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. 2019. hal-01844419v2

HAL Id: hal-01844419
https://hal.archives-ouvertes.fr/hal-01844419v2
Submitted on 15 May 2019

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$^{1,4,*}$, Abderrahman Skiredj$^{2,3,**}$

1 Université de Lorraine, CNRS, Inria$^1$, IECL$^2$, 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 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 matrix in $\mathbb{R}^p$, $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ 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 $z_1, \ldots, z_n, \ldots$; 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 $(Z_1, \ldots, Z_n, \ldots)$ is an i.i.d sample of a random vector $Z$. Let $\theta$ 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.
\[ \lambda_l. \] For \( l = 1, \ldots, r \), 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 metric. 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.

Many articles have been 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 Dufo [7]. Recall the normed matrix, using an i.i.d. sample of it. See for example the well-known algorithms of Benzécri [1], Krasulina [2], the problem of estimating eigenvectors and eigenvalues in decreasing order of the expectation stochastic approximation processes.

Let \( X_1 \), \( \ldots, X_n \) 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:

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

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 \( \sigma \)-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 \( B \);
- \( B_n = \omega_1 B_n^1 + \omega_2 B_n^2 \) with \( \omega_1 + \omega_2 = 1 \), \( B_n^1 \) and \( B_n^2 \) converging 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 \((X_n^i)\), \( i = 1, \ldots, r \) 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 \( B = M^{1/2} CM^{1/2} \); 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:

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

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 \( \sigma \)-field generated by the events before time \( n \).
\[ X_{n+1} = \frac{(I + a_n B_n) X_n}{\| (I + a_n B_n) X_n \|}. \]

2.1 First case

2.1.1 Theorem of almost sure convergence

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

\((H1a)\) \( B \) is \( Q \)-symmetric.
\((H1b)\) \( B \) has distinct eigenvalues: \( \lambda_1 > \lambda_2 > \ldots > \lambda_p \). Denote \( V_i \) a unit eigenvector of \( B \) corresponding to \( \lambda_i, i = 1, \ldots, p \).

\((H2a)\) 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 \( U_n = \langle X_n, B X_n \rangle, W_n = \langle X_n, B_n X_n \rangle. \)

**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 \( E_j = \{ U_n \rightarrow \lambda_j \} \), \( X_n \) 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 \( \bigcup_{j=2}^{\infty} E_j \), \( \lim \inf E \left[ (B_n X_n, V_1)^2 | T_n \right] > 0 \) \( a.s. \), 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:

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

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 \( \lim \inf E[\epsilon_{n+1}^2 | T_n] > 0 \) \( a.s. \)

Then, \( P(\Gamma) = 0. \)

**Proof of Theorem 1**

**Part 1: expression of \( X_{n+1} \)**

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

\[ \frac{1}{\| (I + a_n B_n) X_n \|} = 1 - a_n W_n - \frac{1}{2} a_n^2 \| B_n X_n \|^2 + \alpha_n, \quad \alpha_n = O(a_n^2). \]
\[ X_{n+1} = (I + a_n B_n) \left( 1 - a_n W_n - \frac{1}{2} a_n^2 \|B_n X_n\|^2 + a_n \right) X_n \]
\[ = (I + a_n (B_n - W_n I) + a_n \beta_n) X_n, \text{ with} \]
\[ \beta_n = -a_n W_n B_n - \frac{1}{2} a_n \|B_n X_n\|^2 I - \frac{1}{2} a_n^2 B_n \|B_n X_n\|^2 + a_n^{-1} a_n I + a_n B_n. \]
\[ X_{n+1} = (I + a_n (B - U_n I) + a_n \Gamma_n) X_n, \text{ with} \]
\[ \Gamma_n = (B_n - B) - (X_n, (B_n - B) X_n) I + \beta_n. \]

**Part 2: convergence of \( U_n \)**

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

As \( \|X_n\| = 1 \), we have with \( \mu_n = 2a_n \langle E[\Gamma_n |T_n], X_n, B X_n \rangle + a_n^2 \eta_n \):
\[ \langle (B - U_n I) X_n, B X_n \rangle = \|B X_n\|^2 - U_n^2 \geq \|B X_n - U_n X_n\|^2 \geq 0. \]
\[ E[U_{n+1}|T_n] \geq U_n + \mu_n \] a.s.
\[ E[U_{n+1} - \sum_{i=1}^{n-1} \mu_i |T_n] \geq U_n - \sum_{i=1}^{n-1} \mu_i \] a.s.

Prove the convergence of the submartingale \( U_n - \sum_{i=1}^{n-1} \mu_i \). By H2a:
\[ \|\beta_n\| \leq \frac{3}{2} a_n \|B_n\|^2 + \frac{1}{2} a_n^2 \|B_n\|^3 + a_n^{-1} |a_n| + |a_n| \|B_n\| = O(a_n) ; \]
\[ \|\Gamma_n\| \leq 2 \|B_n - B\| + \|\beta_n\| = O(1) ; \]
\[ \|\eta_n\| \leq 4 \|B\|^3 + 4 \|B\|^2 \|E[\Gamma_n |T_n]\| + \|B\| E[\|\Gamma_n\|^2 |T_n] = O(1). \]
\[ \|E[\Gamma_n |T_n]\| \leq 2 \|E[B_n |T_n] - B\| + \|E[\beta_n |T_n]\| \] a.s.

