An extended Oja process for streaming canonical analysis

Jean-Marie Monnez

An extended Oja process for streaming canonical analysis

Jean-Marie Monnez [a] *

[a] **Université de Lorraine, CNRS, Inria*[1], *IECL*[2], *F-54000 Nancy, France *jean-marie.monnez@univ-lorraine.fr*

[1] Inria, project-team BIGS

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

An extended Oja process for streaming canonical analysis

Canonical components of the canonical analysis of two random vectors are collinear with principal components of a PCA of the multidimensional linear regression function of one vector with respect to the other. In the context of streaming data, we estimate online in parallel this regression function and components of a canonical correlation analysis, taking into account at each step a mini-batch of current data or all the data up to the current step to have a faster convergence, and using extended Oja processes. We extend this approach to generalized canonical correlation analysis and deal with the cases of streaming factorial correspondence analysis, multiple correspondence analysis and factorial discriminant analysis.

Keywords: incremental learning; projected PCA; stochastic approximation; streaming linear regression, streaming canonical analysis

1. Introduction

Let Q be a positive definite symmetric (p,p) matrix called metric, $\langle .,. \rangle$ be the inner product and $\|.\|$ 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,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 $\mathbb{R}^{p\times q}$ be the space of (p,q) matrices. 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. Let $\mathbb{R}^{p*}$ be the dual space of $\mathbb{R}^p$. The abbreviation a.s. stands for almost surely and Si.j.k for Subsection i.j.k.

In the context of streaming data or big data processed sequentially as a data stream, that can oversize memory storage or computation capacity, stochastic approximation algorithms can be used to estimate online statistical parameters. For example: parameters of a regression function (Ljung, Pflug and Walk 1992, Duarte, Monnez and Albuisson, 2018, Lalloué, Monnez and Albuisson 2022); centers of clusters in unsupervised classification (Monnez 2006, Cardot, Cénac and Monnez 2012); principal components in principal component analysis (PCA) (Benzécri 1969, Duflo 1997, Cardot and Degras 2018, Monnez 2022a and references therein). The incoming observation vectors are 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, for PCA).

To estimate online eigenvectors corresponding to eigenvalues in decreasing order of a Q-symmetric matrix B, Q and B being unknown, we define in (Monnez 2022a) an extension of the Oja process (Oja and Karhunen 1985). In the Oja process, the authors supposed that B is the expectation of a random matrix of which one observation is used at each step and that $Q = I$. A corollary of the almost sure convergence theorem of (Monnez 2022a) recalled in Section 2 (Lemma 1) applies to the case where there exist two sequences of observable random matrices $(V_n, n \geq 1)$ and $(Q_n, n \geq 1)$ converging almost surely to B and Q respectively. If the estimator V_n of B is determined sequentially by a stochastic approximation process, we can perform in parallel the online estimation of B and that of eigenvectors and eigenvalues of B. We apply this principle to the online estimation of canonical components of a streaming canonical analysis (CA) of two random vectors.

CA of Z^1 and Z^2 consists in determining at step $i \in \{1, \dots, r\}$ a couple of canonical components $\left(V_i^1, V_i^2\right)$ that are affine combinations of the components of two random vectors Z^1 and Z^2 respectively, centered, of variance 1, uncorrelated respectively with the previous components, maximizing $E\left[V_i^1 V_i^2\right]$. Particular cases are:

- canonical correlation analysis (CCA) when there is no affine relation between the components of (Z^1, Z^2);
- factorial correspondence analysis (FCA) when the components of Z^1 and Z^2 are respectively the indicators of the exclusive modalities of two categorical variables;
- factorial discriminant analysis (FDA) when there is no affine relation between the components of Z^1 and the components of Z^2 are the indicators of the exclusive modalities of a categorical variable.

Canonical components are collinear with principal components of a PCA of the multidimensional linear regression function of one vector with respect to the other or projected PCA. The stochastic approximation of this regression function and that of the principal components of its PCA can be implemented in parallel (Monnez 2008, 2010, Bar 2013). In these references are defined stochastic approximation processes with one observation per step for estimating a linear regression function and processes of the Krasulina type for estimating principal components. Here we use processes using at each step all observations up to this step (A) or a mini-batch of current observations (B) for linear regression, and extended Oja processes for principal components. Previous experiments conducted in the application to streaming linear regression or streaming PCA led to the conclusion that type A processes generally converge faster than type B processes

(Duarte, Monnez and Albuisson 2018, Monnez and Skiredj 2021). Moreover we study the extension to generalized canonical correlation analysis and multiple correspondence analysis.
Section 3 (S3) is devoted to the estimation of a regression function. We first give two almost sure convergence theorems of stochastic approximation processes to the solution of a linear system $FX = G$, F and G being respectively unknown (p,p) and (p,q) matrices (Theorems 1,2). We apply them to the estimation of the multidimensional linear regression function $\hat{E}[S|R]$ of a random vector S with respect to a random vector R, in the case where there is no affine relation between the components of R (Corollaries 1,2) and the case where the components of R are the indicators of the exclusive modalities of a categorical variable (Corollaries 3,4). Corollaries 1,3 apply when at each step of the process all observations of (R,S) up to this step are used, Corollaries 2,4 when only the current observations are used. In S4, we present algorithms for estimating online canonical components of a CCA (Theorem 3) and general components of a generalized canonical correlation analysis (gCCA) (Theorem 4). We present in S5 algorithms for estimating online canonical components of a FCA (Theorem 5) and general components of a multiple correspondence analysis (MCA) (Theorem 6), and in S6 discriminant components of a FDA (Theorem 7). The proofs of Theorems 1,2 are in S7.

2. An extension of the Oja process

Let B be a (p,p) Q-symmetric matrix, for $i \in \{1,\dots,r\}, r \leq p$, λ_i be its *ith* largest eigenvalue and v_i a Q-normed eigenvector corresponding to λ_i, B and Q being unknown. The Oja process (Oja and Karhunen 1985) for estimating online eigenvectors of B is restricted to the case where the metric Q is known and B is the expectation of a random matrix of which independent observations are used in the construction of the process.

We suppose here that there exist two sequences of observable random matrices $(V_n, n \geq 1)$ and $(Q_n, n \geq 1)$ converging almost surely respectively to B and Q. For a random metric Q_n in $\mathbb{R}^p$, let $\langle .,. \rangle_n$ be the inner product and $\|.\|_n$ the norm induced by Q_n. Let (a_n) be a sequence of positive numbers. 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 $n \geq 1$:

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

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 $(\tilde{Y}_{n+1}^1, \dots, \tilde{Y}_{n+1}^r)$.
Assume, using the same numbering of assumptions as in (Monnez 2022a):

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

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

(H2b) $V_n \longrightarrow B, \sum_1^\infty a_n \|V_n - B\| < \infty$ a.s.

(H3a) $a_n > 0, \sum_1^\infty a_n = \infty, \sum_1^\infty a_n^2 < \infty$.

(H4) $Q_n \longrightarrow Q, \sum_1^\infty a_n \|Q_n - Q\| < \infty$ 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).

Applying Corollary 4 of (Monnez 2022a) yields:

Lemma 1. *Under H1a,b on B, H2a,b on* V_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* $-v_i$*,* $\langle V_n X_n^i, X_n^i \rangle_n$ *to* λ_i*,* $\sum_1^\infty a_n \left| \langle V_n X_n^i, X_n^i \rangle_n - \lambda_i \right| < \infty$ *a.s.*

If a convergent recursive estimator (V_n) of B can be constructed, this extension of the Oja process can be used in parallel to estimate online eigenvectors of B.

3. Stochastic approximation for streaming multidimensional linear regression

3.1 Stochastic approximation of the solution of a linear system

Let F be an unknown positive definite symmetric (p, p) matrix with smallest eigenvalue λ and G an unknown (p, q) matrix. Let $A = F^{-1}G$ be the solution of the linear system $FX = G$, with X a (p, q) matrix. Let $(F_n, n \geq 1)$ and $(G_n, n \geq 1)$ be sequences of (p, p) and (p, q) random matrices respectively, $(a_n, n \geq 1)$ be a sequence of positive numbers. Consider the stochastic approximation process $(A_n, n \geq 1)$ in $\mathbb{R}^{p \times q}$ and assume:

$$A_{n+1} = A_n - a_n(F_n A_n - G_n).$$

(A1) $F_n \longrightarrow F$ a.s. (A2) $\sum_1^\infty a_n \|F_n - F\| < \infty, \sum_1^\infty a_n \|G_n - G\| < \infty$ a.s.

(A3a) $a_n > 0, a_n = o(1), \sum_1^\infty a_n = \infty$.

Theorem 1. *Under A1, A2, A3a,* $A_n \longrightarrow A, \sum_1^\infty a_n \|A_n - A\| < \infty$ *a.s.*

Theorem 1 complements Theorem 9 of (Duarte, Monnez and Albuisson 2018).

Let T_n be the σ-field generated by $A_1, F_1, G_1, \dots, F_{n-1}, G_{n-1}$. Assume:

(B1) $E[F_n | T_n] = F, E[G_n | T_n] = G$ a.s.

