HAL open science

# Approximation stochastique en analyse factorielle multiple

Jean-Marie Monnez

**To cite this version:**

Jean-Marie Monnez. Approximation stochastique en analyse factorielle multiple. 2006. hal-00109529

**HAL Id: hal-00109529**

**https://hal.science/hal-00109529v1**

Preprint submitted on 24 Oct 2006

**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.

# Approximation stochastique en analyse factorielle multiple

Jean-Marie Monnez
*Institut Elie Cartan UMR 7502*
*Nancy-Université, CNRS, INRIA*

**Abstract**

We study the multiple factorial analysis of a random vector $X$ in $\mathbb{R}^p$ as a principal component analysis of $X$ with a particular choice of the inner product in $\mathbb{R}^p$. We use a stochastic approximation process to estimate recursively the principal components of this analysis by means of a sequence of i.i.d. observations of $X$. Stochastic approximations of the symmetrical positive definite matrix $M$ defining the inner product in $\mathbb{R}^p$ and of the principal components are made simultaneously.

**Key Words and Phrases :** Stochastic approximation, principal component analysis.

**AMS A1991 Subject Classifications : primary :** 62 **; secondary :** H25, L20.

## 1 Estimation récursive en analyse des données

### 1.1 Cadre d'application

Dans (MacGregor 1997), l'auteur fait remarquer que de nos jours, des volumes énormes de données portant sur des variables de procédé, comme par exemple des températures, des pressions, des débits dans des processus industriels, sont collectés en ligne par des ordinateurs chaque heure, chaque minute, chaque seconde et souvent non exploités par manque de méthodes adaptées. Ce cas de données obtenues par une collecte électronique est un cas particulier de flux de données (Aguilar-Ruiz 2006) : on a un flux continu, rapide, illimité de données qui sont lues une fois et pour lesquelles on souhaite faire une analyse en temps réel, sachant que l'on dispose d'une mémoire limitée. Le type de méthode d'estimation récursive étudié dans cet article dans le cadre général de l'approximation stochastique et dans le prolongement des travaux de Benzécri (1969), Krasulina (1970), Lebart (1974), permet de traiter des données qui arrivent en ligne.

En outre, on sait que, par exemple dans l'industrie chimique, le nombre de variables de procédé mesurées peut être de plusieurs milliers. On se trouve alors confronté à un double problème : des données nombreuses, ici arrivant en ligne, et de grande dimension (D'Aubigny 2001). Dans la mesure où l'on utilise des métriques diagonales (que

l'on a à inverser), le type de méthode étudié permet l'utilisation d'un grand nombre de variables.

## 1.2 Modélisation probabiliste

On observe des caractères (ou variables) quantitatifs ou qualitatifs sur n individus. On obtient des vecteurs de mesures ou de valeurs d'indicatrices $z_1, z_2, ..., z_n$ dans $\mathbb{R}^p$. On peut effectuer une analyse de ces données à l'aide de diverses méthodes, dépendant de la nature des données, d'analyse factorielle, de classification, d'analyse discriminante, par exemple. On note ici cette analyse ADO (analyse des données observées).

Supposons maintenant que $z_1, z_2, ..., z_n$ constituent un échantillon i.i.d. d'un vecteur aléatoire $Z$ dans $\mathbb{R}^p$ défini sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. $\Omega$ représente une population d'où on a extrait $n$ individus. On peut alors définir une analyse des données de ce vecteur aléatoire, notée ici ADVA, qui représente l'analyse des données effectuée sur la population, dont on va chercher à estimer les résultats. Soit $\theta$ un résultat de cette ADVA, par exemple une valeur propre, un facteur, un centre de classe, appartenant à un espace $\Theta$.

On peut estimer $\theta$ par un résultat $\theta_n$ de l'ADO faite sur l'échantillon de $Z$, réalisation d'une variable aléatoire $\Theta_n$, estimateur d'ordre $n$ de $\theta$, dont on établit alors les propriétés.

On peut aussi effectuer une estimation récursive, en supposant que l'on dispose d'une suite d'observations i.i.d. $(z_n)$ : à partir de $n-1$ observations $z_1, z_2, ..., z_{n-1}$, on détermine une estimation $\theta_n$ de $\theta$ ($\theta_1$ est quelconque, mais peut être une estimation a priori de $\theta$) ; on introduit alors l'observation $z_n$ et on actualise l'estimation de $\theta$. $\theta_{n+1}$ est déterminée récursivement par une équation du type suivant, où $f_n$ est une fonction de $\Theta \times \mathbb{R}^p$ dans $\Theta$ :

$$\theta_{n+1} = f_n(\theta_n \, ; z_n). \tag{1}$$

Un exemple simple est l'estimation récursive d'une moyenne. Si l'on estime $\theta = E(Z)$ par la moyenne $\overline{z_{n-1}}$ d'un échantillon $(z_1, z_2, ..., z_{n-1})$ de $Z$ et si l'on introduit une nouvelle observation $z_n$, la nouvelle estimation est $\overline{z_n}$, que l'on peut écrire

$$\overline{z_n} = \overline{z_{n-1}} - \frac{1}{n}(\overline{z_{n-1}} - z_n).$$

Un avantage de l'estimation récursive est que l'on n'a pas besoin de conserver les observations $z_1, z_2, ..., z_{n-1}$ ; au pas $n$, il suffit de disposer de l'estimation $\theta_n$ et de $z_n$. Cette méthode est donc particulièrement adaptée aux données qui arrivent de façon séquentielle dans le temps. Elle est aussi adaptée aux données dont l'observation est coûteuse : une fois que l'on a constaté que l'algorithme récursif converge, on peut arrêter l'introduction des données. Elle est également utilisable dans le cas de très grands tableaux de données : ainsi, pour effectuer l'analyse d'un tableau de données observées sur un ensemble d'individus de cardinal $N$ très grand, on peut faire une succession de tirages au hasard dans cet ensemble ; pour chaque individu tiré, l'observation $z$ est alors la réalisation d'un vecteur aléatoire $Z$ défini sur l'ensemble des $N$ individus muni de l'équiprobabilité. L'ADVA de $Z$ n'est autre que l'ADO sur cet ensemble, dont on peut estimer les résultats par des méthodes récursives.

Le processus récursif (1) peut être modifié dans les sens suivants.

Au pas $n$, on peut utiliser au lieu d'une seule observation $z_n$ un paquet de $m_n$ observations. On peut aussi utiliser toutes les observations faites jusqu'au pas $n$, à condition de ne pas avoir à les conserver, ce qui annulerait un avantage de la méthode. Par exemple, si l'on cherche un vecteur propre d'une matrice de covariance d'un vecteur aléatoire, on peut utiliser dans la définition du processus une matrice de covariance empirique, mais calculée de façon récursive, sans avoir à la déterminer à chaque pas en fonction de toutes les observations effectuées jusqu'à ce pas.

Dans les arguments de la fonction $f_n$ peuvent intervenir des éléments aléatoires $y_n$ fonctions des observations $z_1, z_2, ..., z_{n-1}$ et eux-mêmes calculés de façon récursive. On a alors :

$$\begin{aligned} \theta_{n+1} &= f_n(\theta_n \; ; \; z_n \; ; \; y_n) \\ y_n &= g_{n-1}(y_{n-1} \; ; z_{n-1}) \end{aligned} \tag{2}$$

Par exemple, dans l'ACP normée du vecteur aléatoire $Z$, on utilise dans $\mathbb{R}^p$ la métrique $M$ des inverses des variances des composantes de $Z$ ; lorsque l'on dispose de $n$ observations $z_1, z_2, ..., z_n$, on peut l'estimer par la métrique $M_n$ des inverses des variances empiriques calculées à partir de ces observations, elle-même calculée de façon récursive. On effectue ainsi une estimation récursive simultanée de $M$ et de $\theta$.

