is thus enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq \left((nq)^{1/2}\sqrt{q}\frac{\lambda}{\sqrt{nq}}\right)^2\right) = \mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq (nq)\left(\frac{\lambda}{\sqrt{n}}\right)^2\right) \to 0, \text{ as } n \to \infty,$$

which holds true by the second condition of (L) and Markov inequality. Hence, the third term of (53) tends to zero as n tends to infinity.

Let us now study the fourth term of (53). By (46), it amounts to prove that

$$\mathbb{P}\left(\|C_{J^c,J}\|_2 \, \|(\widetilde{C}_{J,J})^{-1} - (C_{J,J})^{-1}\|_2\|A\|_2\|Z\|_2 \geq \frac{\lambda}{24\sqrt{nq}}\eta\right) \to 0, \text{ as } n \to \infty.$$

By (54), (42) and (48) it is enough tho prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq (nq)\,(nq)\left(\frac{\lambda}{\sqrt{n}}\right)^2\right) \to 0, \text{ as } n \to \infty,$$

which holds true by the second condition of (L). Hence, the fourth term of (53) tends to zero as n tends to infinity.

Let us now study the fifth term of (53). By (51), proving that the fifth term of (53) tends to 0 amounts to proving that

$$\mathbb{P}\left(\left\|C_{J^c,J} - \widetilde{C}_{J^c,J}\right\|_2 \|(C_{J,J})^{-1}\|_2\|A_1\|_2\|Z\|_2 \geq \frac{\lambda}{24\sqrt{nq}}\eta\right) \to 0, \text{ as } n \to \infty.$$

Let us now bound $\|C_{J^c,J} - \widetilde{C}_{J^c,J}\|_2$.

$$\left\|C_{J^c,J} - \widetilde{C}_{J^c,J}\right\|_2 = \left\|\left(C - \widetilde{C}\right)_{J^c,J}\right\|_2 = \rho\left(\left(C - \widetilde{C}\right)_{J^c,J}\left(C - \widetilde{C}\right)_{J^c,J}\right)^{1/2}$$
$$\leq \left\|\left(C - \widetilde{C}\right)_{J^c,J}\left(C - \widetilde{C}\right)_{J^c,J}\right\|_\infty^{1/2} \leq \left\|\left(C - \widetilde{C}\right)\left(C - \widetilde{C}\right)\right\|_\infty^{1/2} \leq \left\|C - \widetilde{C}\right\|_\infty$$
$$= \frac{1}{q}\left\|\Sigma^{-1} - \widehat{\Sigma}^{-1}\right\|_\infty \left\|\frac{X'X}{n}\right\|_\infty = O_P(q^{-1}(nq)^{-1/2}), \tag{55}$$

as n tends to infinity, where the last equality comes from (A5) and (A9).

By (39), (52) and (55), to prove that the fifth term of (53) tends to zero as n tends to infinity, it is enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq nq\left(\frac{\lambda}{\sqrt{n}}\right)^2\right) \to 0, \text{ as } n \to \infty,$$

which holds using Markov's inequality and the second condition of (L).

Using similar arguments as those used for proving that the second, third and fourth terms of (53) tend to zero, we get that the sixth, seventh and eighth terms of (53) tend to zero, as n tends to infinity, by replacing (54) by (55).

Let us now study the ninth term of (53). Replacing J by J^c in (43), (44), (46, (47) and (48) in order to prove that the ninth term of (53) tends to 0 it is enough to prove that

$$\mathbb{P}\left(\sum_{k=1}^{nq} Z_k^2 \geq nq\left(\frac{\lambda}{\sqrt{n}}\right)^2\right) \to 0, \text{ as } n \to \infty,$$

which holds using Markov's inequality and the second condition of (L).

Let us now study the tenth term of (53). Using the same idea as the one used for proving (38), we get that

$$\begin{aligned}&\mathbb{P}\left(\left\{\left|C_{J^c,J}\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J)\right| \geq \frac{\eta}{12}\right\}\right)\\ &\leq \mathbb{P}\left(\sqrt{|J|}\,\|C_{J^c,J}\|_2\,\left\|(\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right\|_2 \geq \frac{\eta}{12}\right),\end{aligned}$$

which tends to zero as n tends to infinity by (A3), (42), (54) and the fact that $c_1 < 1/2$.

