Adaptive Covariance Estimation with model selection.

Rolando Biscay, Hélène Lescornel, Jean-Michel Loubes

▶ **To cite this version:**

Rolando Biscay, Hélène Lescornel, Jean-Michel Loubes. Adaptive Covariance Estimation with model selection.. Mathematical Methods of Statistics, Allerton Press, Springer (link), 2012, 21 (4), pp.283-297. <10.3103/S1066530712040023>. <hal-00675273>

HAL Id: hal-00675273

https://hal.archives-ouvertes.fr/hal-00675273

Submitted on 29 Feb 2012

Adaptive Covariance Estimation with model selection

Rolando Biscay, Hélène Lescornel and Jean-Michel Loubes

Abstract

We provide in this paper a fully adaptive penalized procedure to select a covariance among a collection of models observing i.i.d replications of the process at fixed observation points. For this we generalize the results of [3] and propose to use a data driven penalty to obtain an oracle inequality for the estimator. We prove that this method is an extension to the matricial regression model of the work by Baraud in [1].

Keywords: covariance estimation, model selection, adaptive procedure.

1 Introduction

Estimating the covariance function of stochastic processes is a fundamental issue in statistics with many applications, ranging from geostatistics, financial series or epidemiology for instance (we refer to [10], [8] or [5] for general references). While parametric methods have been extensively studied in the statistical literature (see [5] for a review), nonparametric procedures have only recently received attention, see for instance [6, 3, 4, 2] and references therein.

In [3], a model selection procedure is proposed to construct a non parametric estimator of the covariance function of a stochastic process under mild assumptions. However their method heavily relies on a prior knowledge of the variance. In this paper, we extend this procedure and propose a fully data driven penalty which leads to select the best covariance among a collection of models. This result constitutes a generalization to the matricial regression model of the selection methodology provided in [1].

Consider a stochastic process $(X(t))_{t \in T}$ taking its values in $\mathbb{R}$ and indexed by $T \subset \mathbb{R}^d$, $d \in \mathbb{N}$. We assume that $\mathbb{E}[X(t)] = 0 \ \forall t \in T$ and we aim at estimating its covariance function $\sigma(s, t) = \mathbb{E}[X(s) X(t)] < \infty$ for all $t, s \in T$. We assume we observe $X_i(t_j)$ where $i \in \{1 \dots n\}$ and $j \in \{1 \dots p\}$. Note that the observation points t_j are fixed and that the X_i's are independent copies of the process X. Set $x_i = (X_i(t_1), \dots, X_i(t_p)) \ \forall i \in \{1 \dots n\}$ and denote by Σ the covariance matrix of X at the observations points $\Sigma = \mathbb{E}(x_i x_i^\top) = (\sigma(t_j, t_k))_{1 \le j \le p, 1 \le k \le p}$.

Following the methodology presented in [3], we approximate the process X by its projection onto some finite dimensional model. For this, consider a countable set of functions $(g_\lambda)_{\lambda \in \Lambda}$ which may be for instance a basis of $L^2(T)$ and choose a collection of models $\mathcal{M} \subset \mathcal{P}(\Lambda)$. For $m \subset \mathcal{M}$, a finite number of indices, the process can be

approximated by

$$X\left(t\right) \approx \sum_{\lambda \in m} a_\lambda g_\lambda\left(t\right).$$

Such an approximation leads to an estimator which depends on the collection of functions m, denoted by $\hat{\Sigma}_m$. Our objective is to select in a data driven way, the best model, i.e. the one close to an oracle m_0 defined as the minimizer of the quadratic risk, namely

$$m_0 \in \underset{m \in \mathcal{M}}{\arg \min} R\left(m\right) = \underset{m \in \mathcal{M}}{\arg \min} \mathbb{E}\left[\left\|\Sigma - \hat{\Sigma}_m\right\|^2\right].$$

This result is achieved using a model selection procedure.

The paper falls into the following parts. The description of the statistical framework of the matrix regression is given in Section 2. Section 3 is devoted to the main statistical results. Namely we recall the results of the estimate given in [3] and prove an oracle inequality with a fully data driven penalty. Section 4 states technical results which are used in all the paper, while the proofs are postponed to the Appendix.

2 Statistical model and notations

We consider an $\mathbb{R}$-valued process $X\left(t\right)$ indexed by T a subset of $\mathbb{R}^d$ with expectation equal to 0. We are interested in its covariance function denoted by $\sigma\left(s, t\right) = \mathbb{E}\left[X\left(s\right) X\left(t\right)\right]$.

We have at hand the observations $x_i = \left(X_i\left(t_1\right), \dots, X_i\left(t_p\right)\right)$ for $1 \leqslant i \leqslant n$ where X_i are independent copies of the process and t_j are deterministic points. We note $\Sigma \in \mathbb{R}^{p \times p}$ the covariance matrix of the vector x_i.

Hence we observe

$$x_i x_i^\top = \Sigma + U_i, \quad 1 \leqslant i \leqslant n \tag{1}$$

where U_i are i.i.d. error matrices with expectation 0. We denote by S the empirical covariance of the sample : $S = \frac{1}{n} \sum_{i=1}^n x_i x_i^\top$.

We use the Frobenius norm $\|\quad\|$ defined by $\|A\|^2 = \mathrm{Tr}\left(AA^\top\right)$ for all matrix A. Recall that for a given matrix $A \in \mathbb{R}^{p \times q}$, $vec(A)$ is the vector in $\in \mathbb{R}^{pq}$ obtained by stacking the columns of A on top of one another. We denote by A^- the reflexive generalized inverse of the matrix A, see for instance in [9] or [7].

The idea is to consider that we have a quite good approximation of the process in the following form

$$X\left(t\right) \approx \sum_{\lambda \in m} a_\lambda g_\lambda\left(t\right), \tag{2}$$

where m is a finite subset of a countable set Λ , $(a_\lambda)_{\lambda \in \Lambda}$ are random coefficients in $\mathbb{R}$ and $(g_\lambda)_{\lambda \in \Lambda}$ are real valued functions. We will consider models m among a finite collection denoted by $\mathcal{M}$.

We note $G_m \in \mathbb{R}^{p \times |m|}$ where $(G_m)_{j\lambda} = g_\lambda\left(t_j\right)$ and a_m the random vector of $\mathbb{R}^{|m|}$ with coefficients $(a_\lambda)_{\lambda \in m}$.

Hence, we obtain the following approximations :

$$x = \left(X\left(t_1\right), .., X\left(t_p\right)\right)^\top \approx G_m a_m$$

$$xx^\top \approx G_m a_m a_m^\top G_m^\top$$

$$\Sigma \approx G_m \mathbb{E}\left[a_m a_m^\top\right] G_m^\top$$

Thus, this point of view leads us to approximate Σ by a matrix in the subset

$$\mathcal{S}\left(G_m\right) = \left\{G_m \Psi G_m^\top / \Psi \text{ symmetric in } \mathbb{R}^{|m| \times |m|}\right\} \subset \mathbb{R}^{p \times p}. \tag{3}$$

Hence, for a model m, a natural estimator for Σ is given by the projection of S onto $\mathcal{S}\left(G_m\right)$. We can prove using standard algebra (see in [3] for a general proof) that it has the following form :

$$\widehat{\Sigma}_m = \Pi_m S \Pi_m \quad m \in \mathcal{M} \quad \in \mathbb{R}^{p \times p}, \tag{4}$$

where

$$\Pi_m = G_m \left(G_m^\top G_m\right)^- G_m^\top \quad \in \mathbb{R}^{p \times p} \tag{5}$$

are orthogonal projection matrices. Set

$$D_m = Tr\left(\Pi_m \otimes \Pi_m\right)$$

which is the dimension of $\mathcal{S}\left(G_m\right)$ assumed to be positive, and $\Sigma_m = \Pi_m \Sigma \Pi_m$ the projection of Σ onto this subspace.

Hence we obtain the model selection procedure defined in [3]. The estimation error for a model $m \in \mathcal{M}$ is given by

$$\mathbb{E}\left(\left\|\Sigma - \widehat{\Sigma}_m\right\|^2\right) = \left\|\Sigma - \Pi_m \Sigma \Pi_m\right\|^2 + \frac{\delta_m^2 D_m}{n}, \tag{6}$$

where

$$\delta_m^2 = \frac{\text{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)}{D_m},$$

$$\Phi = V\left(vec\left(x_1 x_1^\top\right)\right).$$

Given $\theta > 0$, it is thus natural to define the penalized covariance estimator $\widehat{\Sigma} = \widehat{\Sigma}_{\widehat{m}}$ by

$$\widehat{m} = \arg\min_{m \in \mathcal{M}} \left\{\frac{1}{n}\sum_{i=1}^{n}\left\|x_i x_i^\top - \widehat{\Sigma}_m\right\|^2 + pen\left(m\right)\right\},$$

where

$$pen\left(m\right) = \left(1 + \theta\right)\frac{\delta_m^2 D_m}{n}. \tag{7}$$

The following result proved in [3] states an oracle inequality for the estimator $\widehat{\Sigma}$.

