Convergence of a normed eigenvector stochastic approximation process and application to online principal component analysis of a data stream

Jean-Marie Monnez, Abderrahman Skiredj

To cite this version:
Jean-Marie Monnez, Abderrahman Skiredj. Convergence of a normed eigenvector stochastic approximation process and application to online principal component analysis of a data stream. 2018. hal-01844419

HAL Id: hal-01844419
https://hal.archives-ouvertes.fr/hal-01844419
Submitted on 19 Jul 2018

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Convergence of a Normed Eigenvector Stochastic Approximation Process and Application to Online Principal Component Analysis of a Data Stream

Jean-Marie Monnez\textsuperscript{1,4,5*}, Abderrahman Skiredj\textsuperscript{2,3**}

1 Université de Lorraine, CNRS, Inria\textsuperscript{1}, IECL\textsuperscript{2}, F-54000 Nancy, France
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 Université de Lorraine, IUT Nancy-Charlemagne, F 54000 Nancy, France
5 CHRU Nancy, INSERM, CIC\textsuperscript{4}, Plurithématique, F-54000 Nancy, France

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

** Abstract **

Many articles were devoted to the problem of estimating recursively the eigenvectors and eigenvalues in decreasing order of the expectation of a random matrix using an i.i.d. sample of it. The present study makes the following contributions. The convergence of a normed process is proved under more general assumptions: the random matrices are not supposed i.i.d. and a new data mini-batch or all data until the current step are taken into account at each step without storing them; three types of processes are studied; this is 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 metrics 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 or factors, whose estimations 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.

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 \( (\Omega, A, P) \) and that \( (Z_1, ..., Z_n, ...) \) is an i.i.d sample of a random vector \( Z \). Denote \( \theta \) the expectation of \( Z \) and \( C \) its covariance matrix which are unknown in the case of a data stream.

Let \( M \) be a positive definite symmetric \( p \times p \) matrix called metrics. 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 maximal variance, under the normalization constraint \( c_l'M^{-1}c_l = 1 \); \( c_l \) is a \( M^{-1} \)-unit eigenvector of \( M \) corresponding to the \( l \)-th largest eigenvalue \( \lambda_l \). 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 \( \lambda_l \). 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 metrics. But the expectation \( \theta \) 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.

\textsuperscript{1}Inria, project-team BIGS
\textsuperscript{2}Institut Elie Cartan de Lorraine, BP 239, F-54506 Vandœuvre-lès-Nancy, France
\textsuperscript{3}Campus Artem, CS 14234, 92 Rue Sergent Blandan, F-54042 Nancy, France
\textsuperscript{4}Centre d’Investigation Clinique
\textsuperscript{5}Université de Lorraine, CNRS, IECL, F-54000 Nancy, France
Many articles were devoted to this problem when supposing $M$ and $\theta$ 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 Duflou [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 $E[B_n] = B$, $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$.

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. and a new data mini-batch or all data until the current step are taken into account at each step without storing them; this is applied to online estimation of principal components in PCA of a random vector $Z$, when its expectation is unknown, as well as possibly the metrics used, and must be estimated online.

Denote $Q$ a metrics 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 transposed 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 basis of $\mathbb{R}^p$ composed of eigenvectors of $A$. The norm of a matrix $A$ is the spectral norm denoted $\|A\|$.

Let $B$ be a $Q$-symmetric matrix. Denote $T_n$ the $\sigma$-field generated by the events before time $n$.

In the next section, the almost sure (a.s.) convergence of the normed process to eigenvectors of $B$ is studied. Three cases are considered:
- $E[B_n|T_n]$ converges a.s. to $B$;
- $B_n = \omega_1 B_n^1 + \omega_2 B_n^2$ with $\omega_1 n + \omega_2 n = 1$, $B_n^1$ 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 $B$ associated to its greatest eigenvalue is proved with, in the first case, a method following that of [8] 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 $(X_n^i)$, $i = 1, \ldots, r$ of the same type, obtained by Gram-Schmidt orthonormalization, to unit eigenvectors associated 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 in PCA. In order to reduce computing time, particularly in the case of a data stream, and to avoid numerical explosions, it is proposed:
- to estimate the eigenvectors $a_i$ of the symmetric $p \times p$ matrix $B = M^{1/2}CM^{1/2}$ (symmetrization); then the orthonormalization is computed with respect to $I$; estimates of $c_i$ and $u_i$ can be obtained from that of $a_i$;
- 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 (pseudo-centering);
- to use a data mini-batch at step $n$ or all observations until step $n$ without storing them.

This yields the following definitions of $B_n$, $(Z_{n,1}, \ldots, Z_{n,m_n})$ denoting the new observations taken into account at step $n$, $\overline{Z}_{n-1}$ the mean of the sample $(Z_{n,1}, \ldots, Z_{n-1,m_{n-1}})$, $M_{n-1}$ an estimation of $M$ depending on this sample, for $i = 1, \ldots, r$, $Z_{n,i} = Z_{n,i} - m$ and $\overline{Z}_{n} = \overline{Z}_{n} - m$:

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

$$B_n^2 = M_{n-1}^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Z_{i,j}' Z_{i,j} - \overline{Z}_{n} \overline{Z}_{n}' \right) M_{n-1}^{-1/2},$$

and a combination of both.

Three cases are studied: at each step are taken into account:
- a data mini-batch, supposing $Z$ and $M_{n-1}$ uniformly bounded,
- or all observations until this step with different weights for observations in the past, which are not stored, and observations at this step, supposing $Z$ and $M_{n-1}$ uniformly bounded,
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 \(B\) associated to \(\lambda_i, i = 1, ... p\).

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

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

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

Denote \(U_n = (X_n - B X_n), W_n = (X_n, B X_n)\).

**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 \(E_j = \{U_n \rightarrow \lambda_j\}\), \(X_n\) 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 \(\bigcup_{j=2}^{p} E_j\), \(\liminf E \left[ (B_n X_n, V_1)^2 | T_n \right] > 0\) a.s., then \(P(E_1) = 1\).

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:

\[
\langle M \rangle_1 = M_1^2
\]

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

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 \((\gamma_n)\) be a sequence of positive numbers such that \(\sum_{n=1}^{\infty} \gamma_n^2 < \infty\). Let \((Z_n)\) and \((\delta_n)\) be two sequences of random variables adapted to a filtration \((T_n)\), and \(\epsilon_n\) a noise adapted to \((T_n)\).

Suppose on the set \(\Gamma\):

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

Step 1: expression of $X_{n+1}$
Under H2a, as $\| (I + a_n B_n) X_n \|^2 = 1 + 2a_n W_n + a_n^2 \| B_n X_n \|^2$:

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

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

$$
\beta_n = -a_n W_n B_n - \frac{1}{2} a_n \| B_n X_n \|^2 I - \frac{1}{2} a_n^2 B_n \| B_n X_n \|^2 + a_n^{-1} \alpha_n I + a_n B_n.
$$

$X_{n+1} = (I + a_n (B - U_n I) + a_n \Gamma_n) X_n, \text{ with }
\Gamma_n = (B_n - B) - (X_n (B_n - B) X_n) I + \beta_n, \quad ||\beta_n|| = O(a_n).

$\beta_n$ and $\Gamma_n$ are uniformly bounded by assumption H2a.