By H2b and H3:
\[ E \left[ \left| \sum_{i=1}^{n-1} \mu_i \right| \right] \leq 4 \|B\| E \left[ \sum_{i=1}^{\infty} a_i \|E[B_i |T_i] - B\| \right] + 2 \|B\| E \left[ \sum_{i=1}^{\infty} a_i \|E[\beta_i |T_i]\| \right] + E \left[ \sum_{i=1}^{\infty} a_i^2 \|\eta_i\| \right] \]
\[ < \infty. \]

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

**Part 3: convergence of \( X_n; = (X_n, V_j) \)**

Let \( \Gamma_n = (\Gamma_n X_n, V_j) \).
\[ X_{n+1}^j = \langle (I + a_n (B - U_n I) + a_n \Gamma_n) X_n, V_j \rangle = X_n^j (1 + a_n (\lambda_j - U_n)) + a_n \Gamma_n^j. \]
\[ (X_{n+1}^j)^2 = (X_n^j)^2 (1 + 2a_n (\lambda_j - U_n)) + a_n^2 (\lambda_j - U_n)^2 (X_n^j)^2 \]
\[ + 2a_n (1 + a_n (\lambda_j - U_n)) X_n^j \Gamma_n^j + a_n^2 (\Gamma_n^j)^2 \]
\[ = (X_n^j)^2 (1 + 2a_n (\lambda_j - U_n)) + a_n^2 ((\lambda_j - U_n) X_n^j + \Gamma_n^j)^2 + 2a_n X_n^j \Gamma_n^j \]
\[ = (X_n^j)^2 + 2 \sum_{i=1}^{n} a_i (\lambda_j - U_i) (X_i^j)^2 + \sum_{i=1}^{n} a_i^2 ((\lambda_j - U_i) X_i^j + \Gamma_i^j)^2 + 2 \sum_{i=1}^{n} a_i X_i^j \Gamma_i^j \]
\[ = \left( X_n^j \right)^2 + (1) + (2) + (3). \]
Study the convergence of the terms (2), (3), (1) of this decomposition.

(i) \( (2) = \sum_{l=1}^{\infty} a_l^2 \left( (\lambda_j - U_l)X_j^l + \Gamma_j^l \right)^2 \)
converges a.s. by H2a and H3.

(ii) Consider now (3).
\[
\sum_{l=1}^{\infty} a_l |X_l^j E[\Gamma_l^j | T_l]| \leq \sum_{l=1}^{\infty} a_l |(E[\Gamma_l^j | T_l]X_l, V_j)|
\leq \sum_{l=1}^{\infty} a_l |E[\Gamma_l^j | T_l]| \leq \sum_{l=1}^{\infty} a_l (2||E[B_l | T_l] - B|| + ||E[\beta | T_l]||).
\]

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

Let \(M^j_n = \sum_{l=1}^{n-1} a_l X_l^j (\Gamma_l^j - E[\Gamma_l^j | T_l])\); \((M^j_n)\) is a square-integrable martingale adapted to the filtration \((T_n)\); denote \((M^j_n)\) its increasing process. By H2a:
\[
(M^j)_{n+1} - (M^j)_n = E[\{M^j_{n+1} - M^j_n\}^2 | T_n] = a_n^2 E[\{X_l^j\}^2 (\Gamma_l^j - E[\Gamma_l^j | T_n])^2 | T_n]
\leq a_n^2 E[\|\Gamma_l^j\|^2 | T_n] \leq a_n^2 E[\|\Gamma_l^j\|^2 | T_n]
\]
is the general term of a convergent and uniformly bounded series; thus by Lemma 2 \((M^j_n)\) 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. 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 \(\omega\). Thus there exists \(A > 0\) such that:
\[
2 \sum_{l=N}^{\infty} a_l |\lambda_j - U_l| (X_l^j)^2 = 2 \left| \sum_{l=N}^{\infty} a_l (\lambda_j - U_l) (X_l^j)^2 \right| = \left| (X_{n+1}^j)^2 - (X_N^j)^2 - \sum_{l=N}^{\infty} a_l^2 (\lambda_j - U_l)X_l^j + \Gamma_l^j \right|^2 - 2 \sum_{l=N}^{\infty} a_l X_l^j \Gamma_l^j < A.
\]

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

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

**Part 4: convergence of \(X_n\)**

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

For \(j \neq i\), \(X_l^j\) converges to 0, therefore \((X_l^j)^2\) converges to 1 and since
\[
X_{n+1} - X_n = a_n ((B - U_n I) + \Gamma_n) X_n,
\]
\(X_{n+1} - X_n\) converges to 0 and the limit of \(X_n\) is \(V_i\) or \(-V_i\) on \(E_i = \{U_n \rightarrow \lambda_i\}\) (first assertion of Theorem 1).
Consider now the decomposition:
\[ \sum_{n=1}^{\infty} a_n (\lambda_i - W_n) = \sum_{n=1}^{\infty} a_n (\lambda_i - U_n) + \sum_{n=1}^{\infty} a_n \langle X_n, (B_n - E[B_n|T_n])X_n \rangle + \sum_{n=1}^{\infty} a_n \langle X_n, (E[B_n|T_n] - B)X_n \rangle. \]

(i) Using the decomposition of \((X_i^1)^2\) in step 3, the convergence of \((X_i^1)^2\) 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 \(M_n = \sum_{i=1}^{n-1} a_i \langle X_i, (B_i - E[B_i|T_i])X_i \rangle.\) \((M_n)\) is a square-integrable martingale adapted to the filtration \((T_n)\). Its increasing process \(\langle M_n \rangle\) converges; indeed:

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

is the general term of a convergent and uniformly bounded series. Thus \(\langle M \rangle\) 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 \(X_n^1\) to \(\pm V_1\)**

Suppose \(i > 1\).

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

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

(i) \(X_i^1\) is bounded.

