



HAL
open science

Stochastic approximation of eigenvectors and eigenvalues of the Q -symmetric expectation of a random matrix

Jean-Marie Monnez

► **To cite this version:**

Jean-Marie Monnez. Stochastic approximation of eigenvectors and eigenvalues of the Q -symmetric expectation of a random matrix. *Communications in Statistics - Theory and Methods*, 2024, 53 (5), pp.1669-1683. 10.1080/03610926.2022.2107225 . hal-03956687

HAL Id: hal-03956687

<https://hal.science/hal-03956687v1>

Submitted on 25 Jan 2023

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.

Stochastic approximation of eigenvectors and eigenvalues of the Q-symmetric expectation of a random matrix

Jean-Marie Monnez^{a*}

^aUniversité de Lorraine, CNRS, Inria¹, IECL², F-54000 Nancy, France

*jean-marie.monnez@univ-lorraine.fr

Abstract. We establish an almost sure convergence theorem of the stochastic approximation process of Oja for estimating eigenvectors of the Q -symmetric expectation of a random matrix, under a correlation model between the incoming random matrices. This theorem generalizes previous theorems and extends them to the case where the metric Q is unknown and estimated online in parallel. We apply it to streaming PCA of a random vector Z , when a mini-batch of observations of Z is used at each step or all the observations up to the current step. We deal with the case of streaming generalized canonical correlation analysis, with a metric estimated online in parallel.

Keywords: big data; online learning; stochastic approximation; streaming gCCA; streaming PCA

1. Introduction

Let Q be a positive definite symmetric $p \times p$ matrix called metric, $\langle \cdot, \cdot \rangle$ be the inner product and $\|\cdot\|$ the norm induced by Q . For vectors in \mathbb{R}^p , Q -orthogonal and Q -normed respectively denote orthogonal and normed with respect to the metric Q . Recall that a $p \times p$ matrix B is Q -symmetric if QB is symmetric; then B has p real eigenvalues and there exists a Q -orthonormal basis of \mathbb{R}^p comprised of eigenvectors of B . Let \mathbb{R}^{p*} be the dual space of \mathbb{R}^p . Let $\|B\|$ denote the spectral norm of B . Let A^T denote the transpose of a matrix A . Let (\cdot) denote a sequence of matrices or vectors or reals depending on the context. The abbreviation a.s. stands for almost surely.

¹ Inria, project-team BIGS

² Institut Elie Cartan de Lorraine, BP 70239, F-54506 Vandoeuvre-lès-Nancy, France

In the context of streaming data or big data since one can process the latter sequentially as a data stream, stochastic approximation algorithms can be used to estimate online, for example, parameters of a regression function (Lalloué, Monnez and Albuisson 2022), centers of clusters in unsupervised classification (Cardot, Cénac and Monnez 2012), principal components in a principal component analysis (PCA) (Cardot and Degras 2018). Each incoming observation vector is used to update the estimate sequence until the latter converges to the quantity of interest. When using such processes, it is not necessary to store the data and, due to the relative simplicity of the computation involved, a much greater number of data than with non-sequential methods can be taken into account during the same amount of time. Moreover, this type of method uses less memory space than a batch method (Balsubramani, Dasgupta and Freund 2013). In this context, we deal with the problem of estimating by a recursive process eigenvectors and corresponding eigenvalues in decreasing order of the unknown Q -symmetric expectation B of a random $p \times p$ matrix.

Many articles have been devoted to this problem when B is a symmetric matrix ($Q = I$). See among others the algorithms of Benzécri (1969), Krasulina (1970), Oja (1982), Karhunen and Oja (1982), Oja and Karhunen (1985), Brandière (1994, 1995, 1998), Brandière and Duflo (1996) and Duflo (1997) in the case of PCA. Oja and Karhunen (1985) proved the a.s. convergence to a normed eigenvector V_1 , corresponding to the largest eigenvalue λ_1 of B of unit multiplicity, of the process (X_n) defined recursively at step n by $X_{n+1} = \frac{(I+a_n B_n)X_n}{\|(I+a_n B_n)X_n\|}$, assuming the random $p \times p$ matrices $B_n, n \geq 1$, mutually independent with expectation B and a.s. bounded ($\|B_n\| \leq b \in \mathbb{R}^+$), with $a_n > 0, \sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} a_n^2 < \infty$. They proved the convergence of processes $(X_n^i), 1 \leq i \leq p$, of the same type, for estimating, after orthonormalization, eigenvectors V_i corresponding to the eigenvalues λ_i in decreasing order. The assumption of independence of the random matrices B_n may not hold, for instance in the streaming PCA of a random vector Z in \mathbb{R}^p , when the expectation of Z and the metric M in \mathbb{R}^p are estimated online (Section 4). Using different methods of proof, Monnez and Skiredj (2021) studied two cases where the matrices B_n are not independent, the first when $E[B_n|T_n]$ converges a.s. to B , T_n being the σ -field generated by B_1, \dots, B_{n-1} , the second when B_n converges a.s. to B .

The present study extends and generalizes previous results in two directions - the metric Q is supposed unknown and the matrices B_n not independent -, proves in a single theorem

the convergence of processes using a mini-batch of observations at each step and of those using all the observations up to the current step, and gives applications to streaming PCA and generalized canonical correlation analysis (gCCA). The unknown metric Q is estimated online. We use a correlation model between the matrices B_n defined by assumptions on the weighted averaging process (U_n) such that $U_{n+1} = (1 - a_n)U_n + a_n B_n, U_1 = 0$. In the vector case, this model was used for Robbins-Monro type processes for example by Ljung, Pflug and Walk (1992) and Monnez (1991, 2006). This model has for particular cases those studied by Monnez and Skiredj (2021). In Section 2, we define processes $(X_n^i), 1 \leq i \leq r \leq p$, which converge to eigenvectors corresponding to the r largest eigenvalues of B in decreasing order under assumptions on (U_n) ; a Gram-Schmidt orthonormalization is made at step n with respect to a metric Q_{n+1} such that Q_n converges a.s. to Q as n goes to infinity; in the same theorem, we prove the convergence of specific processes to the eigenvalues (Theorem 1). In Section 3, we give sufficient conditions on (B_n) to verify the assumptions on (U_n) and deduce corollaries of Theorem 1. In Section 4, we give another corollary in the application to streaming PCA of a random vector Z , using a mini-batch of current observations of Z at each step of the processes or all the observations up to the current step. In Section 5, we study the application to streaming generalized canonical correlation analysis (gCCA) interpreted as a PCA with a specific metric that is estimated online by a stochastic approximation process parallel to the processes (X_n^i) . All the proofs are in Section 6.

2. Extended almost sure convergence theorem

For a random metric Q_n in \mathbb{R}^p , let $\langle \cdot, \cdot \rangle_n$ be the inner product and $\|\cdot\|_n$ the norm induced by Q_n . For $i \in \{1, \dots, r\}, r \leq p$, we recursively define the processes $(\tilde{X}_n^i), (\tilde{Y}_n^i), (X_n^i)$ in \mathbb{R}^p such that, for $n \geq 1$:

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

The set of vectors $(X_{n+1}^1, \dots, X_{n+1}^r)$ is obtained by Gram-Schmidt orthonormalization with respect to Q_{n+1} of $(\tilde{Y}_{n+1}^1, \dots, \tilde{Y}_{n+1}^r)$. The assumptions on B , respectively $B_n, U_n, a_n, Q_n, X_1^i$, are denoted H1, respectively H2a, H2b, H3, H4, H5. Assume:

(H1a) B is Q -symmetric. (H1b) The r largest eigenvalues of B are simple.

(H2a) For all n , $I + a_n B_n$ is invertible.

(H3a) $a_n > 0$, $\sum_1^\infty a_n = \infty$, $\sum_1^\infty a_n^2 < \infty$. (H4) $Q_n \rightarrow Q$, $\sum_1^\infty a_n \|Q_n - Q\| < \infty$ a.s.

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

We recursively define the weighted averaging process (U_n) in the set of $p \times p$ matrices and for $i \in \{1, \dots, r\}$, $r \leq p$, the process (Λ_n^i) in \mathbb{R} , such that, for $n \geq 1$:

$$U_{n+1} = (1 - a_n)U_n + a_n B_n, U_1 = 0. \text{ For } a_n = \frac{1}{n}, U_{n+1} = \frac{1}{n} \sum_1^n B_i.$$

$$\Lambda_{n+1}^i = \langle U_{n+1} X_{n+1}^i, X_{n+1}^i \rangle_{n+1}.$$

We assume H2b on U_n :

$$(H2b) U_n \rightarrow B, \sum_1^\infty a_n \|U_n - B\| < \infty, \sum_1^\infty a_n \|(U_{n+1} - B)B_n\| < \infty \text{ a.s.}$$

Theorem 1. Under H1a,b on B , H2a on B_n , H2b on U_n , H3a on a_n , H4 on Q_n and H5 on X_1^i , for $i \in \{1, \dots, r\}$, X_n^i converges to V_i or $-V_i$, Λ_n^i converges to λ_i , $\sum_1^\infty a_n |\langle B X_n^i, X_n^i \rangle_n - \lambda_i| < \infty$ and $\sum_1^\infty a_n |\Lambda_n^i - \lambda_i| < \infty$ a.s.

