Analyse des données en flux. Analyse en composantes principales et méthodes dérivées

Jean-Marie Monnez

▶ **To cite this version:**

Jean-Marie Monnez. Analyse des données en flux. Analyse en composantes principales et méthodes dérivées. Doctorat. France. 2022. hal-03690459v2

HAL Id: hal-03690459

https://hal.science/hal-03690459v2

Submitted on 19 Oct 2022 (v2), last revised 19 Jun 2023 (v3)

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.

Analyse des données en flux
Analyse en composantes principales et méthodes dérivées

Jean-Marie Monnez [1]*
Université de Lorraine, CNRS, Inria[2], IECL[3], F-54000 Nancy, France

Table des matières

1. Approximation stochastique de vecteurs et valeurs propres

2. Analyse en composantes principales

3. Analyse canonique généralisée

4. Analyse canonique

Notations

Soit Q une matrice $p \times p$ symétrique définie positive appelée métrique, $<.,.>$ le produit scalaire et $\|.\|$ la norme induite par Q. Pour des vecteurs dans $\mathbb{R}^p$, Q-orthogonaux et Q-normés signifient respectivement orthogonaux et normés par rapport à la métrique Q. On rappelle qu'une matrice $p \times p$ B est Q-symétrique si QB est symétrique ; alors B a p valeurs propres réelles et il existe une base Q-orthonormée de $\mathbb{R}^p$ formée de vecteurs propres de B. Soit $\mathbb{R}^{p*}$ l'espace dual de $\mathbb{R}^p$. Soit $\|B\|$ la norme spectrale de B. On note A^{T} la transposée de la matrice A. On note $(\cdot)$ une suite de matrices ou de vecteurs ou de réels selon le contexte. On utilise l'abréviation p.s. pour "presque sûrement".

[1] *jean-marie.monnez@univ-lorraine.fr
[2] Inria, project-team BIGS
[3] Institut Elie Cartan de Lorraine, BP 70239, F-54506 Vandoeuvre-lès-Nancy, France

1 Approximation stochastique de vecteurs et valeurs propres

1.1 Formulation

Dans le cas d'un flux de données ou de données massives que l'on traite séquentiellement, on peut utiliser des algorithmes d'approximation stochastique pour estimer en ligne, par exemple, des paramètres d'une fonction de régression (par exemple, Lalloué, Monnez, Albuisson, 2022), des centres de classe en classification non supervisée (Cardot, Cénac, Monnez, 2012), des composantes principales d'une ACP (par exemple, Cardot, Degras, 2018). Chaque vecteur de données entrant est utilisé pour actualiser l'estimation courante jusqu'à la convergence vers le paramètre à estimer. Quand on utilise ce type de processus, il n'est pas nécessaire de stocker les données et, en raison de la simplicité relative des calculs à effectuer, on peut utiliser un plus grand nombre de données que pour une méthode non séquentielle dans le même temps de calcul. De plus, ce type de méthode utilise moins d'espace mémoire qu'une méthode non séquentielle (Balsubramini, Dasgupta, Freund, 2013).

Dans ce contexte, considérons le problème de l'estimation par un processus récursif de vecteurs propres et des valeurs propres correspondantes par ordre décroissant de l'espérance mathématique Q-symétrique inconnue B d'une matrice aléatoire $p \times p$.

Beaucoup d'articles ont été consacrés à ce problème dans le cas où B est une matrice symétrique ($Q = I$) ou une matrice Q-symétrique, la métrique Q étant connue. Citons parmi d'autres les algorithmes de Benzécri (1969), Krasulina (1970), Oja (1982), Karhunen et Oja (1982), Oja et Karhunen (1985), Brandière (1994, 1995, 1998), Brandière et Duflo (1996) et Duflo (1997) dans le cas de l'ACP. Oja et Karhunen (1985) ont démontré la convergence p.s vers un vecteur propre unitaire V_1, correspondant à la plus grande valeur propre simple λ_1 de B, du processus (X_n) défini récursivement au pas n par $X_{n+1} = \frac{(I+a_n B_n)X_n}{\|(I+a_n B_n)X_n\|}$, en supposant les matrices aléatoires $B_n, n \geq 1$, mutuellement indépendantes d'espérance mathématique B et presque sûrement bornées ($\|B_n\| \leq b \in \mathbb{R}^+$), $a_n > 0$, $\sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} a_n^2 < \infty$. Ils ont démontré la convergence de processus $\left(X_n^i\right), 1 \leq i \leq p$, du même type pour estimer, après orthonormalisation, des vecteurs propres V_i correspondant aux valeurs propres λ_i de B en ordre décroissant. L'hypothèse d'indépendance des matrices aléatoires B_n peut ne pas être

vérifiée, par exemple dans l'ACP en flux d'un vecteur aléatoire Z dans $\mathbb{R}^p$ étudiée dans le paragraphe 2, lorsque l'espérance mathématique de Z et la métrique M dans $\mathbb{R}^p$ sont estimées en ligne. En utilisant des méthodes de démonstration différentes, Monnez et Skiredj (2021) ont étudié deux cas où les matrices B_n ne sont pas indépendantes, le premier lorsque $E[B_n|T_n]$ converge p.s. vers B, T_n étant la tribu engendrée par $B_1, \dots, B_{n-1}$, le second lorsque B_n converge p.s. vers B ; dans cet article, des résultats d'expériences dans le cas de l'ACP en flux montrent que les processus prenant en compte au pas n toutes les observations de Z jusqu'à ce pas dans la définition de B_n convergent en général plus rapidement que les processus classiques prenant en compte une ou un mini-lot d'observations.

Nous donnons dans le paragraphe 1.2 un théorème de Monnez (2022) qui étend et généralise des résultats précédents dans deux directions, la métrique Q est supposée inconnue et les matrices B_n non indépendantes, et s'applique aux processus prenant en compte à chaque pas un mini-lot d'observations ou toutes les observations jusqu'au pas courant.

1.2 Convergence presque sûre d'un processus d'approximation stochastique

Soit $\lambda_1, \dots, \lambda_p$ les valeurs propres en ordre décroissant de la matrice Q-symétrique B et pour $i \in \{1, \dots, p\}$, V_i un vecteur propre Q-normé correspondant à λ_i. Soit (Q_n) une suite de métriques aléatoires dans $\mathbb{R}^p$, $<.,.>_n$ le produit scalaire et $\|.\|_n$ la norme induites par Q_n. Pour $i \in \{1, \dots, r\}, r \leq p$, considérons les processus $\left(\tilde{X}_n^i\right), \left(\tilde{Y}_n^i\right), \left(X_n^i\right)$ dans $\mathbb{R}^p$ tels que, pour $n \geq 1$:

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

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

Le r-uplet de vecteurs $(X_{n+1}^1, \dots, X_{n+1}^r)$ est obtenu par une orthonormalisation de Gram-Schmidt par rapport à Q_{n+1} de $\left(\tilde{Y}_{n+1}^1, \dots, \tilde{Y}_{n+1}^r\right)$. Les hypothèses sur B, respectivement B_n, a_n, Q_n, X_1^i, sont notées H1, respectivement H2, H3, H4, H5. Supposons :

(H1a) *B est Q-symétrique*. (H1b) *Les r plus grandes valeurs propres de B sont simples.*

(H2a) *Pour tout n, $I + a_n B_n$ est inversible.*

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

(H5) $X_1^i, i \in \{1, \dots, r\}$, *est une variable aléatoire absolument continue par rapport à la mesure de Lebesgue, indépendante de la suite* (B_n).