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

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

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

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

\[ E[[(\Gamma_n X_n, V_1) - (B_n X_n, V_1)]^2 |T_n] \]

\[ \leq 2(X_n^1)^2 E[(\lambda_1 + \langle (B_n - B) X_n, X_n \rangle)^2 |T_n] + 2E[||\beta_n X_n, V_1||^2 |T_n] \]

\[ \leq c(X_n^1)^2 + 2E[||\beta_n||^2 |T_n] \rightarrow 0 \]

since \(||\beta_n|| = O(a_n)\) and \(X_n^1 \rightarrow 0\) a.s. Likewise:

\[ E[[(\Gamma_n X_n, V_1) - (B_n X_n, V_1))] \langle B_n X_n, V_1 \rangle |T_n] \rightarrow 0. \]

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:

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

$$ T_{n+1}^i = Y_{n+1}^i - \sum_{j<i} \langle Y_{n+1}^i, X_{n+1}^j \rangle X_{n+1}^j, \quad X_{n+1}^i = \frac{T_{n+1}^i}{||T_{n+1}^i||}. $$

$(X_{n+1}^1, ..., X_{n+1}^r)$ 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 $\bigcup_{j=i+1}^p \{ X_n^j \rightarrow \pm V_j \}$, $\liminf_{n \to \infty} E[||B_n X_n^i, V_i||^2] > 0$, then $X_n^i$ converges a.s. to $V_i$ or $-V_i$, $\sum_{n=1}^\infty a_n |\lambda_i - \langle BX_n^i, X_n^i \rangle|$ and $\sum_{n=1}^\infty a_n (\lambda_i - \langle B_n X_n^i, X_n^i \rangle)$ 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 $r \leq p$, let $^r \Lambda \mathbb{R}^p$ be the exterior power of order $r$ of $\mathbb{R}^p$, generated by the $C_p^r$ exterior products $e_{i_1} \wedge e_{i_2} \wedge ... \wedge e_{i_r}$, $i_1 < i_2 < ... < i_r \in \{1, ..., p\}$.

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

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

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

Let $||\cdot||$ 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 $^r \Lambda \mathbb{R}^p$.

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

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

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

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

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

3) 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 $\lambda_j$. Then the $C_p^r$ vectors $V_{i_1} \wedge ... \wedge V_{i_r}$, $1 \leq i_1 < ... < i_r \leq p$, are eigenvectors of $^r U$ respectively corresponding to the eigenvalues $\lambda_{i_1} + ... + \lambda_{i_r}$.

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

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

**Proof**

**Part 1**

For $i = 1, ..., r$, it follows from the orthogonality of $T_{n+1}^1, ..., T_{n+1}^i$ that:

$$ ||T_{n+1}^1 \wedge ... \wedge T_{n+1}^i|| = \prod_{l=1}^i ||T_{n+1}^l||. $$
Let \( iX_{n+1} = X_{n+1}^i \wedge \ldots \wedge X_{n+1}^i \) and \( D_n^i = iB_n + \sum_{j=2}^i a_j^i jB_n \). Then:

\[
iX_{n+1} = \frac{T_{n+1}^i \wedge \ldots \wedge T_{n+1}^i}{\|T_{n+1}^i \wedge \ldots \wedge T_{n+1}^i\|} = \frac{Y_{n+1}^i \wedge \ldots \wedge Y_{n+1}^i}{\|Y_{n+1}^i \wedge \ldots \wedge Y_{n+1}^i\|} = i\Lambda (I + a_n B_n) \ iX_n
\]

\[
= \left( I + a_n iB_n + \sum_{j=2}^i a_j^i jB_n \right) \ iX_n
\]

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

Moreover, as \( U \rightarrow i1 \) is a linear application, assumptions H2a,b and H3 yield that:

\[
E \left[ \sum_{n=1}^\infty a_n \|E[D_n|T_{n} - i1B]\| \right] = E \left[ \sum_{n=1}^\infty a_n \left\| E \left[ i1B_n - i1B + \sum_{j=2}^i a_j^i jB_n |T_{n}\right]\right. \right]
\]

\[
\leq E \left. \left\| \sum_{n=1}^\infty a_n \left( \|i1E[B_n - B |T_{n}\| + \sum_{j=2}^i a_j^i jE[\|jB_n\| |T_{n}\| \right] \right. \right] \right]
\]

\[
\leq c_1 E \left[ \sum_{n=1}^\infty a_n \left( \|E[B_n|T_{n}] - B\| + \sum_{j=2}^i a_j^i jE[\|B_n\|^i |T_{n}\| \right. \right] \right] < \infty.
\]

Applying first assertion of Theorem 1 yields that almost surely, \( iX_n \) converges to a unit eigenvector \( \pm V_j \wedge \ldots \wedge V_j \) of \( i1B \),

\[
\sum_{n=1}^\infty a_n (\lambda_{j_1} + \ldots + \lambda_{j_i} - \langle i1B \ iX_n, iX_n \rangle) \text{ and}
\]

\[
\sum_{n=1}^\infty a_n (\lambda_{j_1} + \ldots + \lambda_{j_i} - \langle D_n^i iX_n, iX_n \rangle) \text{ converge.}
\]

Moreover by H2a and H3, \( \sum_{n=1}^\infty a_n (\lambda_{j_1} + \ldots + \lambda_{j_i} - \langle i1B_n \ iX_n, iX_n \rangle) \) converges a.s.

**Part 2**

Suppose that for \( k = 1, \ldots, i-1 \), \( X_n^i \rightarrow \pm V_k \), which is verified for \( k = 1 \), and prove that \( X_n^i \rightarrow \pm V_j \).