## 1.3 Cas de l'analyse factorielle multiple

Comme exemple d'application, on présente ici l'estimation des facteurs d'une analyse factorielle multiple (AFM) d'un vecteur aléatoire $Z$. On présente d'abord dans le paragraphe 2 l'analyse en composantes principales (ACP) de $Z$. La présentation classique de cette ACP peut être trouvée dans (Anderson 1958) et une présentation plus générale faisant intervenir un choix de métrique $M$ dans l'espace $\mathbb{R}^p$ des réalisations de $Z$ dans (Caussinus 1984). L'AFM étant une ACP avec un choix particulier de métrique, c'est l'esprit de cette formulation qui est utile ici ; cependant, on en présente d'abord l'aspect géométrique, puis l'interprétation statistique complétée par un résultat utilisé pour expliquer le choix de la métrique en AFM. La présentation de l'AFM qui suit celle de l'ACP est une extension dans un cadre probabiliste de celle faite par Escofier et Pagès (1988) dans le cadre de l'analyse des données classique.

Les facteurs d'une AFM sont vecteurs propres d'une matrice $M^{-1}$-symétrique. L'étude de processus récursifs d'estimation de vecteurs propres de ce type de matrices fait partie de la théorie de l'approximation stochastique, initialisée par les travaux de Robbins et Monro (1951) et qui a connu un très grand développement depuis. Cependant, le premier article consacré à ce problème particulier (Benzécri 1969) ne fait pas référence aux techniques utilisées classiquement en approximation stochastique, contrairement à celui de Krasulina (1970), qui se place dans le cadre des processus de gradient stochastique. Divers processus ont été proposés depuis par différents auteurs. On pourra consulter une bibliographie dans les articles de Bouamaine et Monnez (1997, 1998), qui ont effectué l'étude d'une classe contenant plusieurs de ces processus, aux résultats de laquelle on fera référence ici.

On présente le principe de l'approximation stochastique de vecteurs propres dans le paragraphe 3. On présente l'approximation stochastique des facteurs de l'AFM et

des valeurs propres dans le paragraphe 4, on conclut dans le paragraphe 5 et on donne les démonstrations dans le paragraphe 6.

On note $A'$ la transposée d'une matrice $A$ et $I$ une matrice-identité. L'abréviation *p.s.* signifie "presque sûrement".

# 2 ACP et AFM d'un vecteur aléatoire

## 2.1 ACP d'un vecteur aléatoire

Soit un vecteur aléatoire $Z$ dans $\mathbb{R}^p$, défini sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, de composantes $Z^1, Z^2, ..., Z^P$ de carré intégrable. On note $Covar(Z)$ ou $C$ la matrice de covariance de $Z$.

On munit $\mathbb{R}^p$ d'une métrique $M$ ; on note $\|.\|$ la norme associée : $\|Z(\omega)\|^2 = Z'(\omega)MZ(\omega)$. A partir de $M$ est définie la distance entre deux réalisations $Z(\omega)$ et $Z(\omega')$ de $Z$ qui est la mesure de la différence vis-à-vis de $Z$ entre les éléments, ou individus, $\omega$ et $\omega'$. Le choix de cette métrique est primordial et conditionne les résultats de l'ACP. Nous en verrons un exemple en AFM.

On désigne par $F_r$ un sous-espace affine de $\mathbb{R}^p$ de dimension $r$ auquel appartient l'espérance mathématique $E(Z)$ de $Z$. On note $\Pi Z$ le vecteur aléatoire dans $\mathbb{R}^p$ qui, à tout $\omega \in \Omega$, fait correspondre la projection orthogonale, au sens de la métrique $M$, $\Pi Z(\omega)$ de $Z(\omega)$ sur $F_r$. On a $E(Z) = E(\Pi Z)$ et :

$$E(\|Z - E(Z)\|^2) = E(\|Z - \Pi Z\|^2) + E(\|\Pi Z - E(Z)\|^2).$$

### 2.1.1 Etude géométrique

L'ACP du vecteur aléatoire $Z$ consiste à déterminer un sous-espace $F_r$ qui restitue au mieux en dimension $r$ la dispersion de $Z$ mesurée par $E(\|Z - E(Z)\|^2)$, donc qui soit tel que $E(\|\Pi Z - E(Z)\|^2)$ soit maximale ou $E(\|Z - \Pi Z\|^2)$ minimale. Si l'on note $(u_1, u_2, ..., u_r)$ une base $M$-orthonormée de $F_r$, on a

$$E\left(\|\Pi Z - E(Z)\|^2\right) = \sum_{k=1}^{r} u_k' MCMu_k.$$

Pour $k = 1, 2, ..., r$, on recherche alors un vecteur $u_k$ qui rend maximale la forme quadratique $u'MCMu$ sous les contraintes d'être $M$-unitaire et $M$-orthogonal aux vecteurs $u_j$, $j = 1, ..., k-1$ ; $u_k$ est vecteur propre de la matrice $CM$ associé à la $k^{ième}$ plus grande valeur propre $\lambda_k$ ; on a $u_k' MCMu_k = \lambda_k$ ; l'axe $(E(Z), u_k)$ est appelé le $k^{ième}$ axe principal de l'ACP de $Z$.

### 2.1.2 Interprétation statistique

La formulation statistique, équivalente à la géométrique, est le cadre usuel de présentation de l'ACP d'un vecteur aléatoire. Soit l'élément $a_k = Mu_k$ du dual $\mathbb{R}^{p*}$ de $\mathbb{R}^p$, appelé $k^{ième}$ facteur principal de l'ACP de $Z$. A partir du critère de détermination de $u_k$, on obtient que $a_k$ rend maximale la forme quadratique $a'Ca$ sous les

contraintes $a'Ca_j = 0$, $j = 1, 2, ..., k-1$ et $a'M^{-1}a = 1$ ; la combinaison linéaire des composantes centrées de $Z$, $C_k = a'_k(Z - E(Z))$, appelée $k^{ième}$ composante principale, est donc de variance maximale sous les contraintes d'être non corrélée aux composantes précédentes et que $a_k$ soit $M^{-1}$-unitaire. $a_k$ est vecteur propre associé à la $k^{ième}$ plus grande valeur propre $\lambda_k$ de la matrice $M^{-1}$-symétrique $MC$ et on a $a'_k C a_k = \lambda_k$.

On a en outre le résultat suivant : $v_k = \frac{a_k}{\sqrt{\lambda_k}}$ rend maximale la forme quadratique $v'CMCv$ sous les contraintes $v'Cv_j = 0$, $j = 1, 2, ..., k-1$ et $v'Cv = 1$. En effet, une solution de ce problème est vecteur propre de $C^{-1}CMC = MC$ associé à sa $k^{ième}$ plus grande valeur propre, comme $a_k$ ; mais $a'_k C a_k = \lambda_k$ ; on peut prendre $v_k = \frac{a_k}{\sqrt{\lambda_k}}$.

### 2.1.3 Cas où l'espace fondamental est fini

On constate que, lorsque $\Omega$ est un ensemble fini de cardinal $N$, la probabilité du $i^{ème}$ élément, ou individu, de $\Omega$ étant notée $p_i$, l'ACP du vecteur aléatoire $Z$ est l'ACP classique, ou descriptive, du tableau des réalisations des $p$ variables aléatoires $Z^1, ..., Z^p$ pour les $N$ individus de $\Omega$ affectés respectivement des poids $p_1, ..., p_N$, la métrique dans $\mathbb{R}^p$ étant $M$.

