Analyse en composantes principales d'un flux de données d'espérance variable dans le temps

Jean-Marie Monnez

Institut Elie Cartan UMR 7502 – Laboratoire de Mathématiques
Nancy-Université, CNRS, INRIA
BP 239 – F 54506 – Vandoeuvre-lès-Nancy Cedex
jean-marie.monnez@iecn.u-nancy.fr

Résumé. On considère un flux de données représenté par une suite de vecteurs de données. On suppose que chaque vecteur de données est une réalisation d'un vecteur aléatoire dont l'espérance mathématique varie dans le temps selon un modèle linéaire pour chacune des composantes. On utilise des processus d'approximation stochastique pour estimer en ligne les paramètres des modèles linéaires et en même temps les facteurs de l'ACP du vecteur aléatoire.

1 Modèle de flux et plan d'étude

Soit un flux de données, représenté par une suite de vecteurs $(z_1, \ldots, z_n, \ldots)$ dans $\mathbb{R}^p$.

1.1 Modèle d'étude et formulation

On suppose que :

- pour tout n, z_n est la réalisation d'un vecteur aléatoire Z_n, d'espérance mathématique variable dans le temps ;
- les vecteurs aléatoires Z_n sont mutuellement indépendants ;
- pour tout n, on a la décomposition $Z_n = \theta_n + R_n$, $\theta_n = (\theta_n^1 \ldots \theta_n^p)'$ étant un vecteur de $\mathbb{R}^p$, la loi du vecteur aléatoire R_n ne dépendant pas de n, $E[R_n] = 0, Covar[R_n] = \Sigma$ (matrice de covariance de R_n) ; ceci revient à supposer que les $R_n = Z_n - \theta_n$ constituent un échantillon i.i.d. d'un vecteur aléatoire R dans $\mathbb{R}^p$ tel que $E[R] = 0, Covar[R] = \Sigma$; on a alors $E[Z_n] = \theta_n, Covar[Z_n] = \Sigma$; $r_n = z_n - E[Z_n]$ représente la donnée z_n centrée ;
- pour $i = 1, \ldots, p$, il existe un vecteur β^i de $\mathbb{R}^{n_i}$ inconnu et , pour tout n, un vecteur U_n^i de $\mathbb{R}^{n_i}$ connu au temps n tels que $\theta_n^i = \langle \beta^i, U_n^i \rangle$, $\langle ., . \rangle$ dé signant le produit scalaire euclidien usuel dans $\mathbb{R}^{n_i}$; U_n^i peut être un vecteur de fonctions connues du temps (θ_n^i est alors une combinaison linéaire de fonctions connues du temps) ou un vecteur de valeurs de variables explicatives contrôlées ; si l'on note Z_n^i, respectivement R_n^i, la $i^{ème}$ composante de Z_n, respectivement R_n, on a alors le modèle de régression linéaire

$$Z_n^i = \langle \beta^i, U_n^i \rangle + R_n^i, \quad i = 1, \ldots, p.$$

On pose le problème suivant : réaliser une analyse en composantes principales (ACP) du vecteur aléatoire R dans $\mathbb{R}^p$ que l'on munit d'une métrique M ; on effectue ainsi une ACP de données corrigées de l'effet du temps ou d'effets explicatifs. On étudie ici l'estimation des facteurs de cette analyse.

On rappelle dans le paragraphe 2 une présentation de l'ACP d'un vecteur aléatoire (Monnez, 2006).

1.2 Principe et plan de l'étude

Les facteurs de l'ACP du vecteur aléatoire R sont vecteurs propres de la matrice $C = M\Sigma$ associés aux valeurs propres rangées par ordre décroissant. On va en effectuer une estimation en ligne, en actualisant au temps n, après avoir introduit l'observation z_n, l'estimation d'un facteur obtenue au temps $n-1$. On fait cette estimation en parallèle avec celle des paramètres β^i, donc des composantes θ_n^i de θ_n. On utilise pour cela des processus d'approximation stochastique de la famille de ceux de Robbins et Monro (1951), Benzécri (1969) et Krasulina (1970).

On définit les processus dans le paragraphe 3, on donne les théorèmes de convergence presque sûre dans le paragraphe 4, on étudie le cas particulier où l'on prend pour métrique M celle de l'ACP normée dans le paragraphe 5, on donne une conclusion dans le paragraphe 6 et les démonstrations dans le paragraphe 7.

2 ACP d'un vecteur aléatoire

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

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

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

$$E\left[\|R - E[R]\|^2\right] = E\left[\|R - \Pi R\|^2\right] + E\left[\|\Pi R - E[R]\|^2\right].$$

2.1 Etude géométrique

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

$$E\left[\|\Pi R - E[R]\|^2\right] = \sum_{k=1}^{r} u_k' M\Sigma M u_k.$$

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

2.2 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 R. À partir du critère de détermination de u_k, on obtient que a_k rend maximale la forme quadratique $a'\Sigma a$ sous les contraintes $a'\Sigma a_j = 0$, $j = 1, 2, \ldots, k-1$ et $a'M^{-1}a = 1$; la combinaison linéaire des composantes centrées de R, $C_k = a_k'(R - E[R])$, 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 λ_k de la matrice M^{-1}-symétrique $M\Sigma$ et on a $a_k'\Sigma a_k = \lambda_k$.

