Convergence of a Normed Eigenvector Stochastic Approximation Process and Application to Online Principal Component Analysis of a Data Stream

Jean-Marie Monnez[1,4,5*], Abderrahman Skiredj[2,3**]

1 Université de Lorraine, CNRS, Inria[1], IECL[2], F-54000 Nancy, France
2 Université de Lorraine, CNRS, IECL, F-54000 Nancy, France
3 Ecole Nationale Supérieure des Mines de Nancy[3], F-54000 Nancy, France
4 CHRU Nancy, INSERM, CIC[4], Plurithématique, F-54000 Nancy, France
5 Université de Lorraine, IUT Nancy-Charlemagne, F 54000 Nancy, France

* Corresponding author, jean-marie.monnez@univ-lorraine.fr **abderrahmanskiredj@mines-nancy.org

Abstract

Many articles have been devoted to the problem of estimating recursively the eigenvectors corresponding to eigenvalues in decreasing order of the expectation of a random matrix A using an i.i.d. sample of it. The present study makes the following contributions: the convergence of a normed process having the same formal definition as that of Oja is proved under more general assumptions, the random matrices used at each step are not supposed i.i.d.; at each step, a data mini-batch or all the data up to the current step can be taken into account without storing them; three types of processes are studied and applied to online principal component analysis of a data stream, assuming that data are realizations of a random vector Z whose expectation is unknown and must be estimated online, as well as possibly the metric used when it depends on unknown characteristics of Z.

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

1 Introduction

Data stream factorial analysis is defined as the factorial analysis of data that arrive continuously such as process control data, web data, telecommunication data, medical data, financial data,.... Recursive stochastic algorithms can be used for observations arriving sequentially to estimate principal components of a principal component analysis (PCA), the estimations of which are updated by each new arriving observation vector. 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 than with other methods can be taken into account during the same duration of time.

Define first some notations. Let Q be a positive definite symmetric $p \times p$ matrix called metric in $\mathbb{R}^p$, $\langle .,. \rangle$ and $\|.\|$ respectively the inner product and the norm induced by Q: $\langle x, y \rangle = x'Qy$, x' denoting the transpose of the column vector x. Remind that a $p \times p$ matrix A is Q-symmetric if $(QA)' = QA$; then A has p real eigenvalues and there exists a Q-orthonormal (i.e. orthonormal with respect to the metric Q) basis of $\mathbb{R}^p$ composed of eigenvectors of A. The norm of a matrix A is the spectral norm denoted $\|A\|$.

Consider the following model: suppose that p quantitative variables are observed on individuals (p may be very large); data vectors in $\mathbb{R}^p$ are thus obtained; considering that z_n is observed at time n (or more generally that several observations, a data mini-batch, are made at time n), there is a sequence of data vectors $z_1, ..., z_n, ...$; assume that, for $n \geq 1$, z_n is a realization of a random variable Z_n defined on a probability space (Ω, A, P) and that $(Z_1, ..., Z_n, ...)$ is an i.i.d sample of a random vector Z. Let θ be the expectation of Z and C its covariance matrix which are unknown in the case of a data stream. Define a metric M in $\mathbb{R}^p$. Recall briefly the principal component analysis (PCA) algorithm of the random vector Z. At step l of PCA is determined a linear combination $c_l'Z$ of the components of Z, called l^{th} principal component, uncorrelated with the previous ones and of maximum variance, under the normalization constraint $c_l'M^{-1}c_l = 1$; c_l is a M^{-1}-unit eigenvector, i.e. of norm 1 with respect to M^{-1}, of MC corresponding to its l^{th} largest eigenvalue

[1] Inria, project-team BIGS
[2] Institut Elie Cartan de Lorraine, BP 239, F-54506 Vandœuvre-lès-Nancy, France
[3] Campus Artem, CS 14234, 92 Rue Sergent Blandan, F-54042 Nancy, France
[4] Centre d'Investigation Clinique

λ_l. For $l = 1, ..., r$, a M-unit 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 use standardized data, i.e. observations of $M(Z-\theta)$, and the identity metric. But the expectation θ and the variances of the components of Z are usually unknown and only raw data are observed. One application of this article is to recursively estimate the c_l or the u_l using stochastic approximation processes.

Many articles have been devoted to this problem when supposing M and θ known or more generally to the problem of estimating eigenvectors and eigenvalues in decreasing order of the expectation B of a random matrix, using an i.i.d. sample of it. See for example the well-known algorithms of Benzécri [1], Krasulina [2], Karhunen and Oja [3], Oja and Karhunen [4], Brandière [5],[6], Brandière and Duflo [7]. Recall the normed process studied in [3][4]:

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

with $E[B_n|T_n] = B$, $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$, T_n the σ-field generated by the events before time n.

This work makes the following contributions. The convergence of this process is proved under more general assumptions: the random matrices B_n are not supposed i.i.d.; this is applied to online estimation of principal components of PCA of a random vector Z, when its expectation is unknown, as well as possibly the metric used, and must be estimated online, and a data mini-batch or all the data up to the current step can be taken into account at each step without storing them, thus using all the information contained in the previous data.

More precisely, let Q be a metric in $\mathbb{R}^p$ and B a Q-symmetric matrix. In the next section, the almost sure (a.s.) convergence of normed processes to eigenvectors of B is studied. Three cases are considered:

- $E[B_n|T_n]$ converges a.s. to B;
- $B_n = \omega_{1n} B_n^1 + \omega_{2n} B_n^2$ with $\omega_{1n} + \omega_{2n} = 1$, B_n^2 is T_n-measurable, $E[B_n^1|T_n]$ and B_n^2 converge a.s. to B;
- B_n converges a.s. to B.

For each case, firstly a theorem of a.s. convergence of (X_n) to a unit eigenvector of B corresponding to its largest eigenvalue is proved with, in the first case, a method following that of [8] (in the case of PCA) under more general assumptions, a corollary in the second case and another method of proof in the third case; secondly, using arguments of exterior algebra, the convergence of processes $\left(X_n^i\right)$, $i = 1, .., r$ of the same type, obtained by Gram-Schmidt orthonormalization, to unit eigenvectors corresponding to eigenvalues of B in decreasing order is proved as a corollary.

Then, in the following section, the whole results are applied to online estimation of principal components of a PCA. In order to reduce computing time, particularly in the case of a data stream, and to avoid possibly numerical explosions, we propose:

a) to estimate the eigenvectors a_l of the symmetric $p \times p$ matrix $B = M^{\frac{1}{2}} C M^{\frac{1}{2}}$; then the orthonormalization is computed with respect to I; estimates of c_l and u_l can be obtained from that of a_l;

b) to replace Z_n by $Z_n - m$, m being an estimation of $E[Z]$ computed in a preliminary phase with a small number of observations e.g. 1000;

c) to use a data mini-batch at step n or all the observations up to step n without storing them.

Three cases are studied: at each step are taken into account

- a data mini-batch,
- or all the observations up to this step with different weights for observations in the past, which are not stored, and observations at this step,
- or all the observations up to this step with uniform weights.

The paper ends with a brief conclusion.

2 Convergence of a normed process

Let (B_n) be a sequence of random $p \times p$ matrices, B a $p \times p$ matrix, (a_n) a sequence of positive numbers, X_1 a random variable of norm 1 in $\mathbb{R}^p$ independent from the sequence of random matrices (B_n) and (X_n) a stochastic process in $\mathbb{R}^p$ recursively defined at step n by:

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

2.1 First case

2.1.1 Theorem of almost sure convergence

Suppose B_n not T_n-measurable (B_n is T_{n+1}-measurable). Make the following assumptions:

(H1a) B is Q-symmetric.

(H1b) B has distinct eigenvalues: $\lambda_1 > \lambda_2 > ... > \lambda_p$. Denote V_i a unit eigenvector of B corresponding to $\lambda_i, i = 1, .., p$.

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

(H2b) $E\left[\sum_{n=1}^{\infty} a_n ||E[B_n|T_n] - B||\right] < \infty \quad a.s.$

(H3) $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$.

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

Theorem 1 *Suppose assumptions H1a,b,H2a,b,H3 hold. Then :*

1) Almost surely, U_n converges to one of the eigenvalues of B; on $E_j = \{U_n \longrightarrow \lambda_j\}$, X_n converges to V_j or $-V_j$, $\sum_{n=1}^{\infty} a_n (\lambda_j - U_n)$ and $\sum_{n=1}^{\infty} a_n (\lambda_j - W_n)$ converge.

2) If moreover on $\overset{p}{\underset{j=2}{\cup}} E_j$, $\liminf E\left[\langle B_n X_n, V_1 \rangle^2 \ |T_n\right] > 0$ a.s., then $P(E_1) = 1$.

Let us state two lemmas of Duflo [8] used in the proof.

Lemma 2 *Let (M_n) be a square-integrable martingale adapted to the filtration (T_n) and $(\langle M \rangle_n)$ its increasing process defined by:*

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

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.