In the proof, we use Theorem 1ii of Monnez and Skiredj (2021) as a lemma:

Lemma 1. Consider the process (X_n) in \mathbb{R}^p such that $X_{n+1} = \frac{(I+a_n B_n)X_n}{\|(I+a_n B_n)X_n\|}$. Under H1a,b for $r = 1$ on B , H2a on B_n , H3a on a_n , H5 on X_1^i and $\sum_1^\infty a_n \|B_n - B\| < \infty$ a.s., X_n converges to V_1 or $-V_1$, $\sum_1^\infty a_n (\lambda_1 - \langle B X_n, X_n \rangle) < \infty$, $\sum_1^\infty a_n |\lambda_1 - \langle B_n X_n, X_n \rangle| < \infty$ a.s.

3. Particular cases of the correlation model

3.1 Lemmas

We establish sufficient conditions on B_n to ensure the assumption H2b on U_n . T_n being the σ -field generated by $X_1, Q_1, B_1^1, B_1^2, Q_2, \dots, B_{n-1}^1, B_{n-1}^2, Q_n$, assume $B_n = B_n^1 + B_n^2$, $E[B_n^2 | T_n] = 0$, $U_n = U_n^1 + U_n^2$, $U_{n+1}^i = (1 - a_n)U_n^i + a_n B_n^i$, $i \in \{1, 2\}$, $U_1^1 + U_1^2 = 0$, and:

$$(H2c) \sum_1^\infty a_n \|B_n^1 - B\| < \infty \text{ a.s.} \quad (H2d) E[B_n^2 | T_n] = 0 \text{ a.s.}, \sup_n E[\|B_n^2\|^2] < \infty.$$

$$(H2e) \sum_{n=1}^\infty a_n \|(U_{n+1}^1 - B)B_n^2\| < \infty \text{ a.s.} \quad (H2f) \sup_n \|B_n^2\| < \infty \text{ a.s.}$$

$$(H3b) a_n > 0, \sum_1^\infty a_n = \infty, \sum_1^\infty a_n^{3/2} < \infty, \frac{a_n}{a_{n+1}} \leq 1 + \gamma a_n + \gamma_n, 0 < \gamma < 2,$$

$$\gamma_n \geq 0, \sum_1^\infty \gamma_n < \infty.$$

Assumption H3b implies H3a. Note that, when $a_n = \frac{a}{n^\alpha}$, H3b is verified from a certain rank n_0 for $\frac{2}{3} < \alpha < 1$ and if $a > \frac{1}{2}$ for $\alpha = 1$.

Lemma 2. *Under H2c on B_n^1 , $a_n > 0$, $a_n = o(1)$ and $\sum_1^\infty a_n = \infty$, we have: $U_n^1 \rightarrow B$, $\sum_1^\infty a_n \|U_n^1 - B\| < \infty$, $\sum_1^\infty a_n \|(U_{n+1}^1 - B)B_n^1\| < \infty$ a.s. For $U_n^2 = 0$, H2b holds.*

Lemma 3. *Under H2d on B_n^2 , H3a, $\frac{a_n}{a_{n+1}} \leq 1 + \gamma a_n + \gamma_n$, $0 < \gamma < 2$, $\gamma_n \geq 0$ and $\sum_1^\infty \gamma_n < \infty$, we have: $U_n^2 \rightarrow 0$ a.s., $\sum_1^\infty a_n E[\|U_n^2\|^2] < \infty$, $\exists b > 0$: $E[\|U_n^2\|^2] \leq ba_n$; for $\sum_1^\infty a_n^{\frac{3}{2}} < \infty$, $\sum_1^\infty a_n \|U_n^2\| < \infty$ a.s.*

Lemma 4. *Under H2c on B_n^1 , H2d on B_n^2 and H3b on a_n , we have: $U_n \rightarrow B$, $\sum_1^\infty a_n \|U_n - B\| < \infty$ a.s.; if H2e or H2f on B_n^2 also holds, $\sum_1^\infty a_n \|(U_{n+1} - B)B_n\| < \infty$ a.s.*

3.2 Corollaries

Let A be the set of the assumptions H1a,b on B, H2a on B_n , H3b on a_n , H4 on Q_n , H5 on X_1^i . By Lemma 4, H2b holds under H2c, H2d, H2e or H2f, H3b; then:

Corollary 1. *Under A, H2c on B_n^1 , H2d on B_n^2 and H2e or H2f on B_n^2 , Theorem 1 holds.*

1) Let $B_n^1 = B$, $B_n^2 = B_n - B$. Let $U_1^1 = B$, then $U_n^1 = B$ for $n \geq 1$. H2c,e hold.

Corollary 2. *Under A, $E[B_n|T_n] = B$ and $\sup_n E[\|B_n - B\|^2] < \infty$, Theorem 1 holds.*

This corollary is an extension of Theorem 2 of Oja and Karhunen (1985), which was demonstrated by the ODE method assuming $Q = I$, the r largest eigenvalues of B are positive and $\|B_n\| \leq b$ a.s.

2) Let $B_n^1 = E[B_n|T_n]$, $B_n^2 = B_n - E[B_n|T_n]$.

Corollary 3. *Under A, $\sum_1^\infty a_n \|E[B_n|T_n] - B\| < \infty$, $\sup_n E[\|B_n - E[B_n|T_n]\|^2] < \infty$ and $\sup_n \|B_n - E[B_n|T_n]\| < \infty$ a.s., Theorem 1 holds.*

This corollary, applying when $E[B_n|T_n]$ converges a.s. to B , is an extension of Theorem 1i and Corollary 2i of Monnez and Skiredj (2021), which was demonstrated using a method of Duflo (1997, 343) and arguments of exterior algebra, assuming Q known,

$\sum_1^\infty a_n E[\|E[B_n|T_n] - B\|] < \infty$, there exists $b > 0$ such that $\sup_n \|B_n\| < b$ a.s. and another assumption not used here.

3) Let $B_n^1 = B_n$, $B_n^2 = 0$, $U_1^2 = 0$, then $U_n^2 = 0$ for $n \geq 1$.

By Lemma 2, H2b on U_n holds under H2c and H3a. Thus $\sum_1^\infty a_n |\langle BX_n^i, X_n^i \rangle_n - \lambda_i| < \infty$ a.s. by Theorem 1; under H2c and H4, $\sum_1^\infty a_n |\langle B_n X_n^i, X_n^i \rangle_n - \lambda_i| < \infty$ a.s.

Corollary 4. *Under A with H3b replaced by H3a and $\sum_1^\infty a_n \|B_n - B\| < \infty$ a.s., Theorem 1 holds and for $i \in \{1, \dots, r\}$, $\sum_1^\infty a_n |\langle B_n X_n^i, X_n^i \rangle_n - \lambda_i| < \infty$ a.s.*

This corollary, applying when B_n converges a.s. to B , is an extension of Theorem 1ii and Corollary 2ii of Monnez and Skiredj (2021) demonstrated using the theorem of Robbins and Siegmund (1971) and arguments of exterior algebra, assuming Q known.

For example, assume that the random matrices B_n are i.i.d. with $E[B_n] = B$, $E[\|B_n\|^2] < \infty$. Let $V_n = \frac{1}{n} \sum_1^n B_i$. Then $E[V_n] = B$, $E[\|V_n - B\|^2] = O\left(\frac{1}{n}\right)$, $V_n \rightarrow B$ a.s. Assuming $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$, we have $\sum_1^\infty a_n \|V_n - B\| < \infty$ a.s. Defining the processes (F_n^i) by replacing B_n by V_n in the definition of the processes (X_n^i) , we can apply Corollary 4 to these new processes.

Likewise, by Lemma 4, under H2c, H2d and H3b, $U_n \rightarrow B$, $\sum_1^\infty a_n \|U_n - B\| < \infty$ a.s. Defining the processes (F_n^i) by replacing B_n by U_n in the definition of the processes (X_n^i) , we can apply Corollary 4 to these new processes.

4. Application to streaming PCA

4.1 Formulation of PCA

Let Z^1, \dots, Z^p be the components of a random vector Z in \mathbb{R}^p defined on a probability space (Ω, \mathcal{A}, P) and $C = E[(Z - E[Z])(Z - E[Z])^T]$ its covariance matrix. Let M be a metric in \mathbb{R}^p defining a distance between two observations of Z .