Theorem 2.1. *Let $q > 0$ be given such that there exists $\kappa > 2\left(1 + q\right)$ satisfying $\mathbb{E}\left\|x_1 x_1^\top\right\|^\kappa < \infty$. Then, for some constants $K\left(\theta\right) > 1$ and $C'\left(\theta, \kappa, q\right) > 0$ we have that*

$$\left(\mathbb{E}\left\|\Sigma - \widehat{\Sigma}\right\|^{2q}\right)^{1/q} \leq 2^{\left(q^{-1} - 1\right)_+}\left[K\left(\theta\right)\inf_{m \in \mathcal{M}}\left(\left\|\Sigma - \Pi_m \Sigma \Pi_m\right\|^2 + \frac{\delta_m^2 D_m}{n}\right) + \frac{\Delta_\kappa}{n}\delta_{\sup}^2\right],$$

3

where

$$\Delta_\kappa^q = C'\left(\theta, \kappa, q\right) \mathbb{E}\left\|x_1 x_1^\top\right\|^\kappa \left(\sum_{m \in \mathcal{M}} \delta_m^{-\kappa} D_m^{-(\kappa/2 - 1 - q)}\right)$$

and

$$\delta_{\sup}^2 = \max\left\{\delta_m^2 : m \in \mathcal{M}\right\}.$$

However the penalty defined here depends on the quantity δ_m which is unknown in practice since it relies on the matrix $\Phi = V\left(vec\left(xx^\top\right)\right)$. Our objective is to study a covariance estimator built with a new penalty involving an estimator of Φ.

More precisely, we will replace $pen(m)$ by an empirical version $\widehat{pen}(m)$, where

$$\widehat{pen}\left(m\right) = \left(1 + \theta\right)\frac{\widehat{\delta}_m^2 D_m}{n}, \tag{8}$$

and

$$\widehat{\delta}_m^2 = \frac{\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\widehat{\Phi}\right)}{D_m},$$

with $\widehat{\Phi}$ an estimator of Φ.

The objective is to generalize Theorem 2.1 and to construct a fully adaptive penalized procedure to estimate the covariance function.

3 Main result : adaptive penalized covariance estimation

Here we state the oracle inequality obtained for the new covariance estimator introduced previously.

Set

$$y_i = vec\left(x_i x_i^\top\right), 1 \leqslant i \leqslant n,$$

which are vectors in $\mathbb{R}^{p^2}$ and denote by $S_{vec} = \frac{1}{n}\sum_{i=1}^n y_i$ their empirical mean. Consider the following constant $C_{inf} = \inf_{m \in \mathcal{M}} \mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)$, and assume that the collection of models is chosen such that $C_{inf} > 0$. Set

$$\widehat{\Phi} = \frac{1}{n}\sum_{i=1}^n \left(y_i y_i^\top - S_{vec} S_{vec}^\top\right),$$

$$\widehat{\delta}_m^2 = \frac{\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\widehat{\Phi}\right)}{D_m}.$$

Given $\theta > 0$, we consider the covariance estimator $\widetilde{\Sigma} = \widehat{\Sigma}_{\widetilde{m}}$ with

$$\widetilde{m} = \arg\min_{m \in \mathcal{M}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left\| x_i x_i^\top - \widehat{\Sigma}_m \right\|^2 + \widehat{pen}\,(m) \right\},$$

where

$$\widehat{pen}\,(m) = (1 + \theta) \frac{\widehat{\delta}_m^2 D_m}{n}. \tag{9}$$

Theorem 3.1. *Let $1 \geqslant q > 0$ be given such that there exists $\beta > \max\left(2\left(1 + 2q\right), 3 + 2q\right)$ satisfying $\mathbb{E}\left\| xx^\top \right\|^\beta < \infty$.*

Then, for a constant C depending on θ, β and q, we have for $n \geqslant n(\beta, \theta, C_{inf}, \Sigma)$, and $\forall \kappa \in\,]2\left(1 + 2q\right); \min\left(\beta, 2\beta - 4\right)[\,$:

$$\left(\mathbb{E}\left[\left\| \Sigma - \widetilde{\Sigma} \right\|^{2q} \right] \right)^{1/q} \leqslant C \inf_{m \in \mathcal{M}} \left(\left\| \Sigma - \Sigma_m \right\|^2 + \frac{\delta_m^2 D_m}{n} \right) \tag{10}$$

$$+ \frac{C}{n} \left[\widetilde{\Delta}_\beta \left[\mathbb{E}\left[\left\| xx^\top \right\|^\beta \right]^{\frac{2}{\beta}} + \left\| \Sigma \right\|^2 \right] + \delta_{sup}^2 \Delta_\kappa \right] \tag{11}$$

where

$$\widetilde{\Delta}_\beta^q = c\left(\theta, \beta, q\right) \left(\mathbb{E}\left[\left\| xx^\top \right\|^\beta \right] \sum_{m \in \mathcal{M}} \delta_m^{-\beta} D_m^{-\beta/2} \right)^{1 - \frac{2q}{\kappa}}$$

$$\Delta_\kappa^q = C\left(\theta, \kappa, q\right) \mathbb{E}\left\| xx^\top \right\|^\kappa \left(\sum_{m \in \mathcal{M}} \delta_m^{-\kappa} D_m^{-(\kappa/2 - 1 - q)} \right)$$

and

$$\delta_{\sup}^2 = \max\left\{ \delta_m^2 : m \in \mathcal{M} \right\}.$$

We have obtained in Theorem 3.1 an oracle inequality since the estimator $\widetilde{\Sigma}$ has the same quadratic risk as the "oracle" estimator except for an additive term of order $O\left(\frac{1}{n}\right)$ and a constant factor. Hence, the selection procedure is optimal in the sense that it behaves as if the true model were at hand.

The proof of this theorem is divided into two parts. First, as in the of Theorem 2.1 proved in [3], we will consider a vectorized version of the model (1). In this technical part we will obtain an oracle inequality under some particular assumptions for a general penalty. In a second part, we will prove that our particular penalty verifies these assumptions by using properties of the estimator $\widehat{\Phi}$.

4 Technical results

4.1 Vectorized model

Here we consider the vectorized version of model (1). In this case, we observe the following vectors in $\mathbb{R}^{p^2}$:

$$y_i = f_i + \varepsilon_i \quad 1 \leqslant i \leqslant n. \tag{12}$$

Here y_i corresponds to $vec\left(x_i x_i^\top\right)$ in the model (1), f_i to $vect\left(\Sigma\right)$ and ε_i to $vec\left(U_i\right)$. We set $f = \left(f_1^\top, \ldots, f_n^\top\right)^\top, y = \left(y_1^\top, \ldots, y_n^\top\right)^\top$ and $\varepsilon = \left(\varepsilon_1^\top, \ldots, \varepsilon_n^\top\right)^\top$, which are vectors in $\mathbb{R}^{np^2}$.

We estimate f by an estimator of the form

$$\widehat{f}_m = P_m y \quad m \in \mathcal{M},$$

where P_m is the orthogonal projection onto a subspace $\mathcal{S}_m$ of dimension D_m. We note $f_m = P_m f$ and we consider the empirical norm $\|f\|_n^2 = \frac{1}{n}\sum_{i=1}^n f_i^\top f_i$ with the corresponding scalar product $\langle \cdot, \cdot \rangle_n$.