Step 2: convergence of $U_n$

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

As $\|X_n\| = 1$, denoting $\mu_n = 2a_n \langle E[\Gamma_n | T_n] X_n, B X_n \rangle + a_n^2 \eta_n$:

$$
\langle (B - U_n I) X_n, B X_n \rangle = \|B X_n\|^2 - U_n^2 = \|B X_n - U_n X_n\|^2 \geq 0.
$$

$$
E[U_{n+1} | T_n] \geq U_n + \mu_n \quad a.s.
$$

$$
E[U_{n+1} - \sum_{i=1}^{n} \mu_i | T_n] \geq U_n - \sum_{i=1}^{n-1} \mu_i \quad a.s.
$$

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

$$
\|\beta_n\| \leq \frac{3}{2} a_n \|B_n\|^2 + \frac{1}{2} a_n^2 \|B_n\|^3 + a_n^{-1} |\alpha_n| + |\alpha_n| \|B_n\| = O(a_n^2);
$$

$$
\|\Gamma_n\| \leq 2 \|B_n - B\| + \|\beta_n\| = O(1);
$$

$$
\|\eta_n\| \leq 4 \|B\|^3 + 4 \|B\|^2 \|E[\Gamma_n | T_n]\| + \|B\| E \left[ \|\Gamma_n\|^2 | T_n \right] = O(1).
$$

$$
\|E[\Gamma_n | T_n]\| \leq 2 \|E[B_n | T_n] - B\| + \|E[\beta_n | T_n]\| \quad a.s.
$$

By H2b and H3:

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

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.

Step 3: convergence of $X_n = \left< X_n, V_j \right>$
Denote $\Gamma_n = \left< \Gamma_n X_n, V_j \right>$.

$$
X_{n+1}^j = \left< (I + a_n (B - U_n I) + a_n \Gamma_n) X_n, V_j \right> = X_n^j (1 + a_n (\lambda_j - U_n)) + a_n \Gamma_n^j.
$$
\[(X_{n+1}^j)^2 = (X_n^j)^2(1 + 2a_n(\lambda_j - U_n)) + a_n^2(\lambda_j - U_n)^2(X_n^j)^2 + 2a_n(1 + a_n(\lambda_j - U_n))X_n^j\Gamma_n + a_n^2(\Gamma_n)^2 \]
\[= (X_n^j)^2(1 + 2a_n(\lambda_j - U_n)) + a_n^2(\lambda_j - U_n)^2 + 2a_nX_n^j\Gamma_n \]
\[= (X_n^j)^2 + 2\sum_{l=1}^{n} a_l(\lambda_j - U_l)(X_l^j)^2 + \sum_{l=1}^{n} a_l^2(\lambda_j - U_l)^2 + 2\sum_{l=1}^{n} a_lX_l^j\Gamma_l \]
\[= (X_n^j)^2 + (1) + (2) + (3). \]

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

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

(ii) Consider now (3).

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

\[\sum_{l=1}^{n} a_l|X_l^j E[\Gamma_l^j|T_l]| \leq \sum_{l=1}^{n} a_l|E[\Gamma_l^j|T_l]|X_l, V_j | \]
\[\leq \sum_{l=1}^{n} a_l|E[\Gamma_l^j|T_l]| \leq \sum_{l=1}^{n} a_l (2||E[B_l|T_l] - B|| + ||E[\beta_l|T_l]|). \]

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

Let \(M_l^j = \sum_{l=1}^{n-1} a_lX_l^j(\Gamma_l^j - E[\Gamma_l^j|T_l]) \); \((M_l^j) \) is a square-integrable martingale adapted to the filtration \((T_l)\); denoting \((M_l^j)\) its increasing process. By H2a:

\[\langle M_l^j \rangle_{n+1} - \langle M_l^j \rangle_n = E[(M_l^j_{n+1} - M_l^j_{n})^2|T_n] = a_n^2 E[(X_n^j)^2(T_n^j - E[\Gamma_l^j|T_n])^2|T_n] \]
\[\leq a_n^2 E[(\Gamma_l^j)^2|T_n] - E[E[\Gamma_l^j|T_n]^2|T_n]) \leq a_n^2 E(||\Gamma_n||^2|T_n) \]

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

Therefore (3) converges a.s.

(iii) Consider finally (1). Let \(\omega \) fixed belonging to the convergence set of \(U_n\). The writing of \(\omega \) will be omitted in the following. Denote \(L \) 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 \(\omega \). Thus there exists \(A > 0 \) such that:

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

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 \(L \neq \lambda_j, (X_j^j)^2 \) converges a.s.

Step 4: convergence of \(X_n \)
If the limit of \( U_n \) is different from \( \lambda_j \), then by convergence of (1) in step 3, \( \sum_{l=1}^{\infty} a_l (X_l^i)^2 < \infty \) and \( X_l^i \) 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 \( B, \lambda_i \).

For \( j \neq i \), \( X_j^i \) converges to 0, therefore \( (X_j^i)^2 \) converges to 1 and since

\[
X_{n+1} - X_n = a_n ((B - U_n) + \Gamma_n) X_n,
\]

\( X_{n+1} - X_n \) 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:

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

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

(iii) Let \( M_n = \sum_{i=1}^{n-1} a_t (X_t, (B_t - E[B_t|T_t])X_t) \). \( (M_n) \) is a square-integrable martingale adapted to the filtration \((T_n)\). Its increasing process \((M_n)\) converges indeed:

\[
\langle M \rangle_{n+1} - \langle M \rangle_n = E[(M_{n+1} - M_n)^2|T_n] = a_n^2 E[|X_n - (B_n - E[B_n|T_n])|^2|T_n] \leq a_n^2 E[|B_n - E[B_n|T_n]|^2|T_n]
\]

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).

**Step 5: convergence of \( X_n \) to \( \pm V_1 \)**

Suppose \( i > 1 \).

\[
X_{n+1}^1 = (1 + a_n (\lambda_1 - U_n)) X_n^1 + a_n (\Gamma_n X_n, V_1) = (1 + a_n (\lambda_1 - \lambda_i) + (\lambda_i - U_n)) X_1^1 + a_n (\Gamma_n X_n, V_1).
\]

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

(i) \( X_n^i \) is bounded.

(ii) \( \sum_{n=1}^{\infty} a_n^2 (\lambda_1 - U_n)^2 < \infty \) 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:

\[
\langle \Gamma_n X_n, V_1 \rangle = \langle (B_n - B) X_n, V_1 \rangle - \langle (B_n - B) X_n, X_n \rangle \langle X_n, V_1 \rangle + \langle \beta_n X_n, V_1 \rangle = \langle B_n X_n, V_1 \rangle - \langle X_n, V_1 \rangle (\lambda_1 + \langle (B_n - B) X_n, X_n \rangle) + \langle \beta_n X_n, V_1 \rangle.
\]

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

\[
E[(\Gamma_n X_n, V_1) - \langle B_n X_n, V_1 \rangle|^2|T_n] \leq 2 (X_n^i)^2 E[(\lambda_1 + \langle (B_n - B) X_n, X_n \rangle)^2|T_n] + 2 E[(\langle \beta_n X_n, V_1 \rangle)^2|T_n] \leq c (X_n^i)^2 + 2 E[|\beta_n|^2|T_n] \rightarrow 0 \text{ as } n \rightarrow \infty.
\]
since \(\|\beta_n\| = O(a_n)\) and \(X_n^{1} \xrightarrow{n \to +\infty} 0\). Likewise:
\[
E[(\langle T_nX_n, V_1 \rangle - \langle B_nX_n, V_1 \rangle) \langle B_nX_n, V_1 \rangle | T_n] \xrightarrow{n \to +\infty} 0.
\]
Then, if \(\lim \inf E[\langle B_nX_n, V_1 \rangle^2 | T_n] > 0\), \(\lim \inf E[\langle T_nX_n, V_1 \rangle^2 | T_n] > 0\).