Formulation of PCA. *For $l \in \{1, \dots, r\}$, $r \leq p$, determine at step l a linear combination of the centered components of Z , $U_l = c_l^T (Z - E[Z])$, $c_l \in \mathbb{R}^{p*}$, called l th principal component, which is uncorrelated with U_1, \dots, U_{l-1} and of maximum variance under the constraint $c_l^T M^{-1} c_l = 1$.*

Thus we have to maximize the quadratic form $c^T C c$ under the constraints $c^T C c_j = 0, j \in \{1, \dots, l-1\}$, and $c^T M^{-1} c = 1$. c_l is an eigenvector of the M^{-1} -symmetric matrix $B = MC$ ($Q = M^{-1}$) corresponding to its l th largest eigenvalue λ_l .

$u_l = M^{-1} c_l$ is an eigenvector of the M -symmetric matrix CM corresponding to λ_l and maximizes the quadratic form $u^T M C M u = u^T M E[(Z - E[Z])(Z - E[Z])^T] M u = E[[(Z - E[Z])^T M u]^2]$ under the constraints $u^T M u_j = 0, j \in \{1, \dots, l-1\}$, and $u^T M u = 1$. The axis $(E[Z], u_l)$ with origin $E[Z]$ and M -unit vector u_l is called the l th principal axis. For all $\omega \in \Omega$, $(Z(\omega) - E[Z])^T M u_l$ is the abscissa of the M -orthogonal projection of $Z(\omega) - E[Z]$ on the axis $(E[Z], u_l)$. The subspace generated by u_1, \dots, u_r maximizes $E[\|\Pi(Z - E[Z])\|^2]$ with Π the M -orthogonal projection operator on a r -dimensional subspace containing $E[Z]$.

A particular case of PCA is normed PCA, in which M is the diagonal matrix of the inverses of the variances of the components of Z . Normed PCA is equivalent to using standardized data, i.e. observations of $M^{1/2}(Z - E[Z])$, and the identity metric, hence the same importance is given to each of the components of Z in this PCA.

4.2 Stochastic approximation of the principal components of PCA

Let $(Z_{11}, \dots, Z_{1m_1}, \dots, Z_{n1}, \dots, Z_{nm_n}, \dots)$ be a sequence of i.i.d. observations of Z , Z_{n1}, \dots, Z_{nm_n} being taken into account at step n of the processes. Let \bar{Z}_n be the mean of the sample $(Z_{11}, \dots, Z_{nm_n})$ and C_n its covariance matrix, $C_n = \frac{1}{\mu_n} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij} Z_{ij}^T - \bar{Z}_n \bar{Z}_n^T$ with $\mu_n = \sum_{i=1}^n m_i$; let a random matrix M_n be an estimator of M recursively computed from this sample.

We define the processes $(\tilde{Y}_n^i), (\tilde{X}_n^i), (X_n^i), (\Lambda_n^i), i \in \{1, \dots, r\}$ as in Section 2:

$$\tilde{Y}_{n+1}^i = (I + a_n B_n) \tilde{X}_n^i, \quad \tilde{X}_{n+1}^i = \tilde{Y}_{n+1}^i - \sum_{j < i} \langle \tilde{Y}_{n+1}^i, X_{n+1}^j \rangle_{n+1} X_{n+1}^j,$$

$$X_{n+1}^i = \frac{\tilde{X}_{n+1}^i}{\|\tilde{X}_{n+1}^i\|_{n+1}}, \quad U_{n+1} = (1 - a_n) U_n + a_n B_n, \quad U_1 = 0,$$

$$\Lambda_{n+1}^i = \langle U_{n+1} X_{n+1}^i, X_{n+1}^i \rangle_{n+1}, \quad \text{with}$$

$$B_n = M_n (\omega_{1n} (S_{1n} - \bar{Z}_n \bar{Z}_n^T) + \omega_{2n} (S_{2n} - \bar{Z}_n \bar{Z}_n^T))$$

$$= M_n (\omega_{1n} S_{1n} + \omega_{2n} S_{2n} - \bar{Z}_n \bar{Z}_n^T), \quad S_{1n} = \frac{1}{\mu_n} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij} Z_{ij}^T, \quad S_{2n} = \frac{1}{m_n} \sum_{j=1}^{m_n} Z_{nj} Z_{nj}^T,$$

$$\omega_{1n} \geq 0, \omega_{2n} \geq 0, \omega_{1n} + \omega_{2n} = 1.$$

At step n , the orthonormalization is made with respect to the metric Q_{n+1} . The matrix B_n is a convex combination of two matrices. Different time-varying weights can be assigned

to the set of past observations and the set of current observations. For $\omega_{2n} = 0$, past and current observations of Z are used at step n ; for $\omega_{1n} = 0$, only the m_n current observations are used at step n . We consider the decomposition:

$$\begin{aligned} B_n - B &= B_n^1 - B + B_n^2 \text{ with } B_n^2 = \omega_{2n}M(S_{2n} - E[ZZ^T]), \\ B_n^1 - B &= (M_n - M)(\omega_{1n}(S_{1n} - \bar{Z}_n\bar{Z}_n^T) + \omega_{2n}(S_{2n} - \bar{Z}_n\bar{Z}_n^T)) \\ &+ \omega_{1n}M(S_{1n} - E[ZZ^T]) - M(\bar{Z}_n - E[Z])\bar{Z}_n^T - ME[Z](\bar{Z}_n - E[Z])^T. \end{aligned}$$

Assume:

$$(H3c) \ a_n > 0, \sum_1^\infty a_n = \infty, \sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty, \sum_1^\infty a_n^2 < \infty.$$

$$(H6a) \ M_n \rightarrow M \text{ a.s.} \quad (H6b) \ \sum_1^\infty a_n \|M_n - M\| < \infty \text{ a.s.}$$

$$(H7a) \ Z \text{ has } 4^{\text{th}} \text{ moments.} \quad (H7b) \ Z \text{ is a.s. bounded.}$$

Under H7a:

$$\bar{Z}_n \rightarrow E[Z], \ S_{1n} \rightarrow E[ZZ^T] \text{ a.s.}, \ E[\|\bar{Z}_n - E[Z]\|] = O(1/\sqrt{n}), \ E[\|S_{1n} - E[ZZ^T]\|] = O(1/\sqrt{n});$$

since $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$, $\sum_1^\infty a_n \|\bar{Z}_n - E[Z]\| < \infty$, $\sum_1^\infty a_n \|S_{1n} - E[ZZ^T]\| < \infty$ a.s.

Then, by H3c, H6b and H7b, $\sum_1^\infty a_n \|B_n^1 - B\| < \infty$ a.s. Moreover, $E[B_n^2 | T_n] = 0$ and, by H7b, $\|B_n^2\|$ is a.s. bounded, thus assumptions H2d,f hold. When $\omega_{2n} = 0$ for all n , H7b can be replaced by H7a, $B_n \rightarrow B$ under H6a and H7a, then $\langle B_n X_n^i, X_n^i \rangle_n \rightarrow \lambda_i$ when $X_n^i \rightarrow \pm c_i$ a.s. Applying Corollaries 1 and 4 yields:

Corollary 5. *Under H1b on B , H2a on B_n , H3b on a_n , H4 on Q_n , H5 on X_1^i , H6a,b on M_n and H7b on Z , for $i \in \{1, \dots, r\}$, X_n^i converges to c_i or $-c_i$, Λ_n^i converges to λ_i , $\sum_1^\infty a_n |\langle BX_n^i, X_n^i \rangle_n - \lambda_i| < \infty$, $\sum_1^\infty a_n |\Lambda_n^i - \lambda_i| < \infty$ a.s. When $\omega_{2n} = 0$ for all n , the same conclusions hold with H3b replaced by H3c and H7b by H7a; moreover, for $i \in \{1, \dots, r\}$, $\langle B_n X_n^i, X_n^i \rangle_n$ converges to λ_i , $\sum_1^\infty a_n |\langle B_n X_n^i, X_n^i \rangle_n - \lambda_i| < \infty$ a.s.*

Note that, when $\omega_{2n} = 0$ for all n , the assumption “ Z is a.s. bounded” is not required.

5. Application to streaming generalized canonical correlation analysis

5.1 Formulation

Consider the generalized canonical correlation analysis (gCCA) defined by Carroll (1968). We use here a probabilistic formulation. Suppose the set of components of a random vector Z in \mathbb{R}^p is partitioned in q subsets of real random variables $\{Z^{k1}, \dots, Z^{kr_k}\}$, $k \in \{1, \dots, q\}$, $\sum_{k=1}^q r_k = p$. Let Z^k be the random vector in \mathbb{R}^{r_k} whose components are

Z^{k_1}, \dots, Z^{k_r} , $C^k = \text{Covar}[Z^k]$ be the covariance matrix of Z^k , $C = \text{Covar}[Z]$ that of Z , $C^{k \cdot} = \text{Covar}[Z^k, Z]$ that of Z^k and Z . Assume:

(H7c) *There is no affine relation between the components of Z .*

Then the matrices C^k , $k \in \{1, \dots, q\}$, and C are invertible.

