Technical appendix to “Adaptive estimation of covariance matrices via Cholesky decomposition”

Nicolas Verzelen

Electronic Journal of Statistics
ISSN: 1935-7524

Technical appendix to "Adaptive estimation of covariance matrices via Cholesky decomposition"

Nicolas Verzelen[*]

INRA, UMR 729 MISTEA,
F-34060 Montpellier, France

SUPAGRO, UMR 729 MISTEA,
F-34060 Montpellier, France
e-mail: nicolas.verzelen@supagro.inra.fr

Abstract:
This is a technical appendix to "Adaptive estimation of covariance matrices via Cholesky decomposition (arXiv:1010.1445).

AMS 2000 subject classifications: Primary 62H12; secondary 62F35, 62J05.
Keywords and phrases: Covariance matrix, banding, Cholesky decomposition, directed graphical models, penalized criterion, minimax rate of estimation.

For the sake of clarity, we emphasize the references to [7] in bold . For instance Eq.(**12**) refers to Eq.(12) in [7], while Eq.(12) stands for Eq.(12) in this paper.

Contents

[*]Research mostly carried out at Univ Paris-Sud (Laboratoire de Matématiques, CNRS-UMR 8628)

1. Additional simulations

We provide here additional simulation results for the complete graph selection problem. We use the same notations and compute the same estimators as in Section **8.2**.

In the third simulation scheme, we consider the case where the "good" ordering is completely unknown. We first sample a precision matrix Ω_1^c according to the first simulation scheme. Then, we sample uniformly a permutation of $\{1,\ldots,p\}$ and reorder the variables according to this permutation to get the precision matrix $\mathbf{\Omega_3^c}$. Its Cholesky factor is generally far less sparse than the Cholesky factor of Ω_1, while Ω_1^c is as sparse as Ω_3^c. The purpose of this scheme is to illustrate the limits of procedures based on the Cholesky factor when no suitable ordering is known. As in the second scheme, we only perform the simulations for $p = 200$, Esp$= 1, 3, 5$, and $n = 100$.

Method	Ledoit	GLasso	Lasso	ChoSelect$^{\text{f}}$
	Kullback discrepancy $\mathcal{K}(\Omega;\widehat{\Omega})$			
Esp=1	20.1 ± 0.2	9.1 ± 0.2	8.2 ± 0.1	7.6 ± 0.1
Esp=3	42.8 ± 1.8	22.0 ± 0.2	24.0 ± 0.4	25.3 ± 0.5
Esp=5	52.6 ± 1.0	35.8 ± 0.3	42.7 ± 0.4	49.1 ± 0.5
	Operator distance $\|\widehat{\Omega}-\Omega\|$			
Esp=1	6.3 ± 0.1	5.6 ± 0.1	4.7 ± 0.1	4.5 ± 0.2
Esp=3	10.0 ± 0.1	10.1 ± 0.1	8.8 ± 0.2	8.0 ± 0.2
Esp=5	14.3 ± 0.2	15.3 ± 0.2	13.0 ± 0.2	12.6 ± 0.2
	Operator distance $\|\widehat{\Omega}^{-1}-\Sigma\|$			
Esp=1	2.7 ± 0.1	1.7 ± 0.1	2.1 ± 0.1	1.7 ± 0.1
Esp=3	8.4 ± 0.5	6.4 ± 0.3	11.5 ± 0.8	10.8 ± 0.8
Esp=5	16.3 ± 0.8	14.7 ± 0.8	26.5 ± 1.6	25.0 ± 1.4

Table 1: Comparison between the procedures for the third covariance model Ω_3^c with $p = 200$.

We observe different results for the Kullback risk depending on the sparsity. When Esp=1, ChoSelect$^{\text{f}}$ still performs better than the other methods. However, the GLasso provides better results than the two other results for Esp=3 and Esp=5. This is not really surprising since the Glasso has been introduced to handle the "unordered" situation. It seems from the case Esp=5, that the Lasso procedure is more robust to a "bad" ordering than ChoSelect$^{\text{f}}$. ChoSelect$^{\text{f}}$ still performs better than the other procedures in terms of the operator distance between precision matrices. Nevertheless, the differences of performance are less obvious than in the previous schemes. Finally, the Glasso and Ledoit and Wolf's method exhibit a smaller operator distance between covariance matrices than the Lasso and ChoSelect$^{\text{f}}$.

2. Proof of the risk upper bounds

Lemma 2.1. *Let V be a χ^2 random variable with $N > 2$ degrees of freedom and let k be some positive integer such that $N > 2k$, then*

$$\mathbb{E}\left[\frac{1}{V^k}\right] = \frac{1}{(N-2)\ldots(N-2k)} \qquad \textit{and} \qquad \mathbb{E}\left[V^k\right] = N(N+2)\ldots(N+2(k-1)) \ .$$

We refer to Lemma 5 in [1] for the proof of slightly more general version of this lemma.

2.1. Proof of Lemma 10.4

Using expression (**31**) of $\mathcal{K}\left[t,s;\widehat{t}_m,\widehat{s}_m\right]$, we derive

$$\begin{aligned} 2(1-\kappa_0)\mathcal{K}\left(t,s;\widetilde{t},\widetilde{s}\right) &= 2\mathcal{K}\left[t,s;\widehat{t}_m,\widehat{s}_m\right]+(1-\kappa_0)\log\left(\frac{\widetilde{s}}{\widehat{s}_m}\right)+(1-\kappa_0)\frac{s+l(\widetilde{t},t)}{\widetilde{s}}\\ &- \frac{s+l(\widehat{t}_m,t)}{\widehat{s}_m}+\kappa_0+\kappa_0\log\left(\frac{s}{\widehat{s}_m}\right)\ . \end{aligned}$$

By definition of $\widehat{m}$, $\log\left(\widetilde{s}/\widehat{s}_m\right)\leq pen(m)-pen(\widehat{m})$. Hence,

$$\begin{aligned} 2(1-\kappa_0)\mathcal{K}\left(t,s;\widetilde{t},\widetilde{s}\right) &\leq 2\mathcal{K}\left(t,s;\widehat{t}_m,\widehat{s}_m\right)+(1-\kappa_0)\left[pen(m)-pen(\widehat{m})\right]\\ &+ \kappa_0\frac{s}{\widehat{s}_m}+\kappa_0\left[-\frac{s}{\widehat{s}_m}+1+\log\left(\frac{s}{\widehat{s}_m}\right)\right]-\frac{s+l(\widehat{t}_m,t)-\|\Pi_m^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2}{\widehat{s}_m}\\ &+ \frac{l(\widetilde{t},t)(1-\kappa_0)+s(1-\kappa_0)-\|\Pi_{\widehat{m}}^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{\widetilde{s}}\ , \end{aligned}$$

since $\widehat{s}_m=\|\Pi_m^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2$ and $\widetilde{s}=\|\Pi_{\widehat{m}}^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2$. As the function $x-\log x-1$ is non-negative, the term $\left[-s/\widehat{s}_m+1+\log\left(s/\widehat{s}_m\right)\right]$ is non-positive. Since $X_{<i}t_m^*$ is the best predictor of X_i given X_m, it follows that $l(\widehat{t}_m,t)=l(\widehat{t}_m,t_m)+l(t_m,t)$. Hence,

$$\kappa_0\frac{s}{\widehat{s}_m}-\frac{s+l(\widehat{t}_m,t)-\|\Pi_m^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2}{\widehat{s}_m}\leq-(1-\kappa_0)\frac{s}{\widehat{s}_m}+\frac{\|\Pi_m^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2-l(t_m,t)}{\widehat{s}_m}\ .$$

In the proof of Lemma 7.5 in [8], we state that

$$l(\widehat{t}_{m'},t_{m'})\leq\varphi_{\max}\left[n\mathbf{Z}_{m'}^*\mathbf{Z}_{m'})^{-1}\right]\|\Pi_{m'}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{m'})\|_n^2\ .$$

This yields

$$\begin{aligned} (1-\kappa_0)\frac{l(\widetilde{t},t)+s}{\widetilde{s}} &\leq (1-\kappa_0)\frac{s+l(t_{\widehat{m}},t)+\kappa_2\varphi_{\max}\left[n(\mathbf{Z}_{\widehat{m}}^*\mathbf{Z}_{\widehat{m}})^{-1}\right]\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{\widetilde{s}}\\ &+ (1-\kappa_0)(1-\kappa_2)\frac{l(\widetilde{t},t_{\widehat{m}})}{\widetilde{s}}\ . \end{aligned}$$

Let us gather all these bounds

$$\begin{aligned} 2(1-\kappa_0)\mathcal{K}\left[t,s;\widetilde{t},\widetilde{s}\right] &\leq 2\mathcal{K}\left[t,s;\widehat{t}_m,\widehat{s}_m\right]+(1-\kappa_0)\left[pen(m)-pen(\widehat{m})\right]\\ &+ (1-\kappa_0)\frac{l(t_{\widehat{m}},t)+(1-\kappa_2)l(\widetilde{t},t_{\widehat{m}})+\kappa_2\varphi_{\max}\left[n(\mathbf{Z}_{\widehat{m}}^*\mathbf{Z}_{\widehat{m}})^{-1}\right]\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{\widetilde{s}}\\ &- \frac{\|\Pi_{\widehat{m}}^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{\widetilde{s}}+s(1-\kappa_0)\left(\frac{1}{\widetilde{s}}-\frac{1}{\widehat{s}_m}\right)+\frac{\|\Pi_m^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2-l(t_m,t)}{\widehat{s}_m}\\ &\leq 2\mathcal{K}\left[t,s;\widehat{t}_m,\widehat{s}_m\right]+(1-\kappa_0)\left[pen(m)-pen(\widehat{m})\right]\\ &+ (1-\kappa_0)\frac{l(t_{\widehat{m}},t)+(1-\kappa_2)l(\widetilde{t},t_{\widehat{m}})+\kappa_2\varphi_{\max}\left[n(\mathbf{Z}_{\widehat{m}}^*\mathbf{Z}_{\widehat{m}})^{-1}\right]\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{\widetilde{s}}\\ &+ \left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)\left(\frac{1}{\widehat{s}_m}-\frac{1}{\widetilde{s}}\right)+\frac{\|\Pi_{\widehat{m}}\boldsymbol{\epsilon}\|_n^2}{\widetilde{s}}+2\frac{\langle\Pi_{\widehat{m}}^{\perp}\boldsymbol{\epsilon},\Pi_{\widehat{m}}^{\perp}\boldsymbol{\epsilon}_{\widehat{m}}\rangle_n}{\widetilde{s}}\\ &- \frac{\|\Pi_{\widehat{m}}^{\perp}\boldsymbol{\epsilon}_{\widehat{m}}\|_n^2}{\widetilde{s}}+2\frac{\langle\Pi_m^{\perp}\boldsymbol{\epsilon},\Pi_m^{\perp}\boldsymbol{\epsilon}_m\rangle_n}{\widehat{s}_m}+\frac{\|\Pi_m^{\perp}\boldsymbol{\epsilon}_m\|_n^2-l(t_m,t)}{\widehat{s}_m}\ . \end{aligned}$$

We then use Condition (**39**) on $pen(\widehat{m})$ and we apply the inequality

$$2\frac{\langle \Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}, \Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}_{\widehat{m}}\rangle_n}{\widetilde{s}} \leq \kappa_1 \frac{l(t_{\widehat{m}},t)}{\widetilde{s}} + \kappa_1^{-1}\frac{s}{\widetilde{s}}\frac{\langle \Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}, \Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}_{\widehat{m}}\rangle_n^2}{sl(t_{\widehat{m}},t)} \ .$$

Hence, we conclude that

$$\begin{aligned} 2(1-\kappa_0)\mathcal{K}\left[t,s;\widetilde{t},\widetilde{s}\right] &\leq 2\mathcal{K}\left[t,s;\widehat{t}_m,\widehat{s}_m\right] + (1-\kappa_0)pen(m) \\ &+ \frac{l(\widetilde{t},t)}{\widetilde{s}}\left[R_1(\widehat{m})\vee(1-\kappa_2)(1-\kappa_0)\right] + R_2(m) + \frac{s}{\widetilde{s}}R_3(\widehat{m}) + R_4(m,\widehat{m}) \ . \end{aligned}$$

2.2. Proof of Lemma **10.5**

This proof follows the same sketch as the proof of Lemma 7.10 in [8]. The main difference lies in the fact that κ_0 is zero in [8]. Let x be a positive number that we shall fix later. For any $k>0$, let us define

$$\delta_k := \sqrt{\frac{\pi}{2k}} + \exp(-k/16) \ .$$

We shall first control the deviations of the random variables involved in $R_1(\widehat{m})$. Applying deviation inequality for χ^2 random variables and largest values of Standard Wishart matrices (see e.g. Lemmas 7.2, 7.3, and 7.4 in [8]) to all models $m\in\mathcal{M}$ ensures that there exists an event $\mathbb{B}_2$ such that $\mathbb{P}(\mathbb{B}_2^c)\leq 4n\exp(-nx)$ and under $\mathbb{B}_2$ it holds that

$$\begin{aligned} \frac{\|\Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}_{\widehat{m}}\|_n^2}{l(t_{\widehat{m}},t)} &\geq \frac{n-|\widehat{m}|}{n}\left[\left(1-\delta_{n-|\widehat{m}|}-\sqrt{\frac{2|\widehat{m}|H(|\widehat{m}|)}{n-|\widehat{m}|}}-\sqrt{\frac{2xn}{n-|\widehat{m}|}}\right)\vee 0\right]^2 \ , \\ \frac{\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{s+l(t_{\widehat{m}},t)} &\leq \frac{2|\widehat{m}|}{n}\left[1+\sqrt{H(|\widehat{m}|)}+H(|\widehat{m}|)\right]+3x \ , \qquad (1) \\ \frac{\|\Pi^{\perp}_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{s+l(t_{\widehat{m}},t)} &\geq \frac{n-|\widehat{m}|}{n}\left[\left(1-\delta_{n-|\widehat{m}|}-\sqrt{\frac{2|\widehat{m}|H(|\widehat{m}|)}{n-|\widehat{m}|}}-\sqrt{\frac{2xn}{n-|\widehat{m}|}}\right)\vee 0\right]^2 \ , \\ n\varphi_{\max}\left[(\mathbf{Z}^*_{\widehat{m}}\mathbf{Z}_{\widehat{m}})^{-1}\right] &\leq \left[\left(1-\left(1+\sqrt{2H(|\widehat{m}|)}\right)\sqrt{\frac{|\widehat{m}|}{n}}-\sqrt{2x}\right)\vee 0\right]^{-2} \ . \end{aligned}$$

By Assumption $(\mathbb{H}^i_{K,\eta})$, the expression $(1+\sqrt{2H(\widehat{m})})\sqrt{|\widehat{m}|/n}$ is bounded by $\sqrt{\eta}$. Moreover, $(\mathbb{H}^i_{K,\eta})$ also ensures that $|\widehat{m}|\leq n/2$. Hence $\delta_{n-|\widehat{m}|}\leq\delta_{n/2}\leq\nu(K)$ for n larger than some quantity $n_0(K)$. Since $\nu(K)\leq 1-\sqrt{\eta}$, we derive that

$$\begin{aligned} \frac{\|\Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}_{\widehat{m}}\|_n^2}{l(t_{\widehat{m}},t)} &\geq \left(1-\frac{|\widehat{m}|}{n}\right)\left[1-\nu(K)-\sqrt{\eta}\right]^2-2\sqrt{2x} \ , \qquad (2) \\ \frac{\|\Pi^{\perp}_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{s+l(t_{\widehat{m}},t)} &\geq \left(1-\frac{|\widehat{m}|}{n}\right)\left[1-\nu(K)-\sqrt{\eta}\right]^2-2\sqrt{2x} \ , \qquad (3) \\ \varphi_{\max}\left[n\left(\mathbf{Z}^*_{\widehat{m}}\mathbf{Z}_{\widehat{m}}\right)^{-1}\right] &\leq \left[\left(1-\sqrt{\eta}-\sqrt{2x}\right)\vee 0\right]^{-2} \ . \end{aligned}$$

Constraining x to be smaller than $\left(1-\sqrt{\eta}\right)^2/8$ ensures that

$$\kappa_2\varphi_{\max}\left[n\left(\mathbf{Z}^*_{\widehat{m}}\mathbf{Z}_{\widehat{m}}\right)^{-1}\right]\mathbf{1}_{\mathbb{B}_1} \leq \frac{(K-1)(1-\sqrt{\eta}-\nu(K))^2}{4} \ . \qquad (4)$$

Gathering the definition of $R_1(.)$, the inequalities (1), (2), (3), and (4), we get

$$R_1(\widehat{m}) \quad \leq \quad \kappa_1 + 1 - [1-\sqrt{\eta}-\nu(K)]^2 + \frac{|\widehat{m}|}{n}(1-\sqrt{\eta}-\nu(K))^2 U_1 + \sqrt{x}U_2 + xU_3 \ ,$$

where U_1, U_2, and U_3 are respectively defined by

$$\left\{\begin{array}{lcl} U_1 & := & -(1-\kappa_0)K\left[1+\sqrt{2H(|\widehat{m}|)}\right]^2 + 1 + (1-\kappa_0)(K-1)/2\left[1+\sqrt{H(|\widehat{m}|)}\right]^2 \leq 0 \\ U_2 & := & 2\sqrt{2}\,[1+K\eta] \\ U_3 & := & \frac{3}{4}(K-1)\left[1-\sqrt{\eta}-\nu(K)\right]^2 \ . \end{array}\right.$$

Since U_1 is non-positive, we obtain an upper bound of $R_1(\widehat{m})$ that does not depend anymore on the model $\widehat{m}$. By assumption $(\mathbb{H}^i_{K,\eta})$, we know that $\eta < (1-\nu(K)-(3/(K+2))^{1/6})^2$. Hence, coming back to the definition of κ_1 allows to prove that κ_1 is strictly smaller than $[1-\sqrt{\eta}-\nu(K)]^2$. Setting

$$x := \left[\frac{\left[1-\sqrt{\eta}-\nu(K)\right]^2-\kappa_1}{4U_2}\right]^2 \wedge \frac{\left[1-\sqrt{\eta}-\nu(K)\right]^2-\kappa_1}{4U_3} \wedge \frac{\left(1-\sqrt{\eta}\right)^2}{8} \ ,$$

we get

$$R_1(\widehat{m}) \leq 1 - \frac{1}{2}\left[(1-\sqrt{\eta}-\nu(K))^2 - \kappa_1\right] < 1 \ ,$$

under the event $\mathbb{B}_2$.

