

Estimation récursive pour les modèles autorégressifs localement stationnaires

Eric MOULINES¹, Pierre PRIOURET², François ROUEFF¹

¹GET / Telecom Paris, CNRS LTCI

46 rue Barrault, 75634 Paris Cedex 13, France

²Laboratoire de Probabilités, CNRS U.M.R. 7599

Université Paris VI - Tour 56 - 3ème étage, 4, Place Jussieu, 75252 Paris cedex 05, France

moulines@tsi.enst.fr, priouret@ccr.jussieu.fr, roueff@tsi.enst.fr

Résumé – Nous nous intéressons dans ce travail à l'estimation des paramètres d'un processus autorégressif localement stationnaire. Dans cette approche, le paramètre d'intérêt est une fonction de $[0,1]$ dans $\mathbb{R}^d \times \mathbb{R}_+$ décrivant l'évolution de la densité spectrale au cours du temps par l'intermédiaire de d coefficients de prédiction linéaire et de la variance de l'innovation. Le problème d'identification du modèle se ramène alors à un problème d'estimation fonctionnelle non-paramétrique. Nous nous intéressons à la mise en oeuvre d'estimateurs récursifs. Nous avons étudié plus particulièrement l'estimateur LMS (Least Mean Square) qui repose sur l'écriture d'un gradient stochastique. Nous avons étudié le comportement asymptotique précis de l'estimateur, montrant que pour des classes de régularité trop forte, il n'atteint pas la vitesse minimax. Dans ce cas, nous appliquons une méthode de correction du biais pour construire un nouvel estimateur récursif de vitesse minimax dans un ensemble plus large de classes de régularité.

Abstract – We consider the problem of estimating the parameters of a locally stationary autoregressive process. This approach models the time evolution of the spectral content of a time series by a $[0, 1] \rightarrow \mathbb{R}^d \times \mathbb{R}_+$ mapping of d linear prediction coefficients and the innovation variance. The identification problem for this model fits the classical non-parametric curve estimation theory. In this contribution we focus on recursive estimators and more particularly on the LMS (Least Mean Square) algorithm. This estimator is based on stochastic gradient approach. A precise study of its asymptotic behaviour is provided. It turns out that this estimator achieves the minimax rate only in alimited range of smoothness classes. We propose a bias reduction method which allows to achieve this rate in a wider range of smoothness classes.

1 Contexte

Nous nous intéressons dans ce travail à l'estimation des paramètres de processus autorégressifs localement stationnaires (voir [5], [3], [4] et [2]). Ces modèles sont définis par l'équation de récurrence, pour $1 \leq k \leq n$:

$$X_{k,n} = \sum_{i=1}^d \theta_i((k-1)/n) X_{k-i,n} + \sigma(k/n) \epsilon_{k,n}, \quad (1)$$

$$= \boldsymbol{\theta}_{k-1,n}^T \mathbf{X}_{k-1,n} + \sigma_{k,n} \epsilon_{k,n} \quad (2)$$

où, pour tout $1 \leq k \leq n$, $\mathbf{X}_{k,n} := [X_{k,n} \cdots X_{k-d+1}]^T$, $\boldsymbol{\theta}(\cdot) := [\theta_1(\cdot) \cdots \theta_d(\cdot)]^T$ et $\sigma(\cdot)$ sont des fonctions définies sur $[0, 1]$ et $(\epsilon_{1,n}, \dots, \epsilon_{n,n})$ sont indépendants, centrées et de variance unité et indépendants de $\{\mathbf{X}_{0,n}, n \geq 1\}$. On a de plus noté $\boldsymbol{\theta}_{k,n} := \boldsymbol{\theta}(k/n)$ et $\sigma_{k,n} := \sigma(k/n)$.