Formulation of gCCA. For $l \in \{1, \dots, r\}$, $r \leq \min(r_1, \dots, r_q)$, determine at step l a linear combination $U_l = \theta_l^\top (Z - E[Z])$ of variance 1 of the centered components of Z , uncorrelated with U_1, \dots, U_{l-1} , called l th general component, and for $k \in \{1, \dots, q\}$, a linear combination $V_l^k = (\eta_l^k)^\top (Z^k - E[Z^k])$ of variance 1 of the centered components of Z^k , called l th canonical component of the k th subset of variables, which maximize $\sum_{k=1}^q \rho^2(U_l, V_l^k)$, ρ denoting the linear correlation coefficient.

5.2 Interpretation of gCCA as a PCA

Let Π^k be the projection operator on the subspace of $L^2(\Omega, \mathcal{A}, P)$ with inner product $\langle X, Y \rangle = E[XY]$ generated by the centered components of Z^k .

Let $\theta_l = ((\theta_l^1)^\top \dots (\theta_l^q)^\top)^\top$, $\theta_l^k \in \mathbb{R}^{r_k}$, $k \in \{1, \dots, q\}$. Let M be the block diagonal metric in \mathbb{R}^p whose k th diagonal diagonal block is $(C^k)^{-1}$.

Suppose U_l is determined; then:

$$\begin{aligned} \sum_{k=1}^q \rho^2(U_l, V_l^k) \max &\Leftrightarrow \rho^2(U_l, V_l^k) = \cos^2(U_l, V_l^k) \max, k \in \{1, \dots, q\} \\ &\Leftrightarrow V_l^k = \frac{\Pi^k U_l}{\|\Pi^k U_l\|}, k \in \{1, \dots, q\}. \end{aligned}$$

It is shown that:

Proposition 1. For $l \in \{1, \dots, r\}$, θ_l is a C -normed eigenvector of the matrix $B = MC$ corresponding to its l th largest eigenvalue λ_l and $V_l^k = \frac{(\theta_l^k)^\top (Z^k - E[Z^k])}{((\theta_l^k)^\top C^k \theta_l^k)^{\frac{1}{2}}}$, $k \in \{1, \dots, q\}$.

Thus for $l \in \{1, \dots, r\}$, the l th general component of the gCCA of Z , $U_l = \theta_l^\top (Z - E[Z])$ of variance 1, is collinear with the l th principal component of the PCA of Z with the metric M , $c_l^\top (Z - E[Z])$ of variance λ_l , and $\theta_l = \frac{c_l}{\sqrt{\lambda_l}}$. A particular case of gCCA is normed PCA when $r_k = 1$ for all k , the diagonal metric M being that of the inverses of the variances of the components of Z .

5.3 Stochastic approximation of the general and canonical components of gCCA

Let $(Z_{11}, \dots, Z_{1m_1}, \dots, Z_{n1}, \dots, Z_{nm_n}, \dots)$ be a sequence of i.i.d. observations of Z with $Z_{ij} = ((Z_{ij}^1)^\top \dots (Z_{ij}^q)^\top)^\top$, $Z_{ij}^k(r_k, 1)$, $k \in \{1, \dots, q\}$, the mini batch $(Z_{n1}, \dots, Z_{nm_n})$ being taken into account at step n of the processes. For $k \in \{1, \dots, q\}$, let \bar{Z}_n^k be the mean of the sample $(Z_{11}^k, \dots, Z_{nm_n}^k)$ of Z^k , $C_n^k = \frac{1}{\mu_n} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^k (Z_{ij}^k)^\top - \bar{Z}_n^k (\bar{Z}_n^k)^\top$ with $\mu_n = \sum_{i=1}^n m_i$ be its covariance matrix, both recursively computed.

Let M_n and Q_{n+1} be the block diagonal matrices of order p whose k th diagonal blocks are respectively M_n^k defined below and C_n^k . The matrix Q_n is positive definite symmetric from a certain rank. Under the assumption H7a, $C_n^k \rightarrow C^k$ a.s., $E[\|C_n^k - C^k\|] = O\left(\frac{1}{\sqrt{\mu_n}}\right) = O\left(\frac{1}{\sqrt{n}}\right)$, since $\mu_n \geq n$. Since $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$ under H3c, $\sum_1^\infty a_n \|C_n^k - C^k\| < \infty$ a.s. Thus $Q_n \rightarrow M^{-1}$, $\sum_1^\infty a_n \|Q_{n+1} - M^{-1}\| < \infty$ and $\sum_1^\infty a_n \|Q_n - M^{-1}\| < \infty$ a.s.; assumption H4 on Q_n holds.

For $k \in \{1, \dots, q\}$, to obtain an estimation of $M^k = (C^k)^{-1}$, we define the stochastic approximation process $(M_n^k, n \geq 1)$ such that:

$$M_n^k = M_{n-1}^k - a_n (C_n^k M_{n-1}^k - I); \text{ thus } M_n = M_{n-1} - a_n (Q_{n+1} M_{n-1} - I).$$

Theorem 2. *Under the assumptions $a_n > 0$, $a_n = o(1)$, $\sum_1^\infty a_n = \infty$ and H7a,c, M_n converges to M and $\sum_1^\infty a_n \|M_n - M\| < \infty$ a.s.*

Thus assumptions H6a,b on M_n hold. We define the processes $(\tilde{Y}_n^i), (\tilde{X}_n^i), (X_n^i), (\Lambda_n^i)$, $i \in \{1, \dots, r\}$, used for PCA (Section 4). Applying Corollary 5 with the additional assumption H7c yields:

Corollary 6. *Under H1b on B , H2a on B_n , H3b on a_n , H5 on X_1^i and H7b,c on Z , for $i \in \{1, \dots, r\}$, X_n^i converges to $c_i = \sqrt{\lambda_i} \theta_i$ or $-c_i$, Λ_n^i converges to λ_i , $\sum_1^\infty a_n |\langle B X_n^i, X_n^i \rangle_n - \lambda_i| < \infty$, $\sum_1^\infty a_n |\Lambda_n^i - \lambda_i| < \infty$ a.s. When $\omega_{2n} = 0$ for all n , the same conclusions hold with H3b replaced by H3c and H7b by H7a; moreover, for $i \in \{1, \dots, r\}$, $\langle B_n X_n^i, X_n^i \rangle_n$ converges to λ_i and $\sum_{n=1}^\infty a_n |\langle B_n X_n^i, X_n^i \rangle_n - \lambda_i| < \infty$ a.s.*

At step n , we obtain an estimation of c_i and λ_i , thus of θ_i . We can estimate $E[Z^k]$ by \bar{Z}_n^k and C^k by C_n^k , thus estimate the general and canonical components of gCCA.

6. Proofs of theorems

6.1 Some concepts of exterior algebra

Let (e_1, \dots, 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 $e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_r}$, $i_1 < i_2 < \dots < i_r \in \{1, \dots, p\}$.

Let Q be a metric in \mathbb{R}^p . The inner product $\langle \cdot, \cdot \rangle$ in ${}^r\Lambda\mathbb{R}^p$ induced by Q is such that

$$\langle e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_r}, e_{k_1} \wedge e_{k_2} \wedge \dots \wedge e_{k_r} \rangle = \sum_{\sigma \in G_r} (-1)^{s(\sigma)} \langle e_{i_1}, e_{\sigma(k_1)} \rangle \dots \langle e_{i_r}, e_{\sigma(k_r)} \rangle,$$

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

Let $\|\cdot\|$ denote the associated norm. Note that if x_1, \dots, x_r are Q -orthogonal, then $\|x_1 \wedge \dots \wedge x_r\| = \prod_{i=1}^r \|x_i\|$ and if (e_1, \dots, e_p) is a Q -orthonormal basis of \mathbb{R}^p , then the set of the C_p^r exterior products $e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_r}$ is an orthonormal basis of ${}^r\Lambda\mathbb{R}^p$.

Let U be an endomorphism in \mathbb{R}^p . For $j \in \{1, \dots, r\}$, the endomorphism ${}^{rj}U$ in ${}^r\Lambda\mathbb{R}^p$ is such that:

$${}^{rj}U(x_1 \wedge \dots \wedge x_r) = \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq r} x_1 \wedge \dots \wedge Ux_{i_1} \wedge \dots \wedge Ux_{i_j} \wedge \dots \wedge x_r;$$

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

The following properties hold.

(P1) Assume U has p real eigenvalues $\lambda_1, \dots, \lambda_p$ and, for $j \in \{1, \dots, p\}$, let V_j be an eigenvector corresponding to λ_j . Then the C_p^r vectors $V_{i_1} \wedge V_{i_2} \wedge \dots \wedge V_{i_r}$, $i_1 < i_2 < \dots < i_r \in \{1, \dots, p\}$, are eigenvectors of ${}^{rj}U$; for $j = 1$, the corresponding eigenvalues are $\lambda_{i_1} + \dots + \lambda_{i_r}$ and for $j = r$, the products $\lambda_{i_1} \cdot \dots \cdot \lambda_{i_r}$; thus if U is invertible, so is ${}^{rr}U$.