(B2) $\exists f, g > 0$: $sup_n\, E[\|F_n - F\|^2 | T_n] < f, sup_n\, E[\|G_n - G\|^2 | T_n] < g$ a.s.

(H3b1) $a_n > 0, \sum_1^\infty a_n = \infty, \sum_1^\infty a_n^{3/2} < \infty, \frac{a_n}{a_{n+1}} \leq 1 + \gamma a_n + \gamma_n, 0 < \gamma < 2\lambda,\ \gamma_n \geq 0,$

$\sum_1^\infty \gamma_n < \infty$.

Theorem 2. *Let* λ *be the smallest eigenvalue of* F*. Under B1, B2, H3a,* $A_n \longrightarrow A$ *a.s. and in q.m.,* $\sum_1^\infty a_n\, E[\|A_n - A\|^2] < \infty$*; H3b1 replacing H3a, there exists* $b > 0$ *such that* $E[\|A_n - A\|^2] \leq b a_n$*,* $\sum_1^\infty a_n \|A_n - A\| < \infty$ *a.s.*

For $a_n = \frac{a}{n^\alpha}$, H3b1 is verified from a certain rank n_0 for $\frac{2}{3} < \alpha < 1$ and if $a > \frac{1}{2\lambda}$ for $\alpha = 1$.

Theorem 2 complements Theorem 2 of (Monnez 2008) for the choice of (a_n).

3.2 Application to streaming least square multidimensional linear regression

Let R and S be two random vectors in $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively, defined on the same probability space. Assume:

(A4a) The *4th moments of* (R,S) *exist.* (A4b) $\|R\|$ *and* $\|S\|$ *are a.s. bounded.*

(A4c) *There is no affine relation between the components of* R.

3.2.1 Case where assumption A4c holds

1) Let $\|.\|$ be the Euclidean norm in $\mathbb{R}^q$. The least square multidimensional linear regression of S with respect to R consists in determining the (p,q) matrix A and the $(q,1)$ matrix D minimizing $E[\|S-A^TR-D\|^2]$. Let the covariance matrices

$$F = Covar\,[R] = E[(R-E[R])(R-E[R])^T],$$
$$G = Covar[R,S] = E[(R-E[R])(S-E[S])^T].$$

Under A4c, the matrix F is positive definite and:

$$A = F^{-1}G\,,\, D = E[S] - A^TE[R].$$

The multidimensional linear regression function of S with respect to R is

$$\hat{E}[S|R] = A^TR + D = Covar[S,R](Covar\,[R])^{-1}(R-E[R]) + E[S].$$

Let $R_1 = \begin{pmatrix} R \\ 1 \end{pmatrix}$ in $\mathbb{R}^{p+1}$, $A_1 = \begin{pmatrix} A \\ D^T \end{pmatrix}$, $F_1 = E[R_1(R_1)^T]$, $G_1 = E[\,R_1S^T\,]$: $F_1A_1 = G_1$,

$$\widehat{E}[S|R] = (A_1)^TR_1 = E[\,S(R_1)^T\,](E[R_1(R_1)^T])^{-1}R_1.$$

2) Let $\left(\left(R_{ij},S_{ij}\right), i \geq 1, j \in \{1,\dots,m_i\}\right)$ be an i.i.d. sample of (R,S), $\mu_n = \sum_1^n m_i$ and:

$$R_{1i} = \begin{pmatrix} R_{ij} \\ 1 \end{pmatrix}, F_{11} = \frac{1}{\mu_n}\sum_{i=1}^n\sum_{j=1}^{m_i} R_{1ij}\left(R_{1ij}\right)^T, G_{11n} = \frac{1}{\mu_n}\sum_{i=1}^n\sum_{j=1}^{m_i} R_{1i}\,\left(S_{ij}\right)^T,$$

$$F_{12n} = \frac{1}{m_n}\sum_{j=1}^{m_n} R_{1nj}\left(R_{1nj}\right)^T, G_{12n} = \frac{1}{m_n}\sum_{j=1}^{m_n} R_{1n}\,\left(S_{nj}\right)^T.$$

For $h \in \{1,2\}$, let the stochastic approximation process $(A_{1hn}, n \geq 1)$ of A_1 in $\mathbb{R}^{(p+1)\times q}$:

$$A_{1h,n+1} = A_{1h} - a_n(F_{1hn}A_{1h} - G_{1hn});\; A_{1hn} = \begin{pmatrix} A_{hn} \\ (D_{hn})^T \end{pmatrix}, A_{hn}(p,q), D_{hn}(q,1).$$

A_{hn} is an estimator of A, D_{hn} of D. Under A4a,c:

$$F_{11n} \longrightarrow F_1, G_{11n} \longrightarrow G_1, E[\|F_{11n}-F_1\|] = O\left(\frac{1}{\sqrt{n}}\right), E[\|G_{11n}-G_1\|] = O\left(\frac{1}{\sqrt{n}}\right).$$

Assume: (A3b) $a_n > 0, a_n = o(1), \sum_1^\infty a_n = \infty\,, \sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$.

Under A3b and A4a,c, the assumptions of Theorem 1 hold for the process (A_{11n}), thus:

Corollary 1. *Under A3b and A4a,c,* $A_{11n} \longrightarrow A_1$, $\sum_1^\infty a_n\|A_{11} - A_1\| < \infty$, $A_{1n} \longrightarrow A$, $\sum_1^\infty a_n\|A_{1n} - A\| < \infty$ *a.s.*

Under H3b1 and A4b,c, the assumptions of Theorem 2 hold for the process (A_{12n}), thus:

Corollary 2. *Let λ be the smallest eigenvalue of $E\left[R_1 {R_1}^T\right]$. Under H3b1 and A4b,c, $A_{12n} \longrightarrow A_1$ a.s. and in q.m., $\sum_1^\infty a_n \|A_{12n} - A_1\| < \infty$ a.s., $A_{2n} \longrightarrow A$ a.s. and in q.m., $\sum_1^\infty a_n \|A_{2n} - A\| < \infty$ a.s.*

3.2.2 Case where the components $R^1, \ldots, R^p$ of R are the indicators of the exclusive modalities of a categorical variable

In this case, the multidimensional linear regression function $\hat{E}[S|R]$ of S with respect to R is equal to the conditional expectation

$$E[S|R] = \sum_{i=1}^p \frac{E[SR^i]}{E[R^i]} R^i = E[SR^T](E[RR^T])^{-1}R = A^T R,\ A = (E[RR^T])^{-1}E[RS^T].$$

Let $\left((R_{ij}, S_{ij}), i \geq 1, j \in \{1, \ldots, m_i\}\right)$ be an i.i.d. sample of (R, S), $\mu_n = \sum_1^n m_i$ and:

$$F_{1n} = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} R_{ij}\left(R_{ij}\right)^T, G_{1n} = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} R_{ij}\left(S_{ij}\right)^T,$$

$$F_{2n} = \frac{1}{m_n}\sum_{j=1}^{m_n} R_{nj}\left(R_{nj}\right)^T, G_{2n} = \frac{1}{m_n}\sum_{j=1}^{m_n} R_{nj}\left(S_{nj}\right)^T.$$

For $h \in \{1,2\}$, let the stochastic approximation process $(A_{hn}, n \geq 1)$ of A in $\mathbb{R}^{p\times q}$:

$$A_{h,n+1} = A_{hn} - a_n(F_{hn}A_{hn} - G_{hn}).$$

Corollary 3. *Under A3b and A4a, $A_{1n} \longrightarrow A$, $\sum_1^\infty a_n \|A_{1n} - A\| < \infty$ a.s.*

Corollary 4. *Let λ be the smallest eigenvalue of $E[RR^T]$. Under H3b1 and A4b, $A_{2n} \longrightarrow A$ a.s. and in q.m., $\sum_1^\infty a_n \|A_{2n} - A\| < \infty$ a.s.*

4. Stochastic approximation for streaming canonical correlation analysis

In this section, the numbering of assumptions is the same as in (Monnez 2022a).

Let Z be a random vector in $\mathbb{R}^p$ of covariance matrix C and M a metric in $\mathbb{R}^p$ possibly depending on characteristics of Z. In the PCA of Z in $(\mathbb{R}^p, M)$ (Monnez 2022a,b) is determined at step i a linear combination $c_i^T(Z - E[Z])$ of the centered components of Z, called i^{th} principal component,

- verifying $c_i^T M^{-1} c_i = 1$ (1),
- uncorrelated with the previous components, $c_i^T C c_j = 0, j \in \{1, \ldots, i-1\}$ (2),
- verifying $c_i^T M^{-1} c_j = 0, j \in \{1, \ldots, i-1\}$ (2'),
- of maximum variance $c_i^T C c_i$ (3).