1) Prove that there exists \( j > i-1 \) such that \( X_n^i \rightarrow \pm V_j \). Suppose that there exists \( k \in \{1, \ldots, i-1\} \) such that, for \( l = 1, \ldots, i, V_{j_l} \neq \pm V_k \); then, for \( l = 1, \ldots, i, \langle X_{n}^{i}, V_{j_l} \rangle \rightarrow 0 \) and \( \langle X_{n}^{i} \Lambda \ldots \Lambda X_{n}^{i}, V_{j_1} \Lambda \ldots \Lambda V_{j_i} \rangle \rightarrow 0 \), a contradiction. Therefore for all \( k \in \{1, \ldots, i-1\} \), there exists \( j_k \) such that \( V_{j_k} = \pm V_k \) and there exists \( j \) such that

\[
iX_n = X_n^i \wedge \ldots \wedge X_n^i \rightarrow \pm V_1 \Lambda \ldots \Lambda V_{i-1} \wedge V_j.
\]

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

\[
\langle X_{n}^{i} \wedge \ldots \wedge X_{n}^{i}, \pm V_1 \wedge \ldots \wedge V_{i-1} \wedge V_j \rangle,
\]

the limit of which is 1 as \( n \rightarrow \infty \), is \( \langle X_{n}^{i}, V_{i} \rangle \langle X_{n}^{i}, V_{i} \rangle \ldots \langle X_{n}^{i}, V_{i} \rangle \langle X_{n}^{i}, V_{j} \rangle \) obtained for \( \sigma = Id \). As for \( k = 1, \ldots, i-1, \langle X_{n}^{k}, V_{k} \rangle \rightarrow \pm 1 \), then \( \langle X_{n}^{i}, V_{j} \rangle \rightarrow \pm 1 \). Therefore \( X_{n}^{i} \rightarrow \pm V_j \).

2) Prove now that \( V_j = \pm V_i \). Suppose \( X_{n}^{i} \rightarrow \pm V_j \neq \pm V_i \).
Let $G_i$ be the set of permutations $\sigma$ of $\{1, ..., i\}$ with $\sigma = (\sigma(1), ..., \sigma(i))$ and $s(\sigma)$ the number of inversions of $\sigma$.

\[
\langle i^1 B_n(X_n^1 \wedge ... \wedge X_n^i), V_i \wedge ... \wedge V_i \rangle = \sum_{l=1}^{i} \langle X_n^1 \wedge ... \wedge B_n X_n^l \wedge ... \wedge X_n^i, V_i \wedge ... \wedge V_i \rangle
\]

\[
= \sum_{l=1}^{i} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle X_n^1, V_{\sigma(1)} \rangle ... \langle B_n X_n^l, V_{\sigma(l)} \rangle ... \langle X_n^i, V_{\sigma(i)} \rangle.
\]

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

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

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

and

\[
\liminf E \left[ \langle i^1 B_n(X_n^1 \wedge ... \wedge X_n^i), V_i \wedge ... \wedge V_i \rangle^2 \mid T_n \right] = \liminf E \left[ \langle B_n X_n^i, V_i \rangle^2 \mid T_n \right] > 0.
\]

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

\[
\liminf E \left[ \left( \sum_{j=2}^{i} a_n^{j-1} i^j B_n(X_n^1 \wedge ... \wedge X_n^j), V_1 \wedge ... \wedge V_i \right)^2 \mid T_n \right] = 0.
\]

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

Applying second assertion of Theorem 1 yields almost surely:

\[
X_n^1 \wedge ... \wedge X_n^i \xrightarrow{n \to +\infty} \pm V_1 \wedge ... \wedge V_i,
\]

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

converge.

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

**Part 3**

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

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 $\sum_{l=1}^{i} \lambda_l$ is the largest eigenvalue of $i^1 B$:

\[
\lambda_i - \langle B X_n^i, X_n^i \rangle = \left( \sum_{l=1}^{i} \lambda_l - \langle i^1 B X_n^i, X_n^i \rangle \right) - \left( \sum_{l=1}^{i-1} \lambda_l - \langle i^{-1,1} B^{-1,1} X_n^{-1,1}, X_n^{-1,1} \rangle \right)
\]

\[
= \left( \sum_{l=1}^{i} \lambda_l - \langle i^1 B X_n^i, X_n^i \rangle \right) - \left( \sum_{l=1}^{i-1} \lambda_l - \langle i^{-1,1} B^{-1,1} X_n^{-1,1}, X_n^{-1,1} \rangle \right).
\]
Almost surely, \( \sum_{n=1}^{\infty} a_n \left| \sum_{i=1}^{n} \lambda_i - \langle iB^j X_n, iX_n \rangle \right| < \infty \), then \( \sum_{n=1}^{\infty} a_n \left| \lambda_i - \langle BX_i, X_n \rangle \right| < \infty \).

Likewise, as almost surely \( \sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^{n} \lambda_i - \langle iB^j X_n, iX_n \rangle \right) \) converges, then \( \sum_{n=1}^{\infty} a_n (\lambda_i - \langle B_n^j X_n, X_n \rangle) \) converges. \( \blacksquare \)

2.2 Second case

Consider the same processes \((X_n, X^i_n)\) 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.
- \((H2b')\) \( E \left[ \sum_{n=1}^{\infty} a_n \|E [B_n^1 | T_n] - B^1 \| \right] < \infty \).
- \((H2c')\) \( B_n^2 \) \( T_n \)-measurable, \( B_n^2 \to B, E \left[ \sum_{n=1}^{\infty} \|B_n^2_n - B\| \right] < \infty \) 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 \( \lim_{n \to \infty} \omega_{1n} = \omega_1 > 0 \) and on \( j \geq 2 \), \( \lim \inf_{n \to \infty} E \left[ (B_n^1 X_n, V_1^j)^2 | T_n \right] > 0 \) 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 \( \lim \inf_{n \to \infty} E \left[ (B_n X_n, V_1)^2 | T_n \right] > 0 \) a.s. when \( \lim_n X_n = \pm V_j \neq \pm V_i \).