Considérons le processus de moyennisation pondérée (U_n) dans l'ensemble des matrices $p \times p$ et pour $i \in \{1, \dots, r\}, r \leq p$, le processus réel $\left(\Lambda_n^i\right)$, tels que, pour $n \geq 1$:

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

$$\Lambda_{n+1}^i = < U_{n+1}X_{n+1}^i, X_{n+1}^i >_{n+1}.$$

Faisons l'hypothèse H2b sur U_n :

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

Un modèle de corrélation entre les matrices aléatoires B_n est défini par cette hypothèse. Notons que, dans le cas vectoriel, ce modèle a été utilisé par exemple par Ljung, Pflug, Walk (1992) et Monnez (1991, 2006).

Théorème 1.1. *Sous les hypothèses H1a,b sur B, H2a sur* B_n*, H2b sur* U_n*, H3a sur* a_n*, H4 sur* Q_n *et H5 sur* X_1^i*, pour* $i \in \{1, \dots, r\}$*,* X_n^i *converge vers* V_i *ou* $-V_i$*,* Λ_n^i *converge vers* λ_i*,* $\sum_1^\infty a_n \left| < BX_n^i, X_n^i >_n - \lambda_i \right| < \infty$ *et* $\sum_1^\infty a_n \left|\Lambda_n^i - \lambda_i\right| < \infty$ *p.s.*

La démonstration est dans (Monnez, 2022). Notons que, dans le cas déterministe où la matrice B est connue et $B_n = B$, il est facile de démontrer que X_n converge vers V_1 ou $-V_1$ lorsque X_1 n'est pas orthogonal à V_1, ce qui est vérifié presque sûrement sous l'hypothèse H5.

1.3 Corollaires

Donnons des conditions suffisantes sur B_n qui impliquent l'hypothèse H2b sur U_n. Les démonstrations des lemmes sont dans (Monnez, 2022).

T_n étant la tribu engendrée par $X_1, Q_1, B_1^1, B_1^2, Q_2, \dots, B_{n-1}^1, B_{n-1}^2, Q_n$, supposons

$$B_n = B_n^1 + B_n^2 \text{ avec } E[B_n^2 | T_n] = 0,$$

$$U_n = U_n^1 + U_n^2 \text{ avec } U_{n+1}^i = (1 - a_n)U_n^i + a_n B_n^i,\ i \in \{1,2\}, U_1^1 + U_1^2 = 0.$$

Faisons les hypothèses :

(H2c) $\sum_1^\infty a_n \|B_n^1 - B\| < \infty$ *p.s.* (H2d) $E[B_n^2 | T_n] = 0$ *p.s.,* $sup_n E[\|B_n^2\|^2] < \infty$.

(H2e) $\sum_{n=1}^{\infty} a_n \left\|(U_{n+1}^1 - B)B_n^2\right\| < \infty$ *p.s.* (H2f) $sup_n \|B_n^2\| < \infty$ *p.s.*

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

Notons que l'hypothèse H3b implique H3a et que, pour $a_n = \frac{a}{n^{\alpha}}$, H3b est vérifiée à partir d'un certain rang n_0 pour $\frac{2}{3} < \alpha < 1$ et si $a > \frac{1}{2}$ pour $\alpha = 1$.

Lemme 1.1. *Sous les hypothèses H2c sur* B_n^1 *et* $a_n > 0$, $a_n = o(1)$, $\sum_1^{\infty} a_n = \infty$, *on a :* $U_n^1 \to B$, $\sum_1^{\infty} a_n \|U_n^1 - B\| < \infty$, $\sum_1^{\infty} a_n \left\|(U_{n+1}^1 - B)B_n^1\right\| < \infty$ *p.s. ; pour* $U_n^2 = 0$, *l'hypothèse H2b est vérifiée.*

Lemme 1.2. *Sous les hypothèses H2d sur* B_n^2, *H3a sur* a_n *et* $\frac{a_n}{a_{n+1}} \leq 1 + \gamma a_n + \gamma_n$, $\gamma \in]0,2[, \gamma_n \geq 0$, $\sum_1^{\infty} \gamma_n < \infty$, *on a :* $U_n^2 \to 0$ *p.s.*, $\sum_1^{\infty} a_n E[\|U_n^2\|^2] < \infty$, $\exists\, b > 0$: $E[\|U_n^2\|^2] \leq b a_n$ *; si* $\sum_1^{\infty} a_n^{3/2} < \infty$, $\sum_1^{\infty} a_n \|U_n^2\| < \infty$ *p.s.*

Lemme 1.3. *Sous les hypothèses H2c sur* B_n^1, *H2d sur* B_n^2, *H3b sur* a_n, *on a :* $U_n \to B$, $\sum_1^{\infty} a_n \|U_n - B\| < \infty$ *p.s. ; si de plus H2e ou H2f sur* B_n^2 *est vérifiée,* $\sum_1^{\infty} a_n \|(U_{n+1} - B)B_n\| < \infty$ *p.s.*

A partir de ces lemmes, on déduit des corollaires du théorème 1.1. Soit A l'ensemble des hypothèses H1a,b sur B, H2a sur B_n, H3b sur a_n, H4 sur Q_n, H5 sur X_1^i.

1) *Cas général*

D'après le lemme 1.3, l'hypothèse H2b est vérifiée sous H2c, H2d, H2e ou H2f, H3b; on a alors :

Corollaire 1.1. *Sous A, H2c sur* B_n^1, *H2d sur* B_n^2, *H2e ou H2f sur* B_n^2, *on a les conclusions du théorème 1.1.*

2) *Cas du théorème de convergence de Oja et Karhunen*

Soit $B_n^1 = B$, $B_n^2 = B_n - B$, $U_1^1 = B$, alors $U_n^1 = B$ for $n \geq 1$. Les hypothèses H2c,e sont vérifiées.

Corollaire 1.2. *Sous les hypothèses A,* $E[B_n|T_n] = B$ *et* $sup_n E[\|B_n - B\|^2] < \infty$, *on a les conclusions du théorème 1.1.*

Ce corollaire est une extension du théorème 2 de Oja et Karhunen (1985), démontré par la méthode de l'EDO (équation différentielle ordinaire) en supposant $Q = I$, les r plus grandes valeurs propres de B sont positives et $\|B_n\| \leq b$ *p.s.*

3) *Cas où* $E[B_n|T_n]$ *converge p.s. vers B*

Soit $B_n^1 = E[B_n|T_n]$, $B_n^2 = B_n - E[B_n|T_n]$. Alors $E[B_n^2|T_n] = 0$.

