Electronic Journal of Probability

Vol. 4 (1999) Paper no. 8, pages 1–28.

Journal URL
http://www.math.washington.edu/~ejpecp/
Paper URL
http://www.math.washington.edu/~ejpecp/EjpVol4/paper8.abs.html

MODERATE DEVIATIONS FOR STABLE MARKOV CHAINS AND REGRESSION MODELS

Julien WORMS

Laboratoire de Mathématiques appliquées,Université de Marne La Vallée,
5, bd Descartes, Champs sur Marne,77454 Marne la Vallée CEDEX, FRANCE
worms@math.univ-mlv.fr

Abstract: We prove moderate deviations principles for

1) unbounded additive functionals of the form $S_n = \sum_{j=1}^{n} g(X_{j-1}^{(p)})$, where $(X_n)_{n\in\mathbb{N}}$ is a stable $\mathbb{R}^d$-valued functional autoregressive model of order p with white noise, and g is an $\mathbb{R}^q$-valued Lipschitz function of order (r, s);

2) the error of the least squares estimator (LSE) of the matrix θ in an $\mathbb{R}^d$-valued regression model $X_n = \theta^t \phi_{n-1} + \epsilon_n$, where (ϵ_n) is a "generalized Gaussian" noise.

We apply these results to study the error of the LSE for a stable $\mathbb{R}^d$-valued linear autoregressive model of order p.

Keywords: Large and Moderate Deviations, Martingales, Markov Chains, Least Squares Estimator for a regression model.

AMS Subject Classification: 60F10 ; 60J10 ; 62J05 ; 62J02.

Submitted to EJP on December 31, 1997. Final version accepted on April 16, 1999.

1 Introduction

- This work is composed of two parts. In the first one we obtain Chernov-type upper bounds and a moderate deviations principle (shortened to MDP) for unbounded continuous additive functionals of a specific class of Markov chains appearing mostly in statistics, namely for stable autoregressive models of order $p \geq 1$ with white noise (results are introduced and stated in Section **2**, the proofs are performed in Section **4**).

  Let $(X_j)_{j>-p}$ be such an $\mathbb{R}^d$-valued stable model, and denote by μ the unique invariant probability measure of the model $(X_{j-p+1}, ..., X_{j-1}, X_j)_{j\geq 0}$; let $g : \mathbb{R}^{dp} \to \mathbb{R}^q$ be a continuous function (with a growth rate related to moments assumptions on the noise), and let

$$S_n = \sum_{j=0}^{n-1} \{ g(X_{j-p+1}, \ldots, X_{j-1}, X_j) - \textstyle\int g \, d\mu \}.$$

  Proposition 1 states the exponential convergence in probability of S_n/n to 0. In Theorem 2 we achieve, for any given speed $(a_n) \uparrow \infty$ such that $a_n = o(n)$, a large deviations principle (LDP) for the sequence $((S_{[nt]}/\sqrt{na_n})_{0\leq t\leq 1})_{n\in\mathbb{N}}$, in the space of càd-làg functions from $[0, 1]$ to $\mathbb{R}^q$. In Section **2.1** we give references of works in which such MDP are achieved for a large class of Markovian models, but always with the assumption that the function g is bounded.

- In the second part of this paper we provide a criteria of obtaining a MDP for the error of the least squares estimator (LSE) in multidimensional linear regression models (Theorem 3). In particular we deal with stable linear autoregressive models (Theorem 4), with noises which are not necessarily white but are assumed to be "generalized Gaussian" (white Gaussian distributed noises and bounded noises satisfy this condition). Moreover (for the autoregressive case), an exponential rate of convergence in probability of the empirical covariance is obtained.

  For such prediction errors, the only LDP the author is aware of concerns stable, unidimensional, linear, and first order autoregressive models with Gaussian white noise: it is obtained in [BerGamRou96], which uses results of [BrySmo93] and [BryDem95]. The proofs in [BerGamRou96] rely strongly on spectral theoretic arguments, and problems had to be solved concerning the steep property of the moment generating function of interest. It does not look very clear whether or not such a LDP can be extended to vector valued models of order greater than 1, and therefore the moderate deviations results obtained in Section **3** can be viewed as a reasonable compromise in this framework.

Notations:

* In this paper, the letter $\mathcal{C}$ denotes a generic positive constant whose exact value is unimportant. A vector $x \in \mathbb{R}^d$ will often be assimilated with the column matrix $(x_1, \ldots, x_d)^t$, and a vector $u \in \mathbb{R}^{dp}$ to $(u_1^t, \ldots, u_p^t)^t$ (where each u_i is in $\mathbb{R}^d$). If (x_n) is a sequence in $\mathbb{R}^d$, and $p \in \mathbb{N}^*$, then $x_n^{(p)}$ denotes the element $(x_n^t, x_{n-1}^t, \ldots, x_{n-p+1}^t)^t$ of $\mathbb{R}^{dp}$.

* On a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_n)$, a d-dimensionnal *noise* is an adapted sequence (ϵ_n) of $\mathbb{R}^d$-valued random vectors such that $\mathbb{E}[\epsilon_n | \mathcal{F}_{n-1}] = 0$. The

noise is said to be *white* when (ϵ_n) is an independent identically distributed (i.i.d.) sequence such that ϵ_n is independent of $\mathcal{F}_{n-1}$. Unlike those in Section 3, noises in Section 2 are always assumed to be white. Generalized Gaussian noises will be considered and defined in Section 3.

∗ In the sequel, a *speed* is a sequence $v = (v_n) \uparrow \infty$ of positive numbers. If E is a Polish space provided with its Borel sigma field $\mathcal{E}$, a *rate function* is a lower semi-continuous function $I : E \mapsto [0, \infty]$ with compact level sets $\{x;\ I(x) \leq \alpha\}$ (we note $I(\Gamma) = \inf_{x\in\Gamma} I(x),\ \forall\Gamma \subset E$). A sequence (Z_n) of E-valued random variables is said to satisfy a *large deviation principle* (shortened as LDP) with speed (v_n) and rate I if

$$\liminf_{n\to\infty} \frac{1}{v_n} \log \mathbb{P}[\, Z_n \in U \,] \geq -I(U) \ \text{ for every open subset } U \subset \mathcal{E}, \tag{1}$$

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mathbb{P}[\, Z_n \in F \,] \leq -I(F) \ \text{ for every closed subset } F \subset \mathcal{E}. \tag{2}$$

We say that (Z_n) satisfies an upper-LDP if only (2) is satisfied. As we will often deal with only one speed, n, we set the following notation, which is used in both Sections 2 and 3: for any sequence $(A_n) \subset \mathcal{E}$,

$$\mathbb{P}[A_n] \ \overset{\text{ld}}{<} \ 0 \qquad \text{means} \qquad \limsup \tfrac{1}{n} \log \mathbb{P}[A_n] \ < \ 0.$$

<u>Remark 1:</u> Frequently, assumptions in this paper will be of this form:

$$\lim \frac{1}{a_n} \log \mathbb{P}\,[\, A_n \,] = -\infty,$$

where (v_n) and (a_n) are speeds such that $a_n = o(v_n)$, and (A_n) is a sequence of events. It may be worth noting beforehand that such a relation is obviously implied by the following stronger one

$$\limsup \frac{1}{v_n} \log \mathbb{P}\,[\, A_n \,] < 0.$$

2 Statement of the results for stable Markov chains

2.1 Introduction

a) Empirical measures of Markov chains

Before stating our results, let us have a look at previous related works. Let E be some Polish state space, provided with its Borel sigma field $\mathcal{E}$, and $\mathcal{P}(E)$ denote the set of probability measures on $(E, \mathcal{E})$. Let (X_n) be a time-homogeneous discrete-time Markov chain with state space E and Fellerian transition kernel π. We say that the chain (X_n) is **stable** if there exists an unique π-invariant probability measure μ such that, for any initial distribution and almost every path, the sequence of the occupation measures

$$\Lambda_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{X_i}$$

converges weakly to μ. When π is strongly Fellerian, this property implies recurrence, but in the general case it does not imply irreducibility (see [MeyTwe], [Duf], or [Bor91] for details on stability).

Under various assumptions which almost always include the irreducibility of the chain, LDP have been proved for the sequence (Λ_n) (see, among other references, [DonVar75a], [DonVar75b], [DonVar76], [Aco90], [DinNey95], [Lip96], [Jai90], [BryDem96], [Ell88]), but stability and the Feller property are sufficient conditions for the sole upper LDP. In this section we are concerned with large and moderate deviations results for empirical means of the type

$$\Lambda_n(g) = \frac{1}{n}\sum_{i=0}^{n-1} g(X_i)$$

where g is an $\mathbb{R}^q$-valued function on E. Using the results cited above, one may achieve the LDP for $(\Lambda_n(g))$ if g is a bounded but not necessarily continuous function; a reference for the case of continuous and unbounded g is [GulLipLot94]. On the other hand, under various weaker conditions, CLT of the following form are proved

$$\sqrt{n}(\Lambda_n(g) - \mu(g)) \longrightarrow \mathcal{N}\left(0, S^2(g)\right) \quad \text{in distribution,}$$

(where $S^2(g)$ is some $q \times q$ covariance matrix related to g). Given a speed (a_n) such that $a_n = o(n)$, it is natural to associate to this CLT a LDP of speed (a_n) and rate $I(x) = \frac{1}{2}x^t(S^2(g))^{-1}x$ for the sequence $(\sqrt{n/a_n}(\Lambda_n(g) - \mu(g)))$.

In this section we achieve this type of MDP for a class of unbounded continuous functions g and a particular type of Markov chain, and assumptions of stability of the chain. During the recent years, several works have been devoted to the study of moderate deviations principles related to Markovian models, the main references being [Wu95], [Gao96], [Aco97], and [AcoChe98]: a large class of models is covered by these works, but (with the exception of [Wu95]) the boundedness of the function g is always assumed. See [Aco97] for a discussion on the various conditions proposed in these references.

b) Functional autoregressive models and Sanov theorem

In Section **2.2**, where our results are stated, we study stable functional autoregressive models of order $p \geq 1$, but in this introductory paragraph, we consider the model of order 1 defined by

$$X_{n+1} = f(X_n) + \sigma(X_n)\epsilon_{n+1}, \tag{3}$$

with $f : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to]0, +\infty[$ and the noise (ϵ_n) satisfying the following assumptions

$$\textbf{(H)} \begin{cases} * & f \text{ is continuous,} \\ * & \text{for some } \rho > 0 \text{ and every } x, \text{ we have } 1/\rho < \sigma(x) < \rho, \\ * & (\epsilon_n) \text{ is i.i.d. and satisfies for some } \tau > 0,\ \mathbb{E}\left[\,\exp \tau\|\epsilon\|\,\right] < \infty, \\ * & \text{the noise is independent of the initial state } X_0; \end{cases}$$

as well as the property, for some norm $|.|$ on $\mathbb{R}^d$,

$$\limsup_{|x|\to\infty} \frac{|f(x)|}{|x|} < \alpha < 1. \tag{4}$$

The transition π is such that $\pi(x, \cdot)$ is the distribution of $f(x) + \sigma(x)\epsilon$, and, if ν is the initial distribution and $g : \mathbb{R}^d \to \mathbb{R}$ is measurable, we will denote by $\Lambda_{n,\nu}$ the occupation measure defined in the previous paragraph, and set $\Lambda_{n,\nu}(g) = \int g \, d\Lambda_{n,\nu}$.

It is known that the stability of this model is achieved under one of the following two commonly encountered assumptions (*see* [Duf]):

- *case (i)* irreducibility of the transition kernel ;
- *case (ii)* a Lipschitz condition (which implies (4)):

  $\exists\, 0 < \alpha < 1$ such that $|f(x) - f(y)| + \mathbb{E}(|\epsilon|)|\sigma(x) - \sigma(y)| \leq \alpha|x - y| \quad (\forall x, y \in \mathbb{R}^d)$.

With property (4), the results from [DupEll], chapter 8, p.299, which are very close to those of Donsker and Varadhan, apply (indeed, the "exponential tightness" Condition 8.2.2 p.280 is satisfied; *cf* lemma 1 part 1 in paragraph **4.2.1**). The following upper LDP for $(\Lambda_{n,\nu})$ is thus valid.