This is enough to prove Lemma **10.5** if we take $\mathbb{B}_1 = \mathbb{B}_2$. In fact, we shall define an event $\mathbb{B}_1$ slightly more restrictive in order to simplify the proof of Lemma **10.6**. Let $\mathbb{B}_3$ be the event defined by

$$\|\boldsymbol{\epsilon}\|_n^2/s \leq \kappa_1^{-1} \ . \tag{5}$$

Since κ_1^{-1} is strictly larger than one and since κ_1 only depends on K and η, it follows that $\mathbb{P}(\mathbb{B}_3^c) \leq \exp(-nL_{K,\eta})$ with $L_{K,\eta} > 0$. Finally we take, $\mathbb{B}_1 := \mathbb{B}_2 \cap \mathbb{B}_3$.

2.3. Proof of Lemma 10.6

The sketch of this proof is similar to the proof of Lemma 7.11 in [8]. First, under the event $\mathbb{B}_1$, it holds that

$$\frac{\|\Pi^{\perp}_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{s+l(t_{\widehat{m}},t)} \geq \left[1-\nu(K)-\sqrt{\eta}\right]^2/4 > 0 \ ,$$

$$\kappa_2\varphi_{\max}\left[n\left(\mathbf{Z}^*_{\widehat{m}}\mathbf{Z}_{\widehat{m}}\right)^{-1}\right] \leq \frac{(K-1)(1-\sqrt{\eta}-\nu(K))^2}{4} \ .$$

This is a consequence of (3), of the choice of x in the previous proof, and of the assumption $(\mathbb{H}^i_{K,\eta})$. Since $\widetilde{s} = |\Pi^{\perp}_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2$, it follows that $s/\widetilde{s}$ is upper bounded under the event $\mathbb{B}_1$. Hence, we only have to upper bound the expectation of $R_3(\widehat{m})$ on $\mathbb{B}_1$.

$$\mathbb{E}\left[\frac{s}{\widetilde{s}}R_3(\widehat{m}\mathbf{1}_{\mathbb{B}_1})\right] \leq L_{K,\eta}\mathbb{E}\left[R_3(\widehat{m}\mathbf{1}_{\mathbb{B}_1})\right] \ . \tag{6}$$

Let us consider the random variables $E_{\widehat{m}}$ defined by

$$E_{\widehat{m}} := \kappa_1^{-1}\frac{\langle\Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon},\Pi^{\perp}_{\widehat{m}}\boldsymbol{\epsilon}_{\widehat{m}}\rangle_n^2}{sl(t_{\widehat{m}},t)} + \frac{\|\Pi_{\widehat{m}}\boldsymbol{\epsilon}\|_n^2}{s} \ .$$

By (5), the random variable $E_{\widehat{m}}$ is upper bounded under $\mathbb{B}_1$ by

$$E_{\widehat{m}} \leq \kappa_1^{-2}\frac{\langle \Pi_{\widehat{m}}^{\perp}\boldsymbol{\epsilon}/\|\Pi_{\widehat{m}}^{\perp}\boldsymbol{\epsilon}\|_n, \Pi_{\widehat{m}}^{\perp}\boldsymbol{\epsilon}_{\widehat{m}}\rangle_n^2}{sl(t_{\widehat{m}},t)} + \frac{\|\Pi_{\widehat{m}}\boldsymbol{\epsilon}\|_n^2}{s}$$

This upper bound follows the distribution of a linear combinations of χ^2 random variables. More details about this observation are given in the proof of Lemma 7.7 in [8]. We shall simultaneously control the deviations of the random variables $E_{\widehat{m}}$, $\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2/[l(t_{\widehat{m}},t)+s]$, and $\|\Pi_{\widehat{m}}^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2/[s+l(t_{\widehat{m}},t)]$ by applying Lemma 1 in [5] and Lemmas 7.2 and 7.3 in [8]. For any $x>0$, we define an event $\mathbb{F}(x)$ such that conditionally on $\mathbb{F}(x)\cap\mathbb{B}_1$,

$$\left\{\begin{array}{ccc} E_{\widehat{m}} & \leq & \frac{|\widehat{m}|+\kappa_1^{-2}}{n} + \frac{2}{n}\sqrt{\left[|\widehat{m}|+\kappa_1^{-4}\right]\left[|\widehat{m}|(\xi+H(|\widehat{m}|))+x\right]} \\ & + & 2\kappa_1^{-2}[\xi(|\widehat{m}|+H(|\widehat{m}|))+x]/n\ , \\ \frac{\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{s+l(t_{\widehat{m}},t)} & \leq & \frac{1}{n}\left[|\widehat{m}|+2\sqrt{|\widehat{m}|\left[|\widehat{m}|(1/16+H(|\widehat{m}|))+x\right]}+2\left[|\widehat{m}|(1/16+H(|\widehat{m}|))+x\right]\right]\ , \\ \frac{\|\Pi_{\widehat{m}}^{\perp}\epsilon_{\widehat{m}}+\epsilon\|_n^2}{s+l(t_{\widehat{m}},t)} & \geq & \frac{n-|\widehat{m}|}{n}\left[\left(1-\delta_{n-|\widehat{m}|}-\sqrt{\frac{|\widehat{m}|(1+2H(|\widehat{m}|))}{n-|\widehat{m}|}}-\sqrt{\frac{2x}{n-|\widehat{m}|}}\right)\vee 0\right]^2\ , \end{array}\right.$$

where δ_k is defined in the previous proof. Then, the probability of $\mathbb{F}(x)$ satisfies

$$\begin{aligned} \mathbb{P}\left[\mathbb{F}(x)^c\right] &\leq e^{-x}\left[\sum_{m\in\mathcal{M}}\exp\left[-|\widehat{m}|H(|\widehat{m}|)\right]\left(e^{-\xi|\widehat{m}|}+e^{-\frac{|\widehat{m}|}{16}}+e^{-\frac{|\widehat{m}|}{2}}\right)\right] \\ &\leq e^{-x}\left(\frac{1}{1-e^{-\xi}}+\frac{1}{1-e^{-1/16}}+\frac{1}{1-e^{-1/2}}\right)\ . \end{aligned}$$

Let us expand the three deviation bounds thanks to the inequality $2ab\leq \tau a^2+\tau^{-1}b^2$:

$$\begin{aligned} E_{\widehat{m}} &\leq \frac{|\widehat{m}|}{n}\left[1+2\sqrt{\xi}+2\kappa_1^{-2}\xi+\tau_1\xi+\tau_2\right]+\frac{x}{n}\left[2\kappa_1^{-2}+\tau_2^{-1}+\tau_1\right] \\ &+ \frac{\kappa_1^{-2}}{n}\left[1+\tau_1^{-1}\kappa_1^{-2}\right]+\frac{|\widehat{m}|H(|\widehat{m}|)}{n}\left[2\kappa_1^{-2}+\tau_1\right]+2\frac{|\widehat{m}|\sqrt{H(|\widehat{m}|)}}{n} \\ &\leq \frac{|\widehat{m}|}{n}\left(1+\sqrt{2H(|\widehat{m}|)}\right)^2\left[\kappa_1^{-2}+2\sqrt{\xi}+2\kappa_1^{-2}\xi+\tau_1\xi+\tau_2\right] \\ &+ \frac{x}{n}\left[2\kappa_1^{-2}+\tau_2^{-1}+\tau_1\right]+\frac{\kappa_1^{-2}}{n}\left[1+\tau_1^{-1}\kappa_1^{-2}\right]\ . \end{aligned}$$

Similarly, we get

$$\frac{\|\Pi_{\widehat{m}}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{l(t_{\widehat{m}},t)+s} \leq 2\frac{|\widehat{m}|}{n}\left[1+\sqrt{2H(|\widehat{m}|)}\right]^2+5\frac{x}{n}\ .$$

We recall that $|m|\leq n/2$ by Assumption $(\mathbb{H}_{K,\eta}^i)$. If n is larger than some quantity $n_0(K)$, then $\delta_{n-|m|}\leq\delta_{n/2}\leq\nu(K)$. Applying again Assumption $(\mathbb{H}_{K,\eta}^i)$, we get

$$\begin{aligned} -K&\frac{|\widehat{m}|}{n-|\widehat{m}|}\left(1+\sqrt{2H(|\widehat{m}|)}\right)^2\frac{\|\Pi_{\widehat{m}}^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_{\widehat{m}})\|_n^2}{l(t_{\widehat{m}},t)+s} \\ &\leq -K\frac{|\widehat{m}|}{n}\left(1+\sqrt{2H(|\widehat{m}|)}\right)^2\left[\left(1-\sqrt{\eta}-\nu(K)-\sqrt{\frac{2x}{n-|\widehat{m}|}}\right)\vee 0\right]^2 \\ &\leq -K\frac{|\widehat{m}|}{n}\left(1+\sqrt{2H(|\widehat{m}|)}\right)^2\left[(1-\sqrt{\eta}-\nu(K))^2-\tau_3\right]+2K\eta\tau_3^{-1}\frac{x}{n}\ . \end{aligned}$$

Let us combine these three bounds with the definitions of $R_3(\widehat{m})$, κ_1, and κ_2, and the bound (4). Hence, under the event $\mathbb{B}_1 \cap \mathbb{F}(x)$, it holds that

$$R_3(\widehat{m}) \leq \frac{|\widehat{m}|}{n}\left[1+\sqrt{2H(|\widehat{m}|)}\right]^2 U_1 + \frac{x}{n}U_2 + \frac{L_{K,\eta}}{n}U_3 \ ,$$

where

$$\left\{\begin{array}{lll} U_1 & := & -\frac{K-1}{10}\left(1-\sqrt{\eta}-\nu(K)\right)^2 + K\tau_3 + 2\sqrt{\xi} + 2\kappa_1^{-2}\xi + \tau_1\xi + \tau_2 \ , \\ U_2 & := & \tau_2^{-1} + \tau_1 + L_{K,\eta}(1+\tau_3^{-1}) \ , \\ U_3 & := & 1+\tau_1^{-1} \ . \end{array}\right.$$

Since $K > 1$, there exists a suitable choice of the constants ξ, τ_1, and τ_2, only depending on K and η that constrains U_1 to be non positive. Hence, under the event $\mathbb{B}_1 \cap \mathbb{F}(x)$,

$$B_{\widehat{m}} \leq \frac{L_{K,\eta}}{n} + L'(K,\eta)\frac{x}{n} \ .$$

Since $\mathbb{P}\left[\mathbb{F}(x)^c\right] \leq e^{-x}L_{K,\eta}$, we conclude by integrating the last expression with respect to x.

2.4. Proof of Lemma **10.7**

Let us assume that $n \geq 17$.

$$R_2(m) = 1 - \frac{l(t_m,t) + \|\Pi_m^{\perp}\boldsymbol{\epsilon}\|_n^2}{\|\Pi_m^{\perp}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m\|_n^2} \ .$$

We shall first upper bound the expectation of $R_2(m)$.

$$\mathbb{E}\left[\frac{l(t_m,t)}{\|\Pi_m^{\perp}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m\|_n^2}\right] = \frac{n}{n-|m|-2}\frac{l(t_m,t)}{[s+l(t_m,t)]} \geq \frac{n-|m|}{n-|m|-2}\frac{l(t_m,t)}{[s+l(t_m,t)]} \tag{7}$$

Let us to the second term.

$$\mathbb{E}\left[\frac{\|\Pi_m^{\perp}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m\|_n^2}{\|\Pi_m^{\perp}\boldsymbol{\epsilon}\|_n^2}\right] = 1 + \frac{l(t_m,t)}{s}\frac{n-|m|}{n-|m|-2} \leq \frac{s+l(t_m,t)}{s}\frac{n-|m|}{n-|m|-2}$$

Applying a convexity argument, we derive that

$$\mathbb{E}\left[\frac{\|\Pi_m^{\perp}\boldsymbol{\epsilon}\|_n^2}{\|\Pi_m^{\perp}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m\|_n^2}\right] \geq \frac{s}{s+l(t_m,t)}\frac{n-|m|-2}{n-|m|} \ . \tag{8}$$

Gathering the inequalities (7) and (8) allows to conclude that

$$\mathbb{E}\left[R_2(m)\right] \leq \frac{2}{n-|m|} \ .$$

We get the upper bound

$$\begin{aligned} \mathbb{E}\left[R_2(m)\mathbf{1}_{\mathbb{B}_1}\right] &= \mathbb{E}\left[R_2(m)\right] - \mathbb{E}\left[R_2(m)\mathbf{1}_{\mathbb{B}_1^c}\right] \\ &\leq \frac{2}{n-|m|} + \sqrt{\mathbb{P}(\mathbb{B}_1^c)}\sqrt{\mathbb{E}\left[R_2^2(m)\right]} \ . \end{aligned} \tag{9}$$

Thus, we need to upper bound the second moment of $R_2(m)$.

$$\begin{aligned}\mathbb{E}\left[R_2^2(m)\right] &\leq 3\left\{1+\mathbb{E}\left[\frac{l^2(t_m,t)}{\|\Pi_m^\perp\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m\|_n^4}\right]+\sqrt{\mathbb{E}\left[\|\Pi_m^\perp\boldsymbol{\epsilon}\|_n^8\right]\mathbb{E}\left[\frac{1}{\|\Pi_m^\perp\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m\|_n^8}\right]}\right\}\\ &\leq 3\left[1+\left(\frac{l(t_m,t)n}{(s+l(t,m,t))(n-|m|-4)}\right)^2+\left(\frac{s(n-|m|+6)}{(s+l(t,m,t))(n-|m|-8)}\right)^2\right] ,\end{aligned}$$

by Lemma 2.1. By Assumption $(\mathbb{H}^i_{K,\eta})$, we have $|m|\leq n/2$. If $n\geq 32$, we can upper bound $\mathbb{E}\left[R_2^2(m)\right]$ by a universal constant. Combining this result with (9) enables to conclude that $\mathbb{E}\left[R_2(m)\mathbf{1}_{\mathbb{B}_1}\right]\leq L_{K,\eta}/n$.

