Stochastic approximation of the factors of a generalized canonical correlation analysis

Jean-Marie Monnez

Stochastic Approximation of the Factors of a Generalized Canonical Correlation Analysis

by

Jean-Marie Monnez

Université Henri Poincaré

Institut Elie Cartan, Laboratoire de Mathématiques

B.P. 239-F 54506 Vandoeuvre lès Nancy Cedex

Abstract

We suppose that data of a generalized canonical correlation analysis are i.i.d. observations of a random vector Z which are taken sequentially. We define a recursive method of sequential estimation of the factors. This can be applied also when there is a great amount of non random data vectors.

Key Words and Phrases: Stochastic approximation, Canonical correlation analysis

AMS 1991 Subject Classifications. Primary: 62**; Secondary:** L20, H20.

1 Introduction

Consider the generalized canonical correlation analysis (gCCA) as defined by Carroll (1968). Suppose that data are i.i.d. observations of a random vector Z which are taken sequentially: at time n, Z_n is observed. We define a recursive method of sequential estimation of the general and canonical factors of gCCA by using a stochastic approximation process derived from that of Krasulina (1970), whose convergence has been studied by Bouamaine and Monnez (1998). This method can be used also when there is a great amount of non random data vectors: at time n, a data vector is randomly chosen.

Numerical results for this type of algorithm in the case of principal component analysis are given by Bouamaine (1996). An implementation is in progress of an algorithm corresponding to a multiple factorial analysis, which treats the same type of data as gCCA, and also a number of simulations are being made.

We present gCCA in the second section, stochastic approximation of the parameter of a linear regression function in the third and stochastic approximation of the factors in the fourth. In the following, ρ denotes a linear correlation coefficient, A' the transposed matrix of A, $\lambda_{\min}(A)$ the smallest eigenvalue of A, $\mathbb{R}^{p*}$ the dual space of $\mathbb{R}^p$. The abbreviation $a.s.$ means almost surely.

2 Generalized canonical correlation analysis of a random vector

Consider two random vectors Z^1 and Z^2 defined on the same probability space and suppose that there is no affine relation between the components of $Z = (Z^1, Z^2)$. The canonical correlation analysis (CCA) consists in determining at step $l \in \{1, ..., r\}$ for $k = 1, 2$ a linear combination of variance 1 of the centered components of Z^k, $V_l^k = (\eta_l^k)'(Z^k - E\left[Z^k\right])$, uncorrelated with $V_{l-1}^k, ..., V_1^k$, which maximizes $\rho(V_l^1, V_l^2)$. This analysis permits one to put into relation two sets of variables or to explain one set of variables by the other.

A generalization of CCA to three or more sets of variables is gCCA. As it will be seen, gCCA can be interpreted as a principal component analysis (PCA) of a random vector $Z = (Z^1, ..., Z^q)$ in $\mathbb{R}^p$, $Z^1, ..., Z^q$ being q random vectors which take in some sense a similar part in the determination of the factors.

More precisely, suppose that the set of components of a random vector Z in $\mathbb{R}^p$ is divided into q sets of real random variables $\left\{Z^{k1}, ..., Z^{km_k}\right\}, k = 1, ..., q, \sum_{k=1}^q m_k = p$. Denote Z^k the random vector in $\mathbb{R}^{m_k}$ whose components are $Z^{k1}, ..., Z^{km_k}$. Suppose that there is no affine relation between the components of Z.

Consider the following problem: for $l = 1, ..., r$, determine at step l a linear combination of all the centered variables Z^{kj}, $U_l = (\theta_l)'\left(Z - E\left[Z\right]\right), \theta_l \in \mathbb{R}^{p*}$, named l^{th} general component, which is of variance 1 and uncorrelated with the preceding ones, and for $k = 1, ..., q$, a linear combination of variance 1 of the centered variables

of the k^{th} set, $V_l^k = (\eta_l^k)'(Z^k - E\left[Z^k\right]), \eta_l^k \in \mathbb{R}^{m_k *}$, named l^{th} canonical component of the k^{th} set, which maximize $\sum_{k=1}^{q} \rho^2(U_l, V_l^k)$. θ_l is named the l^{th} general factor and η_l^k the l^{th} canonical factor of the k^{th} set.

