Widening the scope of an eigenvector stochastic approximation process and application to streaming PCA and related methods

Jean-Marie Monnez, Abderrahman Skiredj

Widening the Scope of an Eigenvector Stochastic Approximation Process and Application to Streaming PCA and Related Methods

Jean-Marie Monnez*,a, Abderrahman Skiredj **,b

[a] *Université de Lorraine, CNRS, Inria*[1]*, IECL*[2]*, F-54000 Nancy, France & CHRU Nancy, INSERM, CIC-P*[3]*, F-54000 Nancy, France*
[b] *Université de Lorraine, CNRS, IECL, F-54000 Nancy, France & Ecole Nationale Supérieure des Mines de Nancy*[4]*, F-54000 Nancy, France*

Abstract

We prove the almost sure convergence of Oja-type processes to eigenvectors of the expectation B of a random matrix while relaxing the i.i.d. assumption on the observed random matrices (B_n) and assuming either (B_n) converges to B or $(E[B_n|T_n])$ converges to B where T_n is the sigma-field generated by the events before time n. As an application of this generalization, the online PCA of a random vector Z can be performed when there is a data stream of i.i.d. observations of Z, even when both the metric M used and the expectation of Z are unknown and estimated online. Moreover, in order to update the stochastic approximation process at each step, we are no longer bound to using only a mini-batch of observations of Z, but all previous observations up to the current step can be used without having to store them. This is useful not only when dealing with streaming data but also with Big Data as one can process the latter sequentially as a data stream. In addition the general framework of this process, unlike other algorithms in the literature, also covers the case of factorial methods related to PCA.

AMS 2020 subject classifications:
Primary: 62L20
Secondary: 60L10, 62H25

Keywords: Big Data, Data Stream, Eigenvectors, Online Estimation, Principal Component Analysis, Stochastic Algorithms, Stochastic Approximation.

1. Introduction

Streaming data are data that arrive continuously such as process control data, web data, telecommunication data, medical data, financial data, etc. In this setting, recursive stochastic algorithms can be used to estimate online, among others, parameters of a regression function (for example, see [12] and references therein) or centers of clusters in unsupervised classification [8] or principal components in a principal component analysis (PCA) (for example, see [13], p. 343). More precisely, each arriving observation vector is used to update the estimate sequence until it converges to the quantity of interest. When using such processes, it is not necessary to store the data and, due to the relative simplicity of the computation involved, much more data can be taken into account during the same duration of time than with non sequential methods. Moreover, this type of method uses less memory space than a batch method [1]. In this article, we propose a general framework of stochastic approximation processes and subsequently apply the latter to the case of streaming PCA. This general framework is sufficiently flexible to cover the case of streaming normed PCA and other related methods while allowing the stochastic algorithm to absorb a greater amount of data at each step or even all of the previously observed data up to the current step without any additional memory storage. This can be extremely efficient for dealing with large data streams.

[1]Inria, project-team BIGS
[2]Institut Elie Cartan de Lorraine, BP 239, F-54506 Vandœuvre-lès-Nancy, France
[3]Centre d'Investigation Clinique Plurithématique
[4]Campus Artem, CS 14234, 92 Rue Sergent Blandan, F-54042 Nancy, France
**Corresponding author. *jean-marie.monnez@univ-lorraine.fr **abderrahman.skiredj@mines-nancy.org

Let us define some notations. Let $A^\top$ denote the transpose of a matrix A. Let Q be a positive definite symmetric (p, p) matrix called metric, $\langle ., . \rangle$ and $\|.\|$ be respectively the inner product and the norm in $\mathbb{R}^p$ induced by Q: $\langle x, y \rangle = x^\top Q y$, $x^\top$ denoting the transpose of the column vector x. For vectors in $\mathbb{R}^p$, Q-orthogonal and Q-normed respectively mean orthogonal and normed with respect to the metric Q. Recall that a (p, p) matrix A is Q-symmetric if $(QA)^\top = QA$; then A has p real eigenvalues and there exists a Q-orthonormal basis of $\mathbb{R}^p$ consisting of eigenvectors of A. The norm of a matrix A is the spectral norm denoted $\|A\|$. The abbreviation a.s. stands for almost surely.

Numerous articles have been devoted to the problem of estimating eigenvectors and corresponding eigenvalues in decreasing order of the expectation B of a random symmetric (p, p) matrix, using an i.i.d. sample of the latter. These include, among others, the algorithms of Benzécri [2], Krasulina [16], Oja [19], Karhunen and Oja [15], Oja and Karhunen [20], Brandière [3–5], Brandière and Duflo [6] and Duflo [13] in the case of PCA. We consider here the commonly used normed process of Oja [15, 20] (X_n), whose rate of convergence is studied in [1], recursively defined by:

$$X_{n+1} = \frac{(I + a_n B_n) X_n}{\|(I + a_n B_n) X_n\|}, \tag{1}$$

the random matrices B_n being mutually independent and a.s. bounded, $E[B_n] = B$, $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$. A commonly used choice of a_n is $\frac{a}{n^\alpha}$, $\frac{1}{2} < \alpha \leq 1$ (for example, see [12]). This process converges a.s. to a normed eigenvector of B corresponding to its largest eigenvalue.

Consider the application presented in Section 3 to PCA of a random vector Z with unknown expectation $E[Z]$ and covariance matrix $C = E\left[(Z - E[Z])(Z - E[Z])^\top\right]$. Suppose there is a data stream $(Z_i, i \geqslant 1)$ of i.i.d. observations of Z and that the metric M used to define the distance between two observations of Z is unknown, for example the diagonal matrix of the inverses of variances of the components of Z in normed PCA. To estimate the principal components, we must estimate eigenvectors of the M^{-1}-symmetric matrix $B = MC$ (or of the symmetric matrix $M^{\frac{1}{2}} C M^{\frac{1}{2}}$, see Section 3). We can estimate online $E[Z]$ and M respectively at step n by $\overline{Z}_{n-1}$, the empirical mean of $(Z_i, i \leq n-1)$, and M_{n-1} depending on $(Z_i, i \leq n-1)$, and define $B_n = M_{n-1}\left(Z_n - \overline{Z}_{n-1}\right)\left(Z_n - \overline{Z}_{n-1}\right)^\top$. It is clear that the random matrices B_n do not satisfy the assumptions of Oja. In fact, these assumptions would be verified if the expectations and variances of the components of Z were known a priori in the case of normed PCA or other characteristics in other types of PCA. This is the case for example if we have at our disposal a massive data vectors set with computed characteristics and we randomly draw a data vector from this data set at each step ([13], p. 343). Here we suppose that data arrive continuously and are drawn from an unknown distribution. The convergence of the process in such a case is not proven.

The general convergence results are presented in Section 2, the application to streaming PCA and related methods in Section 3, the conclusion in Section 4, and the proofs of all theorems in Section 5.

Let Q be a metric in $\mathbb{R}^p$, B a Q-symmetric matrix and, for $n \geqslant 1$, T_n the σ-field generated by the events before time n; $X_1, B_1, \ldots, B_{n-1}$ are T_n-measurable. We prove the almost sure convergence of the process of Oja assuming that the conditional expectation of B_n with respect to σ-field T_n converges almost surely to B as n goes to infinity (Subsection 2.1, Theorem 1, first part). This allows proving the convergence of the process in the case of streaming PCA when a mini-batch of observations of Z is taken into account at each step (Subsection 3.1). A method of Duflo ([13], p. 343) is used in the proof (Section 5), but with more general assumptions. The proof of the convergence of processes $\left(X_n^i\right)$, $i \in \{1, \ldots, r\}$, $r \leq p$, of the same type, obtained by a Gram-Schmidt orthonormalization with respect to Q, to unit eigenvectors of B corresponding to the r largest eigenvalues in decreasing order (Subsection 2.2, Corollary 2, first part), not given in [13], is established in Section 5.

Moreover, we prove the almost sure convergence of the process of Oja, with an entirely different method, in the non-classical case where B_n converges almost surely to B (Subsection 2.1, Theorem 1, second part and Subsection 2.2, Corollary 2, second part; proofs in Section 5). This applies to PCA of a random vector Z while allowing the process to be updated at each step by using all previous observations (Z_i) up to the current step without the need to store the latter (Subsection 3.2). Hence, we define a type of processes different from the classical processes that used a mini-batch of observations at each step. The conducted experiments (see the Appendix), show that these processes are generally faster than the classical processes.

Finally, the scope of these processes is further widened to other factorial methods such as multiple factor analysis [21] or generalized canonical correlation analysis [10] (Subsection 3.3).

2. Theorem of almost sure convergence

2.1. Estimation of an eigenvector corresponding to the largest eigenvalue

We make the assumptions given below:

(H1a) B is Q-symmetric; let $\lambda_1, \lambda_2, \ldots, \lambda_p$, denote its eigenvalues in decreasing order and for $i \in \{1, \ldots, p\}$, V_i a $Q-$ normed eigenvector of B corresponding to λ_i;

(H1b) The largest eigenvalue λ_1 of B is simple;

(H2a) There exists a positive number b such that $sup_n \|B_n\| < b$ a.s.;

(H2b) B_n is not T_n-measurable, $\sum_1^\infty a_n E\left[\|E[B_n|T_n] - B\|\right] < \infty$; (H2b') $\sum_1^\infty a_n \|B_n - B\| < \infty$ a.s.;

(H2c) For all n, $I + a_n B_n$ is invertible;

(H3) $a_n > 0$, $\sum_1^\infty a_n = \infty$, $\sum_1^\infty a_n^2 < \infty$;

(H4) X_1 is a random variable independent of the sequence (B_n);

(H4') X_1 is an absolutely continuous random variable with respect to the Lebesgue measure and independent of the sequence (B_n).

Assumption H2b is obviously verified in the classical case $E[B_n|T_n] = B$ a.s. According to a classical lemma, given that $\sum_1^\infty a_n = \infty$, H2b implies that there exists a subsequence of $(E[B_n|T_n])$ converging a.s. to B. Likewise, H2b' implies that there exists a subsequence of (B_n) converging a.s. to B. H2c is verified in particular when the eigenvalues of B_n are non-negative. Assumption H3 is classical for Robbins-Monro type processes [12]. Note that in the second part of the theorem, since $\omega \in \Omega$ is fixed throughout the proof, a_n can be a positive random variable.

Let $U_n = \langle X_n, BX_n \rangle$ and $W_n = \langle X_n, B_n X_n \rangle$.

Theorem 1. (i) *Suppose assumptions H1a,b, H2a,b, H3, H4 hold. Then,* U_n *converges a.s. to one of the eigenvalues of B; on* $E_j = \left\{U_n \longrightarrow \lambda_j\right\}$, X_n *converges a.s. to* V_j *or* $-V_j$, $\sum_{n=1}^\infty a_n \left(\lambda_j - U_n\right)$ *and* $\sum_{n=1}^\infty a_n \left(\lambda_j - W_n\right)$ *converge a.s. Moreover, if* $\liminf E\left[\langle B_n X_n, V_1 \rangle^2 \ |T_n\right] > 0$ *a.s. on* $\cup_{j=2}^p E_j$, *then* $P(E_1) = 1$;

(ii) *Suppose assumptions H1a,b, H2b',c, H3, H4' hold. Then* X_n *converges a.s. to* V_1 *or* $-V_1$, $\sum_1^\infty a_n (\lambda_1 - U_n)$ *and* $\sum_1^\infty a_n |\lambda_1 - W_n|$ *converge a.s.*

Note that the methods of proofs of the two parts of Theorem 1 provided in Section 5 are entirely different. The first method is that used by Duflo ([13], p. 343) but with more general assumptions; the assumption "on $\cup_{j=2}^p E_j$, $\liminf E\left[\langle B_n X_n, V_1 \rangle^2 \ |T_n\right] > 0$ a.s." is used to avoid the traps (X_n converges to V_j, $j \neq 1$) of the stochastic algorithm; it is verified in the case of PCA of a random vector Z under assumptions H6b and H7b given in Corollary 3.

2.2. Simultaneous estimation of several eigenvectors

For $i \in \{1, \ldots, r\}$, $r \leq p$, recursively define the process $\left(X_n^i\right)$ in $\mathbb{R}^p$ such that:

$$Y_{n+1}^i = (I + a_n B_n) X_n^i, \; T_{n+1}^i = Y_{n+1}^i - \sum_{j<i} \langle Y_{n+1}^i, X_{n+1}^j \rangle X_{n+1}^j, \; X_{n+1}^i = \frac{T_{n+1}^i}{\|T_{n+1}^i\|}.$$

The vector $(X_{n+1}^1, \ldots, X_{n+1}^r)$ is obtained by Gram-Schmidt orthonormalization of $(Y_{n+1}^1, \ldots, Y_{n+1}^r)$. We make the assumptions given below:

(H1b') The eigenvalues of B are distinct;

(H5) For $i \in \{1, \ldots, r\}$, X_1^i is a random variable independent of the sequence (B_n);

(H5') For $i \in \{1, \ldots, r\}$, X_1^i is an absolutely continuous random variable with respect to the Lebesgue measure and independent of the sequence (B_n).

Corollary 2.

(i) *Suppose assumptions H1a,b', H2a,b, H3, H5 hold. Then, for* $i \in \{1, \ldots, r\}$, X_n^i *converges a.s. to one of the eigenvectors of B; moreover, if on* $\cup_{j=i+1}^p \left\{X_n^i \longrightarrow \pm V_j\right\}$, $\liminf E\left[\left\langle B_n X_n^i, V_i \right\rangle^2 \ |T_n\right] > 0$ *a.s., then* X_n^i *converges a.s. to* V_i *or* $-V_i$, $\sum_1^\infty a_n \left|\lambda_i - \left\langle X_n^i, BX_n^i \right\rangle\right|$ *and* $\sum_1^\infty a_n (\lambda_i - \left\langle X_n^i, B_n X_n^i \right\rangle)$ *converge a.s.;*

(ii) *Suppose assumptions H1a,b', H2b',c, H3, H5' hold. Then, for* $i \in \{1, \ldots, r\}$, X_n^i *converges a.s. to* V_i *or* $-V_i$, $\sum_1^\infty a_n \left|\lambda_i - \left\langle X_n^i, BX_n^i \right\rangle\right|$ *and* $\sum_1^\infty a_n \left|\lambda_i - \left\langle X_n^i, B_n X_n^i \right\rangle\right|$ *converge a.s.*

The proofs of the two parts of Corollary 2 are provided in Section 5.