Using the method of Lagrange multipliers gives:

Lemma 2. *Under (1), (2 or 2'), (3), c_i is a M^{-1}-normed eigenvector of the M^{-1}-symmetric matrix MC, corresponding to its i^{th} largest eigenvalue $\lambda_i = c_i^T C c_i$.*

4.1 Streaming CCA

Let Z^1 and Z^2 be random vectors in $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$ respectively, defined on the same probability space $(\Omega, \mathcal{A}, P)$. Assume the components of Z^1 and Z^2 belong to $L^2(\Omega, \mathcal{A}, P)$ equipped with the inner product $\langle V, W \rangle = E[VW]$. Assume there is no affine relation between the components of $Z = \begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix}$. Let $E^k, k \in \{1,2\}$, be the subspace of $L^2(\Omega, \mathcal{A}, P)$ generated by the centered components of Z^k: $E^1 \cap E^2 = \{0\}$. Let for $k \neq l \in \{1,2\}$:

$$C^k = E\left[\left(Z^k - E[Z^k]\right)\left(Z^k - E[Z^k]\right)^T\right], C^{kl} = E\left[\left(Z^k - E[Z^k]\right)\left(Z^l - E[Z^l]\right)^T\right].$$

$$B^{12} = (C^1)^{-1}C^{12}, B^{21} = (C^2)^{-1}C^{21}.$$

Let $\rho(.,.)$ be a linear correlation coefficient. Canonical correlation analysis consists in determining at step $i \in \{1, \dots, r\}, r \leq min(p_1, p_2)$, for $k \in \{1,2\}$, an element of E^k, called i^{th} canonical component in Z^k, $V_i^k = \left(\eta_i^k\right)^T \left(Z^k - E[Z^k]\right)$,

- of variance 1: $E\left[\left(V_i^k\right)^2\right] = \left(\eta_i^k\right)^T C^k \eta_i^k = 1$ (4.1.1),
- uncorrelated with $V_j^k, j \in \{1, \dots, i-1\}$: $E\left[V_i^k V_j^k\right] = \left(\eta_i^k\right)^T C^k \eta_j^k = 0$ (4.1.2),
- maximizing $\rho\left(V_i^1, V_i^2\right) = E\left[V_i^1 V_i^2\right] = \cos\left(V_i^1, V_i^2\right)$ (4.1.3).

For $k \neq l \in \{1,2\}$, for fixed V_i^k of norm 1, V_i^l of norm 1 which maximizes the cosine $\cos\left(V_i^k, V_i^l\right)$ is collinear with the projection of V_i^k on E^l (see S3.2.1),

$$\hat{E}\left[V_i^k | Z^l\right] = Covar\left[V_i^k, Z^l\right]\left(C^l\right)^{-1}\left(Z^l - E[Z^l]\right) = \left(\eta_i^k\right)^T \hat{E}[Z^k - E[Z^k] | Z^l],$$

$$cos^2\left(V_i^k, V_i^l\right) = E\left[\left(\hat{E}\left[V_i^k | Z^l\right]\right)^2\right] = Var\left[\left(\eta_i^k\right)^T \hat{E}[Z^k | Z^l]\right].$$

Under (4.1.1), (4.1.2), (4.1.3), $\hat{E}\left[V_i^k | Z^l\right]$ *is the* i^{th} *principal component of the PCA of* $\hat{E}[Z^k | Z^l]$ *in* $\left(\mathbb{R}^{p_k}, \left(C^k\right)^{-1}\right)$*, of maximum variance* $cos^2\left(V_i^k, V_i^l\right) = \rho^2\left(V_i^1, V_i^2\right)$*, and* V_i^l *of variance* 1 *is equal to* $\frac{1}{\rho\left(V_i^1, V_i^2\right)} \hat{E}\left[V_i^k | Z^l\right]$. By S3.2.1 and Lemma 2:

$$\hat{E}\left[V_i^k | Z^l\right] = \left(\eta_i^k\right)^T \hat{E}[Z^k - E[Z^k] | Z^l] = \left(\eta_i^k\right)^T C^{kl}\left(C^l\right)^{-1}\left(Z^l - E[Z^l]\right),$$

$$\rho^2\left(V_i^1, V_i^2\right) = E\left[\left(\hat{E}\left[V_i^k | Z^l\right]\right)^2\right] = \left(\eta_i^k\right)^T C^{kl}\left(C^l\right)^{-1} C^{lk} \eta_i^k,$$

$$V_i^l = \frac{1}{\rho\left(V_i^1, V_i^2\right)} \hat{E}\left[V_i^k | Z^l\right] = \left(\eta_i^l\right)^T \left(Z^l - E[Z^l]\right) \Leftrightarrow \eta_i^l = \frac{1}{\rho\left(V_i^1, V_i^2\right)}\left(C^l\right)^{-1} C^{lk} \eta_i^k.$$

Proposition 1. *For* $i \in \{1, \dots, r\}$, $r \leq min(p_1, p_2)$, *for* $k \neq l \in \{1,2\}$, η_i^k *is a* C^k*-normed eigenvector of the* C^k*-symmetric matrix* $\left(C^k\right)^{-1} C^{kl}\left(C^l\right)^{-1} C^{lk} = B^{kl} B^{lk}$ *corresponding to its* i^{th} *largest eigenvalue* $\rho^2\left(V_i^1, V_i^2\right)$ *and* $\eta_i^l = \frac{1}{\rho\left(V_i^1, V_i^2\right)}\left(C^l\right)^{-1} C^{lk} \eta_i^k = \frac{1}{\rho\left(V_i^1, V_i^2\right)} B^{lk} \eta_i^k$.

Stochastic approximation of canonical components

Let $SEQ = \left(Z_{11}, \ldots, Z_{1m_1}, \ldots, Z_{n1}, \ldots, Z_{nm_n}, \ldots\right)$ be a sequence of i.i.d. observations of Z, the mini-batch $\left(Z_{n1}, \ldots, Z_{nm_n}\right)$ being introduced at step n of the processes. Let for $i \geq 1$, $j \in \{1, \ldots, m_i\}, k \in \{1,2\}, Z_{ij} = \begin{pmatrix} Z_{ij}^1 \\ Z_{ij}^2 \end{pmatrix}, Z_{ij}^k\ (p_k, 1),\ Z_1^k = \begin{pmatrix} Z^k \\ 1 \end{pmatrix}, Z_{1ij}^k = \begin{pmatrix} Z_{ij}^k \\ 1 \end{pmatrix}, \mu_n = \sum_1^n m_i, \lambda^k$ be the smallest eigenvalue of $E\left[Z_1^k \left(Z_1^k\right)^T\right], k \in \{1,2\}, \lambda = min(\lambda^1, \lambda^2)$. Assume:

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

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

(H3b1) $a_n > 0, \sum_1^\infty a_n = \infty, \sum_1^\infty a_n^{3/2} < \infty, \frac{a_n}{a_{n+1}} \leq 1 + \gamma a_n + \gamma_n, 0 < \gamma < 2\lambda,\ \gamma_n \geq 0,$ $\sum_1^\infty \gamma_n < \infty$.

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

H3b1 implies H3c. In order to estimate online the canonical components, we define parallel stochastic approximation processes of B^{kl} and of eigenvectors of $B^k = B^{kl}B^{lk}$.

For $k \neq l \in \{1,2\},\ h \in \{1,2\}$, we define the random matrices $F_{1h}^k\ , G_{1hn}^{kl}$, and the stochastic approximation processes $\left(B_{1hn}^{kl}\right)$ in $\mathbb{R}^{(p_k+1)\times p_l}$, such that:

$$F_{11n}^k = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} Z_{1ij}^k \left(Z_{1i}^k\right)^T, G_{11n}^{kl} = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} Z_{1ij}^k \left(Z_{ij}^l\right)^T,$$

$$F_{12n}^k = \frac{1}{m_n}\sum_{j=1}^{m_n} Z_{1nj}^k \left(Z_{1nj}^k\right)^T, G_{12n}^{kl} = \frac{1}{m_n}\sum_{j=1}^{m_n} Z_{1nj}^k \left(Z_{nj}^l\right)^T.$$

$$B_{1h,n+1}^{kl} = B_{1hn}^{kl} - a_n\left(F_{1hn}^k B_{1hn}^{kl} - G_{1hn}^{kl}\right), B_{1hn}^{kl} = \begin{pmatrix} B_{hn}^{kl} \\ \left(D_{hn}^{kl}\right)^T \end{pmatrix}, B_{hn}^{kl}\ (p_k, p_l), D_{hn}^{kl}(p_l, 1).$$

We define the convex combination B_n^{kl} in $\mathbb{R}^{p_k \times p_l}$,

$$B_n^{kl} = \omega_{1n} B_{1n}^{kl} + \omega_{2n} B_{2n}^{kl}, \omega_{1n} \geq 0, \omega_{2n} \geq 0, \omega_{1n} + \omega_{2n} = 1.$$

By Corollary 1, under H3c, H7a,c, $B_{1n}^{kl} \longrightarrow B^{kl}, \sum_1^\infty a_n \left\|B_{1n}^{kl} - B^{kl}\right\| < \infty$ a.s.

By Corollary 2, under H3b1, H7b,c, $B_{2n}^{kl} \longrightarrow B^{kl}, \sum_1^\infty a_n \left\|B_{2n}^{kl} - B^{kl}\right\| < \infty$ a.s.

