Numéro d'ordre : —

UNIVERSITÉ BLAISE PASCAL

(UFR Sciences et Technologies)

Laboratoire de Mathématiques – UMR 6620

Habilitation à Diriger des Recherches

présentée par

Hacène Djellout

Tome II – Textes des Articles

Sujet de l'Habilitation :

Quelques Contributions à la Statistique des Processus :
Inégalités de Déviations & Grandes Déviations

Soutenue le devant le jury composé de :

Rapporteurs Fabrice GAMBOA, Université de Toulouse (Toulouse, France)
Marc HOFFMANN, Université Paris-Dauphine (Paris, France)
Christian HOUDRÉ, Georgia Institute of Technology (Atlanta, USA)

Table des matières

I Grandes déviations pour les estimateurs de la (co)-volatilité 1

Large and Moderate Deviations for Estimators of Quadratic Variational Processes of Diffusions (with A. Guillin & L. Wu)
Statistical Inference for Stochastic Processes
vol. 2, pp. 195–225, 1999 **3**

Large and Moderate Deviations for Estimators of realized covolatility (with Y. Samoura)
Statistics and Probability Letters
vol. 86, pp. 30–37, 2014 **35**

Large deviations of the realized (co-) volatility vector (with A. Guillin & Y. Samoura)
Submitted
vol. , pp. – , 2015 **43**

Large deviations of the threshold estimator of integrated (co-) volatility vector in the presence of jumps (with H. Jiang)
Submited
vol. , pp. – , 2015 **81**

II Chaînes de Markov bifurcantes, théorèmes limites 97

Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application (with S. V. Bitseki Penda & A. Guillin)
The Annals of applied Probability
vol. 24, no. 1, pp. 235–291, 2014 **99**

Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models (with S. V. Bitseki Penda)
Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
vol. 50, no. 3, pp. 806–844, 2014 **157**

III Inégalités du coût du transport et applications 197

Transportation cost-information inequalities and applications to random dynamical systems and diffusions (with A. Guillin & L. Wu)
The Annals of Probability
vol. 32, no. 3B, pp. 2702–2732, 2004 **199**

Lipschitzian nom estimate of one-dimensional Poisson equations and applications (with L. Wu)
Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
vol. 47, no. 2, pp. 450–465, 2011 **231**

IV Déviations modérées pour des variables aléatoires dépendantes 247

Moderate deviations of empirical periodogram and non-linear functionals of moving average processes (with A. Guillin & L. Wu)
Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
vol. 42, pp. 393–416, 2006 **249**

Moderate deviations for the Durbin-Watson statistics related to the first-order autoregressive process (with S. V. Bitseki Penda & F. Proïa)
ESAIM - Probability and Statistics
vol. 18, pp. 308–331, 2014 **273**

Moderate deviations for martinagle differences and applications to ϕ-mixing sequences
Stochastics and Stochastics Reports
vol. 73, no. 1-2, pp. 37–63, 2002 **297**

Large and moderate deviations for moving average processes (with A. Guillin)
Annales de la Faculté des Sciences de Toulouse
vol. X, no. 1, pp. 23–31, 2001 **325**

Moderate deviations for Markov chains with atom (with A. Guillin)
Stochastic Processes and their Applications
vol. 95, pp. 203–217, 2001 **335**

Principe de déviations modérées pour le processus empirique fonctionnal d'une chaîne de Markov (with A. Guillin)
C. R. Acad. Sci. Paris, t. 330, Série I
vol. 330, no. 1, pp. 377–380, 2000 **351**

Première Partie

Grandes déviations pour les estimateurs de la (co)-volatilité

Statistical Inference for Stochastic Processes **2:** 195–225, 1999.

Large and Moderate Deviations for Estimators of Quadratic Variational Processes of Diffusions

HACÈNE DJELLOUT, ARNAUD GUILLIN and LIMING WU
Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620 Université Blaise Pascal, 63177 Aubiere, France

Abstract. For a diffusion process $\mathrm{d}X_t = \sigma_t \mathrm{d}B_t + b(t, X_t)\mathrm{d}t$ with (σ_t) unknown, we study the large and moderate deviations of the estimator $\bar{\Theta}_n(t) := \sum_{k=0}^{[nt]} (X_{k/n} - X_{(k-1)/n})^2$ of the quadratic variational process $\Theta(t) = \int_0^t \sigma_s^2 \mathrm{d}s$.

AMS Mathematics Subject Classifications (1991): 60F10, 62J05, 60J05.

Key words: large deviation, moderate deviation, quadratic variational process.

Introduction

Consider the time evolution of some quantity $(X_t)_{0 \leqslant t \leqslant 1}$ in a random environment modelized by the Ito's stochastic differential equation

$$\mathrm{d}X_t = \sigma_t \mathrm{d}B_t + b(t, \omega)\,\mathrm{d}t, \tag{0.1}$$

defined on some filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbf{P})$ (satisfying the usual condition), where $(b(t, \omega))_{t \geqslant 0}$ is an adapted process denoting the *mean forward velocity*, (B_t) is a real standard Brownian motion, and $0 \leqslant \sigma_t \in L^2(\mathbf{R}^+, \mathrm{d}t)$ (deterministic) represents the *strength of the random perturbation (or noise) at time t*. Assume that (σ_t) is unknown and we want to estimate it from a sample $(X_t)_{0 \leqslant t \leqslant 1}$, or more exactly to estimate the unknown quadratic variational process of X,

$$\Theta(t) := [X]_t = \int_0^t \sigma_s^2 \mathrm{d}s, \quad t \in [0, 1]. \tag{0.2}$$

This question appears very naturally in mathematical finances where σ_t is called *volatility*. D. Florens-Zmirou [7] studied it both from the parametric and non-parametric statistic point of view, and she obtained the consistency and the central limit theorem of her estimators. Several further works have been realized by Avesani and Bertrand [1], Bertrand [2] about various statistical questions related to this model. See [2] for relevant references.

By Stochastic analysis and by following [7] and [2], a natural estimator of $\Theta(\cdot) = [X]$, is the *empirical quadratic variational process* of X

$$Q_t^n(X) := \sum_{k=1}^{[nt]} (X_{t_k^n} - X_{t_{k-1}^n})^2, \quad t \in [0,1], \quad n \geqslant 1, \tag{0.3}$$

where $\tau_n = \{t_k^n := \frac{k}{n}; 0 \leqslant k \leqslant n\}$ is the equi-partition of $[0,1]$ into n parts, $[x]$ denotes the integer part of $x \in \mathbf{R}$. If the drift $b(t,\omega)$ is known, we can consider the following variant

$$Q_t^n\,(X - Y) = \sum_{k=1}^{[nt]} \left(X_{t_k^n} - X_{t_{k-1}^n} - \int_{t_{k-1}^n}^{t_k^n} b(s,\omega)\, \mathrm{d}s \right)^2 \tag{0.4}$$

where $Y_\cdot := \int_0^\cdot b(t,\omega)\,\mathrm{d}t$.

If the drift $b(t,\omega) = b(t, X_t(\omega))$ where $b(t,x)$ is some deterministic function (a current situation), (X_t) verifies

$$\mathrm{d}X_t = \sigma_t \mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t. \tag{0.5}$$

When $b(t,x)$ is known, and only the sample $(X_{t_k^n}; k = 0, \dots, n)$ is observed, we can also consider the following estimator

$$\tilde{Q}_t^n(X) := \sum_{k=1}^{[nt]} \left(X_{t_k^n} - X_{t_{k-1}^n} - b(t_{k-1}^n, X_{t_{k-1}^n})(t_k^n - t_{k-1}^n) \right)^2. \tag{0.6}$$

Under the mild condition that $\sup_{0 \leqslant t \leqslant 1} |b(t,\omega)| < +\infty$, $\mathbf{P} - a.s$,

$$Q_t^n(X), \quad Q_t^n\,(X - Y)\,, \quad \tilde{Q}_t^n(X) \longrightarrow \int_0^t \sigma_s^2\,\mathrm{d}s, \quad \mathbf{P} - a.s. \tag{0.7}$$

(well known, see [12]). Thus they all are strongly consistent estimators of $\Theta(t) = [X]_t$.

The purpose of this paper is to furnish some further estimations about these estimators, refining the already known central limit theorem [7, 1, 2]. More precisely, we are interested in the estimations of

$$\mathbf{P}\left(\frac{\sqrt{n}}{b(n)} \left(\bar{\Theta}_n(\cdot) - \Theta(\cdot) \right) \in A \right),$$

where $\bar{\Theta}_n(\cdot)$ denotes one of the three estimators in (0.3), (0.4) and (0.6) above, A is a given domain of deviation, $(b(n) > 0)$ is some sequence denoting the scale of deviation. When $b(n) = 1$, this is exactly the estimation of the central limit theorem. When $b(n) = \sqrt{n}$, it becomes the *large deviations*. And when $1 \ll b(n) \ll \sqrt{n}$, this is the so called *moderate deviations*. The main problem studied in this paper is:

What are the large and moderate deviations estimations of the estimators $\bar{\Theta}_n(t) = Q_t^n(X)$, *or* $Q_t^n\,(X - Y)$, *or* $\tilde{Q}_t^n(X)$?

The above question will be investigated from two points of view : the parametrical statistical one when $t = 1$ is fixed ; and the nonparametrical statistical one when t varies in [0, 1]. Let us regard roughly the feature of this question.

The first point is that the question can be reduced to the *no-drift* (i.e., $b = 0$) gaussian case by approximation technique (see Section 2). And in this simplified case, the exact calculations could be made in principle (but one waits the explicit results, too).

The second point of this object is that $t \to Q_t^n(X - Y)$ has independent increments which are not homogeneous once $t \to \sigma_t$ is not constant. So it is much like partial sums of non-i.i.d.r.v., and we can not directly apply the powerful theory in the i.i.d. case.

Though we have not found the studies exactly on this question in the literatures (it is a surprise to us), but technically we are much inspired from the two lines of the studies:

(1) the work of Lynch and Sethuraman [10] on large deviations of processes with independent increments, and Pukhalski [11] about large deviations of stochastic processes;
(2) the works of Bryc and Dembo [3], of Bercu, Gamboa and Rouault [4] about the large and moderate deviations of quadratic forms of a stationary gaussian process.

Especially we will encounter the same technical difficulties:

(1) a correction (or extra) term appears in the evaluation of the rate function in the process-level large deviations because of the *weak* exponential integrability of $Q^n(X)$, a phenomenon first discovered by Lynch and Sethuraman [10]; and
(2) the Ellis–Gärtner Theorem can not be applied in some situations, as in [3, 4].

This paper is organized as follows. In the next Section we present the main results of this paper. They are established successively in the remaining part of the paper. The related works are often presented in the remarks.

1. Main Results

Let us present now the main results of this paper. We follow [5] for the language of large deviations, throughout this paper.

1.1. LARGE DEVIATIONS AT A FIXED TIME

Our first result is about the large deviations of $Q_1^n(X)$ from $[X]_1$, with time $t = 1$ fixed.

THEOREM 1.1 *Let* (X_t) *satisfy (0.1) and* $Y_\cdot = \int_0^\cdot b(t, \omega)\, \mathrm{d}t$.

(a) For every $\lambda \in \mathbf{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E} \exp\left(\lambda n Q_1^n(X-Y)\right) = \Lambda(\lambda) := \int_0^1 P(\lambda\sigma_t^2)\,\mathrm{d}t \tag{1.1}$$

in $(-\infty, +\infty]$, *where*

$$P(\lambda) = -\tfrac{1}{2}\log(1-2\lambda) \quad \textit{for} \quad \lambda < \tfrac{1}{2} \textit{ and } +\infty \textit{ else}. \tag{1.2}$$

(b) Assume that $\sigma_\cdot \in L^\infty([0,1], \mathrm{d}t)$ *and* $b(\cdot,\cdot) \in L^\infty(\mathrm{d}t \otimes \mathbf{P})$. *Then* $\mathbf{P}(Q_1^n(X) \in \cdot)$ *satisfies the large deviation principle (in abridge: LDP) with speed* n *and with the good rate function* I *given by the Legendre transformation of* Λ, *that is,*

$$I(x) := \Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda);\ \lambda \in \mathbf{R}\}. \tag{1.3}$$

In other words I *is inf-compact and for each Borel subset* A *of* $\mathbf{R}$,

$$-\inf_{x\in A^\circ} I(x) \leqslant \lim_{n\to\infty} \begin{pmatrix}\inf\\ \sup\end{pmatrix} \frac{1}{n} \log \mathbf{P}\left(Q_1^n(X) \in A\right) \leqslant -\inf_{x\in\bar{A}} I(x). \tag{1.4}$$

Remarks 1.1. By the well-known Ellis–Gärtner Theorem ([6, Th.II.6.1] or [5, § 2.3]), we can deduce the LDP (1.4) for $Q_1^n(X-Y)$ (instead of $Q_1^n(X)$) from (1.1) once if

$$\lim_{\lambda\uparrow\lambda_0} \Lambda'(\lambda) = \lim_{\lambda\uparrow\lambda_0} \int_0^1 \frac{\sigma_t^2}{1-2\lambda\sigma_t^2}\,\mathrm{d}t = \int_0^1 \frac{\sigma_t^2}{1-2\lambda_0\sigma_t^2}\,\mathrm{d}t =: A = +\infty, \tag{1.5}$$

where $\lambda_0 := \frac{1}{2\|\sigma^2\|_\infty}$. The main difficulty in this result resides just in the case where $\lambda_0 < +\infty$ and

$$A = \int_0^1 \frac{\sigma_t^2}{1-2\lambda_0\sigma_t^2}\,\mathrm{d}t < +\infty, \tag{1.6}$$

in which Ellis–Gärtner's theorem is not applicable, because Λ is not *steep*, see [5, § 2.3].

Notice that the rate function $\Lambda^*(x)$ is strictly convex on $(0, A)$, $\Lambda^*(x) = +\infty$, $\forall x \leqslant 0$ and $\Lambda^*(x) = \lambda_0 x - \Lambda(\lambda_0)$ (affine) for all $x \geqslant A$. See Proposition 1.5. below for some explicit estimations.

Remarks 1.2. $\sigma_\cdot \in L^\infty([0,1], \mathrm{d}t)$ is a necessary condition to the LDP in (1.4). See Corollary 2.3.

1.2. LARGE DEVIATIONS OF PROCESS-LEVEL

We now extend Theorem 1.1. to the process-level large deviations of $\bar{\Theta}_n(\cdot) := Q_\cdot^n(X)$, which is interesting from the viewpoint of non-parametric statistics.

Let $\mathbf{D}^+[0,1]$ be the real right-continuous-left-limit non-decreasing functions γ with $\gamma(0)$ non-negative. The space $\mathbf{D}^+[0,1]$ of γ, identified in the usual way as the space of non-negative bounded measures $d\gamma$ on [0,1] with $d\gamma[0,t] = \gamma(t)$, will be equipped with the weak convergence topology. The empirical quadratic processes defined in (0.3) and (0.4) are $\mathbf{D}^+([0,1])$-valued random variables with respect to the σ-field $\mathcal{B}^s$ generated by the coordinates $\{\gamma(t); 0 \leqslant t \leqslant 1\}$.

THEOREM 1.2. *Given (X_t) by (0.1) with $\sigma_\cdot \in L^\infty([0,1], dt)$, $b(\cdot,\cdot) \in L^\infty(dt \otimes \mathbf{P})$.*

(a) $\mathbf{P}\left(Q^n_\cdot(X) \in \cdot\right)$ satisfies the LDP on $\mathbf{D}^+([0,1])$ w.r.t. the weak convergence topology, with speed n and with some inf-compact convex rate function $J(\gamma)$.
(b) If moreover $t \to \sigma_t$ is continuous and strictly positive on $[0,1]$, then

$$J(\gamma) = \int_0^1 P^*\left(\frac{\dot\gamma(t)}{\sigma_t^2}\right) dt + \frac{1}{2}\int_0^1 \frac{1}{\sigma_t^2}\, d\gamma^\perp(t), \tag{1.7}$$

where $\dot\gamma(t)\,dt$, $d\gamma^\perp$ are respectively the absolute continuous part and the singular part of the measure $d\gamma$ associated with $\gamma \in \mathbf{D}^+[0,1]$ w.r.t. the Lebesgue measure dt, and

$$P^*(x) = \begin{cases} \frac{1}{2}(x - 1 - \log x) & \text{if} \quad x > 0 \\ +\infty & \text{if} \quad x \leqslant 0 \end{cases} \tag{1.8}$$

which is the Legendre transformation of $P(\lambda)$ given in (1.2).

Remarks 1.3. In the homogenous case $\sigma_t = \sigma$ constant, this theorem is already obtained by [4, Theorem 7].

Remarks 1.4. We emphasize that the LDP in the part (a) holds only w.r.t. the weak convergence topology, not w.r.t. the Skorohod topology even in the simplest case where $\sigma_t = \sigma$ is constant. In fact in this last case, Lynch and Sethuraman [10] found that $J(\gamma)$ given by the right hand side (in short : RHS) of (1.7) is not inf-compact w.r.t. the Skorohod topology. A more direct way to see this point is

$$\lim_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(Q^n_{t+1/n}(X) - Q^n_t(X) > x\right) = -\frac{x}{2\sigma_t^2}, \quad \forall x > 0, \ \ t \in [0,1)$$

where $t \to \sigma_t^2$ is assumed continuous and $b = 0$ (an easy calculation).

Remarks 1.5. In the evaluation of $J(\gamma)$, one finds rather naturally the first term in the RHS of (1.7), but not so easily the second correction term using $\gamma^\perp$. Hence even for 'bad' configurations such as $d\gamma = d[X] + \delta_t$ (δ_t the Dirac measure at $t \in [0,1]$), $Q^n_\cdot(X)$ can fall into any small neighborhood of γ with a non-negligible exponential small probability (this is clear also from the estimation in the remark above).

Notice that this curious phenomenon (with an extra term including $\gamma^\perp$) was at first found by Lynch and Sethuraman [10, Th.3.2] in their investigations of large deviations of processes with (homogeneous) independent increments. As said at the beginning, we are much inspired by their work.

Remarks 1.6. One can weaken slightly the strong condition in (b) : if $\sigma_\cdot \geqslant \epsilon > 0$ is piecewisely continuous on [0, 1] (i.e., has only a finite number of discontinuity points of the first type), then the rate function J in Theorem 1.2.(a) will be instead of (1.7), given by

$$J(\gamma) = \int_0^1 P^*\left(\frac{\dot\gamma(t)}{\sigma_t^2}\right) \mathrm{d}t + \frac{1}{2}\int_0^1 \frac{1}{\overline{\sigma}_t^2}\, \mathrm{d}\gamma^{\perp}(t),$$

where $\overline{\sigma}_t = \max\{\sigma_{t+}, \sigma_{t-}\}$ is the upper semi-continuous version of $\sigma_\cdot$. This can be established by applying Theorem 1.2.(b) to each sub-interval over which $\sigma_\cdot$ is continuous, and using the independence of increments of $Q_\cdot^n(X-Y)$ and the contraction principle.

1.3. MODERATE DEVIATIONS

We discuss now the moderate deviations of $Q^n(X)$. To this purpose, let $(b(n))_{n\geqslant 1}$ be a sequence of positive numbers such that

$$b(n) \longrightarrow +\infty, \quad \frac{b(n)}{\sqrt{n}} \longrightarrow 0. \tag{1.9}$$

Let $\mathbf{D}_0([0,1])$ be the Banach space of real right-continuous-left-limit functions γ on [0, 1] with $\gamma(0) = 0$, equipped with the *uniform sup norm* (it is non-separable!). The associated Borel σ-field is too large. We shall use the σ-field $\mathcal{B}^s$ generated by the coordinates $\{\gamma(t); 0 \leqslant t \leqslant 1\}$.

THEOREM 1.3. *Given* (X_t) *by (0.1) with* $b(t,\omega) \in L^\infty(\mathrm{d}t \otimes \mathbf{P})$. *Assume* $\sigma_t^2 \in L^2([0,1], \mathrm{d}t)$ *and*

$$\sqrt{n}b(n) \max_{1\leqslant k\leqslant n} \int_{(k-1)/n}^{k/n} \sigma_t^2\, \mathrm{d}t \longrightarrow 0, \tag{1.10}$$

or more particularly for some $p > 2$,

$$\sigma_t^2 \in L^p([0,1], \mathrm{d}t) \quad \text{and} \quad b(n) = O\left(n^{\frac{1}{2}-\frac{1}{p}}\right). \tag{1.11}$$

Then for each $\mathcal{B}^s$*-measurable subset* $A \subset \mathbf{D}_0([0,1])$,

$$\begin{aligned} -\inf_{\gamma\in A^\circ} J_m(\gamma) &\leqslant \lim_{n\to\infty} \binom{\inf}{\sup} \frac{1}{b^2(n)} \times \\ &\times \log \mathbf{P}\left(\frac{\sqrt{n}}{b(n)}\left(Q_\cdot^n(X) - [X]_\cdot\right) \in A\right) \leqslant -\inf_{\gamma\in\bar{A}} J_m(\gamma), \end{aligned} \tag{1.12}$$

where A^0, $\bar{A}$ are taken w.r.t. the sup norm topology, and the good rate function J_m is given by

$$J_m(\gamma) = \begin{cases} \displaystyle\int_0^1 \frac{\dot{\gamma}(t)^2}{4\sigma_t^4} 1_{[t:\sigma_t>0]}\, \mathrm{d}t & \text{if } \mathrm{d}\gamma \ll \sigma_t^2\, \mathrm{d}t \\ +\infty & \text{otherwise} \end{cases} \tag{1.13}$$

In particular, $\mathbf{P}\left(\frac{\sqrt{n}}{b(n)}\left(Q_1^n(X) - [X]_1\right) \in \cdot\right)$ *satisfies the LDP on* $\mathbf{R}$ *with speed* $b^2(n)$ *and with the rate function given by*

$$I_m(x) = \frac{x^2}{4\int_0^1 \sigma_s^4\, \mathrm{d}s}, \quad \forall x \in \mathbf{R}. \tag{1.14}$$

Remarks 1.7. In the literature, the large deviation principles in Theorem 1.3 are often called Moderate Deviation Principles (in abridge MDP, see e.g. [5]). Notice an essential difference between the large deviations in Theorem 1.2 and the moderate deviation in Theorem 1.3 – the LDP in Theorem 1.3. holds even w.r.t. the sup norm topology which is much stronger than the Skorohod topology.

We emphasize that our condition (1.11) is inspired by Bryc and Dembo [3, Theorem 2.3].

1.4. THE UNBOUNDED DRIFT CASE

In the previous three results we have imposed the boundedness of $b(t, \omega)$, which allows us to reduce very easily the large and moderate deviations of $Q^n(X)$ to those of $Q^n(X - Y)$ (no drift case). It is very natural to ask whether they continue to hold under the *Lipchitzian condition* or more generally *linear growth* condition of the drift $b(t, x)$, rather than the boundedness. This is the object of the following.

THEOREM 1.4. *Given (X_t) by (0.5) with X_0 bounded. Assume that the drift b satisfies the following uniform linear growth condition:*

$$|b(s,x) - b(t,y)| \leqslant C\,[1 + |x-y| + \eta(|s-t|)(|x| + |y|)]\,, \quad \forall s,t \in [0,1], x,y \in \mathbf{R}, \tag{1.15}$$

where $\eta\colon [0,+\infty) \to [0,+\infty)$ is a continuous non-decreasing function with $\eta(0) = 0$ and $C > 0$ is a constant.

(a) Assume $\sigma_\cdot \in L^\infty([0,1], \mathrm{d}t)$. Then $\tilde{Q}^n_\cdot(X)$ defined by (0.6) satisfies the LDPs in Theorem 1.1 and Theorem 1.2.

(b) Assume (1.10). The MDP in Theorem 1.3. continue to hold for $\tilde{Q}^n_\cdot(X)$ (instead of $Q^n(X)$).

Remarks 1.8. The uniformly linear growth condition (1.15) is satisfied for example under

(1) the usual Lipchitzian condition; or
(2) $b(t,x) = g(x)$ where $g(x)$ satisfies: $|g(x) - g(y)| \leqslant C\,(1 + |x - y|)$; or
(3) $b(t,x) = f(t)g(x)$ with f continuous on [0, 1] and g satisfies (2) above.

Remarks 1.9. As the reader can imagine naturally, the key is to show that the difference between $Q^n(X - Y)$ and $\tilde{Q}^n(X)$ is negligible in the senses of both large deviations and moderate deviations. This will be realized by means of three powerful tools: Gronwall's inequality, Lévy's maximal inequality and an isoperimetric inequality for gaussian processes.

1.5. EXACT DEVIATION INEQUALITIES

The LDP in Theorem 1.2. holds only w.r.t. the weak convergence topology and then it does not give precise estimation about deviation domains such as $\left\{\gamma : \sup_{t\in[0,1]} |\gamma(t) - [X]_t| \geqslant r\right\}$, which are, however, of particularly practical interest in statistics. The following proposition fills this gap.

PROPOSITION 1.5. *Given (X_t) by (0.1), let $Y_\cdot := \int_0^\cdot b(t,\omega)\,\mathrm{d}t$. We have for every $n \geqslant 1$ and $r > 0$,*

$$\mathbf{P}\left(\sup_{t\in[0,1]} [Q_t^n(X - Y) - \mathbf{E}Q_t^n(X - Y)] \geqslant r\right) \leqslant \exp\left(-n\Lambda^*([X]_1 + r)\right)$$
$$\leqslant \exp\left(-\frac{n}{2}\left[\frac{r}{\|\sigma^2\|_\infty} - \log\left(1 + \frac{r}{\|\sigma^2\|_\infty}\right)\right]\right) \tag{1.16}$$

$$\mathbf{P}\left(\inf_{t\in[0,1]} [Q_t^n(X - Y) - \mathbf{E}Q_t^n(X - Y)] \leqslant -r\right) \leqslant \exp\left(-n\Lambda^*([X]_1 - r)\right)$$
$$\leqslant \exp\left(-n\frac{r^2}{4\int_0^1 \sigma_t^4\,\mathrm{d}t}\right) \tag{1.17}$$

and in particular

$$\mathbf{P}\left(\sup_{t\in[0,1]} |Q_t^n(X - Y) - \mathbf{E}Q_t^n(X - Y)| \geqslant r\right)$$
$$\leqslant \exp\left\{-n(\Lambda^*([X]_1 + r) \wedge \Lambda^*([X]_1 - r))\right\}.$$

Moreover $\forall r > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\sup_{t\in[0,1]} \pm(Q_t^n(X - Y) - [X]_t) \geqslant r\right) = -\Lambda^*([X]_1 \pm r). \tag{1.18}$$

Remarks 1.10. The upper bounds in (1.16) and (1.17) hold for arbitrary n and r (not a limit relation, unlike in the previous results), hence they are much more practical (in statistics) and stronger than those given by the LDP in Theorem 1.2 for this special type of deviation domains. Moreover (1.18) means that the exponents in (1.16) and (1.17) are exact for great n.

2. Proof of Theorem 1.1.

For the convenience of the reader, we recall the following ([5, Th. 4.2.13, p114]):

APPROXIMATION LEMMA: *Let* $(Y^n, X^n, n \in \mathbf{N})$ *be a family of random variables valued in a Polish space S with metric* $d(\cdot,\cdot)$*, defined on a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*. Assume*

(i) $\mathbf{P}(Y^n \in \cdot)$ *satisfies as* $n \to +\infty$*, the LDP with speed* $\lambda(n)(\to +\infty)$ *and the good rate function* $I(x)$*;*
(ii) For every $\delta > 0$*,*

$$\limsup_{n\to+\infty} \frac{1}{\lambda(n)} \log \mathbf{P}\left(d(Y^n, X^n) > \delta\right) = -\infty. \tag{2.1}$$

Then $\mathbf{P}(X^n \in \cdot)$*, as* $n \to +\infty$*, satisfies the LDP on S with speed* $\lambda(n)$ *and rate function* $I(x)$.

LEMMA 2.1. *If a real r.v.* ξ *is of law* $N(0,1)$*,*

$$P(\lambda) := \log \mathbf{E} \exp(\lambda\xi^2) = \begin{cases} -\frac{1}{2}\log(1-2\lambda) & \text{if} \quad \lambda < \frac{1}{2} \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.2}$$

Proof. Elementary.

LEMMA 2.2. *Let* $Y_\cdot = \int_0^\cdot b(s,\omega)\,\mathrm{d}s$ *and define* $Q_t^n(X-Y)$ *as in (0.4). For every* $\lambda \in \mathbf{R}$*,*

$$\begin{aligned} \Lambda_n(\lambda) &:= \frac{1}{n}\log \mathbf{E}\exp\left(\lambda n Q_1^n(X-Y)\right) \\ &\leqslant \Lambda(\lambda) := \int_0^1 P(\lambda\sigma_t^2)\,\mathrm{d}t \\ &= \begin{cases} \displaystyle\int_0^1 -\frac{1}{2}\log\left(1-2\lambda\sigma_t^2\right)\,\mathrm{d}t, & \text{if} \quad \lambda \leqslant \dfrac{1}{2\|\sigma^2\|_\infty} \\ +\infty, & \textit{otherwise} \end{cases} \end{aligned} \tag{2.3}$$

and

$$\lim_{n\to\infty} \Lambda_n(\lambda) = \Lambda(\lambda). \tag{2.4}$$

Proof. As $P(\lambda\sigma_t^2) \geqslant -|\lambda|\sigma_t^2$, the integral $\int_0^1 P(\lambda\sigma_t^2)\,\mathrm{d}t$ is well defined with value in $(-\infty, +\infty]$. The last equality in (2.3) is obvious, because for $\lambda > \lambda_0 :=$

$\left(2\|\sigma^2\|_\infty\right)^{-1}$, $\left(t : P(\lambda\sigma_t^2) = +\infty\right)$ has positive Lebesgue measure. We prove now the first inequality in (2.3).

Since

$$Q_1^n(X - Y) = \sum_{k=1}^n \left(\int_{t_{k-1}^n}^{t_k^n} \sigma_s \, \mathrm{d}B_s \right)^2 = \sum_{k=1}^n \xi_k^2 a_k \tag{2.5a}$$

where

$$\xi_k := \frac{\int_{t_{k-1}^n}^{t_k^n} \sigma_s \, \mathrm{d}B_s}{\sqrt{a_k}}, \qquad a_k := \int_{t_{k-1}^n}^{t_k^n} \sigma_s^2 \, \mathrm{d}s. \tag{2.5b}$$

Obviously $(\xi_k)_{k=1,\dots,n}$ are independent and of law $N(0, 1)$. We get by Lemma 2.1

$$\Lambda_n(\lambda) = \frac{1}{n} \sum_{k=1}^n P(\lambda n a_k) = \int_0^1 P(\lambda f_n(t)) \, \mathrm{d}t, \tag{2.6a}$$

where

$$f_n(t) := \sum_{k=1}^n 1_{(t_{k-1}^n, t_k^n]}(t) \frac{\int_{t_{k-1}^n}^{t_k^n} \sigma_s^2 \, \mathrm{d}s}{t_k^n - t_{k-1}^n} = \mathbf{E}^{\mathrm{d}t}(\sigma_\cdot^2 | \mathcal{B}_{\tau_n})(t), \tag{2.6b}$$

is a $\mathrm{d}t$-martingale w.r.t. the partially directed filtration $(\mathcal{B}_{\tau_n} := \sigma((t_{k-1}^n, t_k^n]; k = 1, \dots, n))_n$.

By the convexity of $P(\lambda)$ and Jensen's inequality, we have

$$\int_0^1 P(\lambda f_n(t)) \, \mathrm{d}t \leqslant \int_0^1 \mathbf{E}^{\mathrm{d}t} \left(P(\lambda\sigma_\cdot^2) | \mathcal{B}_{\tau_n} \right)(t) \, \mathrm{d}t = \int_0^1 P(\lambda\sigma_t^2) \, \mathrm{d}t,$$

which implies (2.3).

On the other hand, by the partially ordered martingale convergence theorem (or the classical Lebesgue derivation theorem), $f_n(t) \longrightarrow \sigma_t^2$, $\mathrm{d}t - a.e.$ on $[0, 1]$. Consequently by the continuity of $P : \mathbf{R} \to (-\infty, +\infty]$,

$$P(\lambda f_n(t)) \longrightarrow P(\lambda\sigma_t^2), \quad \mathrm{d}t - a.e. \text{ on } [0, 1].$$

As $P(\lambda f_n(t)) \geqslant -|\lambda| \cdot \sigma_t^2 \in L^1([0, 1], \mathrm{d}t)$, we can apply Fatou's lemma to conclude

$$\liminf_{n\to\infty} \Lambda_n(\lambda) = \liminf_{n\to\infty} \int_0^1 P(\lambda f_n(t)) \, \mathrm{d}t \geqslant \int_0^1 \lim_{n\to\infty} P(\lambda f_n(t)) \, \mathrm{d}t = \int_0^1 P(\lambda\sigma_t^2) \, \mathrm{d}t,$$

the desired result (2.4). □

Note that the lemma above does not require the boundedness of $\sigma_\cdot$.

Proof of Theorem 1.1.

(a) It is contained in Lemma 2.2.
(b) We shall prove it by three steps.

STEP 1. (Reduction to the case where $b = 0$). By the contraction principle, $\mathbf{P}(Q_1^n(X) \in \cdot)$ and $\mathbf{P}(Q_1^n(X-Y) \in \cdot)$ satisfy the same LDP if and only if $\mathbf{P}(\sqrt{Q_1^n(X)} \in \cdot)$ and $\mathbf{P}(\sqrt{Q_1^n(X-Y)} \in \cdot)$ do. Since

$$\begin{aligned} &\left|\sqrt{Q_1^n(X)} - \sqrt{Q_1^n(X-Y)}\right| \\ &\qquad \leqslant \sqrt{Q_1^n(Y)} = \sqrt{\sum_{k=1}^n \left(\int_{t_{k-1}^n}^{t_k^n} b(t,\omega)\,\mathrm{d}t\right)^2} \leqslant \sqrt{n \times \frac{1}{n^2}} \cdot \|b\|_\infty \end{aligned}$$

by the approximation lemma, $\mathbf{P}(\sqrt{Q_1^n(X)} \in \cdot)$ satisfies the same LDP as $\mathbf{P}(\sqrt{Q_1^n(X-Y)} \in \cdot)$. Consequently $\mathbf{P}\left(Q_1^n(X) \in \cdot\right)$ and $\mathbf{P}\left(Q_1^n(X-Y) \in \cdot\right)$ satisfy the same LDP. Hence we can assume in the following that $b = 0$ and $\|\sigma\|_\infty > 0$ (trivial otherwise).

STEP 2. By the part (a) and the Ellis–Gärtner theorem [6, Th. II.6.1.], $\mathbf{P}\left(Q_1^n(X) \in \cdot\right)$ satisfies the upper bound of large deviations (i.e. the RHS inequality in (1.4)) with rate function $I = \Lambda^*$.

But for the lower bound, the same theorem (see e.g. [6, Lemma VII.4.2., (7.16)]) gives only

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}(Q_1^n(X) \in \mathbf{G}) \geqslant -\inf\left\{I(x);\, x \in \mathbf{G} \bigcap [0, A)\right\} \tag{2.7}$$

for any open subset $\mathbf{G} \subset \mathbf{R}$, where $A := \int_0^1 \frac{\sigma_t^2}{1-2\lambda_0\sigma_t^2}\,\mathrm{d}t$ $\left(\lambda_0 = \left(2\|\sigma^2\|\right)^{-1}\right)$. Hence if $A = +\infty$, or equivalently if $\Lambda(\lambda)$ is essentially smooth (or steep in the language of [5]), $\mathbf{P}\left(Q_1^n(X) \in \cdot\right)$ satisfies the lower bound in (1.4).

STEP 3. We turn now to the case where $A < +\infty$ and where the main difficulty of this theorem resides. Our first observation is for (1.4) it is enough to show that for any $r' > r > [X]_1$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(Q_1^n(X) \in (r, r')\right) \geqslant -I(r). \tag{2.8}$$

Indeed since $I(x)$ is increasing on $[\Lambda'(0) = [X]_1, \infty)$ and $[X]_1 < A$, combining (2.7) and (2.8) we get for any open $\mathbf{G}$ of $\mathbf{R}$,

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}(Q_1^n(X) \in \mathbf{G}) \\ &\qquad \geqslant \max\left\{-\inf\left\{I(x);\, x \in \mathbf{G} \bigcap [0, A)\right\},\right. \end{aligned}$$

$$
\begin{aligned}
&-\inf\left\{I(x);\, x\in \mathbf{G}\bigcap([X]_1,+\infty)\right\}\Big\}\\
&=-\inf\{I(x);\, x\in\mathbf{G}\}
\end{aligned}
$$

the desired lower bound (1.4). We show now (2.8).

For each $\varepsilon>0$ and $\varepsilon<\|\sigma\|_\infty$, let $\sigma_t^\varepsilon=\sigma_t\wedge(\|\sigma\|_\infty-\varepsilon)$, and $X_t^\varepsilon=\int_0^t\sigma_t^\varepsilon\,\mathrm{d}B_t$ (below we add the exponent ε to denote the objects associated with X^ε). Since

$$\left[t;\,\sigma_t^\varepsilon=\|\sigma^\varepsilon\|_\infty=(\|\sigma\|_\infty-\varepsilon)\right]$$

has positive Lebesgue measure,

$$A^\varepsilon:=\int_0^1\frac{(\sigma_t^\varepsilon)^2}{1-2\lambda_0^\varepsilon(\sigma_t^\varepsilon)^2}\,\mathrm{d}t=+\infty,\qquad \textit{where}\quad \lambda_0^\varepsilon=\left(2\|\sigma^\varepsilon\|_\infty^2\right)^{-1}.$$

Hence by what was shown above, $\mathbf{P}(Q_1^n(X^\varepsilon)\in\cdot)$ satifies the LDP with the rate function given by $I^\varepsilon(x)=(\Lambda^\varepsilon)^*(x)$.

Now by the decomposition (2.5a,b), for each $r\geqslant 0$,

$$
\begin{aligned}
\mathbf{P}(Q_1^n(X)>r) &= \mathbf{P}\left(\sum_{k=1}^n a_k(\xi_k)^2>r\right)\\
&\geqslant \mathbf{P}\left(\sum_{k=1}^n a_k^\varepsilon(\xi_k^\varepsilon)^2>r\right)=\mathbf{P}(Q_1^n(X^\epsilon)>r)
\end{aligned}
\tag{2.9}
$$

(because $0\leqslant a_k^\varepsilon\leqslant a_k$ and (ξ_k^ε) as well as (ξ_k) are i.i.d. of N(0,1)). Therefore, for any $\varepsilon>0$,

$$l(r):=\liminf_{n\to\infty}\frac{1}{n}\log\mathbf{P}(Q_1^n(X)>r)\geqslant-\inf_{x>r}I^\varepsilon(x).\tag{2.10}$$

Now as Λ^ε is essentially smooth (noted previously), $I^\varepsilon=(\Lambda^\varepsilon)^*$ is strictly convex on $(0,A^\varepsilon)=(0,+\infty)$, then I is strictly increasing and continuous for $r>\Lambda'(0)=[X]_1\geqslant[X^\varepsilon]_1=(\Lambda^\varepsilon)'(0)$. Thus

$$\inf_{x>r}I^\varepsilon(x)=I^\varepsilon(r)=\sup_{\lambda\geqslant 0}(\lambda r-\Lambda^\varepsilon(\lambda)).$$

But since $\Lambda^\varepsilon(\lambda)\uparrow\Lambda(\lambda)$ for all $\lambda\geqslant 0$ and $\Lambda^\varepsilon(\lambda)$, $\varepsilon>0$ are inf-compact for $\lambda\geqslant 0$, hence we can exchange the order below (an exercise in analysis)

$$\lim_{\varepsilon\downarrow 0}I^\varepsilon(r)=\inf_{\varepsilon>0}\sup_{\lambda\geqslant 0}(\lambda r-\Lambda^\varepsilon(\lambda))=\sup_{\lambda\geqslant 0}\inf_{\varepsilon>0}(\lambda r-\Lambda^\varepsilon(\lambda)).$$

This last quantity is exactly

$$\sup_{\lambda\geqslant 0}(\lambda r-\Lambda(\lambda))=\Lambda^*(r)=I(r).$$

So we get from (2.10) that $l(r) \geqslant -I(r), \quad \forall r > [X]_1$. Combining it with the already shown upper bound, we get $\forall r > [X]_1$,

$$\begin{aligned} l(r) &= \lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}(Q_1^n(X) > r) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}(Q_1^n(X) \geqslant r) = -I(r). \end{aligned} \tag{2.11}$$

Now for any $r' > r > [X]_1$, it is easy to see

$$\begin{aligned} l(r) &= \lim_{n\to\infty} \frac{1}{n} \log \left[\mathbf{P}(Q_1^n(X) \geqslant r') + \mathbf{P}(Q_1^n(X) \in (r, r'))\right] \\ &\leqslant \max\left\{ l(r') \; ; \; \liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(Q_1^n(X) \in (r, r')\right) \right\}. \end{aligned} \tag{2.12}$$

As $l(r') < l(r)$ by (2.11) and the strict increasement of I on $[\Lambda'(0), +\infty)$, we obtain (2.8). □

We apply now (2.9) to establish

COROLLARY 2.3. *If $\sigma \notin L^\infty([0, 1], \, \mathrm{d}t)$, then for any $r > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(Q_1^n(X) > r\right) = 0. \tag{2.13}$$

In other words, that $\sigma \in L^\infty([0, 1], \, \mathrm{d}t)$ is a necessary condition to the LDP in Theorem 1.1, as claimed in Remarks (1.2).