Corollaire 1.3. *Sous les hypothèses A,* $\sum_1^\infty a_n \|E[B_n|T_n] - B\| < \infty$ *p.s.*, $sup_n E[\|B_n - E[B_n|T_n]\|^2] < \infty$ et $sup_n \|B_n - E[B_n|T_n]\| < \infty$ *p.s., on a les conclusions du théorème 1.1.*

Ce corollaire, qui s'applique lorsque $E[B_n|T_n]$ converge p.s. vers B, est une extension du théorème 1i et du corollaire 2i de Monnez et Skiredj (2021), qui a été démontré en utilisant une méthode de Duflo (1997, p. 343) et des arguments d'algèbre extérieure, en supposant Q connue, il existe $b > 0$ tel que $sup_n \|B_n\| < b$ p.s., $\sum_1^\infty a_n E[\|E[B_n|T_n] - B\|] < \infty$ et une autre hypothèse non utilisée ici.

4) *Cas où* B_n *converge p.s. vers B*

Soit $B_n^1 = B_n$, $B_n^2 = 0$, $U_1^2 = 0$, alors $U_n^2 = 0$ pour $n \geq 1$. D'après le lemme 1.1, l'hypothèse H2b est vérifiée sous H2c et H3a. D'après le théorème 1.1, $\sum_1^\infty a_n \left| < BX_n^i, X_n^i >_n - \lambda_i \right| < \infty$ p.s.; sous H4, comme $\sum_1^\infty a_n \|B_n - B\| < \infty$ d'après H2c, $\sum_1^\infty a_n \left| < B_n X_n^i, X_n^i >_n - \lambda_i \right| < \infty$ p.s.

Corollaire 1.4. *Sous les hypothèses A avec H3b remplacée par H3a et* $\sum_1^\infty a_n \|B_n - B\| < \infty$ *p.s., on a les conclusions du théorème 1.1 et pour* $i \in \{1, \dots, r\}$, $\sum_1^\infty a_n \left| < B_n X_n^i, X_n^i >_n - \lambda_i \right| < \infty$ *p.s.*

Ce corollaire, qui s'applique lorsque B_n converge p.s. vers B, est une extension du théorème 1ii et du corollaire 2ii de Monnez et Skiredj (2021), qui a été démontré en utilisant le théorème de Robbins et Siegmund (1971) et des arguments d'algèbre extérieure, en supposant Q connue.

5) *Définition d'autres processus*

Par exemple, supposons que les matrices aléatoires B_n soient i.i.d. avec $E[B_n] = B$ et $E[\|B_n\|^2] < \infty$. Soit $V_n = \frac{1}{n}\sum_1^n B_i$. Alors $E[V_n] = B$, $E[\|V_n - B\|^2] = O\left(\frac{1}{n}\right)$, $V_n \to B$ p.s. En supposant $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$, on a $\sum_1^\infty a_n \|V_n - B\| < \infty$ p.s. Si l'on définit les processus $\left(F_n^i\right)$ en remplaçant B_n par V_n dans la définition des processus $\left(X_n^i\right)$, on peut appliquer le corollaire 1.4 à ces nouveaux processus.

De même, d'après le lemme 1.3, sous H2c, H2d et H3b, on a $U_n \to B$, $\sum_1^\infty a_n \|U_n - B\| < \infty$ p.s. Si l'on définit les processus $\left(F_n^i\right)$ en remplaçant B_n par U_n dans la définition des processus $\left(X_n^i\right)$, on peut appliquer le corollaire 1.4 à ces nouveaux processus.

2 Analyse en composantes principales (ACP)

2.1 Formulation

Soit $Z^1, \dots, Z^p$ les composantes d'un vecteur aléatoire Z dans $\mathbb{R}^p$ défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$ et $C = E[(Z - E[Z])(Z - E[Z])^{\mathrm{T}}]$ sa matrice de covariance. Soit M une métrique dans $\mathbb{R}^p$ qui définit une distance entre deux observations de Z. Considérons le problème suivant :

Formulation de l'ACP. *Pour $l \in \{1, \dots, r\}, r \leq p$, déterminer au pas l une combinaison linéaire des composantes centrées de $Z, U_l = c_l^T(Z - E[Z])$, $c_l \in \mathbb{R}^{p*}$, qui soit non corrélée à $U_1, \dots, U_{l-1}$ et de variance maximale sous la contrainte $c_l^T M^{-1} c_l = 1$.*

Définition 2.1. *La combinaison linéaire des composantes centrées de Z déterminée au pas l de l'ACP est appelée la $l^{ième}$ composante principale.*

Nous avons donc à maximiser la forme quadratique $c^T C c$ sous les contraintes $c^T C c_j = 0, j \in \{1, \dots, l-1\}$, et $c^T M^{-1} c = 1$. On établit aisément que :

Proposition 2.1. *c_l est un vecteur propre M^{-1}-normé de la matrice M^{-1}-symétrique $B = MC$ correspondant à sa $l^{ième}$ plus grande valeur propre λ_l. $u_l = M^{-1} c_l$ est un vecteur propre M-normé de la matrice M-symétrique CM correspondant à la même valeur propre λ_l.*

$u_l = M^{-1} c_l$ maximise $u^T MCMu = u^T ME[(Z - E[Z])(Z - E[Z])^{\mathrm{T}}]Mu = E[((Z - E[Z])^T Mu)^2]$ sous les contraintes $u^T M u_j = 0$ (u est M-orthogonal à u_j), $j \in \{1, \dots, l-1\}$, et $u^T Mu = 1$ (u est M-normé). Pour tout $\omega \in \Omega$, $(Z(\omega) - E[Z])^T M u_l$ est l'abscisse de la projection de $Z(\omega) - E[Z]$ sur l'axe $(E[Z], u_l)$.

Définition 2.2. *L'axe $(E[Z], u_l)$ d'origine $E[Z]$ et de vecteur M-normé u_l est appelé le $l^{ième}$ axe principal.*

Proposition 2.2. *Le sous-espace engendré par $u_1, \dots, u_r$ maximise $E[\|\Pi(Z - E[Z])\|^2]$, Π étant l'opérateur de projection M-orthogonale sur un sous-espace de dimension r contenant $E[Z]$.*

Un cas particulier d'ACP est l'ACP normée, lorsque M est la matrice diagonale des inverses des variances des composantes de Z. L'ACP normée revient à utiliser des données standardisées, i.e. des observations de $M^{\frac{1}{2}}(Z - E[Z])$, et la métrique identité, donc la même importance est donnée à chacune des composantes de Z dans cette ACP.

2.2 Approximation stochastique des composantes principales de l'ACP

Soit $(Z_{11}, \dots, Z_{1m_1}, \dots, Z_{n1}, \dots, Z_{nm_n}, \dots)$ une suite d'observations i.i.d. de Z ; $Z_{n1}, \dots, Z_{nm_n}$ sont introduites au pas n des processus. Soit $\bar{Z}_n$ la moyenne de l'échantillon $(Z_{11}, \dots, Z_{nm_n})$ et C_n sa matrice de covariance, $C_n = \frac{1}{\mu_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Z_{ij}Z_{ij}^{\mathrm{T}} - \bar{Z}_n\bar{Z}_n^{\mathrm{T}}$ avec $\mu_n = \sum_{i=1}^{n} m_i$; soit M_n un estimateur de M déterminé récursivement à partir de cet échantillon.