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 \(i = 1, \ldots, r\), \(X_n^i\) does not denote the \(i^{th}\) component of \(X_n\), but a random variable in \(\mathbb{R}^p\) recursively defined by:
\[
Y_{n+1}^i = (I + a_nB_n)X_n^i,
\]
\[
T_{n+1}^i = Y_{n+1}^i - \sum_{j<i} (Y_{n+1}^i, X_n^j)X_n^j + \frac{T_{n+1}^i}{\|T_{n+1}^i\|}.
\]

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

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

1) For \(i = 1, \ldots, r\), almost surely \(X_n^i\) converges to one of the eigenvectors of \(B\).
2) If moreover, for \(i = 1, \ldots, r\), almost surely on
\[
\bigcup_{j=1}^p \{ X_n^i \to \pm V_j \},
\]
\(\lim \inf E[\langle B_nX_n^i, V_1 \rangle^2 | T_n] > 0\), then \(X_n^i\) converges a.s. to \(V_i\) or \(-V_i\), 
\[
\sum_{n=1}^\infty a_n(|\lambda_i - \langle BX_n^i, X_n^i \rangle|) \text{ and } \sum_{n=1}^\infty a_n(|\lambda_i - \langle B_nX_n^i, X_n^i \rangle|)
\]
converge a.s.

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

Let \((e_1, \ldots, e_p)\) be a basis of \(\mathbb{R}^p\). For \(r \leq p\), denote \(\Lambda^r \mathbb{R}^p\) the exterior power of order \(r\) of \(\mathbb{R}^p\), generated by the \(C_r^p\) exterior products \(e_{i1} \wedge e_{i2} \wedge \ldots \wedge e_{ir}\), \(i_1 < i_2 < \ldots < i_r \in \{1, \ldots, p\}\).

a) Let \(Q\) be a metrics in \(\mathbb{R}^p\). Define the inner product in \(\Lambda^r \mathbb{R}^p\) induced by the metrics \(Q\), also denoted \(\langle \cdot, \cdot \rangle\), such that:
\[
\langle e_{i1} \wedge \ldots \wedge e_{ir}, e_{k1} \wedge \ldots \wedge e_{kr} \rangle = \sum_{\sigma \in G_r} (-1)^{s(\sigma)} \langle e_{i\sigma(k_1)} \ldots e_{i\sigma(k_r)}, e_{k\sigma(k_1)} \ldots e_{k\sigma(k_r)} \rangle,
\]
\(G_r\) being the set of permutations \(\sigma\) of \(\{k_1, \ldots, k_r\}\) and \(s(\sigma)\) the number of inversions of \(\sigma\).

Denote also \(\|\cdot\|\) the associated norm. Note that if \(x_1, \ldots, x_r\) are \(Q\)-orthogonal, \(\|x_1 \wedge \ldots \wedge x_r\| = \prod_{i=1}^r \|x_i\|\), and if \((e_1, \ldots, e_p)\) is a \(Q\)-orthonormal basis of \(\mathbb{R}^p\), then the set of the \(C_r^p\) exterior products \(e_{i1} \wedge \ldots \wedge e_{ir}\) is an orthonormal basis of \(\Lambda^r \mathbb{R}^p\).

b) Let \(U\) be an endomorphism in \(\mathbb{R}^p\). Define for \(j = 1, \ldots, r\) the endomorphism \(\Lambda^r U\) in \(\Lambda^p \mathbb{R}^p\) such that:
\[
\Lambda^r U(x_1 \wedge \ldots \wedge x_r) = \sum_{1 \leq i_1 < i_2 < \ldots < i_r \leq r} x_1 \wedge \ldots \wedge Ux_{i_1} \wedge \ldots \wedge Ux_{i_r} \wedge \ldots \wedge x_r.
\]
For \(j = 1, \Lambda^r U(x_1 \wedge \ldots \wedge x_r) = \sum_{i=1}^r x_1 \wedge \ldots \wedge Ux_i \wedge \ldots \wedge x_r.
\]
Denote \(\Lambda^r U\) the endomorphism \(\Lambda^r U\) such that:
\[
\Lambda^r U(x_1 \wedge \ldots \wedge x_r) = Ux_1 \wedge Ux_2 \wedge \ldots \wedge Ux_r.
\]

c) The following properties hold:

(i) Suppose that the eigenvalues \(\lambda_1 > \ldots > \lambda_p\) of \(U\) are distinct and denote for \(j = 1, \ldots, r\), \(V_j\) an eigenvector associated to \(\lambda_j\). Then the \(C_r^p\) vectors \(V_1, \ldots, V_r\), \(1 \leq i_1 < \ldots < i_r \leq p\), are eigenvectors of \(\Lambda^r U\) respectively associated to the eigenvalues \(\lambda_{i_1} + \ldots + \lambda_{i_r}\).
(ii) $r\Lambda(I + U) = I + \sum_{j=1}^{r} r^j U$.

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

**Proof**

**Step 1**

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

$$\|T^1_{n+1} \wedge ... \wedge T^i_{n+1}\| = \prod_{l=1}^{i} \|T^l_{n+1}\|.$$  

Then, denoting $iX_{n+1} = X^1_{n+1} \wedge ... \wedge X^i_{n+1}$ and $D^i_n = i^1 B_n + \sum_{j=2}^{i} a^j_n i^j B_n$:

$$iX_{n+1} = \frac{T^1_{n+1} \wedge ... \wedge T^i_{n+1}}{\|T^1_{n+1} \wedge ... \wedge T^i_{n+1}\|} = \frac{Y^1_{n+1} \wedge ... \wedge Y^i_{n+1}}{\|Y^1_{n+1} \wedge ... \wedge Y^i_{n+1}\|} = (I + a_n i^1 B_n + \sum_{j=2}^{i} a^j_n i^j B_n) iX_n$$

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

Moreover, as $U \mapsto i^1 U$ is a linear application, assumptions H2a,b and H3 yield that:

$$E \left[ \sum_{n=1}^{\infty} \alpha_n \|E[D^i_n T_n] - i^1B\| \right] = E \left[ \sum_{n=1}^{\infty} \alpha_n \left\| i^1 B_n - i^1 B + \sum_{j=2}^{i} a^j_n i^j B_n \left| T_n \right| \right\| \right]$$

$$\leq E \left[ \sum_{n=1}^{\infty} \alpha_n \left( \|E[B_n - B]\| T_n\| + \sum_{j=2}^{i} a^j_n \|E[i^j B_n]\| T_n\| \right) \right]$$

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

Applying first assertion of theorem 1 yields that almost surely, $iX_n$ converges to a unit eigenvector $\pm V_{j_i} \wedge ..., \wedge V_{j_i}$ of $i^1 B$,

$$\sum_{n=1}^{\infty} \alpha_n (\lambda_{j_i} + ... + \lambda_{j_i} - \langle i^1 B iX_n, iX_n \rangle)$$

and

$$\sum_{n=1}^{\infty} \alpha_n (\lambda_{j_i} + ... + \lambda_{j_i} - \langle D^i_n iX_n, iX_n \rangle)$$

converge.