Thus under H3b1, H7b,c, $B_n^{kl} \longrightarrow B^{kl}, \sum_1^\infty a_n \left\|B_n^{kl} - B^{kl}\right\| < \infty$ a.s. and:

$$B_n^k = B_n^{kl} B_n^{lk} \longrightarrow B^k = B^{kl} B^{lk}, \sum_1^\infty a_n \left\|B_n^k - B^k\right\| < \infty \text{ a.s.}$$

For $k \in \{1,2\}$, let $\bar{Z}_n^k$ be the mean of the sample $\left(Z_{11}^k, \ldots, Z_{nm_n}^k\right)$ of Z^k and C_n^k its covariance matrix, $C_n^k = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^k \left(Z_{ij}^k\right)^T - \bar{Z}_n^k \left(\bar{Z}_n^k\right)^T$, both recursively computed. The matrix C_n^k is positive definite from a certain rank. Let $\langle .,. \rangle_{k,n+1}$ be the inner product and $\|.\|_{k,n+1}$ the norm induced by C_n^k in $\mathbb{R}^{p_k}$. Under assumptions H3c, H7a,c:

$$C_n^k \longrightarrow C^k, E\left[\left\|C_n^k - C^k\right\|\right] = 0\left(\frac{1}{\sqrt{n}}\right), \sum_1^\infty a_n \left\|C_n^k - C^k\right\| < \infty \text{ a.s.}$$

Assumption H4 with $Q_{n+1} = C_n^k$ holds. For $i \in \{1, \ldots, r\}, r \leq min(p_1, p_2)$, we define the processes $\left(\tilde{X}_n^{ki}\right), \left(\tilde{Y}_n^{ki}\right), \left(X_n^{ki}\right)$ in $\mathbb{R}^{p_k}$ and apply Lemma 1 with $V_n = B_n^k, B = B^k$:

$\tilde{Y}_{n+1}^{ki} = \left(I + a_n B_n^k\right)\tilde{X}_n^{ki}$, $\tilde{X}_{n+1}^{ki} = \tilde{Y}_{n+1}^{ki} - \sum_{j<i}\langle \tilde{Y}_{n+1}^{ki}, X_{n+1}^{kj}\rangle_{k,n+1} X_{n+1}^{kj}$, $X_{n+1}^{ki} = \frac{\tilde{X}_{n+1}^{ki}}{\|\tilde{X}_{n+1}^{ki}\|_{k,n+1}}$.

Theorem 3. *Let* λ^k *be the smallest eigenvalue of* $E\left[Z_1^k\left(Z_1^k\right)^T\right]$, $k \in \{1,2\}$, $\lambda = min(\lambda^1, \lambda^2)$. *Under H1b, H2a, H3b1, H5, H7b,c, for* $k \neq l \in \{1,2\}$, *for* $i \in \{1, \dots, r\}$, X_n^{ki} *converges to* η_i^k *or* $-\eta_i^k$, $\langle B_n^k X_n^{ki}, X_n^{ki}\rangle_{kn}$ *to* $\rho^2\left(V_i^1, V_i^2\right)$, $\frac{B_n^{lk} X_n^{ki}}{\sqrt{\langle B_n^k X_n^{ki}, X_n^{ki}\rangle_{kn}}}$ *to* η_i^l *or* $-\eta_i^l$, $\sum_1^\infty a_n \left|\langle B_n^k X_n^{ki}, X_n^{ki}\rangle_{kn} - \rho^2\left(V_i^1, V_i^2\right)\right| < \infty$ *a.s. When* $\omega_{2n} = 0$ *for all* n, *the same conclusions hold with H3b1 replaced by H3c and H7b by H7a.*

4.2 Streaming gCCA

Consider the generalized canonical correlation analysis (gCCA) defined by Carroll (1968). Suppose the set of components of a random vector Z in $\mathbb{R}^p$ is partitioned in q subsets of real random variables $\{Z^{k1}, \dots, Z^{kp_k}\}$, $k \in \{1, \dots, q\}, \sum_{k=1}^q p_k = p$. Let Z^k be the random vector in $\mathbb{R}^{p_k}$ whose components are $Z^{k1}, \dots, Z^{kp_k}$, $C^k = Covar[Z^k]$ its covariance matrix, $C = Covar[Z]$ that of Z. Assume there is no affine relation between the components of Z (H7c). Thus the matrices $C^k, k \in \{1, \dots, q\}$, and C are invertible. Let for $k \in \{1, \dots, q\}$, $Z_1^k = \begin{pmatrix} Z^k \\ 1 \end{pmatrix}$, λ^k be the smallest eigenvalue of $E\left[Z_1^k\left(Z_1^k\right)^T\right]$, $\lambda = min(\lambda^1, \dots, \lambda^q)$ in Assumption H3b1.

In gCCA, for $i \in \{1, \dots, r\}, r \leq \min\left(p_1, \dots, p_q\right)$, are determined at step i:

- a linear combination of variance 1 of the centered components of Z, $U_i = \theta_i^{\mathrm{T}}(Z - E[Z])$, uncorrelated with $U_1, \dots, U_{i-1}$, called i^{th} general component,
- for $k \in \{1, \dots, q\}$, a linear combination of variance 1 of the centered components of Z^k, $V_i^k = \left(\eta_i^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right)$, called i^{th} canonical component in Z^k,
- maximizing $\sum_{k=1}^q \rho^2\left(U_i, V_i^k\right)$.

Let M be the block diagonal (p,p) matrix whose k^{th} diagonal block is $\left(C^k\right)^{-1}$.

Let ${\theta_i}^T = \left(\left(\theta_i^1\right)^T \dots \left(\theta_i^q\right)^T\right)$, $dim\left(\theta_i^k\right) = p_k$. It is shown (Monnez 2008, 2022b):

Proposition 2. *For* $i \in \{1, \dots, r\}$, $r \leq \min\left(p_1, \dots, p_q\right)$, θ_i *is a* C*-normed eigenvector of the matrix* $B = MC$ *corresponding to its* i^{th} *largest eigenvalue* $\nu_i = \sum_{k=1}^q \rho^2\left(U_i, V_i^k\right)$, U_i *is collinear with the* i^{th} *principal component* $c_i^{\mathrm{T}}(Z - E[Z])$ *of the PCA of* Z *in* $(\mathbb{R}^p, M)$, $c_i = \sqrt{\nu_i}\theta_i$; *for* $k \in \{1, \dots, q\}$, $\eta_i^k = \frac{\theta_i^k}{\sqrt{\left(\theta_i^k\right)^{\mathrm{T}} C^k \theta_i^k}}$, $U_i = \sum_{k=1}^q \sqrt{\left(\theta_i^k\right)^{\mathrm{T}} C^k \theta_i^k}\ V_i^k$.

The covariance matrix $C = Covar[Z]$ can be subdivided into q^2 (p_k, p_l) covariance matrices $C^{kl} = Covar[Z^k, Z^l]$, $k, l \in \{1, \dots, q\}$. Likewise, the matrix $B = MC$ can be subdivided into q^2 (p_k, p_l) matrices $B^{kl} = \left(C^k\right)^{-1} C^{kl}$ with $B^{kk} = I_k$, the identity matrix of order p_k.

Stochastic approximation of general components of gCCA

Consider the sequence SEQ (S4.1) of i.i.d. observations of Z, with $Z_{ij} = \left(\left(Z_{ij}^1\right)^{\mathrm{T}} \dots \left(Z_{ij}^q\right)^{\mathrm{T}}\right)^{\mathrm{T}}$, $Z_{ij}^k(p_k, 1), k \in \{1, \dots, q\}$. For $k \in \{1, \dots, q\}$, let $\bar{Z}_n^k$ be the mean of the sample $\left(Z_{11}^k, \dots, Z_{nm_n}^k\right)$ of Z^k, $C_n^k = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^k \left(Z_{ij}^k\right)^{\mathrm{T}} - \bar{Z}_n^k \left(\bar{Z}_n^k\right)^{\mathrm{T}}$ with $\mu_n = \sum_{i=1}^n m_i$ its covariance matrix, both recursively computed. Let Q_{n+1} be the block diagonal (p, p) matrix whose k^{th} diagonal block is $C_n^k, k \in \{1, \dots, q\}$. Q_n is positive definite symmetric from a certain rank. Under H3c and H7a:

$$Q_n \to M^{-1}, \sum_1^\infty a_n \|Q_n - M^{-1}\| < \infty \text{ a.s.}$$

Let $\langle ., . \rangle_{n+1}$ be the inner product and $\|.\|_{n+1}$ the norm induced by Q_{n+1} in $\mathbb{R}^p$.