2.5. Proof of Lemma 10.8

We bound the quantity $R_4(m,\widehat{m})$ using the same arguments as in the proof of Theorem 3 in [1]. We first split this quantity into a sum of two terms:

$$\begin{aligned}R_4(m,\widehat{m}) &= \left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)_+\left[\frac{1}{\widehat{s}_m}-\frac{1}{\widetilde{s}}\right]+\left(s(1-\kappa_0)-\|\boldsymbol{\epsilon}\|_n^2\right)_+\left[-\frac{1}{\widehat{s}_m}+\frac{1}{\widetilde{s}}\right]\\ &\leq R_{4,1}(m,\widehat{m})+R_{4,2}(\widehat{m}) ,\end{aligned}$$

where $R_{4,1}(m,\widehat{m})$ and $R_{4,2}(\widehat{m})$ are respectively defined by

$$\begin{aligned}R_{4,1}(m,\widehat{m}) &:= \left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)_+\left[\frac{1}{\widehat{s}_m}-\frac{1}{\widetilde{s}}\right]\\ R_{4,2}(\widehat{m}) &:= \left(s(1-\kappa_0)-\|\boldsymbol{\epsilon}\|_n^2\right)_+\frac{1}{\widetilde{s}} .\end{aligned}$$

By definition, we know that $\log\left(\widetilde{s}/\widehat{s}_m\right)$ is smaller than $pen(m)-pen(\widehat{m})$.

$$\begin{aligned}R_{4,1}(m,\widehat{m}) &\leq \left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)_+\frac{1}{\widehat{s}_m}\log\left(\frac{\widetilde{s}}{\widehat{s}_m}\right)\\ &\leq \left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)_+\frac{1}{\widehat{s}_m}pen(m) .\end{aligned}$$

Applying Cauchy-Schwarz inequality yields

$$\begin{aligned}\mathbb{E}\left[R_{4,1}(m,\widehat{m})\mathbf{1}_{\mathbb{B}_1}\right] &\leq \mathbb{E}\left[\frac{\left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)_+}{\widehat{s}_m}\right]pen(m)\\ &\leq \sqrt{\mathbb{E}\left[\left(\|\boldsymbol{\epsilon}\|_n^2-s(1-\kappa_0)\right)^2\right]\mathbb{E}\left[\frac{1}{\widehat{s}_m^2}\right]}pen(m)\\ &\leq \frac{s}{s_m}\sqrt{\left[\kappa_0^2+\frac{2}{n}\right]\frac{n^2}{(n-|m|-2)(n-|m|-4)}}pen(m)\\ &\leq Lpen(m) ,\end{aligned}$$

since $|m| \leq n/2$ by Assumption $\mathbb{H}_{K,\eta}$ and since $n \geq 17$. Let us turn to $R_{4,2}(\widehat{m})$. We apply Hölder's inequality with $v := \lfloor n/8 \rfloor$ and $u = v/(v-1)$.

$$\begin{aligned}\mathbb{E}\left[R_{4,2}(\widehat{m})\mathbf{1}_{\mathbb{B}_1}\right] &\leq \mathbb{E}\left[\left(s(1-\kappa_0) - \|\boldsymbol{\epsilon}\|_n^2\right)_+ \frac{1}{\widetilde{s}}\right]\\ &\leq \mathbb{E}\left[\mathbf{1}_{s(1-\kappa_0)\geq\|\boldsymbol{\epsilon}\|_n^2}\frac{s}{\widetilde{s}}\right]\\ &\leq \left[\mathbb{P}\left[\|\boldsymbol{\epsilon}\|_n^2 \leq s(1-\kappa_0)\right]\right]^{1/u}\left[\mathbb{E}\left(\frac{s}{\widetilde{s}}\right)^v\right]^{1/v}\\ &\leq \left[\mathbb{P}\left[\|\boldsymbol{\epsilon}\|_n^2 \leq s(1-\kappa_0)\right]\right]^{1/u}\left[\sum_{m\in\mathcal{M}}\mathbb{E}\left(\frac{s}{\widehat{s}_m}\right)^v\right]^{1/v} .\end{aligned}$$

Since v is smaller than $n/8$ and since $|m|$ is smaller than $n/2$ it follows that $n-|m|-2v$ is larger than $n/4$. Hence, we can apply Lemma 2.1 to any model $m \in \mathcal{M}$.

$$\begin{aligned}\mathbb{E}\left[R_{4,2}(\widehat{m})\mathbf{1}_{\mathbb{B}_1}\right] &\leq \exp\left[-n\frac{\kappa_0^2}{4u}\right]\left[\sum_{m\in\mathcal{M}}\frac{n^v}{(n-|m|-2)\ldots(n-|m|-2v)}\right]^{1/v}\\ &\leq \frac{n}{n/2-2v}\exp\left[-n\frac{\kappa_0^2}{4u}\right]|\mathcal{M}|^{1/v}\\ &\leq n\exp\left[-n\frac{\kappa_0^2}{4u}\right]|\mathcal{M}|^{1/v} .\end{aligned}$$

Let us bound the cardinality of the collection $\mathcal{M}$. We recall that the dimension of any model $m \in \mathcal{M}$ is assumed to be smaller than $n/2$ by $(\mathbb{H}_{K,\eta})$. Besides, for any $d \in \{1,\ldots,n/2\}$, there are less than $\exp(dH(d))$ models of dimension d. Hence,

$$\log(|\mathcal{M}|) \leq \log(n) + \sup_{d=1,\ldots,n/2} dH(d) .$$

By assumption $(\mathbb{H}_{K,\eta})$, $dH(d)$ is smaller than $n/2$. Thus, $\log(|\mathcal{M}|) \leq \log(n) + n/2$ and it follows that $|\mathcal{M}|^{1/v}$ is smaller than an universal constant providing that n is larger than 8. All in all, we get

$$\mathbb{E}\left[R_{4,2}(\widehat{m})\mathbf{1}_{\mathbb{B}_1}\right] \leq Ln\exp\left[-n\frac{\kappa_0^2}{4u}\right] .$$

2.6. Proof of Lemma 10.9

For any $x > 0$, the following inequality holds

$$x - 1 - \log(x) \leq \frac{9}{64}\left(x - \frac{1}{x}\right)^2 .$$

This statement is easy to establish by studying the derivative of the associated function. Hence, we upper bound the Kullback divergence

$$\begin{aligned}\mathcal{K}\left[t,s;\widehat{t}_m,\widehat{s}_m\right] &= \frac{s}{\widehat{s}_m} + 1 - \log\left(\frac{s}{\widehat{s}_m}\right) + \frac{l(\widehat{t}_m,t)}{\widehat{s}_m}\\ &\leq \frac{9}{64}\left[\frac{s^2}{\widehat{s}_m^2} + \frac{\widehat{s}_m^2}{s^2}\right] + \frac{l(t_m,t)}{\widehat{s}_m} + \frac{l(\widehat{t}_m,t_m)}{\widehat{s}_m} .\end{aligned}$$

Thanks to Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\mathbb{E}\left[\mathcal{K}\left(t,s;\widetilde{t},\widetilde{s}\right)\mathbf{1}_{\mathbb{B}_1^c}\right] &\leq \mathbb{E}\left[\left(\frac{9}{64}\left[\frac{s^2}{\widetilde{s}^2}+\frac{\widetilde{s}^2}{s^2}\right]+\frac{l(t_m,t)}{\widetilde{s}}+\frac{l(\widetilde{t},t_{\widehat{m}})}{\widetilde{s}}\right)\mathbf{1}_{\mathbb{B}_1^c}\right]\\ &\leq L\sqrt{\mathbb{P}\left[\mathbb{B}_1^c\right]}\sqrt{\sum_{m\in\mathcal{M}}\mathbb{E}\left[\mathbf{1}_{m=\widehat{m}}\left(\frac{s^4}{\widehat{s}_m^4}+\frac{\widehat{s}_m^4}{s^4}+\frac{l^2(t_m,t)}{\widehat{s}_m^2}+\frac{l^2(\widehat{t}_m,t_m)}{\widehat{s}_m^2}\right)\right]}\ .\end{aligned}$$

As in the proof of Lemma **10.8**, we apply Hölder's inequality with $v=\lfloor n/16\rfloor$ and $u=v/(v-1)$. Again, we check that for any model $m\in\mathcal{M}$, $n-|m|-8v\geq 1$.

$$\begin{aligned}&\mathbb{E}\left[\mathbf{1}_{m=\widehat{m}}\left(\frac{s^4}{\widehat{s}_m^4}+\frac{\widehat{s}_m^4}{s^4}+\frac{l^2(t_m,t)}{\widehat{s}_m^2}+\frac{l^2(\widehat{t}_m,t_m}{\widehat{s}_m^2}\right)\right]\\ &\quad\leq\mathbb{P}\left[m=\widehat{m}\right]^{\frac{1}{u}}\left[\mathbb{E}\left(\frac{s^{4v}}{\widehat{s}_m^{4v}}\right)^{\frac{1}{v}}+\mathbb{E}\left(\frac{\widehat{s}_m^{4v}}{s^{4v}}\right)^{\frac{1}{v}}+\mathbb{E}\left(\frac{l^{2v}(t_m,t)}{\widehat{s}_m^{2v}}\right)^{\frac{1}{v}}+\mathbb{E}\left(\frac{l^{2v}(\widehat{t}_m,t_m)}{\widehat{s}_m^{2v}}\right)^{\frac{1}{v}}\right]\ .\end{aligned}$$

We bound the first two terms applying Lemma 2.1 or computing the v-th moment of χ^2 random variable.

$$\begin{aligned}\mathbb{E}\left[\frac{s^{4v}}{\widehat{s}_m^{4v}}\right]^{\frac{1}{v}} &\leq \frac{n^4}{(n-|m|-8v)^4}\ ,\\ \mathbb{E}\left[\frac{\widehat{s}_m^{4v}}{s^{4v}}\right]^{\frac{1}{v}} &= \left(\frac{(n-|m|)(n-|m|+2)\ldots(n-|m|+2(4v-1))(s_m)^{4v}}{(ns)^{4v}}\right)^{\frac{1}{v}}\\ &\leq \frac{(n-|m|+8v)^4\left(s+l(0,t)\right)^4}{n^4s^4}\ .\end{aligned}$$

As $\|\Pi_m^{\perp}(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2$ is independent of the couple $(\|\Pi_m(\boldsymbol{\epsilon}+\boldsymbol{\epsilon}_m)\|_n^2,\mathbf{X}_m)$, the random variables $\widehat{s}_m$ and $l(\widehat{t}_m,t_m)$ are independent. We bound the the l^{2v}-risk of $l(\widehat{t}_m,t)$ thanks to Proposition 7.8 in [8].

$$\begin{aligned}\mathbb{E}\left(\frac{l^{2v}(\widehat{t}_m,t_m)}{\widehat{s}_m^{2v}}\right)^{\frac{1}{v}} &= \left(\mathbb{E}\left[l^{2v}(\widehat{t}_m,t_m)\right]\mathbb{E}\left[\frac{1}{\widehat{s}_m^{2v}}\right]\right)^{\frac{1}{v}}\\ &\leq \frac{Lv^2|m|^2n^4}{(n-|m|-4v)^2}\leq Lv^2|m|^2n^2\frac{n^2}{(n-|m|-4v)^2}\ .\end{aligned}$$

Combining these upper bounds and noting that $n-|m|-8v\geq 1$ and $|m|\leq n/2$ yields

$$\begin{aligned}\mathbb{E}\left[\mathcal{K}\left(t,s;\widetilde{t},\widetilde{s}\right)\mathbf{1}_{\mathbb{B}_1^c}\right] &\leq \left[\frac{2n^2}{(n-|m|-8v)^2}+\frac{Lv|m|n^2}{n-|m|-4v}+\frac{(n-|m|+8v)^2}{n^2}\left(1+\frac{l(0,t)}{s}\right)^2\right]\\ &\quad\times L\sqrt{\mathbb{P}\left[\mathbb{B}_1^c\right]}|\mathcal{M}|^{\frac{1}{2v}}\\ &\leq L_{K,\eta}n^{5/2}\left[1+\frac{l(0,t)}{s}\right]\exp\left[-nL_{K,\eta}\right]\ ,\end{aligned}$$

since $|\mathcal{M}|^{1/2v}$ is smaller than than an universal constant as explained in the proof of Lemma **10.8**. Finally, $l(0,t)/s$ is smaller than $\mathcal{K}(t,s;0,1)$.

2.7. Proof of Corollary 6.1

First, we claim that the penalties (**21**) are lower bounded by penalties defined in (**7**). Suppose that $K > e-1$. Since $\log(1+Kx) \geq \log(1+K)x$ for any x between 0 and 1, it follows that

$$pen_i(m_i) \geq \log(1+K)\frac{|m_i|}{n-|m_i|}\left\{1+\sqrt{2\left[1+\log\left((i-1)/|m_i|\right)\right]^2}\right\} , \tag{10}$$

if $|m_i|/(n-|m_i|)\{1+\sqrt{2[1+\log((i-1)/|m_i|))]^2}\} \leq 1$. If $K \leq e-1$, there exists a positive constant $\zeta(K)$ such that $\log(1+Kx) \geq \sqrt{K}x$, for all $x \leq \zeta(K)$. Hence, we get

$$pen_i(m_i) \geq \sqrt{K}\frac{|m_i|}{n-|m_i|}\left\{1+\sqrt{2\left[1+\log\left((i-1)/|m_i|\right)\right]^2}\right\} , \tag{11}$$

if $K \leq e-1$ and if $|m_i|/(n-|m_i|)[1+\sqrt{2[1+\log((i-1)/|m_i|)]^2}] \leq \zeta(K)$.

Gathering (10) and (11), we get that for any $K > 1$, there exists some $K' > 1$ and some $\zeta(K) > 0$ such that:

$$pen_i(m_i) \geq K'\frac{|m_i|}{n-|m_i|}\left\{1+\sqrt{2\left[1+\log\left((i-1)/|m_i|\right)\right]^2}\right\} ,$$

if $|m_i|/(n-|m_i|)\{1+\sqrt{2[1+\log((i-1)/|m_i|)]^2}\} \leq \zeta(K)$.

For any $2 \leq i \leq p$ and any $1 \leq k \leq (i-1)\wedge d$, $H_i(k)$ is smaller than $1+\log((i-1)/k)$. Hence, $pen_i(m_i)$ is lower bounded by a penalty of the form (**7**) with some $K' > 1$. Assuming that

$$\frac{|m_i|}{n-|m_i|}\left\{1+\sqrt{2\left[1+\log\left((i-1)/|m_i|\right)\right]^2}\right\} \leq \zeta(K)\wedge\eta(K') , \tag{12}$$

we derive that $(\mathbb{H}_{K',\eta})$ is fulfilled and that the risk bound (**23**) holds.

We conclude by observing that Condition (12) is satisfied if

$$d[1+\log(p/d)\vee 0] \leq n\eta'(K) ,$$

for some suitable function $\eta'(K)$.

2.8. Proof of Proposition 4.5

We apply the same arguments as in the proof of Theorem **4.4**, except that we replace $H(|m|)$ by l_m. Then, Lemmas **10.4** and **10.7** are still true.

In the proof of Lemmas **10.8** and **10.9**, the only difference with the previous case concerns the upper bound of $\log(|\mathcal{M}|)$. By definition of l_m,

$$|\mathcal{M}|-1 \leq \sup_{m\in\mathcal{M}\setminus\{\emptyset\}} \exp(|m|l_m) .$$

Hence, $\log(|\mathcal{M}|) \leq 1+\sup_{m\in\mathcal{M}\setminus\{\emptyset\}}|m|l_m$, which is smaller than $1+n/2$ by Assumption $(\mathbb{H}^{bay}_{K,\eta})$.

Lemmas **10.5** and **10.6** also hold when $H(|m|)$ is replaced by l_m as explained in the proof of Proposition 3.5 in [8].

3. Proofs of the minimax bounds

We note $d_H(.,.)$ the Hamming distance between two vectors. The Hamming distance between two matrices of size p is defined as the Hamming distance between their vectorialized version of size p^2. It is also noted $d_H(.,.)$.

3.1. Main lemma

We first state two useful lemmas for proving the minimax lower bounds. The first one is known as Varshamov-Gilbert's lemma, whereas the second one is a modified version of Birgé's lemma for covariance estimation.

Lemma 3.1 (Varshamov-Gilbert's lemma)**.** *Let $\{0,1\}^D$ be equipped with Hamming distance d_H. There exists some subset Θ of $\{0,1\}^D$ with the following properties*

$$d_H(\theta,\theta') > D/4 \text{ for every } (\theta,\theta')\in\Theta^2 \text{ with } \theta\neq\theta' \text{ and } \log|\Theta|\geq D/8 \ .$$

We note $\|t\|_{l_2}$ the Euclidean norm of a vector t.

