



HAL
open science

Fast estimation of the median covariation matrix with application to online robust principal components analysis

Hervé Cardot, Antoine Godichon

► **To cite this version:**

Hervé Cardot, Antoine Godichon. Fast estimation of the median covariation matrix with application to online robust principal components analysis. *Test*, 2017, 26 (3), pp.461-480. 10.1007/s11749-016-0519-x . hal-01482743

HAL Id: hal-01482743

<https://u-bourgogne.hal.science/hal-01482743>

Submitted on 30 Jan 2024

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

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

Fast Estimation of the Median Covariation Matrix with Application to Online Robust Principal Components Analysis

Hervé CARDOT, Antoine GODICHON-BAGGIONI
Institut de Mathématiques de Bourgogne,
Université de Bourgogne Franche-Comté,
9, rue Alain Savary, 21078 Dijon, France

July 12, 2016

Abstract

The geometric median covariation matrix is a robust multivariate indicator of dispersion which can be extended without any difficulty to functional data. We define estimators, based on recursive algorithms, that can be simply updated at each new observation and are able to deal rapidly with large samples of high dimensional data without being obliged to store all the data in memory. Asymptotic convergence properties of the recursive algorithms are studied under weak conditions. The computation of the principal components can also be performed online and this approach can be useful for online outlier detection. A simulation study clearly shows that this robust indicator is a competitive alternative to minimum covariance determinant when the dimension of the data is small and robust principal components analysis based on projection pursuit and spherical projections for high dimension data. An illustration on a large sample and high dimensional dataset consisting of individual TV audiences measured at a minute scale over a period of 24 hours confirms the interest of considering the robust principal components analysis based on the median covariation matrix. All studied algorithms are available in the R package `Gmedian` on CRAN.

Keywords. Averaging, Functional data, Geometric median, Online algorithms, Online principal components, Recursive robust estimation, Stochastic gradient, Weiszfeld's algorithm.

1 Introduction

Principal Components Analysis is one of the most useful statistical tool to extract information by reducing the dimension when one has to analyze large samples of multivariate or functional data (see *e.g.* Jolliffe (2002) or Ramsay and Silverman (2005)). When both the dimension and the sample size are large, outlying observations may be difficult to detect automatically. Principal components, which are derived from the spectral analysis of the

covariance matrix, can be very sensitive to outliers (see Devlin et al. (1981)) and many robust procedures for principal components analysis have been considered in the literature (see Hubert et al. (2008), Huber and Ronchetti (2009) and Maronna et al. (2006)).

The most popular approaches are probably the minimum covariance determinant estimator (see Rousseeuw and van Driessen (1999)) and the robust projection pursuit (see Croux and Ruiz-Gazen (2005) and Croux et al. (2007)). Robust PCA based on projection pursuit has been extended to deal with functional data in Hyndman and Ullah (2007) and Bali et al. (2011). Adopting another point of view, robust modifications of the covariance matrix, based on projection of the data onto the unit sphere, have been proposed in Locantore et al. (1999) (see also Gervini (2008) and Taskinen et al. (2012)).

We consider in this work another robust way of measuring association between variables, that can be extended directly to functional data. It is based on the notion of median covariation matrix (MCM) which is defined as the minimizer of an expected loss criterion based on the Hilbert-Schmidt norm (see Kraus and Panaretos (2012) for a first definition in a more general M -estimation setting). It can be seen as a geometric median (see Kemperman (1987) or Möttönen et al. (2010)) in the particular Hilbert spaces of square matrices (or operators for functional data) equipped with the Frobenius (or Hilbert-Schmidt) norm. The MCM is non negative and unique under weak conditions. As shown in Kraus and Panaretos (2012) it also has the same eigenspace as the usual covariance matrix when the distribution of the data is symmetric and the second order moment is finite. Being a spatial median in a particular Hilbert space of matrices, the MCM is also a robust indicator of central location, among the covariance matrices, which has a 50 % breakdown point (see Kemperman (1987) or Maronna et al. (2006)) as well as a bounded gross sensitivity error (see Cardot et al. (2013)).

The aim of this work is twofold. It provides efficient recursive estimation algorithms of the MCM that are able to deal with large samples of high dimensional data. By this recursive property, these algorithms can naturally deal with data that are observed sequentially and provide a natural update of the estimators at each new observation. Another advantage compared to classical approaches is that such recursive algorithms will not require to store all the data. Secondly, this work also aims at highlighting the interest of considering the median covariation matrix to perform principal components analysis of high dimensional contaminated data.

Different algorithms can be considered to get effective estimators of the MCM. When the dimension of the data is not too high and the sample size is not too large, Weiszfeld's algorithm (see Weiszfeld (1937) and Vardi and Zhang (2000)) can be directly used to estimate effectively both the geometric median and the median covariation matrix. When both the dimension and the sample size are large this static algorithm which requires to store all the data may be inappropriate and ineffective. We show how the algorithm developed by Cardot et al. (2013) for the geometric median in Hilbert spaces can be adapted to estimate recursively and simultaneously the median as well as the median covariation matrix. Then an averaging step (Polyak and Juditsky (1992)) of the two initial recursive

estimators of the median and the MCM permits to improve the accuracy of the initial stochastic gradient algorithms. A simple modification of the stochastic gradient algorithm is proposed in order to ensure that the median covariance estimator is non negative. We also explain how the eigenelements of the estimator of the MCM can be updated online without being obliged to perform a new spectral decomposition at each new observation.

The paper is organized as follows. The median covariation matrix as well as the recursive estimators are defined in Section 2. In Section 3, almost sure and quadratic mean consistency results are given for variables taking values in general separable Hilbert spaces. The proofs, which are based on new induction steps compared to Cardot et al. (2013), allow to get better convergence rates in quadratic mean even if this new framework is much more complicated because two averaged non linear algorithms are running simultaneously. One can also note that the techniques generally employed to deal with two time scale Robbins Monro algorithms (see Mokkadem and Pelletier (2006) for the multivariate case) require assumptions on the rest of the Taylor expansion and the finite dimension of the data that are too restrictive in our framework. In Section 4, a comparison with some classic robust PCA techniques is made on simulated data. The interest of considering the MCM is also highlighted on the analysis of individual TV audiences, a large sample of high dimensional data which, because of its dimension, can not be analyzed in a reasonable time with classical robust PCA approaches. The main parts of the proofs are described in Section 5. Perspectives for future research are discussed in Section 6. Some technical parts of the proofs as well as a description of Weiszfeld's algorithm in our context are gathered in an Appendix.

2 Population point of view and recursive estimators

Let H be a separable Hilbert space (for example $H = \mathbb{R}^d$ or $H = L^2(I)$, for some closed interval $I \subset \mathbb{R}$). We denote by $\langle \cdot, \cdot \rangle$ its inner product and by $\|\cdot\|$ the associated norm.

We consider a random variable X that takes values in H and define its center $m \in H$ as follows:

$$m := \arg \min_{u \in H} \mathbb{E} [\|X - u\| - \|X\|]. \quad (1)$$

The solution $m \in H$ is often called the geometric median of X . It is uniquely defined under broad assumptions on the distribution of X (see Kemperman (1987)) which can be expressed as follows.

Assumption 1. *There exist two linearly independent unit vectors $(u_1, u_2) \in H^2$, such that*

$$\mathbf{Var}(\langle u, X \rangle) > 0, \quad \text{for } u \in \{u_1, u_2\}.$$

If the distribution of $X - m$ is symmetric around zero and if X admits a first moment that is finite then the geometric median is equal to the expectation of X , $m = \mathbb{E}[X]$. Note however that the general definition (1) does not require to assume that the first order moment of $\|X\|$ is finite since $|\mathbb{E} [\|X - u\| - \|X\|] | \leq \|u\|$.

2.1 The (geometric) median covariation matrix (MCM)

We now consider the special vector space, denoted by $\mathcal{S}(H)$, of $d \times d$ matrices when $H = \mathbb{R}^d$, or for general separable Hilbert spaces H , the vector space of linear operators mapping $H \rightarrow H$. Denoting by $\{e_j, j \in J\}$ an orthonormal basis in H , the vector space $\mathcal{S}(H)$ equipped with the following inner product:

$$\langle A, B \rangle_F = \sum_{j \in J} \langle Ae_j, Be_j \rangle \quad (2)$$

is also a separable Hilbert space. In $\mathcal{S}(\mathbb{R}^d)$, we have equivalently

$$\langle A, B \rangle_F = \text{tr}(A^T B), \quad (3)$$

where A^T is the transpose matrix of A . The induced norm is the well known Frobenius norm (also called Hilbert-Schmidt norm) and is denoted by $\|\cdot\|_F$.

When X has finite second order moments, with expectation $\mathbb{E}[X] = \mu$, the covariance matrix of X , $\mathbb{E}[(X - \mu)(X - \mu)^T]$ can be defined as the minimum argument, over all the elements belonging to $\mathcal{S}(H)$, of the functional $G_{\mu,2} : \mathcal{S}(H) \rightarrow \mathbb{R}$,

$$G_{\mu,2}(\Gamma) = \mathbb{E} \left[\left\| (X - \mu)(X - \mu)^T - \Gamma \right\|_F^2 - \left\| (X - \mu)(X - \mu)^T \right\|_F^2 \right].$$

Note that in general Hilbert spaces with inner product $\langle \cdot, \cdot \rangle$, operator $(X - \mu)(X - \mu)^T$ should be understood as the operator $u \in H \mapsto \langle u, X - \mu \rangle (X - \mu)$. The MCM is obtained by removing the squares in previous function in order to get a more robust indicator of "covariation". For $\alpha \in H$, define $G_\alpha : \mathcal{S}(H) \rightarrow \mathbb{R}$ by

$$G_\alpha(V) := \mathbb{E} \left[\left\| (X - \alpha)(X - \alpha)^T - V \right\|_F - \left\| (X - \alpha)(X - \alpha)^T \right\|_F \right]. \quad (4)$$

The median covariation matrix, denoted by Γ_m , is defined as the minimizer of $G_m(V)$ over all elements $V \in \mathcal{S}(H)$. The second term at the right-hand side of (4) prevents from having to introduce hypotheses on the existence of the moments of X . Introducing the random variable $Y := (X - m)(X - m)^T$ that takes values in $\mathcal{S}(H)$, the MCM is unique provided that the support of Y is not concentrated on a line and Assumption 1 can be rephrased as follows in $\mathcal{S}(H)$,

Assumption 2. *There exist two linearly independent unit vectors $(V_1, V_2) \in \mathcal{S}(H)^2$, such that*

$$\mathbf{Var}(\langle V, Y \rangle_F) > 0, \quad \text{for } V \in \{V_1, V_2\}.$$

We can remark that Assumption 1 and Assumption 2 are strongly connected. Indeed, if Assumption 1 holds, then $\mathbf{Var}(\langle u, X \rangle) > 0$ for $u \in \{u_1, u_2\}$. Consider the rank one matrices $V_1 = u_1 u_1^T$ and $V_2 = u_2 u_2^T$, we have $\langle V_1, Y \rangle_F = \langle u_1, X - m \rangle^2$ which has a strictly positive variance when the distribution of X has no atom. More generally $\mathbf{Var}(\langle V_1, Y \rangle_F) > 0$ unless there is a scalar $a > 0$ such that $\mathbb{P}[\langle u_1, X - m \rangle = a] = \mathbb{P}[\langle u_1, X - m \rangle = -a] = \frac{1}{2}$ (assuming also that $\mathbb{P}[X - m = 0] = 0$).

Furthermore it can be deduced easily that the MCM, which is a geometric median in the particular Hilbert spaces of Hilbert-Schmidt operators, is a robust indicator with a 50% breakdown point (see Kemperman (1987)) and a bounded sensitive gross error (see Cardot et al. (2013)).

We also assume that

Assumption 3. *There is a constant C such that for all $h \in H$ and all $V \in \mathcal{S}(H)$*

$$\begin{aligned} (a) : \quad & \mathbb{E} \left[\left\| (X - h)(X - h)^T - V \right\|_F^{-1} \right] \leq C. \\ (b) : \quad & \mathbb{E} \left[\left\| (X - h)(X - h)^T - V \right\|_F^{-2} \right] \leq C. \end{aligned}$$

This assumption implicitly forces the distribution of $(X - h)(X - h)^T$ to have no atoms. It is more "likely" to be satisfied when the dimension d of the data is large (see Chaudhuri (1992) and Cardot et al. (2013) for a discussion). Note that it could be weakened as in Cardot et al. (2013) by allowing points, necessarily different from the MCM Γ_m , to have strictly positive masses. Considering the particular case $V = 0$, Assumption 3(a) implies that for all $h \in H$,

$$\mathbb{E} \left[\frac{1}{\|X - h\|^2} \right] \leq C, \quad (5)$$

and this is not restrictive when the dimension d of H is equal or larger than 3.

Under Assumption 3(a), the functional G_h is twice Fréchet differentiable, with gradient

$$\nabla G_h(V) = -\mathbb{E} \left[\frac{(X - h)(X - h)^T - V}{\left\| (X - h)(X - h)^T - V \right\|_F} \right]. \quad (6)$$

and Hessian operator, $\nabla_h^2 G(V) : \mathcal{S}(H) \rightarrow \mathcal{S}(H)$,

$$\nabla_h^2 G(V) = \mathbb{E} \left[\frac{1}{\|Y(h) - V\|_F} \left(I_{\mathcal{S}(H)} - \frac{(Y(h) - V) \otimes_F (Y(h) - V)}{\|Y(h) - V\|_F^2} \right) \right]. \quad (7)$$

where $Y(h) = (X - h)(X - h)^T$, $I_{\mathcal{S}(H)}$ is the identity operator on $\mathcal{S}(H)$ and $A \otimes_F B(V) = \langle A, V \rangle_F B$ for any elements A, B and V belonging to $\mathcal{S}(H)$.

Furthermore, Γ_m is also defined as the unique zero of the non linear equation:

$$\nabla G_m(\Gamma_m) = 0. \quad (8)$$

Remarking that previous equality can be rewritten as follows,

$$\Gamma_m = \frac{1}{\mathbb{E} \left[\frac{1}{\|(X - m)(X - m)^T - \Gamma_m\|_F} \right]} \mathbb{E} \left[\frac{(X - m)(X - m)^T}{\|(X - m)(X - m)^T - \Gamma_m\|_F} \right], \quad (9)$$

it is clear that Γ_m is a bounded, symmetric and non negative operator in $\mathcal{S}(H)$.

As stated in Proposition 2 of Kraus and Panaretos (2012), operator Γ_m has an important stability property when the distribution of X is symmetric, with finite second moment,

i.e. $\mathbb{E} [\|X\|^2] < \infty$. Indeed, the covariance operator of X , $\Sigma = \mathbb{E} [(X - m)(X - m)^T]$, which is well defined in this case, and Γ_m share the same eigenvectors: if e_j is an eigenvector of Σ with corresponding eigenvalue λ_j , then $\Gamma_m e_j = \tilde{\lambda}_j e_j$, for some non negative value $\tilde{\lambda}_j$. This important result means that for Gaussian and more generally symmetric distribution (with finite second order moments), the covariance operator and the median covariation operator have the same eigenspaces. Note that it is also conjectured in Kraus and Panaretos (2012) that the order of the eigenfunctions is also the same.

2.2 Efficient recursive algorithms

We suppose now that we have i.i.d. copies X_1, \dots, X_n, \dots of random variables with the same law as X .

For simplicity, we temporarily suppose that the median m of X is known. We consider a sequence of (learning) weights $\gamma_n = c_\gamma/n^\alpha$, with $c_\gamma > 0$ and $1/2 < \alpha < 1$ and we define the recursive estimation procedure as follows

$$W_{n+1} = W_n + \gamma_n \frac{(X_{n+1} - m)(X_{n+1} - m)^T - W_n}{\|(X_{n+1} - m)(X_{n+1} - m)^T - W_n\|_F} \quad (10)$$

$$\bar{W}_{n+1} = \bar{W}_n - \frac{1}{n+1} (\bar{W}_n - W_{n+1}). \quad (11)$$

This algorithm can be seen as a particular case of the averaged stochastic gradient algorithm studied in Cardot et al. (2013). Indeed, the first recursive algorithm (10) is a stochastic gradient algorithm,

$$\mathbb{E} \left[\frac{(X_{n+1} - m)(X_{n+1} - m)^T - W_n}{\|(X_{n+1} - m)(X_{n+1} - m)^T - W_n\|_F} \middle| \mathcal{F}_n \right] = \nabla G_m(W_n)$$

where $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ is the σ -algebra generated by X_1, \dots, X_n whereas the final estimator \bar{W}_n is obtained by averaging the past values of the first algorithm. The averaging step (see Polyak and Juditsky (1992)), *i.e.* the computation of the arithmetical mean of the past values of a slowly convergent estimator (see Proposition 3.4 below), permits to obtain a new and efficient estimator converging at a parametric rate, with the same asymptotic variance as the empirical risk minimizer (see Theorem 3.1 below).

In most of the cases the value of m is unknown so that it also required to estimate the median. To build an estimator of Γ_m , it is possible to estimate simultaneously m and Γ_m by considering two averaged stochastic gradient algorithms that are running simultaneously. For $n \geq 1$,

$$m_{n+1} = m_n + \gamma_n^{(m)} \frac{X_{n+1} - m_n}{\|X_{n+1} - m_n\|} \quad (12)$$

$$\bar{m}_{n+1} = \bar{m}_n - \frac{1}{n+1} (\bar{m}_n - m_{n+1}) \quad (13)$$

$$V_{n+1} = V_n + \gamma_n \frac{(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n}{\|(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n\|_F} \quad (14)$$

$$\bar{V}_{n+1} = \bar{V}_n - \frac{1}{n+1} (\bar{V}_n - V_{n+1}), \quad (14)$$

where the averaged recursive estimator \bar{m}_{n+1} of the median m is controlled by a sequence of descent steps $\gamma_n^{(m)}$. The learning rates are generally chosen as follows, $\gamma_n^{(m)} = c_m n^{-\alpha}$, where the tuning constants satisfy $c_m \in [2, 20]$ and $1/2 < \alpha < 1$.

Note that by construction, even if V_n is non negative, V_{n+1} may not be a non negative matrix when the learning steps do not satisfy

$$\frac{\gamma_n}{\|(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n\|_F} \leq 1.$$

Projecting V_{n+1} onto the closed convex cone of non negative operators would require to compute the eigenvalues of V_{n+1} which is time consuming in high dimension even if V_{n+1} is a rank one perturbation to V_n (see Cardot and Degras (2015)). We consider the following simple approximation to this projection which consists in replacing in (13) the descent step γ_n by a thresholded one,

$$\gamma_{n,pos} = \min(\gamma_n, \|(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n\|_F) \quad (15)$$

which ensures that V_{n+1} remains non negative when V_n is non negative. The use of these modified steps and an initialization of the recursive algorithm (13) with a non negative matrix (for example $V_0 = 0$) ensure that for all $n \geq 1$, V_n and \bar{V}_n are non negative.