## 2.2 AFM d'un vecteur aléatoire

On suppose maintenant que l'ensemble des composantes du vecteur aléatoire $Z$ dans $\mathbb{R}^p$ est partitionné en $q$ groupes de variables aléatoires réelles

$$\left\{Z^{k1}, Z^{k2}, ..., Z^{km_k}\right\}, k = 1, 2, ..., q \; ; \; m_1 + m_2 + ... + m_q = p.$$

On note $Z^k$ le vecteur aléatoire de composantes $Z^{k1}, Z^{k2}, ..., Z^{km_k}$.

Le but de l'analyse factorielle multiple est de réaliser une ACP de $Z$ dans laquelle les vecteurs aléatoires $Z^k$ aient un rôle équilibré. On veut éviter que les composantes principales de cette ACP soient principalement déterminées à partir de certains vecteurs $Z^k$, donc à partir de certaines composantes de $Z$ au détriment des autres. Ceci est réalisé en faisant un choix particulier de la métrique $M$ dans $\mathbb{R}^p$.

Pour $k = 1, 2, ..., q$, on munit l'espace $\mathbb{R}^{m_k}$ d'une métrique $M^k$. Il paraît alors naturel de prendre dans $\mathbb{R}^p$ la métrique diagonale par blocs

$$M = \begin{pmatrix} M^1 & & & & \\ & M^2 & & & \\ & & . & & \\ & & & . & \\ & & & & M^q \end{pmatrix}.$$

Le carré de la distance entre les points représentatifs de deux individus dans $\mathbb{R}^p$ est alors égal à la somme sur $k$ des carrés des distances entre les points représentatifs de ces individus dans $\mathbb{R}^{m_k}$.

On note $C^k$ la matrice de covariance de $Z^k$, $\lambda_1^k$ la plus grande valeur propre de $M^k C^k$, à laquelle est associé le premier facteur de l'ACP de $Z^k$, et $C$ la matrice de covariance de $Z$ que l'on suppose de plein rang. On a le résultat suivant dont la démonstration est faite dans le paragraphe 6.

**Théorème 1** *Pour un élément $v$ de l'ensemble $A = \{v \in \mathbb{R}^{p*} : v'Cv = 1\}$, on a $v'CMCv \leq \sum_{k=1}^{q} \lambda_1^k$.*

On note $a_1^1$ le premier facteur de l'ACP de $Z^1$, élément de $\mathbb{R}^{m_1*}$, $w_1^1$ le vecteur de $\mathbb{R}^{p*}$ obtenu en complétant $a_1^1$ par des composantes nulles (on plonge $\mathbb{R}^{m_1*}$ dans $\mathbb{R}^{p*}$) et $v_1^1$ le vecteur $w_1^1$ $C$-normé. On a d'après la démonstration du théorème précédent :

$$\lambda_1^1 \leq (v_1^1)'CMCv_1^1 \leq \sum_{k=1}^{q} \lambda_1^k.$$

Si l'on suppose la valeur propre $\lambda_1^1$ très grande par rapport à $\lambda_1^2, ..., \lambda_1^q$, telle que la somme des $\lambda_1^k$ soit peu différente de $\lambda_1^1$, la valeur maximale de $v'CMCv$ est presque atteinte en prenant $v = v_1^1$. Les composantes de $Z^1$ jouent dans ce cas un rôle prépondérant dans la détermination de la première composante principale de l'ACP de $Z$ : c'est ce que l'on veut éviter.

Pour remédier à ceci, on peut prendre dans $\mathbb{R}^p$ la métrique diagonale par blocs

$$M_1 = \begin{pmatrix} \frac{M^1}{\lambda_1^1} & & & & \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & \frac{M^q}{\lambda_1^q} \end{pmatrix}.$$

Pour tout $k$, la plus grande valeur propre de l'ACP de $Z^k$ avec la métrique $\frac{M^k}{\lambda_1^k}$ est alors égale à 1. Aucune plus grande valeur propre n'est prépondérante. L'ACP de $Z$ avec la métrique $M_1$ est appelée analyse factorielle multiple de $Z$.

Pour effectuer l'estimation récursive des facteurs de cette analyse, on effectuera parallèlement l'estimation récursive des valeurs propres $\lambda_1^k$ des ACP des $Z^k$.

# 3 Approximation stochastique de vecteurs propres

On reprend des éléments de l'étude effectuée par Bouamaine et Monnez (1998). On utilise des notations spécifiques à ce paragraphe.

Soit B une matrice $M$-symétrique d'ordre $p$ qui définit un endomorphisme dans $\mathbb{R}^{p*}$. On veut estimer $r$ vecteurs propres correspondant respectivement aux $r$ plus grandes valeurs propres de $B$. On suppose que la matrice $B$ est inconnue, de même que la métrique $M$ dans $\mathbb{R}^{p*}$, mais que l'on dispose au temps $n$ d'une matrice symétrique définie positive $M_n$ et d'une matrice $B_n$ qui vérifient certaines hypothèses.

## 3.1 Définition du processus d'approximation stochastique

On note $< ., . >_n$, respectivement $\|.\|_n$, le produit scalaire, respectivement la norme, induits dans $\mathbb{R}^{p*}$ par la métrique $M_n$. On définit pour $i = 1, ..., r$, le processus récursif

d'approximation stochastique $(X_n^i)$ dans $\mathbb{R}^{p*}$ de type Krasulina tel que, au temps $n$ :

$$\begin{aligned} F_n(X_n^i) &= \frac{\left\langle B_n X_n^i, X_n^i \right\rangle_{n-1}}{\left\langle X_n^i, X_n^i \right\rangle_{n-1}} \\ Y_{n+1}^i &= X_n^i + a_n(B_n - F_n(X_n^i)I)X_n^i \\ X_{n+1}^i &= orth_{M_n}\,(Y_{n+1}^i). \end{aligned} \tag{3}$$

$X_{n+1}^i = orth_{M_n}\,(Y_{n+1}^i)$ signifie que $(X_{n+1}^1, ..., X_{n+1}^i)$ est obtenu en orthogonalisant par rapport à $M_n$ au sens de Gram-Schmidt $(Y_{n+1}^1, ..., Y_{n+1}^i)$.

## 3.2 Discussion des hypothèses

Les pas positifs $a_n$ sont tels que

$$\sum_1^\infty a_n = \infty, \sum_1^\infty a_n^2 < \infty.$$

La suite $(M_n)$ vérifie les hypothèses

$$M_n \longrightarrow M \quad p.s., \sum_1^\infty a_n \left\| M_n - M \right\| < \infty \quad p.s.$$

Par exemple, dans le cas de l'ACP normée, $M$ est la matrice diagonale des variances des composantes de $Z$, $M_n$ la matrice diagonale des variances empiriques calculées à partir des observations faites jusqu'au temps $n$ ; en particulier, pour $a_n = \frac{1}{n}$, on a $\sum_1^\infty a_n E(\|M_n - M\|) < \infty$ ; les hypothèses sur $(M_n)$ sont donc vérifiées.

On suppose que la suite $(B_n)$ vérifie en particulier l'hypothèse suivante, où $T_n$ désigne la tribu du passé au temps $n$ :

$$\sum_1^\infty a_n \left\| E(B_n/T_n) - B \right\| < \infty \quad p.s.$$

Dans l'application à l'analyse factorielle, on a souvent la décomposition multiplicative $B = CD$ (par exemple, dans l'ACP du vecteur aléatoire $Z$, si $N$ désigne la métrique choisie dans $\mathbb{R}^p$, les facteurs de l'ACP sont vecteurs propres de $B = N\ Covar(Z)$) et la décomposition correspondante $B_n = C_n D_n$. Alors :

$$B_n - B = C_n D_n - CD = C_n(D_n - D) + (C_n - C)D.$$

Si $(C_n)$ converge presque sûrement vers $C$ et si

$$\sum_1^\infty a_n \left\| E(D_n/T_n) - D \right\| < \infty, \quad \sum_1^\infty a_n \left\| E(C_n/T_n) - C \right\| < \infty \quad p.s.,$$

alors l'hypothèse sur $(B_n)$ est vérifiée. $C_n$ (ou $D_n$) peut être lui-même obtenu au temps $n-1$ par un processus d'approximation stochastique et être $T_n$-mesurable ; alors, la condition $\sum_1^\infty a_n \left\| C_n - C \right\| < \infty \quad p.s.$ peut être éventuellement vérifiée en utilisant un résultat de comportement asymptotique du processus $(C_n)$. Cette façon de procéder peut être étendue au cas où $B$ est la résultante multiplicative de plus de deux matrices.