Proof. Notice that the lim sup as $n \to \infty$ in (2.13) is $\leqslant 0$ always. To control the lim inf, let $\sigma_t^\varepsilon = \sigma_t \wedge (1/\varepsilon)$ and define $X^\varepsilon, a_k^\varepsilon, \xi_k^\varepsilon, \Lambda^\varepsilon$ corresponding to this new σ^ε. The inequality (2.9) still holds and we get hence

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}(Q_1^n(X) > r) \geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}(Q_1^n(X^\epsilon) > r) \geqslant -\inf_{x>r} (\Lambda^\varepsilon)^*(x).$$

For $r < [X]_1$, (2.13) is trivial by (0.7). Now fix $r \geqslant [X]_1$. Hence $r \geqslant [X^\varepsilon]_1 = (\Lambda^\varepsilon)'(0)$ and

$$\inf_{x>r} (\Lambda^\varepsilon)^*(x) = (\Lambda^\varepsilon)^*(r) = \sup_{\lambda \geqslant 0} (\lambda r - \Lambda^\varepsilon(\lambda)).$$

As in the proof of Theorem 1.1, we have

$$\inf_{\varepsilon>0} \sup_{\lambda \geqslant 0} (\lambda r - \Lambda^\varepsilon(\lambda)) = \sup_{\lambda \geqslant 0} \inf_{\varepsilon>0} (\lambda r - \Lambda^\varepsilon(\lambda)) = \sup_{\lambda \geqslant 0} (\lambda r - \Lambda(\lambda)) = 0,$$

because $\Lambda(\lambda) = +\infty, \forall \lambda > 0$ (recalling $\|\sigma^2\|_\infty = \infty$ in (2.3) now). (2.13) follows. □

3. Proof of Theorem 1.2

Proof of Part (a). We separate its proof into three steps.

STEP 1. At first for any partition $\mathcal{P} = \{0 = s_0 < s_1 < \cdots < s_m = 1\}$, we shall prove that $\mathbf{P}\left(Q^n_{\mathcal{P}}(X) \in \cdot\right)$, with $Q^n_{\mathcal{P}}(X) := (Q^n_{s_k}(X))_{0 \leqslant k \leqslant m}$, satisfies the LDP on $\mathbf{R}^{\mathcal{P}}$ with speed n and with the rate function given by

$$I^{\mathcal{P}}(x_0, \dots, x_m) = \sup_{(\lambda_1, \dots, \lambda_m) \in \mathbf{R}^m} \left\{ \sum_{k=1}^m (x_k - x_{k-1})\lambda_k - \sum_{k=1}^m \int_{s_{k-1}}^{s_k} P(\lambda_k \sigma_t^2)\, \mathrm{d}t \right\} \tag{3.1}$$

if $x_0 = 0$ and $I^{\mathcal{P}}(x_0, \dots, x_m) = +\infty$ if $x_0 \neq 0$.

As in Step 1 of the proof of Theorem 1.1, we can and will assume that $b = 0$.

In the above case of LDP where $b = 0$, the key observation is that

$$\Delta_k^{\mathcal{P}} Q^n(X) := Q^n_{s_k}(X) - Q^n_{s_{k-1}}(X), \quad k = 1, \dots, m$$

are independent. And by Theorem 1.1, for each $k = 1, \dots, m$ fixed, $\mathbf{P}(\Delta_k^{\mathcal{P}} Q^n(X) \in \cdot)$ satisfies the LDP with speed n and the rate function

$$I_k^{\mathcal{P}}(y_k) = \sup_{\lambda \in \mathbf{R}} \left\{ y_k \lambda - \int_{s_{k-1}}^{s_k} P(\lambda \sigma_t^2)\, \mathrm{d}t \right\}. \tag{3.2}$$

Consequently by [10, Corollary 2.9], $\mathbf{P}\left(\Delta^{\mathcal{P}} Q^n(X) := (\Delta_k^{\mathcal{P}} Q^n(X))_{0 \leqslant k \leqslant m} \in \cdot\right)$ satisfies a LDP with the same speed and with the rate function given by

$$I^{\mathcal{P}}(y_1, \dots, y_m) = \sum_{k=1}^m I_k^{\mathcal{P}}(y_k) = \sup_{(\lambda_1, \dots, \lambda_m) \in \mathbf{R}^m} \left\{ \sum_{k=1}^m y_k \lambda_k - \sum_{k=1}^m \int_{s_{k-1}}^{s_k} P(\lambda_k \sigma_t^2)\, \mathrm{d}t \right\}.$$

Finally the desired LDP in this step follows by the contraction principle.

STEP 2. By [13, Th.I.5.2], the finite dimensional LDP in Step 1 implies that $\mathbf{P}\left(Q^n_{\cdot}(X) \in \cdot\right)$ satisfies the LDP on the topological product space $\mathbf{R}^{[0,1]}$ with the speed n and with the rate function given by

$$I^{\infty}(\gamma) = \sup_{\mathcal{P}} I^{\mathcal{P}}(\gamma(\mathcal{P})), \tag{3.3}$$

where the supremum are taken over all finite partitions $\mathcal{P} = \{0 = s_0 < s_1 < \cdots < s_m = 1\}$ of [0, 1]. And for each such $\mathcal{P}$,

$$I^{\mathcal{P}}(\gamma(\mathcal{P})) = \sum_{k=1}^m \sup_{\lambda_k \in \mathbf{R}} \left\{ \lambda_k(\gamma(s_k) - \gamma(s_{k-1})) - \int_{s_{k-1}}^{s_k} P(\lambda_k \sigma_t^2)\, \mathrm{d}t \right\}. \tag{3.4}$$

by (3.1). Remark that

$$I^{\mathcal{P}}(\gamma(\mathcal{P})) = +\infty, \quad \text{if } \gamma(s_k) \leqslant \gamma(s_{k-1}).$$

Since the space $\mathcal{I}_0[0,1]$ of the non-decreasing functions γ on [0,1] with $\gamma(0) = 0$ equipped with the pointwise convergence topology is a closed subset of $\mathbf{R}^{[0,1]}$, hence $\mathbf{P}\left(Q^n_\cdot(X) \in \cdot\right)$ satisfies the LDP on $\mathcal{I}_0[0,1]$ with the rate function given by (3.3).

STEP 3. For each non-decreasing function $\gamma \in \mathcal{I}_0[0,1]$, let $\tilde{\gamma}(\cdot) = \gamma(\cdot+)$ be its right continuous version. We claim that the mapping $\gamma \longrightarrow \tilde{\gamma}$ from $\mathcal{I}_0[0,1]$ to $\mathbf{D}^+[0,1]$ equipped with the weak convergence topology is continuous. To this end, it is enough to establish that for all $\gamma_0 \in \mathcal{I}_0[0,1]$ fixed, $\forall \varepsilon > 0$, $\forall \mathcal{P} = \{0 = s_0 < s_1 < \cdots < s_m = 1\}$ where s_i, $i = 1, \ldots, m-1$ are continuous points of γ_0, there exists a neighborhood $N(\gamma_0)$ de γ_0 dans $\mathcal{I}_0[0,1]$, such that

$$\forall \gamma \in N(\gamma_0), \quad \forall i = 1, \ldots, m \ : \qquad |\tilde{\gamma}(s_i) - \tilde{\gamma}_0(s_i)| < \varepsilon. \tag{3.5}$$

(It is left to the reader to see why the point $s_0 = 0$ is excluded in (3.5).) In fact, for $i = 1, \ldots, m-1$, since γ_0 is continuous at s_i, we can choose $s_{i-1} < a_i < s_i < b_i < s_{i+1}$ so that

$$\gamma_0(b_i) - \gamma_0(a_i) < \varepsilon/2 \qquad i = 1, \ldots, m-1.$$

Observe that $\tilde{\gamma}(1) = \gamma(1)$ for all $\gamma \in \mathcal{I}_0[0,1]$ and for each $i = 1, \ldots, m-1$

$$\tilde{\gamma}(s_i) - \tilde{\gamma}_0(s_i) \leqslant \gamma(b_i) - \gamma_0(a_i) \leqslant \gamma(b_i) - \gamma_0(b_i) + \frac{\epsilon}{2};$$

$$\tilde{\gamma}(s_i) - \tilde{\gamma}_0(s_i) \geqslant \gamma(a_i) - \gamma_0(b_i) \geqslant \gamma(a_i) - \gamma_0(a_i) - \frac{\epsilon}{2}.$$

We can thus give an expression of $N(\gamma_0)$ which satisfies (3.5)

$$\begin{aligned} N(\gamma_0) \ = \ &\{\gamma;\ |\gamma(a_i) - \gamma_0(a_i)| < \varepsilon/2, |\gamma(b_i) - \\ &-\gamma_0(b_i)| < \varepsilon/2, |\gamma(1) - \gamma_0(1)| < \varepsilon\}. \end{aligned}$$

The continuity is so shown.

Finally by the contraction principle, $\mathbf{P}\left(Q^n_\cdot(X) \in \cdot\right)$ satisfies the LDP on $\mathbf{D}^+[0,1]$ with the speed n and with the rate function given by

$$\tilde{I}^\infty(\gamma) = \inf\left\{I^\infty(f);\ f \in \mathcal{I}_0[0,1] \text{ and } \tilde{f} = \gamma\right\}, \tag{3.6}$$

which is inf-compact and convex. That is the claim of Theorem 1.2.(a).

Proof of Part (b). This part is rather delicate. Let $J(\gamma)$ be the RHS of (1.7). We should establish $J(\gamma) = \tilde{I}^\infty(\gamma)$ defined by (3.6).

For any finite partition $\mathcal{P} = \{0 = s_0 < s_1 < \cdots < s_m = 1\}$ of [0,1], we write

$$\mathcal{B}_\mathcal{P} := \sigma\left([0, s_1],\ (s_{k-1}, s_k],\ 2 \leqslant k \leqslant m\right), \quad \Delta\mathcal{P} := \max_{1 \leqslant k \leqslant m} |s_k - s_{k-1}|.$$

Proof of $\tilde{I}^\infty(\gamma) \leqslant J(\gamma)$. It is enough to show that for any $\gamma \in \mathbf{D}^+[0,1]$,

$$\limsup_{l\to\infty} I^{\mathcal{P}^l}(\gamma^0(\mathcal{P}^l)) \leqslant J(\gamma), \tag{3.7}$$

for any increasing sequence of finite partitions $(\mathcal{P}^l)$ such that $\Delta\mathcal{P}^l \to 0$, where $\gamma^0(0) = 0$, $\gamma^0(t) = \gamma(t)$, $\forall t \in (0,1]$.

To prove (3.7), we can assume $J(\gamma) < +\infty$ (otherwise it is trivial). Since $P^*(0) = +\infty$, the finiteness of $J(\gamma)$ implies $d[X] \ll \mathrm{d}\gamma$ by the expression (1.7).

From (3.4) and the following consequence of Jensen's inequality

$$\frac{1}{s_k - s_{k-1}} \int_{s_{k-1}}^{s_k} P(\lambda_k \sigma_t^2)\,\mathrm{d}t \geqslant P(\lambda_k c_k), \quad \text{where} \quad c_k := \frac{\int_{s_{k-1}}^{s_k} \sigma_t^2\,\mathrm{d}t}{s_k - s_{k-1}}$$

we get

$$\begin{aligned} I^{\mathcal{P}}(\gamma^0(\mathcal{P})) &= \sum_{k=1}^m \sup_{\lambda_k\in\mathbf{R}} \left\{\lambda_k(\gamma^0(s_k) - \gamma^0(s_{k-1})) - \int_{s_{k-1}}^{s_k} P(\lambda_k\sigma_t^2)\,\mathrm{d}t\right\} \\ &\leqslant \sum_{k=1}^m \sup_{\lambda_k\in\mathbf{R}} \left\{\lambda_k \frac{\gamma^0(s_k) - \gamma^0(s_{k-1})}{s_k - s_{k-1}} - P(\lambda_k c_k)\right\}(s_k - s_{k-1}) \\ &= \sum_{k=1}^m (s_k - s_{k-1}) 1_{[c_k>0]} P^*\left(\frac{\gamma^0(s_k) - \gamma^0(s_{k-1})}{(s_k - s_{k-1})c_k}\right). \end{aligned} \tag{3.8}$$

Now introduce

$$g(x) := xP^*(1/x), \quad \text{for} \quad x > 0 \quad \text{and} \quad g(0) := \lim_{x\to 0+} g(x) \tag{3.9}$$

where $P^*(\lambda)$ is given by (1.8). Obviously $g(0) = \frac{1}{2}$ and g is convex.

The key remark is (1.7) can be rewritten as

$$J(\gamma) = \int_0^1 g\left(\frac{\mathrm{d}[X]}{\mathrm{d}\gamma}\right)\frac{1}{\sigma_t^2}\mathrm{d}\gamma \quad \text{if} \quad \mathrm{d}[X] \ll \mathrm{d}\gamma \quad \text{and} \quad +\infty \quad \text{else} \tag{3.10}$$

and (3.8) can be rewritten similarly as

$$I^{\mathcal{P}}(\gamma^0(\mathcal{P})) \leqslant \int_{[0,1]} g\left(\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}\gamma_{\mathcal{P}}}\right)\cdot\frac{\mathrm{d}t_{\mathcal{P}}}{\mathrm{d}[X]_{\mathcal{P}}}\mathrm{d}\gamma. \tag{3.11}$$

(in fact the RHS above coincides with the last line of (3.8)) where $\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}\gamma_{\mathcal{P}}}$, $\frac{\mathrm{d}t_{\mathcal{P}}}{\mathrm{d}[X]_{\mathcal{P}}}$ are Radon–Nykodym densities restricted to $\mathcal{B}_{\mathcal{P}}$. By (3.11) and the following consequence of Jensen's inequality

$$g\left(\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}\gamma_{\mathcal{P}}}\right) \leqslant \mathbf{E}^{\mathrm{d}\gamma}\left(g\left(\frac{\mathrm{d}[X]}{\mathrm{d}\gamma}\right) \mid \mathcal{B}_{\mathcal{P}}\right),$$

we get

$$\limsup_{\mathcal{P}} I^{\mathcal{P}}(\gamma^0(\mathcal{P})) \leqslant \limsup_{\mathcal{P}} \int_0^1 g\left(\frac{\mathrm{d}[X]}{\mathrm{d}\gamma}\right)\left(\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}t_{\mathcal{P}}}\right)^{-1} \mathrm{d}\gamma \leqslant J(\gamma),$$

where the last inequality follows from the dominated convergence and the fact that

$$\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}t_{\mathcal{P}}} \geqslant \inf_{t\in[0,1]} \sigma_t^2 > 0$$

(consequence of our assumption about σ_t^2).

Proof of $J(\gamma) \leqslant \tilde{I}^{\infty}(\gamma)$. To this purpose it is enough to show that for any increasing function $\gamma \in \mathcal{I}_0[0,1]$, there is an increasing sequence $(\mathcal{P}^l)$ of partitions of $[0,1]$ such that

$$\liminf_{l\to\infty} I^{\mathcal{P}^l}(\gamma(\mathcal{P}^l)) \geqslant J(\tilde{\gamma}) \tag{3.12}$$

(because its left hand side (in short : LHS) is smaller than $I^{\infty}(\gamma)$). Actually we choose an (arbitrary) increasing sequence $(\mathcal{P}^l)$ of finite partitions of $[0,1]$ composed only of *continuous points* of γ except 0 and 1 such that $\Delta\mathcal{P}^l \to 0$.

For $\mathcal{P} = \mathcal{P}^l = \{0 = s_0 < s_1 < \cdots < s_m = 1\}$, using the fact that

$$\begin{aligned}&\sup_{\lambda\in\mathbf{R}}\left\{\lambda\,(\gamma(s_k)-\gamma(s_{k-1})) - \int_{s_{k-1}}^{s_k} P(\lambda\sigma_t^2)\,\mathrm{d}t\right\}\\ &\quad= \sup_{\pm\lambda\geqslant 0}\left\{\lambda\,(\gamma(s_k)-\gamma(s_{k-1})) - \int_{s_{k-1}}^{s_k} P(\lambda\sigma_t^2)\,\mathrm{d}t\right\}\end{aligned}$$

according to

$$\pm(\gamma(s_k)-\gamma(s_{k-1})) \geqslant \pm\left(\int_{s_{k-1}}^{s_k} P(\lambda\sigma_t^2)\,\mathrm{d}t\right)'_{\lambda=0} = \pm\int_{s_{k-1}}^{s_k} \sigma_t^2\,\mathrm{d}t,$$

we have from the first line of (3.8)

$$\begin{aligned}I^{\mathcal{P}}(\gamma(\mathcal{P})) \;&\geqslant\; \sum_{k=1}^m \sup_{\lambda\geqslant 0}\left\{\lambda\,(\gamma(s_k)-\gamma(s_{k-1})) - \int_{s_{k-1}}^{s_k} P(\lambda(\bar{\sigma}_k^{\mathcal{P}})^2)\,\mathrm{d}t\right\} 1_{\left[\frac{\gamma(s_k)-\gamma(s_{k-1})}{\int_{s_{k-1}}^{s_k}\sigma_t^2\,\mathrm{d}t}>1\right]} +\\ &\quad+ \sum_{k=1}^m \sup_{\lambda\leqslant 0}\left\{\lambda\,(\gamma(s_k)-\gamma(s_{k-1})) - \int_{s_{k-1}}^{s_k} P(\lambda(\underline{\sigma}_k^{\mathcal{P}})^2)\,\mathrm{d}t\right\} 1_{\left[\frac{\gamma(s_k)-\gamma(s_{k-1})}{\int_{s_{k-1}}^{s_k}\sigma_t^2\,\mathrm{d}t}<1\right]}\\ &=\; \sum_{k=1}^m (s_k-s_{k-1})P^*\left(\frac{\gamma(s_k)-\gamma(s_{k-1})}{(\bar{\sigma}_k^{\mathcal{P}})^2(s_k-s_{k-1})}\right) 1_{\left[\frac{[X]_{s_k}-[X]_{s_{k-1}}}{\gamma(s_k)-\gamma(s_{k-1})}<1\right]} +\\ &\quad+ \sum_{k=1}^m (s_k-s_{k-1})P^*\left(\frac{\gamma(s_k)-\gamma(s_{k-1})}{(\underline{\sigma}_k^{\mathcal{P}})^2(s_k-s_{k-1})}\right) 1_{\left[\frac{[X]_{s_k}-[X]_{s_{k-1}}}{\gamma(s_k)-\gamma(s_{k-1})}>1\right]}\\ &:=\; \int_0^1 g^{\mathcal{P}}(t)\,\mathrm{d}\tilde{\gamma}(t)\end{aligned} \tag{3.13a}$$

where

$$\bar{\sigma}_k^{\mathcal{P}} := \sup_{s\in(s_{k-1},s_k)} \sigma_s, \quad \underline{\sigma}_k^{\mathcal{P}} := \inf_{s\in(s_{k-1},s_k)} \sigma_s$$

and

$$\begin{aligned} g^{\mathcal{P}}(t) &= \sum_{k=1}^{m} 1_{(s_{k-1},s_k]}(t)\frac{1}{(\bar{\sigma}_k^{\mathcal{P}})^2} g\left(\frac{\left(\bar{\sigma}_k^{\mathcal{P}}\right)^2(s_k - s_{k-1})}{\gamma(s_k) - \gamma(s_{k-1})}\right) 1_{\left[\frac{[X]_{s_k}-[X]_{s_{k-1}}}{\gamma(s_k)-\gamma(s_{k-1})}<1\right]} + \\ &\quad + \sum_{k=1}^{m} 1_{(s_{k-1},s_k]}(t)\frac{1}{(\underline{\sigma}_k^{\mathcal{P}})^2} g\left(\frac{(\underline{\sigma}_k^{\mathcal{P}})^2(s_k - s_{k-1})}{\gamma(s_k) - \gamma(s_{k-1})}\right) 1_{\left[\frac{[X]_{s_k}-[X]_{s_{k-1}}}{\gamma(s_k)-\gamma(s_{k-1})}>1\right]}, \end{aligned} \tag{3.13b}$$

where $g(x)$ is given by (3.9).

Now if $\mathrm{d}[X] \ll \mathrm{d}\tilde{\gamma}$, by the continuity of $t \to \sigma_t^2$ and the martingale convergence, we have

$$\begin{aligned} \liminf_{l\to\infty} g^{\mathcal{P}^l}(t) &\geqslant \frac{1}{\sigma_t^2} g\left(\frac{\mathrm{d}[X]}{\mathrm{d}\tilde{\gamma}}\right)\left[1_{\left[\frac{\mathrm{d}[X]}{\mathrm{d}\tilde{\gamma}}<1\right]} + 1_{\left[\frac{\mathrm{d}[X]}{\mathrm{d}\tilde{\gamma}}>1\right]}\right] \\ &= \frac{1}{\sigma_t^2} g\left(\frac{\mathrm{d}[X]}{\mathrm{d}\tilde{\gamma}}\right), \quad \mathrm{d}\tilde{\gamma} - a.e. \end{aligned}$$

(since $g(1) = 0$). Consequently Fatou's lemma implies

$$\liminf_{l\to\infty} I^{\mathcal{P}^l}(\gamma(\mathcal{P}^l)) \geqslant \int_0^1 \frac{1}{\sigma_t^2} g\left(\frac{\mathrm{d}[X]}{\mathrm{d}\tilde{\gamma}}\right) \mathrm{d}\tilde{\gamma}$$

which is exactly $J(\tilde{\gamma})$ by (3.10).

It remains to treat the case where $\mathrm{d}[X]$ is not absolutely continuous w.r.t. $\mathrm{d}\tilde{\gamma}$. By (3.10), we have to show that $\lim_{l\to\infty} I^{\mathcal{P}^l}(\gamma(\mathcal{P}^l)) = +\infty$. By absurd, assume in contrary that

$$\lim_{l\to\infty} I^{\mathcal{P}^l}(\gamma(\mathcal{P}^l)) = \sup_{l\geqslant 1} I^{\mathcal{P}^l}(\gamma(\mathcal{P}^l)) < +\infty.$$

For $\mathcal{P} = \mathcal{P}^l$, since $g(x)$ is increasing for $x \geqslant 1$, we get by (3.13a,b),

$$I^{\mathcal{P}}(\gamma(\mathcal{P})) \geqslant \frac{1}{\|\sigma^2\|_\infty} \int_0^1 1_{\left[\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}\tilde{\gamma}_{\mathcal{P}}}>1\right]} g\left(\frac{\mathrm{d}[X]_{\mathcal{P}}}{\mathrm{d}\tilde{\gamma}_{\mathcal{P}}}\right) \mathrm{d}\tilde{\gamma}.$$

Because $\lim_{x\to+\infty} g(x)/x = \lim_{x\to+\infty} P^*(1/x) = +\infty$, this implies that the $\mathrm{d}\tilde{\gamma}$-martingale

$$\left(\frac{\mathrm{d}[X]_{\mathcal{P}^l}}{\mathrm{d}\tilde{\gamma}_{\mathcal{P}^l}}\right)_{l\geqslant 1}$$

is $\mathrm{d}\tilde{\gamma}$-uniformly integrable. Consequently $\mathrm{d}[X] \ll \mathrm{d}\tilde{\gamma}$, which is in contradiction with our assumption.

The proof of Theorem 1.2. is thus completed. □

4. Proof of Theorem 1.3

The reader can read at first Lemma A.1 in Appendix before the proof below.

Proof. We treat here only the case $b = 0$, and one can prove easily that $Q^n_\cdot(X)$ and $Q^n_\cdot(X - Y)$ satisfy the same MDP by following the proof of Theorem 1.4. in the next section. We separate its proof into five steps. The first three steps consist to show the finite dimensional LDP in condition (i) of Lemma A.1 in the appendix, the fourth step consists to prove condition (ii) of Lemma A.1, and in the last step we identify the rate function.

(1) We check at first why (1.11) implies (1.10). In fact, by Hölder's inequality,

$$\max_{1\leqslant k\leqslant n}\int_{(k-1)/n}^{k/n}\sigma_t^2\,\mathrm{d}t \leqslant n^{\frac{1}{p}-1}\max_{1\leqslant k\leqslant n}\left(\int_{(k-1)/n}^{k/n}\sigma_t^{2p}\,\mathrm{d}t\right)^{1/p}.$$

And by the condition that $\sigma^2 \in L^p$ in (1.11), $\max_{1\leqslant k\leqslant n}\int_{(k-1)/n}^{k/n}\sigma_t^{2p}\,\mathrm{d}t \longrightarrow 0$. Consequently (1.10) follows by the second condition in (1.11).

(2) We shall establish that under the condition (1.10), for each $\lambda \in \mathbf{R}$,

$$\lim_{n\to\infty}\frac{1}{b^2(n)}\log\mathbf{E}\exp\left(\lambda b^2(n)\cdot\frac{\sqrt{n}}{b(n)}\left(Q_1^n(X)-[X]_1\right)\right)=\lambda^2\int_0^1\sigma_t^4\,\mathrm{d}t. \tag{4.1}$$

Since the Legendre transformation of the RHS of (4.1) is exactly $I_m(x)$ given in (1.14), by Ellis–Gärtner theorem, (4.1) implies the last LDP in Theorem 1.3.

Taking the calculations in Lemma 2.2 (and the notations there), we have

$$\begin{aligned} G_n(\lambda) :=\ & \frac{1}{b^2(n)}\log\mathbf{E}\exp\left(\lambda b^2(n)\cdot\frac{\sqrt{n}}{b(n)}\left(Q_1^n(X)-[X]_1\right)\right) \\ =\ & \frac{1}{b^2(n)}\sum_{k=1}^n\left(P\left(\lambda b(n)\sqrt{n}a_k\right)-\lambda b(n)\sqrt{n}[X]_1\right), \end{aligned}\tag{4.2}$$

where

$$a_k = \int_{t_{k-1}^n}^{t_k^n}\sigma_t^2\mathrm{d}t,\quad k=1,\ldots,n.$$

By our condition (1.10),

$$\varepsilon(n) := b(n)\sqrt{n}\max_{1\leqslant k\leqslant n}a_k \longrightarrow 0.$$

By Taylor formula and noting that $P(0)=0,\ P'(0),\ P''(0)=2$, we obtain once if $|\lambda\varepsilon(n)|<\frac{1}{4}$,

$$P(\lambda b(n)\sqrt{n}a_k)=\lambda b(n)\sqrt{n}a_k+(1+\eta(k,n))\cdot\lambda^2b^2(n)na_k^2, \tag{4.3a}$$

where $\eta(k,n)$ satisfies

$$|\eta(k,n)|\leqslant C|\lambda|\varepsilon(n), \tag{4.3b}$$

where $C=\frac{1}{2}\sup_{|\lambda|\leqslant 1/4}|P'''(\lambda)|$.

Substituting now (4.3a) into (4.2) and noting that $\sum_k a_k = [X]_1$, we get

$$\begin{aligned} G_n(\lambda) &= \frac{1}{b^2(n)}\left[\sum_{k=1}^{n}(1+\eta(k,n))\lambda^2 b^2(n) n a_k^2\right] \\ &= \lambda^2\sum_{k=1}^{n}(1+\eta(k,n))\left(\frac{\int_{t_{k-1}^n}^{t_k}\sigma_t^2\,\mathrm{d}t}{t_k^n-t_{k-1}^n}\right)^2 (t_k^n - t_{k-1}^n). \end{aligned} \tag{4.4}$$

Now by (4.3b), to prove (4.1), it is enough to show

$$\sum_{k=1}^{n}\left(\frac{\int_{t_{k-1}^n}^{t_k^n}\sigma_t^2\,\mathrm{d}t}{t_k^n-t_{k-1}^n}\right)^2 (t_k^n - t_{k-1}^n) \longrightarrow \int_0^1 \sigma_t^4\,\mathrm{d}t. \tag{4.5}$$

It follows easily from the L^2-martingale convergence (the detail is omitted).

(3) By the independence of increments of $t \to Q_t^n(X)$, we can apply [10, Corollary 2.9] as in the proof of Theorem 1.2.(a) to get the LDP of

$$\mathbf{P}\left(\frac{\sqrt{n}}{b(n)}\left[Q_{\mathcal{P}}^n(X) - \mathbf{E}Q_{\mathcal{P}}^n(X)\right] \in \cdot\right)$$

on $\mathbf{R}^{\mathcal{P}}$ with speed $b^2(n)$ and with the rate function given by

$$I_m^{\mathcal{P}}(\gamma(\mathcal{P})) = \sum_{k=1}^{m}\frac{(\gamma(s_k)-\gamma(s_{k-1}))^2}{4\int_{s_{k-1}}^{s_k}\sigma_t^4\,\mathrm{d}t}, \quad \text{if } \gamma(0)=0 \text{ and } +\infty \text{ else} \tag{4.6}$$

where $\mathcal{P} = \{0 = s_0 < s_1 < \cdots < s_m\}$ is an arbitrary partition of [0, 1].

(4) In this step, we shall establish for any $\delta > 0$,

$$\begin{aligned} &\sup_{0\leqslant s\leqslant 1}\limsup_{n\to\infty}\frac{1}{b^2(n)}\times \\ &\quad\times\log\mathbf{P}\left(\sup_{s\leqslant t\leqslant s+\varepsilon}\frac{\sqrt{n}}{b(n)}\left|\Delta_s^t\left(Q_\cdot^n(X) - \mathbf{E}Q_\cdot^n(X)\right)\right| > \delta\right) \longrightarrow -\infty \end{aligned} \tag{4.7}$$

as $\varepsilon \to 0+$, where $\Delta_s^t Y_\cdot := Y_t - Y_s$ and $Q_t^n(X) := Q_1^n(X)$ for $t > 1$. This estimation implies condition (ii) in Lemma A.1, then the LDP of $\mathbf{P}\Big(\frac{\sqrt{n}}{b(n)}(Q_\cdot^n(X) - \mathbf{E}Q_\cdot^n(X)) \in \cdot\Big)$ on $\mathbf{D}_0([0,1])$ w.r.t. the sup norm topology, with the speed $b^2(n)$ and with the rate function given by

$$I_m^\infty(\gamma) = \sup_{\mathcal{P}} I_m^{\mathcal{P}}(\gamma(\mathcal{P})), \quad \forall\gamma\in\mathbf{D}_0([0,1]), \tag{4.8}$$

where the supremum is taken over all finite partitions $\mathcal{P}$ of [0, 1].

Since

$$\sqrt{n}\sup_{t\in[0,1]}|\mathbf{E}Q_t^n(X)-[X]_t|\leqslant\sqrt{n}\max_{k\leqslant n}\int_{(k-1)/n}^{k/n}\sigma_s^2\,\mathrm{d}s$$
$$\leqslant\max_{k\leqslant n}\sqrt{\int_{(k-1)/n}^{k/n}\sigma_s^4\,\mathrm{d}s}\to 0,$$

(by Cauchy–Schwarz), the LDP above still holds with $\mathbf{E}Q_t^n(X)$ substituted by $[X]_.$.

We turn now to show (4.7). Remark that $(Q_t^n(X)-\mathbf{E}Q_t^n(X))$ is a $(\mathcal{F}_{[nt]/n})$-martingale. Then

$$\exp\left(\lambda\left[Q_t^n(X)-\mathbf{E}Q_t^n(X)\right]\right)$$

is a sub-martingale. Writing $\Delta_s^t M=M_t-M_s$, by the maximal inequality, we have for any $r,\lambda>0$,

$$\mathbf{P}\left(\sup_{s\leqslant t\leqslant s+\varepsilon}[\Delta_s^t\left(Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)\right)]>r\right)$$
$$=\mathbf{P}\left(\exp\left(\lambda\sup_{s\leqslant t\leqslant s+\varepsilon}\Delta_s^t[Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)]\right)>e^{\lambda r}\right)$$
$$\leqslant e^{-\lambda r}\mathbf{E}\exp\left(\lambda\Delta_s^{s+\varepsilon}\left[Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)\right]\right)\tag{4.9a}$$

and similarly

$$\mathbf{P}\left(\inf_{s\leqslant t\leqslant s+\varepsilon}[\Delta_s^t\left(Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)\right)]<-r\right)$$
$$\leqslant e^{-\lambda r}\mathbf{E}\exp\left(-\lambda\Delta_s^{s+\varepsilon}\left[Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)\right]\right).\tag{4.9b}$$

By Step 2,

$$\lim_{n\to\infty}\frac{1}{b^2(n)}\log\mathbf{E}\exp\left(cb^2(n)\frac{\sqrt{n}}{b(n)}\Delta_s^{s+\varepsilon}\left[Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)\right]\right)$$
$$=c^2\int_s^{s+\varepsilon}\sigma_t^4\,\mathrm{d}t,\qquad\forall c\in\mathbf{R}$$

Therefore taking $r=\delta\frac{b(n)}{\sqrt{n}}$, $\lambda=b(n)\sqrt{n}c$ ($c>0$) in (4.9a), we get

$$\limsup_{n\to\infty}\frac{1}{b^2(n)}\log\mathbf{P}\left(\frac{\sqrt{n}}{b(n)}\sup_{s\leqslant t\leqslant s+\varepsilon}\Delta_s^t[Q_\cdot^n(X)-\mathbf{E}Q_\cdot^n(X)]>\delta\right)$$
$$\leqslant\inf_{c>0}\left\{-c\delta+c^2\int_s^{s+\varepsilon}\sigma_t^4\,\mathrm{d}t\right\}=-\frac{\delta^2}{4\int_s^{s+\varepsilon}\sigma_t^4\,\mathrm{d}t}$$

and similarly by (4.9b),

$$\limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\frac{\sqrt{n}}{b(n)} \inf_{s\leqslant t\leqslant s+\varepsilon} \Delta_s^t[Q^n_\cdot(X) - \mathbf{E}Q^n_\cdot(X)] < -\delta\right)$$
$$\leqslant -\frac{\delta^2}{4\int_s^{s+\varepsilon}\sigma_t^4\,\mathrm{d}t}.$$

(we have used the convention that $\sigma_t = 0$ for $t > 1$). By the integrability of σ^4, we have

$$\lim_{\varepsilon\to 0}\sup_{s\in[0,1]}\int_s^{s+\varepsilon}\sigma_t^4\,\mathrm{d}t = 0.$$

Hence (4.7) follows from the above estimations.

(5) It remains to show that $I_m^\infty(\gamma)$ defined in (4.8) coincides with $J_m(\gamma)$ given by (1.13). To this end, it is enough to show that

$$\lim_{l\to+\infty} I_m^{\mathcal{P}^l}(\gamma(\mathcal{P}^l)) = J_m(\gamma), \tag{4.10}$$

for any increasing sequence $(\mathcal{P}^l)_{l\geqslant 1}$ of partitions of [0, 1] such that $\Delta\mathcal{P}^l \to 0$.

Let $\mu(t) := \int_0^t \sigma_s^4\,\mathrm{d}s$. We can rewrite (4.6) as

$$\begin{aligned} 4I^{\mathcal{P}}(\gamma(\mathcal{P})) &= \sum_{k=1}^m \left(\frac{\gamma(s_k)-\gamma(s_{k-1})}{\mu(s_k)-\mu(s_{k-1})}\right)^2(\mu(s_k)-\mu(s_{k-1})) \\ &= \mathbf{E}^{\mathrm{d}\mu}\left(\frac{\mathrm{d}\gamma_{\mathcal{P}}}{\mathrm{d}\mu_{\mathcal{P}}}\right)^2. \end{aligned} \tag{4.11}$$

If $J_m(\gamma) < +\infty$, then $\mathrm{d}\gamma \ll \mathrm{d}\mu$ and $J_m(\gamma) = \frac{1}{4}\mathbf{E}^{\mathrm{d}\mu}\left(\frac{\mathrm{d}\gamma}{\mathrm{d}\mu}\right)^2$. Hence

$$M_l := \frac{\mathrm{d}\gamma_{\mathcal{P}^l}}{\mathrm{d}\mu_{\mathcal{P}^l}} = \mathbf{E}^{\mathrm{d}\mu}\left(\frac{\mathrm{d}\gamma}{\mathrm{d}\mu}|\mathcal{B}_{\mathcal{P}^l}\right).$$

Then (4.10) follows from the L^2-martingale convergence.

Inversely if the LHS of (4.10) is finite, by the same Doob's theorem,

$$M_l \longrightarrow M_\infty, \quad \text{in} \quad L^2([0,1],\ \mathrm{d}\mu)$$

then in $L^1(\mathrm{d}\mu)$ too. This implies that $\mathrm{d}\gamma \ll \mathrm{d}\mu$ and $\mathrm{d}\gamma = M_\infty\,\mathrm{d}\mu$. Therefore (4.10) follows. □

5. Proof of Theorem 1.4

We shall prove that $\tilde{Q}^n(X)$ and $Q^n(X-Y)$ satisfy the same LDPs and MDP, by means of the approximation lemma. By the elementary inequality $|(a+b)^2 - a^2| \leqslant \varepsilon a^2 + (1+1/\varepsilon)b^2$ where $\varepsilon > 0, a, b \in \mathbf{R}$, we get

$$\sup_{t\in[0,1]} |\tilde{Q}_t^n(X) - Q_t^n(X-Y)| \leqslant \varepsilon(n) Q_1^n(X-Y) + \left(1 + \frac{1}{\varepsilon(n)}\right) Z_n, \tag{5.1}$$

where $\epsilon(n) > 0$ will be chosen later, and

$$Z_n := \sum_{k=1}^n \left[\int_{t_{k-1}^n}^{t_k^n} b(t, X_t)\, \mathrm{d}t - b(t_{k-1}^n, X_{t_{k-1}^n})(t_k^n - t_{k-1}^n) \right]^2 .$$

We have to control the RHS of (5.1). As $Q_1^n(X-Y)$ have been well estimated by Theorem 1.1 and 1.3, it remains to control Z_n. The main idea is to reduce it to the estimations of $X_\cdot^0 = \int_0^\cdot \sigma_s\, \mathrm{d}B_s$, by means of Gronwall's inequality.