Lemma 3.2. *Let A be a subset of $\{1,\ldots,p\}$. For any positive matrices Ω and Ω', we define the function $d(\Omega,\Omega')$ by*

$$d(\Omega,\Omega') := \sum_{i\in A}\log\left[1+\frac{\|t_i-t'_i\|_{l_2}^2}{4}\right] + \sum_{i\in A^c}\frac{s_i}{s'_i}+\log\left(\frac{s_i}{s'_i}\right)-1 \ . \tag{13}$$

Let Υ be a subset of square matrices of size p which satisfies the following assumptions:

1. *For all $\Omega\in\Upsilon$, $\varphi_{max}(\Omega)\leq 2$ and $\varphi_{min}(\Omega)\geq 1/2$.*
2. *There exists $(\mathbf{s}_1,\mathbf{s}_2)\in[1;2]^2$ such that $\forall\Omega\in\Upsilon,\ \forall 1\leq i\leq p$, $s_i\in\{\mathbf{s}_1,\mathbf{s}_2\}$.*

Setting $\delta = \min_{\Omega,\Omega'\in\Upsilon,\Omega\neq\Omega'} d(\Omega,\Omega')$, provided that $\max_{\Omega,\Omega'\in\Upsilon}\mathcal{K}(\mathbb{P}_\Omega^{\otimes n};\mathbb{P}_{\Omega'}^{\otimes n})\leq\kappa_1\log|\Upsilon|$, the following lower bound holds

$$\inf_{\widehat{\Omega}}\sup_{\Omega\in\Upsilon}\mathbb{E}_\Omega\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right]\geq\kappa_2\delta \ .$$

The numerical constants κ_1 and κ_2 are made explicit in the proof.

Proof of Lemma 3.2. This lemma is mainly based on an application of Birgé's version of Fano's lemma [3]. We provide a statement of the result that is taken from [6] Sect.2.4.

Lemma 3.3. *Let $(\mathbb{P}_i)_{0\leq i\leq N}$ be some family of probability distributions and $(A_i)_{0\leq i\leq N}$ be some family of disjoint events. Let $a=\min_{0\leq i\leq N}\mathbb{P}_i(A_i)$, then*

$$a\leq\kappa\vee\left(\frac{\max_{1\leq i\leq N}\mathcal{K}\left(\mathbb{P}_i;\mathbb{P}_0\right)}{\log(1+N)}\right) \ ,$$

where $\kappa=2e/(2e+1)$.

Let $\widehat{\Omega}$ be an estimator of Ω. Let $\widetilde{\Omega}$ be an estimator that takes its values in Υ and satisfies

$$d(\widetilde{\Omega},\widehat{\Omega})=\min_{\Omega'\in\Upsilon}d(\Omega',\widehat{\Omega}) \ .$$

We note $(\widetilde{T},\widetilde{S})$ and $(\widehat{T},\widehat{S})$ the Cholesky decompositions of $\widetilde{\Omega}$ and $\widehat{\Omega}$. Let $i\in\{1,\ldots,p\}$. By the triangle inequality,

$$\frac{\|t_i-\widetilde{t}_i\|_{l_2}^2}{4}\leq 2\left[\frac{\|t_i-\widehat{t}_i\|_{l_2}^2}{4}+\frac{\|\widehat{t}_i-\widetilde{t}_i\|_{l_2}^2}{4}\right]\ .$$

For any positive numbers a and b, $\log(1+a+b)\leq\log(1+a)+\log(1+b)$. Moreover, for any positive number a, we have $\log(1+2a)\leq 2\log(1+a)$ because the log function is concave. Hence, we get

$$\log\left[1+\frac{\|t_i-\widetilde{t}_i\|_{l_2}^2}{4}\right]\leq 2\log\left[1+\frac{\|t_i-\widehat{t}_i\|_{l_2}^2}{4}\right]+2\log\left[1+\frac{\|\widehat{t}_i-\widetilde{t}_i\|_{l_2}^2}{4}\right]\ . \tag{14}$$

Let us define the function f by $f(x):=x-\log(x)-1$ for any $x>0$. We state that there exists some numerical constant L such that

$$f\left(\frac{s_i}{\widetilde{s}_i}\right)\leq L\left[f\left(\frac{s_i}{\widehat{s}_i}\right)+f\left(\frac{\widetilde{s}_i}{\widehat{s}_i}\right)\right]\ . \tag{15}$$

If $s_i=\widetilde{s}_i$, this inequality holds for any $L>0$ since $f(1)=0$ and f is non negative. If $s_i\neq\widetilde{s}_i$, there are two possibilities: either $s_i=\mathbf{s}_1$ and $\widetilde{s}_i=\mathbf{s}_2$ or $s_i=\mathbf{s}_2$ and $\widetilde{s}_i=\mathbf{s}_1$. By deriving $f(\mathbf{s}_1/x)+f(\mathbf{s}_2/x)$, one observes that this sum is minimized for $x=(\mathbf{s}_1+\mathbf{s}_2)/2$ and that this minimum equals $f\left[2/\left(1+\mathbf{s}_1/\mathbf{s}_2\right)\right]+f\left[2/\left(1+\mathbf{s}_2/\mathbf{s}_1\right)\right]$. Hence, we obtain that

$$f\left(\frac{s_i}{\widetilde{s}_i}\right)\leq\frac{f\left(\frac{\mathbf{s}_1}{\mathbf{s}_2}\right)\vee f\left(\frac{\mathbf{s}_2}{\mathbf{s}_1}\right)}{f\left[2/\left(1+\frac{\mathbf{s}_1}{\mathbf{s}_2}\right)\right]+f\left[2/\left(1+\frac{\mathbf{s}_2}{\mathbf{s}_1}\right)\right]}\left[f\left(\frac{s_i}{\widehat{s}_i}\right)+f\left(\frac{\widetilde{s}_i}{\widehat{s}_i}\right)\right]\ .$$

Since $\mathbf{s}_1$ and $\mathbf{s}_2$ lie between one and two, it follows that

$$f\left(\frac{s_i}{\widetilde{s}_i}\right)\leq\sup_{1/2\leq x\leq 2}\frac{f(x)}{f\left[2/(1+x)\right]+f\left[2/(1+\frac{1}{x})\right]}\left[f\left(\frac{s_i}{\widehat{s}_i}\right)+f\left(\frac{\widetilde{s}_i}{\widehat{s}_i}\right)\right]\ . \tag{16}$$

The ratio $f(x)/\left(f[2/(1+x)]+f[2/(1+1/x)]\right)$ is positive and continuous on $[1/2;1[$ and $]1;2]$. By studying the Taylor series of $f(x)$ at x equals one, we observe that $f(x)=(x-1)^2/2+o[(x-1)^2]$, $f(2/(1+x))=(x-1)^2/8+o[(x-1)^2]$, and $f(2/(1+1/x))=(x-1)^2/8+o[(x-1)^2]$. Hence, there exists a continuation of the ratio $f(x)/(f[2/(1+x)]+f[2/(1+1/x)])$ around one. The supremum in (16) is therefore finite and the upper bound (15) holds.

Combining the upper bounds (14) and (15) with the definition of $\widetilde{\Omega}$ yields

$$\begin{aligned}d(\Omega,\widetilde{\Omega}) &\leq 2\sum_{i\in A}\left\{\log\left[1+\frac{\|t_i-\widehat{t}_i\|_{l_2}^2}{4}\right]+\log\left[1+\frac{\|\widehat{t}_i-\widetilde{t}_i\|_{l_2}^2}{4}\right]\right\}+L\sum_{i\in A^c}\left[f\left(\frac{s_i}{\widehat{s}_i}\right)+f\left(\frac{s_i}{\widetilde{s}_i}\right)\right]\\ &\leq L\left[d(\Omega,\widehat{\Omega})+d(\widetilde{\Omega},\widehat{\Omega})\right]\leq Ld(\Omega,\widehat{\Omega})\ .\end{aligned}$$

Hence, one can lower bound the risk of $\widehat{\Omega}$ as follows

$$\sup_{\Omega\in\Upsilon}\mathbb{E}_{\Omega}\left[d\left(\Omega,\widehat{\Omega}\right)\right]\geq L^{-1}\delta\sup_{\Omega\in\Upsilon}\mathbb{P}_{\Omega}\left[\Omega\neq\widetilde{\Omega}\right]=L^{-1}\delta\left(1-\min_{\Omega\in\Upsilon}\mathbb{P}_{\Omega}\left[\Omega=\widehat{\Omega}\right]\right)\ .$$

Applying Lemma 3.3, we conclude that

$$\inf_{\widehat{\Omega}} \sup_{\Omega\in\Upsilon} \mathbb{E}_{\Omega}\left[d\left(\Omega,\widehat{\Omega}\right)\right] \geq L^{-1}(1-\kappa)\delta \ , \tag{17}$$

if $\max_{\Omega,\Omega'\in\Upsilon} \mathcal{K}(\mathbb{P}_{\Omega}^{\otimes n},\mathbb{P}_{\Omega'}^{\otimes n}) \leq \kappa \log|\Upsilon|$.

Let us now express this minimax lower bound in term of Kullback divergence. Thanks to the chain rule and Lemma **10.1**, the Kullback divergence between two positive matrices Ω and Ω' decomposes as

$$\mathcal{K}\left(\Omega;\Omega'\right) = \sum_{i=1}^{p} \frac{1}{2}\left[\log\frac{s_i'}{s_i} + \frac{s_i}{s_i'} - 1 + \frac{l_i(t_i,t_i')}{s_i'}\right] \ .$$

Straightforward computations allow to prove that the function $\log(s_i'/s_i)+s_i/s_i'-1+l_i(t_i,t_i')/s_i'$ is minimized with respect to s_i' when $s_i' = s_i + l_i(t_i,t_i')$. This leads to the lower bound

$$\log\frac{s_i'}{s_i} + \frac{s_i}{s_i'} - 1 + \frac{l_i(t_i,t_i')}{s_i'} \geq \log\left(1+\frac{l_i(t_i,t_i')}{s_i}\right) \ .$$

By Definition (**29**) of $l_i(.,.)$ the quantity $l_i(t_i,t_i')$ is lower bounded by $\left[\varphi_{\max}\left(\Omega\right)\right]^{-1}\|t_i-t_i'\|_{l_2}^2$. By assumption, $\left[\varphi_{\max}\left(\Omega\right)\right]^{-1}$ is larger than $1/2$ for any $\Omega\in\Upsilon$. Moreover, s_i is smaller than 2. We conclude that for any $\Omega\in\Upsilon$ and any positive matrix Ω', the following lower bound holds

$$2\mathcal{K}\left(\Omega;\Omega'\right) \geq \sum_{i\in A}\log\left(1+\frac{\|t_i-t_i'\|_{l_2}^2}{4}\right) + \sum_{i\in A^c}\frac{s_i}{s_i'} - \log\left(\frac{s_i}{s_i'}\right) - 1 = d(\Omega,\Omega') \ .$$

We conclude by gathering this last bound with (17) and setting $\kappa_1 := \kappa$ and $\kappa_2 := L^{-1}(1-\kappa)/2$. □

3.2. Adaptive banding

In order to compute the minimax rates of estimation over ellipsoids, we first need to consider a lower bound over the sets $\mathcal{T}_{\mathrm{ord}}[k_1,\ldots,k_p,r]$ and $\mathcal{U}_{\mathrm{ord}}[k_1,\ldots,k_p,r]$.

Definition 3.4. *Let $(k_1,\ldots,k_p)\in\mathbb{N}^p$ such that $k_i\leq i-1$ and let r be a positive number. We respectively define the sets $\mathcal{T}_{ord}[k_1,\ldots,k_p,r]$ and $\mathcal{U}_{ord}[k_1,\ldots,k_p,r]$ as*

$$\mathcal{T}_{ord}[k_1,\ldots,k_p,r] := \left\{T\in Trig(p)\ s.t.\ \ \forall 2\leq i\leq p, T[i,j] = \begin{array}{ccc} 0 & if & 1\leq j\leq i-k_i-1 \\ 0\ or\ r & if & i-k_i\leq j\leq i-1\end{array}\right\} , \tag{18}$$

$$\mathcal{U}_{ord}[k_1,\ldots,k_p,r] := \left\{T^*S^{-1}T\ ,\ \ T\in\mathcal{T}_{ord}[k_1,\ldots,k_p,r]\ and\ S\in Diag(p)\right\} \ . \tag{19}$$

The set $\mathcal{T}_{\mathrm{ord}}[k_1,\ldots,k_p,r]$ contains lower triangular matrices with unit diagonal such that for each line i between 2 and p, the support of the vector $(T[i,j])_{1\leq j\leq i-1}$ is included in $\{i-k_i, i-k_i+1,\ldots,i-1\}$. We are able to lower bound the minimax rates of estimation over $\mathcal{U}_{ord}[(k_1,\ldots,k_p),r]$.

Proposition 3.5. *Assume that $k := 1\vee\max_{1\leq i\leq p} k_i$ is smaller than $\sqrt{n}$. Then,*

$$\inf_{\widehat{\Omega}} \sup_{\Omega\in\mathcal{U}_{\mathrm{ord}}[(k_1,\ldots,k_p),r]} \mathbb{E}\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right] \geq L\left[\sum_{i=2}^{p} k_i + p\right]\left(r^2\wedge\frac{1}{n}\right)$$

These minimax rates of estimation are not really surprising, since they correspond to the minimax rates of estimation of p different parametric regression problems whose minimax rates is known to be of the order $k_i(r^2 \wedge 1/n)$. We refer for instance to [6] Prop. 4.8. Moreover, the term p/n is due to the diagonal matrices S in $\Omega = T^*S^{-1}T$. We believe that the assumption "k is smaller than $\sqrt{n}$" is not necessary but we do not know how to remove it.

3.2.1. Proof of Proposition **5.3**

The lower bound (**17**) is a consequence of Proposition 3.5. Let k be a positive integer smaller than $\lfloor\sqrt{n}\rfloor \wedge (p-1)$. Given some $0 < r < \sqrt{a_k^2 R^2/k}$, we consider the set $\mathcal{U}_{\mathrm{ord}}[0, 1, \ldots, k-1, k, \ldots, k, r]$. Let (T, S) refer to the Cholesky decomposition of a matrix Ω belonging to this set. By definition of $\mathcal{U}_{\mathrm{ord}}$,

$$\sum_{j=1}^{i-1} \frac{T[i, i-j]^2}{a_j^2} = \sum_{j=1}^{k} \frac{T[i, i-j]^2}{a_j^2} \leq kr^2/a_k^2 \leq R^2 \ .$$

Hence, the set $\mathcal{U}_{\mathrm{ord}}[0, 1, \ldots, k-1, k, \ldots, k, r]$ is included in $\mathcal{E}(a, R, p)$. By Proposition 3.5, we obtain the minimax lower bound.

$$\begin{aligned} \inf_{\widehat{\Omega}} \sup_{\Omega \in \mathcal{E}(a,R,p)} \mathbb{E}\left[\mathcal{K}\left(\Omega; \widehat{\Omega}\right)\right] &\geq Lp(k+1)\left(\frac{a_k^2 R^2}{k} \wedge \frac{1}{n}\right) \\ &\geq Lp\left(a_k^2 R^2 \wedge \frac{k+1}{n}\right) \ . \end{aligned}$$

Similarly if $k = 0$, the set $\mathcal{U}_{\mathrm{ord}}[0, \ldots, 0, +\infty]$ is included in $\mathcal{E}(a, R, p)$ and the minimax rates of estimation over $\mathcal{E}(a, R, p)$ is lower bounded by $Lp(a_0 R^2 \wedge 1/n)$ with the convention $a_0 = +\infty$. Taking the infimum for all non-positive integers k smaller than $\lfloor\sqrt{n}\rfloor \wedge (p-1)$ yields the first result.