Soit (Q_n) une suite de métriques aléatoires dans $\mathbb{R}^p$ convergeant presque sûrement vers $Q = M^{-1}$, $<.,.>_n$ le produit scalaire et $\|.\|_n$ la norme induits par Q_n. Pour $i \in \{1, \dots, r\}$, définissons les processus $(\tilde{Y}_n^i)$, $(\tilde{X}_n^i)$, (X_n^i) et (Λ_n^i), l'orthonormalisation au pas n étant faite par rapport à la métrique Q_{n+1} :

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

$$U_{n+1} = (1 - a_n)U_n + a_n B_n, U_1 = 0, \quad \Lambda_{n+1}^i = < U_{n+1}X_{n+1}^i, X_{n+1}^i >_{n+1}, \text{ avec}$$

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

$$S_{1n} = \frac{1}{\mu_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Z_{ij}Z_{ij}^{\mathrm{T}}, \; S_{2n} = \frac{1}{m_n}\sum_{j=1}^{m_n} Z_{nj}Z_{nj}^{\mathrm{T}}, \; \omega_{1n} \geq 0, \omega_{2n} \geq 0, \omega_{1n} + \omega_{2n} = 1.$$

B_n est une combinaison convexe de deux matrices. Des poids différents variant dans le temps sont affectés à l'ensemble des observations passées et à celui des observations courantes. Pour $\omega_{2n} = 0$, les observations passées et courantes de Z sont prises en compte au pas n, pour $\omega_{1n} = 0$ seulement les m_n observations courantes. Supposons :

(H1a) *B est Q-symétrique*. (H1b) *Les r plus grandes valeurs propres de B sont simples.*

(H2a) *Pour tout n, $I + a_n B_n$ est inversible.*

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

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

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

(H4) $Q_n \to Q, \sum_1^\infty a_n \|Q_n - Q\| < \infty$ *p.s.*

(H5) $X_1^i, i \in \{1, \dots, r\}$, *est une variable aléatoire absolument continue par rapport à la mesure de Lebesgue, indépendante de la suite* (B_n).

(H6a) $M_n \to M$ *p.s.* (H6b) $\sum_1^\infty a_n \|M_n - M\| < \infty$ *p.s.*

(H7a) Z *a des moments d'ordre 4.* (H7b) Z *est p.s. bornée.*

Soit $B_n = B_n^1 + B_n^2$ avec $B_n^2 = \omega_{2n} M(S_{2n} - E[ZZ^{\mathrm{T}}])$,

$$B_n^1 = B + (M_n - M)\big(\omega_{1n}(S_{1n} - \bar{Z}_n \bar{Z}_n^{\mathrm{T}}) + \omega_{2n}(S_{2n} - \bar{Z}_n \bar{Z}_n^{\mathrm{T}})\big) + \omega_{1n} M(S_{1n} - E[ZZ^{\mathrm{T}}]) -$$

$$M(\bar{Z}_n - E[Z])\bar{Z}_n^{\mathrm{T}} - ME[Z](\bar{Z}_n - E[Z])^{\mathrm{T}}.$$

Sous H7a, $\bar{Z}_n \to E[Z]$, $S_{1n} \to E[ZZ^{\mathrm{T}}]$ p.s., $E[\|\bar{Z}_n - E[Z]\|] = O\big(1/\sqrt{n}\big)$, $E[\|S_{1n} - E[ZZ^{\mathrm{T}}]\|] = O\big(1/\sqrt{n}\big)$; comme $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty, \sum_1^\infty a_n \|\bar{Z}_n - E[Z]\| < \infty, \sum_1^\infty a_n \|S_{1n} - E[ZZ^{\mathrm{T}}]\| < \infty$ p.s. Donc, sous H3c, H6b et H7b, $\sum_1^\infty a_n \|B_n^1 - B\| < \infty$ p.s. De plus, $E[B_n^2 | T_n] = 0$ et, sous H7b, $\|B_n^2\|$ est p.s. bornée, donc les hypothèses H2d,f sont vérifiées. Lorsque $\omega_{2n} = 0$ pour tout n, H7b peut être remplacée par H7a, $B_n \to B$ sous H6a et H7a, donc $\langle B_n X_n^i, X_n^i \rangle_n \to \lambda_i$ lorsque $X_n^i \to \pm c_i$ p.s. En appliquant les corollaires 1.1 et 1.4, on obtient :

Théorème 2.1. *Sous H1b sur B, H2a sur* B_n*, H3b,c sur* a_n*, H4 sur* Q_n*, H5 sur* X_1^i*, H6a,b sur* M_n *et H7b sur Z, pour* $i \in \{1, \dots, r\}$*,* X_n^i *converge vers* c_i *ou* $-c_i$*,* Λ_n^i *converge vers* λ_i*,* $\sum_1^\infty a_n \big| < BX_n^i, X_n^i >_n - \lambda_i \big| < \infty$*,* $\sum_1^\infty a_n \big|\Lambda_n^i - \lambda_i\big| < \infty$ *p.s. Lorsque* $\omega_{2n} = 0$ *pour tout* n*, on a les mêmes conclusions sans H3b et avec H7a remplaçant H7b ; de plus, pour* $i \in \{1, \dots, r\}$*,* $\langle B_n X_n^i, X_n^i \rangle_n$ *converge vers* λ_i*,* $\sum_1^\infty a_n \big|\langle B_n X_n^i, X_n^i \rangle_n - \lambda_i\big| < \infty$ *p.s.*

Notons que, lorsque $\omega_{2n} = 0$ pour tout n, l'hypothèse H7b, "Z est p.s. bornée", n'est pas demandée.

3 Analyse canonique généralisée (ACG)

3.1 Formulation

On utilise une formulation probabiliste de l'analyse canonique généralisée (generalized canonical correlation analysis, gCCA) définie par Carroll (1968).

Soit un vecteur aléatoire Z dans $\mathbb{R}^p$, défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$, dont l'ensemble des composantes est partitionné en q sous-ensembles de variables aléatoires réelles $\{Z^{k1}, \dots, Z^{kr_k}\}$, $k \in \{1, \dots, q\}, \sum_{k=1}^{q} r_k = p$. Soit Z^k le vecteur aléatoire dans $\mathbb{R}^{r_k}$ dont les composantes sont $Z^{k1}, \dots, Z^{kr_k}$, et $C^k = Covar[Z^k]$ la matrice de covariance de Z^k, $C = Covar[Z]$ celle de Z, $C^{k\cdot} = Covar[Z^k, Z]$ la matrice de covariance $E\left[\left(Z^k - E[Z^k]\right)(Z - E[Z])^{\mathrm{T}}\right]$. Supposons :

(H7c) *Il n'y a pas de relation affine entre les composantes de Z.*

Alors les matrices $C^k, k \in \{1, \dots, q\}$, et C sont inversibles.

