



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Thèse
présentée pour l'obtention du titre de
Docteur de l'Université de Lorraine
en Mathématiques
par
Romain BAR

**DEVELOPPEMENT DE METHODES D'ANALYSE DE DONNEES EN
LIGNE**

Soutenue publiquement le 29 Novembre 2013

Membres du Jury :

C. BOUVEYRON	Professeur à l'Université Paris Descartes (Examineur)
H. CARDOT	Professeur à l'Université de Bourgogne (Rapporteur)
B. FICHET	Maître de conférences à l'Université d'Aix-Marseille (Examineur)
A. GEGOUT-PETIT	Professeur à l'Université de Lorraine (Examineur)
J.M. MONNEZ	Professeur à l'Université de Lorraine (Directeur)
J. SARACCO	Professeur à l'Institut Polytechnique de Bordeaux (Rapporteur)
P. VALLOIS	Professeur à l'Université de Lorraine (Président)

Institut Élie Cartan de Lorraine, Laboratoire de Mathématiques, Campus des Aiguillettes,
B.P. 70239, 54506 Vandoeuvre-lès-Nancy Cedex

Au premier chef, toute ma gratitude va à Monsieur Jean-Marie MONNEZ qui a accepté de diriger ma thèse et a consciencieusement suivi l'avancée de mes travaux tout au long de ces trois années. Ce dernier a toujours su se montrer disponible et ce, malgré les responsabilités qui lui incombent, ce qui témoigne de son professionnalisme sans faille que je salue avec le plus grand respect. Nos échanges ont toujours été très enrichissants et m'ont permis d'évoluer dans la rigueur, mais toujours sereinement, dans mon travail de recherche. Enfin, je voudrais le remercier de la manière la plus solennelle pour avoir pris le temps et la peine de me transmettre de son savoir mais aussi de sa riche expérience professionnelle, m'éclairant ainsi sur la réalité du métier.

Ensuite, je souhaite vivement remercier Monsieur Pierre VALLOIS pour avoir accepté de présider mon jury de thèse. Je lui dois, en partie, ma formation aux Probabilités et à la Statistique sans laquelle je n'aurais pas pu réaliser cette étude. Je lui suis aussi très reconnaissant pour tous les conseils bienveillants qu'il m'a prodigués tout au long de ma scolarité à l'Université Nancy I.

J'exprime également mes remerciements les plus chaleureux à Messieurs Hervé CARDOT et Jérôme SARACCO qui me font l'honneur d'être rapporteurs de ma thèse. Nous avons échangé à plusieurs reprises sur nos lieux de travail respectifs ou lors de conférences et j'en garde à chaque fois d'excellents souvenirs.

Je voudrais encore adresser mes profonds et sincères remerciements à Madame Anne GEGOUT-PETIT et Messieurs Charles BOUVEYRON et Bernard FICHET pour avoir participé à ce jury et m'avoir fait part de leurs remarques et suggestions sur cette étude.

D'autre part et dans le désordre, je salue et remercie Marc, Angelo, Sophie, Jean-Luc, Geoffrey, Hassan, Paul, Didier, Elodie, Olivier, Roberto, Benoît ainsi que toutes les autres personnes avec qui j'ai partagé de beaux moments de convivialité à l'Université.

Enfin, je ne peux terminer ces remerciements sans avoir une pensée émue pour ma famille et mes proches que je remercie du fond du cœur pour leurs encouragements appuyés et leur soutien inconditionnel.

Table des matières

1	Etat de l'art	8
2	ACP projetée de données séquentielles	24
2.1	Introduction	25
2.1.1	Cadre d'étude	25
2.1.2	Présentation de l'ACP projetée	25
2.1.3	ACP d'un vecteur aléatoire	26
2.1.4	Facteurs de l'ACP projetée	28
2.2	Cas où il n'y a pas de relation affine entre les composantes du vecteur explicatif	28
2.2.1	A.S. d'une fonction de régression linéaire	30
2.2.2	A.S. des facteurs de l'ACP projetée	31
2.2.3	Variantes	35
2.2.4	Cas de l'Analyse Canonique de deux vecteurs aléatoires	37
2.3	Cas où les composantes du vecteur explicatif sont les indicatrices des modalités exclusives d'une variable aléatoire nominale	46
2.3.1	A.S. de la fonction de régression linéaire	49
2.3.2	A.S. des facteurs de l'ACP projetée	52
2.3.3	Cas particulier de l'Analyse Factorielle Discriminante	52
2.3.4	Cas particulier de l'Analyse Factorielle des Correspondances	56
2.4	Annexes du chapitre 2	62
2.4.1	Démonstration du théorème 10	62
2.4.2	Hypothèses	65
3	ACP partielle d'un flux de données d'espérance mathématique variable dans le temps	68
3.1	Cas général de l'ACP partielle	69
3.1.1	Cas général	70
3.1.2	Cas particulier d'un modèle linéaire de variation de l'espérance	73
3.2	Cas particulier : l'ACG partielle	74
3.2.1	Premier cas : utilisation d'un paquet de données	74
3.2.2	Deuxième cas : utilisation d'un processus d'A.S. de C^k	75
3.2.3	Troisième cas : utilisation d'un processus d'A.S. de C^k (cas particulier du précédent)	76
3.3	Annexes du chapitre 3	77
3.3.1	Lemmes techniques	77
3.3.2	Démonstrations des théorèmes	84
3.3.3	Hypothèses	104

4	ACP partielle d'un flux de données d'espérance mathématique et de matrice de covariance variables dans le temps	106
4.1	Cas général de l'ACP partielle	107
4.1.1	Cas général	108
4.1.2	Cas particulier d'un modèle linéaire de variation de l'espérance et de la variance . .	109
4.2	Cas particulier : l'ACG partielle	111
4.3	Annexes du chapitre 4	114
4.3.1	Démonstrations des théorèmes	114
4.3.2	Hypothèses	147
5	Mise en oeuvre informatique	150
5.1	Cadre déterministe	151
5.2	Cadre aléatoire	154
5.2.1	Cas i.i.d.	154
5.2.2	Cas où l'espérance est variable dans le temps	157
5.2.3	Cas où l'espérance et la matrice de covariance sont variables dans le temps	163
5.3	Package R : factas	170
	Conclusion	172
	Bibliographie	176

Chapitre 1

Etat de l'art

Comme il est d'usage pour commencer la présentation des travaux d'une thèse, nous consacrons cette première partie à l'état de l'art. Ainsi, nous allons faire état des connaissances actuelles sur le sujet afin de mieux cerner et situer l'apport de cette thèse d'un point de vue scientifique.

A cet effet, nous ferons référence et expliciterons les travaux antérieurs ayant un lien plus ou moins direct avec le travail accompli et expliquerons en quoi notre contribution, de par son originalité, vise à étendre la théorie actuelle.

Ce travail bibliographique, auquel nous nous sommes majoritairement consacrés durant la première année de recherche, s'est avéré particulièrement utile pour la compréhension du sujet, l'identification des principaux outils mathématiques à maîtriser ainsi que la prise de conscience des enjeux sous-jacents.

En effet, il ne faut pas oublier que la statistique en général et l'analyse des données en particulier ont vocation à être des sciences dites appliquées ; ainsi, retracer chronologiquement l'évolution de cette branche des mathématiques permet de mettre en perspective les travaux de recherche publiés avec les connaissances techniques et technologiques du moment. Aussi, nous nous efforcerons de justifier la pertinence de nos travaux au vu du contexte global actuel.

Il faut d'ailleurs insister sur le fait que, et c'est d'autant plus vrai pour l'analyse des données, la plupart des modèles mathématiques développés pour répondre à une problématique concrète et visant à être utilisés dans quelque domaine que ce soit, se doivent, de par leur essence même, d'être exploitables !

Ainsi, nous verrons que l'avènement de l'informatique a joué un rôle prépondérant dans le développement de l'analyse des données et que, depuis une dizaine d'années, ce dernier explose, devenant même aux dires de certains un enjeu majeur de compétitivité pour l'ensemble des acteurs économiques. Compte tenu de ces considérations, il nous est apparu essentiel de mettre en oeuvre informatiquement les algorithmes conçus de manière théorique, afin de mesurer leur efficacité.

Ceci étant, nous pouvons désormais entrer dans le vif du sujet et présenter le contexte global dans lequel s'inscrit cette thèse.

Le travail que nous présentons fait partie d'une théorie plus globale née de la rencontre heureuse d'un do-

maine de la statistique : l'analyse des données et de techniques d'optimisation stochastique : les méthodes d'approximation stochastique (a.s.).

L'analyse des données au sens où on l'entend, c'est à dire au sens moderne, trouve ses racines au début du XX^{ème} siècle dans les travaux du mathématicien anglais Karl Pearson, fondateur, entre autres, du journal *Biometrika* avec Walter Weldon et Raphael Gatson.

Ses contributions au développement de la statistique furent nombreuses mais c'est sans conteste le développement du test du χ^2 qui marqua le plus les esprits et le rendit célèbre, encore aujourd'hui.

Même si ses idées de l'époque sur l'eugénisme peuvent s'avérer sujettes à polémiques de nos jours, Pearson n'en fut pas moins un grand spécialiste du domaine biologique, étudiant plus spécifiquement la théorie de l'évolution.

Il s'intéressa également largement à la physique et ses travaux sur la relativité furent même repris par un illustre scientifique... Albert Einstein.

C'est donc dans [56] que l'on peut percevoir les prémices de l'actuelle analyse en composantes principales.

Parallèlement et de façon concurrentielle pour ne pas dire conflictuelle, le psychologue anglais Charles Spearman développa la première version d'une analyse factorielle [65], pour sa théorie de l'intelligence, dans l'espoir de déterminer un facteur général de l'intelligence humaine : le facteur "G".

Dans les années 30, Louis Thurstone, fondateur du journal *Psychometrika*, reprend les travaux initiés par Spearman. Il propose pour sa part une approche multi-factorielle de l'intelligence basée sur sept aptitudes primaires de référence : la compréhension orale, la fluidité du langage, le calcul numérique, la représentation spatiale, la mémoire, la vitesse de perception et le raisonnement [67].

A cette époque, et dans un tout autre domaine, celui de l'économie, l'américain Harold Hotelling se distingue par la publication de [28], article dans lequel les bases de l'analyse canonique sont posées.

La même année et non sans lien, Ronald Fisher, sans conteste l'un des plus importants statisticiens du XX^{ème} siècle, applique son modèle d'analyse discriminante au désormais célèbre jeu de données sur les fleurs d'iris dans [18].

Les premières traces des équations de l'analyse des correspondances apparaissent quant à elles dans [27], article de 1935 dû au statisticien germano-américain Hirschfeld (ou Hartley).

Ces travaux furent ensuite repris par Louis Guttman [23] et Cyril Burt [11] dans les années 40, mais c'est dans les années 50 et 60, respectivement sous l'impulsion du japonais Chikio Hayashi [25], [26] et du français Jean-Paul Benzécri [5], [4] que ces techniques exploratoires sont développées et appliquées de manière systématique et intensive.

Il est important de noter que l'essor de l'analyse des données a été nettement facilité par les progrès de l'informatique, comme l'atteste cette phrase de Benzécri : "*Avec l'Analyse des Données fondée sur l'usage de l'ordinateur, c'est une nouvelle méthodologie que la statistique apporte à la science et notamment aux sciences de l'homme*".

Ces années 60 – 70 amorcent un véritable tournant dans la façon de penser l'analyse des données. En effet, l'introduction du modèle probabiliste, sous l'impulsion de l'américain John Tukey [69], [20], va apporter à

l'analyse exploratoire des données une nouvelle dimension inférentielle prometteuse. Sous ce nouvel angle, les techniques descriptives s'enrichissent d'estimations, de convergences, voire de tests d'hypothèses.

Les méthodes d'a.s., quant à elles, ont été introduites en 1951 par Herbert Robbins et Sutton Monro dans [59]. Le contexte du problème qui a motivé ce travail était celui d'expériences chimiques dans lesquelles une certaine dose x d'une substance avait pour effet moyen $M(x)$; le but était alors de trouver la solution $x = x^*$ de l'équation $M(x) = C$, où C était une constante donnée.

Nous présentons maintenant de manière plus formelle le processus de Robbins-Monro :

Soit $(Y(x))_{x \in \mathbb{R}}$ une famille de variables aléatoires réelles observables pour toute valeur du paramètre réel x ; on suppose inconnue la fonction $\mathbb{E}[Y(x)] = M(x)$.

On cherche à estimer la solution x^* , supposée unique, de l'équation $M(x) = 0$.

Pour cela, on définit un processus stochastique (X_n) tel que :

$$X_{n+1} = X_n - a_n Y_n,$$

où Y_n est, pour n fixé, une v.a. réelle dont la loi de probabilité conditionnellement à $(X_1 = x_1, \dots, X_n = x_n)$ est la loi de $Y(x_n)$, et a_n une suite de nombres réels positifs telle que $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$.

Sous certaines hypothèses portant sur la famille $(Y(x))_{x \in \mathbb{R}}$, le processus (X_n) converge presque sûrement (p.s.) vers x^* .

L'année suivante, Jack Kiefer et son directeur de thèse, Jacob Wolfowitz, mirent au point un algorithme légèrement différent de celui de Robbins-Monro, basé sur une forme aux différences finies, pour estimer le maximum x^* , supposé unique, de la fonction $M(x)$ [32].

En 1965, le russe Gladyshev donne une preuve alternative à celle du théorème fondateur de Robbins et Monro [22], faisant intervenir la théorie des martingales, théorie alors à son apogée grâce en particulier aux travaux de l'américain Joseph Doob.

A ce propos, on fera référence à plusieurs reprises, dans le corps de la thèse, à un lemme dû à Robbins et Siegmund tiré de [60] dont voici l'énoncé :

Lemme 1.1.

Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une suite non-décroissante T_n de sous-tribus de Ω .

Soit, pour tout n , a_n , β_n , γ_n , δ_n des variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, T_n -mesurables, non négatives et intégrables telles que :

$$\mathbb{E}[\alpha_{n+1}|T_n] \leq \alpha_n(1 + \beta_n) + \gamma_n - \delta_n, \sum_{n=1}^{+\infty} \beta_n < \infty, \sum_{n=1}^{+\infty} \gamma_n < \infty \text{ p.s.}$$

Alors, la suite (α_n) converge p.s. vers une variable aléatoire réelle α finie et on a $\sum_{n=1}^{+\infty} \delta_n < \infty$ p.s. ■

A partir de là, de nombreuses contributions ont permis d'élargir les connaissances sur le sujet, notamment les travaux de Kushner et ses coauteurs [36], [37], [35], [38], [39], [40].

Polyak et Juditsky proposèrent également un algorithme utilisant un procédé de moyennisation afin d'améliorer la rapidité de convergence des algorithmes originels [58], [29]. Cependant, de telles considérations apparaissent déjà dès 1978 dans les travaux de Nemirovski et Yudin [52].

Enfin, d'autres algorithmes de type Kiefer-Wolfowitz ont été proposés pour accélérer la convergence des algorithmes, le dernier en date et le plus performant étant l'algorithme SPSA (Simultaneous Perturbation Stochastic Approximation) introduit par Spall en 1992 dans [64].

Ceci étant, nous allons maintenant expliquer comment ces techniques ont pu être utilisées pour résoudre des problèmes d'analyse des données.

Il suffit en fait de constater que la mise en oeuvre d'analyses factorielles nécessite l'extraction de valeurs et de vecteurs propres d'une certaine matrice B .

Dans le cas où B est connue, il existe différents processus numériques déterministes qui permettent d'estimer ses éléments propres.

La méthode la plus ancienne est celle de la puissance itérée introduite en 1929 par von Mises et Pollaczek-Geiringer dans [44] qui permet d'approcher le vecteur propre associé à la plus grande valeur propre (v.p.) de la matrice considérée.

La méthode de déflation de Wielandt permet ensuite le calcul des autres v.p. et des vecteurs propres associés.

De nombreuses autres méthodes itératives ont depuis vu le jour avec plus ou moins de succès. On pourra citer par exemple les méthodes d'Arnoldi basées sur les sous-espaces de Krylov [2] ou encore l'algorithme QR [19], [34] qui utilise une décomposition astucieuse de la matrice originale.

Malheureusement, dans le cas où la matrice B est inconnue, les méthodes déterministes deviennent inopérantes et on se tourne alors vers des processus stochastiques.

Le premier à avoir publié en ce sens est Benzécri en 1969 dans [5]. On se propose de reformuler le résultat obtenu :

Soit A une matrice aléatoire carrée. On se place dans le cas où la matrice $B = \mathbb{E}[A]$ est inconnue, M -symétrique avec M connue.

On suppose qu'on dispose d'une suite de matrices aléatoires observables B_n de même loi de probabilité que A .

On définit alors le processus $(X_n^1, \dots, X_n^r)$ dans $(R^p)^r$ tel que, pour $n \geq 1$:

$$Y_{n+1}^i = (I + a_n B_n) X_n^i, \quad i = 1, \dots, r$$

$$(X_{n+1}^1, \dots, X_{n+1}^r) = \text{orth}_M(Y_{n+1}^1, \dots, Y_{n+1}^r)$$

où I désigne la matrice identité d'ordre p et où l'orthogonalisation par rapport à M est effectuée au sens de Gram-Schmidt.

On établit alors, sous certaines hypothèses, la convergence p.s. en direction de ce processus vers r premiers vecteurs propres de la matrice B associés aux r plus grandes v.p. $\lambda_1, \dots, \lambda_r$.

A noter que cette estimation simultanée des r premiers vecteurs propres est basée sur des résultats d'algèbre extérieure.

Un an plus tard et de manière indépendante, Krasulina développe un algorithme d'a.s. basé sur le quotient de Rayleigh pour déterminer un vecteur propre associé à la plus grande v.p. de l'espérance mathématique d'une matrice aléatoire [33].

Shimura et Imai en 1973 [63], puis Oja et Karhunen en 1985 [54], proposent tour à tour des variantes de ces algorithmes. Les deux derniers auteurs utilisent notamment des résultats de Kushner et Clark datant de 1978 [36] dans leurs démonstrations.

Une autre solution du problème est due à Monnez dans sa thèse d'état de 1982 [46] :

On suppose cette fois ci qu'on dispose d'une suite de matrices aléatoires B_n qui converge p.s. vers B .

En définissant un processus similaire à celui de Benzécri, on établit, sous certaines hypothèses, la convergence p.s. de ce processus vers un vecteur propre de la matrice B associé à la plus grande v.p.

Dans le cadre de l'estimation séquentielle des facteurs en analyse factorielle inférentielle, Bouamaine et Monnez considèrent dans [8] que M est inconnue mais supposent disposer d'une suite de matrices aléatoires M_n qui converge p.s. vers M .

Le processus de type Benzécri $(X_n^1, \dots, X_n^r)$ dans $(R^p)^r$ tel que, pour $n \geq 1$:

$$Y_{n+1}^i = (I + a_n B_n) X_n^i, \quad i = 1, \dots, r$$

$$(X_{n+1}^1, \dots, X_{n+1}^r) = \text{orth}_{M_n}(Y_{n+1}^1, \dots, Y_{n+1}^r)$$

est alors défini et, sous certaines hypothèses, la convergence p.s. en direction de ce processus vers r premiers vecteurs propres de la matrice B associés aux r plus grandes v.p. $\lambda_1, \dots, \lambda_r$ est établie.

En 1994 ensuite, Monnez démontre dans [48] un résultat similaire en affaiblissant l'hypothèse effectuée sur la suite de matrices B_n . Cette fois ci, on suppose seulement que $\mathbb{E}[B_n|T_n]$ converge p.s. vers B , T_n étant une suite non-décroissante de sous-tribus.

Dans [10], Bouamaine et Monnez introduisent un processus plus général de la forme :

$$\begin{aligned} Y_{n+1}^i &= X_n^i + a_n(B_n - F_n(X_n^i)I)X_n^i, \quad i = 1, \dots, r \\ (X_{n+1}^1, \dots, X_{n+1}^r) &= \text{orth}_{M_n}(Y_{n+1}^1, \dots, Y_{n+1}^r) \end{aligned}$$

où F_n désigne une fonction aléatoire de $\mathbb{R}^p$ dans $\mathbb{R}$.

Différents choix de cette fonction mènent à des processus étudiés indépendamment par les auteurs cités précédemment, dans le cas où $M_n = M$ est supposée connue et $\mathbb{E}[B_n] = B$.

Pour $F_n = 0$, on retrouve le processus de Benzécri et pour $F_n(x) = \frac{\langle B_n x, x \rangle_{M_n}}{\|x\|_{M_n}^2}$, on a un processus de type Krasulina.

On rappelle ici le théorème 4 de cet article relatif au processus de type Krasulina, auquel on fera référence à de nombreuses reprises ultérieurement :

On note r le nombre de vecteurs propres à estimer.

Soit les hypothèses (avec les numérotations de l'article) :

- (H1) (a) $\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[B_n|T_n] - B\| < \infty$ p.s.
- (c) $\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[\|B_n - B\|^j | T_n] < \infty$ pour $j = 2, \dots, 2r$ p.s.
- (H1') $\sum_{n=1}^{+\infty} a_n \|B_n - B\| < \infty$ p.s.
- (H4') (a) M_n est T_n -mesurable
- (b) $M_n \rightarrow M$ p.s.
- (c) $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.
- (H5) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$.

Avant d'énoncer le théorème, définissons l'ensemble ${}^jE$. Pour plus de détails concernant les résultats d'algèbre extérieure, on pourra consulter [49].

Soit $(e_1, e_2, \dots, e_p)$ une base de $\mathbb{R}^p$. On note ${}^r\Lambda\mathbb{R}^p$ la puissance extérieure d'ordre r de $\mathbb{R}^p$, engendrée par les C_p^r produits extérieurs $e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_r}$, $i_1 < i_2 < \dots < i_r \in \{1, 2, \dots, p\}$.

Soit M une métrique dans $\mathbb{R}^p$. On définit dans ${}^r\Lambda\mathbb{R}^p$ le produit scalaire induit par la métrique M , également noté $\langle \cdot, \cdot \rangle_M : \langle e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_r}, e_{k_1} \wedge e_{k_2} \wedge \dots \wedge e_{k_r} \rangle_M = \sum_{\sigma \in G_r} (-1)^{s(\sigma)} \langle e_{i_1}, e_{\sigma(k_1)} \rangle_M \dots \langle e_{i_r}, e_{\sigma(k_r)} \rangle_M$, G_r étant l'ensemble des permutations σ de $\{k_1, \dots, k_r\}$ et $s(\sigma)$ le nombre d'inversions de σ .

On suppose que les r plus grandes v.p. de l'endomorphisme B dans $\mathbb{R}^p$ sont différentes : $\lambda_1 > \dots > \lambda_r$. On définit pour $j = 1, \dots, r$, l'endomorphisme ${}^jB$ dans ${}^j\Lambda\mathbb{R}^p$ par :

$${}^jB(x^1 \wedge \dots \wedge x^j) = \sum_{h=1}^j x^1 \wedge \dots \wedge Bx^h \wedge \dots \wedge x^j, \quad x^l \in \mathbb{R}^p, \quad l = 1, \dots, j.$$

Soit $\lambda_{1j} = \sum_{l=1}^j \lambda_l$ la plus grande v.p. de ${}^jB$. On note ${}^jS_1$ le sous-espace propre correspondant à λ_{1j} et $({}^jS_1)^\perp$ son supplémentaire orthogonal.

Soit le processus $({}^jX_n)$ dans ${}^j\Lambda\mathbb{R}^p$ défini par ${}^jX_n = X_n^1 \wedge \dots \wedge X_n^j$.

On définit $h_j({}^jx) = \frac{\langle {}^jB {}^jx, {}^jx \rangle_M}{\langle {}^jx, {}^jx \rangle_M}$, ${}^jx \in {}^j\Lambda\mathbb{R}^p$.

On note :

$$\Delta_{nj} = 1 + a_n(\lambda_{1j} - h_j({}^jX_n)) \text{ et }$$

$$Q^j = {}^jX_1 + \sum_{n=1}^{+\infty} \frac{{}^jX_{n+1} - (I + a_n({}^jB - h_j({}^jX_n)I)) {}^jX_n}{\prod_{i=1}^n \Delta_{ij}}.$$

L'ensemble ${}^jE$ est $\{Q^j \notin ({}^jS_1)^\perp\}$.

Théorème 1.2.

Sous $(H1')$, $(H4')(b),(c)$ et $(H5)$, pour $k = 1, \dots, r$, si l'on suppose les r plus grandes v.p. de B distinctes, le processus (X_n^k) converge p.s. dans $\bigcap_{j=1}^k {}^jE$ vers un vecteur propre associé à la $k^{\text{ième}}$ plus grande v.p. λ_k de B

et $\sum_{n=1}^{+\infty} a_n |\lambda_k - \frac{\langle BX_n^k, X_n^k \rangle_M}{\|X_n\|_M^2}| < \infty$ p.s.

On obtient le même résultat sous les hypothèses alternatives $(H1)(a),(c)$, $(H4')$ et $(H5)$. ■

Dans la suite, il sera toujours sous-entendu que la convergence p.s. de (X_n^k) est dans $\bigcap_{j=1}^k E$.

L'analyse des données d'un vecteur aléatoire (v.a.), qui représente l'analyse des données sur toute une population, et l'analyse des données observées sur des individus tirés de la population considérée seront définies par la suite dans différents travaux d'estimation récursive en analyse des données.

Suivra alors une première série d'articles où ces données, notées $z_1, \dots, z_n, \dots$, seront supposées être des observations indépendantes et identiquement distribuées (i.i.d.) d'un v.a. Z :

Dans [49], Monnez présente tout d'abord l'analyse en composantes principales (ACP) d'un v.a. Z . En découle alors l'analyse factorielle multiple (AFM) d'un v.a., en remarquant que l'AFM est une ACP avec un choix particulier de métrique. Enfin, les facteurs de l'AFM ainsi que les v.p. associées sont estimés par a.s. en se basant sur l'article de 1998.

D'autre part, le même type d'étude est réalisé pour l'analyse canonique généralisée (ACG) dans [51].

On se propose de redonner les principaux résultats de cet article, sur lesquelles seront basées certaines études du deuxième chapitre de ce manuscrit de thèse.

1. Approximation Stochastique d'une fonction de régression linéaire

Soit R un v.a. de $\mathbb{R}^r$ et S un v.a. de $\mathbb{R}^s$.

On suppose que :

(H1) Il n'existe pas de relation affine entre les composantes de R .

On note $Cov(R)$ la matrice de covariance de R et $Cov(R, S) = \mathbb{E}[(R - \mathbb{E}[R])(S - \mathbb{E}[S])']$.

Soit A une matrice de dimension (r, s) et D une matrice-colonne de dimension $(s, 1)$.

$(Cov(R))^{-1} Cov(R, S)$ est la solution en A du problème de régression linéaire suivant : trouver (A, D) qui rende minimale $J(A, D) = \mathbb{E}[\|R - A'S - D\|^2]$, $\|\cdot\|$ représentant la norme euclidienne usuelle.

On note $A_1 = \begin{pmatrix} A \\ D' \end{pmatrix}$, matrice de dimension $(r+1, s)$ et $R_1 = \begin{pmatrix} R \\ 1 \end{pmatrix}$, v.a. de $\mathbb{R}^{r+1}$

Sous (H1), la matrice $\mathbb{E}[R_1 R_1']$ est définie positive.

La matrice $(Cov(R))^{-1} Cov(R, S)$ peut alors être obtenue de la manière suivante.

Théorème 1.3.

$A = (\text{Cov}(R))^{-1} \text{Cov}(R, S)$ et $D = \mathbb{E}[S] - A' \mathbb{E}[R]$ forment une solution du système linéaire :

$$\mathbb{E}[R_1 R_1'] A_1 = \mathbb{E}[R_1 S'] \quad \blacksquare$$

On considère désormais un échantillon i.i.d. $((R_n, S_n))$ de (R, S) .

Pour tout n , on note $R_{1n} = \begin{pmatrix} R_n \\ 1 \end{pmatrix}$, v.a. de $\mathbb{R}^{r+1}$

On définit le processus d'a.s. (A_{1n}) de $A_1 = (\mathbb{E}[R_1 R_1'])^{-1} \mathbb{E}[R_1 S']$, dans l'ensemble des matrices de dimension $(r+1, s)$, tel que :

$$A_{1,n+1} = A_{1n} - a_n (R_{1n} R_{1n}' A_{1n} - R_{1n} S_n'), \quad a_n > 0.$$

(A_n) , processus d'a.s. de $A = (\text{Cov}(R))^{-1} \text{Cov}(R, S)$, est alors obtenu en enlevant la dernière ligne du processus A_{1n} pour tout n .

On fait les nouvelles hypothèses :

(H2) (R, S) admet des moments d'ordre 4,

(H3) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$,

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

Théorème 1.4.

Sous (H1)-(H3), $\|A_{1n} - A_1\|$ et $\|A_n - A\|$ convergent vers 0 p.s. et dans L^2 . Si (H3) est remplacée par (H3'), on a en outre : $\sum_{n=1}^{+\infty} a_n \|A_{1n} - A_1\| < \infty$ et $\sum_{n=1}^{+\infty} a_n \|A_n - A\| < +\infty$ p.s. $\blacksquare$

2. Approximation stochastique des facteurs généraux

Soit Z un v.a. dans $\mathbb{R}^p$, décomposé en q sous-vecteurs $Z^1, \dots, Z^q$, de dimensions respectives $m_1, \dots, m_q$, avec $\sum_{k=1}^q m_k = p$.

Soit $((Z_n^1, \dots, Z_n^q))_{n \geq 1}$ un échantillon i.i.d. du v.a. $Z = (Z^1, \dots, Z^q)$ dans $\mathbb{R}^p$.

On note C la matrice de covariance de Z et C^k la matrice de covariance de Z^k , $k = 1, \dots, q$.

Pour $n \geq 1$ et $k = 1, \dots, q$, on définit :

$$\overline{Z}_n^k = \frac{1}{n} \sum_{i=1}^n Z_i^k, \quad M_n^k = \frac{1}{n} \sum_{i=1}^n Z_i^k (Z_i^k)' - \overline{Z}_n^k (\overline{Z}_n^k)';$$

M_n^k est un estimateur convergent de C^k .

Pour $n \geq 1$, on définit la matrice diagonale par blocs carrée d'ordre p ,

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

$$M_n \text{ est un estimateur convergent de } M^{-1} = \begin{pmatrix} C^1 & & \\ & \ddots & \\ & & C^q \end{pmatrix}.$$

Dans la suite, on notera : $C^{ij} = \text{Cov}(Z^i, Z^j)$, $i = 1, \dots, q$, $j \neq i$.

Les facteurs de l'ACG sont vecteurs propres de $B = MC$, que l'on peut décomposer en blocs $B^{ij} = (C^i)^{-1} C^{ij}$. Pour les estimer, on définit les processus d'a.s. (B_{1n}^{ij}) dans l'ensemble des matrices de dimension $(m_i + 1, m_j)$, (F_n^{ij}) dans l'ensemble des matrices de dimension $(m_j, 1)$ et (B_n^{ij}) dans l'ensemble des matrices de dimension (m_i, m_j) :

$$B_{1,n+1}^{ij} = B_{1n}^{ij} - a_n \left(Z_{1n}^i (Z_{1n}^i)' B_{1n}^{ij} - Z_{1n}^i (Z_n^j)' \right),$$

$$Z_{1n}^i = \begin{pmatrix} Z_n^i \\ 1 \end{pmatrix}, \quad B_{1n}^{ij} = \begin{pmatrix} B_n^{ij} \\ (F_n^{ij})' \end{pmatrix}.$$

On définit alors dans l'ensemble des matrices carrées d'ordre p le processus d'a.s. (B_n) de $B = MC$ tel que le bloc (i, j) de B_n soit B_n^{ij} :

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

En vue d'estimer r facteurs généraux de l'ACG, on définit, pour $l = 1, \dots, r$, le processus d'a.s. $(X_n) = ((X_n^1, \dots, X_n^r))$, de type Krasulina, des facteurs généraux $(\theta_1, \dots, \theta_r)$ dans $\mathbb{R}^{p*}$:

$$\begin{aligned} F_n(X_n^l) &= \frac{\langle B_n X_n^l, X_n^l \rangle_{M_n}}{\|X_n^l\|_{M_n}^2}, \\ Y_{n+1}^l &= X_n^l + a_n(B_n - F_n(X_n^l)I)X_n^l, \quad l = 1, \dots, r, \\ X_{n+1} &= \text{orth}_{M_n}(Y_{n+1}). \end{aligned}$$

On suppose que (avec les hypothèses numérotées comme dans l'article) :

(H1') Il n'existe pas de relation affine entre les composantes de Z ,

(H2') Z admet des moments d'ordre 4,

(H3'') $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\min_i \lambda_{\min}(\mathbb{E}[Z_1^i (Z_1^i)'])})$, $Z_1^i = \begin{pmatrix} Z^i \\ 1 \end{pmatrix}$.

Théorème 1.5.

Sous (H1'), (H2') et (H3''), en supposant que les r plus grandes v.p. de B sont différentes, pour $l = 1, \dots, r$, X_n^l converge p.s. vers un vecteur propre de B associé à sa $l^{\text{ième}}$ plus grande v.p. ■

Comme indiqué précédemment, on aura plusieurs fois recours à ce théorème dans le deuxième chapitre de la thèse. Celle-ci est consacrée à l'étude de l'ACP projetée dans le cas où on observe des réalisations i.i.d. d'un v.a. $Z = (R, S)$ partitionné en deux sous-vecteurs.

Ensuite, le cas où les données sont des réalisations d'un v.a. dont l'espérance varie dans le temps a été envisagé par Monnez dans [50].

On se propose de reprendre les parties principales de cet article, sur lesquelles seront basées certaines études des troisième et quatrième chapitres de la thèse.

1. Modèle

On considère un flux de données, représenté par une suite de vecteurs $z_1, \dots, z_n, \dots$ dans $\mathbb{R}^p$.

On suppose alors que :

- pour tout n , z_n est la réalisation d'un v.a. Z_n , d'espérance mathématique variable dans le temps ;
- les v.a. Z_n sont mutuellement indépendants ;
- pour tout n , on a la décomposition :

$$Z_n = \theta_n + R_n ,$$

$\theta_n = (\theta_n^1 \dots \theta_n^p)'$ étant un vecteur de $\mathbb{R}^p$, la loi du v.a. R_n ne dépendant pas de n ,
 $\mathbb{E}[R_n] = 0$, $Cov(R_n) = \Sigma$.

On a alors $\mathbb{E}[Z_n] = \theta_n$, $Cov(Z_n) = \Sigma$; $r_n = z_n - \mathbb{E}[Z_n]$ représente la donnée z_n centrée ; les hypothèses faites reviennent à supposer que les v.a. R_n constituent un échantillon i.i.d. d'un v.a. R , d'espérance mathématique nulle et de matrice de covariance Σ .

· pour $i = 1, \dots, p$, 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 :

$$\theta_n^i = \langle \beta^i, U_n^i \rangle ,$$

$\langle \cdot, \cdot \rangle$ désignant le produit scalaire euclidien usuel dans $\mathbb{R}^{n_i}$.

Si l'on note Z_n^i , respectivement R_n^i , la $i^{\text{ème}}$ composante de Z_n , respectivement R_n , on a alors le modèle de régression linéaire :

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

On souhaite alors réaliser une ACP du v.a. R , aussi appelée ACP partielle, dans $\mathbb{R}^p$ que l'on munit d'une métrique M . On étudie ici l'estimation des facteurs de cette analyse.

2. Définition des processus

Soit M_n un estimateur de M au temps n et $(a_n)_{n \geq 1}$ une suite de nombres réels positifs.

Pour $i = 1, \dots, p$, on définit le processus d'a.s. (B_n^i) de β^i tel que

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

On définit alors :

$$\Theta_n^i = \langle B_n^i, U_n^i \rangle, \quad \Theta_n = (\Theta_n^1 \dots \Theta_n^p)',$$

$$C_n = M_n (Z_n Z_n' - \Theta_n \Theta_n').$$

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

$$\begin{aligned} F_n(X_n^i) &= \frac{\langle C_n X_n^i, X_n^i \rangle_{M_n^{-1}}}{\|X_n^i\|_{M_n^{-1}}^2}, \\ Y_{n+1}^i &= X_n^i + a_n(C_n - F_n(X_n^i)I)X_n^i, \\ X_{n+1} &= \text{orth}_{M_n^{-1}}(Y_{n+1}) \end{aligned}$$

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

3. Hypothèses

(H2) (a) M_n T_n -mesurable,

(b) $M_n \rightarrow M$ p.s.,

(c) $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.,

(H6) (a) $\max_i \sup_n \|U_n^i\| < \infty$,

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

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

(H3') $a_n > 0$, $\sum_{n=1}^{+\infty} \min_{j \in I_{il}} a_j = \infty$ pour $i = 1, \dots, p$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$.

4. Théorèmes

Théorème 1.6.

Sous (H3') et (H6), pour $i = 1, \dots, p$, $B_n^i \rightarrow \beta^i$ et $\Theta_n^i - \theta_n^i \rightarrow 0$ p.s. ■

Théorème 1.7.

Sous (H3) et (H6), pour $i = 1, \dots, p$:

1) pour $(\frac{1}{2} < \alpha < 1)$ ou $(\alpha = 1, \frac{2\lambda_i c}{r_i} > 1)$:

$\overline{\lim} n^\alpha \mathbb{E}[\|B_n^i - \beta^i\|^2] < \infty$ et $\overline{\lim} n^\alpha \mathbb{E}[\|\Theta_n^i - \theta_n^i\|^2] < \infty$;

2) pour $(\alpha = 1, \frac{2\lambda_{ic}}{r_i} = 1)$:

$$\overline{\lim} \frac{n}{\ln(n)} \mathbb{E}[||B_n^i - \beta^i||^2] < \infty \text{ et } \overline{\lim} \frac{n}{\ln(n)} \mathbb{E}[||\Theta_n^i - \theta_n^i||^2] < \infty ;$$

3) pour $(\alpha = 1, \frac{2\lambda_{ic}}{r_i} < 1)$:

$$\overline{\lim} n^{\frac{2\lambda_{ic}}{r_i}} \mathbb{E}[||B_n^i - \beta^i||^2] < \infty \text{ et } \overline{\lim} n^{\frac{2\lambda_{ic}}{r_i}} \mathbb{E}[||\Theta_n^i - \theta_n^i||^2] < \infty. \blacksquare$$

Théorème 1.8.

Sous (H2), (H3) avec $\frac{2}{3} < \alpha \leq 1$ et (H6), si l'on suppose que R admet des moments d'ordre $4r$ et que les r plus grandes v.p. de $C = M\Sigma$ sont distinctes, alors, pour $i = 1, \dots, r$, X_n^i converge p.s. vers un vecteur propre de C associé à sa $i^{\text{ème}}$ plus grande v.p. ■

On fera référence à ces théorèmes dans les troisième et quatrième chapitres de la thèse, respectivement consacrées à l'étude de l'ACP partielle d'un flux de données dont l'espérance varie dans le temps puis à l'ACP partielle d'un flux de données dont l'espérance et la matrice de covariance varient dans le temps.

Pour préciser les choses, l'étude de l'ACP partielle d'un flux de données dont l'espérance varie dans le temps que nous proposons ici a pour objectif de généraliser l'article de (Monnez, 2008) :

- en étudiant d'autres processus (C_n) ,
 - en n'imposant pas de modèle de représentation a priori pour l'espérance θ_n
- (le cas du modèle de variation linéaire sera traité comme cas particulier).

Les trois derniers articles auxquels nous venons de faire référence s'inscrivent dans la nouvelle problématique actuelle des “Big Data”, où le nombre de données récoltées est gigantesque, ces dernières arrivant souvent en temps réel.

Dans le contexte actuel, une prise de décision rapide est souvent indispensable : on doit donc être capable de fournir la meilleure réponse possible en un temps donné.

Actuellement, l'arrivée de ces données massives est, de manière générale, considérée comme une contrainte. De notre côté, nous considérons les données comme des ressources, au même titre que l'espace de stockage ou le temps dont on dispose pour l'analyse. Le fait de disposer d'un grand nombre de données est alors vu comme une chance.

Cette différence de point de vue est ainsi répercutée dans la manière d'envisager le traitement des données : alors que dans la majorité des cas, on se ramène à l'ère “pré-Big Data ” en échantillonnant les données puis en appliquant des algorithmes sophistiqués mais lourds d'un point de vue informatique, nous privilégions des algorithmes récursifs très simples, moins précis mais qui permettent de prendre en compte et tirer profit de plus voire de toutes les données disponibles.

Nous donnons maintenant le plan adopté pour la présentation de ce travail. Comme nous l'avons déjà mentionné, nous nous intéresserons à l'ACP projetée de données séquentielles dans le chapitre 2 de l'étude. Ensuite, le troisième (resp. quatrième) chapitre sera consacrée à l'ACP d'un flux de données d'espérance (resp. d'espérance et de matrice de covariance) variables(s) dans le temps. Enfin, nous donnerons des résultats de simulation dans le cinquième chapitre avant de conclure sur l'apport de cette étude.

Chapitre 2

ACP projetée de données séquentielles

Sommaire

2.1	Introduction	25
2.1.1	Cadre d'étude	25
2.1.2	Présentation de l'ACP projetée	25
2.1.3	ACP d'un vecteur aléatoire	26
2.1.4	Facteurs de l'ACP projetée	28
2.2	Cas où il n'y a pas de relation affine entre les composantes du vecteur explicatif . . .	28
2.2.1	A.S. d'une fonction de régression linéaire	30
2.2.2	A.S. des facteurs de l'ACP projetée	31
2.2.3	Variantes	35
2.2.4	Cas de l'Analyse Canonique de deux vecteurs aléatoires	37
2.3	Cas où les composantes du vecteur explicatif sont les indicatrices des modalités exclusives d'une variable aléatoire nominale	46
2.3.1	A.S. de la fonction de régression linéaire	49
2.3.2	A.S. des facteurs de l'ACP projetée	52
2.3.3	Cas particulier de l'Analyse Factorielle Discriminante	52
2.3.4	Cas particulier de l'Analyse Factorielle des Correspondances	56
2.4	Annexes du chapitre 2	62
2.4.1	Démonstration du théorème 10	62
2.4.2	Hypothèses	65

2.1 Introduction

2.1.1 Cadre d'étude

On observe p caractères quantitatifs $r^1, \dots, r^p$ et q caractères quantitatifs $s^1, \dots, s^q$ sur des individus : on obtient des vecteurs de données (r_i, s_i) dans $\mathbb{R}^{p+q}$. On se place dans le cas où ces vecteurs de données arrivent séquentiellement dans le temps : on observe (r_n, s_n) au temps n ; on a une suite de vecteurs de données $(r_1, s_1), \dots, (r_n, s_n), \dots$.

On suppose que cette suite constitue un échantillon indépendant et identiquement distribué (i.i.d.) d'un vecteur aléatoire (v.a.) (R, S) de dimension $p + q$ défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$: Ω représente une population d'où on extrait un échantillon.

On étudie l'analyse en composantes principales (ACP) projetée de R par rapport à S , c'est-à-dire l'ACP du régressé linéaire au sens des moindres carrés de R par rapport à S , qui représente l'ACP projetée effectuée sur la population Ω , dont on va chercher à estimer au temps n les résultats à partir de l'échantillon dont on dispose à ce temps.

Soit un résultat θ de cette analyse, par exemple une valeur propre (v.p.), un facteur (on considère ici le cas d'un facteur). Plutôt que d'effectuer au temps n une estimation de θ en utilisant toutes les données du passé disponibles, on va effectuer une estimation réursive de θ : disposant d'une estimation θ_n obtenue à partir des observations $(r_1, s_1), \dots, (r_{n-1}, s_{n-1})$, on introduit l'observation (r_n, s_n) et on définit une nouvelle estimation θ_{n+1} fonction de θ_n et de (r_n, s_n) : $\theta_{n+1} = f_n(\theta_n; r_n, s_n)$.

On définit pour cela un processus d'approximation stochastique (a.s.) comme dans [10].

L'ACP projetée a pour cas particulier l'analyse canonique, l'analyse factorielle discriminante ainsi que l'analyse factorielle des correspondances ; pour chacune de ces analyses, on définit des processus spécifiques d'estimation des facteurs.

2.1.2 Présentation de l'ACP projetée

Soit deux v.a. $R = (R^1, \dots, R^p)$ de dimension p et $S = (S^1, \dots, S^q)$ de dimension q .

Dans un premier temps, on effectue la régression linéaire au sens des moindres carrés de R par rapport à S : on détermine la matrice A de dimension (q, p) et la matrice-colonne D de dimension $(p, 1)$ qui rendent minimale $J(A, D) = \mathbb{E}[\|R - A'S - D\|^2]$. On appellera régressé de R par rapport à S le v.a. de dimension p noté $\hat{\mathbb{E}}[R|S] = A'S + D$.

L'ACP projetée de R par rapport à S est l'ACP du régressé $\hat{\mathbb{E}}[R|S] = A'S + D$ dans $\mathbb{R}^p$ muni d'une métrique M .

2.1.3 ACP d'un vecteur aléatoire

Soit Z un v.a. dans $\mathbb{R}^p$, rappelons alors le principe de l'ACP de Z : au l -ième pas, on cherche une combinaison linéaire (c.l.) des composantes centrées de Z , $\theta_l'(Z - \mathbb{E}[Z])$, de variance maximale sous les contraintes de ne pas être corrélée aux c.l. déterminées aux pas $1, \dots, l-1$ et la contrainte d'être M^{-1} -normée.

Autrement dit, on cherche θ_l solution du système :

$$\begin{cases} \theta_l' \text{Cov}(Z) \theta_l \text{ max} \\ \theta_l' \text{Cov}(Z) \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases} \quad (2.1)$$

Un facteur θ_l solution de ce problème est vecteur propre de $M \text{Cov}(Z)$ associé à la l -ième plus grande v.p. λ_l . (On supposera ici et dans toute la suite que la l -ième v.p. λ_l n'est pas nulle.)

Lemme 1. Soit le problème qui consiste à trouver θ_l solution du problème :

$$\begin{cases} \theta_l' \text{Cov}(Z) \theta_l \text{ max} \\ \theta_l' M^{-1} \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases} \quad (2.2)$$

Toute solution θ_l de (2.1) est solution de (2.2), et réciproquement. ■

Démonstration.

Montrons cette proposition par récurrence sur l .

Initialisation : Pour $l = 1$, on n'a pas la contrainte de non-corrélation : les problèmes sont les mêmes et, d'après un théorème de maximisation sous contraintes d'une forme quadratique, la solution est θ_1 , vecteur propre de $MCov(Z)$ associé à la plus grande v.p. λ_1 .

Hérédité : Supposons la proposition vérifiée jusqu'au rang $l - 1$ et montrons qu'elle l'est au rang l .

Soit donc le problème :

$$\begin{cases} \theta_l' Cov(Z) \theta_l \max \\ \theta_l' Cov(Z) \theta_j = 0, \quad j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases}$$

Par hypothèse de récurrence, on sait que : $\forall j = 1, \dots, l-1$, chaque θ_j est vecteur propre de $MCov(Z)$ associé à la j -ième plus grande v.p. non nulle λ_j .

On a donc :

$$\begin{aligned} MCov(Z) \theta_j &= \lambda_j \theta_j \\ \Leftrightarrow Cov(Z) \theta_j &= \lambda_j M^{-1} \theta_j \end{aligned}$$

Ainsi, le problème peut se réécrire :

$$\begin{cases} \theta_l' Cov(Z) \theta_l \max \\ \theta_l' \lambda_j M^{-1} \theta_j = 0, \quad j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases}$$

Comme $\lambda_j \neq 0$, ceci équivaut à :

$$\begin{cases} \theta_l' Cov(Z) \theta_l \max \\ \theta_l' M^{-1} \theta_j = 0, \quad j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases}$$

θ_l , solution de ce problème, est vecteur propre de $MCov(Z)$ associé à la l -ième plus grande v.p. non nulle λ_l .

□

Pour davantage de précisions concernant l'ACP d'un v.a., on pourra consulter de manière détaillée la partie 2.1 de [49].

2.1.4 Facteurs de l'ACP projetée

On déduit des deux sections précédentes que le l -ième facteur θ_l de l'ACP projetée de R par rapport à S est vecteur propre de $B = MCov\left(\hat{\mathbb{E}}[R|S]\right)$ associé à la l -ième plus grande v.p. λ_l .

2.2 Cas où il n'y a pas de relation affine entre les composantes du vecteur explicatif

Dans cette partie, on s'intéresse au cas où il n'y a pas de relation affine entre les composantes de S . Cette hypothèse nous assure en particulier l'inversibilité de la matrice de covariance de S notée dans la suite $Cov(S)$.

Tout d'abord, on revient à la régression linéaire de R par rapport à S : on cherche les matrices A et D .

Proposition 2. *Dans le cas où il n'y a pas de relation affine entre les composantes de S , $A = \left(Cov(S)\right)^{-1} Cov(S, R)$ et $D = \mathbb{E}[R] - A'\mathbb{E}[S]$ rendent minimale $\mathbb{E}\left[||R - A'S - D||^2\right] : \hat{\mathbb{E}}[R|S] = A'(S - \mathbb{E}[S]) + \mathbb{E}[R]$.*

En outre, $Cov\left(\hat{\mathbb{E}}[R|S]\right) = Cov(R, S)\left(Cov(S)\right)^{-1} Cov(S, R)$. ■

Démonstration.

Soit à minimiser : $J(A, D) = \mathbb{E}\left[||R - A'S - D||^2\right]$.

On a à résoudre le système :

$$\begin{cases} \nabla_D(J) = -2\mathbb{E}[R - A'S - D] = 0 \end{cases} \quad (2.3)$$

$$\begin{cases} \nabla_A(J) = -2\mathbb{E}\left[(R - A'S - D)S'\right] = 0 \end{cases} \quad (2.4)$$

$$(2.3) \Leftrightarrow D = \mathbb{E}[R] - A' \mathbb{E}[S]$$

$$\begin{aligned} \text{Alors, (2.4)} \quad &\Leftrightarrow \mathbb{E} \left[\left((R - \mathbb{E}[R]) - A'(S - \mathbb{E}[S]) \right) S' \right] = 0 \\ &\Leftrightarrow A = \left(\mathbb{E} [S(S - \mathbb{E}[S])'] \right)^{-1} \mathbb{E} [S(R - \mathbb{E}[R])'] = \left(\text{Cov}(S) \right)^{-1} \text{Cov}(S, R). \end{aligned}$$

$$\text{En outre, } \text{Cov}(\hat{\mathbb{E}}[R|S]) = A' \text{Cov}(S) A = \text{Cov}(R, S) A = \text{Cov}(R, S) \left(\text{Cov}(S) \right)^{-1} \text{Cov}(S, R).$$

□

Proposition 3. Soit $S_1 = \begin{pmatrix} S \\ 1 \end{pmatrix}$ de dimension $(q+1, 1)$. Alors, $A_1 = \begin{pmatrix} A \\ D' \end{pmatrix}$ de dimension $(q+1, p)$ est solution du système en X : $\mathbb{E}[S_1 S_1' X - S_1 R'] = 0$. ■

Démonstration.

$$\mathbb{E}[S_1 S_1' A_1] = \mathbb{E} \begin{pmatrix} SS' & S \\ S' & 1 \end{pmatrix} \cdot \begin{pmatrix} A \\ D' \end{pmatrix} = \begin{pmatrix} \mathbb{E}[SS']A + \mathbb{E}[S]D' \\ \mathbb{E}[S']A + D' \end{pmatrix}.$$

$$\text{Comme } D' = \mathbb{E}[R'] - \mathbb{E}[S']A,$$

$$\begin{aligned} \mathbb{E}[SS']A + \mathbb{E}[S]D' &= \text{Cov}(S)A + \mathbb{E}[S]\mathbb{E}[R'] \quad \text{et} \quad \mathbb{E}[S']A + D' = \mathbb{E}[R']. \\ &= \text{Cov}(S, R) + \mathbb{E}[S]\mathbb{E}[R'] \\ &= \mathbb{E}[SR'] \end{aligned}$$

$$\text{Ainsi, } \mathbb{E}[S_1 S_1' A_1] = \begin{pmatrix} \mathbb{E}[SR'] \\ \mathbb{E}[R'] \end{pmatrix} = \mathbb{E} \left[\begin{pmatrix} S \\ 1 \end{pmatrix} \cdot R' \right] = \mathbb{E}[S_1 R']. \quad \square$$

Proposition 4. Soit une c.l. des composantes de R , $\theta'R$.

On a : $\hat{\mathbb{E}}[\theta'R|S] = \theta'\hat{\mathbb{E}}[R|S]$. ■

Démonstration.

D'après la proposition 2, $\hat{\mathbb{E}}[\theta'R|S] = \left(\left(\text{Cov}(S) \right)^{-1} \text{Cov}(S, \theta'R) \right)' (S - \mathbb{E}[S]) + \mathbb{E}[\theta'R]$.

Or, $\text{Cov}(S, \theta'R) = \mathbb{E}[(S - \mathbb{E}[S])R'\theta] = \mathbb{E}[(S - \mathbb{E}[S])R']\theta = \text{Cov}(S, R)\theta$.

Donc, $\hat{\mathbb{E}}[\theta'R|S] = \theta' \text{Cov}(R, S) \left(\text{Cov}(S) \right)^{-1} (S - \mathbb{E}[S]) + \theta' \mathbb{E}[R] = \theta' \hat{\mathbb{E}}[R|S]$. □

2.2.1 A.S. d'une fonction de régression linéaire

Comme $A_1 = \begin{pmatrix} A \\ D' \end{pmatrix}$ est solution du système $\mathbb{E}[S_1 S_1' X - S_1 R'] = 0$, on définit le processus d'a.s. (A_{1n}) de A_1 tel que :

$$A_{1,n+1} = A_{1n} - a_n (S_{1n} S_{1n}' A_{1n} - S_{1n} R_n')$$

où (R_n, S_n) est un échantillon i.i.d. de (R, S) et $S_{1n} = \begin{pmatrix} S_n \\ 1 \end{pmatrix}$.