3 Définition des processus d'approximation stochastique

Soit M_n un estimateur de M au temps n.
On note $\langle ., . \rangle_n$ le produit scalaire dans le dual $\mathbb{R}^{p*}$ de $\mathbb{R}^p$ au sens de la métrique M_n^{-1}.
Soit (a_n) une suite de nombres réels positifs.
Pour $i = 1, \ldots, p$, on définit le processus d'approximation stochastique (B_n^i) de β^i tel que

$$B_{n+1}^i = B_n^i - a_n U_n^i (U_n^{i\prime} B_n^i - Z_n^i).$$

On définit :

$$\begin{aligned} \widehat{\Theta}_n^i &= \left\langle B_n^i, U_n^i \right\rangle, \widehat{\Theta}_n = (\widehat{\Theta}_n^1 \ldots \widehat{\Theta}_n^p)'; \\ C_n &= M_{n-1}(Z_n Z_n' - \widehat{\Theta}_n \widehat{\Theta}_n'). \end{aligned}$$

Soit r le nombre de facteurs à estimer. Pour $i = 1, \ldots, r$, on définit les processus (X_n^i) d'estimation des facteurs tels que :

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

I désigne la matrice-identité d'ordre p. $X_{n+1}^i = orth_{M_n^{-1}}(Y_{n+1}^i)$ signifie que $(X_{n+1}^1, \ldots, X_{n+1}^i)$ est obtenu en orthogonalisant par rapport à M_n^{-1} au sens de Gram-Schmidt $(Y_{n+1}^1, \ldots, Y_{n+1}^i)$.

4 Théorèmes de convergence presque sûre

4.1 Principe de l'étude de la convergence des processus d'estimation des facteurs

L'étude de la convergence du type précédent de processus a été faite par Bouamaine et Monnez (1997, 1998) en se plaçant dans l'algèbre extérieure d'ordre j de $\mathbb{R}^{p*}$. On rappelle d'abord quelques éléments théoriques relatifs à cette algèbre.

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

On note $\wedge$ le produit extérieur de vecteurs de $\mathbb{R}^{p*}$ et pour $j = 1, \ldots, p$, ${}^j\wedge\mathbb{R}^{p*}$ l'algèbre extérieure d'ordre j de $\mathbb{R}^{p*}$: $(e_1, \ldots, e_p)$ étant une base de $\mathbb{R}^{p*}$, l'ensemble des C_p^j produits extérieurs $e_{i_1}\wedge\ldots\wedge e_{i_j}$ pour $1 \leq i_1 < \ldots < i_j \leq p$ 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^{-1} ; dans cette définition, G_j est l'ensemble des permutations σ de $\{k_1, \ldots, k_j\}$, $s(\sigma)$ est le nombre d'inversions de la permutation σ et $\varepsilon(\sigma) = (-1)^{s(\sigma)}$:

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

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

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

Si $V^1, \ldots, V^j$ sont des vecteurs propres de C correspondant respectivement à $\lambda_1, \ldots, \lambda_j$, $V^1\wedge\ldots\wedge V^j$ est vecteur propre de ${}^{j1}C$ 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 λ_{1j} et $({}^jS_1)^\perp$ son supplémentaire orthogonal.

4.1.2 Convergence des processus

On effectue la démonstration de la convergence en deux étapes.

Soit le processus $({}^jX_n)$ dans l'algèbre extérieure d'ordre j de $\mathbb{R}^{p*}$ défini par : ${}^jX_n = X_n^1\wedge\ldots\wedge X_n^j$.

On démontre d'abord que, pour $j = 1, \ldots, r$, $\frac{{}^jX_n}{\|{}^jX_n\|}$ converge $p.s.$ dans un ensemble ${}^jE$ vers $V^1\wedge\ldots\wedge V^j \in {}^jS_1$ (on suppose les vecteurs V^l normés). On démontre ensuite que, pour $l = 1, \ldots, r$, $\frac{X_n^l}{\|X_n^l\|}$ 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 C et M^{-1}, on définit

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

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

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

Dans ce cas, ${}^jE$ est l'ensemble $\left\{{}^jX_1 \notin \left({}^jS_1\right)^{\perp}\right\}$; $X_1^1, \ldots, X_1^j$ 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 $({}^jX_n) = (X_n^1 \wedge \ldots \wedge X_n^j)$ peut être considéré comme une perturbation stochastique du processus $\left({}^jU_n\right)$. On note :

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

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

4.2 Hypothèses

On fait les hypothèses suivantes.

(H1) (a) $\max_i \sup_n \left\|U_n^i\right\| < \infty$.

(b) *Pour $i = 1, \ldots, p$, il existe un entier r_i, un réel $\lambda_i > 0$, une suite croissante d'entiers $(n_{il}, l \geq 1)$ tels que $n_{i1} = 1, n_{i,l+1} \leq n_{il} + r_i$, $\lambda_{\min}(\sum_{j\in I_{il}} U_j^i U_j^{i'}) \geq \lambda_i$, avec $I_{il} = \{n_{il}, \ldots, n_{i,l+1} - 1\}$.*

(H2) $M_n \to M \quad p.s.$

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

(H3) $a_n = \frac{c}{n^{\alpha}}, \quad c > 0, \quad \frac{1}{2} < \alpha \leq 1.$

(H3') $a_n > 0, \quad \sum_1^{\infty} \min_{j\in I_l} a_j = \infty, \quad \sum_1^{\infty} a_n^2 < \infty.$

4.3 Théorèmes

On a les énoncés suivants.