Lemma 3 *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).*
Suppose on the set Γ:
1) For every integer n, $Z_{n+1} = Z_n(1 + \delta_n) + \gamma_n \epsilon_{n+1}$;
2) (Z_n) is bounded;
3) $\sum_{n=1}^{\infty} \delta_n^2 < \infty$, $\delta_n \geq 0$ 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;
4) 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, $P(\Gamma) = 0$.

Proof of Theorem 1

Part 1: expression of $\mathbf{X}_{n+1}$

Under H2a, as $||(I + a_n B_n) X_n||^2 = 1 + 2a_n W_n + a_n^2 \|B_n X_n\|^2$:

$$\frac{1}{||(I + a_n B_n) X_n||} = 1 - a_n W_n - \frac{1}{2} a_n^2 \|B_n X_n\|^2 + \alpha_n, \quad \alpha_n = O(a_n^2).$$

$$
\begin{aligned}
X_{n+1} &= (I + a_n B_n) \left(1 - a_n W_n - \frac{1}{2} a_n^2 \left\| B_n X_n \right\|^2 + \alpha_n \right) X_n \\
&= \left(I + a_n \left(B_n - W_n I \right) + a_n \beta_n \right) X_n, \text{ with} \\
\beta_n &= -a_n W_n B_n - \frac{1}{2} a_n \left\| B_n X_n \right\|^2 I - \frac{1}{2} a_n^2 B_n \left\| B_n X_n \right\|^2 + a_n^{-1} \alpha_n I + \alpha_n B_n.
\end{aligned}
$$

$$
\begin{aligned}
X_{n+1} &= \left(I + a_n \left(B - U_n I \right) + a_n \Gamma_n \right) X_n, \text{ with} \\
\Gamma_n &= (B_n - B) - \langle X_n, (B_n - B) X_n \rangle I + \beta_n.
\end{aligned}
$$

Part 2: convergence of $\mathbf{U}_n$

$$
\begin{aligned}
E[U_{n+1}|T_n] &= E \left[\langle \left(I + a_n \left(B - U_n I \right) + a_n \Gamma_n \right) X_n, B \left(I + a_n \left(B - U_n I \right) + a_n \Gamma_n \right) X_n \rangle \ |T_n \right] \\
&= U_n + 2 a_n \langle \left(B - U_n I \right) X_n, B X_n \rangle + 2 a_n E \left[\langle \Gamma_n X_n, B X_n \rangle \ |T_n \right] + a_n^2 \eta_n \\
\text{with} \quad \eta_n &= \langle \left(B - U_n I \right) X_n, B \left(B - U_n I \right) X_n \rangle + 2 E \left[\langle \Gamma_n X_n, B \left(B - U_n I \right) X_n \rangle \ |T_n \right] \\
&\quad + E \left[\langle \Gamma_n X_n, B \Gamma_n X_n \rangle \ |T_n \right] \text{ a.s.}
\end{aligned}
$$

As $\|X_n\| = 1$, we have with $\mu_n = 2 a_n \langle E \left[\Gamma_n \ |T_n \right] X_n, B X_n \rangle + a_n^2 \eta_n$:

$$
\begin{aligned}
\langle \left(B - U_n I \right) X_n, B X_n \rangle &= \left\| B X_n \right\|^2 - U_n^2 = \left\| B X_n - U_n X_n \right\|^2 \geqslant 0. \\
E[U_{n+1}|T_n] &\geq U_n + \mu_n \text{ a.s.} \\
E[U_{n+1} - \sum_{i=1}^{n} \mu_i \ |T_n] &\geq U_n - \sum_{i=1}^{n-1} \mu_i \text{ a.s.}
\end{aligned}
$$

Prove the convergence of the submartingale $U_n - \sum_{i=1}^{n-1} \mu_i$. By H2a:

$$
\begin{aligned}
\|\beta_n\| &\leq \frac{3}{2} a_n \left\| B_n \right\|^2 + \frac{1}{2} a_n^2 \left\| B_n \right\|^3 + a_n^{-1} \left| \alpha_n \right| + \left| \alpha_n \right| \left\| B_n \right\| = O\left(a_n \right) ; \\
\|\Gamma_n\| &\leq 2 \left\| B_n - B \right\| + \left\| \beta_n \right\| = O\left(1 \right); \\
\|\eta_n\| &\leq 4 \left\| B \right\|^3 + 4 \left\| B \right\|^2 \left\| E \left[\Gamma_n \ |T_n \right] \right\| + \left\| B \right\| E \left[\left\| \Gamma_n \ \right\|^2 |T_n \right] = O\left(1 \right).
\end{aligned}
$$

$$
\left\| E \left[\Gamma_n \ |T_n \right] \right\| \leq 2 \left\| E \left[B_n \ |T_n \right] - B \right\| + \left\| E \left[\beta_n \ |T_n \right] \right\| \text{ a.s.}
$$

By H2b and H3:

$$
\begin{aligned}
E \left[\left| \sum_{i=1}^{n-1} \mu_i \right| \right] &\leq 4 \left\| B \right\| E \left[\sum_{i=1}^{\infty} a_i \left\| E \left[B_i \ |T_i \right] - B \right\| \right] + 2 \left\| B \right\| E \left[\sum_{i=1}^{\infty} a_i \left\| E \left[\beta_i \ |T_i \right] \right\| \right] + E \left[\sum_{i=1}^{\infty} a_i^2 \left\| \eta_i \right\| \right] \\
&< \infty.
\end{aligned}
$$

By Doob lemma the submartingale $U_n - \sum_{i=1}^{n-1} \mu_i$ converges a.s. to an integrable random variable. As $\sum_{i=1}^{n-1} \mu_i$ converges, U_n converges a.s.

Part 3: convergence of $\mathbf{X}_n^j = \langle \mathbf{X}_n, \mathbf{V}_j \rangle$

Let $\Gamma_n^j = \langle \Gamma_n X_n, V_j \rangle$.

$$
X_{n+1}^j = \langle \left(I + a_n (B - U_n I) + a_n \Gamma_n \right) X_n, V_j \rangle = X_n^j \left(1 + a_n (\lambda_j - U_n) \right) + a_n \Gamma_n^j.
$$

$$
\begin{aligned}
(X_{n+1}^j)^2 &= (X_n^j)^2 (1 + 2 a_n (\lambda_j - U_n)) + a_n^2 (\lambda_j - U_n)^2 (X_n^j)^2 \\
&\quad + 2 a_n (1 + a_n (\lambda_j - U_n)) X_n^j \Gamma_n^j + a_n^2 (\Gamma_n^j)^2 \\
&= (X_n^j)^2 \left(1 + 2 a_n (\lambda_j - U_n) \right) + a_n^2 \left((\lambda_j - U_n) X_n^j + \Gamma_n^j \right)^2 + 2 a_n X_n^j \Gamma_n^j \\
&= (X_1^j)^2 + 2 \sum_{l=1}^{n} a_l (\lambda_j - U_l) (X_l^j)^2 + \sum_{l=1}^{n} a_l^2 \left((\lambda_j - U_l) X_l^j + \Gamma_l^j \right)^2 + 2 \sum_{l=1}^{n} a_l X_l^j \Gamma_l^j \\
&= \left(X_1^j \right)^2 + (1) + (2) + (3).
\end{aligned}
$$

Study the convergence of the terms (2), (3), (1) of this decomposition.

(i) (2) $= \sum_{l=1}^{n} a_l^2 \left((\lambda_j - U_l) X_l^j + \Gamma_l^j \right)^2$ converges a.s. by H2a and H3.

(ii) Consider now (3).

$$\sum_{l=1}^{n} a_l X_l^j \Gamma_l^j = \sum_{l=1}^{n} a_l X_l^j (\Gamma_l^j - E[\Gamma_l^j | T_l]) + \sum_{l=1}^{n} a_l X_l^j E[\Gamma_l^j | T_l].$$

$$\begin{aligned} \sum_{l=1}^{n} a_l |X_l^j E[\Gamma_l^j | T_l]| &\leq \sum_{l=1}^{n} a_l \left| \langle E[\Gamma_l | T_l] X_l, V_j \rangle \right| \\ &\leq \sum_{l=1}^{n} a_l ||E[\Gamma_l | T_l]|| \leq \sum_{l=1}^{n} a_l \left(2||E[B_l|T_l] - B|| + ||E[\beta_l | T_l]|| \right). \end{aligned}$$

By H2a,b and H3, as $||\beta_n|| = O(a_n)$, $\sum_{l=1}^{+\infty} a_l X_l^j E[\Gamma_l^j | T_l]$ is convergent.

Let $M_n^j = \sum_{l=1}^{n-1} a_l X_l^j (\Gamma_l^j - E[\Gamma_l^j | T_l])$; (M_n^j) is a square-integrable martingale adapted to the filtration (T_n); denote $(\langle M^j \rangle_n)$ its increasing process. By H2a:

$$\begin{aligned} \langle M^j \rangle_{n+1} - \langle M^j \rangle_n &= E[(M_{n+1}^j - M_n^j)^2 | T_n] = a_n^2 E[(X_n^j)^2 (\Gamma_n^j - E[\Gamma_n^j | T_n])^2 | T_n] \\ &\leq a_n^2 \left(E[(\Gamma_n^j)^2 | T_n] - (E[\Gamma_n^j | T_n])^2 | T_n] \right) \leq a_n^2 E[(\|\Gamma_n\|^2 | T_n] \end{aligned}$$

is the general term of a convergent and uniformly bounded series; thus by Lemma 2 (M_n^j) converges a.s. to a finite random variable.