3. Application to streaming PCA and related methods

In this section, we apply the theoretical results given in Section 2 to the estimation of the principal components of a streaming PCA. Classical convergence results exist for the process (1) or for processes of the same type. In these processes, a mini-batch of data is used at each step and it is assumed that the random matrices B_n are independent. However, in practice, the metric M used to define a distance between two observations is a priori unknown and must be estimated at each step using the available data. As a result, the matrices B_n are no longer independent. Corollary 3 in Subsection 3.1 deals with this issue and thus broadens the scope of these processes. In Subsection 3.2, we define a type of processes that can be updated at each step with all previously observed data without the need for their storage. The conducted experiments show that this type of processes performs better than the former.

PCA is essential for data compression and feature extraction and has applications in various fields, such as data mining, engineering [23], face recognition [9], astronomy [7], text analytics [11, 17], etc. Batch PCA is unfeasible with massive datasets or data streams although online PCA algorithms which provide very fast updates can be used in this context [9]. Let $Z^1, \ldots, Z^p$ be the components of a random vector Z and C its covariance matrix. Define a metric M in $\mathbb{R}^p$ and consider the following problem: for $l \in \{1, \ldots, r\}, r \leq p$, determine at step l a linear combination of all the centered components of Z, $U_l = c_l^\top (Z - E[Z])$, named l^{th} principal component, uncorrelated with $U_1, \ldots, U_{l-1}$ and of maximum variance under the constraint $c_l^\top M^{-1} c_l = 1$. It is proven that c_l is an eigenvector of the matrix $B = MC$ corresponding to its l^{th} largest eigenvalue λ_l. For $l \in \{1, \ldots, r\}$, a M-normed direction vector u_l of the l^{th} principal axis is defined as $M^{-1} c_l$; the vectors u_l are M-orthonormal and are eigenvectors of the matrix CM corresponding respectively to the same eigenvalues λ_l. A particular case is normed PCA, where M is the diagonal matrix of the inverses of variances of the p components of Z; this is equivalent to using standardized data, i.e. observations of $M^{\frac{1}{2}} (Z - E[Z])$, and the identity metric, hence the same importance is given to each component of Z. However, in the case of streaming data, the expectation and the variance of the components of Z are unknown. An application of this work is to recursively estimate the c_l or the u_l in this setting using a stochastic approximation process. Another example where the expectation of Z and the metric used are unknown in the context of streaming data is generalized canonical correlation analysis (gCCA) [10, 18], which can be interpreted as a PCA where each pre-defined group of components of Z has the same importance (see Section 5).

In order to reduce computing time and to avoid possible numerical explosions, we propose herein:

(a) to estimate the eigenvectors a_l of the symmetric (p, p) matrix $B = M^{\frac{1}{2}} C M^{\frac{1}{2}}$ (thus the Gram-Schmidt orthonormalization is made with respect to $Q = I$) and then, using an estimation of $M^{\frac{1}{2}}$ or $M^{-\frac{1}{2}}$, to deduce estimations of the eigenvectors c_l of MC or of the eigenvectors u_l of CM, since $c_l = M^{\frac{1}{2}} a_l$ and $u_l = M^{-\frac{1}{2}} a_l$; note that if M is the diagonal matrix of the inverses of variances of the p components of Z, B is the correlation matrix of Z;

(b) to replace Z by $Z^c = Z - \xi$, ξ being an estimation of $E[Z]$ computed in a preliminary phase with a small number of observations e.g. 1000, in order to reduce possible high values of certain components of Z and to possibly avoid a numerical explosion; then, $B = M^{\frac{1}{2}} \left(E\left[Z^c (Z^c)^\top\right] - E[Z^c] E[Z^c]^\top\right) M^{\frac{1}{2}}$;

(c) to use, instead of a mini-batch of observations of Z at step n, all observations up to this step without having to store them; thus information contained in previous data is used at each step; another algorithm with the same goal is History PCA [24].

Let $(Z_{11}, \ldots, Z_{1m_1}, \ldots, Z_{n1}, \ldots, Z_{nm_n}, \ldots)$ be an i.i.d. sample of Z. The variables Z_{nj}, $j \in \{1, \ldots, m_n\}$, are observed at time n. Let $\overline{Z}_{n-1}$ be the mean of the sample $(Z_{11}, \ldots, Z_{n-1,m_{n-1}})$ of Z and M_{n-1} a T_n-measurable estimation of M obtained from this sample. Let $Z_{ni}^c = Z_{ni} - \xi$ and $\overline{Z}_{n-1}^c = \overline{Z}_{n-1} - \xi$.

Given (B_n), we recursively define the processes $\left(X_n^i\right)$, $i \in \{1, \ldots, r\}$, by:

$$Y_{n+1}^i = (I + a_n B_n) X_n^i, \; T_{n+1}^i = Y_{n+1}^i - \sum_{j<i} \langle Y_{n+1}^i, X_{n+1}^j \rangle X_{n+1}^j, \; X_{n+1}^i = \frac{T_{n+1}^i}{\|T_{n+1}^i\|}.$$

As mentioned in (a), the Gram-Schmidt orthonormalization is made with respect to the metric I.

3.1. Use of a data mini-batch at each step

Let $B_n = M_{n-1}^{\frac{1}{2}} \left(\frac{1}{m_n} \sum_{j=1}^{m_n} Z_{nj}^c \left(Z_{nj}^c\right)^\top - \overline{Z}_{n-1}^c \left(\overline{Z}_{n-1}^c\right)^\top\right) M_{n-1}^{\frac{1}{2}}$. Then the conditional expectation

$$E[B_n \mid T_n] = M_{n-1}^{\frac{1}{2}} \left(E\left[Z^c (Z^c)^\top\right] - \overline{Z}_{n-1}^c \left(\overline{Z}_{n-1}^c\right)^\top\right) M_{n-1}^{\frac{1}{2}}$$

is different from B but converges a.s. to B under assumptions H6a and H7b given below:

(H3') $a_n > 0, \sum_1^\infty a_n = \infty$, $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$, $\sum_1^\infty a_n^2 < \infty$;
(H6a) $\|Z\|$ is a.s. bounded;
(H6b) There is no affine or quadratic relationship between the components of Z;
(H7a) There exists a positive number d such that $\sup_n \left\| M_n^{\frac{1}{2}} \right\| < d$ a.s.;
(H7b) $M_n^{\frac{1}{2}} \longrightarrow M^{\frac{1}{2}}$ a.s.;
(H7c) $\sum_1^\infty a_n E\left[\left\| M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}} \right\|\right] < \infty$.

Corollary 3. *Suppose assumptions H1b', H3', H5, H6a,b and H7a,b,c hold. Then, for $i \in \{1, \ldots, r\}$, X_n^i converges a.s. to V_i or $-V_i$, $\sum_{n=1}^\infty a_n \left| \lambda_i - \left(X_n^i\right)' B X_n^i \right|$ and $\sum_{n=1}^\infty a_n \left(\lambda_i - \left(X_n^i\right)' B_n X_n^i\right)$ converge a.s.*

The proof is provided in Section 5.

3.2. Use of all observations up to the current step

Let $B_n = M_n^{\frac{1}{2}} \left(\frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^c \left(Z_{ij}^c\right)^\top - \overline{Z}_n^c \left(\overline{Z}_n^c\right)^\top \right) M_n^{\frac{1}{2}}$. Then the conditional expectation $E[B_n | T_n]$ is different from B, but B_n converges a.s. to B under assumptions H7b and H6c below:
(H6c) Z has 4^{th} order moments;
(H7d) $\sum_1^\infty a_n \left\| M_n^{\frac{1}{2}} - M^{\frac{1}{2}} \right\| < \infty$ a.s.

Corollary 4. *Suppose assumptions H1b', H3', H5', H6c, H7b,d hold. Then, for $i \in \{1, \ldots, r\}$, X_n^i converges a.s. to V_i or $-V_i$, $\sum_1^\infty a_n \left| \lambda_i - \left\langle X_n^i, B X_n^i \right\rangle \right|$ and $\sum_1^\infty a_n \left| \lambda_i - \left\langle X_n^i, B_n X_n^i \right\rangle \right|$ converge a.s.*

The proof is provided in Section 5.

Let M be the diagonal matrix of the inverses of variances of $Z^1, \ldots, Z^p$. Let for $j \in \{1, \ldots, p\}$, V_n^j be the empirical variance of the sample $(Z_{11}^j, \ldots, Z_{n,m_n}^j)$ of Z^j and M_n the diagonal matrix of order p whose element (j, j) is the inverse of $\frac{\mu_n}{\mu_n - 1} V_n^j$, $\mu_n = \sum_1^n m_i$. Under H6c, H7b holds. Moreover, it is established in ([12], Lemma 5) that H7d holds under H6c and H3'.

3.3. Widening to related methods

The calculation of $M_{n-1}^{\frac{1}{2}}$ is straightforward when M_{n-1} as M are diagonal matrices. This is often the case in PCA and related methods such as normed PCA, Multiple Correspondence Analysis (MCA) [14] for categorical variables, Factor Analysis of Mixed Data (FAMD) [21] for continuous or categorical variables, and Multiple Factor Analysis (MFA) [21] for groups of continuous or categorical variables. However, in generalized Canonical Correlation Analysis (gCCA) [10], the metric M is block diagonal (see Section 5). We can define B_n in four different ways,

$$B_n = M_{n-1} \left(\frac{1}{m_n} \sum_{j=1}^{m_n} Z_{nj}^c \left(Z_{nj}^c\right)^\top - \overline{Z}_{n-1}^c \left(\overline{Z}_{n-1}^c\right)^\top \right), B_n = M_n \left(\frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^c \left(Z_{ij}^c\right)^\top - \overline{Z}_n^c \left(\overline{Z}_n^c\right)^\top \right) \quad (2)$$

and achieve the orthogonalization in processes $\left(X_n^i\right)$ with respect to M_{n-1}^{-1} in the first case or M_n^{-1} in the second case to estimate directly the eigenvectors c_i of the M^{-1}- symmetric matrix $B = MC$, or

$$B_n = \left(\frac{1}{m_n} \sum_{j=1}^{m_n} Z_{nj}^c \left(Z_{nj}^c\right)^\top - \overline{Z}_{n-1}^c \left(\overline{Z}_{n-1}^c\right)^\top \right) M_{n-1}, B_n = \left(\frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^c \left(Z_{ij}^c\right)^\top - \overline{Z}_n^c \left(\overline{Z}_n^c\right)^\top \right) M_n \quad (3)$$

and achieve the orthogonalization in processes $\left(X_n^i\right)$ with respect to M_{n-1} in the third case or M_n in the fourth case to estimate directly the eigenvectors u_i of the M-symmetric matrix $B = CM$, $M_n^{\frac{1}{2}}$ being replaced by M_n and $M^{\frac{1}{2}}$ by M in assumptions H7a,b,c,d. This requires an extension of Theorem 1 and Corollary 2, the metric Q (M^{-1} or M) being replaced at step n by an estimated T_n-measurable metric Q_n such that Q_n converges a.s. to Q and $\sum_{n=1}^\infty a_n \|Q_n - Q\| < \infty$ a.s. This will be treated in a forthcoming article.

4. Conclusion

In this article, we have provided almost sure convergence theorems of an Oja-type stochastic approximation process to eigenvectors of a Q-symmetric matrix B corresponding to eigenvalues in decreasing order. These theorems apply to cases where there is a sequence (B_n) such that $E[B_n|T_n]$ or B_n converges a.s. to B. This extends previous results where random matrices B_n are assumed to be i.i.d. and $E[B_n|T_n] = B$.

We subsequently applied these results to the online estimation of principal components in streaming PCA. By the streaming nature of the data, the expectation, the variance of the variables (or other characteristics if necessary) and the metric are generally unknown and need to be estimated online along with the principal components. Classical results of convergence do not apply to this case.

Moreover, we have defined a type of processes which are updated at each step by all of the previously observed data without necessitating their storage in memory. Classical algorithms only use a data mini-batch at each step.

In addition, these results can be applied to other factorial methods such as online gCCA.

Experiments have been conducted to compare processes of the second type (A: using all observations up to the current step) to processes of the first type (B: using a mini-batch of observations) on simulated or real datasets. Results regarding computing time are provided in the Appendix. It appears that processes A are generally faster than processes B. Such behavior is not surprising since, intuitively, a greater quantity of data at each step enables the algorithm to converge faster.

5. Proofs

Generalized Canonical Correlation Analysis. Suppose the set of components of a random vector Z in $\mathbb{R}^p$ is partitioned in q sets of real random variables $\left\{Z^{k1}, \ldots, Z^{km_k}\right\}$, $k \in \{1, \ldots, q\}$. Let Z^k be the random vector in $\mathbb{R}^{m_k}$ whose components are Z^{kj}, $j \in \{1, \ldots, m_k\}$. Assume there is no affine relationship between the components of Z. Let C^k and C be the covariance matrices of Z^k and Z, respectively. Consider the following problem: for $l \in \{1, \ldots, r\}, r \leq p$, determine at step l a linear combination of all the centered components of Z, $U_l = \theta_l^\top (Z - E[Z])$, named l^{th} general component, of variance 1 and uncorrelated with $U_1, \ldots, U_{l-1}$, and, for $k \in \{1, \ldots, q\}$, a linear combination of variance 1 of the centered components of Z^k, $V_l^k = \left(\eta_l^k\right)^\top \left(Z^k - E\left[Z^k\right]\right)$, named l^{th} canonical component of the k^{th} set of variables, which maximize $\sum_{k=1}^q \rho^2\left(U_l, V_l^k\right)$, ρ denoting the linear correlation coefficient. Let M be the unknown block diagonal matrix of order p whose k^{th} diagonal block is $\left(C^k\right)^{-1}$. Let $\theta_l = \left(\left(\theta_l^1\right)^\top, \ldots, \left(\theta_l^q\right)^\top\right)^\top, \theta_l^k \in \mathbb{R}^{m_k}, k \in \{1, \ldots, q\}$. It is proven that θ_l is a C-normed eigenvector of the M^{-1}-symmetric matrix $B = MC$ corresponding to its l^{th} largest eigenvalue and that for $k \in \{1, \ldots, q\}$, there exists $\alpha_l^k \in \mathbb{R}$ such that $\eta_l^k = \alpha_l^k \theta_l^k$. Note that θ_l is collinear with the l^{th} principal component of PCA of Z with the metric M. The objective is then to perform online PCA of Z using at step n a consistent estimator M_n of M.