La plupart des approches développées à ce jour pour estimer les fonctions $t \rightarrow \theta_i(t)$ et $t \rightarrow \sigma(t)$ sont basées sur des approches hors ligne. Nous étudions ici des méthodes d'estimation récursives, étendant ainsi les travaux de [2]. L'algorithme standard d'estimation récursive est donné par :

$$\begin{aligned} \hat{\epsilon}_{k+1,n} &= X_{k+1,n} - \hat{\boldsymbol{\theta}}_{k,n}^T(\mu) \mathbf{X}_{k,n}, \\ \hat{\boldsymbol{\theta}}_{k+1,n}(\mu) &= \hat{\boldsymbol{\theta}}_{k,n}(\mu) + \mu \hat{\epsilon}_{k+1,n} \mathbf{Y}_{\mu,k,n}, \\ \hat{\sigma}_{k+1,n}^2 &= \hat{\sigma}_{k,n}^2 + \mu (\hat{\epsilon}_{k+1,n}^2 - \hat{\sigma}_{k,n}^2). \end{aligned} \quad (3)$$

où μ est le pas d'adaptation et $\mathbf{Y}_{\mu,k,n}$ est un vecteur dépendant de μ et des observations $\mathbf{X}_{l,n}$, $l \leq n$. L'estimée $\hat{\boldsymbol{\theta}}_{0,n}(\mu)$ est

déterministe indépendante de n (typiquement $\hat{\boldsymbol{\theta}}_{0,n}(\mu) = 0$). La méthode la plus élémentaire est l'algorithme LMS, pour lequel $\mathbf{Y}_{\mu,k,n} = \mathbf{X}_{k,n}$. Une variante algorithmique utile est le NLMS, pour lequel nous prendrons $\mathbf{Y}_{\mu,k,n} = (1 + \mu \|\mathbf{X}_{k,n}\|^2)^{-1} \mathbf{X}_{k,n}$, où $\|\cdot\|$ est la norme euclidienne.

2 Principaux résultats

Le comportement de l'algorithme LMS pour les modèles TVAR a récemment été étudié par [2] sous des hypothèses très restrictives (modèle autorégressif d'ordre 1 et innovation bornée). Nous proposons dans ce travail une étude sous des hypothèses beaucoup plus générales. A notre connaissance, bien que les algorithmes d'estimation récursifs soient couramment utilisés pour estimer les paramètres de modèles autorégressifs, il n'existe pas de preuve convaincante de convergence pour l'algorithme à pas fixe (qui ne soient pas basée sur des approximations erronées - hypothèses d'indépendance). La stationnarité locale permet de caractériser la capacité de poursuite des algorithmes dans un cadre tout aussi pertinent que celui utilisé généralement des paramètres diffusifs, mal adaptés aux modèles autorégressifs.

Pour $\beta \in (0, 1]$, la β -Lipschitz semi-norme de la fonction $f : [0, 1] \mapsto \mathbb{R}^d$ est donnée par :

$$\|f\|_{\Lambda, \beta} = \sup_{t \neq s} \frac{|f(t) - f(s)|}{|t - s|^\beta}. \quad (4)$$

Pour tout variable aléatoire Z à valeur dans \mathbb{R}^m , $m \geq 1$, muni de la norme euclidienne $|\cdot|$, on note

$$\|Z\|_p := (\mathbb{E}[|Z|^p])^{1/p}.$$

Considérons les hypothèses suivantes.

(A1) Il existe $\beta \in (0, 1]$ tel que $|\theta|_{\Lambda, \beta} < \infty$ et $|\sigma|_{\Lambda, \beta} < \infty$.

(A2) Il existe $\rho \in (0, 1)$ tel que le polynôme caractéristique $\theta(z; t) = 1 - \sum_{j=1}^p \theta_j(t) z^j$ soit non nul pour tout $t \in [0, 1]$ et tout $|z| \leq \rho^{-1}$.

(A3) $\{\epsilon_{k,n}\}_{n \geq 1, 1 \leq k \leq n}$ est un tableau triangulaire de v.a. indépendantes, centrées et de variance unité et la variance locale est uniformément minorée et majorée, i.e. il existe $0 < \sigma_- < \sigma_+ < \infty$ tels que $\sigma_- \leq \sigma(t) \leq \sigma_+$ pour tout $t \in [0, 1]$.

(A4) Il existe $q \geq 4$ tel que $\sup_{n \geq 1} \|\mathbf{X}_{0,n}\|_q$ et $\sup_{1 \leq k \leq n} \|\epsilon_{k,n}\|_q$

soient finis.

(A'4) Il existe $\alpha > 0$ et $u > 0$ tels que $\sup_{n \geq 1} \mathbb{E} \exp(\alpha |\mathbf{X}_{0,n}|^u)$

et $\sup_{1 \leq k \leq n} \mathbb{E} \exp(\alpha |\epsilon_{k,n}|^u)$ soient finis.