First we state the vectorized form of Theorem 2.1. Write

$$\delta_m^2 = \frac{\mathrm{Tr}\left(P_m\left(I_n \otimes \Phi\right)\right)}{D_m},$$

$$\delta_{\sup}^2 = \max\left\{\delta_m^2 : m \in \mathcal{M}\right\}.$$

Given $\theta > 0$, define the penalized estimator $\widehat{f} = \widehat{f}_{\widehat{m}}$, where

$$\widehat{m} = \arg\min_{m \in \mathcal{M}}\left\{\left\|y - \widehat{f}_m\right\|_n^2 + pen\left(m\right)\right\},$$

with

$$pen\left(m\right) = \left(1 + \theta\right)\frac{\delta_m^2 D_m}{n}.$$

Then, the proof of Theorem 2.1 relies on the following proposition proved in [3]:

Proposition 4.1. : *Let $q > 0$ be given such that there exists $\kappa > 2\left(1 + q\right)$ satisfying $\mathbb{E}\left\|\varepsilon_1\right\|^\kappa < \infty$. Then, for some constants $K\left(\theta\right) > 1$ and $C\left(\theta, \kappa, q\right) > 0$ we have that*

$$\left(\mathbb{E}\left\|f - \widehat{f}\right\|_n^{2q}\right)^{1/q} \leq 2^{\left(q^{-1}-1\right)_+}\left[K\left(\theta\right)\inf_{m \in \mathcal{M}}\left(\left\|f - P_m f\right\|_n^2 + \frac{\delta_m^2 D_m}{n}\right) + \frac{\Delta_\kappa}{n}\delta_{\sup}^2\right], \quad (13)$$

where

$$\Delta_\kappa^q = C\left(\theta, \kappa, q\right)\mathbb{E}\left\|\varepsilon_1\right\|^\kappa\left(\sum_{m \in \mathcal{M}}\delta_m^{-\kappa}D_m^{-\left(\kappa/2 - 1 - q\right)}\right).$$

The new estimator $\widetilde{\Sigma}$ defined previously corresponds here to the estimator $\widetilde{f} = \widehat{f}_{\widetilde{m}}$, where

$$\widetilde{m} = \arg\min_{m \in \mathcal{M}}\left\{\left\|y - \widehat{f}_m\right\|_n^2 + \widehat{pen}\left(m\right)\right\},$$

with

$$\widehat{pen}\left(m\right) = \left(1 + \theta\right)\frac{\widehat{\delta}_m^2 D_m}{n},$$

and $\widehat{\delta}_m^2$ is some estimator of δ_m^2.

Next Proposition gives an oracle inequality for this estimator under new assumptions on the model. As Proposition 4.1, it is inspired by the paper [1].

Proposition 4.2. *Let $1 \geqslant q > 0$ be given such that there exists $\kappa > 2\left(1 + 2q\right)$ satisfying $\mathbb{E}\left\|\varepsilon_1\right\|^\kappa < \infty$.*

For $\alpha \in \,]0;1[$, set $\Omega = \cap_{m \in \mathcal{M}}\left\{\widehat{\delta}_m^2 \geqslant \left(1 - \alpha\right)\delta_m^2\right\}$.

Assume that

A1. $\mathbb{E}\left[\widehat{\delta}_m^2\right] \leqslant \delta_m^2$.

A2. $\mathbb{P}\left(\Omega^c\right) \leqslant \tilde{C}\left(\alpha\right)\frac{1}{n^\gamma}$ *for some $\gamma \geqslant \frac{q}{1 - 2q/\kappa}$.*

Then, for a constant C depending on κ, θ and q, and we have

$$
\left(\mathbb{E}\left[\left\|f - \tilde{f}\right\|_n^{2q}\right]\right)^{1/q} \leqslant C \inf_{m \in \mathcal{M}}\left(\left\|f - P_m f\right\|_n^2 + \frac{\delta_m^2 D_m}{n}\right) \tag{14}
$$

$$
+ \frac{C}{n}\left[\widetilde{\Delta}_\kappa\left[\mathbb{E}\left[\left\|\varepsilon_1\right\|^\kappa\right]^{\frac{2}{\kappa}} + \left\|f\right\|_n^2\right] + \delta_{sup}^2 \Delta_\kappa\right] \tag{15}
$$

where

$$
\widetilde{\Delta}_\kappa^q = \left(\tilde{C}\left(\alpha\right)\right)^{\left(1 - \frac{2q}{\kappa}\right)} \quad \text{with } \alpha = \alpha\left(\theta\right) \text{ is fixed in }]0;1[
$$

and

$$
\Delta_\kappa^q = C\left(\theta, \kappa, q\right)\mathbb{E}\left\|\varepsilon_1\right\|^\kappa\left(\sum_{m \in \mathcal{M}}\delta_m^{-\kappa}D_m^{-(\kappa/2 - 1 - q)}\right)
$$

Theorem 3.1 is thus a direct application of Proposition 4.2. Hence only remain to be checked the two assumptions A1 and A2.

4.2 Auxiliary concentration type lemmas

Here we state some propositions required in the proofs of the previous results.

To our knowledge, the first is due to von Bahr and Esseen in [11].

Lemma 4.3. *Let $U_1, \ldots, U_n$ independent centred variables with values in $\mathbb{R}$. For any $1 \leqslant \kappa \leqslant 2$ we have :*

$$
\mathbb{E}\left[\left|\sum_{i=1}^n U_i\right|^\kappa\right] \leqslant 8\sum_{i=1}^n \mathbb{E}\left[\left|U_i\right|^\kappa\right]
$$

The next proposition is proved in [3].

Proposition 4.4. *Given $N, k \in \mathbb{N}$, let $\widetilde{A} \in \mathbb{R}^{Nk \times Nk}\setminus\{0\}$ be a non-negative definite and symmetric matrix and $\varepsilon_1, \ldots, \varepsilon_N$ i.i.d random vectors in $\mathbb{R}^k$ with $\mathbb{E}\left(\varepsilon_1\right) = 0$ and $V\left(\varepsilon_1\right) = \Phi$. Write $\varepsilon = \left(\varepsilon_1^\top, \ldots, \varepsilon_N^\top\right)^\top$, $\zeta\left(\varepsilon\right) = \sqrt{\varepsilon^\top \widetilde{A}\varepsilon}$, and $\delta^2 = \frac{\mathrm{Tr}\left(\widetilde{A}(I_N \otimes \Phi)\right)}{\mathrm{Tr}\left(\widetilde{A}\right)}$. For all $\beta \geq 2$ such that $\mathbb{E}\left\|\varepsilon_1\right\|^\beta < \infty$ it holds that, for all $x > 0$,*

$$
\mathbb{P}\left(\zeta^2\left(\varepsilon\right) \geq \delta^2\mathrm{Tr}\left(\widetilde{A}\right) + 2\delta^2\sqrt{\mathrm{Tr}\left(\widetilde{A}\right)\rho\left(\widetilde{A}\right)x} + \delta^2\rho\left(\widetilde{A}\right)x\right) \leq C_2\left(\beta\right)\frac{\mathbb{E}\left\|\varepsilon_1\right\|^\beta \mathrm{Tr}\left(\widetilde{A}\right)}{\delta^\beta \rho\left(\widetilde{A}\right)x^{\beta/2}},
\tag{16}
$$

where the constant $C_2\left(\beta\right)$ depends only on β.

5 Appendix

5.1 Proof of Proposition 4.2

This proof follows the guidelines of the proof of Theorem 6.1 in [1]. The following lemma will be helpful for the proof of this proposition

Lemma 5.1. *Choose $\eta = \eta(\theta) > 0$ and $\alpha = \alpha(\theta) \in \,]0;1[$ such that $(1+\theta)(1-\alpha) \geqslant (1+2\eta)$. Set $H_m(f) = \left\{ \left\| f - \tilde{f} \right\|_n^2 - \tilde{\kappa}(\theta) \left[\|f - f_m\|_n^2 + \frac{D_m}{n}\hat{\delta}_m^2 \right] \right\}_+$ where $\tilde{\kappa}(\theta) = \left(2 + \frac{4}{\eta}\right)(1+\theta)$.*
Then, for m_0 minimizing $m \mapsto \|f - f_m\|_n^2 + \frac{D_m}{n}\delta_m^2$ in $m \in \mathcal{M}$

$$\mathbb{E}\left[H_{m_0}(f)^q \, \mathbf{1}_\Omega \right] \leqslant \Delta_\kappa^q \delta_{sup}^{2q} \frac{1}{n^q}. \tag{17}$$

where Δ_κ was defined in Proposition 4.2.

Proof. **Lemma 5.1**

First, remark that on the set Ω, for all $m \in \mathcal{M}$

$$\widehat{pen}(m) \geqslant (1-\alpha)(1+\theta)\frac{\delta_m^2 D_m}{n} \geqslant (1+2\eta)\frac{\delta_m^2 D_m}{n}.$$

Set $pen(m) = (1+2\eta)\frac{\delta_m^2 D_m}{n}$, which corresponds to the penalty of Proposition 4.1.