## 3.3 Principe de l'étude de la convergence

### 3.3.1 Algèbre extérieure d'ordre $j$ de $\mathbb{R}^{p*}$

On note $\wedge$ le produit extérieur de vecteurs de $\mathbb{R}^{p*}$ et pour $j = 1, ..., p$, ${}^j\wedge\mathbb{R}^{p*}$ l'algèbre extérieure d'ordre $j$ de $\mathbb{R}^{p*}$ : $(e_1, ..., e_p)$ étant une base de $\mathbb{R}^{p*}$, l'ensemble des $C_p^j$ produits extérieurs $e_{i_1}\wedge...\wedge e_{i_j}$ pour $1 \leq i_1 < ... < i_j \leq p$ est une base de ${}^j\wedge\mathbb{R}^{p*}$.

On définit un produit scalaire dans ${}^j\wedge\mathbb{R}^{p*}$ à partir de celui dans $\mathbb{R}^{p*}$ induit par la métrique $M$ ; dans cette définition, $G_j$ est l'ensemble des permutations $\sigma$ de $\{k_1, ..., k_j\}$, $s(\sigma)$ est le nombre d'inversions de la permutation $\sigma$ et $\varepsilon(\sigma) = (-1)^{s(\sigma)}$ :

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

On suppose que les $r$ plus grandes valeurs propres de l'endomorphisme $B$ dans $\mathbb{R}^{p*}$ sont différentes : $\lambda_1 > ... > \lambda_r$. On définit pour $j = 1, ..., r$, l'endomorphisme ${}^{j1}B$ dans ${}^j\wedge\mathbb{R}^{p*}$ par :

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

Si $V^1, ..., V^j$ sont des vecteurs propres de $B$ correspondant respectivement à $\lambda_1, ..., \lambda_j$, $V^1\wedge...\wedge V^j$ est vecteur propre de ${}^{j1}B$ correspondant à la plus grande valeur propre $\lambda_{1j} = \sum_{l=1}^{j} \lambda_l$. On note ${}^jS_1$ le sous-espace propre correspondant à $\lambda_{1j}$ et $({}^jS_1)^\perp$ son supplémentaire orthogonal.

### 3.3.2 Convergence du processus

On effectue la démonstration de la convergence en deux étapes.

Soit ${}^jX_n = X_n^1\wedge...\wedge X_n^j$. On démontre d'abord que, pour $j = 1, ..., r$, $\frac{{}^jX_n}{\|{}^jX_n\|}$ converge $p.s.$ dans un ensemble ${}^jE$ vers $V^1\wedge...\wedge V^j \in {}^jS_1$ (on suppose les vecteurs $V^l$ normés). On démontre ensuite que, pour $l = 1, ..., r$, $\frac{X_n^l}{\|X_n^l\|}$ converge $p.s.$ dans $\cap_{j=1}^{l} {}^jE$ vers $V^l$. On donne ci-dessous la définition de l'ensemble ${}^jE$.

Dans le cas où l'on connaît $B$ et $M$, on définit

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

et le processus $\left({}^jU_n\right)$ dans ${}^j\wedge\mathbb{R}^{p*}$ par

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

Dans ce cas, ${}^jE$ est l'ensemble $\left\{{}^jX_1 \notin \left({}^jS_1\right)^\perp\right\}$ ; $X_1^1, ..., X_1^j$ ne doivent pas être orthogonaux au sous-espace engendré par les vecteurs propres de $B$ correspondant à ses $j$ plus grandes valeurs propres.

Le processus $({}^jX_n) = (X_n^1 \wedge ... \wedge X_n^j)$ peut être considéré comme une perturbation stochastique du processus $\left({}^jU_n\right)$. On note :

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

L'ensemble ${}^jE$ est $\left\{Q^j \notin ({}^jS_1)^\perp\right\}$. On remarque que $Q^j = {}^jX_1$ pour $({}^jX_n) = \left({}^jU_n\right)$.

# 4 Approximation stochastique des facteurs de l'AFM

On note $r$ le nombre de facteurs de l'AFM de $Z$ que l'on veut estimer.

Soit $(Z_1, Z_2, ..., Z_n, ...)$ une suite de vecteurs aléatoires constituant un échantillon i.i.d. de $Z$. $Z_n$ a pour composantes les variables aléatoires réelles $Z_n^{kj}$, $j = 1, 2, ..., m_k, k = 1, 2, ..., q$ ; pour $k$ fixé, les vecteurs aléatoires $Z_n^k$ de composantes $Z_n^{kj}, j = 1, 2, ..., m_k$, constituent un échantillon i.i.d. de $Z^k$. On note $\overline{Z_n^k} = \frac{1}{n}\sum_{i=1}^n Z_i^k$ la moyenne d'ordre $n$ des $Z_i^k$ et $\overline{Z_n}$ celle des $Z_i$.

Soit pour $k = 1, 2, ..., q$, $(M_n^k)$ une suite de matrices symétriques définies positives d'ordre $m_k$.

On considère les hypothèses suivantes.

(H1) *$Z$ admet des moments d'ordre $4r$.*

(H2a) *Pour $k = 1, 2, ..., q$, $M_n^k \longrightarrow M^k$ p.s.*

(H2b) *Pour $k = 1, 2, ..., q$, $\sum_1^\infty a_n \left\|M_{n-1}^k - M^k\right\| < \infty$ p.s.*

(H3) $a_n > 0$ ; $\sum_1^\infty a_n = \infty$ ; $\sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty$ ; $\sum_1^\infty a_n^2 < \infty$.

## 4.1 Définition et convergence d'un processus d'estimation du premier facteur de l'ACP de $\mathbf{Z}^k$

Le premier facteur dans $\mathbb{R}^{m_k *}$ de l'ACP du vecteur aléatoire $Z^k$ dans $\mathbb{R}^{m_k}$ muni de la métrique $M^k$ est vecteur propre associé à la plus grande valeur propre de

$$B^k = M^k C^k = M^k(E(Z^k Z^{k\,\prime}) - E(Z^k)E(Z^{k\,\prime})).$$

On suppose que l'on dispose d'un estimateur convergent $(M_n^k)$ de $M^k$. Par exemple, si $M^k$ est l'inverse de la matrice diagonale des variances des composantes de $Z^k$ (c'est le cas de l'ACP normée), on peut l'estimer à partir de l'échantillon $(Z_1^k, ..., Z_{n-1}^k)$ de $Z^k$ par l'inverse de la matrice diagonale des variances empiriques ; on démontre alors que l'hypothèse H2 est vérifiée (Bouamaine 1996, VI, lemme 1.1.b).