Theorem 1 (Donsker-Varadhan, Dupuis-Ellis)
In one of the two cases stated above, and with assumptions (H) and (4), the Markov chain defined by (3) is stable. If μ denotes its unique invariant probability measure, and if we set, for any $\xi \in \mathcal{P}(\mathbb{R}^d)$,

$$I(\xi) = \inf\left\{ \int R(\ q(x,\cdot) \mid \pi(x,\cdot)\ )\ d\xi(x)\ ;\ q \textit{ transition probability kernel s.t. } \xi q = \xi \right\}$$

where $R(.|.)$ denotes the usual relative entropy between two probability measures, then

1. *I is a good rate function and $I(\xi) = 0 \Leftrightarrow \xi = \mu$.*

2. *For any family Φ of initial distributions such that $\sup_{\nu\in\Phi} \int e^{|x|} d\nu(x) < \infty$, the sequence $(\Lambda_{n,\nu})$ satisfies an upper-LDP of speed (n) and rate I, uniformly for $\nu \in \Phi$.*

   Consequently, for any bounded continuous $g : \mathbb{R}^d \to \mathbb{R}^q$, we have

$$\limsup \frac{1}{n} \log \sup_{\nu\in\Phi} \mathbb{P}\ [\ \|\Lambda_{n,\nu}(g) - \mu(g)\| > r\ ] < 0. \tag{5}$$

2.2 Statement of the results

2.2.1 Convergence with exponential speed

Relation (5) is this "convergence in probability with exponential speed" of $\Lambda_{n,\nu}(g)$ to $\mu(g)$ we generalize for some unbounded g and the following functional autoregressive model of order p

$$X_{n+1} = f(X_n^{(p)}) + \sigma(X_n^{(p)})\epsilon_{n+1} \tag{6}$$

in the following framework: functions $f : \mathbb{R}^{dp} \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^{dp} \to]0, +\infty[$ satisfy assumption (H) above, as well as one of the following two conditions, with $\alpha_1, \ldots, \alpha_p$ being ≥ 0 and such that $0 < \alpha_1 + \ldots + \alpha_p < 1$:

Case (i) the noise's distribution has a strictly positive density w.r.t. the Lebesgue measure (irreducibility), and f satisfies for some norm $|.|$ on $\mathbb{R}^d$, $c \geq 0$, and any $x \in \mathbb{R}^{dp}$:

$$|f(x)| \leq \alpha_1|x_1|+\ldots+\alpha_p|x_p|+c. \qquad \textbf{(R)}$$

Case (ii) (*Lipschitz model*) for some norm $|.|$ on $\mathbb{R}^d$, f and σ satisfy the following relation for $\beta \geq 1$ and any $x, y \in \mathbb{R}^{dp}$

$$|f(x)-f(y)|+\mathbb{E}(|\epsilon|^\beta)^{1/\beta}|\sigma(x)-\sigma(y)| \leq \alpha_1|x_1-y_1|+\ldots+\alpha_p|x_p-y_p|. \qquad \textbf{(L)}$$

This assumption provides the existence of a norm $|||.|||$ on $\mathbb{R}^{dp}$ such that, if $X_n^{(p)x}$ denotes the vector $X_n^{(p)}$ for the initial state x, we have for any $x, y \in \mathbb{R}^{dp}$

$$\mathbb{E}\,[\;|||X_n^{(p)x} - X_n^{(p)y}|||^\beta\;]^{1/\beta} \leq c_p\,\alpha^n|||x-y||| \tag{7}$$

where $0 < \alpha < 1$ and $c_p > 0$ is a constant. The proof that (L) implies (7) is given in Section **4.1**.

If ν denotes the distribution of $X_0^{(p)}$, we will denote by $(X_{n,\nu}^{(p)})$ the corresponding Markov chain. Under one of these two sets of conditions, this Markov chain is stable because (R) ensures the a.s. tightness of (Λ_n), whereas the uniqueness of limit points is given by the irreductibility of (i) or by (L) (*see* e.g. [BenMetPri], [Duf], and [Bor91] for more details): hence Theorem 1 applies to this model.

With $\|.\|$ denoting the Euclidean norm on $\mathbb{R}^{dp}$ and $\mathbb{R}^q$, $|.|$ the norm on $\mathbb{R}^d$ for which f satisfies (R) or (L), and β in (8) matching with β in (L), we achieve the following result (proved in Sections **4.1** and **4.2**).

Proposition 1 *("exponential convergence in probability")*
In one of these two frameworks, assume that $\tau > 0$, $\beta > 1$, and a family $\Phi \subset \mathcal{P}(\mathbb{R}^{dp})$ are such that

$$\mathbb{E}\left[\exp(\tau|\epsilon|^\beta)\right] < \infty \quad and \quad E(\Phi) := \sup_{\nu\in\Phi}\int \exp(\tau\|x\|^\beta)\,d\nu(x) < \infty. \tag{8}$$

Let $\beta' \in [1, \beta[$. The following properties hold.

1)

$$\limsup_{R\to+\infty}\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\Phi}\mathbb{P}\left[\frac{1}{n}\sum_{j=0}^{n-1}(1+\|X_{j,\nu}^{(p)}\|^{\beta'})\,\mathbb{I}_{\{\|X_{j,\nu}^{(p)}\|^{\beta'}\geq R\}} > r\right] = -\infty \quad (\forall r > 0). \tag{9}$$

2) If $g : \mathbb{R}^{dp} \to \mathbb{R}^q$ is a continuous function satisfying

$$\limsup_{\|x\|\to\infty}\frac{\|g(x)\|}{\|x\|^{\beta'}} < \infty,$$

then (5) holds, i.e. $\Lambda_{n,\nu}(g) = \frac{1}{n}\sum_{j=0}^{n-1} g(X_{j,\nu}^{(p)})$ converges to $\mu(g)$ with exponential speed, uniformly for $\nu \in \Phi$.

3) For any speed (a_n) such that $a_n = o(n)$,

$$\limsup_{R\to+\infty} \limsup_{n\to\infty} \frac{1}{a_n} \log \sup_{\nu\in\Phi} \mathbb{P}\left[\frac{1}{\sqrt{na_n}} \sup_{k\leq n} \|X_{k,\nu}^{(p)}\|^\beta \geq R\right] = -\infty. \tag{10}$$

For instance Φ can be taken as a set of deterministic initial states located in a compact subset of $\mathbb{R}^{dp}$.

2.2.2 Moderate deviation principle

If (X_n) is a Markov chain with transition π and invariant distribution μ, and g is a μ-integrable function such that the **Poisson equation**

$$g - \mu(g) = G - \pi(G)$$

admits a solution G, then the sequence

$$M_n = \sum_{j=1}^{n} (G(X_j) - \pi G(X_{j-1})) = \Lambda_n(g) - \mu(g) + G(X_0) - G(X_n)$$

defines a martingale such that, when G is μ-square-integrable,

$$\frac{1}{n} < M >_n \stackrel{n\to\infty}{\longrightarrow} S^2(g) = \int (GG^t - (\pi G)(\pi G)^t)\, d\mu \quad \text{a.s.}$$

Hence a CLT is at hand, as well as a MDP for $(\sqrt{n}(\Lambda_n(g) - \mu(g)))$ as soon as G is bounded (as in [Gao96]) and applying Dembo's result in [Dem96].

Let (X_n) denote the autoregressive model defined in **2.2.1** in the Lipschitz model framework (ii). We adapt the method outlined above to achieve this MDP, for a function $g : \mathbb{R}^{dp} \to \mathbb{R}^q$ in the set $Li(r,s)$ of *Lipschitz of order* (r,s) functions (with $s > 0$ and $r \geq 0$), *i.e.* that satisfies (where $\|.\|$ denotes at the same time the Euclidean norm on $\mathbb{R}^{dp}$ and on $\mathbb{R}^q$)

$$\|g(x) - g(y)\| \leq \mathcal{C}\|x-y\|^s(\|x\|^r + \|y\|^r + 1).$$

Indeed, under the "Lipschitz mixing property" (7) and some assumptions (which are largely satisfied in our framework) on r, s, and the moments of the noise, it is proved in [BenMetPri] (part II, chapter 2) or in [Duf] (chapter 6) that the Poisson equation admits a solution G which is also in $Li(r,s)$.

Let $\mathcal{D}_1(\mathbb{R}^q)$ denote the set of càd-làg functions from $[0,1]$ to $\mathbb{R}^q$, provided with the usual Skorohod topology. We achieve the following MDP, which proof relies mainly on Puhalskii's criterion developped in [Puh94] and is achieved in Section **4.3** (and **4.1**).

Theorem 2 *Let (X_n) denote the functional autoregressive model (6) satisfying the Lipschitz condition (L) above and the assumptions of Proposition 1, for some $\beta > 2$.*
Let $g \in Li(r,s)$. If $1 \leq r+s < \beta/2$, then, for any given speed (a_n) such that $a_n = o(n)$, the sequence $(\sqrt{\frac{n}{a_n}}(\Lambda_n(g) - \mu(g)))_{n\in\mathbb{N}}$ satisfies in $\mathbb{R}^q$ a LDP with speed (a_n) and rate

$$I(u) = \sup_{\theta\in\mathbb{R}^q} \left\{ < \theta, u > -\frac{1}{2}\theta^t S^2(g)\theta \right\} \qquad \textit{where} \qquad S^2(g) = \int (GG^t - (\pi G)(\pi G)^t)\, d\mu.$$

Moreover the sequence of processes $((\tilde{X}_t^{(n)})_{0\leq t\leq 1})_n$ *defined by*

$$\tilde{X}_t^{(n)} := \frac{1}{\sqrt{na_n}} \sum_{j=1}^{[nt]} (g(X_{j-1}^{(p)}) - \mu(g)),$$

satisfies in $\mathcal{D}_1(\mathbb{R}^q)$ *a LDP with speed* (a_n) *and the associated functional rate defined by*

$$J(\psi) = \begin{cases} \int_0^1 I(\dot{\psi}(t))dt & \textit{if } \psi(0) = 0 \textit{ and } \psi \textit{ is absolutely continous,} \\ \infty & \textit{otherwise.} \end{cases} \tag{11}$$

<u>Remark 2:</u> If $S^2(g)$ is invertible, then $I(u) = \frac{1}{2}u^t(S^2(g))^{-1}u$; if not, we have

$$I(u) = \begin{cases} \frac{1}{2}u^t S^2(g)^- u & \text{if } u \perp Ker(S^2(g)), \\ \infty & \text{otherwise,} \end{cases}$$

where $S^2(g)^-$ is the "generalized inverse" of $S^2(g)$.

3 Statement of the results on least squares estimation

3.1 Preliminary definitions

- **Generalized Gaussian noise:** in this section, MDP are achieved under the assumption that the noises (ϵ_n) handled are *generalized Gaussian, dominated by a centered Gaussian distribution of covariance* L, *i.e* such that there exists a $d \times d$ covariance matrix L satisfying, $\forall \theta \in \mathbb{R}^d, n \in \mathbb{N}$,

$$\mathbb{E}\left[\, \exp < \theta, \epsilon_n > \mid \mathcal{F}_{n-1} \,\right] \leq \exp\left(\tfrac{1}{2}\theta^t L \theta\right).$$

Examples of generalized Gaussian noises are centered bounded noises, and white centered noises with a Laplace transform $\phi(\theta)$ finite for every $\theta \in \mathbb{R}^d$ and such that

$$\limsup_{\|\theta\| \to \infty} \log \phi(\theta) / \|\theta\|^2 < \infty.$$

See [Pet] lemma 2.5. Moreover, to achieve not only the upper MDP but the full MDP, we will be led to assume that the generalized Gaussian noises satisfy the following property:

Property ($\mathbf{P}_\Gamma$): for some convex functions ψ_+ and ψ_-, of class C^3 in a neighbourhood of $0 \in \mathbb{R}^d$, we have

$$\log \mathbb{E}\left[\, \exp < u, \epsilon_n > \mid \mathcal{F}_{n-1} \,\right] \leq \psi_+(u) \tag{12}$$

$$\log \mathbb{E}\left[\, \exp < u, \epsilon_n > \mid \mathcal{F}_{n-1} \,\right] \geq \psi_-(u) \tag{13}$$

with $\psi_\pm(0) = 0$, $\nabla\psi_\pm(0) = 0$, and $D^2\psi_\pm(0) = \Gamma$, and Γ being some symmetric positive definite invertible $d \times d$ matrix.

Property (P_Γ) implies $\mathbb{E}\left[\epsilon_n \epsilon_n^t \middle| \mathcal{F}_{n-1}\right] = \Gamma$ for every $n \geq 1$. Any white noise with a Laplace transform $\phi(\theta) = \mathbb{E}\left[\exp < \theta, \epsilon_n >\right]$ defined in a neighbourhood of 0 satisfies (P_Γ) with $\psi_+ = \psi_- = \log \phi$. Any bounded noise such that $\mathbb{E}\left[\, \epsilon_n \epsilon_n^t \middle| \mathcal{F}_{n-1} \,\right] = \Gamma$ also satisfies (P_Γ).