To this last end, we have at first for all $t \in [0,1]$

$$\begin{aligned} |X_t| &\leqslant |X_0| + C \int_0^t (1 + [1+\eta(s)]|X_s|)\, \mathrm{d}s + \sup_{s\leqslant t} |X_s^0| \\ &\leqslant \left(\|X_0\|_\infty + C + \sup_{s\leqslant 1} |X_s^0| \right) + C_1 \int_0^t |X_s|\, \mathrm{d}s, \end{aligned}$$

where $C_1 = C(1+\eta(1))$ (below we will use C_k to denote positive constants depending only of $C, \|X_0\|_\infty, \eta(1)$). It follows by Gronwall's inequality

$$|X_t| \leqslant \left(C + \|X_0\|_\infty + \sup_{s\leqslant 1} |X_s^0| \right) e^{C_1 t}, \quad \forall t \in [0,1]. \tag{5.2}$$

Next by our condition (1.15), for any $s \in [0,1]$, $u > 0$

$$\begin{aligned} \sup_{s\leqslant t\leqslant s+u} |X_t - X_s| &\leqslant \sup_{s\leqslant t\leqslant s+u} |X_t^0 - X_s^0| + u \sup_{s\leqslant t\leqslant s+u} |b(t, X_t)| \\ &\leqslant \sup_{s\leqslant t\leqslant s+u} |X_t^0 - X_s^0| + uC_2 \left(\sup_{0\leqslant t\leqslant 1} |X_t| + 1 \right). \end{aligned} \tag{5.3}$$

Consequently by (1.15), Cauchy–Schwarz and (5.2), (5.3), we get

$$\begin{aligned}
&\left[\int_{t_{k-1}^n}^{t_k^n} b(t, X_t)\,\mathrm{d}t - b(t_{k-1}^n, X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)\right]^2 \\
&\quad \leqslant \left(\frac{1}{n} C\left(1 + \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |X_t - X_{t_{k-1}^n}| + 2\eta\left(\frac{1}{n}\right) \sup_{0 \leqslant t \leqslant 1} |X_t|\right)\right)^2 \\
&\quad \leqslant \frac{C_3}{n^2}\left[1 + \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |X_t^0 - X_{t_{k-1}^n}^0|^2 + \frac{1}{n^2} \sup_{0 \leqslant t \leqslant 1} |X_t^0|^2 + \right. \\
&\qquad \left. + \frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2 \sup_{0 \leqslant t \leqslant 1} |X_t^0|^2\right]. \qquad (5.4)
\end{aligned}$$

Having this estimation we can now prove

(a): Choose $\epsilon(n) > 0$ so that

$$\epsilon(n) \to 0 \quad \text{but} \quad \frac{\eta^2(1/n) + (1/n)^2}{\epsilon(n)} \longrightarrow 0. \qquad (5.5)$$

Hence $\epsilon(n) Q_1^n(X - Y)$ is negligible in the sense of large deviation, by Theorem 1.1. It remains to show that the second term in (5.1) is negligible in the sense of large deviation, that is, for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{\epsilon(n)} Z_n > \delta\right) = -\infty.$$

By (5.4) and the definition of Z_n, the LHS above is majorized by the maximum of

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{\epsilon(n)n} \max_{k \leqslant n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |X_t^0 - X_{t_{k-1}^n}^0|^2 > C_4\delta\right) \qquad (5.6a)$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{\epsilon(n)n}\left(\frac{1}{n^2} + \eta^2\left(\frac{1}{n}\right)\right) \sup_{0 \leqslant t \leqslant 1} |X_t^0|^2 > C_5\delta\right). \qquad (5.6b)$$

Since $\int_0^t \sigma_s \,\mathrm{d}B_s = B'_{\int_0^t \sigma_s^2 \,\mathrm{d}s}$ where B' is another Brownian motion, and by Lévy's inequality

$$\mathbf{P}\left(\sup_{0 \leqslant t \leqslant T} |B'_t| > r\right) \leqslant 4 \exp\left(-\frac{r^2}{2T}\right), \qquad (5.7)$$

the limit (5.6a) is smaller than

$$\limsup_{n\to\infty} \frac{1}{n} \log\left[n \times \mathbf{P}\left(\frac{1}{\epsilon(n)n} \sup_{0 \leqslant t \leqslant \|\sigma^2\|_\infty/n} |B'_t|^2 > C_4\delta\right)\right] = -\infty.$$

The limit (5.6b) is also $-\infty$ by (5.7) and our choice (5.5) of $\epsilon(n)$.

(b): This time instead of (5.5), we choose $\epsilon(n) > 0$ so that

$$\frac{\epsilon(n)\sqrt{n}}{b(n)} \longrightarrow 0 \quad \text{but} \quad \epsilon(n)\sqrt{n} \to +\infty, \quad \frac{\eta^2\left(\frac{1}{n}\right)\cdot b(n)}{\epsilon(n)\sqrt{n}} \longrightarrow 0. \tag{5.8}$$

(e.g. $\epsilon(n) = [\eta(1/n) + 1/c(n)]b(n)/\sqrt{n}$, where $c(n)$ satisfies $b(n) \gg c(n) \to +\infty$).

By the first condition in (5.8),

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\epsilon(n)\frac{\sqrt{n}}{b(n)} Q_1^n(X-Y) > \delta\right) \\
&\quad \leqslant \limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\epsilon(n)\frac{\sqrt{n}}{b(n)} \left(Q_1^n(X-Y) - [X]_1\right) > \frac{\delta}{2}\right) \\
&\quad = -\infty
\end{aligned}$$

by Theorem 1.3. By the approximation lemma and (5.1), it remains to show

$$\limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\frac{1}{\epsilon(n)}\frac{\sqrt{n}}{b(n)} Z_n > \delta\right) = -\infty, \quad \forall \delta > 0. \tag{5.9}$$

By (5.4), the LHS above is majorized by the maximum of the following three limits

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\frac{\sqrt{n}}{\epsilon(n)b(n)} \cdot \frac{1}{n^2}\sum_{k=1}^{n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |X_t^0 - X_{t_{k-1}^n}^0|^2 > C_5\delta\right) \qquad (5.10a) \\
&\quad \leqslant \limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\frac{1}{\epsilon(n)b(n)\sqrt{n}} \max_{1\leqslant k\leqslant n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |X_t^0 - X_{t_{k-1}^n}^0|^2 > C_5\delta\right)
\end{aligned}$$

and

$$\limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\frac{1}{\epsilon(n)b(n)n^{5/2}} \sup_{0\leqslant t\leqslant 1} |X_t^0|^2 > C_6\delta\right) \tag{5.10b}$$

and

$$\limsup_{n\to\infty} \frac{1}{b^2(n)} \log \mathbf{P}\left(\frac{\eta(1/n)^2}{\epsilon(n)b(n)\sqrt{n}} \sup_{0\leqslant t\leqslant 1} |X_t^0|^2 > C_7\delta\right). \tag{5.10c}$$

By Lévy's inequality (5.7) and our choice (5.8) of $\epsilon(n)$, the limits (5.10b) and (5.10c) are both $-\infty$.

The estimation of (5.10a) is a little more difficult and we can not estimate it roughly as in the control of (5.6a) above (see the remark below). The key is the following isoperimetric inequality [8, p17, (1.24)] (with a small abuse : there is no absolute value in his formula, but the same proof works)

$$\mathbf{P}\left(\max_{k\leqslant n}\sup_{t_{k-1}^n\leqslant t\leqslant t_k^n}|X_t^0-X_{t_{k-1}^n}^0|>m(n)+r\right)\leqslant\exp\left(-\frac{r^2}{2\Sigma(n)}\right),\tag{5.11}$$

where $m(n)=\mathbf{E}\max_{k\leqslant n}\sup_{t_{k-1}^n\leqslant t\leqslant t_k^n}|X_t^0-X_{t_{k-1}^n}^0|$ and

$$\Sigma(n)=\max_{k\leqslant n}\sup_{t_{k-1}^n\leqslant t\leqslant t_k^n}\mathbf{E}|X_t^0-X_{t_{k-1}^n}^0|^2.\tag{5.12}$$

At first by our condition (1.10) on $\sigma.$,

$$\sqrt{n}b(n)\Sigma(n)\longrightarrow 0.\tag{5.13}$$

Next by Jensen's inequality and Doob's inequality

$$\begin{aligned}m(n)&\leqslant\left(\mathbf{E}\max_{k\leqslant n}\sup_{t_{k-1}^n\leqslant t\leqslant t_k^n}|X_t^0-X_{t_{k-1}^n}^0|^4\right)^{1/4}\\&\leqslant\left(\sum_{k=1}^n\mathbf{E}\sup_{t_{k-1}^n\leqslant t\leqslant t_k^n}|X_t^0-X_{t_{k-1}^n}^0|^4\right)^{1/4}\\&\leqslant\frac{4}{3}\left(\sum_{k=1}^n\mathbf{E}|X_{t_k^n}^0-X_{t_{k-1}^n}^0|^4\right)^{1/4}\\&\leqslant C_8\cdot n^{1/4}\sqrt{\Sigma(n)}\longrightarrow 0,\end{aligned}\tag{5.14}$$

where $C_8=4\|\xi\|_4/3=4\sqrt{2}/3$ (ξ is of law $N(0,1)$), and the last relation follows from (5.13).

Consequently by (5.11), the limit (5.10a) is smaller than

$$\limsup_{n\to\infty}-\frac{1}{b^2(n)}\cdot\frac{\epsilon(n)b(n)\sqrt{n}}{2\Sigma(n)}C_9\delta=-C_9\delta\cdot\liminf_{n\to\infty}\frac{\epsilon(n)n}{2\sqrt{n}b(n)\Sigma(n)}$$

which is $-\infty$ by (5.13) and (5.8).

The proof is completed. □

Remarks 5.1. If one estimates (5.10a) roughly as in the control of (5.6a), we would get

$$\begin{aligned}&\mathbf{P}\left(\max_{k\leqslant n}\sup_{t_{k-1}^n\leqslant t\leqslant t_k^n}|X_t^0-X_{t_{k-1}^n}^0|>r\right)\\&\quad\leqslant n\cdot\mathbf{P}\left(\sup_{0\leqslant t\leqslant\Sigma(n)}|B_t'|^2>r\right)\\&\quad\leqslant 4n\cdot\exp\left(-\frac{r^2}{2\Sigma(n)}\right),\end{aligned}\tag{5.15}$$

where the last inequality follows by Lévy's inequality (5.7). It has an extra factor n w.r.t. (5.11). When $b^2(n) \ll \log n$, this estimation is not enough to conclude the negligibility of (5.10a).

6. Proof of Proposition 1.5

To prove (1.16) and (1.17), we can assume without loss of generality that $b = 0$. Remark that $(Q_t^n(X) - \mathbf{E} Q_t^n(X))$ is a $(\mathcal{F}_{[nt]/n})$-martingale. Then

$$\exp\left(\lambda\left[Q_t^n(X) - \mathbf{E} Q_t^n(X)\right]\right)$$

is a sub-martingale for each $\lambda \in \mathbf{R}$. Let

$$T := \inf\left\{0 \leqslant t \leqslant 1 : \left[Q_t^n(X) - \mathbf{E} Q_t^n(X)\right] \geqslant r\right\} \quad (\inf \emptyset := +\infty).$$

For all $r, \lambda > 0$, we have by Chebychev inequality and the stopping time theorem of Doob,

$$\begin{aligned}
&\mathbf{P}\left(\sup_{0\leqslant t\leqslant 1}[Q_t^n(X) - \mathbf{E} Q_t^n(X)] \geqslant r\right) \\
&\quad = \mathbf{P}\left(T \leqslant 1\right) \\
&\quad \leqslant e^{-\lambda n r} \mathbf{E} 1_{[T\leqslant 1]} \exp\left(n\lambda\left[Q_T^n(X) - \mathbf{E} Q_T^n(X)\right]\right) \\
&\quad \leqslant e^{-\lambda n r} \mathbf{E} 1_{[T\leqslant 1]} \exp\left(\lambda n\left[Q_1^n(X) - \mathbf{E} Q_1^n(X)\right]\right).
\end{aligned} \tag{6.1}$$

But by (2.3) in Lemma 2.2, this last quantity is smaller than

$$\exp\left(-n\left(\lambda(r + [X]_1) - \Lambda(\lambda)\right)\right).$$

Taking the infinimum of this quantity over $\lambda \geqslant 0$, we get

$$\exp\left(-n\Lambda^*(r + [X]_1)\right)$$

which is the first inequality in (1.16). For the second explicit inequality in (1.16), observe for each $\lambda \in [0, \lambda_0)$ (recall $\lambda_0 = \frac{1}{2\|\sigma^2\|_\infty}$)

$$\Lambda''(\lambda) = \int_0^1 \frac{2\sigma_t^4}{\left(1 - 2\lambda\sigma_t^2\right)^2}\, \mathrm{d}t \leqslant \int_0^1 \frac{2\|\sigma^2\|_\infty^2}{(1 - 2\lambda\|\sigma^2\|_\infty)^2}\, \mathrm{d}t = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} P(\lambda\|\sigma^2\|_\infty).$$

where $P(\lambda)$ is given in Lemma 2.1. Consequently by Taylor formula: $f(x) = f(0) + f'(0)x + \int_0^x f''(y)(x-y)\,\mathrm{d}y$, we have for each $\lambda \in [0, \lambda_0)$,

$$\Lambda(\lambda) - \Lambda'(0)\lambda \leqslant P(\lambda\|\sigma^2\|_\infty) - P'(0)\lambda\|\sigma^2\|_\infty.$$

It follows that (note $\Lambda'(0) = [X]_1$, $P'(0) = 1$)

$$\begin{aligned}
\Lambda^*([X]_1 + r) &= \sup_{0<\lambda<\lambda_0} \{\lambda([X]_1 + r) - \Lambda(\lambda)\} \\
&\geqslant \sup_{0<\lambda<\lambda_0} \left\{\lambda(\|\sigma^2\|_\infty + r) - P(\lambda\|\sigma^2\|_\infty)\right\} \\
&= P^*\left(\frac{\|\sigma^2\|_\infty + r}{\|\sigma^2\|_\infty}\right),
\end{aligned}$$

where P^* given by (1.8) is the Legendre transformation of $P(\lambda)$. So the second inequality in (1.16) follows.

The proof of the first inequality in (1.17) is similar to that of (1.16). We turn now to prove the second *explicit* inequality in (1.17). To this end, we assume $\sigma \in L^4(\mathrm{d}t)$ (trivial otherwise). Since for $\lambda < 0$,

$$\Lambda''(\lambda) = \int_0^1 \frac{2\sigma_t^4}{\left(1 - 2\lambda\sigma_t^2\right)^2}\,\mathrm{d}t \tag{6.2}$$

is increasing in λ, then by Taylor's formula,

$$\Lambda(\lambda) \leqslant \Lambda'_-(0)\lambda + \frac{1}{2}\lim_{c\to 0-}\Lambda''(c)\lambda^2 = [X]_1\lambda + \int_0^1 \sigma_t^4\,\mathrm{d}t \cdot \lambda^2.$$

Hence

$$\Lambda^*([X]_1 - r) \geqslant \sup_{\lambda<0}\left(-\lambda r - \int_0^1 \sigma_t^4\,\mathrm{d}t \cdot \lambda^2\right) = \frac{r^2}{4\int_0^1 \sigma_s^4\,\mathrm{d}s},$$

where (1.17) follows. Finally by Theorem 1.1,

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{n}\log\mathbf{P}\left(\sup_{t\in[0,1]} \pm(Q_t^n(X) - [X]_t) \geqslant r\right)\\
&\quad\geqslant \liminf_{n\to\infty}\frac{1}{n}\log\mathbf{P}\left(\pm(Q_1^n(X) - [X]_1) \geqslant r\right)\\
&\quad= -\inf\left\{\Lambda^*(x); \pm(x - [X]_1) \geqslant r\right\} = -\Lambda^*\left([X]_1 \pm r\right).
\end{aligned}$$

Combining it with the upper bounds in (1.16) and (1.17), we obtain (1.18). □

Acknowledgements

The authors want to thank M. Bouaziz and P.Bertrand for fruitful discussions. We are grateful to the anonymous referees for their valuable comments which improve the presentation of this work.

Appendix

The proof of the process-level LDP in Theorem 1.3 is based on the following:

LEMMA A.1. *Let* $(X^n(t)_{0\leqslant t\leqslant 1})_{n\geqslant 0}$ *be a sequence of real right continuous left limit processes defined on* $(\Omega, \mathcal{F}, \mathbf{P})$. *Let* $(\lambda(n))_{n\geqslant 0}$ *be a sequence of positive numbers tending to infinity, and* $\mathbf{D}[0,1]$ *be the space of real right continuous left limit functions but equipped with the uniform convergence topology and with the* σ*-field* $\mathcal{B}^s$. *Assume* *(i) For every finite partition* $\mathcal{P}$ *of* $[0,1]$, $\mathbf{P}\left(X^n_{\mathcal{P}} \in \cdot\right)$ *satisfies the LDP on* $\mathbf{R}^{\mathcal{P}}$ *with speed* $\lambda(n)$ *and with the rate function* $I^{\mathcal{P}}$*;*

(ii) $\forall\delta > 0$,

$$\lim_{\varepsilon\to 0}\sup_{0\leqslant s\leqslant 1}\limsup_{n\to\infty}\frac{1}{\lambda(n)}\log\mathbf{P}\left(\sup_{s\leqslant t\leqslant s+\varepsilon}|X^n(t)-X^n(s)|>\delta\right)=-\infty, \quad \text{(A.1)}$$

(Convention: $\forall t > 1,\ X^n(t) := X^n(1)$*). Then* $\mathbf{P}\left(X^n \in \cdot\right)$ *satisfies on* $\mathbf{D}[0,1]$ *w.r.t the sup norm topology with the same speed* $\lambda(n)$ *and with the rate function given by*

$$I(\gamma) = \sup_{\mathcal{P}} I^{\mathcal{P}}(\gamma(\mathcal{P})). \quad \text{(A.2)}$$

where the supremum is taken over all finite partitions of [0,1]. Moreover $[I < +\infty]$ *is a subset of the space* $C[0,1]$ *of continuous functions.*

Remarks. In the works of Liptser and Pukhalskii [9, Th.3.1] and Pukhalskii [11], they show that

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{\lambda(n)}\log\sup_{\tau\in T_1(\mathcal{F}^n)}\mathbf{P}\left(\sup_{\tau\leqslant t\leqslant\tau+\varepsilon}|X^n(t)-X^n(\tau)|>\delta\right)=-\infty, \quad \text{(A.3)}$$

is sufficient to the so called *exponential tightness* on $\mathbf{D}[0,1]$ w.r.t. the Skorohod topology, where $\mathcal{F}^n$ is a filtration w.r.t. which X^n is adapted and $T_1(\mathcal{F}^n)$ is the family of all $\mathcal{F}^n$-stopping times less than 1. Remark that (A.1) is much weaker than (A.3), especially the supremum over $\tau \in T_1(\mathcal{F}^n)$ *inside* the limit of $n \to \infty$ in (A.3) appears now *outside* that limit in (A.1) (this is crucial for Step 4 in our proof of Th.1.3). Moreover the LDP on $\mathbf{D}[0,1]$ w.r.t. the sup norm topology in this lemma is much stronger, and it implies the exponential tightness on $\mathbf{D}[0,1]$ w.r.t. the Skorohod topology.

This lemma is taken from [13, Prop.I.5.6] and we reproduce here its proof (as [13] is not available for many readers, and its proof is short).

Proof. At first condition (i) implies the LDP of $\mathbf{P}(X^n \in \cdot)$ on $\mathbf{R}^{[0,1]}$ w.r.t. the pointwise convergence topology, with the rate function $I(\gamma)$ given by (A.2) (see [13] or applying [5,§4.6]). Next let $\mathcal{P}^k = \{i/k; i = 0,\dots,k\}$ and consider the application $f^k\colon \mathbf{R}^{[0,1]} \to \mathbf{D}[0,1]$ such that the graph of $f^k(\gamma)$ on $[0,1]$ is the polygon linking the points $(i/k, \gamma(i/k)), i = 0,1,\dots,k$.

We shall establish the following two facts :

$$\limsup_{n\to\infty}\frac{1}{\lambda(n)}\log\mathbf{P}\left(\sup_{0\leqslant t\leqslant 1}|X^n(t)-f^k(X^n)(t)|>\delta\right)\longrightarrow -\infty; \quad \text{(A.4)}$$

as k tends to infinity, for any $\delta > 0$; and for each $L \geqslant 0$,

$$\lim_{k\to\infty}\sup_{\gamma\in[I\leqslant L]}\sup_{0\leqslant t\leqslant 1}|f^k(\gamma)(t)-\gamma(t)| = 0. \quad \text{(A.5)}$$

To show (A.4), it is enough to notice that its LHS is smaller than

$$\max_{0 \leqslant i \leqslant k-1} \limsup_{n\to\infty} \frac{1}{\lambda(n)} \log \mathbf{P}\left(\sup_{\frac{i}{k} \leqslant t \leqslant \frac{i+1}{k}} |X^n(t) - X^n(i/k)| > \delta \right)$$

which tends to $-\infty$ as $k \to \infty$, by (A.1).

To show (A.5), consider the set

$$A(k,\delta) := \left\{ \gamma \in \mathbf{R}^{[0,1]}\, ;\ \sup_{0 \leqslant t \leqslant 1} |\gamma(t) - f^k(\gamma)(t)| > \delta \right\}.$$

where $\delta > 0$ fixed, $k \geqslant 1$. It is an open subset in $\mathbf{R}^{[0,1]}$ on which the lower bound of large deviation below holds (as noted at the beginning):

$$\liminf_{n\to\infty} \frac{1}{\lambda(n)} \log \mathbf{P}\left(X^n \in A(k,\delta)\right) \geqslant -\inf\{I(\gamma)\, ;\, \gamma \in A(k,\delta)\}. \tag{A.6}$$

But for any $\delta > 0$, the LHS above tends to $-\infty$ by (A.4) as $k \to \infty$. Therefore $[I \leqslant L] \subset \bigcap_{k \geqslant N} A(k,\delta)^c$ for some N large enough, by (A.6). Then (A.5) follows.

By condition (i), $\{\gamma(0);\ I(\gamma) \leqslant L\}$ is bounded and (A.5) implies the equicontinuity of $[I \leqslant L]$. Then $[I \leqslant L]$ is compact in $C[0,1]$ for any $L > 0$ and $[I < +\infty] \subset C[0,1] \subset \mathbf{D}[0,1]$. Consequently $\mathbf{P}(X^n \in \cdot)$ satisfies the LDP on $\mathbf{D}[0,1]$ w.r.t. the pointwise convergence topology. Finally by an approximation lemma in [5, Th.4.2.23] (with some abuse : their approximation lemma is stated in the framework of Polish space, but it can be easily translated into the actual context), (A.4)+(A.5) implies the LDP of $\mathbf{P}(X^n \in \cdot)$ on $\mathbf{D}[0,1]$ w.r.t. the uniform convergence topology. □

References

1. Avesani, R. G. and Bertrand, P.: Does volatility jump or just diffuse: a statistical approach. In: L.C.G. Rogers and D. Talay (eds), *Numerical Methods in Finances*, Cambridge Univ. Press, pp. 270–289, 1997.
2. Bertrand, P.: Quelques applications de processus stochastiques: contrôle adaptatif, statistique des processus, détection de ruptures..., HDR Université Blaise-Pascal, Clermont-Ferrand, 1997.
3. Bryc, W. and Dembo, A.: Large deviations for quadratic functionals of Gaussian processes, *J. Theoret. Probab.* **10** (1997), 307–332.
4. Bercu, B., Gamboa, F. and Rouault, A.: Large deviations for quadratic forms of Gaussian stationnary processes, *Stoch. Proc. Appli.* **71** (1997), 75–90.
5. Dembo, A. and Zeitouni, O.: *Large deviations techniques and applications*, Jones and Bartlett, Boston, MA. 1993.
6. Ellis, R. S.: *Entropy, Large Deviations and Statistical Mechanics*, Springer-Verlag, New York, 1985.
7. Florens-Zmirou, D.: On estimating the variance of diffusion processes, *J. Appl. Prob.* **30** (1993), 790–804.
8. Ledoux, M.: *Concentration of Measure and Logarithmic Sobolev Inequalities*, Preprint (Berlin), 1997.

9. Liptser, R. S. and Pukhalski, A. A.: Limit theorems on large deviations for semimartingales, *Stoch. Stoch. Reports* **38** (1992), 201–249.
10. Lynch, J. and Sethuraman, J.: Large deviations for processes with independent increments, *The Annals of Prob.* **15**(2) (1987), 610–627.
11. Pukhalskii, A. A.: Large deviations of semimartingales via convergence of the predictable characteristics, *Stoch. and Stoch. Reports* **49** (1994), 27–85.
12. Revuz, D. and Yor, M.: *Continuous Martingales and Brownian Motion*, Springer-Verlag, 1991.
13. Wu, L.: An introduction to large deviations (in chinese). In: J. A. Yan, S. Peng, S. Fang and L. Wu (eds), *Several Topics in Stochastic Analysis*, Academic Press of China, Beijing, pp. 225–336, 1997.

Statistics and Probability Letters 86 (2014) 30–37

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Large and moderate deviations of realized covolatility

Hacène Djellout*, Yacouba Samoura

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais, BP80026, 63171 Aubière Cedex, France

ARTICLE INFO

Article history:
Received 17 June 2013
Received in revised form 3 December 2013
Accepted 3 December 2013
Available online 25 December 2013

MSC:
60F10
62J05
60J05

Keywords:
Deviation inequalities
Large and moderate deviation principle
Diffusion
Discrete-time observation
Realized volatility

ABSTRACT

In this note, we consider the large and moderate deviation principle of the estimators of the integrated covariance of two-dimensional diffusion processes when they are observed only at discrete times in a synchronous manner. The proof is extremely simple. It is essentially an application of the contraction principle for the results given in the case of the volatility by Djellout et al. (1999).

1. Motivation and context

Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, let $(X_{1,t}, X_{2,t})$ be a two-dimensional diffusion process given by

$$\begin{cases} dX_{1,t} = u_{1,t}(X_{1,t})dt + \sigma_{1,t}dB_{1,t} \\ dX_{2,t} = u_{2,t}(X_{2,t})dt + \sigma_{2,t}dB_{2,t} \end{cases} \tag{1.1}$$

where $((B_{1,t}, B_{2,t}), t \geq 0)$ is a two-dimensional Gaussian process with independent increments, zero mean and covariance matrix

$$\begin{pmatrix} t & \int_0^t \rho_s ds \\ \int_0^t \rho_s ds & t \end{pmatrix} \quad \forall t \geq 0.$$

In (1.1), (u_1, u_2) is a progressively measurable process (possibly unknown). In what follows, we restrict our attention to the case when σ_1, σ_2 and ρ are deterministic functions; the functions σ_i, $i = 1, 2$ take positive values while ρ takes values in the interval $[-1, 1]$. Note that the marginal processes B_1 and B_2 are Brownian motions (BM). Moreover, we can define a process B_t^* such that $(B_{1,t}, B_t^*)_{t\geq 0}$ is a two-dimensional BM and $dB_{2,t} = \rho_t dB_{1,t} + \sqrt{1-\rho_t^2}dB_t^*$ for every $t \geq 0$.

* Corresponding author. Fax: +33 4 73 40 70 64.
E-mail addresses: Hacene.Djellout@math.univ-bpclermont.fr (H. Djellout), Samoura.Yacouba@math.univ-bpclermont.fr (Y. Samoura).

0167-7152/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.12.003

In this note, the parameter of interest is the (deterministic) covariance of X_1 and X_2

$$\langle X_1, X_2\rangle_t = \int_0^t \sigma_{1,t}\sigma_{2,t}\rho_t dt. \tag{1.2}$$

In finance, $\langle X_1, X_2\rangle_.$ is the integrated covariance (over [0, 1]) of the logarithmic prices X_1 and X_2 of two securities. It is an essential quantity to be measured for risk management purposes. The covariance for multiple price processes is of great interest in many financial applications. The naive estimator is the realized covariance, which is the analogue of realized variance for a single process.

Typically $X_{1,t}$ and $X_{2,t}$ are not observed in continuous time but we have only discrete time observations. Given discrete equally spaced observation $(X_{1,t_k^n}, X_{2,t_k^n}, k = 1, \dots, n)$ in the interval [0, 1] (with $t_k = k/n$), a commonly used approach to estimate is to take the sum of cross products

$$\mathbf{C}_t^n := \sum_{k=1}^{[nt]} \left(X_{1,t_k^n} - X_{1,t_{k-1}^n}\right)\left(X_{2,t_k^n} - X_{2,t_{k-1}^n}\right), \tag{1.3}$$

where $[x]$ denotes the integer part of $x \in \mathbb{R}$.

When the drift is known, we can also consider the following estimator:

$$\bar{\mathbf{C}}_t^n := \sum_{k=1}^{[nt]} \left(X_{1,t_k^n} - X_{1,t_{k-1}^n} - \int_{t_{k-1}^n}^{t_k^n} u_{1,t}(X_{1,t})dt\right)\left(X_{2,t_k^n} - X_{2,t_{k-1}^n} - \int_{t_{k-1}^n}^{t_k^n} u_{2,t}(X_{2,t})dt\right).$$

In the unidimensional case and in the case that X have non-jump, this question has been well investigated—see Djellout et al. (1999) for relevant references. In Djellout et al. (1999) and recently in Kanaya and Otsu (2012), the authors obtained the large and moderate deviations for the realized volatility. The results of Djellout et al. (1999) are extended to jump-diffusion processes. Mancini (2008) established the large deviation result for the threshold estimator for the constant volatility. Jiang (2010) derived a moderate deviation result for the threshold estimator for the quadratic variational process.

In the bivariate case, Hayashi and Yoshida (2011) considered the problem of estimating the covariation of two diffusion processes under a non-synchronous sampling scheme. They proposed an alternative estimator and they investigated the asymptotic distributions. In Dalalyan and Yoshida (2011), the authors complement the results in Hayashi and Yoshida (2011) by establishing a second-order asymptotic expansion for the distribution of the estimator in a fairly general setup, including random sampling schemes and (possibly random) drift terms. Several further works have been realized when data on two securities are observed non-synchronously, see also Aït-Sahalia et al. (2010). Here we do not consider the asynchronous case. In the bivariate case we also mention the work of Mancini and Gobbi (2012) which deals with the problem of distinguishing the Brownian covariation from the co-jumps using a discrete set of observations.

The purpose of this note is to furnish some further estimations about the estimator (1.3), refining the already known central limit theorem. More precisely, we are interested in the estimations of

$$\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left(\mathbf{C}_t^n - \int_0^t \sigma_{1,t}\sigma_{2,t}\rho_t dt\right) \in A\right),$$

where A is a given domain of deviation, and $(b_n)_{n>0}$ is some sequence denoting the scale of the deviation. When $b_n = 1$, this is exactly the estimation of the central limit theorem. When $b_n = \sqrt{n}$, it becomes the *large deviations*. And when $1 \ll b_n \ll \sqrt{n}$, this is the so-called *moderate deviations*. The main problem studied in this paper is the large and moderate deviation estimations of the estimator. In this bivariate case things are not complicated.

We refer to Dembo and Zeitouni (1998) for an exposition of the general theory of large deviation and limit ourself to the statement of the some basic definitions. Let $\{\mu_T, T > 0\}$ be a family of probability on a topological space $(S, \mathcal{S})$ where $\mathcal{S}$ is a σ-algebra on S and $v(T)$ a non-negative function on $[1, \infty)$, such that $\lim_{T\to\infty} v(T) = +\infty$. A function $I : S \to [0, \infty]$ is said to be a rate function if it is lower semicontinuous and it is said to be a good rate function if its level set $\{x \in S : I(x) \le a\}$ is compact for all $a \ge 0$. $\{\mu_T\}$ is said to satisfy a large deviation principle (LDP) with speed $v(T)$ and rate function $I(x)$ if for any set $A \in \mathcal{S}$

$$-\inf_{x\in A^\circ} I(x) \le \lim_{T\to\infty} \begin{pmatrix}\inf\\ \sup\end{pmatrix} \frac{1}{v(T)} \log \mu_T(A) \le -\inf_{x\in\bar{A}} I(x),$$

where $A^0, \bar{A}$ are the interior and the closure of A respectively.

This paper is organized as follows. In the next section we present the main results of this paper. They are established in the last section.

2. Main results

Our first result is about the LDP of $\mathbb{P}(\mathbf{C}_1^n \in \cdot)$, with time $t = 1$ fixed.

Proposition 2.1. *Let* $(X_{1,t}, X_{2,t})$ *be given by* (1.1).

(1) *For every* $\lambda \in \mathbb{R}$

$$\Lambda_n(\lambda) := \frac{1}{n}\log \mathbb{E}(\exp(\lambda n \bar{\mathbf{C}}_1^n)) \le \Lambda(\lambda) := \begin{cases} \int_0^1 -\frac{1}{2}\log(1-\lambda\sigma_{1,t}\sigma_{2,t}(1+\rho_t)) - \frac{1}{2}\log(1+\lambda\sigma_{1,t}\sigma_{2,t}(1-\rho_t))dt \\ \quad \text{if } -\frac{1}{\|\sigma_1\sigma_2(1-\rho)\|} \le \lambda \le \frac{1}{\|\sigma_1\sigma_2(1+\rho)\|} \\ +\infty, \quad \text{otherwise} \end{cases}$$

and

$$\lim_{n\to\infty} \Lambda_n(\lambda) = \Lambda(\lambda).$$

(2) *Assume that* $\sigma_{1,\cdot}\sigma_{2,\cdot}(1 \pm \rho_\cdot) \in L^\infty([0,1], dt)$ *and* $u_{l,\cdot}(\cdot) \in L^\infty(dt \otimes \mathbb{P})$, *for* $l = 1, 2$. *Then* $\mathbb{P}(\mathbf{C}_1^n \in \cdot)$ *satisfies the LDP on* $\mathbb{R}$ *with speed n and with the good rate function given by the Legendre transformation of* Λ, *that is*

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\}. \tag{2.1}$$

We now extend Proposition 2.1 to the process-level large deviations of $\mathbb{P}(\mathbf{C}_\cdot^n \in \cdot)$, which is interesting from the viewpoint of the non-parametric statistics.

Let $\mathbb{D}_b([0,1])$ be the real right-continuous–left-limit and bounded variation functions γ. The space $\mathbb{D}_b([0,1])$ of γ, identified in the usual way as the space of bounded measures $d\gamma$ on $[0,1]$, with $d\gamma[0,t] = \gamma(t)$ and $d\gamma(0) = \gamma(0)$, will be equipped with the weak convergence topology and the σ-field $\mathcal{B}^s$ generated by the coordinate $\{\gamma(t), 0 \le t \le 1\}$. We denote by $\dot{\gamma}(t)dt$ and $d\gamma^\perp$ respectively the absolute continuous part and the singular part of the measure $d\gamma$ associated with $\gamma \in \mathbb{D}_b[0,1]$ w.r.t. the Lebesgue measure dt. The signed measure γ has a unique decomposition into a difference $\gamma = \gamma_+ - \gamma_-$ of two positive measures γ_+ and γ_-. In the paper, we denote by P^* the function

$$P^*(x) = \begin{cases} \frac{1}{2}(x - 1 - \log x) & \text{if } x > 0 \\ +\infty & \text{if } x \le 0, \end{cases} \tag{2.2}$$

which is the Legendre transformation of P given by

$$P(\lambda) = \begin{cases} -\frac{1}{2}\log(1-2\lambda) & \text{if } \lambda < \frac{1}{2} \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.3}$$

Theorem 2.2. *Let* $(X_{1,t}, X_{2,t})$ *be given by* (1.1). *Assume that* $\sigma_{1,\cdot}\sigma_{2,\cdot}(1\pm\rho_\cdot) \in L^\infty([0,1], dt)$ *and* $u_{l,\cdot}(\cdot) = u_l(\cdot,\cdot) \in L^\infty(dt\otimes\mathbb{P})$, *for* $l = 1, 2$. *Then*

(1) $\mathbb{P}(\mathbf{C}_\cdot^n \in \cdot)$ *satisfies the LDP on* $\mathbb{D}_b([0,1])$ *w.r.t. the weak convergence topology, with speed n and with some inf-compact convex rate function* $J(\gamma)$.

(2) *If moreover* $t \to \sigma_{1,t}\sigma_{2,t}(1 \pm \rho_t)$ *is continuous and strictly positive on* $[0,1]$, *then*

$$J(\gamma) = J_+^{abs}(\gamma_+ + \beta) + J_-^{abs}(\gamma_- + \beta) + J_+^\perp(\gamma_+) + J_-^\perp(\gamma_-), \tag{2.4}$$

where β *is absolutely continuous with respect to the Lebesgue measure and given by*

$$\dot{\beta}(t) = \frac{\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2) - (\dot{\gamma}_+(t) + \dot{\gamma}_-(t))}{2} + \frac{\sqrt{[\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2) - (\dot{\gamma}_+(t)+\dot{\gamma}_-(t))]^2 + (\dot{\gamma}_+(t)+\dot{\gamma}_-(t))\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2)}}{2},$$

and

$$J_\pm^\perp(\gamma) = \int_0^1 \frac{1}{\sigma_{1,t}\sigma_{2,t}(1\pm\rho_t)} d\gamma^\perp,$$

and

$$J_\pm^{abs}(\gamma) = \int_0^1 P^*\left(\frac{2\dot{\gamma}(t)}{\sigma_{1,t}\sigma_{2,t}(1\pm\rho_t)}\right)dt,$$

where P^* *is given in* (2.2).

We discuss now the moderate deviation principle. To this purpose, let $(b_n)_{n\geq 1}$ be a sequence of positive numbers such that

$$b_n \to \infty \quad \text{and} \quad \frac{b_n}{\sqrt{n}} \to 0 \quad \text{as } n \to \infty.$$

Let $\mathbb{D}_0[0,1]$ be the Banach space of real right-continuous–left-limit functions γ on $[0,1]$ with $\gamma(0)=0$, equipped with the uniform sup norm and the σ-field $\mathcal{B}^s$ generated by the coordinate $\{\gamma(t), 0 \leq t \leq 1\}$.

Theorem 2.3. *Given $(X_{1,t}, X_{2,t})$ by (1.1) with $u_{l,\cdot}(\cdot) = u_l(\cdot,\cdot) \in L^\infty(dt \otimes \mathbb{P})$, for $l = 1, 2$. Assume that $\sigma_{1,\cdot}\sigma_{2,\cdot}(1 \pm \rho_\cdot) \in L^2([0,1], dt)$ and*

$$\sqrt{n} b_n \max_{1\leq k\leq n} \int_{(k-1)/n}^{k/n} \sigma_{1,t}\sigma_{2,t}(1 \pm \rho_t)dt \longrightarrow 0. \tag{2.5}$$

Then $\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left(\mathbf{C}^n_\cdot - \langle X_1, X_2\rangle_\cdot\right) \in \cdot\right)$ satisfies the LDP on $\mathbb{D}_0([0,1])$ with speed b_n^2 and with the good rate function J_m given by

$$J_m(\gamma) = \begin{cases} \displaystyle\int_0^1 \frac{\dot{\gamma}(t)^2}{2\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)} 1_{[t:\sigma_{1,t}\sigma_{2,t}>0]}dt & \text{if } d\gamma \ll \sigma_{1,t}\sigma_{2,t}\sqrt{1+\rho_t^2}dt \\ +\infty & \text{otherwise.} \end{cases} \tag{2.6}$$

Remark 2.4. In particular, $\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left(\mathbf{C}^n_1 - \langle X_1, X_2\rangle_1\right) \in \cdot\right)$ satisfies the LDP on $\mathbb{R}$ with speed b_n^2 and with the rate function given by

$$I_m(x) = \frac{x^2}{2\int_0^1 \sigma_{1,s}^2\sigma_{2,s}^2(1+\rho_s^2)ds}, \quad \forall x \in \mathbb{R}.$$

Remark 2.5. If for some $p > 2$,

$$\sigma_{1,\cdot}\sigma_{2,\cdot}(1 \pm \rho_\cdot) \in L^p([0,1], dt) \quad \text{and} \quad b_n = O\left(n^{\frac{1}{2}-\frac{1}{p}}\right),$$

we obtain (2.5).