Therefore (3) converges a.s.

(iii) Consider finally (1). Let ω 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. Thus there exists $A > 0$ such that:

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

Then for $L \neq \lambda_j$, $2 \sum_{l=N}^{n} a_l |\lambda_j - U_l| (X_l^j)^2$ converges.

It follows from the convergence of (1), (2) and (3) that for $L \neq \lambda_j$, $(X_n^j)^2$ converges a.s.

Part 4: convergence of $\mathbf{X}_n$

If the limit of U_n is different from λ_j, then by convergence of (1) in step 3, $\sum_{l=1}^{\infty} a_l (X_l^j)^2 < \infty$ and X_n^j converges a.s. to 0. As $||X_n|| = 1$, this can not be true for every j.

Thus the limit of U_n is one of the eigenvalues of B, λ_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 $E_i = \{U_n \longrightarrow \lambda_i\}$ (first assertion of Theorem 1).

Consider now the decomposition:

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

(i) Using the decomposition of $(X_n^i)^2$ in step 3, the convergence of $(X_n^i)^2$ and of (2) and (3) yields that $\sum_{n=1}^{\infty} a_n(\lambda_i - U_n)$ converges a.s. (second assertion of Theorem 1).

(ii) By H2b: $\sum_{n=1}^{\infty} a_n \langle X_n, (E[B_n|T_n] - B)X_n\rangle$ converges a.s.

(iii) 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:

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

is the general term of a convergent and uniformly bounded series. Thus (M_n) converges a.s. to a finite random variable.

Therefore by (i), (ii) and (iii), $\sum_{n=1}^{\infty} a_n(\lambda_i - W_n)$ converges (third assertion of Theorem 1).

Part 5: convergence of X_n to $\pm V_1$

Suppose $i > 1$.

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

In the following, apply Lemma 3 to the sequence $\left(X_n^1\right)$ on $E_i = \{X_n \longrightarrow V_i\}, i > 1$, with $\gamma_n = a_n$, $\delta_n = a_n(\lambda_1 - U_n) \geqslant 0$, $b_n = a_n(\lambda_1 - \lambda_i) > 0$, $\epsilon_{n+1} = \langle \Gamma_n X_n, V_1\rangle$.

(i) X_n^1 is bounded.

(ii) $\sum_{n=1}^{\infty} a_n^2 < \infty$, $\sum_{n=1}^{\infty} a_n^2(\lambda_1 - U_n)^2 < \infty$ by H3, $\sum_{n=1}^{\infty} a_n(\lambda_1 - \lambda_i) = \infty$, $\sum_{n=1}^{\infty} a_n(\lambda_i - U_n)$ converges a.s.

(iii) Consider the decomposition:

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

By H2a, there exists a positive number c such that a.s.:

$$\begin{aligned}&E[(\langle \Gamma_n X_n, V_1\rangle - \langle B_n X_n, V_1\rangle)^2 \ |T_n] \\ \leq\ & 2(X_n^1)^2 E[(\lambda_1 + \langle (B_n - B)X_n, X_n\rangle)^2 \ |T_n] + 2E[(\langle \beta_n X_n, V_1\rangle)^2 \ |T_n] \\ \leq\ & c(X_n^1)^2 + 2E[||\beta_n||^2 \ |T_n] \underset{n\to+\infty}{\longrightarrow} 0\end{aligned}$$

since $||\beta_n|| = O(a_n)$ and $X_n^1 \underset{n\to+\infty}{\longrightarrow} 0$. Likewise:

$$E[(\langle \Gamma_n X_n, V_1\rangle - \langle B_n X_n, V_1\rangle)\langle B_n X_n, V_1\rangle \ |T_n] \underset{n\to+\infty}{\longrightarrow} 0.$$

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 3, under (i), (ii), (iii), $P(E_i) = 0$, $i > 1$.

Then, $P(E_1) = 1$ (fourth assertion of Theorem 1). ■

2.1.2 Simultaneous estimation of several eigenvectors

In this part, for $i = 1, ..., r$, X_n^i does not represent the i^{th} component of X_n, but a random variable in $\mathbb{R}^p$ recursively defined by:

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

$(X_{n+1}^1, ..., X_{n+1}^r)$ is obtained by Gram-Schmidt orthonormalization of $(Y_{n+1}^1, ..., Y_{n+1}^r)$.

Corollary 4 *Suppose assumptions H1a,b, H2a,b and H3 hold.*

1) For $i = 1, ..., r$, almost surely X_n^i converges to one of the eigenvectors of B.

2) If moreover, for $i = 1, ..., r$, almost surely on $\bigcup_{j=i+1}^{p} \left\{ X_n^i \longrightarrow \pm V_j \right\}$, $\liminf E[\left\langle B_n X_n^i, V_i \right\rangle^2 \; |T_n] > 0$, then X_n^i converges a.s. to V_i or $-V_i$, $\sum_{n=1}^{\infty} a_n \left| \lambda_i - \left\langle B X_n^i, X_n^i \right\rangle \right|$ and $\sum_{n=1}^{\infty} a_n (\lambda_i - \left\langle B_n X_n^i, X_n^i \right\rangle)$ converge a.s.

Before the proof, some concepts of exterior algebra are reminded.

Let $(e_1, ..., e_p)$ be a basis of $\mathbb{R}^p$. For $r \le p$, 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 $e_{i_1} \wedge e_{i_2} \wedge ... \wedge e_{i_r}$, $i_1 < i_2 < ... < i_r \in \{1, ..., p\}$.

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 by:

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

G_r being the set of permutations σ of $\{k_1, ..., k_r\}$ and $s(\sigma)$ the number of inversions of σ.

Let $\|.\|$ be the associated norm. Note that if $x_1, ..., x_r$ are Q-orthogonal, $||x_1 \wedge ... \wedge x_r|| = \prod_{i=1}^{r} ||x_i||$, and if $(e_1, ..., e_p)$ is a Q-orthonormal basis of $\mathbb{R}^p$, then the set of the C_p^r exterior products $e_{i_1} \wedge ... \wedge e_{i_r}$ is an orthonormal basis of ${}^r\Lambda\mathbb{R}^p$.

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

$${}^{rj}U(x_1 \wedge ... \wedge x_r) = \sum_{1 \le i_1 < i_2 < ... < i_j \le r} x_1 \wedge ... \wedge U x_{i_1} \wedge ... \wedge U x_{i_j} \wedge ... \wedge x_r.$$

.

$$\text{For j=1}, {}^{r1}U(x_1 \wedge ... \wedge x_r) = \sum_{i=1}^{r} x_1 \wedge ... \wedge U x_i \wedge ... \wedge x_r.$$

Let ${}^r\Lambda U$ be the endomorphism ${}^{rr}U$ such that:

$${}^r\Lambda U(x_1 \wedge ... \wedge x_r) = U x_1 \wedge U x_2 \wedge ... \wedge U x_r.$$

c) The following properties hold:

(i) Suppose that the eigenvalues $\lambda_1 > ... > \lambda_p$ of U are distinct and denote for $j = 1, ..., r$, V_j an eigenvector corresponding to λ_j. Then the C_p^r vectors $V_{i_1} \wedge ... \wedge V_{i_r}$, $1 \le i_1 < ... < i_r \le p$, are eigenvectors of ${}^{r1}U$ respectively corresponding to the eigenvalues $\lambda_{i_1} + ... + \lambda_{i_r}$.

(ii) ${}^r\Lambda(I + U) = I + \sum_{j=1}^{r} {}^{rj}U$.

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

Proof

Part 1

For $i = 1, ..., r$, it follows from the orthogonality of $T_n^1, ..., T_n^i$ that:

$$||T_{n+1}^1 \wedge ... \wedge T_{n+1}^i|| = \prod_{l=1}^{i} ||T_{n+1}^l||.$$

Let ${}^iX_{n+1} = X^1_{n+1} \wedge ... \wedge X^i_{n+1}$ and $D^i_n = {}^{i1}B_n + \sum\limits_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n$. Then:

$$
\begin{aligned}
{}^iX_{n+1} &= \frac{T^1_{n+1} \wedge ... \wedge T^i_{n+1}}{||T^1_{n+1} \wedge ... \wedge T^i_{n+1}||} = \frac{Y^1_{n+1} \wedge ... \wedge Y^i_{n+1}}{||Y^1_{n+1} \wedge ... \wedge Y^i_{n+1}||} = \frac{{}^i\Lambda(I + a_n B_n)\ {}^iX_n}{||{}^i\Lambda(I + a_n B_n)\ {}^iX_n||} \\
&= \frac{\left(I + a_n\ {}^{i1}B_n + \sum\limits_{j=2}^{i} a_n^j\, {}^{ij}B_n\right)\ {}^iX_n}{\left\|\left(I + a_n\ {}^{i1}B_n + \sum\limits_{j=2}^{i} a_n^j\, {}^{ij}B_n\right)\ {}^iX_n\right\|} = \frac{\left(I + a_n\ D^i_n\right)\ {}^iX_n}{||\left(I + a_n\ D^i_n\right)\ {}^iX_n||}.
\end{aligned}
$$

As $||{}^{ij}B_n|| \leq c(i)||B_n||^j$, assumptions H2a and H3 yield that there exists $b_1 > 0$ such that for all n, $||D^i_n|| \leq b_1$.