Let us now turn to the second part of the proposition. We fix some $\gamma > 2$. The matrices Ω considered in the proof of Proposition 3.5 for bounding the minimax rates of estimation over sets of the type $\mathcal{U}_{\mathrm{ord}}[0, 1, \ldots, k-1, k, \ldots, k, r]$ have their eigenvalues between $1/2$ and 2. Hence, the previous lower bound still holds and we get

$$\inf_{\widehat{\Omega}} \sup_{\Omega \in \mathcal{E}(a,R,p) \cap \mathcal{B}_{\mathrm{op}}(\gamma)} \mathbb{E}\left[\mathcal{K}\left(\Omega; \widehat{\Omega}\right)\right] \geq Lp \sup_{k=0,\ldots,\lfloor\sqrt{n}\rfloor \wedge (p-1)} \left(a_k^2 R^2 \wedge \frac{k+1}{n}\right) \ . \qquad (20)$$

Let k be a non-negative integer smaller or equal to $d \wedge (p-1)$, where d is the maximal dimension of the models defining the estimator $\widetilde{\Omega}_{\mathrm{ord}}^d$. We consider the model $m \in \mathcal{M}_{\mathrm{ord}}^d$ defined by

$$m := (\emptyset, \{1\}, \ldots, \{i-1, \ldots, i-k\}, \ldots, \{p-1, \ldots, p-k\}) \ .$$

This model corresponds to estimating a k-th banded Cholesky factor. By Corollary **5.1**, the risk of $\widetilde{\Omega}_{\mathrm{ord}}^d$ is upper bounded by

$$\mathbb{E}\left[\mathcal{K}\left(\Omega; \widetilde{\Omega}_{\mathrm{ord}}^d\right)\right] \leq L_{K,\eta}\left[\frac{p(k+1)}{n} + \mathcal{K}\left(\Omega; \Omega_m\right)\right] + \tau_n(\Omega, K, \eta) \ . \qquad (21)$$

Let us upper bound the bias term $\mathcal{K}(\Omega; \Omega_m)$. By Equation (**31**), it decomposes as

$$\begin{aligned} 2\mathcal{K}\left(\Omega; \Omega_m\right) &= \sum_{i=1}^{p} \frac{s_i}{s_{i,m_i}} - \log\left(\frac{s_i}{s_{i,m_i}}\right) - 1 + \frac{l_i(t_i, t_{i,m_i})}{s_{i,m_i}} \\ &= \sum_{i=1}^{p} \log\left(1 + \frac{l_i(t_i, t_{i,m_i})}{s_i}\right) \leq \sum_{i=1}^{p} \frac{l_i(t_i, t_{i,m_i})}{s_i} \ , \end{aligned}$$

since we have mentioned in the proof of Lemma 3.2 that $s_{i,m_i} = s_i + l_i(t_i, t_{i,m_i})$.

Let i be an integer between $k+2$ and p (if there exists one). We define t^P_{i,m_i} as the orthogonal projection of t_i with respect to the Euclidean norm in $\mathbb{R}^{i-1}$. Since Ω belongs to $\mathcal{B}_{\text{op}}(\gamma)$, it follows that s_i is larger than $1/\gamma$ and that the largest eigenvalue of $\varphi_{\max}(\Omega^{-1}) \leq \gamma$. Hence, we obtain that

$$\begin{aligned}\frac{l_i(t_i, t_{i,m_i})}{s_i} \leq \gamma l_i(t_i, t_{i,m_i}) &\leq \gamma^2 \left[\sum_{j=1}^{i-1} \left(t_i[i-j] - t^P_{i,m_i}[i-j]\right)^2\right] \\ &= \gamma^2 \left[\sum_{j=k+1}^{i-1} t_i^2[i-j]\right] \leq \gamma^2 a^2_{k+1} R^2 \ .\end{aligned}$$

If $i \leq k+2$, then $l_i(t_i, t_{i,m_i}) = 0$. Combining this upper bound with (21), we get

$$\mathbb{E}\left[\mathcal{K}\left(\Omega; \widetilde{\Omega}^d_{\text{ord}}\right)\right] \leq L_{K,\eta,\gamma} p \left[\frac{(k+1)}{n} + a_{k+1} R^2\right] + \tau_n(\Omega, K, \eta) \ .$$

Let us note $(\varphi_1(\Omega), \ldots, \varphi_p(\Omega))$ the eigenvalues of Ω. Since Ω belong to $\mathcal{B}_{\text{op}}(\gamma)$,

$$\begin{aligned}2\mathcal{K}(\Omega; I_p)/p &= 1/p \sum_{i=1}^{p} \left[\varphi_i(\Omega) - \log\left(\varphi_i(\Omega)\right) - 1\right] \\ &\leq \left[\varphi_{\min}(\Omega) - \log\left(\varphi_{\min}(\Omega)\right) - 1\right] \vee \left[\varphi_{\max}(\Omega) - \log\left(\varphi_{\max}(\Omega)\right) - 1\right] \leq L_\gamma \ .\end{aligned}$$

Hence, the term $\tau_n(\Omega, K, \eta)$ is smaller than some $L_{K,\eta,\gamma} p/n$. For n larger than some universal constant, the largest dimension d in the model collection that defines $\widetilde{\Omega}^d_{\text{ord}}$ is larger than $\lfloor\sqrt{n}\rfloor$. Taking the infimum over k in $0, \ldots, \lfloor\sqrt{n}\rfloor \wedge (p-1)$, we conclude that

$$\mathbb{E}\left[\mathcal{K}\left(\Omega; \widetilde{\Omega}^d_{\text{ord}}\right)\right] \leq L_{K,\eta,\gamma,\beta} p \inf_{k=1,\ldots,\lfloor\sqrt{n}\rfloor \wedge (p-1)} \left(a^2_{k+1} R^2 + \frac{k+1}{n}\right) \ .$$

Let us define $d^* := \sup\left\{d' \geq 0 \text{ s.t. } (d'+1)/n \leq a_{d'} R^2\right\}$. By assumption, d^* is smaller or equal to $\lfloor\sqrt{n}\rfloor$. Hence,

$$\begin{aligned}\mathbb{E}\left[\mathcal{K}\left(\Omega; \widetilde{\Omega}^d_{\text{ord}}\right)\right] &\leq L_{K,\eta,\gamma,\beta} p \left(a^2_{d^*+1} R^2 + \frac{d^*+1}{n}\right) \\ &\leq L_{K,\eta,\gamma,\beta} \inf_{\widehat{\Omega}} \sup_{\Omega \in \mathcal{E}(a,R,p) \cap \mathcal{B}_{\text{op}}(\gamma)} \mathbb{E}\left[\mathcal{K}\left(\Omega; \widehat{\Omega}\right)\right] \ ,\end{aligned}$$

thanks to Equation (20).

3.2.2. Proof of Proposition 3.5

Given $r > 0$, let $\mathcal{T}'_{\text{ord}}[k_1, \ldots, k_p, r]$ be a maximal subset of $\mathcal{T}_{\text{ord}}[k_1, \ldots, k_p, r]$ which satisfies the property: "for any two different elements T and T' of $\mathcal{T}'_{\text{ord}}[k_1, \ldots, k_p, r]$, the Hamming distance $d_H(T, T')$ is larger than $\sum_{1 \leq i \leq p} k_i/4$".

By Lemma 3.1, there exists such a set $\mathcal{T}'_{\text{ord}}[k_1, \ldots, k_p, r]$ which satisfies $\log|\mathcal{T}'_{\text{ord}}[k_1, \ldots, k_p, r]| \geq \sum_{2 \leq i \leq p} k_i/8$. Let T be a matrix in $\mathcal{T}'_{\text{ord}}[k_1, \ldots, k_p, r]$. Standard computations allow to prove that the diagonal elements of T^*T lie between 1 and $1+kr^2$. Besides, the sum of the absolute values of the off-diagonal elements on each line is upper bounded as follows.

$$\begin{aligned}\sum_{j\neq i}|T^*T[i,j]| = \sum_{l=1}^{p}\sum_{j\neq i}|T[l,i]T[l,j]| &\leq \sum_{j\neq i}T[i,j]+\sum_{j\neq i}T[j,i]+\sum_{j\neq i}\sum_{j\neq l\neq i}T[l,i]T[l,j] \\ &\leq 2kr+k^2r^2 \ .\end{aligned}$$

If r is smaller than $1/(8k)$, the matrices T^*T are diagonally dominant and their eigenvalues lie between 5/8 and 1.3. Let us choose the subset A of $\{1,\ldots,p\}$ defined by $A:=\{i,\ k_i>0\}$. Then, we introduce the subset $\mathcal{S}[A,p,r]$ as

$$\mathcal{S}[A,p,r] := \{S\in \text{Diag}(p),\ S[i,i]=1 \text{ if } i\in A \text{ and } S[i,i]=1 \text{ or } 1+r \text{ if } i\in A^c\}\ .$$

Applying again Lemma 3.1, we define a subset $\mathcal{S}'[A,p,r]$ of $\mathcal{S}[A,p,r]$ such that $\log|\mathcal{S}'[A,p,r]|\geq \log\left(|A^c|\right)/8$ and such that its elements are $|A^c|/4$-separated with respect to the Hamming distance. If r is smaller than 0.5, then the eigenvalues of any matrix in $\mathcal{S}'[A,p,r]$ are between 1 and 1.5. Finally, we define the set $\mathcal{U}'_{\text{ord}}[k_1,\ldots,k_p,r]$ as

$$\mathcal{U}'_{\text{ord}}[k_1,\ldots,k_p,r] := \{T^*ST,\ T\in\mathcal{T}'_{\text{ord}}[k_1,\ldots,k_p,r] \text{ and } S\in\mathcal{S}'[A,p,r]\}\ .$$

We therefore lower bound its cardinality

$$\log|\mathcal{U}'_{\text{ord}}[k_1,\ldots,k_p,r]| \geq \left(|A^c|+\sum_{i=1}^{p}k_i\right)/8 \geq \left(p+\sum_{i=1}^{p}k_i\right)/16\ .$$

Moreover, if $r\leq 1/(8k)$, the eigenvalues of any matrix in this set are between 1/2 and 2. Let us upper bound the Kullback entropy between any two elements $\Omega=T^*S^{-1}T$ and $\Omega'=T'^*S'^{-1}T'$ of $\mathcal{U}'_{\text{ord}}[k_1,\ldots,k_p,r]$.

$$2\mathcal{K}\left(\Omega;\Omega'\right) = \sum_{i\in A}\frac{l_i\left(t_i,t'_i\right)}{s'_i} + \sum_{i\in A^c}\frac{s_i}{s'_i}+\log\left(\frac{s_i}{s'_i}\right)-1\ .$$

If $i\in A$, then $s'_i=1$. Besides, $l_i(t_i,t'_i)\leq[\varphi_{\min}(\Omega)]^{-1}\|t_i-t'_i\|_{l_2}^2\leq 2k_ir^2$. Recalling that the function f is defined by $f(x)=x-\log x-1$ and that $r\leq 1/8$, straightforward computations lead to $f\left(s'_i/s_i\right)\leq Lr^2$. Hence, for any $(\Omega_1,\Omega_2)\in\mathcal{U}'_{\text{ord}}[k_1,\ldots,k_p,r]$, it holds that

$$\mathcal{K}\left(\mathbb{P}_{\Omega_1}^{\otimes n};\mathbb{P}_{\Omega_2}^{\otimes n}\right)\leq L\left(p+\sum_{i=2}^{p}k_i\right)r^2\ .$$

Moreover, we have $f(1+r)\geq Lr^2$ and $f[(1+r)^{-1}]\geq Lr^2$ since $f(1+x)=x^2/2+o(x^2)$ and $r\leq 1/8$. If $\Omega_1\neq\Omega_2$, then $d(\Omega_1,\Omega_2)$ is lower bounded as follows

$$\begin{aligned}d(\Omega_1,\Omega_2) &\geq \sum_{i\in A}\log\left(1+\frac{k_ir^2}{16}\right)+\frac{|A^c|}{4}\left[f(1+r)\wedge f(1/(1+r))\right] \\ &\geq L\left[\sum_{i\in A}k_i+|A^c|\right]r^2 \geq L\left[\sum_{i=1}^{p}k_i+p\right]r^2\ ,\end{aligned}$$

since r is smaller than 1/8 and $k_ir^2/16$ is smaller than 1/64. Hence, as long as $r\leq L1/\sqrt{n}\wedge 1/8k\wedge 1/2$, one can apply Lemma 3.2.

$$\inf_{\widehat{\Omega}}\sup_{\Omega\in\mathcal{U}_{ord}[k_1,\ldots,k_p,r]}\mathbb{E}\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right]\geq L\left[\sum_{i=2}^{p}k_i+p\right]\left(r^2\wedge\frac{1}{n}\wedge\frac{1}{k^2}\right)\ .$$

By assumption, $1/k^2$ is larger than $1/n$ and the result follows.

3.3. Complete graph selection (proof of Proposition **6.2***)*

3.3.1. First case: minimax rate over $\mathcal{U}_1(k,p)$

Let $T \in \mathcal{T}_1[k,p,r]$ be the set of lower triangular matrices of size p with a unit diagonal and such that each line contains at most k non-zero off-diagonal entries. These entries are also smaller than r in absolute value. We first provide a minimax lower bound on the minimax risk over $T \in \mathcal{T}_1[k,p,r]$.

Proposition 3.6. *Let k and p be two positive integers such that $k \leq p$. Assume that $n \geq Lk^2[1+\log(p/k)]$, where L is some universal constant exhibited in the proof. Then, for any $r > 0$, the minimax rates of estimation over the set $\mathcal{T}_1[k,p,r]$ is lower bounded as follows*

$$\inf_{\widehat{\Omega}} \sup_{T \in \mathcal{T}_1[k,p,r]} \mathcal{K}\left(T^*T; \widehat{\Omega}\right) \geq Lkp\left[r^2 \wedge \frac{1+\log\left(\frac{p}{k}\right)}{n}\right] . \tag{22}$$

We believe that the condition $n \geq Lk^2[1+\log(p/k)]$ is essentially technical but we do not know how to remove it. Thanks to Corollary **6.1**, we can easily derive the minimax rates of estimation over the sets $\mathcal{U}_1[k,p]$. Let us first provide the proof of Proposition 3.6 and then derive the proof of Proposition **6.2**.

Proof of Proposition 3.6. Assume first that k is a power of 2, that $2k$ divides p and that $\log(p/k) \geq 19$. Let us consider the set $\mathcal{T}_1^{(1)}[k,p,r]$ of lower triangular square matrices T of size p such that:

1. the diagonal of T is made of 1,
2. the lower left submatrix of T of size $p/2$ contains exactly k entries that equal r on each line and on each column,
3. every other entry of T is zero.

Clearly, $\mathcal{T}_1^{(1)}[k,p,r]$ is in one to one correspondence with the set $\Theta[k,p/2]$ of binary square matrices of size $p/2$ that contain exactly k non-zero coefficients on each line and each column.

Consider $T \in \mathcal{T}_1^{(1)}[k,p,r]$. We claim that as long as r is smaller than $1/8k$, the eigenvalues of T^*T are between $1/2$ and 2. Indeed, the diagonal elements of T^*T are all between 1 and $1+kr^2$. Besides, the sum of the off-diagonal elements is upper bounded by

$$\begin{aligned}\sum_{j\neq i} |T^*T[i,j]| = \sum_{l=1}^{p}\sum_{j\neq i} |T[l,i]T[l,j]| &\leq \sum_{j\neq i} T[i,j] + \sum_{j\neq i} T[j,i] + \sum_{j\neq i}\sum_{j\neq l\neq i} T[l,i]T[l,j] \\ &\leq 2kr + k^2r^2 .\end{aligned}$$

Hence, if $r \leq 1/8k$, the matrix T is diagonally dominant and the sum of the off diagonal terms is smaller than $3/8$ whereas the diagonal term is between 1 and $1+1/8$.

Let T and T' be two elements of $\mathcal{T}_1^{(1)}[k,p,r]$. Let us upper bound the Kullback entropy between the corresponding precision matrices.

$$2\mathcal{K}\left(T^*T; T'^*T'\right) = \sum_{i=1}^{p} l_i(t_i, t_i') \leq \sum_{i=p/2+1}^{p} \varphi_{\max}(T^*T)\|t_i - t_i'\|_{l_2}^2 \leq 4kpr^2 . \tag{23}$$

Lemma 3.7. *Assume that* $\log(p/k) \geq 19$. *Let* $\Theta[k,p/2]$ *be equipped with Hamming distance* d_H. *There exists some subset* $\Theta'[k,p/2]$ *of* $\Theta[k,p/2]$ *with the following properties*

$$d_H(\theta,\theta') > pk/4 \text{ for every } (\theta,\theta') \in \Theta'^2 \text{ with } \theta \neq \theta' \text{ and } \log|\Theta'| \geq kp/20 \log\left(\frac{p}{k}\right) \ . \tag{24}$$

The proof of this lemma is postponed to the end of this subsection. By Lemma 3.7, there exists some subset $\mathcal{T}_1^{(2)}[k,p,r]$ of $\mathcal{T}_1^{(1)}[k,p,r]$ such that $d_H(T,T') \geq pk/4$ for every $(T,T') \in \mathcal{T}_1^{(2)}[k,p,r]$ with $T \neq T'$ and

$$\log|\mathcal{T}_1^{(2)}[k,p,r]| \geq kp/20 \log\left(\frac{p}{k}\right) \ . \tag{25}$$

Let us take $A = \{1,\ldots p\}$ and let us consider the function $d(.,.)$ defined in Lemma 3.2. Observe that $2kr^2 \leq 1/32$ since $r \leq 1/(8k)$. By the mean value theorem, we obtain that $\log(1+x/4) \geq x/8$ for any positive number x smaller than $2kr^2$. Hence, we get

$$\begin{aligned} d(T^*T, T'^*T') &= \sum_{i=1}^p \log\left[1 + \frac{d_H(t_i,t_i')r^2}{4}\right] \geq \sum_{i=1}^p \frac{d_H(t_i,t_i')r^2}{8} \\ &\geq d_H(T,T')\frac{r^2}{8} \geq \frac{pkr^2}{32} \ , \end{aligned} \tag{26}$$

for any $T \neq T'$ in $\mathcal{T}_1^{(2)}[k,p,r]$. We are now in position to apply Lemma 3.2 to $\mathcal{T}_1^{(2)}[k,p,r]$ with the bounds (23), (25), and (26).

$$\inf_{\widehat{\Omega}} \sup_{T\in\mathcal{T}_1^{(1)}[k,p,r]} \mathbb{E}\left[\mathcal{K}\left(T^*T;\widehat{\Omega}\right)\right] \geq \frac{\kappa_2}{64}pkr^2 \ ,$$

as long as $2kpnr^2 \leq \kappa_1 kp/20 \log(p/k)$ and $r \leq 1/8k$. This yields

$$\begin{aligned} \inf_{\widehat{\Omega}} \sup_{T\in\mathcal{T}_1[k,p,r]} \mathbb{E}\left[\mathcal{K}\left(T^*T;\widehat{\Omega}\right)\right] &\geq Lpk\left[r^2 \wedge \frac{\log\left(\frac{p}{k}\right)}{n}\right] \\ &\geq Lpk\left[r^2 \wedge \frac{1+\log\left(\frac{p}{k}\right)}{n}\right] \ , \end{aligned}$$

since $n \geq k^2[1+\log(p/k)]$ and $\log(p/k) \geq 19$.