- **Algebraic notations** (borrowed from [HorJoh])

  Let $\mathcal{M}_{\delta,d}$ design the space of $\delta \times d$ matrices. Any $\theta \in \mathcal{M}_{\delta,d}$ can be associated to the following element of $\mathbb{R}^{\delta d}$
  $$vec\,\theta = (\theta_{11}, \dots, \theta_{\delta 1}, \theta_{12}, \dots, \theta_{\delta d})^t.$$
  The map vec is one to one from $\mathcal{M}_{\delta,d}$ to $\mathbb{R}^{\delta d}$, and we denote by $mat : \mathbb{R}^{\delta d} \to \mathcal{M}_{\delta,d}$ its inverse. If $C \in \mathcal{M}_{\delta,\delta}$ and $L \in \mathcal{M}_{d,d}$, we define the Kronecker product of C and L as the $\delta d \times \delta d$ matrix $C \otimes L$ such that
  $$(C \otimes L)_{d(i-1)+j, d(I-1)+J} = C_{iI} L_{jJ} \qquad (\forall 1 \leq i, I \leq \delta,\ \forall 1 \leq j, J \leq d)$$
  i.e. $\xi^t.C \otimes L.\xi = \sum_{i,I=1}^{\delta} \sum_{j,J=1}^{d} (mat\,\xi)_{ij} (mat\,\xi)_{IJ} C_{iI} L_{jJ}$ for every $\xi \in \mathbb{R}^{\delta d}$. If $\delta = 1$ then C is a real number and $C \otimes L = CL$, and if C and L are invertible then $(C \otimes L)^{-1} = C^{-1} \otimes L^{-1}$.

3.2 MDP for least-squares estimators

3.2.1 General linear regression models

We first deal with the d-dimensional regression model (X_n) defined on the space $(\Omega, \mathcal{A}, \mathbb{P})$ provided with a filtration $(\mathcal{F}_n)$, by
$$X_n = \theta^t \phi_{n-1} + \epsilon_n \tag{14}$$
where θ is the $\delta \times d$ matrix to be estimated, (ϕ_n) is the $\mathbb{R}^{\delta}$-valued sequence of *explicative variables*, and (ϵ_n) is an $\mathbb{R}^d$-valued noise. We use the natural filtration $(\mathcal{F}_n) = (\sigma\{\phi_k, \epsilon_k\}_{k \leq n})$. The error $\tilde{\theta}_n$ of the least squares estimator $\hat{\theta}_n$ of θ is
$$\tilde{\theta}_n \;=\; \hat{\theta}_n - \theta \;=\; Q_n^{-1} \left(\sum_{k=1}^{n} \phi_{k-1} X_k^t \right) - \theta \;=\; -Q_n^{-1}\theta + Q_n^{-1} M_n,$$
where $C_n = \sum_{k=1}^n \phi_{k-1}\phi_{k-1}^t$, $Q_n = C_n + I_d$, and $M_n = \sum_{k=1}^n \phi_{k-1}\epsilon_k^t$. The sequences (C_n) and (Q_n) are both previsible ones. In Theorem 3 below, the results are stated for the vector sequence $(vec\,\tilde{\theta}_n)$ in order to avoid any confusion with handling MDPs in spaces such as $\mathcal{M}_{\delta,d}$.

Let (v_n) and (a_n) be speed sequences such that $a_n = o(v_n)$. The assumptions on the noise and the explicative variables are the following.

(N1) the noise (ϵ_n) is generalized Gaussian, dominated by a Gaussian distribution of invertible covariance L.

(N2) the noise (ϵ_n) satisfies property (P_Γ) (defined in Section **3.1**).

(C1) there exists some invertible $C \in \mathcal{M}_{\delta,\delta}$ such that, for every $r > 0$,
$$\lim \frac{1}{a_n} \log \mathbb{P}\left[\left\| \frac{C_n}{v_n} - C \right\| > r \right] = -\infty.$$

(C2) (ϕ_n) satisfies the exponential Lindeberg condition: $\forall r > 0$, $\forall \rho > 0$,
$$\lim \frac{1}{a_n} \log \mathbb{P}\left[\frac{1}{v_n} \sum_{j=1}^{n} \|\phi_{j-1}\|^2 \, \mathbb{I}_{\{\|\phi_{j-1}\| > r\sqrt{v_n/a_n}\}} > \rho \right] = -\infty.$$

The following theorem relies on a moderate deviations result for regressive sequences such as (M_n)

above, which proof may be found in [Wor98a] (see also [Wor98b]) and is based on the method of cumulants developped in [Puh94] by Puhalskii. The proof is performed in Section **5**.

Theorem 3 *Let $\tilde{\theta}_n = \hat{\theta}_n - \theta$ denote the error of the least squares estimator of θ in the regression model defined by (14), and (a_n) be a speed satisfying $a_n = o(v_n)$.*

a) Under **(N1)** *and* **(C1)**, $(\sqrt{\frac{v_n}{a_n}} vec\, \tilde{\theta}_n)_{n\in\mathbb{N}}$ *satisfies in $\mathbb{R}^{\delta d}$ an upper-LDP of speed (a_n) and rate*

$$J_{C,L}(\xi) = \frac{1}{2}\xi^t C \otimes L^{-1} \xi \quad (\forall \xi \in \mathbb{R}^{\delta d}). \tag{15}$$

b) In addition, if the noise is white with distribution $\mathcal{N}(0,\Gamma)$, or if **(N2)** *and* **(C2)** *are also valid, then $(\sqrt{\frac{v_n}{a_n}} vec\, \tilde{\theta}_n)_{n\in\mathbb{N}}$ satisfies a LDP of speed (a_n) and rate*

$$J_{C,\Gamma}(\xi) = \frac{1}{2}\xi^t C \otimes \Gamma^{-1} \xi \quad (\forall \xi \in \mathbb{R}^{\delta d}). \tag{16}$$

Corollary 1 *The latter theorem applies with $v_n = n$ when (ϕ_n) is an i.i.d. square-integrable sequence independent of the noise (ϵ_n), with previous assumptions on this noise.*

Remark 3: It is straightforward to check that the following "exponential Lyapunov condition" implies the "exponential Lindeberg condition" **(C2)**:

there exist a $\beta > 2$ and a $C > 0$ such that $\limsup \frac{1}{v_n} \log \mathbb{P}\left[\frac{1}{v_n}\sum_{j=1}^{n} \|\phi_{j-1}\|^\beta > C\right] < 0.$

Both conditions are naturally satisfied if (ϕ_j) is a bounded sequence.

3.2.2 Stable linear autoregressive models

Let (X_n) be the stable autoregressive model of order p (AR(p)) and dimension d defined by

$$X_n = A_1 X_{n-1} + A_2 X_{n-2} + \ldots + A_p X_{n-p} + \epsilon_n \tag{I}$$

with an initial state $X_0^{(p)} = (X_0, X_{-1}, \ldots, X_{-p+1})$ independent from the noise. By stable model we mean that all roots of the polynomial $z \mapsto \det(I - A_1 z - \cdots A_p z^p)$ have their modulus > 1. For the noise (ϵ_n), the assumptions may be of two types, with Γ below designing an invertible $d \times d$ covariance matrix:

Case 1: (ϵ_n) is a white generalized Gaussian noise with covariance Γ, and for some $\tau > 0$,

$$\mathbb{E}\left[\exp(\tau\|\epsilon\|^2)\right] < \infty \quad \text{and} \quad \mathbb{E}\left[\exp(\tau\|X_0^{(p)}\|^2)\right] < \infty.$$

Case 2: (ϵ_n) is a generalized Gaussian noise such that $\mathbb{E}\left[\epsilon_n\epsilon_n^t | \mathcal{F}_{n-1}\right] = \Gamma$ $(\forall n \geq 1)$

and for some $\beta > 2$, $\tau > 0$, and $E > 0$,

$$\sup_n \mathbb{E}\left[\exp(\tau\|\epsilon_n\|^\beta) \,|\, \mathcal{F}_{n-1}\right] \leq E < \infty \quad \text{and} \quad \mathbb{E}\left[\exp(\tau\|X_0^{(p)}\|^\beta)\right] < \infty.$$

Moreover, we assume that for every $r > 0$

$$\mathbb{P}\left[\left\|\frac{1}{n}\sum_{j=1}^{n}(\epsilon_j\epsilon_j^t - \Gamma)\right\| > r\right] \overset{\text{ld}}{<} 0.$$

The model $(X_n^{(p)})_{n\in\mathbb{N}}$ is itself a stable AR(1) model

$$X_n^{(p)} = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_d & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & I_d & 0 \end{bmatrix} X_{n-1}^{(p)} + \begin{bmatrix} \epsilon_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = B^t X_{n-1}^{(p)} + \eta_n, \tag{II}$$

and its "stationary covariance" is $C = \sum_{j=0}^{\infty} B^j \bar{\Gamma}(B^j)^t$, where $\bar{\Gamma} = \begin{bmatrix} \Gamma & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}$.

Theorem 4 *1. For every $r > 0$, we have*

$$\mathbb{P}\left[\left\|\frac{1}{n}\sum_{k=1}^{n} X_k^{(p)}(X_k^{(p)})^t - C\right\| \geq r\right] \overset{ld}{<} 0.$$

Moreover, there exists $R > 0$ sufficiently large such that

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^{n} \|X_k^{(p)}\|^{\beta} > R\right] \overset{ld}{<} 0,$$

and, in case 1, for any $r > 0$,

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^{n}\|X_k^{(p)}\|^2\,\mathbb{I}_{\|X_k^{(p)}\|>R} > r\right] < 0.$$

2. We assume that C is invertible (e.g. A_p is invertible). Let $\tilde{\theta}_n = \hat{\theta}_n - \theta$ be the error of the least squares estimator $\hat{\theta}_n$ of $\theta = [A_1\ A_2\ \cdots\ A_p]^t$ for the regression model (I)

$$X_n = [A_1\ A_2\ \cdots\ A_p]X_{n-1}^{(p)} + \epsilon_n. \tag{I}$$

For any given speed (a_n) such that $a_n = o(n)$, the sequence $(\sqrt{v_n/a_n}\,vec\,\tilde{\theta}_n)_{n\in\mathbb{N}}$ satisfies in $\mathbb{R}^{dp\times d}$ an upper LDP with speed (a_n) and rate (15) (with $\delta = dp$), and it obeys the full LDP with rate (16) if in addition (ϵ_n) satisfies property (P_Γ) (which is the case in case 1).

In order to establish Theorem 4, we shall prove the following result (which is a version of part 1 of Theorem 4 when $p = 1$).

Proposition 2 *Let A be a stable (i.e. of spectral radius < 1) $d\times d$ matrix, and (X_n) be the stable AR(1) model defined by*

$$X_n = AX_{n-1} + \epsilon_n.$$

For the noise (ϵ_n) and the initial state X_0, we take the assumptions of theorem 4 (case 1 or case 2). Then for every $r > 0$,

$$\mathbb{P}\left[\left\|\tfrac{1}{n}\sum_{k=1}^{n} X_k(X_k)^t - C\right\| \geq r\right] \overset{ld}{<} 0, \tag{17}$$

where C is the stationary covariance of the model, and there exists $R > 0$ sufficiently large such that

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^{n}\|X_k\|^{\beta} > R\right] \overset{ld}{<} 0, \tag{18}$$

and, in case 1, we have for every $r > 0$

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^{n}\|X_k\|^2\mathbb{I}_{\|X_k\|>R} > r\right] < 0. \tag{19}$$

If we apply this proposition to model (II), we achieve part 1 of Theorem 4, and then part 2 comes easily by application of theorem 3 to the regression model (X_n) with explicative variables sequence $(X_n^{(p)})$. We prove Proposition 2 (and therefore Theorem 4) in Section **6.**

The most simple framework in which Theorem 4 applies is when (ϵ_n) is a generalized Gaussian white noise with covariance Γ (*i.e.* case 1). Another simple application is when (ϵ_n) is a bounded martingale increment sequence with conditional covariance Γ.

4 Proof of proposition 1 and theorem 2

4.1 Restriction to the case $p = 1$

- We want to induce the general case $p \geq 1$ from the case $p = 1$. Let us consider the following associated autoregressive model $(X_n^{(p)})_{n\in\mathbb{N}}$ of order 1,

$$X_{n+1}^{(p)} = F(X_n^{(p)}) + \sigma(X_n^{(p)})\zeta_n = \begin{bmatrix} f(X_n^{(p)}) \\ X_n \\ \vdots \\ X_{n-p+2} \end{bmatrix} + \sigma(X_n^{(p)})\begin{bmatrix} \epsilon_{n+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

  and construct a norm on $\mathbb{R}^{dp}$ for which F would satisfy the assumptions (either (i) or (ii)). As in **2.2.1** (L) implies (R) and we only need relation (R) to prove Proposition 1, we will deal with (R) first. We will then turn to Theorem 2, which applies only if the model is "Lipschitz-mixing" (that is, in case (ii)), as it needs the existence of a solution (in $Li(r,s)$) of the Poisson equation associated to g, relying on relation (7).