Moreover, as $U \mapsto^{i1} U$ is a linear application, assumptions H2a,b and H3 yield that:

$$
\begin{aligned}
E\left[\sum_{n=1}^{\infty} a_n ||E[D^i_n|T_n] -\ {}^{i1}B||\right] &= E\left[\sum_{n=1}^{\infty} a_n \left\|E\left[{}^{i1}B_n - {}^{i1}B + \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n\ |T_n\right]\right\|\right] \\
\leq\ & E\left[\sum_{n=1}^{\infty} a_n \left(||{}^{i1}E[B_n - B\ |T_n]|| + \sum_{j=2}^{i} a_n^{j-1} E[||{}^{ij}B_n||\ |T_n]\right)\right] \\
\leq\ & c(i)E\left[\sum_{n=1}^{\infty} a_n \left(||E\left[B_n|T_n\right] - B]|| + \sum_{j=2}^{i} a_n^{j-1} E[||B_n||^j\ |T_n]\right)\right] < \infty.
\end{aligned}
$$

Applying first assertion of Theorem 1 yields that almost surely,
${}^iX_n$ converges to a unit eigenvector $\pm V_{j_1} \wedge ... \wedge V_{j_i}$ of ${}^{i1}B$,
$\sum\limits_{n=1}^{\infty} a_n(\lambda_{j_1} + ... + \lambda_{j_i} - \langle {}^{i1}B\ {}^iX_n, {}^iX_n\rangle)$ and
$\sum\limits_{n=1}^{\infty} a_n(\lambda_{j_1} + ... + \lambda_{j_i} - \langle D^i_n\ {}^iX_n, {}^iX_n\rangle)$ converge.
Moreover by H2a and H3, $\sum\limits_{n=1}^{\infty} a_n(\lambda_{j_1} + ... + \lambda_{j_i} - \langle {}^{i1}B_n\ {}^iX_n, {}^iX_n\rangle)$ converges a.s.

Part 2

Suppose that for $k = 1, ..., i-1$, $X^k_n \underset{n\to+\infty}{\longrightarrow} \pm V_k$, which is verified for $k = 1$, and prove that $X^i_n \underset{n\to+\infty}{\longrightarrow} \pm V_i$.

1) Prove that there exists $j > i-1$ such that $X^i_n \underset{n\to+\infty}{\longrightarrow} \pm V_j$. Suppose that there exists $k \in \{1, ..., i-1\}$ such that, for $l = 1, ..., i$, $V_{j_l} \neq \pm V_k$; then, for $l = 1, ..., i$, $\langle X^k_n, V_{j_l}\rangle \underset{n\to+\infty}{\longrightarrow} 0$ and $\langle X^1_n \Lambda ... \Lambda X^i_n, V_{j_1}\Lambda ... \Lambda V_{j_i}\rangle \underset{n\to+\infty}{\longrightarrow} 0$, a contradiction. Therefore for all $k \in \{1, ..., i-1\}$, there exists j_l such that $V_{j_l} = \pm V_k$ and there exists j such that

$$
{}^iX_n = X^1_n \wedge ... \wedge X^i_n \underset{n\to+\infty}{\longrightarrow} \pm V_1 \Lambda ... \Lambda V_{i-1} \wedge V_j.
$$

The only term which has a non-zero limit in the development of

$$
\left\langle X^1_n \wedge ... \wedge X^i_n, \pm V_1 \wedge ... \wedge V_{i-1} \wedge V_j \right\rangle,
$$

the limit of which is 1 as $n \longrightarrow \infty$, is $\langle X^1_n, V_1\rangle\langle X^2_n, V_2\rangle ... \langle X^{i-1}_n, V_{i-1}\rangle \left\langle X^i_n, V_j\right\rangle$ obtained for $\sigma = Id$. As for $k = 1, ..., i-1$, $\langle X^k_n, V_k\rangle \underset{n\to+\infty}{\longrightarrow} \pm 1$, then $\left\langle X^i_n, V_j \right\rangle \underset{n\to+\infty}{\longrightarrow} \pm 1$. Therefore $X^i_n \underset{n\to+\infty}{\longrightarrow} \pm V_j$.

2) Prove now that $V_j = \pm V_i$. Suppose $X^i_n \underset{n\to+\infty}{\longrightarrow} \pm V_j \neq \pm V_i$.

Let G_i be the set of permutations σ of $\{1,...,i\}$ with $\sigma = (\sigma(1), ..., \sigma(i))$ and $s(\sigma)$ the number of inversions of σ.

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

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

As for $k = 1, ..., i-1$, $X_n^k \underset{n \to +\infty}{\longrightarrow} \pm V_k$, the only term with a non-zero limit in the development of this conditional expectation is

$$\langle X_n^1, V_1\rangle^2 ... \langle X_n^{i-1}, V_{i-1}\rangle^2 E\left[\langle B_n X_n^i, V_i\rangle^2 \ | T_n\right]$$

and

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

Moreover, by H2a and $\underset{n \longrightarrow \infty}{lim} a_n = 0$:

$$\begin{aligned}& \liminf E\left[\left\langle \sum_{j=2}^{i} a_n^{j-1} \, {}^{ij}B_n(X_n^1 \wedge ... \wedge X_n^i), V_1 \wedge ... \wedge V_i\right\rangle^2 \ | T_n\right] \\ = \ & \liminf E\left[\left(\sum_{j=2}^{i} a_n^{j-1} \left\langle {}^{ij}B_n(X_n^1 \wedge ... \wedge X_n^i), V_1 \wedge ... \wedge V_i\right\rangle\right)^2 \ | T_n\right] = 0.\end{aligned}$$

Then $\liminf E\left[\langle D_n^i(X_n^1 \wedge ... \wedge X_n^i), V_1 \wedge ... \wedge V_i\rangle^2 | T_n\right] > 0$.

Applying second assertion of Theorem 1 yields almost surely:

$$X_n^1 \wedge ... \wedge X_n^i \underset{n \to +\infty}{\longrightarrow} \pm V_1 \wedge ... \wedge V_i, \text{ therefore } X_n^i \underset{n \to +\infty}{\longrightarrow} \pm V_i,$$

$$\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B \ {}^iX_n, {}^iX_n\rangle\right) \text{ and } \sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle D_n^i \ {}^iX_n, {}^iX_n\rangle\right) \text{ converge,}$$

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

Part 3

$$\langle {}^{i1}B \ {}^iX_n, {}^iX_n\rangle = \sum_{k=1}^{i} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle X_n^1, X_n^{\sigma(1)}\rangle ... \langle B X_n^k, X_n^{\sigma(k)}\rangle ... \langle X_n^i, X_n^{\sigma(i)}\rangle.$$

As $X_n^1, ..., X_n^i$ are orthonormal, this sum is equal to

$$\sum_{k=1}^{i} \langle X_n^1, X_n^1\rangle ... \langle B X_n^k, X_n^k\rangle ... \langle X_n^i, X_n^i\rangle = \sum_{k=1}^{i} \langle B X_n^k, X_n^k\rangle.$$

Then, as $\sum\limits_{l=1}^{i} \lambda_l$ is the largest eigenvalue of ${}^{i1}B$:

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

Almost surely, $\sum_{n=1}^{\infty} a_n \left| \sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B^i X_n, {}^i X_n \rangle \right| < \infty$, then $\sum_{n=1}^{\infty} a_n \left| \lambda_i - \left\langle B X_n^i, X_n^i \right\rangle \right| < \infty$.

Likewise, as almost surely $\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \langle {}^{i1}B_n\ {}^i X_n, {}^i X_n \rangle \right)$ converges, then $\sum_{n=1}^{\infty} a_n \left(\lambda_i - \left\langle B_n X_n^i, X_n^i \right\rangle \right)$ converges. ■

2.2 Second case

Consider the same processes (X_n) and $\left(X_n^i\right)$ as in the first case.

Suppose now $B_n = \omega_{1n} B_n^1 + \omega_{2n} B_n^2$ with $\omega_{1n} > 0, \omega_{2n} \geqslant 0, \omega_{1n} + \omega_{2n} = 1, B_n^2$ T_n-measurable.

2.2.1 Theorem of almost sure convergence

Make the following assumptions:

(H2a') There exists a positive number b_1 such that $\sup_n \left\| B_n^1 \right\| < b_1$ a.s.

(H2b') $E\left[\sum_{n=1}^{\infty} a_n \left\| E\left[B_n^1 | T_n \right] - B \right\| \right] < \infty$.

(H2c') B_n^2 T_n-measurable, $B_n^2 \underset{n \to +\infty}{\longrightarrow} B$, $E\left[\sum_{n=1}^{\infty} a_n || B_n^2 - B || \right] < \infty$ a.s.