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

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

\[
\lim \inf_{n \to \infty} E \left[ (\omega_{1n} B_n^1 + \omega_{2n} B_n^2) X_n, V_1 \right] = (\omega_{1n})^2 \lim \inf_{n \to \infty} E \left[ (B_n^1 X_n, V_1)^2 | T_n \right] > 0 \text{ a.s.} \quad \blacksquare
\]

2.2.2 Simultaneous estimation of several eigenvectors

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

1. For \( i = 1, \ldots, r \), 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 \( i = 1, \ldots, r \), a.s. on \( \bigcup_{j=1}^{r} \{ X_n^i \to \pm V_j \} \), \( \lim \inf_{n \to \infty} E \left[ (B_n^1 X_n^i, V_1)^2 | T_n \right] > 0 \), 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 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 $(\bar{X}_n)$ such that

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

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

$$\bar{U}_{n+1} = \frac{\bar{X}_{n+1}}{\prod_{i=1}^{n} (1 + \lambda_1 a_i)} = \frac{I + a_n B_n \bar{U}_n}{1 + \lambda_1 a_n}$$

$$= \bar{U}_n + \frac{a_n}{1 + \lambda_1 a_n} (B_n \bar{U}_n - \lambda_1 \bar{U}_n), \quad \bar{U}_1 = \bar{X}_1.$$

Note that $\frac{\bar{U}_n}{\|\bar{U}_n\|} = \frac{\bar{X}_n}{\|\bar{X}_n\|} = X_n$. 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 $B_1, \ldots, B_n, \ldots$.

**Theorem 7** Suppose assumptions H1a,b,c, H2c,d, H3 and H5 hold. Almost surely, $\bar{U}_n$ converges to a random vector colinear to $V_1$, therefore $X_n$ converges to $\pm V_1$, $\sum_{n=1}^{\infty} a_n (\lambda_1 - \langle B X_n, X_n \rangle)$ 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 $n$, $(z_n)$, $(\alpha_n)$, $(\beta_n)$ and $(\gamma_n)$ are four sequences of non-negative numbers such that:

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

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

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

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

**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.

**Part 1**

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

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

$$- 2 \frac{a_n}{1 + \lambda_1 a_n} \langle \bar{U}_n, (\lambda_1 I - B) \bar{U}_n \rangle.$$
\[ \lambda_1 I - B \text{ is a non-negative } Q\text{-symmetric matrix, with eigenvalues } 0, \lambda_1 - \lambda_2, ..., \lambda_1 - \lambda_p. \]

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

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

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

By assumptions H2c and H3, applying Lemma 9 yields:

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

As \( \sum_{n=1}^{\infty} a_n = \infty \), either \( ||\tilde{U}_n|| \rightarrow 0 \) or \( \sum_{n=1}^{\infty} a_n (\lambda_1 - \langle X_n, B X_n \rangle) < \infty. \)

**Part 2: convergence of \( \tilde{U}_n = \langle \tilde{U}_n, V_j \rangle \)**

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

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

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

\[ \left| \tilde{U}_{n+1}^j \right| \leq \frac{1 + \lambda_1 a_n}{1 + \lambda_1 a_n} \left| \tilde{U}_n^j \right| + a_n \|B_n - B\| \left| \tilde{U}_n \right| \]

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

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

\[ \left| \tilde{U}_n^j \right| \rightarrow \sigma, \quad \sum_{n=1}^{\infty} a_n \left| \tilde{U}_n^j \right| < \infty. \] As \( \sum_{n=1}^{\infty} a_n = \infty, \tilde{U}^j = 0. \)

b) For \( j = 1, \) by H2c and \( ||\tilde{U}_n|| \rightarrow \sqrt{U}; \)

\[ \tilde{U}_{n+1}^1 = \tilde{U}_n^1 + \frac{a_n}{1 + \lambda_1 a_n} \langle V_1, (B_n - B) \tilde{U}_n \rangle = \tilde{U}_1^1 + \sum_{i=1}^{n} \frac{a_i}{1 + \lambda_1 a_i} \langle V_1, (B_i - B) \tilde{U}_i \rangle \]

\[ \sum_{n=1}^{\infty} \tilde{U}_n^1 = \tilde{U}_1^1 + \sum_{i=1}^{\infty} \frac{a_i}{1 + \lambda_1 a_i} \langle V_1, (B_i - B) \tilde{U}_i \rangle. \]

Now:

\[ \tilde{U}_{n+1}^1 = \langle V_1, \tilde{U}_{n+1} \rangle = \langle V_1, \prod_{i=1}^{n} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \tilde{U}_1 \rangle \rightarrow \langle V_1, \prod_{i=1}^{\infty} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \tilde{U}_1 \rangle \]

\[ = V_1^i U_1 \text{ with } S = \prod_{i=1}^{\infty} \frac{I + a_i B_i}{1 + \lambda_1 a_i}. \]

As \( U_1 \) is absolutely continuous, if \( V_1^i QS \neq 0, P \left( V_1^i QS U_1 = 0 \mid S \right) = 0, \) then \( P \left( \tilde{U}^1 = 0 \right) = 0. \) Prove that \( V_1^i QS \neq 0. \)

**Part 3**
Denote $C_2 = \{ \bar{U}_1 \neq 0 \}$. Suppose $\omega \in C_1 \cap C_2$.

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

$$V'_1QS = V'_1 QR \prod_{i=N}^{\infty} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \prod_{i=1}^{N-1} \frac{I + a_i B_i}{1 + \lambda_1 a_i}$$

$$= V'_1 QR \prod_{i=1}^{N-1} \frac{I + a_i B_i}{1 + \lambda_1 a_i} \text{ with } R = \prod_{i=N}^{\infty} \frac{I + a_i B_i}{1 + \lambda_1 a_i}.$$  

Under H2d, $V'_1QS \neq 0 \Leftrightarrow V'_1QR \neq 0$.