Remark 2.6. Theorems 2.2 and 2.3 continue to hold under the linear growth condition of the drift u_l $(l = 1, 2)$ rather than the boundedness. More precisely assume that

$$|u_{l,s}(x) - u_{l,t}(y)| \leq \alpha_l \left[1 + |x-y| + \eta_l(|s-t|)(|x|+|y|)\right], \quad \forall s,t \in [0,1],\ x,y \in \mathbb{R},$$

where $\eta_l : [0,+\infty) \to [0,+\infty)$ is a continuous nondecreasing function with $\eta_l(0) = 0$ and $\alpha_l > 0$ is a constant. Then the LDP of Theorems 2.2 and 2.3 continue to hold for $\mathbb{P}(\tilde{\mathbf{C}}^n_\cdot \in \cdot)$, where $\tilde{\mathbf{C}}^n_\cdot$ is given by

$$\tilde{\mathbf{C}}^n_t := \sum_{k=1}^{[nt]} \left(X_{1,t_k^n} - X_{1,t_{k-1}^n} - u_{1,t_{k-1}^n}(X_{1,t_{k-1}^n})(t_k^n - t_{k-1}^n)\right)\left(X_{2,t_k^n} - X_{2,t_{k-1}^n} - u_{2,t_{k-1}^n}(X_{2,t_{k-1}^n})(t_k^n - t_{k-1}^n)\right).$$

We introduce the following function:

$$\Lambda_\pm^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda_\pm(\lambda)\}, \tag{2.7}$$

which is the Legendre transformation of $\Lambda_\pm$ given by

$$\Lambda_\pm(\lambda) := \int_0^1 P\left(\pm\frac{\lambda\sigma_{1,t}\sigma_{2,t}(1\pm\rho_t)}{2}\right)dt \tag{2.8}$$

and we denote

$$\alpha_{\pm,t} = \frac{1}{2}\int_0^t \sigma_{1,s}\sigma_{2,s}(1\pm\rho_s)ds. \tag{2.9}$$

An easy application of deviation inequalities given in Proposition 1.5 in Djellout et al. (1999) gives the following proposition.

Proposition 2.7. *We have for every $n \geq 1$ and $r > 0$,*

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,1]}[\bar{\mathbf{C}}_t^n - \mathbb{E}\bar{\mathbf{C}}_t^n] \geq r\right) &\leq \exp\left(-n\Lambda_+^*\left(\alpha_+ + \frac{r}{2}\right)\right) + \exp\left(-n\Lambda_-^*\left(\alpha_- - \frac{r}{2}\right)\right)\\ &\leq \exp\left(-\frac{n}{2}\left[\frac{r}{\|\sigma_1\sigma_2(1+\rho)\|_\infty} - \log\left(1+\frac{r}{\|\sigma_1\sigma_2(1+\rho)\|_\infty}\right)\right]\right)\\ &\quad + \exp\left(-n\frac{r^2}{4\int_0^1[\sigma_1\sigma_2(1-\rho)]^2dt}\right),\\ \mathbb{P}\left(\inf_{t\in[0,1]}[\bar{\mathbf{C}}_t^n - \mathbb{E}\bar{\mathbf{C}}_t^n] \leq -r\right) &\leq \exp\left(-n\Lambda_+^*\left(\alpha_+ - \frac{r}{2}\right)\right) + \exp\left(-n\Lambda_-^*\left(\alpha_- + \frac{r}{2}\right)\right)\\ &\leq \exp\left(-n\frac{r^2}{4\int_0^1[\sigma_1\sigma_2(1+\rho)]^2dt}\right)\\ &\quad + \exp\left(-\frac{n}{2}\left[\frac{r}{\|\sigma_1\sigma_2(1-\rho)\|_\infty} - \log\left(1+\frac{r}{\|\sigma_1\sigma_2(1-\rho)\|_\infty}\right)\right]\right),\end{aligned}$$

where $\Lambda_\pm^*$ and $\alpha_\pm$ are given in (2.7) and (2.9) respectively.

3. Proof

In this section, we will give some hints for the proof of the main results. We have

$$\bar{\mathbf{C}}_t^n = \sum_{k=1}^{[nt]}\left(\int_{t_{k-1}^n}^{t_k^n}\sigma_{1,s}dB_{1,s}\right)\left(\int_{t_{k-1}^n}^{t_k^n}\sigma_{2,s}dB_{2,s}\right) = \sum_{k=1}^{[nt]}\sqrt{a_k}\sqrt{a_k'}\xi_k\xi_k',$$

where

$$\xi_k = \frac{\int_{t_{k-1}^n}^{t_k^n}\sigma_{1,s}dB_{1,s}}{\sqrt{a_k}} \qquad \xi_k' = \frac{\int_{t_{k-1}^n}^{t_k^n}\sigma_{2,s}dB_{2,s}}{\sqrt{a_k'}} \quad \text{with } a_k = \int_{t_{k-1}^n}^{t_k^n}\sigma_{1,t}^2dt\ a_k' = \int_{t_{k-1}^n}^{t_k^n}\sigma_{2,t}^2dt.$$

Obviously $((\xi_k, \xi_k'))_{k=1,\dots,n}$ are independent centered Gaussian random vector with covariance

$$\frac{1}{\sqrt{a_k}\sqrt{a_k'}}\begin{pmatrix}\sqrt{a_k}\sqrt{a_k'} & \int_{t_{k-1}^n}^{t_k^n}\sigma_{1,s}\sigma_{2,s}\rho_s ds\\ \int_{t_{k-1}^n}^{t_k^n}\sigma_{1,s}\sigma_{2,s}\rho_s ds & \sqrt{a_k}\sqrt{a_k'}\end{pmatrix}.$$

Let us introduce the following notation:

$$\mathbf{Q}_{\pm,t}^n = \frac{1}{4}\sum_{k=1}^{[nt]}\sqrt{a_k}\sqrt{a_k'}(\xi_k \pm \xi_k')^2.$$

The proof relies on the following decomposition:

$$\bar{\mathbf{C}}_t^n = \mathbf{Q}_{+,t}^n - \mathbf{Q}_{-,t}^n.$$

Proof of Proposition 2.1. By the independence of $\mathbf{Q}_{+,1}^n$ and $\mathbf{Q}_{-,1}^n$, we obtain that

$$\begin{aligned}\Lambda_n(\lambda) &= \frac{1}{n}\log\mathbb{E}(\exp(\lambda n\bar{\mathbf{C}}_1^n)) = \frac{1}{n}\log\mathbb{E}(\exp(\lambda n(\mathbf{Q}_{+,1}^n - \mathbf{Q}_{-,1}^n)))\\ &= \frac{1}{n}\log\mathbb{E}(\exp(\lambda n\mathbf{Q}_{+,1}^n)) + \frac{1}{n}\log\mathbb{E}(\exp(-\lambda n\mathbf{Q}_{-,1}^n)) := \Lambda_{n,+}(\lambda) + \Lambda_{n,-}(\lambda).\end{aligned}$$

Let us deal with $\Lambda_{n,+}$. We have that

$$\begin{aligned}\Lambda_{n,+}(\lambda) &= \frac{1}{n}\log\mathbb{E}(\exp(\lambda n\mathbf{Q}_{+,1}^n))\\ &= \frac{1}{n}\sum_{k=1}^n P\left(n\frac{\lambda}{2}\left(\sqrt{\int_{t_{k-1}^n}^{t_k^n}\sigma_{1,s}^2ds}\sqrt{\int_{t_{k-1}^n}^{t_k^n}\sigma_{2,s}^2ds} + \int_{t_{k-1}^n}^{t_k^n}\sigma_{1,s}\sigma_{2,s}\rho_s ds\right)\right)\\ &= \int_0^1 P\left(\frac{\lambda}{2}f_n(t)\right)dt,\end{aligned}$$

where P is given in (2.3) and

$$f_n(t) := \sum_{k=1}^{n} 1_{(t_{k-1}^n, t_k^n]}(t) \frac{\sqrt{\int_{t_{k-1}^n}^{t_k^n} \sigma_{1,s}^2 ds}\sqrt{\int_{t_{k-1}^n}^{t_k^n} \sigma_{2,s}^2 ds} + \int_{t_{k-1}^n}^{t_k^n} \sigma_{1,s}\sigma_{2,s}\rho_s ds}{t_k^n - t_{k-1}^n}.$$

Let us remark that we have

$$f_n(t) = \sqrt{\sum_{k=1}^{n} 1_{(t_{k-1}^n, t_k^n]}(t) \frac{\int_{t_{k-1}^n}^{t_k^n} \sigma_{1,s}^2 ds \int_{t_{k-1}^n}^{t_k^n} \sigma_{2,s}^2 ds}{(t_k^n - t_{k-1}^n)^2}} + \sum_{k=1}^{n} 1_{(t_{k-1}^n, t_k^n]}(t) \frac{\int_{t_{k-1}^n}^{t_k^n} \sigma_{1,s}\sigma_{2,s}\rho_s ds}{t_k^n - t_{k-1}^n}.$$

Clearly, $f_n(t)$ is a dt martingale w.r.t. the partially directed filtration $(\mathcal{B}_{\tau_n} := \sigma((t_{k-1}^n, t_k^n], k = 1, \dots, n))_n$. By the convexity of P and Jensen inequality, we obtain that

$$\int_0^1 P\left(\frac{\lambda}{2} f_n(t)\right) dt \le \int_0^1 P\left(\frac{\lambda\sigma_{1,t}\sigma_{2,t}(1+\rho_t)}{2}\right) dt = \Lambda_+(\lambda).$$

On the other hand, by the classical Lebesgue derivation theorem, we have that

$$f_n(t) \longrightarrow f(t) := \sigma_{1,t}\sigma_{2,t} + \sigma_{1,t}\sigma_{2,t}\rho_t, \quad dt - a.e. \text{ on } [0, 1].$$

The continuity of $P : \mathbb{R} \to (-\infty, +\infty]$ gives

$$P\left(\frac{\lambda}{2} f_n(t)\right) \longrightarrow P\left(\frac{\lambda}{2} f(t)\right), \quad dt - a.e. \text{ on } [0, 1].$$

As $P\left(\frac{\lambda}{2} f_n(t)\right) \ge -\frac{|\lambda|}{2}\sigma_{1,t}\sigma_{2,t} \in L^1([0, 1], dt)$, we can apply Fatou's lemma to conclude that

$$\liminf_{n\to\infty} \Lambda_{n,+}(\lambda) = \liminf_{n\to\infty} \int_0^1 P\left(\frac{\lambda}{2} f_n(t)\right) dt \ge \int_0^1 \liminf_{n\to\infty} P\left(\frac{\lambda}{2} f_n(t)\right) dt = \Lambda_+(\lambda).$$

Doing the same calculations with $\Lambda_{n,-}$, we obtain that

$$\Lambda_{n,-}(\lambda) \le \int_0^1 P\left(-\frac{\lambda\sigma_{1,t}\sigma_{2,t}(1-\rho_t)}{2}\right) dt = \Lambda_-(\lambda),$$

and

$$\liminf_{n\to\infty} \Lambda_{n,-}(\lambda) \ge \Lambda_-(\lambda).$$

From below, we conclude that

$$\Lambda_n(\lambda) \le \Lambda_+(\lambda) + \Lambda_-(\lambda) := \Lambda(\lambda),$$

and

$$\liminf_{n\to\infty} \Lambda_n(\lambda) \ge \Lambda(\lambda),$$

which implies that

$$\lim_{n\to\infty} \Lambda_n(\lambda) = \lim_{n\to\infty} (\Lambda_{n,+}(\lambda) + \Lambda_{n,-}(\lambda)) = \Lambda(\lambda),$$

which ends the proof of first part of Proposition 2.1.

For the second part of Proposition 2.1, first we will reduce the study to the case $u_l = 0$ for $l = 1, 2$. Let $\beta = \max(||u_1||_\infty, ||u_2||_\infty)$. Since

$$\left|\mathbf{C}_1^n - \bar{\mathbf{C}}_1^n\right| \le \frac{\beta^2}{n} + \frac{\beta}{n} \sum_{k=1}^{n} \left(\left|\int_{t_{k-1}^n}^{t_k^n} \sigma_{1,s} dB_{1,s}\right| + \left|\int_{t_{k-1}^n}^{t_k^n} \sigma_{2,s} dB_{2,s}\right|\right).$$

For $l = 1, 2$, we have for all $\lambda > 0$ and all $\delta > 0$

$$\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} \left|\int_{t_{k-1}^n}^{t_k^n} \sigma_{l,s} dB_{l,s}\right| \ge \delta\right) \le -\delta\lambda + \frac{\lambda^2}{2n} \int_0^1 \sigma_{l,s}^2 ds.$$

Letting n go to infinity and then λ to infinity we get that for all $\delta > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\left|\int_{t_{k-1}^n}^{t_k^n} \sigma_{l,s} dB_{l,s}\right| \geq \delta\right) = -\infty.$$

By the approximation technique (Theorem 4.2.13 in Dembo and Zeitouni (1998)), we deduce that $\mathbb{P}\left(\mathbf{C}_1^n \in \cdot\right)$ satisfies the same LDP as $\mathbb{P}\left(\bar{\mathbf{C}}_1^n \in \cdot\right)$. Hence we can assume that $u_l = 0$ for $l = 1, 2$.

Now, by inspection of the proof of Theorem 1.1 in Djellout et al. (1999), we deduce that the sequence $\mathbb{P}\left(\mathbf{Q}_{\pm,1}^n \in \cdot\right)$ satisfies the LDP on $\mathbb{R}$ with speed n and rate function given by

$$\Lambda_{\pm}^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda_{\pm}(\lambda)\}.$$

By the independence of the sequences $\mathbf{Q}_{+,1}^n$ and $\mathbf{Q}_{-,1}^n$, and the contraction principle, see Exercise 4.2.7 in Dembo and Zeitouni (1998), we deduce that $\mathbb{P}\left(\bar{\mathbf{C}}_1^n \in \cdot\right)$ satisfies the LDP with rate function

$$\Lambda^*(x) = \inf_{x=x_1-x_2}\left\{\Lambda_+^*(x_1) + \Lambda_-^*(x_2)\right\}.$$

As we have also determined explicitly the logarithm of the moment generating function Λ, the rate function is also given by (2.1).

Proof of Theorem 2.2. The proof of the first part is very similar to Proposition 2.1. It is a consequence of Theorem 1.2 in Djellout et al. (1999) and the contraction principle. For the second part of Theorem 2.2, the same arguments give the large deviation with the rate function

$$I(\gamma) = \inf_{\gamma=\gamma_1-\gamma_2}\{I_+(\gamma_1) + I_-(\gamma_2)\},$$

where

$$I_{\pm}(\gamma) = \int_0^1 P^*\left(\frac{2\dot{\gamma}(t)}{\sigma_{1,t}\sigma_{2,t}(1\pm\rho_t)}\right)dt + \int_0^1 \frac{1}{\sigma_{1,t}\sigma_{2,t}(1\pm\rho_t)} d\gamma^{\perp}.$$

An easy variational calculus gives the identification of the rate function in (2.4).

Proof of Theorem 2.3. As before, we treat only the case $u_l = 0$ for $l = 1, 2$. We have the following decomposition:

$$\mathbf{C}_{\cdot}^n - \langle X_1, X_2\rangle_{\cdot} = (\mathbf{Q}_{+,\cdot}^n - \alpha_{+,\cdot}) - (\mathbf{Q}_{-,\cdot}^n - \alpha_{-,\cdot}),$$

where the definition of $\alpha_{\pm}$ is given in (2.9).

Now using Theorem 1.3 in Djellout et al. (1999), we deduce that $\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(\mathbf{Q}_{\pm,\cdot}^n - \alpha_{\pm,\cdot}) \in \cdot\right)$ satisfies the LDP on $\mathbb{D}_0([0,1])$ with speed b_n^2 and with the good rate function $J_{\pm,m}$ given by

$$J_{\pm,m}(\gamma) = \begin{cases} \displaystyle\int_0^1 \frac{\dot{\gamma}(t)^2}{\sigma_{1,t}^2\sigma_{2,t}^2(1\pm\rho_t)^2} 1_{[t:\sigma_{1,t}\sigma_{2,t}>0]}dt & \text{if } d\gamma \ll \sigma_{1,t}\sigma_{2,t}(1\pm\rho_t)dt \\ +\infty & \text{otherwise.}\end{cases}$$

By the same argument as before, we deduce that $\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(\mathbf{C}_{\cdot}^n - \langle X_1, X_2\rangle_{\cdot}) \in \cdot\right)$ satisfies the LDP on $\mathbb{D}_0([0,1])$ with speed b_n^2 and with the good rate function J_m given by

$$J_m(\gamma) = \inf_{\gamma=\gamma_1-\gamma_2}\left\{J_{+,m}(\gamma_1) + J_{-,m}(\gamma_2)\right\}.$$

An easy calculation gives the identification of the rate function in (2.6).

Acknowledgments

The authors want to thank Professor Liming Wu for fruitful discussions. We are grateful to the anonymous referee for the valuable comments.

References

Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency covariance estimates with noisy and asynchronous financial data. J. Amer. Statist. Assoc. 105 (492), 1504–1517.

Dalalyan, A., Yoshida, N., 2011. Second-order asymptotic expansion for a non-synchronous covariation estimator. Ann. Inst. Henri Poincaré Probab. Stat. 47 (3), 748–789.

Dembo, A., Zeitouni, O., 1998. Large Deviations Techniques and Their Applications, 2nd ed.. In: Applications of Mathematics, vol. 38. Springer-Verlag, New York.
Djellout, H., Guillin, A., Wu, L., 1999. Large and moderate deviations for estimators of quadratic variational processes of diffusions. Stat. Inference Stoch. Process. 2, 195–225.
Hayashi, T., Yoshida, N., 2011. Nonsynchronous covariation process and limit theorems. Stochastic Process. Appl. 121, 2416–2454.
Jiang, H., 2010. Moderate deviations for estimators of quadratic variational process of diffusion with compound Poisson jumps. Statist. Probab. Lett. 80, 1297–1305.
Kanaya, S., Otsu, T., 2012. Large deviations of realized volatility. Stochastic Process. Appl. 122, 546–581.
Mancini, C., 2008. Large deviation principle for an estimator of the diffusion coefficient in a jump-diffusion process. Statist. Probab. Lett. 78, 869–879.
Mancini, C., Gobbi, F., 2012. Identifying the brownian covariation from the co-jumps given discrete observations. Econometric Theory 28 (2), 249–273.

LARGE DEVIATIONS OF THE REALIZED (CO-)VOLATILITY VECTOR

HACÈNE DJELLOUT, ARNAUD GUILLIN, AND YACOUBA SAMOURA

ABSTRACT. Realized statistics based on high frequency returns have become very popular in financial economics. In recent years, different non-parametric estimators of the variation of a log-price process have appeared. These were developed by many authors and were motivated by the existence of complete records of price data. Among them are the realized quadratic (co-)variation which is perhaps the most well known example, providing a consistent estimator of the integrated (co-)volatility when the logarithmic price process is continuous. Limit results such as the weak law of large numbers or the central limit theorem have been proved in different contexts. In this paper, we propose to study the large deviation properties of realized (co-)volatility (i.e., when the number of high frequency observations in a fixed time interval increases to infinity. More specifically, we consider a bivariate model with synchronous observation schemes and correlated Brownian motions of the following form: $dX_{\ell,t} = \sigma_{\ell,t}dB_{\ell,t} + b_\ell(t,\omega)dt$ for $\ell = 1, 2$, where X_ℓ denotes the log-price, we are concerned with the large deviation estimation of the vector $V_t^n(X) = \left(Q_{1,t}^n(X), Q_{2,t}^n(X), C_t^n(X)\right)$ where $Q_{\ell,t}^n(X)$ and $C_t^n(X)$ represente the estimator of the quadratic variational processes $Q_{\ell,t} = \int_0^t \sigma_{\ell,s}^2 ds$ and the integrated covariance $C_t = \int_0^t \sigma_{1,s}\sigma_{2,s}\rho_s ds$ respectively, with $\rho_t = cov(B_{1,t}, B_{2,t})$. Our main motivation is to improve upon the existing limit theorems. Our large deviations results can be used to evaluate and approximate tail probabilities of realized (co-)volatility. As an application we provide the large deviation for the standard dependence measures between the two assets returns such as the realized regression coefficients up to time t, or the realized correlation. Our study should contribute to the recent trend of research on the (co-)variance estimation problems, which are quite often discussed in high-frequency financial data analysis.

AMS 2000 subject classifications: 60F10, 60G42, 62M10, 62G05.

1. Introduction, Model and Notations

In the last decade there has been a considerable development of the asymptotic theory for processes observed at a high frequency. This was mainly motivated by financial applications, where the data, such as stock prices or currencies, are observed very frequently.

Asset returns covariance and its related statistics play a prominent role in many important theoretical as well as practical problems in finance. Analogous to the realized volatility approach, the idea of employing high frequency data in the computation of daily (or lower frequency) covariance between two assets leads to the concept of realized covariance (or covariation). The key role of quantifying integrated (co-)volatilities in portfolio optimization and risk management has stimulated an increasing interest in estimation methods for these models.

It is quite natural to use the asymptotic framework when the number of high frequency observations in a fixed time interval (say, a day) increases to infinity. Thus Barndorff-Nielsen

Date: November 14, 2014.

Key words and phrases. Realised Volatility and covolatility, large deviations, diffusion, discrete-time observation.

and Shephard [6] established a law of large numbers and the corresponding fluctuations for realized volatility, also extended to more general setups and statistics by Barndorff-Nielsen et al. [5] and [4]. Dovonon, Gonçalves, and Meddahi [16] considered Edgeworth expansions for the realized volatility statistic and its bootstrap analog. These results are crucial to explore asymptotic behaviors of realized (co-)volatility, in particular around the center of its distribution. There are also different estimation approaches for the integrated covolatility in multidimensional models and limit theorem, and we can refer to Barndorff-Nielsen et al. [7] and [5] where the authors present, in an unified way, a weak law of large numbers and a central limit theorem for a general estimator, called realized generalized bipower variation.

For related work concerning bivariate case under a non-synchronous sampling scheme, see Hayashi and Yoshida [19], Bibinger [8], Dalalyan and Yoshida [12], see also Aït-Sahalia et al. [1] and the references therein. Estimation of the covariance of log-price processes in the presence of market microstructure noise, we refer to Bibinger and Reiß [9], Robert and Rosenbaum [30], Zhang et al. [37] and [38]. See also Gloter, or Comte et al. [11] for non parametric estimation in the case of a stochastic volatility model.

We model the evolution of an observable state variable by a stochastic process $X_t = (X_{1,t}, X_{2,t}), t \in [0,1]$. In financial applications, X_t can be thought of as the short interest rate, a foreign exchange rate, or the logarithm of an asset price or of a stock index. Suppose both $X_{1,t}$ and $X_{2,t}$ are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and follow an Itô process, namely,

$$\begin{cases} dX_{1,t} &= \sigma_{1,t} dB_{1,t} + b_1(t,\omega)dt \\ dX_{2,t} &= \sigma_{2,t} dB_{2,t} + b_2(t,\omega)dt \end{cases} \tag{1.1}$$

where B_1 and B_2 are standart Brownian motions, with correlation $Corr(B_{1,t}, B_{2,t}) = \rho_t$. We can write $dB_{2,t} = \rho_t dB_{1,t} + \sqrt{1-\rho_t^2} dB_{3,t}$, where $B_1 = (B_{1,t})_{t\in[0,1]}$ and $B_3 = (B_{3,t})_{t\in[0,1]}$ are independent Brownian processes.

We will suppose of course existence and uniqueness of strong solutions, and in what follows, the drift coefficient b_1 and b_2 are assumed to satisfy an uniform linear growth condition and we limit our attention to the case when σ_1, σ_2 and ρ are deterministic functions. The functions σ_ℓ, $\ell = 1,2$ take positive values while ρ takes values in the interval $]-1,1[$.

In this paper, our interest is to estimate the (co-)variation vector

$$[V]_t = ([X_1]_t, [X_2]_t, \langle X_1, X_2\rangle_t)^T \tag{1.2}$$

between two returns in a fixed time period $[0;1]$ when $X_{1,t}$ and $X_{2,t}$ are observed synchronously, $[X_\ell]_t, \ell = 1,2$ represente the quadratic variational process of X_ℓ and $\langle X_1, X_2\rangle_t$ the (deterministic) covariance of X_1 and X_2:

$$[X_\ell]_t = \int_0^t \sigma_{\ell,s}^2 \mathrm{d}s, \qquad \langle X_1, X_2\rangle_t = \int_0^t \sigma_{1,s}\sigma_{2,s}\rho_s \mathrm{d}s.$$

Inference for (1.2) is a well-understood problem if $X_{1,t}$ and $X_{2,t}$ are observed simultaneously. Note that $X_{1,t}$ and $X_{2,t}$ are not observed in continuous time but we have only discrete time observations. Given discrete equally space observation $(X_{1,t_k^n}, X_{2,t_k^n}, k = 1, \cdots, n)$ in the interval $[0,1]$ (with $t_k^n = k/n$), a limit theorem in stochastic processes states that

$$V_t^n(X) = \left(Q_{1,t}^n(X), Q_{2,t}^n(X), C_t^n(X)\right)^T$$

commonly called realized (co-)variance, is a consistent estimator for $[V]_t$, with, for $\ell = 1,2$

$$Q_{\ell,t}^n(X) = \sum_{k=1}^{[nt]} (\Delta_k^n X_\ell)^2 \qquad C_t^n(X) = \sum_{k=1}^{[nt]} (\Delta_k^n X_1)(\Delta_k^n X_2)$$

where $[x]$ denote the integer part of $x \in \mathbb{R}$ and $\Delta_k^n X_\ell = X_{\ell,t_k^n} - X_{\ell,t_{k-1}^n}$.

When the drift $b_\ell(t,\omega)$ is known, we can consider the following variant

$$V_t^n(X-Y) = \left(Q_{1,t}^n(X-Y), Q_{2,t}^n(X-Y), C_t^n(X-Y)\right)^T$$

with for $\ell = 1,2$ and $Y_{\ell,t} := \int_0^t b_\ell(t,\omega)\mathrm{d}t$,

$$Q_{\ell,t}^n(X-Y) = \sum_{k=1}^{[nt]} (\Delta_k^n X_\ell - \Delta_k^n Y_\ell)^2,$$

and

$$C_t^n(X-Y) = \sum_{k=1}^{[nt]} (\Delta_k^n X_1 - \Delta_k^n Y_1)(\Delta_k^n X_2 - \Delta_k^n Y_2).$$

If the drift $b_\ell(t,\omega) := b_\ell(t, X_{1,t}(\omega), X_{2,t}(\omega))$, where $b_\ell(t,x_1,x_2)$ is some deterministic function (a current situation), $X_t = (X_{1,t}, X_{2,t})$ verifies

$$\begin{cases} dX_{1,t} &= \sigma_{1,t} dB_{1,t} &+ b_1(t,X_t)dt \\ dX_{2,t} &= \sigma_{2,t} dB_{2,t} &+ b_2(t,X_t)dt. \end{cases} \tag{1.3}$$

When $b_\ell(t,x)$ is known, and only the sample $X_{t_{k-1}^n} = (X_{1,t_{k-1}^n}, X_{2,t_{k-1}^n}), k = 0 \cdots n$ is observed, we can also consider the following estimator $\tilde{V_t^n}(X) = \left(\tilde{Q_{1,t}^n}(X), \tilde{Q_{2,t}^n}(X), \tilde{C_t^n}(X)\right)^T$ with for $\ell = 1,2$

$$\tilde{Q_{\ell,t}^n}(X) = \sum_{k=1}^{[nt]} \left(\Delta_k^n X_\ell - b_{\ell,t_{k-1}^n}(X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)\right)^2,$$

$$\tilde{C_t^n}(X) = \sum_{k=1}^{[nt]} \left(\Delta_k^n X_1 - b_{1,t_{k-1}^n}(X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)\right)\left(\Delta_k^n X_2 - b_{2,t_{k-1}^n}(X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)\right).$$

In the aforementionned papers, and under quite weak assumptions, it is proved the following consistency

$$V_1^n(X),\, V_1^n(X-Y),\, \tilde{V_1^n}(X) \longrightarrow [V]_1 \qquad a.s.$$

and the corresponding fluctuations

$$\sqrt{n}(V_1^n(X) - [V]_1),\ \sqrt{n}(V_1^n(X-T) - [V]_1),\ \sqrt{n}(\tilde{V_1^n}(X) - [V]_1), \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma).$$

The purpose of this paper is to furnish some further trajectorial estimations about the estimator $V^n_\cdot$, deepening the law of large numbers and central limit theorem. More precisely, we are interested in the estimation of

$$\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(V^n_\cdot(X) - [V]_\cdot) \in A\right),$$

where A is a given domain of deviation, and $(b_n)_{n\geqslant 0}$ is some sequence denoting the scale of the deviation.

When $b_n = 1$, this is exactly the estimation of the central limit theorem. When $b_n = \sqrt{n}$, it becomes the large deviations. And when $1 << b_n << \sqrt{n}$, it is called moderate deviations. In other words, the moderate deviations investigate the convergence speed between the large deviations and central limit theorem.

The large deviations and moderate deviations problems arise in the theory of statistical inference quite naturally. For estimation of unknown parameters and functions, it is first of all important to minimize the risk of wrong decisions implied by deviations of the observed values of estimators from the true values of parameters or functions to be estimated. Such important errors are precisely the subject of large deviation theory. The large deviation and moderate deviation results of estimators can provide us with the rates of convergence and an useful method for constructing asymptotic confidence intervals.

The aim of this paper is then to focus on the large and moderate deviation estimations of the estimators of volatility and co-volatility. Despite the fact that these statistics are nearly 20 years old, there has been remarkably few result in this direction, it is a surprise to us. The answer may however be the following: the usual techniques (such as Gärtner-Ellis method) do not work and a very particular treatment has to be considered for this problem. Recently, however, some papers considered the unidimensional case. Djellout et al. [14] and recently Shin and Otsu [33] obtained the large and moderate deviations for the realized volatility. In the bivariate case Djellout and Yacouba [15], obtained the large and moderate deviations for the realized covolatility. The large deviation for threshold estimator for the constant volatility was established by Mancini [23] in jumps case. And the moderate deviation for threshold estimator for the quadratic variational process was derived by Jiang [20]. Let us mention that the problem of the large deviation for threshold estimator vector, in the presence of jumps, will be considered in a forthcoming paper, consistency, efficience and robustnesse were proved in Mancini and Gobbi [24]. The case of asynchronous sampling scheme, or in the presence of micro-structure noise is also outside the scope of the present paper but are currently under investigations.

Two economically interesting functions of the realized covariance vector are the realized correlation and the realized regression coefficients. In particular, realized regression coefficients are obtained by regressing high frequency returns for one asset on high frequency returns for another asset. When one of the assets is the market portfolio, the result is a realized beta coefficient. A beta coefficient measures the assets systematic risk as assessed by its correlation with the market portfolio. Recent examples of papers that have obtained empirical estimates of realized betas include Andersen, Bollerslev, Diebold and Wu [2], Todorov and Bollerslev [34], Dovonon, Gonçalves and Meddahi [16], Mancini and Gobbi [24].

Let us stress that large deviations for the realized correlation can not be deduced from unidimensional quantities and were thus largely ignored. As an application of our main results, we provide a large and moderate deviation principle for the realized correlation and the realized regression coefficients in some special cases. The realized regression coefficient from regressing is $\beta^n_{\ell,t}(X) = \frac{C^n_t(X)}{Q^n_{\ell,t}(X)}$ which consistently estimates $\beta_{\ell,t} = \frac{C_t}{Q_{\ell,t}}$ and the realized correlation coefficient is $\varrho^n_t(X) = \frac{C^n_t(X)}{\sqrt{Q^n_{1,t}(X)Q^n_{2,t}(X)}}$ which estimates $\varrho_t = \frac{C_t}{\sqrt{Q_{1,t}Q_{2,t}}}$. The application will be based essentially on an application of the delta method, developped by Gao and Zhao ([17]).

As in Djellout et al. [14], Shin and Otsu [33], it should be noted that the proof strategy of Gärtner and Ellis large deviation theorem can not be adapted here int he large deviations case. We will encounter the same technical difficulties as in the papers of Bercu et al. [3] and Bryc and Dembo [10] where they established the large deviation principle for quadratic forms of Gaussian processes. Since we cannot determine the limiting behavior of the cumulant generating function at some boundary point, we will use an other approach based on the results of Najim [26], [27] and [25], where the steepness assumption concerning the cumulant generating function is relaxed. It has to be noted that the form of the large deviations rate function is also original: at the process level, and because of the weak exponential integrability of V^n_t, a correction (or extra) term appears in rate function, a phenomenon first discovered by Lynch and Sethuraman [22].

To be complete, let us now recall some basic definitions of the large deviations theory (c.f [13]). Let $(\lambda_n)_{n\geq 1}$ be a sequence of nonnegative real number such that $\lim_{n\to\infty}\lambda_n = +\infty$. We say that a sequence of a random variables $(M_n)_n$ with topological state space $(S, \mathbb{S})$, where $\mathbb{S}$ is a $\sigma - algebra$ on S, satisfies a large deviation principle with speed λ_n and rate function $I : S \to [0, +\infty]$ if, for each $A \in \mathbb{S}$,

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{\lambda_n}\log\mathbb{P}\Big(M_n \in A\Big) \leq \limsup_{n\to\infty}\frac{1}{\lambda_n}\log\mathbb{P}\Big(M_n\in A\Big) \leq -\inf_{x\in\bar{A}} I(x)$$

where A^o and $\bar{A}$ denote the interior and the closure of A, respectively.

The rate function I is lower semicontinuous, *i.e.* all the sub-level sets $\{x \in S \mid I(x) \leq c\}$ are closed, for $c \geq 0$. If these level sets are compact, then I is said to be a good rate function. When the speed of the large deviation principle correspond to the regime between the central limit theorem and the law of large numbers, we talk of moderate deviation principle.

Notations. *In the whole paper, for any matrix M, M^T and $\|M\|$ stand for the transpose and the euclidean norm of M, respectively. For any square matrix M,* $\det(M)$ *is the determinant of M. Moreover, we will shorten* large deviation principle *by LDP and* moderate deviation principle *by MDP. We denote by $\langle\cdot,\cdot\rangle$ the usual scalar product. For any process Z_t, $\Delta^n_k Z$ stands for the increment $Z_{t^n_k} - Z_{t^n_{k-1}}$. In addition, for a sequence of random variables $(Z_n)_n$ on $\mathbb{R}^{d\times p}$, we say that $(Z_n)_n$ converges $(\lambda_n)-$superexponentially fast in probability to some random variable Z if, for all $\delta > 0$,*

$$\limsup_{n\to\infty}\frac{1}{\lambda_n}\log\mathbb{P}\Big(\;\|Z_n - Z\| > \delta\Big) = -\infty.$$

This exponential convergence *with speed* λ_n *will be shortened as*

$$Z_n \quad \xrightarrow[\lambda_n]{\text{superexp}} Z.$$

The article is arranged in three upcoming sections and an appendix comprising some theorems used intensively in the paper, we have included them here for completeness. Section 2 is devoted to our main results on the LDP and MDP for the (co-)volatility vector. In Section 3, we deduce applications for the realized correlation and the realized regression coefficients, when σ_ℓ, for $\ell = 1, 2$ are constants. In section 4, we give the proof of these theorems.

2. Main results

Let $X_t = (X_{1,t}, X_{2,t})$ be given by (1.1), and $Y_t = (Y_{1,t}, Y_{2,t})$ where for $\ell = 1, 2 \quad Y_{\ell,t} := \int_0^t b_\ell(t, \omega)dt$. We introduce the following conditions

(B) for $\ell = 1, 2$ $b(\cdot, \cdot) \in L^\infty(dt \otimes \mathbb{P})$

(LDP) Assume that for $\ell = 1, 2$

- $\sigma^2_{\ell,t}(1 - \rho_t^2)$ and $\sigma_{1,t}\sigma_{2,t}(1 - \rho_t^2) \in L^\infty([0, 1], dt)$.
- the functions $t \to \sigma_{\ell,t}$ and $t \to \rho_t$ are continuous.

(MDP) Assume that for $\ell = 1, 2$

- $\sigma^2_{\ell,t}(1 - \rho_t^2)$ and $\sigma_{1,t}\sigma_{2,t}(1 - \rho_t^2) \in L^2([0, 1], dt)$.
- Let $(b_n)_{n\geqslant 1}$ be a sequence of positive numbers such that

$$\begin{aligned} & b_n \xrightarrow[n\to\infty]{} \infty \quad \text{and} \quad \frac{b_n}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0 \\ & \text{and for} \quad \ell = 1, 2 \quad \sqrt{n} b_n \max_{1\leqslant k\leqslant n} \int_{(k-1)/n}^{k/n} \sigma^2_{\ell,t} \mathrm{d}t \xrightarrow[n\to\infty]{} 0. \end{aligned} \tag{2.1}$$

We introduce the following function, which will play a crucial role in the calculation of the moment generating function: for $-1 < c < 1$ let for any $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$

$$P_c(\lambda) := \begin{cases} -\dfrac{1}{2}\log\left(\dfrac{(1-2\lambda_1(1-c^2))(1-2\lambda_2(1-c^2)) - (\lambda_3(1-c^2)+c)^2}{1-c^2}\right) \\ \qquad\qquad\qquad\qquad if \qquad \lambda \in \mathcal{D} \\ +\infty, \quad otherwise \end{cases} \tag{2.2}$$

where

$$\mathcal{D}_c = \left\{\lambda \in \mathbb{R}^3,\ \max_{\ell=1,2} \lambda_\ell < \frac{1}{2(1-c^2)} \text{ and } \prod_{\ell=1}^{2} \left(1 - 2\lambda_\ell(1-c^2)\right) > \left(\lambda_3(1-c^2)+c\right)^2\right\}. \tag{2.3}$$

Let us present now the main results.

2.1. **Large deviation.** Our first result is about the large deviation of $V_1^n(X)$, i.e. at fixed time.

Theorem 2.1. *Let $t = 1$ be fixed.*

(1) *For every $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}(\exp(n \langle \lambda, V_1^n(X-Y)\rangle)) = \Lambda(\lambda) := \int_0^1 P_{\rho_t}(\lambda_1\sigma_{1,t}^2, \lambda_2\sigma_{2,t}^2, \lambda_3\sigma_{1,t}\sigma_{2,t})\mathrm{d}t,$$

where the function P_c is given in (2.2).

(2) *Under the conditions* **(LDP)** *and* **(B)** *, the sequence $V_1^n(X)$ satisfies the LDP on $\mathbb{R}^3$ with speed n and with the good rate function given by the legendre transformation of Λ, that is*

$$I_{ldp}(x) = \sup_{\lambda\in\mathbb{R}^3} \left(\langle \lambda, x\rangle - \Lambda(\lambda)\right). \tag{2.4}$$

Let us consider the case where diffusion and correlation coefficients are constant, the rate function being easier to read (see also [32] in the purely Gaussian case, i.e. $b = 0$). Before that let us introduce the function P_c^* which is the Legendre transformation of P_c given in (2.2), for all $x = (x_1, x_2, x_3)$

$$P_c^*(x) := \begin{cases} \log\left(\dfrac{\sqrt{1-c^2}}{\sqrt{x_1x_2 - x_3^2}}\right) - 1 + \dfrac{x_1 + x_2 - 2cx_3}{2(1-c^2)} \\ \qquad\qquad\qquad\qquad if \qquad x_1 > 0,\ x_2 > 0,\ x_1x_2 > x_3^2 \\ +\infty, \quad otherwise. \end{cases} \tag{2.5}$$

Corollary 2.2. *We assume that for $\ell = 1, 2$ σ_ℓ and ρ are constants. Under the condition* **(B)***, we obtain that $V_1^n(X)$ satisfies the LDP on $\mathbb{R}^3$ with speed n and with the good rate function I_{ldp}^V given by*

$$I_{ldp}^V(x_1, x_2, x_3) = P_\rho^*\left(\frac{x_1}{\sigma_1^2}, \frac{x_2}{\sigma_2^2}, \frac{x_3}{\sigma_1\sigma_2}\right), \tag{2.6}$$

where P_c^* is given in (2.5).

Now, we shall extend the Theorem 2.1 to the process-level large deviations, i.e. for trajectories $(V_t^n(X))_{0\le t\le 1}$, which is interesting from the viewpoint of non-parametric statistics.