We define stochastic approximation processes $\left(B_n^{kl}\right)$ of B^{kl}, $k, l \in \{1, \dots, q\}$, similar to those defined for $q = 2$ in the case of CCA (S4.1) and for $n \geq 1$, the (p, p) random matrix B_n subdivided into the q^2 (p_k, p_l) blocks B_n^{kl} . Under H3b1 and H7b,c:

$$B_n^{kl} \longrightarrow B^{kl}, \sum_1^\infty a_n \left\|B_n^{kl} - B^{kl}\right\| < \infty \text{ a.s., } B_n \to B, \sum_1^\infty a_n \|B_n - B\| < \infty \text{ a.s.}$$

For $i \in \{1, \dots, r\}$, $r \leq \min\left(p_1, \dots, p_q\right)$, we define the processes $\left(\tilde{X}_n^i\right), \left(\tilde{Y}_n^i\right), \left(X_n^i\right)$ in $\mathbb{R}^p$ and apply Lemma 1 with $V_n = B_n$:

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

Theorem 4. *Let λ^k be the smallest eigenvalue of $E\left[Z_1^k \left(Z_1^k\right)^T\right]$, $k \in \{1, \dots, q\}$, $\lambda = min(\lambda^1, \dots, \lambda^q)$. Under H1b, H2a, H3b1, H5, H7b,c, for $i \in \{1, \dots, r\}$, X_n^i converges to $c_i = \sqrt{\nu_i}\theta_i$ or $-c_i$, $\langle B_n X_n^i, X_n^i \rangle_n$ to $\nu_i, \sum_1^\infty a_n \left|\langle B_n X_n^i, X_n^i \rangle_n - \nu_i\right| < \infty$ a.s. When $\omega_{2n} = 0$ for all n, the same conclusions hold with H3b1 replaced by H3c and H7b by H7a.*

CCA as particular case of gCCA

For $q = 2$: $Z = \begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix}, B = MC = \begin{pmatrix} (C^1)^{-1} & 0 \\ 0 & (C^2)^{-1} \end{pmatrix} \begin{pmatrix} C^1 & C^{12} \\ C^{21} & C^2 \end{pmatrix} = \begin{pmatrix} I_{p_1} & B^{12} \\ B^{21} & I_{p_2} \end{pmatrix}$.

For CCA, under the conditions (4.1.1), (4.1.2), (4.1.3), with $\eta_i = \begin{pmatrix} \eta_i^1 \\ \eta_i^2 \end{pmatrix}$, we have:

$$(\eta_i)^T M^{-1} \eta_i = 2, \ (\eta_i)^T M^{-1} \eta_j = 0, j \in \{1, \dots, i-1\}, (\eta_i)^T C \eta_i = 2 + 2\cos\left(V_i^1, V_i^2\right).$$

We have to maximize $(\eta_i)^T C \eta_i$. By Lemma 2:

Proposition 3. *η_i, of squared M^{-1}-norm 2, is an eigenvector of $B = MC$ corresponding to its i^{th} largest eigenvalue $\nu_i = 1 + \rho\left(V_i^1, V_i^2\right)$.*

Thus $\frac{1}{\sqrt{2}}(\eta_i)^T (Z - E[Z])$ is the i$^{\text{th}}$ principal component of the PCA of Z in $(\mathbb{R}^{\mathrm{p}}, \mathrm{M})$ of variance ν_i, collinear with the i^{th} general component $\mathrm{U_i}$ of the gCCA of Z of variance 1,

$$\mathrm{U_i} = \frac{1}{\sqrt{2\nu_i}}(\eta_i)^T(Z - E[Z]) = \frac{1}{\sqrt{2\nu_i}}\left(\mathrm{V_i^1} + \mathrm{V_i^2}\right), \nu_i = 1 + \rho\left(V_i^1, V_i^2\right) = \sum_{k=1}^{2}\rho^2\left(U_i, V_i^k\right).$$

V_i^1 and V_i^2 are the first canonical components of the gCCA of Z.

We can use the stochastic approximation processes $\left(X_n^i\right)$ of $c_i = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_i^1\\ \eta_i^2\end{pmatrix}$ in $\mathbb{R}^p$, $i \in \{1, \dots, r\}$, defined for gCCA, and apply Theorem 4 for $q = 2$.

5. Stochastic approximation for streaming factorial correspondence analysis

5.1 Streaming FCA

In the case of FCA, Z^1 and Z^2 are random vectors in $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$ respectively, defined on the same probability space $(\Omega, \mathcal{A}, P)$, whose respective components $Z^{11}, \dots, Z^{1p_1}$ and $Z^{21}, \dots, Z^{2p_2}$ are the indicators of the exclusive modalities of two categorical variables. We have $\left(u^k\right)^T Z^k = 1$, $k \in \{1,2\}$, u^k being the vector in $\mathbb{R}^{p_k}$ whose all components are equal to 1; a linear combination of the components of $Z^k - E[Z^k]$ is a linear combination of the components of Z^k orthogonal to 1. Let E^k be the subspace of $L^2(\Omega, \mathcal{A}, P)$ generated by the components of $Z^k, k \in \{1,2\}$; $1 \in E^k$. Let for $k \neq l \in \{1,2\}$:

$$C^k = E\left[Z^k\left(Z^k\right)^T\right], C^k(i,i) = P\left(Z^{ki} = 1\right), C^{kl} = E\left[Z^k\left(Z^l\right)^T\right],\ B^{kl} = \left(C^k\right)^{-1}C^{kl},$$

$$C^{kl}(i,j) = P\left(Z^{ki} = 1, Z^{lj} = 1\right), B^{kl}(i,j) = P\left(Z^{lj} = 1 | Z^{ki} = 1\right).$$

FCA consists in determining at step $i \in \{1, \dots, r\}, r \leq min(p_1 - 1, p_2 - 1)$, for $k \in \{1,2\}$, an element of E^k, called i^{th} canonical component in Z^k, $V_i^k = \left(\eta_i^k\right)^T Z^k$,

- centered: $\left(\eta_i^k\right)^T E[Z^k] = \left(\eta_i^k\right)^T C^k u^k = 0$ (a),
- of variance 1: $E\left[\left(V_i^k\right)^2\right] = \left(\eta_i^k\right)^T C^k \eta_i^k = 1$ (5.1.1),
- uncorrelated with $V_j^k, j \in \{1, \dots, i-1\}$: $E\left[V_i^k V_j^k\right] = \left(\eta_i^k\right)^T C^k \eta_j^k = 0$ (5.1.2),
- maximizing $\rho\left(V_i^1, V_i^2\right) = E\left[V_i^1 V_i^2\right] = \cos\left(V_i^1, V_i^2\right)$ (5.1.3).

A trivial solution to (5.1.3) is $V_0^k = \left(u^k\right)^T Z^k = 1$, $k \in \{1,2\}$, $E[V_0^1 V_0^2] = 1$.

For $k \neq l \in \{1,2\}$, for fixed V_i^k of norm 1, V_i^l of norm 1 which maximizes $\cos\left(V_i^k, V_i^l\right)$ is collinear with $\hat{E}\left[V_i^k | Z^l\right] = E\left[V_i^k | Z^l\right] = \left(\eta_i^k\right)^T E[Z^k | Z^l]$ which is centered, and:

$$cos^2\left(V_i^k, V_i^l\right) = E\left[\left(E\left[V_i^k | Z^l\right]\right)^2\right] = Var\left[E\left[V_i^k | Z^l\right]\right] = r_c^2\left(V_i^k | Z^l\right),$$

the squared correlation ratio of V_i^k with respect to Z^l.

Under (a), (5.1.1), (5.1.2), (5.1.3), $E\left[V_i^k|Z^l\right]$ *is the* i^{th} *principal component of the PCA of* $E[Z^k|Z^l]$ *in* $\left(\mathbb{R}^{p_k},\left(C^k\right)^{-1}\right)$*, of maximum variance* $r_c^2\left(V_i^k|Z^l\right)$*, and* V_i^l *of variance* 1 *is equal to* $\frac{1}{r_c\left(V_i^k|Z^l\right)}E\left[V_i^k|Z^l\right]$. By S3.2.2 and Lemma 2:

$$E\left[V_i^k|Z^l\right] = \left(\eta_i^k\right)^T C^{kl}\left(C^l\right)^{-1}Z^l,$$

$$r_c^2\left(V_i^k|Z^l\right) = E\left[\left(E\left[V_i^k|Z^l\right]\right)^2\right] = \left(\eta_i^k\right)^T C^{kl}\left(C^l\right)^{-1}C^{lk}\eta_i^k.$$

η_i^k is an eigenvector of the C^k-symmetric matrix $\left(C^k\right)^{-1}C^{kl}\left(C^l\right)^{-1}C^{lk} = B^{kl}B^{lk}$, as u^k (since $B^{lk}u^k = u^l$) that corresponds to the largest eigenvalue 1 to be removed. The other eigenvectors are C^k-orthogonal to u^k, thus condition (a) holds.

Proposition 4. *For* $i \in \{1,\dots,r\}$, $r \le min(p_1 - 1, p_2 - 1)$*, for* $k \ne l \in \{1,2\}$, η_i^k *is a* C^k*-normed eigenvector of the* C^k*-symmetric matrix* $\left(C^k\right)^{-1}C^{kl}\left(C^l\right)^{-1}C^{lk} = B^{kl}B^{lk}$ *corresponding to its* i^{th} *largest eigenvalue* $r_c^2\left(V_i^k|Z^l\right)$ *after removing the eigenvalue 1 and* $\eta_i^l = \frac{1}{r_c\left(V_i^k|Z^l\right)}\left(C^l\right)^{-1}C^{lk}\eta_i^k = \frac{1}{r_c\left(V_i^k|Z^l\right)}B^{lk}\,\eta_i^k$.