Commentons succinctement ces hypothèses. L'hypothèse (A1) est une hypothèse classique d'estimation fonctionnelle. Elle se comprend aisément puisqu'elle contrôle l'évolution des paramètres dans le temps. Plus l'indice β est grand et plus cette évolution est régulière et l'estimation est facilitée. L'hypothèse (A2) permet d'établir la stabilité du modèle. Pensons au cas où le paramètre est constant au cours du temps. Dans ce cas, le modèle suit l'équation de récurrence d'un processus AR stationnaire et (A2) revient à la condition habituelle d'un AR stable causal. Dans le cas d'un processus localement stationnaire, cette hypothèse continue d'opérer mais dans le contexte d'une chaîne inhomogène. Nous ne nous étendrons pas ici sur la notion de stabilité dans ce contexte et renvoyons à [10] pour une étude détaillée. L'hypothèse (A3) est une hypothèse classique d'innovations de type bruit blanc avec une variance évoluant au cours du temps. La condition de minoration uniforme de cette variance par une constante positive est cruciale pour démontrer la persistance d'excitation des algorithmes récursifs qui est un ingrédient fondamental des systèmes adaptatifs (voir [9]). L'hypothèse (A4) est une hypothèse de majoration uniforme sur les moments des innovations. C'est une hypothèse essentielle et extrêmement faible comparée à celle utilisée par [2] qui suppose les innovations bornées. C'est l'avantage apporté par la normalisation de l'algorithme NLMS. Néanmoins, même dans le cas du LMS, l'hypothèse de [2] peut être largement améliorée, puisque dans ce cas, nous nous contenterons de (A'4), c'est-à-dire d'un contrôle uniforme des moments exponentiels des innovations.

Nous supposons (A1)-(A4). Une conséquence importante de ces hypothèses est la stabilité du processus défini par (1) (voir [10]), en particulier :

$$X_q^* := \sup_{k,n} \|X_{k,n}\|_q < \infty. \quad (5)$$

Dans le cadre du LMS l'hypothèse (A4) est renforcée par l'hypothèse (A'4) mais ceci n'est pas nécessaire dans le cadre du NLMS. On obtient alors le résultat suivant.

Théorème 1 Pour tout $p \in [1, q/3]$, il existe $M, \delta > 0$ et $\mu_0 > 0$ tels que pour tout $\mu \in (0, \mu_0]$, pour tout $n \geq 1$ et pour tout

$k = 1, \dots, n$,

$$\|\widehat{\theta}_{k,n}(\mu) - \theta_{k,n}\|_p \leq M \left(|\widehat{\theta}_{0,n}(\mu) - \theta(0)| (1 - \delta \mu)^k + \sqrt{\mu} + (n\mu)^{-\beta} \right). \quad (6)$$

Résumé de la preuve. Notons $\delta_{k,n} = \widehat{\theta}_{k,n} - \theta_{k,n}$ l'erreur d'estimation. D'après (3), cette erreur satisfait l'équation de récurrence

$$\delta_{k+1,n} = (I - \mu \mathbf{F}_{k,n}) \delta_{k,n} + \xi_{k,n}, \quad 0 \leq k \leq n-1,$$

où $\xi_{k,n} := \theta_{k,n} - \theta_{k+1,n} + \mu \sigma_{k+1,n} \epsilon_{k+1,n} \mathbf{Y}_{k,n}$ et $\mathbf{F}_{k,n} := \mathbf{Y}_{k,n} \mathbf{X}_{k,n}^T$. Notons enfin, pour tout $-1 \leq j \leq k \leq n$,

$$\Psi_n(k, j; \mu) := (I - \mu \mathbf{F}_{k,n})(I - \mu \mathbf{F}_{k-1,n}) \cdots (I - \mu \mathbf{F}_{j+1,n}),$$

avec la convention $\Psi_n(k, k; \mu) := I$. D'après l'équation de récurrence vérifiée ci-dessus, on obtient, pour tout $0 \leq k \leq n$,

$$\delta_{k,n} = \Psi_n(k-1, -1; \mu) \delta_{0,n} + \sum_{j=0}^{k-1} \Psi_n(k-1, j; \mu) \xi_{j,n}. \quad (7)$$

La majoration (6) repose alors essentiellement sur le résultat suivant.