On définit les hypothèses :

(H1) Il n'y a pas de relation affine entre les composantes de S ,

(H2) (R, S) admet des moments d'ordre 4,

(H3) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$,

(H3') $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[S_1 S_1'])})$.

On note A_n la matrice de dimension (q, p) obtenue, pour tout n , en enlevant la dernière ligne de A_{1n} .

Théorème 5. *Sous (H1), (H2) et (H3), (A_n) converge vers A presque sûrement (p.s.) et dans L^2 et, si on remplace (H3) par (H3'), on a en outre : $\sum_{n=1}^{+\infty} a_n \|A_n - A\| < \infty$ p.s. ■*

Démonstration.

Il suffit d'appliquer le théorème 3 de [51] (rappelé dans le chapitre 1 p.18) en permutant les rôles de R et S . $\square$

2.2.2 A.S. des facteurs de l'ACP projetée

On effectue maintenant l'ACP du v.a. $\hat{\mathbb{E}}[R|S]$ dans $\mathbb{R}^p$ muni de la métrique M ; le l -ième facteur θ_l , dans $(\mathbb{R}^p)^*$, est vecteur propre de $B = MCov(\hat{\mathbb{E}}[R|S]) = MCov(R, S)A$ associé à la l -ième plus grande v.p. λ_l .

On suppose que l'on dispose d'une suite d'estimateurs convergents p.s. (M_n) et (N_n) de M et M^{-1} respectivement.

On note $\bar{R}_{n-1}$ et $\bar{S}_{n-1}$ les moyennes d'ordre $n-1$ des observations R_i et S_i respectivement.

On définit alors le processus d'a.s. $(X_n) = ((X_n^1, \dots, X_n^r))$ des r premiers facteurs de l'ACP projetée de R par rapport à S :

$$\begin{cases} B_n = M_n(R_n S_n' - \bar{R}_{n-1} \bar{S}_{n-1}') A_n \\ F_n(X_n^l) = \frac{\langle B_n X_n^l, X_n^l \rangle_{N_n}}{\langle X_n^l, X_n^l \rangle_{N_n}} \\ Y_{n+1}^l = X_n^l + a_n (B_n - F_n(X_n^l) I) X_n^l, l = 1, \dots, r \\ X_{n+1} = orth_{N_n}(Y_{n+1}) \end{cases}$$

$X_{n+1} = orth_{N_n}(Y_{n+1})$ signifie que, pour obtenir $X_{n+1} = (X_{n+1}^1, \dots, X_{n+1}^r)$, on orthogonalise au sens de Gram-Schmidt par rapport à N_n le r -uplet $Y_{n+1} = (Y_{n+1}^1, \dots, Y_{n+1}^r)$.

Dans toute la suite, on supposera que les r plus grandes v.p. de B sont distinctes.

On définit les hypothèses :

(H2') (R, S) admet des moments d'ordre $2r$.

Soit T_n la tribu du passé au temps n , engendrée par $R_1, \dots, R_{n-1}, S_1, \dots, S_{n-1}, A_{11}$ et X_1 , alors :

- (H4) Pour tout n , (a) M_n est T_n -mesurable,
 (b) M_n converge vers M p.s.,
 (c) $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.,

- (H4') Pour tout n , (a) N_n est T_n -mesurable,
 (b) N_n converge vers N p.s.,
 (c) $\sum_{n=1}^{+\infty} a_n ||N_n - N|| < \infty$ p.s.

Théorème 6. Sous (H1), (H2'), (H3'), (H4) et (H4'), X_n^l converge p.s. vers un vecteur propre de $B = MCov(R, S)A$ associé à la l -ième plus grande v.p. λ_l . ■

Démonstration.

On souhaite appliquer le théorème 4 de [10], on vérifie alors que la première série d'hypothèses nécessaire à son application est satisfaite.

Avant cela, rappelons ces hypothèses :

- (1)(a) $\sum_{n=1}^{+\infty} a_n ||\mathbb{E}[B_n|T_n] - B|| < \infty$ p.s.,
 (c) $\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||B_n - B||^j | T_n] < \infty$ p.s. pour $j = 2, 3, \dots, 2r$,
 (4') Pour tout n , (a) N_n est T_n -mesurable,
 (b) N_n converge vers N p.s.,
 (c) $\sum_{n=1}^{+\infty} a_n ||N_n - N|| < \infty$ p.s.,
 (5) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$ et $\sum_{n=1}^{+\infty} a_n^2 < \infty$.

Montrons que (1)(a) est vérifiée :

On pose $B_n - B = M_n C_n A_n - MCA$ avec : $\begin{cases} C_n = R_n S'_n - \bar{R}_{n-1} \bar{S}'_{n-1} \\ C = \mathbb{E}[RS'] - \mathbb{E}[R]\mathbb{E}[S'] \end{cases}$

Dans la suite, on utilisera la décomposition :

$$B_n - B = M_n C(A_n - A) + M_n (C_n - C)A_n + (M_n - M)CA.$$

Comme M_n et A_n sont T_n -mesurables respectivement sous (H4)(a) et par définition,

$$\mathbb{E}[B_n|T_n] - B = M_n C(A_n - A) + M_n (\mathbb{E}[C_n|T_n] - C)A_n + (M_n - M)CA.$$

a) M_n tendant p.s. vers M sous (H4)(b) , il existe une variable aléatoire (v.a.) D positive, finie p.s. telle que :

$$\sum_{n=1}^{+\infty} a_n \|M_n C(A_n - A)\| \leq D \sum_{n=1}^{+\infty} a_n \|A_n - A\| < \infty \text{ d'après le théorème 5 sous (H1), (H2')} \left((H2') \Rightarrow (H2) \right) \text{ et (H3')}.$$

b) D'autre part, $\sum_{n=1}^{+\infty} a_n \|(M_n - M)CA\| \leq \|CA\| \sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ sous (H4)(c).

c) Commençons par un résultat dont on se servira à plusieurs reprises dans la suite :

Lemme 7. $\mathbb{E} \left[\|\mathbb{E}[R] - \bar{R}_{n-1}\| \right] = O\left(\frac{1}{\sqrt{n}}\right).$ ■

Démonstration.

$$\mathbb{E} \left[\|\mathbb{E}[R] - \bar{R}_{n-1}\| \right] = \mathbb{E} \left[\left(\sum_{i=1}^p (\mathbb{E}[R^i] - \bar{R}_{n-1}^i)^2 \right)^{\frac{1}{2}} \right] \text{ en prenant la norme de Hilbert-Schmidt} \quad (2.5)$$

$$\leq \left(\mathbb{E} \left[\sum_{i=1}^p (\mathbb{E}[R^i] - \bar{R}_{n-1}^i)^2 \right] \right)^{\frac{1}{2}} \text{ par l'inégalité de Jensen} \quad (2.6)$$

$$\leq \left(\sum_{i=1}^p \text{Var}[\bar{R}_{n-1}^i] \right)^{\frac{1}{2}} \text{ par linéarité de l'espérance} \quad (2.7)$$

$$\leq \left(\sum_{i=1}^p \frac{\sigma_i^2}{n-1} \right)^{\frac{1}{2}} \text{ avec } \sigma_i^2 = \text{Var}[R^i] < \infty \text{ sous (H2')} \quad (2.8)$$

$$= O\left(\frac{1}{\sqrt{n}}\right).$$

□

$$\|\mathbb{E}[C_n - C|T_n]\| = \|E[R_n S'_n - \bar{R}_{n-1} \bar{S}'_{n-1} | T_n] - \mathbb{E}[R S'] + \mathbb{E}[R] \mathbb{E}[S']\|.$$

Mais R_n et S_n sont indépendants de T_n tandis que $R_n \stackrel{\text{loi}}{=} R$ et $S_n \stackrel{\text{loi}}{=} S$ donc :

$$\|\mathbb{E}[C_n - C|T_n]\| = \|\mathbb{E}[R] \mathbb{E}[S'] - \bar{R}_{n-1} \bar{S}'_{n-1}\| \quad (2.9)$$

$$\begin{aligned} &\leq \|(\mathbb{E}[R] - \bar{R}_{n-1}) \mathbb{E}[S']\| + \|\bar{R}_{n-1} (\mathbb{E}[S'] - \bar{S}'_{n-1})\| \\ &\leq c \|\mathbb{E}[R] - \bar{R}_{n-1}\| + \|\bar{R}_{n-1}\| \cdot \|\mathbb{E}[S'] - \bar{S}'_{n-1}\| \text{ sous (H2')}. \end{aligned} \quad (2.10)$$

Pour le premier terme, on a :

$$\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\left| \mathbb{E}[R] - \bar{R}_{n-1} \right| \right] \leq \sum_{n=1}^{+\infty} a_n O\left(\frac{1}{\sqrt{n}}\right) \text{ d'après le lemme précédent} \quad (2.11)$$

$$< \infty \text{ sous (H3')}.$$

Donc : $\sum_{n=1}^{+\infty} a_n \left| \mathbb{E}[R] - \bar{R}_{n-1} \right| < \infty$ p.s. (1)

De même, pour le second terme, $\sum_{n=1}^{+\infty} a_n \left| \mathbb{E}[S'] - \bar{S}'_{n-1} \right| < \infty$ p.s.

Comme $\bar{R}_{n-1} \rightarrow \mathbb{E}[R]$ p.s., on a : $\sum_{n=1}^{+\infty} a_n \left| \bar{R}_{n-1} \right| \cdot \left| \mathbb{E}[S'] - \bar{S}'_{n-1} \right| < \infty$ p.s.

On déduit de ces deux résultats que : $\sum_{n=1}^{+\infty} a_n \left| \mathbb{E}[C_n - C|T_n] \right| < \infty$ p.s.

Enfin, comme $M_n \rightarrow M$ p.s. et $A_n \rightarrow A$ p.s., on a : $\sum_{n=1}^{+\infty} a_n \left| M_n \right| \cdot \left| \mathbb{E}[C_n - C|T_n] \right| \cdot \left| A_n \right| < \infty$ p.s.

D'après les conclusions de a), b), c), on obtient finalement que :

$$\sum_{n=1}^{+\infty} a_n \left| \mathbb{E}[B_n|T_n] - B \right| < \infty \text{ p.s.}$$

(1)(a) est vérifiée ; reste à montrer que (1)(c) l'est également.

Tout d'abord,

$$\left| B_n - B \right|^j \leq 3^{j-1} \left(\left| M_n C(A_n - A) \right|^j + \left| M_n (C_n - C) A_n \right|^j + \left| (M_n - M) C A \right|^j \right).$$

Comme $M_n \rightarrow M$ p.s. (resp. $A_n \rightarrow A$ p.s.) et M_n (resp. A_n) est T_n -mesurable, il existe deux v.a. G et H positives et finies p.s. telles que :

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[\left| M_n C(A_n - A) \right|^j | T_n \right] \leq G \sum_{n=1}^{+\infty} a_n^j < \infty \text{ p.s. sous (H3')} \text{ et}$$

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[\left| (M_n - M) C A \right|^j | T_n \right] \leq H \sum_{n=1}^{+\infty} a_n^j < \infty \text{ p.s. sous (H3')}.$$

D'autre part,

$$\left| C_n - C \right|^j = \left| R_n S'_n - \mathbb{E}[R S'] + \mathbb{E}[R] \mathbb{E}[S'] - \bar{R}_{n-1} \bar{S}'_{n-1} \right|^j \quad (2.12)$$

$$= \left| R_n S'_n - \mathbb{E}[R S'] + (\mathbb{E}[R] - \bar{R}_{n-1}) \mathbb{E}[S'] + \bar{R}_{n-1} (\mathbb{E}[S'] - \bar{S}'_{n-1}) \right|^j \quad (2.13)$$

$$\leq 3^{j-1} \left(\left| R_n S'_n - \mathbb{E}[R S'] \right|^j + \left| (\mathbb{E}[R] - \bar{R}_{n-1}) \mathbb{E}[S'] \right|^j + \left| \bar{R}_{n-1} (\mathbb{E}[S'] - \bar{S}'_{n-1}) \right|^j \right).$$

Pour le premier terme, comme R_n et S_n sont indépendants de T_n tandis que $R_n \stackrel{\text{loi}}{=} R$ et $S_n \stackrel{\text{loi}}{=} S$, on a :

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[||R_n S'_n - \mathbb{E}[RS']||^j | T_n \right] = \sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[||RS' - \mathbb{E}[RS']||^j \right] < \infty \text{ p.s. d'après (H2')} \text{ et (H3')}.$$

Pour le deuxième terme, comme $\bar{R}_{n-1}$ est T_n -mesurable et $\bar{R}_{n-1} \rightarrow \mathbb{E}[R]$ p.s., on a :

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[||(\mathbb{E}[R] - \bar{R}_{n-1}) \mathbb{E}[S']||^j | T_n \right] &= \sum_{n=1}^{+\infty} a_n^j ||(\mathbb{E}[R] - \bar{R}_{n-1}) \mathbb{E}[S']||^j \text{ car } \bar{R}_{n-1} \text{ est } T_n\text{-mesurable} \\ &\leq c \sum_{n=1}^{+\infty} a_n^j ||(\mathbb{E}[R] - \bar{R}_{n-1})||^j \text{ sous (H2')} < \infty \text{ d'après (1)}. \end{aligned}$$

En raisonnant de la même manière pour le troisième terme, on obtient que :

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[||M_n(C_n - C)A_n||^j | T_n \right] < \infty \text{ p.s. et par suite } \sum_{n=1}^{+\infty} a_n^j \mathbb{E} [||B_n - B||^j | T_n] < \infty \text{ p.s.}$$

ce qui signifie que (H1)(c) est vérifiée. □

2.2.3 Variantes

On rappelle que $B = MCov(R, S)A = M(\mathbb{E}[RS'] - \mathbb{E}[R]\mathbb{E}[S'])A$.

On peut définir des variantes du processus X_n , en utilisant au pas n , au lieu d'une seule observation (R_n, S_n) , plusieurs observations de (R, S) ou même toutes les observations faites jusqu'à ce pas.

Dans ce dernier cas, on introduit $(R_1, \dots, R_n, \dots)$ (resp. $(S_1, \dots, S_n, \dots)$) échantillon i.i.d. de R (resp. S) et le processus $W_n = \frac{1}{n} \sum_{i=1}^n R_i S'_i$ et on définit $B_n = M_n(W_{n-1} - \bar{R}_{n-1} \bar{S}'_{n-1})A_n$.

On définit alors le nouveau processus d'a.s. $(X_n) = ((X_n^1, \dots, X_n^r))$ des r premiers facteurs de l'ACP projetée de R par rapport à S :

$$\begin{cases} B_n = M_n(W_{n-1} - \bar{R}_{n-1} \bar{S}'_{n-1})A_n \\ F_n(X_n^l) = \frac{\langle B_n X_n^l, X_n^l \rangle_{N_n}}{\langle X_n^l, X_n^l \rangle_{N_n}} \\ Y_{n+1}^l = X_n^l + a_n (B_n - F_n(X_n^l)I) X_n^l, l = 1, \dots, r \\ X_{n+1} = orth_{N_n}(Y_{n+1}) \end{cases}$$

Théorème 8. *Sous les hypothèses (H1), (H2), (H3'), (H4')(b),(c) et (H4)(b),(c), X_n^l converge p.s. vers un vecteur propre de $B = MCov(R, S)A$ associé à la l-ième plus grande v.p. λ_l . ■*

Démonstration.

On souhaite encore appliquer le théorème 4 de [10], on vérifie cette fois-ci que la seconde série d'hypothèses nécessaire à son application est satisfaite.

Avant cela, rappelons ces hypothèses :

$$(1') \sum_{n=1}^{+\infty} a_n \|B_n - B\| < \infty \text{ p.s.,}$$

$$(4') \text{ Pour tout } n, \quad (b) N_n \text{ converge vers } N \text{ p.s.,}$$

$$(c) \sum_{n=1}^{+\infty} a_n \|N_n - N\| < \infty \text{ p.s.,}$$

$$(5) a_n > 0, \sum_{n=1}^{+\infty} a_n = \infty \text{ et } \sum_{n=1}^{+\infty} a_n^2 < \infty.$$

Il reste donc à montrer que (1') est vérifiée. Pour cela, on utilise la même décomposition que dans la démonstration du théorème 6 :

$$B_n - B = M_n C(A_n - A) + M_n(C_{n-1} - C)A_n + (M_n - M)CA \text{ avec : } \begin{cases} C_{n-1} = W_{n-1} - \bar{R}_{n-1} \bar{S}'_{n-1} \\ C = \mathbb{E}[RS'] - \mathbb{E}[R]\mathbb{E}[S'] \end{cases}$$

On sait déjà que :

$$\sum_{n=1}^{+\infty} a_n \|A_n - A\| < \infty \text{ p.s. sous (H1), (H2) et (H3')}$$

d'après le théorème 5 et que :

$$M_n \rightarrow M \text{ p.s. et } \sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty \text{ p.s. sous (H4)(b),(c).}$$

En outre,

$$\sum_{n=1}^{+\infty} a_n \|C_{n-1} - C\| \leq \sum_{n=1}^{+\infty} a_n \|W_{n-1} - \mathbb{E}[RS']\| + \sum_{n=1}^{+\infty} a_n \|(\mathbb{E}[R] - \bar{R}_{n-1})\mathbb{E}[S']\| + \sum_{n=1}^{+\infty} a_n \|\bar{R}_{n-1}(\mathbb{E}[S'] - \bar{S}'_{n-1})\|.$$

D'après le lemme 7, ces trois termes sont finis p.s.

Ainsi,

$$\sum_{n=1}^{+\infty} a_n \|B_n - B\| < \infty.$$

□

2.2.4 Cas de l'Analyse Canonique de deux vecteurs aléatoires

Soit deux v.a. R et S définis sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

L'analyse canonique (AC) consiste à déterminer au pas l , pour $l = 1, \dots, r$, une c.l. des composantes centrées de R ,

$$U_l = \theta'_l (R - \mathbb{E}[R])$$

et une c.l. des composantes centrées de S ,

$$V_l = \eta'_l (S - \mathbb{E}[S])$$

telles que le coefficient de corrélation linéaire $\rho(U_l, V_l)$ soit maximal, sous les contraintes :

$$\begin{aligned} \text{Var}(U_l) &= 1, & \text{Var}(V_l) &= 1, \\ \text{Cov}(U_l, U_j) &= 0 \text{ pour } j = 1, \dots, l-1, & \text{Cov}(V_l, V_j) &= 0 \text{ pour } j = 1, \dots, l-1. \end{aligned}$$

Les θ_l (resp. η_l) sont appelés facteurs canoniques en R (resp. en S).

Dans un premier temps, montrons que l'analyse canonique peut être vue comme un cas particulier de l'ACP projetée.

Interprétation en tant qu'ACP projetée

On se place dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ muni du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$ et de la norme associée $\|X\| = \left(\mathbb{E}[X^2] \right)^{\frac{1}{2}}$.

Ainsi, pour X et Y v.a. réelles centrées, $\langle X, Y \rangle = \text{Cov}(X, Y)$, $\|X\|^2 = \text{Var}(X)$ et donc,

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \cos(X, Y).$$

A U_l fixé, V_l qui maximise $\rho(U_l, V_l)$ est colinéaire à la projection de U_l sur le sous espace vectoriel engendré par les composantes centrées de S , c'est-à-dire colinéaire au régressé linéaire de U_l par rapport aux composantes centrées de S .

Ainsi, on peut considérer V_l de la forme $\mu_l \hat{\mathbb{E}}[U_l|S]$, $\mu_l \in \mathbb{R}$.

Pour que la contrainte $Var(V_l) = 1$ soit vérifiée, on prend $\mu_l = \frac{1}{\sqrt{Var(\hat{\mathbb{E}}[U_l|S])}}$.

Dans ce contexte, le problème consiste à trouver θ_l , l -ième facteur canonique en R , tel que :

$$\begin{aligned} & \begin{cases} \rho(U_l, \hat{\mathbb{E}}[U_l|S]) \max \\ Cov(U_l, U_j) = 0, j = 1, \dots, l-1 \\ Var(U_l) = 1 \end{cases} \\ \Leftrightarrow & \begin{cases} \rho(U_l, \hat{\mathbb{E}}[U_l|S]) \max \\ Cov(\theta'_l R, \theta'_j R) = 0, j = 1, \dots, l-1 \\ Var(\theta'_l R) = 1 \end{cases} \end{aligned} \quad (2.14)$$

Or,

$$\rho^2(U_l, \hat{\mathbb{E}}[U_l|S]) = \cos^2(U_l, \hat{\mathbb{E}}[U_l|S]) \quad (2.15)$$

$$= \frac{||\hat{\mathbb{E}}[U_l|S]||^2}{||U_l||^2} \quad (2.16)$$

$$\begin{aligned} &= \frac{Var(\hat{\mathbb{E}}[U_l|S])}{Var(U_l)} \\ &= Var(\hat{\mathbb{E}}[U_l|S]) \text{ car } Var(U_l) = 1. \end{aligned} \quad (2.17)$$

Donc : $Var(\hat{\mathbb{E}}[U_l|S]) \max \Leftrightarrow \rho^2(U_l, \hat{\mathbb{E}}[U_l|S]) \max$.

Trouver la solution de (2.5) est donc équivalent à trouver celle de :

$$\begin{cases} Var(\theta'_l \hat{\mathbb{E}}[R|S]) \max \\ Cov(\theta'_l R, \theta'_j R) = 0, j = 1, \dots, l-1 \\ Var(\theta'_l R) = 1 \end{cases}$$

ou encore de :

$$\begin{cases} \theta'_l Cov(\hat{\mathbb{E}}[R|S]) \theta_l \max \\ \theta'_l Cov(R) \theta_j = 0, j = 1, \dots, l-1 \\ \theta'_l Cov(R) \theta_l = 1 \end{cases}$$

Or, d'après la section 2.1.3 sur l'ACP d'un v.a., l'ACP de $\hat{\mathbb{E}}[R|S]$ consiste, au l -ième pas, à chercher θ_l solution du problème :

$$\begin{cases} \theta_l' Cov(\hat{\mathbb{E}}[R|S]) \theta_l \max \\ \theta_l' M^{-1} \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases} \quad (2.18)$$

En prenant $M = (Cov(R))^{-1}$, (2.6) équivaut à (2.5).

Ainsi, θ_l est vecteur propre de $MCov(\hat{\mathbb{E}}[R|S]) = (Cov(R))^{-1} Cov(R, S) (Cov(S))^{-1} Cov(S, R)$.

Montrons alors que V_l vérifie la contrainte $Cov(V_l, V_j) = 0, j = 1, \dots, l-1$:

$$Cov(V_l, V_j) = Cov(\mu_l \hat{\mathbb{E}}[U_l|S], \mu_j \hat{\mathbb{E}}[U_j|S])$$

Comme, d'après les propositions 2 et 4, $\hat{\mathbb{E}}[\theta_j'(R - \mathbb{E}[R])|S] = \theta_j' Cov(R, S) (Cov(S))^{-1} (S - \mathbb{E}[S])$, on a :

$$\begin{aligned} Cov(V_l, V_j) &= \mu_l \mu_j \theta_l' Cov(R, S) (Cov(S))^{-1} \mathbb{E}[(S - \mathbb{E}[S])(S - \mathbb{E}[S])'] (Cov(S))^{-1} Cov(S, R) \theta_j \\ &= \mu_l \mu_j \theta_l' Cov(R, S) (Cov(S))^{-1} Cov(S, R) \theta_j. \end{aligned} \quad (2.19)$$

Or, $(Cov(R))^{-1} Cov(R, S) (Cov(S))^{-1} Cov(S, R) \theta_j = \lambda_j \theta_j$, donc :

$$Cov(V_l, V_j) = \mu_l \mu_j \lambda_j \theta_l' Cov(R) \theta_j.$$

Or, $\theta_l' Cov(R) \theta_j = 0$ pour $j = 1, \dots, l-1$.

Par conséquent, $Cov(V_l, V_j) = 0$.

Ainsi, on a montré que U_l , solution de l'AC, est solution du problème de l'ACP projetée avec la métrique $M = (Cov(R))^{-1}$, et réciproquement, la solution du problème de l'AC étant colinéaire à $\hat{\mathbb{E}}[U_l|S]$.

Dans cette prochaine section, on propose plusieurs méthodes d'estimation de ces facteurs pour l'AC.

Application de la méthode générale

On se ramène à l'a.s. des facteurs de l'ACP projetée.

Le processus d'a.s. de A_1 reste le même :

$$A_{1,n+1} = A_{1n} - a_n(S_{1n}S'_{1n}A_{1n} - S_{1n}R'_n).$$

On définit alors A_n , processus d'a.s. de A , en ôtant pour tout n la dernière ligne de A_{1n} .

On définit ensuite deux processus récurrents (N_n) et (M_n) qui soient tels que les hypothèses (H4) et (H4') ou (H4)(b),(c) et (H4')(b),(c) (selon le processus (X_n) que l'on souhaite mettre en oeuvre) soient vérifiées pour estimer respectivement $N = Cov(R)$ et $M = (Cov(R))^{-1}$.

1) Estimation de N

a) Un choix simple et naturel de N_n est la matrice de covariance empirique de l'échantillon :

$$N_n = \frac{1}{n-1} \sum_{i=1}^{n-1} R_i R'_i - \frac{1}{(n-1)^2} \sum_{i=1}^{n-1} R_i \sum_{i=1}^{n-1} R'_i,$$

que l'on peut facilement mettre à jour récursivement.

Soit l'hypothèse :

(H2'') R admet des moments d'ordre 4.

Sous les hypothèses (H2'') et (H3') par exemple, la loi forte des grands nombres nous assure que N_n tend vers $N = Cov(R)$ p.s. On a également : $\sum_{n=1}^{+\infty} a_n ||N_n - N|| < \infty$ p.s par application du lemme 7.

b) Une autre manière d'estimer la matrice de covariance consiste, sachant que $\mathbb{E}[N - (R - \mathbb{E}[R])R'] = 0$, à utiliser une méthode d'a.s. en définissant le processus (N_n) tel que :

$$N_{n+1} = N_n - a_n(N_n - (R_n - \bar{R}_{n-1})R'_n).$$

Soit l'hypothèse :

(H3'') $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{3}{4} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2})$.

Théorème 9. Sous (H2'') et (H3''), N_n tend vers $N = Cov(R)$ p.s. et $\sum_{n=1}^{+\infty} a_n ||N_n - N|| < \infty$ p.s. ■

Démonstration.

cf. chapitre 3, première partie de la démonstration du théorème 2 avec $\theta_n = \mathbb{E}[R]$ et $\Theta_n = \bar{R}_{n-1}$ (p.89). □

2) Estimation de $M = N^{-1}$

a) On calcule $M_n = (N_n)^{-1}$ pour tout n ou, pour éviter d'avoir à inverser la matrice de covariance empirique N_n , manipulation qui prend beaucoup de temps de calcul en grande dimension, on calcule $(N_n)^{-1}$ par le processus récursif exact qu'on retrouve dans [53] et dont on rappelle l'expression :

$$N_{n+1}^{-1} = \frac{n+1}{n} N_n^{-1} - \frac{n+1}{n(n+1+\rho_{n+1})} N_n^{-1} \phi_{n+1} \phi_{n+1}' N_n^{-1}$$

avec $\phi_{n+1} = R_{n+1} - \bar{R}_n$ et $\rho_{n+1} = \phi_{n+1}' N_n^{-1} \phi_{n+1}$.

Comme N_n tend vers N p.s. et $\sum_{n=1}^{+\infty} a_n \|N_n - N\| < \infty$ p.s., il vient que M_n tend vers $M = (Cov(R))^{-1}$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

b) On peut également choisir d'estimer l'inverse de la matrice de covariance en utilisant à nouveau une méthode d'a.s. :

Proposition 10. $M_1 = \begin{pmatrix} (Cov(R))^{-1} \\ -\mathbb{E}[R'] (Cov(R))^{-1} \end{pmatrix}$ est solution du système linéaire

en $X : \mathbb{E}[R_1 R_1'] X = J = \begin{pmatrix} I_p \\ 0 \end{pmatrix}$ avec $R_1 = \begin{pmatrix} R \\ 1 \end{pmatrix}$. ■

Démonstration.

$$\mathbb{E}[R_1 R_1'] \begin{pmatrix} (Cov(R))^{-1} \\ -\mathbb{E}[R'] (Cov(R))^{-1} \end{pmatrix} = \begin{pmatrix} \mathbb{E}[R R'] & \mathbb{E}[R] \\ \mathbb{E}[R'] & 1 \end{pmatrix} \begin{pmatrix} (Cov(R))^{-1} \\ -\mathbb{E}[R'] (Cov(R))^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbb{E}[RR'] \left(\text{Cov}(R) \right)^{-1} - \mathbb{E}[R] \mathbb{E}[R'] \left(\text{Cov}(R) \right)^{-1} \\ \mathbb{E}[R'] \left(\text{Cov}(R) \right)^{-1} - \mathbb{E}[R'] \left(\text{Cov}(R) \right)^{-1} \end{pmatrix} = \begin{pmatrix} I_p \\ 0 \end{pmatrix}.$$

□

On définit alors le processus (M_{1n}) dans l'ensemble des matrices $(p+1, p)$ tel que :

$$M_{1,n+1} = M_{1n} - a_n(R_{1n}R'_{1n}M_{1n} - J).$$

M_n est obtenu en enlevant, pour tout n , la dernière ligne de M_{1n} .

Soit les deux hypothèses :

(H1') Il n'y a pas de relation affine entre les composantes de R ,

(H3⁽³⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[R_1R'_1])})$.

Théorème 11. *Sous (H1'), (H2'') et (H3), on a :*

$$\|M_n - M\| \rightarrow 0 \text{ p.s. ; } \sum_{n=1}^{+\infty} a_n \|M_n - M\|^2 < \infty \text{ p.s.}$$

En remplaçant (H3) par (H3⁽³⁾), on a en outre : $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty \text{ p.s. } \blacksquare$

Démonstration.

Voir paragraphe 2.4.1 (p.65) : A.S. de l'inverse d'une matrice de covariance lorsque l'espérance ne varie pas dans le temps en introduisant une seule donnée.

□

On pose $B_n = M_n(R_nS'_n - \bar{R}_{n-1}\bar{S}'_{n-1})A_n$ (resp. $B_n = M_n(W_{n-1} - \bar{R}_{n-1}\bar{S}'_{n-1})A_n$) qu'on introduit dans la définition du processus (X_n) du paragraphe 2.2.2 (p.33) (resp. paragraphe 2.2.3 (p.37)).

Théorème 12. *Sous (H1), (H2'), (H3'), (H4) et (H4') (resp. (H1), (H2), (H3'), (H4)(b),(c) et (H4')(b),(c)), X_n^l converge p.s. vers un vecteur propre de $B = \left(\text{Cov}(R) \right)^{-1} \text{Cov}(\hat{\mathbb{E}}[R|S])$ associé à la l-ième plus grande v.p. λ_l . ■*

Démonstration.

Application directe du théorème 6 (resp. du théorème 8).

□

Méthode spécifique

Les facteurs canoniques en R , θ_l , sont vecteurs propres de $B = \left(Cov(R)\right)^{-1} Cov(R, S) \left(Cov(S)\right)^{-1} Cov(S, R)$ associés à la l -ième plus grande v.p. λ_l .

Rappelons que $A = \left(Cov(S)\right)^{-1} Cov(S, R)$ et posons $G = \left(Cov(R)\right)^{-1} Cov(R, S)$.

Les θ_l sont ainsi vecteurs propres de $B = GA$ associés à la l -ième plus grande v.p. λ_l .

D'autre part, on a montré au début de la section (p.39) que $V_l = \eta'_l(S - \mathbb{E}[S])$ est colinéaire à $\hat{\mathbb{E}}[U_l|S]$.

Or,

$$\hat{\mathbb{E}}[U_l|S] = \hat{\mathbb{E}}[\theta'_l(R - \mathbb{E}[R])|S] \quad (2.20)$$

$$= \theta'_l \hat{\mathbb{E}}[R - \mathbb{E}[R]|S] \text{ d'après la proposition 4} \quad (2.21)$$

$$= \theta'_l A'(S - \mathbb{E}[S]) \text{ d'après la proposition 2} \quad (2.22)$$

$$= (A\theta_l)'(S - \mathbb{E}[S])$$

Donc, η_l est colinéaire à $A\theta_l$.

De plus, comme θ_l est vecteur propre de GA associé à la l -ième plus grande v.p. λ_l , on a : $GA\theta_l = \lambda_l\theta_l$.

Donc : $AGA\theta_l = \lambda_l A\theta_l$.

Enfin, puisque η_l est colinéaire à $A\theta_l$, on a : $AG\eta_l = \lambda_l\eta_l$.

Toute v.p. de GA est v.p. de AG , et réciproquement.

Donc, les facteurs canoniques en S , η_l , sont vecteurs propres de AG associés à la l -ième plus grande v.p. λ_l .

De la même façon qu'on a défini le processus (A_{1n}) pour estimer A , on définit le processus G_{1n} tel que :

$$G_{1,n+1} = G_{1n} - a_n(R_{1n}R'_{1n}G_{1n} - R_{1n}S'_n)$$

pour estimer G .

Le processus G_n est alors obtenu en enlevant, pour tout n , la dernière ligne de G_{1n} .

A partir du théorème 5, en permutant le rôle de R et S , on obtient le résultat suivant :

Théorème 13. *Sous $(H1')$, $(H2)$ et $(H3)$, $G_n \rightarrow G$ p.s. En remplaçant $(H3)$ par $(H3^{(3)})$, on a en outre :*

$$\sum_{n=1}^{+\infty} a_n \|G_n - G\| < \infty \text{ p.s. } \blacksquare$$

Soit N_n un estimateur de $\left(\text{Cov}(R)\right)^{-1}$.

On définit alors $B_n = G_n A_n$ puis le processus d'a.s. (X_n) des r premiers facteurs canoniques en R :

$$\begin{cases} B_n = G_n A_n \\ F_n(X_n^l) = \frac{\langle B_n X_n^l, X_n^l \rangle_{N_n}}{\langle X_n^l, X_n^l \rangle_{N_n}} \\ Y_{n+1}^l = X_n^l + a_n (B_n - F_n(X_n^l) I) X_n^l, l = 1, \dots, r \\ X_{n+1} = \text{orth}_{N_n}(Y_{n+1}) \end{cases}$$

Soit la nouvelle hypothèse :

$$(H3^{(4)}) \left(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1 \right) \text{ ou } \left(a_n = \frac{a}{n}, a > \frac{1}{2 \min \left(\lambda_{\min}(\mathbb{E}[S_1 S_1']), \lambda_{\min}(\mathbb{E}[R_1 R_1']) \right)} \right).$$

Théorème 14. *Sous $(H1)$, $(H1')$, $(H2)$, $(H3^{(4)})$ et $(H4')(b),(c)$, X_n^l converge p.s. vers un vecteur propre de $B = \left(\text{Cov}(R)\right)^{-1} \text{Cov}\left(\hat{\mathbb{E}}[R|S]\right)$ associé à la l -ième plus grande v.p. λ_l . $\blacksquare$*

Démonstration.

Sous $(H1)$, $(H1')$, $(H2)$ et $(H3^{(4)})$, $A_n \rightarrow A$ p.s. et $G_n \rightarrow G$ p.s. donc $G_n A_n \rightarrow GA$ p.s.

On a également, sous ces mêmes hypothèses :

$$\sum_{n=1}^{+\infty} a_n \|G_n - G\| < \infty \text{ p.s. et } \sum_{n=1}^{+\infty} a_n \|A_n - A\| < \infty \text{ p.s.}$$

$$\text{D'autre part, } \|B_n - B\| = \|G_n A_n - GA\| \leq \|(G_n - G)A_n\| + \|G(A_n - A)\|.$$

Puisque $A_n \rightarrow A$ p.s., il existe une v.a. positive et finie p.s. D telle que :

$$\sum_{n=1}^{+\infty} a_n \|(G_n - G)A_n\| \leq D \sum_{n=1}^{+\infty} a_n \|G_n - G\| < \infty \text{ p.s.}$$

$$\text{Comme } \sum_{n=1}^{+\infty} a_n \|G(A_n - A)\| \leq \|G\| \sum_{n=1}^{+\infty} a_n \|A_n - A\| < \infty \text{ p.s.,}$$

finalement $\sum_{n=1}^{+\infty} a_n ||B_n - B|| < \infty$ p.s.

On applique dès lors le théorème 4 de [10] avec la deuxième série d'hypothèses pour terminer la preuve. $\square$

Interprétation en tant qu'ACG et estimation simultanée des facteurs en R et en S

Comme θ_l est vecteur propre de GA et η_l est vecteur propre de AG , on obtient les relations de transition suivantes :

$$\eta_l = \frac{1}{\sqrt{\lambda_l}} A \theta_l \text{ et } \theta_l = \frac{1}{\sqrt{\lambda_l}} G \eta_l.$$

On note $\hat{B} = \begin{pmatrix} I_p & G \\ A & I_q \end{pmatrix}$ et $\xi_l = \begin{pmatrix} \theta_l \\ \eta_l \end{pmatrix}$.

Ainsi, on remarque que $\hat{B}\xi_l = \begin{pmatrix} \theta_l + G\eta_l \\ A\theta_l + \eta_l \end{pmatrix} = (1 + \sqrt{\lambda_l}) \begin{pmatrix} \theta_l \\ \eta_l \end{pmatrix} = (1 + \sqrt{\lambda_l})\xi_l$.

ξ_l est donc vecteur propre de $\hat{B}$ associé à la l -ième plus grande v.p. $(1 + \sqrt{\lambda_l})$.

Essayons maintenant d'établir un lien entre analyse canonique et analyse canonique généralisée (ACG).

Dans l'ACG du vecteur $Z = (R, S)$ (c.f. chapitre 1 p.18), on observe que le l -ième facteur général est vecteur propre de la matrice B où :

$$B = MC = \begin{pmatrix} (C^1)^{-1} & 0 \\ 0 & (C^2)^{-1} \end{pmatrix} \cdot \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} = \begin{pmatrix} (C^1)^{-1}C^{11} & (C^1)^{-1}C^{12} \\ (C^2)^{-1}C^{21} & (C^2)^{-1}C^{22} \end{pmatrix}$$

avec $C^1 = C^{11} = \text{Cov}(R)$, $C^2 = C^{22} = \text{Cov}(S)$, $C^{12} = \text{Cov}(R, S)$ et $C^{21} = (C^{12})'$.

$$\text{Ainsi } B = \begin{pmatrix} I_p & \left(\text{Cov}(R)\right)^{-1} \text{Cov}(R, S) \\ \left(\text{Cov}(S)\right)^{-1} \text{Cov}(S, R) & I_q \end{pmatrix} = \begin{pmatrix} I_p & G \\ A & I_q \end{pmatrix} = \hat{B},$$

ξ_l peut ainsi être vu comme le l -ième facteur général de l'ACG en dimension 2.

Soit $B_n = \begin{pmatrix} I_p & G_n \\ A_n & I_q \end{pmatrix}$ et N_n un estimateur de $N = M^{-1} = \begin{pmatrix} C^1 & 0 \\ 0 & C^2 \end{pmatrix}$; on définit alors un processus d'a.s. (X_n^l)

de ξ_l :

$$\begin{cases} B_n = \begin{pmatrix} I_p & G_n \\ A_n & I_q \end{pmatrix} \\ F_n(X_n^l) = \frac{\langle B_n X_n^l, X_n^l \rangle_{N_n}}{\langle X_n^l, X_n^l \rangle_{N_n}} \\ Y_{n+1}^l = X_n^l + a_n(B_n - F_n(X_n^l)I)X_n^l, l = 1, \dots, r. \\ X_{n+1} = \text{orth}_{N_n}(Y_{n+1}) \end{cases}$$

$X_{n+1} = \text{orth}_{N_n}(Y_{n+1})$ signifie que, pour obtenir $(X_{n+1}^1, \dots, X_{n+1}^r)$, on orthogonalise au sens de Gram-Schmidt par rapport à N_n le r-uplet $(Y_{n+1}^1, \dots, Y_{n+1}^r)$.

L'intérêt de cette méthode est que l'on estime de manière simultanée, via l'estimation de ξ_l , les facteurs canoniques θ_l en R et η_l en S .

Théorème 15. *Sous $(H1)$, $(H1')$, $(H2)$, $(H3^{(4)})$ et $(H4')(b),(c)$, X_n^l converge p.s. vers un vecteur propre de B associé à la l -ième plus grande v.p. $(1 + \sqrt{\lambda_l})$. ■*

Démonstration.

D'une part, $N_n \rightarrow N$ p.s. et $\sum_{n=1}^{+\infty} a_n \|N_n - N\| < \infty$ p.s.

D'autre part, $A_n \rightarrow A$ p.s., $G_n \rightarrow G$ p.s. et $\sum_{n=1}^{+\infty} a_n \|A_n - A\| < \infty$ p.s., $\sum_{n=1}^{+\infty} a_n \|G_n - G\| < \infty$ p.s.

On en déduit que $B_n \rightarrow B$ p.s. et $\sum_{n=1}^{+\infty} a_n \|B_n - B\| < \infty$ p.s.

On applique alors le théorème 4 de [10] avec la deuxième série d'hypothèses.

□

2.3 Cas où les composantes du vecteur explicatif sont les indicatrices des modalités exclusives d'une variable aléatoire nominale

Dans cette deuxième partie, on s'intéresse au cas où les composantes de S sont les indicatrices des modalités exclusives d'une v.a. nominale.

Cette fois-ci, il existe une relation affine entre les composantes de S (la somme de ces dernières vaut 1) et la matrice $Cov(S)$ n'est plus inversible.

On revient alors à la régression linéaire de R par rapport à S : on cherche les matrices A_2 et D_2 qui rendent minimale $J(A_2, D_2) = \mathbb{E} \left[\|R - A_2' S - D_2\|^2 \right]$. On appellera régressé de R par rapport à S le v.a. de dimension p noté $\hat{\mathbb{E}}[R|S]$.

Proposition 16. $\hat{\mathbb{E}}[R|S] = \mathbb{E}[R|S] = A_2' S$ avec $A_2 = \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR']$. ■

Démonstration.

Comme les composantes S^k , $k = 1, \dots, q$, de S sont les indicatrices des modalités exclusives d'une variable aléatoire nominale, pour toute composante R^i , $i = 1, \dots, p$, de R , il existe des nombres $a_{1i}, \dots, a_{qi}$ tels que $\mathbb{E}[R^i|S] = \sum_{k=1}^q a_{ki} S^k$.

Donc, il existe une matrice A_2 telle que $\mathbb{E}[R|S] = A_2' S$ et par conséquent $\hat{\mathbb{E}}[R|S] = \mathbb{E}[R|S] = A_2' S$.

Ainsi, $\mathbb{E}[SR'|S] = S \mathbb{E}[R'|S] = SS' A_2$.

On en déduit que : $\mathbb{E}[SR'] = \mathbb{E}[SS'] A_2 \Leftrightarrow A_2 = \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR']$.

□

En outre,

Proposition 17. $Cov \left(\mathbb{E}[R|S] \right) = Cov(R, S) \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR']$. ■

Démonstration.

$$\begin{aligned} Cov \left(\mathbb{E}[R|S] \right) &= \mathbb{E} \left[\left(\mathbb{E}[R|S] - \mathbb{E}[\mathbb{E}[R|S]] \right) \left(\mathbb{E}[R|S] - \mathbb{E}[\mathbb{E}[R|S]] \right)' \right] \\ &= \mathbb{E} \left[\left(\mathbb{E}[(R - \mathbb{E}[R])|S] \right) \left(\mathbb{E}[(R - \mathbb{E}[R])|S] \right)' \right]. \end{aligned} \quad (2.23)$$

Or, $\mathbb{E} \left[(R - \mathbb{E}[R])|S \right] = \mathbb{E} \left[(R - \mathbb{E}[R])S' \right] \left(\mathbb{E}[SS'] \right)^{-1} S$ d'après (2.7).

Donc,

$$Cov(\mathbb{E}[R|S]) = \mathbb{E} \left[\left(\mathbb{E}[(R - \mathbb{E}[R])S'] (\mathbb{E}[SS'])^{-1} S \right) \left(S' (\mathbb{E}[SS'])^{-1} \mathbb{E}[S(R - \mathbb{E}[R])'] \right) \right] \quad (2.24)$$

$$= \mathbb{E} \left[Cov(R, S) (\mathbb{E}[SS'])^{-1} SS' (\mathbb{E}[SS'])^{-1} Cov(S, R) \right] \quad (2.25)$$

$$= Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[SS'] (\mathbb{E}[SS'])^{-1} Cov(S, R) \quad (2.26)$$

$$= Cov(R, S) (\mathbb{E}[SS'])^{-1} Cov(S, R) \quad (2.27)$$

$$= Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[S(R - \mathbb{E}[R])'] \quad (2.28)$$

$$= Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] - Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[S] \mathbb{E}[R'].$$

Or,

$$(\mathbb{E}[SS'])^{-1} \mathbb{E}[S] = \begin{pmatrix} 1 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{pmatrix} \begin{matrix} \uparrow \\ \text{q fois} \\ \downarrow \end{matrix} = u_q.$$

Donc,

$$Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[S] \mathbb{E}[R'] = \mathbb{E}[(R - \mathbb{E}[R])S'] u_q \mathbb{E}[R'].$$

Or, $S' u_q = 1$. Donc,

$$\begin{aligned} Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[S] \mathbb{E}[R'] &= \mathbb{E}[R - \mathbb{E}[R]] \mathbb{E}[R'] \\ &= 0. \end{aligned} \quad (2.29)$$

et par suite,

$$Cov(\mathbb{E}[R|S]) = Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'].$$

□

On a ainsi montré que, dans le cas où les composantes de S sont les indicatrices des modalités exclusives d'une v.a. nominale,

$$\begin{cases} \mathbb{E}[R|S] = \mathbb{E}[RS'] (\mathbb{E}[SS'])^{-1} S = A_2' S \text{ avec } A_2 = (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] \\ Cov(\mathbb{E}[R|S]) = Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] = Cov(R, S) A_2 \end{cases}$$

2.3.1 A.S. de la fonction de régression linéaire

$A_2 = (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR']$ est solution du système en X : $\mathbb{E}[SS'X - SR'] = 0$.

1ère méthode : processus d'a.s. avec pas déterministe

On définit alors le processus d'a.s. (A_{2n}) de A_2 , dans l'ensemble des matrices de dimension (q, p) , tel que :

$$A_{2,n+1} = A_{2n} - a_n(S_n S_n' A_{2n} - S_n R_n').$$

Etant donné que les composantes du vecteur explicatif S sont les indicatrices des modalités exclusives d'une v.a. nominale, on peut simplifier cette mise à jour de la manière suivante.

Supposons que le n -ième individu ait la modalité j de S ($j \in 1, \dots, q$), alors, pour $i = 1, \dots, p$, on a :

$$\begin{aligned} A_{2,n+1}(j, i) &= A_{2n}(j, i) - a_n(A_{2n}(j, i) - R_n^i) \\ A_{2,n+1}(k, i) &= A_{2n}(k, i), k = 1, \dots, q; k \neq j, \end{aligned}$$

c'est-à-dire que seule la ligne j de A_{2n} est modifiée.

Soit l'hypothèse :

$$(H3^{(5)}) \left(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1 \right) \text{ ou } \left(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[SS'])} = \frac{1}{2\min_k \mathbb{P}(\Omega_k)} \right).$$

Théorème 18. *Sous $(H2'')$ et $(H3)$, $A_{2n} \rightarrow A_2$ p.s. En remplaçant $(H3)$ par $(H3^{(5)})$, on a en outre : $\sum_{n=1}^{+\infty} a_n \|A_{2n} - A_2\| < \infty$ p.s. ■*

Démonstration.

Application du théorème 5 en remplaçant S_1 par S (l'hypothèse $(H1)$ est vérifiée car il n'existe pas de relation linéaire entre les composantes de S).

□

2ème méthode : processus d'a.s. avec pas aléatoire

Soit $P_n = \left(\sum_{l=1}^n S_l S_l' \right)^{-1}$.

Cette matrice aléatoire va jouer le rôle du pas dans la définition du processus d'a.s. (A_{2n}) de A_2 , dans l'ensemble des matrices de dimension (q, p) :

$$A_{2,n+1} = A_{2n} - P_n (S_n S_n' A_{2n} - S_n R_n').$$

De façon générale, pour $k = 1, \dots, q$ et $i = 1, \dots, p$:

$$\begin{aligned} A_{2,n+1}(k, i) &= A_{2n}(k, i) - \frac{1}{\sum_{l=1}^n S_l^k} (S_n^k A_{2n}(k, i) - S_n^k R_n^i) \\ &= \frac{\sum_{l=1}^{n-1} S_l^k}{\sum_{l=1}^n S_l^k} A_{2n}(k, i) + \frac{S_n^k R_n^i}{\sum_{l=1}^n S_l^k}. \end{aligned} \tag{2.30}$$

Par récurrence, on peut montrer que : $A_{2n}(k, i) = \frac{\sum_{l=1}^{n-1} S_l^k R_l^i}{\sum_{l=1}^{n-1} S_l^k}$, $A_{21}(k, i) = 0$.

Théorème 19. Sous $(H2'')$ et $(H3^{(5)})$, $A_{2n} \rightarrow A_2$ p.s. et $\sum_{n=1}^{+\infty} a_n \|A_{2n} - A_2\| < \infty$ p.s. ■

Démonstration.

D'après la loi forte des grands nombres :

$$A_{2n}(k, i) = \frac{\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i}{\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k} \rightarrow \frac{\mathbb{E}[S^k R^i]}{\mathbb{E}[S^k]} (= \mathbb{E}[R^i | S^k = 1]) = A_2(k, i) \text{ p.s.}$$

Donc, $A_{2n} \rightarrow A_2 = (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR']$ p.s.

D'autre part,

$$A_{2n}(k, i) - A_2(k, i) = \frac{\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i}{\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k} - \frac{\mathbb{E}[S^k R^i]}{\mathbb{E}[S^k]} \quad (2.31)$$

$$\begin{aligned} &= \frac{\left(\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i \right) \mathbb{E}[S^k] - \mathbb{E}[S^k R^i] \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k}{\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k \mathbb{E}[S^k]} \\ &= \frac{\left(\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i - \mathbb{E}[S^k R^i] \right) \mathbb{E}[S^k] - \mathbb{E}[S^k R^i] \left(\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k - \mathbb{E}[S^k] \right)}{\left(\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k \right) \mathbb{E}[S^k]}. \end{aligned} \quad (2.32)$$

Or, d'après le lemme 7, sous (H2''), on a :

$$\mathbb{E} \left[\left| \mathbb{E}[S^k R^i] - \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i \right| \right] = O\left(\frac{1}{\sqrt{n}}\right) \text{ et } \mathbb{E} \left[\left| \mathbb{E}[S^k] - \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k \right| \right] = O\left(\frac{1}{\sqrt{n}}\right).$$

Ainsi, $\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\left| \mathbb{E}[S^k R^i] - \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i \right| \right] < \infty$ et $\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\left| \mathbb{E}[S^k] - \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k \right| \right] < \infty$ sous (H3⁽⁵⁾). (a_n du processus (X_n) d'a.s. des vecteurs propres)

Donc, $\sum_{n=1}^{+\infty} a_n \left| \mathbb{E}[S^k R^i] - \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k R_l^i \right| < \infty$ p.s. et $\sum_{n=1}^{+\infty} a_n \left| \mathbb{E}[S^k] - \frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k \right| < \infty$ p.s.

Enfin, $\frac{1}{n-1} \sum_{l=1}^{n-1} S_l^k \rightarrow \mathbb{E}[S^k]$ p.s. et $\mathbb{E}[S^k R^i] < \infty$ sous (H2'').

Donc, $\sum_{n=1}^{+\infty} a_n |A_{2n}(j, i) - A_2(j, i)| < \infty$ p.s.

Enfin,

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n \|A_{2n} - A_2\| &= \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^q \sum_{i=1}^p (A_{2n}(j, i) - A_2(j, i))^2 \right)^{\frac{1}{2}} \text{ en prenant la norme de Hilbert - Schmidt} \\ &\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^q \sum_{i=1}^p |A_{2n}(j, i) - A_2(j, i)| \right) \\ &\leq \sum_{j=1}^q \sum_{i=1}^p \left(\sum_{n=1}^{+\infty} a_n |A_{2n}(j, i) - A_2(j, i)| \right) \\ &< \infty. \end{aligned}$$

□

2.3.2 A.S. des facteurs de l'ACP projetée

Soit l'hypothèse :

(H2⁽³⁾) R admet des moments d'ordre $2r$.

Soit M_n et N_n deux estimateurs respectifs de M et N .

Ceci étant, on pose $B_n = M_n(R_n S'_n - \bar{R}_{n-1} \bar{S}'_{n-1}) A_n$ (resp. $B_n = M_n(W_{n-1} - \bar{R}_{n-1} \bar{S}'_{n-1}) A_n$) qu'on introduit dans la définition du processus (X_n) du paragraphe 2.2.2 (p.33) (resp. paragraphe 2.2.3 (p.37)). Les deux processus que l'on vient de définir peuvent être utilisés pour (A_n) .

On a le résultat suivant :

Théorème 20. *Sous (H2⁽³⁾), (H3⁽⁵⁾), (H4) et (H4') (resp. (H2''), (H3⁽⁵⁾), (H4)(b),(c) et (H4')(b),(c)), X_n^l converge p.s. vers un vecteur propre de $B = MCov(R, S)A_2$ associé à la l -ième plus grande v.p. λ_l . ■*

Démonstration.