2.3 Online estimation of the principal components

It is also possible to approximate recursively the q eigenvectors (unique up to sign) of Γ_m associated to the q largest eigenvalues without being obliged to perform a spectral decomposition of \bar{V}_{n+1} at each new observation. Many recursive strategies can be employed (see Cardot and Degras (2015) for a review on various recursive estimation procedures of the eigenlements of a covariance matrix). Because of its simplicity and its accuracy, we consider the following one:

$$u_{j,n+1} = u_{j,n} + \frac{1}{n+1} \left(\bar{V}_{n+1} \frac{u_{j,n}}{\|u_{j,n}\|} - u_{j,n} \right), \quad j = 1, \dots, q \quad (16)$$

combined with an orthogonalization by deflation of $u_{1,n+1}, \dots, u_{q,n+1}$. This recursive algorithm is based on ideas developed by Weng et al. (2003) that are related to the power method for extracting eigenvectors. If we assume that the q first eigenvalues $\lambda_1 > \dots > \lambda_q$ are distinct, the estimated eigenvectors $u_{1,n+1}, \dots, u_{q,n+1}$, which are uniquely determined up to sign change, tend to $\lambda_1 u_1, \dots, \lambda_q u_q$.

Once the eigenvectors are computed, it is possible to compute the principal components as well as indices of outlyingness for each new observation (see Hubert et al. (2008) for a review of outliers detection with multivariate approaches).

2.4 Practical issues, complexity and memory

The recursive algorithms (13) and (14) require each $O(d^2)$ elementary operations at each update. With the additional online estimation given in (16) of the q eigenvectors associated

to the q largest eigenvalues, $O(qd^2)$ additional operations are required. The orthogonalization procedure only requires $O(q^2d)$ elementary operations.

Note that the use of classical Newton-Raphson algorithms for estimating the MCM (see Fritz et al. (2012)) can not be envisaged for high dimensional data since the computation or the approximation of the Hessian matrix would require $O(d^4)$ elementary operations. The well known and fast Weiszfeld's algorithm requires $O(nd^2)$ elementary operations for each sample with size n . However, the estimation cannot be updated automatically if the data arrive sequentially. Another drawback compared to the recursive algorithms studied in this paper is that all the data must be stored in memory, which is of order $O(nd^2)$ elements whereas the recursive technique require an amount of memory of order $O(d^2)$.

The performances of the recursive algorithms depend on the values of tuning parameters c_γ , c_m and α . The value of parameter α is often chosen to be $\alpha = 2/3$ or $\alpha = 3/4$. Previous empirical studies (see Cardot et al. (2013) and Cardot et al. (2010)) have shown that, thanks to the averaging step, estimator \bar{m}_n performs well and is not too sensitive to the choice of c_m , provided that the value of c_m is not too small. An intuitive explanation could be that here the recursive process is in some sense "self-normalized" since the deviations at each iteration in (10) have unit norm and finding some universal values for c_m is possible. Usual values for c_m and c_γ are in the interval $[2, 20]$. When n is fixed, this averaged recursive algorithm is about 30 times faster than the Weiszfeld's approach (see Cardot et al. (2013)).

3 Asymptotic properties

When m is known, \bar{W}_n can be seen as an averaged stochastic gradient estimator of the geometric median in a particular Hilbert space and the asymptotic weak convergence of such estimator has been studied in Cardot et al. (2013). They have shown that:

Theorem 3.1. (Cardot et al. (2013), Theorem 3.4).

If assumptions 1-3(a) hold, then as n tends to infinity,

$$\sqrt{n} (\bar{W}_n - \Gamma_m) \rightsquigarrow \mathcal{N}(0, \Delta)$$

where \rightsquigarrow stands for convergence in distribution and $\Delta = (\nabla_m^2(\Gamma_m))^{-1} \Psi (\nabla_m^2(\Gamma_m))^{-1}$ is the limiting covariance operator, with $\Psi = \mathbb{E} \left[\frac{(Y(m) - \Gamma_m) \otimes_F (Y(m) - \Gamma_m)}{\|Y(m) - \Gamma_m\|_F^2} \right]$.

As explained in Cardot et al. (2013), the estimator \bar{W}_n is efficient in the sense that it has the same asymptotic distribution as the empirical risk minimizer related to $G_m(V)$ (see for the derivation of its asymptotic normality in Möttönen et al. (2010) in the multivariate case and Chakraborty and Chaudhuri (2014) in a more general functional framework).

Using the delta method for weak convergence in Hilbert spaces (see Dauxois et al. (1982) or Cupidon et al. (2007)), one can deduce, from Theorem 3.1, the asymptotic normality of the estimated eigenvectors of \bar{W}_n . It can also be proven (see Godichon-Baggioni (2016)), under Assumptions 1-3, that there is a positive constant K such that for all $n \geq 1$,

$$\mathbb{E} \left[\|\bar{W}_n - \Gamma_m\|_F^2 \right] \leq \frac{K}{n}.$$

Note finally that non asymptotic bounds for the deviation of \bar{W}_n around Γ_m can be derived readily with the general results given in Cardot et al. (2016).

The more realistic case in which m must also be estimated is more complicated because \bar{V}_n depends on \bar{m}_n which is also estimated recursively with the same data. We first state the strong consistency of the estimators V_n and \bar{V}_n .

Theorem 3.2. *If assumptions 1-3(b) hold, we have*

$$\lim_{n \rightarrow \infty} \|V_n - \Gamma_m\|_F = 0 \quad a.s.$$

and

$$\lim_{n \rightarrow \infty} \|\bar{V}_n - \Gamma_m\|_F = 0 \quad a.s.$$

The obtention of the rate convergence of the averaged recursive algorithm relies on a fine control of the asymptotic behavior of the Robbins-Monro algorithms, as stated in the following proposition.

Theorem 3.3. *If assumptions 1-3(b) hold, there is a positive constant C' , and for all $\beta \in (\alpha, 2\alpha)$, there is a positive constant C_β such that for all $n \geq 1$,*

$$\begin{aligned} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] &\leq \frac{C'}{n^\alpha}, \\ \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] &\leq \frac{C''}{n^\beta}. \end{aligned}$$

The obtention of an upper bound for the rate of convergence at the order four of the Robbins-Monro algorithm is crucial in the proofs. Furthermore, the following proposition ensures that the exhibited rate in quadratic mean is the optimal one.

Proposition 3.4. *Under assumptions 1-3(b), there is a positive constant c' such that for all $n \geq 1$,*

$$\mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] \geq \frac{c'}{n^\alpha}.$$

Finally, the following theorem is the most important theoretical result of this work. It shows that, in spite of the fact that it only considers the observed data one by one, the averaged recursive estimation procedure gives an estimator which has a classical parametric \sqrt{n} rate of convergence in the Hilbert-Schmidt norm.

Theorem 3.5. *Under Assumptions 1-3(b), there is a positive constant K' such that for all $n \geq 1$,*

$$\mathbb{E} \left[\|\bar{V}_n - \Gamma_m\|_F^2 \right] \leq \frac{K'}{n}.$$

Assuming the eigenvalues of Γ_m are of multiplicity one, it can be deduced from Theorem 3.5 and Lemma 4.3 in Bosq (2000), the convergence in quadratic mean of the eigenvectors of \bar{V}_n towards the corresponding (up to sign) eigenvector of Γ_m .

4 An illustration on simulated and real data

A small comparison with other classical robust PCA techniques is performed in this section considering data in relatively high dimension but samples with moderate sizes. This permits to compare our approach with classical robust PCA techniques, which are generally not designed to deal with large samples of high dimensional data. In our comparison, we have employed the following well known robust techniques: robust projection pursuit (see Croux and Ruiz-Gazen (2005) and Croux et al. (2007)), minimum covariance determinant (MCD, see Rousseeuw and van Driessen (1999)) and spherical PCA (see Locantore et al. (1999)). The computations were made in the R language (R Development Core Team (2010)), with the help of packages `pcaPP` and `rrcov`. For reproducible research, our codes for computing the MCM have been posted on CRAN in the `Gmedian` package. We will denote by $\text{MCM}(\text{R})$ the recursive estimator \bar{V}_n defined in (14) and $\text{MCM}(\text{R}+)$ its non negative modification whose learning weights are defined in (15).

If the size of the data $n \times d$ is not too large, an effective way for estimating Γ_m is to employ Weiszfeld's algorithm (see Weiszfeld (1937) and Vardi and Zhang (2000) as well the Supplementary file for a description of the algorithms in our particular situation). The estimate obtained thanks to Weiszfeld's algorithm is denoted by $\text{MCM}(\text{W})$ in the following. Note that other optimization algorithms which may be preferred in small dimension (see Fritz et al. (2012)) have not been considered here since they would require the computation of the Hessian matrix, whose size is d^4 , and this would lead to much slower algorithms. Note finally that all these alternative algorithms do not admit a natural updating scheme when the data arrive sequentially so that they should be completely ran again at each new observation.

4.1 Simulation protocol

Independent realizations of a random variable $Y \in \mathbb{R}^d$ are drawn, where

$$Y = (1 - O(\delta))X + O(\delta)\epsilon, \quad (17)$$

is a mixture of two distributions and X, O and ϵ are independent random variables. The random vector X has a centered Gaussian distribution in \mathbb{R}^d with covariance matrix $[\Sigma]_{\ell,j} = \min(\ell, j)/d$ and can be thought as a discretized version of a Brownian sample path in $[0, 1]$. The multivariate contamination comes from ϵ , with different rates of contamination controlled by the Bernoulli variable $O(\delta)$, independent from X and ϵ , with $\mathbb{P}(O(\delta) = 1) = \delta$ and $\mathbb{P}(O(\delta) = 0) = 1 - \delta$. Three different scenarios (see Figure 1) are considered for the distribution of ϵ :

- The elements of vector ϵ are d independent realizations of a Student t distribution with one degree of freedom. This means that the first moment of Y is not defined when $\delta > 0$.
- The elements of vector ϵ are d independent realizations of a Student t distribution

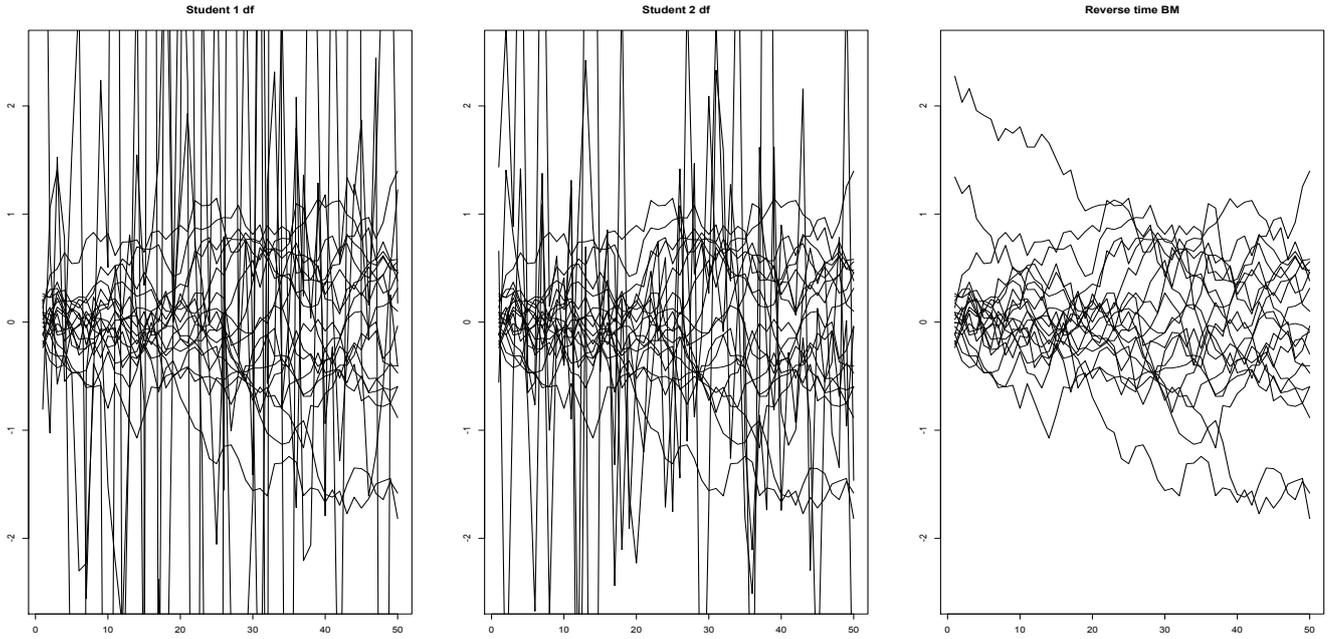


Figure 1: A sample of $n = 20$ trajectories when $d = 50$ and $\delta = 0.10$ for the three different contamination scenarios: Student t with 1 degree of freedom, Student t with 2 degrees of freedom and reverse time Brownian motion (from left to right).

with two degrees of freedom. This means that the second moment of Y is not defined when $\delta > 0$.

- The vector ϵ is distributed as a "reverse time" Brownian motion. It has a Gaussian centered distribution, with covariance matrix $[\Sigma_\epsilon]_{\ell,j} = 2 \min(d - \ell, d - j)/d$. The covariance matrix of Y is $(1 - \delta)\Sigma + \delta\Sigma_\epsilon$.

For the averaged recursive algorithms, we have considered tuning coefficients $c_m = c_\gamma = 2$ and a speed rate of $\alpha = 3/4$. Note that the values of these tuning parameters have not been particularly optimised. We have noted that the simulation results were very stable, and did not depend much on the value of c_m and c_γ for $c_m, c_\gamma \in [1, 20]$.

The estimation error of the eigenspaces associated to the largest eigenvalues is evaluated by considering the squared Frobenius norm between the associated orthogonal projectors. Denoting by \mathbf{P}_q the orthogonal projector onto the space generated by the q eigenvectors of the covariance matrix Σ associated to the q largest eigenvalues and by $\widehat{\mathbf{P}}_q$ an estimation, we consider the following loss criterion,

$$\begin{aligned}
 R(\widehat{\mathbf{P}}_q, \mathbf{P}_q) &= \text{tr} \left[\left(\widehat{\mathbf{P}}_q - \mathbf{P}_q \right)^T \left(\widehat{\mathbf{P}}_q - \mathbf{P}_q \right) \right] \\
 &= 2q - 2\text{tr} \left[\widehat{\mathbf{P}}_q \mathbf{P}_q \right].
 \end{aligned} \tag{18}$$

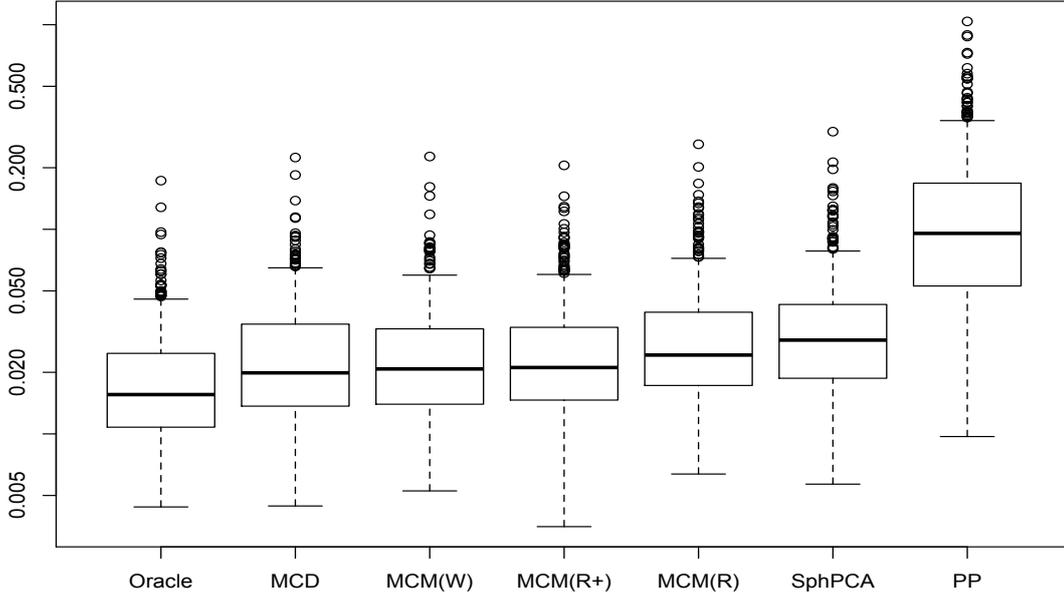


Figure 2: Estimation errors (at a logarithmic scale) over 500 Monte Carlo replications, for $n = 200$, $d = 50$ with no contamination ($\delta = 0$). MCM(W) stands for the estimation performed by the Weiszfeld's algorithm whereas MCM(R) denotes the averaged recursive approach and MCM(R+) its non negative modification (see equation 15).

Note that we always have $R(\hat{\mathbf{P}}_q, \mathbf{P}_q) \leq 2q$ and $R(\hat{\mathbf{P}}_q, \mathbf{P}_q) = 2q$ means that the eigenspaces generated by the true and the estimated eigenvectors are orthogonal.

4.2 Comparison with classical robust PCA techniques

We first compare the performances of the two estimators of the MCM based on the Weiszfeld's algorithm and the recursive algorithms (see (14)) with more classical robust PCA techniques.

We generated samples of Y with size $n = 500$ and dimension $d \in \{50, 200\}$, over 500 replications. Different levels of contamination are considered : $\delta \in \{0, 0.02, 0.05, 0.10, 0.20\}$. For both dimensions $d = 50$ and $d = 200$, the first eigenvalue of the covariance matrix of X represents about 81 % of the total variance, and the second one about 9 %.