Théorème 1. *Sous H1 et H3', pour $i = 1, \ldots, p$, $B_n^i \to \beta^i$ et $\widehat{\Theta}_n^i - \theta_n^i \to 0 \quad p.s.$*

Théorème 2. *Sous H1 et H3, pour $i = 1, \ldots, p$:*

1. *pour $\frac{1}{2} < \alpha \leq 1$ ou $(\alpha = 1$ et $\frac{2\lambda_i c}{r_i} > 1)$:*

$$\overline{\lim}\ n^{\alpha} E\left[\left\|B_n^i - \beta^i\right\|^2\right] < \infty \text{ et } \overline{\lim}\ n^{\alpha} E\left[\left\|\widehat{\Theta}_n^i - \theta_n^i\right\|^2\right] < \infty\ ;$$

2. *pour* $\alpha = 1$ *et* $\frac{2\lambda_i c}{r_i} = 1$ *:*

$$\overline{\lim}\, \frac{n}{\ln n} E\left[\left\|B_n^i - \beta^i\right\|^2\right] < \infty \; et \; \overline{\lim}\, \frac{n}{\ln n} E\left[\left\|\widehat{\Theta}_n^i - \theta_n^i\right\|^2\right] < \infty \; ;$$

3. *pour* $\alpha = 1$ *et* $\frac{2\lambda_i c}{r_i} < 1$ *:*

$$\overline{\lim}\, n^{\frac{2\lambda_i c}{r_i}} E\left[\left\|B_n^i - \beta^i\right\|^2\right] < \infty \; et \; \overline{\lim}\, n^{\frac{2\lambda_i c}{r_i}} E\left[\left\|\widehat{\Theta}_n^i - \theta_n^i\right\|^2\right] < \infty.$$

Théorème 3. *Sous H1, H2, H3, si l'on suppose que* R *admet des moments d'ordre* $4r$*, pour*$j = 1, \ldots, r$*,* Q^j *converge presque sûrement et, si l'on suppose que les* r *plus grandes valeurs propres de* $C = M\Sigma$ *sont distinctes, alors, pour* $i = 1, \ldots, r$*,* X_n^i *converge presque sûrement dans* $\cap_{j=1}^i {}^j E$ *vers un vecteur propre de* C *associé à sa* $i^{ème}$ *plus grande valeur propre.*

5 Cas particulier de la métrique de l'ACP normée

Soit $\left(\sigma^i\right)^2 = Var\left[R^i\right]$, $i = 1, \ldots, p$. La métrique de l'ACP normée dans $\mathbb{R}^p$ est la métrique diagonale M des $\frac{1}{(\sigma^i)^2}$.

On considère l'estimateur $\frac{1}{M_n^i} = \frac{1}{n}\sum_{j=1}^n \left(Z_j^i - \widehat{\Theta}_j^i\right)^2$ de $\left(\sigma^i\right)^2$ et la métrique diagonale M_n dans $\mathbb{R}^p$ des M_n^i. On peut calculer $\frac{1}{M_n^i}$ de façon récursive.

Théorème 4. *On suppose que* R *admet des moments d'ordre* 4*. Alors, sous H1 et H3 avec* $\alpha = 1$*,* $M_n \to M$ *et* $\sum_1^\infty a_n \left\|M_n - M\right\| < \infty \; p.s.$ *; donc, l'hypothèse H2 est vérifiée.*

6 Conclusion et extensions

On a traité ici l'estimation en ligne des facteurs de l'ACP d'un flux de données. On peut également estimer en ligne les valeurs propres associées (Bouamaine et Monnez, 1998), les corrélations entre les variables et les facteurs ; on peut aussi estimer en ligne la valeur d'un facteur pour un individu et procéder éventuellement à une classification des individus.

Cette étude introductive sera développée dans les directions suivantes :

1. étude d'un modèle où l'espérance et la variance varient dans le temps ;
2. étude de modèles non linéaires de variation des paramètres ;
3. autres choix de métriques M ;
4. application à d'autres méthodes d'analyse factorielle.

7 Démonstrations

7.1 Démonstration du théorème 1

On a défini : $B_{n+1}^i = B_n^i - a_n U_n^i (U_n^{i\prime} B_n^i - Z_n^i)$.
Comme $Z_n^i = \theta_n^i + R_n^i = U_n^{i\prime}\beta^i + R_n^i$, on a :

$$B_{n+1}^i - \beta^i = B_n^i - \beta^i - a_n U_n^i U_n^{i\prime}(B_n^i - \beta^i) + a_n U_n^i R_n^i.$$

A i fixé, notons : $Y_n = B_n^i - \beta^i, V_n = U_n^i, S_n = R_n^i, n_l = n_{il}, r = r_i$. On a :

$$Y_{n+1} = Y_n - a_n V_n V_n' Y_n + a_n V_n S_n.$$

Pour établir la convergence presque sûre, on utilise le lemme suivant (Robbins et Siegmund, 1971).