The proof of this lemma is based on the proof of Proposition 4.1 in [3]. In fact, it is sufficient to prove that for each $x > 0$ and $\kappa \geq 2$

$$\mathbb{P}\left(\mathcal{H}(f)\mathbf{1}_\Omega \geq \left(1 + \frac{2}{\eta}\right)\frac{x}{n}\delta_m^2 \right) \leq c(\kappa,\eta)\mathbb{E}\|\varepsilon_1\|^\kappa \sum_{m \in \mathcal{M}} \frac{1}{\delta_m^\kappa} \frac{D_m \vee 1}{(\eta D_m + x)^{\kappa/2}}, \tag{18}$$

where we have set

$$\mathcal{H}(f) = \left[\left\| f - \tilde{f} \right\|_n^2 - \left(2 + \frac{4}{\eta}\right)\left\{ \|f - f_{m_0}\|_n^2 + \widehat{pen}(m_0) \right\} \right]_+.$$

Indeed, for each $m \in \mathcal{M}$,

$$\|f - f_{m_0}\|_n^2 + \widehat{pen}(m_0) = \|f - f_{m_0}\|_n^2 + (1+\theta)\frac{\hat{\delta}_{m_0}^2}{n}D_{m_0}$$

$$\leq (1+\theta)\left(\|f - f_{m_0}\|_n^2 + \frac{\hat{\delta}_{m_0}^2}{n}D_{m_0} \right)$$

then we get that for all $q > 0$,

$$\mathcal{H}^q(f)\mathbf{1}_\Omega \geq H_{m_0}^q(f)\mathbf{1}_\Omega \tag{19}$$

Using the equality

$$\mathbb{E}\left[\mathcal{H}^q(f)\mathbf{1}_\Omega \right] = \int_0^\infty q u^{q-1}\mathbb{P}\left(\mathcal{H}^q(f)\mathbf{1}_\Omega > u \right) du$$

8

and following the proof of Propositon 4.1 in [3] we obtain the upper bound (17) of Lemma 5.1.

Now we turn to the proof of (18). For any $g \in \mathbb{R}^{np^2}$ we define the empirical quadratic loss function by

$$\gamma_n(g) = \|y - g\|_n^2.$$

Using the definition of γ_n we have that for all $g \in \mathbb{R}^{np^2}$,

$$\|f - g\|_n^2 = \gamma_n(g) + 2\langle g - y, \ \varepsilon \rangle_n + \|\varepsilon\|_n^2$$

and therefore

$$\left\| f - \widetilde{f} \right\|_n^2 - \|f - P_{m_0}f\|_n^2 = \gamma_n\left(\widetilde{f} \right) - \gamma_n(P_{m_0}f) + 2\left\langle \widetilde{f} - P_{m_0}f, \ \varepsilon \right\rangle_n. \tag{20}$$

Using the definition of $\widetilde{f}$, we know that

$$\gamma_n\left(\widetilde{f} \right) + \widehat{pen}(\widetilde{m}) \le \gamma_n(g) + \widehat{pen}(m_0)$$

for all $g \in \mathcal{S}_{m_0}$. Then

$$\gamma_n\left(\widetilde{f} \right) - \gamma_n(P_{m_0}f) \le \widehat{pen}(m_0) - \widehat{pen}(\widetilde{m}). \tag{21}$$

So we get from (20) and (21) that

$$\left\| f - \widetilde{f} \right\|_n^2 \le \|f - P_{m_0}f\|_n^2 + \widehat{pen}(m_0) - \widehat{pen}(\widetilde{m})$$
$$+ 2\langle f - P_{m_0}f, \ \varepsilon \rangle_n + 2\langle P_{\widetilde{m}}f - f, \ \varepsilon \rangle_n + 2\left\langle \widetilde{f} - P_{\widetilde{m}}f, \ \varepsilon \right\rangle_n. \tag{22}$$

In the following we set for each $m' \in \mathcal{M}$,

$$\mathcal{B}_{m'} = \{g \in \mathcal{S}_{m'} : \|g\|_n \le 1\},$$
$$G_{m'} = \sup_{t \in \mathcal{B}_{m'}} \langle g, \ \varepsilon \rangle_n = \|P_{m'}\varepsilon\|_n,$$
$$u_{m'} = \begin{cases} \frac{P_{m'}f - f}{\|P_{m'}f - f\|_n} & \text{if } \|P_{m'}f - f\|_n \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Since $\widetilde{f} = P_{\widetilde{m}}f + P_{\widetilde{m}}\varepsilon$, (22) gives

$$\left\| f - \widetilde{f} \right\|_n^2 \le \|f - P_{m_0}f\|_n^2 + \widehat{pen}(m_0) - \widehat{pen}(\widetilde{m})$$

$$+ 2\|f - P_{m_0}f\|_n |\langle u_{m_0}, \ \varepsilon \rangle_n| + 2\|f - P_{\widetilde{m}}f\|_n |\langle u_{\widetilde{m}}, \ \varepsilon \rangle_n| + 2G_{\widetilde{m}}^2. \tag{23}$$

Using repeatedly the following elementary inequality that holds for all positive numbers ν, x, z

$$2xz \le \nu x^2 + \frac{1}{\nu}z^2 \tag{24}$$

9

we get for any $m' \in \mathcal{M}$

$$2 \left\| f - P_{m'} f \right\|_n \left| \langle u_{m'}, \varepsilon \rangle_n \right| \leq \nu \left\| f - P_{m'} f \right\|_n^2 + \frac{1}{\nu} \left| \langle u_{m'}, \varepsilon \rangle_n \right|^2. \tag{25}$$

By Pythagora's Theorem we have

$$\left\| f - \widetilde{f} \right\|_n^2 = \left\| f - P_{\widetilde{m}} f \right\|_n^2 + \left\| P_{\widetilde{m}} f - \widetilde{f} \right\|_n^2$$
$$= \left\| f - P_{\widetilde{m}} f \right\|_n^2 + G_{\widetilde{m}}^2. \tag{26}$$

We derive from (23) and (25) that for any $\nu > 0$

$$\left\| f - \widetilde{f} \right\|_n^2 \leq \left\| f - P_{m_0} f \right\|_n^2 + \nu \left\| f - P_{m_0} f \right\|_n^2 + \frac{1}{\nu} \langle u_{m_0}, \varepsilon \rangle_n^2$$

$$+ \nu \left\| f - P_{\widetilde{m}} f \right\|_n^2 + \frac{1}{\nu} \langle u_{\widetilde{m}}, \varepsilon \rangle_n^2 + 2 G_{\widetilde{m}}^2 + \widehat{pen}\,(m_0) - \widehat{pen}\,(\widetilde{m}).$$

Now taking into account that by equation (26) $\left\| f - P_{\widetilde{m}} f \right\|_n^2 = \left\| f - \widetilde{f} \right\|_n^2 - G_{\widetilde{m}}^2$ the above inequality is equivalent to

$$(1 - \nu) \left\| f - \widetilde{f} \right\|_n^2 \leq (1 + \nu) \left\| f - P_{m_0} f \right\|_n^2 + \frac{1}{\nu} \langle u_{m_0}, \varepsilon \rangle_n^2$$

$$+ \frac{1}{\nu} \langle u_{\widetilde{m}}, \varepsilon \rangle_n^2 + (2 - \nu) G_{\widetilde{m}}^2 + \widehat{pen}\,(m_0) - \widehat{pen}\,(\widetilde{m}). \tag{27}$$

We choose $\nu = \frac{2}{2+\eta} \in]0, 1[$, but for sake of simplicity we keep using the notation ν. Let $\widetilde{p}_1$ and $\widetilde{p}_2$ be two functions depending on ν mapping $\mathcal{M}$ into $\mathbb{R}_+$. They will be specified as in [3] to satisfy

$$pen\,(m') \geq (2 - \nu) \widetilde{p}_1\,(m') + \frac{1}{\nu} \widetilde{p}_2\,(m') \ \ \forall (m') \in \mathcal{M}. \tag{28}$$