We now turn to the case where k is not a power of 2 or $2k$ does not divide p. We only assume that $\log(p/k)$ is larger than $19 + \log(2)$. Let us define $k' := 2^{\lfloor \log_2 k \rfloor}$ and p' as the largest integer that is divided by $2k'$ and is smaller than p. Here $\log_2$ refers to the function $\log(.)/\log(2)$. It follows that $k/2 \leq k' \leq k$ and $p/2 \leq p' \leq p$. Consequently, $\log(p'/k')$ is larger than $\log(p/2k) \geq 19$. Let $\mathcal{T}_1^{(1)}[k',p',p,r]$ denote the set of lower triangular matrices T such that the diagonal elements of T equal 1, the lower left submatrix of T of size $p'/2$ contains exactly k' entries that equal r on each line and on each column and such that every other entry of T is zero. Arguing as in the first case, we obtain that

$$\begin{aligned} \inf_{\widehat{\Omega}} \sup_{T\in\mathcal{T}_1[k,p,r]} \mathbb{E}\left[\mathcal{K}\left(T^*T;\widehat{\Omega}\right)\right] &\geq Lp'k'\left[r^2 \wedge \frac{1+\log\left(\frac{p'}{k'}\right)}{n}\right] \\ &\geq Lpk\left[r^2 \wedge \frac{1+\log\left(\frac{p}{k}\right)}{n}\right] \ . \end{aligned}$$

Finally, we consider the situation where the ratio $\log(p/k)$ is smaller than $19+\log(2)$. The set $\mathcal{T}_{\text{ord}}[(0,1,\ldots,k-1,k,\ldots,k),r]$ is included in $\mathcal{T}_1[k,p,r]$. If we choose the set A to be $\{1,\ldots,p\}$, then a slight modification in the proof of Proposition 3.5 allows to show the minimax lower bound:

$$\inf_{\widehat{\Omega}} \sup_{T\in\mathcal{T}_{\text{ord}}[(0,\ldots,k,\ldots,k),r]} \mathcal{K}\left[T^*T;\widehat{\Omega}\right] \geq Lkp\left[r^2\wedge\frac{1}{n}\right] ,$$

as long as $k\leq\sqrt{n}$. Hence, it follows that

$$\inf_{\widehat{\Omega}} \sup_{T\in\mathcal{T}_1[k,p,r]} \mathcal{K}\left[T^*T;\widehat{\Omega}\right] \quad\geq\quad Lkp\left[r^2\wedge\frac{1}{n}\right]\geq Lkp\left[r^2\wedge\frac{1+\log\left(\frac{p}{k}\right)}{n}\right] ,$$

since $\log(p/k)$ is smaller than $19+\log(2)$.

□

Proof of Proposition **6.2**. We derive from the proof of Proposition 3.6 a minimax lower bound over $\mathcal{U}_1[k,p]\cap\mathcal{B}_{\mathcal{K}}(n^\beta)$. First, we consider the case where k is a power of 2, $2k$ divides p and $\log(p/k)$ is larger than 19. Take $r^2=[(1+\log(p/k))/n]\wedge(8k)^{-2}$. In the previous proof, we have shown that

$$\inf_{\widehat{\Omega}} \sup_{T\in\mathcal{T}_1^{(1)}[k,p,r]} \mathbb{E}\left[\mathcal{K}\left(T^*T;\widehat{\Omega}\right)\right]\geq Lkp\left(r^2\wedge\frac{1+\log(p/k)}{n}\right)\geq Lkp\frac{1+\log(p/k)}{n} ,$$

since $n\geq k^2(1+\log(p/k)$. Moreover, we have mentioned that for any matrix T in $\mathcal{T}_1^{(1)}[k,p,r]$, $\varphi_{\min}(T^*T)\geq 1/2$. Let us now upper bound the Kullback divergence with the identity matrix.

$$\begin{aligned}\mathcal{K}\left(T^*T;I_p\right) &\leq \frac{1}{2}\sum_{i=2}^p l_i(t_i,0_{i-1})\leq\frac{\varphi_{\min}^{-1}(T^*T)}{2}\|T-I_p\|_F^2\\ &\leq kpr^2\leq p\leq pn^\beta .\end{aligned}$$

Hence, the set $\{T^*T,\,T\in\mathcal{T}_1^{(1)}[k,p,r]\}$ is included in $\mathcal{U}_1[k,p]\cap\mathcal{B}_{\mathcal{K}}(n^\beta)$ and the lower bound follows.

The case where k is not a power of 2 or $2k$ does not divide p is handled similarly if one uses the set $\mathcal{T}_1^{(1)}[k',p',p,r]$ defined in the proof of Proposition 3.6. Finally, one uses the set $\mathcal{T}_{\text{ord}}[(0,\ldots,k,\ldots,k),r]$ if $\log(p/k)\leq 19+\log(2)$.

Let us turn to the upper bound on the risk. By Corollary **6.1**, the estimator $\widetilde{\Omega}_{\text{co}}^d$ satisfies

$$\begin{aligned}\mathbb{E}\left[\mathcal{K}\left(\Omega;\widetilde{\Omega}_{\text{co}}^d\right)\right] &\leq L_{K,\eta}pk\frac{1+\log\left(\frac{p}{k}\right)}{n}+L_{K,\eta}pn^{5/2}\left[1+\mathcal{K}(\Omega;I_p)\right]\exp\left[-nL_{K,\eta}\right]\\ &\leq L_{K,\eta,\beta}pk\frac{1+\log\left(\frac{p}{k}\right)}{n} ,\end{aligned}$$

for any $\Omega\in\mathcal{U}_1[k,p]\cap\mathcal{B}_{\mathcal{K}}(n^\beta)$. We conclude by gathering the lower and the upper bounds.

□

3.3.2. Second case: minimax rate over $\mathcal{U}_2(k,p)$

This proof follows the same sketch as the proof of Proposition 3.6. Let k be an integer smaller than $p/2$. It is sufficient to prove the result (**25**) for all k smaller than $p/2$ since this lower bound holds up to a multiplicative numerical constant. Assume first that $\log(p) \geq 21$.

Let us take $A := \{p-k+1,\ldots,p\}$. We need to build a suitable subset of $\mathcal{U}_2[k,p]$ that is well separated with respect to the function $d(.,.)$ introduced in Lemma 3.2. Let r_1 and r_2 be two positive numbers respectively smaller than $1/4$ and $1/8$. We shall fix them later.

Let us introduce the set $\mathcal{S}[A^c,p,r_1]$ of diagonal matrices S such that $S[i,i]=1$ if $i\in A$ and $S[i,i]$ is either 1 or $1+r_1$ if $i\in A^c$. The cardinality of this set is $2^{|A^c|}$. Applying Lemma 3.1, there exists a subset $\mathcal{S}'[A^c,p,r_1]$ that satisfies $\log|\mathcal{S}'[A^c,p,r_1]|\geq |A^c|/8$ and such that any two elements of $\mathcal{S}'[A^c,p,r_1]$ are $|A^c|/4$ separated with respect to the Hamming distance d_H.

Let us consider the set $\mathcal{T}_2^{(1)}[k,p,r]$ of lower triangular square matrices T of size p such that:

1. the diagonal of T is made of 1,
2. the lower left submatrix of T of size $k\times\lfloor p/2\rfloor$ contains exactly one entry that equals r_2 on each line and at most one on each column.
3. every other entry of T is zero.

Lemma 3.8. *Assume that $\log p\geq 21$. There exists some subset $\mathcal{T}_2^{(2)}[k,p,r_2]$ of $\mathcal{T}_2^{(1)}[k,p,r_2]$ such that the Hamming distance between any two different elements of $\mathcal{T}_2^{(2)}[k,p,r_2]$ is larger than $k/2$ and such that $\log|\mathcal{T}_2^{(2)}[k,p,r_2]|\geq k\log(p)/10$.*

We now define the subset $\mathcal{U}_2'[k,p,r_1,r_2]$ of $\mathcal{U}_2[k,p]$ by

$$\mathcal{U}_2'[k,p,r_1,r_2] := \left\{\Omega = T^*S^{-1}T\ ,\quad T\in\mathcal{T}_2^{(2)}[k,p,r_2] \text{ and } S\in\mathcal{S}'[A^c,p,r_1]\right\}\ .$$

In order to apply Lemma 3.2, we need to bound the eigenvalues of the matrices in $\mathcal{U}_2'[k,p,r_1,r_2]$, lower bound the function $d(.,.)$ defined in (13), upper bound the Kullback divergence between elements of $\mathcal{U}_2'[k,p,r_1,r_2]$, and lower bound the cardinality of $\mathcal{U}_2'[k,p,r_1,r_2]$.

1. Let us first bound the smallest and the largest eigenvalues of the matrices Ω in this set. Let T and S correspond to the Cholesky decomposition of Ω. Straightforward computations allow to prove that each diagonal element of T^*T is between 1 and $1+r_2^2$ and the sum of the absolute value of the off-diagonal elements of T^*T on each line is smaller than $2r_2$. Hence, T^*T is diagonally dominant and $\varphi_{\max}(T^*T)\leq(1+r_2)^2$ and $\varphi_{\min}(T^*T)\geq 1-2r_2$. Since r_2 is constrained to be smaller than $1/8$ than the eigenvalues of T^*T are between $3/4$ and $3/2$. The eigenvalues of S are between 1 and $5/4$, because r_1 is constrained to be smaller than $1/4$. The eigenvalues of Ω are bounded as follows: $\varphi_{\max}(\Omega)\leq\varphi_{\max}(T^*T)\varphi_{\max}(S^{-1})$ and $\varphi_{\min}(\Omega)\geq\varphi_{\min}(T^*T)\varphi_{\min}(S^{-1})$. Hence, we conclude that the eigenvalues of Ω are between $1/2$ and 2.
2. Let us now lower bound $d(\Omega,\Omega')$ if $\Omega\neq\Omega'$. The quantity $d_H(t_i,t_i')r_2^2/4$ is smaller than 2. Hence, by the mean value theorem $\log(1+d_H(t_i,t_i')r_2^2/4)$ is larger than $d_H(t_i,t_i')r_2^2/8$. By

definition of the sets $\mathcal{T}_2^{(2)}[k,p,r_2]$ and $\mathcal{S}'[A^c,p,r_1]$, we get

$$\begin{aligned} d(\Omega,\Omega') &= \sum_{i=1}^{p-k} f\left(\frac{s_i}{s'_i}\right) + \sum_{i=p-k+1}^{k} \log\left(1+\frac{d_H(t_i,t'_i)r_2^2}{4}\right) \\ &\geq \frac{p-k}{4}\left[f(1+r_1)\wedge f(1/(1+r_1))\right] + \frac{d_H(T,T')r_2^2}{8} \\ &\geq L\left[(p-k)r_1^2 + k\log(p)r_2^2\right] \\ &\geq L\left[pr_1^2 + k\log(p)r_2^2\right] \ , \end{aligned}$$

since k is assumed to be smaller than $p/2$.

3. Let us upper bound the Kullback divergence between two element Ω and Ω' in $\mathcal{U}_2'[k,p,r_1,r_2]$

$$\begin{aligned} 2\mathcal{K}\left(\Omega;\Omega'\right) &= \sum_{i=1}^{p} \frac{s_i}{s'_i} - \log\left(\frac{s_i}{s'_i}\right) - 1 + \frac{l_i(t_i,t'_i)}{s'_i} \\ &= \sum_{i=1}^{p-k} \frac{s_i}{s'_i} - \log\left(\frac{s_i}{s'_i}\right) - 1 + \sum_{i=p-k+1}^{k} l_i(t_i,t'_i) \ . \end{aligned}$$

Since the smallest eigenvalue of Ω is smaller than $1/2$, it follows that $l_i(t_i.t'_i)$ is smaller than $2\|t_i-t'_i\|_{l_2}^2$ which is smaller than $4r_2^2$ by definition of $\mathcal{T}_2^{(2)}[k,p,r_2]$. Let us recall that the function f defined by $f(x)=x-1-\log(x)$ is positive and equivalent to $(x-1)^2$ when x is close to one. Since r_1 is smaller than $1/4$, there exists some numerical constant L such that $f(s_i/s'_i)\leq Lr_1^2$. All in all, we obtain the upper bound

$$\mathcal{K}\left(\Omega;\Omega'\right) \leq L\left[(p-k)r_1^2 + kr_2^2\right] \leq L\left[pr_1^2+kr_2^2\right] .$$

4. Finally, we lower bound the cardinality of $\mathcal{U}_2'[k,p,r_1,r_2]$.

$$\log|\mathcal{U}_2'[k,p,r_1,r_2]| \geq \frac{p-k}{8} + \frac{k\log p}{8} \geq L\left[p + k\log(p)\right] \ .$$

Applying Lemma 3.2, we conclude that

$$\inf_{\widehat{\Omega}} \sup_{\Omega\in\mathcal{U}_2'[k,p,r_1,r_2]} \mathbb{E}\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right] \geq L\left[pr_1^2 + k\log(p)r_2^2\right] \ ,$$

provided that $r_1\leq 1/4$, $r_2\leq 1/8$, and $n\left[pr_1^2+kr_2^2\right]\leq L_1\left[p+k\log(p)\right]$. Choosing $r_1^2 = 1/16\wedge (L_1\wedge 1)/n$ and $r_2^2 = 1/64\wedge (L_1\wedge 1)\log(p)/n$ yields

$$\inf_{\widehat{\Omega}} \sup_{\Omega\in\mathcal{U}_2'[k,p,r_1,r_2]} \mathbb{E}\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right] \geq L\frac{p+k\log p}{n} \ ,$$

since we assume that $n\geq\log(p)$. Let us now prove that the set $\mathcal{U}_2'[k,p,r_1,r_2]$ is included in $\mathcal{B}_{\mathcal{K}}(1)$.

$$\mathcal{K}\left(\Omega;I_p\right) \quad \leq \quad \left[\varphi_{\min}(\Omega)\right]^{-1}\frac{kr_2^2}{2} + \frac{p-k}{2}f(1+r_1) \leq \frac{k\log p}{n} + p - k \leq p \ ,$$

since $f(5/4)\leq 1$ and $n\geq\log(p)$. Hence, the following minimax lower bound also holds

$$\inf_{\widehat{\Omega}} \sup_{\Omega\in\mathcal{U}_2[k,p]\cap\mathcal{B}_{\mathcal{K}}(n^\beta)} \mathbb{E}\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right] \geq L\frac{p+k\log p}{n} \ .$$

If $\log p \leq 21$, we consider the set $\mathcal{U}_{\text{ord}}[0,\ldots,0,1,\ldots,1,r]$ where there are $p-k$ times 0 and k times 1. It is included in $\mathcal{U}_2[k,p]$ and by Proposition 3.5, we conclude that

$$\inf_{\widehat{\Omega}} \sup_{\Omega \in \mathcal{U}_2[k,p]} \mathbb{E}\left[\mathcal{K}\left(\Omega;\widehat{\Omega}\right)\right] \geq L\frac{p+k}{n} \geq L\frac{p+k\log p}{n} ,$$

since $\log p \leq 21$. Besides, one can prove that the set $\mathcal{U}_{\text{ord}}[0,\ldots,0,1,\ldots,1,r]$ is included in $\mathcal{B}_{\mathcal{K}}(1)$ if r is smaller than $1/\sqrt{n}\wedge 1/8$. Hence, the same minimax lower bound holds on $\mathcal{U}_2[k,p]\cap\mathcal{B}_{\mathcal{K}}(n^{\beta})$