Remark that when $m_k = 1$ for $k = 1, ..., p$ $(q = p)$, this analysis is equivalent to the normed principal component analysis of Z i.e. PCA of Z with the inner product in $\mathbb{R}^p$ defined by the weight matrix

$$\begin{pmatrix} \frac{1}{(\sigma^1)^2} & & & \\ & . & & \\ & & . & \\ & & & . \\ & & & & \frac{1}{(\sigma^p)^2} \end{pmatrix}, \left(\sigma^k\right)^2 = E\left[\left(Z^k - E\left[Z^k\right]\right)^2\right].$$

In this PCA, the random variables $Z^1, ..., Z^p$ play in some sense a similar part in the definition of the squared distance in $\mathbb{R}^p$ between two data vectors $Z(\omega_1)$ and $Z(\omega_2)$:

$$d^2(Z(\omega_1), Z(\omega_2)) = \sum_{k=1}^{p} \left(\frac{Z^k(\omega_1) - Z^k(\omega_2)}{\sigma^k}\right)^2.$$

In the general case, denote C^k the covariance matrix of Z^k, C that of Z and define the covariance matrices

$$\begin{aligned} C^{kl} &= E\left[(Z^k - E\left[Z^k\right])(Z^l - E\left[Z^l\right])'\right] \\ C^{k.} &= E\left[(Z^k - E\left[Z^k\right])(Z - E[Z])'\right] = (C^{k1}...C^{kq}). \end{aligned}$$

Let M be the block diagonal matrix

$$M = \begin{pmatrix} (C^1)^{-1} & & & \\ & . & & \\ & & . & \\ & & & . \\ & & & & (C^q)^{-1} \end{pmatrix}.$$

Let $\theta_l = (\theta_l^{1\prime}, ..., \theta_l^{q\prime})', \theta_l^k \in \mathbb{R}^{m_k *}$ for $k = 1, ..., q$. Consider the matrix $B = MC$ which can be divided into q^2 blocks, the block (k, l) being the (m_k, m_l) matrix $(C^k)^{-1} C^{kl}$.

Theorem 1 *θ_l is a C-normed eigenvector of the matrix $B = MC$ corresponding to the l^{th} greatest eigenvalue λ_l and for $k = 1, ..., q$, there exists $\alpha_l^k \in \mathbb{R}$ such that $\eta_l^k = \alpha_l^k \theta_l^k$.*

Proof. Assume U_l to be determined. Then: $\sum_{k=1}^q \rho^2(U_l, V_l^k)$ max $\Longleftrightarrow$ $\rho^2(U_l, V_l^k)$ max, $k = 1, ..., q$. Thus V_l^k is obtained by making the linear regression of U_l with respect to Z^k. Therefore there exists $\beta_l^k \in \mathbb{R}$ such that

$$\begin{aligned} V_l^k &= \beta_l^k \Pi^k U_l, \qquad \Pi^k U_l = (Z^k - E\left[Z^k\right])' b_l^k, \\ b_l^k &= (C^k)^{-1} E\left[(Z^k - E\left[Z^k\right]) U_l\right] = (C^k)^{-1} C^{k\cdot} \theta_l. \\ \sum_{k=1}^q \rho^2(U_l, V_l^k) &= \sum_{k=1}^q E\left[U_l \Pi^k U_l\right] = \sum_{k=1}^q (\theta_l)' (C^{k\cdot})' (C^k)^{-1} C^{k\cdot} \theta_l = (\theta_l)' C M C \theta_l. \end{aligned}$$

θ_l maximizes $(\theta_l)'CMC\theta_l$ under the constraints $\theta_l'C\theta_j = 0,\ j = 1,\ 2,\ ..., l-1$ and $\theta_l'C\theta_l = 1$. Then θ_l is eigenvector of MC corresponding to the eigenvalue λ_l and $\eta_l^k = \beta_l^k b_l^k = \beta_l^k (C^k)^{-1} C^{k\cdot}\theta_l = \beta_l^k \lambda_l \theta_l^k$. ■

Remark that θ_l is the l^{th} factor of the principal component analysis of Z, the inner product in $\mathbb{R}^p$ being defined by M. In this PCA, the random vectors $Z^1, ..., Z^q$ play in some sense a similar part in the definition of the squared distance in $\mathbb{R}^p$ between two data vectors $Z(\omega_1)$ and $Z(\omega_2)$:

$$d^2(Z(\omega_1), Z(\omega_2)) = \sum_{k=1}^{q} (Z^k(\omega_1) - Z^k(\omega_2))' \left(C^k\right)^{-1} (Z^k(\omega_1) - Z^k(\omega_2)).$$