Lemme 2 Il existe des constantes $M, \delta > 0$ et $\mu_0 > 0$ telles que, pour tout $\mu \in [0, \mu_0]$ et pour tout $-1 \leq j \leq k \leq n$,

$$\|\Psi_n(k, j; \mu)\|_p \leq M(1 - \mu \delta)^{k-j}. \quad (8)$$

La preuve de ce lemme requiert une étude approfondie. Elle repose sur l'utilisation d'idées propres aux algorithmes récursifs (qui remontent à [6], [11], [7]) remises à jour par l'utilisation de techniques simples basées sur les propriétés de stabilité des chaînes de Markov inhomogènes. Elle comporte essentiellement trois étapes. Une première étape concerne l'étude de la stabilité du modèle, c'est-à-dire du processus $\{X_{k,n}, 1 \leq k \leq n\}$. La deuxième étape consiste à prouver la persistance d'excitation, qui permet de minorer *en moyenne* le gradient stochastique à chaque itération de l'algorithme (3). Enfin l'étape la plus difficile consiste à montrer un résultat de stabilité sur un produit de matrices aléatoires sous des hypothèses adaptées aux chaînes inhomogènes. Ce résultat est donné en appendice A. Nous renvoyons à [10] pour une étude détaillée des propriétés de stabilité du modèle, des résultats sur la persistance d'excitation, sur l'établissement du théorème 5 et son application dans le cadre de l'algorithme NLMS. L'adaptation au LMS est en cours de rédaction. Concluons maintenant sur l'utilisation du lemme 2 dans le cadre du théorème. L'équation (7) et ce lemme donne directement

$$\|\delta_{k,n}\|_p \leq M |\delta_{0,n}| (1 - \delta \mu)^k + \left\| \sum_{j=0}^{k-1} \Psi_n(k-1, j; \mu) \xi_{j,n} \right\|_p.$$

Le premier terme de cette borne donne le premier terme de (6). Le second terme de cette borne est contrôlé par une transformation d'Abel. On remarque que les sommes partielles vérifient, d'une part, d'après (A1),

$$\left| \sum_{i=j}^{k-1} (\theta_{i,n} - \theta_{i+1,n}) \right| = |\theta(i/n) - \theta(k/n)| \leq \frac{|\theta|_{\Lambda, \beta}}{n} |k - i|^\beta,$$

d'autre part, d'après (A3), $\{\sigma_{i+1,n} \mathbf{Y}_{i,n} \epsilon_{i+1,n}, i \geq 0\}$ est un processus d'incrément de martingale, et l'inégalité de Burkholder de [8, Theorem 2.12] donne, sous (A4),

$$\left\| \sum_{i=j}^{k-1} \mu \sigma_{i+1,n} \mathbf{Y}_{i,n} \epsilon_{i+1,n} \right\|_q \leq \mu \sigma_+ \mathbf{Y}_q^* \epsilon_q^* \sqrt{k-j},$$

où $\epsilon_q^* := \sup_{1 \leq k \leq n} \|\epsilon_{k,n}\|_q$ et $\mathbf{Y}_q^* := \sup_{\mu, k, n} \|\mathbf{Y}_{k,n}\|_q$. Enfin, on a, par inégalité d'Hölder, pour tout $s \geq 1$,

$$\|\Psi_n(k-1, j; \mu) - \Psi_n(k-1, j-1; \mu)\|_{sq/(2s+q)} \leq \mu \mathbf{F}_{q/2}^* \|\Psi_n(k-1, j-1; \mu)\|_s,$$

où $\mathbf{F}_{q/2}^* := \sup_{\mu, k, n} \|\mathbf{F}_{k,n}\|_{q/2}$. L'éq. (5) implique $\mathbf{Y}_q^* < \infty$ et $\mathbf{F}_{q/2}^* < \infty$. Après transformation d'Abel, les trois dernières équations ainsi que (8) donnent (6) pour $p \in [1, q/3]$ en choisissant s suffisamment grand et en utilisant le lemme 6 mis en appendice B. \square