Lemme 5. *Soit (Ω, A, P) un espace probabilisé, (T_n) une suite croissante de sous-tribus de A. Soit, pour tout n, $\alpha_n, \beta_n, \gamma_n$ des variables aléatoires réelles T_n - mesurables, non négatives, intégrables, telles que*

$$\begin{aligned} E\left[\alpha_{n+1} \mid T_n\right] &\leq \alpha_n(1+\beta_n) + \gamma_n - \delta_n, \\ \sum_1^\infty \beta_n &< \infty, \quad \sum_1^\infty \gamma_n < \infty \quad p.s. \end{aligned}$$

Alors, la suite (α_n) converge presque sûrement vers une variable aléatoire α finie et on a $\sum_1^\infty \delta_n < \infty \quad p.s.$

1. On a :

$$\begin{aligned} \|Y_{n+1}\|^2 &= \|Y_n\|^2 + a_n^2 \|V_n V_n' Y_n - V_n S_n\|^2 + 2a_n \langle Y_n, V_n S_n\rangle \\ &\quad -2a_n \langle Y_n, V_n V_n' Y_n\rangle \\ &\leq \|Y_n\|^2 + 2a_n^2 \|V_n\|^2 \|Y_n\|^2 + 2a_n^2 \|V_n\|^2 S_n^2 + 2a_n \langle Y_n, V_n S_n\rangle \\ &\quad -2a_n \langle Y_n, V_n V_n' Y_n\rangle . \end{aligned}$$

Soit T_n la tribu du passé au temps n, par rapport à laquelle $Y_1, \ldots, Y_n$ sont mesurables. On a $E\left[S_n \mid T_n\right] = E\left[R^i\right] = 0$, $E\left[S_n^2 \mid T_n\right] = E\left[(R^i)^2\right]$.

$$\begin{aligned} E\left[\|Y_{n+1}\|^2 \mid T_n\right] &\leq (1 + 2a_n^2 \|V_n\|^2) \|Y_n\|^2 + 2a_n^2 \|V_n\|^2 E\left[S^2\right] \\ &\quad -2a_n \langle Y_n, V_n V_n' Y_n\rangle \quad p.s. \end{aligned}$$

Sous les hypothèses H1a et H3', on a : $\sum_1^\infty a_n^2 \|V_n\|^2 < \infty$. D'après le lemme 5 :

$$\exists T \geq 0 : \|Y_n\|^2 \to T \quad p.s.; \sum_1^\infty a_n \langle Y_n, V_n V_n' Y_n\rangle < \infty \quad p.s.$$

2. On a : $\|Y_{n+1} - Y_n\| \leq a_n \|V_n\|^2 \|Y_n\| + a_n \|V_n S_n\|$.
   $E\left[\sum_1^\infty a_n^2 \|V_n S_n\|^2\right] = \sum_1^\infty a_n^2 \|V_n\|^2 E\left[S^2\right] < \infty$.
   Donc : $\sum_1^\infty a_n^2 \|V_n S_n\|^2 < \infty \quad p.s.$; $a_n \|V_n S_n\| \to 0 \quad p.s.$
   Sous H1a et H3, $a_n \|V_n\|^2 \to 0 \quad p.s.$
   On en déduit que $\|Y_{n+1} - Y_n\| \to 0 \quad p.s.$
3. On raisonne à ω fixé, appartenant à l'intersection des ensembles de convergence presque sûre définis. Supposons $T(\omega) \neq 0$.
   On supprime dans la suite l'écriture de ω.
   Alors : $\exists 0 < \epsilon_1 < 1, \exists N(\epsilon_1) : \forall n > N(\epsilon_1), \epsilon_1 < \|Y_n\| < \frac{1}{\epsilon_1}$.
   Donc, sous H1b, à partir d'un certain rang L, on a :

$$\left\langle Y_{n_l}, \sum_{j \in I_l} V_j V_j' Y_{n_l} \right\rangle \geq \lambda \epsilon_1^2.$$

   On en déduit qu'il existe un entier $m_l \in I_l$ tel que :

$$\left\langle Y_{n_l}, V_{m_l} V_{m_l}' Y_{n_l} \right\rangle \geq \frac{\lambda \epsilon_1^2}{r}.$$

   On considère la décomposition :

$$\left\langle Y_{m_l}, V_{m_l} V_{m_l}' Y_{m_l} \right\rangle = \left\langle Y_{m_l} + Y_{n_l}, V_{m_l} V_{m_l}' (Y_{m_l} - Y_{n_l}) \right\rangle + \left\langle Y_{n_l}, V_{m_l} V_{m_l}' Y_{n_l} \right\rangle.$$

   Soit $\epsilon > 0$ tel que $\epsilon \leq \frac{\lambda \epsilon_1^3}{4 r^2 C^2}$, avec $C = \sup_n \|V_n\| < \infty$ sous H1a. A partir d'un certain rang : $\|Y_{n+1} - Y_n\| < \epsilon$; donc : $\|Y_{m_l} - Y_{n_l}\| < r\epsilon$;

$$\begin{aligned} &\implies \left|\left\langle Y_{m_l} + Y_{n_l}, V_{m_l} V_{m_l}' (Y_{m_l} - Y_{n_l}) \right\rangle\right| < 2 \frac{1}{\epsilon_1} C^2 r \epsilon \leq \frac{\lambda \epsilon_1^2}{2r} \\ &\implies \left\langle Y_{m_l}, V_{m_l} V_{m_l}' Y_{m_l} \right\rangle > \frac{\lambda \epsilon_1^2}{2r}. \end{aligned}$$