Let us now study the eleventh term of (53). Using the same idea as the one used for proving (38), we get that

$$\begin{aligned}&\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)(C_{J,J})^{-1}\text{sign}(\mathcal{B}_J)\right| \geq \frac{\eta}{12}\right\}\right)\\ &\leq \mathbb{P}\left(\sqrt{|J|}\left\|\widetilde{C}_{J^c,J}-C_{J^c,J}\right\|_2\|(C_{J,J})^{-1}\|_2 \geq \frac{\eta}{12}\right),\end{aligned}$$

which tends to zero as n tends to infinity by (A3), (39) and (55) and the fact that $c_1 < 1/2$.

Finally, the twelfth term of (53) can be bounded as follows:

$$\begin{aligned}&\mathbb{P}\left(\left\{\left|\left(\widetilde{C}_{J^c,J}-C_{J^c,J}\right)\left((\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right)\text{sign}(\mathcal{B}_J)\right| \geq \frac{\eta}{12}\right\}\right)\\ &\leq \mathbb{P}\left(\sqrt{|J|}\left\|\widetilde{C}_{J^c,J}-C_{J^c,J}\right\|_2\left\|(\widetilde{C}_{J,J})^{-1}-(C_{J,J})^{-1}\right\|_2 \geq \frac{\eta}{12}\right),\end{aligned}$$

which tends to zero as n tends to infinity by (A3), (42) and (55) and the fact that $c_1 < 1/2$. □

Proof of Proposition 6. Observe that

$$\Sigma^{-1} = \begin{pmatrix} 1 & -\phi_1 & 0 & \cdots & 0 \\ -\phi_1 & 1+\phi_1^2 & -\phi_1 & \cdots & 0 \\ 0 & -\phi_1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1+\phi_1^2 & -\phi_1 \\ 0 & 0 & \cdots & -\phi_1 & 1 \end{pmatrix}. \tag{56}$$

Let $S = \mathcal{X}'\mathcal{X} = \Sigma^{-1} \otimes X'X$. Then,

$$S_{i,j} = \begin{cases} n_{r_i+1} & \text{if } j = i \text{ and } k_i \in \{0, q-1\} \\ (1+\phi_1^2)n_{r_i+1} & \text{if } j = i \text{ and } k_i \notin \{0, q-1\} \\ -\phi_1 n_{r_i+1} & \text{if } j = i+p \text{ or if } j = i-p \\ 0 & \text{otherwise} \end{cases},$$

where $i-1 = (p-1)k_i + r_i$ corresponds to the Euclidean division of $(i-1)$ by $(p-1)$.

In order to prove (IC), it is enough to prove that

$$\|S_{J^c,J}(S_{J,J})^{-1}\|_\infty \leq 1-\eta,$$

where $\eta \in (0,1)$.

Since for all j, $(j-p) \in J^c$ or $(j+p) \in J^c$,

$$\|S_{J^c,J}\|_\infty = \nu|\phi_1|.$$

Let $A = S_{J,J}$. Since $A = (a_{i,j})$ is a diagonally dominant matrix, then, by Theorem 1 of Varah (1975),

$$\|A^{-1}\|_\infty \leq \frac{1}{\min_k(a_{k,k} - \sum_{\substack{1\leq j\leq |J| \\ j\neq k}} a_{k,j})}.$$

Using that for all j, $(j-p) \in J^c$ or $(j+p) \in J^c$,

$$\sum_{\substack{1\leq j\leq |J| \\ j\neq k}} a_{k,j} \leq \nu|\phi_1|.$$

If $k \in J$ then $k > p$ and $k < pq - p$. Thus,

$$a_{k,k} \geq \nu(1+\phi_1^2).$$

Hence,

$$\|A^{-1}\|_\infty \leq \frac{1}{\nu(1+\phi_1^2 - |\phi_1|)}$$

and

$$\|S_{J^c,J}(S_{J,J})^{-1}\|_\infty \leq \|S_{J^c,J}\|_\infty \|(S_{J,J})^{-1}\|_\infty \leq \frac{|\phi_1|}{1+\phi_1^2-|\phi_1|}.$$

Since $|\phi_1| < 1$, the strong Irrepresentability Condition holds when

$$|\phi_1| \leq (1-\eta)(1+|\phi_1|^2 - |\phi_1|),$$

which is true for a small enough η. □

Proof of Proposition 7. Since $|\phi_1| < 1$,

$$\|\Sigma^{-1}\|_\infty \leq |\phi_1| + |1+\phi_1^2| \leq 3,$$

which gives (A7) by Theorem 5.6.9 of Horn and Johnson (1986).