The median errors of estimation of the eigenspace generated by the first two eigenvectors ($q = 2$), according to criterion (18), are given in Table 1 for non contaminated data ($\delta = 0$). The distribution of the estimation error $R(\hat{\mathbf{P}}_q, \mathbf{P}_q)$ is drawn for the different approaches in Figure 2 when the dimension is not large ($d = 50$). As expected, the "Oracle", which is the classical PCA in this situation, provides the best estimations of the eigenspaces. Then, the MCD and the median covariation matrix, estimated by the Weiszfeld algorithm

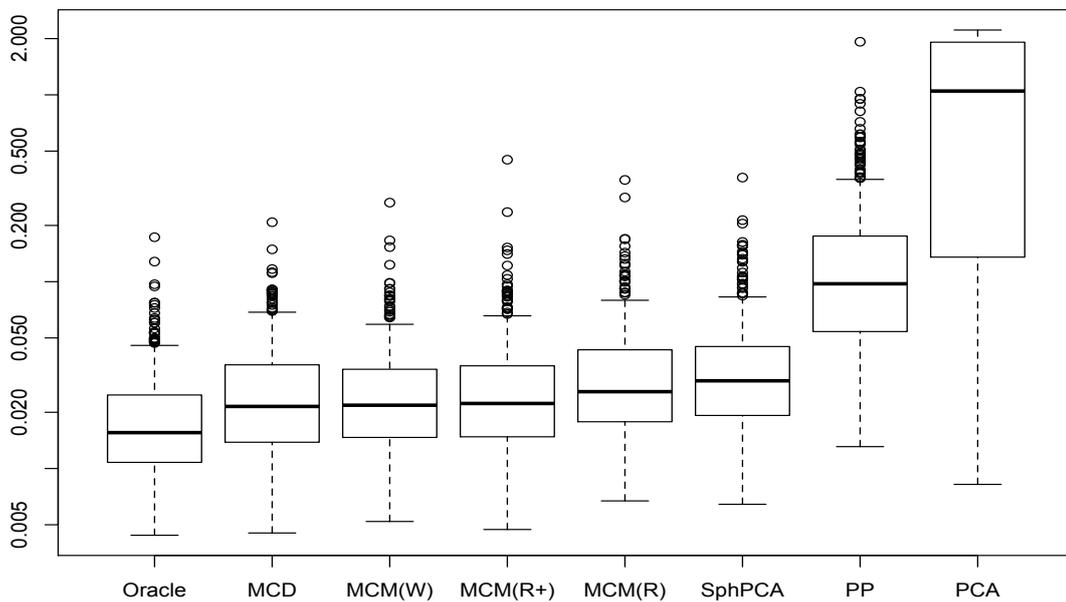


Figure 3: Estimation errors (at a logarithmic scale) over 500 Monte Carlo replications, for $n = 200$, $d = 50$ and a contamination by a t distribution with 2 degrees of freedom with $\delta = 0.02$. MCM(W) stands for the estimation performed by the Weiszfeld's algorithm whereas MCM(R) denotes the averaged recursive approach and MCM(R+), its non negative modification with learning steps as in (15).

| | PCA | MCD | MCM(W) | MCM(R+) | MCM(R) | SphPCA | PP |
|-------|--------|--------|--------|---------|--------|--------|--------|
| d=50 | 0.0156 | 0.0199 | 0.0208 | 0.0211 | 0.0243 | 0.0287 | 0.0955 |
| d=200 | 0.0148 | - | 0.0200 | 0.0209 | 0.0246 | 0.0275 | 0.0895 |

Table 1: Median estimation errors, according to criterion $R(\hat{\mathbf{P}}_q, \mathbf{P}_q)$ with a dimension $q = 2$, for non contaminated samples of size $n = 200$, over 500 Monte Carlo experiments.

or the modified MCM(R+) recursive estimator, behave well and similarly. Note that when the dimension gets larger, the MCD cannot be used anymore and the MCM is the more effective robust estimator of the eigenspaces.

When the data are contaminated, the median errors of estimation of the eigenspace generated by the first two eigenvectors ($q = 2$), according to criterion (18), are given in Table 2. In Figure 3, the distribution of the estimation error $R(\hat{\mathbf{P}}_q, \mathbf{P}_q)$ is drawn for the different approaches.

We can make the following remarks. At first note that even when the level of contamination is small (2% and 5%), the performances of classical PCA are strongly affected by the presence of outlying values in such (large) dimensions. When $d = 50$, the MCD algorithm and the MCM estimation provide the best estimations of the original two dimensional eigenspace, whereas when d gets larger ($d = n = 200$), the MCD estimator can not be used anymore (by construction) and the MCM estimators, obtained with Weiszfeld's and the non negative recursive algorithm, remain the most accurate. We can also remark that the recursive MCM algorithms, which are designed to deal with very large samples, performs well even for such moderate sample sizes (see also Figure 3). The modification of the descent step suggested in (15), which corresponds to estimator MCM(R+), permits to improve the accuracy the initial MCM estimator, specially when the noise level is not small. The performances of the spherical PCA are slightly less accurate whereas the median error of the robust PP is always the largest among the robust estimators. When, the contamination is highly structured temporally and the level of contamination is not small (contamination by a reverse time Brownian motion, with $\delta = 0.20$), the behavior of the MCM is different from the other robust estimators and, with our criterion, it can appear as less effective. However, one can think that we are in presence of two different populations with completely different multivariate correlation structure and the MCD completely ignores that part of the data, which is not necessarily a better behavior.

4.3 Online estimation of the principal components

We now consider an experiment in high dimension, $d = 1000$, and evaluate the ability of the recursive algorithms defined in (16) to estimate recursively the eigenvectors of Γ_m associated to the largest eigenvalues. Note that due to the high dimension of the data and limited computation time, we only make comparison of the recursive robust techniques with the classical PCA. For this we generate growing samples and compute, for each sample size the approximation error of the different (fast) strategies to the true eigenspace generated

| δ | Method | t 1 df | t 2 df | inv. B. | t 1 df | t 2 df | inv. B. |
|----------|-----------------|----------|----------|---------|----------|----------|---------|
| | | d = 50 | | | d = 200 | | |
| 2% | PCA | 3.13 | 1.04 | 0.698 | 3.95 | 1.87 | 0.731 |
| | PP | 0.086 | 0.097 | 0.090 | 0.085 | 0.094 | 0.084 |
| | MCD | 0.022 | 0.021 | 0.021 | – | – | – |
| | Sph. PCA | 0.028 | 0.029 | 0.027 | 0.027 | 0.030 | 0.028 |
| | MCM (Weiszfeld) | 0.021 | 0.021 | 0.021 | 0.021 | 0.022 | 0.022 |
| | MCM (R+) | 0.022 | 0.022 | 0.024 | 0.023 | 0.023 | 0.025 |
| | MCM (R) | 0.026 | 0.025 | 0.027 | 0.026 | 0.027 | 0.028 |
| 5% | PCA | 3.82 | 1.91 | 0.862 | 3.96 | 1.98 | 0.910 |
| | PP | 0.090 | 0.103 | 0.093 | 0.089 | 0.098 | 0.087 |
| | MCD | 0.022 | 0.023 | 0.021 | – | – | – |
| | Sph. PCA | 0.029 | 0.031 | 0.033 | 0.029 | 0.031 | 0.034 |
| | MCM (Weiszfeld) | 0.023 | 0.023 | 0.028 | 0.022 | 0.023 | 0.030 |
| | MCM (R+) | 0.025 | 0.024 | 0.035 | 0.024 | 0.024 | 0.039 |
| | MCM (R) | 0.029 | 0.027 | 0.037 | 0.028 | 0.028 | 0.040 |
| 10% | PCA | 3.83 | 1.96 | 1.03 | 3.96 | 1.99 | 1.10 |
| | PP | 0.107 | 0.108 | 0.099 | 0.088 | 0.101 | 0.097 |
| | MCD | 0.023 | 0.022 | 0.023 | – | – | – |
| | Sph. PCA | 0.033 | 0.033 | 0.054 | 0.031 | 0.033 | 0.057 |
| | MCM (Weiszfeld) | 0.025 | 0.026 | 0.059 | 0.023 | 0.024 | 0.056 |
| | MCM (R+) | 0.030 | 0.027 | 0.089 | 0.027 | 0.027 | 0.086 |
| | MCM (R) | 0.035 | 0.032 | 0.088 | 0.032 | 0.031 | 0.086 |
| 20% | PCA | 3.84 | 2.02 | 1.19 | 3.96 | 2.01 | 1.25 |
| | PP | 0.110 | 0.135 | 0.138 | 0.091 | 0.122 | 0.137 |
| | MCD | 0.025 | 0.026 | 0.026 | – | – | – |
| | Sph. PCA | 0.037 | 0.038 | 0.140 | 0.034 | 0.037 | 0.150 |
| | MCM (Weiszfeld) | 0.030 | 0.030 | 0.174 | 0.026 | 0.028 | 0.181 |
| | MCM (R+) | 0.044 | 0.036 | 0.255 | 0.038 | 0.032 | 0.256 |
| | MCM (R) | 0.050 | 0.041 | 0.251 | 0.042 | 0.037 | 0.256 |

Table 2: Median estimation errors, according to criterion $R(\hat{\mathbf{P}}_q, \mathbf{P}_q)$ with a dimension $q = 2$, for datasets with a sample size $n = 200$, over 500 Monte Carlo experiments.

by the q eigenvectors associated to the q largest eigenvalues of Γ_m .

We have drawn in Figure 4, the evolution of the mean (over 100 replications) approximation error $R(\mathbf{P}_q, \hat{\mathbf{P}}_q)$, for a dimension $q = 3$, as a function of the sample size for samples contaminated by a 2 degrees of freedom Student t distribution with a rate $\delta = 0.1$. An important fact is that the recursive algorithm which approximates recursively the eigenelements behaves very well and we can see nearly no difference between the spectral decomposition of \bar{V}_n (denoted by MCM in Figure 4) and the estimates produced with the sequential algorithm (16) for sample sizes larger than a few hundreds. We can also note that the error made by the classical PCA is always very high and does not decrease with the sample size.

4.4 Robust PCA of TV audience

The last example is a high dimension and large sample case. Individual TV audiences are measured, by the French company Médiamétrie, every minutes for a panel of $n = 5422$ people over a period of 24 hours, $d = 1440$ (see Cardot et al. (2012) for a more detailed presentation of the data). With a classical PCA, the first eigenspace represents 24.4% of the total variability, whereas the second one reproduces 13.5% of the total variance, the third one 9.64% and the fourth one 6.79%. Thus, more than 54% of the variability of the data can be captured in a four dimensional space. Taking account of the large dimension of the data, these values indicate a high temporal correlation.

Because of the large dimension of the data, the Weiszfeld's algorithm as well as the other robust PCA techniques can not be used anymore in a reasonable time with a personal computer. The MCM has been computed thanks to the recursive algorithm given in (14) in approximately 3 minutes on a laptop in the R language (without any specific C routine).

As seen in Figure 5, the first two eigenvectors obtained by a classical PCA and the robust PCA based on the MCM are rather different. This is confirmed by the relatively large distance between the two corresponding eigenspaces, $R(\hat{P}_2^{PCA}, \hat{P}_2^{MCM}) = 0.56$. The first robust eigenvector puts the stress on the time period comprised between 1000 minutes and 1200 minutes whereas the first non robust eigenvector focuses, with a smaller intensity, on a larger period of time comprised between 600 and 1200 minutes. The second robust eigenvector differentiates between people watching TV during the period between 890 and 1050 minutes (negative value of the second principal component) and people watching TV between minutes 1090 and 1220 (positive value of the second principal component). Rather surprisingly, the third and fourth eigenvectors of the non robust and robust covariance matrices look quite similar (see Figure 6).

5 Proofs

We give in this Section the proofs of Theorems 3.2, 3.3 and 3.5. These proofs rely on several technical Lemmas whose proofs are given in the Supplementary file.

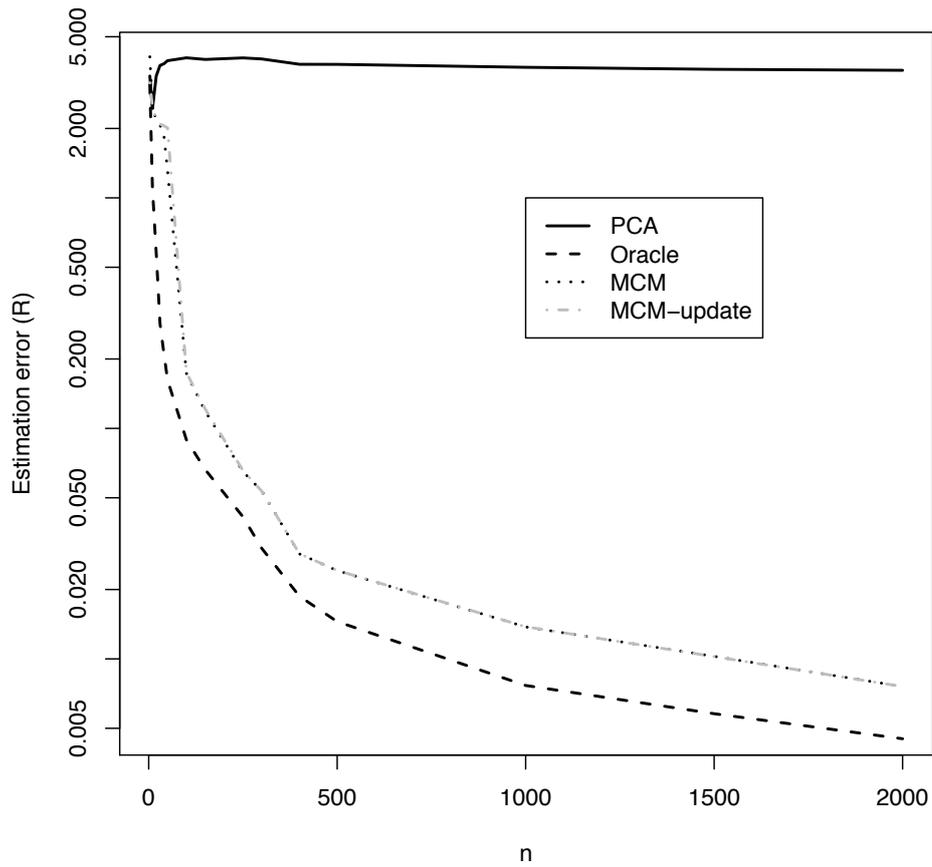


Figure 4: Estimation errors of the eigenspaces (criterion $R(\hat{\mathbf{P}}_q)$) with $d = 1000$ and $q = 3$ for classical PCA, the oracle PCA and the recursive MCM estimator with recursive estimation of the eigenelements (MCM-update) and with static estimation (based on the spectral decomposition of \bar{V}_n) of the eigenelements (MCM).

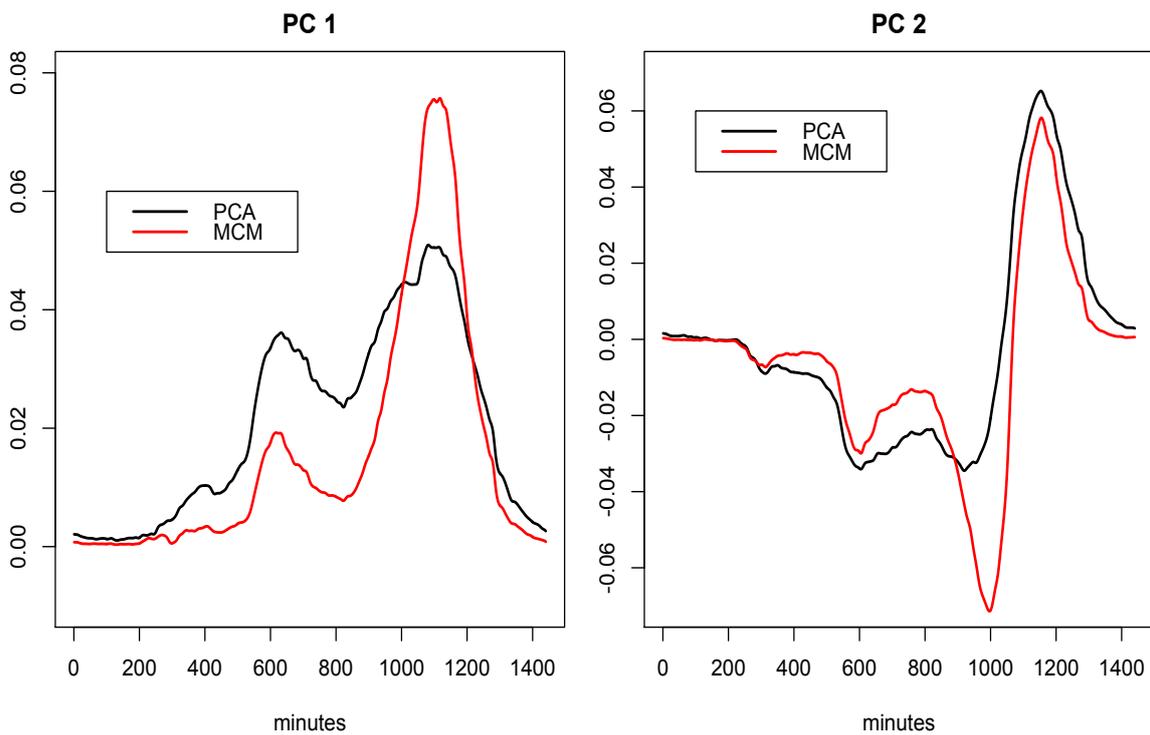


Figure 5: TV audience data measured the 6th September 2010, at the minute scale. Comparison of the principal components of the classical PCA (black) and robust PCA based on the Median Covariation Matrix (red). First eigenvectors on the left, second eigenvectors on the right.

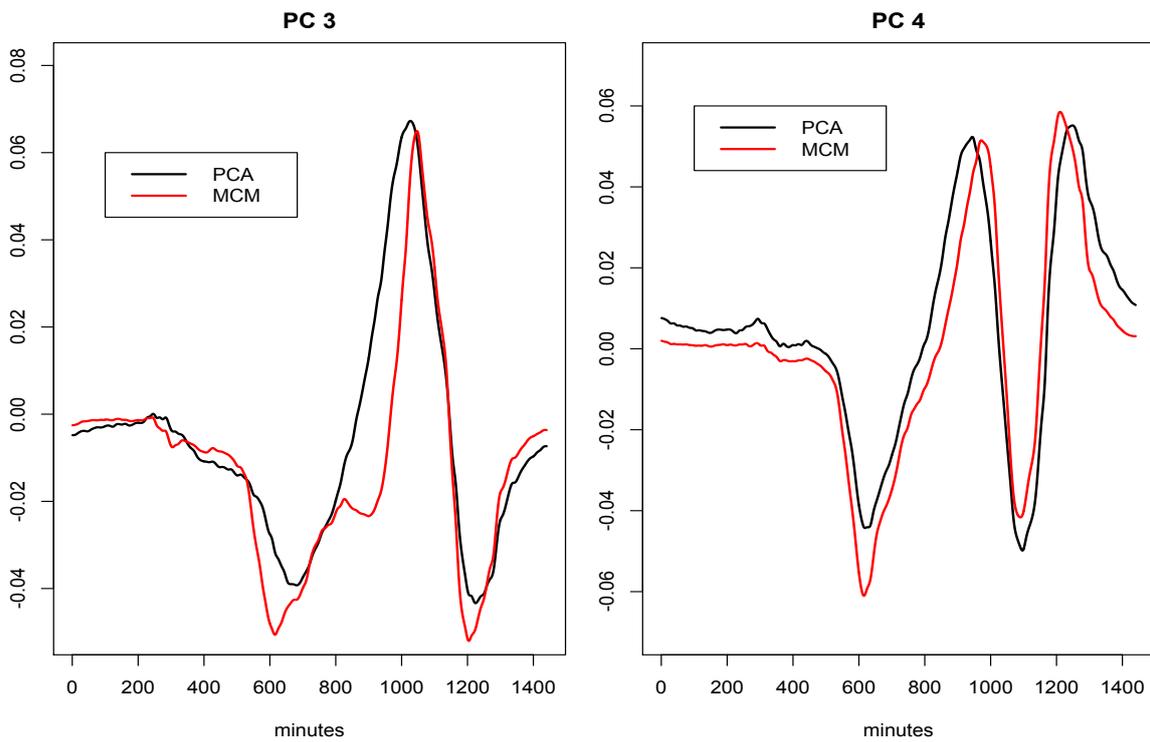


Figure 6: TV audience data measured the 6th September 2010, at the minute scale. Comparison of the principal components of the classical PCA (black) and robust PCA based on the MCM (red). Third eigenvectors on the left, fourth eigenvectors on the right.