   Sous H3', on a alors : $\sum_{l=1}^\infty a_{m_l} \left\langle Y_{m_l}, V_{m_l} V_{m_l}' Y_{m_l} \right\rangle = \infty$. Donc :
   $\sum_{n=1}^\infty a_n \langle Y_n, V_n V_n' Y_n \rangle = \infty$; il y a contradiction. Par conséquent : $T(\omega) = 0$.■

7.2 Démonstration du théorème 2

1. On reprend les notations du théorème 1. D'après la partie 1 de sa démonstration, on a :

$$\begin{aligned} E\left[\|Y_{n+1}\|^2\right] \quad \leq \quad & (1 + 2a_n^2 \|V_n\|^2) E\left[\|Y_n\|^2\right] + 2a_n^2 \|V_n\|^2 E\left[S^2\right] \\ & -2a_n E\left[\langle Y_n, V_n V_n' Y_n \rangle\right]. \end{aligned}$$

   D'après le lemme 5 :

$$\exists t \geq 0 : E\left[\|Y_n\|^2\right] \to t \quad ; \quad \sum_1^\infty a_n E\left[\langle Y_n, V_n V_n' Y_n \rangle\right] < \infty.$$

2. Donc, il existe $b > 0$ tel que, avec $\mu_l = \min_{j \in I_l} a_j$:

$$\begin{aligned} E\left[\|Y_{n+1}\|^2\right] &\leq E\left[\|Y_n\|^2\right] + ba_n^2 \|V_n\|^2 - 2a_n E\left[\langle Y_n, V_n V_n' Y_n\rangle\right] \\ E\left[\left\|Y_{n_{l+1}}\right\|^2\right] &\leq E\left[\|Y_{n_l}\|^2\right] + b\sum_{j\in I_l} a_j^2 \|V_j\|^2 - 2\mu_l \sum_{j\in I_l} E\left[\left\langle Y_j, V_j V_j' Y_j\right\rangle\right]. \end{aligned}$$

$$\begin{aligned} \sum_{j\in I_l} E\left[\langle Y_j, V_j V_j' Y_j\rangle\right] &= \sum_{j\in I_l} E\left[\langle Y_{n_l}, V_j V_j' Y_{n_l}\rangle\right] \\ &\quad + \sum_{j\in I_l} E\left[\left\langle Y_j + Y_{n_l}, V_j V_j'(Y_j - Y_{n_l})\right\rangle\right] \end{aligned}$$

Sous H1b : $\sum_{j\in I_l} E\left[\langle Y_{n_l}, V_j V_j' Y_{n_l}\rangle\right] \geq \lambda E\left[\|Y_{n_l}\|^2\right]$.
On note $C = \sup_n \|V_n\|$. Il existe $a > 0$ tel que :

$$\begin{aligned} &\sum_{j\in I_l} \left|E\left[\left\langle Y_j + Y_{n_l}, V_j V_j'(Y_j - Y_{n_l})\right\rangle\right]\right| \\ \leq\ &\sum_{j\in I_l} E\left[\|Y_j + Y_{n_l}\| \|V_j\|^2 \|Y_j - Y_{n_l}\|\right] \\ \leq\ &C^2 \sum_{j\in I_l} \left(E\left[\|Y_j + Y_{n_l}\|^2\right]\right)^{\frac{1}{2}} \left(E\left[\|Y_j - Y_{n_l}\|^2\right]\right)^{\frac{1}{2}} \\ \leq\ &a^2 \sum_{j\in I_l} \left(E\left[\|Y_j - Y_{n_l}\|^2\right]\right)^{\frac{1}{2}} \end{aligned}$$

Or : $Y_j - Y_{n_l} = \sum_{k=n_l}^{j-1}(-a_k V_k V_k' Y_k + a_k V_k S_k)$.
$\|Y_j - Y_{n_l}\| \leq \sum_{k=n_l}^{j-1}(a_k C^2 \|Y_k\| + a_k C \|S_k\|)$.
Il existe $d > 0$ tel que : $E\left[\|Y_j - Y_{n_l}\|^2\right] \leq d\sum_{k\in I_l} a_k^2 \leq dr \max_{k\in I_l} a_k^2$.
Donc, il existe $f > 0$ tel que :

$$\sum_{j\in I_l} \left|E\left[\left\langle Y_j + Y_{n_l}, V_j V_j'(Y_j - Y_{n_l})\right\rangle\right]\right| \leq f \max_{k\in I_l} a_k.$$

Par conséquent, il existe $g > 0$ tel que :

$$\begin{aligned} E\left[\left\|Y_{n_{l+1}}\right\|^2\right] &\leq (1 - 2\lambda\mu_l) E\left[\|Y_{n_l}\|^2\right] + brC^2 \max_{k\in I_l} a_k^2 + 2f\mu_l \max_{k\in I_l} a_k \\ &\leq (1 - 2\lambda\mu_l) E\left[\|Y_{n_l}\|^2\right] + g \max_{k\in I_l} a_k^2. \end{aligned}$$

Or : $\mu_l = \frac{c}{(n_{l+1}-1)^\alpha} \geq \frac{c}{(lr)^\alpha}$, $\max_{k\in I_l} a_k = \frac{c}{(n_l)^\alpha} \leq \frac{c}{l^\alpha}$.
Donc, il existe $h > 0$ tel que :

$$E\left[\left\|Y_{n_{l+1}}\right\|^2\right] \leq \left(1 - 2\frac{\lambda c}{r^\alpha}\frac{1}{l^\alpha}\right) E\left[\|Y_{n_l}\|^2\right] + \frac{h}{l^{2\alpha}}.$$