(P2) If U is Q -symmetric, ${}^{rj}U$ is symmetric with respect to the metric induced by Q in ${}^r\Lambda\mathbb{R}^p$.

$$(P3) {}^{rr}(I + U) = I + \sum_{j=1}^r {}^{rj}U.$$

(P4) There exists $c(r) > 0$ such that, for $1 \leq j \leq r$, $\|{}^{rj}U\| \leq c(r) \|U\|^j$.

6.2 Proof of Theorem 1

The writing of fixed ω belonging to the intersection of the a.s. convergence sets is omitted. Let $W_n = U_n - B$; $W_{n+1} = (1 - a_n)W_n + a_n(B_n - B)$. Since W_n converges to 0 by H2b, $I - W_n$ is invertible from a certain rank, assumed equal to 1 without loss of generality. We use Lemmas 1 and 5. Let ${}^i\tilde{X}_n = \tilde{X}_n^1 \wedge \dots \wedge \tilde{X}_n^i$.

Lemma 5. $\frac{\|\tilde{X}_n^j\|}{\|\tilde{X}_n^i\|} \rightarrow 1$, $\frac{\|\tilde{X}_n^i\|}{\|\tilde{X}_n^1\| \cdots \|\tilde{X}_n^i\|} \rightarrow 1$, $\sum_{n=1}^{\infty} a_n \left| \frac{\langle \tilde{X}_n, {}^{i1}B \tilde{X}_n \rangle}{\|\tilde{X}_n\|^2} - \frac{\langle \tilde{X}_n, {}^{i1}B \tilde{X}_n \rangle_n}{\|\tilde{X}_n\|_n^2} \right| < \infty$ under H4.

Outline of the proof. In Step 1, we prove Lemma 5. In Step 2, defining, for $i \in \{1, \dots, r\}$,

$$\tilde{Z}_n^i = (I - W_n)\tilde{X}_n^i, \quad {}^i\tilde{Z}_n = \tilde{Z}_n^1 \Lambda \cdots \Lambda \tilde{Z}_n^i,$$

we show that there exists a $p \times p$ random matrix C_n such that

$${}^i\tilde{Z}_{n+1} = {}^{ii}(I + a_n C_n) {}^i\tilde{Z}_n,$$

$C_n \rightarrow B$ and $\sum_1^{\infty} a_n \|C_n - B\| < \infty$ under H2b.

In Steps 3, 4, using a proof similar to that of Corollary 2ii of Monnez and Skiredj (2021), we deduce by Lemma 1, with $\lambda_{1i} = \lambda_1 + \cdots + \lambda_i$,

$$\frac{{}^i\tilde{Z}_n}{\|{}^i\tilde{Z}_n\|} \rightarrow \pm V_{1i} = \pm V_1 \Lambda \cdots \Lambda V_i, \quad \sum_1^{\infty} a_n \left(\lambda_{1i} - \left\langle \frac{{}^i\tilde{Z}_n}{\|{}^i\tilde{Z}_n\|}, {}^{i1}B \frac{{}^i\tilde{Z}_n}{\|{}^i\tilde{Z}_n\|} \right\rangle \right) < \infty,$$

then $\frac{{}^i\tilde{X}_n}{\|{}^i\tilde{X}_n\|} \rightarrow \pm V_{1i}$, $X_n^i \rightarrow \pm V_i$, $\Lambda_n^i \rightarrow \lambda_i$.

In Step 5, we infer that $\sum_1^{\infty} a_n \left(\lambda_{1i} - \left\langle \frac{{}^i\tilde{X}_n}{\|{}^i\tilde{X}_n\|}, {}^{i1}B \frac{{}^i\tilde{X}_n}{\|{}^i\tilde{X}_n\|} \right\rangle \right) < \infty$, then by Lemma 5,

$$\sum_1^{\infty} a_n |\lambda_{1i} - \langle {}^iX_n, {}^{i1}B {}^iX_n \rangle_n| < \infty,$$

therefore $\sum_1^{\infty} a_n |\lambda_i - \langle X_n^i, B X_n^i \rangle_n| < \infty$ and $\sum_1^{\infty} a_n |\Lambda_n^i - \lambda_i| < \infty$.

Step 1. Proof of Lemma 5

1) Since $Q_n \rightarrow Q$, $\frac{\langle \tilde{X}_n^j, \tilde{X}_n^k \rangle - \langle \tilde{X}_n^j, \tilde{X}_n^k \rangle_n}{\|\tilde{X}_n^j\| \|\tilde{X}_n^k\|} = \frac{(\tilde{X}_n^j)^T (Q - Q_n) \tilde{X}_n^k}{\|\tilde{X}_n^j\| \|\tilde{X}_n^k\|} \rightarrow 0$; thus, for $j = k$, $\frac{\|\tilde{X}_n^j\|}{\|\tilde{X}_n^j\|} \rightarrow 1$, and,

since the vectors \tilde{X}_n^j are Q_n -orthogonal, $\frac{\langle \tilde{X}_n^j, \tilde{X}_n^k \rangle}{\|\tilde{X}_n^j\| \|\tilde{X}_n^k\|} \rightarrow \delta^{jk}$ (Kronecker symbol). Therefore

$$\frac{\|\tilde{X}_n^i\|^2}{\|\tilde{X}_n^1\|^2 \cdots \|\tilde{X}_n^i\|^2} = \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \frac{\langle \tilde{X}_n^1, \tilde{X}_n^{\sigma(1)} \rangle}{\|\tilde{X}_n^1\| \|\tilde{X}_n^{\sigma(1)}\|} \cdots \frac{\langle \tilde{X}_n^i, \tilde{X}_n^{\sigma(i)} \rangle}{\|\tilde{X}_n^i\| \|\tilde{X}_n^{\sigma(i)}\|} \rightarrow 1,$$

G_i being the set of permutations σ of $\{1, \dots, i\}$ and $s(\sigma)$ the number of inversions of σ .

$$\begin{aligned} 2) \langle \tilde{X}_n, {}^{i1}B \tilde{X}_n \rangle &= \sum_{k=1}^i \langle \tilde{X}_n^1 \Lambda \cdots \Lambda \tilde{X}_n^i, \tilde{X}_n^1 \Lambda \cdots \Lambda B \tilde{X}_n^k \Lambda \cdots \Lambda \tilde{X}_n^i \rangle \\ &= \sum_{k=1}^i \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \langle \tilde{X}_n^1, \tilde{X}_n^{\sigma(1)} \rangle \cdots \langle \tilde{X}_n^k, B \tilde{X}_n^{\sigma(k)} \rangle \cdots \langle \tilde{X}_n^i, \tilde{X}_n^{\sigma(i)} \rangle. \end{aligned}$$

Replacing each inner product $\langle \tilde{X}_n^j, \tilde{X}_n^{\sigma(j)} \rangle$ by $\langle \tilde{X}_n^j, \tilde{X}_n^{\sigma(j)} \rangle_n + (\tilde{X}_n^j)^T (Q - Q_n) \tilde{X}_n^{\sigma(j)}$, and using Landau notation $O(\cdot)$,

$$\frac{\langle \tilde{X}_n^j, \tilde{X}_n^{\sigma(j)} \rangle}{\|\tilde{X}_n^j\|_n \|\tilde{X}_n^{\sigma(j)}\|_n} = \frac{\langle \tilde{X}_n^j, \tilde{X}_n^{\sigma(j)} \rangle_n}{\|\tilde{X}_n^j\|_n \|\tilde{X}_n^{\sigma(j)}\|_n} + O(\|Q_n - Q\|) = \delta^{j\sigma(j)} + O(\|Q_n - Q\|)$$

$$\text{and } \frac{\langle \tilde{X}_n^k, B\tilde{X}_n^{\sigma(k)} \rangle}{\|\tilde{X}_n^k\|_n \|\tilde{X}_n^{\sigma(k)}\|_n} = \frac{\langle \tilde{X}_n^k, B\tilde{X}_n^{\sigma(k)} \rangle_n}{\|\tilde{X}_n^k\|_n \|\tilde{X}_n^{\sigma(k)}\|_n} + O(\|Q_n - Q\|); \text{ then:}$$

$$\frac{\langle \tilde{X}_n^1, \tilde{X}_n^{\sigma(1)} \rangle \dots \langle \tilde{X}_n^k, B\tilde{X}_n^{\sigma(k)} \rangle \dots \langle \tilde{X}_n^i, \tilde{X}_n^{\sigma(i)} \rangle}{\|\tilde{X}_n^1\|_n^2 \dots \|\tilde{X}_n^i\|_n^2} = \frac{\langle \tilde{X}_n^1, \tilde{X}_n^{\sigma(1)} \rangle_n \dots \langle \tilde{X}_n^k, B\tilde{X}_n^{\sigma(k)} \rangle_n \dots \langle \tilde{X}_n^i, \tilde{X}_n^{\sigma(i)} \rangle_n}{\|\tilde{X}_n^1\|_n^2 \dots \|\tilde{X}_n^i\|_n^2} + O(\|Q_n - Q\|).$$