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

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

$$W_{n+1}^1 = \langle V_1, W_{n+1} \rangle = 1 + \sum_{i=N}^{n} \frac{a_i}{1 + \lambda_1 a_i} \langle V_1, (B_i - B) W_i \rangle \geq 1 - \sum_{i=N}^{n} C_i \|W_i\|.$$  

$$\|W_i\| \leq \frac{\|I + a_{i-1} B_i - B\|}{1 + \lambda_1 a_{i-1}} \|W_{i-1}\|$$

$$\leq \frac{\|I + a_{i-1} B + a_{i-1} B_{i-1} - B\|}{1 + \lambda_1 a_{i-1}} \|W_{i-1}\| = (1 + C_{i-1}) \|W_{i-1}\|$$

$$\leq \prod_{i=N}^{n} (1 + C_i), \ i = N + 1, \ldots, n.$$

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

$$W_{n+1}^1 \geq 1 - \sum_{i=N}^{n} C_i \prod_{l=N}^{i-1} (1 + C_l) = 1 - \left( \prod_{l=N}^{n} (1 + C_l) - 1 \right)$$

$$= 2 - \prod_{i=N}^{n} (1 + C_l) \geq 2 - e^{\sum_{i=N}^{n} C_i} \geq 2 - e^{\sum_{i=N}^{\infty} C_i} > 0.$$  

By Part 2, $W_{n}^1$ converges to $\langle V_1, RV_1 \rangle = V'_1QRV_1$ which is therefore strictly positive, thus $V'_1QR \neq 0$.

**Part 4: conclusion**

It follows that $(\bar{U}_n)$ converges to $\bar{U}_1 V_1 \neq 0$, therefore $\|\bar{U}_n\| = X_n$ converges to $\pm V_1$, and by the conclusion of Part 1, $\sum_{n=1}^{\infty} a_n (\lambda_1 - \langle X_n, BX_n \rangle) < \infty$.

Moreover by H2c:

$$\sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n, B_n X_n \rangle| = \sum_{n=1}^{\infty} a_n |\lambda_1 - \langle X_n, (B_n - B) X_n \rangle - \langle X_n, BX_n \rangle|$$

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

**Remark 10** Part 1 can be replaced by:

$$\|\bar{U}_{n+1}\| \leq \frac{\|I + a_n B_n\|}{1 + \lambda_1 a_n} \|\bar{U}_n\| \leq \left( 1 + \frac{a_n \|B_n - B\|}{1 + \lambda_1 a_n} \right) \|\bar{U}_n\|.$$  

Under H2c, $\|\bar{U}_n\|$ 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 $\sum_{n=1}^{\infty} a_n (\lambda_1 - \langle BX_n, X_n \rangle)$ and $\sum_{n=1}^{\infty} a_n |\lambda_1 - \langle B_nX_n, X_n \rangle|$ is not proved.
2.3.2 Simultaneous estimation of several eigenvectors

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

\[
\begin{align*}
\tilde{Y}_{n+1}^i &= (I + a_n B_n) \tilde{X}^i_n, \\
\tilde{X}_{n+1}^i &= \tilde{Y}_{n+1}^i - \sum_{j<i} \left( \frac{\tilde{Y}_{n+1}^i}{\| \tilde{X}^j_{n+1} \|} - \frac{\tilde{X}_{n+1}^j}{\| \tilde{X}^j_{n+1} \|} \right) \tilde{X}^j_{n+1}.
\end{align*}
\]

Note that $\frac{\tilde{X}^i_n}{\| \tilde{X}^i_n \|} = X^i_n$.

Let $D_n^i = i^1 B_n + \sum_{j=2}^{i} a^{i-1}_{j} ij B_n$.

Make the following assumptions:

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

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

(H5') For $i = 1, ..., r$, $X^i_1$ 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 $X^i_n$ converges to $\pm V_i$, $\sum_{n=1}^{\infty} a_n |\lambda_i - \langle X^i_n, B X^i_n \rangle|$ and $\sum_{n=1}^{\infty} a_n |\lambda_i - \langle X^i_n, B_n X^i_n \rangle|$ converge.

**Proof**

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

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

\[
i \tilde{X}_{n+1}^i = \tilde{X}_{n+1}^1 \land ... \land \tilde{X}_{n+1}^i = \tilde{Y}_{n+1}^1 \land ... \land \tilde{Y}_{n+1}^i = (I + a_n B_n) \tilde{X}^i_n,
\]

By H2c and H3:

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

As $B$ is $Q$-symmetric with distinct eigenvalues, $V^1 \land ... \land V^i$ is an eigenvector corresponding to its largest eigenvalue $\lambda_1 + ... + \lambda_i$. Applying Theorem 7 yields that:

\[
i X^i_n \text{ converges to } \pm V^1 \land ... \land V^i,
\]

\[
\sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^{i} \lambda_i - \langle i^1 B^i X_n, i^i X_n \rangle \right) \text{ and } \sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^{i} \lambda_i - \langle D_n^{i} X^i_n, i^i X_n \rangle \right)
\]

which implies that $\sum_{n=1}^{\infty} a_n \left( \sum_{i=1}^{i} \lambda_i - \langle i^1 B^i X_n, i^i X_n \rangle \right)$ converges.