Moreover by H2a and H3, $\sum_{n=1}^{\infty} \alpha_n (\lambda_{j_i} + ... + \lambda_{j_i} - \langle i^1 B_n iX_n, iX_n \rangle)$ converges a.s.

**Step 2**

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

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

$$iX_n = X^1_n \wedge ... \wedge X^i_n \mathop{\longrightarrow}\limits_{n\to\infty} \pm V_1 \Lambda ... \Lambda V_{i-1} \wedge V_j.$$
The only term which has a non-zero limit in the development of
\[ \left< X_n^1 \land \ldots \land X_n^i, \pm V_1 \land \ldots \land V_{i-1} \land V_j \right>, \]
whose limit is 1 as \( n \to \infty \), is \( \left< X_n^1, V_1 \right> \left< X_n^2, V_2 \right> \ldots \left< X_n^{i-1}, V_{i-1} \right> \left< X_n^i, V_j \right> \) obtained for \( \sigma = Id \). As for \( k = 1, \ldots, i-1 \), \( \left< X_n^k, V_k \right> \to \pm 1 \), then \( \left< X_n^i, V_j \right> \to \pm 1 \). Therefore \( X_n^i \to \pm V_j \).

2) Prove now that \( V_j = \pm V_i \). Suppose \( X_n^i \to \pm V_j \neq \pm V_i \).

Denote \( G_i \) the set of permutations \( \sigma \) of \( \{1, \ldots, i\} \) with \( \sigma = (\sigma(1), \ldots, \sigma(i)) \) and \( s(\sigma) \) the number of inversions of \( \sigma \).

\[
\langle \overset{i}{\overset{1}{B}}_n(X_n^1 \land \ldots \land X_n^i), V_1 \land \ldots \land V_i \rangle = \sum_{l=1}^{i} \langle X_n^1 \land \ldots \land B_n X_n^l \land \ldots \land X_n^i, V_1 \land \ldots \land V_i \rangle
\]

\[
= \sum_{l=1}^{i} E \left[ \langle \overset{i}{\overset{l}{B}}_n(X_n^1 \land \ldots \land X_n^l), V_1 \land \ldots \land V_l \rangle^2 | T_n \right]
\]

As for \( k = 1, \ldots, i-1 \), \( X_n^k \to \pm V_k \), the only term with a non-zero limit in the development of this conditional expectation is
\[
\langle X_n^1, V_1 \rangle \ldots \langle X_n^{i-1}, V_{i-1} \rangle \left< [B_n X_n^i, V_i] \right>^2 | T_n \]
and
\[
\lim_{n \to \infty} E \left[ \langle \overset{i}{\overset{i}{B}}_n(X_n^1 \land \ldots \land X_n^i), V_1 \land \ldots \land V_i \rangle^2 | T_n \right] = \lim_{n \to \infty} E \left[ (B_n X_n^i, V_i)^2 | T_n \right] > 0.

Moreover, by H2a and \( \lim_{n \to \infty} a_n = 0 \):

\[
\lim_{n \to \infty} E \left[ \left( \overset{i}{\overset{j}{B}}_n(X_n^1 \land \ldots \land X_n^j), V_1 \land \ldots \land V_j \right)^2 | T_n \right] = 0.
\]

Then \( \lim_{n \to \infty} E \left[ (D_n^i(X_n^1 \land \ldots \land X_n^i), V_1 \land \ldots \land V_i)^2 | T_n \right] > 0 \).

Applying second assertion of theorem 1 yields almost surely:

\[
X_n^1 \land \ldots \land X_n^i \to \pm V_1 \land \ldots \land V_i, \text{ therefore } X_n^i \to \pm V_i,
\]

\[
\sum_{n=1}^{\infty} a_n \left( \sum_{l=1}^{i} \lambda_l - \langle \overset{i}{\overset{i}{B}}_n X_n^i, X_n^i \rangle \right) \text{ and } \sum_{n=1}^{\infty} a_n \left( \sum_{l=1}^{i} \lambda_l - \langle D_n^i X_n^i, X_n^i \rangle \right)
\]
converge,

then by H2a and H3, \( \sum_{n=1}^{\infty} a_n \left( \sum_{l=1}^{i} \lambda_l - \langle \overset{i}{\overset{i}{B}}_n X_n^i, X_n^i \rangle \right) \) converges.

**Step 3**

\[
\langle \overset{i}{\overset{i}{B}}(X_n^i)^i, X_n^i \rangle = \sum_{k=1}^{i} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle X_n^1, X_n^{\sigma(1)} \rangle \ldots \langle B X_n^k, X_n^{\sigma(k)} \rangle \ldots \langle X_n^i, X_n^{\sigma(i)} \rangle.
\]
As \( X_n^1, \ldots, X_n^i \) are orthonormal, this sum is equal to
\[
\sum_{k=1}^{i} \langle X_n^k, X_n^i \rangle = \sum_{k=1}^{i} \langle BX_n^k, X_n^i \rangle = \sum_{k=1}^{i} \langle BX_n^k, X_n^k \rangle.
\]

Then, as \( \sum_{i=1}^{i} \lambda_i \) is the greatest eigenvalue of \( i^1 B \):
\[
\lambda_i - \langle BX_n^i, X_n^i \rangle = \left( \sum_{i=1}^n \lambda_i - \langle i^1 B^i X_n^i, X_n^i \rangle \right) - \left( \sum_{i=1}^{i-1} \lambda_i - \langle i^{-1} B^{i-1} X_n^i, X_n^{i-1} \rangle \right)
\]
\[
= \left| \sum_{i=1}^n \lambda_i - \langle i^1 B^i X_n^i, X_n^i \rangle \right| - \left| \sum_{i=1}^{i-1} \lambda_i - \langle i^{-1} B^{i-1} X_n^i, X_n^{i-1} \rangle \right|.
\]

Almost surely, \( \sum_{n=1}^{\infty} a_n \left| \sum_{i=1}^n \lambda_i - \langle i^1 B^i X_n^i, X_n^i \rangle \right| < \infty \), then \( \sum_{n=1}^{\infty} a_n \lambda_i - \langle BX_n^i, X_n^i \rangle \) converges, then
\[
\sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^n \lambda_i - \langle i^1 B^i X_n^i, X_n^i \rangle \right) \text{ converges.}
\]

2.2 Second case

Consider the same processes \((X_n)\) and \((X_n^i)\) as in the first case. Suppose now \( B_n = \omega_1 B_n^1 + \omega_2 B_n^2 \) with \( \omega_1 > 0, \omega_2 \geq 0, \omega_1 + \omega_2 = 1, B_n^2 T_n \)-measurable.