La démarche est identique à celle du théorème 6 (resp. théorème 8) en remplaçant A_n par A_{2n} . □

2.3.3 Cas particulier de l'Analyse Factorielle Discriminante

L'analyse factorielle discriminante (AFD) consiste à déterminer, pour $l = 1, \dots, r$, un facteur $U_l = \theta'_l(R - \mathbb{E}[R])$ tel que :

$$\begin{cases} r_c^2(U_l/S) \max \\ Cov(U_l, U_j) = 0, j = 1, \dots, l-1 \\ Var(U_l) = 1 \end{cases} \quad (2.33)$$

où $r_c(U_l/S)$ est le rapport de corrélation de U_l par rapport à S .

En d'autres termes, au l -ième pas, on cherche une c.l. des composantes centrées de R , la plus liée possible au caractère qualitatif S qui détermine les classes, et non corrélée aux c.l. déterminées antérieurement.

Tout d'abord, on montre que l'analyse factorielle discriminante peut être vue comme un cas particulier de l'ACP projetée.

Interprétation en tant qu'ACP projetée

Rappelons que, en s'appuyant sur le lemme 1, on peut écrire le problème de l'ACP de $\mathbb{E}[R|S]$ sous la forme :

Au l -ième pas, on cherche θ_l tel que :

$$\begin{cases} \theta_l' Cov(\mathbb{E}[R|S]) \theta_l \max \\ \theta_l' M^{-1} \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' M^{-1} \theta_l = 1 \end{cases}$$

Par définition du rapport de corrélation,

$$\begin{aligned} (2.8) \Leftrightarrow \begin{cases} \frac{Var(\mathbb{E}[U_l|S])}{Var(U_l)} \max \\ \theta_l' Cov(R) \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' Cov(R) \theta_l = 1 \end{cases} &\Leftrightarrow \begin{cases} Var(\mathbb{E}[\theta_l'(R - \mathbb{E}[R])|S]) \max \\ \theta_l' Cov(R) \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' Cov(R) \theta_l = 1 \end{cases} \\ &\Leftrightarrow \begin{cases} \theta_l' Cov(\mathbb{E}[R|S]) \theta_l \max \\ \theta_l' Cov(R) \theta_j = 0, j = 1, \dots, l-1 \\ \theta_l' Cov(R) \theta_l = 1 \end{cases} \end{aligned}$$

Ce dernier critère n'est autre que celui de l'ACP de $\mathbb{E}[R|S]$ avec la métrique $M = (Cov(R))^{-1}$ et l'AFD peut bien être vue comme un cas particulier d'ACP projetée.

Le facteur θ_l de l'AFD est donc vecteur propre de $B = (Cov(R))^{-1} Cov(\hat{\mathbb{E}}[R|S]) = (Cov(R))^{-1} Cov(R, S) (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR']$
 $(Cov(R))^{-1} Cov(R, S) A_2$.

Application de la méthode générale

Soit (M_n) un estimateur de $M = \left(\text{Cov}(R) \right)^{-1}$ et N_n un estimateur de $N = M^{-1} = \text{Cov}(R)$.

On définit $B_n = M_n(R_n S'_n - \bar{R}_{n-1} \bar{S}'_{n-1}) A_{2n}$ (resp. $B_n = M_n(W_{n-1} - \bar{R}_{n-1} \bar{S}'_{n-1}) A_{2n}$) et on applique le théorème 20 pour établir la convergence p.s. de X_n^l vers un vecteur propre de $B = M \text{Cov}(R, S) A_2$ associé à la l -ième plus grande v.p. λ_l .

Méthode spécifique

On peut toutefois utiliser une méthode d'estimation plus spécifique à ce problème particulier en remarquant que :

$$B = G A_2 \text{ avec } G = \left(\text{Cov}(R) \right)^{-1} \text{Cov}(R, S) \text{ et } A_2 = \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR'].$$

On définit alors $B_n = G_n A_{2n}$ où :

– G_n est le processus d'a.s. de G défini au paragraphe 2.2.4

– A_{2n} est le processus d'a.s. de A_2 défini au paragraphe 2.3.1.

On considère également l'hypothèse suivante :

$$(H3^{(6)}) \left(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1 \right) \text{ ou } \left(a_n = \frac{a}{n}, a > \frac{1}{2 \min \left(\lambda_{\min}(\mathbb{E}[SS']), \lambda_{\min}(\mathbb{E}[R_1 R_1']) \right)} \right).$$

Soit N_n un estimateur de $N = M^{-1} = \text{Cov}(R)$.

Théorème 21. Sous $(H1')$, $(H2'')$, $(H3^{(6)})$ et $(H4')(b),(c)$, X_n^l converge p.s. vers un vecteur propre de la matrice $B = \left(\text{Cov}(R) \right)^{-1} \text{Cov}(R, S) \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR'] = G A_2$ associé à la l -ième plus grande v.p. λ_l . ■

Démonstration.

Identique à celle du théorème 14 en remplaçant A_n par A_{2n} .

□

Interprétation en tant qu'Analyse Canonique

Soit $\tilde{\theta}_l$ (resp. $\tilde{\eta}_l$) l -ième facteur canonique en R (resp. en S) de l'AC formelle des v.a. R et S .

Nous allons montrer que θ_l est colinéaire à $\tilde{\theta}_l$.

Au 1-ième pas, l'AC consiste à chercher
$$\begin{cases} U_l = \tilde{\theta}_l'(R - \mathbb{E}[R]) \\ V_l = \tilde{\eta}_l'(S - \mathbb{E}[S]) \end{cases}$$

tels que le coefficient de corrélation linéaire $\rho(U_l, V_l)$ soit maximal, sous les contraintes :

$$\begin{aligned} \text{Var}(U_l) &= 1, & \text{Var}(V_l) &= 1, \\ \text{Cov}(U_l, U_j) &= 0 \text{ pour } j = 1, \dots, l-1, & \text{Cov}(V_l, V_j) &= 0 \text{ pour } j = 1, \dots, l-1. \end{aligned}$$

V_l qui maximise $\rho(U_l, V_l)$ est colinéaire à $\mathbb{E}[U_l|S]$.

D'après un résultat de la section 2.2.4 (p.41), on établit que $\tilde{\theta}_l$ est vecteur propre de $MCov(\mathbb{E}[R|S]) = \left(\text{Cov}(R) \right)^{-1} \text{Cov}(R, S) \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR']$ (d'après la proposition 17) associé à la l -ième plus grande v.p. λ_l et que $\text{Cov}(V_l, V_j) = 0$ pour $j = 1, \dots, l-1$.

Comme θ_l l'est aussi, on en déduit que ces deux vecteurs sont colinéaires.

Estimation simultanée des facteurs en R et en S

Pour estimer les facteurs de l'AFD, on peut donc utiliser la méthode d'estimation simultanée précédemment développée pour l'AC.

Dans notre cas particulier, les relations de transition deviennent :

$$\eta_l = \frac{1}{\sqrt{\lambda_l}} A_2 \theta_l \text{ et } \theta_l = \frac{1}{\sqrt{\lambda_l}} G \eta_l$$

avec

$$A_2 = \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR'] \text{ et } G = \left(\text{Cov}(R) \right)^{-1} \text{Cov}(R, S).$$

Ainsi, on remarque que $\xi_l = \begin{pmatrix} \theta_l \\ \eta_l \end{pmatrix}$, vecteur propre de $B = \begin{pmatrix} I_p & G \\ A_2 & I_q \end{pmatrix}$ est le l -ième facteur général de l'ACG de $Z = (R, S)$.

On utilise alors à nouveau les deux processus A_{2n} et G_n pour définir $B_n = \begin{pmatrix} I_p & G_n \\ A_{2n} & I_q \end{pmatrix}$

Soit N_n un estimateur de $N = M^{-1} = \begin{pmatrix} \text{Cov}(R) & 0 \\ 0 & \mathbb{E}[SS'] \end{pmatrix}$.

Théorème 22. *Sous les hypothèses $(H1')$, $(H2'')$, $(H3^{(6)})$ et $(H4')(b),(c)$, X_n^l converge p.s. vers un vecteur propre de la matrice B associé à la l -ième plus grande v.p. $(1 + \sqrt{\lambda_l})$. ■*

Démonstration.

La démarche est identique à celle du théorème 15 en remplaçant A_n par A_{2n} . □

2.3.4 Cas particulier de l'Analyse Factorielle des Correspondances

Les composantes des vecteurs R et S sont respectivement les indicatrices des modalités exclusives de deux variables aléatoires nominales.

L'analyse factorielle des correspondances (AFC) consiste à rechercher pour $l = 1, \dots, r = \min(p-1, q-1)$ une c.l. $V_l = \eta'_l(S - \mathbb{E}[S])$ et une c.l. $U_l = \theta'_l(R - \mathbb{E}[R])$ telles que :

$$\begin{cases} r_c^2(V_l|R) \max \\ \text{Cov}(V_l, V_j) = 0, j = 1, \dots, l-1 \\ \text{Var}(V_l) = 1 \end{cases} \quad \begin{cases} r_c^2(U_l|S) \max \\ \text{Cov}(U_l, U_j) = 0, j = 1, \dots, l-1 \\ \text{Var}(U_l) = 1 \end{cases}$$

Interprétation en tant qu'ACP projetée

Nous allons désormais présenter l'AFC sous un angle nouveau. En effet, cette analyse peut être interprétée comme une ACP projetée munie d'une métrique M particulière :

$$V_l = \eta'_l(S - \mathbb{E}[S]) = \eta'_l(S - \mathbb{E}[S]u'_q S) = \eta'_l(I_q - \mathbb{E}[S]u'_q)S = \mu'_l S.$$

Donc : $V_l = \mu'_l S$, μ_l vérifiant $\mu'_l \mathbb{E}[S] = 0$, car V_l est centré.

$$\text{Var}(V_l) = \mathbb{E}[(\mu'_l S)^2] = \mu'_l \mathbb{E}[SS'] \mu_l.$$

La première partie de l'AFC consiste à rechercher pour $l = 1, \dots, r = \min(p-1, q-1)$ une c.l. $V_l = \mu'_l S$ telle que :

$$\begin{cases} r_c^2(V_l|R) \max \\ \text{Cov}(V_l, V_j) = 0, j = 1, \dots, l-1 \\ \text{Var}(V_l) = 1 \\ \mathbb{E}[V_l] = 0 \end{cases} \Leftrightarrow \begin{cases} \text{Var}(\mathbb{E}[V_l|R]) \max \\ \mathbb{E}[V_l V_j] = 0, j = 1, \dots, l-1 \\ \mu'_l \mathbb{E}[SS'] \mu_l = 1 \\ \mu'_l \mathbb{E}[S] = 0 \end{cases} \Leftrightarrow \begin{cases} \mu'_l \text{Cov}(E[S|R]) \mu_l \max \\ \mu'_l \mathbb{E}[SS'] \mu_j = 0, j = 1, \dots, l-1 \\ \mu'_l \mathbb{E}[SS'] \mu_l = 1 \\ \mu'_l \mathbb{E}[S] = 0 \end{cases}$$

Donc, la première partie de l'AFC est l'ACP projetée de S par rapport à R en utilisant dans $\mathbb{R}^q$ la métrique $(\mathbb{E}[SS'])^{-1}$, avec la contrainte $\mu'_l \mathbb{E}[S] = 0$.

Sans tenir compte de la dernière contrainte, μ_l est vecteur propre de la matrice $(\mathbb{E}[SS'])^{-1} \text{Cov}(\mathbb{E}[S/R]) = (\mathbb{E}[SS'])^{-1} \text{Cov}(S, R) (\mathbb{E}[RR'])^{-1} \mathbb{E}[RS']$ (proposition 17) associé à la l -ième plus grande v.p. λ_l .

Proposition 23. u_q est vecteur propre de $(\mathbb{E}[SS'])^{-1} \text{Cov}(S, R) (\mathbb{E}[RR'])^{-1} \mathbb{E}[RS']$ associé à la v.p. 0. ■

Démonstration.

On montre que u_q est vecteur propre des matrices $(\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] (\mathbb{E}[RR'])^{-1} \mathbb{E}[RS']$ et $(\mathbb{E}[SS'])^{-1} \mathbb{E}[S] \mathbb{E}[R'] (\mathbb{E}[RR'])^{-1} \mathbb{E}[R]$ associé à la v.p. 1.

D'une part, on a :

$$(\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] (\mathbb{E}[RR'])^{-1} \mathbb{E}[RS'] u_q = (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] (\mathbb{E}[RR'])^{-1} \mathbb{E}[R] \text{ puisque } S' u_q = 1 \quad (2.34)$$

$$= (\mathbb{E}[SS'])^{-1} \mathbb{E}[SR'] u_p \text{ car } (\mathbb{E}[RR'])^{-1} \mathbb{E}[R] = u_p \quad (2.35)$$

$$= (\mathbb{E}[SS'])^{-1} \mathbb{E}[S] \text{ car } R' u_p = 1 \quad (2.36)$$

$$= u_q.$$

D'autre part, on a :

$$\left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[S] \mathbb{E}[R'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS'] u_q = \left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[S] \mathbb{E}[R'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[R] \text{ puisque } S' u_q = \mathbb{E}[R] \text{ (2.37)}$$

$$= \left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[S] \mathbb{E}[R'] u_p \quad (2.38)$$

$$= \left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[S] \text{ car } R' u_p = 1 \quad (2.39)$$

$$= u_q.$$

Ainsi, $\left(\mathbb{E}[SS']\right)^{-1} \text{Cov}(S, R) \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS'] u_q = \left(\left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[SR'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS'] - \left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[S] \mathbb{E}[R'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[R]\right) u_q - u_q = 0$.

□

μ_l et u_q sont tous deux vecteurs propres de la matrice $\mathbb{E}[SS']$ -symétrique $\left(\mathbb{E}[SS']\right)^{-1} \text{Cov}(S, R) \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS']$.
Donc, pour tout l , $\mu_l' \mathbb{E}[SS'] u_q = 0 \Leftrightarrow \mu_l' \mathbb{E}[S] = 0$ puisque $S' u_q = 1$.

μ_l vérifie donc la dernière contrainte $\mu_l' \mathbb{E}[S] = 0$ et est donc solution du problème de l'ACP projetée.

Proposition 24. μ_l est vecteur propre de la matrice $\left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[SR'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS']$ associé à la même v.p. λ_l . ■

Démonstration.

On a :

$$\left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[S] \mathbb{E}[R'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS'] \mu_l = u_q \mu_l' \mathbb{E}[RS'] \mu_l \quad (2.40)$$

$$= u_q \mathbb{E}[S'] \mu_l \quad (2.41)$$

$$= 0 \text{ car } \mu_l' \mathbb{E}[S] = 0$$

Donc :

$$\left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[SR'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS'] \mu_l = \left(\mathbb{E}[SS']\right)^{-1} \text{Cov}(S, R) \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS'] \mu_l \quad (2.42)$$

$$= \lambda_l \mu_l.$$

□

Cette matrice a aussi pour vecteur propre u_q associé à la v.p. 1 (cf. démonstration de la proposition 23).

Donc, μ_l est vecteur propre de la matrice $\left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[SR'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS']$ associé à la l -ième plus grande v.p., en ne tenant pas compte de la v.p. 1.

On définit le processus (G_{2n}) d'a.s. de $G_2 = \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS']$ tel que :

$$G_{2,n+1} = G_{2n} - a_n(R_n R_n' G_{2n} - R_n S_n') \text{ (pas déterministe)}$$

ou

$$G_{2,n+1} = G_{2n} - Q_n(R_n R_n' G_{2n} - R_n S_n') \text{ où } Q_n = \left(\sum_{l=1}^n R_l R_l'\right)^{-1} \text{ (pas aléatoire).}$$

On peut ainsi estimer les facteurs grâce à une méthode plus spécifique en approchant $B = A_2 G_2$ par $B_n = A_{2n} G_{2n}$. Les deux processus (pas déterministe ou pas aléatoire) peuvent être utilisés pour A_{2n} et G_{2n} .

Théorème 25. *Sous $(H3^{(8)})$ et $(H4')$ (b),(c), X_n^l converge p.s. vers un vecteur propre de la matrice $B = \left(\mathbb{E}[SS']\right)^{-1} \mathbb{E}[SR'] \left(\mathbb{E}[RR']\right)^{-1} \mathbb{E}[RS']$ associé à la l -ième plus grande v.p. λ_l . ■*

Démonstration.

La démarche est identique à celle du théorème 14 en remplaçant A_n par A_{2n} et G_n par G_{2n} . □

De même, θ_l est vecteur propre de $G_2 A_2$ associé à la l -ième plus grande v.p. λ_l .

Interprétation en tant qu'Analyse Canonique

Dans ce paragraphe, on montre que θ_l (resp. η_l) est colinéaire au l -ième facteur canonique en R , $\tilde{\theta}_l$, (resp. en S , $\tilde{\eta}_l$) de l'analyse canonique des vecteurs R et S .

En reprenant la présentation de l'AC du paragraphe 2.2.4 (p.39), dans notre contexte, cette dernière consiste à chercher, pour $l = 1, \dots, r$:

$$\begin{cases} U_l = \tilde{\theta}_l' R \\ V_l = \tilde{\eta}_l' S \end{cases}$$

tels que le coefficient de corrélation linéaire $\rho(U_l, V_l)$ soit maximal, sous les contraintes :

$$\begin{aligned} \text{Var}(U_l) &= 1, & \text{Var}(V_l) &= 1, \\ \text{Cov}(U_l, U_j) &= 0 \text{ pour } j = 1, \dots, l-1, & \text{Cov}(V_l, V_j) &= 0 \text{ pour } j = 1, \dots, l-1, \\ \tilde{\theta}'_l \mathbb{E}[R] &= 0 & \tilde{\eta}'_l \mathbb{E}[S] &= 0 \end{aligned}$$

V_l qui maximise $\rho(U_l, V_l)$ est colinéaire à $\mathbb{E}[U_l|S] = \mathbb{E}[\tilde{\theta}'_l R|S] = \tilde{\theta}'_l \mathbb{E}[R|S] = \tilde{\theta}'_l \mathbb{E}[RS'] \left(\mathbb{E}[SS'] \right)^{-1} S$.

Ainsi, $\rho^2(U_l, V_l) = r_c^2(U_l|S)$ et on établit que $\tilde{\theta}_l$ est vecteur propre de la matrice $\left(\mathbb{E}[RR'] \right)^{-1} \text{Cov}(\mathbb{E}[R|S]) = \left(\mathbb{E}[RR'] \right)^{-1} \text{Cov}(R, S) \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[RS']$ associé à la 1-ième plus grande v.p. λ_l et vérifie $\tilde{\theta}'_l \mathbb{E}[R] = 0$.

Il vient alors que $\tilde{\eta}_l$, colinéaire à $\left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR'] \tilde{\theta}_l = \left(\mathbb{E}[SS'] \right)^{-1} \text{Cov}(S, R) \tilde{\theta}_l$, est vecteur propre de la matrice $\mathbb{E}[SS']$ -symétrique $\left(\mathbb{E}[SS'] \right)^{-1} \text{Cov}(S, R) \left(\mathbb{E}[RR'] \right)^{-1} \mathbb{E}[RS']$ associé à la 1-ième plus grande v.p. λ_l et vérifie $\tilde{\eta}'_l \mathbb{E}[S] = 0$ et $\text{Cov}(V_l, V_j) = \tilde{\eta}'_l \mathbb{E}[SS'] \tilde{\eta}_j = 0$.

Donc, effectuer l'analyse canonique de R et S revient à affectuer l'AFC de R et S .

Estimation simultanée des facteurs en R et en S

Pour estimer les facteurs de l'AFC, on peut ainsi utiliser la méthode d'estimation simultanée développée précédemment pour l'AC.

Dans notre cadre, on a les relations de transition :

$$\mu_l = \frac{1}{\sqrt{\lambda_l}} A_2 \theta_l \text{ et } \theta_l = \frac{1}{\sqrt{\lambda_l}} G_2 \eta_l \text{ où } A_2 = \left(\mathbb{E}[SS'] \right)^{-1} \mathbb{E}[SR'] \text{ et } G_2 = \left(\mathbb{E}[RR'] \right)^{-1} \mathbb{E}[RS'].$$

On peut d'ailleurs montrer que $\xi_l = \begin{pmatrix} \theta_l \\ \eta_l \end{pmatrix}$, vecteur propre de la matrice $B = \begin{pmatrix} I_p & G_2 \\ A_2 & I_q \end{pmatrix}$ en ne tenant pas compte de la v.p. 1, est colinéaire au l -ième facteur général de l'ACG de $Z = (R, S)$.

On pose alors $B_n = \begin{pmatrix} I_p & G_{2n} \\ A_{2n} & I_q \end{pmatrix}$ et on définit N_n , estimateur de $N = M^{-1} = \begin{pmatrix} \mathbb{E}[RR'] & 0 \\ 0 & \mathbb{E}[SS'] \end{pmatrix}$.

Une fois encore, les deux processus (pas déterministe ou pas aléatoire) peuvent être utilisés pour A_{2n} et G_{2n} .

Théorème 26. *Sous $(H3^{(8)})$ et $(H4')(b), (c)$, X_n^l converge p.s. vers un vecteur propre de la matrice B associé à la l -ième plus grande v.p. $(1 + \sqrt{\lambda_l})$. ■*

Démonstration.

La démarche est identique à celle du théorème 15 en remplaçant A_n par A_{2n} et G_n par G_{2n} .

□

2.4 Annexes du chapitre 2

2.4.1 Démonstration du théorème 10

A.S. de l'inverse d'une matrice de covariance lorsque l'espérance ne varie pas dans le temps en introduisant une seule donnée.

Soit le processus (M_{1n}) dans l'ensemble des matrices $(p+1, p)$ tel que :

$$M_{1,n+1} = M_{1n} - a_n(R_{1n}R'_{1n}M_{1n} - J).$$

M_n est obtenu en enlevant, pour tout n , la dernière ligne de M_{1n} .

Sous les hypothèses :

(H1') Il n'existe pas de relation affine entre les composantes de R ,

(H2'') R admet des moments d'ordre 4,

(H3) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$,

on a : $\|M_n - M\| \rightarrow 0$ p.s.

Si, en outre, l'hypothèse suivante est vérifiée :

(H3⁽³⁾) $(a_n = \frac{a}{n^\alpha}$ avec $a > 0$, $\frac{2}{3} < \alpha < 1$) ou $(a_n = \frac{a}{n}$ et $a > \frac{1}{2\lambda_{\min}(\mathbb{E}[R_1 R'_1])})$,

on a : $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

Démonstration.

Comme $\mathbb{E}[R_1 R'_1] M_1 = J$, on a :

$$\begin{aligned} M_{1,n+1} - M_1 &= M_{1n} - M_1 - a_n \left(\mathbb{E}[R_1 R'_1] (M_{1n} - M_1) + (R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]) M_{1n} \right) \\ &= \left(I - a_n \mathbb{E}[R_1 R'_1] \right) (M_{1n} - M_1) - a_n \left(R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1] \right) M_{1n}. \end{aligned}$$

D'où :

$$\begin{aligned}
||M_{1,n+1} - M_1||^2 &= ||(I - a_n \mathbb{E}[R_1 R'_1])(M_{1n} - M_1)||^2 - \\
&\quad 2a_n \left\langle (I - a_n \mathbb{E}[R_1 R'_1])(M_{1n} - M_1), (R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]) M_{1n} \right\rangle + \\
&\quad a_n^2 ||(R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]) M_{1n}||^2.
\end{aligned}$$

En prenant l'espérance conditionnellement à T_n , il vient que :

$$\begin{aligned}
\mathbb{E} \left[||M_{1,n+1} - M_1||^2 | T_n \right] &= ||(I - a_n \mathbb{E}[R_1 R'_1])(M_{1n} - M_1)||^2 + \\
&\quad a_n^2 \mathbb{E} \left[||(R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]) M_{1n}||^2 | T_n \right] \\
&= ||M_{1n} - M_1||^2 + a_n^2 ||\mathbb{E}[R_1 R'_1](M_{1n} - M_1)||^2 - \\
&\quad 2a_n \left\langle M_{1n} - M_1, \mathbb{E}[R_1 R'_1](M_{1n} - M_1) \right\rangle + \\
&\quad a_n^2 \mathbb{E} \left[||(R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]) M_{1n}||^2 | T_n \right].
\end{aligned}$$

Pour le dernier terme,

$$\begin{aligned}
\mathbb{E} \left[||R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]||^2 / T_n \right] &= \mathbb{E} \left[||R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]||^2 \right] \\
&= \mathbb{E} \left[||R_1 R'_1 - \mathbb{E}[R_1 R'_1]||^2 \right] \\
&= \mathbb{E} [||R_1 R'_1||^2] - ||\mathbb{E}[R_1 R'_1]||^2 \\
&\leq \mathbb{E} [||R_1 R'_1||^2].
\end{aligned}$$

$$\text{et } ||M_{1n}||^2 \leq 2||M_{1n} - M_1||^2 + 2||M_1||^2.$$

Donc,

$$a_n^2 \mathbb{E} \left[||(R_{1n} R'_{1n} - \mathbb{E}[R_1 R'_1]) M_{1n}||^2 | T_n \right] \leq 2a_n^2 \mathbb{E} [||R_1 R'_1||^2] (||M_{1n} - M_1||^2 + ||M_1||^2).$$

Ainsi,

$$\begin{aligned}
\mathbb{E} \left[||M_{1,n+1} - M_1||^2 | T_n \right] &\leq \left(1 + 3a_n^2 \mathbb{E} [||R_1 R'_1||^2] \right) ||M_{1n} - M_1||^2 + 2a_n^2 \mathbb{E} [||R_1 R'_1||^2] \cdot ||M_1||^2 - \\
&\quad 2a_n \left\langle M_{1n} - M_1, \mathbb{E}[R_1 R'_1](M_{1n} - M_1) \right\rangle \\
&\leq (1 + 3a_n^2 \mathbb{E} [||R_1 R'_1||^2]) ||M_{1n} - M_1||^2 + 2a_n^2 \mathbb{E} [||R_1 R'_1||^2] ||M_1||^2 - \\
&\quad 2a_n \lambda_{\min}(\mathbb{E}[R_1 R'_1]) ||M_{1n} - M_1||^2 \quad (\star).
\end{aligned}$$

Sous (H1'), (H2'') et (H3), on peut appliquer le lemme de Robbins-Siegmund :

$$||M_{1n} - M_1||^2 \text{ converge p.s. et } \sum_{n=1}^{+\infty} a_n ||M_{1n} - M_1||^2 < \infty \text{ p.s.}$$

$$\text{Comme } \sum_{n=1}^{+\infty} a_n = \infty, ||M_{1n} - M_1|| \rightarrow 0 \text{ p.s.}$$

En prenant l'espérance dans $(\star)$, on obtient que $\mathbb{E} [||M_{1n} - M_1||^2] \rightarrow 0$.

Ainsi, il existe $f \geq 0$ tel que :

$$\begin{aligned} \mathbb{E}[||M_{1,n+1} - M_1||^2] &\leq 1 - 2a_n \lambda_{\min}(\mathbb{E}[R_1 R_1']) \mathbb{E}[||M_{1n} - M_1||^2] + \\ &\quad a_n^2 \mathbb{E}[||R_1 R_1'||^2] (3\mathbb{E}[||M_{1n} - M_1||^2] + 2||M_1||^2) \\ &\leq \left(1 - 2a_n \lambda_{\min}(\mathbb{E}[R_1 R_1'])\right) \mathbb{E}[||M_{1n} - M_1||^2] + f a_n^2 \text{ sous (H2'')}. \end{aligned}$$

Sous (H3⁽³⁾), le lemme 3 p.83 nous assure que : $\overline{\lim} \frac{1}{a_n} \mathbb{E}[||M_{1n} - M_1||^2] < \infty$.

Alors, à partir d'un certain rang, $\mathbb{E}[||M_{1n} - M_1||^2] \leq a_n$.

Or, l'inégalité de Cauchy-Schwartz nous donne :

$$\sum_{n=1}^{+\infty} a_n \mathbb{E}[||M_{1n} - M_1||] \leq \sum_{n=1}^{+\infty} a_n \left(\mathbb{E}[||M_{1n} - M_1||^2] \right)^{\frac{1}{2}}.$$

Ainsi, $\sum_{n=1}^{+\infty} a_n \mathbb{E}[||M_{1n} - M_1||] \leq \sum_{n=1}^{+\infty} a_n^{\frac{3}{2}} < \infty$ sous (H3⁽³⁾).

Par suite, $\sum_{n=1}^{+\infty} a_n ||M_{1n} - M_1|| < \infty$ p.s.

En enlevant la dernière ligne pour tout n , on a :

$$||M_n - M|| \rightarrow 0 \text{ p.s. et } \sum_{n=1}^{+\infty} a_n ||M_n - M|| < \infty \text{ p.s.}$$

□

2.4.2 Hypothèses

(H1) Il n'y a pas de relation affine entre les composantes de S ,
(H1') Il n'y a pas de relation affine entre les composantes de R ,

(H2) (R, S) admet des moments d'ordre 4,
(H2') (R, S) admet des moments d'ordre $2r$,
(H2'') R admet des moments d'ordre 4,
(H2⁽³⁾) R admet des moments d'ordre $2r$,

(H3) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$,

(H3') $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[S_1 S_1'])})$,

(H3'') $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{3}{4} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2})$,

(H3⁽³⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[R_1 R_1'])})$,

(H3⁽⁴⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\min(\lambda_{\min}(\mathbb{E}[S_1 S_1']), \lambda_{\min}(\mathbb{E}[R_1 R_1'])})})$,

(H3⁽⁵⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[SS'])})$,

(H3⁽⁶⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\min(\lambda_{\min}(\mathbb{E}[SS']), \lambda_{\min}(\mathbb{E}[R_1 R_1'])})})$,

(H3⁽⁷⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\lambda_{\min}(\mathbb{E}[RR'])})$,

(H3⁽⁸⁾) $(a_n = \frac{a}{n^\alpha}, a > 0, \frac{2}{3} < \alpha < 1)$ ou $(a_n = \frac{a}{n}, a > \frac{1}{2\min(\lambda_{\min}(\mathbb{E}[SS']), \lambda_{\min}(\mathbb{E}[RR'])})})$,

Soit T_n la tribu du passé au temps n , engendrée par $R_1, \dots, R_{n-1}, S_1, \dots, S_{n-1}, A_{11}$ et X_1 , alors :

(H4) Pour tout n , (a) M_n est T_n -mesurable,
(b) M_n converge vers M p.s.,
(c) $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.,

(H4') Pour tout n , (a) N_n est T_n -mesurable,
(b) N_n converge vers N p.s.,
(c) $\sum_{n=1}^{+\infty} a_n \|N_n - N\| < \infty$ p.s.,

(H4) Pour tout n , (a) $\tilde{M}_n$ est T_n -mesurable,
(b) $\tilde{M}_n$ converge vers $\tilde{M}$ p.s.,
(c) $\sum_{n=1}^{+\infty} a_n \|\tilde{M}_n - \tilde{M}\| < \infty$ p.s.,

(H4') Pour tout n , (a) $\tilde{N}_n$ est T_n -mesurable,
(b) $\tilde{N}_n$ converge vers $\tilde{N}$ p.s.,
(c) $\sum_{n=1}^{+\infty} a_n \|\tilde{N}_n - \tilde{N}\| < \infty$ p.s.

Chapitre 3

ACP partielle d'un flux de données d'espérance mathématique variable dans le temps

Sommaire

3.1	Cas général de l'ACP partielle	69
3.1.1	Cas général	70
3.1.2	Cas particulier d'un modèle linéaire de variation de l'espérance	73
3.2	Cas particulier : l'ACG partielle	74
3.2.1	Premier cas : utilisation d'un paquet de données	74
3.2.2	Deuxième cas : utilisation d'un processus d'A.S. de C^k	75
3.2.3	Troisième cas : utilisation d'un processus d'A.S. de C^k (cas particulier du précédent)	76
3.3	Annexes du chapitre 3	77
3.3.1	Lemmes techniques	77
3.3.2	Démonstrations des théorèmes	84
3.3.3	Hypothèses	104

3.1 Cas général de l'ACP partielle

On observe p caractères quantitatifs sur des individus : on obtient des vecteurs de données z_i dans $\mathbb{R}^p$.

On se place ici dans le cas où ces vecteurs arrivent séquentiellement dans le temps (données en ligne) : on observe z_n au temps n ; on a une suite de vecteurs de données $z_1, \dots, z_n, \dots$.

On suppose que, pour tout $n \geq 1$, z_n est la réalisation d'un vecteur aléatoire (v.a.) Z_n dans $\mathbb{R}^p$, défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, d'espérance mathématique θ_n variable dans le temps :

$$Z_n = \theta_n + \tilde{Z}_n,$$

les $\tilde{Z}_n$ constituant un échantillon indépendant et identiquement distribué (i.i.d.) d'un v.a. $\tilde{Z}$ de $\mathbb{R}^p$ d'espérance nulle et de matrice de covariance C .

On a alors, pour tout $n \geq 1$, $\theta_n = \mathbb{E}[Z_n]$ et $C = \mathbb{E}[(Z_n - \theta_n)Z_n']$.

On veut effectuer l'ACP de $\tilde{Z}$, appelée ACP partielle, dans $\mathbb{R}^p$ muni de la métrique M .

Des vecteurs directeurs v_l des axes principaux sont vecteurs propres de la matrice M -symétrique $B = CM$. On pose le problème d'estimer ces vecteurs v_l pour $l = 1, \dots, r$.

On suppose que l'on dispose au temps n d'un estimateur M_n de M tel que $M_n \rightarrow M$ presque sûrement (p.s.) et $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

Le processus (M_n) peut être un processus d'approximation stochastique (a.s.), mis en œuvre parallèlement au processus (X_n) dont la définition suit. On étudiera ultérieurement la définition de (M_n) selon la nature de M .

On définit récursivement un processus d'a.s. $(X_n) = ((X_n^1, \dots, X_n^r))$ de $(v_1, \dots, v_r)$ par :

$$\begin{aligned} B_n &= C_n M_n, \\ F_n(X_n^l) &= \frac{\langle B_n X_n^l, X_n^l \rangle_{M_n}}{\|X_n^l\|_{M_n}^2}, \\ Y_{n+1}^l &= X_n^l + a_n (B_n - F_n(X_n^l)I) X_n^l, \quad l = 1, \dots, r, \\ X_{n+1} &= \text{orth}_{M_n}(Y_{n+1}). \end{aligned}$$

$X_{n+1} = \text{orth}_{M_n}(Y_{n+1})$ signifie que, pour obtenir $X_{n+1} = (X_{n+1}^1, \dots, X_{n+1}^r)$, on orthogonalise au sens de Gram-Schmidt par rapport à M_n le r -uplet $Y_{n+1} = (Y_{n+1}^1, \dots, Y_{n+1}^r)$.

Dans la suite, on étudie trois définitions différentes du processus (C_n) en supposant que l'on dispose au temps n d'un estimateur Θ_n de θ_n .

Dans le cas particulier où l'espérance θ_n peut être représentée par un modèle linéaire, on définit un estimateur Θ_n en utilisant un processus d'a.s. qui sera mis en œuvre parallèlement au processus (X_n) .

On désigne par T_n la tribu du passé au temps n .

3.1.1 Cas général

Dans tout ce paragraphe, on ne définit pas de modèle a priori pour θ_n . On donne trois définitions différentes du processus C_n .

Premier cas : utilisation d'un paquet de données

Dans ce premier cas, on suppose qu'on fait au temps n une ou plusieurs observations ("paquet de données"); on utilise ces observations pour définir C_n .

On suppose que, au temps n :

- on dispose d'un bloc de r_n nouvelles observations indépendantes $Z_{R_{n-1}+1}, \dots, Z_{R_n}$, avec $R_0 = 0$ et
$$R_n = \sum_{j=1}^n r_j,$$
- $R = \sup_n r_n < \infty$,
- on dispose d'estimateurs $(\Theta_{R_{n-1}+1}, \dots, \Theta_{R_n})$ de $(\theta_{R_{n-1}+1}, \dots, \theta_{R_n})$.

On note $I_n = \{R_{n-1} + 1, \dots, R_n\}$ l'ensemble des indices.

On définit $C_n = \frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z_i'$.

Théorème 3.1.

Sous les hypothèses :

(H1) (b) $\tilde{Z}$ admet des moments d'ordre $4r$,

(H2) $\sup_{l \in \mathbb{N}} \|\theta_l\| < \infty$,

(H3) (a) $a_n > 0$, $\sum_{n=1}^{+\infty} a_n = \infty$, $\sum_{n=1}^{+\infty} a_n^2 < \infty$,

(H4) (a) M_n est T_n -mesurable,

(b) $M_n \rightarrow M$ p.s.,

(c) $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.,

(H5) (a) Θ_n est T_n -mesurable,

(b) $\|\Theta_n - \theta_n\| \rightarrow 0$ p.s.,

(c) $\sum_{n=1}^{+\infty} a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| < \infty$ p.s.,

X_n^l converge p.s., pour $l = 1, \dots, r$, vers un vecteur propre de $B = CM$ associé à la l -ième plus grande valeur propre (v.p.) λ_l . ■

Deuxième cas : utilisation d'un processus d'A.S. de C

Dans ce deuxième cas, on définit un processus (C_n) d'a.s. de $C = \text{Cov}(\tilde{Z})$.

Pour tout n , C est solution de l'équation en X , matrice carrée d'ordre p :

$$\mathbb{E} \left[X - (Z_n - \theta_n) Z_n' \right] = 0.$$

Par ailleurs, on suppose qu'au temps n , on dispose d'un estimateur Θ_n de θ_n .

On définit alors le processus (C_n) d'a.s. de C , de type Robbins-Monro, dans l'ensemble des matrices carrées d'ordre p tel que :

$$C_{n+1} = C_n - a_n \left(C_n - (Z_n - \Theta_n) Z_n' \right).$$

Soit les hypothèses :

(H1) (c) $\tilde{Z}$ admet des moments d'ordre 4,

(H3) (b) $(a_n = \frac{a}{n^\alpha}$ avec $a > 0$, $\frac{2}{3} < \alpha < 1$) ou $(\alpha = 1$ i.e. $a_n = \frac{a}{n}$ et $a > 1$),

(H5) (d) $\mathbb{E}[\|\theta_n - \Theta_n\|] = O(\frac{1}{n^\gamma})$, $\gamma > 2(1 - \alpha)$.

Théorème 3.2.

Sous les hypothèses (H1)(c), (H2), (H3)(b) et (H5)(a),(b),(d), C_n converge p.s. vers C et $\sum_{n=1}^{+\infty} a_n \|C_n - C\| < \infty$ p.s.

Sous ces hypothèses et (H4)(b),(c), X_n^l converge p.s., pour $l = 1, \dots, r$, vers un vecteur propre de $B = CM$ associé à la l -ième plus grande v.p. λ_l . ■

Troisième cas : utilisation d'un processus d'A.S. de C (cas particulier du précédent)

Dans ce troisième cas, on se propose d'étudier le même processus que précédemment, dans le cas particulier où $a_n = \frac{1}{n}$.

On peut alors récrire le processus C_n de la manière suivante :

$$C_1 = 0, \quad C_n = \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \Theta_i)(Z_i)'$$

On définit les nouvelles hypothèses :

(H3) (c) $a_n = \frac{1}{n}$,

(H5) (e) $\sum_{n=1}^{+\infty} \frac{1}{n} \left(\mathbb{E}[\|\theta_n - \Theta_n\|^2] \right)^{\frac{1}{2}} < \infty$.

Théorème 3.3.

Sous les hypothèses (H1)(c), (H2), (H3)(c) et (H5)(e), C_n converge p.s. vers C et $\sum_{n=1}^{+\infty} a_n \|C_n - C\| < \infty$ p.s.

Sous ces hypothèses et (H4)(b),(c), X_n^l converge p.s., pour $l = 1, \dots, r$, vers un vecteur propre de $B = CM$ associé à la l -ième plus grande v.p. λ_l . ■

3.1.2 Cas particulier d'un modèle linéaire de variation de l'espérance

Dans ce paragraphe, pour tout n , on choisit un modèle linéaire pour représenter l'espérance θ_n et on définit un processus (Θ_n) d'a.s. de (θ_n) .

Plus précisément, pour $i = 1, \dots, p$, on suppose qu'il existe un vecteur β^i inconnu de $\mathbb{R}^{n_i}$ et, pour tout n , un vecteur U_n^i de $\mathbb{R}^{n_i}$ connu au temps n tels que :

$$\theta_n^i = \langle \beta^i, U_n^i \rangle.$$

On définit le processus d'a.s. (B_n^i) de β^i tel que :

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

On définit alors comme estimateur de θ_n^i la statistique $\Theta_n^i = \langle B_n^i, U_n^i \rangle$ et comme estimateur de θ_n la statistique $\Theta_n = (\Theta_n^1 \dots \Theta_n^p)'$.

On fait l'hypothèse :

$$(H6) \text{ (a) } \max_i \sup_n \|U_n^i\| < \infty,$$

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

Théorème 3.4.

On définit : $a_n = \frac{a}{n^\alpha}$ avec $a > 0$. Alors, sous (H6)(a),(b) :

_pour $\frac{2}{3} < \alpha \leq 1$, les hypothèses concernant Θ_n du théorème 3.1 sont vérifiées ;

_pour $\frac{4}{5} < \alpha \leq 1$, les hypothèses concernant Θ_n du théorème 3.2 sont vérifiées ;

_pour $\alpha = 1$, les hypothèses concernant Θ_n du théorème 3.3 sont vérifiées. ■

3.2 Cas particulier : l'ACG partielle

Comme dans [51], on se place dans le cas où le v.a. $\tilde{Z}$ est partitionné en sous-vecteurs $\tilde{Z}^{(1)}, \dots, \tilde{Z}^{(q)}$.

Pour $k = 1, \dots, q$, $\tilde{Z}^{(k)}$ est un v.a. dans $\mathbb{R}^{m_k}$, de composantes $\tilde{Z}^{k1}, \dots, \tilde{Z}^{km_k}$.

On note également $\tilde{Z}^{(k)} = \{\tilde{Z}^i, i \in J_k\}$, $\text{card}(J_k) = m_k$, $\sum_{k=1}^q m_k = p$.

On définit $C^k = \mathbb{E}[\tilde{Z}^{(k)}(\tilde{Z}^{(k)})']$, matrice de covariance de $\tilde{Z}^{(k)}$.

On souhaite effectuer une ACP de $\tilde{Z}$ dans laquelle les v.a. $\tilde{Z}^{(k)}$ aient un rôle équilibré : on veut éviter que les premiers facteurs soient principalement déterminés à partir de certains vecteurs $\tilde{Z}^{(k)}$.

L'analyse canonique généralisée du v.a. $\tilde{Z}$ fournit une solution à ce problème grâce à un choix particulier de métrique. L'ACG de $\tilde{Z}$ est une ACP avec la métrique diagonale par blocs d'ordre p , M :

$$M = \begin{pmatrix} (C^1)^{-1} & & & \\ & \cdot & & 0 \\ & & \cdot & \\ 0 & & & \cdot \\ & & & & (C^q)^{-1} \end{pmatrix}$$

Le v.a. Z_n observé est partitionné en sous-vecteurs $Z_n^{(1)}, \dots, Z_n^{(q)}$, de dimensions respectives $m_1, \dots, m_q$, avec :

$$Z_n^{(k)} = \theta_n^{(k)} + \tilde{Z}_n^{(k)},$$

$\theta_n^{(k)}$ étant le vecteur des θ_n^i , $i \in J_k$, et les $\tilde{Z}_n^{(k)}$ constituant un échantillon i.i.d. de $\tilde{Z}^{(k)}$.

Soit $\Theta_n^{(k)}$ le vecteur des Θ_n^i , $i \in J_k$.

Dans la suite, pour tout n et pour $k = 1, \dots, q$, on étudie trois processus d'a.s. différents (M_n^k) de $(C^k)^{-1}$.

On définit alors comme estimateur de M au pas n la matrice diagonale par blocs M_n qui a pour k -ième bloc diagonal M_n^k .

3.2.1 Premier cas : utilisation d'un paquet de données

Pour tout n , pour $k = 1, \dots, q$, $(C^k)^{-1}$ est solution de l'équation en X , matrice carrée d'ordre m_k :

$$\mathbb{E} \left[(Z_n^{(k)} - \theta_n^{(k)})(Z_n^{(k)})' X - I \right] = 0$$

où I est la matrice-identité d'ordre m_k .

Dans la suite, on remplace les (k) par des k (sans parenthèses) pour simplifier les notations.

On définit :

$$M_{n+1}^k = M_n^k - a_n \left(\left(\frac{1}{r_n} \sum_{i \in I_n} (Z_i^k - \Theta_i^k)(Z_i^k)' \right) M_n^k - I \right).$$

Soit les hypothèses :

(H1) (a) Il n'existe pas de relation affine entre les composantes de $\tilde{Z}$,

(H3) (d) ($a_n = \frac{a}{n^\alpha}$ avec $a > 0$, $\frac{2}{3} < \alpha < 1$) ou ($\alpha = 1$ i.e. $a_n = \frac{a}{n}$ et $a > \frac{1}{\min_k \lambda_{\min}(C^k)}$).

Théorème 3.5.

Sous les hypothèses (H1)(a),(c), (H2), (H3)(d) et (H5)(a),(b),(d), les hypothèses concernant M_n du théorème 1 sont vérifiées. ■

3.2.2 Deuxième cas : utilisation d'un processus d'A.S. de C^k

C^k est solution de l'équation en X , matrice carrée d'ordre m_k :

$$\mathbb{E} \left[X - (Z_n^k - \theta_n^k)(Z_n^k)' \right] = 0.$$

On définit :

$$C_{n+1}^k = C_n^k - a_n \left(C_n^k - (Z_n^k - \Theta_n^k)(Z_n^k)' \right).$$

$(C^k)^{-1}$ est solution de l'équation en X , matrice carrée d'ordre m_k : $C^k X - I = 0$.

On définit :

$$M_{n+1}^k = M_n^k - a_n (C_n^k M_n^k - I).$$

Théorème 3.6.

Sous les hypothèses (H1)(a),(c), (H2), (H3)(b) et (H5)(a),(b),(d), les hypothèses concernant M_n des théorèmes 3.2 et 3.3 sont vérifiées. ■

3.2.3 Troisième cas : utilisation d'un processus d'A.S. de C^k (cas particulier du précédent)

On considère le cas particulier du précédent où $a_n = \frac{1}{n}$, avec $C_1^k = 0$.

On a alors :

$$C_n^k = \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i^k - \Theta_i^k)(Z_i^k)'$$

et

$$M_{n+1}^k = M_n^k - a_n \left(\frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i^k - \Theta_i^k)(Z_i^k)' M_n^k - I \right).$$

Théorème 3.7.

Sous les hypothèses (H1)(a),(c), (H2), (H3)(c) et (H5)(e), les hypothèses concernant M_n des théorèmes 3.2 et 3.3 sont vérifiées. ■

3.3 Annexes du chapitre 3

3.3.1 Lemmes techniques

Lemme 3.8.

Soit $C_n = \frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z_i'$.

Sous (H2) et (H5)(a),(c), on a $\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| < \infty$ p.s. ■

Démonstration.

$$\begin{aligned}
\mathbb{E}[C_n|T_n] - C &= \mathbb{E}\left[\frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z_i' | T_n\right] - C \\
&= \frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}\left[(Z_i - \Theta_i) Z_i' | T_n\right] - C \\
&= \frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}\left[(Z_i - \theta_i + \theta_i - \Theta_i)(Z_i - \theta_i + \theta_i)' | T_n\right] - C \\
&= \frac{1}{r_n} \sum_{i \in I_n} \left(\mathbb{E}\left[(Z_i - \theta_i)(Z_i - \theta_i)' | T_n\right] + \mathbb{E}\left[(Z_i - \theta_i)\theta_i' | T_n\right] + (\theta_i - \Theta_i)\mathbb{E}[Z_i' | T_n] \right) - C \text{ sous (H5)(a)} \\
&= \frac{1}{r_n} \sum_{i \in I_n} \left(\mathbb{E}[\tilde{Z}\tilde{Z}'] + 0 + (\theta_i - \Theta_i)\theta_i' \right) - C \text{ car } Z_i, i \in I_n, \text{ ne dépend pas du passé} \\
&= \frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i)\theta_i'.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
\|\mathbb{E}[C_n|T_n] - C\| &= \left\| \frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i)\theta_i' \right\| \\
&\leq c_1 \sum_{i \in I_n} \|\theta_i - \Theta_i\| \text{ sous (H2)}.
\end{aligned}$$

et

$$\begin{aligned}
\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| &\leq c_1 \sum_{n=1}^{+\infty} a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| \\
&< \infty \text{ sous (H5)(c)}.
\end{aligned}$$

Par conséquent, $\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| < \infty$ p.s.

□

Lemme 3.9.

Soit $C_n = \frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z'_i$.

Sous (H1)(b), (H2), (H3)(a) et (H5)(a),(b) $\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||C_n - C||^j | T_n] < \infty$ p.s. pour $j = 2, \dots, 2r$. ■

Démonstration.

$$\begin{aligned}
\mathbb{E}[||C_n - C||^j | T_n] &= \mathbb{E}\left[||\frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z'_i - C||^j | T_n\right] \\
&\leq 2^{j-1} \mathbb{E}\left[||\frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z'_i||^j + ||C||^j | T_n\right] \text{ par inégalité de convexité} \\
&\leq 2^{j-1} \left(\frac{1}{r_n^j} \mathbb{E}\left[||\sum_{i \in I_n} (Z_i - \Theta_i) Z'_i||^j | T_n\right] + ||C||^j \right) \\
&\leq 2^{j-1} \left(\frac{1}{r_n^j} r_n^{j-1} \mathbb{E}\left[\sum_{i \in I_n} ||(Z_i - \Theta_i) Z'_i||^j | T_n\right] + ||C||^j \right) \text{ par inégalité de convexité à nouveau} \\
&\leq 2^{j-1} \left(\frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}\left[||Z_i - \Theta_i||^j ||Z'_i||^j | T_n\right] + ||C||^j \right) \text{ par linéarité de l'espérance} \\
&\leq 2^{j-1} c_1 \left(\sum_{i \in I_n} \mathbb{E}\left[||Z_i - \Theta_i||^j ||Z_i||^j | T_n\right] + 1 \right).
\end{aligned}$$

Ensuite, en décomposant $Z_i - \Theta_i$, on obtient :

$$\begin{aligned}
\mathbb{E}[||C_n - C||^j | T_n] &\leq 2^{j-1} c_1 \left(\sum_{i \in I_n} \mathbb{E}\left[||Z_i - \theta_i + \theta_i - \Theta_i||^j ||Z_i||^j | T_n\right] + 1 \right) \\
&\leq c_2 \left(\sum_{i \in I_n} \left(\mathbb{E}\left[||Z_i - \theta_i||^j ||Z_i||^j | T_n\right] + \mathbb{E}\left[||\theta_i - \Theta_i||^j ||Z_i||^j | T_n\right] \right) + 1 \right)
\end{aligned}$$

Alors, en décomposant Z_i , on trouve :

$$\begin{aligned}
\mathbb{E}[||C_n - C||^j | T_n] &\leq c_2 \left(\sum_{i \in I_n} \left(\mathbb{E}\left[||Z_i - \theta_i||^j ||Z_i - \theta_i + \theta_i||^j | T_n\right] + \mathbb{E}\left[||\theta_i - \Theta_i||^j ||Z_i - \theta_i + \theta_i||^j | T_n\right] \right) + 1 \right) \\
&\leq c_3 \left(\sum_{i \in I_n} \left(\mathbb{E}\left[||Z_i - \theta_i||^{2j} | T_n\right] + \mathbb{E}\left[||Z_i - \theta_i||^j ||\theta_i||^j | T_n\right] + ||\theta_i - \Theta_i||^j \mathbb{E}\left[||Z_i - \theta_i||^j | T_n\right] \right. \right. \\
&\quad \left. \left. + ||\theta_i - \Theta_i||^j ||\theta_i||^j \right) + 1 \right) \text{ sous (H5)(a)} \\
&\leq c_3 \left(\sum_{i \in I_n} \left(\mathbb{E}\left[||\tilde{Z}_i||^{2j}\right] + \mathbb{E}\left[||\tilde{Z}_i||^j ||\theta_i||^j\right] + ||\theta_i - \Theta_i||^j \mathbb{E}\left[||\tilde{Z}_i||^j\right] + ||\theta_i - \Theta_i||^j ||\theta_i||^j \right) + 1 \right)
\end{aligned}$$

Comme les $\tilde{Z}_i$ sont i.i.d., on peut écrire :

$$\begin{aligned}
\mathbb{E}[||C_n - C||^j | T_n] &\leq c_3 \left(\sum_{i \in I_n} \left(\mathbb{E}[||\tilde{Z}||^{2j}] + \mathbb{E}[||\tilde{Z}||^j ||\boldsymbol{\theta}_i||^j] + ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j \mathbb{E}[||\tilde{Z}||^j] + ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j ||\boldsymbol{\theta}_i||^j \right) + 1 \right) \\
&\leq c_4 \left(\sum_{i \in I_n} \left(\mathbb{E}[||\tilde{Z}||^{2j}] + \mathbb{E}[||\tilde{Z}||^j] + ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j \mathbb{E}[||\tilde{Z}||^j] + ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j \right) + 1 \right) \text{ sous (H2)} \\
&\leq c_5 \sum_{i \in I_n} (1 + ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j) \text{ sous (H1)(b)} \\
&\leq c_5 (r_n + \sum_{i \in I_n} ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j) \\
&\leq c_6 (1 + \sum_{i \in I_n} ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j) \text{ car } \sup_n r_n = R < \infty.
\end{aligned}$$

donc,

$$\begin{aligned}
\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||C_n - C||^j | T_n] &\leq c_7 \sum_{n=1}^{+\infty} a_n^j + c_4 \sum_{n=1}^{+\infty} a_n^j \sum_{i \in I_n} ||\boldsymbol{\theta}_i - \boldsymbol{\Theta}_i||^j \\
&< \infty \text{ sous (H3)(a) et (H5)(b) .}
\end{aligned}$$

□

Tout d'abord, on rappelle un lemme de [47] :

Lemme 3.10.