Stochastic approximation of canonical components of FCA

Consider the sequence SEQ (S4.1) of i.i.d. observations of $Z = \begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix}$ with $Z_{ij} = \begin{pmatrix} Z_{ij}^1 \\ Z_{ij}^2 \end{pmatrix}$, $Z_{ij}^k(p_k, 1)$, $k \in \{1,2\}$. Let $\mu_n = \sum_{i=1}^n m_i$. Let λ^k be the smallest eigenvalue of C^k, $k \in \{1,2\}$, $\lambda = min(\lambda^1, \lambda^2)$ in Assumption H3b1. For $k \ne l \in \{1,2\}$, we define:

$$F_{1n}^k = \frac{1}{\mu_n}\sum_{i=1}^n\sum_{j=1}^{m_i} Z_{ij}^k\left(Z_{ij}^k\right)^T, G_{1n}^{kl} = \frac{1}{\mu_n}\sum_{i=1}^n\sum_{j=1}^{m_i} Z_{ij}^k\left(Z_{ij}^l\right)^T,$$

$$F_{2n}^k = \frac{1}{m_n}\sum_{j=1}^{m_n} Z_{nj}^k\left(Z_{nj}^k\right)^T, G_{2n}^{kl} = \frac{1}{m_n}\sum_{j=1}^{m_n} Z_{nj}^k\left(Z_{nj}^l\right)^T,$$

and the stochastic approximation processes $\left(B_{hn}^{kl}\right)$, $h \in \{1,2\}$, and $\left(B_n^{kl}\right)$ in $\mathbb{R}^{p_k \times p_l}$:

$$B_{h,n+1}^{kl} = B_{hn}^{kl} - a_n\left(F_{hn}^k B_{hn}^{kl} - G_{hn}^{kl}\right),$$

$$B_n^{kl} = \omega_{1n}B_{1n}^{kl} + \omega_{2n}B_{2n}^{kl}, \omega_{1n} \ge 0, \omega_{2n} \ge 0, \omega_{1n} + \omega_{2n} = 1.$$

By Corollaries 3, 4, under H3b1, $B_n^{kl} \longrightarrow B^{kl}$, $\sum_1^\infty a_n\left\|B_n^{kl} - B^{kl}\right\| < \infty$ a.s. and:

$$B_n^k = B_n^{kl}B_n^{lk} \longrightarrow B^k = B^{kl}B^{lk}, \sum_1^\infty a_n\left\|B_n^k - B^k\right\| < \infty \text{ a.s.}$$

When $\omega_{2n} = 0$ for all n, H3b1 can be replaced by H3c.

Let $\langle .,. \rangle_{k,n+1}$ be the inner product and $\|.\|_{k,n+1}$ the norm induced by F_{1n}^k in $\mathbb{R}^{p_k}$. Under H3c: $F_{1n}^k \longrightarrow C^k, \sum_1^\infty a_n\left\|F_{1n}^k - C^k\right\| < \infty$ a.s. For $k \in \{1,2\}$, $i \in \{1,\dots,r+1\}$, we define the processes $\left(\tilde{X}_n^{ki}\right), \left(\tilde{Y}_n^{ki}\right), \left(X_n^{ki}\right)$ in $\mathbb{R}^{p_k}$ and apply Lemma 1 with $V_n = B_n^k, B = B^k$:

$\tilde{Y}_{n+1}^{ki} = \left(I + a_n B_n^k\right)\tilde{X}_n^{ki}$, $\tilde{X}_{n+1}^{ki} = \tilde{Y}_{n+1}^{ki} - \sum_{j<i}\langle\tilde{Y}_{n+1}^{ki}, X_{n+1}^{kj}\rangle_{k,n+1}X_{n+1}^{kj}$, $X_{n+1}^{ki} = \frac{\tilde{X}_{n+1}^{ki}}{\left\|\tilde{X}_{n+1}^{ki}\right\|_{k,n+1}}$.

Theorem 5. *Let λ^k be the smallest eigenvalue of $C^k, k \in \{1,2\}$, $\lambda = min(\lambda^1, \lambda^2)$. Under H1b, H2a, H3b1, H5, for $k \neq l \in \{1,2\}$, for $i \in \{1, \dots, r\}$, $X_n^{k,i+1}$ converges to η_i^k or $-\eta_i^k$, $\langle B_n^k X_n^{k,i+1}, X_n^{k,i+1} \rangle_{kn}$ converges to $r_c^2(V_i^k|Z^l)$, $\frac{B_n^{lk} X_n^{k,i+1}}{\sqrt{\langle B_n^k X_n^{ki}, X_n^{ki} \rangle_{kn}}}$ to η_i^l or $-\eta_i^l$, $\sum_1^\infty a_n \left| \langle B_n^k X_n^{k,i+1}, X_n^{k,i+1} \rangle_{kn} - r_c^2(V_i^k|Z^l) \right| < \infty$ a.s. When $\omega_{2n} = 0$ for all n, the same conclusions hold with H3b1 replaced by H3c.*

5.2 Streaming MCA

Multiple correspondence analysis (MCA) can be considered as a particular case of gCCA of a random vector $Z = ((Z^1)^T, \dots, (Z^q)^T)^T$ (Subsection 4.2) with, for $k \in \{1, \dots, q\}$:

- the p_k components of Z^k are the indicators of the exclusive modalities of a categorical variable, $(Z^k)^T u^k = 1$, $Z^T u = q$, all components of u^k in $\mathbb{R}^{p_k}$ and u in $\mathbb{R}^p$ being 1;
- $C^k = E\left[Z^k (Z^k)^T\right]$, $C = E[ZZ^T]$, $C^{kl} = E\left[Z^k (Z^l)^T\right]$, λ^k the smallest eigenvalue of $E\left[Z^k (Z^k)^T\right]$, M the block diagonal matrix whose k^{th} diagonal block is $(C^k)^{-1}$;
- the centered i^{th} general component of variance 1 $U_i = (\theta_i)^T Z$, $(\theta_i)^T C u = 0$, and the centered canonical component in Z^k, $V_i^k = (\eta_i^k)^T Z^k$, $(\eta_i^k)^T C^k u^k = 0$, are such that:

$$\sum_{k=1}^q \rho^2(U_i, V_i^k) \max \Leftrightarrow V_i^k \text{ collinear with } E[U_i|Z^k], k \in \{1, \dots, q\}$$
$$\Leftrightarrow \sum_{k=1}^q E\left[\left(E[U_i|Z^k]\right)^2\right] \max \Leftrightarrow \sum_{k=1}^q r_c^2(U_i|Z^k) \; max;$$

- applying Proposition 2, θ_i is a C-normed eigenvector of the matrix $B = MC$; since $B^{kl} u^l = u^k$, the largest eigenvalue of $B = MC$ is q, corresponding to the eigenvector u and to the trivial solution $U_0 = u^T Z = 1$, $V_0^k = (u^k)^T Z^k = 1$, $k \in \{1, \dots, q\}$; the eigenvectors θ_i are M^{-1}-orthogonal to u, thus $(\theta_i)^T C u = 0$, and:

Proposition 5. *For $i \in \{1, \dots, r\}$, $r \leq \min(p_1 - 1, \dots, p_q - 1)$, θ_i is a C-normed eigenvector of the matrix $B = MC$ corresponding to its i^{th} largest eigenvalue $\nu_i = \sum_{k=1}^q \rho^2(U_i, V_i^k)$ after removing the eigenvalue q, U_i is collinear with the i^{th} principal component $c_i^T(Z - E[Z])$ of the PCA of Z in $(\mathbb{R}^p, M)$, $c_i = \sqrt{\nu_i}\theta_i$; for $k \in \{1, \dots, q\}$, $\eta_i^k = \frac{\theta_i^k}{\sqrt{(\theta_i^k)^T C^k \theta_i^k}}$,*

$U_i = \sum_{k=1}^q \sqrt{(\theta_i^k)^T C^k \theta_i^k}\; V_i^k$.

Stochastic approximation of general components of MCA

Defining the processes (B_n^{kl}), $k, l \in \{1, \dots, q\}$, as in S5.1 for FCA, the processes (X_n^i) in $\mathbb{R}^p$ as in S4.2 for gCCA, we have by Theorem 4:

Theorem 6. *Let* $\lambda = min(\lambda^1, \dots, \lambda^q)$. *Under H1b, H2a, H3b1, H5, for* $i \in \{1, \dots, r\}$, X_n^{i+1} *converges to* $c_i = \sqrt{\nu_i}\theta_i$ *or* $-c_i$, $\langle B_n X_n^{i+1}, X_n^{i+1}\rangle_n$ *to* $\nu_i, \sum_1^\infty a_n \left|\langle B_n X_n^{i+1}, X_n^{i+1}\rangle_n - \nu_i\right| < \infty$ *a.s. When* $\omega_{2n} = 0$ *for all* n, *the same conclusions hold with H3b1 replaced by H3c.*

FCA as particular case of MCA

For $q = 2$: $Z = \begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix}, B = MC = \begin{pmatrix} (C^1)^{-1} & 0 \\ 0 & (C^2)^{-1} \end{pmatrix}\begin{pmatrix} C^1 & C^{12} \\ C^{21} & C^2 \end{pmatrix} = \begin{pmatrix} I_{p_1} & B^{12} \\ B^{21} & I_{p_2} \end{pmatrix}$.