Formulation de l'ACG. *Pour $l \in \{1, \dots, r\}, r \leq \min(r_1, \dots, r_q)$, déterminer au pas l une combinaison linéaire de variance 1 des composantes centrées de Z, $U_l = \theta_l^{\mathrm{T}}(Z - E[Z])$, non corrélée à $U_1, \dots, U_{l-1}$, appelée $l^{ième}$ composante générale, et pour $k \in \{1, \dots, q\}$, une combinaison linéaire de variance 1 des composantes centrées de Z^k, $V_l^k = \left(\eta_l^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right)$, appelée $l^{ième}$ composante canonique du $k^{ième}$ sous-ensemble de variables, qui maximisent $\sum_{k=1}^{q} \rho^2\left(U_l, V_l^k\right)$, ρ désignant un coefficient de corrélation linéaire.*

Nous verrons que deux cas particuliers de l'ACG sont l'ACP normée ($r_k = 1$ pour tout k) et l'analyse canonique ($q = 2$).

3.2 Interprétation de l'ACG comme une ACP

L'espace $L^2(\Omega, \mathcal{A}, \mathrm{P})$ est muni du produit scalaire $\langle X, Y \rangle = E[XY]$. Pour $k \in \{1, \dots, q\}$, soit Π^k l'opérateur de projection orthogonale sur le sous-espace de $L^2(\Omega, \mathcal{A}, \mathrm{P})$ engendré par les composantes

centrées de Z^k. Soit $\theta_l = \left(\left(\theta_l^1\right)^{\mathrm{T}} \ldots \left(\theta_l^q\right)^{\mathrm{T}}\right)^{\mathrm{T}}$, $\theta_l^k \in \mathbb{R}^{r_k}, k \in \{1, \ldots, q\}$. Soit M la métrique diagonale par blocs dans $\mathbb{R}^p$ dont le k[ième] bloc diagonal est $\left(C^k\right)^{-1}$. Supposons que U_l soit déterminée ; alors :

$$\sum_{k=1}^{q} \rho^2\left(U_l, V_l^k\right) \max \Leftrightarrow \rho^2\left(U_l, V_l^k\right) = cos^2\left(U^l, V_l^k\right) \max, k \in \{1, \ldots, q\}$$

$$\Leftrightarrow V_l^k = \frac{\Pi^k U_l}{\left\|\Pi^k U_l\right\|}, k \in \{1, \ldots, q\}.$$

Proposition 3.1. *Pour* $l \in \{1, \ldots, r\}$, θ_l *est un vecteur propre* C*-normé de la matrice* $B = MC$ *correspondant à sa l[ième] plus grande valeur propre* λ_l *et, pour* $k \in \{1, \ldots, q\}$, $\left(\left(\theta_l^k\right)^{\mathrm{T}} C^k \theta_l^k\right)^{1/2} V_l^k = \left(\theta_l^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right)$. *On a* $U_l = \sum_{k=1}^{q}\left(\left(\theta_l^k\right)^{\mathrm{T}} C^k \theta_l^k\right)^{1/2} V_l^k$.

La démonstration est dans le paragraphe 3.4.

Donc pour $l \in \{1, \ldots, r\}$, la *l[ième]* composante générale de l'ACG de Z, $U_l = \theta_l^{\mathrm{T}}(Z - E[Z])$, de variance 1, est colinéaire à la *l[ième]* composante principale de l'ACP de Z avec la métrique M, $c_l^{\mathrm{T}}(Z - E[Z])$, de variance λ_l, et $\theta_l = \frac{c_l}{\sqrt{\lambda_l}}$. Un cas particulier de l'ACG est l'ACP normée lorsque $r_k = 1$ pour tout k, la métrique diagonale M étant alors celle des inverses des variances des composantes de Z.

3.3 Approximation stochastique des composantes générales et canoniques de l'ACG

Soit $\left(Z_{11}, \ldots, Z_{1m_1}, \ldots, Z_{n1}, \ldots, Z_{nm_n}, \ldots\right)$ une suite d'observations i.i.d. de Z avec $Z_{ij} = \left(\left(Z_{ij}^1\right)^{\mathrm{T}} \ldots \left(Z_{ij}^q\right)^{\mathrm{T}}\right)^{\mathrm{T}}$, $Z_{ij}^k(r_k, 1), k \in \{1, \ldots, q\}$; le mini-lot $\left(Z_{n1}, \ldots, Z_{nm_n}\right)$ est introduit au pas n des processus. Pour $k \in \{1, \ldots, q\}$, soit $\bar{Z}_n^k$ la moyenne de l'échantillon $\left(Z_{11}^k, \ldots, Z_{nm_n}^k\right)$ de Z^k, $C_n^k = \frac{1}{\mu_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Z_{ij}^k\left(Z_{ij}^k\right)^{\mathrm{T}} - \bar{Z}_n^k\left(\bar{Z}_n^k\right)^{\mathrm{T}}$, avec $\mu_n = \sum_{i=1}^{n} m_i$, sa matrice de covariance, calculées récursivement.

Soit M_n et Q_{n+1} les matrices diagonales par blocs d'ordre p dont les k[ièmes] blocs diagonaux sont respectivement M_n^k défini ci-dessous et C_n^k. La matrice Q_n est symétrique définie positive à partir d'un certain rang. Sous l'hypothèse H7a, $C_n^k \to C^k$ a.s., $E\left[\left\|C_n^k - C^k\right\|\right] = O\left(\frac{1}{\sqrt{\mu_n}}\right) = O\left(\frac{1}{\sqrt{n}}\right)$, car $\mu_n \geq n$.

Comme $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$ sous H3c, $\sum_1^\infty a_n \left\| C_n^k - C^k \right\| < \infty$ p.s. Donc $Q_n \to M^{-1}, \sum_1^\infty a_n \| Q_{n+1} - M^{-1} \| < \infty$ et $\sum_1^\infty a_n \| Q_n - M^{-1} \| < \infty$ p.s. ; l'hypothèse H4 est vérifiée.

Pour $k \in \{1, \dots, q\}$, pour obtenir une estimation de $M^k = \left(C^k\right)^{-1}$, définissons le processus d'approximation stochastique $\left(M_n^k, n \geq 1\right)$ tel que :

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

Théorème 3.1. *Sous les hypothèses* $a_n > 0$, $a_n = o(1), \sum_1^\infty a_n = \infty$ *et H7a,c,* $M_n \to M$ *et* $\sum_1^\infty a_n \| M_n - M \| < \infty$ *p.s.*

Les hypothèses H6a,b sur M_n sont vérifiées (Monnez, 2022, Théorème 2).