- If $|.|$ denotes the norm on $\mathbb{R}^d$ for which either (R) or (L) is satisfied, we adopt the notations $|x| = (|x_1|,\ldots,|x_p|)^t \in \mathbb{R}^p$ for $x \in \mathbb{R}^{dp}$, and for v and w vectors of $\mathbb{R}^p$, we set $v \leq w \Leftrightarrow \forall i = 1,\ldots,p,\ (w-v)_i \geq 0$. We then have

$$|F(x)| \leq C.|x| + \eta \quad \text{where} \quad C = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad \text{and} \quad \eta = \begin{bmatrix} c \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}.$$

  As $\alpha_1 + \ldots \alpha_p < 1$, it is known that C is an irreducible positive matrix to which the Perron Froebenius theorem applies, *i.e.* the spectral radius $\rho(C)$ of C is an eigenvalue of C and

there exists an eigenvector $v > 0$ associated to $\lambda = \rho(C) < 1$. We can assume $v_1 = 1$, and consequently $v = (1, \lambda^{-1}, \dots, \lambda^{-p+1})$.
Hence, for $x \in \mathbb{R}^{dp}$, if we set $|||x||| = \sup_{i=1\dots p} \lambda^{i-1}|x_i|$, then $|x| \leq |||x|||v$ hence $C|x| \leq \lambda |||x|||v$ and we get

$$|||F(x)||| \;\leq\; \sup_{i=1\dots p} \frac{(C.|x|)_i}{v_i} + c \;\leq\; \lambda |||x||| + c.$$

As $0 < \lambda < 1$, property (R) is valid for F instead of f. Therefore, for Proposition 1, we can restrict the proof to the case $p = 1$.

- Now, we have to prove that (L) implies (7) (and consequently that Theorem 2 is proved for $p \geq 1$ if it is for $p = 1$). To the notations above, we add the following ones: if $z \in \mathbb{R}^p$ is ≥ 0 (*i.e.* $z_i \geq 0,\ \forall i$), we set $z^\beta = (z_1^\beta, \dots, z_p^\beta)^t$. Hence, as $(\sum_{i=1}^p \alpha_i z_i)^\beta \leq (\sum_{i=1}^p \alpha_1)^{\beta-1}(\sum_{i=1}^p \alpha_i z_i^\beta)$ and $\sum_{i=1}^p \alpha_i < 1$, we have $(Cz)^\beta \leq Cz^\beta$ for any $z \geq 0$ in $\mathbb{R}^p$. Thus, (L) implies
$$\mathbb{E}\,[\,|X_1^{(p)x} - X_1^{(p)y}|^\beta\,] \leq C|x-y|^\beta \qquad (\forall x, y\ \in \mathbb{R}^{dp}),$$
and consequently for every $n \geq 1$
$$\mathbb{E}\,[\,|X_n^{(p)x} - X_n^{(p)y}|^\beta\,] \leq C^n|x-y|^\beta \qquad (\forall x, y\ \in \mathbb{R}^{dp}).$$
For any $z \geq 0$ in $\mathbb{R}^p$, we have $z \leq v \sup_{i\leq p} \lambda^{i-1} z_i$, hence $C^n z \leq \lambda^n(\sup_{i\leq p} \lambda^{i-1} z_i)v$, and, with $z = |x-y|^\beta$ and $x, y\ \in \mathbb{R}^{dp}$,
$$\sup_{i\leq p} \lambda^{i-1}(C^n|x-y|^\beta)_i \;\leq\; \lambda^n \sup_{i\leq p} \lambda^{i-1}(|x-y|^\beta)_i \;\leq\; \lambda^n \lambda^{(p-1)(1-\beta)} |||x-y|||^\beta.$$
Moreover,
$$\mathbb{E}\,[\,|||X_n^{(p)x} - X_n^{(p)y}|||^\beta\,] \leq p\ \sup_{i\leq p} (\lambda^{i-1})^\beta \mathbb{E}\,[\,|X_n^{(p)x} - X_n^{(p)y}|_i^\beta\,] \leq p\ \sup_{i\leq p} \lambda^{i-1}(C^n|x-y|^\beta)_i,$$
hence (7) is proved with $c_p = p\lambda^{(p-1)(1-\beta)}$.

4.2 Proof of proposition 1 (with $p = 1$)

It is straightforward to see that the results of Proposition 1 extend to $q > 1$ if they are proved for $q = 1$. We will therefore assume that $q = 1$ in this proof. On the other hand, as we restricted the proof to the case $p = 1$, we will use the norm $|.|$ for which (R) or (L) holds as the norm on $\mathbb{R}^{dp} = \mathbb{R}^d$ (instead of the Euclidean norm appearing in each of the relations (8), (9), and (10)).

Finally, to lighten the notations, in this and later proofs, we will often omit the subscript ν in $X_{n,\nu}$.

4.2.1 Proof of 1)

- First note that (L) implies relation (R) with $c = |f(0)|$. Hence, in either case, if we set $\eta_k = c + |\epsilon_k|$, it comes
$$|X_n| \leq \alpha^n|X_0| + \sum_{k=1}^n \alpha^{n-k}\eta_k.$$

Applying the Hölder inequality to the measure on $\{0,\ldots,n\}$ giving the weight α^{n-k} to k, and to the function ψ defined by $\psi(0)=|X_0|$ and $\psi(k)=\eta_k$, we get (as $0<\alpha<1$)

$$\begin{aligned} |X_n| &\leq (\alpha^n|X_0|^\beta+\sum_{k=1}^n \alpha^{n-k}\eta_k^\beta)^{1/\beta}.\left(\frac{1-\alpha^{n+1}}{1-\alpha}\right)^{\frac{\beta-1}{\beta}} \\ (1-\alpha)^{\beta-1}|X_n|^\beta &\leq \alpha^n|X_0|^\beta+\sum_{k=1}^n \alpha^{n-k}\eta_k^\beta \leq |X_0|^\beta+\frac{1}{1-\alpha}\sup_{k\leq n}\eta_k^\beta. \end{aligned} \tag{20}$$

As for $x_j>0$, $\sum_{k=1}^n\sum_{j=1}^k \alpha^{k-j}x_j=\sum_{j=1}^n(\sum_{i=1}^{n-j}\alpha^i)x_j\leq\frac{1}{1-\alpha}\sum_{j=1}^n x_j$, (20) entails

$$(1-\alpha)^\beta\sum_{k=1}^n|X_k|^\beta\leq|X_0|^\beta+\sum_{k=1}^n\eta_k^\beta$$

hence, X_0 being independent from (η_n), we have for $\lambda=\tau(1-\alpha)^\beta$

$$\sup_{\nu\in\Phi}\mathbb{E}\,[\exp\lambda\sum_{k=1}^n|X_{k,\nu}|^\beta\,]\;\leq\;\sup_{\nu\in\Phi}\mathbb{E}\,[\exp\tau|X_{0,\nu}|^\beta].\mathbb{E}\,[\exp\tau\eta^\beta\,]^n\leq\mathcal{C}e^{nL}$$

where $L=\log\mathbb{E}\,[\exp\tau|\epsilon|^\beta\,]$. Hence if $R>0$

$$\sup_{\nu\in\Phi}\mathbb{P}\left[\tfrac{1}{n}\textstyle\sum_{k=1}^n|X_{k,\nu}|^\beta\geq R\right]=\sup_{\nu\in\Phi}\mathbb{P}\left[\exp\lambda\textstyle\sum_{k=1}^n|X_{k,\nu}|^\beta\geq\exp\lambda Rn\right]\leq\mathcal{C}e^{-\lambda Rn+nL}$$

thus

$$\limsup_{R\to+\infty}\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\Phi}\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^n|X_{k,\nu}|^\beta\geq R\right]\;\leq\;\lim_{R\to+\infty}-\lambda R+L=-\infty. \tag{21}$$

- Let $\tau'>0$. In [DonVar76], [GulLipLot94], and other papers related to Sanov theorem for Markov chains (*see* **2.1.1**), the following functions are introduced (with $\pi(x,dy)$ denoting the transition probability of the Markovian model (X_n), and $|.|$ the norm for which (R) holds)

$$u(x)=\tau'|x|^\beta,\qquad V(x)=u(x)-\log\int e^{u(y)}\,\pi(x,dy).$$

We state the following lemma, part 2 of which will also prove useful in Section **4.3.2**.

Lemma 1 *1. For some real positive constants V^* and τ'' (with $0<\tau''<\tau'$)*

$$V(x)\leq\tau'|x|^\beta\qquad\textit{and}\qquad\tau''|x|^\beta\leq V(x)+V^*\quad(\forall x\in\mathbb{R}^d).$$

2. We have $\sup_{\nu\in\Phi}\sup_n\mathbb{E}\,[\exp\sum_{i=0}^{n-1}V(X_{i,\nu})\,]\leq E(\Phi):=\sup_{\nu\in\Phi}\int e^{\tau|x|^\beta}\,d\nu(x)$, and, for any $\gamma\in[1,\beta[$, any speed (v_n), and every $R\geq1$, $r>0$,

$$\sup_{\nu\in\Phi}\mathbb{P}\left[\frac{1}{v_n}\sum_{i=0}^{n-1}(1+|X_{i,\nu}|^\gamma)\,\mathbb{I}_{\{|X_{i,\nu}|^\gamma>R\}}>r\right]\;\leq\;E(\Phi)\exp\left(-rv_n\frac{\tau''}{2}R^{(\beta-\gamma)/\gamma}+nV^*\right). \tag{22}$$

With $v_n=n$ and $\gamma=\beta'$, (22) implies (9): that ends the proof of part 1 of Proposition 1.

- *Proof of lemma 1: a) Proof of 1.* Jensen's inequality entails the first assertion. For the second one, as $\beta \geq 1$ and $(a+b)^\beta \leq \lambda^{1-\beta}a^\beta + (1-\lambda)^{1-\beta}b^\beta$ for any $\lambda \in]0,1[$ and $a, b \in \mathbb{R}^+$, it comes
$$V(x) = \tau'|x|^\beta - \log \mathbb{E}\,[\exp \tau'|f(x) + \sigma(x)\epsilon|^\beta\,] \geq \tau'|x|^\beta(1 - \lambda^{1-\beta}\alpha^\beta) - V^*$$
with $V^* = \mathcal{C} + \log \mathbb{E}\,[\exp \tau|\epsilon|^\beta\,] > 0$ and $\tau' = (1-\lambda)^{\beta-1}\rho^{-\beta}\tau$. Therefore, taking $\lambda = \left(\frac{\alpha+1}{2}\right)^{\beta/(\beta-1)}$ and $\tau'' = \tau'(1 - \lambda^{1-\beta}\alpha^\beta)$, the second inequality of part 1 comes.
b) Proof of 2. The first assertion is easy to check (by the definition of V), and, as $1 \leq \gamma < \beta$ and $R \geq 1$, we have for every $x \in \mathbb{R}^d$
$$(1 + |x|^\gamma)\mathbb{I}_{\{|x|^\gamma \geq R\}} \leq 2|x|^\beta \left(R^{1/\gamma}\right)^{\gamma-\beta} \leq \frac{2}{\tau''}(V(x) + V^*)R^{(\gamma-\beta)/\gamma}$$
hence for every $r > 0$
$$\mathbb{P}\left[\tfrac{1}{v_n}\textstyle\sum_{i=0}^{n-1}(1 + |X_{i,\nu}|^\gamma)\,\mathbb{I}_{\{|X_{i,\nu}|^\gamma > R\}} > r\right] \;\leq\; \mathbb{P}\left[\textstyle\sum_{i=0}^{n-1} V(X_{i,\nu}) > rv_n\frac{\tau''}{2}R^{(\beta-\gamma)/\gamma} - nV^*\right].$$

4.2.2 Proof of 2)

It comes easily from relation (9) and from Sanov-type results such as Theorem 1 (part 3) stated in the introduction. Indeed, by the assumption on g, there exists some constant $A > 0$ such that
$$|g(x)| \leq A(1 + |x|^{\beta'}) \quad (\forall x \in \mathbb{R}^d).$$
For every $R > 0$, we define the following real valued bounded continuous function $g^{(R)}$ on $\mathbb{R}^d$
$$g^{(R)}(x) = \begin{cases} g(x) & \text{if } |g(x)| \leq R, \\ R & \text{if } g(x) \geq R, \\ -R & \text{if } g(x) \leq -R. \end{cases}$$
We have $|(g - g^{(R)})(x)| \leq |g(x)|\,\mathbb{I}_{\{|g(x)| \geq R\}} \xrightarrow{R\to\infty} 0$ $(\forall x)$. Hence, for any given $r > 0$, if $R > 0$ is taken sufficiently large such that $|\mu(g - g^{(R)})| < \frac{r}{3}$, we have
$$\begin{aligned}\mathbb{P}\,[\,|\Lambda_{n,\nu}(g) - \mu(g)| \geq r\,] &\leq \mathbb{P}\,[\,|\Lambda_{n,\nu}(g - g^{(R)})| \geq \tfrac{r}{3}\,] + \mathbb{P}\,[\,|\Lambda_{n,\nu}(g^{(R)}) - \mu(g^{(R)})| \geq \tfrac{r}{3}\,] \\ &\leq \mathbb{P}\left[\tfrac{1}{n}\textstyle\sum_{i=0}^{n-1}(1 + |X_{i,\nu}|^{\beta'})\mathbb{I}_{\{|X_{i,\nu}|^{\beta'} > \frac{R}{A} - 1\}} > \tfrac{r}{3A}\right] + \mathbb{P}\,[\,|\Lambda_{n,\nu}(g^{(R)}) - \mu(g^{(R)})| \geq \tfrac{r}{3}\,]\end{aligned}$$
and thus, by (9) and by (5) applied to $g^{(R)}$, proof of part 2 is completed.