Since the vectors \tilde{X}_n^j are Q_n -orthogonal, $\|{}^i\tilde{X}_n\|_n^2 = \|\tilde{X}_n^1\|_n^2 \dots \|\tilde{X}_n^i\|_n^2$. Thus:

$$\frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|{}^i\tilde{X}_n\|_n^2} = \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|\tilde{X}_n^1\|_n^2 \dots \|\tilde{X}_n^i\|_n^2} = \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle_n}{\|{}^i\tilde{X}_n\|_n^2} + O(\|Q_n - Q\|).$$

Likewise, $\frac{\|{}^i\tilde{X}_n\|_n^2}{\|{}^i\tilde{X}_n\|_n^2} = 1 + O(\|Q_n - Q\|)$. Then: $\sum_1^\infty a_n \left| \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|{}^i\tilde{X}_n\|_n^2} - \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle_n}{\|{}^i\tilde{X}_n\|_n^2} \right|$

$$= \sum_1^\infty a_n \left| \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|{}^i\tilde{X}_n\|_n^2} \left(1 - \frac{\|{}^i\tilde{X}_n\|_n^2}{\|{}^i\tilde{X}_n\|_n^2} \right) + \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|{}^i\tilde{X}_n\|_n^2} - \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle_n}{\|{}^i\tilde{X}_n\|_n^2} \right|$$

$$= \sum_1^\infty a_n O(\|Q_n - Q\|) < \infty \text{ by H4.}$$

Step 2. Let, for $i \in \{1, \dots, r\}$, $\tilde{Z}_n^i = (I - W_n)\tilde{X}_n^i$.

$$\tilde{Z}_{n+1}^i = (I - W_{n+1})\tilde{X}_{n+1}^i = (I - W_{n+1})\tilde{Y}_{n+1}^i - \sum_{j < i} \frac{\langle \tilde{Y}_{n+1}^i, \tilde{X}_{n+1}^j \rangle_{n+1}}{\|\tilde{X}_{n+1}^j\|_{n+1}^2} \tilde{Z}_{n+1}^j.$$

$${}^i\tilde{Z}_{n+1} = \tilde{Z}_{n+1}^1 \Lambda \dots \Lambda \tilde{Z}_{n+1}^i = (I - W_{n+1})\tilde{Y}_{n+1}^1 \Lambda \dots \Lambda (I - W_{n+1})\tilde{Y}_{n+1}^i.$$

$$(I - W_{n+1})\tilde{Y}_{n+1}^j = (I - W_{n+1})(I + a_n B_n)\tilde{X}_n^j = (I - W_{n+1})(I + a_n B_n)(I - W_n)^{-1}\tilde{Z}_n^j$$

$$= (I - W_n + a_n W_n + a_n B - a_n W_{n+1} B_n)(I - W_n)^{-1}\tilde{Z}_n^j = (I + a_n C_n)\tilde{Z}_n^j,$$

$$C_n = (B + W_n - W_{n+1} B_n)(I - W_n)^{-1} = B + (B W_n + W_n - W_{n+1} B_n)(I - W_n)^{-1}.$$

By H2a,b, $\sum_1^\infty a_n \|C_n - B\| < \infty$, $I + a_n C_n = (I - W_{n+1})(I + a_n B_n)(I - W_n)^{-1}$ is invertible.

$${}^i\tilde{Z}_{n+1} = (I + a_n C_n)\tilde{Z}_n^1 \Lambda \dots \Lambda (I + a_n C_n)\tilde{Z}_n^i = {}^{ii}(I + a_n C_n) {}^i\tilde{Z}_n = (I + a_n D_n^i) {}^i\tilde{Z}_n,$$

$$D_n^i = {}^{i1}C_n + \sum_{j=2}^i a_n^{j-1} {}^{ij}C_n \text{ by P3.}$$

Step 3. With $S_n = \frac{{}^i\tilde{Z}_n}{\|{}^i\tilde{Z}_n\|}$, we have $S_{n+1} = \frac{(I + a_n D_n^i) S_n}{\|(I + a_n D_n^i) S_n\|}$. Let us show that the process (S_n)

verifies the assumptions of Lemma 1 with B_n replaced by D_n^i and B by ${}^{i1}B$.

By P2, since B is Q -symmetric, ${}^{i1}B$ is symmetric with respect to the metric induced by Q in ${}^i\mathbb{R}^p$. By P1, under H1b, the largest eigenvalue $\lambda_{1i} = \lambda_1 + \dots + \lambda_i$ of ${}^{i1}B$ is simple;

a corresponding normed eigenvector is $V_{1i} = V_1 \Lambda \cdots \Lambda V_i$. Thus assumptions H1a and H1b of Lemma 1 hold for ${}^i 1B$.

By P1, since $I + a_n C_n$ is invertible, so is ${}^{ii}(I + a_n C_n) = I + a_n D_n^i$; H2a holds.

By P4, there exists $c(i) > 0$: for $1 \leq j \leq i$, $\|{}^{ij}(C_n - B)\| \leq c(i)\|C_n - B\|^j$;

by P3, since $\sum_1^\infty a_n \|C_n - B\| < \infty$ and $\|C_n\|^j \leq 2^{j-1}(\|C_n - B\|^j + \|B\|^j)$:

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

H2b holds. Then by Lemma 1:

$$S_n = \frac{{}^i \tilde{Z}_n}{\|{}^i \tilde{Z}_n\|} \rightarrow \pm V_{1i} = \pm V_1 \Lambda \cdots \Lambda V_i, \sum_1^\infty a_n (\lambda_{1i} - \langle S_n, {}^i 1B S_n \rangle) < \infty.$$

By P3, ${}^i \tilde{Z}_n = {}^{ii}(I - W_n) {}^i \tilde{X}_n = (I + \sum_{j=1}^i (-1)^j {}^{ij} W_n) {}^i \tilde{X}_n$;

by P4, $\|{}^{ij} W_n\| \leq c(i)\|W_n\|^j \rightarrow 0$, then $\left\| \frac{{}^i \tilde{Z}_n - {}^i \tilde{X}_n}{\|{}^i \tilde{X}_n\|} \right\| \rightarrow 0$, $\frac{\|{}^i \tilde{Z}_n\|}{\|{}^i \tilde{X}_n\|} \rightarrow 1$, $\frac{{}^i \tilde{X}_n}{\|{}^i \tilde{X}_n\|} \rightarrow \pm V_{1i}$.

Step 4. Assume $\frac{{}^i \tilde{X}_n}{\|{}^i \tilde{X}_n\|} \rightarrow V_{1i}$, $\frac{\tilde{X}_n^1}{\|\tilde{X}_n^1\|} \rightarrow V_1$, $\frac{\tilde{X}_n^j}{\|\tilde{X}_n^j\|} \rightarrow V_j$, $j \in \{1, \dots, i-1\}$, prove $\frac{\tilde{X}_n^i}{\|\tilde{X}_n^i\|} \rightarrow V_i$.

$$\left\langle \frac{{}^i \tilde{X}_n}{\|{}^i \tilde{X}_n\|}, V_{1i} \right\rangle = \frac{\|\tilde{X}_n^1\| \cdots \|\tilde{X}_n^i\|}{\|{}^i \tilde{X}_n\|} \sum_{\sigma \in G_i} (-1)^{s(\sigma)} \left\langle \frac{\tilde{X}_n^1}{\|\tilde{X}_n^1\|}, V_{\sigma(1)} \right\rangle \cdots \left\langle \frac{\tilde{X}_n^i}{\|\tilde{X}_n^i\|}, V_{\sigma(i)} \right\rangle \rightarrow 1.$$

$$\left\langle \frac{\tilde{X}_n^j}{\|\tilde{X}_n^j\|}, V_{\sigma(j)} \right\rangle \rightarrow \delta^{j\sigma(j)}, j \in \{1, \dots, i-1\}; \text{ by Lemma 5, } \frac{\|\tilde{X}_n^1\| \cdots \|\tilde{X}_n^i\|}{\|{}^i \tilde{X}_n\|} \rightarrow 1, \text{ then:}$$

$$\left\langle \frac{\tilde{X}_n^i}{\|\tilde{X}_n^i\|}, V_i \right\rangle \rightarrow 1, \frac{\tilde{X}_n^i}{\|\tilde{X}_n^i\|} \rightarrow V_i, \text{ and since } \frac{\|\tilde{X}_n^j\|}{\|\tilde{X}_n^j\|} \rightarrow 1, X_n^i = \frac{\tilde{X}_n^i}{\|\tilde{X}_n^i\|} \rightarrow V_i \text{ and by H2b and H4,}$$

$$\Lambda_n^i = \langle U_n X_n^i, X_n^i \rangle_n \rightarrow \lambda_i.$$

Step 5. Let $R_n = \frac{{}^i \tilde{X}_n}{\|{}^i \tilde{X}_n\|}$. By Step 3, $\sum_1^\infty a_n (\lambda_{1i} - \langle S_n, {}^i 1B S_n \rangle) < \infty$. Let us show that

$$\sum_1^\infty a_n |\langle S_n, {}^i 1B S_n \rangle - \langle R_n, {}^i 1B R_n \rangle| < \infty. \text{ Let } E_n^i = \sum_{j=1}^i (-1)^j {}^{ij} W_n.$$

$$\text{By Step 3, } {}^i \tilde{Z}_n = (I + E_n^i) {}^i \tilde{X}_n, \text{ then } S_n = \frac{{}^i \tilde{Z}_n}{\|{}^i \tilde{Z}_n\|} = \frac{(I + E_n^i) R_n}{\|(I + E_n^i) R_n\|}.$$