Soit (a_n) et (g_n) deux suites de réels positifs, (b_n) et (ψ_n) deux suites de réels non négatifs, λ, μ deux réels positifs, Ψ un réel inférieur à λ , tels que :

- 1) $\forall n, b_{n+1} \leq (1 - \lambda a_n) b_n + \mu g_n$,
- 2) $a_n = o(1)$ et $\sum_{n=1}^{+\infty} a_n = \infty$,
- 3) $\forall n, \frac{a_{n+1}}{g_{n+1}} \leq \frac{a_n}{g_n} \left(1 + \Psi a_n + \psi_n + o(a_n) \right)$,
- 4) $\sum_{n=1}^{+\infty} \psi_n < \infty$,

Alors, $\overline{\lim}_{g_n} \frac{a_n}{b_n} \leq \frac{\mu}{\lambda - \Psi}$. ■

Lemme 3.11.

Soit (a_n) , (h_n) deux suites de réels positifs, (b_n) une suite de réels non négatifs, λ, μ deux réels positifs tels que :

- 1') $\forall n, b_{n+1} \leq (1 - \lambda a_n) b_n + \mu a_n h_n$,
- 2') $h_n = \frac{1}{n^\gamma}$, $\gamma > 0$,
- 3') $(a_n = \frac{a}{n^\alpha}$ avec $a > 0$, $0 < \alpha < 1$) ou $(\alpha = 1$ i.e. $a_n = \frac{a}{n}$ et $a > \frac{\gamma}{\lambda})$,

Alors, $\overline{\lim}_{h_n} \frac{1}{b_n} < \infty$. ■

Démonstration. On vérifie les hypothèses d'application du lemme 3.10 avec $\psi_n = 0$ et $g_n = a_n h_n$:

$$\begin{aligned} \frac{h_n}{h_{n+1}} &= \frac{\frac{1}{n^\gamma}}{\frac{1}{(n+1)^\gamma}} \\ &= \left(1 + \frac{1}{n}\right)^\gamma \\ &= 1 + \frac{\gamma}{n} + o\left(\frac{1}{n}\right) \\ &\leq 1 + \gamma a_n \left(\frac{1}{an^{1-\alpha}}\right) + o(a_n). \end{aligned}$$

Si $\alpha < 1$, $\frac{\gamma}{an^{1-\alpha}} < \lambda$ à partir d'un certain rang sous la condition 3').

Si $\alpha = 1$, $\frac{\gamma}{a} < \lambda$ sous la condition 3').

Dans les deux cas, on peut appliquer le lemme 3.10 ce qui donne le résultat.

□

Lemme 3.12.

Soit une suite de matrices aléatoires i.i.d. (M_n) de dimension (p, q) telle que $M_n \stackrel{loi}{=} M$, $\mathbb{E}[M] = A$.

Soit $\overline{M}_n = \frac{1}{n} \sum_{i=1}^n M_i$.

On suppose que pour tout $k = 1, \dots, p$ et pour tout $l = 1, \dots, q$, $\text{Var}[M(k, l)]$ existe.

Alors, on a : $\mathbb{E}[|\overline{M}_n - A|] = O(\frac{1}{\sqrt{n}})$. ■

Démonstration.

$$\begin{aligned}
 \mathbb{E}[|\overline{M}_n - A|^2] &= \mathbb{E}\left[\sum_{k=1}^p \sum_{l=1}^q \left(\overline{M}_n(k, l) - A(k, l)\right)^2\right] \text{ en utilisant la norme de Hilbert-Schmidt} \\
 &= \sum_{k=1}^p \sum_{l=1}^q \text{Var}[\overline{M}_n(k, l)] \\
 &= \frac{1}{n} \sum_{k=1}^p \sum_{l=1}^q \text{Var}[M(k, l)] \\
 &= O(\frac{1}{n}).
 \end{aligned}$$

Ainsi, $\mathbb{E}[|\overline{M}_n - A|] = O(\frac{1}{\sqrt{n}})$.

□

Lemme 3.13.

On définit une matrice carrée d'ordre p qu'on note $H_i = (Z_i - \mathbb{E}[Z_i])\mathbb{E}[Z_i']$ d'éléments $H_i^{jk} = (Z_i^j - \mathbb{E}[Z_i^j])\mathbb{E}[Z_i^k] = (\tilde{Z}_i^j)\theta_i^k$.

Soit $\overline{H}_n = \frac{1}{n} \sum_{i=1}^n H_i$.

Sous (H1)(c) et (H2), on a : $\sum_{n=1}^{+\infty} a_n \mathbb{E}[|\overline{H}_n|] = O\left(\sum_{n=1}^{+\infty} \frac{a_n}{\sqrt{n}}\right)$. ■

Démonstration.

$$\begin{aligned}
\sum_{n=1}^{+\infty} a_n \mathbb{E}[|\overline{H}_n|] &\leq \sum_{n=1}^{+\infty} a_n \left(\mathbb{E}[|\overline{H}_n|^2] \right)^{\frac{1}{2}} \\
&\leq \sum_{n=1}^{+\infty} a_n \left(\mathbb{E} \left[\sum_{j=1}^p \sum_{k=1}^p (\overline{H}_n^{jk})^2 \right] \right)^{\frac{1}{2}} \text{ en utilisant la norme de Hilbert-Schmidt} \\
&\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^p \sum_{k=1}^p \mathbb{E} \left[(\overline{H}_n^{jk})^2 \right] \right)^{\frac{1}{2}} \\
&\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^p \sum_{k=1}^p \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n (Z_i^j - \mathbb{E}[Z_i^j]) \mathbb{E}[Z_i^k] \right)^2 \right] \right)^{\frac{1}{2}} \\
&\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^p \sum_{k=1}^p \text{Var} \left(\frac{1}{n} \sum_{i=1}^n (Z_i^j - \mathbb{E}[Z_i^j]) \mathbb{E}[Z_i^k] \right) \right)^{\frac{1}{2}} \text{ car } \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n (Z_i^j - \mathbb{E}[Z_i^j]) \mathbb{E}[Z_i^k] \right] = 0 \\
&\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^p \sum_{k=1}^p \text{Var} \left(\frac{1}{n} \sum_{i=1}^n \tilde{Z}_i^j \theta_i^k \right) \right)^{\frac{1}{2}} \\
&\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{j=1}^p \sum_{k=1}^p \frac{1}{n^2} \sum_{i=1}^n \text{Var}[\tilde{Z}_i^j] (\theta_i^k)^2 \right)^{\frac{1}{2}} \text{ puisque les } \tilde{Z}_i^j \text{ sont i.i.d.} \\
&\leq c_1 \sum_{n=1}^{+\infty} \frac{a_n}{n} \left(\sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^p \text{Var}[\tilde{Z}_i^j] \right)^{\frac{1}{2}} \text{ sous (H2)} \\
&\leq c_1 \sum_{n=1}^{+\infty} \frac{a_n}{n} \left(\sum_{i=1}^n c_2 \right)^{\frac{1}{2}} \text{ sous (H1)(c)} \\
&\leq c_3 \sum_{n=1}^{+\infty} \frac{a_n}{n} \cdot \sqrt{n} \\
&\leq c_3 \sum_{n=1}^{+\infty} \frac{a_n}{\sqrt{n}}.
\end{aligned}$$

□

Lemme 3.14.

Soit, pour tout $n \geq 1$, w_n une suite de nombres positifs ou nuls tels que :

$$w_{n+1} = (1 - a_n)w_n + a_n u_n, \quad w_1 = 0, \quad u_n > 0, \quad 0 < a_n < 1.$$

Si $\sum_{n=1}^{+\infty} a_n = \infty$ et $\sum_{n=1}^{+\infty} a_n u_n < \infty$, alors $\sum_{n=1}^{+\infty} a_n w_n < \infty$ et $w_n \rightarrow 0$.

On remarque que, pour $a_n = \frac{1}{n}$, $w_{n+1} = \frac{1}{n} \sum_{j=1}^n u_j$. ■

Démonstration.

Pour $n > 1$, $w_n > 0$ ($a_n u_n > 0$, $1 - a_n > 0$).

Alors, d'après le lemme de Robbins-Siegmund, il existe $w \geq 0$: $w_n \rightarrow w$ et $\sum_{n=1}^{+\infty} a_n w_n < \infty$.

Comme $\sum_{n=1}^{+\infty} a_n = \infty$, on a $w = 0$.

□

3.3.2 Démonstrations des théorèmes

Démonstration du théorème 3.1 :

On vérifie les hypothèses d'application du théorème 1.2.

1) Sous (H4), M_n est T_n -mesurable, $M_n \rightarrow M$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

2) On se propose maintenant de montrer que $\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[B_n|T_n] - B\| < \infty$ p.s.

On commence par écrire la décomposition :

$$B_n - B = C_n M_n - CM = (C_n - C)M_n + C(M_n - M)$$

dont on déduit que $\mathbb{E}[B_n|T_n] - B = (\mathbb{E}[C_n|T_n] - C)M_n + C(M_n - M)$, M_n étant T_n -mesurable,

$$\text{puis que } \sum_{n=1}^{+\infty} a_n \|\mathbb{E}[B_n|T_n] - B\| \leq \sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| \cdot \|M_n\| + \|C\| \sum_{n=1}^{+\infty} a_n \|M_n - M\|.$$

Comme sous (H4)(b) $M_n \rightarrow M$ p.s., il existe une v.a. P positive et finie p.s. telle que :

$$\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[B_n|T_n] - B\| \leq P \left(\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| + \sum_{n=1}^{+\infty} a_n \|M_n - M\| \right).$$

Sous (H4)(c), le second terme du second membre est fini p.s.

D'après le lemme 3.8, le premier terme l'est également sous (H2) et (H5)(a),(c).

Ainsi, on peut conclure que $\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[B_n|T_n] - B\| < \infty$ p.s.

3) On montre enfin que, pour $j = 2, \dots, 2r$, $\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||B_n - B||^j | T_n] < \infty$ p.s.

Tout d'abord,

$$\begin{aligned} ||B_n - B||^j &= ||(C_n - C)M_n + C(M_n - M)||^j \\ &\leq 2^{j-1} (||C_n - C||^j ||M_n||^j + ||C||^j ||M_n - M||^j). \end{aligned}$$

Alors,

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||B_n - B||^j | T_n] \leq c_1 \left(\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||C_n - C||^j | T_n] \cdot ||M_n||^j + ||C||^j \sum_{n=1}^{+\infty} a_n^j ||M_n - M||^j \right)$$

puisque M_n est T_n -mesurable

et comme $M_n \rightarrow M$ p.s., il existe une v.a. Q positive et finie p.s. telle que :

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||B_n - B||^j | T_n] \leq Q \left(\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||C_n - C||^j | T_n] + \sum_{n=1}^{+\infty} a_n^j ||M_n - M||^j \right).$$

Or, l'hypothèse (H4)(c) nous assure que le second terme du second membre est fini p.s. tandis que le lemme 3.9 nous donne que le premier l'est également.

On en déduit que, pour $j = 2, \dots, 2r$, $\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[||B_n - B||^j | T_n] < \infty$ p.s.

Démonstration du théorème 3.2 :

On vérifie les hypothèses d'application du théorème 1.2.

1) Approximation stochastique de la matrice de covariance C :

a) Montrons en premier lieu que $C_n \rightarrow C$ p.s.

Tout d'abord,

$$\begin{aligned} C_{n+1} - C &= (C_n - C) - a_n \left((C_n - C) + C - ((Z_n - \theta_n) + (\theta_n - \Theta_n)) Z'_n \right) \\ &= (1 - a_n)(C_n - C) - a_n U_n \text{ avec } U_n = C - (Z_n - \theta_n) Z'_n - (\theta_n - \Theta_n) Z'_n. \end{aligned}$$

Ainsi,

$$\begin{aligned} \|C_{n+1} - C\|^2 &= \|(1 - a_n)(C_n - C) - a_n U_n\|^2 \\ &\leq (1 - a_n)^2 \|C_n - C\|^2 + a_n^2 \|U_n\|^2 - 2a_n(1 - a_n) \langle C_n - C, U_n \rangle. \end{aligned}$$

et par suite,

$$\mathbb{E}[\|C_{n+1} - C\|^2 | T_n] \leq (1 - a_n)^2 \|C_n - C\|^2 + a_n^2 \mathbb{E}[\|U_n\|^2 | T_n] - 2a_n(1 - a_n) \langle C_n - C, \mathbb{E}[U_n | T_n] \rangle.$$

Or, si on étudie plus en détails le dernier terme du second membre, on remarque que :

$$\begin{aligned} \mathbb{E}[U_n | T_n] &= \mathbb{E}[C - (Z_n - \theta_n) Z'_n - (\theta_n - \Theta_n) Z'_n | T_n] \\ &= C - \mathbb{E}[(Z_n - \theta_n) Z'_n] - (\theta_n - \Theta_n) \mathbb{E}[Z'_n] \text{ sous (H5)(a)} \\ &= C - C - (\theta_n - \Theta_n) \theta'_n \\ &= -(\theta_n - \Theta_n) \theta'_n. \end{aligned}$$

Ainsi,

$$\begin{aligned} -2a_n(1 - a_n) \langle C_n - C, \mathbb{E}[U_n | T_n] \rangle &= -2a_n(1 - a_n) \langle C_n - C, -(\theta_n - \Theta_n) \theta'_n \rangle \\ &\leq 2a_n(1 - a_n) \|C_n - C\| \cdot \|\theta_n - \Theta_n\| \cdot \|\theta'_n\| \\ &\leq 2a_n \|C_n - C\| \cdot \|\theta_n - \Theta_n\| \cdot \|\theta_n\| \\ &\leq c_1 a_n \|C_n - C\| \cdot \|\theta_n - \Theta_n\| \text{ sous (H2)} \\ &\leq c_1 a_n (1 + \|C_n - C\|^2) \|\theta_n - \Theta_n\|. \end{aligned}$$

De même, pour le deuxième terme du second membre, on trouve que :

$$\begin{aligned}
\mathbb{E}[|U_n|^2 | T_n] &= \mathbb{E}[|C - (Z_n - \theta_n)Z'_n - (\theta_n - \Theta_n)Z'_n|^2 | T_n] \\
&\leq \mathbb{E}\left[2(|C - (Z_n - \theta_n)Z'_n|^2 + |\theta_n - \Theta_n|^2 |Z'_n|^2) | T_n\right] \\
&\leq 2\left(\mathbb{E}[|\mathbb{E}[(Z_n - \theta_n)Z'_n] - (Z_n - \theta_n)Z'_n|^2] + |\theta_n - \Theta_n|^2 \mathbb{E}[|Z'_n|^2]\right) \\
&\leq c_2(1 + |\theta_n - \Theta_n|^2) \text{ sous (H1)(c), (H2) et (H5)(a) .}
\end{aligned}$$

On déduit des deux derniers développements que :

$$\begin{aligned}
\mathbb{E}[|C_{n+1} - C|^2 | T_n] &\leq (1 - a_n)^2 |C_n - C|^2 + a_n^2 c_2 (1 + |\theta_n - \Theta_n|^2) + c_1 a_n (1 + |C_n - C|^2) |\theta_n - \Theta_n| \\
&\leq (1 + a_n^2 + c_1 a_n |\theta_n - \Theta_n|) |C_n - C|^2 + c_2 a_n^2 (1 + |\theta_n - \Theta_n|^2) + c_1 a_n |\theta_n - \Theta_n| - 2a_n |C_n - C|^2.
\end{aligned}$$

On démontre maintenant que $\sum_{n=1}^{+\infty} a_n |\theta_n - \Theta_n| < \infty$ p.s.

$$\begin{aligned}
\sum_{n=1}^{+\infty} a_n \mathbb{E}[|\theta_n - \Theta_n|] &= \sum_{n=1}^{+\infty} a_n O\left(\frac{1}{n^\gamma}\right) \text{ sous (H5)(d)} \\
&= \sum_{n=1}^{+\infty} O\left(\frac{1}{n^{\alpha+\gamma}}\right) \text{ sous (H3)(b).}
\end{aligned} \tag{3.1}$$

Sous (H3)(b) et (H5)(d), $\alpha + \gamma > \alpha + 2(1 - \alpha) = 2 - \alpha \geq 1$.

Ainsi, $\sum_{n=1}^{+\infty} a_n \mathbb{E}[|\theta_n - \Theta_n|] < \infty$ et par suite, $\sum_{n=1}^{+\infty} a_n |\theta_n - \Theta_n| < \infty$ p.s.

On peut alors appliquer le lemme de Robbins-Siegmund :

Il existe une v.a. $T \geq 0$ telle que $|C_n - C| \rightarrow T$ p.s. et $\sum_{n=1}^{+\infty} a_n |C_n - C|^2 < \infty$ p.s.

Or, sous (H3)(b), $\sum_{n=1}^{+\infty} a_n = \infty$, donc $T = 0$ et $C_n \rightarrow C$ p.s.

b) Montrons maintenant que $\sum_{n=1}^{+\infty} a_n |C_n - C| < \infty$ p.s.

Tout d'abord, remarquons que sous (H5)(b), $\|\theta_n - \Theta_n\| \rightarrow 0$ p.s.

Ainsi, $\forall \delta > 0, \forall \varepsilon > 0, \exists N(\delta, \varepsilon), \forall n \geq N(\delta, \varepsilon), \mathbb{P}(\sup_{k > n} \|\theta_k - \Theta_k\| \leq \delta) > 1 - \varepsilon$.

On fixe alors δ et ε on note $N_0 = N(\delta, \varepsilon)$.

On pose :

$$F = \left\{ \sup_{k > N_0} \|\theta_k - \Theta_k\| \leq \delta \right\},$$

$$F_n = \left\{ \sup_{N_0 < k \leq n} \|\theta_k - \Theta_k\| \leq \delta \right\}.$$

On peut déjà remarquer que $\mathbb{P}(F) > 1 - \varepsilon$ et que F_n est T_n -mesurable.

De plus, on a $F \Rightarrow F_{n+1} \Rightarrow F_n$ i.e. $F \subset F_{n+1} \subset F_n$ ou encore $\mathbb{1}_F \leq \mathbb{1}_{F_{n+1}} \leq \mathbb{1}_{F_n} \leq 1$.

Donc, pour $n > N_0$:

$$\begin{aligned} \mathbb{E}[\|C_{n+1} - C\|^2 \mathbb{1}_{F_{n+1}} | T_n] &\leq \mathbb{E}[\|C_{n+1} - C\|^2 \mathbb{1}_{F_n} | T_n] \\ &\leq \mathbb{E}[\|C_{n+1} - C\|^2 | T_n] \mathbb{1}_{F_n}. \end{aligned}$$

De la majoration de $\mathbb{E}[\|C_{n+1} - C\|^2 | T_n]$ établie précédemment, on déduit que :

$$\begin{aligned} \mathbb{E}[\|C_{n+1} - C\|^2 \mathbb{1}_{F_{n+1}} | T_n] &\leq (1 + a_n^2 + c_1 a_n \|\theta_n - \Theta_n\|) \|C_n - C\|^2 \mathbb{1}_{F_n} + \\ &\quad c_1 a_n \|\theta_n - \Theta_n\| + c_2 a_n^2 (1 + \|\theta_n - \Theta_n\|^2) \mathbb{1}_{F_n} - 2a_n \|C_n - C\|^2 \mathbb{1}_{F_n}. \end{aligned}$$

Or, dans F_n , $\|\theta_n - \Theta_n\| \leq \delta$, donc :

$$c_1 a_n \|\theta_n - \Theta_n\| \cdot \|C_n - C\|^2 \mathbb{1}_{F_n} \leq c_1 a_n \delta \|C_n - C\|^2 \mathbb{1}_{F_n}.$$

On choisit δ tel que $e\delta < 1$.

On peut alors écrire,

$$c_1 a_n \|\theta_n - \Theta_n\| \cdot \|C_n - C\|^2 \mathbb{1}_{F_n} \leq a_n \|C_n - C\|^2 \mathbb{1}_{F_n},$$

et par suite,

$$\mathbb{E}[\|C_{n+1} - C\|^2 \mathbb{1}_{F_{n+1}} | T_n] \leq (1 + a_n^2) \|C_n - C\|^2 \mathbb{1}_{F_n} + c_1 a_n \|\theta_n - \Theta_n\| + c_2 (1 + \delta^2) a_n^2 - a_n \|C_n - C\|^2 \mathbb{1}_{F_n}.$$

En passant à l'espérance mathématique, on obtient :

$$\mathbb{E}[||C_{n+1} - C||^2 \mathbb{1}_{F_{n+1}}] \leq (1 + a_n^2) \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] + c_1 a_n \mathbb{E}[||\theta_n - \Theta_n||] + c_2 (1 + \delta^2) a_n^2 - a_n \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}].$$

Sous (H3)(b) et (H5)(d), on peut de nouveau appliquer le lemme de Robbins-Siegmund dont on déduit que :

$$\text{Il existe un réel } t \geq 0 \text{ tel que } \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] \rightarrow t \text{ et } \sum_{n=1}^{+\infty} a_n \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] < \infty.$$

$$\text{Comme } \sum_{n=1}^{+\infty} a_n = \infty, t = 0.$$

Ainsi, il existe $\mu > 0$ tel que :

$$\begin{aligned} \mathbb{E}[||C_{n+1} - C||^2 \mathbb{1}_{F_{n+1}}] &\leq (1 - a_n) \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] + a_n g_n \text{ avec } g_n = O(a_n + \mathbb{E}[||\theta_n - \Theta_n||]) = O\left(\frac{1}{n^{\min(\alpha, \gamma)}}\right) \\ &\leq (1 - a_n) \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] + \mu a_n h_n \text{ avec } h_n = \frac{1}{n^{\min(\alpha, \gamma)}}. \end{aligned}$$

Sous les hypothèses (H3)(b) et (H5)(d), d'après le lemme 3.10, on a :

$$\overline{\lim} \frac{1}{h_n} \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] < \infty.$$

$$\text{Alors, } \mathbb{E}[||C_n - C||^2 \mathbb{1}_{F_n}] = O(h_n) \text{ d'où } \mathbb{E}[||C_n - C|| \mathbb{1}_{F_n}] = O(\sqrt{h_n}).$$

$$\text{On en déduit que sous (H3)(b) et (H5)(d), } \sum_{n=1}^{+\infty} a_n \mathbb{E}[||C_n - C|| \mathbb{1}_{F_n}] < \infty \text{ puis } \sum_{n=1}^{+\infty} a_n ||C_n - C|| \mathbb{1}_{F_n} < \infty \text{ p.s.}$$

$$\text{Comme } \mathbb{1}_F \leq \mathbb{1}_{F_n}, \text{ on peut écrire que } \sum_{n=1}^{+\infty} a_n ||C_n - C|| \mathbb{1}_F < \infty \text{ p.s.}$$

Ainsi, d'un point de vue logique,

$$\{F\} \Rightarrow \left\{ \sum_{n=1}^{+\infty} a_n ||C_n - C|| < \infty \right\}$$

ce qui implique, d'un point de vue probabiliste, que :

$$\mathbb{P}\left(\sum_{n=1}^{+\infty} a_n ||C_n - C|| < \infty\right) \geq \mathbb{P}(F) > 1 - \varepsilon$$

selon la remarque du début de partie.

$$\varepsilon \text{ étant quelconque, on en déduit finalement que } \sum_{n=1}^{+\infty} a_n ||C_n - C|| < \infty \text{ p.s.}$$

2) Sous (H4), $M_n \rightarrow M$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

3) On montre finalement que $\sum_{n=1}^{+\infty} a_n \|B_n - B\| < \infty$ p.s.

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n \|B_n - B\| &= \sum_{n=1}^{+\infty} a_n \|C_n M_n - C M\| \\ &\leq \sum_{n=1}^{+\infty} a_n \|C_n\| \cdot \|M_n - M\| + \sum_{n=1}^{+\infty} a_n \|C_n - C\| \cdot \|M\|. \end{aligned}$$

Comme $C_n \rightarrow C$ p.s., il existe une v.a. F positive et finie p.s. telle que :

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n \|B_n - B\| &\leq F \left(\sum_{n=1}^{+\infty} a_n \|M_n - M\| + \sum_{n=1}^{+\infty} a_n \|C_n - C\| \right) \\ &< \infty \text{ d'après les conclusions des deux premières parties de la démonstration.} \end{aligned}$$

Démonstration du théorème 3.3 :

On vérifie les hypothèses d'application du théorème 1.2.

On rappelle que dans toute la suite : $a_n = \frac{1}{n}$.

1) Tout d'abord, on montre que $\sum_{n=1}^{+\infty} a_n \|C_n - C\| < \infty$ p.s.

On commence par écrire la décomposition suivante :

$$\begin{aligned} C_n &= \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \Theta_i) Z_i' \\ &= \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i) Z_i' + \frac{1}{n-1} \sum_{i=1}^{n-1} (\theta_i - \Theta_i) Z_i' \\ &= \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i)(Z_i - \theta_i)' + \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i) \theta_i' + \frac{1}{n-1} \sum_{i=1}^{n-1} (\theta_i - \Theta_i) Z_i'. \end{aligned}$$

Ainsi,

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \Theta_i) Z_i' - C \right\| &\leq \sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i)(Z_i - \theta_i)' - C \right\| + \\ &\quad \sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i) \theta_i' \right\| + \sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (\theta_i - \Theta_i) Z_i' \right\|. \end{aligned}$$

a) Pour le premier terme du second membre, en passant à l'espérance mathématique, on obtient :

$$\begin{aligned} \mathbb{E} \left[\sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i)(Z_i - \theta_i)' - C \right\| \right] &= \sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\left\| \frac{1}{n-1} \sum_{i=1}^{n-1} \tilde{Z}_i \tilde{Z}_i' - \mathbb{E}[\tilde{Z} \tilde{Z}'] \right\| \right] \\ &= \sum_{n=1}^{+\infty} a_n O\left(\frac{1}{\sqrt{n}}\right) \text{ sous (H1)(c) d'après le lemme 3.11} \\ &< \infty \text{ sous (H3)(c) .} \end{aligned}$$

b) Pour le deuxième terme du second membre, on obtient de manière similaire :

$$\begin{aligned} \mathbb{E} \left[\sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (Z_i - \theta_i) \theta_i' \right\| \right] &\leq \sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\left\| \frac{1}{n-1} \sum_{i=1}^{n-1} H_i \right\| \right] \text{ avec } H_i = (Z_i - \theta_i) \theta_i' \\ &\leq \sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\left\| \bar{H}_{n-1} \right\| \right] \\ &\leq c_1 \sum_{n=1}^{+\infty} \frac{a_n}{\sqrt{n}} \text{ sous (H1)(c) et (H2) d'après le lemme 3.12} \\ &< \infty \text{ sous (H3)(c) .} \end{aligned}$$

c) pour le dernier terme du second membre, avec la même technique, on trouve que :

$$\begin{aligned}
\mathbb{E} \left[\sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (\theta_i - \Theta_i) Z'_i \right\| \right] &\leq \mathbb{E} \left[\sum_{n=1}^{+\infty} \frac{a_n}{n-1} \sum_{i=1}^{n-1} \|(\theta_i - \Theta_i) Z'_i\| \right] \\
&\leq \sum_{n=1}^{+\infty} \frac{a_n}{n-1} \sum_{i=1}^{n-1} \mathbb{E} [\|\theta_i - \Theta_i\| \cdot \|Z_i\|] \\
&\leq \sum_{n=1}^{+\infty} \frac{a_n}{n-1} \sum_{i=1}^{n-1} \left(\mathbb{E} [\|\theta_i - \Theta_i\|^2] \right)^{\frac{1}{2}} \left(\mathbb{E} [\|Z_i\|^2] \right)^{\frac{1}{2}} \\
&\leq \sum_{n=1}^{+\infty} \frac{a_n}{n-1} \sum_{i=1}^{n-1} \left(\mathbb{E} [\|\theta_i - \Theta_i\|^2] \right)^{\frac{1}{2}} \left(\mathbb{E} [\|\tilde{Z}\|^2] + \|\theta_i\|^2 \right)^{\frac{1}{2}} \\
&\leq c_2 \sum_{n=1}^{+\infty} \frac{a_n}{n-1} \sum_{i=1}^{n-1} \left(\mathbb{E} [\|\theta_i - \Theta_i\|^2] \right)^{\frac{1}{2}} \text{ sous (H1)(c) et (H2) .}
\end{aligned}$$

Sous (H5)(e), on peut appliquer le lemme 3.13 avec $u_n = \left(\mathbb{E} [\|\theta_n - \Theta_n\|^2] \right)^{\frac{1}{2}}$.

Ainsi, d'après les conclusions de ce lemme, $\sum_{n=1}^{+\infty} \frac{a_n}{n-1} \sum_{i=1}^{n-1} \left(\mathbb{E} [\|\theta_i - \Theta_i\|^2] \right)^{\frac{1}{2}} < \infty$. Par suite, $\sum_{n=1}^{+\infty} \frac{a_n}{n} \sum_{i=1}^{n-1} \left(\mathbb{E} [\|\theta_i - \Theta_i\|^2] \right)^{\frac{1}{2}} < \infty$ puis $\mathbb{E} \left[\sum_{n=1}^{+\infty} a_n \left\| \frac{1}{n-1} \sum_{i=1}^{n-1} (\theta_i - \Theta_i) Z'_i \right\| \right] < \infty$.

On obtient que $\sum_{n=1}^{+\infty} a_n \|C_n - C\| < \infty$ p.s. en regroupant a), b) et c).

2) Sous (H4), $M_n \rightarrow M$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

3) On montre finalement que $\sum_{n=1}^{+\infty} a_n \|B_n - B\| < \infty$ p.s.

$$\begin{aligned}
\sum_{n=1}^{+\infty} a_n \|B_n - B\| &= \sum_{n=1}^{+\infty} a_n \|C_n M_n - C M\| \\
&\leq \sum_{n=1}^{+\infty} a_n \|C_n\| \cdot \|M_n - M\| + \sum_{n=1}^{+\infty} a_n \|C_n - C\| \cdot \|M\| \\
&= \sum_{n=1}^{+\infty} a_n \|(C_n - C) + C\| \cdot \|M_n - M\| + \|M\| \sum_{n=1}^{+\infty} a_n \|C_n - C\| \\
&\leq \sum_{n=1}^{+\infty} a_n \|C_n - C\| \cdot \|M_n - M\| + \|C\| \sum_{n=1}^{+\infty} a_n \|M_n - M\| + \|M\| \sum_{n=1}^{+\infty} a_n \|C_n - C\|.
\end{aligned}$$

Comme $M_n \rightarrow M$ p.s., il existe une v.a. F positive et finie p.s. telle que :

$$\begin{aligned}
\sum_{n=1}^{+\infty} a_n \|B_n - B\| &\leq F\left(\sum_{n=1}^{+\infty} a_n \|C_n - C\| + \sum_{n=1}^{+\infty} a_n \|M_n - M\|\right) \\
&< \infty \text{ d'après les conclusions des deux premières parties de la démonstration.}
\end{aligned}$$

Démonstration du théorème 3.4 :

On vérifie les hypothèses concernant Θ_n des théorèmes ad hoc :

Sous (H6)(a),(b) et pour $a_n = \frac{a}{n^\alpha}$, $a > 0$, $\frac{1}{2} < \alpha \leq 1$, les hypothèses d'application des théorèmes 1.6 et 1.7 sont vérifiées. D'après les conclusions de ces théorèmes, (H5)(b) est vérifiée et, à partir d'un certain rang, pour $i = 1, \dots, p$, on a :

- pour $(\frac{1}{2} < \alpha < 1)$ ou $(\alpha = 1, \frac{2\lambda_i a}{r_i} > 1)$, $\mathbb{E}[||\Theta_n^i - \theta_n^i||^2] = O(\frac{1}{n^\alpha})$,
- pour $(\alpha = 1, \frac{2\lambda_i a}{r_i} = 1)$, $\mathbb{E}[||\Theta_n^i - \theta_n^i||^2] = O(\frac{\ln(n)}{n})$,
- pour $(\alpha = 1, \frac{2\lambda_i a}{r_i} < 1)$, $\mathbb{E}[||\Theta_n^i - \theta_n^i||^2] = O(\frac{1}{n^{\frac{2\lambda_i a}{r_i}}})$, $\frac{2\lambda_i a}{r_i} > 0$.

On remarque que, pour $\alpha = 1$, dans les trois cas :

$$\exists \delta > 0, \text{ pour } i = 1, \dots, p, \mathbb{E}[||\Theta_n^i - \theta_n^i||^2] = O(\frac{1}{n^\delta}).$$

1) On vérifie l'hypothèse (H5)(c) du théorème 3.1 : $\sum_{n=1}^{+\infty} a_n \sum_{j \in I_n} \mathbb{E}[||\Theta_j - \theta_j||] < \infty$.

- pour $\frac{2}{3} < \alpha < 1$, on a : $\mathbb{E}[||\Theta_n^i - \theta_n^i||^2] = O(\frac{1}{n^\alpha})$, qui implique : $\mathbb{E}[||\Theta_n - \theta_n||^2] = O(\frac{1}{n^\alpha})$.

$$\text{Donc, } \sum_{j \in I_n} \mathbb{E}[||\Theta_j - \theta_j||] \leq c \sum_{j \in I_n} \frac{1}{j^{\frac{\alpha}{2}}} \leq c \sum_{j \in I_n} \frac{1}{(R_{n-1}+1)^{\frac{\alpha}{2}}} \leq \frac{Rc}{(R_{n-1}+1)^{\frac{\alpha}{2}}} \leq \frac{Rc}{n^{\frac{\alpha}{2}}}.$$

$$\text{Ainsi, } \sum_{n=1}^{+\infty} a_n \sum_{j \in I_n} \mathbb{E}[||\Theta_j - \theta_j||] \leq Rc \sum_{n=1}^{+\infty} \frac{1}{n^{\frac{3\alpha}{2}}} < \infty.$$

- pour $\alpha = 1$, on a dans les trois cas :

$$\exists \delta > 0, \text{ pour } i = 1, \dots, p, \mathbb{E}[||\Theta_n^i - \theta_n^i||^2] = O(\frac{1}{n^\delta}), \text{ ce qui implique : } \exists \delta > 0 : \text{ pour } i = 1, \dots, p, \mathbb{E}[||\Theta_n - \theta_n||^2] = O(\frac{1}{n^\delta}).$$

$$\text{Donc, } \sum_{j \in I_n} \mathbb{E}[||\Theta_j - \theta_j||] \leq c \sum_{j \in I_n} \frac{1}{j^{\frac{\delta}{2}}} \leq c \sum_{j \in I_n} \frac{1}{(R_{n-1}+1)^{\frac{\delta}{2}}} \leq \frac{Rc}{(R_{n-1}+1)^{\frac{\delta}{2}}} \leq \frac{Rc}{n^{\frac{\delta}{2}}}.$$

$$\text{Ainsi, } \sum_{n=1}^{+\infty} a_n \sum_{j \in I_n} \mathbb{E}[||\Theta_j - \theta_j||] \leq Rc \sum_{n=1}^{+\infty} \frac{1}{n^{1+\frac{\delta}{2}}} < \infty.$$

Dans tous les cas, l'hypothèse (H5)(c) du théorème 3.1 est vérifiée.

2) On vérifie l'hypothèse (H5)(d) du théorème 3.2 : $\mathbb{E} \left[\|\boldsymbol{\theta}_n - \boldsymbol{\Theta}_n\| \right] = O\left(\frac{1}{n^\gamma}\right), \gamma > 2(1 - \alpha).$

· pour $\frac{4}{5} < \alpha < 1$, on a : $\mathbb{E}[\|\boldsymbol{\Theta}_n - \boldsymbol{\theta}_n\|^2] = O\left(\frac{1}{n^\alpha}\right).$

Donc, $\mathbb{E}[\|\boldsymbol{\Theta}_n - \boldsymbol{\theta}_n\|] = O\left(\frac{1}{n^{\frac{\alpha}{2}}}\right) = O\left(\frac{1}{n^\gamma}\right)$ avec $\gamma = \frac{\alpha}{2}.$

Comme $\frac{4}{5} < \alpha, \gamma = \frac{\alpha}{2} > 2(1 - \alpha).$

· pour $\alpha = 1$, on a dans les trois cas :

$\exists \delta > 0$, pour $i = 1, \dots, p$, $\mathbb{E}[\|\boldsymbol{\Theta}_n - \boldsymbol{\theta}_n\|^2] = O\left(\frac{1}{n^\delta}\right).$

Donc, $\mathbb{E}[\|\boldsymbol{\Theta}_n - \boldsymbol{\theta}_n\|] = O\left(\frac{1}{n^{\frac{\delta}{2}}}\right) = O\left(\frac{1}{n^\gamma}\right)$ avec $\gamma = \frac{\delta}{2}.$

On a : $\gamma = \frac{\delta}{2} > 0 = 2(1 - \alpha).$

Dans tous les cas, l'hypothèse (H5)(d) du théorème 3.2 est vérifiée.

3) On vérifie l'hypothèse (H5)(e) du théorème 3.3 : $\sum_{n=1}^{+\infty} \frac{1}{n} \left(\mathbb{E}[\|\boldsymbol{\theta}_n - \boldsymbol{\Theta}_n\|^2] \right)^{\frac{1}{2}} < \infty.$

· pour $\alpha = 1$, on a dans les trois cas :

$\exists \delta > 0$, pour $i = 1, \dots, p$, $\mathbb{E}[\|\boldsymbol{\Theta}_n - \boldsymbol{\theta}_n\|^2] = O\left(\frac{1}{n^\delta}\right).$

Donc $\sum_{n=1}^{+\infty} \frac{1}{n} \left(\mathbb{E}[\|\boldsymbol{\Theta}_n - \boldsymbol{\theta}_n\|^2] \right)^{\frac{1}{2}} \leq \sum_{n=1}^{\infty} O\left(\frac{1}{n^{1+\frac{\delta}{2}}}\right) < \infty.$

Dans tous les cas, l'hypothèse (H5)(e) du théorème 3.3 est vérifiée.

Démonstration du théorème 3.5 :

On rappelle que les (k) sont remplacés par des k pour simplifier les notations.

Sous (H5)(a),(b) et (d), on a, pour tout $k = 1, \dots, q$:

$$\Theta_n^k \text{ } T_n\text{-mesurable ; } \|\Theta_n^k - \theta_n^k\| \rightarrow 0 \text{ p.s. ; } \mathbb{E}[\|\Theta_n^k - \theta_n^k\|] = O\left(\frac{1}{n^\gamma}\right), \gamma > 2(1 - \alpha).$$

1) Montrons d'abord que $M_n \rightarrow M$ p.s.

Pour ce faire, on montre que pour $k = 1, \dots, q$, $M_n^k \rightarrow (C^k)^{-1}$ p.s.

Pour simplifier, on supprime l'écriture des k .

Sous (H1)(a), on peut écrire :

$$\begin{aligned} M_{n+1} - C^{-1} &= M_n - C^{-1} - a_n \left(\left(\frac{1}{r_n} \sum_{i \in I_n} (Z_i - \Theta_i) Z_i' \right) M_n - I \right) \\ &= (M_n - C^{-1}) - a_n \left(C(M_n - C^{-1}) + \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n \right) \\ &\quad \text{avec } B_i = (Z_i - \theta_i) Z_i' \\ &= (I - a_n C)(M_n - C^{-1}) - a_n \left(\left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n \right). \end{aligned}$$

D'où,

$$\begin{aligned} \|M_{n+1} - C^{-1}\|^2 &= \|(I - a_n C)(M_n - C^{-1})\|^2 + a_n^2 \left\| \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n \right\|^2 - \\ &\quad 2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n \right\rangle. \end{aligned}$$

puis, en prenant l'espérance mathématique conditionnellement à T_n , tribu du passé au temps n , on obtient :

$$\begin{aligned} \mathbb{E}[\|M_{n+1} - C^{-1}\|^2 | T_n] &= \|(I - a_n C)(M_n - C^{-1})\|^2 + a_n^2 \mathbb{E} \left[\left\| \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n \right\|^2 | T_n \right] \\ &\quad - 2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \mathbb{E} \left[\left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n | T_n \right] \right\rangle. \end{aligned}$$

a) Pour le premier terme du second membre,

$$\begin{aligned} \|(I - a_n C)(M_n - C^{-1})\|^2 &= \|M_n - C^{-1}\|^2 - 2a_n \langle M_n - C^{-1}, C(M_n - C^{-1}) \rangle + a_n^2 \|C(M_n - C^{-1})\|^2 \\ &\leq (1 + a_n^2 \|C\|^2) \|M_n - C^{-1}\|^2 - 2a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2. \end{aligned}$$

b) Pour le troisième et dernier terme du second membre,

$$\begin{aligned}
& -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \mathbb{E} \left[\left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z'_i \right) M_n \middle| T_n \right] \right\rangle \\
&= -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(\frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}[B_i | T_n] - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) \mathbb{E}[Z'_i | T_n] \right) M_n \right\rangle \\
&\quad \text{puisque } M_n \text{ et les } \Theta_i, \text{ sont } T_n - \text{mesurables} \\
&= -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(\frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}[B_i] - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) \mathbb{E}[Z'_i] \right) M_n \right\rangle \\
&\quad \text{puisque les } B_i \text{ et les } Z_i, i \in I_n, \text{ sont indépendants de } T_n \\
&= -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) \theta'_i \right) M_n \right\rangle \text{ puisque } \mathbb{E}[B_i] = C \\
&\leq da_n \left\| \frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) \theta'_i \right\| \cdot \|M_n - C^{-1}\| \cdot \|M_n\| \\
&\leq d' a_n \left(\frac{1}{r_n} \sum_{i \in I_n} \|\theta_i - \Theta_i\| \cdot \|\theta_i\| \right) (\|M_n - C^{-1}\|^2 + 1) \\
&\leq ea_n \|M_n - C^{-1}\|^2 \frac{1}{r_n} \sum_{i \in I_n} \|\theta_i - \Theta_i\| + ea_n \frac{1}{r_n} \sum_{i \in I_n} \|\theta_i - \Theta_i\| \text{ sous (H2).}
\end{aligned}$$

c) Enfin, pour le deuxième terme du second membre,

$$\begin{aligned}
& a_n^2 \mathbb{E} \left[\left\| \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z'_i \right) M_n \right\|^2 \middle| T_n \right] \\
&\leq a_n^2 \mathbb{E} \left[\left\| \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C + \frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z'_i \right) \right\|^2 \middle| T_n \right] \cdot \|M_n\|^2 \\
&\leq 3a_n^2 \mathbb{E} \left[\left\| \frac{1}{r_n} \sum_{i \in I_n} B_i \right\|^2 + \|C\|^2 + \left\| \frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z'_i \right\|^2 \middle| T_n \right] \cdot \|M_n\|^2 \\
&\leq 3a_n^2 \left(\frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}[\|B_i\|^2 | T_n] + \|C\|^2 + \frac{1}{r_n} \sum_{i \in I_n} \|\theta_i - \Theta_i\|^2 \mathbb{E}[\|Z_i\|^2 | T_n] \right) \cdot \|M_n\|^2 \\
&\leq 3a_n^2 \left(\frac{1}{r_n} \sum_{i \in I_n} \mathbb{E}[\|B_i\|^2] + \|C\|^2 + \frac{1}{r_n} \sum_{i \in I_n} \|\theta_i - \Theta_i\|^2 \mathbb{E}[\|Z_i\|^2] \right) \cdot \|M_n\|^2.
\end{aligned}$$

Or,

$$\begin{aligned}
\mathbb{E}[||Z_i||^2] &= \mathbb{E}[||\theta_i + \tilde{Z}_i||^2] \\
&\leq 2\left(||\theta_i||^2 + \mathbb{E}[||\tilde{Z}_i||^2]\right) \\
&\leq 2\left(\sup_i ||\theta_i||^2 + \mathbb{E}[||\tilde{Z}||^2]\right) \\
&< \infty \text{ sous (H1)(c) et (H2)}
\end{aligned}$$

et

$$\begin{aligned}
\mathbb{E}[||B_i||^2] &= \mathbb{E}[||(Z_i - \theta_i)Z_i'|^2] \\
&= \mathbb{E}[|(Z_i - \theta_i)(Z_i - \theta_i)' + (Z_i - \theta_i)\theta_i'|^2] \\
&\leq 2\left(\mathbb{E}[||\tilde{Z}||^4] + \mathbb{E}[||\tilde{Z}||^2] \sup_i ||\theta_i||^2\right) \\
&< \infty \text{ sous (H1)(c) et (H2)}.
\end{aligned}$$

Alors,

$$\begin{aligned}
&a_n^2 \mathbb{E}\left[\left| \left(\frac{1}{r_n} \sum_{i \in I_n} B_i - C \right) M_n + \left(\frac{1}{r_n} \sum_{i \in I_n} (\theta_i - \Theta_i) Z_i' \right) M_n \right|^2 \middle| T_n \right] \\
&\leq c_2 a_n^2 \left(1 + \frac{1}{r_n} \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 \right) ||M_n||^2 \\
&\leq f a_n^2 \left(1 + \frac{1}{r_n} \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 \right) ||M_n - C^{-1}||^2 + f a_n^2 \left(1 + \frac{1}{r_n} \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 \right).
\end{aligned}$$

De a), b) et c), on d duit que :

$$\begin{aligned}
\mathbb{E}\left[||M_{n+1} - C^{-1}||^2 \middle| T_n \right] &\leq \left(1 + a_n^2 ||C||^2 \right) ||M_n - C^{-1}||^2 - 2a_n \lambda_{\min}(C) ||M_n - C^{-1}||^2 + \\
&\quad e a_n ||M_n - C^{-1}||^2 \sum_{i \in I_n} ||\theta_i - \Theta_i|| + e a_n \sum_{i \in I_n} ||\theta_i - \Theta_i|| + \\
&\quad f a_n^2 \left(1 + \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 \right) ||M_n - C^{-1}||^2 + f a_n^2 \left(1 + \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 \right) \\
&\leq \left(1 + a_n^2 ||C||^2 + e a_n \sum_{i \in I_n} ||\theta_i - \Theta_i|| + f a_n^2 + f a_n^2 \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 \right) ||M_n - C^{-1}||^2 + \\
&\quad e a_n \sum_{i \in I_n} ||\theta_i - \Theta_i|| + f a_n^2 + f a_n^2 \sum_{i \in I_n} ||\theta_i - \Theta_i||^2 - 2a_n \lambda_{\min}(C) ||M_n - C^{-1}||^2.
\end{aligned}$$

Or,

$$\begin{aligned}
\mathbb{E} \left[\sum_{n=1}^{+\infty} a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| \right] &= \sum_{n=1}^{+\infty} a_n \sum_{i \in I_n} \mathbb{E} \left[\|\theta_i - \Theta_i\| \right] \\
&\leq \sum_{n=1}^{+\infty} a_n R \max_{i \in I_n} \mathbb{E} \left[\|\theta_i - \Theta_i\| \right] \\
&\leq \sum_{n=1}^{+\infty} a_n R O\left(\frac{1}{n^\gamma}\right) \text{ sous (H5)(d)} \\
&\leq R' \sum_{n=1}^{+\infty} \frac{a_n}{n^\gamma} \\
&< \infty \text{ sous (H3)(d) et (H5)(d).}
\end{aligned}$$

On en déduit que $\sum_{n=1}^{+\infty} a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| < \infty$ a.s.

Comme $\sum_{i \in I_n} \|\theta_i - \Theta_i\|^2 \leq \left(\sum_{i \in I_n} \|\theta_i - \Theta_i\| \right)^2$, on a : $\sum_{n=1}^{+\infty} a_n^2 \sum_{i \in I_n} \|\theta_i - \Theta_i\|^2 < \infty$ a.s. ; on peut alors appliquer le lemme de Robbins-Siegmund :

Il existe une v.a. $T \geq 0$ telle que $\|M_n - C^{-1}\| \rightarrow T$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\|^2 < \infty$ p.s.

Comme sous (H3)(d) $\sum_{n=1}^{+\infty} a_n = \infty$, $T = 0$ et, en réécrivant les k , $M_n^k \rightarrow (C^k)^{-1}$ p.s.

Ceci étant vrai pour tout $k = 1, \dots, q$ et puisque M_n a pour k-ième bloc diagonal M_n^k , $\|M_n - M\| \rightarrow 0$ p.s.

2) Montrons maintenant que $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

Pour ce faire, on montre que pour $k = 1, \dots, q$, $\sum_{n=1}^{+\infty} a_n \|M_n^k - (C^k)^{-1}\| < \infty$ p.s.

Dans la suite, on supprime à nouveau l'écriture des k .

Tout d'abord, remarquons que sous (H5)(b), $\max_{i \in I_n} \|\theta_i - \Theta_i\| \rightarrow 0$ p.s.

Ainsi, $\forall \delta > 0, \forall \varepsilon > 0, \exists N(\delta, \varepsilon), \forall n \geq N(\delta, \varepsilon), \mathbb{P}(\sup_{j > n} (\max_{i \in I_j} \|\theta_i - \Theta_i\|) \leq \delta) > 1 - \varepsilon$.

On fixe alors δ et ε on note $N_0 = N(\delta, \varepsilon)$.

On pose :

$$\begin{aligned} F &= \left\{ \sup_{j > N_0} (\max_{i \in I_j} \|\theta_i - \Theta_i\|) \leq \delta \right\}, \\ F_n &= \left\{ \sup_{N_0 < j \leq n} (\max_{i \in I_j} \|\theta_i - \Theta_i\|) \leq \delta \right\}. \end{aligned}$$

On peut déjà remarquer que $\mathbb{P}(F) > 1 - \varepsilon$ et que F_n est T_n -mesurable.

De plus, on a $F \Rightarrow F_{n+1} \Rightarrow F_n$ i.e. $F \subset F_{n+1} \subset F_n$ ou encore $\mathbb{1}_F \leq \mathbb{1}_{F_{n+1}} \leq \mathbb{1}_{F_n} \leq 1$.

Donc, pour $n > N_0$:

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n \right] &\leq \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_n} | T_n \right] \\ &\leq \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 | T_n \right] \mathbb{1}_{F_n}. \end{aligned}$$

De la majoration de $\mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 | T_n \right]$ établie précédemment, on déduit que :

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n \right] &\leq \left(1 + a_n^2 \|C\|^2 + f a_n^2 + f a_n^2 \sum_{i \in I_n} \|\theta_i - \Theta_i\|^2 \right) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} + \\ &\quad e a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| + e a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| \cdot \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} + \\ &\quad f a_n^2 + f a_n^2 \sum_{i \in I_n} \|\theta_i - \Theta_i\|^2 \mathbb{1}_{F_n} - 2 a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}. \end{aligned}$$

Or, dans F_n , $\max_{i \in I_n} \|\theta_i - \Theta_i\| \leq \delta$, donc :

$$\begin{aligned} e a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| \cdot \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} &\leq e a_n r \max_{i \in I_n} \|\theta_i - \Theta_i\| \cdot \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \\ &\leq e a_n r \delta \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}. \end{aligned}$$

On choisit δ tel que $e r \delta < \lambda_{\min}(C)$.

On peut alors écrire,

$$ea_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| \cdot \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \leq a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n},$$

et par suite,

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n \right] &\leq \left(1 + a_n^2 \|C\|^2 + f(1 + r\delta^2) a_n^2 \right) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} + ea_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| + \\ &\quad f(1 + r\delta^2) a_n^2 - a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}. \end{aligned}$$

En passant à l'espérance mathématique, on obtient :

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} \right] &\leq \left(1 + a_n^2 \|C\|^2 + f(1 + r\delta^2) a_n^2 \right) \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] + ea_n \sum_{i \in I_n} \mathbb{E} \left[\|\theta_i - \Theta_i\| \right] + \\ &\quad f(1 + r\delta^2) a_n^2 - a_n \lambda_{\min}(C) \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right]. \end{aligned}$$

Sous (H5)(d) et (H3)(d), on applique de nouveau le lemme de Robbins-Siegmund et on déduit que :

$$\text{Il existe un réel } t \geq 0 \text{ tel que } \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] \rightarrow t \text{ et } \sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] < \infty.$$

Comme $\sum_{n=1}^{+\infty} a_n = \infty, t = 0$.

Ainsi, il existe $\mu > 0$ tel que :

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} \right] &\leq \left(1 - a_n \lambda_{\min}(C) \right) \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] + a_n g_n \\ &\quad \text{avec } g_n = O \left(a_n + \max_{i \in I_n} \mathbb{E} \left[\|\theta_i - \Theta_i\| \right] \right) = O \left(\frac{1}{n^{\min(\alpha, \gamma)}} \right) \\ &\leq \left(1 - a_n \lambda_{\min}(C) \right) \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] + \mu a_n h_n \text{ avec } h_n = \frac{1}{n^{\min(\alpha, \gamma)}}. \end{aligned}$$

Sous les hypothèses (H3)(d) et (H5)(d), d'après le lemme 3.10, on a :

$$\overline{\lim} \frac{1}{h_n} \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] < \infty.$$

$$\text{Alors, } \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] = O(h_n) \text{ d'où } \mathbb{E} \left[\|M_n - C^{-1}\| \mathbb{1}_{F_n} \right] = O(\sqrt{h_n}).$$

On en déduit que sous (H3)(d) et (H5)(d), $\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\|M_n - C^{-1}\| \mathbb{1}_{F_n} \right] < \infty$ puis $\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| \mathbb{1}_{F_n} < \infty$ p.s.