Suppose that, for $k = 1, ..., i - 1$, $X^k_n$ 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 \( \langle X_n^i, V_i \rangle \) converges to \( \pm 1 \).

Applying the same proof as that of Corollary 4, Part 3, yields:

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

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

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

Let \( Z_{11}, \ldots, Z_{m_1}, Z_{21}, \ldots, Z_{m_2}, \ldots, Z_{n1}, \ldots, Z_{n_m}, \ldots \) be an i.i.d sample of a random vector \( Z \) in \( \mathbb{R}^p \) whose components are denoted \( Z_1, \ldots, Z_p \). 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 \( Z' = Z - m \), we have:

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

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

#### 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 \( \langle X_n^i \rangle, i = 1, \ldots, r \), by

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

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

Let \( Z_{ni}^c = Z_{ni} - m, Z_{n-1} = Z_{n-1} - m \). We define:

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

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 \( \sup_n \| M_n^\frac{1}{2} \| < d \).

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

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

**Corollary 12.** Suppose assumptions H1b, H3', H4a,b and H6a,b,c hold. Then \( X_n^i \) converges a.s. to \( \pm V_i \),

\[
\sum_{n=1}^{\infty} a_n \left| \lambda_i - (X_n^i)' B X_n^i \right| \quad \text{and} \quad \sum_{n=1}^{\infty} a_n \left( \lambda_i - (X_n^i)' B_n X_n^i \right) \quad \text{converge a.s. for} \quad i = 1, \ldots, r.
\]
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:

\[
E [B_n \mid T_n] = B
= E \left[ M_{n-1}^{\hat{1}} \left( \frac{1}{m_n} \sum_{i=1}^{m_n} Z_{ni}^c Z_{ni}^c \right) \mid T_n \right]
- M_{n-1}^{\hat{1}} \left( \left[ Z^c Z^c \right] - E[Z^c]E[Z^c]' \right) M_{n-1}^{\hat{1}}
= M_{n-1}^{\hat{1}} \left( \left[ Z^c Z^c \right] - E[Z^c]E[Z^c]' \right) M_{n-1}^{\hat{1}} - M_{n-1}^{\hat{1}} \left( \left[ Z^c Z^c \right] - E[Z^c]E[Z^c]' \right) M_{n-1}^{\hat{1}}
- (M_{n-1}^{\hat{1}} - M_{n-1}^{\hat{1}}) \left( \left[ Z^c Z^c \right] - E[Z^c]E[Z^c]' \right) M_{n-1}^{\hat{1}}
+ M_{n-1}^{\hat{1}} \left( \left[ Z^c Z^c \right] - E[Z^c]E[Z^c]' \right) \left( M_{n-1}^{\hat{1}} - M_{n-1}^{\hat{1}} \right)
- M_{n-1}^{\hat{1}} \left( \left[ Z^c Z^c \right] - E[Z^c] \right) M_{n-1}^{\hat{1}} - M_{n-1}^{\hat{1}} \left[ E[Z^c] \left( \left[ Z^c \right] - E[Z^c] \right) \right] M_{n-1}^{\hat{1}}.
\]

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

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

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

By Corollary 4, for $k = 1, \ldots, r$, $X_n^k$ converges a.s. to one of the eigenvectors of $B$.

Prove now that $\lim_{n \to \infty} E[(X_n^kB_nV_k)^2 \mid T_n] > 0$ 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 $E[(X_n^kB_nV_k)^2 \mid T_n]$ into the sum of three terms (1),(2),(3):

\[
E \left[ \left( X_n^t M_{n-1}^{\hat{1}} \left( \frac{1}{m_n} \sum_{i=1}^{m_n} Z_{ni} Z_{ni}^{\prime} \right) M_{n-1}^{\hat{1}} V_k \right)^2 \mid T_n \right]
= E \left[ \left( \frac{1}{m_n} \sum_{i=1}^{m_n} X_n^t M_{n-1}^{\hat{1}} Z_{ni}^c \left( Z_{ni}^c M_{n-1}^{\hat{1}} V_k \right) - \left( X_n^t M_{n-1}^{\hat{1}} Z_{n-1}^c \right) \left( Z_{n-1}^c M_{n-1}^{\hat{1}} V_k \right) \right)^2 \mid T_n \right]
= E \left[ \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \left( X_n^t M_{n-1}^{\hat{1}} Z_{ni}^c \right) \left( Z_{n-1}^c M_{n-1}^{\hat{1}} V_k \right) \right)^2 \mid T_n \right] \quad (1)
-2 \left( X_n^t M_{n-1}^{\hat{1}} Z_{n-1}^c \right) \left( Z_{n-1}^c M_{n-1}^{\hat{1}} V_k \right) \frac{1}{m_n} \sum_{i=1}^{m_n} E \left[ \left( X_n^t M_{n-1}^{\hat{1}} Z_{ni}^c \right) \left( Z_{ni}^c M_{n-1}^{\hat{1}} V_k \right) \mid T_n \right] \quad (2)
+ \left( X_n^t M_{n-1}^{\hat{1}} Z_{n-1}^c \right)^2 \left( Z_{n-1}^c M_{n-1}^{\hat{1}} V_k \right)^2. \quad (3)
\]