3.3.3. *Technical lemmas*

Proof of Lemma 3.7. Let $\Theta'[k,p/2]$ be a maximal subset of $\Theta[k,p/2]$ which is $pk/4$-separated with respect to the Hamming distance. Then, the closed Hamming balls $\mathcal{B}_H(x,pk/4)$ centered at the elements of $\Theta'[k,p/2]$ and with radius $kp/4s$ are covering $\Theta[k,p/2]$. Hence,

$$|\Theta[k,p/2]| \leq \sum_{x\in\Theta'[k,p/2]} \left|\mathcal{B}_H\left(x, 0.5k\frac{p}{2}\right)\right| .$$

The balls $\mathcal{B}_H(x,kp/4)$ can also be considered as subsets of the set $\{0,1\}^{(p/2)^2}_{kp/2}$ of binary sequences of size $(p/2)^2$ with exactly $kp/2$ non-zero coefficients. In the proof of Lemma 4.10 in [6], Massart shows that if $p\geq 8k$

$$\left|\mathcal{B}_H\left(x,k\frac{p}{4}\right)\right| \leq \binom{(p/2)^2}{kp/2}\left(\frac{p}{2k}\right)^{-\rho k(p/2)} ,$$

where $\rho \geq 0.23$. Since we assume that $\log(p/k)\geq 19$, we can apply this result. If follows that

$$\log|\Theta'[k,p/2]| \geq \frac{\rho}{2}kp\log\left(\frac{p}{2k}\right) + \log|\Theta[k,p/2]| - \log\binom{(p/2)^2}{kp/2} . \tag{27}$$

Let us now lower bound the cardinality of $\Theta[k,p/2]$. Observe that $|\Theta[2k,2p]|\geq|\Theta[k,p]|^4$. Let us indeed cut the square matrix of size $2p$ into four square matrices of size p. Then, any combination of any four elements of $\Theta[k,p]$ yields an unique element of $\Theta[2k,2p]$. Since $k=2^s$ for some integer $s>0$ and since $2k$ divides p, one concludes by straightforward induction that

$$\log|\Theta[k,p/2]| \geq k^2\log|\Theta[1,p/(2k)]| .$$

Moreover, $\Theta[1,p/2k]$ is in correspondence with the set of permutations of $p/2k$ elements. Thus,

$$\log|\Theta[1,p/(2k)]| \geq \left(\frac{p}{2k}\right)! \geq \frac{p}{2k}\log\left(\frac{p}{2ek}\right) ,$$

since $a!\geq(a/e)^a$ for any positive integer a. If follows that $\log|\Theta[k,p/2]|\geq pk/2\log[p/(2ek)]$. In contrast, $\log\binom{(p/2)^2}{kp/2}$ is upper bounded by $kp/2\log[pe/(2k)]$ since $\binom{a}{b}\leq(ae/b)^b$ for any positive integers a and b. Gathering these bounds with (27) yields

$$\log|\Theta'[k,p/2]| \geq \frac{pk}{2}\left[\rho\log\left(\frac{p}{k}\right) - \rho\log 2 - 2\right] \geq \frac{\rho}{4}kp\log\left(\frac{p}{k}\right) ,$$

since $\log(p/k)$ is assumed to be larger than 19 which is larger than $2\log 2 + 4/\rho$.

□

Proof of Lemma 3.8. The set $\mathcal{T}_2^{(1)}[k,p,r_2]$ is in one to one correspondence with the set $\Theta_2[k,\lfloor p/2\rfloor]$ of binary matrices of size $k\times\lfloor p/2\rfloor$ with exactly one non-zero entry on each line and at most one on each column. The proof is then quite similar to the proof of Lemma 3.7. Let $\Theta'_2[k,\lfloor p/2\rfloor]$ be a maximal subset of $\Theta_2[k,\lfloor p/2\rfloor]$ such that the Hamming distance between any two different elements of $\Theta'_2[k,\lfloor p/2\rfloor]$ is larger than $k/2$. Then, we take the set $\mathcal{T}_2^{(2)}[k,p,r_2]$ that corresponds to $\Theta'_2[k,\lfloor p/2\rfloor]$.

Let us lower bound the cardinality of $\Theta'_2[k,\lfloor p/2\rfloor]$. Since the closed Hamming balls with radius $k/2$ and centered at the elements of $\Theta'_2[k,\lfloor p/2\rfloor]$ cover $\Theta_2[k,\lfloor p/2\rfloor]$, we get

$$|\Theta_2[k,\lfloor p/2\rfloor]|\geq\sum_{x\in\Theta'_2[k,\lfloor p/2\rfloor]}|\mathcal{B}_H(x,k/2)|\ .$$

One can consider these balls as subsets of the set $\{0,1\}_k^{k\lfloor p/2\rfloor}$of binary sequence of size $k\lfloor p/2\rfloor$ with exactly k non-zero coefficients. We use the same lower bound for the Hamming balls as in the previous proof:

$$|\mathcal{B}_H(x,k/2)|\leq\binom{k\lfloor p/2\rfloor}{k}\left(\lfloor p/2\rfloor\right)^{-\rho k}\ ,$$

if $p\geq 8$. We recall that $\rho\geq 0.23$. We can apply this result since $\log(p)\geq 21$. It follows that

$$\log|\Theta'_2[k,\lfloor p/2\rfloor]|\geq\rho k\log\left(\lfloor p/2\rfloor\right)+\log|\Theta_2[k,\lfloor p/2\rfloor]|-\log\binom{k\lfloor p/2\rfloor}{k}\ .$$

The cardinality of $\Theta_2[k,\lfloor p/2\rfloor]$ is $\lfloor p/2\rfloor!/(\lfloor p/2\rfloor-k)!$. For any positive integers a and b, it holds that $\binom{a}{b}\geq(a/b)^b$ and $a!\geq(a/e)^a$. Hence, we obtain

$$\log|\Theta_2[k,\lfloor p/2\rfloor]|\geq\log\binom{\lfloor p/2\rfloor}{k}+\log(k!)\geq k\log\left(\frac{\lfloor p/2\rfloor}{e}\right)\ .$$

Let us combine the previous bounds and and let us apply the inequality $\binom{a}{b}\leq(ae/b)^b$ which holds for any positive integer a and b. Hence, we get

$$\log|\Theta'_2[k,\lfloor p/2\rfloor]|\geq\rho k\log\left(\lfloor p/2\rfloor\right)-2k\geq k\left[\rho\log(p)-\rho\log(4)-2\right]\geq\frac{\rho k}{2}\log(p)\ ,$$

since $\log(p)\geq 21$.

□

4. Proof of the Frobenius bounds

4.1. Proof of Corollary 5.4

For any symmetric matrix A, we denote $\{\varphi_i(A)\}_{1\leq i\leq p_n}$ the set of its eigenvalues. Since $x-\log x-1$ is equivalent to $(x-1)^2$ when x goes to one, the Kullback-Leibler divergence $\mathcal{K}\left(\Omega;\Omega'\right)$ decomposes as

$$\begin{aligned}\mathcal{K}(\Omega;\Omega') &= \frac{1}{2}\left[tr\left(\Omega'\Sigma\right)-\log\left(|\Omega'\Sigma|\right)-p_n\right]\\ &= \frac{1}{2}\sum_{i=1}^{p_n}\left\{\varphi_i\left(\sqrt{\Sigma}\Omega'\sqrt{\Sigma}\right)-\log\left[\varphi_i\left(\sqrt{\Sigma}\Omega'\sqrt{\Sigma}\right)\right]-1\right\}\\ &= \frac{1}{4}\sum_{i=1}^{p_n}\left[\varphi_i\left(\sqrt{\Sigma}\Omega'\sqrt{\Sigma}\right)-1\right]^2+o\left[\mathcal{K}(\Omega;\Omega')\right]\\ &= \frac{1}{4}\sum_{i=1}^{p_n}\varphi_i^2\left(\sqrt{\Sigma}\Omega'\sqrt{\Sigma}-I_{p_n}\right)+o\left[\mathcal{K}(\Omega;\Omega')\right]\ ,\end{aligned}$$

when $\mathcal{K}(\Omega;\Omega')$ is close to 0. This last sum corresponds to the Frobenius norm of $\sqrt{\Sigma}\Omega'\sqrt{\Sigma}-I_{p_n}$. Hence, we get

$$\|\sqrt{\Sigma}\Omega'\sqrt{\Sigma}-I_{p_n}\|_F^2=4\left[\mathcal{K}(\Omega;\Omega')\right]+o\left[\mathcal{K}(\Omega;\Omega')\right]\ , \tag{28}$$

when $\mathcal{K}(\Omega;\Omega')$ is close to 0. Let us come back to the Frobenius distance between Ω' and Ω,

$$\begin{aligned}\|\Omega'-\Omega\|_F^2 &= tr\left[\sqrt{\Omega}\left(\sqrt{\Sigma}\Omega'\sqrt{\Sigma}-I_{p_n}\right)\Omega\left(\sqrt{\Sigma}\Omega'\sqrt{\Sigma}-I_{p_n}\right)\sqrt{\Omega}\right]\\ &\leq \varphi_{\max}^2(\Omega)\,\|\sqrt{\Sigma}\Omega'\sqrt{\Sigma}-I_{p_n}\|_F^2\ .\end{aligned}$$

Gathering this upper bound with the preceding result yields

$$\|\Omega'-\Omega\|_F^2\leq 4\varphi_{\max}^2(\Omega)\left[\mathcal{K}(\Omega;\Omega')+o\left(\mathcal{K}(\Omega;\Omega')\right)\right]\ , \tag{29}$$

when $\mathcal{K}(\Omega;\Omega')$ is close to 0. By Corollary **5.1**, the risk of $\widetilde{\Omega}^d_{\text{ord}}$ on $\mathcal{U}_{\text{ord}}[k_1,,\ldots,k_p,+\infty]\cap\mathcal{B}_{\text{op}}(\gamma)$ is upper bounded

$$\mathbb{E}\left[\mathcal{K}\left(\Omega;\widetilde{\Omega}^d_{\text{ord}}\right)\right]\leq L_{K,\eta}\frac{p_n+\sum_{i=1}^{p_n}k_i}{n}+\tau_n(K,\eta,\Omega)\ .$$

The Kullback divergence $\mathcal{K}(\Omega;I_{p_n})/p_n$ is upper bounded by $\varphi_{\max}(\Omega)\vee(\log[1/\varphi_{\min}(\Omega)]-1)\leq L_\gamma$. Hence, the term $\tau_n(K,\eta,\Omega)$ is upper bounded by $L_{K,\eta,\gamma}p_n/n$. We conclude that

$$\mathbb{E}\left[\mathcal{K}\left(\Omega;\widetilde{\Omega}^d_{\text{ord}}\right)\right]\leq L_{K,\eta,\gamma}\frac{p_n+\sum_{i=1}^{p_n}k_i}{n}\ .$$

Gathering this upper bound with (29) yields the first result. By Proposition **5.3**, we know that

$$\mathbb{E}\left[\mathcal{K}\left(\Omega;\widetilde{\Omega}^d_{\text{ord}}\right)\right]\leq L_{K,\eta,\gamma}p_n\left(R^{\frac{2}{2s+1}}n^{-\frac{2s}{2s+1}}\wedge\frac{p_n}{n}\right)\ .$$

We prove the second result using this last bound and (29).

The corresponding minimax lower bounds are proved as Propositions 3.5 and **5.3**. Indeed, we consider again the set $\mathcal{U}'_{\text{ord}}[k_1,\ldots,k_{p_n},r]$ defined in the proof of proposition 3.5 with $r\leq(8k)^{-1}\wedge 1/n$. We recall that this set belongs to $\mathcal{B}_{\text{op}}(2)$. For any two matrices $\Omega_1\neq\Omega_2$ in this set,

$$\mathcal{K}(\Omega_1;\Omega_2)\geq 2d(\Omega_1,\Omega_2)\geq L\left[\sum_{i=1}^{p_n}k_i+p_n\right]r^2\ ,$$

where $d(.,.)$ is introduced in Lemma 3.2. The second lower bound is given at the end of the proof of Proposition 3.5. We also have stated the converse upper bound

$$\mathcal{K}(\Omega_1;\Omega_2) \leq L\left[\sum_{i=1}^{p_n} k_i + p_n\right] r^2 \ .$$

Arguing as previously, we connect the Frobenius distance between Ω_1 and Ω_2 with the Kullback entropy.

$$\begin{aligned}\|\Omega_1-\Omega_2\|_F^2 &\geq \varphi^2_{\min}(\Omega_1)\,\|\sqrt{\Omega_1}^{-1}\Omega_2\sqrt{\Omega_1}^{-1}-I_{p_n}\|_F^2\\ &\geq 4\varphi^2_{\min}(\Omega_1)\,\mathcal{K}(\Omega_1;\Omega_2)+o\left[\mathcal{K}(\Omega_1;\Omega_2)\right]\\ &\geq \mathcal{K}(\Omega_1;\Omega_2)+o\left[\mathcal{K}(\Omega_1;\Omega_2)\right]\ ,\end{aligned}$$

because $\varphi_{\min}(\Omega_1)$ is larger than $1/2$. Since r^2 is assumed to be smaller than $1/n$ and since $\sum_{i=1}^{p_n} k_i + p_n = o(n)$, $\mathcal{K}(\Omega_1;\Omega_2)$ goes to 0 when n goes to infinity. Hence, for n sufficiently large,

$$\|\Omega_1-\Omega_2\|_F^2 \geq \frac{1}{2}\mathcal{K}(\Omega_1;\Omega_2) \geq L\left[\sum_{i=1}^{p_n} k_i + p_n\right] r^2 \ .$$

Applying suitably Lemma 3.3 yields

$$\inf_{\widehat{\theta}}\ \sup_{\Omega\in\mathcal{U}_{\mathrm{ord}}[k_1,\ldots,k_{p_n},r]\cap\mathcal{B}_{\mathrm{op}}(\gamma)} \mathbb{E}\left[\|\Omega-\widehat{\Omega}\|_F^2\right] \geq L\left[\sum_{i=1}^{p_n} k_i + p_n\right]\left(r^2\wedge\frac{1}{n}\right)\ ,$$

as long as n is large enough. This proves the first minimax lower bound.

Let us define $k_n := (R^2n)^{1/(2s+1)} \wedge (p-1)$ and $r_n = 1/(8\sqrt{n})$. Since $s > 1/4$, k_n is smaller than $\lfloor\sqrt{n}\rfloor$ for n large enough. We straightforwardly check as in the proof of Proposition **5.3** that $\mathcal{U}_{\mathrm{ord}}[0,1,\ldots,k_n,\ldots k_n,r_n]$ is included in $\mathcal{E}'[s,p_n,R]\cap\mathcal{B}_{\mathrm{op}}(2)$. Using the last minimax lower bound, we conclude that

$$\inf_{\widehat{\theta}}\ \sup_{\Omega\in\mathcal{E}'[s,p_n,R]\cap\mathcal{B}_{\mathrm{op}}(2)} \mathbb{E}\left[\|\Omega-\widehat{\Omega}\|_F^2\right] \geq L\frac{p_nk_n}{n} \geq Lp_n\left(\left(\frac{R}{n^s}\right)^{\frac{2}{2s+1}}\wedge\frac{p_n-1}{n}\right),$$

for n large enough.

4.2. Proof of Corollary 6.3

From the previous proof, we know that for any estimator $\widehat{\Omega}$ such that $\mathcal{K}(\Omega;\widehat{\Omega}) = o_P(1)$ satisfies $\|\widehat{\Omega}-\Omega\|_F^2 = \mathcal{O}_P[\mathcal{K}(\Omega;\widehat{\Omega})]$. Let us apply Corollary **6.1**:

$$\mathbb{E}\left[\mathcal{K}\left(\Omega;\widetilde{\Omega}\right)\right] \leq L_{K,\eta}\frac{(k_n+1)\log p_n}{n} + L_{K,\eta}n^{5/2}\left[p+\mathcal{K}\left(\Omega;I_{p_n}\right)\right]\exp\left[-nL_{K,\eta}\right]\ .$$

The Kullback divergence $\mathcal{K}(\Omega;I_{p_n})$ is upper bounded by $p_n\left[\varphi_{\max}(\Omega)\vee(\log[1/\varphi_{\min}(\Omega)]-1)\right]$. Hence, we get

$$\mathbb{E}\left[\mathcal{K}\left(\Omega;\widetilde{\Omega}\right)\right] \leq L_{K,\eta,\gamma}\frac{p_n+k_n\log p_n}{n}\left[1+o(1)\right]\ .$$

Gathering this last upper bound with (28) yields the first result. Since the Frobenius norm dominates the operator norm, the second result follows.