Comme $\mathbb{1}_F \leq \mathbb{1}_{F_n}$, on peut écrire que $\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| \mathbb{1}_F < \infty$ p.s.

Ainsi, d'un point de vue logique,

$$\{F\} \Rightarrow \left\{ \sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| < \infty \right\}$$

ce qui implique, d'un point de vue probabiliste, que :

$$\mathbb{P}\left(\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| < \infty\right) \geq \mathbb{P}(F) > 1 - \varepsilon$$

selon la remarque du début de partie.

ε étant quelconque, on en déduit finalement, en réécrivant les k , que pour $k = 1, \dots, q$, $\sum_{n=1}^{+\infty} a_n \|M_n^k - (C^k)^{-1}\| < \infty$ p.s.

L'ensemble des valeurs propres de la matrice diagonale par blocs $(M_n - M)$ étant la réunion des ensembles des valeurs propres des matrices $(M_n^k - (C^k)^{-1})$, $k = 1, \dots, q$, on a : $\|M_n - M\| = \max_k \|M_n^k - (C^k)^{-1}\| \leq \sum_{k=1}^q \|M_n^k - (C^k)^{-1}\|$.

On en déduit que $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

Démonstration du théorème 3.6 (resp. 3.7) :

Tout d'abord, sous (H1)(a) et pour $k = 1, \dots, q$, on peut écrire :

$$\begin{aligned}
 M_{n+1}^k - (C^k)^{-1} &= M_n^k - (C^k)^{-1} - a_n (C_n^k M_n^k - I) \\
 &= (I - a_n C^k) (M_n^k - (C^k)^{-1}) + a_n C^k M_n^k - a_n C_n^k M_n^k \\
 &= (I - a_n C^k) (M_n^k - (C^k)^{-1}) - a_n (C_n^k - C^k) (M_n^k - (C^k)^{-1}) - a_n (C_n^k - C^k) (C^k)^{-1}.
 \end{aligned}$$

D'où,

$$\begin{aligned}
 \|M_{n+1}^k - (C^k)^{-1}\| &\leq \|I - a_n C^k\| \cdot \|M_n^k - (C^k)^{-1}\| + a_n \|C_n^k - C^k\| \cdot \|M_n^k - (C^k)^{-1}\| + a_n \|C_n^k - C^k\| \cdot \|(C^k)^{-1}\| \\
 &\leq (1 - \lambda_{\min}(C^k) a_n) \|M_n^k - (C^k)^{-1}\| + a_n \|C_n^k - C^k\| \cdot \|M_n^k - (C^k)^{-1}\| + \\
 &\quad a_n \|C_n^k - C^k\| \cdot \|(C^k)^{-1}\| \\
 &\leq (1 + a_n \|C_n^k - C^k\|) \|M_n^k - (C^k)^{-1}\| + a_n \|C_n^k - C^k\| \cdot \|(C^k)^{-1}\| - \\
 &\quad a_n \lambda_{\min}(C^k) \cdot \|M_n^k - (C^k)^{-1}\|.
 \end{aligned}$$

Ensuite, en reprenant la partie 1 de la démonstration du théorème 3.2 (resp. 3.3), on montre que :

$$\sum_{n=1}^{+\infty} a_n \|C_n^k - C^k\| < \infty \text{ p.s.}$$

Dès lors, on peut appliquer le lemme de Robbins-Siegmund :

Il existe une v.a. $T \geq 0$ telle que $\|M_n^k - (C^k)^{-1}\| \rightarrow T$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n^k - (C^k)^{-1}\| < \infty$ p.s.

Comme $\sum_{n=1}^{+\infty} a_n = \infty$, $T = 0$ et $M_n^k \rightarrow (C^k)^{-1}$ p.s.

Ceci étant vrai pour tout $k = 1, \dots, q$ et puisque M_n a pour k-ième bloc diagonal M_n^k , $\|M_n - M\| \rightarrow 0$ p.s.

En outre, l'ensemble des valeurs propres de la matrice diagonale par blocs $(M_n - M)$ étant la réunion des ensembles des valeurs propres des matrices $(M_n^k - (C^k)^{-1})$, $k = 1, \dots, q$, on a : $\|M_n - M\| = \max_k \|M_n^k - (C^k)^{-1}\|$

$$\leq \sum_{k=1}^q \|M_n^k - (C^k)^{-1}\|.$$

On en déduit que $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

3.3.3 Hypothèses

- (H1) (a) Il n'existe pas de relation affine entre les composantes de $\tilde{Z}$,
 (H1) (b) $\tilde{Z}$ admet des moments d'ordre $4r$,
 (H1) (c) $\tilde{Z}$ admet des moments d'ordre 4,

$$(H2) \sup_{l \in \mathbb{N}} \|\theta_l\| < \infty,$$

$$(H3) (a) a_n > 0, \sum_{n=1}^{+\infty} a_n = \infty, \sum_{n=1}^{+\infty} a_n^2 < \infty,$$

$$(H3) (b) (a_n = \frac{a}{n^\alpha} \text{ avec } a > 0, \frac{2}{3} < \alpha < 1) \text{ ou } (\alpha = 1 \text{ i.e. } a_n = \frac{a}{n} \text{ et } a > 1),$$

$$(H3) (c) a_n = \frac{1}{n},$$

$$(H3) (d) (a_n = \frac{a}{n^\alpha} \text{ avec } a > 0, \frac{2}{3} < \alpha < 1) \text{ ou } (\alpha = 1 \text{ i.e. } a_n = \frac{a}{n} \text{ et } a > \frac{1}{\min_k \lambda_{\min}(C^k)}),$$

$$(H4) (a) M_n \text{ est } T_n\text{-mesurable},$$

$$(H4) (b) M_n \rightarrow M \text{ p.s.},$$

$$(H4) (c) \sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty \text{ p.s.},$$

$$(H5) (a) \Theta_n \text{ est } T_n\text{-mesurable},$$

$$(H5) (b) \|\Theta_n - \theta_n\| \rightarrow 0 \text{ p.s.},$$

$$(H5) (c) \sum_{n=1}^{+\infty} a_n \sum_{i \in I_n} \|\theta_i - \Theta_i\| < \infty \text{ p.s.},$$

$$(H5) (d) \mathbb{E} \left[\|\theta_n - \Theta_n\| \right] = O\left(\frac{1}{n^\gamma}\right), \gamma > 2(1 - \alpha),$$

$$(H5) (e) \sum_{n=1}^{+\infty} \frac{1}{n} \left(\mathbb{E} [\|\theta_n - \Theta_n\|^2] \right)^{\frac{1}{2}} < \infty,$$

$$(H6) (a) \max_i \sup_n \|U_n^i\| < \infty,$$

$$(H6) (b) \text{ Pour } i = 1, \dots, p, \text{ il existe un entier } r_i, \text{ un réel } \lambda_i > 0, \text{ une suite croissante d'entiers } (n_{il}, l \geq 1) \text{ tels que } n_{i1} = 1, n_{i,l+1} \leq n_{il} + r_i, \lambda_{\min} \left(\sum_{j \in J_{il}} U_j^i (U_j^i)^\top \right) \geq \lambda_i, \text{ avec } J_{il} = \{n_{il}, \dots, n_{i,l+1} - 1\}.$$

Chapitre 4

ACP partielle d'un flux de données d'espérance mathématique et de matrice de covariance variables dans le temps

Sommaire

4.1	Cas général de l'ACP partielle	107
4.1.1	Cas général	108
4.1.2	Cas particulier d'un modèle linéaire de variation de l'espérance et de la variance .	109
4.2	Cas particulier : l'ACG partielle	111
4.3	Annexes du chapitre 4	114
4.3.1	Démonstrations des théorèmes	114
4.3.2	Hypothèses	147

4.1 Cas général de l'ACP partielle

On observe p caractères quantitatifs sur des individus : on obtient des vecteurs de données z_i dans $\mathbb{R}^p$.

On se place ici dans le cas où ces vecteurs arrivent séquentiellement dans le temps (données en ligne) : on observe z_n au temps n ; on a une suite de vecteurs de données $z_1, \dots, z_n, \dots$.

On suppose que, pour tout $n \geq 1$, z_n est la réalisation d'un vecteur aléatoire (v.a.) Z_n dans $\mathbb{R}^p$, défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, d'espérance mathématique et de matrice de covariance variables dans le temps, les v.a. Z_n étant mutuellement indépendants.

On suppose que, pour tout n , on a la décomposition :

$$Z_n = \theta_n + \Delta_n R_n \quad \text{où :}$$

– $\theta_n = (\theta_n^1 \dots \theta_n^p)'$ est un vecteur de $\mathbb{R}^p$,

– $\Delta_n = \begin{pmatrix} \delta_n^1 & & \\ & \ddots & \\ & & \delta_n^p \end{pmatrix}$ est une matrice diagonale d'ordre p d'éléments non nuls.

– la loi du v.a. R_n ne dépend pas de n , $\mathbb{E}[R_n] = 0$, $Cov(R_n) = C$.

Ceci revient à supposer que les $R_n = \Delta_n^{-1}(Z_n - \theta_n)$ constituent un échantillon indépendant et identiquement distribué (i.i.d.) d'un v.a. R dans $\mathbb{R}^p$ tel que $\mathbb{E}[R] = 0$, $Cov(R) = C$.

On a alors :

$$\mathbb{E}[Z_n] = \theta_n \text{ et } Cov(Z_n) = \Delta_n C \Delta_n.$$

Pour $i = 1, \dots, p$, 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 :

$$Z_n^i = \theta_n^i + \delta_n^i R_n^i, \quad i = 1, \dots, p$$

avec $\mathbb{E}[R_n^i] = \mathbb{E}[R^i] = 0$, les R_n^i constituant un échantillon i.i.d. de R^i .

Dans la suite, on suppose sans perte de généralité que $Var(R_n^i) = Var(R^i) = 1$.

On pose le problème suivant : réaliser une analyse en composantes principales (ACP) du v.a. R dans $\mathbb{R}^p$ que l'on munit d'une métrique M . On étudie en particulier l'estimation de vecteurs directeurs des r premiers axes principaux de cette analyse.

On est ainsi amené à chercher les r premiers vecteurs propres $v_1, \dots, v_r$ de la matrice M -symétrique $B = CM$ associés aux r plus grandes valeurs propres (v.p.) $\lambda_1, \dots, \lambda_r$.

Comme $C = Cov(R) = Cov(R_n) = Cov(\Delta_n^{-1}(Z_n - \Theta_n)) = \Delta_n^{-1} \mathbb{E}[(Z_n - \Theta_n)Z_n'] \Delta_n^{-1}$, on considère les données $\Delta_n^{-1}(Z_n - \Theta_n)$, sur lesquelles on effectue une ACP avec la métrique M .

Soit M_n, D_n^{-1} et Θ_n trois estimateurs convergents respectivement de M, Δ_n^{-1} et Θ_n .

On définit alors $B_n = C_n M_n = D_n^{-1}(Z_n - \Theta_n)Z_n' D_n^{-1} M_n$ puis récursivement un processus d'approximation stochastique (a.s.) $(X_n) = (X_n^1, \dots, X_n^r)$ de $(v_1, \dots, v_r)$ par :

$$\begin{aligned} F_n(X_n^l) &= \frac{\langle B_n X_n^l, X_n^l \rangle_{M_n}}{\|X_n^l\|_{M_n}^2}, \\ Y_{n+1}^l &= X_n^l + \frac{a}{n^\alpha} (B_n - F_n(X_n^l)I) X_n^l, \quad l = 1, \dots, r, \\ X_{n+1} &= \text{orth}_{M_n}(Y_{n+1}). \end{aligned}$$

Pour obtenir X_{n+1} , on effectue une orthogonalisation au sens de Gram-Schmidt par rapport à M_n de $Y_{n+1} = (Y_{n+1}^1, \dots, Y_{n+1}^r)$.

4.1.1 Cas général

Dans tout ce paragraphe, on ne définit pas de modèle a priori pour Θ_n et pour Δ_n .

Théorème 4.1.

Sous les hypothèses :

(H1) (b) R admet des moments d'ordre $4r$,

$$(H2) \sup_{l \in \mathbb{N}} \|\theta_l\| < \infty,$$

$$(H2') (a) \sup_{l \in \mathbb{N}} \max_i |\delta_l^i| < \infty, \\ (b) \inf_{l \in \mathbb{N}} \min_i |\delta_l^i| > 0,$$

$$(H3) (a) a_n > 0, \quad \sum_{n=1}^{+\infty} a_n = \infty, \quad \sum_{n=1}^{+\infty} a_n^2 < \infty,$$

$$(H4) (a) M_n \text{ est } T_n\text{-mesurable}, \\ (b) M_n \rightarrow M \text{ presque sûrement (p.s.)}, \\ (c) \sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty \text{ p.s.},$$

$$(H5) (a) \Theta_n \text{ est } T_n\text{-mesurable}, \\ (b) \|\theta_n - \Theta_n\| \rightarrow 0 \text{ p.s.}, \\ (c) \sum_{n=1}^{+\infty} a_n \|\theta_n - \Theta_n\| < \infty \text{ p.s.},$$

$$(H5') (a) D_n \text{ est } T_n\text{-mesurable}, \\ (b) \|\Delta_n - D_n\| \rightarrow 0 \text{ p.s.}, \\ (c) \sum_{n=1}^{+\infty} a_n \|\Delta_n - D_n\| < \infty \text{ p.s.},$$

X_n^i converge p.s., pour $i = 1, \dots, r$, vers un vecteur propre de $B = CM$ associé à la i -ème plus grande v.p. λ_i . ■

4.1.2 Cas particulier d'un modèle linéaire de variation de l'espérance et de la variance

On définit un modèle linéaire pour représenter l'espérance et la variance de Z_n .

On note $\delta_n = (\delta_n^1 \dots \delta_n^p)'$.

Pour $i = 1, \dots, p$, il existe un vecteur β^i (resp. γ^i) inconnu de $\mathbb{R}^{n_i}$ (resp. $\mathbb{R}^{m_i}$) et, pour tout n , un vecteur U_n^i (resp. V_n^i) de $\mathbb{R}^{n_i}$ (resp. $\mathbb{R}^{m_i}$) connu au temps n tels que la $i^{\text{ème}}$ composante réelle de θ_n (resp. δ_n) s'écrive :

$$\theta_n^i = \langle \beta^i, U_n^i \rangle \left(\text{resp. } \delta_n^i = \left(\langle \gamma^i, V_n^i \rangle \right)^{\frac{1}{2}} \right).$$

De façon générale, on note Π_E l'opérateur de projection sur un sous-espace E .

On définit le processus d'a.s. (B_n^i) de β^i tel que :

$$B_{n+1}^i = \Pi_{K_n^i} \left(B_n^i - a_n U_n^i \left((U_n^i)' B_n^i - Z_n^i \right) \right)$$

avec $K_n^i = \left\{ x \in \mathbb{R}^{n_i}, \|x\| < k_n^i \right\}$ où $k_n^i = O(n^\zeta)$, $\zeta > 0$.

On définit comme estimateur de θ_n^i , $\Theta_n^i = \langle B_n^i, U_n^i \rangle$ et comme estimateur de θ_n , $\Theta_n = (\Theta_n^1 \dots \Theta_n^p)'$.

On définit le processus d'a.s. (C_n^i) de γ_i tel que :

$$C_{n+1}^i = \Pi_{L_n^i} \left(C_n^i - a_n V_n^i \left((V_n^i)' C_n^i - (Z_n^i - \Theta_n^i)^2 \right) \right)$$

avec $L_n^i = \left\{ x \in \mathbb{R}^{m_i}, \|x\| < l_n^i \right\}$ où $l_n^i = O(n^\gamma)$, $\gamma > 0$.

On définit enfin comme estimateur de δ_n^i , $D_n^i = \left(\langle C_n^i, V_n^i \rangle \right)^{\frac{1}{2}}$ et comme estimateur de Δ_n , $D_n = \begin{pmatrix} D_n^1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & D_n^p \end{pmatrix}$.

On fait également les nouvelles hypothèses :

(H1) (c) R admet des moments d'ordre 4,

(H3) (b) $(a_n = \frac{a}{n^\alpha}$ avec $a > 0$, $\frac{2}{3} < \alpha \leq 1$),

(H6) (a) $\max_i \sup_n \|U_n^i\| < \infty$,

(a') $\max_i \sup_n \|V_n^i\| < \infty$,

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

(H7) Pour $\frac{1}{2} < \alpha < 1$, on suppose :

- (a) $\zeta < 2\alpha - 1$,
- (b) $\gamma < \frac{3}{2}\alpha - 1$,

(H8) Pour $\alpha = 1$, pour $i = 1, \dots, p$:

(a) si $\frac{2\lambda_i a}{r_i} \geq 1$, on suppose :

- (i) $\zeta < 2\alpha - 1$,
- (ii) $\gamma < \frac{3}{2}\alpha - 1$,

(b) si $\frac{2\lambda_i a}{r_i} < 1$, on suppose :

- (i) $\zeta < \frac{1}{2} + \frac{\lambda_i a}{r_i}$,
- (ii) $\gamma < \frac{\lambda_i a}{r_i}$.

Théorème 4.2.

Sous (H1)(c), (H3)(b), (H6), (H7) et (H8), les hypothèses concernant Θ_n et D_n du théorème 4.1 sont vérifiées. ■

Remarques : cas $\alpha = 1$.

1) Si l'on impose à a la condition : $a \geq \max_i \frac{r_i}{2\lambda_i}$, difficile à réaliser, alors $\frac{2\lambda_i a}{r_i} \geq 1$ pour tout i et il suffit de prendre $\zeta < 2\alpha - 1$ et $\gamma < \frac{3}{2}\alpha - 1$.

2) D'après la forme de l'hypothèse (H6)(b), λ_i peut être pris aussi petit que l'on veut. Donc, on peut toujours supposer que $\lambda_i < \frac{r_i}{2a}$ pour tout $i \Leftrightarrow \frac{2\lambda_i a}{r_i} < 1$ pour tout i . Mais la condition sur γ : $\gamma < \frac{\lambda_i a}{r_i}$ pour tout $i = 1 \Leftrightarrow \gamma < \min_i \frac{\lambda_i a}{r_i}$ est alors difficile à réaliser.

4.2 Cas particulier : l'ACG partielle

Comme dans [51], on se place dans le cas où le v.a. R est partitionné en q sous-vecteurs $R^{(1)}, \dots, R^{(q)}$.

Pour $k = 1, \dots, q$, $R^{(k)}$ est un v.a. dans $\mathbb{R}^{m_k}$, de composantes $R^{k1}, \dots, R^{km_k}$.

On note également $R^{(k)} = \{R^i, i \in J_k\}$, $\text{card}(J_k) = m_k$, $\sum_{k=1}^q m_k = p$.

On définit $C^k = \mathbb{E}[R^{(k)}(R^{(k)})']$, matrice de covariance de $R^{(k)}$.

On souhaite effectuer une ACP de R dans laquelle les v.a. $R^{(k)}$ aient un rôle équilibré : on veut éviter que les premiers facteurs soient principalement déterminés à partir de certains vecteurs $R^{(k)}$.

L'analyse canonique généralisée (ACG) du v.a. R fournit une solution à ce problème grâce à un choix particulier de métrique. L'ACG de R est une ACP avec la métrique diagonale par blocs d'ordre p , M :

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

Le v.a. Z_n observé est partitionné en sous-vecteurs $Z_n^{(1)}, \dots, Z_n^{(q)}$, de dimensions respectives $m_1, \dots, m_q$, avec :

$$Z_n^{(k)} = \theta_n^{(k)} + \Delta_n^k R_n^{(k)},$$

$\theta_n^{(k)}$ étant le vecteur des θ_n^i , $i \in J_k$,

Δ_n^k étant la matrice diagonale d'ordre m_k des δ_n^i ,

les $R_n^{(k)}$ constituant un échantillon i.i.d. de $R^{(k)}$,

Pour tout n , pour $k = 1, \dots, q$, $(C^k)^{-1}$ est solution de l'équation en X , matrice carrée d'ordre m_k :

$$\mathbb{E} \left[(\Delta_n^k)^{-1} (Z_n^{(k)} - \theta_n^{(k)}) (Z_n^{(k)})' (\Delta_n^k)^{-1} X - I \right] = 0$$

où I est la matrice-identité d'ordre m_k .

On suppose qu'on dispose d'estimateurs convergents $\Theta_n^{(k)}$ et D_n^k respectivement de $\theta_n^{(k)}$ et Δ_n^k .

Comme dans le chapitre 3, on simplifie les notations dans la suite en remplaçant les (k) par des k .

Pour estimer $(C^k)^{-1}$, on définit récursivement le processus d'a.s. de $(C^k)^{-1}$, (M_n^k) , par :

$$M_{n+1}^k = M_n^k - a_n \left((D_n^k)^{-1} (Z_n^k - \Theta_n^k) (Z_n^k)' (D_n^k)^{-1} M_n^k - I \right).$$

On définit alors comme estimateur de M au pas n la matrice diagonale par blocs M_n qui a pour $k^{\text{ième}}$ bloc diagonal M_n^k .

Soit les hypothèses :

(H1) (a) Il n'existe pas de relation affine entre les composantes de R ,

(H3) (c) ($a_n = \frac{a}{n^\alpha}$ avec $a > 0$, $\frac{2}{3} < \alpha < 1$) ou ($\alpha = 1$ i.e. $a_n = \frac{a}{n}$ et $a > \frac{1}{\min_k \lambda_{\min}(C^k)}$),

(H5) (d) $\mathbb{E}[||\theta_n - \Theta_n||] = O(\frac{1}{n^{\gamma_1}})$, $\gamma_1 > 2(1 - \alpha)$,

(H5') (d) $\mathbb{E}[||\Delta_n - D_n||] = O(\frac{1}{n^{\gamma_2}})$, $\gamma_2 > 2(1 - \alpha)$.

Théorème 4.3.

Sous les hypothèses (H1)(a),(b), (H2')(a),(b), (H3)(c), (H5)(a),(b),(d), (H5')(a),(b),(d) et (H6)(a), les hypothèses concernant M_n du théorème 4.1 sont vérifiées.

On peut alors appliquer directement le théorème 4.1. ■

Remarques :

1) L'hypothèse (H5)(d) (resp. (H5')(d)) implique, sous (H3)(c), l'hypothèse (H5)(c) (resp. (H5')(c)).

2) Dans le cas linéaire, sous les hypothèses du théorème 4.2,

· (H5)(d) est vérifiée pour $\frac{4}{5} < \alpha \leq 1$ car :

$$\mathbb{E}[||\theta_n - \Theta_n||] = O(\frac{1}{n^{\frac{\alpha}{2}}}) ; 2(1 - \alpha) < \frac{\alpha}{2} \text{ pour } \frac{4}{5} < \alpha < 1 \text{ et}$$

$$\mathbb{E}[||\theta_n - \Theta_n||] = O(\frac{1}{n^{\delta_2}}), \delta_2 > 0 \text{ pour } \alpha = 1,$$

· (H5')(d) est vérifiée pour $\alpha = 1$ car :

$$\mathbb{E}[||C_n^i - \gamma^i||] = O(\frac{1}{n^{\delta_3}}), \delta_3 > 0 \text{ donc } \mathbb{E}[||\Delta_n - D_n||] = O(\frac{1}{n^{\delta_3}}), \delta_3 > 0 \text{ pour } \alpha = 1.$$

4.3 Annexes du chapitre 4

4.3.1 Démonstrations des théorèmes

Démonstration du théorème 4.1 :

On vérifie les hypothèses d'application du théorème 1.2.

1) Dans un premier temps, on montre que $\sum_{n=1}^{+\infty} a_n ||\mathbb{E}[B_n|T_n] - B|| < \infty$ p.s.

$B_n = C_n M_n$. On a la décomposition suivante de $C_n - C$:

$$C_n - C = D_n^{-1}(Z_n Z_n' - \Theta_n \Theta_n') D_n^{-1} - \Delta_n^{-1}(\mathbb{E}[Z_n Z_n'] - \theta_n \theta_n') \Delta_n^{-1}$$

D'où, en passant à l'espérance conditionnellement à T_n ,

$$\mathbb{E}[C_n|T_n] - C = D_n^{-1}(\mathbb{E}[Z_n Z_n'] - \Theta_n \theta_n') D_n^{-1} - \Delta_n^{-1} \text{Cov}(Z_n) \Delta_n^{-1} \text{ sous (H5)(a) et (H5')(a) .} \quad (4.1)$$

$$(4.2)$$

$$= D_n^{-1}(\text{Cov}(Z_n) + \theta_n \theta_n' - \Theta_n \theta_n') D_n^{-1} - \Delta_n^{-1} \text{Cov}(Z_n) \Delta_n^{-1} \quad (4.3)$$

$$(4.4)$$

$$= D_n^{-1} \text{Cov}(Z_n) D_n^{-1} - \Delta_n^{-1} \text{Cov}(Z_n) \Delta_n^{-1} + D_n^{-1}(\theta_n - \Theta_n) \theta_n' D_n^{-1} \quad (4.5)$$

$$(4.6)$$

$$= (D_n^{-1} - \Delta_n^{-1}) \text{Cov}(Z_n) D_n^{-1} + \Delta_n^{-1} \text{Cov}(Z_n) (D_n^{-1} - \Delta_n^{-1}) + \quad (4.7)$$

$$(D_n^{-1} - \Delta_n^{-1})(\theta_n - \Theta_n) \theta_n' D_n^{-1} + \Delta_n^{-1}(\theta_n - \Theta_n) \theta_n' D_n^{-1} \quad (4.8)$$

$$(4.9)$$

$$= (D_n^{-1} - \Delta_n^{-1}) \text{Cov}(Z_n) (D_n^{-1} - \Delta_n^{-1}) + (D_n^{-1} - \Delta_n^{-1}) \text{Cov}(Z_n) \Delta_n^{-1} + \quad (4.10)$$

$$\Delta_n^{-1} \text{Cov}(Z_n) (D_n^{-1} - \Delta_n^{-1}) + (D_n^{-1} - \Delta_n^{-1})(\theta_n - \Theta_n) \theta_n' (D_n^{-1} - \Delta_n^{-1}) + \quad (4.11)$$

$$(D_n^{-1} - \Delta_n^{-1})(\theta_n - \Theta_n) \theta_n' \Delta_n^{-1} + \Delta_n^{-1}(\theta_n - \Theta_n) \theta_n' (D_n^{-1} - \Delta_n^{-1}) + \quad (4.12)$$

$$\Delta_n^{-1}(\theta_n - \Theta_n) \theta_n' \Delta_n^{-1}.$$

Alors,

$$\mathbb{E}[C_n|T_n] - C \leq \|D_n^{-1} - \Delta_n^{-1}\|^2 \|Cov(Z_n)\| + 2 \|D_n^{-1} - \Delta_n^{-1}\| \cdot \|Cov(Z_n)\| \cdot \|\Delta_n^{-1}\| + \quad (4.13)$$

$$\|D_n^{-1} - \Delta_n^{-1}\|^2 \|\theta_n - \Theta_n\| \cdot \|\theta_n\| + 2 \|D_n^{-1} - \Delta_n^{-1}\| \cdot \|\theta_n - \Theta_n\| \cdot \|\theta_n\| \cdot \|\Delta_n^{-1}\| + \quad (4.14)$$

$$\|\Delta_n^{-1}\|^2 \|\theta_n - \Theta_n\| \cdot \|\theta_n\|. \quad (4.15)$$

$$\leq \|D_n^{-1} - \Delta_n^{-1}\| \left(\|D_n^{-1} - \Delta_n^{-1}\| \cdot \|Cov(Z_n)\| + 2 \|Cov(Z_n)\| \cdot \|\Delta_n^{-1}\| + \quad (4.16)$$

$$\|D_n^{-1} - \Delta_n^{-1}\| \cdot \|\theta_n - \Theta_n\| \cdot \|\theta_n\| + 2 \|\theta_n - \Theta_n\| \cdot \|\theta_n\| \cdot \|\Delta_n^{-1}\| \right) + \quad (4.17)$$

$$\|\Delta_n^{-1}\|^2 \|\theta_n - \Theta_n\| \cdot \|\theta_n\|.$$

Or,

$$\cdot \|Cov(Z_n)\| = \|\Delta_n C \Delta_n\| \leq \|C\| \cdot \|\Delta_n\|^2 \leq \|C\| \sup_l \max_i (\delta_l^i)^2 < \infty \text{ sous (H2')}(a),$$

$$\cdot \|\Delta_n^{-1}\| = \frac{1}{\min_i |\delta_n^i|} < \infty \text{ sous (H2')}(b),$$

$$\cdot \|\theta_n\| \leq \sup_l \|\theta_l\| < \infty \text{ sous (H2)},$$

$$\cdot \|\theta_n - \Theta_n\| \rightarrow 0 \text{ p.s. sous (H5)(b)},$$

$$\cdot \|D_n^{-1} - \Delta_n^{-1}\| = \|D_n^{-1}(\Delta_n - D_n)\Delta_n^{-1}\| \leq \|\Delta_n^{-1}\| \cdot \|D_n^{-1}\| \cdot \|\Delta_n - D_n\|.$$

Comme $\Delta_n - D_n \rightarrow 0$ p.s., pour $i = 1, \dots, p$, $D_n^i - \delta_n^i \rightarrow 0$ p.s.

Soit $0 < \varepsilon < \inf_l \min_i |\delta_l^i|$,

$$\|D_n^{-1}\| = \frac{1}{\min_i |D_n^i|} < \frac{1}{\inf_l \min_i |\delta_l^i| - \varepsilon}, \text{ à partir d'un certain rang.}$$

Comme, sous (H5')(b), $\|\Delta_n - D_n\| \rightarrow 0$ p.s., on a : $\|D_n^{-1} - \Delta_n^{-1}\| \rightarrow 0$ p.s.

Ainsi, il existe deux v.a. P et Q positives et finies p.s. telles que :

$$\|\mathbb{E}[C_n|T_n] - C\| \leq P \|D_n - \Delta_n\| + Q \|\theta_n - \Theta_n\|.$$

puis sous (H5)(c) et (H5')(c) : $\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| < \infty$ p.s.

D'autre part,

$$\begin{aligned}\mathbb{E}[B_n|T_n] - B &= \mathbb{E}[C_n|T_n]M_n - CM \text{ sous (H4)(a)} \\ &= (\mathbb{E}[C_n|T_n] - C)M_n + C(M_n - M)\end{aligned}\quad (4.18)$$

d'où :

$$\begin{aligned}\sum_{n=1}^{+\infty} a_n \|\mathbb{E}[B_n|T_n] - B\| &\leq \sum_{n=1}^{+\infty} a_n \|\mathbb{E}[C_n|T_n] - C\| \cdot \|M_n\| + \sum_{n=1}^{+\infty} a_n \|C\| \cdot \|M_n - M\| \\ &< \infty \text{ sous (H4)(b),(c)}.\end{aligned}\quad (4.19)$$

2) On montre maintenant que, pour $j = 2, \dots, 2r$, $\sum_{n=1}^{+\infty} a_n^j \mathbb{E}[\|B_n - B\|^j | T_n] < \infty$ p.s.

$$C_n - C = D_n^{-1}(Z_n Z'_n - \Theta_n Z'_n) D_n^{-1} - C.$$

d'où :

$$\begin{aligned}\|C_n - C\|^j &\leq 2^{j-1} (\|D_n^{-1}(Z_n Z'_n - \Theta_n Z'_n) D_n^{-1}\|^j + \|C\|^j) \\ &\leq 2^{j-1} (\|D_n^{-1}\|^{2j} \|Z_n Z'_n - \Theta_n Z'_n\|^j + \|C\|^j)\end{aligned}\quad (4.20)$$

puis, en passant à l'espérance conditionnellement à T_n :

$$\begin{aligned}\mathbb{E}[\|C_n - C\|^j | T_n] &\leq 2^{j-1} \|D_n^{-1}\|^{2j} \mathbb{E}[\|Z_n Z'_n - \Theta_n Z'_n\|^j | T_n] + 2^{j-1} \|C\|^j \\ &\leq 2^{j-1} (\|D_n^{-1} - \Delta_n^{-1}\|^{2j} + \|\Delta_n^{-1}\|^{2j}) 2^{2j-1} \mathbb{E}[\|Z_n Z'_n - \Theta_n Z'_n\|^j | T_n] + 2^{j-1} \|C\|^j.\end{aligned}\quad (4.21)$$

Or,

$$\|Z_n Z'_n - \Theta_n Z'_n\| = \|(\Theta_n + \Delta_n R_n)(\Theta_n + \Delta_n R_n)' - \Theta_n(\Theta_n + \Delta_n R_n)'\| \quad (4.22)$$

$$= \|\Theta_n \Theta'_n + \Delta_n R_n \Theta'_n + \Theta_n R'_n \Delta_n + \Delta_n R_n R'_n \Delta_n - \Theta_n \Theta'_n - \Theta_n \Theta'_n - (\Theta_n - \Theta_n) R'_n \Delta_n - \Theta_n R'_n \Delta_n\| \quad (4.23)$$

$$\begin{aligned}&= \|\Delta_n R_n \Theta'_n + \Theta_n R'_n \Delta_n + \Delta_n R_n R'_n \Delta_n\| \\ &\leq c(\|R_n\| + \|R_n\|^2 + \|\Theta_n - \Theta_n\| + \|\Theta_n - \Theta_n\| \cdot \|R_n\|).\end{aligned}\quad (4.24)$$

Ainsi,

$$\begin{aligned}\mathbb{E}[\|Z_n Z'_n - \Theta_n Z'_n\|^j | T_n] &\leq c 4^{j-1} \left(\mathbb{E}[\|R\|^j] + \mathbb{E}[\|R\|^{2j}] + \|\Theta_n - \Theta_n\|^j + \|\Theta_n - \Theta_n\|^j \mathbb{E}[\|R\|^4] \right) \\ &\leq d (1 + \|\Theta_n - \Theta_n\|^j) \text{ sous (H1)(b)}\end{aligned}\quad (4.25)$$

et finalement :

$$\mathbb{E}[\|C_n - C\|^j | T_n] \leq e (1 + \|D_n^{-1} - \Delta_n^{-1}\|^{2j}) (1 + \|\Theta_n - \Theta_n\|^j) + 2^{j-1} \|C\|^j. \quad (4.26)$$

Alors, sous (H2')(b), (H5')(b), (H5)(c) et (H5')(c), il existe une v.a. S positive et finie p.s. telle que :

$$\sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[\|C_n - C\|^j \middle| T_n \right] < \infty.$$

En utilisant la décomposition : $B_n - B = C_n M_n - CM = (C_n - C)M_n + C(M_n - M)$, on peut écrire :

$$\|B_n - B\|^j \leq 2^{j-1} \left(\|C_n - C\|^j \|M_n\|^j + \|C\|^j \|M_n - M\|^j \right) \text{ et alors :}$$

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[\|B_n - B\|^j \middle| T_n \right] &\leq 2^{j-1} \left(\sum_{n=1}^{+\infty} a_n^j \mathbb{E} \left[\|C_n - C\|^j \middle| T_n \right] \cdot \|M_n\|^j + \sum_{n=1}^{+\infty} a_n^j \|C\|^j \|M_n - M\|^j \right) \text{ sous (H4)(a)} \\ &< \infty \text{ sous (H4)(b),(c).} \end{aligned}$$

Démonstration du théorème 4.2 :

1) On montre que sous (H3)(b) et (H6), $\|\theta_n - \Theta_n\| \rightarrow 0$ p.s.

a) On a : $B_{n+1}^i = \Pi_{K_n^i} \left(B_n^i - a_n U_n^i \left((U_n^i)' B_n^i - Z_n^i \right) \right)$

$$\begin{aligned} B_{n+1}^i - \beta^i &= \Pi_{K_n^i} \left(B_n^i - a_n U_n^i \left((U_n^i)' B_n^i - Z_n^i \right) \right) - \beta^i \\ &= \Pi_{K_n^i} \left(B_n^i - a_n U_n^i \left((U_n^i)' B_n^i - Z_n^i \right) \right) - \Pi_{K_n^i}(\beta^i) \text{ à partir d'un certain rang car } \zeta > 0. \end{aligned}$$

On enlève dans la suite l'écriture de i .

$$\begin{aligned} \|B_{n+1} - \beta\| &\leq \|B_n - a_n U_n (U_n' B_n - Z_n) - \beta\| \\ &\leq \|B_n - \beta - a_n U_n (U_n' (B_n - \beta) - \delta_n R_n)\|, \end{aligned}$$

puis, en posant $Y_n = B_n - \beta$, on trouve :

$$\begin{aligned} \|Y_{n+1}\|^2 &\leq \|Y_n\|^2 + a_n^2 \|U_n U_n' Y_n - U_n \delta_n R_n\|^2 - 2a_n \langle Y_n, U_n (U_n' Y_n - \delta_n R_n) \rangle \\ &\leq \|Y_n\|^2 + a_n^2 \|U_n U_n' Y_n - U_n \delta_n R_n\|^2 - 2a_n \langle Y_n, U_n U_n' Y_n \rangle + 2a_n \langle Y_n, U_n \delta_n R_n \rangle \\ &\leq \|Y_n\|^2 + 2a_n^2 \|U_n\|^4 \|Y_n\|^2 + 2a_n^2 \delta_n^2 \|U_n\|^2 R_n^2 - 2a_n \langle Y_n, U_n U_n' Y_n \rangle + 2a_n \delta_n \langle Y_n, U_n R_n \rangle. \end{aligned}$$

En prenant l'espérance conditionnellement à T_n , on obtient alors :

$$\mathbb{E}[\|Y_{n+1}\|^2 | T_n] \leq \|Y_n\|^2 \left(1 + 2a_n^2 \|U_n\|^4 \right) + 2a_n^2 \delta_n^2 \|U_n\|^2 \mathbb{E}[R^2] - 2a_n \langle Y_n, U_n U_n' Y_n \rangle.$$

Sous les hypothèses (H6)(a) et (H3)(b), on peut appliquer le lemme de Robbins-Siegmund :

$$\exists T > 0, \|Y_n\|^2 \rightarrow T \text{ p.s. et } \sum_{n=1}^{+\infty} a_n \langle Y_n, U_n U_n' Y_n \rangle < \infty \text{ p.s.}$$

b) On a :

$$\begin{aligned} Y_{n+1} - Y_n &= B_{n+1} - B_n \\ &= \Pi_{K_n} \left(B_n - a_n U_n (U_n' B_n - Z_n) \right) - B_n \\ &= \Pi_{K_n} \left(B_n - a_n U_n (U_n' B_n - Z_n) \right) - \Pi_{K_n}(B_n) \text{ car } B_n \in K_{n-1} \subset K_n. \end{aligned}$$

Comme $Z_n = U_n' \beta + R_n$,

$$\begin{aligned} \|Y_{n+1} - Y_n\| &\leq \| -a_n U_n (U_n' Y_n) + a_n U_n \delta_n R_n \| \\ &\leq a_n \|U_n\|^2 \|Y_n\| + a_n \|U_n\| \cdot |\delta_n| \cdot |R_n|. \end{aligned}$$

Or,

$$\begin{aligned} \mathbb{E} \left[\sum_{n=1}^{+\infty} a_n^2 \|U_n\|^2 \delta_n^2 R_n^2 \right] &= \sum_{n=1}^{+\infty} a_n^2 \|U_n\|^2 \delta_n^2 \mathbb{E}[R^2] \\ &< \infty \quad (\text{H3})(b), (\text{H6})(a) \text{ et } (\text{H6})(a'). \end{aligned}$$

Donc : $\sum_{n=1}^{+\infty} a_n^2 \|U_n\|^2 \delta_n^2 R_n^2 < \infty$ p.s. et par suite $a_n \|U_n\| \cdot |\delta_n| \cdot |R_n| \rightarrow 0$ p.s.

D'autre part, comme $\|Y_n\|^2 \rightarrow T$ p.s., $a_n \|U_n\|^2 \|Y_n\| \rightarrow 0$ p.s. sous (H3)(b) et (H6)(a).

On en déduit que : $\|Y_{n+1} - Y_n\| \rightarrow 0$ p.s.

c) On raisonne à ω fixé, appartenant à l'intersection des ensembles de convergence p.s. définis. Supposons $T(\omega) \neq 0$.

On supprime dans la suite l'écriture de ω .

Comme $\|Y_n\| \rightarrow \sqrt{T}$ p.s.,

$$\exists 0 < \varepsilon_1 < 1, \exists N(\varepsilon_1) : \forall n > N(\varepsilon_1), \varepsilon_1 < \|Y_n\| < \frac{1}{\varepsilon_1}.$$

Donc, sous (H6)(b), à partir d'un certain rang L , on a :

$$\langle Y_{n_l}, \sum_{j \in I_l} U_j U_j' Y_{n_l} \rangle \geq \lambda \varepsilon_1^2.$$

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

$$\langle Y_{n_l}, U_{p_l} U_{p_l}' Y_{n_l} \rangle \geq \frac{\lambda \varepsilon_1^2}{r}. \quad (1)$$

On considère la décomposition :

$$\langle Y_{p_l}, U_{p_l} U_{p_l}' Y_{p_l} \rangle = \langle Y_{p_l} + Y_{n_l}, U_{p_l} U_{p_l}' (Y_{p_l} - Y_{n_l}) \rangle + \langle Y_{n_l}, U_{p_l} U_{p_l}' Y_{n_l} \rangle. \quad (2)$$

Soit $\varepsilon > 0$ tel que $\varepsilon \leq \frac{\lambda \varepsilon_1^3}{4r^2 C^2}$, avec $C = \sup_n \|U_n\| < \infty$ sous (H6)(a).

A partir d'un certain rang : $\|Y_{n+1} - Y_n\| < \varepsilon$, donc : $\|Y_{p_l} - Y_{n_l}\| < r\varepsilon$;

$$\begin{aligned}
\left| \left\langle Y_{p_l} + Y_{n_l}, U_{p_l} U'_{p_l} (Y_{p_l} - Y_{n_l}) \right\rangle \right| &< \left(\|Y_{p_l}\| + \|Y_{n_l}\| \right) \|U_{p_l}\|^2 \|Y_{p_l} - Y_{n_l}\| \\
&< \frac{2}{\varepsilon_1} C^2 r \varepsilon \\
&\leq \frac{\lambda \varepsilon_1^2}{2r}.
\end{aligned}$$

Ainsi, $-\frac{\lambda \varepsilon_1^2}{2r} < \left\langle (Y_{p_l} + Y_{n_l}, U_{p_l} U'_{p_l} (Y_{p_l} - Y_{n_l})) \right\rangle < +\frac{\lambda \varepsilon_1^2}{2r}. \quad (3)$

On déduit de (1), (2) et (3) que :

$$\begin{aligned}
\langle Y_{p_l}, U_{p_l} U'_{p_l} Y_{p_l} \rangle &> \frac{\lambda \varepsilon_1^2}{r} - \frac{\lambda \varepsilon_1^2}{2r} \\
&\geq \frac{\lambda \varepsilon_1^2}{2r}.
\end{aligned}$$

Sous (H3)(b), on a alors :

$$\begin{aligned}
\sum_{l=1}^{+\infty} a_{p_l} \langle Y_{p_l}^i, U_{p_l} U'_{p_l} Y_{p_l} \rangle &\geq \frac{\lambda \varepsilon_1^2}{2r} \sum_{l=1}^{+\infty} a_{p_l} \\
&\geq c_1 \sum_{l=1}^{+\infty} \min_{j \in I_l} a_j.
\end{aligned}$$

Or, sous (H3)(b), $\min_{j \in I_l} a_j = \frac{a}{(n_{l+1} - 1)^\alpha} \geq \frac{a}{(lr)^\alpha}.$

Donc : $\sum_{l=1}^{+\infty} \min_{j \in I_l} a_j = \infty$ (car $\alpha \leq 1$).

$\sum_{n=1}^{+\infty} a_n \langle Y_n, U_n U'_n Y_n \rangle \geq \sum_{l=1}^{+\infty} a_{p_l} \langle Y_{p_l}, U_{p_l} U'_{p_l} Y_{p_l} \rangle = \infty$; il y a contradiction avec la conclusion de a).

Par conséquent, $T(\omega) = 0$ et en réécrivant les i , on obtient : $B_n^i \rightarrow \beta^i$ p.s.

Par suite, $|\theta_n^i - \Theta_n^i| \rightarrow 0$ p.s. et enfin $\|\theta_n - \Theta_n\| \rightarrow 0$ p.s.

2) On montre que sous (H3)(b) et (H6), $\sum_{n=1}^{+\infty} a_n \|\theta_n - \Theta_n\| < \infty$ p.s.

On supprime à nouveau l'écriture des i et on pose $Y_n = B_n - \beta$.

a) D'après l'étude de la convergence p.s. de (B_n) (partie 1)a) de la démonstration),

$$\mathbb{E} \left[\|Y_{n+1}\|^2 \right] \leq \mathbb{E} \left[\|Y_n\|^2 \right] \cdot (1 + 2a_n^2 \|U_n\|^4) + 2a_n^2 \delta_n^2 \|U_n\|^2 \mathbb{E}[R^2] - 2a_n \mathbb{E} \left[\langle Y_n, U_n U_n' Y_n \rangle \right].$$

Donc, sous (H3)(b), (H6)(a) et (H6)(a'), $\mathbb{E} \left[\|Y_n\|^2 \right] \rightarrow t$ et $\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\langle Y_n, U_n U_n' Y_n \rangle \right] < \infty$.

b) Ainsi, il existe $b > 0$ tel que :

$$\mathbb{E} \left[\|Y_{n+1}\|^2 \right] \leq \mathbb{E} \left[\|Y_n\|^2 \right] + ba_n^2 - 2a_n \mathbb{E} \left[\langle Y_n, U_n U_n' Y_n \rangle \right]$$

et

$$\mathbb{E} \left[\|Y_{n_l+1}\|^2 \right] \leq \mathbb{E} \left[\|Y_{n_l}\|^2 \right] + b \sum_{j \in I_l} a_j^2 - 2\mu_l \sum_{j \in I_l} \mathbb{E} \left[\langle Y_j, U_j U_j' Y_j \rangle \right] \text{ avec } \mu_l = \min_{j \in I_l} a_j.$$

On décompose le dernier terme du second membre de la manière suivante :

$$\sum_{j \in I_l} \mathbb{E} \left[\langle Y_j, U_j U_j' Y_j \rangle \right] = \sum_{j \in I_l} \mathbb{E} \left[\langle Y_{n_l}, U_j U_j' Y_{n_l} \rangle \right] + \sum_{j \in I_l} \mathbb{E} \left[\langle Y_j + Y_{n_l}, U_j U_j' (Y_j - Y_{n_l}) \rangle \right]. \quad (4)$$

$$\text{Sous (H6)(b), } \sum_{j \in I_l} \mathbb{E} \left[\langle Y_{n_l}, U_j U_j' Y_{n_l} \rangle \right] \geq \lambda \mathbb{E} \left[\|Y_{n_l}\|^2 \right]. \quad (5)$$

On note $C_1 = \sup_n \|U_n\|$.

$$\begin{aligned} \sum_{j \in I_l} \left| \mathbb{E} \left[\langle Y_j + Y_{n_l}, U_j U_j' (Y_j - Y_{n_l}) \rangle \right] \right| &\leq \sum_{j \in I_l} \mathbb{E} \left[\|Y_j + Y_{n_l}\| \cdot \|U_j\|^2 \|Y_j - Y_{n_l}\| \right] \\ &\leq C_1^2 \sum_{j \in I_l} \left(\mathbb{E} \left[\|Y_j + Y_{n_l}\|^2 \right] \right)^{\frac{1}{2}} \left(\mathbb{E} \left[\|Y_j - Y_{n_l}\|^2 \right] \right)^{\frac{1}{2}} \\ &\leq C_2 \sum_{j \in I_l} (\mathbb{E} \left[\|Y_j - Y_{n_l}\|^2 \right])^{\frac{1}{2}} \text{ car } \mathbb{E} \left[\|Y_n\|^2 \right] \rightarrow t. \end{aligned}$$

Or, $Y_j - Y_{n_l} = \sum_{k=n_l}^{j-1} (Y_{k+1} - Y_k)$, donc :

$$\begin{aligned} \|Y_j - Y_{n_l}\| &\leq \sum_{k=n_l}^{j-1} \|(B_{k+1} - \beta) - (B_k - \beta)\| \\ &\leq \sum_{k=n_l}^{j-1} \|-a_k U_k U_k' Y_k + a_k U_k \delta_k R_k\| \\ &\leq \sum_{k=n_l}^{j-1} (a_k C_1^2 \|Y_k\| + a_k C_3 \cdot |R_k|) \text{ sous (H6)(a')}. \end{aligned}$$

Alors, $\|Y_j - Y_{n_l}\|^2 \leq c \sum_{k=n_l}^{j-1} (a_k^2 \|Y_k\|^2 + a_k^2 R_k^2)$ et $\mathbb{E} \left[\|Y_j - Y_{n_l}\|^2 \right] \leq d \max_{k \in I_l} a_k^2$ puisque $\mathbb{E} \left[\|Y_n\|^2 \right] \rightarrow t$ p.s.

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

puis, avec (5), (4) devient : $\sum_{j \in I_l} \mathbb{E} \left[\langle Y_j, U_j U'_j Y_j \rangle \right] \geq \lambda \mathbb{E} \left[\|Y_{n_l}\|^2 \right] - f \max_{k \in I_l} a_k.$

Par conséquent,

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

Or, sous (H3)(b) et (H6)(b), $\mu_l \geq \frac{a}{(lr)^\alpha}$ et $\max_{k \in I_l} a_k = \frac{a}{n_l^\alpha} \leq \frac{c}{l^\alpha}$,

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

c) Dans le cas $\frac{1}{2} < \alpha < 1$, on applique un lemme de Schemetterer (1969) [62] :

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

Dans le cas $\alpha = 1$, on applique un lemme de Venter (1966) [70] :

- pour $\frac{2\lambda a}{r} > 1$, $\overline{\lim} l \mathbb{E} \left[\|Y_{n_l}\|^2 \right] < \infty$;
- pour $\frac{2\lambda a}{r} = 1$, $\overline{\lim} \frac{l}{\ln(l)} \mathbb{E} \left[\|Y_{n_l}\|^2 \right] < \infty$;
- pour $\frac{2\lambda a}{r} < 1$, $\overline{\lim} l^{\frac{2\lambda a}{r}} \mathbb{E} \left[\|Y_{n_l}\|^2 \right] < \infty$.