Let $\mathcal{B}V([0,1], \mathbb{R}^3)$ (shortened in $\mathcal{B}V$) be the space of functions of bounded variation on $[0,1]$. We identify $\mathcal{B}V$ with $\mathcal{M}_3([0,1])$, the set of vector measures with value in $\mathbb{R}^3$. This is done in the usual manner: to $f \in \mathcal{B}V$ there corresponds μ^f caracterized by $\mu^f([0,t]) = f(t)$. Up to this identification, $\mathcal{C}_3([0,1])$ the set of $\mathbb{R}^3$-valued continuous bounded functions on $[0,1])$, is the topological dual of $\mathcal{B}V$. We endow $\mathcal{B}V$ with the weak-* convergence topology $\sigma(\mathcal{B}V, \mathcal{C}_3([0,1]))$ (shortened σ_w) and with the associated Borel $\sigma-$field $\mathcal{B}_w$. Let $f \in \mathcal{B}V$ and μ^f the associated measure in $\mathcal{M}_3([0,1])$. Consider the Lebesgue decomposition of μ^f, $\mu^f = \mu_a^f + \mu_s^f$ where μ_a^f denotes the absolutely continous part of μ^f with respect to dx and μ_s^f its singular part. We denote by $f_a(t) = \mu_a^f([0,t])$ and by $f_s(t) = \mu_s^f([0,t])$.

Theorem 2.3. *Under the conditions* **(LDP)** *and* **(B)**, *the sequence* $V^n_\cdot(X)$ *satisfies the LDP on* $\mathcal{B}V$ *with speed* n *and with the rate function* J_{ldp} *given for any* $f=(f_1,f_2,f_3)\in\mathcal{B}V$ *by*

$$\begin{aligned} J_{ldp}(f) &= \int_0^1 P^*_{\rho_t}\left(\frac{f'_{1,a}(t)}{\sigma^2_{1,t}},\frac{f'_{2,a}(t)}{\sigma^2_{2,t}},\frac{f'_{3,a}(t)}{\sigma_{1,t}\sigma_{2,t}}\right)\mathrm{d}t \\ &\quad+\int_0^1 \frac{\sigma^2_{2,t}f'_{1,s}(t)+\sigma^2_{1,t}f'_{2,s}(t)-2\rho_t\sigma_{1,t}\sigma_{2,t}f'_{3,s}(t)}{2\sigma^2_{1,t}\sigma^2_{2,t}(1-\rho_t^2)}1_{[t;f'_{1,s}>0,f'_{2,s}>0,(f'_{3,s})^2<f'_{1,s}f'_{2,s}]}\,\mathrm{d}\theta(t), \end{aligned}$$

where P^*_c *is given in (2.5) and* θ *is any real-valued nonnegative measure with respect to which* μ^f_s *is absolutely continuous and* $f'_s=d\mu^f_s/d\theta=(f'_{1,s},f'_{2,s},f'_{3,s})$.

Remark 2.1. *Note that the definition of* f'_s *is* $\theta-$*dependent. However, by homogeneity,* J_{ldp} *does not depend upon* θ. *One can choose* $\theta=|f_{1,s}|+|f_{2,s}|+|f_{3,s}|$, *with* $|f_{i,s}|=f^+_{i,s}+f^-_{i,s}$, *where* $f_{i,s}=f^+_{i,s}-f^-_{i,s}$ *by the Hahn-Jordan decomposition.*

Remark 2.2. *As stated above, the problem of the LDP for* $Q^n_{\ell,\cdot}(X)$ *and* $C^n_\cdot(X)$ *was alreay studied by Djellout et al.* [14] *and* [15], *and the rate function is given explicitly in the last case. This is the first time that the LDP is investigated for the vector of the (co-)volatility.*

Remark 2.3. *By using the contraction principle, and if* σ_ℓ *is strictly positive, we may find back the result of* [14], *i.e. that* $Q^n_{\ell,\cdot}$ *satisfies a LDP with speed* n *and rate function*

$$J^{\sigma_\ell}_{ldp}(f)=\int_0^1\mathcal{P}^*\left(\frac{f'_a(t)}{\sigma^2_{\ell,t}}\right)dt+\frac{1}{2}\int_0^1\frac{1}{\sigma^2_{\ell,t}}d|f_s|(t)$$

where $\mathcal{P}^*(x)=\frac{1}{2}(x-1-\log(x))$ *when* x *is positive and infinite if non positive, using the same notation as in the theorem (with* $\theta=|f_s|$*). One may also obtain the LDP for* $C^n_\cdot$ *by the contraction principle, recovering the result of Djellout-Yacouba* [15] *(see there for the quite explicit complicated rate function).*

Remark 2.4. *The continuity assumptions in* **(LDP)** *of* $\sigma_{\ell,\cdot}$ *and* $\rho_\cdot$ *is not necessary, but in this case we have to consider another strategy of the proof, more technical and relying on Dawson-Gärtner type theorem, which moreover does not enable to get other precision on the rate function that the fact it is a good rate function.*
However it is not hard to adapt our proof to the case where $\sigma_{\ell,\cdot}$ *and* $\rho_\cdot$ *have only a finite number of discontinuity points (of the first type). This can be done by applying the previous theorem to each subinterval where all functions are continuous and using the independence of the increments of* $V^n_t(X-Y)$.

2.2. **Moderate deviation.** Let us now considered the intermediate scale between the central limit theorem and the law of large numbers.

Theorem 2.4. *For t=1 fixed. Under the conditions* **(MDP)** *and* **(B)** , *the sequence*

$$\frac{\sqrt{n}}{b_n}\left(V^n_1(X)-[V]_1\right)$$

satisfies the LDP on $\mathbb{R}^3$ *with speed* b_n^2 *and with the rate function given by*

$$I_{mdp}(x) = \sup_{\lambda\in\mathbb{R}^3}\left(\langle\lambda, x\rangle - \frac{1}{2}\langle\lambda, \Sigma_1\cdot\lambda\rangle\right) = \frac{1}{2}\left\langle x, \Sigma_1^{-1}\cdot x\right\rangle \tag{2.7}$$

with

$$\Sigma_1 = \begin{pmatrix} \int_0^1\sigma_{1,t}^4\mathrm{d}t & \int_0^1\sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2\mathrm{d}t & \int_0^1\sigma_{1,t}^3\sigma_{2,t}\rho_t\mathrm{d}t \\ \int_0^1\sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2\mathrm{d}t & \int_0^1\sigma_{2,t}^4\mathrm{d}t & \int_0^1\sigma_{1,t}\sigma_{2,t}^3\rho_t\mathrm{d}t \\ \int_0^1\sigma_{1,t}^3\sigma_{2,t}\rho_t\mathrm{d}t & \int_0^1\sigma_{1,t}\sigma_{2,t}^3\rho_t\mathrm{d}t & \int_0^1\frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t \end{pmatrix}.$$

Remark 2.5. *If for some* $p>2$, $\sigma_{1,t}^2$, $\sigma_{2,t}^2$ *and* $\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2)\in L^p([0,1])$ *and* $b_n = O(n^{\frac{1}{2}-\frac{1}{p}})$, *the condition* (2.1) *in* **(MDP)** *is verified.*

Let $\mathcal{H}$ be the banach space of $\mathbb{R}^3$-valued right-continuous-left-limit non decreasing functions γ on $[0,1]$ with $\gamma(0)=0$, equipped with the uniform norm and the $\sigma-$field $\mathcal{B}^s$ generated by the coordinate $\{\gamma(t), 0\leqslant t\leqslant 1\}$.

Theorem 2.5. *Under the conditions* **(MDP)** *and* **(B)**, *the sequence*

$$\frac{\sqrt{n}}{b_n}\left(V_\cdot^n(X) - [V]_\cdot\right)$$

satisfies the LDP on $\mathcal{H}$ *with speed* b_n^2 *and with the rate function given by*

$$J_{mdp}(\phi) = \begin{cases} \displaystyle\int_0^1 \frac{1}{2}\left\langle\dot\phi(t), \Sigma_t^{-1}\cdot\dot\phi(t)\right\rangle dt & \text{if} \quad \phi\in\mathcal{AC}_0([0,1]) \\ +\infty, \qquad \text{otherwise}, \end{cases} \tag{2.8}$$

where

$$\Sigma_t = \begin{pmatrix} \sigma_{1,t}^4 & \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 & \sigma_{1,t}^3\sigma_{2,t}\rho_t \\ \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 & \sigma_{2,t}^4 & \sigma_{1,t}\sigma_{2,t}^3\rho_t \\ \sigma_{1,t}^3\sigma_{2,t}\rho_t & \sigma_{1,t}\sigma_{2,t}^3\rho_t & \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2) \end{pmatrix}$$

is invertible and Σ_t^{-1} *his inverse such that*

$$\Sigma_t^{-1} = \frac{1}{\det(\Sigma_t)}\begin{pmatrix} \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^6(1-\rho_t^2) & \frac{1}{2}\sigma_{1,t}^4\sigma_{2,t}^4\rho_t^2(1-\rho_t^2) & -\sigma_{1,t}^3\sigma_{2,t}^5\rho_t(1-\rho_t^2) \\ \frac{1}{2}\sigma_{1,t}^4\sigma_{2,t}^4\rho_t^2(1-\rho_t^2) & \frac{1}{2}\sigma_{1,t}^6\sigma_{2,t}^2(1-\rho_t^2) & -\sigma_{1,t}^5\sigma_{2,t}^3\rho_t(1-\rho_t^2) \\ -\sigma_{1,t}^3\sigma_{2,t}^5\rho_t(1-\rho_t^2) & -\sigma_{1,t}^5\sigma_{2,t}^3\rho_t(1-\rho_t^2) & \sigma_{1,t}^4\sigma_{2,t}^4(1-\rho_t^4) \end{pmatrix},$$

$$\textit{with} \qquad \det(\Sigma_t) = \frac{1}{2}\sigma_{1,t}^6\sigma_{2,t}^6(1-\rho_t^2)^3,$$

and $\mathcal{AC}_0 = \{\phi:[0,1]\to\mathbb{R}^3$ *is absolutely continuous with* $\phi(0)=0\}$.

Let us note once again that it is the first time the MDP is considered for the vector of (co)-volatility.

In the previous results, we have imposed the boundedness of $b(t,\omega)$ which allows us to reduce quite easily the LDP and MDP of $V^n(X)$to those of $V^n(X-Y)$ (no drift case). It is very natural to ask whether they continue to hold under a Lipchitzian condition or more generally linear growth condition of the drift $b(t,x)$, rather than the boundedness. This is the object of the following

Theorem 2.6. *Let $X_t = (X_{1,t}, X_{2,t})$ be given by* (1.3)*, with $(X_{1,0}, X_{2,0})$ bounded. We assume that the drift b_ℓ satisfies the following uniform linear growth condition: $\forall s,t \in [0,1], x,y \in \mathbb{R}^2$*

$$|b_\ell(t,x) - b_\ell(s,y)| \leqslant C[1 + \|x-y\| + \eta(|t-s|)(\|x\| + \|y\|)], \tag{2.9}$$

where $C > 0$ is a constant and $\eta : [0,\infty) \to [0,\infty)$ is a continuous non-decreasing function with $\eta(0) = 0$.

(1) *Under the condition* **(LDP)***, the sequence $\tilde{V}^n_\cdot(X)$ satisfies the LDPs in Theorem 2.1 and Theorem 2.3.*
(2) *Under the condition* **(MDP)***, the sequence $\dfrac{\sqrt{n}}{b_n}(\tilde{V}^n_\cdot(X) - [V]_\cdot)$ satisfies the MDPs in Theorem 2.4 and Theorem 2.5.*

As it can be remarked, the LDP and the MDP are established here for $\tilde{V}^n$ instead of V^n. If we conjecture that the MDP may still be valid in this case with V^n, we do not believe it should be the case for the LDP, and it is thus a challenging and interesting question to establish the LDP in this case for V^n. However for the statistical purpose, if the drift b is known, the previous result is perfectly satisfactory.

3. Applications: Large deviations for the realized correlation and the realized regression coefficients

In this section we apply our results to obtain the LDP and MDP for the standard dependence measures between the two assets returns such as the realized regression coefficients up to time 1, $\beta_{\ell,1} = \frac{C_1}{Q_{\ell,1}}$ for $\ell = 1,2$ and the realized correlation $\varrho_1 = \frac{C_1}{\sqrt{Q_{1,1}Q_{2,1}}}$ which are estimated by $\beta^n_{\ell,1}(X) = \frac{C^n_1(X)}{Q^n_{\ell,1}(X)}$ and $\varrho^n_1(X) = \frac{C^n_1(X)}{\sqrt{Q^n_{1,1}(X)Q^n_{2,1}(X)}}$ respectively. To simplify the argument, we focus in the case where σ_ℓ for $\ell = 1,2$ are constants and we denote $\varrho := \int_0^1 \rho_t dt$. The consistency and the central limit theorem for these estimators were already studied see for example Mancini and Gobbi [24]. Up to our knowledge, however no results are known for the large and moderate deviation principle.

3.1. Correlation coefficient.

Proposition 3.1. *Let for $\ell = 1,2$, σ_ℓ are constants and $\varrho := \int_0^1 \rho_t dt$. Under the conditions* **(LDP)** *and* **(B)***, the sequence $\varrho^n_1(X)$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given by*

$$I^\varrho_{ldp}(u) = \inf_{\{(x,y,z)\in\mathbb{R}^3 : u = \frac{z}{\sqrt{xy}}\}} I_{ldp}(x,y,z)$$

where I_{ldp} is given in (2.4).

Once again, let us specify the rate function in the case of constant correlation.

Corollary 3.2. *We suppose that for $\ell = 1, 2$, σ_ℓ and ρ are constant. Under the condition* **(B)**, *we obtain that $\varrho_1^n(X)$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given by*

$$I_{ldp}^{\rho}(u) = \begin{cases} \log\left(\dfrac{\sqrt{1-\rho u}}{\sqrt{1-\rho^2}\sqrt{1-u^2}}\right) - 1 + \dfrac{\sigma_1^4 + \sigma_2^4 - 2\rho\sigma_1^2\sigma_2^2 u}{2\sigma_1^2\sigma_2^2(1-\rho u)}, & -1 < u < 1 \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.1}$$

As the reader can imagine from the rate function expression, it is quite a simple application of the contraction principle starting from the LDP of the realized (co)-volatility. As will be seen from the proof, in this case, the MDP is harder to establish and requires a more subtle technology: large deviations for the delta-method.

Proposition 3.3. *Let for $\ell = 1, 2$, σ_ℓ are constants and $\varrho := \int_0^1 \rho_t dt$. Under the conditions* **(MDP)** *and* **(B)**, *the sequence $\dfrac{\sqrt{n}}{b_n}(\varrho_1^n(X) - \varrho)$ satisfies the LDP on $\mathbb{R}$ with speed b_n^2 and with the rate function given by*

$$I_{mdp}^{\varrho}(u) = \inf_{\{(x,y,z)\in\mathbb{R}^3 : u = \frac{z}{\sigma_1\sigma_2} - \varrho\frac{\sigma_1^2 y + x\sigma_2^2}{2\sigma_1^2\sigma_2^2}\}} I_{mdp}(x, y, z)$$

where I_{mdp} is given in (2.7).

Corollary 3.4. *We suppose that for $\ell = 1, 2$, σ_ℓ and ρ are constant. Under the condition* **(B)**, *we obtain that $\dfrac{\sqrt{n}}{b_n}(\varrho_1^n(X) - \rho)$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given for all $u \in \mathbb{R}$ by*

$$I_{mdp}^{\rho}(u) = \frac{2u^2}{(1-\rho^2)^2}. \tag{3.2}$$

3.2. **Regression coefficient.** The strategy initiated for the correlation coefficient is even simpler in the case of regression coefficient.

Proposition 3.5. *Let for $\ell = 1, 2$, σ_ℓ are constants . Under the conditions* **(LDP)** *and* **(B)**, *for $\ell = 1$ or 2, the sequence $\beta_{l,1}^n(X)$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given by*

$$I_{ldp}^{\beta_{\ell,1}}(u) = \inf_{\{(x_1,x_2,x_3)\in\mathbb{R}^3 : u = \frac{x_3}{x_\ell}\}} I_{ldp}(x_1, x_2, x_3)$$

where I_{ldp} is given in (2.4).

Once again, this Proposition is a simple application of the contraction principle. Let us specify the rate function in the case of constant correlation.

Corollary 3.6. *We suppose that for $\ell = 1, 2$, σ_ℓ and ρ are constant. Under the condition* **(B)***, we obtain that $\beta^n_{l,1}(X)$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given for $\iota = 1, 2$ with $\ell \neq \iota$ and for all $u \in \mathbb{R}$ by*

$$I^{\beta_l}_{ldp}(u) = \frac{1}{2}\log\left(1 + \frac{(\sigma_\ell u - \rho\sigma_\iota)^2}{\sigma_\iota^2(1-\rho^2)}\right). \tag{3.3}$$

We may also consider the MDP.

Proposition 3.7. *Let for $\ell = 1, 2$, σ_ℓ are constants and $\varrho := \int_0^1 \rho_t dt$. Under the conditions* **(MDP)** *and* **(B)** *and for $\ell, \iota \in \{1, 2\}$ with $\ell \neq \iota$, the sequence $\frac{\sqrt{n}}{b_n}(\beta^n_{\ell,1}(X) - \varrho\frac{\sigma_\iota}{\sigma_\ell})$ satisfies the LDP on $\mathbb{R}$ with speed b_n^2 and with the rate function given by*

$$I^{\beta_{\ell,1}}_{mdp}(u) = \inf_{\{(x,y,z)\in\mathbb{R}^3 : u = \frac{z}{\sigma_\ell^2} - \varrho\frac{\sigma_\iota}{\sigma_\ell^3}x\}} I_{mdp}(x, y, z)$$

where I_{mdp} is given in (2.7).

Corollary 3.8. *We suppose that for $\ell = 1, 2$, σ_ℓ and ρ are constant. Under the condition* **(B)** *and for $\ell, \iota \in \{1, 2\}$ with $\ell \neq \iota$, we obtain that $\frac{\sqrt{n}}{b_n}(\beta^n_{\ell,1}(X) - \rho\frac{\sigma_\iota}{\sigma_\ell})$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given for all $u \in \mathbb{R}$ by*

$$I^{\beta^c_{\ell,1}}_{mdp}(u) = \frac{2\sigma_\ell^2 u^2}{\sigma_\iota^2(1-\rho^2)}. \tag{3.4}$$

4. Proof

Let us say a few words on our strategy of proof. As the reader may have guessed, one of the important step is first to consider the no-drift case, where we have to deal with non homogenous quadratic forms of Gaussian processes (in the vector case). In these non essentially smooth case (in the terminology of Gärtner-Ellis), we will use (after some technical approximations) powerful recent results of Najim [26]. In a second step, we see how to reduce the general case to the no-drift case.

4.1. Proof of Theorem 2.1.

Lemma 4.1. *If (ξ, ξ') are independent centered Gaussian random vector with covariance*

$$\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}, -1 < c < 1.$$

Then for all $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$

$$\log \mathbb{E} \exp\left(\lambda_1\xi^2 + \lambda_2\xi'^2 + \lambda_3\xi\xi'\right) = P_c(\lambda_1, \lambda_2, \lambda_3),$$

where the function P_c is given in in (2.2).

Proof : Elementary.

Lemma 4.2. *Let $X_t = (X_{1,t}, X_{2,t})$ given (1.1) and $Y_t = (Y_{1,t}, Y_{2,t})$ where for $\ell = 1, 2$ $\quad Y_{\ell,t} := \int_0^t b_\ell(t, \omega)dt$. We have for every $\lambda \in \mathbb{R}^3$*

$$\Lambda_n(\lambda) := \frac{1}{n}\log \mathbb{E}\left(\exp\left(n\left\langle \lambda, V_1^n(X - Y)\right\rangle\right)\right) \leqslant \Lambda(\lambda) := \int_0^1 P_{\rho_t}(\lambda_1\sigma^2_{1,t}, \lambda_2\sigma^2_{2,t}, \lambda_3\sigma_{1,t}\sigma_{2,t})\mathrm{d}t,$$

where the function P_c is given in in (2.2), and

$$\lim_{n\to\infty} \Lambda_n(\lambda) = \Lambda(\lambda).$$

Proof : For $\ell = 1, 2$, we have

$$Q^n_{\ell,t}(X-Y) = \sum_{k=1}^{[nt]} a_{\ell,k}\xi^2_{\ell,k} \qquad \text{and} \quad C^n_t(X-Y) = \sum_{k=1}^{[nt]} \sqrt{a_{1,k}}\sqrt{a_{2,k}}\xi_{1,k}\xi_{2,k}$$

where

$$\xi_{\ell,k} := \frac{\int_{t_{k-1}}^{t_k} \sigma_{\ell,s}\mathrm{d}B_{\ell,s}}{\sqrt{a_{\ell,k}}} \quad \text{and} \quad a_{\ell,k} := \int_{t^n_{k-1}}^{t^n_k} \sigma^2_{\ell,s}\mathrm{d}s. \tag{4.1}$$

Obviously $((\xi_{1,k}, \xi_{2,k}))_{k=1\cdots n}$ are independent centered Gaussian random vector with covariance matrix

$$\begin{pmatrix} 1 & c^n_k \\ c^n_k & 1 \end{pmatrix}$$

where

$$c^n_k := \frac{\vartheta^n_k}{\sqrt{a_{1,k}}\sqrt{a_{2,k}}} \qquad \text{and} \quad \vartheta^n_k := \int_{t^n_{k-1}}^{t^n_k} \sigma_{1,s}\sigma_{2,s}\rho_s\mathrm{d}s. \tag{4.2}$$

We use the lemma 4.1 and the martingale convergence theorem (or the classical Lebesgue derivation theorem) to get the final assertions (see for example [14, p.204] for details).

Proof Theorem 2.1.

(1) It is contained in lemma 4.2.

(2) We shall prove it in three steps.

<u>*Part 1*</u>. At first, we consider that the drift $b_\ell = 0$. In this case $V^n_t(X) = V^n_t(X-Y)$. We recall that since $B_{2,t} = \rho_t dB_{1,t} + \sqrt{1-\rho^2_t}dB_{3,t}$, we may rewrite (1.1) as

$$\begin{cases} dX_{1,t} &= \sigma_{1,t}dB_{1,t} \\ dX_{2,t} &= \sigma_{2,t}(\rho_t dB_{1,t} + \sqrt{1-\rho^2_t}dB_{3,t}) \end{cases} \tag{4.3}$$

Using the approximation Lemma in [13], we shall prove that

$$V^n_1(X-Y) = \left(\sum_{k=1}^n (\Delta^n_k X_1)^2, \sum_{k=1}^n (\Delta^n_k X_2)^2, \sum_{k=1}^n (\Delta^n_k X_1)(\Delta^n_k X_2)\right)^T$$

will satisfy the same LDPs as

$$W_1^n := \begin{pmatrix} \frac{1}{n}\sum_{k=1}^n \sigma^2_{1,\frac{k-1}{n}} N^2_{1,k} \\ \frac{1}{n}\sum_{k=1}^n \left(\sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}} N_{1,k} + \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho^2_{\frac{k-1}{n}}}N_{3,k}\right)^2 \\ \frac{1}{n}\sum_{k=1}^n \sigma_{1,\frac{k-1}{n}} N_{1,k}\left(\sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}} N_{1,k} + \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho^2_{\frac{k-1}{n}}}N_{3,k}\right) \end{pmatrix},$$

where $N_{\ell,k} := \int_{t^n_{k-1}}^{t^n_k} \sqrt{n}\mathrm{d}B_{\ell,s}$, for $\ell = 1, 3$.

Let us first focus on the LDP of W_1^n. We will use Najim result (see Lemma 5.1) to prove that.

It is easy to see that W_1^n can be reritten as

$$W_1^n = \frac{1}{n}\sum_{k=1}^n F\left(\frac{k-1}{n}\right) Z_k$$

where

$$F\left(\frac{k}{n}\right) = \begin{pmatrix} f_1(\frac{k}{n}) \\ f_2(\frac{k}{n}) \\ f_3(\frac{k}{n}) \end{pmatrix} = \begin{pmatrix} \sigma^2_{1,\frac{k}{n}} & 0 & 0 \\ \sigma^2_{2,\frac{k}{n}}\rho^2_{\frac{k}{n}} & \sigma^2_{2,\frac{k}{n}}(1-\rho^2_{\frac{k}{n}}) & 2\sigma^2_{2,\frac{k}{n}}\rho_{\frac{k}{n}}\sqrt{1-\rho^2_{\frac{k}{n}}} \\ \sigma_{1,\frac{k}{n}}\sigma_{2,\frac{k}{n}}\rho_{\frac{k}{n}} & 0 & \sigma_{1,\frac{k}{n}}\sigma_{2,\frac{k}{n}}\sqrt{1-\rho^2_{\frac{k}{n}}} \end{pmatrix} \tag{4.4}$$

and

$$Z_j = \left(N^2_{1,j},\ N^2_{3,j},\ N_{1,j}N_{3,j}\right)^T. \tag{4.5}$$

Obviously $(N_{1,k}, N_{3,k})_{k=1\cdots n}$ are independent centered Gaussian random vector with identity covariance matrix.

For the LDP of W_1^n we will use Lemma 5.1, in the case where $\mathcal{X} := [0,1]$ and $R(dx)$ is the Lebesgue measure on $[0,1]$ and $x_i^n := i/n$. One can check that, in this situation, Assumptions (N-2) hold true. The random variables $(Z_k)_{k=1,\cdots,n}$ are independent and identically distributed. By the definition of Z_k, the Assumptions (N-1) hold true also.

So W_1^n satisfies the LDP on $\mathbb{R}^3$ with speed n and with the good rate function given by for all $x \in \mathbb{R}^3$

$$I(x) = \sup_{\lambda\in\mathbb{R}^3}\left(\langle\lambda, x\rangle - \int_0^1 L\left(\sum_{j=1}^3 \lambda_i \cdot f_i(t)\right)dt\right),$$

with

$$\sum_{i=1}^3 \lambda_i f_i(t) = \begin{pmatrix} \lambda_1\sigma^2_{1,t} + \lambda_2\sigma^2_{2,t}\rho^2_t + \lambda_3\sigma_{1,t}\sigma_{2,t}\rho_t \\ \lambda_2\sigma^2_{2,t}(1-\rho^2_t) \\ 2\lambda_2\sigma^2_{2,t}\rho_t\sqrt{1-\rho^2_t} + \lambda_3\sigma_{1,t}\sigma_{2,t}\sqrt{1-\rho^2_t} \end{pmatrix}^T,$$

and for $\lambda \in \mathbb{R}^3$

$$L(\lambda) := \log\mathbb{E}\exp\langle\lambda, Z_1\rangle = P_0(\lambda_1, \lambda_2, \lambda_3),$$

where P_0 is given in (2.2). In this cas it takes a simpler form wich we recall here:

$$P_0(\lambda_1, \lambda_2, \lambda_3) = -\frac{1}{2} \log \left((1-2\lambda_1)(1-2\lambda_2) - \lambda_3^2 \right).$$

An easy calculation gives us that

$$\int_0^1 P_0 \left(\sum_{j=1}^{3} \lambda_i \cdot f_i(t) \right) dt = \int_0^1 P_{\rho_t} \left(\lambda_1 \sigma_{1,t}^2, \lambda_2 \sigma_{2,t}^2, \lambda_3 \sigma_{1,t} \sigma_{2,t} \right) dt,$$

so

$$I(x) = I_{ldp}(x),$$

where I_{ldp} is given in (2.4).

<u>*Part 2*</u>. Now we shall prove that V_1^n and W_1^n satisfy the same LDPs, by means of the approximation Lemma in [14]. We have to prove that

$$V_1^n(X-Y) - W_1^n \quad \underset{n}{\overset{\text{superexp}}{\longrightarrow}} 0.$$

We do this element by element. We will only consider one element, the other terms can be dealt with in the same way. We have to prove that for $q = 1, 2, 3$

$$\mathcal{R}_{q,1}^n \quad \underset{n}{\overset{\text{superexp}}{\longrightarrow}} 0, \tag{4.6}$$

where

$$\mathcal{R}_{1,t}^n := \sum_{k=1}^{[nt]} (\Delta_k^n X_1)^2 - \frac{1}{n} \sum_{k=1}^{[nt]} \sigma_{1,\frac{k-1}{n}}^2 N_{1,k}^2, \tag{4.7}$$

$$\mathcal{R}_{2,t}^n := \sum_{k=1}^{[nt]} (\Delta_k^n X_2)^2 - \frac{1}{n} \sum_{k=1}^{[nt]} \left(\sigma_{2,\frac{k-1}{n}} \rho_{\frac{k-1}{n}} N_{1,k} + \sigma_{2,\frac{k-1}{n}} \sqrt{1-\rho_{\frac{k-1}{n}}^2} N_{3,k} \right)^2, \tag{4.8}$$

and

$$\mathcal{R}_{3,t}^n := \sum_{k=1}^{[nt]} \Delta_k^n X_1 \Delta_k^n X_2 - \frac{1}{n} \sum_{k=1}^{[nt]} \sigma_{1,\frac{k-1}{n}} N_{1,k} \left(\sigma_{2,\frac{k-1}{n}} \rho_{\frac{k-1}{n}} N_{1,k} + \sigma_{2,\frac{k-1}{n}} \sqrt{1-\rho_{\frac{k-1}{n}}^2} N_{3,k} \right). \tag{4.9}$$

At first, we start the negligibility (4.6) with the quantity $\mathcal{R}_{1,1}^n$ which can be rewritten as

$$\mathcal{R}_{1,1}^n = \sum_{k=1}^{n} R_{-,k} R_{+,k},$$

with $R_{\pm,k} := \int_{t_{k-1}^n}^{t_k^n} (\sigma_{1,s} \pm \sigma_{1,\frac{k-1}{n}}) \mathrm{d}B_{1,s}$, where $((R_{-,k}, R_{+,k}))_{k=1\cdots n}$ are independent centered Gaussian random vector with covariance

$$\begin{pmatrix} \varepsilon_{-,k}^n & \eta_k^n \\ \eta_k^n & \varepsilon_{+,k}^n \end{pmatrix}$$

where

$$\varepsilon_{\pm,k}^n = \int_{t_{k-1}^n}^{t_k^n} \left(\sigma_{1,s} \pm \sigma_{1,\frac{k-1}{n}} \right)^2 \mathrm{d}s \qquad \text{and} \qquad \eta_k^n = \int_{t_{k-1}^n}^{t_k^n} \left(\sigma_{1,s}^2 - \sigma_{1,\frac{k-1}{n}}^2 \right) \mathrm{d}s \tag{4.10}$$

So by Chebyshev's inequality, we have for all $r, \lambda > 0$,

$$\frac{1}{n}\log \mathbb{P}\left(\mathcal{R}^n_{1,1} > r\right) \leqslant -r\lambda + \frac{1}{n}\log \mathbb{E}\exp\left(n\lambda\mathcal{R}^n_{1,1}\right) \tag{4.11}$$

A simple calculation gives us

$$\begin{aligned}
\frac{1}{n}\log \mathbb{E}\exp\left(n\lambda\mathcal{R}^n_{1,1}\right) &= \frac{1}{n}\sum_{k=1}^{n}\log \mathbb{E}\exp\left(n\lambda R_{+,k}R_{-,k}\right)\\
&= -\frac{1}{2n}\sum_{k=1}^{n}\log\left[\frac{\varepsilon^n_{+,k}\varepsilon^n_{-,k} - \left(n\lambda(\varepsilon^n_{+,k}\varepsilon^n_{-,k} - (\eta^n_k)^2) + \eta^n_k\right)^2}{\varepsilon^n_{+,k}\varepsilon^n_{-,k} - (\eta^n_k)^2}\right]\\
&= -\frac{1}{2n}\sum_{k=1}^{n}\log\left[1 - n^2\lambda^2(\varepsilon^n_{+,k}\varepsilon^n_{-,k} - (\eta^n_k)^2) - n\lambda\eta^n_k\right]\\
&= \int_0^1 K(f_n(t))\mathrm{d}t
\end{aligned} \tag{4.12}$$

where K is given by

$$K(\lambda) := \begin{cases} -\frac{1}{2}\log\left(1-2\lambda\right) & if \quad \lambda < \frac{1}{2}\\ +\infty, & otherwise,\end{cases}$$

and

$$f_n(t) = \sum_{k=1}^{n} 1_{(t^n_k - t^n_{k-1})}(t)\left[\lambda^2\left(\frac{\varepsilon^n_{+,k}}{t^n_k - t^n_{k-1}}\frac{\varepsilon^n_{-,k}}{t^n_k - t^n_{k-1}} - \left(\frac{\eta^n_k}{t^n_k - t^n_{k-1}}\right)^2\right) + 2\lambda\left(\frac{\eta^n_k}{t^n_k - t^n_{k-1}}\right)\right]$$

where $\varepsilon^n_{\pm,k}$ and η^n_k are given in (4.10).

By the continuity condition of the assumption **(LDP)** and the classical Lebesgue derivation theorem, we have that

$$f_n(t) \longrightarrow 0 \qquad \text{as} \quad n \to \infty.$$

By the classical Lebesgue derivation theorem we have that the right hand of the equality (4.12) goes to 0.

Letting n goes to infinity and than λ goes to infinity in (4.11), we obtain that

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}\left(\mathcal{R}^n_{1,1} > r\right) = -\infty.$$

Doing the same things with $-\mathcal{R}^n_{1,1}$, we obtain (4.6) for $\mathcal{R}^n_{1,1}$.

Now we shall prove (4.6) with $\mathcal{R}^n_{2,1}$. We have

$$\mathcal{R}^n_{2,1} = \sum_{k=1}^{n} E_{-,k}E_{+,k} + \sum_{k=1}^{n} E_{-,k}B_{+,k} + \sum_{k=1}^{n} E_{+,k}B_{-,k} + \sum_{k=1}^{n} B_{-,k}B_{+,k},$$

where

$$E_{\pm,k} := \int_{t^n_{k-1}}^{t^n_k}(\sigma_{2,s}\rho_s \pm \sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}})\mathrm{d}B_{1,s},$$

and

$$B_{\pm,k} := \int_{t_{k-1}^n}^{t_k^n} (\sigma_{2,s}\sqrt{1-\rho_s^2} \pm \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho_{\frac{k-1}{n}}^2})\mathrm{d}B_{3,s}.$$

where $(E_{-,k}, E_{+k})$,$(E_{-,k}, B_{+,k})$,$(E_{+,k}, B_{-,k})$,$(B_{-,k}, B_{+,k})$, $k = 1\cdots n$ are four independent centered Gaussian random vectors with covariances respectively given by

$$\begin{pmatrix} \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\rho_s - \sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}})^2\mathrm{d}s & \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}^2\rho_s^2 - \sigma_{2,\frac{k-1}{n}}^2\rho_{\frac{k-1}{n}}^2)\mathrm{d}s \\ \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}^2\rho_s^2 - \sigma_{2,\frac{k-1}{n}}^2\rho_{\frac{k-1}{n}}^2)\mathrm{d}s & \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\rho_s + \sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}})^2\mathrm{d}s \end{pmatrix}$$

and

$$\begin{pmatrix} \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\rho_s - \sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}})^2\mathrm{d}s & 0 \\ 0 & \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\sqrt{1-\rho_s^2} + \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho_{\frac{k-1}{n}}^2})^2\mathrm{d}s \end{pmatrix}$$

and

$$\begin{pmatrix} \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\rho_s + \sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}})^2\mathrm{d}s & 0 \\ 0 & \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\sqrt{1-\rho_s^2} - \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho_{\frac{k-1}{n}}^2})^2\mathrm{d}s \end{pmatrix}$$

and

$$\begin{pmatrix} \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\sqrt{1-\rho_s^2} - \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho_{\frac{k-1}{n}}^2})^2\mathrm{d}s & \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}^2(1-\rho_s^2) - \sigma_{2,\frac{k-1}{n}}^2(1-\rho_{\frac{k-1}{n}}^2))\mathrm{d}s \\ \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}^2(1-\rho_s^2) - \sigma_{2,\frac{k-1}{n}}^2(1-\rho_{\frac{k-1}{n}}^2))\mathrm{d}s & \int_{t_{k-1}^n}^{t_k^n}(\sigma_{2,s}\sqrt{1-\rho_s^2} + \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho_{\frac{k-1}{n}}^2})^2\mathrm{d}s. \end{pmatrix}$$

So (4.6) for $\mathcal{R}_{2,1}^n$ is deduced if

$$\sum_{k=1}^n E_{-k}E_{+,k} \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0, \qquad \sum_{k=1}^n E_{-,k}B_{+,k} \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0,$$

$$\sum_{k=1}^n E_{+,k}B_{-,k} \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0, \qquad \sum_{k=1}^n B_{-,k}B_{+,k} \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0.$$

Each convergence is deduced by the same calculations as for $(R_{-,k}, R_{+,k})_{k=1\cdots n}$.

Now we shall prove (4.6) with $\mathcal{R}_{3,1}^n$. We have

$$\begin{aligned}\mathcal{R}_{3,1}^n &= \sum_{k=1}^n R_{-,k}E_{-,k} + \sum_{k=1}^n R_{-,k}B_{-,k} + \sum_{k=1}^n R_{-,k}\left(\int_{t_{k-1}^n}^{t_k^n} \sigma_{2,\frac{k-1}{n}}\rho_{\frac{k-1}{n}}\mathrm{d}B_{1,s}\right)\\ &\quad+\sum_{k=1}^n R_{-,k}\left(\int_{t_{k-1}^n}^{t_k^n} \sigma_{2,\frac{k-1}{n}}\sqrt{1-\rho^2_{\frac{k-1}{n}}}\mathrm{d}B_{3,s}\right) + \sum_{k=1}^n E_{-,k}\left(\int_{t_{k-1}^n}^{t_k^n} \sigma_{1,\frac{k-1}{n}}\mathrm{d}B_{1,s}\right)\\ &\quad+\sum_{k=1}^n B_{-,k}\left(\int_{t_{k-1}^n}^{t_k^n} \sigma_{1,\frac{k-1}{n}}\mathrm{d}B_{1,s}\right),\end{aligned}$$

where we have used the same notation as before. By the same calculations used to prove (4.6) for $\mathcal{R}_{1,1}^n$ and $\mathcal{R}_{2,1}^n$, we obtain (4.6) for $\mathcal{R}_{3,1}^n$.

Then $V_1^n(X-Y)$ and W_1^n satisfy the same LDP.

Part 3. We will prove that $V_1^n(X)$ and $V_1^n(X-Y)$ satisfy the same LDP. We need to prove that

$$\|V_1^n(X) - V_1^n(X-Y)\| \quad \underset{n}{\overset{\text{superexp}}{\longrightarrow}} 0.$$

This will be done element by element : for $\ell = 1,2$

$$Q_{\ell,1}^n(X) - Q_{\ell,1}^n(X-Y) \quad \underset{n}{\overset{\text{superexp}}{\longrightarrow}} 0 \quad \text{and} \quad C_1^n(X) - C_1^n(X-Y) \quad \underset{n}{\overset{\text{superexp}}{\longrightarrow}} 0. \tag{4.13}$$

We have

$$\left|Q_{\ell,1}^n(X) - Q_{\ell,1}^n(X-Y)\right| \le \varepsilon(n)Q_{\ell,1}^n(X-Y) + \left(1+\frac{1}{\varepsilon(n)}\right)Z_\ell^n,$$

and

$$|C_1^n(X) - C_1^n(X-Y)| \le \varepsilon(n)\left(Q_{1,1}^n(X-Y) + Q_{2,1}^n(X-Y)\right) + \left(\frac{1}{2}+\frac{1}{\varepsilon(n)}\right)(Z_1^n + Z_2^n).$$

with

$$Z_{\ell,n} = \sum_{k=1}^n\left(\int_{t_{k-1}^n}^{t_k^n} b_\ell(t,\omega)\mathrm{d}t\right)^2 \le \frac{\|b_\ell\|_\infty^2}{n}.$$

We chose $\varepsilon(n)$ such that $n\varepsilon(n)\to\infty$, so (4.13) follows from the LDP of $Q_{\ell,1}(X)$, $C_1^n(X)$ and the estimations above.