Ce résultat prouve en particulier la stabilité en temps infini de l'algorithme. Il est intéressant d'analyser le rôle joué par les différents termes de la borne. Le premier terme est lié à l'oubli des conditions initiales. Les termes $\sqrt{\mu}$ et $(n\mu)^{-\beta}$ sont associés aux fluctuations (erreur de désajustement) et au biais (erreur de poursuite, ce dernier terme s'annulant lorsque les coefficients de régression sont constants). Il est intéressant de remarquer que la forme de cette borne est très similaire (au terme d'oubli des conditions initiales près) aux bornes sur l'erreur L^p que l'on obtient classiquement dans les problèmes d'estimation fonctionnelle non-paramétrique dans des classes de régularité Hölderienne. Concluons cette section par un simple corollaire du théorème 1 concernant l'estimation de θ en tant que fonction $[0, 1] \mapsto \mathbb{R}^d$. Soit

$$\widehat{\theta}_n(t; \mu) := \widehat{\theta}_{[tn], n}(\mu),$$

où $[u]$ désigne la partie entière de u . On a alors, sous les hypothèses du théorème,

Corollaire 3 *Soit $c > 0$ et $\eta \in (0, 1)$. Il existe $M > 0$ tel que, pour tout $n \geq 1$,*

$$\sup_{t \in [\eta, 1]} \|\theta(t) - \widehat{\theta}_n(t; cn^{-2\beta/(1+2\beta)})\|_p \leq Mn^{-\beta/(1+2\beta)}.$$

La vitesse obtenue correspond en fait à la vitesse optimale au sens minimax (voir [10]).

3 Correction du biais

On note $\Sigma(t, \theta, \sigma)$ la matrice $d \times d$ d'auto-covariance locale au temps t du processus $\{X_{k,n}, 1 \leq k \leq n\}$ de paramètre (θ, σ) , c'est-à-dire (voir [3]) la matrice

$$[\Sigma(t, \theta, \sigma)]_{1 \leq k, l \leq d} := \frac{\sigma^2(t)}{2\pi} \int_{-\pi}^{\pi} \frac{\exp(i\lambda(k-l))}{|\theta(\exp(i\lambda(k-l)); t)|^2} d\lambda,$$

où $\theta(z; t) := 1 - \sum_{i=1}^d \theta_i z^i$ est le polynôme de prédiction local au temps t . La semi-norme définie par (9) pour $\beta \in (0, 1]$ se généralise à $\beta > 1$ de la façon suivante. Soient $k \in \mathbb{N}$ et $\alpha \in (0, 1]$ tels que $\beta = k + \alpha$. On définit alors

$$|f|_{\Lambda, \beta} = \sup_{t \neq s} \frac{|f^{(k)}(t) - f^{(k)}(s)|}{|t - s|^\alpha}. \quad (9)$$

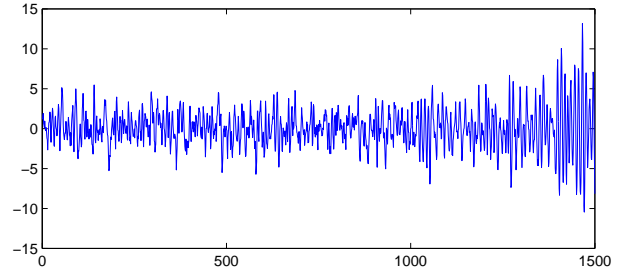


FIG. 1: Simulation d'un processus TVAR(2)

En utilisant les méthodes de développement de perturbation que nous avons adapté de travaux précédents [1], on montre (voir [10]) le résultat suivant.

Théorème 4 *On suppose ici que $q > 4$ et $\beta \in (1, 2]$. Soit $p \in [1, q/4]$. Alors, pour tout $\eta \in (0, 1)$, il existe $M > 0$ et $\mu_0 > 0$ tel que, pour tout $t \in [\eta, 1]$, pour tout $n \geq 1$ et pour tout $\mu \in (0, \mu_0]$,*

$$\left\| \widehat{\theta}_n(t; \mu) - \theta(t) + (\mu n)^{-1} \Sigma^{-1}(t, \theta, \sigma) \dot{\theta}(t) \right\|_p \leq M (\sqrt{\mu} + (\mu n)^{-\beta} + (\mu n)^{-2}). \quad (10)$$