For FCA, under the conditions (5.1.1), (5.1.2), (5.1.3), with $\eta_i = \begin{pmatrix} \eta_i^1 \\ \eta_i^2 \end{pmatrix}$, using the same argumentation as that in the case of CCA particular case of gCCA (S4.2) gives:

Proposition 6. *For* $i \in \{1, \dots, r\}$, $r \leq min(p_1 - 1, p_2 - 1)$, η_i, *of squared* M^{-1}*-norm 2, is an eigenvector of* $B = MC$ *corresponding to its* i^{th} *largest eigenvalue* $\nu_i = 1 + r_c\left(V_i^k | Z^l\right)$ *after removing the eigenvalue 2.*

The i^{th} general component $\mathrm{U_i}$ of variance 1 of the MCA of Z is

$$\mathrm{U_i} = \frac{1}{\sqrt{2\nu_i}}(\eta_i)^T(Z - E[Z]) = \frac{1}{\sqrt{2\nu_i}}\left(\mathrm{V_i^1} + \mathrm{V_i^2}\right), \nu_i = 1 + r_c\left(V_i^k | Z^l\right) = \sum_{k=1}^2 r_c^2\left(\mathrm{U_i} | Z^k\right).$$

V_i^1 and V_i^2 are the first canonical components of the MCA of Z.

We can use the stochastic approximation processes $\left(X_n^i\right)$ of $c_i = \frac{1}{\sqrt{2}}\begin{pmatrix} \eta_i^1 \\ \eta_i^2 \end{pmatrix}$ in $\mathbb{R}^p$ defined for MCA, and apply Theorem 6 for $q = 2$.

6. Stochastic approximation for streaming factorial discriminant analysis

Let Z^1 and Z^2 be random vectors in $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$ respectively, defined on the same probability space $(\Omega, \mathcal{A}, P)$. Assume there is no affine relation between the components of Z^1 and that the components $Z^{21}, \dots, Z^{2p_2}$ of Z^2 are the indicators of the exclusive modalities of a categorical variable. Let u^2 be the vector in $\mathbb{R}^{p_2}$ whose all components are equal to 1; $(u^2)^T Z^2 = 1$; a linear combination of the components of $Z^2 - E[Z^2]$ is a linear combination of the components of Z^2 orthogonal to 1. Let the matrices

$$C^1 = E[(Z^1 - E[Z^1])(Z^1 - E[Z^1])^T], C^2 = E[Z^2(Z^2)^T],$$
$$C^{12} = E[(Z^1 - E[Z^1])(Z^2)^T], C^{21} = (C^{12})^T.$$

Factorial discriminant analysis consists in determining at step $i \in \{1, \dots, r\}, r \leq min(p_1, p_2 - 1)$, a linear combination of the centered components of Z^1, called i^{th} discriminant component, $V_i^1 = \left(\eta_i^1\right)^T(Z^1 - E[Z^1])$,

- of variance 1: $E\left[\left(V_i^1\right)^2\right] = \left(\eta_i^1\right)^T C^1 \eta_i^1 = 1$ (6.1.1),
- uncorrelated with $V_j^1, j \in \{1, \dots, i-1\}$: $E\left[V_i^1 V_j^1\right] = \left(\eta_i^1\right)^T C^1 \eta_j^1 = 0$ (6.1.2),

- maximizing the squared correlation ratio of V_i^1 with respect to Z^2,

$$r_c^2(V_i^1|Z^2) = E\left[\left(E\left[V_i^1|Z^2\right]\right)^2\right] = Var\left[\left(\eta_i^1\right)^T E[Z^1|Z^2]\right] \quad (6.1.3).$$

We use a classical presentation of FDA. This analysis can also be presented, as CCA and FCA, as a canonical analysis consisting in determining at step i a couple $\left(V_i^1, V_i^2\right)$, with $V_i^2 = \left(\eta_i^2\right)^T Z^2$, maximizing $cos\left(V_i^1, V_i^2\right)$, but we are only interested in determining V_i^1.

Under (6.1.1), (6.1.2), (6.1.3), $E\left[V_i^1|Z^2\right]$ *is the* i^{th} *principal component of the PCA of* $E[Z^1|Z^2]$ *in* $(\mathbb{R}^{p_1}, (C^1)^{-1})$. By S3.2.2 and Lemma 2:

$$E\left[V_i^1|Z^2\right] = E\left[\left(\eta_i^1\right)^T E[Z^1 - E[Z^1]|Z^2]\right] = \left(\eta_i^1\right)^T C^{12}(C^2)^{-1}Z^2,$$

$$r_c^2(V_i^1|Z^2) = E\left[\left(E\left[V_i^1|Z^2\right]\right)^2\right] = \left(\eta_i^1\right)^T C^{12}(C^2)^{-1}C^{21}\eta_i^1,$$

η_i^1 is a C^1-normed eigenvector of the C^1-symmetric matrix $(C^1)^{-1}C^{12}(C^2)^{-1}C^{21}$ corresponding to its i^{th} largest eigenvalue $r_c^2\left(V_i^1|Z^2\right)$. But:

$$C^{12}(E[Z^2(Z^2)^T])^{-1}E[Z^2] = C^{12}u^2 = E\left[Z^1 - E[Z^1]\right] = 0$$

$$\Longrightarrow (C^1)^{-1}C^{12}(C^2)^{-1}C^{21} = (C^1)^{-1}C^{12}(C^2)^{-1}\,E[Z^2(Z^1)^T] = B^{12}B^{21},$$

$$B^{12} = (C^1)^{-1}C^{12}, B^{21} = (C^2)^{-1}\,E[Z^2(Z^1)^T].$$

Proposition 7. *For* $i \in \{1, \dots, r\}$, $r \leq min(p_1, p_2 - 1)$, η_i^1 *is a* C^1*-normed eigenvector of the* C^1*-symmetric matrix* $(C^1)^{-1}C^{12}(C^2)^{-1}\,E[Z^2(Z^1)^T] = B^{12}B^{21}$ *corresponding to its* i^{th} *largest eigenvalue* $r_c^2\left(V_i^1|Z^2\right)$.

Stochastic approximation of discriminant components of FDA

Consider the sequence SEQ (S4.1) of i.i.d. observations of $Z = \begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix}$ with $Z_{ij} = \begin{pmatrix} Z_{ij}^1 \\ Z_{ij}^2 \end{pmatrix}$, $Z_{ij}^k(p_k, 1), k \in \{1,2\}$. Let $Z_1^1 = \begin{pmatrix} Z^1 \\ 1 \end{pmatrix}$ and for $i \geq 1$, $j \in \{1, \dots, m_i\}$, $Z_{1ij}^1 = \begin{pmatrix} Z_{ij}^1 \\ 1 \end{pmatrix}$. Let $\mu_n = \sum_{i=1}^n m_i$. Let λ^1 and λ^2 be respectively the smallest eigenvalue of $E[Z_1^1(Z_1^1)^T]$ and C^2, $\lambda = min(\lambda^1, \lambda^2)$ in Assumption H3b1.

We define the process (B_n^{12}) in $\mathbb{R}^{p_1 \times p_2}$ as in S4.1 for CCA and the process (B_n^{21}) in $\mathbb{R}^{p_2 \times p_1}$ as in S5.1 for FCA.

By Corollaries 1, 2, 3, 4, under H3b1 and H7b,c restricted to Z^1, for $k \neq l \in \{1,2\}$:

$$B_n^{kl} \longrightarrow B^{kl}, \sum_1^\infty a_n \left\| B_n^{kl} - B^{kl} \right\| < \infty \text{ a.s.}$$

$$B_n^1 = B_n^{12}B_n^{21} \longrightarrow B^1 = B^{12}B^{21}, \sum_1^\infty a_n \| B_n^1 - B^1 \| < \infty \text{ a.s.}$$

When $\omega_{2n} = 0$ for all n, H3b1 can be replaced by H3c and H7b by H7a.