Proof of Theorem 1, first part. Its plan follows that of ([13], Section 9.4.2, p. 343) in the case of PCA with $E[B_n|T_n] = B$ a.s. Let $X_n^j = \langle X_n, V_j \rangle$. After establishing that there exists $K_2 > 0$ and β_n such that a.s.

$$X_{n+1} = (I + a_n (B_n - W_n I) + a_n \beta_n) X_n, \|\beta_n\| \leq K_2 a_n,$$

we prove that U_n converges a.s., then that if the limit of U_n is different from an eigenvalue λ_j, X_n^j converges a.s. to 0. Since $\|X_n\| = 1$, this cannot be true for every j, therefore the limit of U_n is one of the eigenvalues of B, λ_i, and X_n converges to V_i or $-V_i$ on $E_i = \{U_n \longrightarrow \lambda_i\}$. We then prove that $\sum_{n=1}^\infty a_n(\lambda_i - U_n)$ and $\sum_{n=1}^\infty a_n(\lambda_i - W_n)$ converge on E_i. Applying a lemma of Duflo ([13], p. 342) to the sequence $\left(X_n^1\right)$ on E_i, we prove that, if $\liminf E[\langle B_n X_n, V_1 \rangle^2 |T_n] > 0$, $P(E_i) = 0$ for $i > 1$. Therefore, X_n converges a.s. to V_1.

Let us state two lemmas of Duflo ([13], p. 16, 342) used in this proof.

Lemma 5. *Let (M_n) be a square-integrable martingale adapted to a filtration (T_n) and $(\langle M \rangle_n)$ its increasing process defined recursively by:*

$$\langle M \rangle_1 = M_1^2, \langle M \rangle_{n+1} = \langle M \rangle_n + E[(M_{n+1} - M_n)^2 |T_n] = \langle M \rangle_n + E[M_{n+1}^2 |T_n] - M_n^2.$$

Let $\langle M \rangle_\infty = \lim \langle M \rangle_n$. If $E[\langle M \rangle_\infty] < \infty$, then (M_n) converges a.s. and in mean square to a finite random variable.

In the following lemma, Duflo provides a tool which can be used to prove that given a sufficiently exciting noise (ϵ_n), the traps of recursive algorithms of the type defined in assumption (i) below can be avoided.

Lemma 6. *Let (γ_n) be a sequence of positive numbers such that $\sum_{n=1}^{\infty}\gamma_n^2 < \infty$. Let (Z_n) and (δ_n) be two sequences of random variables adapted to a filtration (T_n), and (ϵ_n) a noise adapted to (T_n).*
Assume on a set Γ:
(i) *for every integer n, $Z_{n+1} = Z_n(1+\delta_n) + \gamma_n\epsilon_{n+1}$;*
(ii) *(Z_n) is bounded;*
(iii) *$\sum_{n=1}^{\infty}\delta_n^2 < \infty$, $\delta_n \geq 0$ for n sufficiently large, and there exists a sequence of positive numbers (b_n) such that $\sum_{n=1}^{\infty} b_n = \infty$ and $\sum_{n=1}^{\infty}(b_n - \delta_n)$ converges;*
(iv) *for an $a > 2$, $E[|\epsilon_{n+1}|^a|T_n] = O(1)$ and $\liminf E[\epsilon_{n+1}^2|T_n] > 0$ a.s.*
Then, $P(\Gamma) = 0$.

Note that, since $\sup_n \|B_n\| < b$ a.s. (H2a) and a_n converges to 0 (H3), $I + a_nB_n$ is invertible from a certain rank N, and if X_N is different from 0, $\|(I + a_nB_N)X_N\| \neq 0$, thus X_n is defined for all $n \geqslant N$ and $\|X_n\| = 1$. We have $\|U_n\| \leq \|B\|, \|W_n\| \leq \|B_n\| \leq b$ a.s. Let $K_i, i = 1,2,3,4,5$ be adequately chosen real numbers. Under H2a, since $\|(I + a_nB_n)X_n\|^2 = 1 + 2a_nW_n + a_n^2\|B_nX_n\|^2$, we have almost surely:

$$\begin{aligned}
\frac{1}{\|(I+a_nB_n)X_n\|} &= 1 - a_nW_n - \frac{1}{2}a_n^2\|B_nX_n\|^2 + \alpha_n, \quad |\alpha_n| \leq K_1a_n^2, \\
X_{n+1} &= (I+a_nB_n)\left(1 - a_nW_n - \frac{1}{2}a_n^2\|B_nX_n\|^2 + \alpha_n\right)X_n = (I + a_n(B_n - W_nI) + a_n\beta_n)X_n, \\
\beta_n &= -a_nW_nB_n - \frac{1}{2}a_n\|B_nX_n\|^2 I - \frac{1}{2}a_n^2B_n\|B_nX_n\|^2 + a_n^{-1}\alpha_nI + \alpha_nB_n, \ \|\beta_n\| \leq K_2a_n, \\
X_{n+1} &= (I + a_n(B - U_nI) + a_n\Gamma_n)X_n, \Gamma_n = (B_n - B) - \langle X_n, (B_n - B)X_n\rangle I + \beta_n, \ \|\Gamma_n\| \leq K_3.
\end{aligned}$$

Step 1: convergence of $\mathbf{U}_n$.

$$\begin{aligned}
U_{n+1} &= \langle (I + a_n(B - U_nI) + a_n\Gamma_n)X_n \ , \ B(I + a_n(B - U_nI) + a_n\Gamma_n)X_n\rangle \\
&= U_n + 2a_n\langle (B - U_nI)X_n \ , \ BX_n\rangle + 2a_n\langle \Gamma_nX_n \ , \ BX_n\rangle + a_n^2\eta_n, \\
\eta_n &= \langle (B - U_nI)X_n \ , \ B(B - U_nI)X_n\rangle + 2\langle \Gamma_nX_n \ , \ B(B - U_nI)X_n\rangle + \langle \Gamma_nX_n \ , \ B\Gamma_nX_n\rangle, \ |\eta_n| \leq K_4 \text{ a.s.}
\end{aligned}$$

Let $\mu_n = 2a_n\langle E[\Gamma_n|T_n]X_n, BX_n\rangle + a_n^2E[\eta_n|T_n]$. Since $\|X_n\| = 1$:

$$\langle (B - U_nI)X_n \ , \ BX_n\rangle = \|BX_n\|^2 - U_n^2 = \|BX_n - U_nX_n\|^2 \geqslant 0.$$

Then, $E[U_{n+1}|T_n] \geq U_n + \mu_n$ and $E[U_{n+1} - \sum_{i=1}^{n}\mu_i|T_n] \geq U_n - \sum_{i=1}^{n-1}\mu_i$ a.s., thus $U_n - \sum_{i=1}^{n-1}\mu_i$ is a submartingale. Since $\|\beta_n\| \leq K_2a_n$ and $|\eta_n| \leq K_4$:

$$\begin{aligned}
\|E[\Gamma_n|T_n]\| &\leq 2\|E[B_n|T_n] - B\| + \left\|E[\beta_n|T_n]\right\|; \\
|\mu_n| &\leq 4a_n\|E[B_n|T_n] - B\| + 2a_n\left\|E[\beta_n|T_n]\right\| + K_4a_n^2 \leq 4a_n\|E[B_n|T_n] - B\| + K_5a_n^2 \text{ a.s.}
\end{aligned}$$

By H2b ($\sum_{n=1}^{\infty}a_nE[\|E[B_n|T_n] - B\|] < \infty$) and H3, we have: $\sup_n E\left[\left|\sum_{i=1}^{n-1}\mu_i\right|\right] < \infty$. By Doob's lemma, the submartingale $U_n - \sum_{i=1}^{n-1}\mu_i$ converges a.s. to an integrable random variable. Since $\sum_{i=1}^{n-1}\mu_i$ converges, U_n converges a.s.

Step 2: convergence of $\mathbf{X}_n^j = \langle \mathbf{X}_n, \mathbf{V}_j\rangle$. Let $\Gamma_n^j = \left\langle \Gamma_nX_n, V_j\right\rangle$. Since B is Q-symmetric, $\left\langle BX_n, V_j\right\rangle = \lambda_jX_n^j$.

$$\begin{aligned}
X_{n+1}^j &= \left\langle (I + a_n(B - U_nI) + a_n\Gamma_n)X_n \ , \ V_j\right\rangle = X_n^j + a_n\left((\lambda_j - U_n)X_n^j + \Gamma_n^j\right); \\
(X_{n+1}^j)^2 &= (X_n^j)^2 + a_n^2\left((\lambda_j - U_n)X_n^j + \Gamma_n^j\right)^2 + 2a_n(\lambda_j - U_n)(X_n^j)^2 + 2a_nX_n^j\Gamma_n^j \\
&= (X_1^j)^2 + \sum_{l=1}^{n}a_l^2\left((\lambda_j - U_l)X_l^j + \Gamma_l^j\right)^2 + 2\sum_{l=1}^{n}a_lX_l^j\Gamma_l^j + 2\sum_{l=1}^{n}a_l(\lambda_j - U_l)(X_l^j)^2.
\end{aligned}$$

Prove the convergence of the three last terms of this decomposition.
(i) Since $\|U_l\| \leq \|B\|$ and $\|\Gamma_l\| \leq K_3$ a.s., $\sum_{l=1}^{\infty} a_l^2 \left((\lambda_j - U_l)X_l^j + \Gamma_l^j\right)^2$ converges a.s. by H3.
(ii) Let $M_n^j = \sum_{l=1}^{n-1} a_l X_l^j (\Gamma_l^j - E[\Gamma_l^j | T_l])$.
$\sum_{l=1}^{n} a_l X_l^j \Gamma_l^j = \sum_{l=1}^{n} a_l X_l^j (\Gamma_l^j - E[\Gamma_l^j|T_l]) + \sum_{l=1}^{n} a_l X_l^j E[\Gamma_l^j|T_l] = \sum_{l=1}^{n} a_l X_l^j E[\Gamma_l^j|T_l] + M_{n+1}^j$.
Firstly: $\sum_{l=1}^{n} a_l |X_l^j E[\Gamma_l^j|T_l]| \leq \sum_{l=1}^{n} a_l \|E[\Gamma_l|T_l]\| \leq \sum_{l=1}^{n} a_l \left(2\|E[B_l|T_l] - B\| + \|E[\beta_l|T_l]\|\right)$.
By H2b ($\sum_{n=1}^{\infty} a_n E\left[\|E\left[B_n | T_n\right] - B\|\right] < \infty$) and H3, since $\|\beta_n\| \leq K_2 a_n$ a.s., $\sum_{l=1}^{\infty} a_l X_l^j E[\Gamma_l^j|T_l]$ converges a.s.
Secondly: (M_n^j) is a square-integrable martingale adapted to the filtration (T_n); let $(\langle M^j \rangle_n)$ be its increasing process.
$\langle M^j \rangle_{n+1} - \langle M^j \rangle_n = E[(M_{n+1}^j - M_n^j)^2|T_n] = a_n^2 E[(X_n^j)^2(\Gamma_n^j - E[\Gamma_n^j|T_n])^2|T_n] \leq K_3 a_n^2$.
Then, applying Lemma 5, since $E\left[\langle M \rangle_\infty\right] < \infty$ by H3, (M_n^j) converges a.s. to a finite random variable. Thus, $\sum_{l=1}^{\infty} a_l X_l^j \Gamma_l^j$ converges a.s.
(iii) Let ω be fixed belonging to the convergence set of U_n. The writing of ω will be omitted in the following. Let L be the limit of U_n. If $L \neq \lambda_j$, the sign of $\lambda_j - U_n$ is constant from a certain rank N depending on ω. Using the decomposition of $(X_{n+1}^j)^2$, there exists $A > 0$ such that:

$$\sum_{l=N}^{n} a_l |\lambda_j - U_l| (X_l^j)^2 = \left|\sum_{l=N}^{n} a_l (\lambda_j - U_l)(X_l^j)^2\right| = \left|(X_{n+1}^j)^2 - (X_N^j)^2 - \sum_{l=N}^{n} a_l^2 \left((\lambda_j - U_l)X_l^j + \Gamma_l^j\right)^2 - 2\sum_{l=N}^{n} a_l X_l^j \Gamma_l^j\right| < A.$$

Then for $L \neq \lambda_j$, $\sum_{l=N}^{\infty} a_l |\lambda_j - U_l| (X_l^j)^2$ and $\sum_{l=N}^{\infty} a_l \left(\lambda_j - U_l\right) (X_l^j)^2$ converge.
It follows from (i), (ii) and (iii) that for $L \neq \lambda_j$, $(X_n^j)^2$ converges a.s. Since by (iii), $\sum_{l=1}^{\infty} a_l (X_l^j)^2 < \infty$, X_n^j converges a.s. to 0. Since $\|X_n\| = 1$, this cannot be true for every j. Thus the limit of U_n is one of the eigenvalues of B, λ_i.

Step 3: convergence of $\mathbf{X}_n$. On $E_i = \{U_n \longrightarrow \lambda_i\}$, for $j \neq i$, X_n^j converges to 0, therefore $(X_n^i)^2$ converges to 1 and since

$$X_{n+1} - X_n = a_n \left((B - U_n I) + \Gamma_n\right) X_n,$$

$X_{n+1} - X_n$ converges to 0 and the limit of X_n is V_i or $-V_i$ on E_i (first assertion of Theorem 1).

Using the decomposition of $(X_n^i)^2$ in Step 2, the convergence of $(X_n^i)^2$ and of (i) and (ii) yields that $\sum_{n=1}^{\infty} a_n(\lambda_i - U_n)$ converges a.s. on E_i (first assertion of Theorem 1). Consider now the decomposition:

$$\sum_{n=1}^{\infty} a_n(\lambda_i - W_n) = \sum_{n=1}^{\infty} a_n(\lambda_i - U_n) + \sum_{n=1}^{\infty} a_n \left\langle X_n, (B_n - E[B_n|T_n])X_n \right\rangle + \sum_{n=1}^{\infty} a_n \left\langle X_n, (E[B_n|T_n] - B)X_n \right\rangle.$$

Firstly, since $\sum_{n=1}^{\infty} a_n E\left[\|E[B_n|T_n] - B\|\right] < \infty$ (H2b), $\sum_{n=1}^{\infty} a_n \langle X_n, (E[B_n|T_n] - B)X_n \rangle$ converges a.s.
Secondly, let $M_n = \sum_{l=1}^{n-1} a_l \langle X_l, (B_l - E[B_l|T_l])X_l \rangle$; (M_n) is a square-integrable martingale adapted to the filtration (T_n). Its increasing process $(\langle M \rangle_n)$ converges: indeed, since $\sup_n \|B_n\| < b$ a.s.,

$$\begin{aligned} \langle M \rangle_{n+1} - \langle M \rangle_n &= E[(M_{n+1} - M_n)^2|T_n] = a_n^2 E[\langle X_n, (B_n - E[B_n|T_n])X_n \rangle^2 |T_n] \\ &\leq a_n^2 E[\|B_n - E[B_n|T_n]\|^2|T_n] \leq a_n^2 E[\|B_n\|^2|T_n] \leq b^2 a_n^2. \end{aligned}$$

Thus by H3 and Lemma 5, (M_n) converges a.s. to a finite random variable. Therefore, $\sum_{n=1}^{\infty} a_n(\lambda_i - W_n)$ converges a.s. on E_i (first assertion of Theorem 1).