Remark too that this analysis in the case $q = 2$ is the usual canonical correlation analysis. $\theta_l = \begin{pmatrix} \theta_l^1 \\ \\ \theta_l^2 \end{pmatrix}$ is eigenvector of $\begin{pmatrix} I & (C^1)^{-1}C^{12} \\ \\ (C^2)^{-1}C^{21} & I \end{pmatrix}$; θ_l^k is colinear to the l^{th} canonical factor in Z^k for $k = 1, 2$.

3 Stochastic approximation of a linear regression function

3.1 Proposition

Let R be a random vector in $\mathbb{R}^r$ and suppose:

(H1) *There is no affine relation between the components of* R.

Let S be a random vector in $\mathbb{R}^s$. Denote $CovarR$ the covariance matrix of R and $Covar(R, S) = E[(R - E[R])(S - E[S])']$. Let A be a (r, s) matrix and D a $(s, 1)$ matrix. $(CovarR)^{-1}Covar(R, S)$ is the solution in A of the following linear regression

problem: find (A, D) such that $E\left[\|S - A'R - D\|^2\right]$ be minimal, $\|.\|$ being the usual euclidean norm.

Denote A_1 the $(r+1, s)$ matrix such that $A_1' = \begin{pmatrix} A' & D \end{pmatrix}$ and R_1 the random vector in $\mathbb{R}^{r+1}$ such that $R_1' = \begin{pmatrix} R' & 1 \end{pmatrix}$. $E\left[R_1 R_1'\right]$ is positive definite by H1. $(CovarR)^{-1}Covar(R, S)$ can be obtained as it follows.

Proposition 2 *$A = (Covar(R))^{-1}Covar(R, S)$ and $D = E[S] - A'E[R]$ form the solution of the linear system $E\left[R_1 R_1'\right] A_1 = E\left[R_1 S'\right]$.*

3.2 Stochastic approximation of $\mathbf{A}_1$

Consider an i.i.d sample $((R_n, S_n))$ of (R, S). Denote for all n R_{1n} the random vector in $\mathbb{R}^{r+1}$ such that $R_{1n}' = \begin{pmatrix} R_n' & 1 \end{pmatrix}$. Consider the stochastic approximation process $(A_{1n}, n \geq 1)$ of $A_1 = \left(E\left[R_1 R_1'\right]\right)^{-1} E\left[R_1 S'\right]$ in the set of $(r+1, s)$ matrices such that

$$A_{1,n+1} = A_{1n} - a_n(R_{1n}R_{1n}'A_{1n} - R_{1n}S_n'), \qquad a_n \in \mathbb{R}^{+*}.$$

Denote $A_{1n} = \binom{A_n}{D_n'}$, A_n and D_n being respectively a (r, s) and a $(s, 1)$ random matrix. $(A_n, n \geq 1)$ is a stochastic approximation process of $A = (Covar(R))^{-1}Covar(R, S)$.

We use in the proof of the convergence the Hilbert-Schmidt inner product $\langle ., . \rangle$ and norm $\|.\|$ in the set of $(r+1, s)$ matrices: $\langle A, B \rangle = Trace(A'B) = \sum_{i=1}^{r+1} \sum_{j=1}^{s} A(i, j)B(i, j)$. Suppose:

(H2) *(R, S) has 4^{th} order moments.*

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

(H3') $\left(a_n = \frac{a}{n^\alpha}, \quad a > 0, \quad \frac{2}{3} < \alpha < 1\right)$ *or* $\left(a_n = \frac{a}{n}, \quad a > \frac{1}{2\lambda_{\min}\left(E\left[R_1 R_1'\right]\right)}\right)$.