Définissons les processus $\left(\tilde{Y}_n^i\right), \left(\tilde{X}_n^i\right), \left(X_n^i\right), \left(\Lambda_n^i\right)$, $i \in \{1, \dots, r\}$, comme pour l'ACP :

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

$$U_{n+1} = (1 - a_n) U_n + a_n B_n, U_1 = 0, \quad \Lambda_{n+1}^i = < U_{n+1} X_{n+1}^i, X_{n+1}^i >_{n+1}, \text{ avec}$$

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

$$S_{1n} = \frac{1}{\mu_n} \sum_{i=1}^n \sum_{j=1}^{m_i} Z_{ij} Z_{ij}^{\mathrm{T}}, \; S_{2n} = \frac{1}{m_n} \sum_{j=1}^{m_n} Z_{nj} Z_{nj}^{\mathrm{T}}, \; \omega_{1n} \geq 0, \omega_{2n} \geq 0, \omega_{1n} + \omega_{2n} = 1.$$

En appliquant le théorème 2.1 avec l'hypothèse complémentaire H7c, on obtient :

Théorème 3.2. *Sous H1b, H2a, H3b,c, H5 et H7b,c, pour* $i \in \{1, \dots, r\}$, X_n^i *converge vers* $c_i = \sqrt{\lambda_i} \theta_i$ *ou* $-c_i, \Lambda_n^i$ *converge vers* λ_i, $\sum_1^\infty a_n \left| < B X_n^i, X_n^i >_n - \lambda_i \right| < \infty$, $\sum_1^\infty a_n \left| \Lambda_n^i - \lambda_i \right| < \infty$ *p.s. Lorsque* $\omega_{2n} = 0$ *pour tout* n*, on a les mêmes conclusions sans H3b et avec H7a remplaçant H7b ; de plus, pour* $i \in \{1, \dots, r\}$, $\langle B_n X_n^i, X_n^i \rangle_n \to \lambda_i$, $\sum_{n=1}^\infty a_n \left| \langle B_n X_n^i, X_n^i \rangle_n - \lambda_i \right| < \infty$ *p.s.*

Au pas n, on obtient une estimation séquentielle de c_i et λ_i, donc de θ_i. On peut estimer $E[Z^k]$ par $\bar{Z}_n^k$ et C^k par C_n^k, donc estimer les composantes générales et canoniques de l'ACG.

3.4 Démonstration de la proposition 3.1

Rappel. Soit $Y, X^1, \dots, X^p$ des éléments de l'espace $L^2(\Omega, \mathcal{A}, \mathrm{P})$ des variables aléatoires réelles de moment d'ordre deux fini, muni du produit scalaire $\langle R, S \rangle = E[RS]$. Soit $X = (X^1 \dots X^p)^{\mathrm{T}}$. Soit $\| . \|$ la

norme vectorielle euclidienne usuelle. Soit Π l'opérateur de projection orthogonale sur le sous-espace de L^2 engendré par $X^1, \dots, X^p$. Dans la régression linéaire au sens des moindres carrés de Y par rapport à $X^1, \dots, X^p$, on détermine le vecteur b de dimension p tel que $E[(Y - b^{\mathrm{T}}X)^2]$ soit minimale. On a :

Lemme 3.1. *On suppose les composantes de X linéairement indépendantes. Alors :*

$$E[(Y - b^{\mathrm{T}}X)^2]\ min \Leftrightarrow b^{\mathrm{T}}X = \Pi Y \ \Leftrightarrow \ b = (E[XX^{\mathrm{T}}])^{-1}E[XY].$$

$\Pi^k U_l$ est obtenu en faisant la régression linéaire de U_l par rapport aux composantes centrées de Z^k: $\Pi^k U_l = \left(b_l^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right)$ avec

$$b_l^k = \left(C^k\right)^{-1}E\left[\left(Z^k - E[Z^k]\right)U_l\right] = \left(C^k\right)^{-1}E\left[\left(Z^k - E[Z^k]\right)(Z - E[Z])^{\mathrm{T}}\right]\theta_l = \left(C^k\right)^{-1}C^{k\cdot}\theta_l.$$

$$\sum_{k=1}^{q} cos^2\left(U_l, V_l^k\right) = \sum_{k=1}^{q} cos^2\left(U_l, \Pi^k U_l\right) = \sum_{k=1}^{q} E[U_l \Pi^k U_l] =$$

$$\sum_{k=1}^{q} \theta_l^{\mathrm{T}} E\left[(Z - E[Z])\left(Z^k - E[Z^k]\right)^{\mathrm{T}}\right] b_l^k = \theta_l^{\mathrm{T}} \sum_{k=1}^{q} \left(C^{k\cdot}\right)^{\mathrm{T}}\left(C^k\right)^{-1} C^{k\cdot}\,\theta_l = \theta_l^{\mathrm{T}} CMC\theta_l.$$

Ainsi θ_l maximise $\theta^{\mathrm{T}}CMC\theta$ sous les contraintes $\theta^{\mathrm{T}}C\theta_j = 0, j \in \{1, \dots, l-1\}$ (U_l n'est pas corrélée à $U_1, \dots, U_{l-1}$) et $\theta^{\mathrm{T}}C\theta = 1$ (U_l est de variance 1). Par conséquent, θ_l est un vecteur propre C-normé de la matrice M^{-1}-symétrique $B = C^{-1}CMC = MC$ correspondant à sa $l^{ième}$ plus grande valeur propre λ_l et $\sum_{k=1}^{q} cos^2\left(U_l, V_l^k\right) = \theta_l^{\mathrm{T}} CMC\theta_l = \lambda_l$. Donc $U_l = \theta_l^{\mathrm{T}}(Z - E[Z])$, de variance 1, est colinéaire à la $l^{ième}$ composante principale $c_l^{\mathrm{T}}(Z - E[Z])$ de l'ACP de Z avec la métrique M de variance λ_l, et $\theta_l = \frac{c_l}{\sqrt{\lambda_l}}$.

Soit $\theta_l = \left(\left(\theta_l^1\right)^{\mathrm{T}} \dots \left(\theta_l^q\right)^{\mathrm{T}}\right)^{\mathrm{T}}, \theta_l^k \in \mathbb{R}^{r_k}, k \in \{1, \dots, q\}$. On a : $V_l^k = \frac{\Pi^k U_l}{\|\Pi^k U_l\|}$,

$\Pi^k U_l = \left(\left(C^k\right)^{-1} C^{k\cdot}\theta_l\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right)$. Comme $MC\theta_l = \lambda_l\theta_l$, $\left(C^k\right)^{-1}C^{k\cdot}\theta_l = \lambda_l\theta_l^k$, $k \in \{1, \dots, q\}$.

$$\Pi^k U_l = \lambda_l\left(\theta_l^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right),\ \left\|\Pi^k U_l\right\|^2 = (\lambda_l)^2\left(\theta_l^k\right)^{\mathrm{T}} C^k \theta_l^k,$$

$V_l^k = \frac{\left(\theta_l^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right)}{\left(\left(\theta_l^k\right)^{\mathrm{T}} C^k \theta_l^k\right)^{\frac{1}{2}}}$, $U_l = \theta_l^{\mathrm{T}}(Z - E[Z]) = \sum_{k=1}^{q}\left(\theta_l^k\right)^{\mathrm{T}}\left(Z^k - E[Z^k]\right) = \sum_{k=1}^{q}\left(\left(\theta_l^k\right)^{\mathrm{T}} C^k \theta_l^k\right)^{\frac{1}{2}} V_l^k$.