Step 4: convergence of $\mathbf{X}_n$ to $\pm\mathbf{V}_1$. Suppose $i > 1$. Recall that, with $\Gamma_n = (B_n - B) - \langle X_n, (B_n - B)X_n \rangle I + \beta_n$:

$$X_{n+1}^1 = (1 + a_n(\lambda_1 - U_n)) X_n^1 + a_n \langle \Gamma_n X_n, V_1 \rangle = (1 + a_n((\lambda_1 - \lambda_i) + (\lambda_i - U_n))) X_n^1 + a_n \langle \Gamma_n X_n, V_1 \rangle.$$

In the following, apply Lemma 6 to the sequence $\left(X_n^1\right)$ on $E_i = \{X_n \longrightarrow V_i\}, i > 1$, with $\gamma_n = a_n$, $\delta_n = a_n(\lambda_1 - U_n)$, $b_n = a_n(\lambda_1 - \lambda_i) > 0$, $\epsilon_{n+1} = \langle \Gamma_n X_n, V_1 \rangle$. Verify the assumptions (ii), (iii) and (iv) of Lemma 6:
(ii) X_n^1 is bounded;
(iii) $\sum_{n=1}^{\infty} a_n^2 < \infty$, $\sum_{n=1}^{\infty} a_n^2(\lambda_1 - U_n)^2 < \infty$, $\sum_{n=1}^{\infty} a_n(\lambda_1 - \lambda_i) = \infty$, $\sum_{n=1}^{\infty} a_n(\lambda_i - U_n)$ converges a.s. on E_i;
(iv) $E\left[\langle \Gamma_n X_n, V_1 \rangle^2 \; |T_n\right] \geqslant \frac{1}{2} E\left[\langle B_n X_n, V_1 \rangle^2 \; |T_n\right] - E\left[\langle (\Gamma_n - B_n) X_n, V_1 \rangle^2 \; |T_n\right]$ a.s.

$$\begin{aligned} \langle (\Gamma_n - B_n) X_n, V_1 \rangle &= -\langle BX_n, V_1 \rangle - \langle (B_n - B)X_n, X_n \rangle \langle X_n, V_1 \rangle + \langle \beta_n X_n, V_1 \rangle \\ &= -X_n^1(\lambda_1 + \langle (B_n - B)X_n, X_n \rangle) + \langle \beta_n X_n, V_1 \rangle. \end{aligned}$$

Since $\|\Gamma_n\| \leq K_3$ and $\|\beta_n\| \leq K_2 a_n$ a.s., there exists a positive number c such that a.s. on E_i:
$E\left[\langle(\Gamma_n - B_n)X_n, V_1\rangle^2 \ |T_n\right] \leq c(X_n^1)^2 + 2E[\|\beta_n\|^2 \ |T_n] \leq c(X_n^1)^2 + 2K_2^2 a_n^2 \underset{n\to+\infty}{\longrightarrow} 0$ since $X_n^1 \underset{n\to+\infty}{\longrightarrow} 0$ on E_i.
Then, if $\liminf E[\langle B_n X_n, V_1\rangle^2 \ |T_n] > 0$, $\liminf E[\langle \Gamma_n X_n, V_1\rangle^2 \ |T_n] > 0$. By Lemma 6, $P(E_i) = 0$, $i > 1$.
Thus $P(E_1) = 1$ (second assertion of Theorem 1). □

Proof of Theorem 1, second part. Recursively define the processes $(\widetilde{X}_n)$ and $\left(\widetilde{U}_n\right)$ such that

$$\begin{aligned}
\widetilde{X}_{n+1} &= (I + a_n B_n)\widetilde{X}_n,\ \widetilde{X}_1 = X_1, \qquad (4)\\
\widetilde{U}_{n+1} &= \frac{\widetilde{X}_{n+1}}{\prod_{i=1}^{n}(1+\lambda_1 a_i)} = \frac{I + a_n B_n}{1+\lambda_1 a_n}\widetilde{U}_n = \widetilde{U}_n + \frac{a_n}{1+\lambda_1 a_n}(B_n\widetilde{U}_n - \lambda_1\widetilde{U}_n),\ \widetilde{U}_1 = X_1.
\end{aligned}$$

Note that $\frac{\widetilde{U}_n}{\|\widetilde{U}_n\|} = \frac{\widetilde{X}_n}{\|\widetilde{X}_n\|} = X_n$.
We prove in the first step that $\left\|\widetilde{U}_n\right\|$ and $\sum_{n=1}^{\infty} a_n \left\|\widetilde{U}_n\right\|^2 (\lambda_1 - \langle BX_n, X_n\rangle)$ converge a.s., in the second step that $\widetilde{U}_n^j = \left\langle \widetilde{U}_n, V_j\right\rangle$ converges a.s. to $\widetilde{U}^j$ and that $\widetilde{U}^j = 0$ for $j > 1$, in the third step that $\widetilde{U}^1 \neq 0$. The conclusion is then immediate.

Lemma 7. *Suppose $(z_n, n \geqslant 1)$, $(\alpha_n, n \geqslant 1)$, $(\beta_n, n \geqslant 1)$ and $(\gamma_n, n \geqslant 1)$ are four sequences of non-negative numbers such that for all $n \geqslant 1$, $z_{n+1} \leq z_n(1+\alpha_n) + \beta_n - \gamma_n$, $\sum_{n=1}^{\infty}\alpha_n < \infty$, $\sum_{n=1}^{\infty}\beta_n < \infty$. Then the sequence (z_n) converges and $\sum_{n=1}^{\infty}\gamma_n < \infty$.*

This is a deterministic form of the Robbins-Siegmund lemma [22]. Its direct proof is based on the convergence of the decreasing sequence (u_n):

$$u_n = \frac{z_n}{\prod_{l=1}^{n-1}(1+\alpha_l)} - \sum_{k=1}^{n-1}\frac{\beta_k - \gamma_k}{\prod_{l=1}^{k}(1+\alpha_l)} \geqslant -\sum_{k=1}^{\infty}\beta_k.$$

Let ω be fixed, belonging to $C_1 = \{\sum_{n=1}^{\infty} a_n\|B_n - B\| < \infty\}$. The writing of ω will be omitted in the following.
Step 1 Since $a_n \longrightarrow 0$ (H3), $1 + \lambda_1 a_n > 0$ from a certain rank N. Suppose $N = 1$ without loss of generality.

$$\begin{aligned}
\|\widetilde{U}_{n+1}\|^2 &= \|\widetilde{U}_n\|^2 + 2\frac{a_n}{1+\lambda_1 a_n}\left\langle \widetilde{U}_n, (B_n - \lambda_1 I)\widetilde{U}_n\right\rangle + \frac{a_n^2}{(1+\lambda_1 a_n)^2}\|(B_n - \lambda_1 I)\widetilde{U}_n\|^2\\
&= \|\widetilde{U}_n\|^2 + 2\frac{a_n}{1+\lambda_1 a_n}\left\langle \widetilde{U}_n, (B_n - B)\widetilde{U}_n\right\rangle + \frac{a_n^2}{(1+\lambda_1 a_n)^2}\|(B_n - \lambda_1 I)\widetilde{U}_n\|^2 - 2\frac{a_n}{1+\lambda_1 a_n}\left\langle \widetilde{U}_n, (\lambda_1 I - B)\widetilde{U}_n\right\rangle.
\end{aligned}$$

$\lambda_1 I - B$ is a non-negative Q-symmetric matrix, with eigenvalues $0, \lambda_1 - \lambda_2, \ldots, \lambda_1 - \lambda_p$.

$$\begin{aligned}
\|B_n - \lambda_1 I\|^2 &\leq 2\|B_n - B\|^2 + 2\|\lambda_1 I - B\|^2.\\
\|\widetilde{U}_{n+1}\|^2 &\leq \|\widetilde{U}_n\|^2\left(1 + 2\frac{a_n}{1+\lambda_1 a_n}\|B_n - B\| + 2\frac{a_n^2}{(1+\lambda_1 a_n)^2}\|B_n - B\|^2 + 2\frac{a_n^2}{(1+\lambda_1 a_n)^2}(\lambda_1 - \lambda_p)^2\right)\\
&\quad -2\frac{a_n}{1+\lambda_1 a_n}\left\langle \widetilde{U}_n, (\lambda_1 I - B)\widetilde{U}_n\right\rangle.
\end{aligned}$$

By assumptions H2b' and H3, applying Lemma 7 yields:

$$\|\widetilde{U}_n\|^2 \underset{n\to+\infty}{\longrightarrow} \widetilde{U}, \quad \sum_{n=1}^{\infty} a_n\left\langle \widetilde{U}_n, (\lambda_1 I - B)\widetilde{U}_n\right\rangle = \sum_{n=1}^{\infty} a_n\|\widetilde{U}_n\|^2\left(\lambda_1 - \frac{\left\langle \widetilde{U}_n, B\widetilde{U}_n\right\rangle}{\|\widetilde{U}_n\|^2}\right) < \infty.$$

Since $\sum_{n=1}^{\infty} a_n = \infty$, either $\|\widetilde{U}_n\| \underset{n\to+\infty}{\longrightarrow} 0$ or $\sum_{n=1}^{\infty} a_n(\lambda_1 - \langle X_n, BX_n\rangle) < \infty$.

Step 2: Let $\widetilde{U}_n^j = \left\langle \widetilde{U}_n, V_j\right\rangle$.

$$\begin{aligned}
\widetilde{U}_{n+1}^j &= \left\langle V_j, \frac{I + a_n B_n}{1+\lambda_1 a_n}\widetilde{U}_n\right\rangle = \left\langle V_j, \frac{1}{1+\lambda_1 a_n}\left(I + a_n B + a_n(B_n - B)\right)\widetilde{U}_n\right\rangle\\
&= \frac{1+\lambda_j a_n}{1+\lambda_1 a_n}\widetilde{U}_n^j + \frac{a_n}{1+\lambda_1 a_n}\left\langle V_j, (B_n - B)\widetilde{U}_n\right\rangle.
\end{aligned}$$

(a) For $j > 1$, since $a_n \underset{n\to+\infty}{\longrightarrow} 0$ and $\lambda_1 - \lambda_j > 0$, there exists $\alpha_n = O(a_n) > 0$ such that for n sufficiently large:

$$\left|\widetilde{U}_{n+1}^j\right| \leq \frac{1+\lambda_j a_n}{1+\lambda_1 a_n}\left|\widetilde{U}_n^j\right| + \frac{a_n}{1+\lambda_1 a_n}\|B_n - B\|\left\|\widetilde{U}_n\right\| \leq (1-\alpha_n)\left|\widetilde{U}_n^j\right| + \frac{a_n}{1+\lambda_1 a_n}\|B_n - B\|\left\|\widetilde{U}_n\right\|.$$

By H2b' and since $\left\|\widetilde{U}_n\right\|$ converges, applying Lemma 7 yields: $\left|\widetilde{U}_n^j\right| \underset{n\to+\infty}{\longrightarrow} \widetilde{U}^j$, $\sum_{n=1}^{\infty}\alpha_n\left|\widetilde{U}_n^j\right| < \infty$.
Since $\sum_{n=1}^{\infty} a_n = \infty$, $\widetilde{U}^j = 0$.
(b) For $j = 1$, by H2b' and since $\|\widetilde{U}_n\| \longrightarrow \sqrt{\widetilde{U}}$ when $n \longrightarrow \infty$:

$$\begin{aligned}\widetilde{U}_{n+1}^1 &= \widetilde{U}_n^1 + \frac{a_n}{1+\lambda_1 a_n}\left\langle V_1, (B_n - B)\,\widetilde{U}_n\right\rangle = \widetilde{U}_1^1 + \sum_{i=1}^{n}\frac{a_i}{1+\lambda_1 a_i}\left\langle V_1, (B_i - B)\,\widetilde{U}_i\right\rangle \\ &\longrightarrow \widetilde{U}^1 = \widetilde{U}_1^1 + \sum_{i=1}^{\infty}\frac{a_i}{1+\lambda_1 a_i}\left\langle V_1, (B_i - B)\,\widetilde{U}_i\right\rangle.\end{aligned}$$

Now:

$$\widetilde{U}_{n+1}^1 = \left\langle V_1, \widetilde{U}_{n+1}\right\rangle = \left\langle V_1, \prod_{i=1}^{n}\frac{I+a_iB_i}{1+\lambda_1 a_i}\widetilde{U}_1\right\rangle \longrightarrow \left\langle V_1, \prod_{i=1}^{\infty}\frac{I+a_iB_i}{1+\lambda_1 a_i}\widetilde{U}_1\right\rangle = V_1'QS\,\widetilde{U}_1 = \widetilde{U}^1,\ S = \prod_{i=1}^{\infty}\frac{I+a_iB_i}{1+\lambda_1 a_i}.$$

Since $\widetilde{U}_1$ is absolutely continuous, if $V_1'QS \neq 0$, $\Pr\left(V_1'QS\,\widetilde{U}_1 = 0 \mid S\right) = 0$, then $\Pr\left(\widetilde{U}^1 = 0\right) = 0$. Prove that $V_1'QS \neq 0$.
Step 3. Denote $C_2 = \left\{\widetilde{U}_1 \neq 0\right\}$. Suppose $\omega \in C_1 \cap C_2$.
Under H2b' and since $a_n \longrightarrow 0$, there exists N such that $\sum_{n=N}^{\infty} a_n\|B_n - B\| < \ln 2$ and all eigenvalues of $I + a_iB$ are positive for $i \geqslant N$, then $\|I + a_iB\| = 1 + \lambda_1 a_i$ and

$$V_1'QS = V_1'Q\prod_{i=N}^{\infty}\frac{I+a_iB_i}{1+\lambda_1 a_i}\prod_{i=1}^{N-1}\frac{I+a_iB_i}{1+\lambda_1 a_i} = V_1'QR\prod_{i=1}^{N-1}\frac{I+a_iB_i}{1+\lambda_1 a_i},\ R = \prod_{i=N}^{\infty}\frac{I+a_iB_i}{1+\lambda_1 a_i}.$$