Since ${}^i 1B$ is symmetric with respect to the metric induced by Q in ${}^i \Lambda \mathbb{R}^p$,

$$\begin{aligned} &\|(I + E_n^i) R_n\|^2 (\langle S_n, {}^i 1B S_n \rangle - \langle R_n, {}^i 1B R_n \rangle) \\ &= \langle (I + E_n^i) R_n, {}^i 1B (I + E_n^i) R_n \rangle - \|(I + E_n^i) R_n\|^2 \langle R_n, {}^i 1B R_n \rangle \\ &= \langle R_n, {}^i 1B R_n \rangle \left(1 - \|(I + E_n^i) R_n\|^2 \right) + 2 \langle E_n^i R_n, {}^i 1B R_n \rangle + \langle E_n^i R_n, {}^i 1B E_n^i R_n \rangle \end{aligned}$$

$$= \langle R_n, {}^{i1}BR_n \rangle \left(-2\langle R_n, E_n^i R_n \rangle - \|E_n^i R_n\|^2 \right) + 2\langle E_n^i R_n, {}^{i1}BR_n \rangle + \langle E_n^i R_n, {}^{i1}BE_n^i R_n \rangle.$$

Since $W_n \rightarrow 0$ and $\sum_1^\infty a_n \|W_n\| < \infty$, $E_n^i \rightarrow 0$ and $\sum_1^\infty a_n \|E_n^i\| < \infty$. Since $\|R_n\| = 1$, $\|(I + E_n^i)R_n\| \rightarrow 1$ and $\sum_1^\infty a_n |\langle S_n, {}^{i1}BS_n \rangle - \langle R_n, {}^{i1}BR_n \rangle| < \infty$. Then:

$$\sum_1^\infty a_n (\lambda_{1i} - \langle R_n, {}^{i1}BR_n \rangle) = \sum_1^\infty a_n \left(\lambda_{1i} - \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|{}^i\tilde{X}_n\|^2} \right) < \infty;$$

by Lemma 5: $\sum_1^\infty a_n \left| \lambda_{1i} - \frac{\langle {}^i\tilde{X}_n, {}^{i1}B {}^i\tilde{X}_n \rangle}{\|{}^i\tilde{X}_n\|^2} \right| < \infty$; since vectors \tilde{X}_n^i are Q_n -orthogonal and

$$X_n^i = \frac{\tilde{X}_n^i}{\|\tilde{X}_n^i\|_n}, \quad \|{}^i\tilde{X}_n\|_n = \|\tilde{X}_n^1\|_n \cdots \|\tilde{X}_n^i\|_n, \quad {}^iX_n = X_n^1 \Lambda \cdots \Lambda X_n^i = \frac{{}^i\tilde{X}_n}{\|{}^i\tilde{X}_n\|_n}, \text{ thus:}$$

$$\sum_1^\infty a_n |\lambda_{1i} - \langle {}^iX_n, {}^{i1}B {}^iX_n \rangle| < \infty.$$

Since $\langle {}^iX_n, {}^{i1}B {}^iX_n \rangle = \sum_{k=1}^i \|X_n^k\|_n^2 \cdots \langle X_n^k, BX_n^k \rangle_n \cdots \|X_n^i\|_n^2 = \sum_{k=1}^i \langle X_n^k, BX_n^k \rangle_n$, for $i \in \{1, \dots, r\}$, $\sum_{n=1}^\infty a_n |\sum_{k=1}^i (\lambda_k - \langle X_n^k, BX_n^k \rangle_n)| < \infty$, then $\sum_{n=1}^\infty a_n |\lambda_i - \langle X_n^i, BX_n^i \rangle_n| < \infty$. By H2b and H4,

$$\sum_{n=1}^\infty a_n |\lambda_i - \langle X_n^i, U_n X_n^i \rangle_n| = \sum_{n=1}^\infty a_n |\Lambda_n^i - \lambda_i| < \infty.$$

6.3 Proofs of Lemmas 2, 3, 4.

Lemma 6 is a deterministic version of the convergence theorem for nonnegative almost supermartingales of Robbins and Siegmund (1971).

Lemma 6. *Assume $(z_n, n \geq 1)$, $(\alpha_n, n \geq 1)$, $(\beta_n, n \geq 1)$ and $(\delta_n, n \geq 1)$ are four sequences of non-negative numbers such that for all $n \geq 1$, $z_{n+1} \leq z_n(1 + \alpha_n) + \beta_n - \delta_n$, $\sum_1^\infty \alpha_n < \infty$, $\sum_1^\infty \beta_n < \infty$. Then the sequence (z_n) converges and $\sum_1^\infty \delta_n < \infty$.*

Assume $a_n < 1$ for $n \geq 1$ without loss of generality. The abbreviation ‘‘a.s.’’ is omitted.

Proof of Lemma 2. The writing of fixed ω belonging to $\{\sum_1^\infty a_n \|B_n^1 - B\| < \infty\}$ is omitted. Since $U_{n+1}^1 - B = (1 - a_n)(U_n^1 - B) + a_n(B_n^1 - B)$,

$$\|U_{n+1}^1 - B\| \leq \|(U_n^1 - B)\| + a_n \|B_n^1 - B\| - a_n \|U_n^1 - B\|.$$

Under H2c, by Lemma 6:

$$\exists T \geq 0 \text{ such that } \|U_n^1 - B\| \rightarrow T \text{ and } \sum_1^\infty a_n \|U_n^1 - B\| < \infty;$$

since $\sum_1^\infty a_n = \infty$, $T = 0$.

$$\sum_1^\infty a_n \|U_{n+1}^1 - B\| \leq \sum_1^\infty a_n \|U_n^1 - B\| + \sum_1^\infty a_n^2 \|B_n^1 - B\| < \infty;$$

$$\sum_1^\infty a_n \|(U_{n+1}^1 - B)B_n^1\| \leq \sum_1^\infty a_n \|U_{n+1}^1 - B\| (\|B_n^1 - B\| + \|B\|) < \infty.$$

Proof of Lemma 3. Since $E[B_n^2|T_n] = 0$ and $U_{n+1}^2 = (1 - a_n)U_n^2 + a_nB_n^2$,

$$\begin{aligned} E[\|U_{n+1}^2\|^2|T_n] &= (1 - a_n)^2\|U_n^2\|^2 + a_n^2E[\|B_n^2\|^2|T_n] \\ &= (1 + a_n^2)\|U_n^2\|^2 + a_n^2E[\|B_n^2\|^2|T_n] - 2a_n\|U_n^2\|^2. \end{aligned}$$

By H2d and H3a, $\sum_1^\infty a_n^2E[\|B_n^2\|^2|T_n] < \infty$. By Robbins-Siegmund theorem (1971),

$$\exists T \geq 0 \text{ such that } \|U_n^2\|^2 \rightarrow T \text{ and } \sum_1^\infty a_n\|U_n^2\|^2 < \infty;$$

since $\sum_1^\infty a_n = \infty$, $T = 0$.