4 Analyse canonique (AC)

4.1 Formulation

On considère l'analyse canonique de deux ensembles de variables aléatoires réelles (canonical correlation analysis, CCA). Soit R et S deux vecteurs aléatoires dans $\mathbb{R}^p$ et $\mathbb{R}^q$ respectivement. Soit les matrices de covariance $C^1 = E[(R - E[R])(R - E[R])^{\mathrm{T}}]$, $C^2 = E[(S - E[S])(S - E[S])^{\mathrm{T}}]$, $C^{12} = E[(R - E[R])(S - E[S])^{\mathrm{T}}]$, $C^{21} = (C^{12})^{\mathrm{T}}$. On suppose qu'il n'y a pas de relation affine entre les composantes de $Z = \begin{pmatrix} R \\ S \end{pmatrix}$.

Formulation de l'AC. *Pour $l \in \{1, \dots, r\}, r \leq min(p, q)$, déterminer au pas l une combinaison linéaire des composantes centrées de $R, V_l = f_l^{\mathrm{T}}(R - E[R])$, et une combinaison linéaire des composantes centrées de $S, W_l = g_l^{\mathrm{T}}(S - E[S])$, de variance 1, non corrélées respectivement à $V_1, \dots, V_{l-1}$ et $W_1, \dots, W_{l-1}$, appelées respectivement $l^{ième}$ composante canonique en R et en S, qui maximisent le coefficient de corrélation linéaire $\rho(V_l, W_l)$.*

Soit $F = (C^1)^{-1}C^{12}$ et $G = (C^2)^{-1}C^{21}$.

Proposition 4.1. *f_l et g_l sont respectivement vecteurs propres de la matrice C^1-symétrique FG et de la matrice C^2-symétrique GF correspondant à la même $l^{ième}$ plus grande valeur propre $\mu_l = \rho^2(U_l, V_l)$ et $f_l = \frac{1}{\sqrt{\mu_l}} F g_l$, $g_l = \frac{1}{\sqrt{\mu_l}} G f_l$.*

La démonstration est dans le paragraphe 4.4.

4.2 Interprétation de l'AC comme une ACG

Considérons l'ACG du vecteur aléatoire $Z = \begin{pmatrix} R \\ S \end{pmatrix}$ dans $\mathbb{R}^{p+q}$. Soit C la matrice de covariance de Z et $U_l = \theta_l^{\mathrm{T}}(Z - E[Z]), l \in \{1, \dots, r\}, r \leq min(p, q)$, la $l^{ième}$ composante générale de variance 1 de l'ACG. D'après la proposition 3.1, θ_l est un vecteur propre C-normé correspondant à la $l^{ième}$ plus grande valeur propre λ_l de la matrice

$$B = MC = \begin{pmatrix} (C^1)^{-1} & 0 \\ 0 & (C^2)^{-1} \end{pmatrix} \begin{pmatrix} C^1 & C^{12} \\ C^{21} & C^2 \end{pmatrix} = \begin{pmatrix} I_p & (C^1)^{-1}C^{12} \\ (C^2)^{-1}C^{21} & I_q \end{pmatrix} = \begin{pmatrix} I_p & F \\ G & I_q \end{pmatrix}.$$

Proposition 4.2. $\begin{pmatrix} f_l \\ g_l \end{pmatrix}, l \leq min(p,q)$, *est un vecteur propre de B correspondant à la* $l^{ième}$ *plus grande valeur propre* λ_l*; on a* $\begin{pmatrix} f_l \\ g_l \end{pmatrix} = \sqrt{2\lambda_l}\,\theta_l$, $V_l + W_l = \sqrt{2\lambda_l}\,U_l$, $\rho(V_l, W_l) = \lambda_l - 1$*;* V_l *et* W_l *sont les composantes canoniques respectivement en R et S de l'ACG de Z.*

La démonstration est dans le paragraphe 4.4.

4.3 Approximation stochastique des composantes canoniques de l'AC

Pour estimer en ligne θ_i et λ_i pour $i \in \{1, \dots, r\}$, on utilise pour l'ACG de $Z = \begin{pmatrix} R \\ S \end{pmatrix}$ les processus $\left(X_n^i\right)$, $\left(\Lambda_n^i\right)$ ou $\left(\langle B_n X_n^i, X_n^i \rangle_n\right)$ lorsque $\omega_{2n} = 0$ pour tout n, définis dans le paragraphe 3.3. Sous les hypothèses du théorème 3.2, X_n^i converge p.s. vers $c_i = \sqrt{\lambda_i}\theta_i = \frac{1}{\sqrt{2}}\begin{pmatrix} f_i \\ g_i \end{pmatrix}$, Λ_n^i ou $\langle B_n X_n^i, X_n^i \rangle_n$ vers $\lambda_i = 1 + \rho(V_i, W_i)$.

4.4 Démonstrations

Démonstration de la proposition 4.1

1) On ne prend pas en compte dans un premier temps les conditions de non-corrélation. Soit Π^1, respectivement Π^2, l'opérateur de projection orthogonale sur le sous-espace de $L^2(\Omega, \mathcal{A}, \mathrm{P})$ engendré par les composantes centrées de R, respectivement S.

Au pas l, on cherche un couple (V_l, W_l) tel que $\rho(V_l, W_l) = cos(V_l, W_l)$ soit maximal.

Pour V_l fixé, $W_l = g_l^{\mathrm{T}}(S - E[S])$ est colinéaire à $\Pi^2 V_l$ obtenu en effectuant la régression linéaire de V_l par rapport à $S - E[S]$: $W_l = \frac{\Pi^2 V_l}{\alpha_l}$ avec $\alpha_l = \|\Pi^2 V_l\| = cos(V_l, W_l) = \rho(V_l, W_l)$; $\Pi^2 V_l = \alpha_l g_l^{\mathrm{T}}(S - E[S])$ avec, d'après le lemme 3.1 :

$$\alpha_l g_l = (C^2)^{-1} E[(S - E[S])V_l] = (C^2)^{-1} E[(S - E[S])(R - E[R])^{\mathrm{T}}] f_l = (C^2)^{-1} C^{21} f_l = G f_l.$$

De même, pour W_l fixé, $V_l = f_l^{\mathrm{T}}(R - E[R])$ est colinéaire à $\Pi^1 W_l$ obtenu en effectuant la régression linéaire de W_l par rapport à $R - E[R]$: $V_l = \frac{\Pi^1 W_l}{\alpha_l}$ avec $\alpha_l f_l = F g_l$. Donc :

$$f_l = \frac{1}{\alpha_l} F g_l = \frac{1}{\alpha_l^2} FG f_l \text{ et } g_l = \frac{1}{\alpha_l^2} GF g_l.$$

f_l et g_l sont respectivement vecteurs propres de FG et GF correspondant à la même valeur propre $\mu_l = \alpha_l^2 = \rho^2(U_l, V_l)$, $f_l = \frac{1}{\sqrt{\mu_l}} F g_l$, $g_l = \frac{1}{\sqrt{\mu_l}} G f_l$.

2) FG et GF sont de rang $\min(p, q)$. Les vecteurs propres de la matrice C^1-symétrique $FG = (C^1)^{-1} C^{12} (C^2)^{-1} C^{21}$ correspondant à des valeurs propres supposées distinctes sont C^1-orthogonaux : pour $l \neq j \leq \min(p, q)$, $f_l^{\mathrm{T}} C^1 f_j = E[V_l V_j] = 0$, V_l et V_j ne sont pas corrélées.