Under H2c, $V_1'QS \neq 0 \Leftrightarrow V_1'QR \neq 0$. Let $C_n = \frac{a_n\|B_n - B\|}{1+\lambda_1 a_n}$ and $(W_n, n \geqslant N)$ be the process $\left(\widetilde{U}_n, n \geqslant N\right)$ with $W_N = V_1$.
By Step 2, since $W_N = V_1$, for $i \in \{N+1, \ldots, n\}$:

$$\begin{aligned}W_{n+1}^1 &= \langle V_1, W_{n+1}\rangle = 1 + \sum_{i=N}^{n}\frac{a_i}{1+\lambda_1 a_i}\langle V_1, (B_i - B)\,W_i\rangle \geqslant 1 - \sum_{i=N}^{n} C_i\,\|W_i\|; \\ \|W_i\| &\leq \frac{\|I + a_{i-1}B_{i-1}\|}{1+\lambda_1 a_{i-1}}\|W_{i-1}\| \leq \frac{\|I + a_{i-1}B\| + a_{i-1}\|B_{i-1} - B\|}{1+\lambda_1 a_{i-1}}\|W_{i-1}\| = (1 + C_{i-1})\|W_{i-1}\| \leq \prod_{l=N}^{i-1}(1 + C_l).\end{aligned}$$

Since $\sum_{n=N}^{\infty} C_n < \ln 2$, it follows that $W_{n+1}^1 \geqslant 1 - \sum_{i=N}^{n} C_i\prod_{l=N}^{i-1}(1 + C_l) = 1 - \left(\prod_{l=N}^{n}(1 + C_l) - 1\right) \geqslant 2 - e^{\sum_{l=N}^{n} C_l} > 0$.
By Step 2, W_n^1 converges to $\langle V_1, RV_1\rangle = V_1'QRV_1$ which is therefore strictly positive, thus $V_1'QR \neq 0$.

Step 4: conclusion. It follows that $\left(\widetilde{U}_n\right)$ converges to $\widetilde{U}^1V_1 \neq 0$, therefore $\frac{\widetilde{U}_n}{\|\widetilde{U}_n\|} = X_n$ converges to $\pm V_1$, and by the conclusion of Step 1, $\sum_{n=1}^{\infty} a_n(\lambda_1 - \langle X_n, BX_n\rangle) < \infty$. Moreover, by H2b':
$\sum_{n=1}^{\infty} a_n|\lambda_1 - \langle X_n, B_nX_n\rangle| = \sum_{n=1}^{\infty} a_n|\lambda_1 - \langle X_n, (B_n - B)\,X_n\rangle - \langle X_n, BX_n\rangle|$
$\leq \sum_{n=1}^{\infty} a_n(\lambda_1 - \langle X_n, BX_n\rangle) + \sum_{n=1}^{\infty} a_n\|B_n - B\| < \infty$.
This concludes the proof. □

Remark. Step 1 can be replaced by: $\|\widetilde{U}_{n+1}\| \leq \frac{\|I+a_nB_n\|}{1+\lambda_1 a_n}\left\|\widetilde{U}_n\right\| \leq \left(1 + \frac{a_n\|B_n - B\|}{1+\lambda_1 a_n}\right)\left\|\widetilde{U}_n\right\|$.Under H2b', $\left\|\widetilde{U}_n\right\|$ converges a.s. Assumption $\sum_{n=1}^{\infty} a_n^2 < \infty$ is not used and can be replaced by $a_n \longrightarrow 0$ when $n \longrightarrow \infty$, but in this case the convergence of $\sum_{n=1}^{\infty} a_n(\lambda_1 - \langle BX_n, X_n\rangle)$ and $\sum_{n=1}^{\infty} a_n|\lambda_1 - \langle B_nX_n, X_n\rangle|$ is not proven.

Proof of Corollary 2, first part. Let us first recall some concepts of exterior algebra used in the proof. Let $(e_1, \ldots, e_p)$ be a basis of $\mathbb{R}^p$. For $r \leq p$, let ${}^r\Lambda\mathbb{R}^p$ be the exterior power of order r of $\mathbb{R}^p$, generated by the C_p^r exterior products $e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_r}$, $i_1 < i_2 < \cdots < i_r \in \{1, \ldots, p\}$.

(a) Let Q be a metric in $\mathbb{R}^p$. Define the inner product $\langle ., .\rangle$ in ${}^r\Lambda\mathbb{R}^p$ induced by the metric Q such that:

$$\langle e_{i_1} \wedge \cdots \wedge e_{i_r}, e_{k_1} \wedge \cdots \wedge e_{k_r}\rangle = \sum_{\sigma\in G_r}(-1)^{s(\sigma)}\langle e_{i_1}, e_{\sigma(k_1)}\rangle \times \cdots \times \langle e_{i_r}, e_{\sigma(k_r)}\rangle,$$

G_r being the set of permutations σ of $\{k_1, \ldots, k_r\}$ and $s(\sigma)$ the number of inversions of σ. The associated norm is also denoted $\|.\|$. Note that if $x_1, \ldots, x_r$ are Q-orthogonal, $\|x_1 \wedge \cdots \wedge x_r\| = \prod_{i=1}^r \|x_i\|$, and if $(e_1, \ldots, e_p)$ is a Q-orthonormal basis of $\mathbb{R}^p$, then the set of the C_p^r exterior products $e_{i_1} \wedge \cdots \wedge e_{i_r}$ is an orthonormal basis of ${}^r\Lambda\mathbb{R}^p$.

(b) Let U be an endomorphism in $\mathbb{R}^p$. Define for $j \in \{1, \ldots, r\}$ the endomorphism ${}^{rj}U$ in ${}^r\Lambda\mathbb{R}^p$ such that:

$$\begin{aligned}
{}^{rj}U(x_1 \wedge \cdots \wedge x_r) &= \sum_{1 \leq i_1 < i_2 < \ldots < i_j \leq r} x_1 \wedge \cdots \wedge Ux_{i_1} \wedge \cdots \wedge Ux_{i_j} \wedge \cdots \wedge x_r; \\
\text{for } j &= 1, {}^{r1}U(x_1 \wedge \cdots \wedge x_r) = \sum_{i=1}^r x_1 \wedge \cdots \wedge Ux_i \wedge \cdots \wedge x_r; \\
\text{for } j &= r, {}^{rr}U(x_1 \wedge \cdots \wedge x_r) = Ux_1 \wedge Ux_2 \wedge \cdots \wedge Ux_r.
\end{aligned}$$

(c) The following properties hold:

(i) Suppose U has p real eigenvalues $\lambda_1, \ldots, \lambda_p$ and, for $j \in \{1, \ldots, p\}$, let V_j be an eigenvector corresponding to λ_j. Then the C_p^r vectors $V_{i_1} \wedge \cdots \wedge V_{i_r}$, $1 \leq i_1 < \cdots < i_r \leq p$, are eigenvectors of ${}^{rj}U$; for $j = 1$, the corresponding eigenvalues are $\lambda_{i_1} + \cdots + \lambda_{i_r}$ and for $j = r$, the products $\lambda_{i_1} \cdots \lambda_{i_r}$; thus if U is invertible, so is ${}^{rr}U$;

(ii) If U is Q-symmetric, ${}^{rj}U$ is symmetric with respect to the metric induced by Q in ${}^r\Lambda\mathbb{R}^p$;

(iii) ${}^{rr}(I + U) = I + \sum_{j=1}^r {}^{rj}U$;

(iv) There exists $c(r) > 0$ such that, for every endomorphism U in $\mathbb{R}^p$ and for $1 \leq j \leq r$, $\|{}^{rj}U\| \leq c(r)\|U\|^j$.

Outline of the proof. Applying the first assertion of Theorem 1, we prove in the first step that for $i \in \{1, \ldots, r\}$, ${}^iX_n = X_n^1 \wedge \cdots \wedge X_n^i$ converges a.s. to an eigenvector $\pm V_{j_1} \wedge \cdots \wedge V_{j_i}$ of ${}^{i1}B$. In the second step, we prove by recurrence on i that X_n^i converges a.s. to $\pm V_i$, by proving first that there exists $j > i - 1$ such that ${}^iX_n$ converges a.s. to $\pm V_1 \wedge \cdots \wedge V_{i-1} \wedge V_j$, thus X_n^i converges a.s. to $\pm V_j$, and then, applying the second assertion of Theorem 1, that $V_j = \pm V_i$. In the third step, we prove that $\sum_{n=1}^\infty a_n \left|\lambda_i - \left\langle BX_n^i, X_n^i \right\rangle\right|$ and $\sum_{n=1}^\infty a_n(\lambda_i - \left\langle B_n X_n^i, X_n^i \right\rangle)$ converge a.s.

Step 1. For $i \in \{1, \ldots, r\}$, it follows from the orthogonality of $T_n^1, \ldots, T_n^i$ that $\|T_{n+1}^1 \wedge \cdots \wedge T_{n+1}^i\| = \prod_{l=1}^i \|T_{n+1}^l\|$. Let ${}^iX_{n+1} = X_{n+1}^1 \wedge \cdots \wedge X_{n+1}^i$ and $D_n^i = {}^{i1}B_n + \sum_{j=2}^i a_n^{j-1} \, {}^{ij}B_n$. Then:

$$\begin{aligned}
{}^iX_{n+1} &= \frac{T_{n+1}^1 \wedge \cdots \wedge T_{n+1}^i}{\|T_{n+1}^1 \wedge \cdots \wedge T_{n+1}^i\|} = \frac{Y_{n+1}^1 \wedge \cdots \wedge Y_{n+1}^i}{\|Y_{n+1}^1 \wedge \cdots \wedge Y_{n+1}^i\|} = \frac{{}^{ii}(I + a_n B_n)\, {}^iX_n}{\|{}^{ii}(I + a_n B_n)\, {}^iX_n\|} \\
&= \frac{\left(I + a_n \, {}^{i1}B_n + \sum_{j=2}^i a_n^j \, {}^{ij}B_n\right) {}^iX_n}{\left\|\left(I + a_n \, {}^{i1}B_n + \sum_{j=2}^i a_n^j \, {}^{ij}B_n\right) {}^iX_n\right\|} = \frac{\left(I + a_n D_n^i\right) {}^iX_n}{\|\left(I + a_n D_n^i\right) {}^iX_n\|}.
\end{aligned}$$

Since $\|{}^{ij}B_n\| \leq c(i)\|B_n\|^j$, assumptions H2a ($\sup_n \|B_n\| < b$ a.s.) and H3 yield that there exists $b_1 > 0$ such that for all n, $\|D_n^i\| \leq b_1$. Moreover, since $U \mapsto {}^{i1}U$ is a linear application, assumptions H2a,b ($\sum_{n=1}^\infty a_n E\left[\|E[B_n|T_n] - B\|\right] < \infty$) and H3 yield that:

$E\left[\sum_{n=1}^\infty a_n \|E[D_n^i|T_n] - {}^{i1}B\|\right] = E\left[\sum_{n=1}^\infty a_n \left\|E\left[{}^{i1}B_n - {}^{i1}B + \sum_{j=2}^i a_n^{j-1} \, {}^{ij}B_n \,|T_n\right]\right\|\right]$

$\leq E\left[\sum_{n=1}^\infty a_n \left(\|{}^{i1}E[B_n - B\,|T_n]\| + \sum_{j=2}^i a_n^{j-1} E[\|{}^{ij}B_n\|\,|T_n]\right)\right]$

$\leq c(i) E\left[\sum_{n=1}^\infty a_n \left(\|E\,[B_n|T_n] - B]\| + \sum_{j=2}^i a_n^{j-1} E[\|B_n\|^j\,|T_n]\right)\right] < \infty.$

${}^{i1}B$ is symmetric with respect to the metric induced by Q in ${}^i\Lambda\mathbb{R}^p$ and by H1b' its largest eigenvalue is simple. Applying the first assertion of Theorem 1 yields that almost surely:

$$\begin{gathered}
{}^iX_n \text{ converges to a normed eigenvector } \pm V_{j_1} \wedge \cdots \wedge V_{j_i} \; of \; {}^{i1}B, \\
\sum_{n=1}^\infty a_n \left(\lambda_{j_1} + \cdots + \lambda_{j_i} - \langle {}^{i1}B \, {}^iX_n, {}^iX_n\rangle\right) < \infty, \sum_{n=1}^\infty a_n \left(\lambda_{j_1} + \cdots + \lambda_{j_i} - \langle D_n^i \, {}^iX_n, {}^iX_n\rangle\right) < \infty.
\end{gathered}$$

Moreover, by H2a and H3, $\sum_{n=1}^\infty a_n \left(\lambda_{j_1} + \cdots + \lambda_{j_i} - \langle {}^{i1}B_n \, {}^iX_n, {}^iX_n\rangle\right) < \infty$.

Step 2. Suppose that for $k \in \{1, \ldots, i-1\}$, when $n \longrightarrow \infty, X_n^k \longrightarrow \pm V_k$, which is verified for $k = 1$ by Theorem 1, and prove that $X_n^i \longrightarrow \pm V_i$.