D'autre part, d'après la partie 2)b) de la démonstration,

$$\begin{aligned} \mathbb{E} \left[\|Y_{n+1}\|^2 \right] &\leq \mathbb{E} \left[\|Y_n\|^2 \right] + ba_n^2 - 2a_n \mathbb{E} \left[\langle Y_n, U_n U'_n Y_n \rangle \right] \\ &\leq \mathbb{E} \left[\|Y_n\|^2 \right] + ba_n^2. \end{aligned}$$

Donc, pour $n \in I_l$,

$$\begin{aligned} \mathbb{E} \left[\|Y_n\|^2 \right] &\leq \mathbb{E} \left[\|Y_{n_l}\|^2 \right] + br \max_{k \in I_l} a_k^2 \\ &\leq \mathbb{E} \left[\|Y_{n_l}\|^2 \right] + \frac{h}{l^{2\alpha}}. \end{aligned}$$

Alors, comme $l \leq n \leq lr$, on a :

- pour $(\frac{1}{2} < \alpha < 1)$ ou $(\alpha = 1 \text{ et } \frac{2\lambda a}{r} > 1)$,

$$\begin{aligned} n^\alpha \mathbb{E} \left[||Y_n||^2 \right] &\leq n^\alpha \mathbb{E} \left[||Y_{n_l}||^2 \right] + \frac{hn^\alpha}{l^{2\alpha}} \\ &\leq r^\alpha l^\alpha \mathbb{E} \left[||Y_{n_l}||^2 \right] + \frac{hr^\alpha}{l^\alpha} \end{aligned}$$

$$\text{et } \overline{\lim} \ n^\alpha \mathbb{E} \left[||Y_n||^2 \right] < \infty ;$$

· pour $(\alpha = 1 \text{ et } \frac{2\lambda a}{r} = 1)$,

$$\begin{aligned} \frac{n}{\ln(n)} \mathbb{E} \left[||Y_n||^2 \right] &\leq \frac{n}{\ln(n)} \mathbb{E} \left[||Y_{n_l}||^2 \right] + \frac{n}{\ln(n)} \frac{h}{l^2} \\ &\leq r \frac{l}{\ln(l)} \mathbb{E} \left[||Y_{n_l}||^2 \right] + \frac{lr}{\ln(l)} \frac{h}{l^2} \\ &\leq r \frac{l}{\ln(l)} \mathbb{E} \left[||Y_{n_l}||^2 \right] + \frac{hr}{l \ln(l)} \end{aligned}$$

$$\text{et } \overline{\lim} \ \frac{n}{\ln(n)} \mathbb{E} \left[||Y_n||^2 \right] < \infty ;$$

· pour $(\alpha = 1 \text{ et } \frac{2\lambda a}{r} < 1)$,

$$\begin{aligned} n^{\frac{2\lambda a}{r}} \mathbb{E} \left[||Y_n||^2 \right] &\leq n^{\frac{2\lambda a}{r}} \mathbb{E} \left[||Y_{n_l}||^2 \right] + n^{\frac{2\lambda a}{r}} \frac{h}{l^2} \\ &\leq r^{\frac{2\lambda a}{r}} l^{\frac{2\lambda a}{r}} \mathbb{E} \left[||Y_{n_l}||^2 \right] + r^{\frac{2\lambda a}{r}} l^{\frac{2\lambda a}{r}} \frac{h}{l^2} \\ &\leq r^{\frac{2\lambda a}{r}} l^{\frac{2\lambda a}{r}} \mathbb{E} \left[||Y_{n_l}||^2 \right] + \frac{hr^{\frac{2\lambda a}{r}}}{l^{2 - \frac{2\lambda a}{r}}} \text{ avec } 2 - \frac{2\lambda a}{r} > 1 \end{aligned}$$

$$\text{et } \overline{\lim} \ n^{\frac{2\lambda a}{r}} \mathbb{E} \left[||Y_n||^2 \right] < \infty .$$

$$\text{d) } \Theta_n - \theta_n = \langle B_n, U_n \rangle - \langle \beta, U_n \rangle = \langle Y_n, U_n \rangle .$$

$$\text{Donc, } (\Theta_n - \theta_n)^2 \leq ||Y_n||^2 ||U_n||^2 \leq C_1^2 ||Y_n||^2 .$$

On a donc les mêmes conclusions pour $\mathbb{E} \left[(\Theta_n - \theta_n)^2 \right]$.

$$\text{Ensuite, } \sum_{n=1}^{+\infty} a_n \mathbb{E} \left[|\theta_n - \Theta_n| \right] \leq \sum_{n=1}^{+\infty} a_n \left(\mathbb{E} \left[(\theta_n - \Theta_n)^2 \right] \right)^{\frac{1}{2}} .$$

Or, à partir d'un certain rang, on a :

· pour $(\frac{1}{2} < \alpha < 1)$ ou $(\alpha = 1 \text{ et } \frac{2\lambda a}{r} > 1)$:

$$\mathbb{E}[(\Theta_n - \theta_n)^2] \leq \frac{c}{n^\alpha} \text{ et donc } \sum_{n=1}^{+\infty} a_n \left(\mathbb{E}[(\theta_n - \Theta_n)^2] \right)^{\frac{1}{2}} \leq \sum_{n=1}^{+\infty} \frac{d}{n^{\frac{3\alpha}{2}}} < \infty \text{ sous (H3)(b)};$$

· pour $(\alpha = 1 \text{ et } \frac{2\lambda a}{r} = 1)$:

$$\mathbb{E}[(\Theta_n - \theta_n)^2] \leq \frac{c \ln(n)}{n} \text{ et donc } \sum_{n=1}^{+\infty} a_n \left(\mathbb{E}[(\theta_n - \Theta_n)^2] \right)^{\frac{1}{2}} \leq \sum_{n=1}^{+\infty} \frac{d (\ln(n))^{\frac{1}{2}}}{n^{\frac{3}{2}}} < \infty;$$

· pour $(\alpha = 1 \text{ et } \frac{2\lambda a}{r} < 1)$:

$$\mathbb{E}[(\Theta_n - \theta_n)^2] \leq \frac{c}{n^{\frac{2\lambda a}{r}}} \text{ et donc } \sum_{n=1}^{+\infty} a_n \left(\mathbb{E}[(\theta_n - \Theta_n)^2] \right)^{\frac{1}{2}} \leq \sum_{n=1}^{+\infty} \frac{d}{n^{1 + \frac{\lambda a}{r}}} < \infty.$$

Dans tous les cas et en réécrivant les i , on a : $\sum_{n=1}^{+\infty} a_n \mathbb{E}[|\theta_n^i - \Theta_n^i|] < \infty$ et par suite $\sum_{n=1}^{+\infty} a_n |\theta_n^i - \Theta_n^i| < \infty$ p.s.

Donc, $\sum_{n=1}^{+\infty} a_n \|\theta_n - \Theta_n\| < \infty$ p.s. et (H3)(a) est vérifiée.

3) On montre que sous (H3)(b) et (H6), $\|D_n^{-1} - \Delta_n^{-1}\| \rightarrow 0$ p.s.

$$\text{a) On a : } C_{n+1}^i = \Pi_{L_n^i} \left(C_n^i - a_n V_n^i \left((V_n^i)' C_n^i - (Z_n^i - \Theta_n^i)^2 \right) \right).$$

Donc,

$$\begin{aligned} C_{n+1}^i - \gamma^i &= \Pi_{L_n^i} \left(C_n^i - a_n V_n^i \left((V_n^i)' C_n^i - (Z_n^i - \Theta_n^i)^2 \right) \right) - \gamma^i \\ &= \Pi_{L_n^i} \left(C_n^i - a_n V_n^i \left((V_n^i)' C_n^i - (Z_n^i - \Theta_n^i)^2 \right) \right) - \Pi_{L_n^i}(\gamma^i) \text{ à partir d'un certain rang car } \gamma > 0. \end{aligned}$$

On enlève dans la suite l'écriture de i .

$$\begin{aligned} \|C_{n+1} - \gamma\| &\leq \|(C_n - \gamma) - a_n V_n \left(V_n' C_n - (Z_n - \Theta_n)^2 \right)\| \\ &\leq \|(C_n - \gamma) - a_n V_n \left(V_n' (C_n - \gamma) + (\delta_n)^2 - (Z_n - \Theta_n)^2 \right)\| \\ &\leq \|(C_n - \gamma) - a_n V_n \left(V_n' (C_n - \gamma) + (\delta_n)^2 - (Z_n - \Theta_n)^2 - 2(Z_n - \Theta_n)(\theta_n - \Theta_n) - (\theta_n - \Theta_n)^2 \right)\|. \end{aligned}$$

Alors, en posant $W_n = C_n - \gamma$, on a :

$$\begin{aligned} ||W_{n+1}||^2 &\leq ||W_n||^2 + 4 a_n^2 ||V_n||^4 ||W_n||^2 + 4 a_n^2 ||V_n||^2 \left(\delta_n^2 - (Z_n - \theta_n)^2 \right)^2 + \\ &\quad 16 a_n^2 ||V_n||^2 (Z_n - \theta_n)^2 (\theta_n - \Theta_n)^2 + 4 a_n^2 ||V_n||^2 (\theta_n - \Theta_n)^4 - 2 a_n \langle W_n, V_n V_n' W_n \rangle - \\ &\quad 2 a_n \langle W_n, V_n \rangle \left(\delta_n^2 - (Z_n - \theta_n)^2 \right) + 4 a_n \langle W_n, V_n \rangle (Z_n - \theta_n) (\theta_n - \Theta_n) + \\ &\quad 2 a_n \langle W_n, V_n \rangle (\theta_n - \Theta_n)^2. \end{aligned}$$

Soit T_n la tribu du passé au temps n .

On a :

i)

$$\begin{aligned} \mathbb{E} \left[(Z_n - \theta_n)^2 | T_n \right] &= \mathbb{E} \left[(Z_n - \theta_n)^2 \right] \\ &= \delta_n^2 \end{aligned}$$

D'autre part,

$$\begin{aligned} \mathbb{E} \left[||V_n (V_n' \gamma - (Z_n - \theta_n)^2)||^2 | T_n \right] &= \mathbb{E} \left[||V_n||^2 \left(V_n' \gamma - (Z_n - \theta_n)^2 \right)^2 | T_n \right] \\ &= ||V_n||^2 \text{Var} \left((Z_n - \theta_n)^2 \right) \\ &= ||V_n||^2 \delta_n^4 \text{Var} (R_n^2) \\ &\leq ||V_n||^4 ||\gamma||^2 \mathbb{E} [R^4]. \end{aligned}$$

ii)

$$\begin{aligned} \mathbb{E} \left[V_n (Z_n - \theta_n) (\theta_n - \Theta_n) | T_n \right] &= V_n \mathbb{E} \left[(Z_n - \theta_n) | T_n \right] (\theta_n - \Theta_n) \\ &= 0. \end{aligned}$$

D'autre part,

$$\begin{aligned} \mathbb{E} \left[||V_n (Z_n - \theta_n) (\theta_n - \Theta_n)||^2 | T_n \right] &= ||V_n||^2 \mathbb{E} \left[(Z_n - \theta_n)^2 | T_n \right] (\theta_n - \Theta_n)^2 \\ &= ||V_n||^2 \delta_n^2 (\theta_n - \Theta_n)^2 \\ &\leq ||V_n||^3 ||\gamma|| (\theta_n - \Theta_n)^2. \end{aligned}$$

iii)

$$\begin{aligned}\langle W_n, V_n \rangle (\theta_n - \Theta_n)^2 &\leq \|W_n\| \cdot \|V_n\| (\theta_n - \Theta_n)^2 \\ &\leq (1 + \|W_n\|^2) \|V_n\| (\theta_n - \Theta_n)^2.\end{aligned}$$

Ainsi, sous (H6)(a'), il existe c, d et $e > 0$ tels que :

$$\begin{aligned}\mathbb{E} \left[\|W_{n+1}\|^2 | T_n \right] &\leq \left(1 + 4a_n^2 \|V_n\|^4 + 2a_n \|V_n\| (\theta_n - \Theta_n)^2 \right) \|W_n\|^2 + \\ &\quad ca_n^2 + da_n^2 (\theta_n - \Theta_n)^2 + ea_n^2 (\theta_n - \Theta_n)^4 - \\ &\quad 2a_n \langle W_n, V_n V_n' W_n \rangle + 2a_n \|V_n\| (\theta_n - \Theta_n)^2.\end{aligned}$$

Comme sous (H3)(b) et (H6), $|\theta_n - \Theta_n| \rightarrow 0$ p.s. et $\sum_{n=1}^{+\infty} a_n |\theta_n - \Theta_n| < \infty$ p.s., alors, $\sum_{n=1}^{+\infty} a_n (\theta_n - \Theta_n)^2 < \infty$ p.s. et on peut appliquer le lemme de Robbins-Siegmund :

Il existe une v.a. $T \geq 0$ telle que : $\|W_n\|^2 \rightarrow T$ p.s. et $\sum_{n=1}^{+\infty} a_n \langle W_n, V_n V_n' W_n \rangle < \infty$ p.s.

b) On a :

$$\begin{aligned}W_{n+1} - W_n &= C_{n+1} - C_n \\ &= \Pi_{L_n} \left(C_n - a_n V_n \left(V_n' C_n - (Z_n - \Theta_n)^2 \right) \right) - C_n \\ &= \Pi_{L_n} \left(C_n - a_n V_n \left(V_n' C_n - (Z_n - \Theta_n)^2 \right) \right) - \Pi_{L_n}(C_n) \text{ car } C_n \in L_{n-1} \subset L_n.\end{aligned}$$

Donc :

$$\begin{aligned}\|W_{n+1} - W_n\| &\leq \| -a_n V_n (V_n' C_n - (Z_n - \Theta_n)^2) \| \\ &\leq \| -a_n V_n (V_n' W_n + \delta_n^2 - ((Z_n - \theta_n) + (\theta_n - \Theta_n))^2) \| \\ &\leq a_n \|V_n\|^2 \|W_n\| + a_n \|V_n\| \delta_n^2 + 2a_n \|V_n\| (Z_n - \theta_n)^2 + 2a_n \|V_n\| (\theta_n - \Theta_n)^2.\end{aligned}$$

Or,

$$\begin{aligned}
\mathbb{E} \left[\sum_{n=1}^{+\infty} a_n^2 \|V_n\|^2 (Z_n - \theta_n)^4 \right] &\leq c_1 \sum_{n=1}^{+\infty} a_n^2 \mathbb{E} \left[(Z_n - \theta_n)^4 \right] \text{ sous (H6)(a')} \\
&\leq c_1 \sum_{n=1}^{+\infty} a_n^2 \mathbb{E} \left[(\delta_n R_n)^4 \right] \\
&\leq c_2 \sum_{n=1}^{+\infty} a_n^2 \mathbb{E} [R_n^4] \text{ sous (H6)(a')} \\
&< \infty \text{ sous (H1)(c)}.
\end{aligned}$$

Donc : $\sum_{n=1}^{+\infty} a_n^2 \|V_n\|^2 (Z_n - \theta_n)^4 < \infty$ p.s. et par suite $a_n \|V_n\| (Z_n - \theta_n)^2 \rightarrow 0$ p.s.

D'autre part, comme $(\theta_n - \Theta_n) \rightarrow 0$ p.s., à partir d'un certain rang :

$$a_n \|V_n\| (\theta_n - \Theta_n)^2 \leq a_n \|V_n\| |\theta_n - \Theta_n|.$$

Or, $\sum_{n=1}^{+\infty} a_n |\theta_n - \Theta_n| < \infty$ p.s. donc $a_n |\theta_n - \Theta_n| \rightarrow 0$ p.s. et $a_n \|V_n\| (\theta_n - \Theta_n)^2 \rightarrow 0$ p.s.

Enfin, sous (H6)(a') et (H3)(b), comme $\|W_n\| \rightarrow T$ p.s., $a_n \|V_n\|^2 \|W_n\| \rightarrow 0$ p.s. et $a_n \|V_n\| \delta_n^2 \rightarrow 0$ p.s.

On en déduit que : $\|W_{n+1} - W_n\| \rightarrow 0$ p.s.

c) On raisonne à ω fixé, appartenant à l'intersection des ensembles de convergence p.s. définis. Supposons $T(\omega) \neq 0$.

On supprime dans la suite l'écriture de ω .

Comme $\|W_n\| \rightarrow \sqrt{T}$ p.s., $\exists 0 < \varepsilon_1 < 1$, $\exists N(\varepsilon_1) : \forall n > N(\varepsilon_1)$, $\varepsilon_1 < \|W_n\| < \frac{1}{\varepsilon_1}$.

Donc, sous (H6)(b), à partir d'un certain rang L , on a :

$$\langle W_{n_l}, \sum_{j \in I_l} V_j V_j' W_{n_l} \rangle \geq \lambda \varepsilon_1^2.$$

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

$$\langle W_{n_l}, V_{p_l} V_{p_l}' W_{n_l} \rangle \geq \frac{\lambda \varepsilon_1^2}{r}. \quad (1)$$

On considère la décomposition :

$$\langle W_{p_l}, V_{p_l} V'_{p_l} W_{p_l} \rangle = \langle W_{p_l} + W_{n_l}, V_{p_l} V'_{p_l} (W_{p_l} - W_{n_l}) \rangle + \langle W_{n_l}, V_{p_l} V'_{p_l} W_{n_l} \rangle. \quad (2)$$

Soit $\varepsilon > 0$ tel que $\varepsilon \leq \frac{\lambda \varepsilon_1^3}{4(r)^2 C^2}$, avec $C = \sup_n \|V_n\| < \infty$ sous (H6)(a').

A partir d'un certain rang : $\|W_{n+1} - W_n\| < \varepsilon$,

donc : $\|W_{p_l} - W_{n_l}\| < r\varepsilon$;

$$\begin{aligned} \left| \langle W_{p_l} + W_{n_l}, V_{p_l} V'_{p_l} (W_{p_l} - W_{n_l}) \rangle \right| &< \left(\|W_{p_l}\| + \|W_{n_l}\| \right) \|V_{p_l}\|^2 \|W_{p_l} - W_{n_l}\| \\ &< \frac{2}{\varepsilon_1} C^2 r \varepsilon \\ &\leq \frac{\lambda \varepsilon_1^2}{2r}. \end{aligned}$$

$$\text{Ainsi, } -\frac{\lambda \varepsilon_1^2}{2r} < \langle W_{p_l} + W_{n_l}, V_{p_l} V'_{p_l} (W_{p_l} - W_{n_l}) \rangle < +\frac{\lambda \varepsilon_1^2}{2r}. \quad (3)$$

On déduit de (1), (2) et (3) que :

$$\begin{aligned} \langle W_{p_l}, V_{p_l} V'_{p_l} W_{p_l} \rangle &> \frac{\lambda \varepsilon_1^2}{r} - \frac{\lambda \varepsilon_1^2}{2r} \\ &\geq \frac{\lambda \varepsilon_1^2}{2r}. \end{aligned}$$

Sous (H3)(b), on a alors :

$$\sum_{l=1}^{+\infty} a_{p_l} \langle W_{p_l}, V_{p_l} V'_{p_l} W_{p_l} \rangle = \infty.$$

Donc : $\sum_{n=1}^{+\infty} a_n \langle W_n, V_n V'_n W_n \rangle = \infty$; il y a contradiction avec la conclusion de la partie 3)a) de la démonstration.

Par conséquent, $T(\omega) = 0$ et, en réécrivant les i , $C_n^i \rightarrow \gamma^i$ p.s.

Par suite, $|D_n^i - \delta_n^i| \rightarrow 0$ p.s. et $\|D_n^{-1} - \Delta_n^{-1}\| \rightarrow 0$ p.s.

4) On montre que sous (H3)(b), (H6), (H7) et (H8), $\sum_{n=1}^{+\infty} a_n \|D_n^{-1} - \Delta_n^{-1}\| < \infty$ p.s.

a) On reprend les notations de la partie 3 de la démonstration et on supprime l'écriture des i .

D'après 3)a)iii) , on a :

$$\begin{aligned}
\mathbb{E}\left[\|W_{n+1}\|^2|T_n\right] &\leq \left(1 + 4a_n^2 \|V_n\|^4 + 2a_n \|V_n\| (\theta_n - \Theta_n)^2\right)\|W_n\|^2 + c a_n^2 + d a_n^2 (\theta_n - \Theta_n)^2 + \\
&\quad e a_n^2 (\theta_n - \Theta_n)^4 + f a_n (\theta_n - \Theta_n)^2 - 2a_n \langle W_n, V_n V_n' W_n \rangle \\
&\leq (1 + 4a_n^2 \|V_n\|^4)\|W_n\|^2 + g a_n (\theta_n - \Theta_n)^2 l_n^2 + c a_n^2 + d a_n^2 (\theta_n - \Theta_n)^2 + \\
&\quad h a_n^2 (\theta_n - \Theta_n)^2 k_n^2 + f a_n (\theta_n - \Theta_n)^2 - 2a_n \langle W_n, V_n V_n' W_n \rangle \\
&\leq (1 + 4a_n^2 \|V_n\|^4)\|W_n\|^2 + g_1 a_n (\theta_n - \Theta_n)^2 l_n^2 + c a_n^2 + e a_n^2 (\theta_n - \Theta_n)^2 k_n^2 - \\
&\quad 2a_n \langle W_n, V_n V_n' W_n \rangle \text{ car } l_n, k_n > 1 \text{ à partir d'un certain rang.}
\end{aligned}$$

En passant à l'espérance mathématique, on obtient :

$$\begin{aligned}
\mathbb{E}\left[\|W_{n+1}\|^2\right] &\leq \left(1 + 4a_n^2 \|V_n\|^4\right)\mathbb{E}\left[\|W_n\|^2\right] + g_1 a_n l_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] + c a_n^2 + e a_n^2 k_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] - \\
&\quad 2a_n \mathbb{E}\left[\langle W_n, V_n V_n' W_n \rangle\right].
\end{aligned}$$

i) 1^{er} cas : ($\frac{1}{2} < \alpha < 1$) ou ($\alpha = 1$ et $\frac{2\lambda a}{r} > 1$)

$$\sum_{n=1}^{+\infty} a_n l_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] \leq c \sum_{n=1}^{+\infty} \frac{1}{n^{2\alpha-2\gamma}} < \infty \text{ sous (H3)(b), (H7)(b) et (H8)(a)(ii).}$$

$$\sum_{n=1}^{+\infty} a_n^2 k_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] \leq c \sum_{n=1}^{+\infty} \frac{1}{n^{3\alpha-2\zeta}} < \infty \text{ sous (H3)(b), (H7)(a) et (H8)(a)(i).}$$

ii) 2^{ème} cas : ($\alpha = 1$ et $\frac{2\lambda a}{r} = 1$) :

Sous (H8)(a)(ii), $2 - 2\gamma > 1$, donc $\exists \varepsilon_1 > 0$, $2 - 2\gamma = 1 + \varepsilon_1$.

$$\text{Ainsi, } \sum_{n=1}^{+\infty} a_n l_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] \leq c \sum_{n=1}^{+\infty} \frac{\ln(n)}{n^{2-2\gamma}} \leq c \sum_{n=1}^{+\infty} \frac{\ln(n)}{n^{1+\varepsilon_1}}.$$

$$\text{Or, } \ln(n) < n^{\frac{\varepsilon_1}{2}} \text{ (à partir d'un certain rang), donc, } \sum_{n=1}^{+\infty} \frac{\ln(n)}{n^{1+\varepsilon_1}} < \sum_{n=1}^{+\infty} \frac{1}{n^{1+\frac{\varepsilon_1}{2}}} < \infty.$$

Sous (H8)(a)(i), $3 - 2\zeta > 1$, donc $\exists \varepsilon_2 > 0$, $3 - 2\zeta = 1 + \varepsilon_2$.

$$\text{Ainsi, } \sum_{n=1}^{+\infty} a_n^2 k_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] \leq c \sum_{n=1}^{+\infty} \frac{\ln(n)}{n^{3-2\zeta}} \leq c \sum_{n=1}^{+\infty} \frac{\ln(n)}{n^{1+\varepsilon_2}} < c \sum_{n=1}^{+\infty} \frac{1}{n^{1+\frac{\varepsilon_2}{2}}} < \infty.$$

iii) 3 ème cas : ($\alpha = 1$ et $\frac{2\lambda a}{r} < 1$) :

$$\sum_{n=1}^{+\infty} a_n l_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] \leq c \sum_{n=1}^{+\infty} \frac{1}{n^{1 + \frac{2\lambda a}{r} - 2\gamma}} < \infty \text{ sous (H8)(b)(ii).}$$

$$\sum_{n=1}^{+\infty} a_n^2 k_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] \leq c \sum_{n=1}^{+\infty} \frac{1}{n^{2 + \frac{2\lambda a}{r} - 2\zeta}} < \infty \text{ sous (H8)(b)(i).}$$

Donc, dans tous les cas, $\exists t \geq 0$, $\mathbb{E}[||W_n||^2] \rightarrow t$; $\sum_{n=1}^{+\infty} a_n \mathbb{E}[\langle W_n, V_n V_n' W_n \rangle] < \infty$.

b) On obtient alors les majorations :

$$\begin{aligned} \mathbb{E}[||W_{n+1}||^2] &\leq \mathbb{E}[||W_n||^2] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] + h_1 a_n^2 k_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] - \\ &\quad 2a_n \mathbb{E}[\langle W_n, V_n V_n' W_n \rangle] \end{aligned}$$

et

$$\begin{aligned} \mathbb{E}[||W_{n_l+1}||^2] &\leq \mathbb{E}[||W_{n_l}||^2] + c_1 \sum_{j \in I_l} a_j^2 + g_1 \sum_{j \in I_l} a_j l_j^2 \mathbb{E}[(\theta_j - \Theta_j)^2] + h_1 \sum_{j \in I_l} a_j^2 k_j^2 \mathbb{E}[(\theta_j - \Theta_j)^2] - \\ &\quad 2\mu_l \sum_{j \in I_l} \mathbb{E}[\langle W_j, V_j V_j' W_j \rangle] \text{ avec } \mu_l = \min_{j \in I_l} a_j. \end{aligned}$$

On décompose le dernier terme du second membre de la manière suivante :

$$\sum_{j \in I_l} \mathbb{E}[\langle W_j, V_j V_j' W_j \rangle] = \sum_{j \in I_l} \mathbb{E}[\langle W_{n_l}, V_j V_j' W_{n_l} \rangle] + \sum_{j \in I_l} \mathbb{E}[\langle W_j + W_{n_l}, V_j V_j' (W_j - W_{n_l}) \rangle]. \quad (4)$$

$$\text{Sous (H6)(b), } \sum_{j \in I_l} \mathbb{E}[\langle W_{n_l}, V_j V_j' W_{n_l} \rangle] \geq \lambda \mathbb{E}[||W_{n_l}||^2]. \quad (5)$$

On note $C_1 = \sup_n ||V_n||$.

$$\begin{aligned} \sum_{j \in I_l} \left| \mathbb{E}[\langle W_j + W_{n_l}, V_j V_j' (W_j - W_{n_l}) \rangle] \right| &\leq \sum_{j \in I_l} \mathbb{E}[||W_j + W_{n_l}|| \cdot ||V_j||^2 ||W_j - W_{n_l}||] \\ &\leq C_1^2 \sum_{j \in I_l} \left(\mathbb{E}[||W_j + W_{n_l}||^2] \right)^{\frac{1}{2}} \left(\mathbb{E}[||W_j - W_{n_l}||^2] \right)^{\frac{1}{2}} \\ &\leq C_2^2 \sum_{j \in I_l} \left(\mathbb{E}[||W_j - W_{n_l}||^2] \right)^{\frac{1}{2}} \text{ car } \mathbb{E}[||W_n||^2] \rightarrow t. \end{aligned}$$

Or, $W_j - W_{n_l} = \sum_{k=n_l}^{j-1} (W_{k+1} - W_k)$, donc :

$$\begin{aligned}
||W_j - W_{n_l}|| &\leq \sum_{k=n_l}^{j-1} ||W_{k+1} - W_k|| \\
&\leq \sum_{k=n_l}^{j-1} \left(||a_k V_k V'_k W_k|| + ||a_k V_k \delta_k^2|| + 2||a_k V_k (Z_k - \theta_k)^2|| + 2||a_k V_k (\theta_k - \Theta_k)^2|| \right) \\
&\leq \sum_{k=n_l}^{j-1} \left(a_k C^2 ||W_k|| + a_k C \delta_k^2 + 2a_k C (Z_k - \theta_k)^2 + 2a_k C (\theta_k - \Theta_k)^2 \right).
\end{aligned}$$

Alors,

$$\begin{aligned}
||W_j - W_{n_l}||^2 &\leq c_2 \sum_{k=n_l}^{j-1} \left(a_k^2 ||W_k||^2 + a_k^2 + a_k^2 (Z_k - \theta_k)^4 + a_k^2 (\theta_k - \Theta_k)^4 \right) \\
&\leq c_3 \sum_{k=n_l}^{j-1} \left(a_k^2 ||W_k||^2 + a_k^2 + a_k^2 (Z_k - \theta_k)^4 + a_k^2 k_k^2 (\theta_k - \Theta_k)^2 \right) \\
&\quad \text{car } |\theta_k - \Theta_k| = |\langle \beta - B_k, U_k \rangle| \leq k_k ||U_k||,
\end{aligned}$$

d'où, en passant à l'espérance, on obtient :

$$\begin{aligned}
\mathbb{E} [||W_j - W_{n_l}||^2] &\leq c_3 \sum_{k=n_l}^{j-1} \left(a_k^2 \mathbb{E} [||W_k||^2] + a_k^2 + a_k^2 \mathbb{E} [(Z_k - \theta_k)^4] + a_k^2 k_k^2 \mathbb{E} [(\theta_k - \Theta_k)^2] \right) \\
&\leq c_3 \sum_{k=n_l}^{j-1} \left(a_k^2 \mathbb{E} [||W_k||^2] + a_k^2 + a_k^2 \delta_k^4 \mathbb{E} [R_k^4] + a_k^2 k_k^2 \mathbb{E} [(\theta_k - \Theta_k)^2] \right) \\
&\leq c_4 \sum_{k=n_l}^{j-1} \left(a_k^2 + a_k^2 k_k^2 \mathbb{E} [(\theta_k - \Theta_k)^2] \right) \\
&\leq c_4 r \left(\max_{k \in I_l} a_k^2 + \max_{k \in I_l} a_k^2 k_k^2 \mathbb{E} [(\theta_k - \Theta_k)^2] \right)
\end{aligned}$$

puis, il vient que :

$$\left(\mathbb{E} [||W_j - W_{n_l}||^2] \right)^{\frac{1}{2}} \leq c_5 \left(\max_{k \in I_l} a_k + \max_{k \in I_l} a_k k_k \left(\mathbb{E} [(\theta_k - \Theta_k)^2] \right)^{\frac{1}{2}} \right).$$

Ainsi,

$$\sum_{j \in I_l} \left| \mathbb{E} [\langle W_j + W_{n_l}, V_j V'_j (W_j - W_{n_l}) \rangle] \right| \leq c_6 \left(\max_{k \in I_l} a_k + \max_{k \in I_l} a_k k_k \left(\mathbb{E} [(\theta_k - \Theta_k)^2] \right)^{\frac{1}{2}} \right).$$

Donc :

$$\begin{aligned}
\mathbb{E}\left[\|W_{n_l+1}\|^2\right] &\leq \mathbb{E}\left[\|W_{n_l}\|^2\right] + c_1 r \max_{k \in I_l} a_k^2 + g_1 r \max_{k \in I_l} a_k l_k^2 \mathbb{E}\left[(\theta_k - \Theta_k)^2\right] + h_1 r \max_{k \in I_l} a_k^2 k_k^2 \mathbb{E}\left[(\theta_k - \Theta_k)^2\right] \\
&\quad - 2\lambda \mu_l \mathbb{E}\left[\|W_{n_l}\|^2\right] + 2c_6 \mu_l \left(\max_{k \in I_l} a_k + \max_{k \in I_l} a_k k_k \left(\mathbb{E}\left[(\theta_k - \Theta_k)^2\right] \right)^{\frac{1}{2}} \right) \\
&\leq (1 - 2\lambda \mu_l) \mathbb{E}\left[\|W_{n_l}\|^2\right] + c_7 \max_{k \in I_l} a_k^2 + g_2 \max_{k \in I_l} a_k l_k^2 \mathbb{E}\left[(\theta_k - \Theta_k)^2\right] + \\
&\quad h_2 \max_{k \in I_l} a_k^2 k_k^2 \mathbb{E}\left[(\theta_k - \Theta_k)^2\right] + c_8 \mu_l \max_{k \in I_l} a_k k_k \left(\mathbb{E}\left[(\theta_k - \Theta_k)^2\right] \right)^{\frac{1}{2}} \\
&\leq (1 - 2\frac{\lambda a}{r^\alpha} \frac{1}{l^\alpha}) \mathbb{E}\left[\|W_{n_l}\|^2\right] + c_7 \max_{k \in I_l} a_k^2 + g_2 \max_{k \in I_l} a_k l_k^2 \mathbb{E}\left[(\theta_k - \Theta_k)^2\right] + \\
&\quad h_2 \max_{k \in I_l} a_k^2 k_k^2 \mathbb{E}\left[(\theta_k - \Theta_k)^2\right] + c_8 \mu_l \max_{k \in I_l} a_k k_k \left(\mathbb{E}\left[(\theta_k - \Theta_k)^2\right] \right)^{\frac{1}{2}}.
\end{aligned}$$

i) 1^{er} cas : ($\frac{1}{2} < \alpha < 1$) ou ($\alpha = 1$ et $\frac{2\lambda a}{r} > 1$).

Alors, $\mathbb{E}\left[(\theta_n - \Theta_n)^2\right] = O(\frac{1}{n^\alpha})$.

$$\mathbb{E}\left[\|W_{n_l+1}\|^2\right] \leq (1 - 2\frac{\lambda a}{r^\alpha} \frac{1}{l^\alpha}) \mathbb{E}\left[\|W_{n_l}\|^2\right] + c_9 \frac{1}{l^{2\alpha}} + g_3 \frac{1}{l^{2\alpha-2\gamma}} + h_3 \frac{1}{l^{3\alpha-2\zeta}} + c_{10} \frac{1}{l^{\frac{3}{2}\alpha-\zeta}}$$

Soit $\delta = \min(\alpha - 2\gamma; 2\alpha - 2\zeta; \frac{3}{2}\alpha - \zeta)$.

D'une part, $\delta < \alpha$ car γ positif.

D'autre part, $\delta > 0$. En effet,

$$\cdot \alpha - 2\gamma > \alpha - 3\alpha + 2 = 2 - 2\alpha > 0,$$

$$\cdot 2\alpha - 2\zeta > 2\alpha - (4\alpha - 2) = 2 - 2\alpha > 0,$$

$$\cdot \frac{3}{2}\alpha - \zeta > \frac{3}{2}\alpha - (2\alpha - 1) = 1 - \frac{\alpha}{2} > 0,$$

$$\mathbb{E}\left[\|W_{n_l+1}\|^2\right] \leq (1 - 2\frac{\lambda a}{r^\alpha} \frac{1}{l^\alpha}) \mathbb{E}\left[\|W_{n_l}\|^2\right] + c_{11} \frac{1}{l^{\alpha+\delta}}.$$

Notons que pour $\alpha = 1$, $\frac{2\lambda a}{r} > 1 \geq \alpha > \delta$.

Ainsi, par application du lemme 3.10 avec $a_l = \frac{a}{l^\alpha}$, $h_l = \frac{1}{l^\delta}$, $b_l = \mathbb{E}\left[\|W_{n_l}\|^2\right]$, λ (du lemme 3.10) $= 2\frac{\lambda}{r^\alpha}$, $\mu = c_{11}$, on obtient : $\overline{\lim} l^\delta \mathbb{E}\left[\|W_{n_l}\|^2\right] < \infty$.

Donc, à partir d'un certain rang, $\exists d > 0$, $\mathbb{E}\left[\|W_{n_l}\|^2\right] < \frac{d}{l^\delta}$.

Or, on a (c.f. partie 4)b) de la démonstration) :

$$\begin{aligned}
\mathbb{E}\left[||W_{n+1}||^2\right] &\leq \mathbb{E}\left[||W_n||^2\right] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] + h_1 a_n^2 k_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] - \\
&\quad 2a_n \mathbb{E}\left[\langle W_n, V_n V_n' W_n \rangle\right] \\
&\leq E\left[||W_n||^2\right] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] + h_1 a_n^2 k_n^2 \mathbb{E}\left[(\theta_n - \Theta_n)^2\right] \\
&\leq E\left[||W_n||^2\right] + c_{12} \frac{1}{n^{2\alpha}} + g_4 \frac{1}{n^{2\alpha-2\gamma}} + h_4 \frac{1}{n^{3\alpha-2\zeta}} \\
&\leq E\left[||W_n||^2\right] + c_{13} \frac{1}{n^{\alpha+\delta_1}} \text{ avec } \delta_1 = \min(\alpha - 2\gamma; 2\alpha - 2\zeta) \geq \delta.
\end{aligned}$$

Pour $n \in I_l$, on peut alors écrire :

$$\begin{aligned}
\mathbb{E}\left[||W_n||^2\right] &\leq \mathbb{E}\left[||W_{n_l}||^2\right] + r c_{13} \frac{1}{n_l^{\alpha+\delta_1}} \\
&\leq \mathbb{E}\left[||W_{n_l}||^2\right] + c_{14} \frac{1}{l^{\alpha+\delta_1}}.
\end{aligned}$$

Donc :

$$\begin{aligned}
n^\delta \mathbb{E}\left[||W_n||^2\right] &\leq n^\delta \mathbb{E}\left[||W_{n_l}||^2\right] + n^\delta c_{14} \frac{1}{l^{\alpha+\delta_1}} \\
&\leq r^\delta l^\delta \mathbb{E}\left[||W_{n_l}||^2\right] + r^\delta l^\delta c_{14} \frac{1}{l^{\alpha+\delta_1}} \\
&\leq r^\delta l^\delta \mathbb{E}\left[||W_{n_l}||^2\right] + r^\delta c_{14} \frac{1}{l^{\alpha+\delta_1-\delta}} \\
&\leq r^\delta d + r^\delta c_{14} \frac{1}{l^\alpha}.
\end{aligned}$$

et $\overline{\lim} n^\delta \mathbb{E}\left[||W_n||^2\right] < \infty$.

dont on déduit que : $\mathbb{E}\left[||W_n||^2\right] \leq \frac{c_{15}}{n^\delta}$.

Ainsi, $\sum_{n=1}^{+\infty} a_n \mathbb{E}\left[||W_n||\right] \leq c_{15} \sum_{n=1}^{+\infty} \frac{1}{n^{\alpha+\frac{1}{2}\delta}}$.

Or, $\alpha + \frac{1}{2}\delta > 1$. En effet, $\alpha + \frac{1}{2}\delta = \min(\frac{3\alpha}{2} - \gamma; 2\alpha - \zeta; \frac{7\alpha}{4} - \frac{\zeta}{2})$ et :

$$\cdot \frac{3\alpha}{2} - \gamma > \frac{3\alpha}{2} - (1 - \frac{3\alpha}{2}) = 1 \text{ sous (H7)(b) ou (H8)(a)(ii),}$$

$$\cdot 2\alpha - \zeta > 2\alpha - (1 - 2\alpha) = 1 \text{ sous (H7)(a) ou (H8)(a)(i),}$$

$$\cdot \frac{7\alpha}{4} - \frac{\zeta}{2} > \frac{7\alpha}{4} - \alpha + \frac{1}{2} \text{ sous (H7)(a) ou (H8)(a)(i)}$$

$$= \frac{3\alpha}{4} + \frac{1}{2}$$

$$> 1 \text{ sous (H3)(b).}$$

Donc : $\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[||W_n|| \right] < \infty$.

ii) 2^{ème} cas : ($\alpha = 1$ et $\frac{2\lambda a}{r} = 1$).

Alors, $\mathbb{E} \left[(\theta_n - \Theta_n)^2 \right] = O\left(\frac{\ln(n)}{n}\right)$.

$$\mathbb{E} \left[||W_{n_l+1}||^2 \right] \leq (1 - \frac{1}{l}) \mathbb{E} \left[||W_{n_l}||^2 \right] + c_{16} \frac{1}{l^2} + g_5 \frac{\ln(l)}{l^{2-2\gamma}} + h_5 \frac{\ln(l)}{l^{3-2\zeta}} + c_{17} \frac{\ln(l)}{l^{\frac{3}{2}-\zeta}}.$$

Soit $\delta' = \min(1 - 2\gamma; 2 - 2\zeta; \frac{3}{2} - \zeta)$.

D'une part, $\delta' < 1$ car γ positif.

D'autre part, $\delta' > 0$ sous (H8)(a).

Alors, $\exists \varepsilon : \delta' > \varepsilon > 0$,

$$\begin{aligned} \mathbb{E} \left[||W_{n_l+1}||^2 \right] &\leq (1 - \frac{1}{l}) \mathbb{E} \left[||W_{n_l}||^2 \right] + c_{18} \frac{\ln(l)}{l^{1+\delta'}} \\ &< (1 - \frac{1}{l}) \mathbb{E} \left[||W_{n_l}||^2 \right] + c_{18} \frac{1}{l^{1+\delta'-\varepsilon}} \text{ (à partir d'un certain rang)} \end{aligned}$$

et $1 > \delta' - \varepsilon$, ce qui nous permet d'appliquer le lemme 3.10 avec $a_l = \frac{a}{l}$, $h_l = \frac{1}{l^{\delta'-\varepsilon}}$, $b_l = \mathbb{E} \left[||W_{n_l}||^2 \right]$, $\lambda = \frac{1}{a}$, $\mu = c_{18}$.

On obtient : $\overline{\lim} l^{\delta'-\varepsilon} \mathbb{E} \left[||W_{n_l}||^2 \right] < \infty$.

Donc, à partir d'un certain rang, $\exists d > 0$, $\mathbb{E} \left[||W_{n_l}||^2 \right] < \frac{d}{l^{\delta'-\varepsilon}}$.

Or, on a (c.f. partie 4)b) de la démonstration) :

$$\begin{aligned} \mathbb{E} \left[||W_{n+1}||^2 \right] &\leq \mathbb{E} \left[||W_n||^2 \right] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E} \left[(\theta_n - \Theta_n)^2 \right] + h_1 a_n^2 k_n^2 \mathbb{E} \left[(\theta_n - \Theta_n)^2 \right] - \\ &\quad 2a_n \mathbb{E} \left[\langle W_n, V_n V_n' W_n \rangle \right] \\ &\leq E \left[||W_n||^2 \right] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E} \left[(\theta_n - \Theta_n)^2 \right] + h_1 a_n^2 k_n^2 \mathbb{E} \left[(\theta_n - \Theta_n)^2 \right] \\ &\leq E \left[||W_n||^2 \right] + c_{19} \frac{1}{n^2} + g_6 \frac{\ln(n)}{n^{2-2\gamma}} + h_6 \frac{\ln(n)}{n^{3-2\zeta}} \\ &\leq E \left[||W_n||^2 \right] + c_{20} \frac{\ln(n)}{n^{1+\delta'_1}} \text{ avec } \delta'_1 = \min(1 - 2\gamma; 2 - 2\zeta) \geq \delta'. \end{aligned}$$

Pour $n \in I_l$, on peut alors écrire :

$$\begin{aligned}\mathbb{E}\left[||W_n||^2\right] &\leq \mathbb{E}\left[||W_{n_l}||^2\right] + c_{21} \frac{\ln(n_{l+1}-1)}{n_l^{1+\delta'_1}} \\ &\leq \mathbb{E}\left[||W_{n_l}||^2\right] + c_{22} \frac{\ln(lr)}{l^{1+\delta'_1}}.\end{aligned}$$

Donc :

$$\begin{aligned}n^{\delta' - \varepsilon} \mathbb{E}\left[||W_n||^2\right] &\leq n^{\delta' - \varepsilon} \mathbb{E}\left[||W_{n_l}||^2\right] + n^{\delta' - \varepsilon} c_{22} \frac{\ln(lr)}{l^{1+\delta'_1}} \\ &\leq r^{\delta' - \varepsilon} l^{\delta' - \varepsilon} \mathbb{E}\left[||W_{n_l}||^2\right] + c_{23} \frac{\ln(lr)}{l^{1+\delta'_1 - \delta' + \varepsilon}} \\ &\leq r^{\delta'} l^{\delta' - \varepsilon} \mathbb{E}\left[||W_{n_l}||^2\right] + c_{24} \frac{1}{l^{1+\delta'_1 - \delta'}} \\ &\leq r^{\delta'} d + c_{25} \frac{1}{l}\end{aligned}$$

dont on déduit que : $\mathbb{E}\left[||W_n||^2\right] \leq c_{25} \frac{1}{n^{\delta' - \varepsilon}}$.

Ainsi, $\sum_{n=1}^{+\infty} a_n \mathbb{E}\left[||W_n||\right] \leq c_{25} \sum_{n=1}^{+\infty} \frac{1}{n^{1 + \frac{\delta' - \varepsilon}{2}}} < \infty$ car $\varepsilon < \delta'$.

iii) 3 ème cas : ($\alpha = 1$ et $\frac{2\lambda a}{r} < 1$).

Alors, $\mathbb{E}\left[(\theta_n - \Theta_n)^2\right] = O\left(\frac{1}{n^{\frac{2\lambda a}{r}}}\right)$.

$$\mathbb{E}\left[||W_{n_{l+1}}||^2\right] \leq (1 - 2\frac{\lambda a}{r} \frac{1}{l}) \mathbb{E}\left[||W_{n_l}||^2\right] + c_{26} \frac{1}{l^2} + g_7 \frac{1}{l^{1 + \frac{2\lambda a}{r} - 2\gamma}} + h_7 \frac{1}{l^{2 + \frac{2\lambda a}{r} - 2\zeta}} + c_{27} \frac{1}{l^{2 + \frac{\lambda a}{r} - \zeta}}.$$

Soit $\delta'' = \min\left(\frac{2\lambda a}{r} - 2\gamma; 1 + \frac{2\lambda a}{r} - 2\zeta; 1 + \frac{\lambda a}{r} - \zeta\right)$.

D'une part, $\delta'' < \frac{2\lambda a}{r} < 1$ car γ positif.

D'autre part, $\delta'' > 0$ sous (H8)(b).

$$\mathbb{E}\left[||W_{n_{l+1}}||^2\right] \leq (1 - 2\frac{\lambda a}{r} \frac{1}{l}) \mathbb{E}\left[||W_{n_l}||^2\right] + c_{28} \frac{1}{l^{1+\delta''}}.$$

Comme $\delta'' < \frac{2\lambda a}{r}$, on peut appliquer le lemme 3.10 avec $a_l = \frac{a}{l}$, $h_l = \frac{1}{l^{\delta''}}$, $b_l = \mathbb{E}\left[||W_{n_l}||^2\right]$, $\lambda = \frac{2\lambda}{r}$, $\mu = c_{28}$.

On obtient : $\overline{\lim} l^{\delta''} \mathbb{E}\left[||W_{n_l}||^2\right] < \infty$.

Donc, à partir d'un certain rang, $\exists d > 0$, $\mathbb{E}\left[||W_{n_l}||^2\right] < \frac{d}{l^{\delta''}}$.

Or, on a (c.f. partie 4)b) de la démonstration) :

$$\begin{aligned}
\mathbb{E}[||W_{n+1}||^2] &\leq \mathbb{E}[||W_n||^2] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] + h_1 a_n^2 k_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] - \\
&\quad 2a_n \mathbb{E}[\langle W_n, V_n V_n' W_n \rangle] \\
&\leq E[||W_n||^2] + c_1 a_n^2 + g_1 a_n l_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] + h_1 a_n^2 k_n^2 \mathbb{E}[(\theta_n - \Theta_n)^2] \\
&\leq E[||W_n||^2] + c_{29} \frac{1}{n^2} + g_8 \frac{1}{n^{1+\frac{2\lambda a}{r}-2\gamma}} + h_8 \frac{1}{n^{2+\frac{2\lambda a}{r}-2\zeta}} \\
&\leq E[||W_n||^2] + c_{30} \frac{1}{n^{1+\delta_1''}}.
\end{aligned}$$

avec $\delta_1'' = \min(1; \frac{2\lambda a}{r} - 2\gamma; 1 + \frac{2\lambda a}{r} - 2\zeta) = \min(\frac{2\lambda a}{r} - 2\gamma; 1 + \frac{2\lambda a}{r} - 2\zeta)$ car $\frac{2\lambda a}{r} < 1$.

On a aussi : $\delta_1'' \geq \delta''$.

Pour $n \in I_l$, on peut alors écrire :

$$\begin{aligned}
\mathbb{E}[||W_n||^2] &\leq \mathbb{E}[||W_{n_l}||^2] + r c_{30} \frac{1}{n_l^{1+\delta_1''}} \\
&\leq \mathbb{E}[||W_{n_l}||^2] + c_{31} \frac{1}{n_l^{1+\delta_1''}}.
\end{aligned}$$

Donc :

$$\begin{aligned}
n^{\delta''} \mathbb{E}[||W_n||^2] &\leq n^{\delta''} \mathbb{E}[||W_{n_l}||^2] + n^{\delta''} c_{31} \frac{1}{l^{1+\delta_1''}} \\
&\leq r^{\delta''} l^{\delta''} \mathbb{E}[||W_{n_l}||^2] + r^{\delta''} l^{\delta''} c_{31} \frac{1}{l^{1+\delta_1''}} \\
&\leq r^{\delta''} l^{\delta''} \mathbb{E}[||W_{n_l}||^2] + r^{\delta''} c_{31} \frac{1}{l^{1+\delta_1''-\delta''}} \\
&\leq r^{\delta''} d + r^{\delta''} c_{31} \frac{1}{l}
\end{aligned}$$

dont on déduit que : $\mathbb{E}[||W_n||^2] \leq c_{32} \frac{1}{n^{\delta''}}$.

Ainsi, $\sum_{n=1}^{+\infty} a_n \mathbb{E}[||W_n||] \leq c_{32} \sum_{n=1}^{+\infty} \frac{1}{n^{1+\frac{1}{2}\delta''}} < \infty$ puisque $\delta'' > 0$.

Donc, dans tous les cas, $\sum_{n=1}^{+\infty} a_n ||W_n|| < \infty$ p.s., ou en réécrivant les i , $\sum_{n=1}^{+\infty} a_n ||C_n^i - \gamma^i|| < \infty$ p.s.

c) On raisonne à ω fixé.

D'une part, $\sum_{n=1}^{+\infty} a_n |(D_n^i)^2 - (\delta_n^i)^2| = \sum_{n=1}^{+\infty} a_n |\langle C_n^i - \gamma^i, V_n^i \rangle| < C \sum_{n=1}^{+\infty} a_n \|C_n^i - \gamma^i\| < \infty$ sous (H6)(a').

D'autre part, d'après la partie 3) de la démonstration, $D_n^i - \delta_n^i \rightarrow 0$ p.s.

Soit $\varepsilon : 0 < \varepsilon < \inf_n \min_i \delta_n^i = d$.

A partir d'un certain rang, on a : $D_n^i > d - \varepsilon$.

Ainsi, $\sum_{n=1}^{+\infty} a_n |D_n^i - \delta_n^i| = \sum_{n=1}^{+\infty} a_n \frac{|(D_n^i)^2 - (\delta_n^i)^2|}{(D_n^i + \delta_n^i)} \leq \sum_{n=1}^{+\infty} a_n \frac{|(D_n^i)^2 - (\delta_n^i)^2|}{2d - \varepsilon} < \infty$.

Ensuite, $\sum_{n=1}^{+\infty} a_n \left| \frac{1}{D_n^i} - \frac{1}{\delta_n^i} \right| = \sum_{n=1}^{+\infty} a_n \frac{|D_n^i - \delta_n^i|}{D_n^i \delta_n^i} \leq \sum_{n=1}^{+\infty} a_n \frac{|D_n^i - \delta_n^i|}{(d - \varepsilon)d} < \infty$.

Enfin,

$$\begin{aligned} \sum_{n=1}^{+\infty} a_n \|D_n^{-1} - \Delta_n^{-1}\| &= \sum_{n=1}^{+\infty} a_n \left(\sum_{i=1}^p \left(\frac{1}{D_n^i} - \frac{1}{\delta_n^i} \right)^2 \right)^{\frac{1}{2}} \text{ en prenant la norme de Hilbert - Schmidt} \\ &\leq \sum_{n=1}^{+\infty} a_n \left(\sum_{i=1}^p \left| \frac{1}{D_n^i} - \frac{1}{\delta_n^i} \right| \right) \\ &\leq \sum_{i=1}^p \left(\sum_{n=1}^{+\infty} a_n \left| \frac{1}{D_n^i} - \frac{1}{\delta_n^i} \right| \right) \\ &< \infty. \end{aligned}$$

Démonstration du théorème 4.3 :

1) Montrons d'abord que $M_n \rightarrow M$ p.s.

Pour ce faire, montrons que, pour $k = 1, \dots, q$, $M_n^k \rightarrow (C^k)^{-1}$ p.s.

Dans la suite, on supprime l'écriture de k .