Pour tout $n$, on note $\langle .,. \rangle_n^k$ le produit scalaire dans $\mathbb{R}^{m_k *}$ muni de la métrique $(M_n^k)^{-1}$ et $\|.\|_n^k$ la norme associée. On définit la matrice $B_n^k$ carrée d'ordre $m_k$ :

$$B_n^k = M_{n-1}^k(Z_n^k Z_n^{k\,\prime} - \overline{Z_{n-1}^k}\,\overline{Z_{n-1}^k}'). \tag{4}$$

On définit le processus $(V_n^k)$ dans $\mathbb{R}^{m_k *}$ de type Krasulina (cf (3)) d'estimation d'un vecteur propre associé à la plus grande valeur propre de la matrice $B^k$ :

$$V_{n+1}^k = V_n^k + a_n(B_n^k - \frac{\left\langle B_n^k V_n^k, V_n^k \right\rangle_{n-1}^k}{\langle V_n^k, V_n^k \rangle_{n-1}^k} I)V_n^k. \tag{5}$$

On note $S_1^k$ le sous-espace propre associé à la plus grande valeur propre $\lambda_1^k$ de $B^k$ et $(S_1^k)^\perp$ l'orthogonal de $S_1^k$. $M$ désignant une métrique quelconque dans $\mathbb{R}^{m_k *}$, A un vecteur de $\mathbb{R}^{m_k *}$ et $B$ une matrice carrée d'ordre $m_k$, on note ci-dessous

$$f(A, B, M) = \frac{\langle BA, A \rangle_M}{\langle A, A \rangle_M}.$$

On note également :

$$\begin{aligned} W^k &= V_1^k + \sum_{n=1}^{\infty} \frac{V_{n+1}^k - (I + a_n(C^k - f(V_n^k, C^k, (M^k)^{-1})I))V_n^k}{\prod_{i=1}^n (1 + a_i(\lambda_1^k - f(V_i^k, C^k, (M^k)^{-1})))} \\ E^k &= \left\{ W^k \notin \left(S_1^k\right)^\perp \right\}. \end{aligned}$$

**Lemme 2** *Sous H1, H2, H3, $W^k$ converge presque sûrement, $V_n^k$ converge presque sûrement dans $E^k$ vers un vecteur propre $U_1^k$ de $B^k = M^k C^k$ associé à la plus grande valeur propre $\lambda_1^k$ et l'on a presque sûrement dans $E^k$*

$$\sum_{n=1}^{\infty} a_n \left| \lambda_1^k - \frac{\left\langle B^k V_n^k, V_n^k \right\rangle_{n-1}^k}{\langle V_n^k, V_n^k \rangle_{n-1}^k} \right| < \infty.$$

On détaille les calculs à effectuer à chaque pas.

$$\begin{aligned} B_n^k V_n^k &= M_{n-1}^k (Z_n^k Z_n^{k\,\prime} - \overline{Z_{n-1}^k}\,\overline{Z_{n-1}^k}^{\prime})V_n^k = M_{n-1}^k (Z_n^k \alpha_n^k - \overline{Z_{n-1}^k}\,\beta_n^k) \\ \alpha_n^k &= Z_n^{k\,\prime} V_n^k, \beta_n^k = \overline{Z_{n-1}^k}^{\prime} V_n^k \in \mathbb{R} \\ F_n^k(V_n^k) &= \frac{\left\langle B_n^k V_n^k, V_n^k \right\rangle_{n-1}^k}{\langle V_n^k, V_n^k \rangle_{n-1}^k} = \frac{\left(\alpha_n^k Z_n^{k\,\prime} - \beta_n^k \overline{Z_{n-1}^k}^{\prime}\right) V_n^k}{V_n^{k\,\prime} \left(M_{n-1}^k\right)^{-1} V_n^k} \\ &= \frac{\left(\alpha_n^k\right)^2 - \left(\beta_n^k\right)^2}{V_n^{k\,\prime} \left(M_{n-1}^k\right)^{-1} V_n^k} \\ V_{n+1}^k &= V_n^k + a_n \left( M_{n-1}^k \left( \alpha_n^k Z_n^k - \beta_n^k \overline{Z_{n-1}^k} \right) - F_n^k(V_n^k) V_n^k \right). \end{aligned}$$

A condition que les matrices $M_n^k$ soient diagonales, on peut facilement mettre en oeuvre le processus précédent dans le cas d'un nombre élevé de composantes de $Z$ (cas du fléau de la dimension).

## 4.2 Définition et convergence d'un processus d'estimation de $\mathbf{B}^k$

On définit pour $k = 1, 2, ..., q$ le processus $(W_n^k)$ :

$$W_n^k = M_n^k(\frac{1}{n}\sum_{i=1}^{n} Z_i^k Z_i^{k\,\prime} - \overline{Z_n^k}\,\overline{Z_n^k}\,'). \tag{6}$$

On peut calculer la matrice symétrique d'ordre $m_k$, $\frac{1}{n}\sum_{i=1}^{n} Z_i^k Z_i^{k\,\prime}$, de façon récursive, ainsi que $\overline{Z_n^k}$.

**Lemme 3** *Sous H1, H2, H3, $W_n^k$ converge presque sûrement vers $B^k$ et*

$$\sum_{n=1}^{\infty} a_n \left\| W_{n-1}^k - B^k \right\| < \infty \;\; p.s.$$

## 4.3 Définition et convergence d'un processus d'estimation de $\lambda_1^k$

On définit pour $k = 1, 2, ..., q$ le processus réel $(\Lambda_n^k)$ :

$$\Lambda_n^k = \frac{\left\langle W_{n-1}^k V_n^k, V_n^k \right\rangle_{n-1}^k}{\left\langle V_n^k, V_n^k \right\rangle_{n-1}^k}. \tag{7}$$

**Lemme 4** *Sous H1, H2, H3, pour $k = 1, 2, ..., q$, $\Lambda_n^k$ converge presque sûrement vers $\lambda_1^k$ et*

$$\sum_{n=1}^{\infty} a_n \left| \Lambda_n^k - \lambda_1^k \right| < \infty \;\; p.s. \text{ dans } E^k.$$

## 4.4 Définition et convergence d'un processus d'estimation de $\mathbf{M}_1$

On définit pour tout $n$ la matrice diagonale par blocs symétrique définie positive $M_{1n}$ qui, pour $k = 1, 2, ..., q$ a pour $k^{ième}$ bloc diagonal $\frac{M_n^k}{\Lambda_{n+1}^k}$.

**Lemme 5** *Sous H1, H2, H3, dans $\cap_{k=1}^{q} E^k$, $M_{1n}$ converge presque sûrement vers $M_1$ et*

$$\sum_{n=1}^{\infty} a_n \left\| M_{1,n-1} - M_1 \right\| < \infty \;\; p.s.$$

## 4.5 Définition et convergence de processus d'estimation des facteurs de l'AFM de Z

### 4.5.1 Définition des processus

Les facteurs dans $\mathbb{R}^{p*}$ de l'AFM de $Z$ sont vecteurs propres associés respectivement aux valeurs propres rangées par ordre décroissant de la matrice

$$B = M_1 Covar(Z) = M_1(E(ZZ') - E(Z)E(Z')).$$

$Z_n$ étant le $n^{ième}$ élément de l'échantillon i.i.d. de $Z$, on définit la matrice $B_n$ carrée d'ordre $p$ :

$$B_n = M_{1,n-1}(Z_n Z_n' - \overline{Z_{n-1}}\,\overline{Z_{n-1}}').$$

On note $\langle .,. \rangle_n$ le produit scalaire dans $\mathbb{R}^{p*}$ muni de la métrique $M_{1n}^{-1}$. On définit alors, pour $i = 1, 2, ..., r$, les processus $(X_n^i)$ dans $\mathbb{R}^{p*}$ de type Krasulina d'estimation des $r$ premiers facteurs de l'AFM de $Z$ (cf (3)) :

$$\begin{aligned} F_n(X_n^i) &= \frac{\langle B_n X_n^i, X_n^i \rangle_{n-1}}{\langle X_n^i, X_n^i \rangle_{n-1}} \\ Y_{n+1}^i &= X_n^i + a_n(B_n - F_n(X_n^i)I)X_n^i \\ X_{n+1}^i &= orth_{M_{1n}^{-1}}\, Y_{n+1}^i. \end{aligned} \tag{8}$$

D'après l'étude faite dans le paragraphe 4.1, à condition que les matrices $M_{1n}$ soient diagonales, on peut facilement mettre en oeuvre les processus précédents dans le cas d'un nombre élevé de composantes de $Z$.