Let $\bar{Z}_n^1$ be the mean of the sample $\left(Z_{11}^1, \dots, Z_{nm_n}^1\right)$ of Z^1 and C_n^1 its covariance matrix, $C_n^1 = \frac{1}{\mu_n}\sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij}^1 \left(Z_{ij}^1\right)^T - \bar{Z}_n^1(\bar{Z}_n^1)^T$, both recursively computed. The matrix C_n^1 is positive definite

from a certain rank. Let $\langle .,. \rangle_{1,n+1}$ be the inner product and $\|.\|_{1,n+1}$ the norm induced by C_n^1 in $\mathbb{R}^{p_1}$. For $i \in \{1,\dots,r\}, r \leq min(p_1, p_2 - 1)$, we recursively define the processes $\left(\tilde{X}_n^i\right), \left(\tilde{Y}_n^i\right), \left(X_n^i\right)$ in $\mathbb{R}^{p_1}$ and apply Lemma 1 with $V_n = B_n^1$, $B = B^1$:

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

Theorem 7. *Let λ^1 and λ^2 be respectively the smallest eigenvalue of $E[Z_1^1 (Z_1^1)^T]$ and C^2, $\lambda = min(\lambda^1, \lambda^2)$. Under H1b, H2a, H3b1, H5, H7b,c restricted to Z^1, for $i \in \{1,\dots,r\}$, X_n^i converges to η_i^1 or $-\eta_i^1$, $\langle B_n^1 X_n^i, X_n^i \rangle_n$ converges to $r_c^2\left(V_i^1 | Z^2\right)$, $\sum_1^\infty a_n \left| \langle B_n^1 X_n^i, X_n^i \rangle_n - r_c^2\left(V_i^1 | Z^2\right) \right| < \infty$ a.s. When $\omega_{2n} = 0$ for all n, the same conclusions hold with H3b1 replaced by H3c and H7b by H7a.*

7. Proofs

We use the Hilbert-Schmidt inner product $\langle .,. \rangle$ and norm $\|.\|$ in a space of matrices. If $\|A\|_2$ is the spectral norm of the matrix A, $\|AC\| \leq \|A\|_2 \|C\| \leq \|A\| \|C\|$.

Lemma A (Robbins-Siegmund 1971). *Let $(\Omega, \mathcal{A}, P)$ be a probability space, (T_n) a non-decreasing sequence of sub-σ-fields of $\mathcal{A}$. Assume for all n, $z_n, \alpha_n, \beta_n, \gamma_n$ are four integrable T_n-measurable random variables defined on $(\Omega, \mathcal{A}, P)$ such that:*

$$E[z_{n+1} | T_n] \leq z_n(1 + \alpha_n) + \beta_n - \gamma_n \text{ a.s.}$$

Then, in the set $\{\sum_1^\infty \alpha_n < \infty, \sum_1^\infty \beta_n < \infty\}$, (z_n) converges to a finite random variable and $\sum_1^\infty \gamma_n < \infty$ a.s.

Lemma B. *Let $\nu > 0, c > 0$, two sequences of positive numbers (a_n) and (d_n) such that:*

$$d_{n+1} \leq (1 - \nu a_n) d_n + c a_n^2,$$

$$\sum_1^\infty a_n = \infty, a_n = o(1), \frac{a_n}{a_{n+1}} \leq 1 + \gamma a_n + \gamma_n, 0 < \gamma < \nu,\ \gamma_n \geq 0, \sum_1^\infty \gamma_n < \infty.$$

Then there exists $b > 0$ such that, for all n, $d_n \leq b a_n$.

The proof of Lemma B for $\nu = 2$ is a part of the proof of Lemma 3 in (Monnez 2022a). The proof for any $\nu > 0$ is similar.

Proof of Theorem 1

The writing of ω belonging to the intersection of the almost sure convergence sets is omitted. Let λ and λ_{max} be respectively the smallest and the largest eigenvalue of F. Since $a_n = o(1)$, from a certain rank $a_n < \frac{1}{\lambda_{max}}$, then all the eigenvalues of $I - a_n F$ are positive and the spectral norm $\|I - a_n F\|_2$ is equal to $1 - a_n \lambda$.

$$A_{n+1} - A = (I - a_n F_n)(A_n - A) - a_n (F_n - F) A + a_n (G_n - G).$$

$$\|A_{n+1} - A\| \leq \|I - a_n F_n\|_2 \|A_n - A\| + a_n \|F_n - F\| \|A\| + a_n \|G_n - G\|.$$

For $0 < \varepsilon < \lambda$, since $\|F_n - F\|_2 = o(1)$ by A1, for n sufficiently large:

$$\|I - a_n F_n\|_2 \leq \|I - a_n F\|_2 + a_n \|F_n - F\|_2 \leq 1 - a_n(\lambda - \varepsilon),$$

$$\|A_{n+1} - A\| \leq \|A_n - A\| + a_n \|F_n - F\| \|A\| + a_n \|G_n - G\| - a_n(\lambda - \varepsilon)\|A_n - A\|.$$

By A2 and A3a, applying Lemma A with $T_n = \mathcal{A}$ for all n yields:

$$\exists T \geq 0\colon \|A_n - A\| \longrightarrow T, \sum\nolimits_1^\infty a_n \|A_n - A\| < \infty \Longrightarrow T = 0.$$

Proof of Theorem 2

$$A_{n+1} - A = (I - a_n F)(A_n - A) - a_n U_n,$$

$$U_n = (F_n - F)(A_n - A) + (F_n - F)A - (G_n - G),\ E[U_n|T_n] = 0 \text{ a.s. (B1)}$$

$$\Longrightarrow E[\|A_{n+1} - A\|^2|T_n] = \|(I - a_n F)(A_n - A)\|^2 + a_n^2 E[\|U_n\|^2|T_n] \text{ a.s.}$$

$$\|(I - a_n F)(A_n - A)\| \leq \|I - a_n F\|_2 \|A_n - A\| = (1 - a_n\lambda)\|A_n - A\|.$$

. Let $d = 3f\|A\|^2 + 3g$. By B2:

$$E[\|U_n\|^2|T_n] \leq 3f\|A_n - A\|^2 + 3f\|A\|^2 + 3g = 3f\|A_n - A\|^2 + d \text{ a.s.}$$

$$E[\|A_{n+1} - A\|^2|T_n] \leq (1 + a_n^2\lambda^2 + 3a_n^2 f)\|A_n - A\|^2 + d\ a_n^2 - 2a_n\lambda\|A_n - A\|^2 \text{ a.s.}$$

By H3a, applying Lemma A yields:

$$\exists T \geq 0\colon \|A_n - A\|^2 \longrightarrow T, \sum\nolimits_1^\infty a_n \|A_n - A\|^2 < \infty \Longrightarrow T = 0 \text{ a.s.}$$

Taking the expectation of $E[\|A_{n+1} - A\|^2|T_n]$ gives:

$$E[\|A_{n+1} - A\|^2] \leq (1 + a_n^2\lambda^2 + 3a_n^2 f)E[\|A_n - A\|^2] + d\ a_n^2 - 2a_n\lambda E[\|A_n - A\|^2].$$

By H3a, applying the deterministic version of Lemma A yields:

$$\exists t \geq 0\colon E[\|A_n - A\|^2] \longrightarrow t, \sum\nolimits_1^\infty a_n E[\|A_n - A\|^2] < \infty \Longrightarrow t = 0.$$

Thus, there exists $c > 0$ such that:

$$E[\|A_{n+1} - A\|^2] \leq (1 - 2a_n\lambda)E[\|A_n - A\|^2] + c\ a_n^2.$$

By H3b1, applying Lemma B yields:

$\exists b > 0\colon E[\|A_n - A\|^2] \leq ba_n \Longrightarrow \sum_1^\infty a_n E[\|A_n - A\|] < \infty, \sum_1^\infty a_n \|A_n - A\| < \infty$ a.s.

References

1. Balsubramani, A., S. Dasgupta and Y. Freund. 2013. The fast convergence of incremental PCA. *Adv. Neural Inf. Syst.* 26:3174-3182.

2. Bar, R. 2013. Développement de méthodes d'analyse de données en ligne. *Thèse de Doctorat de l'Université de Lorraine.* hal.science/tel-01750512

3. Benzécri, J.P. 1969. Approximation stochastique dans une algèbre normée non commutative. *Bull. Soc. Math. France* 97:225-241.

4. 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.

5. Cardot, H. and D. Degras. 2018. Online principal component analysis in high dimension: which algorithm to choose? *Internat. Statist. Rev.* 86:29-50.

6. 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,* 227-228.

7. Duarte, K., J.M. Monnez and E. Albuisson. 2018. Sequential linear regression with online standardized data. *PLoS ONE 13 (1):e0191186.*

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

9. Lalloué, B., J.M. Monnez and E. Albuisson. 2022. Streaming constrained binary logistic regression with online standardized data. *J. Applied Statistics* 49 (6):1519-1539.

10. Ljung, L., G.C. Pflug and H. Walk. 1992. *Stochastic Approximation and Optimization of Random Systems*. DMV Seminar, 17, Birkhaüser, Basel.

11. Monnez, J.M. 2006. Almost sure convergence of stochastic gradient processes with matrix step sizes. *Stat. Proba. Letters* 76:531-536.

12. Monnez, J.M. 2008. Stochastic approximation of the factors of a generalized canonical correlation analysis. *Stat. Proba. Letters* 78:713-722.

13. Monnez, J.M. 2010. ACP projetée de données séquentielles. *42èmes Journées de Statistique, Marseille.* inria-00494695

14. Monnez, J.M. 2022a. Stochastic approximation of eigenvectors and eigenvalues of the Q-symmetric expectation of a random matrix. Accepted in *Communications in Statistics-Theory and Methods.* hal-03956687

15. Monnez, J.M. 2022b. Analyse des données en flux. Analyse en composantes principales et méthodes dérivées. *Lecture.* hal-03690459

16. 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. hal-03038206

17. 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.

18. 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.