Ce résultat montre que le théorème 1 et son corollaire n'est valide que pour $\beta \leq 1$ et que, sous des hypothèses de régularité plus forte correspondant à $\beta > 1$, l'estimateur $\widehat{\theta}_n(t; cn^{-2\beta/(1+2\beta)})$ n'a plus une vitesse minimax. Ceci est causé par la présence du terme en $(\mu n)^{-1}$ dans (10). La méthode de Romberg permet de corriger le biais induit par ce terme. Soit $\gamma \in (0, 1)$ fixé et notons

$$\tilde{\theta}_n(t; \mu, \gamma) := \frac{1}{1-\gamma} \left(\widehat{\theta}_n(t; \mu) - \gamma \widehat{\theta}_n(t; \gamma\mu) \right).$$

On obtient alors facilement, sous les hypothèses du théorème 4,

$$\sup_{t \in [\eta, 1]} \left\| \tilde{\theta}_n(t; cn^{-2\beta/(1+2\beta)}, \gamma) - \theta(t) \right\|_p \leq Mn^{-\frac{\beta}{1+2\beta}}.$$

On voit que cet estimateur permet d'étendre le corollaire 3 pour obtenir la vitesse minimax correspondant au cas $\beta \in (1, 2]$.

4 Simulations

L'estimateur $\widehat{\theta}$ et sa variante $\tilde{\theta}$ ont été implémentés sur des données simulées. Le processus AR(2) localement stationnaire simulé est représenté sur la figure 1. La figure 2 montre l'évolution des deux coordonnées de $\theta(t)$ en fonction de $t \in [0, 1]$ et ses deux valeurs estimées. On remarque le temps d'oubli à l'erreur initiale et la légère amélioration apportée par $\tilde{\theta}$ du point de vue du biais.

Appendice

A Stabilité d'un produit de matrices aléatoires

Théorème 5 *Soit $\{(A_k(\mu), \mathcal{F}_k), k \in \mathbb{N}\}$ une suite de processus adaptés à valeurs dans l'ensemble de matrices symétriques positives $d \times d$. Soient ϕ_k, V_k deux processus \mathcal{F}_k -adaptés*

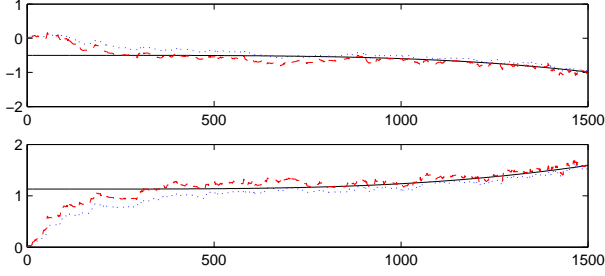


FIG. 2: Estimation NLMS due $\theta_1(t)$ (haut) et $\theta_2(t)$ (bas) à partir des observations simulées. En trait plein est représenté la vraie valeur du paramètre, en trait interrompu rouge la valeur estimée avec $\hat{\theta}$, en trait bleu pointillé la valeur estimée par $\hat{\theta}$.

de nombres positifs avec $V_k \geq 1$. Soit r un entier strictement positif, et R_1 et μ_1 deux constantes strictement positives. Supposons que les propositions suivantes soient vérifiées.

(a) Il existe $\lambda < 1$ et $B \geq 1$ tels que, pour tout $k \in \mathbb{N}$,

$$\mathbb{E}^{\mathcal{F}_k} [V_{k+r}] \leq \lambda V_k \mathbf{I}(\phi_k > R_1) + B V_k \mathbf{I}(\phi_k \leq R_1).$$

(b) Il existe $\alpha_1 > 0$ tel que, pour tout $k \in \mathbb{N}$ et pour tout $\mu \in [0, \mu_1]$,

$$\mathbb{E}^{\mathcal{F}_k} \left[\lambda_{\min} \left(\sum_{l=k+1}^{k+r} A_l(\mu) \right) \right] \geq \alpha_1 \mathbf{I}(\phi_k \leq R_1),$$

Supposons de plus que l'un des deux couples de propositions suivantes soit vérifiée.