Donc, pour $l \in \{1, \ldots, r\}$, f_l vérifiant $f_l^{\mathrm{T}} C^1 f_j = 0, j \in \{1, \ldots, l-1\}$ et maximisant $\rho(U_l, V_l)$ est un vecteur propre C^1-normé de FG correspondant à la $l^{ième}$ plus grande valeur propre $\mu_l = \rho^2(U_l, V_l)$. De même, g_l est un vecteur propre C^2-normé de GF correspondant à la même valeur propre.

Démonstration de la proposition 4.2

Comme $f_l = \frac{1}{\sqrt{\mu_l}} F g_l$ et $g_l = \frac{1}{\sqrt{\mu_l}} G f_l$:

$$B \begin{pmatrix} f_l \\ g_l \end{pmatrix} = \begin{pmatrix} I_p & F \\ G & I_q \end{pmatrix} \begin{pmatrix} f_l \\ g_l \end{pmatrix} = (1 + \sqrt{\mu_l}) \begin{pmatrix} f_l \\ g_l \end{pmatrix}.$$

$\begin{pmatrix} f_l \\ g_l \end{pmatrix}, l \leq \min(p, q)$, est un vecteur propre de B correspondant à la valeur propre $1 + \sqrt{\mu_l}$. Notons que $\begin{pmatrix} -f_l \\ g_l \end{pmatrix}, l \leq \min(p, q)$, est un vecteur propre de B correspondant à $1 - \sqrt{\mu_l}$. Les autres $\max(p, q) - \min(p, q)$ valeurs propres de B sont égales à 1 : pour $(p > q, Gx = 0, x \neq 0)$, $B \begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}$. Donc, la $l^{ième}$ plus grande valeur propre de B est $\lambda_l = 1 + \sqrt{\mu_l}$ et $\begin{pmatrix} f_l \\ g_l \end{pmatrix}$ est colinéaire au vecteur propre C-normé θ_l correspondant à λ_l.

Comme $f_l^{\mathrm{T}} C^1 f_l = E[(V_l)^2] = 1$, $g_l^{\mathrm{T}} C^2 g_l = 1$ et $f_l^{\mathrm{T}} C^{12} g_l = E[V_l W_l] = \rho(V_l, W_l) = \sqrt{\mu_l}$, le carré de la C-norme de $\begin{pmatrix} f_l \\ g_l \end{pmatrix}$ est égal à :

$$(f_l^{\mathrm{T}} \quad g_l^{\mathrm{T}}) \begin{pmatrix} C^1 & C^{12} \\ C^{21} & C^2 \end{pmatrix} \begin{pmatrix} f_l \\ g_l \end{pmatrix} = f_l^{\mathrm{T}} C^1 f_l + 2 f_l^{\mathrm{T}} C^{12} g_l + g_l^{\mathrm{T}} C^2 g_l = 2 + 2\sqrt{\mu_l} = 2\lambda_l.$$

Donc :

$$\begin{pmatrix} f_l \\ g_l \end{pmatrix} = \sqrt{2\lambda_l}\,\theta_l,\ \rho(V_l, W_l) = \sqrt{\mu_l} = \lambda_l - 1,$$

$$U_l = \theta_l^{\mathrm{T}}(Z - E[Z]) = \frac{1}{\sqrt{2\lambda_l}}(f_l^{\mathrm{T}}\ g_l^{\mathrm{T}})\begin{pmatrix} R - E[R] \\ S - E[S] \end{pmatrix} = \frac{1}{\sqrt{2\lambda_l}}(V_l + W_l).$$

$V_l = f_l^{\mathrm{T}}(R - E[R])$, colinéaire à $\Pi^1 W_l$, et $W_l = g_l^{\mathrm{T}}(S - E[S])$, colinéaire à $\Pi^2 V_l$, sont donc colinéaires respectivement à $\Pi^1 U_l$ et à $\Pi^2 U_l$ et sont les composantes canoniques respectivement en R et S de l'ACG du vecteur aléatoire Z.

Bibliographie

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

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

O. Brandière. *Un algorithme du gradient pour l'analyse en composantes principales*. Prépublication 14, Université de Marne-la-Vallée, 1994.

O. Brandière. *Un algorithme pour l'analyse en composantes principales*. C.R. Acad. Sci. 321 (série I) (1995) 233-236.

O. Brandière. *Some pathological traps for stochastic approximation*. SIAM J. Control Optim. 36 (4) (1998) 1293-1314.

O. Brandière, M. Duflo. *Les algorithmes stochastiques contournent-ils les pièges ?* Ann. Inst. Henri Poincaré Probab. Stat. 32 (1996) 395-427.

H. Cardot, P. Cénac, J.M. Monnez. *A fast and recursive algorithm for clustering large datasets with k-medians*. Comput. Statist. Data Anal. 56 (2012) 1434-1449. doi: 10.1016/j.csda.2011.11.019

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

J.D. Carroll. *A generalization of canonical correlation analysis to three or more sets of variables*. Proceedings of the 76th Convention American Psychological Association, 1968, 227-228.

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

J.Karhunen, E. Oja. *New methods for stochastic approximation of truncated Karhunen-Loève expansions*. Proceedings of the 6th International Conference on Pattern Recognition, IEEE Computer Society Press, 1982, 550-553.

T.P. Krasulina. *Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices*. Autom. Remote Control 2 (1970) 215-221.

B. Lalloué, J.M. Monnez, E. Albuisson. *Streaming constrained binary logistic regression with online standardized data*. J. Applied Statistics 49 (6) (2022) 1519-1539.
doi: 10.1080/02664763.2020.1870672

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

J.M. Monnez. *Etude d'un modèle de corrélation en approximation stochastique (Study of a correlation model in stochastic approximation)*. Statistique et Analyse des Données 16 (3) (1991) 143-181. http://www.numdam.org/item?id=SAD_1991__16_3_143_0

J.M. Monnez. *Almost sure convergence of stochastic gradient processes with matrix step sizes*. Stat. Proba. Letters 76 (2006) 531-536. doi: 10.1016/j.spl.2005.09.014

J.M. Monnez. *Stochastic approximation of eigenvectors and eigenvalues of the Q-symmetric expectation of a random matrix*. Communications in Statistics-Theory and Methods (2022). doi: 10.1080/03610926.2022.2107225

J.M. Monnez, A. Skiredj. *Widening the scope of an eigenvector stochastic approximation process and application to streaming PCA and related methods*. J. Multivariate Anal. 182 (2021) 104694. doi: 10.1016/j.jmva.2020.104694.

E. Oja. *A simplified neuron model as a principal component analyser*. J. Math. Biol. 15 (1982) 267-273.

E. Oja, J. Karhunen. *On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix*. J. Math. Anal. Appl. 106 (1985) 69-84.

H. Robbins, D. Siegmund. *A convergence theorem for nonnegative almost supermartingales and some applications*. JS Rustagi (Ed.), Optimizing Methods in Statistics, Academic Press, New-York, 1971, 233-257.