5.1 Proof of Theorem 3.2

Let us recall the Robbins-Monro algorithm, defined recursively by

$$\begin{aligned} V_{n+1} &= V_n + \gamma_n \frac{(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n}{\left\| (X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n \right\|_F} \\ &= V_n - \gamma_n U_{n+1}, \end{aligned}$$

with $U_{n+1} := -\frac{(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n}{\left\| (X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n \right\|_F}$. Since $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$, we have $\mathbb{E}[U_{n+1} | \mathcal{F}_n] = \nabla G_{\bar{m}_n}(V_n)$. Thus $\xi_{n+1} := \nabla_{\bar{m}_n} G(V_n) - U_{n+1}$, (ξ_n) is a sequence of martingale differences adapted to the filtration (\mathcal{F}_n) . Indeed, $\mathbb{E}[\xi_{n+1} | \mathcal{F}_n] = \nabla G_{\bar{m}_n}(V_n) - \mathbb{E}[U_{n+1} | \mathcal{F}_n] = 0$. The algorithm can be written as follows

$$V_{n+1} = V_n - \gamma_n \nabla G_{\bar{m}_n}(V_n) + \gamma_n \xi_{n+1}.$$

Moreover, it can be considered as a stochastic gradient algorithm because it can be decomposed as follows:

$$V_{n+1} = V_n - \gamma_n (\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)) + \gamma_n \xi_{n+1} - \gamma_n r_n, \quad (19)$$

with $r_n := \nabla G_{\bar{m}_n}(\Gamma_m) - \nabla G_m(\Gamma_m)$. Finally, linearizing the gradient,

$$V_{n+1} - \Gamma_m = (I_{S(H)} - \gamma_n \nabla_m^2 G(\Gamma_m)) (V_n - \Gamma_m) + \gamma_n \xi_{n+1} - \gamma_n r_n - \gamma_n r'_n - \gamma_n \delta_n, \quad (20)$$

with

$$\begin{aligned} r'_n &:= (\nabla_{\bar{m}_n}^2 G(\Gamma_m) - \nabla_m^2 G(\Gamma_m)) (V_n - \Gamma_m), \\ \delta_n &:= \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) - \nabla_{\bar{m}_n}^2 G(\Gamma_m) (V_n - \Gamma_m). \end{aligned}$$

The following lemma gives upper bounds of these remainder terms. Its proof is given in the Supplementary file.

Lemma 5.1. *Under assumptions 1-3(b), we can bound the three remainder terms. First,*

$$\|\delta_n\|_F \leq 6C \|V_n - \Gamma_m\|_F^2. \quad (21)$$

In the same way, for all $n \geq 1$,

$$\|r_n\|_F \leq 4 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right) \|\bar{m}_n - m\|. \quad (22)$$

Finally, for all $n \geq 1$,

$$\|r'_n\|_F \leq 12 \left(C \sqrt{\|\Gamma_m\|_F} + C^{3/4} \right) \|\bar{m}_n - m\| \|V_n - \Gamma_m\|_F. \quad (23)$$

We deduce from decomposition (33) that for all $n \geq 1$,

$$\begin{aligned} \|V_{n+1} - \Gamma_m\|_F^2 &= \|V_n - \Gamma_m\|_F^2 - 2\gamma_n \langle V_n - \Gamma_m, \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_F \\ &\quad + \gamma_n^2 \|\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)\|_F^2 \\ &\quad + \gamma_n^2 \|\xi_{n+1}\|_F^2 + 2\gamma_n \langle V_n - \Gamma_m - \gamma_n (\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)), \xi_{n+1} \rangle_F \\ &\quad + \gamma_n^2 \|r_n\|_F^2 - 2\gamma_n \langle r_n, V_n - \Gamma_m \rangle_F - 2\gamma_n^2 \langle r_n, \xi_{n+1} - \nabla G_{\bar{m}_n}(V_n) + \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_F. \end{aligned}$$

Note that for all $h \in H$ and $V \in \mathcal{S}(H)$ we have $\|\nabla G_h(V)\|_F \leq 1$. Furthermore, $\|r_n\|_F \leq 2$ and $\|\xi_{n+1}\|_F \leq 2$. Using the fact that (ξ_n) is a sequence of martingale differences adapted to the filtration (\mathcal{F}_n) ,

$$\begin{aligned} \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^2 \mid \mathcal{F}_n \right] &\leq \|V_n - \Gamma_m\|_F^2 - 2\gamma_n \langle V_n - \Gamma_m, \nabla_{\bar{m}_n} G(V_n) - \nabla_{\bar{m}_n} G(\Gamma_m) \rangle_F \\ &\quad + 28\gamma_n^2 - 2\gamma_n \langle r_n, V_n - \Gamma_m \rangle_F. \end{aligned}$$

Let $\alpha_n = n^{-\beta}$, with $\beta \in (1 - \alpha, \alpha)$, we have

$$\begin{aligned} \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^2 \mid \mathcal{F}_n \right] &\leq (1 + \gamma_n \alpha_n) \|V_n - \Gamma_m\|_F^2 - 2\gamma_n \langle V_n - \Gamma_m, \nabla_{\bar{m}_n} G(V_n) - \nabla_{\bar{m}_n} G(\Gamma_m) \rangle_F \\ &\quad + 28\gamma_n^2 + \frac{\gamma_n}{\alpha_n} \|r_n\|_F^2. \end{aligned} \tag{24}$$

Moreover, applying Lemma 5.1 and Theorem 5.1 in Godichon-Baggioni (2016), we get for all positive constant δ ,

$$\|r_n\|_F^2 = O\left(\|\bar{m}_n - m\|^2\right) = O\left(\frac{(\ln n)^{1+\delta}}{n}\right) \quad a.s.$$

Thus, since $2\gamma_n \langle V_n - \Gamma_m, \nabla_{\bar{m}_n} G(V_n) - \nabla_{\bar{m}_n} G(\Gamma_m) \rangle_F \geq 0$, the Robbins-Siegmund Theorem (see Dufflo (1997) for instance) ensures that $\|V_n - \Gamma_m\|_F$ converges almost surely to a finite random variable and

$$\sum_{n \geq 1} \gamma_n \langle V_n - \Gamma_m, \nabla_{\bar{m}_n} G(V_n) - \nabla_{\bar{m}_n} G(\Gamma_m) \rangle_F < +\infty \quad a.s.$$

Furthermore, by induction, inequality (24) becomes

$$\begin{aligned} \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^2 \right] &\leq \left(\prod_{k=1}^{\infty} (1 + \gamma_k \alpha_k) \right) \mathbb{E} \left[\|V_1 - \Gamma_m\|_F^2 \right] + 28 \left(\prod_{k=1}^{\infty} (1 + \gamma_k \alpha_k) \right) \sum_{k=1}^{\infty} \gamma_k^2 \\ &\quad + \left(\prod_{k=1}^{\infty} (1 + \gamma_k \alpha_k) \right) \sum_{k=1}^{\infty} \frac{\gamma_k}{\alpha_k} \mathbb{E} \left[\|r_k\|_F^2 \right]. \end{aligned}$$

Since $\beta < \alpha$, applying Theorem 4.2 in Godichon-Baggioni (2016) and Lemma 6.1, there is a positive constant C_0 such that

$$\sum_{k=1}^{\infty} \frac{\gamma_k}{\alpha_k} \mathbb{E} \left[\|r_k\|_F^2 \right] = C_0 \sum_{k=1}^{\infty} k^{-\alpha-1-\beta} < +\infty.$$

Thus, there is a positive constant M such that for all $n \geq 1$, $\mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] \leq M$. Since \bar{m}_n converges almost surely to m , one can conclude the proof of the almost sure consistency of V_n with the same arguments as in the proof of Theorem 3.1 in Cardot et al. (2013) and the convexity properties given in the Section B of the supplementary file.

Finally, the almost sure consistency of \bar{V}_n is obtained by a direct application of Topelitz's lemma (see *e.g.* Lemma 2.2.13 in Dufflo (1997)).

5.2 Proof of Theorem 3.3

The proof of Theorem 3.3 relies on properties of the p -th moments of V_n for all $p \geq 1$ given in the following three Lemmas. These properties enable us, with the application of Markov's inequality, to control the probability of the deviations of the Robbins Monro algorithm from Γ_m .

Lemma 5.2. *Under assumptions 1-3(b), for all integer p , there is a positive constant M_p such that for all $n \geq 1$,*

$$\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \leq M_p.$$

Lemma 5.3. *Under assumptions 1-3(b), there are positive constants C_1, C'_1, C_2, C_3 such that for all $n \geq 1$,*

$$\mathbb{E} \left[\|V_n - \Gamma_m\|^2 \right] \leq C_1 e^{-C'_1 n^{1-\alpha}} + \frac{C_2}{n^\alpha} + C_3 \sup_{E(n/2)+1 \leq k \leq n-1} \mathbb{E} \left[\|V_k - \Gamma_m\|^4 \right],$$

where $E(x)$ is the integer part of the real number x .

Lemma 5.4. *Under assumptions 1-3(b), for all integer $p' \geq 1$, there are a rank $n_{p'}$ and positive constants $C_{1,p'}, C_{2,p'}, C_{3,p'}, c_{p'}$ such that for all $n \geq n_{p'}$,*

$$\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] \leq \left(1 - c_{p'} \gamma n^{-\frac{1-\alpha}{p'}} \right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] + \frac{C_{1,p'}}{n^{3\alpha}} + \frac{C_{2,p'}}{n^{2\alpha}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] + \frac{C_{3,p'}}{n^{3\alpha-3\frac{1-\alpha}{p'}}}.$$

We can now prove Theorem 3.3.

Let us choose an integer p' such that $p' > 3/2$. Thus, $2 + \alpha - 3\frac{1-\alpha}{p'} \geq 3\alpha$, and applying Lemma 5.4, there are positive constants $C_{1,p'}, C_{2,p'}, c_{p'}$ and a rank $n_{p'}$ such that for all $n \geq n_{p'}$,

$$\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] \leq \left(1 - c_{p'} \gamma n^{-\frac{1-\alpha}{p'}} \right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] + \frac{C_{1,p'}}{n^{3\alpha}} + \frac{C_{2,p'}}{n^{2\alpha}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right]. \quad (25)$$

Let us now choose $\beta \in (\alpha, 2\alpha)$ and p' such that $p' > \frac{1-\alpha}{2\alpha-\beta}$. Note that $3\alpha - \beta > \alpha + \frac{1-\alpha}{p'}$. One can check that there is a rank $n'_{p'} \geq n_{p'}$ such that for all $n \geq n'_{p'}$,

$$\begin{aligned} (n+1)^\alpha C_1 e^{-C'_1 n^{1-\alpha}} + \frac{1}{2} + C_3 2^{\beta+1} \frac{1}{(n+1)^{\beta-\alpha}} &\leq 1, \\ \left(1 - c_{p'} \gamma n^{-\frac{1-\alpha}{p'}} \right) \left(\frac{n+1}{n} \right)^\beta + 2^{3\alpha} \frac{C_{1,p'} + C_{2,p'}}{(n+1)^{3\alpha-\beta}} &\leq 1. \end{aligned}$$

With the help of a strong induction, we are going to prove the announced results, that is to say that there are positive constants $C_{p'}, C_\beta$ such that $2C_{p'} \geq C_\beta \geq C_{p'} \geq 1$ and $C_{p'} \geq 2^{\alpha+1}C_2$ (with C_2 defined in Lemma 5.3), such that for all $n \geq 1$,

$$\begin{aligned}\mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] &\leq \frac{C_{p'}}{n^\alpha}, \\ \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] &\leq \frac{C_\beta}{n^\beta}.\end{aligned}$$

First, let us choose $C_{p'}$ and C_β such that

$$\begin{aligned}C_{p'} &\geq \max_{k \leq n'_{p'}} \left\{ k^\alpha \mathbb{E} \left[\|V_k - \Gamma_m\|_F^2 \right] \right\}, \\ C_\beta &\geq \max_{k \leq n'_{p'}} \left\{ k^\beta \mathbb{E} \left[\|V_{n'_{p'}} - \Gamma_m\|_F^4 \right] \right\}.\end{aligned}$$

Thus, for all $k \leq n'_{p'}$,

$$\begin{aligned}\mathbb{E} \left[\|V_k - \Gamma_m\|_F^2 \right] &\leq \frac{C_{p'}}{k^\alpha}, \\ \mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right] &\leq \frac{C_\beta}{k^\beta}.\end{aligned}$$

We suppose from now that $n \geq n'_{p'}$ and that previous inequalities are verified for all $k \leq n-1$. Applying Lemma 5.2 and by induction,

$$\begin{aligned}\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^2 \right] &\leq C_1 e^{-C'_1 n^{1-\alpha}} + \frac{C_2}{n^\alpha} + C_3 \sup_{E((n+1)/2)+1 \leq k \leq n} \left\{ \mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right] \right\} \\ &\leq C_1 e^{-C'_1 n^{1-\alpha}} + \frac{C_2}{n^\alpha} + C_3 \sup_{E((n+1)/2)+1 \leq k \leq n} \left\{ \frac{C_\beta}{k^\beta} \right\} \\ &\leq C_1 e^{-C'_1 n^{1-\alpha}} + \frac{C_2}{n^\alpha} + C_3 2^\beta \frac{C_\beta}{n^\beta}.\end{aligned}$$

Since $2C_{p'} \geq C_\beta \geq C_{p'} \geq 1$ and since $C_{p'} \geq 2^{\alpha+1}C_2$, factorizing by $\frac{C_{p'}}{(n+1)^\alpha}$,

$$\begin{aligned}\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^2 \right] &\leq C_{p'} C_1 e^{-C'_1 n^{1-\alpha}} + C_{p'} 2^{-\alpha-1} \frac{1}{n^\alpha} + C_3 2^\beta \frac{2C_{p'}}{n^\beta} \\ &\leq \frac{C_{p'}}{(n+1)^\alpha} (n+1)^\alpha C_1 e^{-C'_1 n^{1-\alpha}} + 2^{-\alpha} \left(\frac{n}{n+1} \right)^\alpha \frac{C_{p'}}{2(n+1)^\alpha} + \frac{C_3 2^{\beta+1}}{(n+1)^{\beta-\alpha}} \frac{C_{p'}}{(n+1)^\alpha} \\ &\leq \frac{C_{p'}}{(n+1)^\alpha} C_1 (n+1)^\alpha e^{-C'_1 n^{1-\alpha}} + \frac{1}{2} \frac{C_{p'}}{(n+1)^\alpha} + C_3 2^{\beta+1} \frac{1}{(n+1)^{\beta-\alpha}} \frac{C_{p'}}{(n+1)^\alpha} \\ &\leq \left((n+1)^\alpha C_1 e^{-C'_1 n^{1-\alpha}} + \frac{1}{2} + C_3 2^{\beta+1} \frac{1}{(n+1)^{\beta-\alpha}} \right) \frac{C_{p'}}{(n+1)^\alpha}.\end{aligned}$$

By definition of $n'_{p'}$,

$$\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^2 \right] \leq \frac{C_{p'}}{(n+1)^\alpha}. \quad (26)$$

In the same way, applying Lemma 5.4 and by induction,

$$\begin{aligned}\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] &\leq \left(1 - c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] + \frac{C_{1,p'}}{n^{3\alpha}} + \frac{C_{2,p'}}{n^{2\alpha}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] \\ &\leq \left(1 - c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \right) \frac{C_\beta}{n^\beta} + \frac{C_{1,p'}}{n^{3\alpha}} + \frac{C_{2,p'}}{n^{2\alpha}} \frac{C_{p'}}{n^\alpha}.\end{aligned}$$

Since $C_\beta \geq C_{p'} \geq 1$, factorizing by $\frac{C_\beta}{(n+1)^\beta}$,

$$\begin{aligned} \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] &\leq \left(1 - c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \frac{C_\beta}{n^\beta} + (C_{1,p'} + C_{2,p'}) \frac{C_\beta}{n^{3\alpha}} \\ &\leq \left(1 - c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \left(\frac{n+1}{n}\right)^\beta \frac{C_\beta}{n^\beta} + 2^{3\alpha} \frac{C_{1,p'} + C_{2,p'}}{(n+1)^{3\alpha-\beta}} \frac{C_\beta}{(n+1)^\beta} \\ &\leq \left(\left(1 - c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \left(\frac{n+1}{n}\right)^\beta + 2^{3\alpha} \frac{C_{1,p'} + C_{2,p'}}{(n+1)^{3\alpha-\beta}} \right) \frac{C_\beta}{(n+1)^\beta}. \end{aligned}$$

By definition of $n'_{p'}$,

$$\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] \leq \frac{C_\beta}{(n+1)^\beta}, \quad (27)$$

which concludes the induction and the proof.

5.3 Proof of Theorem 3.5

In order to prove Theorem 3.5, we first recall the following Lemma.

Lemma 5.5 (Godichon-Baggioni (2016)). *Let Y_1, \dots, Y_n be random variables taking values in a normed vector space such that for all positive constant q and for all $k \geq 1$, $\mathbb{E} [\|Y_k\|^q] < \infty$. Then, for all real numbers a_1, \dots, a_n and for all integer p , we have*

$$\mathbb{E} \left[\left\| \sum_{k=1}^n a_k Y_k \right\|^p \right] \leq \left(\sum_{k=1}^n |a_k| (\mathbb{E} [\|Y_k\|^p])^{\frac{1}{p}} \right)^p \quad (28)$$

We can now prove Theorem 3.5. Let us rewrite decomposition (34) as follows

$$\nabla_m^2 G(\Gamma_m) (V_n - \Gamma_m) = \frac{T_n}{\gamma_n} - \frac{T_{n+1}}{\gamma_n} + \xi_{n+1} - r_n - r'_n - \delta_n, \quad (29)$$

with $T_n := V_n - \Gamma_m$. As in Pelletier (2000), we sum these equalities, apply Abel's transform and divide by n to get

$$\nabla_m^2 G(\Gamma_m) (\bar{V}_n - \Gamma_m) = \frac{1}{n} \left(\frac{T_1}{\gamma_1} - \frac{T_{n+1}}{\gamma_{n+1}} + \sum_{k=2}^n T_k \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right) - \sum_{k=1}^n \delta_k - \sum_{k=1}^n r_k - \sum_{k=1}^n r'_k + \sum_{k=1}^n \xi_{k+1} \right).$$

We now bound the quadratic mean of each term at the right-hand side of previous equality.