Moreover, $E[\|U_{n+1}^2\|^2] = (1 + a_n^2)E[\|U_n^2\|^2] + a_n^2E[\|B_n^2\|^2] - 2a_nE[\|U_n^2\|^2]$. Since $\sum_1^\infty a_n^2 < \infty$ and $\sup_n E[\|B_n^2\|^2] < \infty$, by Lemma 6, $\sum_1^\infty a_nE[\|U_n^2\|^2] < \infty$, $E[\|U_n^2\|^2] \rightarrow 0$, thus there exists $f > 0$ such that:

$$E[\|U_{n+1}^2\|^2] \leq (1 - 2a_n)E[\|U_n^2\|^2] + fa_n^2.$$

Let $\mu = 2 - \gamma > 0$, $c_n = \mu a_n - \gamma_n$, thus $a_n = \frac{1}{\mu}(c_n + \gamma_n)$. Under the assumptions on (a_n) and (γ_n) , $c_n \rightarrow 0$, $\sum_1^\infty c_n = \infty$, $\prod_1^n (1 - c_k) \rightarrow 0$, there exist $c > 0$ and $N_1 \in \mathbb{N}$ such that for $n \geq N_1$, $c_n < 1$ and

$$\begin{aligned} d_{n+1} &= \frac{E[\|U_{n+1}^2\|^2]}{a_{n+1}} \leq (1 + \gamma a_n + \gamma_n)(1 - 2a_n)d_n + f(1 + \gamma a_n + \gamma_n)a_n \\ &\leq (1 - (2 - \gamma)a_n + \gamma_n)d_n + ca_n \leq (1 - c_n)d_n + ca_n \\ &\leq d_{N_1} \prod_{l=N_1}^n (1 - c_l) + \frac{c}{\mu} \sum_{l=N_1}^n (c_l + \gamma_l) \prod_{k=l+1}^n (1 - c_k). \end{aligned}$$

$$\sum_{l=N_1}^n c_l \prod_{k=l+1}^n (1 - c_k) = \sum_{l=N_1}^n (\prod_{k=l+1}^n (1 - c_k) - \prod_{k=l}^n (1 - c_k)) = 1 -$$

$\prod_{k=N_1}^n (1 - c_k) < 1$, then $d_{n+1} \leq d_{N_1} + \frac{c}{\mu} + \frac{c}{\mu} \sum_{l=N_1}^\infty \gamma_l = b_1$; thus, for $n \geq N_1 - 1$,

$E[\|U_n^2\|^2] \leq b_1 a_n$ and there exists $b > 0$ such that $E[\|U_n^2\|^2] \leq b a_n$ for $n \geq 1$. Since $\sum_1^\infty a_n^{3/2} < \infty$ by H3b, $\sum_1^\infty a_n \|U_n^2\| < \infty$.

Proof of Lemma 4. By Lemmas 2 and 3, $U_n - B \rightarrow 0$ and $\sum_1^\infty a_n \|U_n - B\| < \infty$.

$$\sum_1^\infty a_n \|(U_{n+1} - B)B_n\| = \sum_1^\infty a_n \|W_{n+1}B_n\| = \sum_1^\infty a_n \|W_{n+1}(B + B_n^1 - B + B_n^2)\|.$$

By H3b, there exists $d > 0$ such that $a_n \leq da_{n+1}$, then

$$(1) \sum_1^\infty a_n \|W_{n+1}B\| \leq d \sum_1^\infty a_{n+1} \|W_{n+1}\| \|B\| < \infty;$$

$$(2) \sum_1^\infty a_n \|W_{n+1}(B_n^1 - B)\| \leq \sum_1^\infty a_n \|W_{n+1}\| \|B_n^1 - B\| < \infty \text{ by H2c and } W_n \rightarrow 0;$$

$$(3) \sum_1^\infty a_n \|W_{n+1}B_n^2\| \leq d \sum_1^\infty a_{n+1} \|W_{n+1}\| \|B_n^2\| < \infty \text{ by H2f};$$

$$(4) \|W_{n+1}B_n^2\| \leq \|(U_{n+1}^1 - B)B_n^2\| + \|U_{n+1}^2 B_n^2\|; \sum_1^\infty a_n \|(U_{n+1}^1 - B)B_n^2\| < \infty \text{ by H2e,}$$

$$E[\|U_{n+1}^2 B_n^2\|] \leq (E[\|U_{n+1}^2\|^2])^{\frac{1}{2}} (E[\|B_n^2\|^2])^{\frac{1}{2}} = O(\sqrt{a_{n+1}}) \text{ by Lemma 3,}$$

$$\sum_1^\infty a_n \|U_{n+1}^2 B_n^2\| < \infty \text{ by H3b, then } \sum_1^\infty a_n \|W_{n+1}B_n^2\| < \infty.$$

Thus by (1), (2), (3 or 4), $\sum_1^\infty a_n \|W_{n+1} B_n\| < \infty$. H2b holds.

6.4 Proof of Theorem 2

The writing of fixed ω belonging to the set $\{\sum_1^\infty a_n \|Q_{n+1} - M^{-1}\| < \infty\}$ is omitted.

$M_n - M = (I - a_n Q_{n+1})(M_{n-1} - M) - a_n(Q_{n+1} - M^{-1})M$. Let $M_n^1 = M_n - M$.

$$\|M_n^1\| \leq (\|I - a_n M^{-1}\| + a_n \|Q_{n+1} - M^{-1}\|) \|M_{n-1}^1\| + a_n \|Q_{n+1} - M^{-1}\| \|M\|.$$

Let λ_{min} be the lowest eigenvalue of M^{-1} . Since $a_n = o(1)$, $\|I - a_n M^{-1}\| = 1 - a_n \lambda_{min}$ for n sufficiently large. Thus

$$\|M_n^1\| \leq (1 + a_n \|Q_{n+1} - M^{-1}\|) \|M_{n-1}^1\| + a_n \|Q_{n+1} - M^{-1}\| \|M\| - a_n \lambda_{min} \|M_{n-1}^1\|.$$

Since $\sum_1^\infty a_n \|Q_{n+1} - M^{-1}\| < \infty$, by Lemma 6, $\exists T \geq 0$ such that $\|M_{n-1}^1\| \rightarrow T$ and $\sum_1^\infty a_n \|M_{n-1}^1\| < \infty$; since $\sum_1^\infty a_n = \infty$, $T = 0$; note that $\sum_1^\infty a_n \|M_n^1\| < \infty$.

References

1. Balsubramani, A., S. Dasgupta and Y. Freund. 2013. The fast convergence of incremental PCA. *Adv. Neural Inf. Syst.* 26:3174-3182.
2. Benzécri, J.P. 1969. Approximation stochastique dans une algèbre normée non commutative. *Bull. Soc. Math. France* 97:225-241.
3. Brandière, O. 1994. Un algorithme du gradient pour l'analyse en composantes principales. Prépublication 14, Université de Marne-la-Vallée.
4. Brandière, O. 1995. Un algorithme pour l'analyse en composantes principales. *C.R. Acad. Sci.* 321 (série I):233-236.
5. Brandière, O. 1998. Some pathological traps for stochastic approximation. *SIAM J. Control Optim.* 36 (4):1293-1314.
6. Brandière, O. and M. Duflo. 1996. Les algorithmes stochastiques contournent-ils les pièges? *Ann. Inst. Henri Poincaré Probab. Stat.* 32:395-427.
7. Cardot, H., P. Cénac and J.M. Monnez. 2012. A fast and recursive algorithm for clustering large datasets with k-medians. *Comput. Statist. Data Anal.* 56:1434-1449. [doi:10.1016/j.csda.2011.11.019](https://doi.org/10.1016/j.csda.2011.11.019)
8. Cardot, H. and D. Degras. 2018. Online principal component analysis in high dimension: which algorithm to choose? *Internat. Statist. Rev.* 86:29-50.
9. Carroll, J.D. 1968. A generalization of canonical correlation analysis to three or more sets of variables. In *Proceedings of the 76th Convention American Psychological Association*, vol.3, 227-228.

10. Duflo, M. 1997. *Random Iterative Models*. Applications in Mathematics, vol. 34, Springer-Verlag, Berlin.
11. Karhunen, J. and E. Oja. 1982. New methods for stochastic approximation of truncated Karhunen-Loève expansions. In *Proceedings of the 6th International Conference on Pattern Recognition*, IEEE Computer Society Press, 550-553.
12. Krasulina, T.P. 1970. Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices. *Autom. Remote Control* 2:215-221.
13. Lalloué, B., J.M. Monnez and E. Albuissou. 2022. Streaming constrained binary logistic regression with online standardized data. *J. Applied Statistics* 49 (6):1519-1539. [doi: 10.1080/02664763.2020.1870672](https://doi.org/10.1080/02664763.2020.1870672)
14. Ljung, L., G.C. Pflug and H. Walk. 1992. *Stochastic Approximation and Optimization of Random Systems*. DMV Seminar, vol. 17, Birkhäuser, Basel.
15. Monnez, J.M. 1991. Etude d'un modèle de corrélation en approximation stochastique (Study of a correlation model in stochastic approximation). *Statistique et Analyse des Données* 16 (3):143-181.
16. Monnez, J.M. 2006. Almost sure convergence of stochastic gradient processes with matrix step sizes. *Stat. Proba. Letters* 76:531-536. [doi: 10.1016/j.spl.2005.09.014](https://doi.org/10.1016/j.spl.2005.09.014)
17. Monnez, J.M. and A. Skiredj. 2021. Widening the scope of an eigenvector stochastic approximation process and application to streaming PCA and related methods. *J. Multivariate Anal.* 182:104694. [doi: 10.1016/j.jmva.2020.104694](https://doi.org/10.1016/j.jmva.2020.104694)
18. Oja, E. 1982. A simplified neuron model as a principal component analyser. *J. Math. Biol.* 15:267-273.
19. Oja, E. and J. Karhunen. 1985. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *J. Math. Anal. Appl.* 106:69-84.
20. Robbins, H. and D. Siegmund. 1971. A convergence theorem for nonnegative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, ed. J.S. Rustagi, 233-257. New-York: Academic Press.