3. Dans le cas $\frac{1}{2} < \alpha < 1$, on applique un lemme de Schmetterer (1969) :

$$\overline{\lim} l^{\alpha} E\left[\|Y_{n_l}\|^2\right] < \infty.$$

Dans le cas $\alpha = 1$, on applique un lemme de Venter (1966) :

$$\begin{aligned} pour \quad \frac{2\lambda c}{r} &> 1, \overline{\lim} l E\left[\|Y_{n_l}\|^2\right] < \infty \quad ; \\ pour \quad \frac{2\lambda c}{r} &= 1, \overline{\lim} \frac{l}{\ln l} E\left[\|Y_{n_l}\|^2\right] < \infty \quad ; \\ pour \quad \frac{2\lambda c}{r} &< 1, \overline{\lim} l^{\frac{2\lambda c}{r}} E\left[\|Y_{n_l}\|^2\right] < \infty. \end{aligned}$$

Comme $E\left[\|Y_{n+1}\|^2\right] \leq E\left[\|Y_n\|^2\right] + ba_n^2 \|V_n\|^2$, on a pour $n \in I_l$:

$$E\left[\|Y_n\|^2\right] \leq E\left[\|Y_{n_l}\|^2\right] + \frac{h}{l^{2\alpha}}.$$

Comme $l \leq n \leq lr$, on obtient des résutats semblables aux précédents pour $\mathrm{E}\left[\|Y_n\|^2\right]$, en remplaçant l et n_l par n.■

7.3 Démonstration du théorème 3

D'après le théorème 4 de Bouamaine et Monnez (1998), on a les conclusions de ce théorème sous les hypothèses :

1. $\sum_1^{\infty} a_n \|E[C_n \mid T_n] - C\| < \infty \quad p.s.$
2. Pour $j = 2, \ldots, 2r$, $\sum_1^{\infty} a_n^j E\left[\|C_n - C\|^j \mid T_n\right] < \infty \quad p.s.$
3. M_{n-1} est T_n-mesurable ; $M_n \to M \quad p.s.$; $\sum_1^{\infty} a_n \|M_{n-1} - M\| < \infty \quad p.s.$
4. $a_n > 0$, $\sum_1^{\infty} a_n = \infty$, $\sum_1^{\infty} a_n^2 < \infty$.

On vérifie l'hypothèse 1.

$$\begin{aligned} C_n &= M_{n-1}(Z_n Z_n' - \widehat{\Theta}_n \widehat{\Theta}_n') \quad ; \quad C = M\Sigma = M(E[Z_n Z_n'] - \theta_n \theta_n'). \\ C_n - C &= M_{n-1}(Z_n Z_n' - E[Z_n Z_n']) + (M_{n-1} - M)(E[Z_n Z_n'] - \theta_n \theta_n') \\ &\quad - M_{n-1}(\widehat{\Theta}_n - \theta_n)\widehat{\Theta}_n' - M_{n-1}\theta_n(\widehat{\Theta}_n - \theta_n)'. \\ E[C_n \mid T_n] - C &= (M_{n-1} - M)\Sigma - M_{n-1}(\widehat{\Theta}_n - \theta_n)\widehat{\Theta}_n' - M_{n-1}\theta_n(\widehat{\Theta}_n - \theta_n)'. \end{aligned}$$

D'après les conclusions du théorème 2, en utilisant la norme euclidienne usuelle dans $\mathbb{R}^p$, on a dans tous les cas :

$$\begin{aligned} E\left[\sum_1^{\infty} \frac{1}{n^{\alpha}} \left\|\widehat{\Theta}_n - \theta_n\right\|\right] &\leq \sum_1^{\infty} \frac{1}{n^{\alpha}} \left(E\left[\left\|\widehat{\Theta}_n - \theta_n\right\|^2\right]\right)^{\frac{1}{2}} \\ &\leq \sum_{i=1}^{p} \sum_{n=1}^{\infty} \frac{1}{n^{\alpha}} \left(E\left[\left\|\widehat{\Theta}_n^i - \theta_n^i\right\|^2\right]\right)^{\frac{1}{2}} < \infty. \end{aligned}$$

Donc : $\sum_1^\infty \frac{1}{n^\alpha} \left\| \widehat{\Theta}_n - \theta_n \right\| < \infty \quad p.s.$
D'après le théorème 1, $\widehat{\Theta}_n - \theta_n \to 0 \quad p.s.$
Sous H2, on a alors :

$$\sum_1^\infty \frac{1}{n^\alpha} \| E\left[C_n \mid T_n\right] - C \| < \infty \quad p.s.$$

On vérifie l'hypothèse 2 en écrivant que :

$$\begin{aligned} \|C_n - C\|^j \quad \leq \quad & 4^{j-1}(\|M_{n-1}\|^j \|Z_n Z_n' - E\left[Z_n Z_n'\right]\| + \|M_{n-1} - M\|^j \|\Sigma\|^j \\ & + \|M_{n-1}\|^j \left\| \widehat{\Theta}_n - \theta_n \right\|^j \left\| \widehat{\Theta}_n \right\|^j + \|M_{n-1}\|^j \|\theta_n\|^j \left\| \widehat{\Theta}_n - \theta_n \right\|^j), \end{aligned}$$

$$\begin{aligned} \sum_1^\infty a_n^j \|M_n - M\|^j \quad < \quad \infty, \quad \sum_1^\infty a_n^j \left\| \widehat{\Theta}_n - \theta_n \right\|^j < \infty, \quad \sum_1^\infty a_n^j < \infty, \\ M_n \quad \to \quad M, \quad \widehat{\Theta}_n - \theta_n \to 0 \quad p.s. \blacksquare \end{aligned}$$