Note that the two random variables $R = V_j^t M_{n}^{\hat{1}} Z^c$ and $S = V_j^t M_{n}^{\hat{1}} Z^c$ are uncorrelated, then $E[RS] = E[R]E[S]$:

\[
E[(R - E[R])(S - E[S])] = E[V_j^t M_{n}^{\hat{1}} (Z - E[Z]) V_j^t M_{n}^{\hat{1}} (Z - E[Z])]
= V_j^t M_{n}^{\hat{1}} E \left[ (Z - E[Z]) \left( Z - E[Z] \right)^{\prime} \right] M_{n}^{\hat{1}} V_k = \lambda_k V_j^t V_k = 0.
\]
Under H6b, we have:

\[
(1) \quad \frac{1}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} E \left[ \left( X_n' M_n^{-\frac{1}{2}} Z_n^c \right) \left( Z_n^c M_n^{-\frac{1}{2}} V_k \right) \left( X_n' M_n^{-\frac{1}{2}} Z_n^c \right) \left( Z_n^c M_n^{-\frac{1}{2}} V_k \right) \right] | T_n \]
\]
\[
= X_n' M_n^{-\frac{1}{2}} \frac{1}{m_n^2} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} E \left[ \left( V_k' M_n^{-\frac{1}{2}} Z_n^c \right) Z_n^c Z_n^c \left( Z_n^c M_n^{-\frac{1}{2}} V_k \right) \right] | T_n \] \quad \text{for } n \to +\infty
\]
\[
\longrightarrow V_j' M_n^{-\frac{1}{2}} E \left[ \left( V_k' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \left( Z_n^c M_n^{-\frac{1}{2}} V_k \right) \right] M_n^{-\frac{1}{2}} V_j
\]
\[
= E \left[ \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right) \right] \text{a.s.}
\]

(2) \quad \longrightarrow -2E \left[ V_j' M_n^{-\frac{1}{2}} Z_n^c \right] E \left[ Z_n^c \left( V_k' M_n^{-\frac{1}{2}} V_k \right) \right] E \left[ \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right) A S M_n^{-\frac{1}{2}} V_k \right]
\]
\[
= -2E \left[ \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \left( V_k' M_n^{-\frac{1}{2}} Z_n^c \right) \right] \text{a.s.}
\]

(3) \quad \longrightarrow \left( E \left[ V_j' M_n^{-\frac{1}{2}} Z_n^c \right] E \left[ V_k' M_n^{-\frac{1}{2}} Z_n^c \right] \right)^2 \quad = E \left[ \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \left( V_k' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \right] \text{a.s.}
\]

As a result:
\[
E(\left( X_n' B_n V_k \right)^2 | T_n) \quad \longrightarrow \quad E \left[ \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \left( V_k' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \right] - E \left[ \left( V_j' M_n^{-\frac{1}{2}} Z_n^c \right) \left( V_k' M_n^{-\frac{1}{2}} Z_n^c \right)^2 \right] \geq 0 \text{ a.s. by H4b.} \]

### 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 \( \left( X_n^i \right), i = 1, \ldots, r \), we take now:

\[
B_n = \sum_{i=1}^{r} w_i X_n^i + w_2 B_{n-1}, \quad w_1 + w_2 = 1, w_1 > 0, w_2 > 0,
\]

\[
B_n^1 = M_n^{-\frac{1}{2}} \left( \frac{1}{m_n} \sum_{j=1}^{m_n} Z_n Z_n^c - Z_n Z_n^c \right) M_n^{-\frac{1}{2}},
\]

\[
B_n^2 = M_n^{-\frac{1}{2}} \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} Z_n Z_n^c - Z_n Z_n^c \right) M_n^{-\frac{1}{2}}.
\]

**Corollary 13** Suppose assumptions H1b, H3', H4a,b and H6a,b,c hold. Then \( X_n^i \) converges a.s. to \( \pm V_i \),

\[
\sum_{i=1}^{\infty} a_n \left| \lambda_i - (X_n^i)' X_n^i \right| \quad \text{and} \quad \sum_{n=1}^{\infty} a_n (\lambda_i - (X_n^i)' B_n X_n^i) \text{ converge a.s. for } i = 1, \ldots, r.
\]

**Proof**

Verify the assumptions of Corollary 6.

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

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

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

\[
C_{n-1} = \frac{1}{m_n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} Z_n Z_n^c - Z_n Z_n^c = \frac{1}{m_n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} Z_n Z_n^c - Z_n Z_n^c.
\]
When the data arrive continually. In this case, the expectation and the variance of the variables are unknown.

It is assumed that the process to eigenvectors of a

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

\[ \sum_{n=1}^{\infty} b_n E[|Z_n - E[Z]|] = \infty, \sum_{n=1}^{\infty} a_n E[|Z_n Z_n' - E[Z] Z_n'|] = \infty. \]

Under assumptions H3' and H4a:

\[ \sum_{n=1}^{\infty} a_n E[|Z_{n-1} - E[Z]|] < \infty, \sum_{n=1}^{\infty} a_n E[|Z_{n-1}' - E[Z_n']|] < \infty. \]

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:

\[ \lim_{n \to \infty} E[(X_n^k B_n V_k)^2 |T_n] = (\omega_1)^2 \lim_{n \to \infty} E \left[ (X_n^k B_n V_k)^2 |T_n \right] \]

\[ = (\omega_1)^2 \operatorname{Var}[V_j M_\frac{1}{2} Z^c V_j' M_\frac{1}{2} Z^c] > 0 \text{ a.s.} \]

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

We define now:

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

Make the following assumptions:

(H4c) Z has 4th order moments.

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

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

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 \( j = 1, ..., p \), \( V_j \) the variance of the sample \( (Z_{i1}, ..., Z_{im}) \) of \( Z^j \) and \( M_n \) the diagonal matrix of order \( p \) whose element \( (j, j) \) is the inverse of \( \frac{\mu_n}{\mu_n - 1} V_j \) 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 \( E[|B_n - T_n|] \) or \( B_n \) converges a.s. to \( B \). This extends previous results assuming \( B_n \) i.i.d. with \( E[B_n | T_n] = B \).

These results have been applied to online estimation of principal components of PCA of a random vector when the data arrive continually. 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