Sous (H1)(a), on peut écrire :

$$\begin{aligned} M_{n+1} - C^{-1} &= M_n - C^{-1} - a_n \left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} M_n - I \right) \\ &= (M_n - C^{-1}) - a_n \left(C(M_n - C^{-1}) + \left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} - C \right) M_n \right) \\ &= (I - a_n C)(M_n - C^{-1}) - a_n \left(\left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} - C \right) M_n + D_n^{-1} (\Theta_n - \Theta_n) Z_n' D_n^{-1} M_n \right). \end{aligned}$$

D'où,

$$\begin{aligned} \|M_{n+1} - C^{-1}\|^2 &= \|(I - a_n C)(M_n - C^{-1})\|^2 + \\ &\quad a_n^2 \left\| \left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} - C \right) M_n + D_n^{-1} (\Theta_n - \Theta_n) Z_n' D_n^{-1} M_n \right\|^2 - \\ &\quad 2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} - C \right) M_n + D_n^{-1} (\Theta_n - \Theta_n) Z_n' D_n^{-1} M_n \right\rangle. \end{aligned}$$

puis, en prenant l'espérance mathématique conditionnellement à T_n , tribu du passé au temps n , on obtient :

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 | T_n \right] &= \|(I - a_n C)(M_n - C^{-1})\|^2 + \\ &\quad a_n^2 \mathbb{E} \left[\left\| \left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} - C \right) M_n + D_n^{-1} (\Theta_n - \Theta_n) Z_n' D_n^{-1} M_n \right\|^2 \middle| T_n \right] - \\ &\quad 2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \mathbb{E} \left[\left(D_n^{-1} (Z_n - \Theta_n) Z_n' D_n^{-1} - C \right) M_n + \right. \right. \\ &\quad \left. \left. D_n^{-1} (\Theta_n - \Theta_n) Z_n' D_n^{-1} M_n \middle| T_n \right] \right\rangle. \end{aligned}$$

a) Pour le premier terme du second membre,

$$\begin{aligned} \|(I - a_n C)(M_n - C^{-1})\|^2 &= \|M_n - C^{-1}\|^2 - 2a_n \langle M_n - C^{-1}, C(M_n - C^{-1}) \rangle + a_n^2 \|C(M_n - C^{-1})\|^2 \\ &\leq (1 + a_n^2 \|C\|^2) \|M_n - C^{-1}\|^2 - 2a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2. \end{aligned}$$

b) Pour le troisième terme du second membre,

$$\begin{aligned} &-2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \mathbb{E} \left[\left(D_n^{-1} (Z_n - \theta_n) Z_n' D_n^{-1} - C \right) M_n + D_n^{-1} (\theta_n - \Theta_n) Z_n' D_n^{-1} M_n \middle| T_n \right] \right\rangle \\ &= -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), (D_n^{-1} \mathbb{E}[(Z_n - \theta_n) Z_n' | T_n] D_n^{-1} - C) M_n + D_n^{-1} (\theta_n - \Theta_n) \mathbb{E}[Z_n' | T_n] D_n^{-1} M_n \right\rangle \\ &\quad \text{sous (H5)(a) et (H5')(a)} \\ &= -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(D_n^{-1} \mathbb{E}[(Z_n - \theta_n) Z_n'] D_n^{-1} - C \right) M_n + D_n^{-1} (\theta_n - \Theta_n) \mathbb{E}[Z_n'] D_n^{-1} M_n \right\rangle \\ &= -2a_n \left\langle (I - a_n C)(M_n - C^{-1}), \left(D_n^{-1} \text{Cov}(Z_n) D_n^{-1} - C \right) M_n + D_n^{-1} (\theta_n - \Theta_n) (\theta_n - \Theta_n)' D_n^{-1} M_n \right\rangle \\ &\leq 2a_n \|I - a_n C\| \cdot \|M_n - C^{-1}\| \cdot \left(\|D_n^{-1} \text{Cov}(Z_n) D_n^{-1} - C\| + \|D_n^{-1} (\theta_n - \Theta_n) (\theta_n - \Theta_n)' D_n^{-1}\| \right) \cdot \|M_n\|. \end{aligned}$$

Comme $a_n \rightarrow 0$ et C définie positive, à partir d'un certain rang, $\|I - a_n C\| \leq 1$; on continue la majoration :

$$\leq 2a_n \|M_n - C^{-1}\| \cdot \left(\|D_n^{-1} \text{Cov}(Z_n) D_n^{-1} - C\| + \|D_n^{-1}\|^2 \cdot \|\theta_n - \Theta_n\| \cdot \|\theta_n - \Theta_n\| \right) \cdot \|M_n\|$$

i) Or,

$$\begin{aligned} \|M_n - C^{-1}\| \cdot \|M_n\| &= \|M_n - C^{-1}\| \cdot \|(M_n - C^{-1}) + C^{-1}\| \\ &\leq \|M_n - C^{-1}\| \cdot (\|M_n - C^{-1}\| + \|C^{-1}\|) \\ &\leq c_1 (\|M_n - C^{-1}\|^2 + \|M_n - C^{-1}\|) \\ &\leq c_2 (\|M_n - C^{-1}\|^2 + 1) \text{ puisque } \|M_n - C^{-1}\| \leq \|M_n - C^{-1}\|^2 + 1. \end{aligned}$$

ii) D'autre part, on a la décomposition :

$$\begin{aligned}
D_n^{-1}Cov(Z_n)D_n^{-1} - C &= D_n^{-1}Cov(Z_n)D_n^{-1} - \Delta_n^{-1}Cov(Z_n)\Delta_n^{-1} \\
&= (D_n^{-1} - \Delta_n^{-1})Cov(Z_n)\Delta_n^{-1} + (D_n^{-1} - \Delta_n^{-1})Cov(Z_n)(D_n^{-1} - \Delta_n^{-1}) + \\
&\quad \Delta_n^{-1}Cov(Z_n)(D_n^{-1} - \Delta_n^{-1}) \\
&= (D_n^{-1} - \Delta_n^{-1})\Delta_n C + (D_n^{-1} - \Delta_n^{-1})\Delta_n C \Delta_n (D_n^{-1} - \Delta_n^{-1}) + C \Delta_n (D_n^{-1} - \Delta_n^{-1}) \\
&\quad \text{en tenant compte du fait que } Cov(Z_n) = \Delta_n C \Delta_n,
\end{aligned}$$

dont on déduit que :

$$\begin{aligned}
\|D_n^{-1}Cov(Z_n)D_n^{-1} - C\| &\leq \| (D_n^{-1} - \Delta_n^{-1})\Delta_n C \| + \| (D_n^{-1} - \Delta_n^{-1})\Delta_n C \Delta_n (D_n^{-1} - \Delta_n^{-1}) \| + \| C \Delta_n (D_n^{-1} - \Delta_n^{-1}) \| \\
&\leq 2\|D_n^{-1} - \Delta_n^{-1}\| \cdot \|\Delta_n\| \cdot \|C\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 \cdot \|\Delta_n\|^2 \|C\| \\
&\leq c_3 \left(\|D_n^{-1} - \Delta_n^{-1}\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 \right) \text{ sous (H2') (a) .}
\end{aligned}$$

iii) Enfin,

$$\begin{aligned}
\|D_n^{-1}\|^2 &= \|(D_n^{-1} - \Delta_n^{-1}) + \Delta_n^{-1}\|^2 \\
&\leq 2 \left(\|D_n^{-1} - \Delta_n^{-1}\|^2 + \|\Delta_n^{-1}\|^2 \right) \\
&\leq c_3 \left(\|D_n^{-1} - \Delta_n^{-1}\|^2 + 1 \right) \text{ sous (H2') (b).}
\end{aligned}$$

De i), ii), iii), on déduit que :

$$\begin{aligned}
&2a_n \|M_n - C^{-1}\| \cdot (\|D_n^{-1}Cov(Z_n)D_n^{-1} - C\| + \|D_n^{-1}\|^2 \cdot \|\theta_n - \Theta_n\| \cdot \|\theta_n\|) \cdot \|M_n\| \\
&\leq c_4 a_n \left(\|D_n^{-1} - \Delta_n^{-1}\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^2 \cdot \|\theta_n - \Theta_n\| \cdot \|\theta_n\| + \right. \\
&\quad \left. \|\theta_n - \Theta_n\| \cdot \|\theta_n\| \right) \cdot (\|M_n - C^{-1}\|^2 + 1) \\
&\leq c_5 a_n \left(\|D_n^{-1} - \Delta_n^{-1}\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^2 \cdot \|\theta_n - \Theta_n\| + \|\theta_n - \Theta_n\| \right) \cdot (\|M_n - C^{-1}\|^2 + 1) \\
&\quad \text{sous (H6)(a).}
\end{aligned}$$

c) Enfin, pour le deuxième terme du second membre,

$$\begin{aligned}
& a_n^2 \mathbb{E} \left[\left\| \left(D_n^{-1} (Z_n - \theta_n) Z_n' D_n^{-1} - C \right) M_n + D_n^{-1} (\theta_n - \Theta_n) Z_n' D_n^{-1} M_n \right\|^2 \middle| T_n \right] \\
& \leq a_n^2 \mathbb{E} \left[\left\| \left(D_n^{-1} (Z_n - \theta_n) Z_n' D_n^{-1} - C \right) + D_n^{-1} (\theta_n - \Theta_n) Z_n' D_n^{-1} \right\|^2 \middle| T_n \right] \cdot \|M_n\|^2 \\
& \leq 3a_n^2 \mathbb{E} \left[\|D_n^{-1} (Z_n - \theta_n) Z_n' D_n^{-1}\|^2 + \|C\|^2 + \|D_n^{-1} (\theta_n - \Theta_n) Z_n' D_n^{-1}\|^2 \middle| T_n \right] \cdot \|M_n\|^2 \\
& \leq 3a_n^2 \left(\mathbb{E} \left[\|D_n^{-1} (Z_n - \theta_n) Z_n' D_n^{-1}\|^2 \middle| T_n \right] + \|C\|^2 + \mathbb{E} \left[\|D_n^{-1} (\theta_n - \Theta_n) Z_n' D_n^{-1}\|^2 \middle| T_n \right] \right) \cdot \|M_n\|^2 \\
& \leq 3a_n^2 \left(\|D_n^{-1}\|^4 \mathbb{E} \left[\|(Z_n - \theta_n) Z_n'\|^2 \right] + \|C\|^2 + \|D_n^{-1}\|^4 \|\theta_n - \Theta_n\|^2 \mathbb{E} \left[\|Z_n\|^2 \right] \right) \cdot \|M_n\|^2
\end{aligned}$$

D'une part,

$$\begin{aligned}
\mathbb{E} \left[\|(Z_n - \theta_n) Z_n'\|^2 \right] &= \mathbb{E} \left[\|(Z_n - \theta_n) ((Z_n - \theta_n) + \theta_n)'\|^2 \right] \\
&= \mathbb{E} \left[\|\Delta_n R_n (\Delta_n R_n + \theta_n)'\|^2 \right] \\
&\leq 2 \left(\mathbb{E} \left[\|\Delta_n R_n\|^4 + \|\Delta_n R_n \theta_n'\|^2 \right] \right) \\
&\leq 2 \left(\|\Delta_n\|^4 \mathbb{E} \left[\|R_n\|^4 \right] + \|\Delta_n\|^2 \|\theta_n\|^2 \mathbb{E} \left[\|R_n\|^2 \right] \right) \\
&\leq c_6 \left(\mathbb{E} \left[\|R_n\|^4 \right] + \mathbb{E} \left[\|R_n\|^2 \right] \right) \text{ sous (H6)(a) et (H2')(a)} \\
&= c_6 (\mathbb{E} \left[\|R\|^4 \right] + \mathbb{E} \left[\|R\|^2 \right]) \text{ puisque les } R_n \text{ sont i.i.d.} \\
&\leq c_7 \text{ sous (H1)(b)}
\end{aligned}$$

et

$$\begin{aligned}
\mathbb{E}\left[\|Z_n\|^2\right] &= \mathbb{E}\left[\|(Z_n - \theta_n) + \theta_n\|^2\right] \\
&= \mathbb{E}\left[\|\Delta_n R_n + \theta_n\|^2\right] \\
&\leq 2\left(\|\Delta_n\|^2 \mathbb{E}\left[\|R_n\|^2\right] + \|\theta_n\|^2\right) \\
&\leq c_8\left(\mathbb{E}\left[\|R_n\|^2\right] + 1\right) \text{ sous (H6)(a) et (H2')(a)} \\
&= c_8\left(\mathbb{E}\left[\|R\|^2\right] + 1\right) \text{ puisque les } R_n \text{ sont i.i.d.} \\
&\leq c_9.
\end{aligned}$$

D'autre part,

$$\begin{aligned}
\|D_n^{-1}\|^4 &= \|(D_n^{-1} - \Delta_n^{-1}) + \Delta_n^{-1}\|^4 \\
&\leq 8\left(\|D_n^{-1} - \Delta_n^{-1}\|^4 + \|\Delta_n^{-1}\|^4\right) \\
&\leq c_{10}\left(\|D_n^{-1} - \Delta_n^{-1}\|^4 + 1\right) \text{ sous (H2')(b)}
\end{aligned}$$

et

$$\begin{aligned}
\|M_n\|^2 &= \|(M_n - C^{-1}) + C^{-1}\|^2 \\
&\leq 2\left(\|M_n - C^{-1}\|^2 + \|C^{-1}\|^2\right) \\
&\leq c_{11}\left(\|M_n - C^{-1}\|^2 + 1\right).
\end{aligned}$$

Ainsi,

$$\begin{aligned}
&3a_n^2\left(\|D_n^{-1}\|^4 \mathbb{E}\left[\|(Z_n - \theta_n)Z_n'\|^2\right] + \|C\|^2 + \|D_n^{-1}\|^4 \|\theta_n - \Theta_n\|^2 \mathbb{E}\left[\|Z_n\|^2\right]\right) \|M_n\|^2 \\
&\leq c_{12}a_n^2\left(\|D_n^{-1} - \Delta_n^{-1}\|^4 \cdot \|\theta_n - \Theta_n\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^4 + \|\theta_n - \Theta_n\|^2 + 1\right) \cdot \left(\|M_n - C^{-1}\|^2 + 1\right).
\end{aligned}$$

De a), b), c), on déduit que :

$$\begin{aligned}
& \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \middle| T_n \right] \\
& \leq \left(1 + a_n^2 \|C\|^2 \right) \|M_n - C^{-1}\|^2 + \\
& \quad c_5 a_n \left(\|D_n^{-1} - \Delta_n^{-1}\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^2 \|\theta_n - \Theta_n\| + \|\theta_n - \Theta_n\| \right) \|M_n - C^{-1}\|^2 + \\
& \quad c_5 a_n \left(\|D_n^{-1} - \Delta_n^{-1}\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^2 \|\theta_n - \Theta_n\| + \|\theta_n - \Theta_n\| \right) + \\
& \quad c_{12} a_n^2 \|M_n - C^{-1}\|^2 \left(\|D_n^{-1} - \Delta_n^{-1}\|^4 \|\theta_n - \Theta_n\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^4 + \|\theta_n - \Theta_n\|^2 + 1 \right) + \\
& \quad c_{12} a_n^2 \left(\|D_n^{-1} - \Delta_n^{-1}\|^4 \|\theta_n - \Theta_n\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^4 + \|\theta_n - \Theta_n\|^2 + 1 \right) - \\
& \quad 2a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2 \\
& \leq \left(1 + a_n^2 \|C\|^2 + \alpha_n \right) \|M_n - C^{-1}\|^2 + \alpha_n - 2a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2
\end{aligned}$$

avec $\alpha_n = c_5 a_n \left(\|D_n^{-1} - \Delta_n^{-1}\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 + \|\theta_n - \Theta_n\| + \|D_n^{-1} - \Delta_n^{-1}\|^2 \|\theta_n - \Theta_n\| \right) + c_{12} a_n^2 \left(\|D_n^{-1} - \Delta_n^{-1}\|^4 + \|\theta_n - \Theta_n\|^2 + \|D_n^{-1} - \Delta_n^{-1}\|^4 \|\theta_n - \Theta_n\|^2 + 1 \right)$.

Sous (H5)(b), (H5')(b), (H3)(c), (H5)(c) (impliquée par (H5)(d) sous (H3)(c)) et (H5')(c) (impliquée par (H5')(d) sous (H3)(c)), $\sum_{n=1}^{+\infty} \alpha_n < \infty$ p.s.

On peut alors appliquer le lemme de Robbins-Siegmund :

Il existe une v.a. $T \geq 0$ telle que $\|M_n - C^{-1}\| \rightarrow T$ p.s. et $\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\|^2 < \infty$ p.s.

Comme sous (H3)(c) $\sum_{n=1}^{+\infty} a_n = \infty$, $T = 0$ et, en reprenant l'écriture des k , $M_n^k \rightarrow (C^k)^{-1}$ p.s.

Ceci étant vrai pour tout $k = 1, \dots, q$ et puisque M_n a pour k -ième bloc diagonal M_n^k , $\|M_n - M\| \rightarrow 0$ p.s.

2) Montrons maintenant que $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

Pour ce faire, on montre que, pour $k = 1, \dots, q$, $\sum_{n=1}^{+\infty} a_n \|M_n^k - (C^k)^{-1}\| < \infty$ p.s.

Dans la suite, on supprime l'écriture des k .

Tout d'abord, remarquons que sous (H5)(b), $\|\theta_n - \Theta_n\| \rightarrow 0$ p.s. et que sous (H2')(b) et (H5')(b), $\|D_n^{-1} - \Delta_n^{-1}\| \rightarrow 0$ p.s. (c.f. partie 1) de la démonstration du théorème 4.1).

Ainsi,

$$\forall \delta_1 > 0, \forall \varepsilon > 0, \exists N(\delta_1, \varepsilon), \forall n \geq N(\delta_1, \varepsilon), \mathbb{P}(\sup_{j>n} \|\theta_j - \Theta_j\| \leq \delta_1) > 1 - \varepsilon,$$

$$\forall \delta_2 > 0, \forall \varepsilon > 0, \exists N(\delta_2, \varepsilon), \forall n \geq N(\delta_2, \varepsilon), \mathbb{P}(\sup_{j>n} \|D_j^{-1} - \Delta_j^{-1}\| \leq \delta_2) > 1 - \varepsilon.$$

On fixe alors δ_1, δ_2 puis on note $N_0 = \max(N(\delta_1, \varepsilon), N(\delta_2, \varepsilon))$.

On pose :

$$F = \left\{ \sup_{j>N_0} \|\theta_j - \Theta_j\| \leq \delta_1, \sup_{j>N_0} \|D_j^{-1} - \Delta_j^{-1}\| \leq \delta_2 \right\},$$

$$F_n = \left\{ \sup_{N_0 < j \leq n} \|\theta_j - \Theta_j\| \leq \delta_1, \sup_{N_0 < j \leq n} \|D_j^{-1} - \Delta_j^{-1}\| \leq \delta_2 \right\}.$$

On peut déjà remarquer que $\mathbb{P}(F) > 1 - 2\varepsilon$ et que F_n est T_n -mesurable.

De plus, on a $F \Rightarrow F_{n+1} \Rightarrow F_n$ i.e. $F \subset F_{n+1} \subset F_n$ ou encore $\mathbb{1}_F \leq \mathbb{1}_{F_{n+1}} \leq \mathbb{1}_{F_n} \leq 1$.

$$\text{Donc : } \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n \right] \leq \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_n} | T_n \right] \leq \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 | T_n \right] \mathbb{1}_{F_n}.$$

Alors, pour $n > N_0$, de la majoration de $\mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 | T_n \right]$ établie dans la partie 1) de la démonstration, on déduit que :

$$\mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n \right] \leq (1 + a_n^2 \|C\|^2 + \alpha_n) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} + \alpha_n \mathbb{1}_{F_n} - 2a_n \lambda_{\min}(C) \|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}.$$

Dans F_n , $\|\theta_n - \Theta_n\| \leq \delta_1$ et $\|D_n^{-1} - \Delta_n^{-1}\| \leq \delta_2$, on pose alors $\delta = \delta_1 = \delta_2 \leq \min(1 ; \frac{\lambda_{\min}(C)}{4c_5})$.

Avec ce choix de δ , on trouve que :

$$\begin{aligned}\alpha_n &\leq 2c_5a_n(\|D_n^{-1} - \Delta_n^{-1}\| + \|\theta_n - \Theta_n\|) + 4c_{12}a_n^2 & (1) \\ &\leq 4\delta c_5a_n + 4c_{12}a_n^2 & (2)\end{aligned}$$

Ainsi, avec (1) et (2), on a :

$$\mathbb{E}\left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n\right] \leq \left(1 + (\|C^2\| + 4c_{12})a_n^2 + 4c_5\delta a_n\right)\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} + 2c_5a_n\left(\|D_n^{-1} - \Delta_n^{-1}\| + \|\theta_n - \Theta_n\|\right)\mathbb{1}_{F_n} + 4c_{12}a_n^2 - 2a_n\lambda_{\min}(C)\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}.$$

$$\|D_n^{-1} - \Delta_n^{-1}\| = \|D_n^{-1}(\Delta_n - D_n)\Delta_n^{-1}\| \leq \|D_n^{-1}\| \cdot \|\Delta_n - D_n\| \cdot \|\Delta_n^{-1}\|.$$

$$\text{Dans } F_n, \left| \|D_n^{-1}\| - \|\Delta_n^{-1}\| \right| \leq \|D_n^{-1} - \Delta_n^{-1}\| < \delta.$$

$$\text{Donc, } \|D_n^{-1}\| < \|\Delta_n^{-1}\| + \delta < \frac{1}{\inf_n \min_i \delta_n^i} + \delta.$$

$$\text{On en d duit que : } \|D_n^{-1} - \Delta_n^{-1}\| < c_{13}\|D_n - \Delta_n\|.$$

$$\text{D'autre part, comme } 4c_5\delta < \lambda_{\min}(C), 4c_5\delta a_n\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \leq a_n\lambda_{\min}(C)\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}.$$

On a alors :

$$\mathbb{E}\left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} | T_n\right] \leq \left(1 + c_{14}a_n^2\right)\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} + c_{15}a_n\left(\|D_n - \Delta_n\| + \|\theta_n - \Theta_n\|\right) + c_{16}a_n^2 - a_n\lambda_{\min}(C)\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}.$$

En passant   l'esp rance math matique, on obtient :

$$\mathbb{E}\left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}}\right] \leq \left(1 + c_{14}a_n^2\right)\mathbb{E}\left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}\right] + c_{15}a_n\mathbb{E}\left[\|D_n - \Delta_n\| + \|\theta_n - \Theta_n\|\right] + c_{16}a_n^2 - a_n\lambda_{\min}(C)\mathbb{E}\left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}\right].$$

Sous (H3)(c), (H5)(d) et (H5')(d), on peut de nouveau appliquer le lemme de Robbins-Siegmund :

$$\text{Il existe un r el } t \geq 0 \text{ tel que } \mathbb{E}\left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}\right] \rightarrow t \text{ et } \sum_{n=1}^{+\infty} a_n \mathbb{E}\left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n}\right] < \infty.$$

$$\text{Comme sous (H3)(c) } \sum_{n=1}^{+\infty} a_n = \infty, t = 0.$$

Ainsi, il existe c_{17} , $\mu > 0$ tels que :

$$\begin{aligned} \mathbb{E} \left[\|M_{n+1} - C^{-1}\|^2 \mathbb{1}_{F_{n+1}} \right] &\leq \left(1 - a_n \lambda_{\min}(C) \right) \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] + c_{17} a_n g_n \\ &\quad \text{avec } g_n = a_n + \mathbb{E} \left[\|D_n - \Delta_n\| \right] + \mathbb{E} \left[\|\theta_n - \Theta_n\| \right] = O\left(\frac{1}{n^{\min(\alpha, \gamma_2, \gamma_1)}}\right) \\ &\leq \left(1 - a_n \lambda_{\min}(C) \right) \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] + \mu a_n h_n \text{ avec } h_n = \frac{1}{n^{\min(\alpha, \gamma_2, \gamma_1)}}. \end{aligned}$$

Sous (H5)(d), (H5')(d) et (H3)(c), d'après le lemme 3.10, on a :

$$\overline{\lim} \frac{1}{h_n} \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] < \infty.$$

$$\text{Alors, } \mathbb{E} \left[\|M_n - C^{-1}\|^2 \mathbb{1}_{F_n} \right] = O(h_n) \text{ d'où } \mathbb{E} \left[\|M_n - C^{-1}\| \mathbb{1}_{F_n} \right] = O(\sqrt{h_n}).$$

$$\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\|M_n - C^{-1}\| \mathbb{1}_{F_n} \right] = \sum_{n=1}^{+\infty} a_n O(\sqrt{h_n}) \leq c \sum_{n=1}^{+\infty} \frac{1}{n^\alpha} \frac{1}{n^{\min(\frac{\alpha}{2}, \frac{\gamma_2}{2}, \frac{\gamma_1}{2})}} \text{ sous (H3)(c).}$$

Or, sous (H3)(c), (H5)(d) et (H5')(d), $\frac{\alpha}{2} > 1 - \alpha$; $\frac{\gamma_1}{2} > 1 - \alpha$ et $\frac{\gamma_2}{2} > 1 - \alpha$.

Ainsi, $\min(\frac{\alpha}{2}, \frac{\gamma_2}{2}, \frac{\gamma_1}{2}) > 1 - \alpha$: $\exists \delta > 0$, $\min(\frac{\alpha}{2}, \frac{\gamma_2}{2}, \frac{\gamma_1}{2}) = 1 - \alpha + \delta$.

On en déduit que $\sum_{n=1}^{+\infty} a_n \mathbb{E} \left[\|M_n - C^{-1}\| \mathbb{1}_{F_n} \right] = \sum_{n=1}^{+\infty} \frac{1}{n^{\alpha+1-\alpha+\delta}} = \sum_{n=1}^{+\infty} \frac{1}{n^{1+\delta}} < \infty$ puis que $\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| \mathbb{1}_{F_n} < \infty$ p.s.

Comme $\mathbb{1}_F \leq \mathbb{1}_{F_n}$, on peut écrire que $\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| \mathbb{1}_F < \infty$ p.s.

Ainsi, d'un point de vue logique, $\{F\} \Rightarrow \left\{ \sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| < \infty \right\}$ ce qui implique, d'un point de vue probabiliste, que : $\mathbb{P}\left(\sum_{n=1}^{+\infty} a_n \|M_n - C^{-1}\| < \infty\right) \geq \mathbb{P}(F) > 1 - \varepsilon$.

ε étant quelconque, on en déduit finalement que, en réécrivant les k :

$$\sum_{n=1}^{+\infty} a_n \|M_n^k - (C^k)^{-1}\| < \infty \text{ p.s.}$$

L'ensemble des v.p. de la matrice diagonale par blocs $(M_n - M)$ étant la réunion des ensembles des v.p. des matrices $(M_n^k - (C^k)^{-1})$, $k = 1, \dots, q$, on a : $\|M_n - M\| = \max_k \|M_n^k - (C^k)^{-1}\| \leq \sum_{k=1}^q \|M_n^k - (C^k)^{-1}\|$.

On en déduit que $\sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty$ p.s.

4.3.2 Hypothèses

- (H1)(a) Il n'existe pas de relation affine entre les composantes de R ,
(H1)(b) R admet des moments d'ordre $4r$,
(H1)(c) R admet des moments d'ordre 4,

$$(H2) \sup_{l \in \mathbb{N}} \|\theta_l\| < \infty,$$

$$(H2')(a) \sup_{l \in \mathbb{N}} \max_i |\delta_l^i| < \infty,$$

$$(H2')(b) \inf_{l \in \mathbb{N}} \min_i |\delta_l^i| > 0,$$

$$(H3) (a) a_n > 0, \sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} a_n^2 < \infty,$$

$$(H3) (b) (a_n = \frac{a}{n^\alpha} \text{ avec } a > 0, \frac{2}{3} < \alpha \leq 1),$$

$$(H3) (c) (a_n = \frac{a}{n^\alpha} \text{ avec } a > 0, \frac{2}{3} < \alpha \leq 1) \text{ ou } (\alpha = 1 \text{ i.e. } a_n = \frac{a}{n} \text{ et } a > \frac{1}{\min_k \lambda_{\min}(C^k)}),$$

$$(H4) (a) M_n \text{ est } T_n\text{-mesurable},$$

$$(H4) (b) M_n \rightarrow M \text{ p.s.},$$

$$(H4) (c) \sum_{n=1}^{+\infty} a_n \|M_n - M\| < \infty \text{ p.s.},$$

$$(H5) (a) \Theta_n \text{ est } T_n\text{-mesurable},$$

$$(H5) (b) \|\theta_n - \Theta_n\| \rightarrow 0 \text{ p.s.},$$

$$(H5) (c) \sum_{n=1}^{+\infty} a_n \|\theta_n - \Theta_n\| < \infty \text{ p.s.},$$

$$(H5) (d) \mathbb{E}[\|\theta_n - \Theta_n\|] = O(\frac{1}{n^\gamma}), \gamma_1 > 2(1 - \alpha),$$

$$(H5') (a) D_n \text{ est } T_n\text{-mesurable},$$

$$(H5') (b) \|\Delta_n - D_n\| \rightarrow 0 \text{ p.s.},$$

$$(H5') (c) \sum_{n=1}^{+\infty} a_n \|\Delta_n - D_n\| < \infty \text{ p.s.},$$

$$(H5') (d) \mathbb{E}[\|\Delta_n - D_n\|] = O(\frac{1}{n^{\gamma_2}}), \gamma_2 > 2(1 - \alpha),$$

$$(H6) (a) \max_i \sup_n \|U_n^i\| < \infty,$$

$$(H6) (a') \max_i \sup_n \|V_n^i\| < \infty,$$

(H6) (b) Pour $i = 1, \dots, p$, il existe un entier r_i , un réel $\lambda_i > 0$, une suite croissante d'entiers $(n_{il}, l \geq 1)$ tels que $n_{i1} = 1$, $n_{i,l+1} \leq n_{il} + r_i$, $\min \left(\lambda_{\min} \left(\sum_{j \in I_{il}} V_j^i (V_j^i)' \right); \lambda_{\min} \left(\sum_{j \in I_{il}} U_j^i (U_j^i)' \right) \right) \geq \lambda_i$, avec $I_{il} = \{n_{il}, \dots, n_{i,l+1} - 1\}$,

(H7) Pour $\frac{1}{2} < \alpha < 1$, on suppose :

(a) $\zeta < 2\alpha - 1$,

(b) $\gamma < \frac{3}{2}\alpha - 1$,

(H8) Pour $\alpha = 1$, pour $i = 1, \dots, p$:

(a) si $\frac{2\lambda_i a}{r_i} \geq 1$, on suppose :

(i) $\zeta < 2\alpha - 1$,

(ii) $\gamma < \frac{3}{2}\alpha - 1$,

(b) si $\frac{2\lambda_i a}{r_i} < 1$, on suppose :

(i) $\zeta < \frac{1}{2} + \frac{\lambda_i a}{r_i}$,

(ii) $\gamma < \frac{\lambda_i a}{r_i}$.

Chapitre 5

Mise en oeuvre informatique

Sommaire

5.1	Cadre déterministe	151
5.2	Cadre aléatoire	154
5.2.1	Cas i.i.d.	154
5.2.2	Cas où l'espérance est variable dans le temps	157
5.2.3	Cas où l'espérance et la matrice de covariance sont variables dans le temps	163
5.3	Package R : factas	170

Cette dernière partie est consacrée aux simulations informatiques, réalisées avec le logiciel R.

L'objectif est de mettre en oeuvre les processus théoriques définis dans les parties précédentes et de comparer leurs performances face à d'autres méthodes d'estimation.

On a choisi de se concentrer, sans perte de généralité, sur des applications à l'Analyse Canonique Généralisée (ACG).

Dans un premier temps, on considèrera un cadre déterministe (population finie) puis on se placera dans un cadre aléatoire où les données arrivent sous forme de flux, c'est-à-dire en ligne et de manière infinie.

5.1 Cadre déterministe

On dispose d'un tableau de données composé de n lignes, chaque ligne représentant un individu, et p colonnes, chaque colonne représentant un caractère de l'individu (variable quantitative).

Ainsi, en considérant ce tableau comme une matrice X , l'élément (i, j) , $i = 1, \dots, n$, $j = 1, \dots, p$ représentera la mesure du caractère j sur l'individu i .

Comme on travaille dans le cadre de l'ACG, on considère que ces p variables sont scindées en q sous-groupes de variables.

On cherche à calculer les r premiers facteurs généraux de l'ACG de ce tableau de données.

On note C la matrice de covariance de X et $M = \begin{pmatrix} (C^1)^{-1} & & \\ & \ddots & \\ & & (C^q)^{-1} \end{pmatrix}$

où, pour $k = 1, \dots, q$, C^k est la matrice de covariance de la matrice X dont on ne conserve que les colonnes du k -ième sous-groupe de variables.

L'ensemble des observations étant fini, la valeur exacte de ces r facteurs correspond à la valeur des r vecteurs propres de la matrice $B = MC$ associés aux r plus grandes valeurs propres (v.p.) $\lambda_1, \dots, \lambda_r$.

D'une part, on utilise la fonction *eigen* de R, qui fait appel à une routine Lapack implémentée en Fortran, pour effectuer le calcul.

D'autre part, on utilise, pour $l = 1, \dots, r$, le processus d'approximation stochastique (a.s.) (X_n^l) de type Krasulina [33], défini dans [51] et redonné dans l'état de l'art, tel que :

$$\begin{aligned}
F_n(X_n^l) &= \frac{\langle B_n X_n^l, X_n^l \rangle_{M_n}}{\|X_n^l\|_{M_n}^2}, \\
Y_{n+1}^l &= X_n^l + a_n (B_n - F_n(X_n^l)I) X_n^l, \quad l = 1, \dots, r, \\
X_{n+1} &= \text{orth}_{M_n}(Y_{n+1}).
\end{aligned}$$

Dans ce cadre, $M_n = M^{-1}$ et $B_n = B$ sont connus, la procédure est uniquement itérative.

Dans ce premier exemple, on fixe $r = 3$, $n = 10000$ et on fait varier le nombre de colonnes p .

On construit alors un tableau avec des données simulées.

Après quelques essais, le pas $a_n = (\frac{100}{n+100})^{0.6}$ du type $(\frac{a}{n+b})^c$ nous a paru satisfaisant pour obtenir des résultats convaincants quant au problème considéré.

Il faut noter qu'un choix approprié du pas a_n est souvent déterminant pour obtenir de bonnes performances des processus d'a.s.

Plus précisément, le coefficient a a une influence sur la vitesse asymptotique de l'algorithme : son effet multiplicatif sur n peut plus ou moins accélérer la convergence ; le coefficient b permet de régler les problèmes dans la phase transitoire de l'algorithme : au cours des premières itérations, si le terme n est petit devant le terme additif b , le pas est sensiblement égal au ratio $(\frac{a}{b})^c$; ce rapport sert donc à déterminer un pas "acceptable" en début d'algorithme : un pas trop petit pénalise la vitesse de convergence tandis qu'un pas trop grand provoque des explosions numériques durant les premières itérations.

Il existe de nombreux articles décrivant des stratégies de mise à jour des pas. On pourra par exemple consulter [13], [31], [57]. Cependant, la mise en oeuvre d'un algorithme d'a.s. nécessite souvent un certain nombre d'expérimentations numériques avant de donner des résultats satisfaisants.

La fonction R donne la valeur exacte des 3 facteurs tandis que le processus d'a.s. nous en donne seulement une estimation, estimation qui s'affine avec le nombre d'itérations de l'algorithme.

On choisit d'arrêter l'algorithme dès lors que la valeur absolue du cosinus de l'angle formé par chacun des facteurs théoriques avec les facteurs calculés associés est supérieure à 0.999 ce qui correspond à un angle entre les deux vecteurs inférieur à $2,6^\circ$.

On compare alors le temps mis en secondes par chacune des 2 méthodes pour effectuer les calculs.

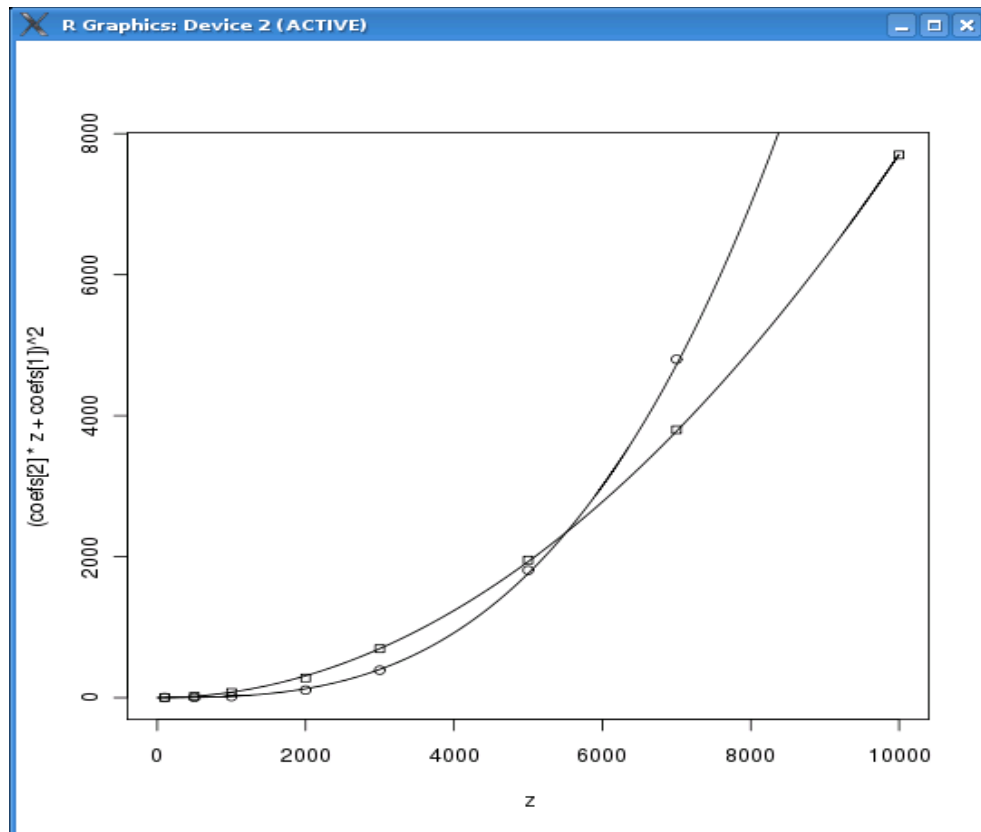
On récapitule ici, sous la forme d'un tableau, les résultats obtenus pour différentes valeurs de p :

TABLE 5.1: Temps de calcul des 2 méthodes

p	temps 1	temps 2
100	1.207	3.509
500	2.500	17.879
1000	15.765	75.461
2000	108.402	279.749
3000	389.321	695.465
5000	1807.646	1951.847
7000	4811.364	3801.275
10000	13543.859	7794.836

On donne une représentation graphique de ces résultats :

FIGURE 5.1: Temps des 2 méthodes en fonction du nombre de colonnes p



On remarque que plus le nombre de colonnes du tableau est élevé, meilleure est la méthode utilisant le processus d'a.s. en terme de temps de calcul relativement à l'autre méthode.

On peut donc envisager d'avoir recours à ce type de méthode pour l'analyse de grands tableaux de données.

5.2 Cadre aléatoire

On considère maintenant le cas où les données arrivent par flux c'est-à-dire en ligne, de manière ininterrompue et infinie : à chaque instant n , on dispose d'une ou plusieurs observations d'un vecteur aléatoire (v.a.) Z_n , partitionné en q sous-vecteurs $(Z_n^1, \dots, Z_n^q)$.

Dans toute la suite, les v.a. Z_n sont supposés indépendants.

5.2.1 Cas i.i.d.

Dans ce premier cas, on suppose que, pour tout n , les Z_n sont identiquement distribués suivant la loi d'un v.a. Z .

On souhaite pouvoir disposer, à tout instant, d'une estimation de vecteurs directeurs des r premiers axes principaux de l'ACG de ce v.a. Z .

Voici les principales étapes du programme de test :

1) On fixe les paramètres du programme en choisissant le temps durant lequel va tourner l'algorithme (en supposant que le flux de données est continu), la dimension du vecteur Z dont on observe des réalisations et le nombre r de vecteurs à estimer.

2) Initialisation : on prend en compte un petit nombre d'observations afin de calculer une première estimation de la matrice de covariance C , C_0 , de la métrique M , M_0 , et de vecteurs directeurs $v^1, \dots, v^r$ des axes principaux, $v_0^1, \dots, v_0^r$.

3) Pas n : On introduit un nouveau vecteur d'observations puis on met à jour la matrice de covariance empirique C_n et la métrique empirique M_n à l'aide de formules récursives.

D'une part, on calcule grâce à la fonction *eigen* de R les r premiers vecteurs propres de la matrice $C_n M_n$ qui sont des estimations de vecteurs directeurs des axes principaux de l'ACG de Z .

D'autre part, on calcule grâce à la fonction *irlba* de R, basée sur une méthode de type Lanczos, les r premiers vecteurs propres de la matrice $C_n M_n$ qui sont des estimations de vecteurs directeurs des axes principaux de l'ACG de Z .

Enfin, on met en oeuvre, pour $l = 1, \dots, r$, le processus (X_n^l) tel que décrit dans le deuxième cas de la partie 3.1.1 (i.e. en utilisant un processus d'a.s. de C) :

$$\begin{aligned} B_n &= C_n M_n, \\ F_n(X_n^l) &= \frac{\langle B_n X_n^l, X_n^l \rangle_{M_n}}{\|X_n^l\|_{M_n}^2}, \\ Y_{n+1}^l &= X_n^l + a_n (B_n - F_n(X_n^l) I) X_n^l, \quad l = 1, \dots, r, \\ X_{n+1} &= \text{orth}_{M_n}(Y_{n+1}). \end{aligned}$$

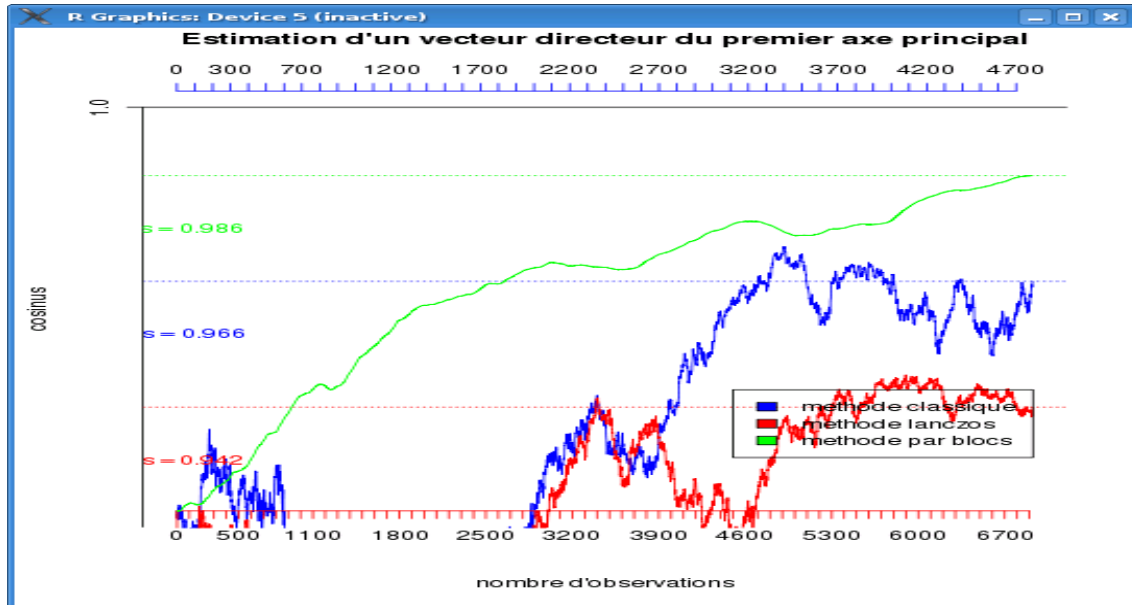
On choisit également d'utiliser, pour $k = 1, \dots, q$, un processus d'a.s. de C^k pour estimer M (c.f. partie 3.2.2).

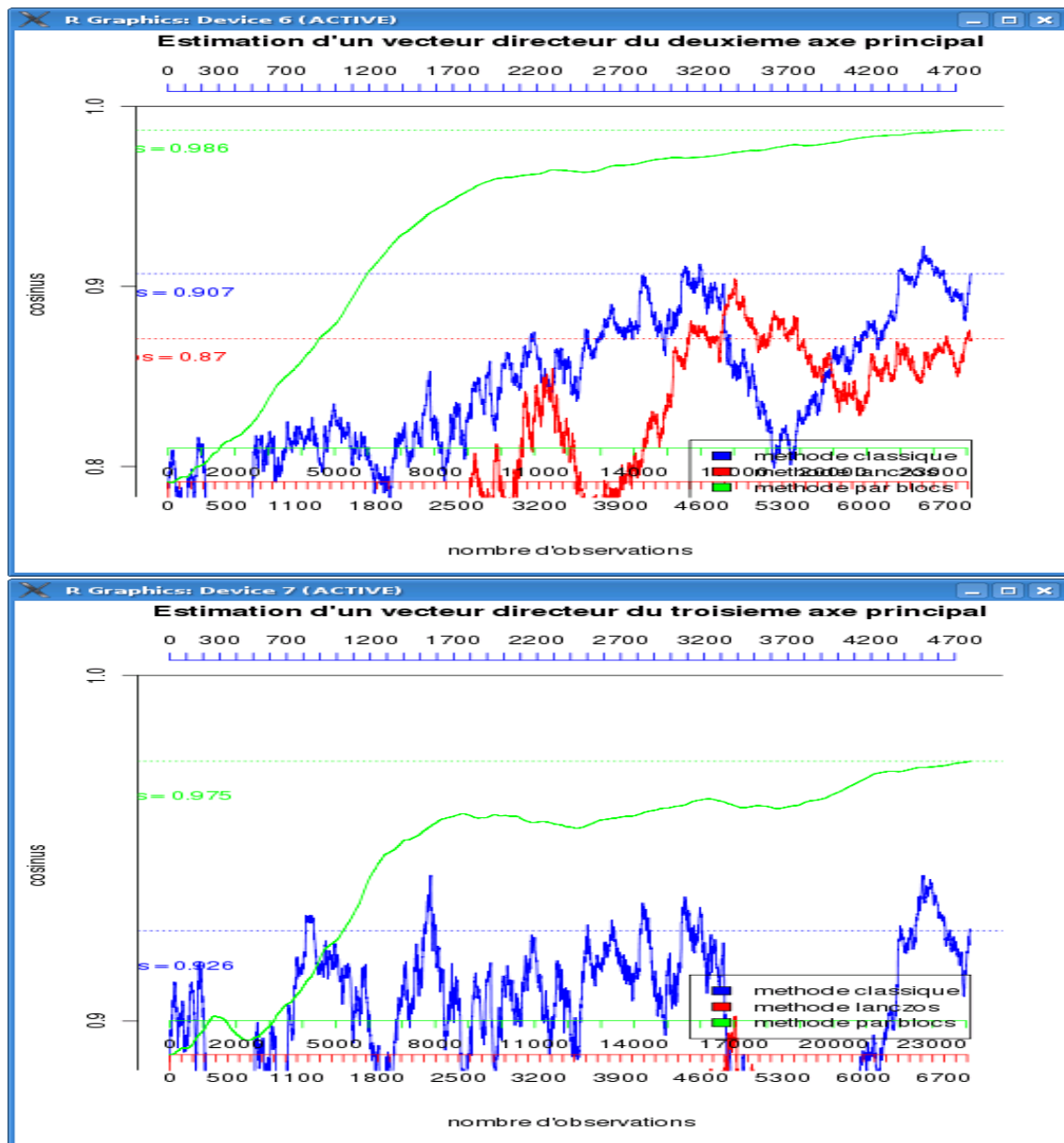
Enfin, on prend $\Theta_n = \frac{1}{n-1} \sum_{i=1}^{n-1} Z_i$ pour estimer θ_n .

Ainsi, on obtient, pour $l = 1, \dots, r$, une estimation d'un vecteur directeur du l -ième axe principal de l'ACG de Z .

4) Pour un même temps d'exécution, on compare alors la précision des trois méthodes via la valeur du cosinus de l'angle formé par les vecteurs théoriques et calculés en fonction du nombre d'observations prises en compte.

Voici par exemple les résultats obtenus pour l'estimation de vecteurs directeurs des trois premiers axes principaux de l'ACG d'un vecteur Z de dimension 190 réparti en 6 sous-vecteurs de dimension respective $(25, 25, 37, 33, 38, 32)$. Après quelques essais, on choisit d'utiliser le processus moyennisé (c.f. [58]) avec un pas $a_n = (\frac{10}{n+10})^{0.9999}$ et ce, pendant une durée de 800 secondes.





1er vecteur directeur :

cos₁ = 0.966
angle₁ = 15°
cos₂ = 0.942
angle₂ = 19.6°
cos₃ = 0.986
angle₃ = 9.6°

2ème vecteur directeur :

cos₁ = 0.907
angle₁ = 24.9 °
cos₂ = 0.87
angle₂ = 29.5°
cos₃ = 0.986
angle₃ = 9.6°

3ème vecteur directeur :

cos₁ = 0.926
angle₁ = 22.2°
cos₂ = 0.83
angle₂ = 33.9°
cos₃ = 0.975
angle₃ = 12.8°

On remarque que l'estimation de vecteurs directeurs des trois premiers axes principaux est plus précise avec la méthode d'a.s. (méthode 3) pour un même temps de calcul.

Cependant, ces résultats sont à nuancer dans la mesure où on observe une certaine variabilité selon le jeu de données considéré.

5.2.2 Cas où l'espérance est variable dans le temps

On rappelle le modèle correspondant :

$$Z_n = \theta_n + \tilde{Z}_n,$$

les $\tilde{Z}_n$ constituant un échantillon i.i.d. d'un v.a. $\tilde{Z}$ de $\mathbb{R}^p$ d'espérance nulle et de matrice de covariance C .

On choisit un modèle linéaire pour représenter l'espérance.

Plus précisément, on suppose que, pour $i = 1, \dots, p$, il existe un vecteur β^i inconnu de $\mathbb{R}^{n_i}$ et, pour tout n , un vecteur U_n^i de $\mathbb{R}^{n_i}$ connu au temps n tels que la $i^{\text{ème}}$ composante réelle de θ_n , θ_n^i s'écrive sous la forme $\langle \beta^i, U_n^i \rangle$.

On définit le processus d'a.s. (B_n^i) de β^i tel que :

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

On définit alors comme estimateur de θ_n^i , $\Theta_n^i = \langle B_n^i, U_n^i \rangle$ et comme estimateur de θ_n , $\Theta_n = (\Theta_n^1 \dots \Theta_n^p)'$.

Poursuite de l'espérance :

Soit $p = 5$.

Pour simplifier, on choisit : $\forall i, n_i = 1$ ce qui signifie que les β^i et les U_n^i sont unidimensionnels :

$$\beta^1 = 1$$

$$\beta^2 = 0.5$$

$$\beta^3 = 5$$

$$\beta^4 = 3.14$$

$$\beta^5 = 4$$

$$U_n^1 = 2$$

$$U_n^2 = \cos(n\pi/16)$$

$$U_n^3 = 1 + \exp(-n/1000)$$

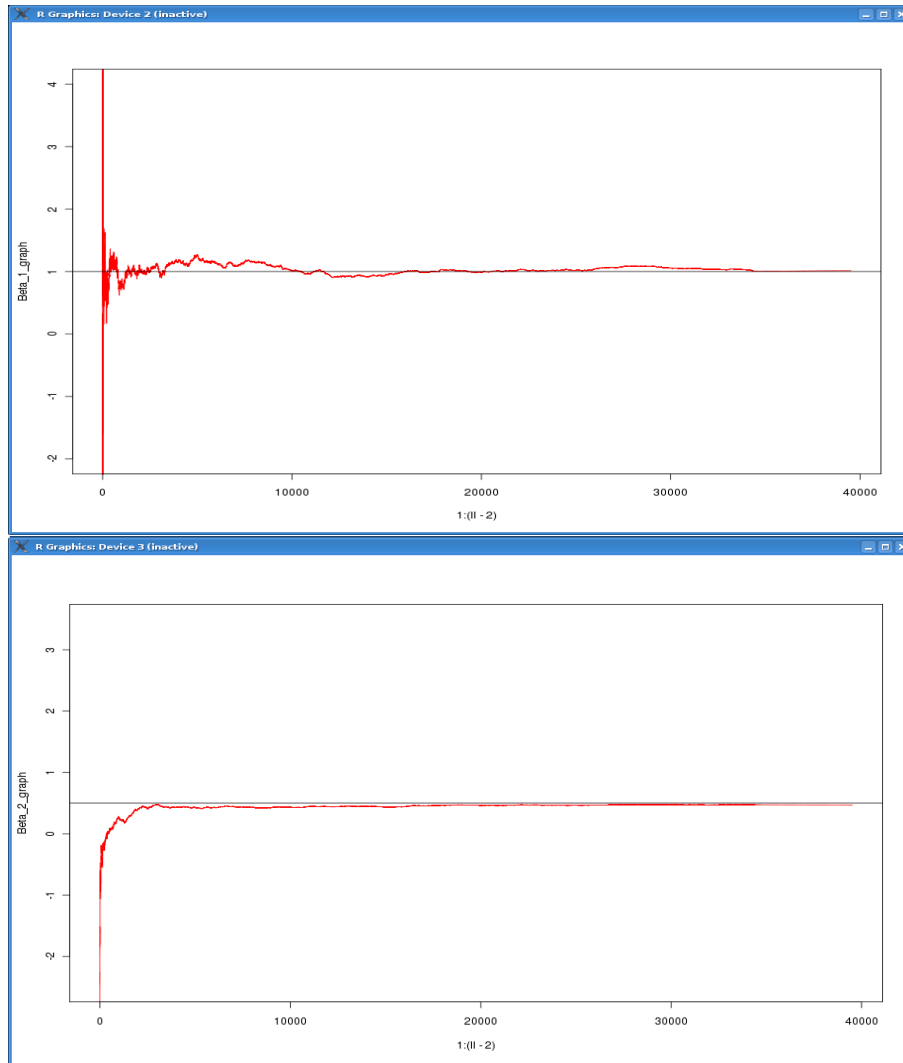
$$U_n^4 = \text{rnorm}(1, 1, 1) \text{ (réalisation d'une } \mathcal{N}(1, 1) \text{)}$$

$$U_n^5 = \text{runif}(1, \min = 0, \max = 1) \text{ (réalisation d'une } \mathcal{U}_{[0,1]} \text{)}$$

$$\theta_n^i = \langle \beta^i, U_n^i \rangle = \beta^i \cdot U_n^i$$

$$\Theta_n^i = \langle B_n^i, U_n^i \rangle = B_n^i \cdot U_n^i.$$

Résultats :



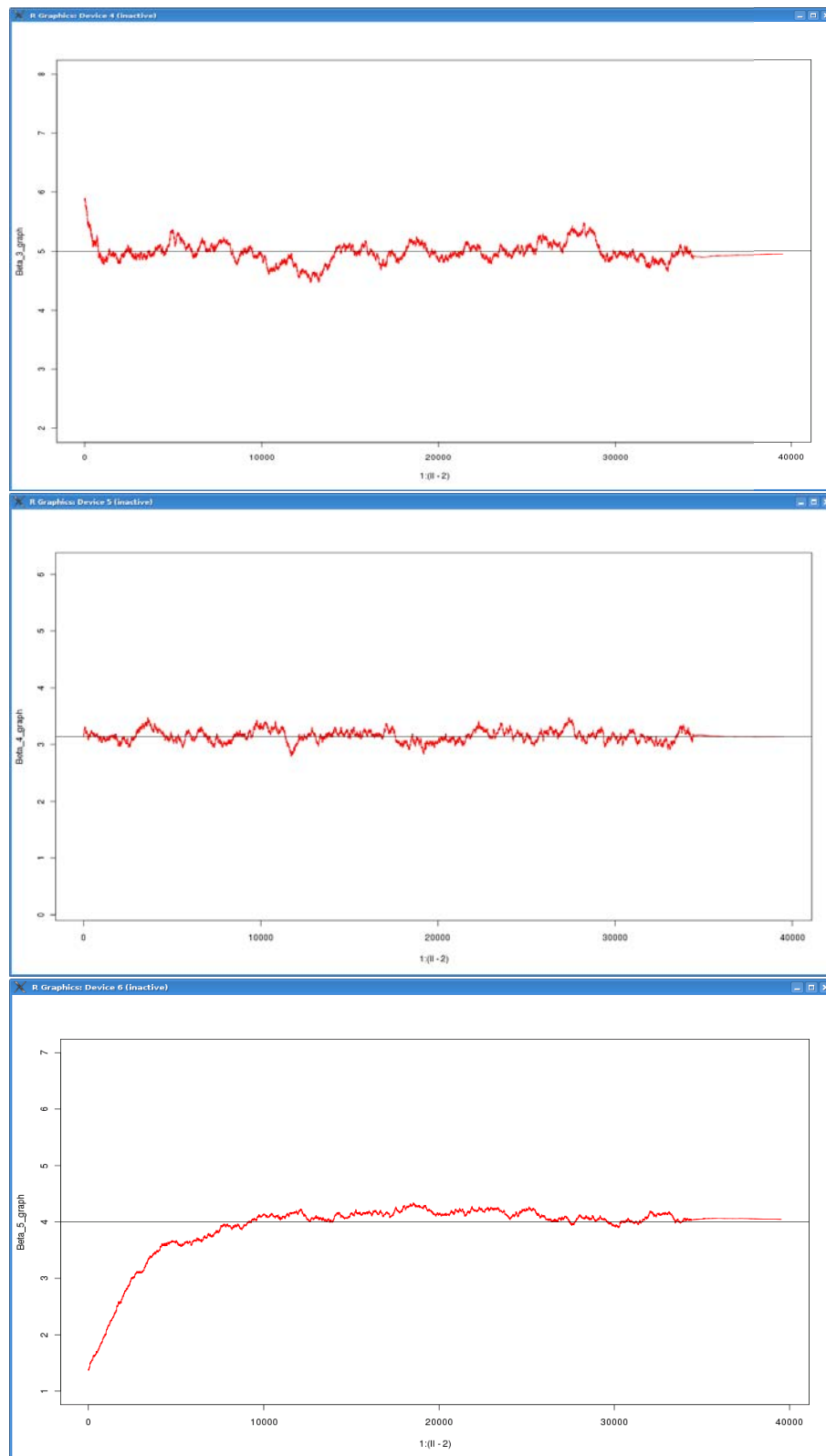
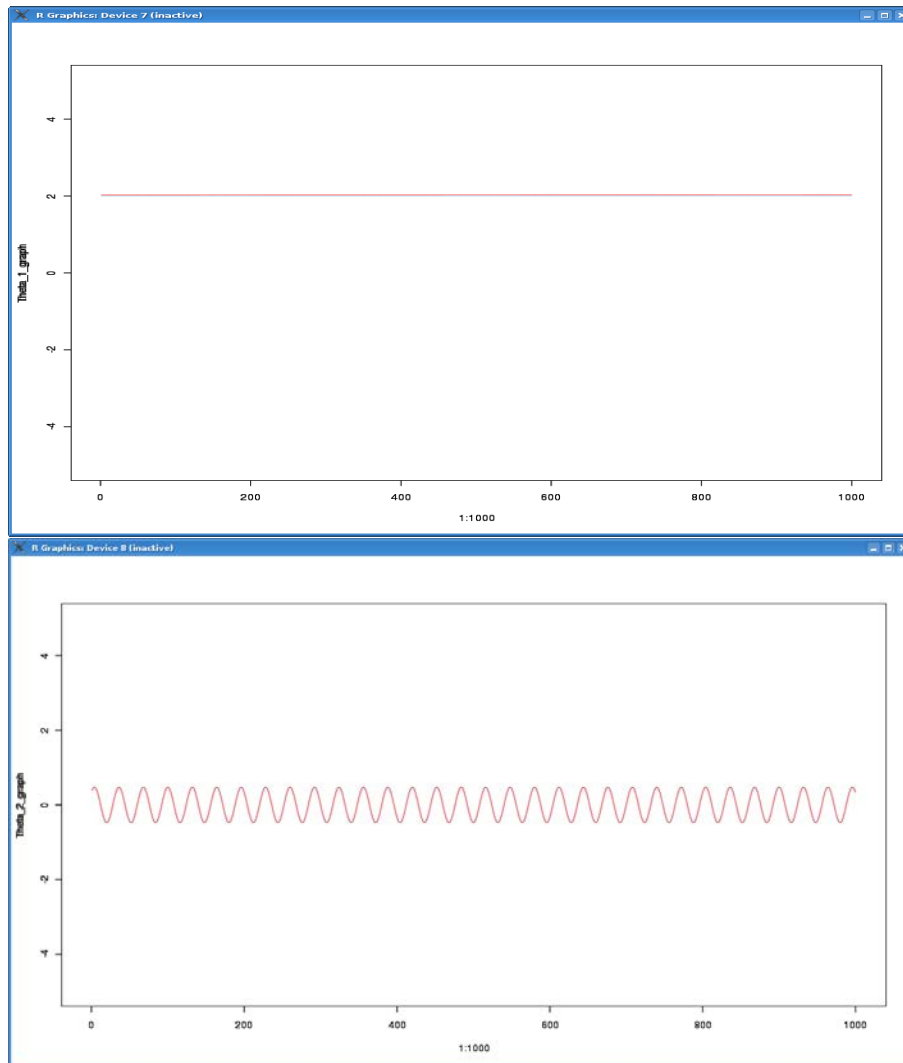


FIGURE 5.2: Estimation des β_i

Les graphiques sont centrés en ordonnée sur la valeur des β^i (ligne noire); l'évolution du processus B_n^i au cours du temps est représentée en rouge.