Remember that on Ω, $\widehat{pen}(m) \geqslant pen(m) \ \forall m \in \mathcal{M}$. Since $\frac{1}{\nu} \widetilde{p}_2\,(m') \leq pen\,(m')$ and $1 + \nu \leq 2$, we get from (27) and (28) that on the set Ω

$$(1 - \nu) \left\| f - \widetilde{f} \right\|_n^2 \leq (1 + \nu) \left\| f - P_{m_0} f \right\|_n^2 + \widehat{pen}\,(m_0) + \frac{1}{\nu} \widetilde{p}_2\,(m_0) + (2 - \nu) \left(G_{\widetilde{m}}^2 - \widetilde{p}_1\,(\widetilde{m}) \right)$$

$$+ \frac{1}{\nu} \left(\langle u_{\widetilde{m}}, \varepsilon \rangle_n^2 - \widetilde{p}_2\,(\widetilde{m}) \right) + \frac{1}{\nu} \left(\langle u_{m_0}, \varepsilon \rangle_n^2 - \widetilde{p}_2\,(m_0) \right)$$

$$\leq 2 \left(\left\| f - P_{m_0} f \right\|_n^2 + \widehat{pen}\,(m_0) \right) + (2 - \nu) \left(G_{\widetilde{m}}^2 - \widetilde{p}_1\,(\widetilde{m}) \right)$$

$$+ \frac{1}{\nu} \left(\langle u_{\widetilde{m}}, \varepsilon \rangle_n^2 - \widetilde{p}_2\,(\widetilde{m}) \right) + \frac{1}{\nu} \left(\langle u_{m_0}, \varepsilon \rangle_n^2 - \widetilde{p}_2\,(m_0) \right). \tag{29}$$

As $\frac{2}{1-\nu} = 2 + \frac{4}{\eta}$ we obtain that

$$(1 - \nu) \mathcal{H}\,(f) \mathbf{1}_\Omega = \left\{ (1 - \nu) \left\| f - \widetilde{f} \right\|_n^2 - (1 - \nu) \left(2 + \frac{4}{\eta} \right) \left(\left\| f - P_{m_0} f \right\|_n^2 + \widehat{pen}\,(m_0) \right) \right\}_+ \mathbf{1}_\Omega$$

$$= \left\{ (1 - \nu) \left\| f - \widetilde{f} \right\|_n^2 - 2 \left(\left\| f - P_{m_0} f \right\|_n^2 + \widehat{pen}\,(m_0) \right) \right\}_+ \mathbf{1}_\Omega$$

$$\leq \left\{ (2 - \nu) \left(G_{\widetilde{m}}^2 - \widetilde{p}_1\,(\widetilde{m}) \right) + \frac{1}{\nu} \left(\langle u_{\widetilde{m}}, \varepsilon \rangle_n^2 - \widetilde{p}_2\,(\widetilde{m}) \right) + \frac{1}{\nu} \left(\langle u_m, \varepsilon \rangle_n^2 - \widetilde{p}_2\,(m_0) \right) \right\}_+$$

For any $x > 0$,

$$\mathbb{P}\left((1-\nu)\,\mathcal{H}(f)\,\mathbf{1}_\Omega \geq \frac{x\delta_m^2}{n}\right) \leq \mathbb{P}\left(\exists m' \in \mathcal{M} : (2-\nu)\left(G_{m'}^2 - \widetilde{p}_1(m')\right) \geq \frac{x\delta_{m'}^2}{3n}\right)$$

$$+ \mathbb{P}\left(\exists m' \in \mathcal{M} : \frac{1}{\nu}\left(\langle u_{m'}, \varepsilon\rangle_n^2 - \widetilde{p}_2(m')\right) \geq \frac{x\delta_{m'}^2}{3n}\right)$$

$$\leq \sum_{m'\in\mathcal{M}} \mathbb{P}\left((2-\nu)\left(\|P_{m'}\varepsilon\|_n^2 - \widetilde{p}_1(m')\right) \geq \frac{x\delta_{m'}^2}{3n}\right)$$

$$+ \sum_{m'\in\mathcal{M}} \mathbb{P}\left(\frac{1}{\nu}\left(\langle u_{m'}, \varepsilon\rangle_n^2 - \widetilde{p}_2(m')\right) \geq \frac{x\delta_{m'}^2}{3n}\right)$$

$$:= \sum_{m'\in\mathcal{M}} P_{1,m'}(x) + \sum_{m'\in\mathcal{M}} P_{2,m'}(x). \tag{30}$$

From now on, the proof of Lemma 5.1 is exactly the same as the end of the proof of Proposition 4.1 in [3] with $L_m = \nu$. $\qquad\square$

Proof. **Proposition 4.2**

We first provide an upper bound for $\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q}\mathbf{1}_\Omega\right]$, where the set Ω depends on α chose as in Lemma 5.1.

As $q \leq 1$, we have $(a+b)^q \leq a^q + b^q$. Together with Lemma 5.1 we deduce that

$$\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q}\mathbf{1}_\Omega\right] \leq \Delta_\kappa^q \delta_{sup}^{2q}\frac{1}{n^q} + \mathbb{E}\left[\widetilde{\kappa}(\theta)^q\left[\|f - f_{m_0}\|_n^2 + \frac{D_{m_0}}{n}\widehat{\delta}_{m_0}^2\right]^q\right].$$

Using the convexity of $x \mapsto x^{\frac{1}{q}}$ together with the Jensen inequality, we obtain

$$\left(\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q}\mathbf{1}_\Omega\right]\right)^{\frac{1}{q}} \leq 2^{1/q-1}\Delta_\kappa\delta_{sup}^2\frac{1}{n} + 2^{1/q-1}\mathbb{E}\left[\widetilde{\kappa}(\theta)\left[\|f - f_{m_0}\|_n^2 + \frac{D_{m_0}}{n}\widehat{\delta}_{m_0}^2\right]\right],$$

and by using the assumption A1 we have that

$$\left(\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q}\mathbf{1}_\Omega\right]\right)^{\frac{1}{q}} \leq 2^{1/q-1}\Delta_\kappa\delta_{sup}^2\frac{1}{n} + 2^{1/q-1}\widetilde{\kappa}(\theta)\left[\|f - f_{m_0}\|_n^2 + \frac{D_{m_0}}{n}\delta_{m_0}^2\right]. \tag{31}$$

Now we need to find an upper bound for the quantity $\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q}\mathbf{1}_{\Omega^c}\right]$.

First, remark that

$$\left\|f - \widetilde{f}\right\|_n^2 = \|f - P_{\widetilde{m}}y\|_n^2 = \|f - P_{\widetilde{m}}f\|_n^2 + \|P_{\widetilde{m}}(f - y)\|_n^2$$

$$\leq \|f - P_{\widetilde{m}}f\|_n^2 + \|\varepsilon\|_n^2 = \|f\|_n^2 - \|P_{\widetilde{m}}f\|_n^2 + \|\varepsilon\|_n^2$$

And thus

$$\left\|f - \widetilde{f}\right\|_n^2 \leq \|f\|_n^2 + \|\varepsilon\|_n^2.$$

So we have

$$\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q} \mathbf{1}_{\Omega^c}\right] \leqslant \|f\|_n^{2q} \mathbb{P}(\Omega^c) + \mathbb{E}\left[\|\varepsilon\|_n^{2q} \mathbf{1}_{\Omega^c}\right].$$

Using Hölder's inequality with $\frac{\kappa}{2q} > 1$ we obtain

$$\mathbb{E}\left[\|\varepsilon\|_n^{2q} \mathbf{1}_{\Omega^c}\right] \leqslant \mathbb{E}\left[\|\varepsilon\|_n^{\kappa}\right]^{\frac{2q}{\kappa}} \mathbb{P}(\Omega^c)^{\left(1 - \frac{2q}{\kappa}\right)}.$$

But

$$\mathbb{E}\left[\|\varepsilon\|_n^{\kappa}\right] = \frac{1}{n^{\frac{\kappa}{2}}} \mathbb{E}\left[\left(\sum_{i=1}^{n} \|\varepsilon_i\|^2\right)^{\frac{\kappa}{2}}\right],$$

and as $\kappa \geqslant 2$, we can use Minkowsky's inequality to obtain

$$\mathbb{E}\left[\|\varepsilon\|_n^{\kappa}\right] \leqslant \frac{1}{n^{\frac{\kappa}{2}}} \left(\sum_{i=1}^{n} \left(E\left[\|\varepsilon_i\|^{\kappa}\right]\right)^{\frac{2}{\kappa}}\right)^{\frac{\kappa}{2}} = \frac{1}{n^{\frac{\kappa}{2}}} \left(n \left(E\left[\|\varepsilon_1\|^{\kappa}\right]\right)^{\frac{2}{\kappa}}\right)^{\frac{\kappa}{2}},$$

that is

$$\mathbb{E}\left[\|\varepsilon\|_n^{\kappa}\right] \leqslant \mathbb{E}\left[\|\varepsilon_1\|^{\kappa}\right].$$