Theorem 5 *Suppose assumptions H1a,b, H2a',b',c' and H3 hold. Then :*

1) Almost surely, U_n converges to one of the eigenvalues of B; on $E_j = \{U_n \longrightarrow \lambda_j\}$, X_n converges to V_j or $-V_j$, $\sum_{n=1}^{\infty} a_n (\lambda_j - U_n)$ and $\sum_{n=1}^{\infty} a_n (\lambda_j - W_n)$ converge.

2) If moreover $\underset{n \longrightarrow \infty}{lim} \omega_{1n} = \omega_1 > 0$ and on $\bigcup_{j=2}^{p} E_j$, $\liminf E\left[\left\langle B_n^1 X_n, V_1 \right\rangle^2 | T_n \right] > 0$ a.s., then $P(E_1) = 1$.

Proof

Apply Theorem 1.

Under assumptions H2a',b',c', assumptions H2a,b are verified. Thus first part of Theorem 1 holds.

Prove that $\liminf E[\langle B_n X_n, V_1 \rangle^2 \ |T_n] > 0$ a.s. when $\underset{n \longrightarrow \infty}{lim} X_n = \pm V_j \neq \pm V_1$.

$$\begin{aligned} & E\left[\left\langle \left(\omega_{1n} B_n^1 + \omega_{2n} B_n^2 \right) X_n, V_1 \right\rangle^2 | T_n \right] \\ = \; & (\omega_{2n})^2 \left\langle B_n^2 X_n, V_1 \right\rangle^2 + 2\omega_{1n}\omega_{2n} \left\langle B_n^2 X_n, V_1 \right\rangle E\left[\left\langle B_n^1 X_n, V_1 \right\rangle | T_n \right] \\ & + (\omega_{1n})^2 E\left[\left\langle B_n^1 X_n, V_1 \right\rangle^2 | T_n \right] \text{ a.s.} \end{aligned}$$

When $\underset{n \longrightarrow \infty}{lim} X_n = \pm V_j \neq \pm V_1$, $\underset{n \longrightarrow \infty}{lim} \left\langle B_n^2 X_n, V_1 \right\rangle = \pm \left\langle B V_j, V_1 \right\rangle = \pm \lambda_j \langle V_j, V_1 \rangle = 0$. Then:

$\liminf E\left[\left\langle \left(\omega_{1n} B_n^1 + \omega_{2n} B_n^2 \right) X_n, V_1 \right\rangle^2 \ | T_n \right] = (\omega_1)^2 \liminf E\left[\left\langle B_n^1 X_n, V_1 \right\rangle^2 \ | T_n \right] > 0$ a.s. ■

2.2.2 Simultaneous estimation of several eigenvectors

Corollary 6 *Suppose assumptions H1a,b, H2a',b' ,c' and H3 hold. Then:*

1) For $i = 1, ..., r$, almost surely X_n^i converges to one of the eigenvectors of B.

2) If moreover $\underset{n \longrightarrow \infty}{lim} \omega_{1n} = \omega_1 > 0$ and for $i = 1, ..., r$, a.s. on $\bigcup_{j=i+1}^{p} \left\{ X_n^i \longrightarrow \pm V_j \right\}$, $\liminf E\left[\left\langle B_n^1 X_n^i, V_i \right\rangle^2 | T_n \right] > 0$, then almost surely X_n^i converges to $\pm V_i$, $\sum_{n=1}^{\infty} a_n \left| \lambda_i - \left\langle B X_n^i, X_n^i \right\rangle \right|$ and $\sum_{n=1}^{\infty} a_n (\lambda_i - \left\langle B_n X_n^i, X_n^i \right\rangle)$ converge.

It is a direct application of Corollary 4 assumptions of which are verified as proved above.

2.3 Third case

It is assumed in the second case that $\omega_1 > 0$. Now assume $\omega_{1n} = \omega_1 = 0$.

2.3.1 Theorem of almost sure convergence

Recursively define the process $(\widetilde{X}_n)$ such that

$$\widetilde{X}_{n+1} = (I + a_n B_n)\widetilde{X}_n$$

and the process $\left(\widetilde{U}_n\right)$ such that

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

Note that $\frac{\widetilde{U}_n}{\|\widetilde{U}_n\|} = \frac{\widetilde{X}_n}{\|\widetilde{X}_n\|} = X_n$. Make the following assumptions:
(H1c) $\|B\| = \lambda_1$.
(H2c) $\sum_{n=1}^{\infty} a_n \|B_n - B\| < \infty$ a.s.
(H2d) For all n, $I + a_n B_n$ is invertible (especially verified if B_n is non-negative).
(H5) $\widetilde{X}_1$ is an absolutely continuous random variable, independent from $B_1, ..., B_n, ...$

Theorem 7 *Suppose assumptions H1a,b,c, H2c,d, H3 and H5 hold. Almost surely, $\widetilde{U}_n$ converges to a random vector colinear to V_1, therefore X_n converges to $\pm V_1$, $\sum_{n=1}^{\infty} a_n(\lambda_1 - \langle BX_n, X_n\rangle)$ and $\sum_{n=1}^{\infty} a_n |\lambda_1 - \langle B_n X_n, X_n\rangle|$ converge.*

Remark 8 *1) Note that assumption H2a is not required.*
2) Since $\omega \in \Omega$ is fixed throughout the following proof, a_n can be a positive random variable. ■

Lemma 9 *Suppose for all n, (z_n), (α_n), (β_n) and (γ_n) are four sequences of non-negative numbers such that:*

$$\textit{for all } n \geqslant 1,\ z_{n+1} \le z_n(1+\alpha_n) + \beta_n - \gamma_n, \sum_{n=1}^{\infty}\alpha_n < \infty, \sum_{n=1}^{\infty}\beta_n < \infty.$$

Then the sequence (z_n) converges and $\sum_{n=1}^{\infty}\gamma_n < \infty$.

This is a deterministic form of the Robbins-Siegmund lemma [10], whose proof is based on the convergence of the decreasing sequence (u_n):

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

Proof

Let ω be fixed, belonging to $C_1 = \left\{\sum_{n=1}^{\infty} a_n\|B_n - B\| < \infty\right\}$. The writing of ω will be omitted in the following.

Part 1

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

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

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

By assumptions H2c and H3, applying Lemma 9 yields:

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

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

Part 2: convergence of $\widetilde{U}_n^j = \left\langle \widetilde{U}_n, V_j\right\rangle$

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

a) For $j > 1$, as $a_n \underset{n\to+\infty}{\longrightarrow} 0$, there exists $\alpha_n = O\left(a_n\right) > 0$ such that for n sufficiently large:

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

By H2c and as $\left\|\widetilde{U}_n\right\|$ converges, applying Lemma 9 yields:

$$\left|\widetilde{U}_n^j\right| \underset{n\to+\infty}{\longrightarrow} \widetilde{U}^j, \ \sum_{n=1}^{\infty}\alpha_n\left|\widetilde{U}_n^j\right| < \infty. \text{ As } \sum_{n=1}^{\infty} a_n = \infty, \ \widetilde{U}^j = 0.$$

b) For $j = 1$, by H2c and $||\widetilde{U}_n|| \underset{n\to+\infty}{\longrightarrow} \sqrt{\widetilde{U}}$:

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

Now:

$$\begin{aligned}
\widetilde{U}_{n+1}^1 &= \left\langle V_1, \widetilde{U}_{n+1}\right\rangle = \left\langle V_1, \prod_{i=1}^{n}\frac{I + a_i B_i}{1+\lambda_1 a_i}\widetilde{U}_1\right\rangle \underset{n\to+\infty}{\longrightarrow} \left\langle V_1, \prod_{i=1}^{\infty}\frac{I + a_i B_i}{1+\lambda_1 a_i}\widetilde{U}_1\right\rangle \\
&= V_1'QS\widetilde{U}_1 = \widetilde{U}^1 \text{ with } S = \prod_{i=1}^{\infty}\frac{I + a_i B_i}{1+\lambda_1 a_i}.
\end{aligned}$$

As $\widetilde{U}_1$ is absolutely continuous, if $V_1'QS \neq 0$, $P\left(V_1'QS\widetilde{U}_1 = 0 \mid S\right) = 0$, then $P\left(\widetilde{U}^1 = 0\right) = 0$. Prove that $V_1'QS \neq 0$.

Part 3

Denote $C_2 = \left\{ \widetilde{U}_1 \neq 0 \right\}$. Suppose $\omega \in C_1 \cap C_2$.

Under H2c, there exists N such that $\sum\limits_{n=N}^{\infty} a_n ||B_n - B|| < \ln 2$.

$$\begin{aligned} V_1' Q S &= V_1' Q \prod_{i=N}^{\infty} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \prod_{i=1}^{N-1} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \\ &= V_1' Q R \prod_{i=1}^{N-1} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \text{ with } R = \prod_{i=N}^{\infty} \frac{I + a_i B_i}{1 + \lambda_1 a_i}. \end{aligned}$$

Under H2d, $V_1' Q S \neq 0 \Leftrightarrow V_1' Q R \neq 0$.