First, we have $\frac{1}{n^2} \mathbb{E} \left[\left\| \frac{T_1}{\gamma_1} \right\|_F^2 \right] = o\left(\frac{1}{n}\right)$. Applying Theorem 3.3,

$$\frac{1}{n^2} \mathbb{E} \left[\left\| \frac{T_{n+1}}{\gamma_n} \right\|_F^2 \right] \leq \frac{1}{n^2} \frac{C' c_\gamma^{-2}}{n^{-\alpha}} = o\left(\frac{1}{n}\right).$$

Moreover, since $|\gamma_k^{-1} - \gamma_{k-1}^{-1}| \leq 2\alpha c_\gamma^{-1} k^{\alpha-1}$, the application of Lemma 5.5 and Theorem 3.3

gives

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E} \left[\left\| \sum_{k=2}^n (\gamma_k^{-1} - \gamma_{k-1}^{-1}) T_k \right\|_F^2 \right] &\leq \frac{1}{n^2} \left(\sum_{k=2}^n |\gamma_k^{-1} - \gamma_{k-1}^{-1}| \sqrt{\mathbb{E} [\|T_k\|_F^2]} \right)^2 \\
&\leq \frac{1}{n^2} 4\alpha^2 c_\gamma^{-2} C' \left(\sum_{k=2}^n \frac{1}{k^{1-\alpha/2}} \right)^2 \\
&= O \left(\frac{1}{n^{2-\alpha}} \right) \\
&= o \left(\frac{1}{n} \right),
\end{aligned}$$

since $\alpha < 1$. In the same way, since $\|\delta_n\|_F \leq 6C \|T_n\|_F^2$, applying Lemma 5.5 and Theorem 3.3 with $\beta > 1$,

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E} \left[\left\| \sum_{k=1}^n \delta_k \right\|_F^2 \right] &\leq \frac{1}{n^2} \left(\sum_{k=1}^n \sqrt{\mathbb{E} [\|\delta_k\|_F^2]} \right)^2 \\
&\leq \frac{36C^2}{n^2} \left(\sum_{k=1}^n \sqrt{\mathbb{E} [\|T_k\|_F^4]} \right)^2 \\
&\leq \frac{36C^2 C_\beta}{n^2} \left(\sum_{k=1}^n \frac{1}{k^{\beta/2}} \right)^2 \\
&= O \left(\frac{1}{n^\beta} \right) \\
&= o \left(\frac{1}{n} \right),
\end{aligned}$$

Moreover, let $D := 12 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)$. Since $\|r_n\|_F \leq D \|\bar{m}_n - m\|$, and since there is a positive constant C'' such that for all $n \geq 1$, $\mathbb{E} [\|\bar{m}_n - m\|^2] \leq C'' n^{-1}$,

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E} \left[\left\| \sum_{k=1}^n r_k \right\|_F^2 \right] &\leq \frac{1}{n^2} \left(\sum_{k=1}^n \sqrt{\mathbb{E} [\|r_k\|_F^2]} \right)^2 \\
&\leq \frac{D^2}{n^2} \left(\sum_{k=1}^n \sqrt{\mathbb{E} [\|\bar{m}_n - m\|^2]} \right)^2 \\
&\leq \frac{D^2 C''}{n^2} \left(\sum_{k=1}^n \frac{1}{k^{1/2}} \right)^2 \\
&= O \left(\frac{1}{n} \right).
\end{aligned}$$

Since $\|r'_n\|_F \leq C_0 \|\bar{m}_n - m\| \|V_n - \Gamma_m\|_F^2$ with $C_0 := 12 \left(C \sqrt{\|\Gamma_m\|_F} + C^{3/4} \right)$, Cauchy-

Schwarz's inequality and Lemma 5.5 give

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E} \left[\left\| \sum_{k=1}^n r'_n \right\|_F^2 \right] &\leq \frac{1}{n^2} \left(\sum_{k=1}^n \sqrt{\mathbb{E} \left[\|r'_n\|_F^2 \right]} \right)^2 \\
&\leq \frac{C_0^2}{n^2} \left(\sum_{k=1}^n \sqrt{\mathbb{E} \left[\|\bar{m}_n - m\|^2 \|V_n - \Gamma_m\|_F^2 \right]} \right)^2 \\
&\leq \frac{C_0^2}{n^2} \left(\sum_{k=1}^n \left(\mathbb{E} \left[\|\bar{m}_n - m\|^4 \right] \right)^{\frac{1}{4}} \left(\mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \right)^{\frac{1}{4}} \right)^2.
\end{aligned}$$

Applying Theorem 4.2 in Godichon-Baggioni (2016) and Theorem 3.3,

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E} \left[\left\| \sum_{k=1}^n r'_n \right\|_F^2 \right] &\leq \frac{C_0^2 \sqrt{C_\beta} \sqrt{K_2}}{n^2} \left(\sum_{k=1}^n \frac{1}{k^{\beta/4+1/2}} \right)^2 \\
&= O \left(\frac{1}{n^{1+\beta/2}} \right) \\
&= o \left(\frac{1}{n} \right),
\end{aligned}$$

since $\beta > 0$. Finally, one can easily check that $\mathbb{E} \left[\|\xi_{n+1}\|_F^2 \right] \leq 1$, and since (ξ_n) is a sequence of martingale differences adapted to the filtration (\mathcal{F}_n) ,

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E} \left[\left\| \sum_{k=1}^n \xi_{k+1} \right\|_F^2 \right] &= \frac{1}{n^2} \left(\sum_{k=1}^n \mathbb{E} \left[\|\xi_{k+1}\|_F^2 \right] + 2 \sum_{k=1}^n \sum_{k'=k+1}^n \mathbb{E} \left[\langle \xi_{k+1}, \xi_{k'+1} \rangle_F \right] \right) \\
&= \frac{1}{n^2} \left(\sum_{k=1}^n \mathbb{E} \left[\|\xi_{k+1}\|_F^2 \right] + 2 \sum_{k=1}^n \sum_{k'=k+1}^n \mathbb{E} \left[\langle \xi_{k+1}, \mathbb{E} \left[\xi_{k'+1} \mid \mathcal{F}_{k'} \right] \rangle_F \right] \right) \\
&= \frac{1}{n^2} \sum_{k=1}^n \mathbb{E} \left[\|\xi_{k+1}\|_F^2 \right] \\
&\leq \frac{1}{n}.
\end{aligned}$$

Thus, there is a positive constant K such that for all $n \geq 1$,

$$\mathbb{E} \left[\left\| \nabla_m^2 G(\Gamma_m) (\bar{V}_n - \Gamma_m) \right\|_F^2 \right] \leq \frac{K}{n}.$$

Let λ_{\min} be the smallest eigenvalue of $\nabla_m^2 G(\Gamma_m)$. We have, with Proposition B.1 in the supplementary file, that $\lambda_{\min} > 0$ and the announced result is proven,

$$\mathbb{E} \left[\left\| \bar{V}_n - \Gamma_m \right\|_F^2 \right] \leq \frac{K}{\lambda_{\min}^2 n}.$$

6 Concluding remarks

The simulation study and the illustration on real data indicate that performing robust principal components analysis via the median covariation matrix, which can bring new

information compared to classical PCA, is an interesting alternative to more classical robust principal components analysis techniques. The use of recursive algorithms permits to perform robust PCA on very large datasets, in which outlying observations may be hard to detect. Another interest of the use of such sequential algorithms is that estimation of the median covariation matrix as well as the principal components can be performed online with automatic update at each new observation and without being obliged to store all the data in memory. A simple modification of the averaged stochastic gradient algorithm is proposed that ensures non negativeness of the estimated covariation matrices. This modified algorithms has better performances on our simulated data.

A deeper study of the asymptotic behaviour of the recursive algorithms would certainly deserve further investigations. Proving the asymptotic normality and obtaining the limiting variance of the sequence of estimators \bar{V}_n when m is unknown would be of great interest. This is a challenging issue that is beyond the scope of the paper and would require to study the joint weak convergence of the two simultaneous recursive averaged estimators of m and Γ_m .

The use of the MCM could be interesting to robustify the estimation in many different statistical models, particularly with functional data. For example, it could be employed as an alternative to robust functional projection pursuit in robust functional time series prediction or for robust estimation in functional linear regression, with the introduction of the median cross-covariation matrix.

Acknowledgements. We thank the company Médiamétrie for allowing us to illustrate our methodologies with their data. We also thank Dr. Peggy Cénac for a careful reading of the proofs.

A Estimating the median covariation matrix with Weiszfeld's algorithm

Suppose we have a fixed size sample X_1, \dots, X_n and we want to estimate the geometric median.

The iterative Weiszfeld's algorithm relies on the fact that the solution m_n^* of the following optimization problem

$$\min_{\mu \in H} \sum_{i=1}^n \|X_i - \mu\|$$

satisfies, when $m_n^* \neq X_i$, for all $i = 1, \dots, n$

$$m_n^* = \sum_{i=1}^n w_i(m_n^*) X_i$$

where the weights $w_i(x)$ are defined by

$$w_i(x) = \frac{\|X_i - x\|^{-1}}{\sum_{j=1}^n \|X_j - x\|^{-1}}.$$

Weiszfeld's algorithm is based on the following iterative scheme. Consider first a pilot estimator $\widehat{m}^{(0)}$ of m . At step (e) , a new approximation $\widehat{m}_n^{(e+1)}$ to m is given by

$$\widehat{m}_n^{(e+1)} = \sum_{i=1}^n w_i \left(\widehat{m}_n^{(e)} \right) X_i. \quad (30)$$

The iterative procedure is stopped when $\left\| \widehat{m}_n^{(e+1)} - \widehat{m}_n^{(e)} \right\| \leq \epsilon$, for some precision ϵ known in advance. The final value of the algorithm is denoted by \widehat{m}_n .

The estimator of the MCM is computed similarly. Suppose $\widehat{\Gamma}^{(e)}$ has been calculated at step (e) , then at step $(e+1)$, the new approximation $\widehat{\Gamma}^{(e+1)}$ to Γ_m is defined by

$$\widehat{\Gamma}_n^{(e+1)} = \sum_{i=1}^n W_i \left(\widehat{\Gamma}^{(e)} \right) (X_i - \widehat{m}_n)(X_i - \widehat{m}_n)^T. \quad (31)$$

The procedure is stopped when $\left\| \widehat{\Gamma}^{(e+1)} - \widehat{\Gamma}^{(e)} \right\|_F \leq \epsilon$, for some precision ϵ fixed in advance.

Note that by construction, this algorithm leads to an estimated median covariation matrix that is always non negative.

B Convexity results

In this section, we first give and recall some convexity properties of functional G_h . The following one gives some information on the spectrum of the Hessian of G .

Proposition B.1. *Under assumptions 1-3(b), for all $h \in H$ and $V \in \mathcal{S}(H)$, $\mathcal{S}(H)$ admits an orthonormal basis composed of eigenvectors of $\nabla_h^2 G(V)$. Let us denote by $\{\lambda_{h,V,i}, i \in \mathbb{N}\}$ the set of eigenvalues of $\nabla_h^2 G(V)$. For all $i \in \mathbb{N}$,*

$$0 \leq \lambda_{h,V,i} \leq C.$$

Moreover, there is a positive constant c_m such that for all $i \in \mathbb{N}$,

$$0 < c_m \leq \lambda_{m,\Gamma_m,i} \leq C.$$

Finally, by continuity, there are positive constants ϵ, ϵ' such that for all $h \in \mathcal{B}(m, \epsilon)$ and $V \in \mathcal{B}(\Gamma_m, \epsilon')$, and for all $i \in \mathbb{N}$,

$$\frac{1}{2}c_m \leq \lambda_{h,V,i} \leq C.$$

The proof is very similar to the one in Cardot et al. (2013) and consequently it is not given here. Furthermore, as in Cardot et al. (2016), it ensures the local strong convexity as shown in the following corollary.

Corollary B.2. *Under assumptions 1-3(b), for all positive constant A , there is a positive constant c_A such that for all $V \in \mathcal{B}(\Gamma_m, A)$ and $h \in \mathcal{B}(m, \epsilon)$,*

$$\langle \nabla_h G(V) - \nabla_h G(\Gamma_m), V - \Gamma_m \rangle_H \geq c_A \|V - \Gamma_m\|_F^2.$$

Finally, the following lemma gives an upper bound on the remainder term in the Taylor's expansion of the gradient.

Lemma B.3. *Under assumptions 1-3(b), for all $h \in H$ and $V \in \mathcal{S}(H)$,*

$$\left\| \nabla G_h(V) - \nabla G_h(\Gamma_m) - \nabla_h^2 G(\Gamma_m)(V - \Gamma_m) \right\|_F \leq 6C \|V - \Gamma_m\|_F^2. \quad (32)$$

Proof of Lemma B.3. Let $\delta_{V,h} := \nabla G_h(V) - \nabla G_h(\Gamma_m) - \nabla_h^2 G(\Gamma_m)(V - \Gamma_m)$, since $\nabla G_h(V) - \nabla G_h(\Gamma_m) = \int_0^1 \nabla_h^2 G(\Gamma_m + t(V - \Gamma_m))(V - \Gamma_m) dt$, we have

$$\begin{aligned} \|\delta_{V,h}\|_F &= \left\| \int_0^1 \nabla_h^2 G(\Gamma_m + t(V - \Gamma_m))(V - \Gamma_m) dt - \nabla_h^2 G(\Gamma_m)(V - \Gamma_m) \right\|_F \\ &\leq \int_0^1 \left\| \nabla_h^2 G(\Gamma_m + t(V - \Gamma_m))(V - \Gamma_m) - \nabla_h^2 G(\Gamma_m)(V - \Gamma_m) \right\|_F dt. \end{aligned}$$

As in the proof of Lemma 5.1 in Cardot et al. (2016), under assumptions 1-3(b), one can check that for all $h \in H$, and $t \in [0, 1]$,

$$\left\| \nabla_h^2 G(\Gamma_m + t(V - \Gamma_m))(V - \Gamma_m) - \nabla_h^2 G(\Gamma_m)(V - \Gamma_m) \right\|_F \leq 6C \|V - \Gamma_m\|_F^2,$$

which concludes the proof. \square

C Decompositions of the Robbins-Monro algorithm and proof of Lemma 5.1

Let us recall that the Robbins-Monro algorithm is defined recursively by

$$\begin{aligned} V_{n+1} &= V_n + \gamma_n \frac{(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n}{\left\| (X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n \right\|_F} \\ &= V_n - \gamma_n U_{n+1}, \end{aligned}$$

with $U_{n+1} := -\frac{(X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n}{\left\| (X_{n+1} - \bar{m}_n)(X_{n+1} - \bar{m}_n)^T - V_n \right\|_F}$. Let us remark that $\xi_{n+1} := \nabla_{\bar{m}_n} G(V_n) - U_{n+1}$, (ξ_n) is a sequence of martingale differences adapted to the filtration (\mathcal{F}_n) and the algorithm can be written as follows

$$V_{n+1} = V_n - \gamma_n (\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)) + \gamma_n \xi_{n+1} - \gamma_n r_n, \quad (33)$$

with $r_n := \nabla G_{\bar{m}_n}(\Gamma_m) - \nabla G_m(\Gamma_m)$. Finally, let us consider the following linearization of the gradient,

$$V_{n+1} - \Gamma_m = (I_{\mathcal{S}(H)} - \gamma_n \nabla_m^2 G(\Gamma_m))(V_n - \Gamma_m) + \gamma_n \xi_{n+1} - \gamma_n r_n - \gamma_n r'_n - \gamma_n \delta_n, \quad (34)$$

with

$$\begin{aligned} r'_n &:= (\nabla_{\bar{m}_n}^2 G(\Gamma_m) - \nabla_m^2 G(\Gamma_m))(V_n - \Gamma_m), \\ \delta_n &:= \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) - \nabla_{\bar{m}_n}^2 G(\Gamma_m)(V_n - \Gamma_m). \end{aligned}$$

Proof of Lemma 5.1. The bound of $\|\delta_n\|$ is a corollary of Lemma B.3.

Bounding $\|r_n\|$

Let us recall that for all $h \in H$, $Y(h) := (X - h)(X - h)^T$. We now define for all $h \in H$ the random function $\varphi_h : [0, 1] \rightarrow \mathcal{S}(H)$ defined for all $t \in [0, 1]$ by

$$\varphi_h(t) := \frac{Y(m + th) - \Gamma_m}{\|Y(m + th) - \Gamma_m\|_F}.$$

Note that $r_n = \mathbb{E} \left[\varphi_{\bar{m}_n - m}(0) - \varphi_{\bar{m}_n - m}(1) \middle| \mathcal{F}_n \right]$. Thus, by dominated convergence,

$$\|r_n\|_F \leq \sup_{t \in [0, 1]} \mathbb{E} \left[\|\varphi'_{\bar{m}_n - m}(t)\|_F \middle| \mathcal{F}_n \right].$$

Moreover, one can check that for all $h \in H$,

$$\begin{aligned} \varphi'_h(t) &= -\frac{h(X - m - th)^T}{\|Y(m + th) - \Gamma_m\|_F} - \frac{(X - m - th)h^T}{\|Y(m + th) - \Gamma_m\|_F} \\ &\quad + \left\langle Y(m + th) - \Gamma_m, h(X - m - th)^T \right\rangle_F \frac{Y(m + th) - \Gamma_m}{\|Y(m + th) - \Gamma_m\|_F^3} \\ &\quad + \left\langle Y(m + th) - \Gamma_m, (X - m - th)h^T \right\rangle_F \frac{Y(m + th) - \Gamma_m}{\|Y(m + th) - \Gamma_m\|_F^3}. \end{aligned}$$

We now bound each term on the right-hand side of previous equality. First, applying Cauchy-Schwarz's inequality and using the fact that for all $h, h' \in H$, $\|hh'^T\|_F = \|h\| \|h'\|$,

$$\begin{aligned} \mathbb{E} \left[\left\| \frac{h(X - m - th)^T}{\|Y(m + th) - \Gamma_m\|_F} \right\|_F \right] &\leq \|h\| \mathbb{E} \left[\frac{\|X - m - th\|}{\|Y(m + th) - \Gamma_m\|_F} \right] \\ &\leq \|h\| \mathbb{E} \left[\frac{\sqrt{\|Y(m + th)\|_F}}{\|Y(m + th) - \Gamma_m\|_F} \right] \\ &\leq \|h\| \left(\mathbb{E} \left[\frac{\sqrt{\|\Gamma_m\|_F}}{\|Y(m + th) - \Gamma_m\|_F} \right] + \mathbb{E} \left[\frac{1}{\sqrt{\|Y(m + th) - \Gamma_m\|_F}} \right] \right). \end{aligned}$$

Thus, since $\mathbb{E} \left[\frac{1}{\|Y(m + th) - \Gamma_m\|_F} \right] \leq C$,

$$\mathbb{E} \left[\left\| \frac{h(X - m - th)^T}{\|Y(m + th) - \Gamma_m\|_F} \right\|_F \right] \leq \|h\| \left(C\sqrt{\|\Gamma_m\|_F} + \sqrt{C} \right). \quad (35)$$