So we have

$$\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q} \mathbf{1}_{\Omega^c}\right] \leqslant \left[\mathbb{E}\left[\|\varepsilon_1\|^{\kappa}\right]^{\frac{2q}{\kappa}} + \|f\|_n^{2q}\right] \mathbb{P}(\Omega^c)^{\left(1 - \frac{2q}{\kappa}\right)},$$

and with assumption A2

$$\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q} \mathbf{1}_{\Omega^c}\right] \leqslant \left[\mathbb{E}\left[\|\varepsilon_1\|^{\kappa}\right]^{\frac{2q}{\kappa}} + \|f\|_n^{2q}\right] \left(\tilde{C}(\alpha) \frac{1}{n^{\gamma}}\right)^{\left(1 - \frac{2q}{\kappa}\right)}.$$

As $\gamma \geqslant \frac{q}{1 - 2q/\kappa}$, we deduce that

$$\left(\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q} \mathbf{1}_{\Omega^c}\right]\right)^{\frac{1}{q}} \leqslant 2^{\frac{1}{q} - 1} \left[\mathbb{E}\left[\|\varepsilon_1\|^{\kappa}\right]^{\frac{2}{\kappa}} + \|f\|_n^2\right] \tilde{C}(\alpha)^{\frac{1 - \frac{2q}{\kappa}}{q}} \frac{1}{n} \tag{32}$$

To conclude, we use again the convexity of $x \mapsto x^{\frac{1}{q}}$ and the inequality (31) to get

$$\left(\mathbb{E}\left[\left\|f - \widetilde{f}\right\|_n^{2q}\right]\right)^{\frac{1}{q}} \leqslant 4^{\frac{1}{q} - 1} \left[\mathbb{E}\left[\|\varepsilon_1\|^{\kappa}\right]^{\frac{2}{\kappa}} + \|f\|_n^2\right] \tilde{C}(\alpha)^{\frac{1 - \frac{2q}{\kappa}}{q}} \frac{1}{n}$$

$$+ 4^{1/q - 1} \Delta_{\kappa} \delta_{sup}^2 \frac{1}{n}$$

$$+ 4^{1/q - 1} \tilde{\kappa}(\theta) \left[\|f - f_{m_0}\|_n^2 + \frac{D_{m_0}}{n} \delta_{m_0}^2\right]$$

$\square$

5.2   Proof of Theorem 3.1

Recall that $\beta > \max\left(2\left(1+2q\right), 3+2q\right)$ and $\kappa \in \left]2\left(1+2q\right); \min\left(\beta, 2\beta-4\right)\right[$. In order to use Proposition 4.2 , we need to prove the following inequalities :

A1. $\mathbb{E}\left[\widehat{\delta}_m^2\right] \leqslant \delta_m^2$

A2. $\mathbb{P}\left(\Omega^c\right) \leqslant \tilde{C}\left(\alpha\right) \frac{1}{n^\gamma}$ for $\gamma \geqslant \frac{q}{1-2q/\kappa}$.

First we prove A1.

Remember that $\widehat{\delta}_m^2 = \frac{\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\widehat{\Phi}\right)}{D_m}$. By using the linearity of the trace and the equality $\mathbb{E}\left[\widehat{\Phi}\right] = \frac{n-1}{n}\Phi$, we obtain that $\mathbb{E}\left[\widehat{\delta}_m^2\right] = \frac{n-1}{n}\delta_m^2$ which proves the result.

For the second, write $\Omega^c = \cup_{m \in \mathcal{M}}\left\{\widehat{\delta}_m^2 \leqslant (1-\alpha)\delta_m^2\right\}$. We bound up the quantity $\mathbb{P}\left(\widehat{\delta}_m^2 \leqslant (1-\alpha)\delta_m^2\right)$ in the following Proposition.

Proposition 5.2. *For all $m \in \mathcal{M}$, $\alpha \in]0;1[$ and $n \geqslant n(\kappa, \beta, \alpha, C_{inf}, \Sigma)$ we have for some constants $C_1\left(\beta\right), C_2\left(\beta\right)$:*

$$\mathbb{P}\left(\widehat{\delta}_m^2 \leqslant (1-\alpha)\delta_m^2\right) \leqslant \frac{1}{n^\gamma}\left(C_2(\beta)2^{\beta+1} + C_1\left(\beta\right)\frac{1}{\alpha^{\frac{\beta}{2}}}\right)\mathbb{E}\left[\left\|xx^\top\right\|^\beta\right]\delta_m^{-\beta}D_m^{-\frac{\beta}{2}},$$

for $\gamma \geqslant \frac{q}{1-2q/\kappa}$.

This Proposition concludes the proof of A2 with

$$\tilde{C}\left(\alpha\right) = \left(C_2(\beta)2^{\beta+1} + C_1\left(\beta\right)\frac{1}{\alpha^{\frac{\beta}{2}}}\right)\mathbb{E}\left[\left\|xx^\top\right\|^\beta\right]\sum_{m \in \mathcal{M}}\delta_m^{-\beta}D_m^{-\frac{\beta}{2}}.$$

Proof. **Proposition 5.2**

We start by dividing $\mathcal{P}_m = \mathbb{P}\left(\widehat{\delta}_m^2 \leqslant (1-\alpha)\delta_m^2\right)$ into two parts with one of them involving a sum of independent variables with expectation equal to 0.

$$\mathcal{P}_m = \mathbb{P}\left(\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\widehat{\Phi}\right) \leqslant (1-\alpha)\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)\right)$$

$$\mathcal{P}_m = \mathbb{P}\left(\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\left(\widehat{\Phi} - \left(\Phi + \mu\mu^\top\right) + \mu\mu^\top\right)\right) \leqslant -\alpha\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)\right)$$

$$\mathcal{P}_m \leqslant \mathbb{P}\left(\left|\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\left(\frac{1}{n}\sum_{i=1}^n\left(y_iy_i^T - \Phi - \mu\mu^\top\right) + \mu\mu^\top - S_{vec}S_{vec}^\top\right)\right)\right| \geqslant \alpha\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)\right)$$

$$\mathcal{P}_m \leqslant \mathbb{P}\left(\left|\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\left(\frac{1}{n}\sum_{i=1}^n\left(y_iy_i^T - \Phi - \mu\mu^\top\right)\right)\right)\right| \geqslant \frac{\alpha}{2}\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)\right)$$

13

$$+\mathbb{P}\left(\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\mu\mu^\top - S_{vec}S_{vec}^\top\right)\right)\right| \geqslant \frac{\alpha}{2}\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\Phi\right)\right)$$

Set

$$Q1 = \mathbb{P}\left(\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\frac{1}{n}\sum_{i=1}^n\left(y_iy_i^T - \Phi - \mu\mu^\top\right)\right)\right)\right| \geqslant \frac{\alpha}{2}\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\Phi\right)\right)$$

and

$$Q2 = \mathbb{P}\left(\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\mu\mu^\top - S_{vec}S_{vec}^\top\right)\right)\right| \geqslant \frac{\alpha}{2}\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\Phi\right)\right)$$

Study of Q1

First we use Markov's inequality to obtain

$$A1 \leqslant \frac{2^{\frac{\beta}{2}}\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\frac{1}{n}\sum_{i=1}^n\left(y_iy_i^T - \Phi - \mu\mu^\top\right)\right)\right)\right|^{\frac{\beta}{2}}\right]}{\left(\alpha\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\Phi\right)\right)^{\frac{\beta}{2}}}.$$

We must consider the two following cases :

- If $\frac{\beta}{2} \geqslant 2$, Rosenthal's inequality gives

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\left(y_iy_i^T - \Phi - \mu\mu^\top\right)\right)\right)\right|^{\frac{\beta}{2}}\right]$$

$$\leqslant C\left(\frac{\beta}{2}\right)\frac{1}{n^{\frac{\beta}{2}-1}}\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1y_1^\top - \Phi - \mu\mu^\top\right)\right)\right|^{\frac{\beta}{2}}\right]$$

$$+C\left(\frac{\beta}{2}\right)\left(\frac{1}{n}\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1y_1^\top - \Phi - \mu\mu^\top\right)\right)\right|^2\right]\right)^{\frac{\beta}{4}}.$$

As $\frac{\beta}{2} \geqslant 2$, $\frac{1}{n^{\frac{\beta}{2}-1}} \leqslant \frac{1}{n^{\frac{\beta}{4}}}$ and we can use Jensen's inequality on the second term to obtain

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\left(y_iy_i^T - \Phi - \mu\mu^\top\right)\right)\right)\right|^{\frac{\beta}{2}}\right]$$

$$\leqslant C\left(\frac{\beta}{2}\right)\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1y_1^\top - \Phi - \mu\mu^\top\right)\right)\right|^{\frac{\beta}{2}}\right]\frac{2}{n^{\frac{\beta}{4}}}.$$