4.2. **Proof of Corollary 2.2.**

From Theorem 2.1, we obtain that $V_1^n(X-Y)$ satisfies the LDP on $\mathbb{R}^3$ with speed n and with the good rate function given by for all $x\in\mathbb{R}^3$

$$I_{ldp}^V(x) = \sup_{\lambda\in\mathbb{R}^3}\left(\langle\lambda,x\rangle - P_\rho(\sigma_1^2\lambda_1,\sigma_2^2\lambda_2,\sigma_1\sigma_2\lambda_3)\right) = P_\rho^*\left(\frac{\lambda_1}{\sigma_1^2},\frac{\lambda_2}{\sigma_2^2},\frac{\lambda_3}{\sigma_1\sigma_2}\right).$$

where P_ρ and P_ρ^* are given in (2.2) and (2.5) respectively. So we get the expression of I_{ldp}^V given in (2.6).

The Legendre transformation of P_c is defined by

$$P_\rho^*(x) := \sup_{\lambda\in\mathbb{R}^3}\left(\langle\lambda,x\rangle - P_\rho(\lambda)\right).$$

The function $\lambda \to \langle \lambda, x\rangle - P_\rho(\lambda)$ reaches the supremum at the point $\lambda^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*)$ such as

$$\begin{cases} \lambda_1^* = \frac{1}{2}\dfrac{x_1x_2 - (1-\rho^2)x_2 - x_3^2}{2(1-\rho^2)(x_1x_2 - x_3^2)}, \\ \lambda_2^* = \frac{1}{2}\dfrac{x_1x_2 - (1-\rho^2)x_1 - x_3^2}{2(1-\rho^2)(x_1x_2 - x_3^2)}, \\ \lambda_3^* = \dfrac{x_3^2\rho - x_1x_2\rho + (1-\rho^2)x_3}{(1-\rho^2)(x_1x_2 - x_3^2)}. \end{cases}$$

So we get the expression of the Legendre transformation P_ρ^* given in (2.5).

4.3. **Proof of Theorem 2.3.**

Now we shall prove the Theorem 2.3 in two steps.
Step 1. We start by proving that the LDP holds for

$$W_t^n = \frac{1}{n}\sum_{k1}^{[nt]} F\left(\frac{k-1}{n}\right) Z_k,$$

where F is given in (4.4) and Z_k is given in (4.5).

This result come from an application of LDP of Lemma 5.2 derived in the case where $\mathcal{X} := [0,1]$ and $R(dx)$ is the Lebesgue measure on $[0,1]$ and $x_i^n := i/n$. One can check that, in this situation, Assumptions (N-2) and (N-3) hold true.

The random variables $(Z_k)_{k=1,\cdots,n}$ are independent and identically distributed. And we will apply the Lemma with the random variables $Z_k^n := F\left(\dfrac{k-1}{n}\right) Z_k$. The law of Z_k^n depends on the position $x_i^n := i/n$ This type of model was partially examined by Najim see section 2.4.2 in [27].

By the definition of Z_k^n, the Assumptions (N-1) and (N-4) hold true.

Finally, we just need to verify that if x_i^n and x_j^n are close then so are $\mathcal{L}(Z_i^n)$ and $\mathcal{L}(Z_j^n)$ for the following Wasserstein type distance between probability measures:

$$d_{OW}(P,Q) = \inf_{\eta \in M(P,Q)} \inf\left\{a > 0; \int_{\mathbb{R}^3\times\mathbb{R}^3} \tau\left(\frac{z-z'}{a}\right)\eta(dzdz') \leqslant 1\right\},$$

where η is a probability with given marginals P and Q and $\eta(z) = e^{|z|} - 1$.

In fact, consider the random variables $Y = F(x)Z$ and $\tilde{Y} = F(x')Z$, since F is continuous

$$\mathbb{E}\tau\left(\frac{Y-\tilde{Y}}{\epsilon}\right) = \mathbb{E}\tau\left(\frac{(F(x)-F(x'))Z}{\epsilon}\right) \leq 1$$

for x' close to x. Thus $d_{OW}(\mathcal{L}(Z_i^n), \mathcal{L}(Z_j^n)) \leq \epsilon$. This gives the Assumption (N-5).

So we deduce that the sequence $W_\cdot^n$ satisfies the LDP on $\mathcal{B}V$ with speed n and with the rate function J_{ldp} given by

$$J_{ldp}(f) = \int_0^1 \Lambda_t^*(f_a^{'}(t))\mathrm{d}t + \int_0^1 \hbar_t(f_s^{'}(t))\mathrm{d}\theta(t).$$

where for all $z \in \mathbb{R}^3$ and all $t \in [0,1]$

$$\Lambda_t^*(z) = \sup_{\lambda \in \mathbb{R}^3} \left(\langle \lambda, z \rangle - \Lambda_t(\lambda) \right),$$

with

$$\Lambda_t(\lambda) = \log \int_{\mathbb{R}^3} e^{\langle \lambda, z \rangle} P(t, dz) = P_{\rho_t}(\lambda_1 \sigma_{1,t}^2, \lambda_2 \sigma_{2,t}^2, \lambda_3 \sigma_{1,t}\sigma_{2,t}),$$

so Λ_t^* coincide with $P_{\rho_t}^*$ given in Theorem 2.3.

And θ is any real-valued nonnegative measure with respect to which μ_s^f is absolutely continuous and $f_s' = d\mu_s^f/d\theta$ and for all $z \in \mathbb{R}^3$ and all $t \in [0,1]$ the recession function $\hbar_t$ of Λ_t^* defined by $\hbar_t(z) = \sup\{\langle \lambda, z \rangle, \lambda \in \mathcal{D}_{\Lambda_t}\}$ with $\mathcal{D}_{\Lambda_t} = \{\lambda \in \mathbb{R}^3, \Lambda_t(\lambda) < \infty\} = \{\lambda \in \mathbb{R}^3, P_{\rho_t}(\lambda_1 \sigma_{1,t}^2, \lambda_2 \sigma_{2,t}^2, \lambda_3 \sigma_{1,t}\sigma_{2,t}), < \infty\}$.

The recession function α of P_c^* , see Theorem 13.3 in [31] is given by

$$\alpha(z) := \lim_{h \to \infty} \frac{P_c^*(hz)}{h} = \frac{z_1 + z_2 - 2cz_3}{2(1-c^2)} 1_{[z_1>0, z_2>0, z_3^2<z_1 z_2]}.$$

Using this expression, we obtain the rate function given in the Theorem 2.3.

Step 2. Now we have to prove that

$$\sup_{t \in [0,1]} \|V_t^n(X-Y) - W_t^n\| \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0.$$

To do that, we have to prove that for $q = 1, 2, 3$

$$\sup_{t \in [0,1]} \left| \mathcal{R}_{q,t}^n \right| \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0, \tag{4.14}$$

where the definition of $\mathcal{R}_{q,t}^n$ arge given in (4.7), (4.8) and (4.9) for $q = 1, 2, 3$ respectively.

We start by proving (4.14) for $q = 1$, the other terms for $q = 2, 3$ follow the same line of proof.

We remark that $(\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n))$ is a $(\mathcal{F}_{[nt]/n})$-martingale. Then

$$\exp(\lambda n \left[\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n) \right])$$

is a sub-martingale. By the maximal inequality, we have for any $r, \lambda > 0$,

$$\begin{aligned} \mathbb{P}\left(\sup_{t \in [0,1]} \left[\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n) \right] > r \right) &= \mathbb{P}\left(\exp\left(\lambda n \sup_{t \in [0,1]} \left[\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n) \right] \right) > e^{n\lambda r} \right) \\ &\leq e^{-n\lambda r} \mathbb{E}\left(\exp\left(\lambda n \left[\mathcal{R}_{1,1}^n - \mathbb{E}(\mathcal{R}_{1,1}^n) \right] \right) \right) \end{aligned}$$

and similarly

$$\mathbb{P}\left(\inf_{t \in [0,1]} \left[\mathcal{R}_{q,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n) \right] < -r \right) \leq e^{-n\lambda r} \mathbb{E}\left(\exp\left(-\lambda n \left[\mathcal{R}_{1,1}^n - \mathbb{E}(\mathcal{R}_{1,1}^n) \right] \right) \right).$$

So we get

$$\frac{1}{n} \log \mathbb{P}\left(\sup_{t \in [0,1]} \left[\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n) \right] > r \right) \leq -\lambda r + \frac{1}{n} \log \mathbb{E}\left(e^{\lambda n \mathcal{R}_{1,1}^n} \right) - \lambda \mathbb{E}(\mathcal{R}_{1,1}^n).$$

It is easy to see that $\mathbb{E}(\mathcal{R}_{1,1}^n) \to 0$ as n goes to infinity. We have already seen in (4.12) that $\frac{1}{n}\log \mathbb{E}\left(e^{\lambda n \mathcal{R}_{1,1}^n}\right) \to 0$ as n gos to infinity. So we obtain for all $\lambda > 0$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sup_{t\in[0,1]} \left[\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n)\right] > r\right) \leq -\lambda r.$$

Letting $\lambda > 0$ goes to infinity, we obtain that the left term in the last inequality goes to $-\infty$.

And similarly, by doing the same calculations with

$$\mathbb{P}\left(\inf_{t\in[0,1]} \left[\mathcal{R}_{q,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n)\right] < -r\right),$$

we obtain that

$$\sup_{t\in[0,1]} \left|\mathcal{R}_{1,t}^n - \mathbb{E}(\mathcal{R}_{1,t}^n)\right| \quad \xrightarrow[n]{\text{superexp}} 0.$$

Since

$$\mathbb{E}(\mathcal{R}_{1,1}^n) \quad \xrightarrow[n]{\text{superexp}} 0,$$

we obtain (4.14) for $q = 1$.

4.4. **Proof of Theorem 2.4.**

As is usual, the proof of the MDP is somewhat simpler than the LDP, relying on the same line of proof than the one for the CLT. Namely, a good control of the asymptotic of the moment generating functions, and Gärtner-Ellis theorem. We shall then prove that for all $\lambda \in \mathbb{R}^3$

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\exp\left(b_n^2 \frac{\sqrt{n}}{b_n} \langle \lambda, V_1^n - [V]_1 \rangle\right) = \frac{1}{2} \langle \lambda, \Sigma_1 \cdot \lambda \rangle . \tag{4.15}$$

Taking the calculation in (4.2),we have

$$\frac{1}{b_n^2} \log \mathbb{E}\exp\left(b_n^2 \frac{\sqrt{n}}{b_n} \langle \lambda, V_1^n - [V]_1 \rangle\right) = \frac{1}{b_n^2} \sum_{k=1}^n \left[H_k^n(\lambda) - b_n\sqrt{n}\, \langle \lambda, [V]_1 \rangle\right],$$

with

$$H_k^n(\lambda) := P_{c_k^n}\left(\lambda_1 b_n \sqrt{n} a_{1,k}, \lambda_2 b_n \sqrt{n} a_{2,k}, \lambda_3 b_n \sqrt{n} \sqrt{a_{1,k}}\sqrt{a_{2,k}}\right),$$

where $a_{\ell,k}$ are given in (4.1) and c_k^n is given in (4.2).

By our condition (2.1),

$$\varepsilon(n) := \sqrt{n} b_n \max_{1\leqslant k\leqslant n} \max_{\ell=1,2} a_{\ell,k} \xrightarrow[n\to\infty]{} 0.$$

By multidimensional Taylor formula and noting that $P_{c_k^n}(0,0,0) = 0$, $\nabla P_{c_k^n}(0,0,0) = (1,1,c_k^n)^T$ and the Hessian matrix

$$H(P_{c_k^n})(0,0,0) = \begin{pmatrix} 2 & 2(c_k^n)^2 & 2c_k^n \\ 2(c_k^n)^2 & 2 & 2c_k^n \\ 2c_k^n & 2c_k^n & 1+(c_k^n)^2 \end{pmatrix},$$

and after an easy calculations, we obtain once if $||\lambda|| \cdot |\varepsilon(n)| < \frac{1}{4}$, i.e. for n large enough,

$$H_k^n(\lambda) = b_n\sqrt{n}\,\langle\lambda, [V]_1\rangle + nb_n^2\frac{1}{2}\langle\lambda, \Sigma_k^n \cdot \lambda\rangle + nb_n^2\nu(k,n),$$

where

$$\Sigma_k^n := \begin{pmatrix} a_{1,k}^2 & (\vartheta_k^n)^2 & a_{1,k}\vartheta_k^n \\ (\vartheta_k^n)^2 & a_{2,k}^2 & a_{2,k}\vartheta_k^n \\ a_{1,k}\vartheta_k^n & a_{2,k}\vartheta_k^n & \frac{1}{2}(a_{1,k}a_{2,k} + (\vartheta_k^n)^2) \end{pmatrix},$$

where ϑ_k^n is given in (4.2), and $\nu(k,n)$ satisfies

$$|\nu(k,n)| \leqslant C||\lambda|| \cdot |\varepsilon(n)|,$$

where $C = \frac{1}{6}\sup_{||\lambda||\leq 1/4}\left|\frac{\partial^3 P(\lambda_1,\lambda_2,\lambda_3)}{\partial^3\lambda}\right|$.

On the other hand, by the classical Lebesgue derivation theorem see [14], we have

$$\sum_{k=1}^n \left(\frac{\int_{t_{k-1}^n}^{t_k^n} g(s)\mathrm{d}s}{t_k^n - t_{k-1}^n}\right)\left(\frac{\int_{t_{k-1}^n}^{t_k^n} h(s)\mathrm{d}s}{t_k^n - t_{k-1}^n}\right)(t_k^n - t_{k-1}^n) \to \int_0^1 g(s)h(s)\mathrm{d}s$$

by taking different chosse of g and h: once $g(s) = h(s) = \sigma_{\ell,s}^2$ or $g(s) = h(s) = \sigma_{1,s}\sigma_{1,s}\rho_s$, or $g(s) = \sigma_{\ell,s}^2$ and $h(s) = \sigma_{\ell',s}^2$ $\ell \neq \ell'$, and $g(s) = \sigma_{\ell,s}^2$ and $h(s) = \sigma_{1,s}\sigma_{1,s}\rho_s$, we obtain that

$$n\sum_{k=1}^n \Sigma_k^n \to_{n\to\infty} \Sigma_1 = \begin{pmatrix} \int_0^1 \sigma_{1,t}^4\mathrm{d}t & \int_0^1 \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2\mathrm{d}t & \int_0^1 \sigma_{1,t}^3\sigma_{2,t}\rho_t\mathrm{d}t \\ \int_0^1 \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2\mathrm{d}t & \int_0^1 \sigma_{2,t}^4\mathrm{d}t & \int_0^1 \sigma_{1,t}\sigma_{2,t}^3\rho_t\mathrm{d}t \\ \int_0^1 \sigma_{1,t}^3\sigma_{2,t}\rho_t\mathrm{d}t & \int_0^1 \sigma_{1,t}\sigma_{2,t}^3\rho_t\mathrm{d}t & \int_0^1 \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t \end{pmatrix} = \Sigma_1.$$

Then the (4.15) follows.

Hence (2.4) follows from the Gärtner-Ellis theorem.

4.5. **Proof of Theorem 2.5.**

It is well known that the LDP of finite dimensional vector

$$\left(\frac{\sqrt{n}}{b_n}(V_{s_1}^n(X) - [V]_{s_1}, \cdots, V_{s_k}^n(X) - [V]_{s_k})\right), 0 < s_1 < \cdots < s_k \leqslant 1, k \geqslant 1$$

and the following exponential tightness: for any $s \in [0,1]$ and $\eta > 0$

$$\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\sup_{s\leqslant t\leqslant s+\varepsilon}\|\Delta_s^t V_\cdot^n(X) - \Delta_s^t[V]_\cdot\| > \eta\right) = -\infty$$

with $\Delta_s^t V_\cdot^n = V_t^n - V_s^n$, are sufficient for the LDP of $\frac{\sqrt{n}}{b_n}(V_\cdot^n(X) - [V]_\cdot)$ for the sup-norm topology (cf. [13],[14]).

Under the assumption of Theorem 2.5, we have:

$$\lim_{n\to\infty} \sqrt{n} \sup_{t\in[0,1]} \|\mathbb{E}V_t^n(X) - [V]_t\| = 0. \tag{4.16}$$

In fact, we have:

$$\begin{aligned}
&\sqrt{n} \sup_{t\in[0,1]} \|\mathbb{E}V_t^n(X) - [V]_t\| \\
&\leqslant 3\sqrt{n} \max\left(\max_{\ell=1,2} \sup_{t\in[0,1]} |\mathbb{E}Q_{\ell,t}^n(X) - [X_\ell]_t|, \sqrt{n} \sup_{t\in[0,1]} |\mathbb{E}C_t^n(X) - \langle X_1, X_2\rangle_t| \right) \\
&\leqslant 3 \max\left(\max_{\ell=1,2} \max_{k\leqslant n} \sqrt{\int_{(k-1)/n}^{k/n} \sigma_{\ell,t}^4 \mathrm{d}t}, \max_{k\leqslant n} \sqrt{\int_{(k-1)/n}^{k/n} \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t} \right).
\end{aligned}$$

By our condition (2.1), we obtain (4.16).

Now, we show that for any partition $0 < s_1 < \cdots < s_k \leqslant 1, k \geqslant 1$ of $[0,1]$

$$\frac{\sqrt{n}}{b_n}\left(V_{s_1}^n(X) - [V]_{s_1}, \cdots, V_{s_k}^n(X) - [V]_{s_k}\right)$$

satisfies the LDP on $\mathbb{R}^k$ with speed b_n^2 and with the rate function given by

$$I_{s_1,\cdots,s_k}(x_1,\cdots,x_k) = \frac{1}{2}\sum_{i=1}^{k} \langle (x_i - x_{i-1}), (\Sigma_{s_{i-1}}^{s_i})^{-1} \cdot (x_i - x_{i-1})\rangle \tag{4.17}$$

where

$$\Sigma_s^u = \begin{pmatrix} \int_s^u \sigma_{1,t}^4 \mathrm{d}t & \int_s^u \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t & \int_s^u \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t \\ \int_s^u \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t & \int_s^u \sigma_{2,t}^4 \mathrm{d}t & \int_s^u \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t \\ \int_s^u \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t & \int_s^u \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t & \int_s^u \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2) \mathrm{d}t \end{pmatrix}$$

is invertible and

$$\begin{aligned}
\det(Q_s^u) = &\left(\int_s^u \sigma_{1,t}^4 \mathrm{d}t\right)\left(\int_s^u \sigma_{2,t}^4 \mathrm{d}t\right)\left(\int_s^u \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2) \mathrm{d}t\right) \\
&+ 2\left(\int_s^u \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t\right)\left(\int_s^u \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t\right)\left(\int_s^u \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t\right) \\
&- \left(\int_s^u \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2) \mathrm{d}t\right)\left(\int_s^u \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t\right)^2 \\
&- \left(\int_s^u \sigma_{1,t}^4 \mathrm{d}t\right)\left(\int_s^u \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t\right)^2 - \left(\int_s^u \sigma_{2,t}^4 \mathrm{d}t\right)\left(\int_s^u \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t\right)^2
\end{aligned}$$

and $(\Sigma_s^u)^{-1}$ his inverse.

For n large enough we have $1 < [nt_1] < \cdots < [nt_k] < n$, so by applying the homeomorphism

$$\Upsilon : (x_1,\cdots,x_k) \to (x_1, x_2 - x_1, \cdots, x_k - x_{k-1})$$

$Z_n = (V_{s_1}^n(X) - [V]_{s_1}, \cdots, V_{s_k}^n(X) - [V]_{s_k})$ can be mapped to $U_n = \Upsilon Z_n$ with independent components.

Then we consider the LDP of $\frac{\sqrt{n}}{b_n}(U_n - \mathbb{E}U_n)$.

For any $\theta = (\theta_1, \cdots, \theta_k) \in (\mathbb{R}^3)^k$,

$$\Lambda s_1, \cdots, s_k(\theta) = \lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E} \exp\left(b_n\sqrt{n}\,\langle \lambda, U_n - EU_n\rangle\right) = \sum_{i=1}^k \frac{1}{2}\langle \lambda_i, \Sigma_{s_{i-1}}^{s_i} \cdot \lambda_i\rangle.$$

By Gärtner-Ellis theorem, $\frac{\sqrt{n}}{b_n}(U_n - \mathbb{E}U_n)$ satifies the LDP in $(\mathbb{R}^3)^k$ with speed b_n^2 and with the good rate function

$$\Lambda^* s_1, \cdots, s_k(x) = \frac{1}{2}\sum_{i=1}^k \langle x_i, (\Sigma_{s_{i-1}}^{s_i})^{-1} \cdot x_i\rangle.$$

Then by the inverse contraction principle, we have $\frac{\sqrt{n}}{b_n}(Z_n - \mathbb{E}Z_n)$ satisfies the LDP with speed b_n^2 and with the rate function $I_{s_1,\cdots,s_k}(x)$ given in (4.17).

Now, we shall prove that for any $\eta > 0$, $s \in [0,1]$

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} \sup_{s\leqslant t\leqslant s+\varepsilon} \|\Delta_s^t V_\cdot^n(X) - \mathbb{E}\Delta_s^t V_\cdot^n(X)\| > \eta\right) = -\infty. \tag{4.18}$$

For that we need to prove that for $\ell = 1,2$ and for all $\eta > 0$ and $s \in [0,1]$

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} \sup_{s\leqslant t\leqslant s+\varepsilon} |\Delta_s^t Q_{\ell,\cdot}(X) - \mathbb{E}\Delta_s^t Q_{\ell,\cdot}(X)| > \eta\right) = -\infty, \tag{4.19}$$

and

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} \sup_{s\leqslant t\leqslant s+\varepsilon} |\Delta_s^t C_\cdot(X) - \mathbb{E}\Delta_s^t C_\cdot(X)| > \eta\right) = -\infty. \tag{4.20}$$

In fact (4.19) can be done in the same way than in Djellout et al.[14]. It remains to show (4.20). This will be done following the same technique as for the proof of (4.19) and using a result of [15]. Remark that $(C_t^n(X) - \mathbb{E}C_t^n(X))$ is an $\mathcal{F}_{[nt]/n}$-martingale. Then

$$\exp(\lambda[\Delta_s^t(C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X))])$$

is a sub-martingale. By the maximal inequality, we have for any $\eta, \lambda > 0$

$$\begin{aligned}\mathbb{P}\left(\sup_{s\leqslant t\leqslant s+\varepsilon} \Delta_s^t\left[C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X)\right] > \eta\right) &= \mathbb{P}\left(\exp(\lambda \sup_{s\leqslant t\leqslant s+\varepsilon} \Delta_s^t\left[C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X)\right] > e^{\lambda\eta}\right)\\ &\leqslant e^{-\lambda\eta}\mathbb{E}\exp\left(\lambda\Delta_s^{s+\varepsilon}\left[C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X)\right]\right),\end{aligned} \tag{4.21}$$

and similary,

$$\mathbb{P}\left(\inf_{s\leqslant t\leqslant s+\varepsilon} \Delta_s^t\left[C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X)\right] < -\eta\right) \leqslant e^{-\lambda\eta}\mathbb{E}\exp\left(-\lambda[\Delta_s^{s+\varepsilon}\left[C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X)\right]\right). \tag{4.22}$$

Using Remark 2.4 in [15], we have that for all $c \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\exp\left(cb_n^2 \frac{\sqrt{n}}{b_n}\Delta_s^{s+\epsilon}\left[C_\cdot^n(X) - \mathbb{E}C_\cdot^n(X)\right]\right) = \frac{1}{2}c^2 \int_s^{s+\epsilon} \sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t.$$

Therefore taking $\eta = \delta\dfrac{b_n}{n}$, $\lambda = b_n\sqrt{n}c$ $(c > 0)$ in (4.21), we get

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\sup_{s\leqslant t\leqslant s+\varepsilon}\Delta_s^t\left[C_\cdot^n(X)-\mathbb{E}C_\cdot^n(X)\right]>\delta\right)\\&\leqslant\inf_{c>0}\left\{-c\delta+\frac{1}{2}c^2\int_s^{s+\epsilon}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t\right\}=-\frac{\delta^2}{2\int_s^{s+\epsilon}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t},\end{aligned}$$

and similary by (4.22),

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\inf_{s\leqslant t\leqslant s+\varepsilon}\Delta_s^t\left[C_\cdot^n(X)-\mathbb{E}C_\cdot^n(X)\right]<-\delta\right)\leqslant-\frac{\delta^2}{2\int_s^{s+\epsilon}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t}.$$

By the integrability of $\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)$, we have

$$\lim_{\varepsilon\downarrow 0}\sup_{s\in[0,1]}\int_s^{s+\epsilon}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t=0.$$

Hence (4.20) follows from the above estimations. So we have (4.18).

By (4.17) and (4.18), $\dfrac{\sqrt{n}}{b_n}(V_\cdot^n-[V]_\cdot)$ satifies the LDP with the speed b_n^2 and the good rate function

$$I_{sup}(x)=\sup\left\{I_{s_1,\cdots,s_k}(x(s_1),\cdots,x(s_k));0<s_1<\cdots<s_k\leqslant 1,k\geqslant 1\right\},$$

where

$$I_{s_1,\cdots,s_k}(x(s_1),\cdots,x(s_k))=\frac{1}{2}\sum_{i=1}^k\langle x(s_i)-x(s_{i-1}),(\Sigma_{s_{i-1}}^{s_i})^{-1}\cdot(x(s_i)-x(s_{i-1}))\rangle$$

It remains to prove that $I_{sup}(x)=I_{mdp}(x)$.

We shall prove that $I_{sup}(x)\leqslant I_{mdp}(x)$.

For this, we treat the first element of the matrix $(\Sigma_{s_{i-1}}^{s_i})^{-1}$ which is denoted $(\Sigma_{s_{i-1}}^{s_i})_{1,1}^{-1}$ and we prove that

$$\sum_{i=1}^k(x_1(s_i)-x_1(s_{i-1}))^2.(\Sigma_{s_{i-1}}^{s_i})_{1,1}^{-1}\leqslant\int_0^1(x_1'(t))^2.(\Sigma_t^{-1})_{1,1}\mathrm{d}t,$$

where $(\Sigma_t^{-1})_{1,1}$ represente the first element of the matrix Σ_t^{-1}.

We have

$$(\Sigma_{s_{i-1}}^{s_i})_{1,1}^{-1}=\frac{1}{\det(\Sigma_{s_{i-1}}^{s_i})}\left(\left(\int_{s_{i-1}}^{s_i}\sigma_{2,t}^4\mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i}\frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t\right)-\left(\int_{s_{i-1}}^{s_i}\sigma_{1,t}\sigma_{2,t}^3\rho_t\mathrm{d}t\right)^2\right).$$

By [20, p.1305], for $x:=(x_1,x_2,x_3)\in\mathcal{H}$, if $I_{sup}(x)<+\infty$, then for $0<s_1<\cdots<s_k\leqslant 1$, Then

$$\begin{aligned}\sum_{i=1}^k(x_1(s_i)-x_1(s_{i-1}))^2.(\Sigma_{s_{i-1}}^{s_i})_{1,1}^{-1}&\leqslant\int_0^1(x_1'(t))^2.\frac{\frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^6(1+\rho_t^2)-\sigma_{1,t}^2\sigma_{2,t}^6\rho_t^2}{\det(\Sigma_t)}\mathrm{d}t\\&=\int_0^1(x_1'(t))^2.(\Sigma_t^{-1})_{1,1}\mathrm{d}t,\end{aligned}$$

The same calculation with the other terms of the matrix given in the following, implies that $I_{sup}(x) \leqslant I_{mdp}(x)$:

$$(\Sigma_{s_{i-1}}^{s_i})^{-1}_{2,2} = \frac{1}{\det(\Sigma_{s_{i-1}}^{s_i})}\left[\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^4 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t\right) - \left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t\right)^2\right],$$

$$(\Sigma_{s_{i-1}}^{s_i})^{-1}_{3,3} = \frac{1}{\det(\Sigma_{s_{i-1}}^{s_i})}\left[\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^4 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \sigma_{2,t}^4 \mathrm{d}t\right) - \left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t\right)^2\right],$$

$$\begin{aligned}(\Sigma_{s_{i-1}}^{s_i})^{-1}_{1,2} = (\Sigma_{s_{i-1}}^{s_i})^{-1}_{2,1} = \frac{1}{\det(\Sigma_{s_{i-1}}^{s_i})}&\left[\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t\right)\right.\\ &\left.-\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t\right)\right],\end{aligned}$$

$$\begin{aligned}(\Sigma_{s_{i-1}}^{s_i})^{-1}_{1,3} = (\Sigma_{s_{i-1}}^{s_i})^{-1}_{3,1} = \frac{1}{\det(\Sigma_{s_{i-1}}^{s_i})}&\left[\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t\right)\right.\\ &\left.-\left(\int_{s_{i-1}}^{s_i} \sigma_{2,t}^4 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t\right)\right],\end{aligned}$$

$$\begin{aligned}(\Sigma_{s_{i-1}}^{s_i})^{-1}_{2,3} = (\Sigma_{s_{i-1}}^{s_i})^{-1}_{3,2} = \frac{1}{\det(\Sigma_{s_{i-1}}^{s_i})}&\left[\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t\right)\right.\\ &\left.-\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}^4 \mathrm{d}t\right)\left(\int_{s_{i-1}}^{s_i} \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t\right)\right].\end{aligned}$$

On the other hand, by the convergence of martingales and Fatou's lemma,

$$I_{mdp}(x) < +\infty, \qquad \text{and} \qquad I_{mdp}(x) \leqslant I_{sup}(x).$$

So we have $I_{sup}(x) = I_{mdp}(x)$.

4.6. **Proof of Theorem 2.6.**

Step 1. We shall prove that $\tilde{V}_\cdot^n$ and $V_\cdot^n(X-Y)$ satisfy the same LDP, by means of the approximation Lemma in [14]. So we shall prove that $\tilde{Q}_{\ell,\cdot}^n$ and $Q_{\ell,\cdot}^n(X-Y)$ satisfy the same LDP and idem for $\tilde{C}_\cdot^n$ and $C_\cdot^n(X-Y)$.

We have

$$\sup_{t\in[0,1]} |\tilde{Q}_{\ell,t}^n(X) - Q_{\ell,t}^n(X-Y)| \leqslant \varepsilon(n) Q_{\ell,t}^n(X-Y) + \left(1 + \frac{1}{\varepsilon(n)}\right) Z_{\ell,n} \tag{4.23}$$

and

$$\sup_{t\in[0,1]} |\tilde{C}_{1,t}^n(X) - C_{1,t}^n(X-Y)| \leqslant \varepsilon(n) \sum_{\ell=1}^2 Q_{\ell,t}^n(X-Y) + \left(\frac{1}{2} + \frac{1}{\varepsilon(n)}\right) \sum_{\ell=1}^2 Z_{\ell,n}, \tag{4.24}$$

where the sequence $\varepsilon(n) > 0$ will be selected later, and $Z_{\ell,n}$ is given

$$Z_{\ell,n} = \sum_{k=1}^n \left(\int_{t_{k-1}^n}^{t_k^n} b_{\ell,t}(X_t)\mathrm{d}t - b_{\ell,t_{k-1}^n}(X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)\right)^2, \tag{4.25}$$

with $X_t = (X_{1,t}, X_{2,t})$.

For $Q^n_{\ell,t}(X-Y)$, being a Gaussian process, Theorem 1.1 in [14] may be used. It remains to control $Z_{\ell,n}$. For this we just need to prove that:

$$\frac{1}{\varepsilon(n)} Z_{\ell,n} \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0. \tag{4.26}$$

The main idea is to reduce it to estimations of $M_{\ell,t} = \int_0^t \sigma_{\ell,s} \mathrm{d}B_{\ell,s}$, by means of Gronwall's inequality. So, we have at first for all $t \in [0,1]$

$$\begin{aligned}
\|X_t\| &\leqslant \|X_0\| + C\int_0^t (1+(1+\eta(s))\|X_s\|)\mathrm{d}s + \sup_{s\leqslant t}\|M_s\| \\
&\leqslant \left(C + \|X_{1,0}\| + \|X_{2,0}\| + \sup_{s\leqslant 1}\|M_s\|\right) + C_1\int_0^t \|X_s\|\mathrm{d}s.
\end{aligned}$$

where $C_1 = C(1+\eta(1))$. Hence, by Gronwall's inequality

$$\|X_t\| \leqslant \left(C + \|X_{1,0}\| + \|X_{2,0}\| + \sup_{s\leqslant 1}\|M_s\|\right) e^{C_1 t}, \qquad \forall t \in [0,1] \tag{4.27}$$

For any $s \in [0,1]$, $v > 0$

$$\begin{aligned}
\sup_{s\leqslant t\leqslant s+v} \|X_t - X_s\| &\leqslant \sup_{s\leqslant t\leqslant s+v}\|M_t - M_s\| + v.\sup_{s\leqslant t\leqslant s+v}\|b(t,X_t)\| \\
&\leqslant \sup_{s\leqslant t\leqslant s+v}\|M_t - M_s\| + vC_2\left(\sup_{0\leqslant t\leqslant 1}\|X_t\| + 1\right)
\end{aligned} \tag{4.28}$$

We get by (2.9), (4.27),(4.28) and Cauchy-Schwarz's inequality

$$\begin{aligned}
&\left(\int_{t^n_{k-1}}^{t^n_k} b_\ell(t,X_t)\mathrm{d}t - b_\ell(t^n_{k-1}, X_{t^n_{k-1}})(t^n_k - t^n_{k-1})\right)^2 \\
\leqslant\ & \left(\frac{1}{n}C\left(1 + \sup_{t^n_{k-1}\leqslant t\leqslant t^n_k}\|X_t - X_{t^n_{k-1}}\| + 2\eta\left(\frac{1}{n}\right)\sup_{0\leqslant t\leqslant 1}\|X_t\|\right)\right)^2 \\
\leqslant\ & \frac{C_3}{n^2}\left(1 + \sup_{t^n_{k-1}\leqslant t\leqslant t^n_k}\|M_t - M_{t^n_{k-1}}\|^2 + \left(\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2\right)\sup_{0\leqslant t\leqslant 1}\|M_t\|^2\right)
\end{aligned} \tag{4.29}$$

Chose $\varepsilon(n) > 0$ so that

$$\varepsilon(n) \to 0 \quad \text{but} \quad \frac{\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2}{\varepsilon(n)} \to 0 \tag{4.30}$$

By (4.29) and the definition of $Z_{\ell,n}$, we have that

$$\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left(\frac{1}{\varepsilon(n)}Z_{\ell,n} > \delta\right) \leqslant \max(A,B)$$

where

$$A = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{\varepsilon(n)n} \max_{k\leqslant n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} \|M_t - M_{t_{k-1}^n}\|^2 > C_4\delta \right)$$

$$\leqslant 2 \max_{\ell=1,2} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{\varepsilon(n)n} \max_{k\leqslant n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |M_{\ell,t} - M_{\ell,t_{k-1}^n}|^2 > C_4\delta \right), \quad (4.31)$$

and

$$B = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{\varepsilon(n)^2 n} \left(\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2 \right) \sup_{0\leqslant t\leqslant 1} \|M_t\|^2 > C_5\delta \right)$$

$$\leqslant 2 \max_{\ell=1,2} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{\varepsilon(n)n} \left(\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2 \right) \sup_{0\leqslant t\leqslant 1} |M_{\ell,t}|^2 > C_5\delta \right). \quad (4.32)$$

By Lévy's inequality for a Brownian motion and our choice (4.30) of $\varepsilon(n)$, the limits (4.31) and (4.32) are both $-\infty$. Limit (4.26) follows.

Step 2. We shall prove that $\frac{\sqrt{n}}{b_n}(\tilde{V}_\cdot^n - [V]_\cdot)$ and $\frac{\sqrt{n}}{b_n}(V_\cdot^n(X-Y) - [V]_\cdot)$ satisfy the same LDP, by means of the approximation lemma in [14] and of three strong tools: Gronwall's inequality, Lévy's inequality and an isoperimetric inequality for gaussian processes. By the estimation above (4.23) and (4.24), and as $Q_{\ell,t}^n(X-Y)$ was also estimated in the proof Theorem 1.3 in [14]. It remains to control $Z_{\ell,n}$ given in (4.25) . For this we just need to prove that:

$$\frac{1}{\varepsilon(n)} \frac{\sqrt{n}}{b_n} Z_{\ell,n} \quad \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0. \quad (4.33)$$

Chose $\varepsilon(n) > 0$ so that

$$\frac{\varepsilon(n)\sqrt{n}}{b_n} \to 0 \quad \text{but} \quad \frac{\left(\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2\right) b_n}{\varepsilon(n)\sqrt{n}} \to 0. \quad (4.34)$$

By (4.29) and the definition of $Z_{\ell,n}$ given in (4.25), we have that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{\varepsilon(n)} \frac{\sqrt{n}}{b_n} Z_{\ell,n} > \delta \right) \leqslant \max(A, B)$$

where

$$A = \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{\varepsilon(n) b_n \sqrt{n}} \max_{k\leqslant n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} \|M_t - M_{t_{k-1}^n}\|^2 > C_4\delta \right)$$

$$\leqslant 2 \max_{\ell=1,2} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{\varepsilon(n) b_n \sqrt{n}} \max_{k\leqslant n} \sup_{t_{k-1}^n \leqslant t \leqslant t_k^n} |M_{\ell,t} - M_{\ell,t_{k-1}^n}|^2 > C_4\delta \right), \quad (4.35)$$

and

$$B = \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{\varepsilon(n) b_n \sqrt{n}} \left(\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2 \right) \sup_{0\leqslant t\leqslant 1} \|M_t\|^2 > C_5\delta \right)$$

$$\leqslant 2 \max_{\ell=1,2} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{\varepsilon(n) b_n \sqrt{n}} \left(\frac{1}{n^2} + \eta\left(\frac{1}{n}\right)^2 \right) \sup_{0\leqslant t\leqslant 1} |M_{\ell,t}|^2 > C_5\delta \right). \quad (4.36)$$

By Lévy's inequality for a Brownian motion and our choice (4.34) of $\varepsilon(n)$, the limit (4.36) are also $-\infty$. As in [14], it's more little difficult to estimate (4.35). By the isoperimetric inequality [[21], p17,(1.24)] and our choice (4.34), we conclude that the limit (4.35) are both $-\infty$.

4.7. **Proof of Corollary 3.2.**

We have just to do the identification of the rate function. We knew that $\varrho_1^n(X)$ satisfies the LDP on $\mathbb{R}$ with speed n and with the good rate function given by

$$I_{ldp}^{\rho}(u) := \inf_{\{(x_1,x_2,x_3)\in\mathbb{R}^3: x_3=u\sqrt{x_1x_2}, x_1>0, x_2>0\}} I_{ldp}^{V}(x_1,x_2,x_3),$$

where I_{ldp}^{V} is given in (2.6). So

$$I_{ldp}^{\rho}(u) = \inf\left\{\log\left(\frac{\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}{\sqrt{x_1x_2}\sqrt{1-u^2}}\right) - 1 + \frac{\sigma_2^2x_1+\sigma_1^2x_2-2\rho\sigma_1\sigma_2u\sqrt{x_1x_2}}{2\sigma_1^2\sigma_2^2(1-\rho^2)}, \quad x_1>0, x_2>0\right\}.$$

The above infinimum is attained at the point $(x_1,x_2) = \left(\dfrac{\sigma_1^2(1-\rho^2)}{1-\rho u}, \dfrac{\sigma_2^2(1-\rho^2)}{1-\rho u}\right)$, so we obtain (3.1).

4.8. **Proof of Proposition 3.3.**

As said before, quite unusually, the MDP is here a little bit harder to prove, due to the fact that it is not a simple transformation of the MDP of $\frac{\sqrt{n}}{b_n}(V_t^n - [V]_t)$. Therefore we will use the strategy developped for the TCL: the delta-method. Fortunately, Gao and Zhao [17] have developped such a technology at the large deviations level. However it will require to prove quite heavy exponential negligibility to be able to do so. For simplicity we omit X in the notations of $Q_{1,t}^n(X)$ and $C_t^n(X)$.