Denote $C_n = \frac{a_n \|B_n - B\|}{1 + \lambda_1 a_n}$ and $(W_n, n \geqslant N)$ the process $\left(\widetilde{U}_n, n \geqslant N\right)$ with $W_N = V_1$.

As $\|B\| = \lambda_1$, $\|I + a_{i-1} B\| = 1 + \lambda_1 a_{i-1}$. By Part 2, as $W_N = V_1$:

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

As $\sum\limits_{n=N}^{\infty} C_n < \ln 2$, it follows that:

$$\begin{aligned} W_{n+1}^1 &\geqslant 1 - \sum_{i=N}^{n} C_i \prod_{l=N}^{i-1} (1 + C_l) = 1 - \left(\prod_{l=N}^{n} (1 + C_l) - 1 \right) \\ &= 2 - \prod_{l=N}^{n} (1 + C_l) \geqslant 2 - e^{\sum\limits_{l=N}^{n} C_l} \geqslant 2 - e^{\sum\limits_{l=N}^{\infty} C_l} > 0. \end{aligned}$$

By Part 2, W_n^1 converges to $\langle V_1, R V_1 \rangle = V_1' Q R V_1$ which is therefore strictly positive, thus $V_1' Q R \neq 0$.

Part 4: conclusion

It follows that $\left(\widetilde{U}_n\right)$ converges to $\widetilde{U}^1 V_1 \neq 0$, therefore $\frac{\widetilde{U}_n}{\|\widetilde{U}_n\|} = X_n$ converges to $\pm V_1$, and by the conclusion of Part 1, $\sum\limits_{n=1}^{\infty} a_n (\lambda_1 - \langle X_n, B X_n \rangle) < \infty$.

Moreover by H2c:

$$\begin{aligned} \sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n, B_n X_n \rangle| &= \sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n, (B_n - B) X_n \rangle - \langle X_n, B X_n \rangle| \\ &\leq \sum_{n=1}^{\infty} a_n (\lambda_1 - \langle X_n, B X_n \rangle) \ + \sum_{n=1}^{\infty} a_n ||B_n - B|| < \infty. \ \blacksquare \end{aligned}$$

Remark 10 *Part 1 can be replaced by:*
$||\widetilde{U}_{n+1}|| \leq \frac{\|I + a_n B_n\|}{1 + \lambda_1 a_n} \left\| \widetilde{U}_n \right\| \leq \left(1 + \frac{a_n \|B_n - B\|}{1 + \lambda_1 a_n}\right) \left\| \widetilde{U}_n \right\|$.

Under H2c, $\left\| \widetilde{U}_n \right\|$ *converges a.s. Assumption* $\sum\limits_{n=1}^{\infty} a_n^2 < \infty$ *is not used and can be replaced by* $a_n \underset{n \to +\infty}{\longrightarrow} 0$, *but in this case, the convergence of* $\sum\limits_{n=1}^{\infty} a_n (\lambda_1 - \langle B X_n, X_n \rangle)$ *and* $\sum\limits_{n=1}^{\infty} a_n |\lambda_1 - \langle B_n X_n, X_n \rangle|$ *is not proved.* $\blacksquare$

2.3.2 Simultaneous estimation of several eigenvectors

For $i = 1, ..., r$, recursively define the process $\left(\widetilde{X}_n^i\right)$ by:

$$\begin{aligned} \widetilde{Y}_{n+1}^i &= (I + a_n B_n)\widetilde{X}_n^i, \\ \widetilde{X}_{n+1}^i &= \widetilde{Y}_{n+1}^i - \sum_{j<i} \left\langle \widetilde{Y}_{n+1}^i, \frac{\widetilde{X}_{n+1}^j}{\left\|\widetilde{X}_{n+1}^j\right\|} \right\rangle \frac{\widetilde{X}_{n+1}^j}{\left\|\widetilde{X}_{n+1}^j\right\|}. \end{aligned}$$

Note that $\frac{\widetilde{X}_n^i}{\left\|\widetilde{X}_n^i\right\|} = X_n^i$.

Let $D_n^i = {}^{i1}B_n + \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n$.

Make the following assumptions:

(H1c') For $i = 1, ..., r$, $\left\|{}^{i1}B\right\| = \lambda_1 + ... + \lambda_i$.

(H2d') For $i = 1, ..., r$, $I + a_n D_n^i$ is invertible.

(H5') For $i = 1, ..., r$, X_1^i is an absolutely continuous random variable, independent from $B_1, ..., B_n, ...$.

Corollary 11 *Suppose assumptions H1a,b,c', H2c,d', H3 and H5' hold. Then, for $i = 1, ..., r$, almost surely X_n^i converges to $\pm V_i$, $\sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, BX_n^i \right\rangle\right|$ and $\sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, B_n X_n^i \right\rangle\right|$ converge.*

Proof

ω is fixed throughout the proof, belonging to the intersection of the a.s. convergence sets. Its writing will be omitted.

Let $i \in \{1, ..., r\}$.

$$\begin{aligned} {}^i\widetilde{X}_{n+1} &= \widetilde{X}_{n+1}^1 \wedge ... \wedge \widetilde{X}_{n+1}^i = \widetilde{Y}_{n+1}^1 \wedge ... \wedge \widetilde{Y}_{n+1}^i =^i \Lambda(I + a_n B_n)^i \widetilde{X}_n \\ &= \left(I + a_n\ {}^{i1}B_n + a_n \sum_{j=2}^{i} a_n^{j-1}\, {}^{ij}B_n\right)\ {}^i\widetilde{X_n} = \left(I + a_n\ D_n^i\right)\ {}^i\widetilde{X}_n. \end{aligned}$$

Note that $\frac{{}^i\widetilde{X}_{n+1}}{\left\|{}^i\widetilde{X}_{n+1}\right\|} = \frac{\widetilde{X}_{n+1}^1}{\left\|\widetilde{X}_{n+1}^1\right\|} \wedge ... \wedge \frac{\widetilde{X}_{n+1}^i}{\left\|\widetilde{X}_{n+1}^i\right\|} = \frac{{}^i\Lambda(I+a_n B_n)^i \widetilde{X}_n}{\left\|{}^i\Lambda(I+a_n B_n)^i \widetilde{X}_n\right\|} = {}^i X_{n+1}$.

By H2c and H3:

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

As B is Q-symmetric with distinct eigenvalues, $V_1 \wedge .. \wedge V_i$ is an eigenvector corresponding to its largest eigenvalue $\lambda_1 + ... + \lambda_i$. Applying Theorem 7 yields that:

${}^i X_n$ converges to $\pm V_1 \wedge .. \wedge V_i$,

$\sum_{n=1}^{\infty} a_n \left(\sum_{l=1}^{i} \lambda_l - \left\langle {}^{i1}B\ {}^i X_n,\ {}^i X_n \right\rangle\right)$ and $\sum_{n=1}^{\infty} a_n \left|\sum_{l=1}^{i} \lambda_l - \left\langle D_n^i\ {}^i X_n,\ {}^i X_n \right\rangle\right|$ converge,

which implies that $\sum_{n=1}^{\infty} a_n \left|\sum_{l=1}^{i} \lambda_l - \left\langle {}^{i1}B_n\ {}^i X_n,\ {}^i X_n \right\rangle\right|$ converges.

Suppose that, for $k = 1, ..., i-1$, X_n^k converges to $\pm V_k$, which is verified for $k = 1$, and prove that it is true for $k = i$.

In the development of $\left\langle X_n^1 \wedge ... \wedge X_n^i, \pm V_1 \wedge .. \wedge V_i \right\rangle$, which converges to ± 1, the only term which has a non-zero limit is $\left\langle X_n^1, V_1 \right\rangle ... \left\langle X_n^{i-1}, V_{i-1} \right\rangle \left\langle X_n^i, V_i \right\rangle$; since for $k = 1, ..., i-1$, $\left\langle X_n^k, V_k \right\rangle$ converges to ± 1, it follows that $\left\langle X_n^i, V_i \right\rangle$ converges to ± 1.

Applying the same proof as that of Corollary 4, Part 3, yields:
$\sum_{n=1}^{\infty} a_n \left| \lambda_i - \left\langle X_n^i, BX_n^i \right\rangle \right| < \infty$. By H2c:

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

3 Application to sequential principal component analysis of a data stream

Let $Z_{11}, ..., Z_{1m_1}, Z_{21}, ..., Z_{2m_2}, ..., Z_{n1}, ..., Z_{nm_n}, ...$ be an i.i.d sample of a random vector Z in $\mathbb{R}^p$ whose components are denoted $Z^1, ..., Z^p$. Let M be the metric used for PCA and $B = M^{\frac{1}{2}} E\left[(Z - E[Z])(Z - E[Z])'\right] M^{\frac{1}{2}}$. Let m belonging to $\mathbb{R}^p$ (in practice m is an estimation of $E[Z]$); with $Z^c = Z - m$, we have:

$$B = M^{\frac{1}{2}} \left(E\left[Z^c Z^{c'} \right] - E[Z^c] E[Z^c]' \right) M^{\frac{1}{2}}.$$

Let $\overline{Z}_{n-1}$ be the mean of the sample $(Z_{11}, ..., Z_{n-1,m_{n-1}})$ of Z and M_{n-1} a T_n-measurable estimation of M.