- If $1 \leqslant \frac{\beta}{2} \leqslant 2$, we use Lemma 4.3 of subsection 4.2 to get

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\left(y_iy_i^T - \Phi - \mu\mu^\top\right)\right)\right)\right|^{\frac{\beta}{2}}\right]$$

$$\leqslant \frac{8}{n^{\frac{\beta}{2}-1}}\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1y_1^\top - \Phi - \mu\mu^\top\right)\right)\right|^{\frac{\beta}{2}}\right]$$

In both cases, we can use the fact that $x \mapsto x^{\frac{\beta}{2}}$ is a convex and increasing function to obtain

$$\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1 y_1^\top - \Phi - \mu\mu^\top\right)\right)\right|^{\frac{\beta}{2}}\right]$$

$$\leqslant 2^{\frac{\beta}{2}-1}\left[\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1 y_1^\top\right)\right)\right|^{\frac{\beta}{2}}\right] + \left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\Phi + \mu\mu^\top\right)\right)\right|^{\frac{\beta}{2}}\right].$$

And by using the Jensen's inequality on the second term we have that

$$\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1 y_1^\top - \Phi - \mu\mu^\top\right)\right)\right|^{\frac{\beta}{2}}\right]$$

$$\leqslant 2^{\frac{\beta}{2}}\mathbb{E}\left[\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1 y_1^\top\right)\right)\right|^{\frac{\beta}{2}}\right].$$

Now consider the following lemma.

Lemma 5.3. *If Ψ is symmetric non-negative definite, then*

$$\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\Psi\right) \in [0; \mathrm{Tr}(\Psi)] \tag{33}$$

From this fact we get that

$$\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(y_1 y_1^\top\right)\right)\right|^{\frac{\beta}{2}} \leqslant \mathrm{Tr}\left(y_1 y_1^\top\right)^{\frac{\beta}{2}} = \|y_1\|^\beta = \|xx^\top\|^\beta.$$

In conclusion, we have

$$Q1 \leqslant C_1(\beta)\frac{\mathbb{E}\left[\|xx^\top\|^\beta\right]}{\alpha^{\frac{\beta}{2}}\delta_m^\beta D_m^{\frac{\beta}{2}}}\frac{1}{n^\gamma}, \tag{34}$$

with $\gamma = \min\left(\frac{\beta}{4}, \frac{\beta}{2} - 1\right)$ and $C_1(\beta) = 2C\left(\frac{\beta}{2}\right)$ if $\beta \geqslant 4$ where $C\left(\frac{\beta}{2}\right)$ is the constant in Rosenthal's inequality and $C_1(\beta) = 8$ if $2 \leqslant \beta \leqslant 4$. Remark that $\frac{\beta}{4} \geqslant \frac{\kappa}{4}$ and $\frac{\beta}{2} - 1 \geqslant \frac{\kappa}{4}$, so $\gamma \geqslant \frac{\kappa}{4}$.

Study of Q2
Recall that

$$Q2 = \mathbb{P}\left(\left|\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\mu\mu^\top - S_{vec}S_{vec}^\top\right)\right)\right| \geqslant \frac{\alpha}{2}\mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\Phi\right)\right).$$

Set

$$B2 = \mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\mu\mu^\top - S_{vec}S_{vec}^\top\right)\right).$$

Using the properties of the trace, we can write

$$B2 = \mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(\mu\mu^\top\right)\right) - \mathrm{Tr}\left((\Pi_m \otimes \Pi_m)\left(S_{vec}S_{vec}^\top\right)\right)$$

$$= \mathrm{Tr}\left(\mu^\top(\Pi_m \otimes \Pi_m)\mu\right) - \mathrm{Tr}\left(S_{vec}^\top(\Pi_m \otimes \Pi_m)S_{vec}\right).$$

But $\Pi_m \otimes \Pi_m$ is an orthogonal projection matrix, then

$$B2 = \mathrm{Tr}\left(\mu^\top \left(\Pi_m \otimes \Pi_m\right)^\top \left(\Pi_m \otimes \Pi_m\right)\mu\right) - \mathrm{Tr}\left(S_{vec}^\top \left(\Pi_m \otimes \Pi_m\right)^\top \left(\Pi_m \otimes \Pi_m\right)S_{vec}\right)$$

$$B2 = \left\|\left(\Pi_m \otimes \Pi_m\right)\mu\right\|^2 - \left\|\left(\Pi_m \otimes \Pi_m\right)S_{vec}\right\|^2$$

$$B2 = \left(\left\|\left(\Pi_m \otimes \Pi_m\right)\mu\right\| - \left\|\left(\Pi_m \otimes \Pi_m\right)S_{vec}\right\|\right)\left(\left\|\left(\Pi_m \otimes \Pi_m\right)\mu\right\| + \left\|\left(\Pi_m \otimes \Pi_m\right)S_{vec}\right\|\right)$$

Hence

$$|B2| \leqslant \left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|\left(\left\|\left(\Pi_m \otimes \Pi_m\right)\mu\right\| + \left\|\left(\Pi_m \otimes \Pi_m\right)S_{vec}\right\|\right)$$

$$|B2| \leqslant \left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|^2 + 2\left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|\left\|\left(\Pi_m \otimes \Pi_m\right)\mu\right\|$$

$$|B2| \leqslant \left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|^2 + 2\left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|\left\|\mu\right\|$$

Finally

$$Q2 \leqslant \mathbb{P}\left(\left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|^2 \geqslant \frac{\alpha}{4}\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)\right) \tag{35}$$
$$+ \mathbb{P}\left(\left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\| \geqslant \frac{\alpha}{8\left\|\mu\right\|}\mathrm{Tr}\left(\left(\Pi_m \otimes \Pi_m\right)\Phi\right)\right)$$

Now we need to provide an upper bound for the quantities

$$\mathbb{P}\left(\left\|\left(\Pi_m \otimes \Pi_m\right)\left(\mu - S_{vec}\right)\right\|^2 \geqslant t\right).$$

For this we will use the deviation bound provided by Proposition 4.4 stated in subsection 4.2 .

Set

$$G_n = \frac{1}{n}\begin{pmatrix} Id_{p^2} & \ldots & Id_{p^2} \\ Id_{p^2} & \ldots & Id_{p^2} \\ \vdots & \ddots & \vdots \\ Id_{p^2} & \ldots & Id_{p^2} \end{pmatrix} \in \mathbb{R}^{p^2 n \times p^2 n}$$

Then

$$G_n\left(y - f\right) = \mathbf{1_n} \otimes \left(S_{vec} - \mu\right)$$

Now, if

$$H_m = Id_n \otimes \left(\Pi_m \otimes \Pi_m\right) = \begin{pmatrix} \Pi_m \otimes \Pi_m & 0 & \ldots & 0 \\ 0 & \Pi_m \otimes \Pi_m & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Pi_m \otimes \Pi_m \end{pmatrix} \in \mathbb{R}^{p^2 n \times p^2 n},$$

we have

$$H_m\left(\mathbf{1}_n \otimes \left(S_{vec} - \mu\right)\right) = \mathbf{1}_n \otimes \left(\left(\Pi_m \otimes \Pi_m\right)\left(S_{vec} - \mu\right)\right).$$

16

In conclusion, with

$$A_m = H_m G_n = \frac{1}{n} \begin{pmatrix} \Pi_m \otimes \Pi_m & \dots & \Pi_m \otimes \Pi_m \\ \Pi_m \otimes \Pi_m & \dots & \Pi_m \otimes \Pi_m \\ \vdots & \ddots & \vdots \\ \Pi_m \otimes \Pi_m & \dots & \Pi_m \otimes \Pi_m \end{pmatrix} \in \mathbb{R}^{p^2 n \times p^2 n},$$

we have that

$$A_m (y - f) = \mathbf{1}_n \otimes ((\Pi_m \otimes \Pi_m)(S_{vec} - \mu)).$$

Moreover, A_m is an orthogonal projection matrix and we have the following equalities

$$\|A_m (y - f)\|^2 = n \|(\Pi_m \otimes \Pi_m)(S_{vec} - \mu)\|^2 = (y - f)^\top A_m (y - f),$$

$$\mathrm{Tr}(A_m) = \frac{n}{n} \mathrm{Tr}(\Pi_m \otimes \Pi_m) = D_m,$$

$$\mathrm{Tr}(A_m (Id_n \otimes \Phi)) = \frac{n}{n} \mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi) = \mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi).$$

Now we can use Proposition 4.4 with $\widetilde{A} = A_m$, $\varepsilon_i = y_i - \mu$, $\mathrm{Tr}(A_m) = D_m$, $\rho(A_m) = 1$, $\delta^2 = \delta_m^2$ and $\beta \geqslant 2$.