The corresponding asymptotic minimax lower bound is proved as Proposition **6.2** using again the lower bound of the Frobenius distance $\|\Omega-\widehat{\Omega}\|_F^2$ in terms of the Kullback divergence $\mathcal{K}(\Omega;\widehat{\Omega})$.

5. Proof of Lemma 10.2

Let d be a positive integer larger than one. By Jensen's inequality, we first notice that $\Psi(d)$ is non-positive. Using the density of a $\chi^2(d)$ distribution, we obtain

$$\Psi(d) = \int_0^{+\infty} \frac{\log(t)e^{-t}t^{d/2-1}}{2^{d/2}\Gamma(n/2)}dt - \log(d) := I_d - \log(d) \ ,$$

where $\Gamma(.)$ stands for the Gamma function. Let us exhibit a recurrence relation for I_d applying integration by parts:

$$\begin{aligned} I_d &= \int_0^{+\infty} \frac{\log(2t)e^{-t}t^{d/2-1}}{\Gamma(d/2)}dt = 0 + \int_0^{+\infty} e^{-t}t^{d/2-2}\frac{1+\log(2t)(d/2-1)}{\Gamma(d/2)} \\ &= \frac{1}{d/2-1} + I_{d-2} \ . \end{aligned}$$

Hence, we only have to work out I_1 and I_2 in order to compute I_d.

$$\begin{aligned} I_2 &= \log(2) + \Gamma'(1)/\Gamma(1) = \log(2) - \gamma \ , \\ I_1 &= \log(2) + \Gamma'(1/2)/\Gamma(1/2) = -\log(2) - \gamma \ , \end{aligned}$$

where γ is the Euler constant. For any positive integer d, we therefore derive that

$$\begin{aligned} \Psi(2d) &= \sum_{i=1}^{d-1} \frac{1}{i} - \gamma - \log(d) \ , \\ \Psi(2d+1) &= \sum_{i=1}^{d} \frac{2}{2i-1} - \gamma - 2\log(2) - \log(d+1/2) \ . \end{aligned}$$

Using the asymptotic expansion of the harmonic series yields $\Psi(2d) = -1/(2d) + \mathcal{O}\left(1/(2d)^2\right)$. We note $h(d)$ the d-th partial sum of harmonic series. Straightforwards computations lead to

$$\begin{aligned} \Psi(2d+1) &= 2h(2d) - h(d) - \gamma - 2\log(2) - \log(d+1/2) \\ &= \mathcal{O}\left(\frac{1}{d^2}\right) + \log\left(\frac{d}{d+1/2}\right) = \frac{-1}{2d+1} + \mathcal{O}\left(\frac{1}{(2d)^2}\right) \ . \end{aligned}$$

Thus, we obtain the asymptotic expansion $\Psi(d) = -1/d + \mathcal{O}\left(1/d^2\right)$. Let us turn to the lower bound. From now on, we assume that $d \geq 3$. We define the sequence v_d by $v_d := \Psi(d) + 1/(d-2)$. We know that v_d converges to 0 when d goes to infinity. Let us prove that the subsequences $(v_{2d})_{d>1}$ and $(v_{2d+1})_{d\geq 1}$ are decreasing. Since $\log(1-x) \leq -x - x^2/2$ for any $0 \leq x < 1$,

$$\begin{aligned} v_{2d+2} - v_{2d} &= \frac{3}{2d} - \frac{1}{2d-2} + \log\left(1 - \frac{1}{d+1}\right) \\ &\leq \frac{1}{d} - \frac{1}{2d(d-1)} - \frac{1}{d+1} - \frac{1}{2(d+1)^2} \\ &\leq \frac{1}{2d(d+1)^2} - \frac{1}{d(d+1)(d-1)} < 0 \ . \end{aligned}$$

Analogously, we compute

$$\begin{aligned} v_{2d+1} - v_{2d-1} &= \frac{3}{2d-1} - \frac{1}{2d-3} + \log\left(1 - \frac{2}{2d+1}\right) \\ &\leq \frac{4}{(2d-1)(2d+1)^2} - \frac{8}{(2d-3)(2d-1)(2d+1)} < 0 \ . \end{aligned}$$

We conclude that v_d is non-negative for any $d \geq 3$. It follows that $\Psi(d) \geq -1/(d-2)$.

6. Proof of Proposition 7.2

For the sake of clarity, we forget the subscript $p-1$ and p in the collection of models $\mathcal{M}$, the penalty $pen(.)$ and the vector t.

The proof is divided in two steps. First, we explain why the "true" model m_* belongs to $\widehat{\mathcal{M}}$ with high probability. Then, we prove that m_* minimizes the penalized criterion over the whole collection $\mathcal{M}_{\text{co}}^D$ with high probability. The matrix $\Sigma_{1:(p-1)}$ refers to the submatrix of Σ where the last line and the last column are removed.

The restricted eigenvalues of order q^* of the matrix $\Sigma_{1:(p-1)}$ lie between c_* and c^*. Define $\mathbf{Z} = \mathbf{X}_{1:(p-1)}\sqrt{\Sigma}^{-1}_{1:(p-1)}$. The matrix $\mathbf{Z}$ follows a standard Wishart distribution with parameters n and $p-1$. Let us define $V = \cup_{|A|=q^*}\cup_{\text{supp}(u)=A}\sqrt{\Sigma}u$ as the images of q^*-sparse vectors by $\sqrt{\Sigma}$. The set V is the union $\binom{(p-1)}{q^*}$ subspaces of of dimension q^*. Let us call V_1 one of these subspaces. By Theorem 2.13 in [4], it holds that

$$1/2 \leq \frac{u^*\mathbf{Z}^*\mathbf{Z}u}{nu^*u} \leq 2, \qquad \forall u \in V_1$$

with probability larger than $2\exp[-n(1-1/\sqrt{2}-\sqrt{q^*/n})]$. Applying an union bound, we conclude that

$$c_*/2 \leq \frac{u^*\left[\sqrt{\Sigma}\mathbf{Z}^*\mathbf{Z}\sqrt{\Sigma}\right]_A u}{nu^*u} \leq 2c^*, \qquad \forall A \text{ with } |A| = q^* \text{ and } u \in \mathbb{R}^{q^*} ,$$

with probability going to one. Hence, the empirical design $\mathbf{X}_{1:(p-1)}$ satisfies a sparse Riesz condition of order q^* with spectrum bounds $c_*/2$ and $2c^*$ with probability going to 1.

We will apply Theorem 2 in [9]. In order to check the assumptions of this theorem, we shortly use the notations of Zhang and Huang (See Section 2 in [9]). Since t is q-sparse, we have $\eta_1 = \eta_2 = 0$. Moreover, M_1^* and M_3^* only depend on c_* and c^*. As p is large, we can take $\lambda = 4\sqrt{2c^*}\sqrt{s\log(p)n}$. By Theorem 2 in [9], the Lasso estimator $\widehat{t^\lambda}$ selects all non-zero coefficients of t and selects no more than $(M_1^*-1)q$ other variables with high probability. We conclude that m_* belongs to $\widehat{\mathcal{M}}$ with probability going to one.

The notations $o(1)$, $O(1)$ respectively refer to sequences that converge to 0 or stay bounded when n goes to infinity. These sequences may depend on K, but *do not* depend on m_*, on the true covariance Σ, or a particular model m. For any model m of size smaller than D, let us define

$$\Delta(m, m_*) \quad := \quad \|\Pi_m^\perp \mathbf{X_p}\|_n^2 e^{pen(m)} - \|\Pi_{m_*}^\perp \mathbf{X_p}\|_n^2 e^{pen(m_*)} ,$$

where the notation $\Pi_m^\perp$ is defined in Section **10.1**. Observe that $\widehat{m} = m_*$ if $\Delta(m, m_*)$ is positive for any model m.

CASE 1: $m_* \subsetneq m$. We have $\Delta(m, m_*) \geq 0$ if

$$\frac{\|\Pi_{m_*^\perp\cap m}\boldsymbol{\epsilon}\|_n^2/|m\setminus m_*|}{\|\Pi_m^\perp\boldsymbol{\epsilon}\|_n^2/(n-|m|)} \leq \frac{e^{pen(|m|)} - e^{pen(q)}(n-|m|)}{|m\setminus m_*|e^{pen(q)}} \tag{30}$$

Let us call A_1 and A_2 the right and the left expression of (30). By definition (21) of the penalty, we derive that

$$A_1 \geq 2K \log\left(\frac{p}{|m|}\right)(1+o(1)) \ , \tag{31}$$

since $q\log(p)/n$ goes to 0 and $p/|m|$ goes to infinity (uniformly with respect to m such $m \leq D = n/\log^2(p)$.

Let us turn to A_2. Consider $u \in (0,1)$ and let $F^{-1}_{D,N}(u)$ denote the $1-u$ quantile of a Fisher random variable with D and N degrees of freedom. By Lemma 1 in [2], it holds that

$$DF^{-1}_{D,N}(u) \quad \leq \quad D + 2\sqrt{D\left(1+2\frac{D}{N}\right)\log\left(\frac{1}{u}\right)} + \left(1+2\frac{D}{N}\right)\frac{N}{2}\left[\exp\left(\frac{4}{N}\log\left(\frac{1}{u}\right)\right)-1\right] \ .$$

Let us set u to

$$u = \left\{p^{-2}\binom{p-q}{|m\setminus m_*|}\right\}^{-1} \ .$$

Applying the inequality $\binom{r}{s} \leq s\log(er/s)$, we get the upper bound

$$\begin{aligned} A_2 = F^{-1}_{|m\setminus m_*|,n-|m|}(u) \quad &\leq \quad \left[2\log\left(\frac{p}{|m\setminus m_*|}\right) + 2\sqrt{\frac{2\log(p)}{|m\setminus m_*|}} + \frac{2\log(p)}{|m\setminus m_*|}\right](1+o(1)) \ , \\ &\leq \quad 4\log\left(\frac{p}{|m\setminus m_*|}\right)(1+o(1)) \ , \end{aligned} \tag{32}$$

since $|m|\log(p)/n \leq D\log(p)/n$ goes to 0.

Conclusion. Let us compare the lower bound (31) of A_1 with the upper bound (32) of A_2.

- Let us first assume that $|m| \leq 2q$. Then,

$$A_1 \geq 2K \log\left(\frac{p}{q}\right)(1+o(1)) \geq 2K(1-v)\log(p)\,(1-o(1)) \ ,$$

  since $q \leq n^v/\log(p) \leq p^v$. Hence, we get

$$A_2 \leq 4\log\left(\frac{p}{|m\setminus m_*|}\right)(1+o(1)) < A_1,$$

  for n large enough since we assume that $2K(1-v) > 4$.
- If $|m| > 2|m_*|$, we also have

$$A_2 \leq 4\log\left(\frac{p}{|m|}\right)(1+o(1)) < A_1 \ ,$$

  for n large enough since we assume that $2K > 4$.

It follows from Ineq. (30) and the definition of A_1 and A_2 that $\mathbb{P}\left[m_* \subsetneq \widehat{m}\right] \leq L/p$, for n larger than some positive constant that may depend on K, s, but does *not* depend on m_*.

CASE 2: $m_* \nsubseteq m$. The random variable $n\|\Pi_m^\perp \mathbf{X}_p\|_n^2/(s+l(t_m,t))$ follows a χ^2 distribution with $n-|m|$ degrees of freedom. Applying Lemma 1 in [5], we derive that for any model m

$$\|\Pi_m^\perp \mathbf{X}_p\|_n^2 \geq s\left(1-2\sqrt{\frac{|m|\log(ep/|m|)+2\log p}{n}}\right) = s(1+o(1)) \ ,$$

with probability larger than $1-L/p$. Similarly, we get that

$$\|\Pi_{m_*}^\perp \epsilon\|_n^2 \leq s\left[1+4\sqrt{\frac{\log(p)}{n}}\right] = s(1+o(1)) \ ,$$

with probability larger than $1/p$. Let us define the random variable E_m by

$$E_m = \left\langle \Pi_m^\perp \epsilon, \frac{\mathbf{X}(t-t_m)}{\|\mathbf{X}(t-t_m)\|_n^2}\right\rangle_n^2$$

The quantity $\Delta(m,m_*)$ decomposes as

$$\begin{aligned}\Delta(m,m_*) \quad &\geq \quad -\|\Pi_m\epsilon\|_n^2 + \|\Pi_{m_*}\epsilon\|_n^2 - 2E_m + \frac{1}{2}\|\Pi_m^\perp \mathbf{X}(t-t_m)\|_n^2 \\ &+ \quad s[1+o(1)][e^{pen(m)}-1] - s[1+o(1)][e^{pen(m_*)}-1] \ .\end{aligned}$$

The random variables involved in this last expression follow χ^2 distributions. Applying Lemma 1 in [5], we get that for all m,

$$\begin{aligned}\frac{\|\Pi_m\epsilon\|_n^2}{s} \quad &\leq \quad 6\frac{|m|}{n}\log\left(\frac{p}{|m|}\right)[1+o(1)] \\ \frac{E_m}{s} \quad &\leq \quad 6\frac{|m|}{n}\log\left(\frac{p}{|m|}\right)[1+o(1)] \\ \|\Pi_m^\perp \mathbf{X}(t-t_m)\|_n^2 \quad &\geq \quad l(t_m,t)/2 \ ,\end{aligned}$$

with probability larger than $1-L/p$. Hence, with probability larger than $1-L/p$, we get

$$\frac{\Delta(m_*,m)}{s} \quad \geq \quad \frac{l(t_m,t)}{2s} + (K-12)\frac{|m|}{n}\log\left(\frac{p}{|m|}\right)[1+o(1)] - K\frac{q}{n}\log\left(\frac{p}{q}\right)[1+o(1)] \ . \quad (33)$$

CASE 2.A: $|m| > 2q$. In this case, we lower bound the difference $\exp(pen(m)) - \exp(pen(m_*))$ as in (31). Hence, we obtain that

$$\frac{\Delta(m,m_*)}{s} \geq [K(|m|-q) - 12|m|]\log\left(\frac{p}{|m|}\right)(1+o(1)) \ ,$$

which is strictly positive for n large enough since K is larger than 24.

CASE 2.B: $|m| \leq 2q$. In such a case, we derive from (33) that

$$\frac{\Delta(m,m_*)}{s} \quad \geq \quad \frac{l(t_m,t)}{2s} - (K+12)\frac{q}{n}\log\left(\frac{p}{q}\right)[1+o(1)] \ .$$

By Assumptions (**H.1**) and (**H.2**) we derive that

$$l(t_m, t) \geq \frac{c_*}{2} M_2(K, c_*) s \frac{q}{n} \log(p) \ .$$

Since $M_2(K, c_*) > 2(K+12)/c_*$, $\Delta(m, m_*)$ is positive for n large enough.

References

[1] Baraud, Y., Giraud, C., and Huet, S. (2009). Gaussian model selection with an unknown variance. *Ann. Statist.* **37**, 2, 630–672.

[2] Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptative tests of linear hypotheses by model selection. *Ann. Statist.* **31**, 1, 225–251.

[3] Birgé, L. (2005). A new lower bound for multiple hypothese testing. *IEEE Trans. Inf. Theory* **51**, 4, 1611–1615.

[4] Davidson, K. R. and Szarek, S. J. (2001). *Handbook of the geometry of Banach spaces.* Vol. **1**. Elsevier, North-Holland, Amsterdam, Chapter Local operator theory, random matrices and Banach spaces, 317–316.

[5] Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic function by model selection. *Ann. Statist.* **28**, 5, 1302–1338.

[6] Massart, P. (2007). Concentration Inequalities and Model Selection*, École d'été de probabilités de Saint Flour XXXIII.* Lecture Notes in Mathematics, Vol. **1896**. Springer-Verlag.

[7] Verzelen, N. (2010a). Adaptive estimation of covariance matrices via cholesky decomposition. hal-00524307.

[8] Verzelen, N. (2010b). High-dimensional gaussian model selection on a gaussian design. *Ann. Inst. H. Poincaré Probab. Statist.* **46**, 2, 480–524.

[9] Zhang, C.-H. and Huang, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36**, 4, 1567–1594. http://dx.doi.org/10.1214/07-AOS520. MR2435448