Observe that

$$\|\Sigma\|_\infty \leq \frac{1}{1-\phi_1^2}\left(1 + 2\sum_{h=1}^{q-1}|\phi_1|^h\right) \leq \frac{1}{1-\phi_1^2}\left(1+\frac{2}{1-|\phi_1|}\right) = \frac{3-|\phi_1|}{1-\phi_1^2} \leq \frac{3}{1-\phi_1^2},$$

which gives (A8) by Theorem 5.6.9 of Horn and Johnson (1986).

Since $\widehat{\Sigma}^{-1}$ has the same expression as Σ^{-1} defined in (56) except that ϕ_1 is replaced by $\widehat{\phi_1}$ defined in (21), we get that

$$\left\|\Sigma^{-1} - \widehat{\Sigma}^{-1}\right\|_\infty \leq 2\left|\phi_1 - \widehat{\phi}_1\right| + \left(\phi_1 - \widehat{\phi}_1\right)^2,$$

which implies Assumption (A9) of Theorem 5 by Lemma 8.

Let us now check Assumption (A10) of Theorem 5. Since, by Theorem 5.6.9 of Horn and Johnson (1986), $\rho(\Sigma - \widehat{\Sigma}) \leq \|\Sigma - \widehat{\Sigma}\|_\infty$, it is enough to prove that

$$\left\|\Sigma - \widehat{\Sigma}\right\|_\infty = O_P((nq)^{-1/2}), \text{ as } n \to \infty.$$

Observe that

$$\begin{aligned}
\left\|\Sigma - \widehat{\Sigma}\right\|_\infty &\leq \left|\frac{1}{1-\phi_1^2} - \frac{1}{1-\widehat{\phi}_1^2}\right| + 2\sum_{h=1}^{q-1}\left|\frac{\phi_1^h}{1-\phi_1^2} - \frac{\widehat{\phi}_1^h}{1-\widehat{\phi}_1^2}\right| \\
&\leq \left|\frac{\phi_1^2 - \widehat{\phi}_1^2}{(1-\phi_1^2)(1-\widehat{\phi}_1^2)}\right| + 2\sum_{h=1}^{q-1}\left|\frac{\phi_1^h - \widehat{\phi}_1^h}{1-\phi_1^2}\right| + 2\sum_{h=1}^{q-1}\left|\widehat{\phi}_1^h\left(\frac{1}{1-\phi_1^2} - \frac{1}{1-\widehat{\phi}_1^2}\right)\right| \\
&\leq \left|\frac{(\phi_1 - \widehat{\phi}_1)(\phi_1 + \widehat{\phi}_1)}{(1-\phi_1^2)(1-\widehat{\phi}_1^2)}\right| + 2\sum_{h=1}^{q-1}\left|\frac{\phi_1^h - \widehat{\phi}_1^h}{1-\phi_1^2}\right| + 2\sum_{h=1}^{q-1}\left|\left(\widehat{\phi}_1^h - \phi_1^h\right)\left(\frac{1}{1-\phi_1^2} - \frac{1}{1-\widehat{\phi}_1^2}\right)\right| \\
&\qquad + 2\sum_{h=1}^{q-1}\left|\phi_1^h\left(\frac{1}{1-\phi_1^2} - \frac{1}{1-\widehat{\phi}_1^2}\right)\right| \\
&\leq \left|\frac{(\phi_1 - \widehat{\phi}_1)(\phi_1 + \widehat{\phi}_1)}{(1-\phi_1^2)(1-\widehat{\phi}_1^2)}\right|\left(1 + \frac{2}{1-|\phi_1|}\right) + 2\left(\frac{1}{|1-\phi_1^2|} + \left|\frac{(\phi_1 - \widehat{\phi}_1)(\phi_1 + \widehat{\phi}_1)}{(1-\phi_1^2)(1-\widehat{\phi}_1^2)}\right|\right)\sum_{h=1}^{q-1}\left|\widehat{\phi}_1^h - \phi_1^h\right|.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\sum_{h=1}^{q-1}\left|\widehat{\phi}_1^h - \phi_1^h\right| &\leq \left|\widehat{\phi}_1 - \phi_1\right|\sum_{h=1}^{q-1}\sum_{k=0}^{h-1}|\phi_1|^k|\widehat{\phi}_1|^{h-k-1} \leq \left|\widehat{\phi}_1 - \phi_1\right|\left(\frac{1-|\widehat{\phi}_1|^{q-1}}{1-|\widehat{\phi}_1|}\right)\left(\frac{1-|\phi_1|^{q-1}}{1-|\phi_1|}\right) \\
&\leq \left|\widehat{\phi}_1 - \phi_1\right|\left(\frac{1}{1-|\widehat{\phi}_1|}\right)\left(\frac{1}{1-|\phi_1|}\right).
\end{aligned}$$