(a) Prove that there exists $j > i - 1$ such that $X_n^i \longrightarrow \pm V_j$.
${}^iX_n \longrightarrow \pm V_{j_1} \wedge \cdots \wedge V_{j_i}$; suppose that there exists $k \in \{1, \ldots, i-1\}$ such that, for $l \in \{1, \ldots, i\}$, $V_{j_l} \neq \pm V_k$; since $X_n^k \longrightarrow \pm V_k$, $\langle X_n^k, V_{j_l} \rangle \longrightarrow 0$ for $l \in \{1, \ldots, i\}$ and $\langle X_n^1 \Lambda \cdots \Lambda X_n^i, V_{j_1} \Lambda \cdots \Lambda V_{j_i} \rangle \longrightarrow 0$, a contradiction. Therefore for all $k \in \{1, \ldots, i-1\}$, there exists j_l such that $V_{j_l} = \pm V_k$ and there exists j such that ${}^iX_n \longrightarrow \pm V_1 \Lambda \cdots \Lambda V_{i-1} \wedge V_j$.
The only term which has a non-zero limit in the development of $\left\langle X_n^1 \wedge \cdots \wedge X_n^i, \pm V_1 \wedge \cdots \wedge V_{i-1} \wedge V_j \right\rangle$, whose limit is 1 when $n \longrightarrow \infty$, is $\pm \langle X_n^1, V_1 \rangle \langle X_n^2, V_2 \rangle \times \cdots \times \langle X_n^{i-1}, V_{i-1} \rangle \left\langle X_n^i, V_j \right\rangle$ obtained for $\sigma = Id$. Since for $k \in \{1, \ldots, i-1\}$, $\langle X_n^k, V_k \rangle \longrightarrow \pm 1$, then $\left\langle X_n^i, V_j \right\rangle \longrightarrow \pm 1$. Therefore, $X_n^i \longrightarrow \pm V_j$.
(b) Prove now that $V_j = \pm V_i$. Suppose $X_n^i \longrightarrow \pm V_j \neq \pm V_i$.
Let G_i be the set of permutations σ of $\{1, \ldots, i\}$ with $\sigma = (\sigma(1), \ldots, \sigma(i))$ and $s(\sigma)$ the number of inversions of σ.
Prove that $\liminf E\left[\langle D_n^i(X_n^1 \wedge \cdots \wedge X_n^i), V_1 \wedge \cdots \wedge V_i \rangle^2 | T_n\right] > 0, D_n^i = {}^{i1}B_n + \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n$.

$$
\begin{aligned}
\langle {}^{i1}B_n(X_n^1 \wedge \cdots \wedge X_n^i), V_1 \wedge \cdots \wedge V_i \rangle &= \sum_{l=1}^{i} \langle X_n^1 \wedge \cdots \wedge B_n X_n^l \wedge \cdots \wedge X_n^i, V_1 \wedge \cdots \wedge V_i \rangle \\
&= \sum_{l=1}^{i} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle X_n^1, V_{\sigma(1)} \rangle \times \cdots \times \langle B_n X_n^l, V_{\sigma(l)} \rangle \times \cdots \times \langle X_n^i, V_{\sigma(i)} \rangle,
\end{aligned}
$$

$$
E\left[\left\langle {}^{i1}B_n(X_n^1 \wedge \cdots \wedge X_n^i), V_1 \wedge \cdots \wedge V_i \right\rangle^2 \ |T_n\right] = E\left[\left(\sum_{l=1}^{i} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle X_n^1, V_{\sigma(1)} \rangle \cdots \langle B_n X_n^l, V_{\sigma(l)} \rangle \cdots \langle X_n^i, V_{\sigma(i)} \rangle \right)^2 \ |T_n\right].
$$

Since for $k \in \{1, \ldots, i-1\}$, $X_n^k \longrightarrow \pm V_k$, the only term with a non-zero limit in the development of this conditional expectation is $\langle X_n^1, V_1 \rangle^2 \times \cdots \times \langle X_n^{i-1}, V_{i-1} \rangle^2 E\left[\langle B_n X_n^i, V_i \rangle^2 \ |T_n\right]$ and

$$
\liminf E\left[\left\langle {}^{i1}B_n(X_n^1 \wedge \cdots \wedge X_n^i), V_1 \wedge \cdots \wedge V_i \right\rangle^2 \ |T_n\right] = \liminf E\left[\langle B_n X_n^i, V_i \rangle^2 \ |T_n\right] > 0.
$$

Moreover, by H2a and $\lim_{n \longrightarrow \infty} a_n = 0$, $\liminf E\left[\left\langle \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n(X_n^1 \wedge \cdots \wedge X_n^i), V_1 \wedge \cdots \wedge V_i \right\rangle^2 \ |T_n\right] = 0$.
Then $\liminf E\left[\langle D_n^i(X_n^1 \wedge \cdots \wedge X_n^i), V_1 \wedge \cdots \wedge V_i \rangle^2 | T_n\right] > 0$. Applying the second assertion of Theorem 1 yields a.s.:

$$
\begin{gathered}
X_n^1 \wedge \cdots \wedge X_n^i \longrightarrow \pm V_1 \wedge \cdots \wedge V_i, \text{ therefore } X_n^i \longrightarrow \pm V_i, \\
\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B\, {}^iX_n, {}^iX_n \rangle \right) < \infty, \sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle D_n^i\, {}^iX_n, {}^iX_n \rangle \right) < \infty.
\end{gathered}
$$

Then, by H2a and H3, $\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B_n\, {}^iX_n, {}^iX_n \rangle \right) < \infty$.

Step 3. Since $X_n^1, \ldots, X_n^i$ are orthonormal, $\langle {}^{i1}B\, {}^iX_n, {}^iX_n \rangle = \sum_{k=1}^{i} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle X_n^1, X_n^{\sigma(1)} \rangle \times \cdots \times \langle BX_n^k, X_n^{\sigma(k)} \rangle \times \cdots \times \langle X_n^i, X_n^{\sigma(i)} \rangle = \sum_{k=1}^{i} \langle X_n^1, X_n^1 \rangle \times \cdots \times \langle BX_n^k, X_n^k \rangle \times \cdots \times \langle X_n^i, X_n^i \rangle = \sum_{k=1}^{i} \langle BX_n^k, X_n^k \rangle$. Then, since $\sum_{l=1}^{i} \lambda_l$ is the largest eigenvalue of ${}^{i1}B$:

$$
\begin{aligned}
\lambda_i - \left\langle BX_n^i, X_n^i \right\rangle &= \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B\, {}^iX_n, {}^iX_n \rangle \right) - \left(\sum_{l=1}^{i-1} \lambda_l - \langle {}^{i-1,1}B\, {}^{i-1}X_n, {}^{i-1}X_n \rangle \right); \\
\left| \lambda_i - \left\langle BX_n^i, X_n^i \right\rangle \right| &\leq \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B\, {}^iX_n, {}^iX_n \rangle \right) + \left(\sum_{l=1}^{i-1} \lambda_l - \langle {}^{i-1,1}B\, {}^{i-1}X_n, {}^{i-1}X_n \rangle \right).
\end{aligned}
$$

Since $\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B^iX_n, {}^iX_n \rangle \right) < \infty$ a.s., then $\sum_{n=1}^{\infty} a_n \left| \lambda_i - \left\langle BX_n^i, X_n^i \right\rangle \right| < \infty$ a.s. Likewise:
since $\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B_n\, {}^iX_n, {}^iX_n \rangle \right) < \infty$, then $\sum_{n=1}^{\infty} a_n \left(\lambda_i - \left\langle B_n X_n^i, X_n^i \right\rangle \right) < \infty$ a.s. □

Proof of Corollary 2, second part. We prove applying Theorem 1 that ${}^iX_n = X_n^1 \wedge \cdots \wedge X_n^i$ converges a.s. to

$\pm V_1 \wedge \cdots \wedge V_i$, then by recurrence on i that X_n^i converges a.s. to $\pm V_i$.
Let ω be fixed throughout the proof, belonging to the intersection of the a.s. convergence sets. Its writing will be omitted. Let $i \in \{1, \ldots, r\}$. As already seen in the proof of Corollary 2, first part:

$$^{i}X_{n+1} = \frac{\left(I + a_n\, D_n^i\right)\, {}^{i}X_n}{\left\|\left(I + a_n\, D_n^i\right)\, {}^{i}X_n\right\|},\ D_n^i = {}^{i1}B_n + \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n.$$

By H2b' and H3:

$$\sum_{n=1}^{\infty} a_n \left\|D_n^i - {}^{i1}B\right\| = \sum_{n=1}^{\infty} a_n \left\|{}^{i1}\left(B_n - B\right) + \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n\right\| \leq c(i)\left(\sum_{n=1}^{\infty} a_n \left\|B_n - B\right\| + \sum_{j=2}^{i}\sum_{n=1}^{\infty} a_n^j \left\|B_n\right\|^j\right) < \infty.$$

By H2c, $I + a_n B_n$ has no null eigenvalue, then ${}^{ii}(I + a_n B_n)$ has no null eigenvalue and is invertible.
Since B is Q-symmetric with distinct eigenvalues (H1a,b'), ${}^{i1}B$ is symmetric with respect to the metric induced by Q in ${}^{i}\Lambda\mathbb{R}^p$ and its largest eigenvalue $\lambda_1 + \cdots + \lambda_i$ is simple; $V_1 \wedge \cdots \wedge V_i$ is a normed eigenvector corresponding to this eigenvalue. Applying the second part of Theorem 1 yields that:

$$^{i}X_n \longrightarrow \pm V_1 \wedge \cdots \wedge V_i, \sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i}\lambda_l - \left\langle {}^{i1}B\, {}^{i}X_n,\ {}^{i}X_n\right\rangle\right) < \infty,\ \sum_{n=1}^{\infty} a_n \left|\sum_{l=1}^{i}\lambda_l - \left\langle D_n^i\, {}^{i}X_n,\ {}^{i}X_n\right\rangle\right| < \infty.$$

It implies that $\sum_{n=1}^{\infty} a_n \left|\sum_{l=1}^{i} \lambda_l - \left\langle {}^{i1}B_n\, {}^{i}X_n,\ {}^{i}X_n\right\rangle\right| < \infty$ since $D_n^i = {}^{i1}B_n + \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n$ and

$$\sum_{n=1}^{\infty} a_n \left|\left\langle \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n\, {}^{i}X_n, {}^{i}X_n\right\rangle\right| \leq \sum_{j=2}^{i}\sum_{n=1}^{\infty} a_n^j \left\|{}^{ij}B_n\right\| \leq c(i)\sum_{j=2}^{i}\sum_{n=1}^{\infty} a_n^j \left\|B_n\right\|^j \leq c(i)\sum_{j=2}^{i} 2^{j-1}\sum_{n=1}^{\infty} a_n^j \left(\left\|B_n - B\right\|^j + \left\|B\right\|^j\right) < \infty.$$

Suppose that, for $k \in \{1, \ldots, i-1\}$, X_n^k converges to $\pm V_k$, which is verified for $k = 1$ by Theorem 1, and prove that it is true for $k = i$. In the development of $\left\langle X_n^1 \wedge \cdots \wedge X_n^i, \pm V_1 \wedge \cdots \wedge V_i\right\rangle$, which converges to ± 1, the only term which has a non-zero limit is $\left\langle X_n^1, V_1\right\rangle \times \cdots \times \left\langle X_n^{i-1}, V_{i-1}\right\rangle \left\langle X_n^i, V_i\right\rangle$; since for $k \in \{1, \ldots, i-1\}$, $\left\langle X_n^k, V_k\right\rangle$ converges to ± 1, it follows that $\left\langle X_n^i, V_i\right\rangle$ converges to ± 1, thus X_n^i converges to $\pm V_i$.
Applying the same argument as in the previous proof, Step 3, yields: $\sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, BX_n^i\right\rangle\right| < \infty$. By H2b':

$$\sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, B_n X_n^i\right\rangle\right| \leqslant \sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, BX_n^i\right\rangle\right| + \sum_{n=1}^{\infty} a_n \left\|B_n - B\right\| < \infty.$$

□

Proof of Corollary 3. We verify the assumptions of the first part of Corollary 2. B is symmetric (H1a), $\sup_n \|B_n\|$ is a.s. bounded under H6a and H7a; we have almost surely:

$$\begin{aligned}
E\left[B_n \mid T_n\right] - B &= E\left[M_{n-1}^{\frac{1}{2}}\left(\tfrac{1}{m_n}\textstyle\sum_{i=1}^{m_n} Z_{ni}^c \left(Z_{ni}^c\right)^\top - \overline{Z}_{n-1}^c\left(\overline{Z}_{n-1}^c\right)^\top\right) M_{n-1}^{\frac{1}{2}} \mid T_n\right] - M^{\frac{1}{2}}\left(E\left[Z^c\left(Z^c\right)^\top\right] - E[Z^c]E[Z^c]^\top\right) M^{\frac{1}{2}} \\
&= M_{n-1}^{\frac{1}{2}}\left(E\left[Z^c\left(Z^c\right)^\top\right] - \overline{Z}_{n-1}^c\left(\overline{Z}_{n-1}^c\right)^\top\right) M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}}\left(E\left[Z^c\left(Z^c\right)^\top\right] - E[Z^c]E[Z^c]^\top\right) M^{\frac{1}{2}} \\
&= \left(M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}}\right)\left(E\left[Z^c\left(Z^c\right)^\top\right] - E[Z^c]E[Z^c]^\top\right) M_{n-1}^{\frac{1}{2}} + M^{\frac{1}{2}}\left(E\left[Z^c\left(Z^c\right)^\top\right] - E[Z^c]E[Z^c]^\top\right)\left(M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}}\right) \\
&\quad - M_{n-1}^{\frac{1}{2}}\left(\overline{Z}_{n-1}^c - E[Z^c]\right)\left(\overline{Z}_{n-1}^c\right)^\top M_{n-1}^{\frac{1}{2}} - M_{n-1}^{\frac{1}{2}} E[Z^c]\left(\overline{Z}_{n-1}^c - E[Z^c]\right)^\top M_{n-1}^{\frac{1}{2}}.
\end{aligned}$$

If Z has 4^{th} order moments and $a_n > 0$, $\sum_{n=1}^{\infty} \frac{a_n}{\sqrt{n}} < \infty$: $\sum_{n=1}^{\infty} a_n E\left[\|\overline{Z}_{n-1}^c - E[Z^c]\|\right] = \sum_{n=1}^{\infty} a_n E\left[\|\overline{Z}_{n-1} - E[Z]\|\right] < \infty$.
Therefore, under H6a, H7a,c, $E\left[\sum_{n=1}^{\infty} a_n \left\|E\left[B_n \mid T_n\right] - B\right\|\right] < \infty$ (H2b). By Corollary 2, for $k = 1, \ldots, r$, X_n^k converges a.s. to one of the eigenvectors of B.