7.4 Démonstration du théorème 4

1. Montrons que, pour $i = 1, \ldots, p$:

$$\frac{1}{n} \sum_{j=1}^n (Z_j^i - \widehat{\Theta}_j^i)^2 \to (\sigma^i)^2 \quad p.s.$$

$$\frac{1}{n} \sum_{j=1}^n (Z_j^i - \widehat{\Theta}_j^i)^2 = \frac{1}{n} \sum_{j=1}^n ((Z_j^i - \theta_j^i) + (\theta_j^i - \widehat{\Theta}_j^i))^2$$

$$= \frac{1}{n} \sum_{j=1}^n (Z_j^i - \theta_j^i)^2 + \frac{2}{n} \sum_{j=1}^n (Z_j^i - \theta_j^i)(\theta_j^i - \widehat{\Theta}_j^i) + \frac{1}{n} \sum_{j=1}^n (\theta_j^i - \widehat{\Theta}_j^i)^2.$$

Les $Z_j^i - \theta_j^i$ constituent un échantillon i.i.d. de R^i. Donc : $\frac{1}{n} \sum_{j=1}^n (Z_j^i - \theta_j^i)^2 \to (\sigma^i)^2 \quad p.s.$
On a : $\theta_j^i - \widehat{\Theta}_j^i \to 0 \quad p.s.$; donc : $\frac{1}{n} \sum_{j=1}^n (\theta_j^i - \widehat{\Theta}_j^i)^2 \to 0 \quad p.s.$
Notons $V_n^i = (Z_n^i - \theta_n^i)(\theta_n^i - \widehat{\Theta}_n^i)$, $W_{n+1}^i = \frac{1}{n} \sum_{j=1}^n V_j^i$, $W_1^i = 0$. On a :

$$\begin{aligned} W_{n+1}^i \quad &= \quad (1 - \frac{1}{n}) W_n^i + \frac{1}{n} V_n^i \\ (W_{n+1}^i)^2 \quad &= \quad (1 + \frac{1}{n^2})(W_n^i)^2 + 2(1 - \frac{1}{n}) \frac{1}{n} V_n^i W_n^i + \frac{1}{n^2} (V_n^i)^2 - \frac{2}{n} (W_n^i)^2. \end{aligned}$$

Soit T_n la tribu du passé au temps n.

$$E\left[V_n^i \mid T_n\right] = E\left[Z_n^i - \theta_n^i \mid T_n\right](\theta_n^i - \widehat{\Theta}_n^i) = E\left[Z_n^i - \theta_n^i\right](\theta_n^i - \widehat{\Theta}_n^i) = 0.$$

$$E\left[(W_{n+1}^i)^2 \mid T_n\right] = (1 + \frac{1}{n^2})(W_n^i)^2 + \frac{1}{n^2}E\left[(V_n^i)^2 \mid T_n\right] - \frac{2}{n}(W_n^i)^2 \quad p.s.$$

Or : $E\left[(V_n^i)^2 \mid T_n\right] = E\left[(Z_n^i - \theta_n^i)^2 \mid T_n\right](\theta_n^i - \widehat{\Theta}_n^i)^2 = E\left[(R^i)^2\right](\theta_n^i - \widehat{\Theta}_n^i)^2$.
Donc : $\sum_1^\infty \frac{1}{n^2}E\left[(V_n^i)^2 \mid T_n\right] = E\left[(R^i)^2\right]\sum_1^\infty \frac{1}{n^2}(\theta_n^i - \widehat{\Theta}_n^i)^2 < \infty \quad p.s.$
En appliquant le lemme 5, on obtient :

$$\exists T^i \geq 0 : (W_n^i)^2 \to T^i \quad p.s. \quad ; \quad \sum_1^\infty \frac{1}{n}(W_n^i)^2 < \infty \quad p.s. \quad \Longrightarrow W_n^i \to 0 \quad p.s.$$

Par conséquent : $\frac{1}{n}\sum_{j=1}^n (Z_j^i - \widehat{\Theta}_j^i)^2 \to (\sigma^i)^2 \quad p.s.$
On en déduit que : $M_n \to M \quad p.s.$

2. On utilise dans la suite le lemme suivant.

Lemme 6. *Soit, pour tout* $n \geq 1$ *:* $w_{n+1} = (1-a_n)w_n + a_n u_n, \quad w_1 = 0, \quad u_n > 0, \quad 0 < a_n < 1$. *Si* $\sum_1^\infty a_n = \infty$ et $\sum_1^\infty a_n u_n < \infty$, *alors* $\sum_1^\infty a_n w_n < \infty$ *et* $w_n \to 0$.
On remarque que, pour $a_n = \frac{1}{n}$, $w_{n+1} = \frac{1}{n}\sum_{j=1}^n u_j$.

Démonstration. Pour $n > 1$, $w_n > 0$. D'après le lemme 5, il existe $w \geq 0 : w_n \to w$ et $\sum_1^\infty a_n w_n < \infty$; comme $\sum_1^\infty a_n = \infty$, on a $w = 0$.