Thus, by Lemma 8,

$$\left\|\Sigma - \widehat{\Sigma}\right\|_\infty = O_P((nq)^{-1/2}),$$

which implies Assumption (A10) of Theorem 5. □

Proof of Lemma 8. In the following, for notational simplicity, $q = q_n$. Observe that

$$\sqrt{nq}\widehat{\phi}_1 = \frac{\frac{1}{\sqrt{nq}}\sum_{i=1}^n\sum_{\ell=2}^q \widehat{E}_{i,\ell}\widehat{E}_{i,\ell-1}}{\frac{1}{nq}\sum_{i=1}^n\sum_{\ell=1}^{q-1}\widehat{E}_{i,\ell}^2}.$$

By (19),

$$\begin{aligned}\sum_{i=1}^{n}\sum_{\ell=2}^{q}\widehat{E}_{i,\ell}\widehat{E}_{i,\ell-1} &= \sum_{\ell=2}^{q}(\widehat{E}_{\bullet,\ell})'\widehat{E}_{\bullet,\ell-1} = \sum_{\ell=2}^{q}(\Pi E_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1}) \\ &= \sum_{\ell=2}^{q}(\phi_1 \Pi E_{\bullet,\ell-1} + \Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1}) \qquad (57) \\ &= \phi_1 \sum_{\ell=1}^{q-1}(\Pi E_{\bullet,\ell})'(\Pi E_{\bullet,\ell}) + \sum_{\ell=2}^{q}(\Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1}),\end{aligned}$$

where (57) comes from the definition of $(E_{i,t})$.

Hence,

$$\sqrt{nq}(\widehat{\phi}_1 - \phi_1) = \frac{\frac{1}{\sqrt{nq}}\sum_{\ell=2}^{q}(\Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1})}{\frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}\widehat{E}_{i,\ell}^2}.$$

In order to prove that $\sqrt{nq}(\widehat{\phi}_1 - \phi_1) = O_P(1)$, it is enough to prove that

$$\frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}E_{i,\ell}^2 - \frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}\widehat{E}_{i,\ell}^2 = o_P(1), \text{ as } n \to \infty, \qquad (58)$$

by Lemma 10 and

$$\frac{1}{\sqrt{nq}}\sum_{\ell=2}^{q}(\Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1}) = O_P(1), \text{ as } n \to \infty. \qquad (59)$$

Let us first prove (58). By (20),

$$\widehat{\mathcal{E}} = [\mathrm{Id}_{\mathbb{R}^q} \otimes \Pi]\,\mathcal{E} := A\mathcal{E}.$$

Note that

$$\mathrm{Cov}(\widehat{\mathcal{E}}) = A(\Sigma \otimes \mathrm{Id}_{\mathbb{R}^n})A' = \Sigma \otimes \Pi.$$

Hence, for all i

$$\mathrm{Var}(\widehat{\mathcal{E}}_i) \le \lambda_{\max}(\Sigma).$$

Since the covariance matrix of $\mathcal{E}$ is equal to $\Sigma \otimes \mathrm{Id}_{\mathbb{R}^n}$, for all i

$$\mathrm{Var}(\mathcal{E}_i) \le \lambda_{\max}(\Sigma).$$

By Markov inequality,

$$\begin{aligned}\frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}E_{i,\ell}^2 - \frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}\widehat{E}_{i,\ell}^2 &= \frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q}E_{i,\ell}^2 - \frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\widehat{E}_{i,\ell}^2 + o_P(1) \\ &= \frac{1}{nq}\left(\|\mathcal{E}\|_2^2 - \|\widehat{\mathcal{E}}\|_2^2\right) + o_P(1).\end{aligned}$$

Observe that

$$\begin{aligned}&\|\mathcal{E}\|_2^2 - \|\widehat{\mathcal{E}}\|_2^2 = \|\mathcal{E}\|_2^2 - \|A\mathcal{E}\|_2^2 = \mathcal{E}'\mathcal{E} - \mathcal{E}'A'A\mathcal{E} = \mathcal{E}'\left(\mathrm{Id}_{\mathbb{R}^{nq}} - \mathrm{Id}_{\mathbb{R}^q} \otimes \Pi\right)\mathcal{E} \\ &= \mathcal{E}'\left(\mathrm{Id}_{\mathbb{R}^q} \otimes (\mathrm{Id}_{\mathbb{R}^n} - \Pi)\right)\mathcal{E} = \sum_{i=1}^{pq}\widetilde{\mathcal{E}}_i^2,\end{aligned}$$

where $\widetilde{\mathcal{E}} = O\mathcal{E}$, where O is an orthogonal matrix. Using that

$$\mathbb{E}(\widetilde{\mathcal{E}}_i^2) = \mathrm{Cov}(\widetilde{\mathcal{E}})_{i,i} \leq \lambda_{\max}(\Sigma),$$

and Markov inequality, we get (58).