Let introduce $\Xi_{t,n}$ such that $\Xi_t^n := \sqrt{Q_{1,t}^n}\sqrt{Q_{2,t}^n}$. Then by the Lemma 5.3 applied to the functions $g := (x,y,z) \mapsto \sqrt{x}\sqrt{y}$ and $h := (x,y,z) \mapsto \frac{1}{\sqrt{x}\sqrt{y}}$, we deduce that $\frac{\sqrt{n}}{b_n}\Big(\Xi_1^n - \mathbb{E}(\Xi_1^n)\Big)$ and $\frac{\sqrt{n}}{b_n}\Big((\Xi_1^n)^{-1} - (\mathbb{E}(\Xi_1^n))^{-1}\Big)$ satisfies the LDP on $\mathbb{R}$ with the same speed b_n^2 and with the rates functions respectively given by I_{mdp}^{Ξ} and $I_{mdp}^{\Xi^{-1}}$:

$$I_{mdp}^{\Xi}(u) := \inf_{\{(x,y,z)\in\mathbb{R}^3, u=\frac{\sigma_1^2y+\sigma_2^2x}{2\sigma_1\sigma_2}\}} \{I_{mdp}(x,y,z)\},$$

and

$$I_{mdp}^{\Xi^{-1}}(u) := \inf_{\{(x,y,z)\in\mathbb{R}^3, u=-\frac{\sigma_1^2y+\sigma_2^2x}{2\sigma_1^3\sigma_2^3}\}} \{I_{mdp}(x,y,z)\},$$

where I_{mdp} is given in (2.7).

By some simple calculations, we have

$$\varrho_1^n(X) - \varrho = \aleph_1^n + \aleph_2^n + \aleph_3^n + \aleph_4^n - \aleph_5^n - \aleph_6^n, \tag{4.37}$$

where

$$\aleph_1^n := (C_1^n - \mathbb{E}C_1^n)\left(\frac{1}{\Xi_1^n} - \frac{1}{\mathbb{E}(\Xi_1^n)}\right), \qquad \aleph_2^n := (C_1^n - \mathbb{E}C_1^n)\frac{1}{\mathbb{E}\Xi_1^n},$$
$$\aleph_3^n := (\mathbb{E}C_1^n - \varrho\mathbb{E}\Xi_1^n)\left(\frac{1}{\Xi_1^n} - \frac{1}{\mathbb{E}\Xi_1^n}\right), \qquad \aleph_4^n := (\mathbb{E}C_1^n - \varrho\mathbb{E}\Xi_1^n)\frac{1}{\mathbb{E}\Xi_1^n},$$
$$\aleph_5^n := \varrho(\Xi_1^n - \mathbb{E}\Xi_1^n)\left(\frac{1}{\Xi_1^n} - \frac{1}{\mathbb{E}\Xi_1^n}\right), \qquad \aleph_6^n := \varrho(\Xi_1^n - \mathbb{E}\Xi_1^n)\frac{1}{\mathbb{E}\Xi_1^n}.$$

To prove the Theorem 3.3, we have to use the Lemma 5.3 and prove some negligibility in the sence of MDP:

$$\frac{\sqrt{n}}{b_n}\aleph_1^n \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.38}$$
$$\frac{\sqrt{n}}{b_n}\aleph_3^n \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.39}$$
$$\frac{\sqrt{n}}{b_n}\aleph_4^n \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.40}$$
$$\frac{\sqrt{n}}{b_n}\aleph_5^n \xrightarrow[b_n^2]{\text{superexp}} 0. \tag{4.41}$$

Since $\mathbb{E}C_1^n - \varrho\mathbb{E}\Xi_{1,n} \to 0$ as $n \to \infty$, (4.40) follows.

We have for all $\delta > 0$

$$\mathbb{P}\left(\frac{\sqrt{n}}{b_n}|\aleph_1^n| \geqslant \delta\right) \leqslant \mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|C_1^n - \mathbb{E}C_1^n\right| \geqslant \alpha_n\right) + \mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|\frac{1}{\Xi_{1,n}} - \frac{1}{\mathbb{E}\Xi_{1,n}}\right| \geqslant \alpha_n\right),$$

where $\alpha_n = \sqrt{\frac{\sqrt{n}}{b_n}\delta}$.

So, by the Lemma 1.2.15 in [13], we have that for all $\delta > 0$

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}|\aleph_1^n| \geqslant \delta\right)$$

is majorized by the maximum of the following two limits

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|C_1^n - \mathbb{E}C_1^n\right| \geqslant \alpha_n\right), \tag{4.42}$$
$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|\frac{1}{\Xi_{1,n}} - \frac{1}{\mathbb{E}\Xi_{1,n}}\right| \geqslant \alpha_n\right). \tag{4.43}$$

Let $A > 0$ be arbitrary, since $\alpha_n \to \infty$ as $n \to \infty$, so for n large enough we obtain that

$$\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|C_1^n - \mathbb{E}C_1^n\right| \geqslant \alpha_n\right) \leqslant \frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|C_1^n - \mathbb{E}C_1^n\right| \geqslant A\right).$$

By the MDP of $\frac{\sqrt{n}}{b_n}(C_1^n - \mathbb{E}C_1^n)$ obtained in Theorem 2.3 in [15], and by letting n to infinity, we obtain that

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{b_n}\left|C_1^n - \mathbb{E}C_1^n\right| \geqslant \alpha_n\right) \leqslant -\inf_{|x|\geqslant A} I_{mdp}^C(x).$$

Letting A gos to the infinity, we obtain that the term in (4.42) goes to $-\infty$.

By the MDP of $\frac{\sqrt{n}}{b_n}(\frac{1}{\Xi_{1,n}} - \frac{1}{\mathbb{E}(\Xi_{1,n})})$ stated before and in the same way we obtain that the term in (4.43) goes to $-\infty$. So we obtain (4.38).

The same calculations give us (4.39) and (4.41).

So

$$\frac{\sqrt{n}}{b_n}(\varrho_1^n(X)-\varrho)$$

and

$$\frac{\sqrt{n}}{b_n}\left(C_1^n-\mathbb{E}C_1^n-\varrho(\Xi_1^n-\mathbb{E}\Xi_1^n)\right)\frac{1}{\mathbb{E}\Xi_1^n}$$

satisfies the same MDP.

Since $\mathbb{E}(\Xi_{1,n})\longrightarrow\sigma_1\sigma_2$, so

$$\frac{\sqrt{n}}{b_n}(\varrho_1^n-\varrho)\quad\text{and}\quad\frac{\sqrt{n}}{b_n}\left(C_1^n-\mathbb{E}C_1^n-\varrho(\Xi_1^n-\mathbb{E}\Xi_1^n)\right)\frac{1}{\sigma_1\sigma_2}$$

satisfies the same MDP.

Then by the Lemma 5.3 applied to the function $\phi:(x,y,z)\mapsto(z-\varrho\sqrt{x}\sqrt{y})/\sigma_1\sigma_2$, we deduce that $\frac{\sqrt{n}}{b_n}(\varrho_1^n(X)-\varrho)$ satisfies the LDP on $\mathbb{R}$ with speed b_n^2 and with the rate function given by

$$I_{mdp}^{\varrho}(u)=\inf_{\{(x,y,z)\in\mathbb{R}^3:u=\frac{z}{\sigma_1\sigma_2}-\varrho\frac{\sigma_1^2y+x\sigma_2^2}{2\sigma_1^2\sigma_2^2}\}}I_{mdp}(x,y,z),$$

where I_{mdp} is given in (2.7).

4.9. **Proof of Proposition 3.7.**

By the Lemma 5.3 applied to the function $f:x\mapsto\frac{1}{x}$, $\frac{\sqrt{n}}{b_n}(\frac{1}{Q_{1,1}^n}-\frac{1}{\mathbb{E}(Q_{1,1}^n)})$ satisfies the LDP on $\mathbb{R}$ with the same speed b_n^2 and with the rate function given by

$$I_{mdp}^{Q_1^{-1}}(u):=\inf_{\{(x,y,z)\in\mathbb{R}^3:u=-\frac{x}{\sigma_1^4}\}}\{I_{mdp}(x,y,z)\}$$

By some simple calculations, we have

$$\beta_{1,1}^n(X)-\varrho\frac{\sigma_2}{\sigma_1}=\jmath_1^n+\jmath_2^n+\jmath_3^n+\jmath_4^n-\jmath_5^n-\jmath_6^n, \tag{4.44}$$

where

$$\jmath_1^n:=(C_1^n-\mathbb{E}C_1^n)\left(\frac{1}{Q_{1,1}^n}-\frac{1}{\mathbb{E}Q_{1,1}^n}\right),\qquad \jmath_2^n:=(C_1^n-\mathbb{E}C_1^n)\frac{1}{\mathbb{E}Q_{1,1}^n},$$

$$\jmath_3^n:=(\mathbb{E}C_1^n-\varrho\frac{\sigma_2}{\sigma_1}\mathbb{E}Q_{1,1}^n)\left(\frac{1}{Q_{1,1}^n}-\frac{1}{\mathbb{E}Q_{1,1}^n}\right),\qquad \jmath_4^n:=(\mathbb{E}C_1^n-\varrho\frac{\sigma_2}{\sigma_1}\mathbb{E}Q_{1,1}^n)\frac{1}{\mathbb{E}Q_{1,1}^n},$$

$$\jmath_5^n:=\varrho\frac{\sigma_2}{\sigma_1}\left(Q_{1,1}^n-\mathbb{E}Q_{1,1}^n\right)\left(\frac{1}{Q_{1,1}^n}-\frac{1}{\mathbb{E}Q_{1,1}^n}\right),\qquad \jmath_6^n:=\varrho\frac{\sigma_2}{\sigma_1}\left(Q_{1,1}^n-\mathbb{E}Q_{1,1}^n\right)\frac{1}{\mathbb{E}Q_{1,1}^n}.$$

To prove the Theorem 3.7, we have to use the Lemma 5.3 and prove some negligibility in the sense of MDP:

$$\frac{\sqrt{n}}{b_n}\jmath_1^n\xrightarrow[b_n^2]{\text{superexp}}0,\quad\frac{\sqrt{n}}{b_n}\jmath_3^n\xrightarrow[b_n^2]{\text{superexp}}0,\quad\frac{\sqrt{n}}{b_n}\jmath_4^n\xrightarrow[b_n^2]{\text{superexp}}0,\quad\frac{\sqrt{n}}{b_n}\jmath_5^n\xrightarrow[b_n^2]{\text{superexp}}0. \tag{4.45}$$

The same calculations as for the negligibility of $\aleph_j^n$ works here to obtain (4.45).

Since $\mathbb{E}Q^n_{1,1} \longrightarrow \sigma_1^2$, we deduce that

$$\frac{\sqrt{n}}{b_n}(\beta^n_{1,1}(X) - \varrho\frac{\sigma_2}{\sigma_1}))$$

and

$$\frac{\sqrt{n}}{b_n}(\left(C^n_1 - \mathbb{E}C^n_1 - \varrho\frac{\sigma_2}{\sigma_1}(Q^n_{1,1} - \mathbb{E}Q^n_{1,1}\right)\frac{1}{\sigma_1^2}$$

satisfies the same MDP.

Then by the Lemma 5.3 applied to the function $k : (x, y, z) \mapsto (z - \varrho\frac{\sigma_2}{\sigma_1}x)/\sigma_1^2$ we deduce that $\frac{\sqrt{n}}{b_n}(\beta^n_{1,1}(X) - \varrho\frac{\sigma_2}{\sigma_1})$ satisfies the LDP on $\mathbb{R}$ with speed b_n^2 and with the rate function given by

$$I^{\beta_{1,1}}_{mdp}(u) = \inf_{\{(x,y,z)\in\mathbb{R}^3 : u=(z-\varrho\frac{\sigma_2}{\sigma_1}x)/\sigma_1^2\}} I_{mdp}(x, y, z)$$

where I_{mdp} is given in (2.7).

5. Appendix

The proofs of the LDP in Theorems 2.1 and 2.3 are respectively based on the Lemmas 5.1 and 5.2 that we will present here for completeness.

5.1. **Avoiding Gärtner-Ellis theorem by Najim** [26, 25]**.** Let us introduce some notations and assumptions in this section.

Let $\mathcal{X}$ be a topological vector compact space endowed with it's Borel $\sigma-$field $\mathcal{B}(\mathcal{X})$. Let $\mathcal{B}V([0, 1], \mathbb{R}^d)$, (shortened in $\mathcal{B}V$) be a space of functions of bounded variation on $[0, 1]$ endowed with it's Borel $\sigma-$field $\mathcal{B}_w$. Let $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$.

Let $\tau(z) = e^{|z|} - 1, z \in \mathbb{R}^d$ and let us consider

$$\begin{aligned}\mathcal{P}_\tau(\mathbb{R}^d) &= \left\{P \in \mathcal{P}(\mathbb{R}^d), \exists a > 0; \int_{\mathbb{R}^d} \tau\left(\frac{z}{a}\right) P(dz) < \infty\right\} \\ &= \left\{P \in \mathcal{P}(\mathbb{R}^d), \exists \alpha > 0; \int_{\mathbb{R}^d} e^{\alpha|z|} P(dz) < \infty\right\}.\end{aligned}$$

$\mathcal{P}_\tau$ is the set of probability distributions having some exponential moments.We denote by $M(P, Q)$ the set of all laws on $\mathbb{R}^d \times \mathbb{R}^d$ with given marginals P and Q. We introduce the Orlicz-Wasserstein distance defined on $\mathcal{P}_\tau(\mathbb{R}^d)$ by

$$d_{OW}(P, Q) = \inf_{\eta\in M(P,Q)} \inf\left\{a > 0; \int_{\mathbb{R}^d\times\mathbb{R}^d} \tau\left(\frac{z - z'}{a}\right)\eta(dzdz') \leqslant 1\right\}$$

Let $(Z^n_i)_{1\leqslant i\leqslant n, n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d-$valued independent random variables satisfying:

N-1

$$\mathbb{E}e^{\alpha.||Z||} < +\infty \qquad for \quad some \quad \alpha > 0.$$

N-2 Let $(x^n_i, 1 \leqslant i \leqslant n, n \geqslant 1)$ be a $\mathcal{X}-$valued sequence of elements satisfying:

$$\frac{1}{n}\sum_{j=1}^{n} \delta_{x^n_j} \xrightarrow[n\to\infty]{weakly} \quad R.$$

Where R is Assumed to be a strictly positive probability measure, that is $R(U) > 0$ whenever U is a nonempty open subset of $\mathcal{X}$.

N-3 $\mathcal{X}$ is a compact space.

N-4 There exist a family of probability measure $(P(x,\cdot), x \in \mathcal{X})$ over $\mathbb{R}^d$ and a sequence $(x_i^n, 1 \leqslant i \leqslant n, n \geqslant 1)$ with values in $\mathcal{X}$ such that the law of each Z_i^n is given by:

$$\mathcal{L}(Z_i^n) \sim P(x_i^n, dz).$$

We will equally write $P(x,\cdot), P_x \quad or \quad P_x(dz)$.

N-5 Let $(P(x,\cdot), x \in \mathcal{X}) \subset \mathcal{P}_\tau(\mathbb{R}^d)$ be a given distribution. The application $x \mapsto P(x, A)$ is measurable whenever the set $A \subset \mathbb{R}^d$ is borel. Morever, the function

$$\begin{aligned} \Gamma : &\mathcal{X} \to \mathcal{P}_\tau(\mathbb{R}^d) \\ &x \mapsto P(x,\cdot) \end{aligned}$$

is continuous when $\mathcal{P}_\tau(\mathbb{R}^d)$ is endowed with the topology induced by the distance d_{OW}

Lemma 5.1. *Theorem 2.2 in* [26]

Assume that Z_i^n are independent and identically distributed, so we denote Z_i^n by Z_i.

Assume that (N-1) and (N-2) hold. Let $f : \mathcal{X} \to \mathbb{R}^{m\times d}$ be a (matrix-valued) bounded continuous function, such that

$$f(x)\cdot z = \begin{pmatrix} f_1(x)\cdot z \\ \\ \vdots \\ \\ f_m(x)\cdot z \end{pmatrix}$$

where each $f_j \in C_d(\mathcal{X})$ is the j^{th} row of the matrix f.

Then the family of the weighted empirical mean

$$\langle L_n, f\rangle := \frac{1}{n}\sum_{i=1}^{n} f(x_i^n)\cdot Z_i$$

satisfies the LDP in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ with speed n and the good rate function

$$I_f(x) = \sup_{\theta\in\mathbb{R}^m}\{\langle\theta, x\rangle - \int_{\mathcal{X}} \Lambda[\sum_{i=1}^{m}\theta_i\cdot f_i(x)]\mathrm{R}(dx)\} \qquad \forall x\in\mathbb{R}^m$$

where Λ denote the cumulant generating function of Z

$$\Lambda(\lambda) = \log\mathbb{E}e^{\lambda\cdot Z} \qquad for \quad \lambda\in\mathbb{R}^d$$

Lemma 5.2. *Theorem 4.3 in* [27]

Assume that (N-1), (N-2), (N-3), (N-4) and (N-5) hold. Then the family of random functions

$$\overline{Z}_n(t) = \frac{1}{n}\sum_{k=1}^{[nt]} Z_k^n, \quad t \in [0,1]$$

satisfies the LDP in $(\mathcal{BV}, \mathcal{B}_w)$ *with the good rate function*

$$\phi(f) = \int_{[0,1]} \Lambda^*(x, f_a'(x))\mathrm{d}x + \int_{[0,1]} \rho(x, f_s'(x))\mathrm{d}\theta(x)$$

where θ *is any real-valued nonnegative measure with respect to which* μ_s^f *is absolutely continuous and* $f_s' = \frac{d\mu_s^f}{d\theta}$, *where*

$$\Lambda^*(x,z) = \sup_{\lambda\in\mathbb{R}^d}\left\{\langle\lambda, z\rangle - \Lambda(x,\lambda)\right\}, \quad \forall z \in \mathbb{R}^d$$

with $\Lambda(x,\lambda) = \log\int_{\mathbb{R}^d} e^{\langle\lambda,z\rangle}P(x,dz), \quad \forall\lambda\in\mathbb{R}^d$ *and the recession function* $\rho(x,z)$ *of* $\Lambda^*(x,z)$ *defined by:* $\rho(x,z) = \sup\{\langle\lambda, z\rangle, \lambda\in D_x\}$ *with* $D_x = \{\lambda\in\mathbb{R}^d, \Lambda(x,\lambda) < \infty\}$.

5.2. **Delta method for large deviations** [17]**.** In this section, we recall the delta method in large deviation.

Let $\mathcal{X}$ and $\mathcal{Y}$ be two metrizable topological linear spaces. A function ϕ defined on a subset $\mathcal{D}_\phi$ of $\mathcal{X}$ with values on $\mathcal{Y}$ is called Hadamard differentiable at x if there exists a continuous functions $\phi' : \mathcal{X} \mapsto \mathcal{Y}$ such that

$$\lim_{n\to\infty}\frac{\phi(x+t_nh_n)-\phi(x)}{t_n} = \phi'(h) \tag{5.1}$$

holds for all t_n converging to $0+$ and h_n converging to h in $\mathcal{X}$ such that $x + t_nh_n \in \mathcal{D}_\phi$ for every n.

Lemma 5.3. *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be two metrizable topological linear spaces. Let* $\phi : \mathcal{D}\subset\mathcal{X}\mapsto\mathcal{Y}$ *be a Hadamard differentiable at* θ *tangentially to* $\mathcal{D}_0$, *where* $\mathcal{D}_\phi$ *and* $\mathcal{D}_0$ *are two subset of* $\mathcal{X}$. *Let* $\{(\Omega_n, \mathcal{F}_n, \mathbb{P}_n), n\geqslant 1\}$ *be a sequence of probability space and let* $\{X_n, n\geqslant 1\}$ *be a sequence of maps from from* Ω_n *to* $\mathcal{D}_\phi$ *and let* $\{r_n, n\geqslant 1\}$ *be a sequence of positive real numbers satisfying* $r_n\to+\infty$ *and let* $\{\lambda(n), n\geqslant 1\}$ *be a sequence of positive real numbers satisfying* $\lambda(n)\to+\infty$.
If $\{r_n(X_n-\theta), n\geqslant 1\}$ *satifies the LDP with speed* $\lambda(n)$ *and rate function* I *and* $\{I<\infty\}\subset\mathcal{D}_0$, *then* $\{r_n(\phi(X_n)-\phi(\theta)), n\geqslant 1\}$ *satifies the LDP with speed* $\lambda(n)$ *and rate function* $I_{\phi_\theta'}$, *where*

$$I_{\phi_\theta'}(y) = inf\{I(x); \phi_\theta'(x) = y\}, \qquad y\in\mathcal{Y} \tag{5.2}$$

References

[1] Aït-Sahalia, Y., Fan, J., and Xui, D. High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association, 105*, 492 (2010), 1504–1517.

[2] Andersen, T. G., Bollerslev, T., Diebold, T. X., and Wu, G. Realized beta: persistence and predictability. *Econometric analysis of financial and economic time series. Part B, Adv. Econom. 20* (2008), 1–39.

[3] B. Bercu, B., Gamboa, F., and Rouault, A. Large deviations for quadratic forms of stationary gaussian processes. *Stochastic Processes and their Applications, 71* (1997), 75–90.

[4] Barndorff-Nielsen, O. E., Graversen, S., Jacod, J., Podolskij, M., and Shephard, N. A central limit theorem for realised power and bipower variations of continuous semimartingales. *In: Y. Kabanov, R. Lispter (Eds.), Stochastic Analysis to Mathematical Finance, Frestchrift for Albert Shiryaev, Springer.* (2006), 33–68.

[5] Barndorff-Nielsen, O. E., Graversen, S., Jacod, J., and Shephard, N. Limit theorems for bipower variation in financial econometrics. *Econometric Theory, 22* (2006), 677–719.

[6] Barndorff-Nielsen, O. E., and Shephard, N. Econometric analysis of realized volatility and its use in estimating volatility models. *Journal of the Royal Statistical Society, Series B 64* (2002), 253–280.

[7] Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and their Applications, 116* (2006), 796–806.

[8] Bibinger, M. An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: asymptotic distribution theory. *Stochastic Processes and their Applications, 122*, 6 (2012), 2411–2453.

[9] Bibinger, M., and Reiss, M. Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics. Theory and Applications, 41*, 1 (2014), 23–50.

[10] Bryc, W. ans Dembo, A. Large deviations for quadratic functionals of gaussian processes. *Journal of Theoretical Probability, 10* (1997), 307–322.

[11] Comte, F., Genon-Catalot, V., and Rozenholc, Y. Nonparametric estimation for a stochastic volatility model. *Finance Stoch. 14*, 1 (2010), 49–80.

[12] Dalalyan, A., and Yoshida, N. Second-order asymptotic expansion for a non-synchronous covariation estimator. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques 47*, 3 (2011), 748–789.

[13] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications. Second edition,.* Springer, 1998.

[14] Djellout, H., Guillin, A., and Wu, L. Large and moderate deviations for estimators of quadratic variational processes of diffusions. *Statistical Inference for Stochastic Processes,*, 2 (1999), 195–225.

[15] Djellout, H., and Samoura, Y. Large and moderate deviations of realized covolatility. *Statistics & Probability Letters, 86* (2014), 30–37.

[16] Dovonon, P., Gonçalves, S., and Meddahi, N. Bootstrapping realized multivariate volatility measures. *Journal of Econometrics, 172*, 1 (2013), 49–65.

[17] Gao, F., and Zhao, X. Delta method in large deviations and moderate deviations for estimators. *The Annals of Statistics, 39*, 2 (2011), 1211–1240.

[18] Gloter, A. Efficient estimation of drift parameters in stochastic volatility models. *Finance Stoch. 11*, 4 (2007), 495–519.

[19] Hayashi, T., and Yoshida, N. Nonsynchronous covariation process and limit theorems. *Stochastic Processes and their Applications, 121*, 10 (2011), 2416–2454.

[20] Jiang, H. Moderate deviations for estimators of quadratic variational process of diffusion with compound poisson. *Statistics & Probability Letters, 80*, 17-18 (2010), 1297–1305.

[21] Ledoux, M. Concentration of measure and logarithmic sobolev inequalities. *Séminaire de Probabilités, XXXIII, Lecture Notes in Math. Springer, Berlin,*, 1709 (1999), 120–216.

[22] Lynch, J., and Sethuraman, J. Large deviations for processes with independent increments. *The Annals of Probability 15*, 2 (1987), 610–627.

[23] Mancini, C. Large deviation principle for an estimator of the diffusion coefficient in a jump-diffusion process. *Statistics & Probability Letters, 78*, 7 (2008), 869–879.

[24] Mancini, C., and Gobbi, F. Identifying the brownian covariation from the co-jumps given discrete observations. *Econometric Theory, 28*, 2 (2012), 249–273.

[25] Najim, J. Grandes déviations pour certaines mesures empiriques. *PhD thesis, Université Nanterre Paris-10* (2001), 1–130.

[26] Najim, J. A cramer type theorem for weighted random variables. *Electronic Journal of Probability, 7*, 4 (2002), 1–32.

[27] Najim, J. Large deviations for independent random variables application to Erdős-Renyi's functional law of large numbers. *ESAIM. Probability and Statistics, 9* (2005), 116–142 (electronic).

[28] Perrin, O., and Zani, M. Large deviations for sample paths of Gaussian processes quadratic variations. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 328*, Veroyatn. i Stat. 9 (2005), 169–181, 280.

[29] Revuz, C., and Yor, M. *Continuous martingales and Brownian motion. Third edition.* Springer-Verlag, Berlin, 1999.

[30] Robert, C. Y., and Rosenbaum, M. Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance, 22*, 1 (2012), 133–164.

[31] Rockafellar, R. T. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[32] Shen, S. Large deviation for the empirical correlation coefficient of two gaussian random variables. *Acta Mathematica Scientia. Series B. English Edition, 27*, 4 (2007), 821–828.

[33] Shin, K., and Otsu, T. Large deviations of realized volatility. *Stochastic Processes and their Applications, 122*, 2 (2012), 546–581.

[34] Todorov, V., and Bollerslev, T. Jumps and betas: a new framework for disentangling and estimating systematic risks. *Journal of Econometrics, 157*, 2 (2010), 220–235.

[35] Wu, L. An introduction to large deviation. in: J. a. yan, s. peng, s. fang and l. wu (eds). *Several topics in stochastic analysis. Academic Press of China, Beijing* (1997), 225–336.

[36] Zani, M. Grandes déviations pour des fonctionnelles issues de la statistique des processus. *PhD thesis, Université Paris-Sud*, 4 (1999), 1–151.

[37] Zhang, L., Mykland, P. M., and Aït-Sahalia. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association, 100* (2005), 1394–1411.

[38] Zhang, L., Mykland, P. M., and Aït-Sahalia, Y. Edgeworth expansions for realized volatility and related estimators. *Journal of Econometrics, 160* (2011), 190–203.

E-mail address: Hacene.Djellout@math.univ-bpclermont.fr

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais, 63177 Aubière, France.

E-mail address: Arnaud.Guillin@math.univ-bpclermont.fr

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais, 63177 Aubière, France.

E-mail address: Yacouba.Samoura@math.univ-bpclermont.fr

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais, 63177 Aubière, France.

LARGE DEVIATIONS OF THE THRESHOLD ESTIMATOR OF INTEGRATED (CO-)VOLATILITY VECTOR IN THE PRESENCE OF JUMPS

HACÈNE DJELLOUT AND HUI JIANG

ABSTRACT. Recently a considerable interest has been paid on the estimation problem of the realized volatility and covolatility by using high-frequency data of financial price processes in financial econometrics. Threshold estimation is one of the useful techniques in the inference for jump-type stochastic processes from discrete observations. In this paper, we adopt the threshold estimator introduced by Mancini [18] where only the variations under a given threshold function are taken into account. The purpose of this work is to investigate large and moderate deviations for the threshold estimator of the integrated variance-covariance vector. This paper is an extension of the previous work in Djellout et al [11]. where the problem has been studied in absence of the jump component. We will use the approximation lemma to prove the LDP. As the reader can expect we obtain the same results as in the case without jump.

AMS 2000 subject classifications: 60F10, 62J05, 60J05.

1. Motivation and context

On a filtred probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,1]}, \mathbb{P})$, we consider $X_1 = (X_{1,t})_{t\in[0,1]}$ and $X_2 = (X_{2,t})_{t\in[0,1]}$ two real processes defined by a Lévy jump-diffusion constructed via the superposition of a Wiener process with drift and an independent compound Poisson process. This is one of the first and simplest extensions to the classical geometric Brownian motion underlying the famous Black-Scholes-Merton framework for option pricing.

More precisely, $X_1 = (X_{1,t})_{t\in[0,1]}$ and $X_2 = (X_{2,t})_{t\in[0,1]}$ are given by

$$\begin{cases} dX_{1,t} = b_1(t,\omega)dt + \sigma_{1,t}dW_{1,t} + dJ_{1,t} \\ dX_{2,t} = b_2(t,\omega)dt + \sigma_{2,t}dW_{2,t} + dJ_{2,t} \end{cases} \tag{1.1}$$

for $t \in [0,1]$ where $W_1 = (W_{1,t})_{t\in[0,1]}$ and $W_2 = (W_{2,t})_{t\in[0,1]}$ are two correlated Wiener processes, with $\rho_t = \mathrm{Cov}(W_{1,t}, W_{2,t}), t \in [0,1]$. We can write $W_{2,t} = \rho_t dW_{1,t} + \sqrt{1-\rho_t^2}dW_{3,t}$, where $W_1 = (W_{1,t})_{t\in[0,1]}$ and $W_3 = (W_{3,t})_{t\in[0,1]}$ are independent Wiener processes. J_1 and J_2 are possibly correlated pure jump processes. We assume here that J_1 and J_2 have finite jump activity, that is a.s. there are only finitely many jumps on any finite time interval. A general Lévy model would contain also a compensated infinte activity pure jump component.

Under our assumption J_ℓ is necessarily a compound Poisson processe and it can be written as

$$J_{\ell,s} = \sum_{i=1}^{N_{\ell,s}} Y_{\ell,i}, \qquad s \in [0,1].$$

Date: May 1, 2015.

Key words and phrases. Moderate deviation principle, Large deviation principle, Diffusion, Discrete-time observation, Quadratic variation, Realised volatility, Lévy process, Threshold estimator, Jump Poisson.

Here $Y_{\ell,i}$ are i.i.d. real random variables having law ν_ℓ/λ_ℓ, where ν_ℓ is the Lévy measure of X_ℓ normalized by the total mass $\lambda_\ell = \nu_\ell(\mathbb{R} - \{0\}) < +\infty$, and N_ℓ is a poisson process, independent of each $Y_{\ell,i}$, and with constant intensity λ_ℓ.

Such a jump-type stochastic process is recently a standard tool, e.g., for modeling asset values in finance and insurance. The key motivation behind jump-diffusion models is the incorporation of market "stocks", which result in "large" and sudden changes in the price of risky security and which can hardly be modeled by the diffusive component.

In this paper we concentrate on the estimation of

$$[\mathcal{V}]_t = \left(\int_0^t \sigma_{1,s}^2 ds, \int_0^t \sigma_{2,s}^2 ds, \int_0^t \sigma_{1,s}\sigma_{2,s}\rho_s ds\right)$$

Over the last decade, several estimation methods for the integrated variance-covariance $\mathcal{V}_t$ have been proposed. We adopt the threshold estimator which is introduced by Mancini [18] and also by Shimizu and Yoshida [26], independently.

In this method, only the variations under a given threshold function are taken into account. The specific estimator excludes all terms containing jumps from the realized co-variation while remaining consistent, efficient and robust when synchronous data are considered.

Since the seminal work of Mancini [18], several authors have leveraged or extended the thresholding cencept to deal with complex stochastic models, see Shimizu and Yoshida [26], or Ogihara and Yoshida [22]. The similar idea is also used by various authors in different contexts; see, e.g., Aït-Sahalia et al. [1], [2] and [3], Gobbi and Mancini [15] , Cont and Mancini [21] , among others.

So, given the synchronous and evenly-spaced observation of the process $X_{1,t_0}, X_{1,t_1}, \cdots, X_{1,t_n}$, $X_{2,t_0}, X_{2,t_1} \cdots, X_{2,t_n}$ with $t_0 = 0, t_n = 1, n \in \mathbb{N}$, we consider the following statistics

$$\left(\sum_{k=1}^{[nt]} (\Delta_k^n X_1)^2, \sum_{k=1}^{[nt]} (\Delta_k^n X_2)^2, \sum_{k=1}^{[nt]} \Delta_k^n X_1 \Delta_k^n X_2\right)$$

where $\Delta_k^n X_\ell := X_{\ell,t_k} - X_{\ell,t_{k-1}}$. However this estimate can be highly biased when the processes X_ℓ contain jumps, in fact, as $n \to \infty$ such a sum approaches the global quadratic variance-covariation

$$([X_1]_t, [X_2]_t, [X_1, X_2]_t)$$

where

$$[X_\ell]_t := \int_0^t \sigma_{\ell,s}^2 ds + \sum_{s \le t} (\Delta J_{\ell,s})^2, \quad \text{and} \quad [X_1, X_2]_t := \int_0^t \sigma_{1,s}\sigma_{2,s}\rho_s ds + \sum_{s \le t} \Delta J_{1,s} \Delta J_{2,s}.$$

which also contain the co-jumps, where $\Delta J_{\ell,s} = J_{\ell,s} - J_{\ell,s^-}$.

If we take a deterministic function $r(\frac{1}{n})$ at the step $\frac{1}{n}$ between the observations, such that

$$\lim_{n\to\infty} r\left(\frac{1}{n}\right) = 0, \qquad \text{and} \quad \lim_{n\to\infty} \frac{\log n}{nr\left(\frac{1}{n}\right)} = 0.$$

The function $r(\cdot)$ is a threshold such that whenever $|\Delta_k^n X_\ell|^2 > r(\frac{1}{n})$, a jump has to occur within $]t_{k-1}, t_k]$. Hence we can recover $[\mathcal{V}]_t$ using the following threshold estimator

$$\mathcal{V}_t^n(X) = (\mathcal{Q}_{1,t}^n(X), \mathcal{Q}_{2,t}^n(X), \mathcal{C}_t^n(X))$$

where

$$\mathcal{Q}_{\ell,t}^n(X) = \sum_{k=1}^{[nt]} (\Delta_k^n X_\ell)^2 \mathbf{1}_{\{(\Delta_k^n X_\ell)^2 \leq r(\frac{1}{n})\}}$$

and

$$\mathcal{C}_t^n(X) = \sum_{k=1}^{[nt]} \Delta_k^n X_1 \Delta_k^n X_2 \mathbf{1}_{\{\max_{\ell=1}^2 (\Delta_k^n X_\ell)^2 \leq r(\frac{1}{n})\}}$$

In the work [14], the authors determine what constitutes a good threshold sequence r_n and they propose an objective method for selecting such a sequence.

In the case that X_ℓ have no jumps, this question has been well investigated. The problem of the large deviation of the quadratic estimator of the integrated volatility (without jumps and in the case of synchronous sampling scheme) is obtained in the paper by Djellout et al. [12] and recently Djellout and Samoura [13] have studied the large deviation for the covariance estimator. Djellout et al. [11] have also investigated the problem of the large deviation for the realized (co-)volatility vector which allows them to provide the large deviation for the standard dependence measures between the two assets returns such as the realized regression coefficients, or the realized correlation.

However, the inclusion of jumps within financial models seems to be more and more necessary for pratical applications. In this case, Mancini [21] has shown that $\mathcal{V}_t^n$ is a consistent estimators of $\mathcal{V}_t$ and has some asymtotic normality respectively. Furthermore, when $\sigma_t = \sigma$, she [19] studied the large deviation for the threshold estimator. Jiang [16] obtained moderate deviations and functional moderate deviations for threshold estimator. In our paper and by the method as in Mancini [19] and Djellout et al [11], we consider moderate and functionnal moderate deviation for estimators V_t^n and large deviation.

More precisely we are interested in the estimations of

$$\mathbb{P}\left(\frac{\sqrt{n}}{v_n}(\mathcal{V}_t^n(X) - [\mathcal{V}]_t) \in A\right)$$

where A is a given domain of deviation, $(v_n)_{n>0}$ is some sequence denoting the scale of deviation. When $v_n = 1$ this is exactly the estimation of central limit theorem. When $v_n = \sqrt{n}$, it becomes the large deviation. Furthermore, when $v_n \to \infty$ and $v_n = o(\sqrt{n})$, this is the so called moderate deviations. In other words, the moderate deviations investigate the convergence speed between the large deviations and central limit theorem.

Let us recall some basic defintions in large deviations theory. Let $(\mu_t)_{t>0}$ be a family of probability on a topological space $(S, \mathcal{S})$ where $\mathcal{S}$ is a σ-algebra on S and λ_t be a nonnegative function on $[1, +\infty[$ such that $\lim_{t\to\infty} \lambda_t = +\infty$. A function $I : S \to [0, +\infty]$ is said to be a rate function if it is lower semicontinuous and it is said to be a good rate function if its level set $\{x \in S; I(x) \leq a\}$ is a compact for all $a \geq 0$.

(μ_t) is said to satisfy a large deviation principle with speed λ_t and rate function I if for any closed set $F \in \mathcal{S}$

$$\limsup_{t\to\infty} \frac{1}{\lambda_t} \log \mu_t(F) \leq -\inf_{x\in F} I(x)$$

and for any open set $G \in \mathcal{S}$

$$\limsup_{t\to\infty} \frac{1}{\lambda_t} \log \mu_t(G) \geq - \inf_{x\in G} I(x).$$

Notations. *In the whole paper, for any matrix M, M^T and $\|M\|$ stand for the transpose and the euclidean norm of M, respectively. For any square matrix M,* $\det(M)$ *is the determinant of M. Moreover, we will shorten* large deviation principle *by LDP and* moderate deviation principle *by MDP. We denote by $\langle \cdot, \cdot \rangle$ the usual scalar product. For any process Z_t, $\Delta_s^t Z$ stands for the increment $Z_t - Z_s$. We use $\Delta_k^n Z$ for $\Delta_{t_{k-1}^n}^{t_k^n} Z$. In addition, for a sequence of random variables $(Z_n)_n$ on $\mathbb{R}^{d\times p}$, we say that $(Z_n)_n$ converges $(\lambda_n)-$superexponentially fast in probability to some random variable Z if, for all $\delta > 0$,*

$$\limsup_{n\to\infty} \frac{1}{\lambda_n} \log \mathbb{P}\Big(\|Z_n - Z\| > \delta \Big) = -\infty.$$

This exponential convergence *with speed λ_n will be shortened as*

$$Z_n \quad \underset{\lambda_n}{\overset{\text{superexp}}{\longrightarrow}} Z.$$

The article is arranged in two upcoming sections. Section 2 is devoted to our main results on the LDP and MDP for the (co-)volatility vector in the presence of jumps. In section 3, we give the proof of these theorems.

2. Main results

Let $X_t = (X_{1,t}, X_{2,t})$ be given by (1.1). We introduce the following conditions

(B) for $\ell = 1, 2$ $b(\cdot, \cdot) \in L^\infty(dt \otimes \mathbb{P})$

(LDP) Assume that for $\ell = 1, 2$

- $\sigma_{\ell,t}^2(1-\rho_t^2)$ and $\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2) \in L^\infty([0,1], dt)$.
- the functions $t \to \sigma_{\ell,t}$ and $t \to \rho_t$ are continuous.
- let r such that

$$r\left(\frac{1}{n}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad nr\left(\frac{1}{n}\right) \underset{n\to\infty}{\longrightarrow} \infty.$$

(MDP) Assume that for $\ell = 1, 2$

- $\sigma_{\ell,t}^2(1-\rho_t^2)$ and $\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2) \in L^2([0,1], dt)$.
- Let $(v_n)_{n\geqslant 1}$ be a sequence of positive numbers such that

$$v_n \underset{n\to\infty}{\longrightarrow} \infty \quad \text{and} \quad \frac{v_n}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \sqrt{n} v_n r\left(\frac{1}{n}\right) = O(1)$$

$$\text{and for} \quad \ell = 1, 2 \quad \frac{r\left(\frac{1}{n}\right)}{\log\left(\frac{n}{v_n^2}\right) \max_{k=1}^{n} \int_{t_{k-1}}^{t_k} \sigma_{\ell,s}^2 ds} \longrightarrow +\infty. \tag{2.1}$$

We introduce the following function, which will play a crucial role in the calculation of the moment generating function: for $-1 < c < 1$ let for any $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$

$$P_c(\lambda) := \begin{cases} -\frac{1}{2}\log\left(\frac{(1-2\lambda_1(1-c^2))(1-2\lambda_2(1-c^2)) - (\lambda_3(1-c^2)+c)^2}{1-c^2}\right) \\ \qquad\qquad if \qquad \lambda \in \mathcal{D} \\ +\infty, \quad otherwise \end{cases} \tag{2.2}$$

where

$$\mathcal{D}_c = \left\{\lambda \in \mathbb{R}^3,\ \max_{\ell=1,2} \lambda_\ell < \frac{1}{2(1-c^2)} \text{ and } \prod_{\ell=1}^{2}\left(1-2\lambda_\ell(1-c^2)\right) > \left(\lambda_3(1-c^2)+c\right)^2\right\}. \tag{2.3}$$

Let us present now the main results.