This gives for all $x > 0$

$$\mathbb{P}\left((y - f)^\top A_m (y - f) \geqslant \mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)\left[1 + \sqrt{\frac{x}{D_m}}\right]^2\right) \leqslant C_2(\beta) \frac{\mathbb{E}\left[\|y_1 - \mu\|^\beta\right] D_m^{\frac{\beta}{2}+1}}{\mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)^{\frac{\beta}{2}} x^{\frac{\beta}{2}}},$$

that is

$$\mathbb{P}\left(\|(\Pi_m \otimes \Pi_m)(S_{vec} - \mu)\|^2 \geqslant \frac{1}{n}\mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)\left[1 + \sqrt{\frac{x}{D_m}}\right]^2\right) \leqslant C_2(\beta) \frac{\mathbb{E}\left[\|y_1 - \mu\|^\beta\right] D_m^{\frac{\beta}{2}+1}}{\mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)^{\frac{\beta}{2}} x^{\frac{\beta}{2}}}.$$

In order to use this deviation bound to obtain the inequalities

$$\mathbb{P}\left(\|(\Pi_m \otimes \Pi_m)(\mu - S_{vec})\|^2 \geqslant \frac{\alpha}{4}\mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)\right) \leqslant \tilde{C}\frac{1}{n^\gamma}$$

and

$$\mathbb{P}\left(\|(\Pi_m \otimes \Pi_m)(\mu - S_{vec})\| \geqslant \frac{\alpha}{8\|\mu\|}\mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)\right) \leqslant \tilde{C}\frac{1}{n^\gamma}$$

with $\gamma \geqslant \frac{q}{1 - 2q/\kappa}$, we need to find $x > 0$ satisfying the three following facts

$$\forall m \in \mathcal{M} \quad \frac{\alpha}{4} \geqslant \frac{1}{n}\left(1 + \sqrt{\frac{x}{D_m}}\right)^2 \tag{36}$$

$$\forall m \in \mathcal{M} \quad \left(\frac{\alpha}{8\|\mu\|}\right)^2 \mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi) \geqslant \left(\frac{\alpha}{8\|\mu\|}\right)^2 C_{inf} \geqslant \frac{1}{n}\left(1 + \sqrt{\frac{x}{D_m}}\right)^2 \tag{37}$$

$$\frac{D_m^{\frac{\beta}{2}+1}}{\mathrm{Tr}((\Pi_m \otimes \Pi_m)\Phi)^{\frac{\beta}{2}} x^{\frac{\beta}{2}}} = \frac{D_m}{\delta_m^\beta x^{\frac{\beta}{2}}} \leqslant C\frac{1}{n^\gamma}. \tag{38}$$

17

(36) and (37) hold for the choice $x = D_m n^r$ with $r < 1$ and if n is large enough to have

$$\frac{1}{n} \left(1 + \sqrt{n^r} \right)^2 \leqslant \frac{\alpha}{4} \tag{39}$$

and

$$\frac{1}{n} \left(1 + \sqrt{n^r} \right)^2 \leqslant \left(\frac{\alpha}{8 \, \|\mu\|} \right)^2 C_{inf} \tag{40}$$

In order to obtain (38) with $x = D_m n^r$, we use the inequality $D_m \leqslant n$ which gives

$$\frac{D_m}{\delta_m^\beta x^{\frac{\beta}{2}}} \leqslant \frac{1}{\delta_m^\beta D_m^{\frac{\beta}{2}} n^{r\beta/2-1}}.$$

Moreover

$$\mathbb{E} \left[\|y_1 - \mu\|^\beta \right] \leqslant \mathbb{E} \left[(\|y_1\| + \|\mu\|)^\beta \right],$$

and by using properties of convexity we obtain

$$\mathbb{E} \left[\|y_1 - \mu\|^\beta \right] \leqslant 2^{\beta-1} \left(\mathbb{E} \left[\|y_1\|^\beta \right] + \|\mu\|^\beta \right).$$

With the Jensen's inequality we get:

$$\mathbb{E} \left[\|y_1 - \mu\|^\beta \right] \leqslant 2^\beta \mathbb{E} \left[\|y_1\|^\beta \right].$$

In conclusion, with $r = \frac{\frac{\kappa}{2}+2}{\beta} < 1$ we obtain for $n \geqslant n(\kappa, \beta, \alpha, C_{inf}, \Sigma)$

$$Q2 \leqslant 2^{\beta+1} C_2(\beta) \frac{\mathbb{E} \left[\|xx^\top\|^\beta \right]}{\delta_m^\beta D_m^{\frac{\beta}{2}}} \frac{1}{n^{\kappa/4}} \tag{41}$$

where $C_2(\beta)$ is the constant which appears in Proposition 4.4.

In conclusion, combining (34) and (41)

$$\mathbb{P} \left(\widehat{\delta}_m^2 \leqslant (1 - \alpha) \delta_m^2 \right) \leqslant \frac{1}{n^{\kappa/4}} \left(C_2(\beta) 2^{\beta+1} + C_1(\beta) \frac{1}{\alpha^{\frac{\beta}{2}}} \right) \mathbb{E} \left[\|xx^\top\|^\beta \right] \delta_m^{-\beta} D_m^{-\frac{\beta}{2}}$$

for $n \geqslant n(\kappa, \beta, \alpha, C_{inf}, \Sigma)$.

To conclude, remark that $\frac{\kappa}{4} (1 - 2q/\kappa) = \frac{\kappa - 2q}{4} > \frac{2+2q}{4} \geqslant q$ as $q \leqslant 1$.

$\square$

Proof. **Lemma 5.3**

Recall that $\Pi_m \otimes \Pi_m$ is an orthogonal projection matrix. Hence there exists an orthogonal matrix P_m such that $P_m^\top (\Pi_m \otimes \Pi_m) P_m = D$, with D a diagonal matrix with $D_{ii} = 1$ if $i \leqslant D_m$, and $D_{ii} = 0$ otherwise. Then if Ψ is symmetric non-negative definite we have :

$$\mathrm{Tr} \left((\Pi_m \otimes \Pi_m) \Psi \right) = \mathrm{Tr} \left(D P_m^\top \Psi P_m \right)$$

$$= \sum_{l=1}^{p^2} \sum_{k=1}^{p^2} D_{kl} \left(P_m^\top \Psi P_m \right)_{kl} = \sum_{l=1}^{p^2} D_{ll} \left(P_m^\top \Psi P_m \right)_{ll}$$

$$= \sum_{l=1}^{D_m} \left(P_m^\top \Psi P_m \right)_{ll} \in [0; \mathrm{Tr}\,(\Psi)]\,.$$

Indeed, $P_m^\top \Psi P_m$ is non-negative definite so all its diagonal entries are non-negative. $\qquad \square$

References

[1] Y. Baraud. Model selection for regression on a fixed design. *Probability theory related fields*, 117(4):467–493, 2000.

[2] J. Bigot, R. Biscay, J.-M. Loubes, and L. M. Alvarez. Group lasso estimation of high-dimensional covariance matrices. *Journal of Machine Learning Resarch*, 2011.

[3] J. Bigot, R. Biscay, J.-M. Loubes, and L. Muñiz-Alvarez. Nonparametric estimation of covariance functions by model selection. *Electron. J. Stat.*, 4:822–855, 2010.

[4] J. Bigot, R. Biscay Lirio, J.-M. Loubes, and L. Muniz Alvarez. Adaptive estimation of spectral densities via wavelet thresholding and information projection. *preprint hal-00440424*, May 2010.

[5] N. A. C. Cressie. *Statistics for spatial data.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1993. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

[6] S. N. Elogne, O. Perrin, and C. Thomas-Agnan. Non parametric estimation of smooth stationary covariance functions by interpolation methods. *Stat. Inference Stoch. Process.*, 11(2):177–205, 2008.

[7] H. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[8] A. G. Journel. Kriging in terms of projections. *J. Internat. Assoc. Mathematical Geol.*, 9(6):563–586, 1977.

[9] G. A. F. Seber. *A matrix handbook for statisticians*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2008.

[10] M. L. Stein. *Interpolation of spatial data.* Springer Series in Statistics. Springer-Verlag, New York, 1999. Some theory for Kriging.

[11] B. von Bahr and C.-G. Esseen. Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$. *Ann. Math. Statist*, 36:299–303, 1965.