Let us now prove (59). By definition of $(E_{i,t})$ and since $|\phi_1| < 1$, $\mathbb{E}[(\Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1})] = 0$. Moreover,

$$\mathbb{E}\left[\left(\sum_{\ell=2}^{q}(\Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1})\right)^2\right] = \mathbb{E}\left[\left(\sum_{\ell=2}^{q}\sum_{i=1}^{n}\left(\sum_{k=1}^{n}\Pi_{i,k}Z_{k,\ell}\right)\left(\sum_{j=1}^{n}\Pi_{i,j}E_{j,\ell-1}\right)\right)^2\right]$$

$$= \sum_{2\leq\ell,\ell'\leq q}\ \sum_{1\leq i,j,k,i',j',k'\leq n} \Pi_{i,k}\Pi_{i',k'}\Pi_{i,j}\Pi_{i',j'}\mathbb{E}\left(Z_{k,\ell}Z_{k',\ell'}E_{j,\ell-1}E_{j',\ell'-1}\right)$$

$$= \sum_{2\leq\ell,\ell'\leq q}\ \sum_{1\leq i,j,k,i',j',k'\leq n} \Pi_{i,k}\Pi_{i',k'}\Pi_{i,j}\Pi_{i',j'}\sum_{r,s\geq 0}\phi_1^r\phi_1^s\mathbb{E}\left(Z_{k,\ell}Z_{k',\ell'}Z_{j,\ell-1-r}Z_{j',\ell'-1-s}\right),$$

since the $(E_{i,t})$ are AR(1) processes with $|\phi_1| < 1$. Note that $\mathbb{E}(Z_{k,\ell}Z_{k',\ell'}Z_{j,\ell-1-r}Z_{j',\ell'-1-s}) = 0$ except when $\ell = \ell'$, $k = k'$, $j = j'$ and $r = s$.

Thus,

$$\mathbb{E}\left[\left(\sum_{\ell=2}^{q}(\Pi Z_{\bullet,\ell})'(\Pi E_{\bullet,\ell-1})\right)^2\right] = \sigma^4\left(\sum_{r\geq 0}\phi_1^{2r}\right)\sum_{\ell=2}^{q}\ \sum_{1\leq i,j,k,i'\leq n}\Pi_{i,k}\Pi_{i',k}\Pi_{i,j}\Pi_{i',j}$$

$$= \frac{q\sigma^4}{1-\phi_1^2}\mathrm{Tr}(\Pi) \leq \frac{nq\sigma^4}{1-\phi_1^2},$$

where $\mathrm{Tr}(\Pi)$ denotes the trace of Π, which concludes the proof of (59) by Markov inequality. □

5. Technical lemmas

Lemma 9. *Let $A \in \mathcal{M}_n(\mathbb{R})$ and Π an orthogonal projection matrix. For any j in $\{1,\ldots,n\}$*

$$(A'\Pi A)_{jj} \geq 0.$$

Proof of Lemma 9. Observe that

$$(A'\Pi A) = A'\Pi'\Pi A = (\Pi A)'(\Pi A),$$

since Π is an orthogonal projection matrix. Moreover,

$$(A'\Pi A)_{jj} = e_j'(\Pi A)'(\Pi A)e_j \geq 0,$$

since $(\Pi A)'(\Pi A)$ is a positive semidefinite symmetric matrix, where e_j is a vector containing null entries except the jth entry which is equal to 1. □

Lemma 10. *Assume that $(E_{1,t})_t$, $(E_{2,t})_t$, ..., $(E_{n,t})_t$ are independent AR(1) processes satisfying:*

$$E_{i,t} - \phi_1 E_{i,t-1} = Z_{i,t},\ \forall i \in \{1,\ldots,n\},$$

where the $Z_{i,t}$'s are zero-mean i.i.d. Gaussian random variables with variance σ^2 and $|\phi_1| < 1$. Then,

$$\frac{1}{nq_n}\sum_{i=1}^{n}\sum_{\ell=1}^{q_n-1}E_{i,\ell}^2 \xrightarrow{P} \frac{\sigma^2}{1-\phi_1^2},\ \text{as } n\to\infty.$$