Remark 4: On the basis of the upper LDP stated in Theorem 1 and of relation (9), we could easily follow the approach of [GulLipLot94] and establish an upper LDP for the sequence $(\Lambda_n(g))$, using Lemma 2.1.4 of [DeuStr].

4.2.3 Proof of 3)

Let (a_n) be such that $a_n = o(n)$. By relation (20) and easy calculations on the initial state, we see that (10) will result from the following statement, which we prove in this paragraph: if $R > 0$,
$$\limsup \frac{1}{a_n} \log \mathbb{P}\left[\frac{1}{\sqrt{na_n}} \sup_{k \leq n} |\eta_k| > R\right] \leq -\frac{R^2}{8c^2}, \tag{23}$$
where c^2 denotes the variance of the white noise (η_k).

Let $Y_t^{(n)} = \sum_{j=1}^{[nt]} \eta_j$. It is known that the sequence $\left(\frac{1}{\sqrt{na_n}}(Y_t^{(n)} - nt\mathbb{E}(\eta))\right)$ satisfies in $\mathcal{D}_1(\mathbb{R})$ the LDP of speed (a_n) and rate function

$$J(\phi) = \int_0^1 (\overset{\bullet}{\phi}(t))^2/2c^2\, dt \quad \text{if } \phi \in \mathcal{D}_1(\mathbb{R}) \text{ is absolutely continuous and } \phi(0)=0,$$

and $J(\phi) = \infty$ otherwise. Moreover, it is shown in [DesPic79] that

$$J\left(\left\{\, \phi \in \mathcal{D}_1(\mathbb{R});\ \phi(0) = 0 \text{ and } \sup_{0\le t\le 1}\phi(t) > R \,\right\}\right) = \inf_{0<t\le 1}\left\{ t\frac{(R/t)^2}{2c^2}\right\} = \frac{R^2}{2c^2}.$$

Consequently, as $\left(\frac{1}{\sqrt{na_n}}(nt\mathbb{E}(\eta) - Y_t^{(n)})\right)$ satisfies the same LDP, we obtain

$$\lim_{n\to\infty}\frac{1}{a_n}\log\mathbb{P}\left[\frac{1}{\sqrt{na_n}}\sup_{0\le t\le 1}|Y_t^{(n)} - nt\mathbb{E}(\eta)| > R\right] = -\frac{R^2}{2c^2}.$$

Hence, (23) comes as we have

$$\sup_{k\le n}|\eta_k| = \sup_{0\le t\le 1}|Y_t^{(n)} - Y_{t-}^{(n)}| \le 2\sup_{0\le t\le 1}|Y_t^{(n)} - nt\mathbb{E}(\eta)|.$$

4.3 Proof of theorem 2 (with $p = 1$)

4.3.1 Preliminary calculations

- Let $g \in Li(r,s)$ with r,s being such that $1 \le r+s < \frac{\beta}{2}$. Thanks to property (7), there exists a solution G to the Poisson equation associated to g, with $G \in Li(r,s)$ as well. Thus, with $\|.\|$ designing the Euclidean norm on $\mathbb{R}^q$, there exist some positive constants A and A' such that

  $$\|G(x)\| \le A|x|^{r+s} + A' \quad (\forall x \in \mathbb{R}^d).$$

  For every $R > 0$, we denote by G_R the following functional

  $$G_R(x) = G(x)\mathbb{I}_{\{\|G(x)\|\le R\}} \qquad \text{hence} \qquad (G - G_R)(x) = G(x)\mathbb{I}_{\{\|G(x)\|>R\}}.$$

- We now prove the following relation: there exists some constant $\mathcal{C} > 0$ such that

  $$\frac{1}{\mathcal{C}}\|\pi(G - G_R)(x)\| \;\le\; (1 + |x|^{r+s})\,\mathbb{I}_{\{|x|^{r+s}>R'\}} + \delta_R \tag{24}$$

  where δ_R and $R' = R'(R,G)$ are real numbers such that $\begin{cases} \delta_R \to 0 \\ R' \to +\infty \end{cases}$ when $R \to +\infty$.

  We set $\gamma = r+s$ and $S(R,x) := \{\, \|G(f(x) + \sigma(x)\epsilon)\| > R \,\}$ (for $R > 0$ and $x \in \mathbb{R}^d$). As $\gamma \ge 1$, $(a+b+c)^\gamma \le 3^{\gamma-1}(a^\gamma + b^\gamma + c^\gamma)$ $(\forall a,b,c \in \mathbb{R}_+)$; hence, for R such that $\frac{R-A'}{(3\alpha)^\gamma A} > |f(0)|^\gamma$, if we set

  $$R' = R'(R,\gamma,G) = \tfrac{R-A'}{A3^\gamma}(\max\{a,\rho\})^{-\gamma},$$

it comes $\mathbb{I}_{S(R,x)} \leq \mathbb{I}_{\{|x|^\gamma > R'\}} + \mathbb{I}_{\{|\epsilon|^\gamma > R'\}}$. Consequently,

$$\begin{aligned}\|\pi(G-G_R)(x)\| &= \mathbb{E}\left[\,\|G(f(x)+\sigma(x)\epsilon)\|\,\mathbb{I}_{S(R,x)}\right] \leq \mathbb{E}\left[\,(A|f(x)+\sigma(x)\epsilon|^\gamma + A')\mathbb{I}_{S(R,x)}\right]\\ &\leq \mathbb{E}\left[\,(A3^{\gamma-1}(\alpha^\gamma|x|^\gamma + |f(0)|^\gamma + \rho^\gamma|\epsilon|^\gamma) + A')\mathbb{I}_{S(R,x)}\right]\\ &\leq \mathcal{C}_1(|x|^\gamma+1)\,\mathbb{I}_{\{|x|^\gamma > R'\}} + \mathcal{C}_2(|x|^\gamma+1)\,\mathbb{P}[|\epsilon|^\gamma > R'] + \mathcal{C}_3\,\mathbb{E}\left[|\epsilon|^\gamma\mathbb{I}_{\{|\epsilon|^\gamma > R'\}}\right]\end{aligned}$$

where $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ are positive constants independent from R and x. As we have

$$(|x|^\gamma+1)\mathbb{P}\left[\,|\epsilon|^\gamma > R'\,\right] \leq (|x|^\gamma+1)\mathbb{I}_{\{|x|^\gamma > R'\}} + (R'+1)\mathbb{E}\left[\,\exp\tau|\epsilon|^\gamma\,\right]\exp(-\tau R'),$$

the desired result comes with $\mathcal{C} = \max\{\mathcal{C}_1+\mathcal{C}_2, \mathcal{C}_3\}$ and

$$\delta_R = (R'+1)\mathbb{E}\left[\,\exp\tau|\epsilon|^\gamma\,\right]\exp(-\tau c_1^{-\gamma}R') \;+\; \mathbb{E}\left[\,|\epsilon|^\gamma\mathbb{I}_{\{|\epsilon|^\gamma > R'\}}\,\right] \overset{R\to+\infty}{\longrightarrow} 0.$$

- Associated to (9), relation (24) implies the following property

$$\limsup_{R\to+\infty}\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\Phi}\mathbb{P}\left[\frac{1}{n}\sum_{j=1}^n\|\pi(G-G_R)(X_{j-1,\nu})| > r\right] = -\infty \quad (\forall r>0), \qquad (25)$$

and the same is obviously true replacing the functional $\pi(G-G_R)$ by $(G-G_R)$. Moreover, as $G \in Li(r,s)$, the functions

$$(G-G_R)(G-G_R)^t,\;\; \pi(G-G_R)(G-G_R)^t,\;\; \text{and}\;\; (\pi(G-G_R))(\pi(G-G_R))^t$$

are all Lipschitz of order $(r, 2s+r)$: as $2(r+s) < \beta$, it is therefore easy to check that we also achieved

$$\limsup_{R\to+\infty}\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\Phi}\mathbb{P}\left[\frac{1}{n}\sum_{j=1}^n\|\pi[(G-G_R)(G-G_R)^t](X_{j-1,\nu})\| > r\right] = -\infty \quad (\forall r>0), \qquad (26)$$

as well as the corresponding relations with $(G-G_R)(G-G_R)^t$ or $(\pi(G-G_R))(\pi(G-G_R))^t$ instead of $\pi(G-G_R)(G-G_R)^t$. All these properties will prove useful in the next paragraph, and we will refer to all of them simply by "relation (25)" or "lation (26)"ote that these relations remain true if we remove "$\limsup_{R\to+\infty}$" and replace R by $R(n)$, where $R(n)$ stands for any sequence such that $R(n) \uparrow \infty$.

4.3.2 Proof of the MDP

- We set $H = \pi(GG^t) - (\pi G)(\pi G)^t \in Li(r, 2s+r)$ and

$$M_n = \sum_{j=1}^n G(X_j) - \pi G(X_{j-1}), \qquad \text{hence} \qquad <M>_n = \sum_{j=1}^n H(X_{j-1}).$$

As $2(r+s) < \beta$, Proposition 1 part 2 applies to H instead of g: consequently, as $\mu(H) = S^2(g)$, we obtain

$$\mathbb{P}\left[\left\|\frac{<M>_n}{n} - S^2(g)\right\| > r\right] \overset{\text{ld}}{<} 0 \quad (\forall r>0),$$

and thanks to Lemma 4 from [Dem96], this implies

$$\mathbb{P}\left[\sup_{t\le 1}\left\|\frac{<M>_{[nt]}}{n} - tS^2(g)\right\| > r\right] \overset{\text{ld}}{<} 0 \quad (\forall r>0). \tag{27}$$

- Let $(a_n)\uparrow\infty$ be a speed such that $a_n = o(n)$. We introduce the sequences

$$N_k^{(n)} = \sum_{j=1}^{k} G_{R(n)}(X_j) - \pi G_{R(n)}(X_{j-1}) \qquad \text{et} \qquad P_k^{(n)} = M_k - N_k^{(n)},$$

where $R(n) = \left(\frac{n}{a_n}\right)^{\delta}$ and $\delta \in]0,\frac{1}{2}[$ to be specified later on.

In order to achieve the required LDP for the sequence $(\tilde{X}_\bullet^{(n)})$ introduced in the statement of the theorem, the first step is to prove the so-called *C-exponential contiguity* (see [LipPuh] for details on this subject, or [Puh97]) in $\mathcal{D}_1(\mathbb{R}^q)$ between the sequences $(\tilde{X}_\bullet^{(n)})$ and $(M_{[n\bullet]}/\sqrt{na_n})$, and between the latter and $(N_{[n\bullet]}^{(n)}/\sqrt{na_n})$.

- For any $r>0$, we have

$$\begin{aligned}
&\mathbb{P}\left[\sup_{t\le 1}|\tfrac{1}{\sqrt{na_n}}(\textstyle\sum_{j=1}^{[nt]}(g(X_{j-1})-\mu(g)) - M_{[nt]})\| > r\right]\\
=\;& \mathbb{P}\left[\sup_{t\le 1}\|G(X_0)-G(X_{[nt]})\| > r\sqrt{na_n}\right]\\
\le\;& \mathbb{P}\left[|X_0|^{r+s} > \tfrac{r}{2A}\sqrt{na_n} - \tfrac{A'}{A}\right] + \mathbb{P}\left[\sup_{k\le n}|X_k|^{r+s} > \tfrac{r}{2A}\sqrt{na_n} - \tfrac{A'}{A}\right]\\
\le\;& E(\Phi)\exp(-\tau(\tfrac{r}{2A}\sqrt{na_n} - \tfrac{A'}{A})) + \mathbb{P}\left[\sup_{k\le n}|X_k|^{r+s} > \tfrac{r}{2A}\sqrt{na_n} - \tfrac{A'}{A}\right].
\end{aligned}$$

But $r+s<\beta$, hence (10) implies that, for every $r>0$,

$$\limsup_{n\to\infty}\tfrac{1}{a_n}\log\mathbb{P}\left[\tfrac{1}{\sqrt{na_n}}\sup_{k\le n}|\eta_k|^{r+s} > r\right] = -\infty,$$

and consequently, as $\sqrt{n/a_n}\overset{n\to\infty}{\longrightarrow}+\infty$, the first C-exponential contiguity is achieved:

$$\limsup_{n\to\infty}\frac{1}{a_n}\log\mathbb{P}\left[\sup_{t\le 1}\left\|\tfrac{1}{\sqrt{na_n}}\left(\textstyle\sum_{j=1}^{[nt]}(g(X_{j-1})-\mu(g)) - M_{[nt]}\right)\right\| > r\right] = -\infty \quad (\forall r>0). \tag{28}$$

- The second contiguity we have to prove is the following relation: for any $r>0$,

$$\limsup\frac{1}{a_n}\log\mathbb{P}\left[\sup_{t\le 1}\left\|M_{[nt]} - N_{[nt]}^{(n)}\right\| > r\sqrt{na_n}\right] = -\infty. \tag{29}$$

We have

$$\sup_{t\le 1}\left\|M_{[nt]} - N_{[nt]}^{(n)}\right\| = \sup_{k\le n}\|P_k^{(n)}\| \le \sum_{j=1}^{n}\|(G-G_{R(n)})(X_j)\| + \|\pi(G-G_{R(n)})(X_{j-1})\|. \tag{30}$$

If we apply part 2 of Lemma 1 to $v_n = \sqrt{na_n}$, $\gamma = r+s$, and $R = R(n) = (n/a_n)^{\delta}$, we obtain

$$\frac{1}{a_n}\log\mathbb{P}\left[\frac{1}{\sqrt{na_n}}\sum_{i=0}^{n-1}(1+|X_{i,\nu}|^{r+s})\,\mathbb{I}_{\{|X_{i,\nu}|^{r+s}>R(n)\}} > r\right] \le -\frac{r\tau''}{2}\left(\frac{n}{a_n}\right)^{\delta'} + \frac{n}{a_n}V^* + \frac{\log E(\Phi)}{a_n} \tag{31}$$

with $\delta' = \frac{1}{2} + \delta\frac{\beta-(r+s)}{r+s}$; as $r+s < \frac{\beta}{2}$, we may choose δ such that $\frac{r+s}{2(\beta-r-s)} < \delta < \frac{1}{2}$, hence $\delta' > 1$, and thus (29) results from (24), (30), and (31).