3. Montrons que, pour $i = 1, \ldots, p$, $\sum_{n=1}^\infty \frac{1}{n}\left|\frac{1}{n}\sum_{j=1}^n (Z_j^i - \widehat{\Theta}_j^i)^2 - (\sigma^i)^2\right| < \infty \; p.s.$

On utilise la décomposition de $\frac{1}{n}\sum_{j=1}^n (Z_j^i - \widehat{\Theta}_j^i)^2$ vue dans la première partie de la démonstration. Soit μ_4^i le moment centré d'ordre 4 de R^i. On a :

$$E\left[(\frac{1}{n}\sum_{j=1}^n (Z_j^i - \theta_j^i)^2 - (\sigma^i)^2)^2\right] = \frac{1}{n^2}\sum_{j=1}^n Var\left[(Z_j^i - \theta_j^i)^2\right] = \frac{\mu_4^i - (\sigma^i)^4}{n},$$

$$\sum_1^\infty \frac{1}{n}E\left[\left|\frac{1}{n}\sum_{j=1}^n (Z_j^i - \theta_j^i)^2 - (\sigma^i)^2\right|\right] < \infty.$$

D'après le théorème 2, on a : $\sum_1^\infty \frac{1}{n}E\left[(\theta_n^i - \widehat{\Theta}_n^i)^2\right] < \infty$. On déduit du lemme 6 que : $\sum_{n=1}^\infty \frac{1}{n}E\left[\frac{1}{n}\sum_{j=1}^n (\theta_j^i - \widehat{\Theta}_j^i)^2\right] < \infty$.

$$\begin{aligned}\sum_1^\infty \frac{1}{n}E\left[\left|(Z_n^i - \theta_n^i)(\theta_n^i - \widehat{\Theta}_n^i)\right|\right] &\leq \sum_1^\infty \frac{1}{n}(E\left[(Z_n^i - \theta_n^i)^2\right])^{\frac{1}{2}}\left(E\left[(\theta_n^i - \widehat{\Theta}_n^i)^2\right]\right)^{\frac{1}{2}} \\ &\leq (E\left[(R^i)^2\right])^{\frac{1}{2}}\sum_1^\infty \frac{1}{n}\left(E\left[(\theta_n^i - \widehat{\Theta}_n^i)^2\right]\right)^{\frac{1}{2}} < \infty.\end{aligned}$$

On déduit du lemme 6 que :

$$\sum_{n=1}^{\infty} \frac{1}{n} E\left[\frac{1}{n}\sum_{j=1}^{n}\left|(Z_j^i-\theta_j^i)(\theta_j^i-\widehat{\Theta}_j^i)\right|\right] < \infty \quad p.s.$$

On déduit des trois conclusions précédentes que :

$$\sum_{n=1}^{\infty} \frac{1}{n} E\left[\left|\frac{1}{n}\sum_{j=1}^{n}(Z_j^i-\widehat{\Theta}_j^i)^2-(\sigma^i)^2\right|\right] < \infty$$

$$\sum_{n=1}^{\infty} \frac{1}{n}\left|\frac{1}{n}\sum_{j=1}^{n}(Z_j^i-\widehat{\Theta}_j^i)^2-(\sigma^i)^2\right| < \infty \quad p.s.$$

Par conséquent, presque sûrement, $\sum_{n=1}^{\infty}\frac{1}{n}\left\|M_n^{-1}-M^{-1}\right\| < \infty$ et

$$\sum_{n=1}^{\infty}\frac{1}{n}\left\|M_n-M\right\| = \sum_{n=1}^{\infty}\frac{1}{n}\left\|M_n(M_n^{-1}-M^{-1})M\right\| < \infty.\blacksquare$$

Références

Aguilar-Ruiz, J. (2006). Recent advances in data stream mining. In *38èmes Journées de Statistique de la SFDS*, Clamart.

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

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

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

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

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

Krasulina, T. (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.

Lebart, L. (1974). On the benzécri's method for computing eigenvectors by stochastic approximation (the case of binary data). In P. Verlag (Ed.), *Proceedings in Computational Statistics*, Vienne, pp. 202–211.

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

Monnez, J. (1994). Convergence d'un processus d'approximation stochastique en analyse factorielle. *Pub. Inst. Stat. Univ. Paris XXXVIII*(1), 37–56.

Monnez, J. (2006). Approximation stochastique en analyse factorielle multiple. *Pub. Inst. Stat. Univ. Paris L*(3), 27–45.

Robbins, H. et S. Monro (1951). A stochastic approximation method. *Ann. Math. Stat. 22*, 400–407.

Robbins, H. et D. Siegmund (1971). *A convergence theorem for nonnegative almost supermartingales and some applications* (Rustagi, J.S. ed.)., pp. 233–257. Academic Press, New York.

Schmetterer, L. (1969). Multidimensional stochastic approximation. In A. Press (Ed.), *Multivariate Analysis II, Proc. 2nd Int. Symp.*, Dayton, Ohio, pp. 443–460.

Venter, J. (1966). On dvoretzky stochastic approximation theorems. *Ann. Math. Stat. 37*, 1534–1544.

Summary

We consider a data stream and suppose that each data is a realization of a random vector whose expectation varies in time according to a linear model for each component. We use stochastic approximation processes to estimate on line the parameters of the linear models and simultaneously the principal components of the data.