### 4.5.2 Définition des événements ${}^jE$

On reprend les définitions et les notations du paragraphe 3.3 (ici la métrique $M$ est $M_1^{-1}$).

On note ${}^jS_1$ le sous-espace propre associé à la plus grande valeur propre $\lambda_{1j}$ de ${}^{j1}B$, somme des $j$ plus grandes valeurs propres de $B$.

On note $< .,. >$ le produit scalaire dans ${}^j\wedge\mathbb{R}^{p*}$ défini à partir du produit scalaire induit par la métrique $M_1^{-1}$ dans $\mathbb{R}^{p*}$. On définit :

$$\begin{aligned} h_j(x^1, ..., x^j) &= \frac{\langle {}^{j1}B\, {}^jx, {}^jx \rangle}{\langle {}^jx, {}^jx \rangle},\ {}^jx = x^1 \Lambda ... \Lambda x^j \\ \Delta_{nj} &= 1 + a_n(\lambda_{1j} - h_j(X_n^1, ..., X_n^j)) \\ Q^j &= {}^jX_1 + \sum_{n=1}^{\infty} \frac{{}^jX_{n+1} - (I + a_n({}^{j1}B - h_j(X_n^1, ..., X_n^j)I))\, {}^jX_n}{\prod_{i=1}^n \Delta_{ij}} \\ {}^jE &= \left\{ Q^j \notin ({}^jS_1)^{\perp} \right\}. \end{aligned}$$

### 4.5.3 Convergence des processus

**Théorème 6** *Sous H1, H2, H3, pour $j = 1, ..., r$, $Q^j$ converge presque sûrement et, si l'on suppose les $r$ plus grandes valeurs propres de $B$ distinctes, alors, pour*

*$i = 1, ..., r$, $X_n^i$ converge presque sûrement dans $\cap_{j=1}^{i} {}^{j}E \cap_{k=1}^{q} E^k$ vers un vecteur propre de $B$ associé à sa $i^{ème}$ plus grande valeur propre.*

On applique le théorème 4 de (Bouamaine et Monnez 1998, p. 24) : l'écriture de ${}^{J}E$ est équivalente à celle donnée pour ce théorème ; d'après le lemme 5, l'hypothèse H4' du théorème 4 est vérifiée dans $\cap_{k=1}^{q} E^k$ ; pour vérifier les hypothèses H1a, c, on utilise, comme pour établir le lemme 2, la démonstration du corollaire 3 de (Monnez 1994, pp. 52-53) ; l'hypothèse H5 est vérifiée pour la suite $(a_n)$. On peut alors appliquer ce théorème dont les conclusions 1 et 3 donnent l'énoncé ci-dessus.

## 4.6 Définition et convergence de processus d'estimation des valeurs propres de l'AFM de Z

On définit, pour $i = 1, ..., r$, les processus $(\Lambda_n^{1i})$ dans $\mathbb{R}$ :

$$\Lambda_{n+1}^{1i} = \Lambda_n^{1i} - a_n(\Lambda_n^{1i} - F_n(X_n^i)). \tag{9}$$

**Théorème 7** *Sous H1, H2, H3, pour $i = 1, ..., r$, $\Lambda_n^{1i}$ converge presque sûrement dans $\cap_{j=1}^{i} {}^{j}E \cap_{k=1}^{q} E^k$ vers la $i^{ème}$ plus grande valeur propre de $B$.*

Ceci est une application directe du théorème 6 de (Bouamaine et Monnez 1998, p. 26).

## 4.7 Cas où l'on utilise à chaque pas toutes les données observées

On suppose dans ce paragraphe que l'on dispose du tableau de toutes les données observées à partir duquel on détermine les métriques $M^k$, les matrices de covariance $C^k$, les matrices $B^k = M^k C^k$, la matrice de covariance $C$ que l'on utilise alors directement dans les processus précédemment définis. La méthode utilisée devient dans ce cas une méthode itérative de détermination des facteurs de l'AFM du tableau des données.

On note $< .,. >^k$ le produit scalaire dans $\mathbb{R}^{m_k *}$ muni de la métrique $(M^k)^{-1}$. On construit les processus $(V_n^k), (\Lambda_n^k), k = 1, ..., r, (M_{1n}), (B_n), (Y_n^i), \left(X_n^i\right), (\Lambda_n^{1i}), i = 1, ..., r$, tels que $M_{1n}$ ait pour $k^{ième}$ bloc diagonal $\frac{M^k}{\Lambda_{n+1}^k}$ et, $< .,. >_n$ désignant le produit scalaire dans $\mathbb{R}^{p*}$ muni de la métrique $M_{1n}^{-1}$, on ait :

$$\begin{aligned}
V_{n+1}^k &= V_n^k + a_n(B^k - \frac{\left\langle B^k V_n^k, V_n^k \right\rangle^k}{\left\langle V_n^k, V_n^k \right\rangle^k} I) V_n^k \\
\Lambda_n^k &= \frac{\left\langle B^k V_n^k, V_n^k \right\rangle^k}{\left\langle V_n^k, V_n^k \right\rangle^k} \\
B_n &= M_{1,n-1} C \\
F_n(X_n^i) &= \frac{\left\langle B_n X_n^i, X_n^i \right\rangle_{n-1}}{\left\langle X_n^i, X_n^i \right\rangle_{n-1}} \\
Y_{n+1}^i &= X_n^i + a_n(B_n - F_n(X_n^i) I) X_n^i \\
X_{n+1}^i &= orth_{M_{1n}^{-1}} Y_{n+1}^i \\
\Lambda_{n+1}^{1i} &= \Lambda_n^{1i} - a_n(\Lambda_n^{1i} - F_n(X_n^i)).
\end{aligned}$$

Ces processus peuvent être utilisés dans une phase d'initialisation : ayant effectué un certain nombre d'observations de $Z$, on estime les facteurs de l'AFM du tableau de données constitué à partir de ces observations ; on utilise ces facteurs estimés pour initialiser les processus d'estimation des facteurs de l'AFM de $Z$.

# 5 Conclusion

Le type de méthode étudié dans cet article peut permettre de traiter des données en ligne ou coûteuses ou nombreuses ou de grande dimension.

On a traité comme exemple l'estimation récursive des facteurs d'une analyse factorielle multiple. D'autres exemples d'application de l'estimation récursive en analyse des données ont été étudiés dans (Bouamaine 1996) ; les cas de l'ACP projetée et de l'ACP partielle sont traités dans (Monnez 1998, 2002). Remarquons que rentre dans le cadre de l'estimation récursive la méthode des k-means de MacQueen (1967) utilisée en classification, que l'on peut interpréter comme un processus de gradient stochastique (Bouamaine 1996) et dont on peut étudier la convergence par les techniques correspondantes.

Nous ne disposons pas actuellement de résultats théoriques concernant la rapidité de convergence des processus définis. Cependant, des études empiriques ont été effectuées (Lebart et al. 1977, Bouamaine 1996). Dans cette dernière référence, l'auteur a comparé la rapidité de convergence de plusieurs processus d'estimation récursive des facteurs d'une ACP. L'implémentation des algorithmes étudiés dans cet article est en cours de réalisation ainsi que des simulations. Comme il est indiqué dans le paragraphe 4.7, on estime d'abord les facteurs de l'AFM d'un tableau de données constitué à partir d'un nombre restreint d'observations de $Z$, qui vont servir à l'initialisation des algorithmes d'estimation des facteurs de l'AFM de $Z$.