- Next we show that (27) extends to the martingale $(N^{(n)})$. If we set $\tau_n = Tr < P^{(n)} >_n$, as $(< M >_k)_k$ and $(< P^{(n)} >_k)_k$ are both increasing sequences, we have
$$\sup_{k\leq n} \| < M >_k - < N^{(n)} >_k \| \ \leq\ \tau_n + 2\sqrt{\tau_n < M >_n},$$
(*see* [Duf] Proposition 4.2.11). Hence for any $r > 0$ there exists a $r' > 0$ such that
$$\begin{aligned} \mathbb{P}\left[\sup_{t\leq 1}\left\|\frac{<M>_{[nt]}}{n} - \frac{<N^{(n)}>_{[nt]}}{n}\right\| > r\right] &= \mathbb{P}\left[\sup_{k\leq n}\left\|<M>_k - <N^{(n)}>_k\right\| > nr\right] \\ &\leq \mathbb{P}\left[\left\|\frac{<!M>_n}{n} - S^2(g)\right\| > r'\right] + \mathbb{P}\left[\frac{\tau_n}{n} > r'\right] \\ &\overset{\mathrm{ld}}{<} 0 \end{aligned}$$
as soon as we prove that $\mathbb{P}\left[\tau_n/n > r'\right] \overset{\mathrm{ld}}{<} 0$ $(\forall r' > 0)$. This is the case thanks to relation (26) (with $R(n)$ instead of R), as we have
$$\tau_n = Tr\left[\sum_{j=1}^n \pi[(G - G_{R(n)})(G - G_{R(n)})^t](X_{j-1}) - (\pi(G - G_{R(n)})(\pi(G - G_{R(n)}))^t)(X_{j-1})\right].$$
From (27) it then comes
$$\mathbb{P}\left[\sup_{t\leq 1}\left\|\frac{< N^{(n)} >_{[nt]}}{n} - tS^2(g)\right\| > r\right] \overset{\mathrm{ld}}{<} 0 \quad (\forall r > 0). \tag{32}$$

- It finally remains to prove that the processes sequence $(N^{(n)}_{[n\bullet]}/\sqrt{n})$ satisfies the LDP stated in the theorem. The proof relies heavily on Puhalskii's criterion of obtaining a LDP for càd-làg processes sequences (*see* [Puh94] Theorem 2.1), as it is the case in [Dem96], where the MDP for martingales with bounded jumps is proved.

  In [Dem96], the functions $v(t) = \frac{2}{t^2}(e^t - 1 - t)$ and $w(t) = v(-t) - \frac{1}{2}t^2(v(-t))^2$ are introduced (and both have 1 as limit in 0): consequently, as for any given $\theta \in \mathbb{R}^q$ the sequence $(< \theta, N^{(n)}_k > \sqrt{a_n/n})_k$ has jumps bounded by $2\|\theta\|R(n)\sqrt{a_n/n} = 2\|\theta\|(n/a_n)^{\delta-1/2}$, the following sequences are respectively a supermartingale and a submartingale for every $\theta \in \mathbb{R}^q$:
$$\begin{aligned} (Z^{(\theta,n)}_t) &= \left(\exp\left[< \theta, N^{(n)}_{[nt]} > \sqrt{\tfrac{a_n}{n}} - \tfrac{a_n}{2n}\theta^t < N^{(n)} >_{[nt]} \theta\ v\left(2\|\theta\|(n/a_n)^{\delta-1/2}\right)\right]\right), \\ (Y^{(\theta,n)}_t) &= \left(\exp\left[< \theta, N^{(n)}_{[nt]} > \sqrt{\tfrac{a_n}{n}} - \tfrac{a_n}{2n}\theta^t < N^{(n)} >_{[nt]} \theta\ w\left(2\|\theta\|(n/a_n)^{\delta-1/2}\right)\right]\right). \end{aligned}$$
  Therefore, as $2\|\theta\|(n/a_n)^{\delta-1/2} \overset{n\to\infty}{\longrightarrow} 0$, (32) entails (for every $r > 0$ and $\theta \in \mathbb{R}^q$)
$$\begin{aligned} \limsup \tfrac{1}{a_n}\log\mathbb{P}\left[\sup_{t\leq 1}\left|\frac{\theta^t<N^{(n)}>_{[nt]}\theta}{2n}\ v(2\|\theta\|(n/a_n)^{\delta-1/2}) - \tfrac{t}{2}\theta^t S^2(g)\theta\right| > r\right] &= -\infty, \\ \limsup \tfrac{1}{a_n}\log\mathbb{P}\left[\sup_{t\leq 1}\left|\frac{\theta^t<N^{(n)}>_{[nt]}\theta}{2n}\ w(2\|\theta\|(n/a_n)^{\delta-1/2}) - \tfrac{t}{2}\theta^t S^2(g)\theta\right| > r\right] &= -\infty. \end{aligned}$$
  From Puhalskii's criterion we induce for $(N^{(n)}_{[n\bullet]}/\sqrt{na_n})$ a LDP of speed (a_n) and rate J defined in (11): thanks to (28) and (29), this LDP is also satisfied by the sequence $(\tilde{X}^{(n)}_\bullet)$, hence

Theorem 2 is proved.

5 Proof of theorem 3

- Let us first prove the following result.

Lemma 2 *Let (M_n) be an adapted sequence with values in $\mathcal{M}_{\delta,d}$, C an invertible $\delta \times \delta$ covariance matrix, and (C_n) a sequence of $\delta \times \delta$ symmetric random matrices such that, for some speed sequences (v_n) and (a_n) with $a_n = o(v_n)$, we have for every $r > 0$*

$$\lim \frac{1}{a_n} \log \mathbb{P}\left[\left\|\frac{C_n}{v_n} - C\right\| > r\right] = -\infty. \tag{33}$$

a) If $Q_n = I_d + C_n$, then for any $r > 0$

$$\lim \frac{1}{a_n} \log \mathbb{P}\left[\, \|v_n Q_n^{-1} - C^{-1}\| \geq r \,\right] = -\infty.$$

b) Let I be a rate on $\mathbb{R}^{\delta d}$ such that $\lim_{\|x\|\to\infty} I(x) = +\infty$. If $(vec\, M_n/\sqrt{a_n v_n})$ satisfies a LDP (resp. an upper LDP) of speed (a_n) and rate function I, then $(\sqrt{v_n/a_n} vec(Q_n^{-1}M_n))$ satisfies a LDP (resp. an upper LDP) of same speed and of rate function defined by

$$J(x) = I(vec\,(C mat\, x)) \qquad (\forall x \in \mathbb{R}^{\delta d}).$$

Proof:

a) Let A be a symmetric positive-definite matrix, with eigenvalues $\lambda_1, \ldots, \lambda_\delta$, and an orthonormal base of eigenvectors $e_1, \ldots, e_\delta$. We then have

$$\|A - I\| = \sup_i |\lambda_i - 1| \quad \text{et} \quad \|A^{-1} - I\| = \sup_i \frac{1}{\lambda_i}|\lambda_i - 1| \leq \|A - I\|/\lambda_{\min} A.$$

Hence, as $\|AB\| \leq \|A\|\|B\|$, and setting $R_n = C^{-1/2}\frac{Q_n}{v_n}C^{-1/2}$,

$$\limsup \frac{1}{v_n} \log \mathbb{P}\left[\, \|R_n - I\| \geq r \,\right] \leq \limsup \frac{1}{v_n} \log \mathbb{P}\left[\, \|\frac{Q_n}{v_n} - C\| \geq r\lambda_{\min} C \right] < 0,$$

for all $r > 0$, as $\|C^{-1}\| = 1/\lambda_{\min} C$. Moreover,

$$\mathbb{P}\left[\, \|v_n Q_n^{-1} - C^{-1}\| \geq r \,\right] \leq \mathbb{P}\left[\, \|R_n^{-1} - I\| \geq \frac{r}{\|C^{-1}\|} \right] \leq \mathbb{P}\left[\, \|R_n - I\| \geq r \frac{\lambda_{\min} R_n}{\lambda_{\min} C} \right],$$

hence, as $\lambda_{\min} R_n \leq 1/2$ implies that $\|R_n - I\| \geq 1/2$, we obtain

$$\mathbb{P}\left[\, \|v_n Q_n^{-1} - C^{-1}\| \geq r \,\right] \leq \mathbb{P}\left[\, \|R_n - I\| \geq 1/2 \,\right] + \mathbb{P}\left[\, \|R_n - I\| \geq r/2\lambda_{\min} C \,\right]$$

and we have proved *a)*.

b) Let us study the case of a full MDP (as the following argument only relies on contiguity in large deviations, the case of an upper MDP follows easily). The function $x \in \mathbb{R}^{\delta d} \mapsto vec\,(C\, mat\, x) \in \mathbb{R}^{\delta d}$ is continuous and one to one, of inverse $y \mapsto vec\,(C\, mat\, y)$: the contraction principle thus entails that the sequence $(vec\,(C^{-1}M_n)/\sqrt{a_n v_n})$ satisfies a LDP of speed (a_n) and rate function $J(\cdot)$. Therefore it remains to prove the exponential contiguity of speed (a_n) between $(vec\,(C^{-1}M_n)/\sqrt{a_n v_n})$ and $(\sqrt{v_n/a_n}\,vec\,(Q_n^{-1}M_n))$; in other words, we must prove that, for every $\rho > 0$,

$$\lim \frac{1}{a_n} \log \mathbb{P}\left[\left\| \sqrt{\frac{v_n}{a_n}} Q_n^{-1} M_n - \frac{C^{-1}M_n}{\sqrt{a_n v_n}} \right\| \geq \rho \right] = -\infty. \tag{34}$$

If $\rho > 0$ and $r > 0$, we have

$$\mathbb{P}\left[\| \sqrt{\frac{v_n}{a_n}} Q_n^{-1} M_n - \frac{C^{-1}M_n}{\sqrt{a_n v_n}} \| \geq \rho \right] \leq \mathbb{P}\left[\|v_n Q_n^{-1} - C^{-1}\| \geq r \right] + \mathbb{P}\left[\| \frac{M_n}{\sqrt{a_n v_n}} \| \geq \frac{\rho}{r} \right].$$

As by assumption $(vec\,M_n/\sqrt{a_n v_n})$ satisfies a LDP of rate $I(\cdot)$ and speed (a_n), we have

$$\limsup \frac{1}{a_n} \log \mathbb{P}\left[\| \frac{M_n}{\sqrt{a_n v_n}} \| \geq \frac{\rho}{r} \right] \leq -I(A_r)$$

for every given $r > 0$, where $A_r = \{x \in \mathbb{R}^{\delta d} / \|x\| \geq \rho/r\}$. By assumption, we have $\lim_{r\to 0} I(A_r) = +\infty$, hence by *a)*, making $r \to 0$, (34) comes and the proof is complete.