Proof of Lemma 10. In the following, for notational simplicity, $q = q_n$. Since $\mathbb{E}(E_{i,\ell}^2) = \sigma^2/(1-\phi_1^2)$, it is enough to prove that

$$\frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}\left(E_{i,\ell}^2 - \mathbb{E}(E_{i,\ell}^2)\right) \xrightarrow{P} 0, \text{ as } n \to \infty.$$

Since

$$E_{i,\ell}^2 = \left(\sum_{j\geq 0}\phi_1^j Z_{i,\ell-j}\right)^2 = \sum_{j,j'\geq 0}\phi_1^j\phi_1^{j'} Z_{i,\ell-j}Z_{i,\ell-j'},$$

$$\mathrm{Var}\left(\frac{1}{nq}\sum_{i=1}^{n}\sum_{\ell=1}^{q-1}\left(E_{i,\ell}^2 - \mathbb{E}(E_{i,\ell}^2)\right)\right) = \frac{1}{(nq)^2}\sum_{i=1}^{n}\sum_{1\leq \ell,\ell'\leq q-1}\mathrm{Cov}(E_{i,\ell}^2; E_{i,\ell'}^2)$$

$$= \frac{1}{(nq)^2}\sum_{i=1}^{n}\sum_{1\leq \ell,\ell'\leq q-1}\sum_{j,j'\geq 0}\sum_{k,k'\geq 0}\phi_1^j\phi_1^{j'}\phi_1^k\phi_1^{k'}\,\mathrm{Cov}(Z_{i,\ell-j}Z_{i,\ell-j'}; Z_{i,\ell'-k}Z_{i,\ell'-k'}). \quad (60)$$

By Cauchy-Schwarz inequality $|\,\mathrm{Cov}(Z_{i,\ell-j}Z_{i,\ell-j'}; Z_{i,\ell'-k}Z_{i,\ell'-k'})|$ is bounded by a positive constant. Moreover $\sum_{j\geq 0}|\phi_1|^j < \infty$, hence (60) tends to zero as n tend to infinity, which concludes the proof of the lemma. □

References

Alquier, P. and P. Doukhan (2011). Sparsity considerations for dependent variables. *Electron. J. Statist. 5*, 750–774.

Banerjee, O., L. E. Ghaoui, and A. D'aspremont (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research 9*, 485–516.

Brockwell, P. J. and R. A. Davis (1990). *Time Series: Theory and Methods.* New York, NY, USA: Springer-Verlag New York, Inc.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432.

Horn, R. A. and C. R. Johnson (1986). *Matrix Analysis.* New York, NY, USA: Cambridge University Press.

Lee, W. and Y. Liu (2012). Simultaneous Multiple Response Regression and Inverse Covariance Matrix Estimation via Penalized Gaussian Maximum Likelihood. *J. Multivar. Anal. 111*, 241–255.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Berlin: Springer.

Mehmood, T., K. H. Liland, L. Snipen, and S. Saebo (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems 118*, 62 – 69.

Meng, C., B. Kuster, A. C. Culhane, and A. M. Gholami (2014). A multivariate approach to the integration of multi-omics datasets. *BMC Bioinformatics 15*(1), 162.

Perrot-Dockès, M., C. Lévy-Leduc, J. Chiquet, L. Sansonnet, M. Brégère, M. P. Étienne, S. Robin, and G. Genta-Jouve (2017). A multivariate variable selection approach for analyzing LC-MS metabolomics data. arXiv:1704.00076.

Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation.* Wiley Series in Probability and Statistics.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electron. J. Statist. 2*, 494–515.

Rothman, A. J., E. Levina, and J. Zhu (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics 19*(4), 947–962.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc B. 58*(1), 267–288.

Varah, J. (1975). A lower bound for the smallest singular value of a matrix. *Linear Algebra and its Applications 11*(1), 3 – 5.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the gaussian graphical model. *Biometrika 94*(1), 19–35.

Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *Journal of Machine Learning Research 7*, 2541–2563.

UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France
E-mail address: marie.perrot-dockes@agroparistech.fr

UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France
E-mail address: celine.levy-leduc@agroparistech.fr

UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France
E-mail address: laure.sansonnet@agroparistech.fr

UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France
E-mail address: julien.chiquet@agroparistech.fr