Prove now that $\lim\limits_{n\to\infty} E[((X_n^k)^\top B_n V_k)^2 \mid T_n] > 0$ a.s. on the set $\left\{X_n^k \longrightarrow V_j\right\}$ for $j \neq k$. In the remainder of the proof, X_n^k is denoted X_n. Decompose $E[(X_n^\top B_n V_k)^2 \mid T_n]$ into the sum of three terms:

$$E\left[\left(X_n^\top M_{n-1}^{\frac{1}{2}}\left(\tfrac{1}{m_n}\textstyle\sum_{i=1}^{m_n} Z_{ni}^c\left(Z_{ni}^c\right)^\top - \overline{Z}_{n-1}^c\left(\overline{Z}_{n-1}^c\right)^\top\right) M_{n-1}^{\frac{1}{2}} V_k\right)^2 \mid T_n\right]$$

$= E\left[\left(\frac{1}{m_n}\sum_{i=1}^{m_n}\left(X_n^\top M_{n-1}^{\frac{1}{2}}Z_{ni}^c\right)\left(\left(Z_{ni}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right)-\left(X_n^\top M_{n-1}^{\frac{1}{2}}\overline{Z}_{n-1}^c\right)\left(\left(\overline{Z}_{n-1}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right)\right)^2 \mid T_n\right] = A_n + B_n + C_n,$

$A_n = E\left[\left(\frac{1}{m_n}\sum_{i=1}^{m_n}\left(X_n^\top M_{n-1}^{\frac{1}{2}}Z_{ni}^c\right)\left(\left(Z_{ni}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right)\right)^2 \mid T_n\right],$

$B_n = -2\left(X_n^\top M_{n-1}^{\frac{1}{2}}\overline{Z}_{n-1}^c\right)\left(\left(\overline{Z}_{n-1}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right)\frac{1}{m_n}\sum_{i=1}^{m_n}E\left[\left(X_n^\top M_{n-1}^{\frac{1}{2}}Z_{ni}^c\right)\left(\left(Z_{ni}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right) \mid T_n\right],$

$C_n = \left(X_n^\top M_{n-1}^{\frac{1}{2}}\overline{Z}_{n-1}^c\right)^2\left(\left(\overline{Z}_{n-1}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right)^2.$

Note that the two random variables $R = V_j^\top M^{\frac{1}{2}}Z^c$ and $S = V_k^\top M^{\frac{1}{2}}Z^c$ are uncorrelated:

$$\begin{aligned} E[(R-E[R])(S-E[S])] &= E[V_j^\top M^{\frac{1}{2}}(Z-E[Z]).V_k^\top M^{\frac{1}{2}}(Z-E[Z])] \\ &= V_j^\top M^{\frac{1}{2}}E\left[(Z-E[Z])(Z-E[Z])^\top\right]M^{\frac{1}{2}}V_k = \lambda_k V_j^\top V_k = 0. \end{aligned}$$

Then, $E[V_j^\top M^{\frac{1}{2}}Z^c.V_k^\top M^{\frac{1}{2}}Z^c] = E[V_j^\top M^{\frac{1}{2}}Z^c]E[V_k^\top M^{\frac{1}{2}}Z^c]$. Under H7b, almost surely, when $n \longrightarrow \infty$, we have:

$A_n = \frac{1}{m_n^2}\sum_{i=1}^{m_n}\sum_{l=1}^{m_n}E\left[\left(X_n^\top M_{n-1}^{\frac{1}{2}}Z_{ni}^c\right)\left(\left(Z_{ni}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right)\left(X_n^\top M_{n-1}^{\frac{1}{2}}Z_{nl}^c\right)\left(\left(Z_{nl}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right) | T_n\right]$

$= X_n^\top M_{n-1}^{\frac{1}{2}}\frac{1}{m_n^2}\sum_{i=1}^{m_n}\sum_{l=1}^{m_n}E\left[\left(V_k^\top M_{n-1}^{\frac{1}{2}}Z_{ni}^c\right)Z_{ni}^c\left(Z_{nl}^c\right)^\top\left(\left(Z_{nl}^c\right)^\top M_{n-1}^{\frac{1}{2}}V_k\right) \mid T_n\right]M_{n-1}^{\frac{1}{2}}X_n$

$\longrightarrow V_j^\top M^{\frac{1}{2}}E\left[\left(V_k^\top M^{\frac{1}{2}}Z^c\right)Z^c(Z^c)^\top\left((Z^c)^\top M^{\frac{1}{2}}V_k\right)\right]M^{\frac{1}{2}}V_j = E\left[\left(V_k^\top M^{\frac{1}{2}}Z^c\right)^2\left(V_j^\top M^{\frac{1}{2}}Z^c\right)^2\right];$

$B_n \longrightarrow -2E\left[V_j^\top M^{\frac{1}{2}}Z^c\right]E\left[(Z^c)^\top M^{\frac{1}{2}}V_k\right]E\left[\left(V_j^\top M^{\frac{1}{2}}Z^c\right)\left((Z^c)^\top M^{\frac{1}{2}}V_k\right)\right] = -2E\left[\left(V_j^\top M^{\frac{1}{2}}Z^c\right)\left(V_k^\top M^{\frac{1}{2}}Z^c\right)\right]^2;$

$C_n \longrightarrow \left(E\left[V_j^\top M^{\frac{1}{2}}Z^c\right]E\left[V_k' M^{\frac{1}{2}}Z^c\right]\right)^2 = E\left[\left(V_j^\top M^{\frac{1}{2}}Z^c\right)\left(V_k^\top M^{\frac{1}{2}}Z^c\right)\right]^2;$

$E[(X_n^\top B_n V_k)^2 | T_n] = A_n + B_n + C_n \longrightarrow E\left[\left(V_j^\top M^{\frac{1}{2}}Z^c\right)^2\left(V_k^\top M^{\frac{1}{2}}Z^c\right)^2\right] - E\left[\left(V_j^\top M^{\frac{1}{2}}Z^c\right)\left(V_k^\top M^{\frac{1}{2}}Z^c\right)\right]^2$

$= Var[V_j^\top M^{\frac{1}{2}}Z^c.V_k^\top M^{\frac{1}{2}}Z^c] > 0$ a.s. by H6b. □

Proof of Corollary 4. Let us verify the assumptions of the second part of Corollary 2. H2c is verified since the eigenvalues of B_n are non-negative. Prove now that $\sum_{n=1}^\infty a_n \|B_n - B\| < \infty$ a.s.

$B = M^{\frac{1}{2}}CM^{\frac{1}{2}}, C = E[ZZ^\top] - E[Z]E[Z]^\top;$

$B_n = M_n^{\frac{1}{2}}C_n M_n^{\frac{1}{2}}, C_n = \frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i}Z_{ij}^c\left(Z_{ij}^c\right)^\top - \overline{Z}_n^c\left(\overline{Z}_n^c\right)^\top = \frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i}Z_{ij}\left(Z_{ij}\right)^\top - \overline{Z}_n\left(\overline{Z}_n\right)^\top;$

$C_n - C = \frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i}Z_{ij}\left(Z_{ij}\right)^\top - E[ZZ^\top] - \left(\overline{Z}_n - E[Z]\right)\left(\overline{Z}_n\right)^\top - E[Z](\overline{Z}_n - E[Z])^\top;$

$B_n - B = M_n^{\frac{1}{2}}C_n M_n^{\frac{1}{2}} - M^{\frac{1}{2}}CM^{\frac{1}{2}} = (M_n^{\frac{1}{2}} - M^{\frac{1}{2}})C_n M_n^{\frac{1}{2}} + M^{\frac{1}{2}}(C_n - C)M_n^{\frac{1}{2}} + M^{\frac{1}{2}}C(M_n^{\frac{1}{2}} - M^{\frac{1}{2}}).$

Under H3' and H6c, $\sum_{n=1}^\infty a_n\left\|\overline{Z}_n - E[Z]\right\| < \infty$, $\sum_{n=1}^\infty a_n\left\|\frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i}Z_{ij}\left(Z_{ij}\right)^\top - E[ZZ^\top]\right\| < \infty$.
Therefore, under H6c and H7b,d, $\sum_{n=1}^\infty a_n\|B_n - B\| < \infty$. □

Acknowledgments

We thank the Editor, Associate Editor and referees for their valuable comments and suggestions which have helped us improve the presentation of the manuscript. Results incorporated in this article received funding from the investments for the Future programme under grant agreement No ANR-15-RHU-0004.

Appendix. Numerical results

Experiments

Two types of data:

- Simulated data: At each iteration a batch of observations having a multivariate normal distribution are simulated.
- Fixed datasets: At each iteration a random sample of lines is drawn from the dataset.

Algorithms: Two types

- Batch 'B': a mini-batch of observations is used at each iteration.
- All observations 'A': all observations up to the current iteration are used.

Algorithms are also specified by batch size. For example, Algorithm 'A_50' is the algorithm using all observations with a batch size of 50 at each iteration.
For each dataset and each algorithm, we chose either to approximate $r = 3$ or $r = 5$ principal components.

Algorithms initialization: The main process is initialized by choosing $X_1^1, \cdots, X_1^r$ to be a realization of a standard multivariate normal distribution.

Stepsize: $a_n = \frac{20}{n+1}$ where n is the iteration index of the algorithm.

Stopping criteria:

- A maximum number of iterations $N = 500000$.
- For $i = 1, \cdots, r$, the cosine between each X_n^i and the true principal factor V^i is greater or equal 0.99.

Results:

- For each dataset (simulated or fixed) and each algorithm, in the case of convergence, the needed computing time t (in seconds) and the needed number of iterations n to convergence are reported.
- For each dataset (simulated or fixed), the ranks of the algorithms regarding their needed computing time or their needed number of iterations before convergence are reported.
- When an algorithm does not converge after $N = 500000$ iterations, both the needed computing time and number of iterations to convergence are considered to be "Inf".

Simulated data

$Z \in \mathbb{R}^p$ is simulated following multivariate normal law $\mathcal{N}(\mu, \Sigma)$ with different p: $p = 50, 100, 500$ and two choices of Σ referred to as Simu1 and Simu2, respectively:

- a deterministic choice: $\Sigma = A^\top A$ with

$$A = \begin{pmatrix} log(2) & log(2) & \cdots & \cdots & log(2) \\ & log(3) & \cdots & \cdots & log(3) \\ & & \ddots & & \\ & & & \ddots & \\ & & & & log(p+1) \end{pmatrix}$$

  A with this form was chosen since it has a natural incremental form and the log has been applied to its elements such that Σ will not have big values.

- Σ positive and definite matrix randomly generated using the make_sparse_spd_matrix method of the sklearn.datasets package in Python.

The following results were obtained:
Needed computing time before convergence

Algo	Simu1 p=50 r=3	Simu1 p=50 r=5	Simu1 p=100 r=3	Simu1 p=100 r=5	Simu1 p=500 r=3	Simu1 p=500 r=5	Simu2 p=50 r=3	Simu2 p=50 r=5	Simu2 p=100 r=3	Simu2 p=100 r=5	Simu2 p=500 r=3	Simu2 p=500 r=5
B_1	1.07	1.5	2.59	3.96	352.01	496.72	24.27	Inf	49.93	85.31	5146.88	5991.51
B_10	0.16	0.31	0.52	0.52	76.83	93.44	3.53	Inf	8.58	13.09	1356.23	1550.46
B_20	0.11	0.22	0.19	0.45	42.58	87.03	2.43	387.33	6.51	9.03	1107.07	1257.54
B_30	0.05	0.06	0.33	0.5	39.31	79.28	2.02	319.07	10.03	13.06	1100.27	1172.56
B_40	0.05	0.06	0.34	0.5	25.21	73.27	1.34	261.82	8.75	11.53	993.1	1170.65
B_50	0.05	0.05	0.3	0.5	19.81	73.02	1.27	241.09	8.34	11.04	936.33	1150.79
A_1	0.11	1.11	0.15	0.83	3.07	25.03	1.64	Inf	5.55	31.85	337.08	1224.85
A_10	0.02	0.2	0.02	0.12	1.02	6.06	0.27	207.7	0.94	4.87	79.08	284.03
A_20	0.02	0.07	0.02	0.08	0.74	4.87	0.16	129.7	0.75	3.44	67.74	238.63
A_30	0.02	0.08	0.02	0.11	0.88	4.65	0.16	106.43	0.94	5.08	61.68	219.58
A_40	0.02	0.08	0.03	0.12	1.05	4.46	0.12	93.12	1.06	4.39	55.42	203.94
A_50	0.0	0.08	0.02	0.12	1.1	1.46	0.12	85.5	0.72	4.08	52.6	199.8

Needed number of iterations before convergence

Algo	Simu1 p=50 r=3	Simu1 p=50 r=5	Simu1 p=100 r=3	Simu1 p=100 r=5	Simu1 p=500 r=3	Simu1 p=500 r=5	Simu2 p=50 r=3	Simu2 p=50 r=5	Simu2 p=100 r=3	Simu2 p=100 r=5	Simu2 p=500 r=3	Simu2 p=500 r=5
B_1	2496	2532	4421	5529	21799	31046	57660	Inf	96492	127235	358538	407489
B_10	250	363	557	553	1875	2180	5766	Inf	9650	12729	35854	40750
B_20	127	217	133	302	575	1251	2883	399044	4825	6365	17927	20375
B_30	60	60	143	185	383	834	1922	261661	3217	4239	11952	13583
B_40	43	45	77	151	189	599	1157	189616	2422	3183	8964	10188
B_50	27	36	60	111	122	458	924	154072	1930	2546	7172	8150
A_1	280	1956	222	1131	171	1326	4133	Inf	10542	46623	19313	68970
A_10	29	210	24	115	24	134	415	270500	1106	4667	1934	6903
A_20	15	106	12	58	11	69	208	133996	572	2336	969	3455
A_30	11	72	9	39	9	47	139	91106	390	1559	647	2305
A_40	9	54	7	31	8	36	105	67898	293	1170	487	1731
A_50	8	44	5	25	7	9	92	54637	160	937	390	1386