In the same way,

$$\mathbb{E} \left[\left\| \frac{(X - m - th)h^T}{\|Y(m + th) - \Gamma_m\|_F} \right\|_F \right] \leq \|h\| \left(C\sqrt{\|\Gamma_m\|_F} + \sqrt{C} \right). \quad (36)$$

Applying Cauchy-Schwarz's inequality,

$$\begin{aligned} \mathbb{E} \left[\left| \left\langle Y(m+th) - \Gamma_m, h(X - m - th)^T \right\rangle_F \frac{\|Y(m+th) - \Gamma_m\|_F}{\|Y(m+th) - \Gamma_m\|_F^3} \right] &\leq \mathbb{E} \left[\frac{\|h(X - m - th)^T\|_F}{\|Y(m+th) - \Gamma_m\|_F} \right] \\ &\leq \|h\| \mathbb{E} \left[\frac{\|X - m - th\|}{\|Y(m+th) - \Gamma_m\|_F} \right] \\ &\leq \|h\| \mathbb{E} \left[\frac{\sqrt{\|Y(m+th)\|_F}}{\|Y(m+th) - \Gamma_m\|_F} \right]. \end{aligned}$$

Thus, since $\mathbb{E} \left[\frac{1}{\|Y(m+th) - \Gamma_m\|_F} \right] \leq C$, and since for all positive constants a, b , $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$,

$$\begin{aligned} \|h\| \mathbb{E} \left[\frac{\sqrt{\|Y(m+th)\|_F}}{\|Y(m+th) - \Gamma_m\|_F} \right] &\leq \|h\| \left(\mathbb{E} \left[\frac{\sqrt{\|\Gamma_m\|_F}}{\|Y(m+th) - \Gamma_m\|_F} \right] + \mathbb{E} \left[\frac{1}{\sqrt{\|Y(m+th) - \Gamma_m\|_F}} \right] \right) \\ &\leq \|h\| \left(C\sqrt{\|\Gamma_m\|_F} + \sqrt{C} \right). \end{aligned}$$

Finally,

$$\mathbb{E} \left[\left| \left\langle Y(m+th) - \Gamma_m, h(X - m - th)^T \right\rangle_F \frac{\|Y(m+th) - \Gamma_m\|_F}{\|Y(m+th) - \Gamma_m\|_F^3} \right] \leq \|h\| \left(C\sqrt{\|\Gamma_m\|_F} + \sqrt{C} \right), \quad (37)$$

$$\mathbb{E} \left[\left| \left\langle Y(m+th) - \Gamma_m, (X - m - th) h^T \right\rangle_F \frac{\|Y(m+th) - \Gamma_m\|_F}{\|Y(m+th) - \Gamma_m\|_F^3} \right] \leq \|h\| \left(C\sqrt{\|\Gamma_m\|_F} + \sqrt{C} \right). \quad (38)$$

Applying inequalities (35) to (38) with $h = \bar{m}_n - m$, the announced result is proven,

$$\|r_n\|_F \leq 4 \left(\sqrt{C} + C\sqrt{\|\Gamma_m\|_F} \right) \|\bar{m}_n - m\|.$$

Bounding $\|r'_n\|$

For all $h \in H$ and $V \in \mathcal{S}(H)$, we define the random function $\varphi_{h,V} : [0, 1] \rightarrow \mathcal{S}(H)$ such that for all $t \in [0, 1]$,

$$\varphi_{h,V}(t) := \frac{1}{\|Y(m+th) - \Gamma_m\|_F} \left(I_{\mathcal{S}(H)} - \frac{(Y(m+th) - \Gamma_m) \otimes_F (Y(m+th) - \Gamma_m)}{\|Y(m+th) - \Gamma_m\|_F^2} \right) (V).$$

Note that $r'_n = \mathbb{E} \left[\varphi_{\bar{m}_n - m, V_n - \Gamma_m}(1) - \varphi_{\bar{m}_n - m, V_n - \Gamma_m}(0) \middle| \mathcal{F}_n \right]$. By dominated convergence,

$$\|r'_n\|_F \leq \sup_{t \in [0, 1]} \mathbb{E} \left[\left\| \varphi'_{\bar{m}_n - m, V_n - \Gamma_m}(t) \right\|_F \middle| \mathcal{F}_n \right].$$

Moreover, as for the bound of $\|r_n\|$, one can check, with an application of Cauchy-Schwarz's

inequality, that for all $h \in H$, $V \in \mathcal{S}(H)$, and $t \in [0, 1]$,

$$\begin{aligned} \varphi'_{h,V}(t) &\leq 6 \frac{\|Y(m+th) - \Gamma_m\|_F \|h^T(X - m - th)\|_F}{\|Y(m+th) - \Gamma_m\|_F^3} \|V\|_F \\ &\quad + 6 \frac{\|Y(m+th) - \Gamma_m\|_F \|h(X - m - th)^T\|_F}{\|Y(m+th) - \Gamma_m\|_F^5} \|(Y(m+th) - \Gamma_m) \otimes_F (Y(m+th) - \Gamma_m)(V)\|_F \\ &\leq 12 \frac{\|h(X - m - th)^T\|_F}{\|Y(m+th) - \Gamma_m\|_F^2} \|V\|_F. \end{aligned}$$

Finally,

$$\begin{aligned} \mathbb{E} \left[\frac{\|h(X - m - th)^T\|_F}{\|Y(m+th) - \Gamma_m\|_F^2} \|V\|_F \right] &\leq \mathbb{E} \left[\frac{\|h\| \|X - m - th\|}{\|Y(m+th) - \Gamma_m\|_F^2} \|V\|_F \right] \\ &\leq \|h\| \|V\|_F \mathbb{E} \left[\frac{\sqrt{\|\Gamma_m\|_F}}{\|Y(m+th) - \Gamma_m\|_F^2} \right] \\ &\quad + \|h\| \|V\|_F \mathbb{E} \left[\frac{1}{\|Y(m+th) - \Gamma_m\|_F^{3/2}} \right] \\ &\leq \left(C\sqrt{\|\Gamma_m\|_F} + C^{3/4} \right) \|h\| \|V\|_F. \end{aligned} \quad (39)$$

Then the announced result follows from an application of inequality (39) with $h = \bar{m}_n - m$ and $V = V_n - \Gamma_m$,

$$\|r'_n\| \leq 12 \left(C\sqrt{\|\Gamma_m\|_F} + C^{3/4} \right) \|\bar{m}_n - m\| \|V_n - \Gamma_m\|_F.$$

□

D Proofs of Lemma 5.2, 5.3 and 5.4

Proof of Lemma 5.2. Using decomposition (33),

$$\begin{aligned} \|V_{n+1} - \Gamma_m\|_F^2 &= \|V_n - \Gamma_m\|_F^2 - 2\gamma_n \langle V_n - \Gamma_m, \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_F \\ &\quad + \gamma_n^2 \|\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)\|_F^2 \\ &\quad + \gamma_n^2 \|\xi_{n+1}\|_F^2 + 2\gamma_n \langle V_n - \Gamma_m - \gamma_n (\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)), \xi_{n+1} \rangle_F \\ &\quad + \gamma_n^2 \|r_n\|_F^2 - 2\gamma_n \langle r_n, V_n - \Gamma_m \rangle_F - 2\gamma_n^2 \langle r_n, \xi_{n+1} - \nabla G_{\bar{m}_n}(V_n) + \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_F. \end{aligned}$$

Note that for all $h \in H$ and $V \in \mathcal{S}(H)$ we have $\|\nabla G_h(V)\|_F \leq 1$. Moreover, $\|r_n\|_F \leq 2$ and $\|\xi_{n+1}\|_F \leq 2$. Since for all $h \in H$, G_h is a convex function, we get with Cauchy-Schwarz's inequality,

$$\|V_{n+1} - \Gamma_m\|_F^2 a \leq \|V_n - \Gamma_m\|_F^2 + 36\gamma_n^2 + 2\gamma_n \langle \xi_{n+1}, V_n - \Gamma_m \rangle_F - 2\gamma_n \langle r_n, V_n - \Gamma_m \rangle_F. \quad (40)$$

Let $C' := 4 \left(\sqrt{C} + C\sqrt{\|\Gamma_m\|_F} \right)$, let us recall that $\|r_n\|_F \leq C' \|\bar{m}_n - m\|$. We now prove by induction that for all integer $p \geq 1$, there is a positive constant M_p such that for all $n \geq 1$, $\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \leq M_p$.

The case $p = 1$ has been studied in the proof of Theorem 3.2. Let $p \geq 2$ and suppose from now that for all $k \leq p - 1$, there is a positive constant M_k such that for all $n \geq 1$,

$$\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2k} \right] \leq M_k.$$

Bounding $\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right]$.

Let us apply inequality (40), for all $p \geq 2$ and use the fact that (ξ_n) is a sequence of martingales differences adapted to the filtration (\mathcal{F}_n) ,

$$\begin{aligned} & \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^{2p} \right] \leq \mathbb{E} \left[\left(\|V_n - \Gamma_m\|_F^2 + 36\gamma_n^2 + 2\gamma_n \|r_n\|_F \|V_n - \Gamma_m\|_F \right)^p \right] \\ & + \sum_{k=2}^p \binom{p}{k} \mathbb{E} \left[(2\gamma_n \langle V_n - \Gamma_m, \xi_{n+1} \rangle_F)^k \left(\|V_n - \Gamma_m\|_F^2 + 36\gamma_n^2 + 2\gamma_n \|r_n\|_F \|V_n - \Gamma_m\|_F \right)^{p-k} \right]. \end{aligned} \quad (41)$$

Let us denote by $(*)$ the second term on the right-hand side of inequality (41). Applying Cauchy-Schwarz's inequality and since $\|\xi_{n+1}\|_F \leq 2$,

$$\begin{aligned} (*) & = \sum_{k=2}^p \binom{p}{k} \mathbb{E} \left[(2\gamma_n \langle V_n - \Gamma_m, \xi_{n+1} \rangle_F)^k \left(\|V_n - \Gamma_m\|_F^2 + 36\gamma_n^2 + 2\gamma_n \|r_n\|_F \|V_n - \Gamma_m\|_F \right)^{p-k} \right] \\ & \leq \sum_{k=2}^p \binom{p}{k} 2^{2k} \gamma_n^k \mathbb{E} \left[\|V_n - \Gamma_m\|_F^k \left(\|V_n - \Gamma_m\|_F^2 + 36\gamma_n^2 + 2\gamma_n \|r_n\|_F \|V_n - \Gamma_m\|_F \right)^{p-k} \right]. \end{aligned}$$

With the help of Lemma E.1,

$$\begin{aligned} (*) & \leq \sum_{k=2}^p 2^{2k} 3^{p-k-1} \gamma_n^k \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p-k} \right] + \sum_{k=2}^p 2^{2k} 3^{p-k-1} 36^{p-k} \gamma_n^{2p-k} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^k \right] \\ & \quad + \sum_{k=2}^p 2^{p+k} 3^{p-k-1} \gamma_n^p \mathbb{E} \left[\|r_n\|_F^{p-k} \|V_n - \Gamma_m\|_F^p \right]. \end{aligned}$$

Applying Cauchy-Schwarz's inequality,

$$\begin{aligned} \sum_{k=2}^p 2^{2k} 3^{p-k-1} \gamma_n^k \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p-k} \right] & = \sum_{k=2}^p 2^{2k} 3^{p-k-1} \gamma_n^k \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{p-1} \|V_n - \Gamma_m\|_F^{p+1-k} \right] \\ & \leq \sum_{k=2}^p 2^{2k} 3^{p-k-1} \gamma_n^k \sqrt{\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2(p-1)} \right]} \sqrt{\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2(p+1-k)} \right]}. \end{aligned}$$

By induction,

$$\begin{aligned} \sum_{k=2}^p 2^{2k} 3^{p-k-1} \gamma_n^k \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p-k} \right] & \leq \sum_{k=2}^p 2^{2k} 3^{p-k-1} \gamma_n^k \sqrt{M_{p-1}} \sqrt{M_{p+1-k}} \\ & = O(\gamma_n^2). \end{aligned} \quad (42)$$

In the same way, applying Cauchy-Schwarz's inequality and by induction,

$$\begin{aligned}
\sum_{k=2}^p 2^{2k} 3^{p-k-1} 36^{p-k} \gamma_n^{2p-k} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^k \right] &= \sum_{k=2}^p 2^{2k} 3^{p-k-1} 36^{p-k} \gamma_n^{2p-k} \mathbb{E} \left[\|V_n - \Gamma_m\|_F \|V_n - \Gamma_m\|_F^{k-1} \right] \\
&\leq \sum_{k=2}^p 2^{2k} 3^{p-k-1} 36^{p-k} \gamma_n^{2p-k} \sqrt{M_1} \sqrt{M_{k-1}} \\
&= O(\gamma_n^2), \tag{43}
\end{aligned}$$

since $p \geq 2$. Similarly, since $\|r_n\|_F \leq 2$ and since $p \geq 2$, applying Cauchy-Schwarz's inequality and by induction,

$$\begin{aligned}
\sum_{k=2}^p 2^{p+k} 3^{p-k-1} \gamma_n^p \mathbb{E} \left[\|r_n\|_F^{p-k} \|V_n - \Gamma_m\|_F^k \right] &\leq \sum_{k=2}^p 2^{2p} 3^{p-k-1} \gamma_n^p \mathbb{E} \left[\|V_n - \Gamma_m\|_F^k \right] \\
&\leq \sum_{k=2}^p 2^{2p} 3^{p-k-1} \gamma_n^p \sqrt{M_1} \sqrt{M_{p-1}} \\
&= O(\gamma_n^2). \tag{44}
\end{aligned}$$

Finally, applying inequalities (42) to (44), there is a positive constant A'_1 such that for all $n \geq 1$,

$$\mathbb{E} \left[\sum_{k=2}^p \binom{p}{k} (2\gamma_n \langle V_n - \Gamma_m, \xi_{n+1} \rangle_F)^k \left(\|V_n - \Gamma_m\|_F^2 + 36\gamma_n^2 + 2\gamma_n \|r_n\|_F \|V_n - \Gamma_m\|_F \right)^{p-k} \right] \leq A'_1 \gamma_n^2. \tag{45}$$

We now denote by (**) the first term at the right-hand side of inequality (41). With the help of Lemma E.1 and applying Cauchy-Schwarz's inequality,

$$\begin{aligned}
(**) &\leq \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] + \sum_{k=1}^p \binom{p}{k} \mathbb{E} \left[(36\gamma_n^2 + 2\gamma_n \langle r_n, V_n - \Gamma_m \rangle_F)^k \|V_n - \Gamma_m\|_F^{2p-2k} \right] \\
&\leq \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] + \sum_{k=1}^p \binom{p}{k} 2^{k-1} \mathbb{E} \left[(36^k \gamma_n^{2k} + 2^k \gamma_n^k \|r_n\|_F^k \|V_n - \Gamma_m\|_F^k) \|V_n - \Gamma_m\|_F^{2p-2k} \right].
\end{aligned}$$

Moreover, let

$$\begin{aligned}
(***) &:= \sum_{k=1}^p \binom{p}{k} 2^{k-1} \mathbb{E} \left[(36^k \gamma_n^{2k} + 2^k \gamma_n^k \|r_n\|_F^k \|V_n - \Gamma_m\|_F^k) \|V_n - \Gamma_m\|_F^{2p-2k} \right] \\
&= \sum_{k=1}^p \binom{p}{k} 2^{k-1} 36^k \gamma_n^{2k} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p-2k} \right] + \sum_{k=1}^p \binom{p}{k} 2^{2k-1} \gamma_n^k \mathbb{E} \left[\|r_n\|_F^k \|V_n - \Gamma_m\|_F^{2p-k} \right].
\end{aligned}$$

By induction,

$$\begin{aligned}
\sum_{k=1}^p \binom{p}{k} 2^{k-1} 36^k \gamma_n^{2k} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p-2k} \right] &= \sum_{k=1}^p \binom{p}{k} 2^{k-1} 36^k \gamma_n^{2k} M_{p-k} \\
&= O(\gamma_n^2).
\end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{k=1}^p \binom{p}{k} 2^{2k-1} \gamma_n^k \mathbb{E} \left[\|r_n\|_F^k \|V_n - \Gamma_m\|_F^{2p-k} \right] &= \sum_{k=2}^p \binom{p}{k} 2^{2k-1} \gamma_n^k \mathbb{E} \left[\|r_n\|_F^k \|V_n - \Gamma_m\|_F^{2p-k} \right] \\ &\quad + 2p\gamma_n \mathbb{E} \left[\|r_n\|_F \|V_n - \Gamma_m\|_F^{2p-1} \right]. \end{aligned}$$

Applying Cauchy-Schwarz's inequality and by induction, since $\|r_n\|_F \leq 2$,

$$\begin{aligned} \sum_{k=2}^p \binom{p}{k} 2^{2k-1} \gamma_n^k \mathbb{E} \left[\|r_n\|_F^k \|V_n - \Gamma_m\|_F^{2p-k} \right] &\leq \sum_{k=2}^p \binom{p}{k} 2^{3k-1} \gamma_n^k \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p-k} \right] \\ &\leq \sum_{k=2}^p \binom{p}{k} 2^{3k-1} \gamma_n^k \sqrt{M_{p+1-k}} \sqrt{M_{p-1}} \\ &= O(\gamma_n^2). \end{aligned}$$

Moreover, applying Theorem 4.2 in Godichon-Baggioni (2016) and Hölder's inequality, since $\|r_n\|_F \leq C' \|\bar{m}_n - m\|$,

$$\begin{aligned} 2p\gamma_n \mathbb{E} \left[\|r_n\|_F \|V_n - \Gamma_m\|_F^{2p-1} \right] &\leq 2C' p \gamma_n \mathbb{E} \left[\|\bar{m}_n - m\| \|V_n - \Gamma_m\|_F^{2p-1} \right] \\ &\leq 2C' p \gamma_n \left(\mathbb{E} \left[\|\bar{m}_n - m\|^{2p} \right] \right)^{\frac{1}{2p}} \left(\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \right)^{\frac{2p-1}{2p}} \\ &\leq 2C' p \gamma_n \frac{K_p^{\frac{1}{2p}}}{n^{1/2}} \left(\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \right)^{\frac{2p-1}{2p}}. \end{aligned}$$

Finally,

$$\begin{aligned} 2C' p \gamma_n \frac{K_p^{\frac{1}{2p}}}{n^{1/2}} \left(\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \right)^{\frac{2p-1}{2p}} &\leq 2C' p \gamma_n \frac{K_p^{\frac{1}{2p}}}{n^{1/2}} \max \left\{ 1, \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \right\} \\ &\leq 2C' p \gamma_n \frac{K_p^{\frac{1}{2p}}}{n^{1/2}} \left(1 + \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] \right). \end{aligned}$$