2.2.1 Theorem of almost sure convergence

Make the following assumptions:
(H2a') There exists a positive number \( b_1 \) such that \( \sup_n \| B_n^1 \| < b_1 \) a.s.
(H2b') \( E \left[ \sum_{n=1}^{\infty} a_n \| E \left[ B_n^1 \mid T_n \right] - B \| \right] < \infty. \)
(H2c') \( B_n^2 T_n \)-measurable, \( B_n^2 \xrightarrow{n \to \infty} B, E \left[ \sum_{n=1}^{\infty} \| B_n^2 - B \| \right] < \infty \) 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 \( E_j = \{ U_n \to \lambda_j \} \), \( X_n \) 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 \( \lim_{n \to \infty} \omega_1 = 1 > 0 \) and on \( \bigcap_{j=2}^{p} E_j \), \( \inf \| E \left[ (B_n^1 X_n, V_1) \right| T_n \| > 0 \) 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 \( \inf \| E \left[ (B_n X_n, V_1) \mid T_n \right] > 0 \) a.s. when \( \lim_{n \to \infty} X_n = \pm V_j \neq \pm V_1. \)

\[
E \left[ (\omega_1 B_n^1 + \omega_2 B_n^2 X_n, V_1) \mid T_n \right]
\]
\[
= (\omega_1)^2 \left( B_n^2 X_n, V_1 \right) + 2\omega_1 \omega_2 \left( B_n^2 X_n, V_1 \right) E \left[ (B_n^1 X_n, V_1) \mid T_n \right]
\]
\[
+ (\omega_1)^2 E \left[ (B_n^1 X_n, V_1) \mid T_n \right] \text{ a.s.}
\]

When \( \lim_{n \to \infty} X_n = \pm V_j \neq \pm V_1, \lim_{n \to \infty} \left( B_n^2 X_n, V_1 \right) = \pm \left( BV_j, V_1 \right) = \pm \lambda_j \left( V_j, V_1 \right) = 0. \) Then:

\[
\inf E \left[ (\omega_1 B_n^1 + \omega_2 B_n^2 X_n, V_1) \mid T_n \right] = (\omega_1)^2 \inf E \left[ (B_n^1 X_n, V_1) \mid 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 $\lim_{n \to \infty} \omega_{1n} = \omega_1 > 0$ and for i = 1, ..., r, a.s. on $\bigcup_{j=1}^{n} \{X_n^i \to \pm V_j\}$, $\liminf E[(B_n X_n^i, V_i)2|T_n] > 0$, then almost surely $X_n^i$ converges to $\pm V_i$, $\sum_{n=1}^{\infty} a_n |\lambda_i - \langle BX_n^i, X_n^i \rangle|$ and $\sum_{n=1}^{\infty} a_n (\lambda_i - \langle B_n X_n^i, X_n^i \rangle)$ converge.

It is a direct application of corollary 4 whose assumptions 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 $(\tilde{X}_n)$ such that

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

and the process $(\tilde{U}_n)$ such that

$$\tilde{U}_{n+1} = \frac{\tilde{X}_{n+1}}{\prod_{i=1}^{n} (1 + \lambda_i a_i)} = \frac{I + a_n B_n \tilde{U}_n}{1 + \lambda_i a_n}$$

Note that $\frac{\tilde{U}_n}{\|\tilde{U}_n\|} = \frac{\tilde{X}_n}{\|\tilde{X}_n\|} = X_n$. 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) $X_1$ 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, $\tilde{U}_n$ 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 $\sum_{n=1}^{\infty} a_n (\lambda_i - \langle B_n X_n, X_n \rangle)$ 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, $(z_n)$, $(\alpha_n)$, $(\beta_n)$ and $(\gamma_n)$ are four sequences of non-negative numbers such that:

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

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

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

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

Step 1

\[
||\bar{U}_{n+1}||^2 = ||\bar{U}_n||^2 + 2 \frac{a_n}{1 + \lambda_1 a_n} \left\langle \bar{U}_n, (B_n - \lambda_1 I)\bar{U}_n \right\rangle + \frac{a_n^2}{(1 + \lambda_1 a_n)^2} \|B_n - \lambda_1 I\|\bar{U}_n\|^2
\]

\[
= ||\bar{U}_n||^2 + 2 \frac{a_n}{1 + \lambda_1 a_n} \left\langle \bar{U}_n, (B_n - B)\bar{U}_n \right\rangle + \frac{a_n^2}{(1 + \lambda_1 a_n)^2} \|B_n - \lambda_1 I\|\bar{U}_n\|^2
\]

\[-2 \frac{a_n}{1 + \lambda_1 a_n} \left\langle \bar{U}_n, (\lambda_1 I - B)\bar{U}_n \right\rangle.
\]

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

\[
||B_n - \lambda_1 I||^2 \leq 2 \|B_n - B\|^2 + 2 ||\lambda_1 I - B||^2.
\]

\[
||\bar{U}_{n+1}||^2 \leq ||\bar{U}_n||^2 (1 + 2a_n \|B_n - B\| + 2a_n^2 \|B_n - B\|^2 + 2a_n^2 (\lambda_1 - \lambda_p)^2)
\]

\[-2 \frac{a_n}{1 + \lambda_1 a_n} \left\langle \bar{U}_n, (\lambda_1 I - B)\bar{U}_n \right\rangle.
\]

By assumptions H2c and H3, applying lemma 9 yields:

\[
||\bar{U}_n||^2 \longrightarrow \bar{U}, \quad \sum_{n=1}^{\infty} a_n \left\langle \bar{U}_n, (\lambda_1 I - B)\bar{U}_n \right\rangle = \sum_{n=1}^{\infty} a_n ||\bar{U}_n||^2 (\lambda_1 - \frac{\left\langle \bar{U}_n, B\bar{U}_n \right\rangle}{||\bar{U}_n||^2}) < \infty.
\]

As \( \sum_{n=1}^{\infty} a_n = \infty \), either \( ||\bar{U}_n|| \longrightarrow \infty \) or \( \sum_{n=1}^{\infty} a_n (\lambda_1 - \left\langle X_n, BX_n \right\rangle) < \infty \).

Step 2: convergence of \( \bar{U}_j = \left\langle \bar{U}_n, V_j \right\rangle \)

\[
\bar{U}_{n+1}^{(j)} = \left\langle V_j, I + a_n B \bar{U}_n \right\rangle = \left\langle V_j, \frac{1}{1 + \lambda_1 a_n} (I + a_n B + a_n (B_n - B)) \bar{U}_n \right\rangle
\]

\[
= \frac{1 + \lambda_1 a_n}{1 + \lambda_1 a_n} \bar{U}_n^{(j)} + \frac{a_n}{1 + \lambda_1 a_n} \left\langle V_j, (B_n - B) \bar{U}_n \right\rangle.
\]

a) For \( j > 1 \), as \( a_n \longrightarrow 0 \), there exists \( \alpha_n = O(a_n) > 0 \) such that for \( n \) sufficiently large:

\[
||\bar{U}_{n+1}^{(j)}|| \leq \frac{1 + \lambda_j a_n}{1 + \lambda_1 a_n} ||\bar{U}_n^{(j)}|| + a_n ||B_n - B|| ||\bar{U}_n||
\]

\[
\leq (1 - |\alpha_n|) ||\bar{U}_n^{(j)}|| + a_n ||B_n - B|| ||\bar{U}_n||.
\]