Algorithms ranking by needed computing time before convergence

Algo	Simu1 p=50 r=3	Simu1 p=50 r=5	Simu1 p=100 r=3	Simu1 p=100 r=5	Simu1 p=500 r=3	Simu1 p=500 r=5	Simu2 p=50 r=3	Simu2 p=50 r=5	Simu2 p=100 r=3	Simu2 p=100 r=5	Simu2 p=500 r=3	Simu2 p=500 r=5	mean rank
A_50	1.0	7.0	3.0	3.0	5.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	2.25
A_20	5.0	4.0	2.0	1.0	1.0	4.0	3.0	4.0	2.0	1.0	4.0	4.0	2.92
A_30	2.0	6.0	4.0	2.0	2.0	3.0	4.0	3.0	3.0	5.0	3.0	3.0	3.33
A_40	4.0	5.0	5.0	4.0	4.0	2.0	2.0	2.0	5.0	3.0	2.0	2.0	3.33
A_10	3.0	8.0	1.0	5.0	3.0	5.0	5.0	5.0	4.0	4.0	5.0	5.0	4.42
B_50	8.0	1.0	8.0	7.0	7.0	7.0	6.0	6.0	8.0	7.0	7.0	6.0	6.5
B_40	7.0	2.0	10.0	8.0	8.0	8.0	7.0	7.0	10.0	8.0	8.0	7.0	7.5
B_30	6.0	3.0	9.0	9.0	9.0	9.0	9.0	8.0	11.0	9.0	9.0	8.0	8.25
A_1	10.0	11.0	6.0	11.0	6.0	6.0	8.0	10.0	6.0	11.0	6.0	9.0	8.33
B_20	9.0	9.0	7.0	6.0	10.0	10.0	10.0	9.0	7.0	6.0	10.0	10.0	8.58
B_10	11.0	10.0	11.0	10.0	11.0	11.0	11.0	10.0	9.0	10.0	11.0	11.0	10.5
B_1	12.0	12.0	12.0	12.0	12.0	12.0	12.0	10.0	12.0	12.0	12.0	12.0	11.83

Algorithms ranking by needed number of iterations before convergence

Algo	Simu1 p=50 r=3	Simu1 p=50 r=5	Simu1 p=100 r=3	Simu1 p=100 r=5	Simu1 p=500 r=3	Simu1 p=500 r=5	Simu2 p=50 r=3	Simu2 p=50 r=5	Simu2 p=100 r=3	Simu2 p=100 r=5	Simu2 p=500 r=3	Simu2 p=500 r=5	mean rank
A_50	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.08
A_40	2.0	4.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.17
A_30	3.0	6.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.25
A_20	4.0	7.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.25
B_50	5.0	1.0	6.0	5.0	6.0	6.0	6.0	5.0	6.0	5.0	6.0	6.0	5.25
A_10	6.0	8.0	5.0	6.0	5.0	5.0	5.0	8.0	5.0	8.0	5.0	5.0	5.92
B_40	7.0	3.0	7.0	7.0	8.0	7.0	7.0	6.0	7.0	6.0	7.0	7.0	6.58
B_30	8.0	5.0	9.0	8.0	9.0	8.0	8.0	7.0	8.0	7.0	8.0	8.0	7.75
B_20	9.0	9.0	8.0	9.0	10.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0	9.0
A_1	11.0	11.0	10.0	11.0	7.0	10.0	10.0	10.0	11.0	11.0	10.0	11.0	10.25
B_10	10.0	10.0	11.0	10.0	11.0	11.0	11.0	10.0	10.0	10.0	11.0	10.0	10.42
B_1	12.0	12.0	12.0	12.0	12.0	12.0	12.0	10.0	12.0	12.0	12.0	12.0	11.83

Fixed datasets

Datasets description

3 datasets were used: Breast_cancer, California_housing and Bio_train, all of which have continuous values:

Dataset	Number of columns	Number of lines	Link
Breast_cancer	8	569	https://archive.ics.uci.edu/ml/datasets/Breast+Cancer +Wisconsin+%28Diagnostic%29
California_housing	9	20 640	https://www.dcc.fc.up.pt/~ltorgo/Regression/DataSets.html
Bio_train	77	145 751	http://osmot.cs.cornell.edu/kddcup/datasets.html

The following results were obtained :

Needed computing time before convergence

Algorithms	breast_cancer r=3	breast_cancer r=5	cal_housing r=3	cal_housing r=5	bio_train r=3	bio_train r=5
B_1	20.13	165.91	Inf	Inf	92.77	173.3
B_10	4.94	19.77	Inf	86.69	14.95	30.59
B_20	4.3	13.23	163.83	268.39	10.45	23.99
B_30	2.79	9.57	159.09	45.25	9.19	18.24
B_40	2.01	8.7	139.1	162.02	8.44	16.44
B_50	2.59	7.68	28.14	31.85	13.03	14.56
A_1	0.24	1.37	0.48	0.57	2.23	3.22
A_10	0.04	0.19	0.06	0.07	0.31	0.48
A_20	0.03	0.11	0.07	0.06	0.28	0.27
A_30	0.03	0.09	0.05	0.02	0.17	0.27
A_40	0.02	0.09	0.03	0.03	0.05	0.25
A_50	0.02	0.08	0.02	0.14	0.08	0.23

Needed number of iterations before convergence

Algorithms	breast_cancer r=3	breast_cancer r=5	cal_housing r=3	cal_housing r=5	bio_train r=3	bio_train r=5
B_1	49220	317329	Inf	Inf	184678	261092
B_10	7046	31734	Inf	148624	18470	26108
B_20	3523	15867	279499	336046	9235	13055
B_30	2349	10581	224031	49529	6156	8704
B_40	1762	7936	168026	168023	4617	6527
B_50	1756	6359	29720	29718	3694	5222
A_1	538	2457	1301	1262	4467	4505
A_10	55	248	134	130	448	452
A_20	28	126	69	67	225	226
A_30	20	85	49	21	150	151
A_40	15	64	39	43	38	114
A_50	12	53	39	72	31	91

Algorithms ranking by needed computing time before convergence

Algorithms	breast_cancer r=3	breast_cancer r=5	cal_housing r=3	cal_housing r=5	bio_train r=3	bio_train r=5	mean rank
A_40	1.0	3.0	2.0	2.0	1.0	2.0	1.83
A_50	2.0	1.0	1.0	5.0	2.0	1.0	2.0
A_30	3.0	2.0	3.0	1.0	3.0	3.0	2.5
A_20	4.0	4.0	5.0	3.0	4.0	4.0	4.0
A_10	5.0	5.0	4.0	4.0	5.0	5.0	4.67
A_1	6.0	6.0	6.0	6.0	6.0	6.0	6.0
B_50	8.0	7.0	7.0	7.0	10.0	7.0	7.67
B_40	7.0	8.0	8.0	10.0	7.0	8.0	8.0
B_30	9.0	9.0	9.0	8.0	8.0	9.0	8.67
B_20	10.0	10.0	10.0	11.0	9.0	10.0	10.0
B_10	11.0	11.0	11.0	9.0	11.0	11.0	10.67
B_1	12.0	12.0	11.0	12.0	12.0	12.0	11.83

Algorithms ranking by needed number of iterations before convergence

Algorithms	breast_cancer r=3	breast_cancer r=5	cal_housing r=3	cal_housing r=5	bio_train r=3	bio_train r=5	mean rank
A_50	1.0	1.0	1.0	4.0	1.0	1.0	1.5
A_40	2.0	2.0	1.0	2.0	2.0	2.0	1.83
A_30	3.0	3.0	3.0	1.0	3.0	3.0	2.67
A_20	4.0	4.0	4.0	3.0	4.0	4.0	3.83
A_10	5.0	5.0	5.0	5.0	5.0	5.0	5.0
A_1	6.0	6.0	6.0	6.0	7.0	6.0	6.17
B_50	7.0	7.0	7.0	7.0	6.0	7.0	6.83
B_40	8.0	8.0	8.0	10.0	8.0	8.0	8.33
B_30	9.0	9.0	9.0	8.0	9.0	9.0	8.83
B_20	10.0	10.0	10.0	11.0	10.0	10.0	10.17
B_10	11.0	11.0	11.0	9.0	11.0	11.0	10.67
B_1	12.0	12.0	11.0	12.0	12.0	12.0	11.83

Recap: Important rankings (computing time)

Algo	Simu1 p=50 r=3	Simu1 p=50 r=5	Simu1 p=100 r=3	Simu1 p=100 r=5	Simu1 p=500 r=3	Simu1 p=500 r=5	Simu2 p=50 r=3	Simu2 p=50 r=5	Simu2 p=100 r=3	Simu2 p=100 r=5	Simu2 p=500 r=3	Simu2 p=500 r=5	mean rank
A_50	1.0	7.0	3.0	3.0	5.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	2.25
A_20	5.0	4.0	2.0	1.0	1.0	4.0	3.0	4.0	2.0	1.0	4.0	4.0	2.92
A_30	2.0	6.0	4.0	2.0	2.0	3.0	4.0	3.0	3.0	5.0	3.0	3.0	3.33
A_40	4.0	5.0	5.0	4.0	4.0	2.0	2.0	2.0	5.0	3.0	2.0	2.0	3.33
A_10	3.0	8.0	1.0	5.0	3.0	5.0	5.0	5.0	4.0	4.0	5.0	5.0	4.42
B_50	8.0	1.0	8.0	7.0	7.0	7.0	6.0	6.0	8.0	7.0	7.0	6.0	6.5
B_40	7.0	2.0	10.0	8.0	8.0	8.0	7.0	7.0	10.0	8.0	8.0	7.0	7.5
B_30	6.0	3.0	9.0	9.0	9.0	9.0	9.0	8.0	11.0	9.0	9.0	8.0	8.25
A_1	10.0	11.0	6.0	11.0	6.0	6.0	8.0	10.0	6.0	11.0	6.0	9.0	8.33
B_20	9.0	9.0	7.0	6.0	10.0	10.0	10.0	9.0	7.0	6.0	10.0	10.0	8.58
B_10	11.0	10.0	11.0	10.0	11.0	11.0	11.0	10.0	9.0	10.0	11.0	11.0	10.5
B_1	12.0	12.0	12.0	12.0	12.0	12.0	12.0	10.0	12.0	12.0	12.0	12.0	11.83

Algorithms	breast_cancer r=3	breast_cancer r=5	cal_housing r=3	cal_housing r=5	bio_train r=3	bio_train r=5	mean rank
A_40	1.0	3.0	2.0	2.0	1.0	2.0	1.83
A_50	2.0	1.0	1.0	5.0	2.0	1.0	2.0
A_30	3.0	2.0	3.0	1.0	3.0	3.0	2.5
A_20	4.0	4.0	5.0	3.0	4.0	4.0	4.0
A_10	5.0	5.0	4.0	4.0	5.0	5.0	4.67
A_1	6.0	6.0	6.0	6.0	6.0	6.0	6.0
B_50	8.0	7.0	7.0	7.0	10.0	7.0	7.67
B_40	7.0	8.0	8.0	10.0	7.0	8.0	8.0
B_30	9.0	9.0	9.0	8.0	8.0	9.0	8.67
B_20	10.0	10.0	10.0	11.0	9.0	10.0	10.0
B_10	11.0	11.0	11.0	9.0	11.0	11.0	10.67
B_1	12.0	12.0	11.0	12.0	12.0	12.0	11.83

References

[1] A. Balsubramani, S. Dasgupta, Y. Freund, The fast convergence of incremental PCA, Advances in Neural Information Processing Systems 26 (2013) 3174-3182.

[2] J.P. Benzécri, Approximation stochastique dans une algèbre normée non commutative, Bulletin de la Société Mathématique de France 97 (1969) 225-241.

[3] O. Brandière, Un algorithme du gradient pour l'analyse en composantes principales, Prépublication 14, Université de Marne-la-Vallée (1994).

[4] O. Brandière, Un algorithme pour l'analyse en composantes principales, Comptes Rendus de l'Académie des Sciences 321 (série I) (1995) 233-236.

[5] O. Brandière, Some pathological traps for stochastic approximation, SIAM J. Control Optim. 36 (4) (1998) 1293-1314.

[6] O. Brandière, M. Duflo, Les algorithmes stochastiques contournent-ils les pièges ? Annales de l'Institut Henri Poincaré, Probabilités et Statistique 32 (1996) 395-427.

[7] T. Budavari, V. Wild, a.s. Szalay, L. Dobos, C.W. Yip, Reliable eigenspectra for new generation surveys, Monthly Notices of the Royal Astronomical Society 394 (2009) 1496-1502.

[8] H. Cardot, P. Cénac, J.M. Monnez, A fast and recursive algorithm for clustering large datasets with k-medians, Computational Statistics and Data Analysis 56 (2012) 1434-1449.

[9] H. Cardot, D. Degras, Online principal component analysis in high dimension: which algorithm to choose ? International Statistical Review 86 (2018) 29-50.

[10] J.D. Carroll, A generalization of canonical correlation analysis to three or more sets of variables, Proceedings of the 76th Convention American. Psychological Association (1968) 227-228.

[11] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, R. Harshman, Indexing by latent semantic analysis, Journal of the American Society for Information Science 41 (6) (1990) 391-407.

[12] K. Duarte, J.M. Monnez, E. Albuisson, Sequential linear regression with online standardized data, PLoS ONE, Public Library of Science (2018) 1-27, https://doi.org/10.1371/journal.pone.0191186.

[13] M. Duflo, Random Iterative Models, Applications in Mathematics, 34, Springer-Verlag, Berlin, 1997.

[14] M. Greenacre, J. Blasius, ed., Multiple Correspondence Analysis and Related Methods, Chapman & Hall/CRC, New-York, 2006.

[15] J. Karhunen, E. Oja, New methods for stochastic approximation of truncated Karhunen-Loève expansions, Proceedings of the 6th International Conference on Pattern Recognition, IEEE Computer Society Press (1982) 550-553.

[16] T.P. Krasulina, Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices, Automation and Remote Control 2 (1970) 215-221.

[17] C.D. Manning, P. Raghavan P, H. Schütze, Matrix decompositions and latent semantic indexing, Introduction to Information Retrieval, Cambridge University Press, Cambridge, 2008.

[18] A. Markos, A.I. D'Enza, Incremental generalized canonical correlation analysis, Analysis of Large and Complex Data, Wilhem AFX, Kestler HA (eds), Studies in Classification, Data Analysis and Knowledge Organization, Springer, 2016, 185-194.

[19] E. Oja, A simplified neuron model as a principal component analyser, Journal of Mathematical Biology 15 (1982) 267-273.

[20] E. Oja E, J. Karhunen, On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix, Journal of Mathematical Analysis and Applications 106 (1985) 69-84.

[21] J. Pagès, Multiple Factor Analysis by Example Using R, The R Series, Chapman & Hall/CRC, London, 2014.

[22] H. Robbins, D. Siegmund, A convergence theorem for nonnegative almost supermartingales and some applications, Optimizing Methods in Statistics, J.S. Rustagi (ed.), Academic Press, New York, 1971, 233-257.

[23] P. Sanguansat, Principal Component Analysis-Engineering Applications, InTech, Rijeka, 2012.

[24] P. Yang, C.J. Hsieh, J.L. Wang, History PCA: a new algorithm for streaming PCA, arXiv:1802.05447v1 (2018).