3.1 Use of a data mini-batch at each step

Note that the metric used for orthonormalization is the identity.

Recursively define the processes $\left(X_n^i\right), i = 1, ..., r$, by

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

.

Let $Z_{ni}^c = Z_{ni} - m$, $\overline{Z}_{n-1}^c = \overline{Z}_{n-1} - m$. We define:

$$B_n = M_{n-1}^{\frac{1}{2}} \left(\frac{1}{m_n} \sum_{i=1}^{m_n} Z_{ni}^c Z_{ni}^{c\,\prime} - \overline{Z}_{n-1}^c \left(\overline{Z}_{n-1}^c \right)' \right) M_{n-1}^{\frac{1}{2}}.$$

Make the following assumptions:

(H3') $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} \frac{a_n}{\sqrt{n}} < \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$.

(H4a) $\|Z\|$ is a.s. bounded.

(H4b) There is no affine or quadratic relation between the components of Z.

(H6a) There exists a positive number d such that $\sup_n \left\| M_n^{\frac{1}{2}} \right\| < d$.

(H6b) $M_n^{\frac{1}{2}} \longrightarrow M^{\frac{1}{2}}$ a.s.

(H6c) $E\left[\sum_{n=1}^{\infty} a_n \left\| M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}} \right\| \right] < \infty$.

Corollary 12 *Suppose assumptions H1b, H3', H4a,b and H6a,b,c hold. Then* X_n^i *converges a.s. to* $\pm V_i$, $\sum_{n=1}^{\infty} a_n \left| \lambda_i - \left(X_n^i\right)' BX_n^i \right|$ *and* $\sum_{n=1}^{\infty} a_n \left(\lambda_i - \left(X_n^i\right)' B_n X_n^i \right)$ *converge a.s. for* $i = 1, ..., r$.

Proof
Verify the assumptions of Corollary 4.
(H1a) B is symmetric.
(H2a) Under H4a and H6a, $\sup_n \|B_n\|$ is a.s. uniformly bounded.
(H2b) Almost surely:

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

If Z has 4^{th} order moments and $a_n>0$, $\sum_{n=1}^{\infty}\frac{a_n}{\sqrt{n}}<\infty$:

$$\sum_{n=1}^{\infty} a_n E\left[||\overline{Z}_{n-1}^c - E[Z^c]||\right] = \sum_{n=1}^{\infty} a_n E\left[||\overline{Z}_{n-1} - E[Z]||\right] < \infty. \text{ [9]}$$

Therefore, under H4a, H6a,c, $E\left[\sum_{n=1}^{\infty} a_n \left\|E\left[B_n \mid T_n\right]-B\right\|\right]<\infty$.
By Corollary 4, for $k=1,...,r$, X_n^k converges a.s. to one of the eigenvectors of B.
Prove now that $\underset{n\longrightarrow\infty}{lim} E[({X_n^k}' B_n V_k)^2 \ |T_n]>0$ a.s. on the set $\{X_n \longrightarrow V_j\}$ for $j\neq k$ to apply second part of Corollary 4.
In the following of the proof, X_n^k is denoted X_n.
Decompose $E[(X_n' B_n V_k)^2 \ |T_n]$ into the sum of three terms (1),(2),(3):

$$\begin{aligned}
& E\left[\left(X_n' M_{n-1}^{\frac{1}{2}}\left(\frac{1}{m_n}\sum_{i=1}^{m_n} Z_{ni}^c {Z_{ni}^c}' - \overline{Z}_{n-1}^c {\overline{Z}_{n-1}^c}'\right) M_{n-1}^{\frac{1}{2}} V_k\right)^2 \mid T_n\right] \\
= & E\left[\left(\frac{1}{m_n}\sum_{i=1}^{m_n}\left(X_n' M_{n-1}^{\frac{1}{2}} Z_{ni}^c\right)\left({Z_{ni}^c}' M_{n-1}^{\frac{1}{2}} V_k\right) - \left(X_n' M_{n-1}^{\frac{1}{2}} \overline{Z}_{n-1}^c\right)\left({\overline{Z}_{n-1}^c}' M_{n-1}^{\frac{1}{2}} V_k\right)\right)^2 \mid T_n\right] \\
= & E\left[\left(\frac{1}{m_n}\sum_{i=1}^{m_n}\left(X_n' M_{n-1}^{\frac{1}{2}} Z_{ni}^c\right)\left({Z_{ni}^c}' M_{n-1}^{\frac{1}{2}} V_k\right)\right)^2 \mid T_n\right] \ (1) \\
& -2\left(X_n' M_{n-1}^{\frac{1}{2}} \overline{Z}_{n-1}^c\right)\left({\overline{Z}_{n-1}^c}' M_{n-1}^{\frac{1}{2}} V_k\right)\frac{1}{m_n}\sum_{i=1}^{m_n} E\left[\left(X_n' M_{n-1}^{\frac{1}{2}} Z_{ni}^c\right)\left({Z_{ni}^c}' M_{n-1}^{\frac{1}{2}} V_k\right) \mid T_n\right] \ (2) \\
& +\left(X_n' M_{n-1}^{\frac{1}{2}} \overline{Z}_{n-1}^c\right)^2\left({\overline{Z}_{n-1}^c}' M_{n-1}^{\frac{1}{2}} V_k\right)^2. \ (3)
\end{aligned}$$

Note that the two random variables $R=V_j' M^{\frac{1}{2}} Z^c$ and $S=V_k' M^{\frac{1}{2}} Z^c$ are uncorrelated, then $E[RS]=E[R]E[S]$:

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

Under H6b, we have:

$$
\begin{aligned}
(1) \quad &= \frac{1}{m_n^2}\sum_{i=1}^{m_n}\sum_{l=1}^{m_n} E\left[\left(X_n' M_{n-1}^{\frac{1}{2}} Z_{ni}^c\right)\left(Z_{ni}^c{}' M_{n-1}^{\frac{1}{2}} V_k\right)\left(X_n' M_{n-1}^{\frac{1}{2}} Z_{nl}^c\right)\left(Z_{nl}^c{}' M_{n-1}^{\frac{1}{2}} V_k\right) \ |T_n\right] \\
&= X_n' M_{n-1}^{\frac{1}{2}} \frac{1}{m_n^2}\sum_{i=1}^{m_n}\sum_{l=1}^{m_n} E\left[\left(V_k' M_{n-1}^{\frac{1}{2}} Z_{ni}^c\right) Z_{ni}^c Z_{nl}^c{}' \left(Z_{nl}^c{}' M_{n-1}^{\frac{1}{2}} V_k\right) \mid T_n\right] M_{n-1}^{\frac{1}{2}} X_n \\
\underset{n\to+\infty}{\longrightarrow} \quad & V_j' M^{\frac{1}{2}} E\left[\left(V_k' M^{\frac{1}{2}} Z^c\right) Z^c Z^{c\prime} \left(Z^{c\prime} M^{\frac{1}{2}} V_k\right)\right] M^{\frac{1}{2}} V_j \\
&= E\left[\left(V_k' M^{\frac{1}{2}} Z^c\right)^2 \left(V_j' M^{\frac{1}{2}} Z^c\right)^2\right] \text{ a.s.}
\end{aligned}
$$

$$
(2) \underset{n\to+\infty}{\longrightarrow} -2E\left[V_j' M^{\frac{1}{2}} Z^c\right] E\left[Z^{c\prime} M^{\frac{1}{2}} V_k\right] E\left[\left(V_j' M^{\frac{1}{2}} Z^c\right)\left(Z^{c\prime} M^{\frac{1}{2}} V_k\right)\right]
$$

$$
= -2E\left[\left(V_j' M^{\frac{1}{2}} Z^c\right)\left(V_k' M^{\frac{1}{2}} Z^c\right)\right]^2 \text{ a.s.}
$$

$$
(3) \underset{n\to+\infty}{\longrightarrow} \left(E\left[V_j' M^{\frac{1}{2}} Z^c\right] E\left[V_k' M^{\frac{1}{2}} Z^c\right]\right)^2 = E\left[\left(V_j' M^{\frac{1}{2}} Z^c\right)\left(V_k' M^{\frac{1}{2}} Z^c\right)\right]^2 \text{ a.s.}
$$

As a result:

$$
E[(X_n' B_n V_k)^2 \ |T_n] \underset{n\to+\infty}{\longrightarrow} E\left[\left(V_j' M^{\frac{1}{2}} Z^c\right)^2 \left(V_k' M^{\frac{1}{2}} Z^c\right)^2\right] - E\left[\left(V_j' M^{\frac{1}{2}} Z^c\right)\left(V_k' M^{\frac{1}{2}} Z^c\right)\right]^2
$$

$$
= Var[V_j' M^{\frac{1}{2}} Z^c . V_k' M^{\frac{1}{2}} Z^c] > 0 \text{ a.s. by H4b.} \ \blacksquare
$$

3.2 Use of all observations up to the current step with different weights

At each step, all the observations up to the current step are taken into account but with different weights for observations at the current step and observations in the past.