Theorem 3 *Assume H1, H2, H3; then* $\|A_{1n} - A_1\|$ *and* $\|A_n - A\|$ *converge to 0 a.s. and in* L^2*; if H3 is replaced by H3', then moreover* $\sum_1^\infty a_n \|A_{1n} - A_1\| < \infty$ *and* $\sum_1^\infty a_n \|A_n - A\| < \infty$ *a.s.*

Proof. Denote $B_n = R_{1n}R_{1n}'$, $F_n = R_{1n}S_n'$, $B = E\left[R_1 R_1'\right]$, $F = E\left[R_1 S'\right]$. As $A_1 = B^{-1}F$, we have:

$$\begin{aligned}
A_{1,n+1} - A_1 &= A_{1n} - A_1 - a_n B(A_{1n} - A_1) - a_n U_n\,, \\
U_n &= (B_n - B)A_{1n} + F - F_n. \\
\|A_{1,n+1} - A_1\|^2 &= \|A_{1n} - A_1\|^2 + a_n^2 \|B(A_{1n} - A_1)\|^2 + a_n^2 \|U_n\|^2 \\
&\quad -2a_n \langle A_{1n} - A_1, B(A_{1n} - A_1)\rangle - 2a_n \langle A_{1n} - A_1, U_n\rangle \\
&\quad +2a_n^2 \langle B(A_{1n} - A_1), U_n\rangle\,.
\end{aligned}$$

Denote $\mathcal{T}_n$ the σ-field generated by (R_j, S_j), $j < n$ and A_{11}. As $E(U_n \mid \mathcal{T}_n) = 0$, we have:

$$\begin{aligned}
E\left(\|A_{1,n+1} - A_1\|^2 \mid \mathcal{T}_n\right) &\leq \|A_{1n} - A_1\|^2 \left(1 + a_n^2 \|B\|^2\right) + a_n^2 E\left(\|U_n\|^2 \mid \mathcal{T}_n\right) \\
&\quad -2a_n \langle A_{1n} - A_1, B(A_{1n} - A_1)\rangle\,.
\end{aligned}$$

By H2, there exist c and $d > 0$ such that:

$$E(\|U_n\|^2 \mid \mathcal{T}_n) = E\left(\|(B_n - B)A_{1n} + (F - F_n)\|^2 \mid \mathcal{T}_n\right) \leq c\,\|A_{1n} - A_1\|^2 + d.$$

Denote $M^{(j)}$ the j^{th} column of a matrix M and λ the smallest eigenvalue of B. Then:

$$\begin{aligned}\langle A_{1n} - A_1, B(A_{1n} - A_1)\rangle &= \sum_{j=1}^{s} \left\langle A_{1n}^{(j)} - A_1^{(j)}, B(A_{1n}^{(j)} - A_1^{(j)})\right\rangle \\ &\geq \lambda \sum_{j=1}^{s} \left\| A_{1n}^{(j)} - A_1^{(j)} \right\|^2 = \lambda \left\| A_{1n} - A_1 \right\|^2.\end{aligned}$$

Therefore:

$$E\left(\left\|A_{1,n+1} - A_1\right\|^2 \mid \mathcal{T}_n\right) \leq (1 + \|B\|^2 a_n^2 + c a_n^2) \left\|A_{1n} - A_1\right\|^2 + d a_n^2 - 2\lambda a_n \left\|A_{1n} - A_1\right\|^2.$$

By H3, applying the Robbins-Siegmund lemma (1971) gives the $a.s.$ convergence of $\|A_{1n} - A_1\|$ and of $\sum_1^\infty a_n \|A_{1n} - A_1\|^2$. As $\sum_1^\infty a_n = \infty$, we have $\|A_{1n} - A_1\| \longrightarrow 0$ $a.s.$ Taking the expectation of $E\left(\|A_{1,n+1} - A_1\|^2 \mid \mathcal{T}_n\right)$ gives in the same way $E\left[\|A_{1n} - A_1\|^2\right] \longrightarrow 0$.

Thus there exist $N \in \mathbb{N}$ and $f > 0$ such that for $n > N$:

$$E\left[\|A_{1,n+1} - A_1\|^2\right] \leq (1 - 2\lambda a_n) E\left[\|A_{1n} - A_1\|^2\right] + f a_n^2.$$