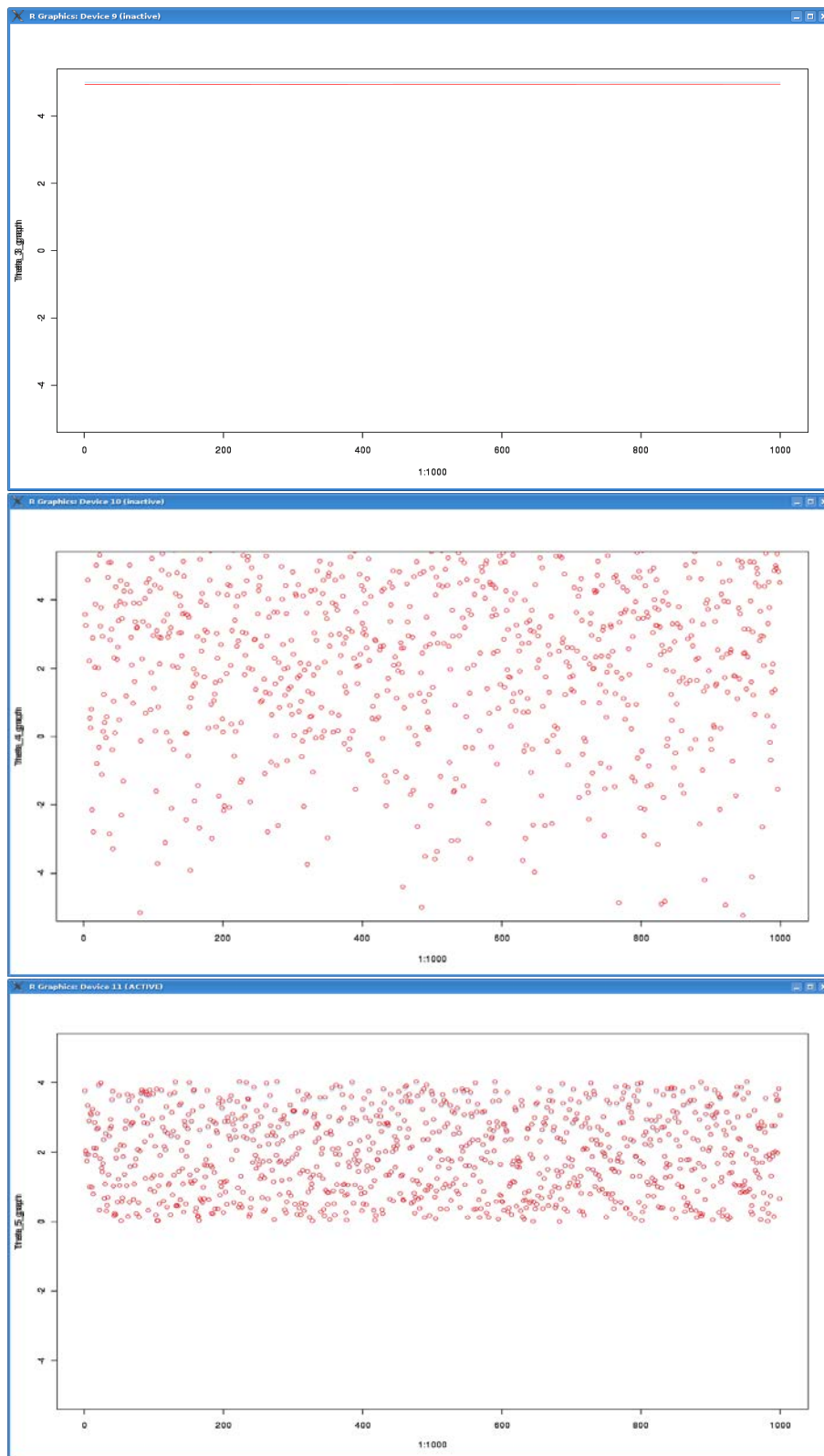


FIGURE 5.3: Estimation des θ_n^i

L'évolution du processus déterministe de l'espérance θ_n^i est d'abord tracée en bleu puis, l'évolution du processus aléatoire de poursuite de l'espérance Θ_n^i est représentée sur le même graphique en rouge. On remarque que le trait rouge recouvre presque parfaitement le tracé bleu à la fin de la simulation (seules les 1000 dernières itérations sont représentées ici) ; ceci témoigne de la bonne estimation de l'espérance par le processus d'a.s.

Remarque :

Voici une alternative à cette étape de poursuite de l'espérance :

Matriciellement, on peut écrire le tableau de toutes les données observées jusqu'à l'instant n de la manière suivante.

Pour $i = 1, \dots, p$, on a :

$$\begin{pmatrix} Z_1^i \\ \vdots \\ Z_n^i \end{pmatrix} = \begin{pmatrix} (U_1^i)^1 & \cdot & \cdot & \cdot & (U_1^i)^{n_i} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ (U_n^i)^1 & \cdot & \cdot & \cdot & (U_n^i)^{n_i} \end{pmatrix} \cdot \begin{pmatrix} (\beta^i)^1 \\ \vdots \\ \vdots \\ \vdots \\ (\beta^i)^{n_i} \end{pmatrix} + \begin{pmatrix} \tilde{Z}_1^i \\ \vdots \\ \vdots \\ \vdots \\ \tilde{Z}_n^i \end{pmatrix}$$

qu'on récrit $Z^i = M_{U^i} \cdot \beta^i + \tilde{Z}^i$ pour simplifier.

Pour $i = 1, \dots, p$, on calcule une estimation des β^i au sens des moindres carrés :

$$\hat{\beta}^i = (M_{U^i}' M_{U^i})^{-1} M_{U^i}' Z^i.$$

Cette estimation peut se faire de manière récursive en remarquant que l'on peut calculer de manière récursive $(M_{U^i}' M_{U^i})^{-1}$, on pourra consulter [53] à ce propos, et $M_{U^i}' Z^i$ également de manière récursive.

Pour tout n , on calcule alors une estimation des θ_n^i correspondants, $\hat{\theta}_n^i$.

Mise en oeuvre de l'ACG

Les étapes sont les mêmes que dans le cas précédent (section 5.2.1).

On présente les résultats obtenus pour l'estimation de vecteurs directeurs des trois premiers axes principaux de l'ACG d'un vecteur Z de dimension 15 pendant une durée de 60 secondes.

Résultats :

— 1er vecteur — 2ème vecteur — 3ème vecteur

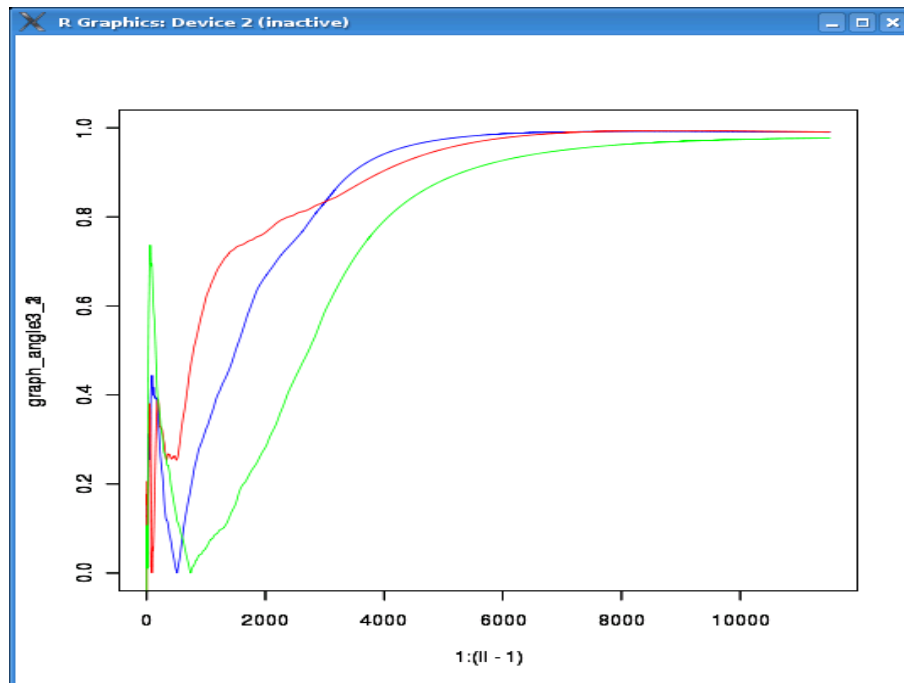


FIGURE 5.4: Cosinus en fonction du temps

On remarque qu'on a bien convergence des vecteurs calculés vers les vecteurs théoriques. De plus, l'estimation est satisfaisante à partir d'environ 6000 itérations pour les 2 premiers vecteurs tandis qu'il faut attendre environ 8000 itérations pour avoir une bonne estimation du dernier vecteur.

5.2.3 Cas où l'espérance et la matrice de covariance sont variables dans le temps

On rappelle le modèle correspondant :

On suppose que, pour tout n , on a la décomposition :

$$Z_n = \theta_n + \Delta_n R_n \quad \text{où :}$$

· $\theta_n = (\theta_n^1 \dots \theta_n^p)'$ est un vecteur de $\mathbb{R}^p$,

$$\Delta_n = \begin{pmatrix} \delta_n^1 & & \\ & \ddots & \\ & & \delta_n^p \end{pmatrix} \text{ est une matrice diagonale d'ordre } p \text{ d'éléments non nuls.}$$

· la loi du v.a. R_n ne dépend pas de n , $\mathbb{E}[R_n] = 0, \text{Cov}(R_n) = C$.

On définit un modèle linéaire pour représenter l'espérance et la variance de Z_n .

On note $\delta_n = (\delta_n^1 \dots \delta_n^p)'$.

Pour $i = 1, \dots, p$, il existe un vecteur β^i (resp. γ^i) inconnu de $\mathbb{R}^{n_i}$ (resp. $\mathbb{R}^{m_i}$) et, pour tout n , un vecteur U_n^i (resp. V_n^i) de $\mathbb{R}^{n_i}$ (resp. $\mathbb{R}^{m_i}$) connu au temps n tels que la $i^{\text{ème}}$ composante réelle de θ_n (resp. δ_n) s'écrit :

$$\theta_n^i = \langle \beta^i, U_n^i \rangle \left(\text{resp. } \delta_n^i = \left(\langle \gamma^i, V_n^i \rangle \right)^{\frac{1}{2}} \right).$$

De façon générale, on note Π_E l'opérateur de projection sur un sous-espace E .

On définit le processus d'a.s. (B_n^i) de β^i tel que :

$$B_{n+1}^i = \Pi_{K_n^i} \left(B_n^i - a_n U_n^i \left((U_n^i)' B_n^i - Z_n^i \right) \right)$$

avec $K_n^i = \left\{ x \in \mathbb{R}^{n_i}, \|x\| < k_n^i \right\}$ où $k_n^i = O(n^\zeta)$, $\zeta > 0$.

On définit comme estimateur de θ_n^i , $\Theta_n^i = \langle B_n^i, U_n^i \rangle$, et comme estimateur de θ_n , $\Theta_n = (\Theta_n^1 \dots \Theta_n^p)'$.

On définit le processus d'a.s. (C_n^i) de γ^i tel que :

$$C_{n+1}^i = \Pi_{L_n^i} \left(C_n^i - a_n V_n^i \left((V_n^i)' C_n^i - (Z_n^i - \Theta_n^i)^2 \right) \right)$$

avec $L_n^i = \left\{ x \in \mathbb{R}^{m_i}, \|x\| < l_n^i \right\}$ où $l_n^i = O(n^\gamma)$, $\gamma > 0$.

On définit enfin comme estimateur de δ_n^i , $D_n^i = \left(\langle C_n^i, V_n^i \rangle \right)^{\frac{1}{2}}$, puis comme estimateur de Δ_n , $D_n = \begin{pmatrix} D_n^1 & & \\ & \ddots & \\ & & D_n^p \end{pmatrix}$.

Poursuite de l'espérance :

Soit $p = 5$, on choisit les mêmes β^i et U_n^i que dans le paragraphe précédent.

Malgré l'opération de projection supplémentaire, les résultats obtenus pour la poursuite de l'espérance dans ce nouveau contexte sont sensiblement identiques à ceux obtenus dans le paragraphe précédent.

Poursuite de la variance :

En remarquant que, pour $i = 1, \dots, p$, le processus Θ_n^i intervient dans la définition du processus C_n^i , on débute le calcul récursif de ce dernier de manière décalée dans le temps afin de disposer déjà d'une estimation convenable de Θ_n^i .

Pour simplifier, on choisit $m_i = 1$ pour tout i , ce qui signifie que les γ^i et les V_n^i sont unidimensionnels :

$$\gamma^1 = 1$$

$$\gamma^2 = 10$$

$$\gamma^3 = 5$$

$$\gamma^4 = 2.718$$

$$\gamma^5 = 4$$

$$V_n^1 = 2$$

$$V_n^2 = 4 + \sin(n\pi/16)$$

$$V_n^3 = 1 + \ln\left(1 + \frac{10}{n}\right)$$

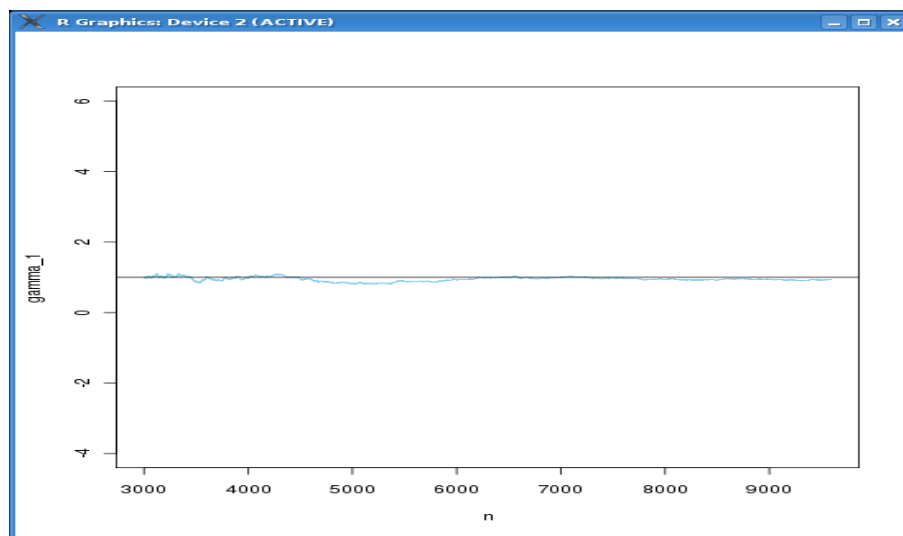
$$V_n^4 = 3 + \text{rexp}(1, 1) \left(3 + \mathcal{E}(1)\right)$$

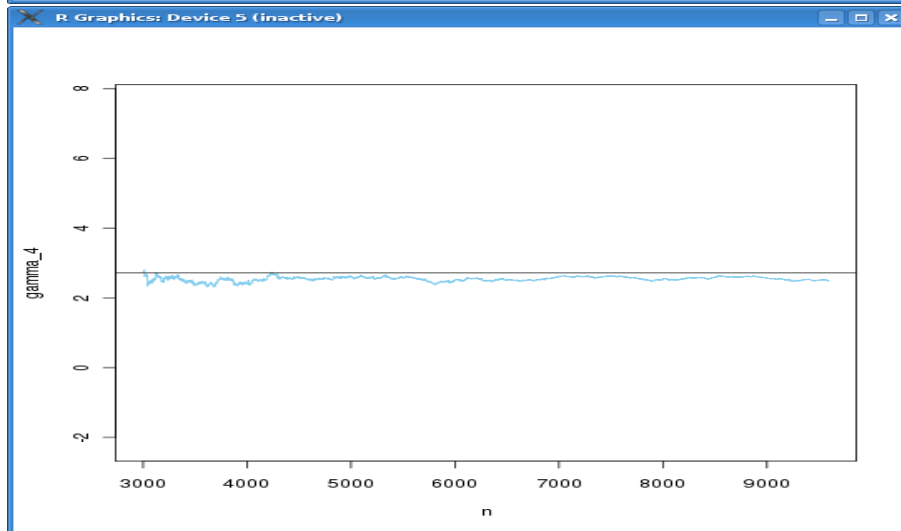
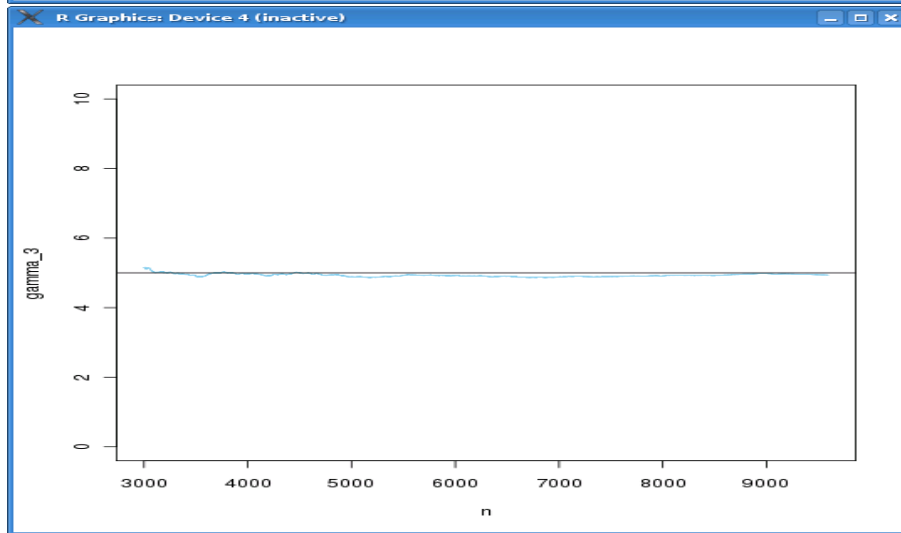
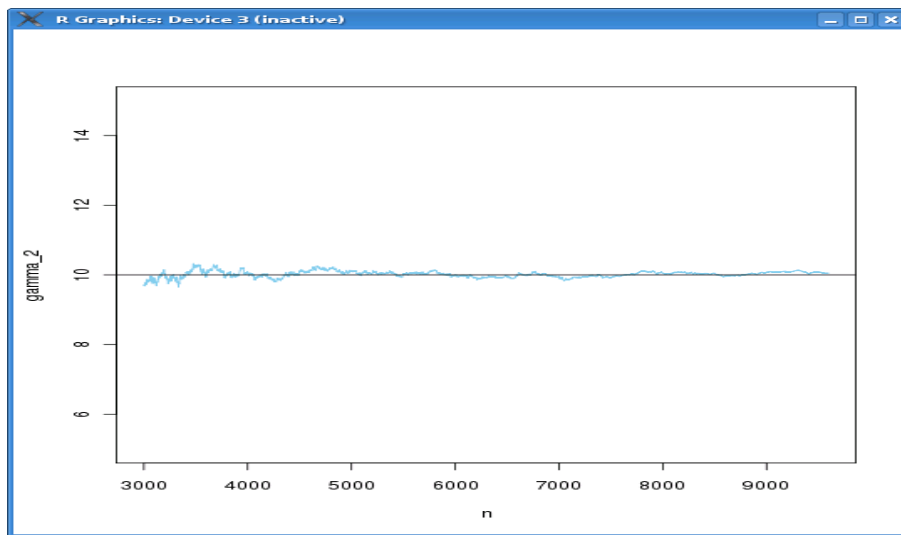
$$V_n^5 = 2 + \text{rbinom}(1, 1, 0.5) \left(2 + \mathcal{B}(1, 0.5)\right)$$

$$\delta_n^i = \left(\gamma^i, V_n^i\right)^{\frac{1}{2}} = (\gamma^i \cdot V_n^i)^{\frac{1}{2}}$$

$$D_n^i = \left(\langle C_n^i, V_n^i \rangle\right)^{\frac{1}{2}} = (C_n^i \cdot V_n^i)^{\frac{1}{2}}.$$

Résultats :





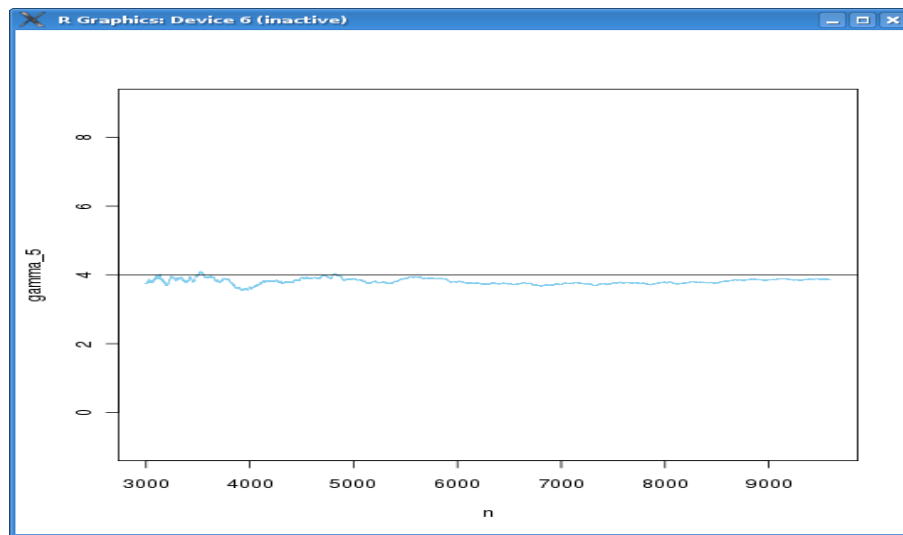
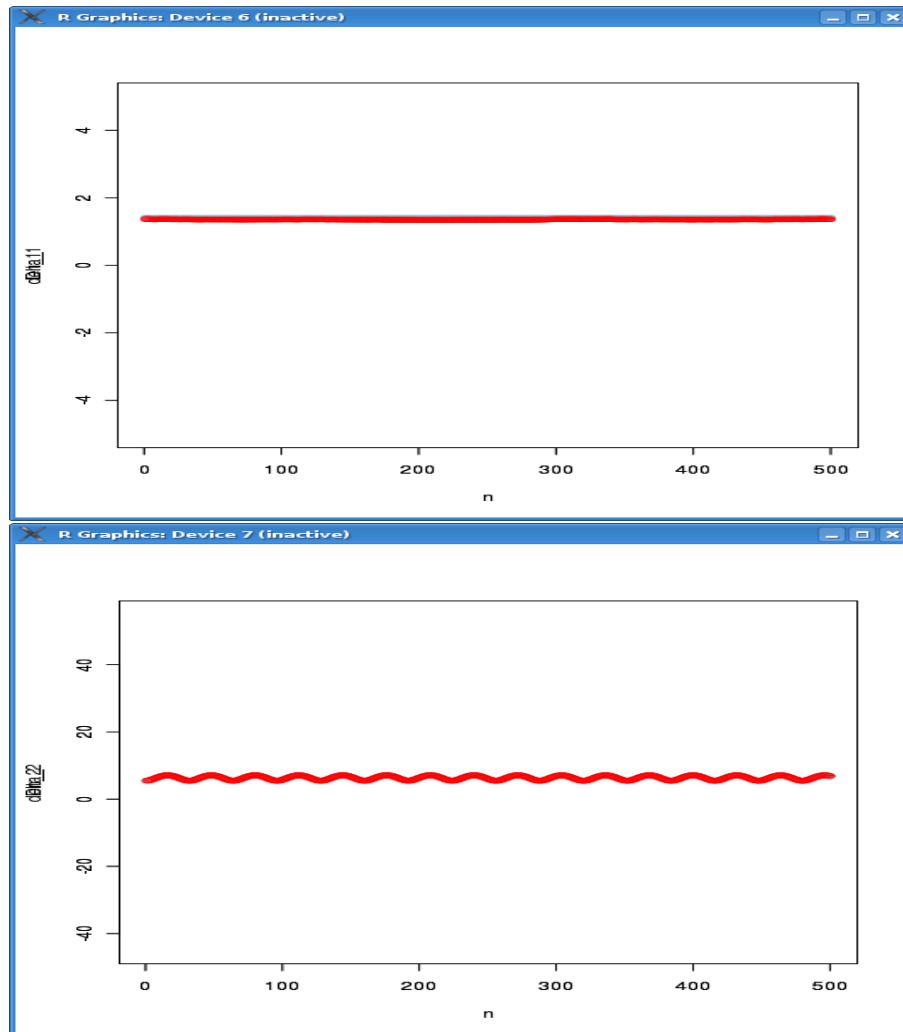


FIGURE 5.5: Estimation des γ_i

Comme pour l'estimation des β^i , les graphiques sont centrés en ordonnée sur la valeur des γ^i (ligne noire) ; l'évolution du processus C_n^i au cours du temps est représentée en bleu.



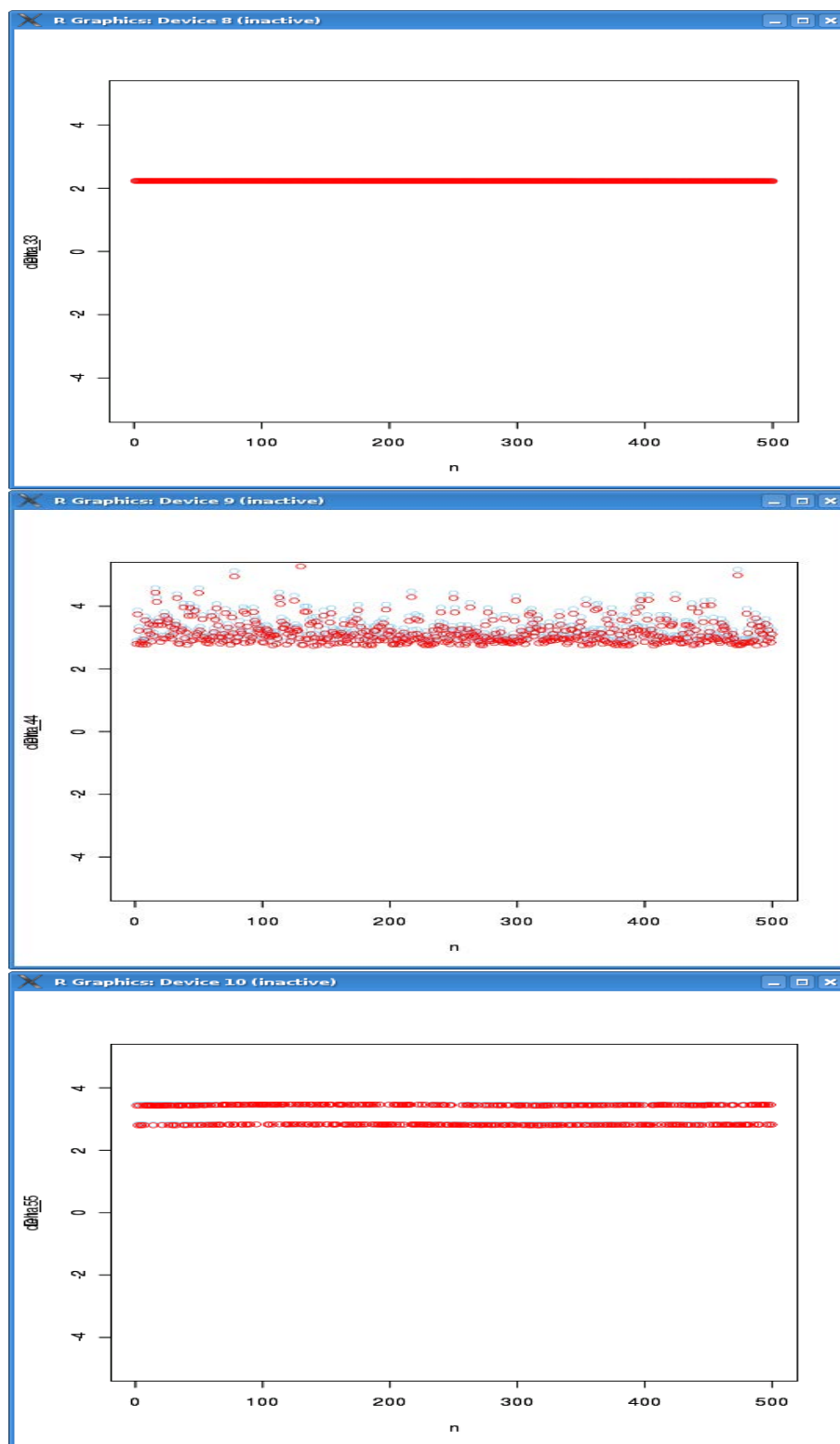


FIGURE 5.6: Estimation des δ_n^i

L'évolution du processus déterministe de l'écart-type δ_n^i est d'abord tracée en bleu puis, l'évolution du processus aléatoire de poursuite D_n^i est représentée sur le même graphique en rouge. On remarque que les tracés rouge et bleu sont quasiment superposés à la fin de la simulation (seules les 500 dernières itérations sont représentées ici), ceci témoigne de la bonne estimation de l'écart-type par le processus d'a.s.

Mise en oeuvre de l'ACG

Les étapes 1, 2 et 4 du cas précédent sont les mêmes.

On précise alors l'étape 3 :

3) Mise à jour au pas n :

On introduit tous les vecteurs d'observations depuis le pas précédent et on met en oeuvre, pour $l = 1, \dots, r$, le processus (X_n^l) décrit dans le chapitre 4 tel que :

$$\begin{aligned} F_n(X_n^l) &= \frac{\langle B_n X_n^l, X_n^l \rangle_{M_n}}{\|X_n^l\|_{M_n}^2}, \\ Y_{n+1}^l &= X_n^l + \frac{a}{n^\alpha} (B_n - F_n(X_n^l)I) X_n^l, \quad l = 1, \dots, r, \\ X_{n+1} &= \text{orth}_{M_n}(Y_{n+1}). \end{aligned}$$

Ainsi, on obtient, pour $l = 1, \dots, r$, une estimation de v_l .

Voici les résultats obtenus pour l'estimation de vecteurs directeurs des trois premiers axes principaux de l'ACG d'un vecteur Z de dimension 15 pendant une durée de 60 secondes.

Résultats :

— 1er vecteur — 2ème vecteur — 3ème vecteur

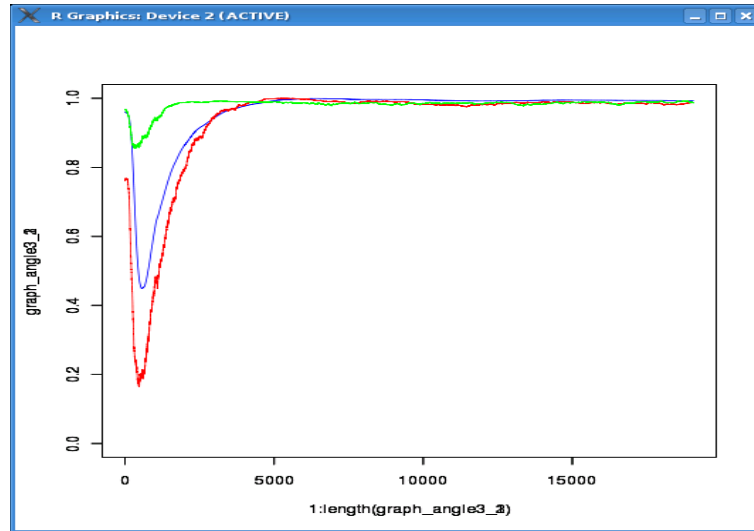


FIGURE 5.7: Cosinus en fonction du temps

On remarque qu'on a bien convergence des vecteurs calculés vers les vecteurs théoriques. De plus, l'estimation est satisfaisante à partir d'environ 5000 itérations pour les 3 vecteurs.

5.3 Package R : *factas*

Un package R intitulé *factas* a également été conçu pour permettre au plus grand nombre d'utiliser les méthodes d'analyse de données en ligne basées sur l'a.s. présentées dans ce travail de thèse. Celui-ci est disponible sur le site internet du CRAN à l'adresse : <http://cran.r-project.org/web/packages/factas/index.html>.

Le package permet la mise en oeuvre de la plupart des analyses factorielles : Analyse en Composantes Principales, Analyse Factorielle Multiple, Analyse Canonique, Analyse Canonique Généralisée, Analyse Factorielle Discriminante, Analyse Factorielle des Correspondances, Analyse Factorielle des Correspondances Multiples sur un tableau de données (cadre déterministe) ou sur un flux de données simulé dans le cas où les observations sont supposées indépendantes et identiquement distribuées.

La documentation complète (en anglais) est disponible à l'adresse web suivante : <http://cran.r-project.org/web/packages/factas/factas.pdf>.

Conclusion

L'étude menée durant cette thèse de doctorat a été largement motivée par les problématiques actuelles de données massives (Big Data) et les enjeux qui en découlent. L'objectif poursuivi tout au long de cette étude a été de développer des méthodes visant à analyser des flux de données d'un point de vue théorique et informatique. Plusieurs raisons nous ont poussés à utiliser des techniques d'approximation stochastique pour traiter ce problème :

- on ne dispose d'aucune indication sur la loi suivie par les données observées,
- on souhaite pouvoir disposer de résultats à tout instant,
- on veut prendre en compte le maximum de données possibles.

Ainsi, récursivité et simplicité des calculs ont été privilégiées pour répondre à ces contraintes.

Dans toute l'étude, nous avons considéré que ces données étaient des observations indépendantes d'un vecteur aléatoire.

Dans un premier temps, on a supposé que ces dernières étaient aussi identiquement distribuées. Dans ce contexte, on a étendu les travaux déjà réalisés sur plusieurs analyses factorielles au cas de l'ACP projetée. Plus précisément, on a défini une méthode récursive d'estimation en temps réel des r premiers facteurs de cette analyse. Ensuite, l'analyse canonique, l'analyse discriminante et enfin l'analyse factorielle des correspondances ont été traitées comme cas particuliers d'ACP projetée avec une métrique particulière. Dans chacun de ces cas, on a défini plusieurs processus spécifiques à l'analyse envisagée dans l'optique d'optimiser la mise en oeuvre informatique. Le chapitre 2, très largement indépendant des autres, est consacré à cette étude.

Ensuite, dans les chapitres 3 et 4, on s'est intéressé au cas où des éléments caractéristiques de la loi du vecteur aléatoire dont on observe des réalisations varient dans le temps. D'autre part, l'objectif est désormais d'estimer des vecteurs directeurs des r premiers axes principaux des analyses envisagées.

Tout d'abord, dans le chapitre 3, on a considéré l'ACP d'un flux de données dont l'espérance varie dans le temps. Dans cette étude on a principalement généralisé des résultats de [50], d'une part en proposant des processus alternatifs (en particulier des processus prenant en compte des paquets de données), d'autre part en n'imposant aucun modèle de représentation a priori pour l'espérance θ_n . Le cas de l'ACG a été traité comme cas particulier en introduisant l'approximation stochastique de l'inverse d'une matrice de covariance pour l'estimation de la métrique.

Enfin, le chapitre 4 a été consacré à l'ACP d'un flux de données dont l'espérance et la matrice de covari-

ance varient dans le temps. Cette démarche originale a abouti à la définition d'un processus d'estimation de vecteurs directeurs des r premiers axes principaux de l'ACP, processus prenant en compte une nouvelle donnée à chaque itération. Comme pour le chapitre précédent, le cas de l'ACG a été traité comme cas particulier en introduisant l'approximation stochastique de l'inverse d'une matrice de covariance pour l'estimation de la métrique.

On a dédié le chapitre 5 aux simulations informatiques, permettant ainsi de confirmer les résultats de convergence obtenus et de comparer les processus définis avec d'autres issus de méthodes déjà existantes. Enfin, on fait référence dans ce chapitre au package R : *factas* développé pendant la thèse et disponible sur les dépôts du CRAN, package qui permet de réaliser diverses analyses factorielles sur des flux de données i.i.d simulés.

Chacun des chapitres 2, 3 et 4 devrait donner lieu à une publication dans le courant de l'année à venir.

Les perspectives de cette étude sont multiples. On pourrait par exemple :

- étendre à d'autres analyses factorielles les résultats des chapitres 3 et 4,
- adapter les processus définis dans le chapitre 2 au contexte du chapitre 4,
- considérer des modèles de variation de l'espérance et/ou de la variance non linéaires,
- étendre le package *factas* au cas où l'espérance et/ou la matrice de covariance varient dans le temps,
- appliquer les processus à des jeux de données réelles...

Bibliographie

- [1] Albert A. and Gardner L. *Stochastic Approximation and nonlinear Regression*. 1967.
- [2] J. Baglama and L. Reichel. Augmented implicitly restarted lanczos bidiagonalization methods. *SIAM J. Scientific Computing*, 27(1) :19–42, 2005.
- [3] A. Benveniste, M. Metivier, and Priouret P. *Adaptive Algorithms and Stochastic Approximation*. Springer Verlag, 1990.
- [4] J.P. Benzecri. *L'analyse des données*. 1973.
- [5] J.P. Benzecri. Approximation stochastique dans une algèbre normée non commutative. *Bulletin de la Société Mathématique de France*, 97 :225–241, 1969.
- [6] A. Bordes, L. Bottou, and P. Gallinari. Careful quasi- newton stochastic gradient descent. *Journal of Machine Learning Research*, 10 :1737–1754, 2009.
- [7] L. Bottou and Y. Lecun. On-line learning for very large datasets. *Applied Stochastic Models in Business and Industry*, 21(2) :137–151, 2004.
- [8] A. Bouamaine. *Analyse factorielle séquentielle par approximation stochastique*. PhD thesis, Université de Nancy, 1986.
- [9] A. Bouamaine. Méthodes d'approximation stochastique en analyse des données. *Thèse de doctorat d'Etat ès Sciences Appliquées, Université Mohammed V, EMI, Rabat*, 1996.
- [10] A. Bouamaine and J.M. Monnez. Stochastic approximation of eigenvectors and eigenvalues. (approximation stochastique de vecteurs et valeurs propres.). *Annales de l'I.S.U.P.*, 42(2-3) :15–38, 1998.
- [11] C. Burt. The factorial analysis of qualitative data. *British J. of Statist. psychol.*, 3(3) :166–185, 1950.
- [12] J. Carroll. Generalization of canonical correlation to three or more sets of variables. *Proc. Amer. Psychological Assoc.*, pages 227–228, 1968.
- [13] H.F. Chen, L. Guo, and A.J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stoch. Proc. Appl.*, 27(2) :217–231, 1988.
- [14] X. Doukopoulos and G. Moustakides. Fast and stable subspace tracking. *IEEE Transactions on Signal Processing*, 56(4) :1452–1465, 2008.
- [15] M. Duflo. *Méthodes Récursives Aléatoires*. Springer Verlag, 1990.
- [16] M. Duflo. *Algorithmes Stochastiques*. Springer Verlag, 1996.
- [17] B. Escofier and J. Pagès. *Analyses factorielles simples et multiples. 3e édition*. 1998.
- [18] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2) :179–188, 1936.
- [19] J. Francis. The qr transformation, i. *The Computer Journal*, 4(3) :265–271, 1961.

- [20] J. Friedman and J. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23(9) :881–890, 1974.
- [21] R. Gemulla, E. Nijkamp, P. Haas, and Y. Sismanis. Large-scale matrix factorization with distributed stochastic gradient descent. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 69–77, New York, NY, USA, 2011. ACM.
- [22] E. Gladyshev. On stochastic approximations. *Theory of Probability and its Applications*, 10 :275–278, 1965.
- [23] L. Guttman. The quantification of a class of attributes : a theory and method of a scale construction. *SSCR New York*, pages 321–348, 1941.
- [24] P. Hall, D. Marshall, and R. Martin. Incremental eigenanalysis for classification. In *British Machine Vision Conference*, pages 286–295, 1998.
- [25] C. Hayashi. Multidimensional quantification. II. *Proc. Japan Acad.*, 30 :165–169, 1954.
- [26] C. Hayashi. *Quantification Method*. 1972.
- [27] H.O. Hirschfeld. A connection between correlation and contingency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4) :520–524, 1935.
- [28] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4) :321–377, 1936.
- [29] A. Juditsky. A stochastic estimation algorithm with observation averaging. *IEEE Trans. Automatic Control*, 38 :794–798, 1993.
- [30] H. Kargupta, R. Bhargava, K. Liu, Powers M., Blair P., S. Bushra, Sarkar K. Dull, J., M. Klein, M. Vasa, and Handy D. Vedas : A mobile and distributed data stream mining system for real-time vehicle monitoring. In *Proceedings of the Fourth SIAM International Conference on Data Mining*, KDD '11, pages 300–311, 2004.
- [31] H. Kesten. Accelerated stochastic approximation. *Annals Math. Statist.*, 29 :41–59, 1958.
- [32] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23(3) :462–466, 1952.
- [33] T.P. Krasulina. A method of stochastic approximation for the determination of the least eigenvalue of a symmetric matrix. *U.S.S.R. Comput. Math. Math. Phys.*, 9(6) :189–195, 1969.
- [34] V. Kublanovskaya. On some algorithms for the solution of the complete eigenvalue problem. *USSR Computational Mathematics and Mathematical Physics*, 1(3) :555–570, 1961.
- [35] H. Kushner. *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*. 1984.
- [36] H. Kushner and D. Clark. *Stochastic Approximation for Constrained and Unconstrained Systems*. 1978.
- [37] H. Kushner and J. Tukey. Rates of convergence for stochastic approximation type algorithms. *SIAM J. Control Optim.*, 17 :607–617, 1979.
- [38] H. Kushner and J. Yang. Stochastic approximation with averaging of the iterates : Optimal asymptotic rates of convergence for general processes. *SIAM J. Control Optim.*, 31 :1045–1062, 1993.
- [39] H. Kushner and J. Yang. Analysis of adaptive step size sa algorithms for parameter tracking. *IEEE Trans. Automatic Control*, AC-40 :1403–1410, 1995.
- [40] H. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.

- [41] L. Lebart. On the Benzécri's method for finding eigenvectors by stochastic approximation. *COMP-STAT, Physica verlag, Vienne*, pages 202–211, 1974.
- [42] L. Lebart, A. Morineau, and N. Tabard. *Techniques de la description statistique*. 1977.
- [43] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.
- [44] R. von Mises and H. Pollaczek-Geiringer. Praktische Verfahren der Gleichungsaufösung. *Z. Angew. Math. Mech.*, 9 :58–77, 1929.
- [45] J.M. Monnez. *Approximation stochastique : Le processus de Robbins-Monro. Mise à jour des résultats et quelques compléments*, Thèse de Doctorat ès Sciences Mathématiques. PhD thesis, Université de Nancy I, 1975.
- [46] J.M. Monnez. Etude d'un processus général multidimensionnel d'approximation stochastique sous contraintes convexes. Applications à l'estimation statistique. *Thèse de Doctorat d'Etat ès Sciences Mathématiques*, Université de Nancy I, 1982.
- [47] J.M. Monnez. Etude d'un modèle de corrélation en approximation stochastique. *Statistique et analyse des données*, 16(3) :143–181, 1992.
- [48] J.M. Monnez. Convergence d'un processus d'approximation stochastique en analyse factorielle. *Publ. de l'ISUP*, pages 37–56, 1994.
- [49] J.M. Monnez. Approximation stochastique en analyse factorielle multiple. *Publ. Inst. Stat. Univ. Paris*, 50(3) :27–45, 2005.
- [50] J.M. Monnez. Analyse en composantes principales d'un flux de données d'espérance variable dans le temps. *Revue des Nouvelles Technologies de l'Information*, (C-2) :43–56, 2008.
- [51] J.M. Monnez. Stochastic approximation of the factors of a generalized canonical correlation analysis. *Statistics and Probability Letters*, 78(14) :2210, 2009.
- [52] A. Nemirovski and D. Yudin. On cezari's convergence of the steepest descent method for approximating saddle point of convex-concave functions. *Soviet Math. Dokl.*, 1978.
- [53] T.M.N. Nguyen and J. Saracco. Recursive estimation for sliced inverse regression. *Journal de la Société Française de Statistique*, 151(2) :19–46, 2010.
- [54] E. Oja and J. Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *J. Math. Anal. Appl.*, 106 :69–84, 1985.
- [55] T. Oyama, S. Karungaru, S. Tsuge, Y. Mitsukura, and M. Fukumi. Incremental learning method of simple-pca. In Ignac Lovrek, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems, 12th International Conference, KES 2008, Zagreb, Croatia, September 3-5, 2008, Proceedings, Part II*, volume 5178 of *Lecture Notes in Computer Science*, pages 403–410. Springer, 2008.
- [56] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6) :559–572, 1901.
- [57] A. Plakhov and P. Cruz. A stochastic approximation algorithm with multiplicative step size adaptation. *arXiv*, math.ST/0503434, 2005.
- [58] B. Polyak and A. Juditsky. Acceleration of stochastic approximation by averaging. *Journal of Machine Learning Research*, 30(4) :838–855, 1992.

- [59] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22 :400–407, 1951.
- [60] H. Robbins and D. Siegmund. A convergence theorem for non negative almost super-martingales and some applications. *Optimizing Methods in Statistics*, pages 233–256, 1971.
- [61] Y. Saad. *Numerical Methods for Large Eigenvalue Problems, Second edition*. SIAM, 2011.
- [62] L. Schmetterer. Multidimensional stochastic approximation. *Multivariate Analysis II*, pages 443–460, 1969.
- [63] M. Shimura and T. Imai. Nonsupervised classification using the principal component. *Pattern Recognition*, 15 :353–363, 1973.
- [64] J. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37 :332–341, 1992.
- [65] C. Spearman. General intelligence, objectively determined and measured. *Amer. Journal of Psychology*, 15 :201–293, 1904.
- [66] P. Teixeira and R. Milidiú. Data stream anomaly detection through principal subspace tracking. In Sung Y. Shin, Sascha Ossowski, Michael Schumacher, Mathew J. Palakal, and Chih-Cheng Hung, editors, *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC), Sierre, Switzerland, March 22-26, 2010*, pages 1609–1616. ACM, 2010.
- [67] L. Thurstone. The vectors of mind. *Psychological Review*, 41(1) :1–32, 1934.
- [68] L. Thurstone. The principal components of scale analysis. *Measurement and prediction, Princeton University Press*, 1947.
- [69] J. Tukey. *Exploratory Data Analysis*. 1977.
- [70] J.H. Venter. On Dvoretzky stochastic approximation theorems. *Ann. Math. Stat.*, 37 :1534–1544, 1966.

Résumé. On suppose que des vecteurs de données de grande dimension arrivant en ligne sont des observations indépendantes d'un vecteur aléatoire.

Dans le second chapitre, ce dernier, noté Z , est partitionné en deux vecteurs R et S et les observations sont supposées identiquement distribuées. On définit alors une méthode récursive d'estimation séquentielle des r premiers facteurs de l'ACP projetée de R par rapport à S . On étudie ensuite le cas particulier de l'analyse canonique, puis de l'analyse factorielle discriminante et enfin de l'analyse factorielle des correspondances. Dans chacun de ces cas, on définit plusieurs processus spécifiques à l'analyse envisagée.

Dans le troisième chapitre, on suppose que l'espérance θ_n du vecteur aléatoire Z_n dont sont issues les observations varie dans le temps. On note $\tilde{Z}_n = Z_n - \theta_n$ et on suppose que les vecteurs $\tilde{Z}_n$ forment un échantillon indépendant et identiquement distribué d'un vecteur aléatoire $\tilde{Z}$. On définit plusieurs processus d'approximation stochastique pour estimer des vecteurs directeurs des axes principaux d'une analyse en composantes principales (ACP) partielle de $\tilde{Z}$. On applique ensuite ce résultat au cas particulier de l'analyse canonique généralisée (ACG) partielle après avoir défini un processus d'approximation stochastique de type Robbins-Monro de l'inverse d'une matrice de covariance.

Dans le quatrième chapitre, on considère le cas où à la fois l'espérance et la matrice de covariance de Z_n varient dans le temps.

On donne finalement des résultats de simulation dans le chapitre 5.

Mots-clés. Big Data, flux de données, Analyse en composantes principales (ACP), ACP projetée, analyse canonique généralisée (ACG), approximation stochastique.

Abstract. High dimensional data are supposed to be independent on-line observations of a random vector.

In the second chapter, the latter is denoted by Z and sliced into two random vectors R et S and data are supposed to be identically distributed. A recursive method of sequential estimation of the factors of the projected PCA of R with respect to S is defined. Next, some particular cases are investigated : canonical correlation analysis, canonical discriminant analysis and canonical correspondence analysis ; in each case, several specific methods for the estimation of the factors are proposed.

In the third chapter, data are observations of the random vector Z_n whose expectation θ_n varies with time. Let $\tilde{Z}_n = Z_n - \theta_n$ and suppose that the vectors $\tilde{Z}_n$ form an independent and identically distributed sample of a random vector $\tilde{Z}$. Stochastic approximation processes are used to estimate on-line direction vectors of the principal axes of a partial principal components analysis (PCA) of $\tilde{Z}$. This is applied next to the particular case of a partial generalized canonical correlation analysis (gCCA) after defining a stochastic approximation process of the Robbins-Monro type to estimate recursively the inverse of a covariance matrix.

In the fourth chapter, the case when both expectation and covariance matrix of Z_n vary with time n is considered.

Finally, simulation results are given in chapter 5.

Keywords. Big data, data streams, Principal components analysis (PCA), projected PCA, generalized canonical correlation analysis (gCCA), stochastic approximation.