By H2c and as \( ||\bar{U}_n|| \) converges, applying lemma 9 yields:

\[
||\bar{U}_n^{(j)}|| \longrightarrow \bar{U}, \quad \sum_{n=1}^{\infty} \alpha_n ||\bar{U}_n^{(j)}|| < \infty. \quad \text{As } \sum_{n=1}^{\infty} a_n = \infty, \bar{U}^{(j)} = 0.
\]

b) For \( j = 1 \), by H2c and \( ||\bar{U}_n|| \longrightarrow \sqrt{U} \):

\[
\bar{U}_{n+1}^1 = \bar{U}_n^1 + \frac{a_n}{1 + \lambda_1 a_n} \left\langle V_1, (B_n - B) \bar{U}_n \right\rangle = \bar{U}_n^1 + \sum_{i=1}^{n} \frac{a_i}{1 + \lambda_1 a_i} \left\langle V_1, (B_i - B) \bar{U}_i \right\rangle
\]

\[
\longrightarrow \bar{U}_1^1 = \bar{U}_1^1 + \sum_{i=1}^{\infty} \frac{a_i}{1 + \lambda_1 a_i} \left\langle V_1, (B_i - B) \bar{U}_i \right\rangle.
\]
Now:
\[
\tilde{U}_{n+1}^1 = \left\langle V_1, \tilde{U}_{n+1} \right\rangle = \left\langle V_1, \prod_{i=1}^{n} \frac{I + a_i B_i}{1 + \lambda_i a_i} \tilde{U}_1 \right\rangle \overset{n \to \infty}{\longrightarrow} \left\langle V_1, \prod_{i=1}^{\infty} \frac{I + a_i B_i}{1 + \lambda_i a_i} \tilde{U}_1 \right\rangle = V'_1 Q S \tilde{U}_1 = \tilde{U}^1 \text{ with } S = \prod_{i=1}^{\infty} \frac{I + a_i B_i}{1 + \lambda_i a_i}.
\]

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

**Step 3**

Denote \(C_2 = \\{ \tilde{U}_1 \neq 0 \}\). Suppose \(\omega \in C_1 \cap C_2\).

Under H2c, there exists \(N\) such that \(\sum_{n=N}^{\infty} a_n \|B_n - B\| < 2\).

\[
V'_1 Q S = V'_1 Q \prod_{i=N}^{\infty} \frac{I + a_i B_i}{1 + \lambda_i a_i} \prod_{i=1}^{N-1} \frac{I + a_i B_i}{1 + \lambda_i a_i} = V'_1 Q R \prod_{i=1}^{N-1} \frac{I + a_i B_i}{1 + \lambda_i a_i} \text{ with } R = \prod_{i=N}^{\infty} \frac{I + a_i B_i}{1 + \lambda_i a_i}.
\]

Under H2d, \(V'_1 Q S \neq 0 \iff V'_1 Q R \neq 0\).

Denote \(C_n = \frac{a_n \|B_n - B\|}{1 + \lambda_i a_i}\) and \((W_n, n \geq N)\) the process \(\left( \tilde{U}_n, n \geq N \right)\) with \(W_N = V_1\).

As \(\|B\| = \lambda_1\), \(\|I + a_i B_i\| = 1 + \lambda_i a_i\). By step 2, as \(W_N = V_1\):

\[
W_{n+1}^1 = \left\langle V_1, W_{n+1} \right\rangle = 1 + \sum_{i=N}^{n} \frac{a_i}{1 + \lambda_i a_i} \left\langle V_1, (B_i - B) W_i \right\rangle \geq 1 - \sum_{i=N}^{n} C_i \|W_i\|.
\]

\[
\|W_i\| \leq \frac{\|I + a_i B_i\|}{1 + \lambda_i a_i} \|W_{i-1}\| \\
\leq \frac{\|I + a_i B_i\| + a_i \|B_i - B\|}{1 + \lambda_i a_i} \|W_{i-1}\| = (1 + C_{i-1}) \|W_{i-1}\| \\
\leq \prod_{i=N}^{n} (1 + C_i), \ i = N + 1, \ldots, n.
\]

As \(\sum_{n=N}^{\infty} C_n < \ln 2\), it follows that:

\[
W_{n+1}^1 \geq 1 - \sum_{i=N}^{n} C_i \prod_{l=N}^{i-1} (1 + C_l) = 1 - \left( \prod_{l=N}^{n} (1 + C_l) - 1 \right) \\
= 2 - \sum_{l=N}^{n} (1 + C_l) \geq 2 - \frac{\sum_{l=N}^{n} C_l}{1 + \lambda_i a_i} \geq 2 - \frac{\sum_{l=N}^{n} C_l}{1 + \lambda_i a_i} > 0.
\]

By step 2, \(W_n^1\) converges to \(\left\langle V_1, RV_1 \right\rangle = V'_1 Q R V_1\) which is therefore strictly positive, thus \(V'_1 Q R \neq 0\).

**Step 4: conclusion**

It follows that \(\left( \tilde{U}_n \right)\) converges to \(\tilde{U}^1 V_1 \neq 0\), therefore \(\frac{\tilde{U}_n}{\|\tilde{U}_n\|} = X_n\) converges to \(\pm V_1\), and by the conclusion of step 1, \(\sum_{n=1}^{\infty} a_n \left( \lambda_i - \langle X_n, B X_n \rangle \right) < \infty\).
Corollary 11

Moreover by H2c:
\[
\sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n, B_n X_n \rangle| = \sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n, (B_n - B) X_n \rangle - \langle X_n, B X_n \rangle| \\
\leq \sum_{n=1}^{\infty} a_n (\lambda_1 - \langle X_n, B X_n \rangle) + \sum_{n=1}^{\infty} a_n \|B_n - B\| < \infty. \quad \blacksquare
\]

Remark 10 Step 1 can be replaced by:
\[
\|\tilde{U}_{n+1}\| \leq \frac{\|I + a_n B_n\|}{1 + a_n \gamma_n} \|\tilde{U}_n\| \leq \left(1 + \frac{a_n \|B_n - B\|}{1 + a_n \gamma_n}\right) \|\tilde{U}_n\|.
\]
Under H2c, \(\|\tilde{U}_n\|\) converges a.s. Assumption \(\sum_{n=1}^{\infty} a_n^2 < \infty\) is not used and can be replaced by \(a_n \rightarrow 0\), but in this case, the convergence of \(\sum_{n=1}^{\infty} a_n (\lambda_1 - \langle B X_n, X_n \rangle)\) and \(\sum_{n=1}^{\infty} a_n |\lambda_1 - \langle B_n X_n, X_n \rangle|\) is not proved. \(\blacksquare\)

2.3.2 Simultaneous estimation of several eigenvectors

For \(i = 1, \ldots, r\), recursively define the process \((\tilde{X}_n^i)\) by:
\[
\tilde{Y}_{n+1}^i = (I + a_n B_n) \tilde{X}_n^i, \\
\tilde{X}_{n+1}^i = \tilde{Y}_{n+1}^i - \sum_{j<i} \left(\tilde{Y}_{n+1}^i, \tilde{X}_{n+1}^j \right) \frac{\tilde{X}_{n+1}^j}{\|\tilde{X}_{n+1}^j\|}.
\]

Note that \(\frac{\tilde{X}_{n+1}^i}{\|\tilde{X}_{n+1}^i\|} = X_n^i\).

Denote \(D_n^i = i B_n + \sum_{j=2}^{i} a_n^{j-1} i j B_n\).

Make the following assumptions:

(H1c') For \(i = 1, \ldots, r\), \(\|i B\| = \lambda_1 + \ldots + \lambda_i\).

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

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

Corollary 11 Suppose assumptions H1a,b,c', H2c,d', H3 and H5' hold. Then, for \(i = 1, \ldots, r\), almost surely \(X_n^i\) converges to \(\pm V_i \sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n^i, B X_n^i \rangle|\) and \(\sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n^i, B_n X_n^i \rangle|\) converge.