- The following result is the central result of [Wor98a]:

Theorem 5 *Let (Y_n) be an adapted sequence with values in $\mathbb{R}^\delta$, and (ϵ_n) a subGaussian noise of dimension d, dominated by a centered Gaussian distribution of covariance L. We suppose that (Y_n) satisfies, for some $\delta \times \delta$ covariance matrix $C \neq 0$ and some speed sequences (v_n) and (a_n) such that $a_n = o(v_n)$, the exponential convergence*

$$\lim \frac{1}{a_n} \log \mathbb{P}\left[\left\| \frac{1}{v_n} \sum_{j=1}^{n} (Y_{j-1})(Y_{j-1})^t - C \right\| \geq r \right] = -\infty, \quad \textit{for any } r > 0. \tag{35}$$

We consider the regressive series

$$M_n = \sum_{j=1}^{n} Y_{j-1} \epsilon_j^t.$$

a) The sequence $(vec\,M_n/\sqrt{a_n v_n})_{n\in\mathbb{N}}$ satisfies in $\mathbb{R}^{\delta d}$ an upper-LDP of speed (a_n) and rate

$$I_{C,L}(\xi) = \sup_{\bar\theta \in \mathbb{R}^{\delta d}} \{ < \bar\theta, \xi > -\tfrac{1}{2} \bar\theta^t C \otimes L \bar\theta \} \quad (\forall \xi \in \mathbb{R}^{\delta d}); \tag{36}$$

if C and L are invertible, then $I_{C,L}(\xi) = \frac{1}{2}\xi^t (C\otimes L)^{-1}\xi = \frac{1}{2}\xi^t C^{-1} \otimes L^{-1} \xi$.

b) In addition, if the noise is white with distribution $\mathcal{N}(0,\Gamma)$, **or** *if the noise satisfies property (P_Γ) and (Y_n) satisfies the exponential Lindeberg condition* **(C2)**, *then the sequence $(vec\,M_n/\sqrt{a_n v_n})_{n\in\mathbb{N}}$ satisfies a LDP of speed (a_n) and rate*

$$I_{C,\Gamma}(\xi) = \sup_{\bar\theta \in \mathbb{R}^{\delta d}} \{ < \bar\theta, \xi > -\tfrac{1}{2} \bar\theta^t C \otimes \Gamma \bar\theta \} \quad (\forall \xi \in \mathbb{R}^{\delta d}) \tag{37}$$

and $I_{C,\Gamma}(\xi) = \frac{1}{2}\xi^t C^{-1} \otimes \Gamma^{-1} \xi$ if C and Γ are invertible.

- Theorem 3 is now ready to be proved. Part *a)* of Theorem 5 above applies with $Y_j = \Phi_j$, hence the sequence $(vec\,M_n/\sqrt{a_n v_n})$ satisfies an upper-LDP of speed (a_n) and rate $I_{C,L}(\cdot)$ defined in (36): the application of part b) of Lemma 2 entails that $(\sqrt{v_n/a_n} vec\,(Q_n^{-1}M_n))$

satisfies the upper-LDP of speed (a_n) and rate

$$J_{C,L}(\xi) = I_{C,L}\left(vec\,(C\,mat\,\xi)\right) = \frac{1}{2}(vec\,(C\,mat\,\xi))^t C^{-1}\otimes L^{-1}(vec\,(C\,mat\,\xi)) \;=\; \xi^t C\otimes L^{-1}\xi.$$

In order to transfer this LDP to the sequence $(\sqrt{v_n/a_n}vec\,\tilde{\theta}_n)$, it thus suffices to prove the exponential contiguity of speed (a_n) between this sequence and $(\sqrt{v_n/a_n}vec\,(Q_n^{-1}M_n))$. This contiguity is provided by part $a)$ of Lemma 2: indeed, for $\rho > 0$ and $r > 0$, there exists some $n_0 = n_0(\rho, r)$ such that for $n \geq n_0$, $\frac{r+\|C^{-1}\|}{\|\theta\|\sqrt{a_n v_n}} \leq \rho$,

$$\begin{aligned}
\mathbb{P}[\,\|\sqrt{v_n/a_n}\tilde{\theta}_n - \sqrt{v_n/a_n}Q_n^{-1}M_n\| > \rho\,] &= \mathbb{P}[\,\|\sqrt{v_n/a_n}Q_n^{-1}\theta\| > \rho\,] \\
&\leq \mathbb{P}[\,\|v_nQ_n^{-1}\| > \rho\|\theta\|\sqrt{a_nv_n}\,] \\
&\leq \mathbb{P}[\,\|v_nQ_n^{-1}\| > r + \|C^{-1}\|\,] \\
&\leq \mathbb{P}[\,\|v_nQ_n^{-1} - C^{-1}\| > r\,]
\end{aligned}$$

and the latter tends "(a_n)-exponentially fast" to 0 thanks to part $a)$ of Lemma 2.

The proof of the full MDP (part $b)$) follows the same scheme as above, just by replacing L by Γ at each occurence, and using part $b)$ of Theorem 5.

6 Proof of theorem 4

As it was outlined at the end of **3.2.2**, it suffices to prove Proposition 2. As a matter of fact, when the latter is applied to the AR(1) model (II), then part 1) of Theorem 4 is achieved, and part 2) results form the application of Theorem 3 to the regression model (I), whose sequence of explicative variables is the sequence $(X_n^{(p)})$: indeed, relations (17), (18), and (19) correspond respectively to the conditions **(C1)**, the exponential Lyapunov condition (in case 2), and an obvious sufficient condition to the exponential Lindeberg condition (in case 1).

6.1 Preliminaries

- As the spectral radius $\rho(A)$ of A is such that $0 < \rho(A) < a < 1$ for some a, there exists a norm $|.|$ that satisfies
$$|Ax| \leq |A||x| \leq a|x|.$$
This norm will be much more convenient to use here, and we have the relations
$$c_1|\cdot| \leq \|\cdot\| \leq c_2|\cdot|(\text{with } 0 < c_1 < 1 < c_2),$$
where $\|.\|$ is the usual Euclidian matrix norm. Note that for some k, $\|A^k\| < a < 1$.

- Let C denote the *stationary covariance* of this stable AR(1) model, *i.e.*
$$C = \sum_{j=0}^{\infty} A^j\Gamma(A^j)^t = \lim_{k\to\infty} C(k) \quad \text{where} \quad C(k) = \sum_{j=0}^{k-1} A^j\Gamma(A^j)^t$$
(in case 1, Γ is the noise's covariance and C the covariance of the stationary distribution of the Markov model, whereas in case 2, $\Gamma = \mathbb{E}\,[\,\epsilon_n\epsilon_n^t|\,\mathcal{F}_{n-1}\,])$. Here we set $C_n = \sum_{j=0}^{n} X_jX_j^t$

and have to show that

$$\mathbb{P}[\ \|C_n/n - C\| > \epsilon\ ] \overset{\text{ld}}{<} 0 \qquad \text{for every } \epsilon > 0.$$

In **6.2** we prove the latter relation (which is a rewriting of (17)) and relation (18), the proof being valid in both case 1 and case 2. At the end of the section, we will finaly prove, in case 1 (white noise setting with $\beta = 2$), that relation (19) is a consequence of (17).

Remark 5: If, in case 1, we had assumed an exponential moment of order β strictly greater than 2, then (17) would have resulted from the application of Proposition 1 proved in Section **4.2** (with $g(x) = xx^t$, $\Lambda_n(g) = C_n/n$, and $\mu(g) = C$).

6.2 Proof of proposition 2

- **Exponential upper bound of $\|X_n\|/\sqrt{n}$:** we have for every θ and n,

$$X_n = A^n X_0 + \sum_{j=1}^{n} A^{n-j}\epsilon_j.$$

The noise being generalized Gaussian, $\exists \Lambda > 0$ such that $\mathbb{E}[\exp < u, \epsilon > | \mathcal{F}_{j-1}] \le \exp \Lambda \|u\|^2$, and we have for every θ

$$\begin{aligned} \mathbb{E}\left[\exp < n\theta, \tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^{n} A^{n-j}\epsilon_j >\right] &\le \textstyle\prod_{j=1}^{n} \exp n\Lambda \|(A^t)^{n-j}\theta\|^2 \\ &\le \exp n\Lambda (\tfrac{c_2}{c_1})^2 \|\theta\|^2 (\textstyle\sum_{i=0}^{n-1} a^i) \\ &\le \exp n\Lambda' \|\theta\|^2 \end{aligned}$$

with $\Lambda' = \Lambda(\frac{c_2}{c_1})^2/(1-a)$. Therefore the Gärtner-Ellis theorem implies that the sequence $\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} A^{n-j}\epsilon_j\right)$ satisfies the upper LDP of speed (n) and rate defined by $J(x) = \frac{1}{4\Lambda'}\|x\|^2$; hence, for every $r > 0$,

$$\limsup \frac{1}{n} \log \mathbb{P}\ [\ \|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} A^{n-j}\epsilon_j\| \ge r\ ] \le -r^2/4\Lambda'.$$

Moreover, with $E_0 = \mathbb{E}[\exp \tau \|X_0\|^2\ ]$, it comes

$$\frac{1}{n} \log \mathbb{P}\ \left[\ \|X_0\|/\sqrt{n} \ge r\ \right] \le \frac{\log E_0}{n} - \tau r^2.$$

We thus have established the following result.

Proposition 3 *In the framework of proposition 2, there exists a constant K such that*

$$\limsup \frac{1}{n} \log \mathbb{P}\ \left[\ \|X_n\|/\sqrt{n} \ge r\ \right] \le -Kr^2 \quad (\forall r > 0). \tag{38}$$

- **Exponential upper bound for TrC_n:** Proceeding as in **4.2.1**, we have

$$(1-a)^\beta \sum_{k=1}^{n} |X_k|^\beta \le |X_0|^\beta + \sum_{k=1}^{n} |\epsilon_k|^\beta.$$

With $E \geq \sup_n \mathbb{E}[\exp\tau\|\epsilon_n\|^\beta|\mathcal{F}_{n-1}]$, $\lambda \leq (1-a)^\beta c_1^\beta\tau$, and R sufficiently large, we have

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^n |X_k|^\beta \geq R\right] \leq E_0 \exp n(-\lambda R + \log E) \overset{\text{ld}}{<} 0.$$

Hence (18) is proved, as well as the following relation, taking $\beta = 2$ in the previous calculations:

$$\mathbb{P}\left[\frac{TrC_n}{n} > R\right] \overset{\text{ld}}{<} 0 \quad (\forall r > 0). \tag{39}$$

- **A splitting of $C_n/n - C$:** let $R > 0$ such that (39) is verified, and $\epsilon > 0$ be given. We have

$$\frac{C_n}{n} - C = \frac{C_n - A^k C_n (A^k)^t}{n} - C(k) + \frac{A^k C_n (A^k)^t}{n} + C(k) - C.$$

If we choose k sufficiently large such that $\|C(k) - C\| < \frac{\epsilon}{3}$ and $\|A^k\|^2 < \frac{\epsilon}{3R}$, then by (39), as

$$\left\{\frac{\|A^k C_n (A^k)^t\|}{n} > \frac{\epsilon}{3}\right\} \subset \left\{\frac{TrC_n}{n} > \frac{\epsilon}{3\|A^k\|^2}\right\} \subset \left\{\frac{TrC_n}{n} > R\right\},$$

all we need to prove to obtain (17) is that for every $r > 0$

$$\mathbb{P}\left[\ \|\frac{C_n - A^k C_n (A^k)^t}{n} - C(k)\| > r\right] \overset{\text{ld}}{<} 0. \tag{40}$$

- **Exponential upper bound of the distance from $(C_n - AC_nA^t)/n$ to Γ**

We have

$$X_j X_j^t = AX_{j-1}X_{j-1}^t A^t + \epsilon_j\epsilon_j^t + AX_{j-1}\epsilon_j^t + \epsilon_j(AX_{j-1})^t,$$

hence, taking the sum from $j = 1$ to n,

$$C_n - AC_nA^t = \Delta_n + D_n + n\Gamma$$

$$\text{where} \begin{cases} \Delta_n = X_0X_0^t - AX_nX_n^tA^t + \sum_{j=1}^n(\epsilon_j\epsilon_j^t - \Gamma), \\ D_n = \sum_{j=1}^n(AX_{j-1}\epsilon_j^t + \epsilon_j(AX_{j-1})^t). \end{cases}$$

By our assumption on $(\epsilon_n\epsilon_n^t - \Gamma)$ in case 2 (which is automatically satisfied in case 1, by the Cramer Theorem), and by (38). we have for any $r > 0$,

$$\mathbb{P}\left[\frac{\|\Delta_n\|}{n} \geq r\right] \overset{\text{ld}}{<} 0.$$