Applying for instance lemmas 2 and 3 in Monnez (1992) gives for $a_n = \frac{a}{n^\alpha}$ with $0 < \alpha < 1$, or $\alpha = 1$ and $a > \frac{1}{2\lambda}$, $\overline{\lim} \frac{1}{a_n} E\left[\|A_{1n} - A_1\|^2\right] < \infty$. Hence there exists $g > 0$ such that

$$\sum_1^\infty a_n E\left[\|A_{1n} - A_1\|\right] \leq g \sum_1^\infty a_n^{\frac{3}{2}} < \infty$$

for $\alpha > \frac{2}{3}$; it follows that $\sum_1^\infty a_n \|A_{1n} - A_1\| < \infty$ $a.s.$ ■

4 Stochastic approximation of general factors

4.1 Definition of the process

Let $((Z_n^1, ..., Z_n^q), n \geq 1)$ be an i.i.d. sample of the random vector $Z = (Z^1, ..., Z^q)$ in $\mathbb{R}^p$. For $k = 1, ..., q$ and $n \geq 1$, let

$$\overline{Z_n^k} = \frac{1}{n} \sum_{i=1}^{n} Z_i^k, \quad M_n^k = \frac{1}{n} \sum_{i=1}^{n} Z_i^k Z_i^{k\,\prime} - \overline{Z_n^k} \left(\overline{Z_n^k} \right)';$$

M_n^k is a consistent estimator of the covariance matrix C^k of Z^k. For $n \geq 1$, define the block-diagonal (p, p) matrix

$$M_n = \begin{pmatrix} M_n^1 & & & \\ & . & & \\ & & . & \\ & & & . \\ & & & M_n^q \end{pmatrix};$$

M_n is a consistent estimator of M^{-1}. Denote $\langle ., . \rangle_n$ the inner product defined by M_n in the dual space $\mathbb{R}^{p*}$ of $\mathbb{R}^p$ and $\|.\|_n$ the associated norm.

In order to estimate $B^{ij} = (C^i)^{-1} C^{ij}$ for $i = 1, ..., q$ and $j \neq i$, define the stochastic approximation processes (B_{1n}^{ij}) in the set of $(m_i + 1, m_j)$ matrices in the same way as in paragraph 3.2, (F_n^{ij}) in the set of $(m_j, 1)$ matrices and (B_n^{ij}) in the set of (m_i, m_j)

matrices:

$$\begin{aligned} B_{1,n+1}^{ij} &= B_{1n}^{ij} - a_n \left(Z_{1n}^i (Z_{1n}^i)' B_{1n}^{ij} - Z_{1n}^i (Z_n^j)' \right), \\ Z_{1n}^i &= \begin{pmatrix} Z_n^i \\ 1 \end{pmatrix}, B_{1n}^{ij} = \begin{pmatrix} B_n^{ij} \\ (F_n^{ij})' \end{pmatrix}. \end{aligned}$$

Define in the set of (p,p) matrices the stochastic approximation process (B_n) of the matrix $B = MC$ such that the block (i,j) of B_n is B_n^{ij}:

$$B_n = \begin{pmatrix} I & B_n^{12} & . & . & B_n^{1q} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ B_n^{q1} & B_n^{q2} & . & . & I \end{pmatrix}.$$

In order to estimate r general factors of GCCA, define for $l = 1, 2, ..., r$, the Krasulina-type stochastic approximation process (X_n^l) of the general factor θ_l in $\mathbb{R}^{p*}$ (Krasulina, 1970, Bouamaine and Monnez, 1998):

$$\begin{aligned} X_{n+1}^l &= orth_{M_n}(Y_{n+1}^l) \\ Y_{n+1}^l &= X_n^l + a_n \left(B_n - F_n(X_n^l) I \right) X_n^l \\ F_n(X_n^l) &= \frac{\left\langle B_n X_n^l, X_n^l \right\rangle_{n-1}}{\|X_n^l\|_{n-1}^2}; \end{aligned}$$

$X_{n+1}^l = orth_{M_n}(Y_{n+1}^l)$ means that $(X_{n+1}^1, ..., X_{n+1}^l)$ is obtained by a Gram-Schmidt orthogonalization with respect to M_n of $(Y_{n+1}^1, ..., Y_{n+1}^l)$.

4.2 Principle of the study of convergence

The details of the proof of convergence of this type of process are given by Bouamaine and Monnez (1998).