Proof

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

Let \(i \in \{1, \ldots, r\}\).

\[
i \tilde{X}_{n+1}^i = \tilde{X}_{n+1}^1 \wedge \ldots \wedge \tilde{X}_{n+1}^i = \tilde{Y}_{n+1}^1 \wedge \ldots \wedge \tilde{Y}_{n+1}^i = i (I + a_n B_n)^i \tilde{X}_n \\
= \left(I + a_n i^1 B_n + a_n \sum_{j=2}^{i} a_n^{j-1} i j B_n\right)^i \tilde{X}_n = (I + a_n D_n^i)^i \tilde{X}_n.
\]

By H2c and H3:
\[
\sum_{n=1}^{\infty} a_n \|D_n^i - i B\| = \sum_{n=1}^{\infty} a_n \|i^1 (B_n - B) + \sum_{j=2}^{i} a_n^{j-1} i j B_n\| \\
\leq C(i) \left(\sum_{n=1}^{\infty} a_n \|B_n - B\| + \sum_{j=2}^{i} \sum_{n=1}^{\infty} a_n^{j-1} \|B_n\|^j\right) < \infty.
\]
As $B$ is $Q$-symmetric with distinct eigenvalues, $\{1\}B$ has the same properties; $V_1 \wedge \ldots \wedge V_i$ is an eigenvector associated to its greatest eigenvalue $\lambda_1 + \ldots + \lambda_i$. Applying theorem 7 yields that:

$$iX_n \text{ converges to } \pm V_1 \wedge \ldots \wedge V_i,$$

$$\sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^{\infty} \lambda_i - \langle iB_i X_n, iX_n \rangle \right) \text{ and } \sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^{\infty} \lambda_i - \langle D_n^i X_n, iX_n \rangle \right) \text{ converge,}$$

which implies that $\sum_{n=1}^{\infty} a_n \sum_{i=1}^{\infty} \langle iB_n iX_n, iX_n \rangle$ converges.

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

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

Applying the same proof as that of corollary 4, step 3, yields:

$$\sum_{n=1}^{\infty} a_n \left| \lambda_i - \langle X_n^i, B X_n^i \rangle \right| < \infty. \text{ By H2c:}$$

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

### 3 Application to sequential principal component analysis of a data stream

Let $Z_{11}, \ldots, Z_{m_1}, Z_{21}, \ldots, Z_{m_2}, \ldots, Z_{n_1}, \ldots, Z_{n_{m_n}}, \ldots$ be an i.i.d sample of a random vector $Z$ in $\mathbb{R}^p$ whose components are denoted $Z^1, \ldots, Z^p$. Denote $M$ the metrics used for PCA and $B = M^{-\frac{1}{2}} E[(Z - E[Z])(Z - E[Z])'] M^{-\frac{1}{2}}$. Let $m$ belonging to $\mathbb{R}^p$ (in practice $m$ is an estimation of $E[Z]$); denoting $Z^c = Z - m$:

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

Denote $Z_{n-1}$ 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 $M$.

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

Note that the metrics used for orthonormalization is the identity because of the symmetrization.

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

$$Y_{n+1}^i = (I + a_n B_n) X_n^i,$$

$$T_{n+1}^i = Y_{n+1}^i - \sum_{j \leq i} (Y_{n+1}^j, X_{n+1}^j) X_{n+1}^j, \quad X_{n+1}^i = \frac{T_{n+1}^i}{\|T_{n+1}^i\|}.$$

Denote $Z_{ni}^c = Z_{ni} - m$, $Z_{n-1}^c = Z_{n-1} - m$. Take

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

Make the following assumptions:

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

15
Suppose assumptions H1b, H3’, H4a,b and H6a,b,c hold. Then

\[ \sum_{n=1}^{\infty} a_n \left| M_{n-1}^{\frac{1}{2}} \right| < \infty. \]

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

**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:

\[
E [B_n | T_n] - B = E \left[ M_{n-1}^{\frac{1}{2}} \left( \frac{1}{m_n} \sum_{i=1}^{m_n} Z_n^i Z_n^i' - Z_n^{-1} (Z_n^{-1})' \right) M_{n-1}^{\frac{1}{2}} | T_n \right]
- M^{\frac{1}{2}} \left( E [Z^e Z^c] - E[Z^e]E[Z^c]' \right) M^{\frac{1}{2}}
= M_{n-1}^{\frac{1}{2}} \left( E [Z^e Z^c] - Z_n^{-1} Z_n^{-1}' \right) M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}} \left( E [Z^e Z^c] - E[Z^e]E[Z^c]' \right) M_{n-1}^{\frac{1}{2}}
+ M^{\frac{1}{2}} \left( E [Z^e Z^c] - E[Z^e]E[Z^c]' \right) \left( M_{n-1}^{\frac{1}{2}} - M^{\frac{1}{2}} \right)
- M_{n-1}^{\frac{1}{2}} (Z_n^{-1} - E[Z^c]) Z_n^{-1} M_{n-1}^{\frac{1}{2}} - M_{n-1}^{\frac{1}{2}} E[Z^c] \left( Z_n^{-1} - E[Z^c] \right)' M_{n-1}^{\frac{1}{2}}.
\]

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

\[
\sum_{n=1}^{\infty} a_n E \left[ |Z_{n-1}^c - E[Z^c]| \right] = \sum_{n=1}^{\infty} a_n E \left[ |Z_{n-1} - E[Z]| \right] < \infty. \quad [9]
\]

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

By corollary 4, for \( k = 1, \ldots, r \), almost surely, \( X_n^k \) converges a.s. to one of the eigenvectors of \( B \).

Prove now that \( \lim_{n \to \infty} E[(X_n^k B_n V_k)^2 | T_n] > 0 \) a.s. on the set \( \{ X_n \to V_j \} \) 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 \( E[(X'_nB_nV_k)^2 \mid T_n] \) into the sum of three terms (1),(2),(3):

\[
E \left[ \left( X'_nM^{' \frac{1}{2}}_{n-1} - \sum_{i=1}^{m_n} Z^{c'}_{n_i} Z^{c'}_{n_i} - \sum_{n_i=1}^{m_n} \overline{Z}^{c'}_{n_i} \overline{Z}^{c'}_{n_i} \right) M^{' \frac{1}{2}}_{n-1} V_k \right] \mid T_n 
\]

\[
= E \left[ \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) - \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) \right)^2 \mid T_n \right] 
\]

\[
= E \left[ \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) \right)^2 \mid T_n \right] 
\]

(1)

\[
-2 \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'} \right) \left( Z^{c'}_{n_i=1} M^{' \frac{1}{2}}_{n-1} V_k \right) \frac{1}{m_n} \sum_{i=1}^{m_n} E \left[ \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) \mid T_n \right] 
\]

(2)

\[
+ \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'} \right) \left( Z^{c'}_{n_i=1} M^{' \frac{1}{2}}_{n-1} V_k \right) \right] \left( Z^{c'}_{n_i=1} M^{' \frac{1}{2}}_{n-1} V_k \right)^2 \times (3)
\]

Note that the two random variables \( R = V'_jM^{' \frac{1}{2}} Z^c \) and \( S = V'_iM^{' \frac{1}{2}} Z^c \) are uncorrelated, then \( E[RS] = E[R]E[S] \):

\[
E[(R - E[R])(S - E[S])] = E[V'_jM^{' \frac{1}{2}} (Z - E[Z])\cdot V'_iM^{' \frac{1}{2}} (Z - E[Z])]
\]

\[
= V'_jM^{' \frac{1}{2}} E \left[ (Z - E[Z])^2 (Z - E[Z]) \right] M^{' \frac{1}{2}} V_k = \lambda V'_jV_k = 0.
\]