Let us show that it is also true with D_n. If $u, v \in \mathbb{R}^d$, we have

$$\begin{aligned} u^t D_n v &= \textstyle\sum_{j=1}^n (u^t AX_{j-1}\epsilon_j^t v + u^t\epsilon_j(AX_{j-1})^t v) \\ &= \textstyle\sum_{j=1}^n (< u, AX_{j-1} >< v, \epsilon_j > + < u, \epsilon_j >< v, AX_{j-1} >); \end{aligned}$$

but $\| < v, AX_{j-1} > u\|^2 \leq \|u\|^2\|v\|^2\|A\|^2 Tr(X_{j-1}X_{j-1}^t)$, and the noise is generalized Gaussian, hence if $\Lambda > 0$ is such that $\mathbb{E}[\exp < u, \epsilon > |\mathcal{F}_{j-1}] \leq \exp\Lambda\|u\|^2$ $(\forall u \in \mathbb{R}^d)$, then

$$\left(\exp(u^t D_n v - 2\|u\|^2\|v\|^2\|A\|^2\Lambda Tr(C_{k-1}))\right)_{k\geq 1}$$

is a positive supermartingale for every $u, v \in \mathbb{R}^d$. We take $u = e_i$ and $v = te_j$ where $t > 0$ and $(e_i)_{i=1}^n$ denotes the canonical base of $\mathbb{R}^d$, and consider the supermartingale

$$(Y_k)_{k\geq 1} = \big(\exp(te_i^t D_n e_j - 2t^2\|A\|^2 \Lambda Tr(C_{k-1}))\big)_{k\geq 1}$$

and the stopping time

$$\tau(n) = \inf\{k \geq 1; \frac{1}{n} TrC_k \geq R\}.$$

The sequence $(TrC_k)_{k\geq 1}$ is an increasing sequence, hence $\{\tau(n) \leq n\} \subset \{\frac{1}{n}TrC_n > R\}$ (as well as $\{\tau(n) > n\} \subset \{TrC_{n-1} \leq nR\}$), and (39) entails

$$\mathbb{P}[\,\tau(n) \leq n\,] \overset{\mathrm{ld}}{<} 0.$$

For $r > 0$, we then write

$$\begin{aligned} \mathbb{P}\left[\frac{e_i^t D_n e_j}{n} > r\right] &\leq \mathbb{P}[\,\tau(n) \leq n\,] + \mathbb{P}[\,Y_n > \exp n(rt - 2t^2\Lambda\|A\|^2(R+1))\,] \\ &\leq \mathbb{P}[\,\tau(n) \leq n\,] + \exp -n(rt - 2t^2\Lambda\|A\|^2(R+1)), \end{aligned}$$

and thus, taking $0 < t < r/2\Lambda\|A\|^2(R+1)$, we obtain for every $1 \leq i, j \leq n$ and every $r > 0$,

$$\mathbb{P}\left[\frac{e_i^t D_n e_j}{n} > r\right] \overset{\mathrm{ld}}{<} 0.$$

Proceeding in the same manner for $\mathbb{P}[\,e_i D_n e_j/n < -r\,]$ (with $u = e_i$ and $v = -te_j$) we achieve $\mathbb{P}[\,\|D_n\| > r\,] \overset{\mathrm{ld}}{<} 0$ for every $r > 0$, and consequently

$$\mathbb{P}\left[\|\frac{C_n - AC_nA^t}{n} - \Gamma\| > r\right] \overset{\mathrm{ld}}{<} 0. \tag{41}$$

- **Exponential upper bound for the distance from $(C_n - A^kC_n(A^k)^t)/n$ to $C(k)$**

Let us introduce the notation

$$C_j^{k,r} = \sum_{i=1}^{j} X_{r+ik}X_{r+ik}^t$$

and consider, with k and r fixed, the sequence $(X_{nk+r})_{n\geq 0}$, that satisfies to the relation

$$\begin{aligned} X_{(n+1)k+r} &= A^kX_{nk+r} + \epsilon_{(n+1)k+r} + A\epsilon_{(n+1)k+r-1} + \ldots + A^{k-1}\epsilon_{nk+r+1} \\ &= A^kX_{nk+r} + \eta_{n+1}^{(k,r)}. \end{aligned}$$

It is indeed a stable autoregressive model of noise $(\eta_n^{(k,r)})_{n\geq 1}$ still generalized Gaussian, with $C(k)$ as its stationary covariance, and to which we can apply relation (41): in other words we have

$$\mathbb{P}\left[\|\frac{C_n^{k,r} - A^kC_n^{k,r}(A^k)^t}{n} - C(k)\| > \rho\right] \overset{\mathrm{ld}}{<} 0 \quad (\forall \rho > 0).$$

But $C_{(n+1)k-1} = C_n^{k,0} + \ldots + C_n^{k,k-1} = \sum_{r=0}^{k-1} C_n^{k,r}$, hence

$$\mathbb{P}\left[\|\frac{C_{(n+1)k-1} - A^kC_{(n+1)k-1}(A^k)^t}{nk} - C(k)\| > \rho\right] \overset{\mathrm{ld}}{<} 0 \quad (\forall \rho > 0),$$

and, by Proposition 3, we see that we can replace in this last line $C_{(n+1)k-1}$ by $C_{(n+1)k}$. Consequently, it results rather easily that

$$\mathbb{P}\left[\|\frac{C_{[n/k]k}-A^kC_{[n/k]k}(A^k)^t}{[n/k]k}-C(k)\|>\eta\right]\overset{\mathrm{ld}}{<}0\quad(\forall\eta>0). \tag{42}$$

For every $n\in\mathbb{N}$, $n=[n/k]k+(n \bmod k)=p_nk+r_n$ (where $0\le r_n<k,\forall n$), and

$$C_n=C_{[n/k]k}+\sum_{j=1}^{r_n}X_{[n/k]k+j}X_{[n/k]k+j}^t=C_{[n/k]k}+R_n.$$

But relation (38) of Proposition 3 implies that

$$\mathbb{P}\left[\|\frac{R_n}{n}\|>\eta\right]\overset{\mathrm{ld}}{<}0\quad(\forall\eta>0), \tag{43}$$

and $\frac{[n/k]k}{n}\overset{n\to\infty}{\longrightarrow}1$, hence we finally achieve, putting (42) and (43) together, relation (40) and put an end to the proof of (17). The proof is thus completed in case 2.

- **Proof of (19) in case 1**

  We assume that the noise is white, and denote by μ the stationary distribution of the stable Markov model (X_n), which satisfies the conditions detailed in Section **1.2.1** (case (ii)). We end the proof of the theorem in case 1 by proving the following statement.

Lemma 3 *If the noise is white, then relation (17) implies property (19).*

Proof: We use the notations of Section **1.1.1**, and set $F(x)=\|x\|^2$,

$$F^{(R)}(x)=F(x)\min\{1\,;\,(F(x)-(R-1))_+\},\qquad\text{and}\qquad\tilde F^{(R)}(x)=\min\{F(x);R\}.$$

We have $0\le F(x)\,\mathbb{I}_{\{F(x)\ge R\}}\le F^{(R)}(x)\le F(x)$, hence

$$0\ \le\ \Lambda_n(F\,\mathbb{I}_{F\ge R})\ \le\ \Lambda_n(F^{(R)}-\tilde F^{(R)})\ +\ \Lambda(\tilde F^{(R)})-\mu(\tilde F^{(R)})\ +\ \mu(\tilde F^{(R)})$$

where

$$F^{(R)}-\tilde F^{(R)}\ =\ (F^{(R)}-R)\,\mathbb{I}_{F^{(R)}\ge R}\ \le\ (F-R)\,\mathbb{I}_{F\ge R}\ =\ F-(F\wedge R).$$

Let $r>0$. The functions $\tilde F^{(R)}$ and $F-(F\wedge R)$ being both continuous, bounded by F (hence μ-integrable), and with pointwise limit 0 as $R\to+\infty$, by Lebesgue's theorem there exists some $R>0$ sufficiently large such that $\mu(\tilde F^{(R)})+\mu(F-(F\wedge R))<\frac{r}{4}$, and consequently it comes from the bounds above

$$\mathbb{P}\left[\tfrac{1}{n}\textstyle\sum_{j=0}^{n-1}\|X_j\|^2\mathbb{I}_{\{\|X_j\|^2\ge R\}}>r\right]$$
$$\le\ \mathbb{P}\,[\,\Lambda_n(F)-\mu(F)>\tfrac{r}{4}\,]+\mathbb{P}\,[\,\Lambda_n(F\wedge R)-\mu(F\wedge R)>\tfrac{r}{4}\,]+\mathbb{P}\,[\,\Lambda_n(\tilde F^{(R)})-\mu(\tilde F^{(R)})>\tfrac{r}{4}].$$

For the second and third term, as the functions $F\wedge R$ and $\tilde F^{(R)}$ are bounded and continuous, we apply relation (5) to both of them, and for the first term we use (17): relation (19) is thus proved. $\square$

Acknowledgements

I would like to thank the referee for his careful reading of the paper, as well as for drawing my attention on an error which appeared in the first version of the paper. Furthermore I am grateful to Professor Marie Duflo for her very helpful discussion, and her suggestions leading to a substantially improved presentation.

References

[Aco90] A. De Acosta. Large deviations for empirical measures of Markov chains. *J. Theoretical Prob.*, 3:395–431, 1990.

[Aco97] A. De Acosta. Moderate deviations for empirical measures of Markov chains: lower bounds. *Ann. Probability*, 25:259–284, 1997.

[AcoChe98] A. De Acosta and X. Chen. Moderate deviations for empirical measures of Markov chains: upper bounds. *J. Theoretical Prob.*, 11:1075–1110, 1998.

[BenMetPri] A. Benveniste M. Metivier and P. Priouret. *Adaptive algorithms and stochastic appproximation.* Springer, 1990.

[BerGamRou96] B. Bercu F. Gamboa and A. Rouault. Large deviations for quadratic forms of stationary Gaussian processes. *Stoch. processes and their appl.*, 71:75–90, 1997.

[Bor91] A. A. Borovkov. Lyapunov functions and multidimensionnal Markov chains. *Theory of probability and applications*, 36:1–18, 1991.

[BryDem95] W. Bryc and A. Dembo. On large deviations of empirical measures for stationary Gaussian processes. *Stochastic processes and appl.*, 58:23–34, 1995.

[BryDem96] W. Bryc and A. Dembo. Large deviations and strong mixing. *Ann. Inst. Henri Poincaré, Probab. stat.*, 32:549–569, 1996.

[BrySmo93] W. Bryc and W. Smolenski. On large deviations of empirical measures for quadratic functional of the autoregressive process. *Stat. and probability letters*, 17:281–285, 1993.

[Dem96] A. Dembo. Moderate deviations for martingales with bounded jumps. *Electronic Communication in Probab.*, 1:11–17, 1996.

[DesPic79] J. Deshayes and D. Picard. Grandes et moyennes déviations pour les marches aléatoires. *Astérisque*, 68:53–70, 1979.

[DeuStr] J. D. Deuschel and D. W. Stroock. *Large deviations.* Academic press, 1989.

[DinNey95] I. H. Dinwoodie and P. Ney. Occupation measures of Markov chains. *J. Theoretical Prob.*, 8(3):679–691, 1995.

[DonVar75a] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time I. *Commun. Pure Appl. Math.*, 28:1–47, 1975.

[DonVar75b] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time II. *Commun. Pure Appl. Math.*, 28:279–301, 1975.

[DonVar76] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time III. *Commun. Pure Appl. Math.*, 29:389–461, 1976.

[Duf] M. Duflo. *Random Iterative Models.* Springer, 1997.

[DupEll] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley, 1997.

[Ell88] R.S. Ellis. Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Ann. Probability*, 16:1496–1508, 1988.

[Gao96] F.Q. Gao. Moderate deviations for martingales and mixing processes. *Stochastic processes and appl.*, 61:263–277, 1996.

[GulLipLot94] O. V. Gulinskii R. S. Liptser S. V. Lototskii. Large deviations for unbounded additive functionals of a Markov process with discrete time (non compact case). *J. Appl. Math. Stoch. Anal.*, 7(3):423–436, 1994.

[HorJoh] R.A. Horn and C.R. Johnson. *Topics in matrix analysis.* Cambridge university press, 1991.

[Jai90] N.C. Jain. Large deviation lower bounds for additive functionals of Markov processes: discrete time, non compact case. *Ann. probability*, 18:1071–1098, 1990.

[Lip96] R. S. Liptser. Large deviation property for occupation measures of Markov processes: discrete time, non compact case. *Theory probability and appl.*, 41:35–54, 1996.

[LipPuh] R. S. Liptser and A. A. Puhalskii. Limit theorems on large deviations for semi-martingales. *Stochastics and stochastic reports*, 38:201–249, 1992.

[MeyTwe] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, 1993.

[Pet] V. V. Petrov. *Limit theorems of probability theory, sequences of independent random variables.* Clarendon Press, Oxford, 1995.

[Puh94] A. A. Puhalskii. The method of stochastic exponentials for large deviations. *Stochastic processes and appl.*, 54:45–70, 1994.

[Puh97] A. A. Puhalskii. Large deviations of semimartingales: a maxingale problem approach 1. limits as solutions to a maxingale problem. *Stochastics and Stochastics Reports*, 61:141–243, 1997.

[Wor98a] J. Worms. Moderate deviations for regression models. *Preprint, Université de Marne-la-Vallée, number 1*, 1998.

[Wor98b] J. Worms. Principes de déviations modérées pour des modèles autorégressifs d'ordre p (moderate deviations principles for autoregressive models of order p). *Compte Rendu de l'Académie des Sciences, série 1*, 328(1):67–72, 1999.

[Wu95] L. Wu. Moderate deviations of dependent random variables related to a CLT. *Ann. of probability*, 23:420–445, 1995.