Thus, there are positive constants A_0'', A_1'' such that

$$(**) \leq \left(1 + A_0'' \frac{1}{n^{\alpha+1/2}} \right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] + A_1'' \frac{1}{n^{\alpha+1/2}}. \quad (46)$$

Finally, thanks to inequalities (45) and (46), there are positive constants A'_0, A'_1 such that

$$\begin{aligned} \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^{2p} \right] &\leq \left(1 + A'_0 \frac{1}{n^{\alpha+1/2}} \right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2p} \right] + A'_1 \frac{1}{n^{\alpha+1/2}} \\ &\leq \prod_{k=1}^n \left(1 + A'_0 \frac{1}{k^{\alpha+1/2}} \right) \mathbb{E} \left[\|V_1 - \Gamma_m\|_F^{2p} \right] + \sum_{k=1}^n \prod_{j=k+1}^n \left(1 + A'_0 \frac{1}{j^{\alpha+1/2}} \right) A'_1 \frac{1}{k^{\alpha+1/2}} \\ &\leq \prod_{k=1}^{\infty} \left(1 + A'_0 \frac{1}{k^{\alpha+1/2}} \right) \mathbb{E} \left[\|V_1 - \Gamma_m\|_F^{2p} \right] + \prod_{j=1}^{\infty} \left(1 + A'_0 \frac{1}{j^{\alpha+1/2}} \right) \sum_{k=1}^{\infty} A'_1 \frac{1}{k^{\alpha+1/2}} \\ &\leq M_p, \end{aligned}$$

which concludes the induction and the proof. \square

Proof of Lemma 5.3. Let us define the following linear operators:

$$\begin{aligned}\alpha_n &:= I_{S(H)} - \gamma_n \nabla_m^2 G(\Gamma_m), \\ \beta_n &:= \prod_{k=1}^n \alpha_k = \prod_{k=1}^n (I_{S(H)} - \gamma_k \nabla_m^2 G(\Gamma_m)), \\ \beta_0 &:= I_{S(H)}.\end{aligned}$$

Using decomposition (34) and by induction, for all $n \geq 1$,

$$V_n - \Gamma_m = \beta_{n-1}(V_1 - \Gamma_m) + \beta_{n-1}M_n - \beta_{n-1}R_n - \beta_{n-1}R'_n - \beta_{n-1}\Delta_n, \quad (47)$$

with

$$\begin{aligned}M_n &:= \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \xi_{k+1}, & R_n &:= \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} r_k, \\ R'_n &:= \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} r'_k, & \Delta_n &:= \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \delta_k.\end{aligned}$$

We now study the asymptotic behavior of the linear operators β_n and $\beta_{n-1}\beta_k^{-1}$. As in Cardot et al. (2013), one can check that there are positive constants c_0, c_1 such that for all integers $k, n \geq 1$ with $k \leq n-1$,

$$\|\beta_{n-1}\|_{op} \leq c_0 e^{-\lambda \min \sum_{k=1}^n \gamma_k}, \quad \|\beta_{n-1}\beta_k^{-1}\|_{op} \leq c_1 e^{-\lambda \min \sum_{j=k}^n \gamma_j}, \quad (48)$$

where $\|\cdot\|_{op}$ is the usual spectral norm for linear operators. We now bound the quadratic mean of each term in decomposition (47).

Step 1: the quasi deterministic term $\beta_{n-1}(V_1 - \Gamma_m)$.

Applying inequality (48), there is a positive constant c'_0 such that

$$\begin{aligned}\mathbb{E} \left[\|\beta_{n-1}(V_1 - \Gamma_m)\|_F^2 \right] &\leq \|\beta_{n-1}\|_{op}^2 \mathbb{E} \left[\|V_1 - \Gamma_m\|_F^2 \right] \\ &\leq c_0 e^{-2\lambda \min \sum_{k=1}^n \gamma_k} \mathbb{E} \left[\|V_1 - \Gamma_m\|_F^2 \right] \\ &\leq c_0 e^{-c'_0 n^{1-\alpha}} \mathbb{E} \left[\|V_1 - \Gamma_m\|_F^2 \right].\end{aligned} \quad (49)$$

This term converges exponentially fast to 0.

Step 2: the martingale term $\beta_{n-1}M_n$.

Since (ξ_n) is a sequence of martingale differences adapted to the filtration (\mathcal{F}_n) ,

$$\begin{aligned}\mathbb{E} \left[\|\beta_{n-1}M_n\|_F^2 \right] &= \sum_{k=1}^{n-1} \mathbb{E} \left[\|\beta_{n-1}\beta_k^{-1} \gamma_k \xi_{k+1}\|_F^2 \right] + 2 \sum_{k=1}^{n-1} \sum_{k'=k+1}^{n-1} \gamma_k \gamma_{k'} \mathbb{E} \left[\langle \beta_{n-1}\beta_k^{-1} \xi_{k+1}, \beta_{n-1}\beta_{k'}^{-1} \xi_{k'+1} \rangle_F \right] \\ &= \sum_{k=1}^{n-1} \mathbb{E} \left[\|\beta_{n-1}\beta_k^{-1} \gamma_k \xi_{k+1}\|_F^2 \right] + 2 \sum_{k=1}^{n-1} \sum_{k'=k+1}^{n-1} \gamma_k \gamma_{k'} \mathbb{E} \left[\langle \beta_{n-1}\beta_k^{-1} \xi_{k+1}, \beta_{n-1}\beta_{k'}^{-1} \mathbb{E}[\xi_{k'+1} | \mathcal{F}_{k'}] \rangle_F \right] \\ &= \sum_{k=1}^{n-1} \mathbb{E} \left[\|\beta_{n-1}\beta_k^{-1} \gamma_k \xi_{k+1}\|_F^2 \right].\end{aligned}$$

Moreover, as in Cardot et al. (2016), Lemma E.2 ensures that there is a positive constant C'_1 such that for all $n \geq 1$,

$$\mathbb{E} \left[\|\beta_{n-1} M_n\|_F^2 \right] \leq \frac{C'_1}{n^\alpha}. \quad (50)$$

Step 3: the first remainder term $\beta_{n-1} R_n$.

Remarking that $\|r_n\|_F \leq 4 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right) \|\bar{m}_n - m\|$,

$$\begin{aligned} \mathbb{E} \left[\|\beta_{n-1} R_n\|_F^2 \right] &\leq \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \|r_k\|_F \right)^2 \right] \\ &\leq 16 \left(\sqrt{C} + \sqrt{\|\Gamma_m\|_F} \right)^2 \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \|\bar{m}_k - m\| \right)^2 \right]. \end{aligned}$$

Applying Lemma 4.3 and Theorem 4.2 in Godichon-Baggioni (2016),

$$\begin{aligned} \mathbb{E} \left[\|\beta_{n-1} R_n\|_F^2 \right] &\leq 16 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)^2 \left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \sqrt{\mathbb{E} \left[\|\bar{m}_k - m\|^2 \right]} \right)^2 \\ &\leq 16 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)^2 K_1 \left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \frac{1}{k^{1/2}} \right)^2. \end{aligned}$$

Applying inequality (48),

$$\begin{aligned} \mathbb{E} \left[\|\beta_{n-1} R_n\|_F^2 \right] &\leq 16 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)^2 K_1 \left(\sum_{k=1}^{n-1} \gamma_k e^{-\sum_{j=k}^n \gamma_j} \frac{1}{k^{1/2}} \right)^2 \\ &\leq 16 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)^2 K_1 \left(\sum_{k=1}^n \gamma_k e^{-\sum_{j=k}^n \gamma_j} \frac{1}{k^{1/2}} \right)^2. \end{aligned}$$

Splitting the sum into two parts and applying Lemma E.2, we have

$$\begin{aligned} \mathbb{E} \left[\|\beta_{n-1} R_n\|_F^2 \right] &\leq 32 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)^2 K_1 \left(\sum_{k=1}^{E(n/2)} \gamma_k e^{-\sum_{j=k}^n \gamma_j} \frac{1}{k^{1/2}} \right)^2 \\ &\quad + 32 \left(\sqrt{C} + C \sqrt{\|\Gamma_m\|_F} \right)^2 K_1 \left(\sum_{k=E(n/2)+1}^n \gamma_k e^{-\sum_{j=k}^n \gamma_j} \frac{1}{k^{1/2}} \right)^2 \\ &= O \left(\frac{1}{n} \right). \end{aligned}$$

Thus, there is a positive constant C'_2 such that for all $n \geq 1$,

$$\mathbb{E} \left[\|\beta_{n-1} R_n\|_F^2 \right] \leq \frac{C'_2}{n}. \quad (51)$$

Step 4: the second remainder term $\beta_{n-1} R'_n$.

Let us recall that for all $n \geq 1$, $\|r'_n\|_F \leq 12D \|\bar{m}_n - m\| \|V_n - \Gamma_m\|_F$ with $D :=$

$C\sqrt{\|\Gamma_m\|_F} + C^{3/4}$. Thus,

$$\begin{aligned}\mathbb{E} \left[\|\beta_{n-1} R'_n\|_F^2 \right] &\leq \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \|r'_k\|_F \right)^2 \right] \\ &\leq 144D^2 \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \|\bar{m}_k - m\| \|V_k - \Gamma_m\|_F \right)^2 \right].\end{aligned}$$

Applying Lemma 4.3 in Godichon-Baggioni (2016),

$$\mathbb{E} \left[\|\beta_{n-1} R'_n\|_F^2 \right] \leq 144D^2 \left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \sqrt{\mathbb{E} \left[\|\bar{m}_k - m\|^2 \|V_k - \Gamma_m\|_F^2 \right]} \right)^2.$$

Thanks to Lemma 5.2, there is a positive constant M_2 such that for all $n \geq 1$, $\mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \leq M_2$. Thus, applying Cauchy-Schwarz's inequality and Theorem 4.2 in Godichon-Baggioni (2016),

$$\begin{aligned}\mathbb{E} \left[\|\beta_{n-1} R'_n\|_F^2 \right] &\leq 144D^2 \left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \left(\mathbb{E} \left[\|\bar{m}_k - m\|^4 \right] \right)^{\frac{1}{4}} \left(\mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right] \right)^{\frac{1}{4}} \right)^2 \\ &\leq 144D^2 \sqrt{M_2 K_2} \left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \frac{1}{k^{1/2}} \right)^2.\end{aligned}$$

As in step 3, splitting the sum into two parts, one can check that there is a positive constant C_1'' such that for all $n \geq 1$,

$$\mathbb{E} \left[\|\beta_{n-1} R'_n\|_F^2 \right] \leq \frac{C_1''}{n}. \quad (52)$$

Step 5: the third remainder term: $\beta_{n-1} \Delta_n$

Since $\|\delta_n\|_F \leq 6C \|V_n - \Gamma_m\|_F^2$, applying Lemma 4.3 in Godichon-Baggioni (2016),

$$\begin{aligned}\mathbb{E} \left[\|\beta_{n-1} \Delta_n\|_F^2 \right] &\leq \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \|\delta_k\|_F \right)^2 \right] \\ &\leq 36C^2 \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \|V_k - \Gamma_m\|_F^2 \right)^2 \right] \\ &\leq 36C^2 \left(\sum_{k=1}^{n-1} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_{op} \sqrt{\mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right]} \right)^2.\end{aligned}$$

Thanks to Lemma 5.2, there is a positive constant M_2 such that for all $n \geq 1$, $\mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \leq M_2$. Thus, splitting the sum into two parts and applying inequalities (48) and Lemma E.2, there

are positive constant c'_0, C'_2 such that for all $n \geq 1$,

$$\begin{aligned} \mathbb{E} \left[\|\beta_{n-1} \Delta_n\|_F^2 \right] &\leq 72C^2 M_2^2 \left(\sum_{k=1}^{E(n/2)} \gamma_k e^{-\sum_{j=k}^n \gamma_j} \right)^2 \\ &\quad + 72C^2 \sup_{E(n/2)+1 \leq k \leq n-1} \left\{ \mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right] \right\} \left(\sum_{k=E(n/2)+1}^n \gamma_k e^{-\sum_{j=k}^n \gamma_j} \right)^2 \\ &\leq C'_2 \sup_{E(n/2)+1 \leq k \leq n-1} \left\{ \mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right] \right\} + O \left(e^{-2c'_0 n^{1-\alpha}} \right). \end{aligned}$$

Thus, there is a positive constant C'_0 such that for all $n \geq 1$,

$$\mathbb{E} \left[\|\beta_{n-1} \Delta_n\|_F^2 \right] \leq C'_0 e^{-2c'_0 n^{1-\alpha}} + C'_2 \sup_{E(n/2)+1 \leq k \leq n-1} \left\{ \mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right] \right\}. \quad (53)$$

Conclusion:

Applying Lemma E.1 and decomposition (47), for all $n \geq 1$,

$$\begin{aligned} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] &\leq 5\mathbb{E} \left[\|\beta_{n-1} (V_1 - \Gamma_m)\|_F^2 \right] + 5\mathbb{E} \left[\|\beta_{n-1} M_n\|_F^2 \right] + 5\mathbb{E} \left[\|\beta_{n-1} R_n\|_F^2 \right] \\ &\quad + 5\mathbb{E} \left[\|\beta_{n-1} R'_n\|_F^2 \right] + 5\mathbb{E} \left[\|\beta_{n-1} \Delta_n\|_F^2 \right]. \end{aligned}$$

Applying inequalities (49) to (53), there are positive constants C_1, C'_1, C_2, C_3 such that for all $n \geq 1$,

$$\mathbb{E} \left[\|V_n - \Gamma_m\|^2 \right] \leq C_1 e^{-C'_1 n^{1-\alpha}} + \frac{C_2}{n^\alpha} + C_3 \sup_{E(n/2)+1 \leq k \leq n-1} \mathbb{E} \left[\|V_k - \Gamma_m\|_F^4 \right].$$

□

Proof of Lemma 5.4. Let us define $W_n := V_n - \Gamma_m - \gamma_n (\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m))$ and use decomposition (33),

$$\begin{aligned} \|V_{n+1} - \Gamma_m\|_F^2 &= \|W_n\|_F^2 + \gamma_n^2 \|\xi_{n+1}\|_F^2 + \gamma_n^2 \|r_n\|_F^2 + 2\gamma_n \langle \xi_{n+1}, V_n - \Gamma_m \rangle_F + 2\gamma_n^2 \langle \xi_{n+1}, \nabla G_{\bar{m}_n}(V_n) \rangle_F \\ &\quad - 2\gamma_n^2 \langle r_n, \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_F - 2\gamma_n \langle r_n, V_n - \Gamma_m \rangle_F. \end{aligned}$$

Since $\|\xi_{n+1}\|_F \leq 2$, $\|r_n\|_F \leq 2$ and the fact that for all $h \in H$, $V \in \mathcal{S}(H)$, $\nabla_h G(V) \leq 1$, we get with an application of Cauchy-Schwarz's inequality

$$\|V_{n+1} - \Gamma_m\|_F^2 \leq \|W_n\|_F^2 + 2\gamma_n \langle \xi_{n+1}, V_n - \Gamma_m \rangle_F + 2\gamma_n \|r_n\|_F \|V_n - \Gamma_m\|_F + 20\gamma_n^2.$$

Thus, since (ξ_n) is a sequence of martingale differences adapted to the filtration (\mathcal{F}_n) , and since $\|W_n\|_F^2 \leq (1 + C^2 c_\gamma^2) \|V_n - \Gamma_m\|_F^2$ (this inequality follows from Proposition B.1 and from the fact that for all $h \in H$, G_h is a convex application),

$$\begin{aligned} \mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] &\leq \mathbb{E} \left[\|W_n\|_F^4 \right] + 2\gamma_n \mathbb{E} \left[\|r_n\|_F \|W_n\|_F^2 \|V_n - \Gamma_m\|_F \right] \\ &\quad + 40 (1 + C^2 c_\gamma^2) \gamma_n^2 \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] \\ &\quad + 4\gamma_n^2 \mathbb{E} \left[\langle \xi_{n+1}, V_n - \Gamma_m \rangle_F^2 \right] + 400\gamma_n^4 + 40\gamma_n^3 \mathbb{E} \left[\|r_n\|_F \|V_n - \Gamma_m\|_F^2 \right] \\ &\quad + 4\gamma_n^2 \mathbb{E} \left[\|r_n\|_F^2 \|V_n - \Gamma_m\|_F^2 \right]. \end{aligned}$$

Since $\|\xi_{n+1}\|_F \leq 2$ and $\|r_n\|_F \leq 2$, applying Cauchy-Schwarz's inequality, there are positive constants C'_1, C'_2 such that for all $n \geq 1$,

$$\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] \leq \mathbb{E} \left[\|W_n\|_F^4 \right] + 2\gamma_n \mathbb{E} \left[\|r_n\|_F \|W_n\|_F^2 \|V_n - \Gamma_m\|_F \right] + \frac{C'_1}{n^{3\alpha}} + \frac{C'_2}{n^{2\alpha}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right]. \quad (54)$$

We now bound the two first terms at the right-hand side of inequality (54).

Step 1: bounding $\mathbb{E} \left[\|W_n\|_F^4 \right]$.