Denote $\wedge$ the exterior product of elements of $\mathbb{R}^{p*}$ and for $j = 1, ..., p$, ${}^{j}\wedge\mathbb{R}^{p*}$ the exterior algebra of order j of $\mathbb{R}^{p*}$: $(e_1, ..., e_p)$ being a basis of $\mathbb{R}^{p*}$, the set of C_p^j exterior products $e_{i_1}\wedge...\wedge e_{i_j}$ with $1 \leq i_1 < ... < i_j \leq p$ is a basis of ${}^{j}\wedge\mathbb{R}^{p*}$.

Define the inner product in ${}^{j}\wedge\mathbb{R}^{p*}$ derived from that in $\mathbb{R}^{p*}$ defined by M^{-1}; in this definition G_j is the set of permutations of $\{k_1, ..., k_j\}$, $s(\sigma)$ is the number of inversions in the permutation σ and $\varepsilon(\sigma) = (-1)^{s(\sigma)}$:

$$\left\langle e_{i_1}\wedge...\wedge e_{i_j}, e_{k_1}\wedge...\wedge e_{k_j}\right\rangle = \sum_{\sigma\in G_j} \varepsilon(\sigma) \left\langle e_{i_1}, e_{\sigma(k_1)}\right\rangle_{M^{-1}} ... \left\langle e_{i_j}, e_{\sigma(k_j)}\right\rangle_{M^{-1}}.$$

Suppose that the r greatest eigenvalues of the endomorphism $B = MC$ in $\mathbb{R}^{p*}$ are different: $\lambda_1 > ... > \lambda_r$. Define for $j = 1, ..., r$, the endomorphism ${}^{j1}B$ in ${}^{j}\wedge\mathbb{R}^{p*}$ by:

$${}^{j1}B(x^1\wedge.....\wedge x^j) = \sum_{h=1}^{j} x^1\wedge...\wedge Bx^h\wedge...\wedge x^j, x^l \in \mathbb{R}^{p*}, l = 1, ..., j.$$

If $V^1, ..., V^j$ are eigenvectors of B corresponding respectively to $\lambda_1, ..., \lambda_j$, $V^1\wedge...\wedge V^j$ is eigenvector of ${}^{j1}B$ corresponding to the greatest eigenvalue $\lambda_{1j} = \sum_{l=1}^{j} \lambda_l$. Denote ${}^{j}S_1$ the eigensubspace corresponding to λ_{1j} and $({}^{j}S_1)^{\perp}$ its orthogonal.

The proof of convergence is made in two steps. Denote ${}^{j}X_n = X_n^1\wedge...\wedge X_n^j$. It is proved first that for $j = 1, ..., r$, $\frac{{}^{j}X_n}{\|{}^{j}X_n\|}$ converges $a.s.$ in a set ${}^{j}E$ to $V^1\wedge...\wedge V^j \in {}^{j}S_1$.

It is proved then that, for $l = 1, ..., r$, $\frac{X_n^l}{\left\|X_n^l\right\|}$ converges $a.s.$ in $\cap_{j=1}^{l} {}^jE$ to V^l. The definition of the convergence set ${}^jE$ is as follows.

In the deterministic case $B_n = B$, $M_n = M^{-1}$, define

$$h_j({}^jx) = \frac{\langle {}^{j1}B\, {}^jx, {}^jx\rangle}{\langle {}^jx, {}^jx\rangle}, {}^jx \in {}^j\wedge\mathbb{R}^{p*},$$

and the process $({}^jU_n)$ in ${}^j\wedge\mathbb{R}^{p*}$ by

$${}^jU_{n+1} = \left(I + a_n\left({}^{j1}B - h_j({}^jU_n)I\right)\right)\ {}^jU_n.$$

In this case ${}^jE$ is the set $\left\{{}^jX_1 \notin ({}^jS_1)^{\perp}\right\}$; $X_1^1, ..., X_1^j$ may not be orthogonal to the subspace generated by the eigenvectors of B corresponding to its j greatest eigenvalues.

The process $({}^jX_n) = (X_n^1 \wedge ... \wedge X_n^j)$ can be considered as resulting from a stochastic perturbation of $({}^jU_n)$; denote

$$\begin{aligned}
\Delta_{nj} &= 1 + a_n(\lambda_{1j} - h_j({}^jX_n)) \\
Q^j &= {}^jX_1 + \sum_{n=1}^{\infty} \frac{{}^jX_{n+1} - \left(I + a_n\left({}^{j1}B - h_j({}^jX_n)I\right)\right)\ {}^jX_n}{\prod_{i=1}^{n}\Delta_{ij}}.
\end{aligned}$$

Then ${}^jE = \left\{Q^j \notin ({}^jS_1)^{\perp}\right\}$. Remark that $Q^j = {}^jX_1$ in the case of the process $({}^jU_n)$.