# 6 Démonstrations

## 6.1 Démonstration du théorème 1

Pour simplifier l'écriture, on suppose les variables aléatoires $Z^{kj}$ centrées. Alors :

$$C = E(ZZ') = \begin{pmatrix} E(ZZ^{1\,\prime}) & . & . & . & E(ZZ^{q\,\prime}) \end{pmatrix} = \begin{pmatrix} E(Z^1Z') \\ . \\ . \\ . \\ E(Z^qZ') \end{pmatrix}$$

$$CMC = \sum_{k=1}^{q} E(ZZ^{k\,\prime})M^kE(Z^kZ').$$

Pour $v \in A = \{v \in \mathbb{R}^{p*} : v'Cv = 1\}$, on a :

$$v'CMCv \leq \sum_{k=1}^{q} \max_A v'E(ZZ^{k\,\prime})M^kE(Z^kZ')v.$$

La valeur maximale dans $A$ de la forme quadratique $v'E(ZZ^{k\,\prime})M^kE(Z^kZ')v$ est la plus grande valeur propre de la matrice $C^{-1}E(ZZ^{k\,\prime})M^kE(Z^kZ')$ qui est de rang $m_k$.

On considère alors une valeur propre $\mu$ de la matrice carrée d'ordre $m_k$, $M^kC^k = M^kE(Z^kZ^{k\,\prime})$ et un vecteur propre $u \in \mathbb{R}^{m_k*}$ associé. Soit $w$ le vecteur de $\mathbb{R}^{p*}$ tel que $w' = (0...u'...0)$, les composantes de $u$ occupant les positions $\sum_{j=1}^{k-1} m_j + 1$ à $\sum_{j=1}^{k} m_j$. On a alors $Z'w = Z^{k\,\prime}u$ ; en outre :

$$\begin{aligned} C^{-1}E(ZZ^{k\,\prime})M^kE(Z^kZ')w &= C^{-1}E(ZZ^{k\,\prime})M^kE(Z^kZ^{k\,\prime})u \\ &= \mu C^{-1}E(ZZ^{k\,\prime})u = \mu C^{-1}E(ZZ')w = \mu w. \end{aligned}$$

Donc, la matrice $C^{-1}E(ZZ^{k\,\prime})M^kE(Z^kZ')$ de rang $m_k$ admet pour valeurs propres non nulles celles de $M^kC^k$ ; sa plus grande valeur propre est $\lambda_1^k$. Par conséquent :

$$\max_A \; v'CMCv \leq \sum_{k=1}^{q} \lambda_1^k.$$

On a également :

$$\begin{aligned} w'CMCw &= \sum_{l=1}^{q} w'E(ZZ^{l\,\prime})M^lE(Z^lZ')w \\ &\geq w'E(ZZ^{k\,\prime})M^kE(Z^kZ')w = \mu w'Cw. \end{aligned}$$

Pour $w'Cw = 1$, on a $w'CMCw \geq \mu$.

## 6.2 Démonstration du lemme 2

On note $T_n$ la tribu engendrée par

$$V_1^k, k = 1, 2, ..., q, Z_1, ..., Z_{n-1}, M_1^k, ..., M_{n-1}^k, k = 1, ..., q.$$

On démontre que, sous H1, H2, H3 (cf Monnez 1994, pp. 52-53) :

$$\begin{aligned} \sum_{n=1}^{\infty} a_n \left\| E(B_n^k/T_n) - B^k \right\| &< \infty \; p.s. \\ \sum_{n=1}^{\infty} a_n^2 E(\left\| B_n^k - B^k \right\|^2 / T_n) &< \infty \; p.s. \end{aligned}$$

On applique alors le théorème 4 de (Bouamaine et Monnez 1998, p. 24) : $W^k$ converge presque sûrement, $V_n^k$ converge presque sûrement dans $E^k$ vers un vecteur propre de $B^k$ associé à sa plus grande valeur propre $\lambda_1^k$ et on a presque sûrement dans $E^k$

$$\sum_{n=1}^{\infty} a_n (\lambda_1^k - \frac{\left\langle B^kV_n^k, V_n^k \right\rangle^k}{\left\langle V_n^k, V_n^k \right\rangle^k}) < \infty,$$

$\langle ., . \rangle^k$ désignant le produit scalaire dans $\mathbb{R}^{m_k*}$ muni de la métrique $(M^k)^{-1}$ et $\|.\|^k$ la norme associée.

Soit

$$\begin{aligned}
A_n^k &= \frac{\left\langle B^k V_n^k, V_n^k \right\rangle^k}{\langle V_n^k, V_n^k \rangle^k} - \frac{\left\langle B^k V_n^k, V_n^k \right\rangle_{n-1}^k}{\langle V_n^k, V_n^k \rangle_{n-1}^k} \\
&= \frac{\left\langle B^k V_n^k, V_n^k \right\rangle^k (V_n^k)'((M_{n-1}^k)^{-1} - (M^k)^{-1}) V_n^k}{\langle V_n^k, V_n^k \rangle^k \langle V_n^k, V_n^k \rangle_{n-1}^k} \\
&\quad + \frac{(V_n^k)' B^k ((M^k)^{-1} - (M_{n-1}^k)^{-1}) V_n^k}{\langle V_n^k, V_n^k \rangle_{n-1}^k}.
\end{aligned}$$

Sous l'hypothèse H2, $\sum_{n=1}^{\infty} a_n \left\| (M_{n-1}^k)^{-1} - (M^k)^{-1} \right\| < \infty \ p.s.$ (Bouamaine 1996, X, corollaire e). On en déduit que $\sum_{n=1}^{\infty} a_n \left| A_n^k \right| < \infty \ p.s.$ et on obtient la conclusion du lemme.

## 6.3 Démonstration du lemme 3

Sous H1 et H2, $W_n^k$ converge presque sûrement vers $B^k$.

$$\begin{aligned}
W_{n-1}^k - B^k &= M_{n-1}^k \Big( \frac{1}{n-1} \sum_{i=1}^{n-1} Z_i^k Z_i^{k\,\prime} - E(Z^k Z^{k\,\prime}) \Big) \\
&\quad + (M_{n-1}^k - M^k)(E(Z^k Z^{k\,\prime}) - E(Z^k) E(Z^{k\,\prime})) \\
&\quad - M_{n-1}^k (\overline{Z_{n-1}^k} - E(Z^k))(\overline{Z_{n-1}^k})' \\
&\quad + M_{n-1}^k E(Z^k)(E(Z^{k\,\prime}) - (\overline{Z_{n-1}^k})').
\end{aligned}$$

Sous H1 et H3, on a :

$$\begin{aligned}
E\Big(\sum_{n=1}^{\infty} a_n \left\| \overline{Z_{n-1}^k} - E(Z^k) \right\| \Big) &\leq \sum_{n=1}^{\infty} a_n E\Big( \left\| \overline{Z_{n-1}^k} - E(Z^k) \right\|^2 \Big)^{\frac{1}{2}} \\
&\leq \sum_{n=1}^{\infty} a_n O\Big(\frac{1}{\sqrt{n}}\Big) < \infty. \\
\sum_{n=1}^{\infty} a_n \left\| \overline{Z_{n-1}^k} - E(Z^k) \right\| &< \infty \ p.s.
\end{aligned}$$

De même :

$$\sum_{n=1}^{\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} Z_i^k Z_i^{k\,\prime} - E(Z^k Z^{k\,\prime}) \right\| < \infty \ p.s.$$