Since $\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) = \int_0^1 \nabla_{\bar{m}_n}^2 G(\Gamma_m + t(V_n - \Gamma_m))(V_n - \Gamma_m) dt$, applying Proposition B.1, one can check that

$$\begin{aligned} \|W_n\|^2 &= \|V_n - \Gamma_m\|_F^2 - 2\gamma_n \langle V_n - \Gamma_m, \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_H + \gamma_n^2 \|\nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m)\|_F^2 \\ &\leq (1 + C^2 \gamma_n^2) \|V_n - \Gamma_m\|_F^2 - 2\gamma_n \langle V_n - \Gamma_m, \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m) \rangle_H. \end{aligned}$$

Since for all $h \in H$, G_h is a convex application, $\|W_n\|_F^2 \leq (1 + c_\gamma^2 C^2) \|V_n - \Gamma_m\|_F^2$. Let p' be a positive integer. We now introduce the sequence of events $(A_{n,p'})_{n \in \mathbb{N}}$ defined for all $n \geq 1$ by

$$A_{n,p'} := \left\{ \omega \in \Omega, \quad \|V_n(\omega) - \Gamma_m\|_F \leq n^{-\frac{1-\alpha}{p'}}, \quad \text{and} \quad \|\bar{m}_n(\omega) - m\| \leq \epsilon \right\}, \quad (55)$$

with ϵ defined in Proposition B.1. For the sake of simplicity, we consider that ϵ' defined in Proposition B.1 verifies $\epsilon' \leq 1$. Applying Proposition B.1, let

$$\begin{aligned} B_n &:= \langle \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m), V_n - \Gamma_m \rangle_F \mathbf{1}_{A_{n,p'}} \mathbf{1}_{\{\|V_n - \Gamma_m\|_F \leq \epsilon'\}} \\ &= \int_0^1 \langle \nabla_{\bar{m}_n}^2 G(\Gamma_m + t(V_n - \Gamma_m))(V_n - \Gamma_m), V_n - \Gamma_m \rangle_F \mathbf{1}_{\{\|V_n - \Gamma_m\|_F \leq \epsilon'\}} \mathbf{1}_{A_{n,p'}} dt \\ &\geq \frac{1}{2} c_m \|V_n - \Gamma_m\|_F^2 \mathbf{1}_{\{\|V_n - \Gamma_m\|_F \leq \epsilon'\}} \mathbf{1}_{A_{n,p'}}. \end{aligned} \quad (56)$$

In the same way, since $G_{\bar{m}_n}$ is convex, let

$$\begin{aligned} B'_n &:= \langle \nabla G_{\bar{m}_n}(V_n) - \nabla G_{\bar{m}_n}(\Gamma_m), V_n - \Gamma_m \rangle_F \mathbf{1}_{A_{n,p'}} \mathbf{1}_{\{\|V_n - \Gamma_m\|_F > \epsilon'\}} \\ &= \int_0^1 \langle \nabla_{\bar{m}_n}^2 G(\Gamma_m + t(V_n - \Gamma_m))(V_n - \Gamma_m), V_n - \Gamma_m \rangle_F \mathbf{1}_{\{\|V_n - \Gamma_m\|_F > \epsilon'\}} \mathbf{1}_{A_{n,p'}} dt \\ &\geq \int_0^{\frac{\epsilon'}{\|V_n - \Gamma_m\|_F}} \langle \nabla_{\bar{m}_n}^2 G(\Gamma_m + t(V_n - \Gamma_m))(V_n - \Gamma_m), V_n - \Gamma_m \rangle_F \mathbf{1}_{\{\|V_n - \Gamma_m\|_F > \epsilon'\}} \mathbf{1}_{A_{n,p'}} dt \end{aligned}$$

Applying Proposition B.1,

$$\begin{aligned} B'_n &\geq \int_0^{\frac{\epsilon'}{\|V_n - \Gamma_m\|_F}} \frac{1}{2} c_m \|V_n - \Gamma_m\|_F^2 \mathbf{1}_{\{\|V_n - \Gamma_m\|_F > \epsilon'\}} \mathbf{1}_{A_{n,p'}} dt \\ &\geq \frac{\epsilon' c_m}{2 \|V_n - \Gamma_m\|_F} \|V_n - \Gamma_m\|_F^2 \mathbf{1}_{\{\|V_n - \Gamma_m\|_F > \epsilon'\}} \mathbf{1}_{A_{n,p'}} \\ &\geq \frac{\epsilon' c_m}{2} n^{-\frac{1-\alpha}{p'}} \|V_n - \Gamma_m\|_F^2 \mathbf{1}_{\{\|V_n - \Gamma_m\|_F > \epsilon'\}} \mathbf{1}_{A_{n,p'}}. \end{aligned} \quad (57)$$

There is a rank $n'_{p'}$ such that for all $n \geq n'_{p'}$, we have $\frac{\epsilon' c_m}{2} n^{-\frac{1-\alpha}{p'}} \leq \frac{1}{2} c_m$. Thus, applying inequalities (56) and (57), for all $n \geq n'_{p'}$,

$$\|W_n\|_F^2 \mathbf{1}_{A_{n,p'}} \leq \left(1 - \frac{\epsilon' c_m}{2} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \|V_n - \Gamma_m\|_F^2 \mathbf{1}_{A_{n,p'}}.$$

Thus, there are a positive constant $c_{p'}$ and a rank $n_{p'}$ such that for all $n \geq n_{p'}$,

$$\begin{aligned} \mathbb{E} \left[\|W_n\|_F^4 \mathbf{1}_{A_{n,p'}} \right] &\leq \left(1 - \frac{\epsilon' c_m}{2} \gamma_n n^{-\frac{1-\alpha}{p'}}\right)^2 \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \mathbf{1}_{A_{n,p'}} \right] \\ &\leq \left(1 - 2c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right]. \end{aligned} \quad (58)$$

Now, we must get an upper bound for $\mathbb{E} \left[\|W_n\|_F^4 \mathbf{1}_{A_{n,p'}^c} \right]$. Since $\|W_n\|_F^2 \leq (1 + c_\gamma^2 C^2) \|V_n - \Gamma_m\|_F^2$ and since there is a positive constant c_0 such that for all $n \geq 1$,

$$\|V_n - \Gamma_m\|_F \leq \|V_1 - \Gamma_m\|_F + \sum_{k=1}^n \gamma_k \leq c_0 n^{1-\alpha}$$

we have

$$\begin{aligned} \mathbb{E} \left[\|W_n\|_F^4 \mathbf{1}_{A_{n,p'}^c} \right] &\leq (1 + c_\gamma^2 C^2)^2 \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \mathbf{1}_{A_{n,p'}^c} \right] \\ &\leq (1 + c_\gamma^2 C^2)^2 c_0^4 n^{4-4\alpha} \mathbb{P} \left[A_{n,p'}^c \right] \\ &\leq (1 + c_\gamma^2 C^2)^2 c_0^4 n^{4-4\alpha} \left(\mathbb{P} \left[\|\bar{m}_n - m\| \geq \epsilon \right] + \mathbb{P} \left[\|V_n - \Gamma_m\|_F \geq n^{\frac{1-\alpha}{p'}} \right] \right). \end{aligned}$$

Applying Markov's inequality, Theorem 4.2 in Godichon-Baggioni (2016) and Lemma 5.2,

$$\begin{aligned} \mathbb{E} \left[\|W_n\|_F^4 \mathbf{1}_{A_{n,p'}^c} \right] &\leq (1 + c_\gamma^2 C^2)^2 c_0^4 n^{4-4\alpha} \left(\frac{\mathbb{E} \left[\|\bar{m}_n - m\|^{2p''} \right]}{\epsilon^{2p''}} + \frac{\mathbb{E} \left[\|V_n - \Gamma_m\|_F^{2q} \right]}{n^{2q \frac{1-\alpha}{p'}}} \right) \\ &\leq \frac{K p''}{\epsilon^{2p''}} (1 + c_\gamma^2 C^2)^2 c_0^4 n^{4-4\alpha-p''} + (1 + c_\gamma^2 C^2)^2 c_0^4 M_q n^{4-4\alpha-2q \frac{1-\alpha}{p'}}. \end{aligned}$$

Taking $p'' \geq 4 - \alpha$ and $q \geq p' \frac{4-\alpha}{2(1-\alpha)}$,

$$\mathbb{E} \left[\|W_n\|_F^4 \mathbf{1}_{A_{n,p'}^c} \right] = O \left(\frac{1}{n^{3\alpha}} \right). \quad (59)$$

Thus, applying inequalities (58) and (59), there are positive constants $c_{p'}$, $C_{1,p'}$ and a rank $n_{p'}$ such that for all $n \geq n_{p'}$,

$$\mathbb{E} \left[\|W_n\|_F^4 \right] \leq \left(1 - 2c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] + \frac{C_{1,p'}}{n^{3\alpha}}. \quad (60)$$

Step 2: bounding $2\gamma_n \mathbb{E} \left[\|r_n\|_F \|W_n\|_F^2 \|V_n - \Gamma_m\|_F \right]$.

Since $\|W_n\|_F^2 \leq (1 + c_\gamma^2 C^2) \|V_n - \Gamma_m\|_F^2$, applying Lemma E.1, let

$$\begin{aligned} D_n &:= 2\gamma_n \mathbb{E} \left[\|r_n\|_F \|W_n\|_F^2 \|V_n - \Gamma_m\|_F \right] \\ &\leq 2(1 + c_\gamma^2 C^2) \gamma_n \mathbb{E} \left[\|r_n\|_F \|V_n - \Gamma_m\|_F^3 \right] \\ &\leq \frac{2}{c_{p'}} (1 + c_\gamma^2 C^2)^2 \gamma_n n^{\frac{1-\alpha}{p'}} \mathbb{E} \left[\|r_n\|_F^2 \|V_n - \Gamma_m\|_F^2 \right] + \frac{1}{2} c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \\ &\leq \frac{2}{c_{p'}^2} (1 + c_\gamma^2 C^2)^4 \gamma_n n^{3 \frac{1-\alpha}{p'}} \mathbb{E} \left[\|r_n\|_F^4 \right] + c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right]. \end{aligned}$$

Since $\|r_n\|_F \leq \left(\sqrt{C} + C\sqrt{\|\Gamma_m\|_F}\right) \|\bar{m}_n - m\|_F$ and applying Theorem 4.2 in Godichon-Baggioni (2016),

$$\begin{aligned}
D_n &\leq \frac{2}{c_{p'}^2} (1 + c_\gamma^2 C^2)^4 \left(\sqrt{C} + C\sqrt{\|\Gamma_m\|_F}\right)^4 \gamma_n n^{3\frac{1-\alpha}{p'}} \mathbb{E} \left[\|\bar{m}_n - m\|^4 \right] + c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \\
&\leq \frac{2}{c_{p'}^2} K_2 (1 + c_\gamma^2 C^2)^4 \left(\sqrt{C} + C\sqrt{\|\Gamma_m\|_F}\right)^4 \gamma_n n^{3\frac{1-\alpha}{p'}} \frac{1}{n^2} + c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \\
&= c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] + O \left(\frac{1}{n^{2+\alpha-3(1-\alpha)/p'}} \right). \tag{61}
\end{aligned}$$

Step 3: Conclusion.

Applying inequalities (54), (60) and (61), there are a rank $n_{p'}$ and positive constants $c_{p'}, C_{1,p'}, C_{2,p'}, C_{3,p'}$ such that for all $n \geq n_{p'}$,

$$\mathbb{E} \left[\|V_{n+1} - \Gamma_m\|_F^4 \right] \leq \left(1 - c_{p'} \gamma_n n^{-\frac{1-\alpha}{p'}}\right) \mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] + \frac{C_{1,p'}}{n^{3\alpha}} + \frac{C_{2,p'}}{n^{2\alpha}} \mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] + \frac{C_{3,p'}}{n^{2+\alpha-3\frac{1-\alpha}{p'}}}.$$

□

E Some technical inequalities

First, the following lemma recalls some well-known inequalities.

Lemma E.1. *Let a, b, c be positive constants. Then,*

$$\begin{aligned}
ab &\leq \frac{a^2}{2c} + \frac{b^2 c}{2}, \\
a &\leq \frac{c}{2} + \frac{a^2}{2c}.
\end{aligned}$$

Moreover, let k, p be positive integers and a_1, \dots, a_p be positive constants. Then,

$$\left(\sum_{j=1}^p a_j \right)^k \leq p^{k-1} \sum_{j=1}^p a_j^k.$$

The following lemma gives the asymptotic behavior for some specific sequences of descent steps.

Lemma E.2. *Let α, β be non-negative constants such that $0 < \alpha < 1$, and $(u_n), (v_n)$ be two sequences defined for all $n \geq 1$ by*

$$u_n := \frac{c_u}{n^\alpha}, \quad v_n := \frac{c_v}{n^\beta},$$

with $c_u, c_v > 0$. Thus, there is a positive constant c_0 such that for all $n \geq 1$,

$$\sum_{k=1}^{E(n/2)} e^{-\sum_{j=k}^n u_j} u_k v_k = O \left(e^{-c_0 n^{1-\alpha}} \right), \tag{62}$$

$$\sum_{k=E(n/2)+1}^n e^{-\sum_{j=k}^n u_j} u_k v_k = O(v_n), \tag{63}$$

where $E(\cdot)$ is the integer part function.

Proof of Lemma E.2. We first prove inequality (62). For all $n \geq 1$,

$$\begin{aligned} \sum_{k=1}^{E(n/2)} e^{-\sum_{j=k}^n u_j} u_k v_k &= c_u c_v \sum_{k=1}^{E(n/2)} e^{-\sum_{j=k}^n u_j} \frac{1}{k^{\alpha+\beta}} \\ &\leq c_u c_v \sum_{k=1}^{E(n/2)} e^{-c_u \sum_{j=k}^n \frac{1}{j^\alpha}}. \end{aligned}$$

Moreover, for all $k \leq E(n/2)$,

$$\begin{aligned} c_u \sum_{j=k}^n \frac{1}{j^\alpha} &\geq c_u \frac{n}{2} \frac{1}{n^\alpha} \\ &\geq \frac{c_u}{2} n^{1-\alpha}. \end{aligned}$$

Thus,

$$\sum_{k=1}^{E(n/2)} e^{-\sum_{j=k}^n u_j} u_k v_k \leq c_u c_v n e^{-\frac{c_u}{2} n^{1-\alpha}}.$$

We now prove inequality (63). With the help of an integral test for convergence,

$$\begin{aligned} \sum_{j=k}^n u_j &= c_u \sum_{j=k}^n \frac{1}{j^\alpha} \\ &\geq c_u \int_k^{n+1} \frac{1}{t^\alpha} dt \\ &\geq \frac{c_u}{1-\alpha} \left((n+1)^{1-\alpha} - k^{-\alpha} \right). \end{aligned}$$

Thus,

$$\sum_{k=E(n/2)+1}^n e^{-\sum_{j=k}^n u_j} u_k v_k \leq c_u c_v e^{-(n+1)^{1-\alpha}} \sum_{k=E(n/2)+1}^n e^{k^{1-\alpha}} k^{-\alpha-\beta}$$

With the help of an integral test for convergence, there is a rank $n_{u,v}$ (for sake of simplicity, we consider that $n_{u,v} = 1$) such that for all $n \geq n_{u,v}$,

$$\begin{aligned} \sum_{k=E(n/2)+1}^n e^{k^{1-\alpha}} k^{-\alpha-\beta} &\leq \int_{E(n/2)+1}^{n+1} e^{t^{1-\alpha}} t^{-\alpha-\beta} dt \\ &\leq \frac{1}{1-\alpha} \left[e^{t^{1-\alpha}} t^{-\beta} \right]_{E(n/2)+1}^n + \beta \int_{E(n/2)+1}^n e^{t^{1-\alpha}} t^{-1-\beta} dt \\ &= e^{(n+1)^{1-\alpha}} (n+1)^{-\beta} + o \left(\int_{E(n/2)+1}^{n+1} e^{t^{1-\alpha}} t^{-\alpha-\beta} dt \right), \end{aligned}$$

since $\alpha < 1$. Thus,

$$\sum_{k=E(n/2)+1}^n e^{k^{1-\alpha}} k^{-\alpha-\beta} = O \left(e^{n^{1-\alpha}} n^{-\beta} \right).$$

As a conclusion, we have

$$\begin{aligned} \sum_{k=E(n/2)+1}^n e^{-\sum_{j=k}^n u_j} u_k v_k &= O \left(e^{-(n+1)^{1-\alpha} + n^{1-\alpha}} v_n \right) \\ &= O(v_n). \end{aligned}$$

References

- Bali, J.-L., Boente, G., Tyler, D.-E., and Wang, J.-L. (2011). Robust functional principal components: a projection-pursuit approach. *The Annals of Statistics*, 39:2852–2882.
- Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications.
- Cardot, H., Cénac, P., and Chaouch, M. (2010). Stochastic approximation to the multivariate and the functional median. In Lechevallier, Y. and Saporta, G., editors, *Compstat 2010*, pages 421–428. Physica Verlag, Springer.
- Cardot, H., Cénac, P., and Godichon-Baggioni, A. (2016). Online estimation of the geometric median in Hilbert spaces: non asymptotic confidence balls. *The Annals of Statistics (to appear)*.
- Cardot, H., Cénac, P., and Monnez, J.-M. (2012). A fast and recursive algorithm for clustering large datasets with k-medians. *Computational Statistics and Data Analysis*, 56:1434–1449.
- Cardot, H., Cénac, P., and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19:18–43.
- Cardot, H. and Degras, D. (2015). Online principal components analysis: which algorithm to choose ? Technical report, arXiv:1511.03688.
- Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics*, 42:1203–1231.
- Chaudhuri, P. (1992). Multivariate location estimation using extension of R -estimates through U -statistics type approach. *Ann. Statist.*, 20(2):897–916.
- Croux, C., Filzmoser, P., and Oliveira, M. (2007). Algorithms for projection-pursuit robust principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 87:218–225.
- Croux, C. and Ruiz-Gazen, A. (2005). High breakdown estimators for principal components: the projection-pursuit approach revisited. *J. Multivariate Anal.*, 95:206–226.
- Cupidon, J., Gilliam, D., Eubank, R., and Ruymgaart, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli*, 13:1179–1194.

- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for principal components analysis of a random vector function: some applications to statistical inference. *Journal of Multivariate Analysis*, 12:136–154.
- Devlin, S., Gnanadesikan, R., and Kettenring, J. (1981). Robust estimation of dispersion matrices and principal components. *J. Amer. Statist. Assoc.*, 76:354–362.
- Duflo, M. (1997). *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.
- Fritz, H., Filzmoser, P., and Croux, C. (2012). A comparison of algorithms for the multivariate L_1 -median. *Comput. Stat.*, 27:393–410.
- Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3):587–600.
- Godichon-Baggioni, A. (2016). Estimating the geometric median in Hilbert spaces with stochastic gradient algorithms; L^p and almost sure rates of convergence. *J. of Multivariate Analysis*, 146:209–222.
- Huber, P. and Ronchetti, E. (2009). *Robust Statistics*. John Wiley and Sons, second edition.
- Hubert, M., Rousseeuw, P., and Van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, 13:92–119.
- Hyndman, R. and Ullah, S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51:4942–4956.
- Jolliffe, I. (2002). *Principal Components Analysis*. Springer Verlag, New York, second edition.
- Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical data analysis based on the L_1 -norm and related methods (Neuchâtel, 1987)*, pages 217–230. North-Holland, Amsterdam.
- Kraus, D. and Panaretos, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99:813–832.
- Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., and Cohen, K. (1999). Robust principal components for functional data. *Test*, 8:1–73.
- Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Theory and methods.
- Mokkadem, A. and Pelletier, M. (2006). Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *Ann. Appl. Probab.*, 16(3):1671–1702.

- Möttönen, J., Nordhausen, K., and Oja, H. (2010). Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, volume 7, pages 182–193. IMS Collection.
- Pelletier, M. (2000). Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1):49–72.
- Polyak, B. and Juditsky, A. (1992). Acceleration of stochastic approximation. *SIAM J. Control and Optimization*, 30:838–855.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, second edition.
- Rousseeuw, P. and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223.
- Taskinen, S., Koch, I., and Oja, H. (2012). Robustifying principal components analysis with spatial sign vectors. *Statist. and Probability Letters*, 82:765–774.
- Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1 -median and associated data depth. *Proc. Natl. Acad. Sci. USA*, 97(4):1423–1426.
- Weiszfeld, E. (1937). On the point for which the sum of the distances to n given points is minimum. *Tohoku Math. J.*, 43:355–386.
- Weng, J., Zhang, Y., and Hwang, W.-S. (2003). Candid covariance-free incremental principal component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25:1034–1040.