In the definition of processes $\left(X_n^i\right), i = 1, ..., r$, we take now:

$$
\begin{aligned}
B_n &= w_1 B_n^1 + w_2 B_n^2, w_1 + w_2 = 1, w_1 > 0, w_2 \geqslant 0, \text{ with} \\
B_n^1 &= M_{n-1}^{\frac{1}{2}} \left(\frac{1}{m_n}\sum_{j=1}^{m_n} Z_{nj}^c Z_{nj}^c{}' - \overline{Z}_{n-1}^c \overline{Z}_{n-1}^c{}'\right) M_{n-1}^{\frac{1}{2}}, \\
B_n^2 &= M_{n-1}^{\frac{1}{2}} \left(\frac{1}{\sum\limits_{i=1}^{n-1} m_i}\sum_{i=1}^{n-1}\sum_{j=1}^{m_i} Z_{ij}^c Z_{ij}^c{}' - \overline{Z}_{n-1}^c \overline{Z}_{n-1}^c{}'\right) M_{n-1}^{\frac{1}{2}}.
\end{aligned}
$$

Corollary 13 *Suppose assumptions H1b, H3', H4a,b and H6a,b,c hold. Then X_n^i converges a.s. to $\pm V_i$, $\sum\limits_{n=1}^{\infty} a_n \left|\lambda_i - \left(X_n^i\right)' B X_n^i\right|$ and $\sum\limits_{n=1}^{\infty} a_n (\lambda_i - \left(X_n^i\right)' B_n X_n^i)$ converge a.s. for $i = 1, ..., r$.*

Proof

Verify the assumptions of Corollary 6.

(i) It is established in the proof of Corollary 12 that $E\left[\sum\limits_{n=1}^{\infty} a_n \left\|E[B_n^1|T_n] - B\right\|\right] < \infty$ a.s. under assumptions H3', H4a and H6a,c.

(ii) Prove now that $\mathrm{E}\left[\sum\limits_{n=1}^{\infty} a_n \left\|B_n^2 - B\right\|\right] < \infty$ a.s.

$$
\begin{aligned}
B_n^2 &= M_{n-1}^{\frac{1}{2}} C_{n-1} M_{n-1}^{\frac{1}{2}} \text{ with} \\
C_{n-1} &= \frac{1}{\sum\limits_{i=1}^{n-1} m_i}\sum_{i=1}^{n-1}\sum_{j=1}^{m_i} Z_{ij}^c Z_{ij}^c{}' - \overline{Z}_{n-1}^c \overline{Z}_{n-1}^c{}' = \frac{1}{\sum\limits_{i=1}^{n-1} m_i}\sum_{i=1}^{n-1}\sum_{j=1}^{m_i} Z_{ij} Z_{ij}{}' - \overline{Z}_{n-1} \overline{Z}_{n-1}{}'.
\end{aligned}
$$

$$
\begin{aligned}
B &= M^{\frac{1}{2}} C M^{\frac{1}{2}} \text{ with } C = E[ZZ'] - E[Z]E[Z']. \\
B_n^2 - B &= M_{n-1}^{\frac{1}{2}} C_{n-1} M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}} C M^{\frac{1}{2}} \\
&= (M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}}) C_{n-1} M_{n-1}^{\frac{1}{2}} + M^{\frac{1}{2}} (C_{n-1} - C) M_{n-1}^{\frac{1}{2}} + M^{\frac{1}{2}} C (M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}}). \\
C_{n-1} - C &= \frac{1}{\sum_{i=1}^{n-1} m_i} \sum_{i=1}^{n-1} \sum_{j=1}^{m_i} Z_{ij} Z'_{ij} - E[ZZ'] - \left(\overline{Z}_{n-1} - E\left[Z\right]\right) \overline{Z}_{n-1}{}' - E[Z](\overline{Z}_{n-1} - E\left[Z\right])'.
\end{aligned}
$$

Under assumptions H3' and H4a:

$\sum_{n=1}^{\infty} a_n E\left[\left\|\overline{Z}_{n-1} - E\left[Z\right]\right\|\right] < \infty$, $\sum_{n=1}^{\infty} a_n E\left[\left\|\overline{ZZ'}_{n-1} - E[ZZ']\right\|\right] < \infty$ [9].

Therefore, under H4a and H6a,c, $E\left[\sum_{n=1}^{\infty} a_n ||B_n^2 - B||\right] < \infty$.

(iii) Prove finally that $\lim_{n\longrightarrow\infty} E[(X_n^{k\prime} B_n V_k)^2 \; |T_n] > 0$ when $\lim_{n\longrightarrow\infty} X_n^k = \pm V_j \neq \pm V_k$ a.s. By the proof of Corollary 12, as $\lim_{n\longrightarrow\infty} X_n^{k\prime} B_n^2 V_k = \pm V_j' B V_k = 0$, under H4b and H6b:

$$
\begin{aligned}
\lim_{n\longrightarrow\infty} E[(X_n^{k\prime} B_n V_k)^2 | T_n] &= (\omega_1)^2 \lim_{n\longrightarrow\infty} E\left[\left(X_n^{k\prime} B_n^1 V_k\right)^2 \; |T_n\right] \\
&= (\omega_1)^2 Var[V_j' M^{\frac{1}{2}} Z^c . V_k' M^{\frac{1}{2}} Z^c] > 0 \text{ a.s. } \blacksquare
\end{aligned}
$$

3.3 Use of all observations up to the current step with uniform weights

We define now:

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

Make the following assumptions:

(H4c) Z has 4^{th} order moments.

(H6d) $\sum_{n=1}^{\infty} a_n \left\| M_n^{\frac{1}{2}} - M^{\frac{1}{2}} \right\| < \infty$ a.s.

Corollary 14 *Suppose assumptions H1b, H3', H4c, H5' and H6b,d hold. Then, for $i = 1, ..., r$, almost surely X_n^i converges to $\pm V_i$, $\sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, B X_n^i \right\rangle\right|$ and $\sum_{n=1}^{\infty} a_n \left|\lambda_i - \left\langle X_n^i, B_n X_n^i \right\rangle\right|$ converge.*

Proof

It suffices to verify assumption H2c, $\sum_{n=1}^{\infty} a_n ||B_n - B|| < \infty$ a.s. to apply Corollary 11. Under assumptions H4c and H6b,d, the proof is similar to that of Corollary 13 for B_n^2 without taking expectation. $\blacksquare$

In the particular case of normed principal component analysis, M is the diagonal matrix of the inverses of variances of $Z^1, ..., Z^p$. Denote for $j = 1, ..., p$, V_n^j the variance of the sample $(Z_{11}^j, ..., Z_{n,m_n}^j)$ of Z^j and M_n the diagonal matrix of order p whose element (j, j) is the inverse of $\frac{\mu_n}{\mu_n - 1} V_n^j$ with $\mu_n = \sum_{i=1}^{n} m_i$. Under H4c, H6b holds; it is established in [9] (lemma5) that H6d holds under H4c and H3'.

4 Conclusion

In this article we have given theorems of almost sure convergence of a normed stochastic approximation process to eigenvectors of a Q-symmetric matrix B corresponding to eigenvalues in decreasing order, assuming that $E\left[B_n \mid T_n\right]$ or B_n converges a.s. to B. This extends previous results assuming B_n i.i.d. with $E\left[B_n \mid T_n\right] = B$.

These results have been applied to online estimation of principal components of PCA of a random vector when the data arrive continously. In this case, the expectation and the variance of the variables are unknown and are estimated online along with the estimation of principal components. Classical results do not apply to this case. Moreover we can use at each step a data mini-batch as usual, but also all the observations obtained up to this step to take into account all the information contained in the previous data.

We have made a first set of experiments: several processes, with different numbers of observations used at each step or with all the observations up to the current step, have been compared on datasets or simulations (data not shown). It appears that processes that use all observations up to the current step yield the best results.

Acknowledgment *Results incorporated in this article received funding from the investments for the Future program under grant agreement No ANR-15-RHU-0004.*

References

[1] Benzécri JP. Approximation stochastique dans une algèbre normée non commutative. Bull.Soc.Math.France. 1969; 97: 225-241.

[2] Krasulina TP. Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices. Automation and Remote Control. 1970; 2:215-221.

[3] Karhunen J and Oja E. New methods for stochastic approximation of truncated Karhunen-Loève expansions. Proc. 6th Internat. Conf. on Pattern Recognition, IEEE Computer Society Press. 1982; 550-553.

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

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

[6] Brandière O. Un algorithme pour l'analyse en composantes principales. C.R. Acad. Sci. Paris. 1995; 321, série I: 233-236.

[7] Brandière O, Duflo M. Les algorithmes stochastiques contournent-ils les pièges ? Ann. Inst. Henri Poincaré, Probabilités et Statistique. 1996; 32(3): 395-427.

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

[9] Duarte K, Monnez JM, Albuisson E. Sequential linear regression with online standardized data. PLoS ONE. 2018; 13(1):e0191186. https://doi.org/10.1371/journal.pone.0191186 .

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