(c₁) (i) pour tout $k \in \mathbb{N}$ et $\mu \in [0, \mu_1]$, $|\mu A_k(\mu)| \leq 1$
(ii) il existe $s > 1$ et $C_1 > 0$ tels que, pour tout $k \in \mathbb{N}$, $\mu \in [0, \mu_1]$ et $l = 1, \dots, r$,

$$\mathbf{I}(\phi_k \leq R_1) \mathbb{E}^{\mathcal{F}_k} [|A_{k+l}(\mu)|^s] \leq C_1,$$

(c₂) (i) Il existe $\rho > 0$, $u > 0$ et $C_2 > 0$ tels que, pour tout $k \in \mathbb{N}$, $\mu \in [0, \mu_1]$ et $l = 1, \dots, r$,

$$\mathbf{I}(\phi_k \leq R_1) \mathbb{E}^{\mathcal{F}_k} [\exp(\rho |A_{k+l}(\mu)|^u)] \leq C_2,$$

(ii) il existe $\xi > 1$, $v > 0$, $\eta > 0$ et $C_3 > 0$ tels que, pour tout entier $n \geq 1$ et pour tout $\mu \in [0, \mu_1]$,

$$\mathbb{E}^{\mathcal{F}_0} \left[\exp \left\{ \xi \sum_{l=1}^n \log |A_l(\mu)| \mathbf{I}(\mu |A_l(\mu)| > 1) \right\} \right] \leq (1 + C_3 \mu^{1+\eta})^n V_0^v.$$

Alors, pour tout $p \geq 1$, il existe $C_0 > 0$, $\delta_0 > 0$ et $\mu_0 > 0$ dépendant uniquement des constantes ci-dessus tels que, pour tout $\mu \in [0, \mu_0]$ et pour tout entier $n \geq 1$,

$$\mathbb{E}^{\mathcal{F}_0} \left[\left| \prod_{i=1}^n (I - \mu A_i) \right|^p \right] \leq C_0 (1 - \delta_0 \mu)^n V_0^w,$$

où $w = 1$ sous (c₁) et $w = 1 + v/\xi$ sous (c₂).

Commentons brièvement les hypothèses de ce théorème. L'hypothèse (a) est liée à la stabilité du processus, (b) est une propriété de type persistence d'excitation. L'utilisation des hypothèses (a), (b) et (c1) est explicitée dans [10]. L'hypothèse (c2) est nécessaire à l'adaptation de ces résultats dans le cadre du LMS (voir [11]).

B Lemme technique

Lemme 6 Soit $\beta \geq 0$ et $\mu \in (0, 1)$. Alors il existe des constantes C_1, C_2 dépendant uniquement de β telles que

$$\sum_{s=1}^{\infty} (1 - \mu)^s s^\beta \leq C_1 \mu^{-(1+\beta)},$$

$$\sup_{t>0} t^\beta (1 - \mu)^t \leq C_2 \mu^{-\beta}.$$

Références

- [1] Rafik Aguech, Eric Moulines, and Pierre Priouret. On a perturbation approach for the analysis of stochastic tracking algorithm. *SIAM Journal on Control and Optimization*, 39:872–899, 2000.
- [2] E. Belitser. Recursive estimation of a drifted autoregressive parameter. *The Annals of Statistics*, 28:860–870, 2000.
- [3] R. Dahlhaus. Fitting time series models to non-stationary processes. *Annals of Statistics*, 25:1–37, 1997.
- [4] R. Dahlhaus and L. Giraitis. On the optimal segment length for parameter estimates for locally stationary processes. *Annals of Statistics*, 19:629–655, 1998.
- [5] Y. Grenier. Time-dependent ARMA modeling of nonstationary signals. *IEEE Transactions on ASSP*, 31(4):899–911, 1983.
- [6] L. Guo. Stability of recursive stochastic tracking algorithms. *SIAM J. on Control and Optimization*, 32:1195–1125, 1994.
- [7] L. Guo and L. Ljung. Exponential stability of general tracking algorithms. *IEEE Trans. on Automatic Control*, 40:1376–1387, 1995.
- [8] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press Inc., New York, 1980.
- [9] T. Kailath. *Linear Systems*. Prentice Hall, 1980.
- [10] E. Moulines, P. Priouret, and F. Roueff. On recursive estimation for locally stationary time varying autoregressive processes. *Prépublication du laboratoire de Probabilité de Paris 6*, may 2003.
- [11] P. Priouret and A. Yu. Veretennikov. A remark on stability of the L.M.S. tracking algorithm. *Stochastic Analysis and its Applications*, 16(1):118–128, 1995.