Consider (1). Under H6b:

\[
(1) \quad = \frac{1}{m_n} \sum_{i=1}^{m_n} E \left[ \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) \left( X'_nM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) \mid T_n \right] 
\]

\[
= X'_nM^{' \frac{1}{2}}_{n-1} \frac{1}{m_n} \sum_{i=1}^{m_n} E \left[ \left( V'_iM^{' \frac{1}{2}}_{n-1} Z^{c'}_{n_i} \right) Z^{c'}_{n_i} Z^{c'}_{n_i} \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) \mid T_n \right] \left( Z^{c'}_{n_i} M^{' \frac{1}{2}}_{n-1} V_k \right) 
\]

\[
\stackrel{n \rightarrow +\infty}{\longrightarrow} V'_jM^{' \frac{1}{2}} E \left[ \left( V'_iM^{' \frac{1}{2}} Z^c \right) Z^c Z^c V_k \right] M^{' \frac{1}{2}} V_k
\]

\[
= E \left[ \left( V'_iM^{' \frac{1}{2}} Z^c \right)^2 \left( V'_jM^{' \frac{1}{2}} Z^c \right)^2 \right] a.s.
\]

Consider (2):

\[
(2) \rightarrow -2E \left[ V'_jM^{' \frac{1}{2}} Z^c \right] E \left[ Z^{c'} V'_iM^{' \frac{1}{2}} \right] E \left[ \left( V'_iM^{' \frac{1}{2}} Z^c \right) \left( Z^{c'} M^{' \frac{1}{2}} V_k \right) \right]
\]

\[
= -2E \left[ \left( V'_jM^{' \frac{1}{2}} Z^c \right) \left( V'_iM^{' \frac{1}{2}} Z^c \right) \right]^2 a.s.
\]

Consider (3):

\[
(3) \rightarrow \left( E \left[ V'_jM^{' \frac{1}{2}} Z^c \right] E \left[ V'_iM^{' \frac{1}{2}} Z^c \right] \right)^2 = E \left[ \left( V'_jM^{' \frac{1}{2}} Z^c \right) \left( V'_iM^{' \frac{1}{2}} Z^c \right) \right]^2 a.s.
\]

As a result:

\[
E[(X'_nB_nV_k)^2 \mid T_n] \stackrel{n \rightarrow +\infty}{\longrightarrow} E \left[ \left( V'_jM^{' \frac{1}{2}} Z^c \right)^2 \left( V'_iM^{' \frac{1}{2}} Z^c \right)^2 \right] - E \left[ \left( V'_jM^{' \frac{1}{2}} Z^c \right) \left( V'_iM^{' \frac{1}{2}} Z^c \right) \right]^2
\]

\[
= \text{Var} \left[ V'_jM^{' \frac{1}{2}} Z^c, V'_iM^{' \frac{1}{2}} Z^c \right] > 0 \text{ a.s. by H4b.}
\]
3.2 Use of all observations until the current step with different weights

At each step, all observations until 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 \(X_n^i\), \(i = 1, \ldots, r\), take now

\[
B_n = w_1 B_n^1 + w_2 B_n^2, \quad w_1 + w_2 = 1, \quad w_1 > 0, \quad w_2 \geq 0,
\]

with

\[
B_n^1 = M_{n-1}^\frac{1}{2} \left( \frac{1}{m_n} \sum_{j=1}^{m_n} Z_{nj}^c Z_{nj}' - Z_{n-1}^c Z_{n-1}' \right) M_{n-1}^\frac{1}{2},
\]

\[
B_n^2 = M_{n-1}^\frac{1}{2} \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} Z_{ij}^c Z_{ij}' - Z_{n-1}^c Z_{n-1}' \right) M_{n-1}^\frac{1}{2}.
\]

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

**Proof**

Verify the assumptions of corollary 6.

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

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

\[
B_n^2 = M_{n-1}^\frac{1}{2} C_{n-1} M_{n-1}^\frac{1}{2}
\]

with \(C = E[Z'] - E[Z] E[Z']\).

\[
B_n^2 - B = M_{n-1}^\frac{1}{2} C_{n-1} M_{n-1}^\frac{1}{2} - M^\frac{1}{2} C M^\frac{1}{2}
\]

\[
= (M_{n-1}^\frac{1}{2} - M^\frac{1}{2}) C_{n-1} M_{n-1}^\frac{1}{2} + M^\frac{1}{2} (C_{n-1} - C) M_{n-1}^\frac{1}{2} + M^\frac{1}{2} C (M_{n-1}^\frac{1}{2} - M^\frac{1}{2}).
\]

\[
C_{n-1} - C = 1 \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} Z_{ij} Z_{ij}' - E[ZZ'] - (Z_{n-1} - E[Z]) Z_{n-1}' - E[Z] (Z_{n-1} - E[Z])'.
\]

Under assumptions H3’ and H4a:

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

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

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

\[
\lim_{n \to \infty} E[(X_n^k B_n V_k)^2|T_n] = (\omega_1)^2 \lim_{n \to \infty} E \left[ \left( X_n^k B_n^2 V_k \right)^2 |T_n \right] = (\omega_1)^2 \text{Var}[V_j^M M^\frac{1}{2} Z^c, V_k^M M^\frac{1}{2} Z^c] > 0 \text{ a.s.} \]
3.3 Use of all observations until the current step with uniform weights

Take now

\[ B_n = M_n^\frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Z_{ij}^c Z_{ij}^c \cdot - Z_n^c Z_n^c \right) M_n^{-\frac{1}{2}}. \]

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 \( i = 1, \ldots, r \), almost surely \( X_n^i \) converges to \( \pm V_i, \sum_{n=1}^{\infty} a_n |\lambda_i - \langle X_n^i, BX_n^i \rangle| \) and \( \sum_{n=1}^{\infty} a_n |\lambda_i - \langle X_n^i, B_n X_n^i \rangle| \) 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, proof is similar to that of corollary 13 for \( B_n^2 \) without taking expectation.

In the particular case of normed principal component analysis, \( M \) is the diagonal matrix of the inverses of variances of \( Z_1, \ldots, Z_p \). Denote for \( j = 1, \ldots, p \), \( V_n^j \) the variance of the sample \( (Z_{11}^j, \ldots, Z_{m,n}^j) \) of \( Z^j \) and \( M_n \) the diagonal matrix of order \( p \) whose element \((j, j)\) is the inverse of \( \frac{\mu_n}{\sum_{i=1}^{n} m_i} V_n^j \) 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 gave theorems of almost sure convergence of a normed stochastic approximation process to eigenvectors of a \( Q \)-symmetric matrix \( B \) associated to eigenvalues in decreasing order, assuming that \( E[B_n | T_n] \) or \( B_n \) converges a.s. to \( B \). This extends previous results assuming \( B_n \) i.i.d. with \( E[B_n] = B \).

Several observations can be used at each step or all observations until the current step. These results are applied to online estimation of principal components in PCA when the data arrive continuously. In this case, the expectation and the variance of the variables are unknown and are estimated online in parallel with the estimation of principal components. To reduce the computing time and to avoid numerical explosions, we proposed to use symmetrisation (\( B \) is \( I \)-symmetric) and pseudo-centering with respect to a preliminary estimation of the expectation.

We made a first set of experiments: several processes, with or without symmetrization, with or without pseudo-centering, with different numbers of observations used at each step or with all observations until the current step, were compared on datasets or simulations (data not shown). It appeared that processes with symmetrization, pseudo-centering and use of all observations until the current step typically 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**