4.3 Convergence of the process

Suppose:

(H1') *There is no affine relation between the components of Z.*

(H2') *Z has* 4^{th} *order moments.*

(H3") $\left(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1\right)$ *or* $\left(a_n = \frac{a}{n}, a > \frac{1}{2 \min_i \lambda_{\min} E\left[Z_1^i Z_1^{i\prime}\right]}, Z_1^i = \begin{pmatrix} Z^i \\ \\ 1 \end{pmatrix} \right)$.

Theorem 4 *Assume H1', H2', H3"; then, for* $l = 1, ..., r$, Q^l *converges a.s. Suppose moreover that the* r *greatest eigenvalues of* B *are different; then, for* $l = 1, ..., r$, X_n^l *converges a.s. in* $\cap_{j=1}^{l} {}^{j}E$ *to a random eigenvector of* B *corresponding to its* l^{th} *greatest eigenvalue.*

Proof. The convergence of this type of process is proved in theorem 4 of Bouamaine and Monnez (1998, p. 24) under the following assumptions:

(1) $\sum_1^\infty a_n \left\| B_n - B \right\| < \infty$ *a.s.*

(2) $M_n \longrightarrow M^{-1}, \sum_1^\infty a_n \left\| M_{n-1} - M^{-1} \right\| < \infty$ *a.s.*

(3) $a_n > 0, \sum_1^\infty a_n = \infty, \sum_1^\infty a_n^2 < \infty$.

It suffices to verify (1) and (2).

By H1', H2', H3" and theorem 3, $\sum_{n=1}^\infty a_n \left\| B_n^{ij} - B^{ij} \right\| < \infty$ *a.s.* for every pair (i, j); then (1) is verified. For $k = 1, ..., q$:

$$\left\| M_{n-1}^k - C^k \right\| \leq \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} Z_i^k Z_i^{k\prime} - E\left[Z^k Z^{k\prime}\right] \right\| + \left\| \overline{Z_n^k} - E\left[Z^k\right] \right\| \left(\left\| \overline{Z_n^k} \right\| + \left\| E\left[Z^k\right] \right\| \right)$$

$$\overline{Z_n^k} \longrightarrow E\left[Z^k\right] \text{ a.s.}; \sum_{n=1}^{\infty} a_n E\left[\left\| \overline{Z_n^k} - E\left[Z^k\right] \right\| \right] = \sum_{n=1}^{\infty} a_n O\left(\frac{1}{\sqrt{n}}\right) < \infty.$$

Then: $\sum_{n=1}^\infty a_n \left\| M_{n-1}^k - C^k \right\| < \infty$ for $k = 1, ..., q$, $\sum_{n=1}^\infty a_n \left\| M_{n-1} - M^{-1} \right\| < \infty$ *a.s.*

(2) is verified. ■

References

Bouamaine, A., 1996. Méthodes d'approximation stochastique en analyse des données. Thèse de Doctorat d'Etat ès Sciences Appliquées, Université Mohamed V, EMI, Rabat.

Bouamaine, A., Monnez, J.M., 1998. Approximation stochastique de vecteurs et valeurs propres. Pub. Inst. Stat. Univ. Paris 42, n° 2-3, pp. 15-38.

Carroll, J.D., 1968. A generalization of canonical correlation analysis to three or more sets of variables. Proceedings of the 76^{th} Convention Amer. Psycho. Assoc., pp. 227-228.

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

Monnez, J.M., 1992. Etude d'un modèle de corrélation en approximation stochastique. Statistique et Analyse des Données 16, n° 3, pp. 143-181.

Monnez, J.M., 1994. Convergence d'un processus d'approximation stochastique en analyse factorielle. Pub. Inst. Stat. Univ. Paris 38, n° 1, pp. 37-56.

Robbins, H., Siegmund, D., 1971. A convergence theorem for non negative al-

most supermartingales and some applications. In: Rustagi, J.S. (Ed.), Optimizing Methods in Statistics, Academic Press, New-York, pp. 233-257.