Sous H2, on a alors $\sum_{n=1}^{\infty} a_n \left\| W_{n-1}^k - B^k \right\| < \infty \ p.s.$

## 6.4 Démonstration du lemme 4

D'après les lemmes 2 et 3, $W_n^k$ converge $p.s.$ vers $B^k$ et $\frac{V_n^k}{\|V_n^k\|^k}$ converge $p.s.$ dans $E^k$ vers $U_1^k$, vecteur propre $(M^k)^{-1}$-unitaire de $B^k$ associé à $\lambda_1^k$. Alors on a presque

sûrement dans $E^k$ :

$$\Lambda_n^k = \frac{(V_n^k)'(W_{n-1}^k)'(M_{n-1}^k)^{-1}V_n^k}{\langle V_n^k, V_n^k\rangle^k}\frac{\langle V_n^k, V_n^k\rangle^k}{\langle V_n^k, V_n^k\rangle_{n-1}^k} \longrightarrow (U_1^k)'(B^k)'(M^k)^{-1}U_1^k = \lambda_1^k$$

$$\begin{aligned}
\left|\Lambda_n^k - \lambda_1^k\right| &= \left|\frac{\langle W_{n-1}^k V_n^k, V_n^k\rangle_{n-1}^k}{\langle V_n^k, V_n^k\rangle_{n-1}^k} - \lambda_1^k\right| \\
&\leq \left|\frac{\langle (W_{n-1}^k - B^k)V_n^k, V_n^k\rangle_{n-1}^k}{\langle V_n^k, V_n^k\rangle_{n-1}^k}\right| + \left|\frac{\langle B^k V_n^k, V_n^k\rangle_{n-1}^k}{\langle V_n^k, V_n^k\rangle_{n-1}^k} - \lambda_1^k\right| \\
&= D_n^k + F_n^k.
\end{aligned}$$

Comme, d'après les lemmes 2 et 3, $\frac{V_n^k}{\|V_n^k\|^k}$ converge $p.s.$ dans $E^k$ vers $U_1^k$ et $\sum_{n=1}^{\infty} a_n \left\|W_{n-1}^k - B^k\right\| < \infty$ $p.s.$, et comme $M_n^k$ converge $p.s.$ vers $M^k$, on a $\sum_{n=1}^{\infty} a_n D_n^k < \infty$ $p.s.$ dans $E^k$ ; en outre, d'après le lemme 2, $\sum_{n=1}^{\infty} a_n F_n^k < \infty$ $p.s.$ dans $E^k$ ; donc, $\sum_{n=1}^{\infty} a_n \left|\Lambda_n^k - \lambda_1^k\right| < \infty$ $p.s.$ dans $E^k$.

## 6.5 Démonstration du lemme 5

Comme pour tout $k$, $M_n^k \longrightarrow M^k$ et $\Lambda_n^k \longrightarrow \lambda_1^k$ $p.s.$ dans $E^k$, on a : $M_{1n} \longrightarrow M_1$ $p.s.$ dans $\cap_{k=1}^q E^k$.

Dans la suite, on prend pour norme matricielle la norme spectrale. Comme $M_{1n} - M_1$ est diagonale par blocs, on a :

$$\begin{aligned}
\|M_{1,n-1} - M_1\| &= \max_k \left\|\frac{M_{n-1}^k}{\Lambda_n^k} - \frac{M^k}{\lambda_1^k}\right\| \\
&\leq \max_k \left(\frac{1}{\Lambda_n^k}\left\|M_{n-1}^k - M^k\right\| + \left|\frac{1}{\Lambda_n^k} - \frac{1}{\lambda_1^k}\right|\left\|M^k\right\|\right) \\
&\leq \sum_{k=1}^{q}\frac{1}{\Lambda_n^k}\left\|M_{n-1}^k - M^k\right\| + \sum_{k=1}^{q}\frac{\left|\Lambda_n^k - \lambda_1^k\right|}{\Lambda_n^k\lambda_1^k}\left\|M^k\right\|.
\end{aligned}$$

Comme $\Lambda_n^k \longrightarrow \lambda_1^k$ $p.s.$, on a sous H2 : pour $k = 1, ..., q$,

$$\sum_{n=1}^{\infty} a_n \frac{1}{\Lambda_n^k}\left\|M_{n-1}^k - M^k\right\| < \infty \ p.s.$$

En outre, $\sum_{n=1}^{\infty} a_n \left|\Lambda_n^k - \lambda_1^k\right| < \infty$ $p.s.$ dans $E^k$. Donc, $\sum_{n=1}^{\infty} a_n \|M_{1,n-1} - M_1\| < \infty$ $p.s.$ dans $\cap_{k=1}^q E^k$.

## Bibliographie


[1] Aguilar-Ruiz, J.S. (2006), Recent advances in data stream mining, *38$^{èmes}$ Journées de Statistique de la SFDS* (Clamart).

[2] Anderson, T.W. (1958), *An Introduction to Multivariate Statistical Analysis* (Wiley, New-York).

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

[4] Bouamaine, A. (1996), *Méthodes d'approximation stochastique en analyse des données*, thèse de doctorat d'Etat ès Sciences Appliquées (Université Mohammed V, EMI, Rabat).

[5] Bouamaine, A., Monnez, J.M. (1997), Convergence d'une classe de processus d'approximation stochastique de vecteurs propres, *Pub. Inst. Stat. Univ. Paris* **XXXXI**, fasc. 1-2, 97-117.

[6] Bouamaine, A., Monnez, J.M. (1998), Approximation stochastique de vecteurs et valeurs propres, *Pub. Inst. Stat. Univ. Paris* **XXXXII**, fasc. 2-3, 15-38.

[7] Caussinus, H. (1984), *Analyses en composantes principales, quelques réflexions sur la part des modèles probabilistes* (Université Paul Sabatier, Toulouse).

[8] D'Aubigny, G. (2001), Data mining et statistique, discussion et commentaires, *Journal de la Société Française de Statistique* **142**, 1, 37-52.

[9] Escofier, B., Pagès, J. (1988), *Analyses factorielles simples et multiples, objectifs, méthodes et interprétation* (Dunod, Paris).

[10] Krasulina, T.P. (1970), Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices, *Automation and Remote Control* **2**, 215-221.

[11] Lebart, L. (1974), On the Benzécri's method for computing eigenvectors by stochastic approximation (the case of binary data), *Proceedings in Computational Statistics*, Physica Verlag, Vienne, 202-211.

[12] Lebart, L., Morineau, A., Tabard, N. (1977), *Techniques de la description statistique* (Dunod, Paris).

[13] MacGregor, J.F. (1997), Using on-line process data to improve quality : challenges for statisticians, *International Statistical Review* **65**, 3, 309-323.

[14] MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, *Proceedings of the $5^{th}$ Berkeley Sumposium on Probability and Statistics*, 281-297.

[15] Monnez, J.M. (1994), Convergence d'un processus d'approximation stochastique en analyse factorielle, *Pub. Inst. Stat. Univ. Paris* **XXXVIII**, fasc. 1, 37-56.

[16] Monnez, J.M. (1998), Approximation stochastique d'analyses en composantes principales, *CIMASI'98* (Casablanca).

[17] Monnez, J.M. (2002), Méthode séquentielle d'analyse en composantes principales partielle, *CIMASI'2002* (Casablanca).

[18] Robbins, H., Monro, S. (1951), A stochastic approximation method, *Ann. Math. Stat.* **22**, 400-407.

Jean-Marie Monnez
Université Henri Poincaré
Institut Elie Cartan - Laboratoire de Mathématiques
BP 239 - F 54506 - Vandoeuvre lès Nancy Cedex
*monnez@iecn.u-nancy.fr*