2.1. **Moderate deviation.** Let us now consider the intermediate scale between the central limit theorem and the law of large numbers.

Theorem 2.1. *For t=1 fixed. Under the conditions* **(MDP)** *and* **(B)**, *the sequence*

$$\frac{\sqrt{n}}{v_n}\left(\mathcal{V}_1^n(X) - [\mathcal{V}]_1\right)$$

satisfies the LDP on $\mathbb{R}^3$ *with speed* v_n^2 *and with rate function given by*

$$I_{mdp}(x) = \sup_{\lambda\in\mathbb{R}^3}\left(\langle\lambda, x\rangle - \frac{1}{2}\langle\lambda, \Sigma_1\cdot\lambda\rangle\right) = \frac{1}{2}\left\langle x, \Sigma_1^{-1}\cdot x\right\rangle \tag{2.4}$$

with

$$\Sigma_1 = \begin{pmatrix} \int_0^1 \sigma_{1,t}^4 \mathrm{d}t & \int_0^1 \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t & \int_0^1 \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t \\ \int_0^1 \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 \mathrm{d}t & \int_0^1 \sigma_{2,t}^4 \mathrm{d}t & \int_0^1 \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t \\ \int_0^1 \sigma_{1,t}^3\sigma_{2,t}\rho_t \mathrm{d}t & \int_0^1 \sigma_{1,t}\sigma_{2,t}^3\rho_t \mathrm{d}t & \int_0^1 \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2)\mathrm{d}t \end{pmatrix}.$$

Remark 2.1. *Under the condition* $b_\ell = 0$*, we can prove that for all* $\theta \in \mathbb{R}^3$

$$\lim_{n\to\infty} \frac{1}{v_n^2}\log\mathbb{E}\left(e^{\sqrt{n}v_n\langle\theta, \mathcal{V}_1^n(X)-[\mathcal{V}]_1\rangle}\right) = \frac{1}{2} < \theta, \Sigma_1\cdot\theta > .$$

This gives an alternative proof of the moderate deviation using Gärtner-Ellis theorem.

Remark 2.2. *If for some* $p > 2$, $\sigma_{1,t}^2$, $\sigma_{2,t}^2$ *and* $\sigma_{1,t}\sigma_{2,t}(1-\rho_t^2) \in L^p([0,1])$ *and* $v_n = O(n^{\frac{1}{2}-\frac{1}{p}})$, *the condition* (2.1) *in* **(MDP)** *is verified.*

Let $\mathcal{H}$ be the banach space of $\mathbb{R}^3$-valued right-continuous-left-limit non decreasing functions γ on $[0,1]$ with $\gamma(0) = 0$, equipped with the uniform norm and the $\sigma-$field $\mathcal{B}^s$ generated by the coordinate $\{\gamma(t), 0 \leqslant t \leqslant 1\}$.

Theorem 2.2. *Under the conditions* **(MDP)** *and* **(B)**, *the sequence*

$$\frac{\sqrt{n}}{v_n}\left(\mathcal{V}^n_\cdot(X) - [\mathcal{V}]_\cdot\right)$$

satisfies the LDP on $\mathcal{H}$ with speed v_n^2 and with rate function given by

$$J_{mdp}(\phi) = \begin{cases} \displaystyle\int_0^1 \frac{1}{2}\left\langle \dot{\phi}(t), \overline{\Sigma}_t^{-1}\cdot\dot{\phi}(t)\right\rangle dt & if \quad \phi \in \mathcal{AC}_0([0,1]) \\ +\infty, \qquad otherwise, \end{cases} \tag{2.5}$$

where

$$\overline{\Sigma}_t = \begin{pmatrix} \sigma_{1,t}^4 & \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 & \sigma_{1,t}^3\sigma_{2,t}\rho_t \\ \sigma_{1,t}^2\sigma_{2,t}^2\rho_t^2 & \sigma_{2,t}^4 & \sigma_{1,t}\sigma_{2,t}^3\rho_t \\ \sigma_{1,t}^3\sigma_{2,t}\rho_t & \sigma_{1,t}\sigma_{2,t}^3\rho_t & \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^2(1+\rho_t^2) \end{pmatrix}$$

is invertible and $\overline{\Sigma}_t^{-1}$ his inverse such that

$$\overline{\Sigma}_t^{-1} = \frac{1}{\det(\overline{\Sigma}_t)} \begin{pmatrix} \frac{1}{2}\sigma_{1,t}^2\sigma_{2,t}^6(1-\rho_t^2) & \frac{1}{2}\sigma_{1,t}^4\sigma_{2,t}^4\rho_t^2(1-\rho_t^2) & -\sigma_{1,t}^3\sigma_{2,t}^5\rho_t(1-\rho_t^2) \\ \frac{1}{2}\sigma_{1,t}^4\sigma_{2,t}^4\rho_t^2(1-\rho_t^2) & \frac{1}{2}\sigma_{1,t}^6\sigma_{2,t}^2(1-\rho_t^2) & -\sigma_{1,t}^5\sigma_{2,t}^3\rho_t(1-\rho_t^2) \\ -\sigma_{1,t}^3\sigma_{2,t}^5\rho_t(1-\rho_t^2) & -\sigma_{1,t}^5\sigma_{2,t}^3\rho_t(1-\rho_t^2) & \sigma_{1,t}^4\sigma_{2,t}^4(1-\rho_t^4) \end{pmatrix},$$

$$with \qquad \det(\overline{\Sigma}_t) = \frac{1}{2}\sigma_{1,t}^6\sigma_{2,t}^6(1-\rho_t^2)^3,$$

and $\mathcal{AC}_0 = \{\phi : [0,1] \to \mathbb{R}^3$ is absolutely continuous with $\phi(0) = 0\}$.

Remark 2.3. *A similar result for the moderate deviations is obtained by Jiang [16] in the jump case for $\left(\frac{\sqrt{n}}{v_n}\left(\mathcal{Q}^n_{\ell,t} - \int_0^t \sigma^2_{\ell,s}ds\right)\right)_{n\geq 1}$.*

2.2. **Large deviation.** Our second result is about the large deviation of $\mathcal{V}_1^n(X)$, i.e. at fixed time.

Theorem 2.3. *Let $t = 1$ be fixed. Under the conditions* **(LDP)** *and* **(B)** , *the sequence $\mathcal{V}_1^n(X)$ satisfies the LDP on $\mathbb{R}^3$ with speed n and with good rate function given by the legendre transformation of Λ, that is*

$$I_{ldp}(x) = \sup_{\lambda\in\mathbb{R}^3}\left(\langle\lambda, x\rangle - \Lambda(\lambda)\right), \tag{2.6}$$

where $\Lambda(\lambda) = \int_0^1 P_{\rho_t}(\lambda_1\sigma_{1,t}^2, \lambda_2\sigma_{2,t}^2, \lambda_3\sigma_{1,t}\sigma_{2,t})dt$.

Remark 2.4. *Under the condition $b_\ell = 0$, we can calculate the moment generating function of $\mathcal{V}_1^n(X)$. We obtain that for all $\theta = (\theta_1, \theta_2, \theta_3)^T \in \mathcal{D}_{\rho_t}$*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left(e^{n\langle \theta, \mathcal{V}_1^n(X)\rangle}\right) = \int_0^1 P_{\rho_s}\left(\theta_1\sigma_{1,s}^2, \theta_2\sigma_{2,s}^2, \theta_3\sigma_{1,s}\sigma_{2,s}\right) ds.$$

But the study of the steepness is more difficult.

Let us consider the case where diffusion and correlation coefficients are constant, the rate function being easier to read. Before that let us introduce the function P_c^* which is the Legendre transformation of P_c given in (2.2), for all $x = (x_1, x_2, x_3)$

$$P_c^*(x) := \begin{cases} \log\left(\dfrac{\sqrt{1-c^2}}{\sqrt{x_1x_2 - x_3^2}}\right) - 1 + \dfrac{x_1 + x_2 - 2cx_3}{2(1-c^2)} \\ \qquad\qquad\qquad if \qquad x_1 > 0,\ x_2 > 0,\ x_1x_2 > x_3^2 \\ +\infty, \quad otherwise. \end{cases} \tag{2.7}$$

Corollary 2.4. *We assume that for $\ell = 1, 2$ σ_ℓ and ρ are constants. Under the condition* **(B)***, we obtain that $\mathcal{V}_1^n(X)$ satisfies the LDP on $\mathbb{R}^3$ with speed n and with good rate function $I_{ldp}^{\mathcal{V}}$ given by*

$$I_{ldp}^{\mathcal{V}}(x_1, x_2, x_3) = P_\rho^*\left(\frac{x_1}{\sigma_1^2}, \frac{x_2}{\sigma_2^2}, \frac{x_3}{\sigma_1\sigma_2}\right), \tag{2.8}$$

where P_c^* is given in (2.7).

Remark 2.5. *In the case σ_ℓ is constant, a similar result for the large deviations is obtained by Mancini* [19] *in the jump case for $\left(\mathcal{Q}_{\ell,1}^n\right)_{n\geq 1}$*

Now, we shall extend Theorem 2.3 to the process-level large deviations, i.e. for trajectories $(\mathcal{V}_t^n(X), t \in [0,1])$ which is interesting from the viewpoint of non-parametric statistics.

Let $\mathcal{BV}([0,1], \mathbb{R}^3)$ (shorted in $\mathcal{BV}$) be the space of functions of bounded variation on $[0,1]$. We identify $\mathcal{BV}$ with $\mathcal{M}_3([0,1])$, the set of vector measures with value in $\mathbb{R}^3$. This is done in the usual manner: to $f \in \mathcal{BV}$, there corresponds μ^f by $\mu^f([0,t]) = f(t)$. Up to this identification, $\mathcal{C}_3([0,1])$ the set of $\mathbb{R}^3$-valued continuous bounded functions on $[0,1]$, is the topology dual of $\mathcal{BV}$. We endow $\mathcal{BV}$ with the weak-* convergence topology $\sigma(\mathcal{BV}, \mathcal{C}_3([0,1]))$ and with the associated Borel-σ-field $\mathcal{B}_\omega$. Let $f \in \mathcal{BV}$ and μ^f the associated measure in $\mathcal{M}_3([0,1])$. Consider the Lebesgue decomposition of μ^f, $\mu^f = \mu_a^f + \mu_s^f$ where μ_a^f denotes the absolutely continuous part of μ^f with respect to dx and μ_s^f its singular part. We denote by $f_a(t) = \mu_a^f([0,t])$ and by $f_s(t) = \mu_s^f([0,t])$.

Theorem 2.5. *Under the conditions* **(LDP)** *and* **(B)***, the sequence $\mathcal{V}_\cdot^n(X)$ satisfies the LDP on $\mathcal{BV}$ with speed n and rate function J_{ldp} given for any $f = (f_1, f_2, f_3) \in \mathcal{BV}$ by*

$$\begin{aligned} J_{ldp}(f) = &\int_0^1 P_{\rho_t}^*\left(\frac{f'_{1,a}(t)}{\sigma_{1,t}^2}, \frac{f'_{2,a}(t)}{\sigma_{2,t}^2}, \frac{f'_{3,a}(t)}{\sigma_{1,t}\sigma_{2,t}}\right) \\ &+ \int_0^1 \frac{\sigma_{2,t}^2 f'_{1,s}(t) + \sigma_{1,t}^2 f'_{2,s}(t) - 2\rho_t\sigma_{1,t}\sigma_{2,t}f'_{3,s}(t)}{2\sigma_{1,t}^2\sigma_{2,t}^2(1-\rho_t^2)}\mathbf{1}_{[t: f'_{1,s}>0, f'_{2,s}>0, (f'_{3,s})^2 < f'_{1,s}f'_{2,s}]} d\theta(t), \end{aligned} \tag{2.9}$$

where P_c^* is given in (2.7) and θ is any real-valued nonnegative measure with respect to which μ_s^f is absolutely continuous and $f'_s = d\mu_s^f/d\theta = (f'_{1,s}, f'_{2,s}, f'_{3,s})$.

3. Proofs

For the convenience of the reader, we recall the following lemma which is the key of the proofs.

Lemma 3.1. *(Approximation Lemma) Theorem 4.2.13 in* [10]

Let $(Y^n, X^n, n \in \mathbb{N})$ be a family of random varibales valued in a Polish space S with metric $d(\cdot,\cdot)$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume

- *$\mathbb{P}(Y^n \in \cdot)$ satisfies the large deviation principle with speed ϵ_n ($\epsilon_n \to \infty$) and the good rate function I.*
- *for every $\delta > 0$*

$$\limsup_{n\to\infty} \frac{1}{\epsilon_n} \log \mathbb{P}(d(Y^n, X^n) > \delta) = -\infty.$$

Then $\mathbb{P}(X^n \in \cdot)$ satisfies the large deviation principle with speed ϵ_n and the good rate function I.

Before starting the proof, we need to introduce some technical tools. In the case without jumps, we introduce the following diffusion for $\ell = 1, 2$

$$D_{\ell,t} = \int_0^t \sigma_{\ell,s} dW_{\ell,s},$$

where $W_{\ell,s}$ and $\sigma_{\ell,s}$ are defined as before. We introduce the correspondent estimator

$$V_t^n = (Q_{1,t}^n, Q_{2,t}^n, C_t^n)$$

where for $\ell = 1, 2$

$$Q_{\ell,t}^n = \sum_{k=1}^{[nt]} (\Delta_k^n D_\ell)^2 \quad \text{and} \quad C_t^n = \sum_{k=1}^{[nt]} \Delta_k^n D_1 \Delta_k^n D_2.$$

We recall the following results from Djellout et al. [11]

Proposition 3.2. *Under the conditions* **(B)** *and* **(MDP)**,

(1) the sequence

$$\frac{\sqrt{n}}{v_n} (V_1^n - [\mathcal{V}]_1)$$

satisfies the LDP on $\mathbb{R}^3$ with speed v_n^2 and with rate function given by (2.1).

(2) the sequence

$$\frac{\sqrt{n}}{v_n} (V_\cdot^n - [\mathcal{V}]_\cdot)$$

satisfies the LDP on $\mathcal{H}$ with speed v_n^2 and with rate function given by (2.2).

Proposition 3.3. *Under the conditions* **(B)** *and* **(LDP)**,

(1) the sequence V_1^n satisfies the LDP on $\mathbb{R}^3$ with speed n and with good rate function given in (2.6).

(2) the sequence $\mathcal{V}_\cdot^n$ satisfies the LDP on $\mathcal{BV}$ with speed n and rate function J_{ldp} given by (2.9).

3.1. **Proof of Theorem 2.1.**

We will do the proof in two steps.

Part 1 We start with the case $b_\ell = 0$. In this case, $\mathcal{V}_t^n(X) = \mathcal{V}_t^n(X^0)$ with $X^0_{\ell,t} = X_{\ell,t} - \int_0^t b_\ell(s,\omega)ds$ and

$$\mathcal{Q}^n_{\ell,1}(X^0) = \sum_{k=1}^n (\Delta_k^n X_\ell^0)^2 \mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n})\}},\ \ell = 1,2$$

and

$$\mathcal{C}^n_1(X^0) = \sum_{k=1}^n \Delta_k^n X_1^0 \Delta_k^n X_2^0 \mathbf{1}_{\{\max_{\ell=1}^2 (\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n})\}}.$$

We will prove that

$$\frac{\sqrt{n}}{v_n}\left(\mathcal{V}_1^n(X^0) - V_1^n\right) \quad \xrightarrow[v_n^2]{\text{superexp}} 0.$$

For that, we will prove that for $\ell = 1,2$

$$\frac{\sqrt{n}}{v_n}\left(\mathcal{Q}_{\ell,1}^n(X^0) - Q_{\ell,1}^n\right) \quad \xrightarrow[v_n^2]{\text{superexp}} 0, \tag{3.1}$$

and

$$\frac{\sqrt{n}}{v_n}\left(\mathcal{C}_1^n(X^0) - C_1^n\right) \quad \xrightarrow[v_n^2]{\text{superexp}} 0. \tag{3.2}$$

We start by the proof of (3.1). Since the processes X_ℓ^0 and D_ℓ have independent increment, by Chebyshev inequality we obtain for all $\theta > 0$

$$\mathbb{P}\left(\frac{\sqrt{n}}{v_n}\left(\mathcal{Q}_{\ell,1}^n(X^0) - Q_{\ell,1}^n\right) > \delta\right) \le e^{-\theta\delta v_n^2}\prod_{k=1}^n \mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2 \mathbf{1}_{\{(\Delta_k^n X_\ell)^2 \le r(\frac{1}{n})\}} - (\Delta_k^n D_\ell)^2\right]}\right).$$

We have to control each term appearing in the product

$$\mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2 \mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n})\}} - (\Delta_k^n D_\ell)^2\right]}\right) \le \Re_1(k,n) + \Re_2(k,n), \tag{3.3}$$

where

$$\Re_1(k,n) := \mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2 - (\Delta_k^n D_\ell)^2\right]}\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n})\}}\right)$$

and

$$\Re_2(k,n) := \mathbb{P}\left((\Delta_k^n X_\ell^0)^2 > r(\frac{1}{n})\right).$$

For the first term, we write

$$\begin{aligned}\Re_1(k,n) &= \mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2 - (\Delta_k^n D_\ell)^2\right]}\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n})\}}|\Delta_k^n N_\ell = 0\right)\mathbb{P}(\Delta_k^n N_\ell = 0)\\ &\quad + \mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2 - (\Delta_k^n D_\ell)^2\right]}\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n}), \Delta_k^n N_\ell \neq 0\}}\right).\end{aligned} \tag{3.4}$$

Since N_ℓ is independent of W_ℓ, we obtain that

$$\begin{aligned}\Re_1(k,n) &\le \mathbb{P}\left((\Delta_k^n D_\ell)^2 \le r(\frac{1}{n})\right)e^{-\lambda_\ell/n} + e^{\sqrt{n}v_n\theta r(\frac{1}{n})}(1 - e^{-\lambda_\ell/n})\\ &\le 1 + e^{\sqrt{n}v_n\theta r(\frac{1}{n})}(1 - e^{-\lambda_\ell/n}).\end{aligned} \tag{3.5}$$

Now we have to control $\Re_2(k,n)$, by the same argument as before we have

$$\begin{aligned}\Re_2(k,n) &= \mathbb{P}\left((\Delta_k^n X_\ell^0)^2 > r(\frac{1}{n})|\Delta_k^n N_\ell = 0\right)\mathbb{P}(\Delta_k^n N_\ell = 0)\\ &\quad +\mathbb{P}\left(\Delta_k^n X_\ell^0)^2 > r(\frac{1}{n}), \Delta_k^n N_\ell \neq 0\right)\\ &\leq \mathbb{P}\left((\Delta_k^n D_\ell)^2 > r(\frac{1}{n})\right)e^{-\lambda_\ell/n} + (1-e^{-\lambda_\ell/n}).\end{aligned}$$

From exponential inequality for martingales, it follows that for $\ell = 1,2$,

$$\mathbb{P}\left((\Delta_k^n D_\ell)^2 > r\left(\frac{1}{n}\right)\right) \leq \exp\left(-\frac{r(\frac{1}{n})}{2\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2 ds}\right), \tag{3.6}$$

which implies that

$$\Re_2(k,n) \leq \exp\left(-\frac{r(\frac{1}{n})}{2\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2 ds}\right) + (1-e^{-\lambda_\ell/n}). \tag{3.7}$$

From (3.3), (3.5) and (3.7), we obtain that

$$\begin{aligned}\mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2\leq r(\frac{1}{n})\}}-(\Delta_k^n D_\ell)^2\right]}\right) &\leq 1+(1+e^{\sqrt{n}v_n\theta r(\frac{1}{n})})(1-e^{-\lambda_\ell/n})\\ &\quad + \exp\left(-\frac{r(\frac{1}{n})}{2\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2 ds}\right).\end{aligned}$$

Using the hypotheses **(MDP)**, we have

$$\limsup_{n\to\infty}\frac{n}{v_n^2}\max_{k=1}^{n}\log\mathbb{E}\left(e^{\theta\sqrt{n}v_n\left[(\Delta_k^n X_\ell^0)^2\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2\leq r(\frac{1}{n})\}}-(\Delta_k^n D_\ell)^2\right]}\right) = 0. \tag{3.8}$$

So

$$\limsup_{n\to\infty}\frac{1}{v_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{v_n}\left(\mathcal{Q}_{\ell,1}^n(X^0)-Q_{\ell,1}^n\right)>\delta\right)\leq -\lambda\delta.$$

Letting λ goes to infinity, we obtain that the right hand of the last inequality goes to $-\infty$. Proceeding in the same way for $-(\mathcal{Q}_{\ell,t}^n(X^0)-Q_{\ell,t}^n)$ we obtain (3.1).

Now we have to prove (3.2). For that we have the following decompostion

$$\mathcal{C}_1^n(X^0)-C_1^n = \frac{1}{2}\left[\overline{\mathcal{Q}}_{3,1}^n(X^0)-Q_{3,1}^n\right]-\frac{1}{2}\left[\sum_{\ell=1}^{2}\overline{\mathcal{Q}}_{\ell,1}^n(X^0)-Q_{\ell,1}^n\right], \tag{3.9}$$

where

$$Q_{3,1}^n = \sum_{k=1}^n(\Delta_k^n D_1+\Delta_k^n D_2)^2,$$

and for $\ell = 1,2$

$$\overline{\mathcal{Q}}_{\ell,t}^n(X^0) = \sum_{k=1}^n(\Delta_k^n X_\ell^0)^2\mathbf{1}_{\{\max_{\ell=1}^2(\Delta_k^n X_\ell^0)^2\leq r(\frac{1}{n})\}}$$

and

$$\overline{\mathcal{Q}}^n_{3,1}(X^0) = \sum_{k=1}^n (\Delta^n_k X^0_1 + \Delta^n_k X^0_2)^2 \mathbf{1}_{\{\max_{\ell=1}^2 (\Delta^n_k X^0_\ell)^2 \le r(\frac{1}{n})\}}.$$

Remark that $\overline{\mathcal{Q}}^n_{\ell,t}(X^0)$ is a slight modification of $\mathcal{Q}^n_{\ell,t}(X^0)$.

We know that $\Delta^n_k D_1 + \Delta^n_k D_2 \sim \mathcal{N}(0, \beta^2(k,n))$ with

$$\beta^2(k,n) = \int_{t_{k-1}}^{t_k} \sigma^2_{1,s} ds + \int_{t_{k-1}}^{t_k} \sigma^2_{2,s} ds + 2\int_{t_{k-1}}^{t_k} \sigma_{1,s}\sigma_{2,s}\rho_s ds.$$

For all $\delta > 0$, we have

$$\mathbb{P}\left(\frac{\sqrt{n}}{v_n}\left|\mathcal{C}^n_1(X^0) - C^n_1\right| > \delta\right) \le 3 \max_{\ell=1}^3 \mathbb{P}\left(\frac{\sqrt{n}}{v_n}\left|\overline{\mathcal{Q}}^n_{\ell,1}(X^0) - Q^n_{\ell,1}\right| > \frac{2\delta}{3}\right).$$

So we obtain (3.2).

<u>*Part 2*</u> We have to prove that

$$\frac{\sqrt{n}}{v_n}\left(\mathcal{V}^n_1(X) - \mathcal{V}^n_1(X^0)\right) \overset{\text{superexp}}{\underset{v_n^2}{\longrightarrow}} 0.$$

We have that

$$\left|\mathcal{Q}^n_{\ell,1}(X) - \mathcal{Q}^n_{\ell,1}(X^0)\right| \le \varepsilon(n)\mathcal{Q}^n_{\ell,1}(X^0) + \left(1 + \frac{1}{\varepsilon(n)}\right) Z^n_\ell \tag{3.10}$$

and

$$\left|\mathcal{C}^n_1(X) - \mathcal{C}^n_1(X^0)\right| \le \varepsilon(n) \max_{\ell=1}^2 \mathcal{Q}^n_{\ell,1}(X^0) + \left(1 + \frac{1}{\varepsilon(n)}\right) \max_{\ell=1}^2 Z^n_\ell, \tag{3.11}$$

where

$$Z^n_\ell = \sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} b_\ell(s,\omega) ds\right)^2.$$

By the condition **(B)**, we have that $\|Z^n_\ell\| \le \dfrac{1}{n}$. We choose $\varepsilon(n)$ such that

$$\frac{\sqrt{n}}{v_n}\varepsilon(n) \to 0, \quad v_n\sqrt{n}\varepsilon(n) \to \infty,$$

so by the MDP of $\mathcal{Q}^n_{\ell,1}(X^0)$, we obtain the result.

3.2. **Proof of Theorem 2.2.**

Since the sequence $\frac{\sqrt{n}}{v_n}(V^n_\cdot - [\mathcal{V}]_\cdot)$ satisfies the LDP on $\mathcal{H}$ with speed v_n^2 and rate function J_{mdp}, by Lemma 3.1, it is sufficient to show that:

$$\frac{\sqrt{n}}{v_n}\sup_{t\in[0,1]}\left\|\mathcal{V}^n_t(X^0) - V^n_t\right\| \overset{\text{superexp}}{\underset{v_n^2}{\longrightarrow}} 0. \tag{3.12}$$

Lemma 3.4. *Under the condition* **(MDP)***, we have*

$$\lim_{n\to\infty}\frac{\sqrt{n}}{v_n}\sup_{t\in[0,1]}\left\|\mathbb{E}\mathcal{V}^n_t(X^0) - [\mathcal{V}]_t\right\| = 0.$$

Proof We will prove that for $\ell = 1,2$

$$\lim_{n\to\infty} \frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left| \mathbb{E}\mathcal{Q}^n_{\ell,t}(X^0) - \int_0^t \sigma^2_{\ell,s} ds \right| = 0. \tag{3.13}$$

and

$$\lim_{n\to\infty} \frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left| \mathbb{E}\mathcal{C}^n_t(X^0) - \int_0^t \sigma_{1,s}\sigma_{1,s}\rho_s ds \right| = 0. \tag{3.14}$$

In fact, (3.13) can be done in the same way as in Jiang [16]. It remains to show (3.14). Using (3.9), we obtain that

$$\left| \mathbb{E}\mathcal{C}^n_t(X^0) - \int_0^t \sigma_{1,s}\sigma_{1,s}\rho_s ds \right| \le \frac{1}{2} \left| \mathbb{E}\overline{\mathcal{Q}}^n_{3,t}(X^0) - \beta_t \right| + \max_{\ell=1}^{2} \left| \mathbb{E}\overline{\mathcal{Q}}^n_{\ell,t}(X^0) - \int_0^t \sigma^2_{\ell,s} ds \right|,$$

where $\beta_t = \int_0^t \sigma^2_{1,s} ds + \int_0^t \sigma^2_{2,s} ds + 2\int_0^t \sigma_{1,s}\sigma_{2,s}\rho_s ds$. So the proof of (3.14) is a consequence of (3.13) and the fact that

$$\lim_{n\to\infty} \frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left| \mathbb{E}\overline{\mathcal{Q}}^n_{3,t}(X^0) - \beta_t \right| = 0,$$

which is an adaptation of the proof in Jiang [16].

Proof of Theorem 2.2

For (3.12), we will prove that for $\ell = 1,2$

$$\frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left\| \mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t} \right\| \quad \xrightarrow[v_n^2]{\text{superexp}} 0 \quad \text{and} \quad \frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left\| \mathcal{C}^n_t(X^0) - C^n_t \right\| \quad \xrightarrow[v_n^2]{\text{superexp}} 0.$$

From Lemma 3.4, it follows that as $n \to \infty$

$$\frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left(\mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t}) \vee \mathbb{E}(\mathcal{C}^n_t(X^0) - C^n_t) \right) \to 0. \tag{3.15}$$

Then, we only need to prove that

$$\frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left\| \mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t} - \mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t}) \right\| \quad \xrightarrow[v_n^2]{\text{superexp}} 0 \tag{3.16}$$

and

$$\frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left\| \mathcal{C}^n_t(X^0) - C^n_t - \mathbb{E}(\mathcal{C}^n_t(X^0) - C^n_t) \right\| \quad \xrightarrow[v_n^2]{\text{superexp}} 0. \tag{3.17}$$

We start by the proof of (3.16). Remark that $\left(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t} - \mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t})\right)$ is a $\mathcal{F}_{[nt]/n}$-martingale. Then

$$\exp\left(\lambda\left(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t} - \mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t})\right)\right)$$

is a submartigale. By the maximal inequality, we have for any $\eta, \lambda > 0$

$$\begin{aligned} &\mathbb{P}\left(\frac{\sqrt{n}}{v_n} \sup_{t\in[0,1]} \left(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t} - \mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0) - Q^n_{\ell,t})\right) > \eta \right) \\ &\le e^{-\lambda v_n^2 \eta} \mathbb{E}\exp\left(\lambda\sqrt{n}v_n\left(\mathcal{Q}^n_{\ell,1}(X^0) - Q^n_{\ell,1} - \mathbb{E}(\mathcal{Q}^n_{\ell,1}(X^0) - Q^n_{\ell,1})\right)\right) \end{aligned}$$

and

$$\mathbb{P}\left(\frac{\sqrt{n}}{v_n}\inf_{t\in[0,1]}\left(\mathcal{Q}^n_{\ell,t}(X^0)-Q^n_{\ell,t}-\mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0)-Q^n_{\ell,t})\right)<-\eta\right)$$
$$\leq e^{-\lambda v_n^2\eta}\mathbb{E}\exp\left(-\lambda\sqrt{n}v_n\left(\mathcal{Q}^n_{\ell,1}(X^0)-Q^n_{\ell,1}-\mathbb{E}(\mathcal{Q}^n_{\ell,1}(X^0)-Q^n_{\ell,1})\right)\right).$$

Together with (3.8) and (3.15), we have

$$\limsup_{n\to\infty}\frac{1}{v_n^2}\log\mathbb{P}\left(\frac{\sqrt{n}}{v_n}\sup_{t\in[0,1]}\left|\mathcal{Q}^n_{\ell,t}(X^0)-Q^n_{\ell,t}-\mathbb{E}(\mathcal{Q}^n_{\ell,t}(X^0)-Q^n_{\ell,t})\right|>\eta\right)\leq-\lambda\eta.$$

(3.16) can be obtained by letting λ goes to infinity.

Similarly, we can have (3.17) by (3.8), (3.9) and (3.15).

3.3. **Proof of Theorem 2.3.**

We will do the proof in two steps.

<u>*Step 1*</u> We will prove that

$$\mathcal{V}^n_1(X^0)-V^n_1 \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0.$$

For that, we will prove that for $\ell=1,2$

$$\mathcal{Q}^n_{\ell,1}(X^0)-Q^n_{\ell,1} \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0, \tag{3.18}$$

and

$$\mathcal{C}^n_1(X^0)-C^n_1 \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0. \tag{3.19}$$

We start by the proof of (3.18). Since the processes X_ℓ and D_ℓ have independent increment, by Chebyshev inequality we obtain for all $\theta>0$

$$\mathbb{P}\left(\mathcal{Q}^n_{\ell,1}(X^0)-Q^n_{\ell,1}>\delta\right)\leq e^{-\theta n\delta}\prod_{k=1}^n\mathbb{E}\left(e^{\theta n\left[(\Delta^n_kX^0_\ell)^2\mathbf{1}_{\{(\Delta^n_kX^0_\ell)^2\leq r(\frac{1}{n})\}}-(\Delta^n_kD_\ell)^2\right]}\right).$$

Similar to (3.3),

$$\mathbb{E}\left(e^{\theta n\left[(\Delta^n_kX^0_\ell)^2\mathbf{1}_{\{(\Delta^n_kX^0_\ell)^2\leq r(\frac{1}{n})\}}-(\Delta^n_kD_\ell)^2\right]}\right)\leq I_1(k,n)+I_2(k,n),$$

where

$$I_1(k,n):=\mathbb{E}\left(e^{\theta n\left[(\Delta^n_kX^0_\ell)^2-(\Delta^n_kD_\ell)^2\right]}\mathbf{1}_{\{(\Delta^n_kX^0_\ell)^2\leq r(\frac{1}{n})\}}\right)$$

and

$$I_2(k,n):=\mathbb{P}\left((\Delta^n_kX^0_\ell)^2>r(\frac{1}{n})\right)$$

From (3.4), (3.5) and (3.7), it follows that

$$I_2(k,n)\leq\exp\left(-\frac{r(\frac{1}{n})}{2\int_{t_{k-1}}^{t_k}\sigma^2_{\ell,s}ds}\right)+(1-e^{-\lambda_\ell/n}).$$

and

$$I_1(k,n)\leq 1+\mathbb{E}\left(e^{\theta n\left[(\Delta^n_kX^0_\ell)^2-(\Delta^n_kD_\ell)^2\right]}\mathbf{1}_{\{(\Delta^n_kX^0_\ell)^2\leq r(\frac{1}{n}),\Delta^n_kN_\ell\neq0\}}\right).$$

Let (α_n) be a sequence of real numbers such that $\alpha_n \to 0$, which will be chosen latter. We have

$$\mathbb{E}\left(e^{\theta n(\Delta_k^n X_\ell^0)^2}\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2\le r(\frac{1}{n}),\Delta_k^n N_\ell\neq 0\}}\right)=F_1(k,n)+F_2(k,n),$$

where

$$F_1(k,n):=\mathbb{E}\left(e^{\theta n(\Delta_k^n X_\ell^0)^2}\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2\le r(\frac{1}{n}),\Delta_k^n N_\ell\neq 0,|\Delta_k^n J_\ell|\le\alpha_n\}}\right)$$

and

$$F_2(k,n):=\mathbb{E}\left(e^{\theta n(\Delta_k^n X_\ell^0)^2}\mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2\le r(\frac{1}{n}),\Delta_k^n N_\ell\neq 0,|\Delta_k^n J_\ell|>\alpha_n\}}\right).$$

We have to prove that for $\ell=1,2$ $\lim_{n\to\infty}\max_{k=1}^n F_\ell(k,n)\to 0$. We start with $F_2(k,n)$.
From condition (**LDP**), it follows that $n\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds<+\infty$.
So for all $\theta>0$, we choose

$$\alpha_n=\left(2\sqrt{\theta n\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds}+1\right)\sqrt{r(1/n)}.$$

Then it is easy to see that

$$F_2(k,n)\le e^{\theta nr(\frac{1}{n})}\mathbb{P}\left(|Z|\ge\frac{2\sqrt{\theta n\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds}\sqrt{r(\frac{1}{n})}}{\sqrt{\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds}}\right),$$

where Z is a standard Gaussian random variable. As a consequence of the well-known inequality $\int_y^{+\infty}e^{-\frac{z^2}{2}}dz\le(1/y)e^{-\frac{y^2}{2}}$, for all $y>0$, we obtain

$$F_2(k,n)\le e^{\theta nr(\frac{1}{n})}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\theta nr(1/n)}}e^{-2\theta nr(\frac{1}{n})}.$$

So for n large enough and $\theta>1$, we have

$$\max_{k=1}^n F_2(k,n)\le e^{-\theta nr(\frac{1}{n})}\longrightarrow 0\quad\text{as}\quad n\to\infty.$$

Now we will control $F_1(k,n)$. Using the fact that

$$\theta n(\Delta_k^n X_\ell^0)^2\le\theta n\left[\frac{1}{4\theta n\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds}(\Delta_k^n D_\ell)^2+4\theta n\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds(\Delta_k^n J_\ell)^2\right],$$

we have with the same choose of the sequence α_n, by independence of $\Delta_k^n D_\ell$ and $\Delta_k^n J_\ell$ and Cauchy-Schwarz inequality that

$$\begin{aligned}F_1(k,n)&\le\mathbb{E}\left(e^{\frac{(\Delta_k^n D_\ell)^2}{4\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds}}\right)\mathbb{E}\left(e^{4\theta^2\left(n\max_{k=1}^n\int_{t_{k-1}}^{t_k}\sigma_{\ell,s}^2ds\right)n(\Delta_k^n J_\ell)^2}\mathbf{1}_{\{|\Delta_k^n J_\ell|\le\alpha_n\}}\mathbf{1}_{\{\Delta_k^n N_\ell\neq 0\}}\right)\\&\le\mathbb{E}\left(e^{\frac{Z^2}{4}}\right)\mathbb{E}^{\frac{1}{2}}\left(e^{8\theta^2 n(\Delta_k^n J_\ell)^2}\mathbf{1}_{\{|\Delta_k^n J_\ell|\le\alpha_n\}}\right)\mathbb{P}^{\frac{1}{2}}\left(\Delta_k^n N_\ell\neq 0\right).\end{aligned}$$

From Mancini [19] page 877, we conclude that

$$\lim_{n\to\infty}\max_{k=1}^n\mathbb{E}\left(e^{8\theta^2 n(\Delta_k^n J_\ell)^2}\mathbf{1}_{\{|\Delta_k^n J_\ell|\le\alpha_n\}}\right)<\infty.$$

Since Z is a standard Gaussian random variable, we conclude that

$$E\left(e^{\frac{Z^2}{4}}\right) < \infty.$$

So that $\max_{k=1}^n F_1(k,n) \le C(1 - e^{-\lambda_\ell/n}) \longrightarrow 0$ as $n \to \infty$.
Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \log \prod_{k=1}^n \mathbb{E}\left(e^{\theta n\left[(\Delta_k^n X_\ell^0)^2 \mathbf{1}_{\{(\Delta_k^n X_\ell^0)^2 \le r(\frac{1}{n})\}} - (\Delta_k^n D_\ell)^2\right]}\right) = 0,$$

which implies that for any $\theta > 1$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\mathcal{Q}_{\ell,1}^n(X^0) - Q_{\ell,1}^n > \delta\right) \le -\theta\delta.$$

Letting θ goes to infinity, we obtain that the left term in the last inequality goes to $-\infty$. And similarly, by doing the same calculation with

$$\mathbb{P}\left(\mathcal{Q}_{\ell,1}^n(X^0) - Q_{\ell,1}^n < -\delta\right),$$

we can get (3.18).

To prove (3.19), we use the decomposition (3.9) and an adaptation of the proof of (3.18).

Step 2 We will prove that

$$\mathcal{V}_1^n(X) - \mathcal{V}_1^n(X^0) \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0.$$

For that we use (3.10) and (3.11) and we choose $\varepsilon(n)$ such that $n\varepsilon(n) \to 0$ to obtain the result.

3.4. **Proof of Theorem 2.5.**

We will prove that for $\ell = 1, 2$

$$\sup_{t\in[0,1]} \left\|\mathcal{Q}_{\ell,t}^n(X^0) - Q_{\ell,t}^n\right\| \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0 \quad \text{and} \quad \sup_{t\in[0,1]} \left\|\mathcal{C}_t^n(X^0) - C_t^n\right\| \quad \overset{\text{superexp}}{\underset{n}{\longrightarrow}} 0.$$

To do that we use the same argument as in the proof of Theorem 2.2 and the fact that

$$\sup_{t\in[0,1]} \left|\mathbb{E}(\mathcal{Q}_{\ell,t}^n(X^0) - Q_{\ell,t}^n)\right| \longrightarrow 0.$$

References

[1] Aït-Sahalia, Y. Disentangling diffusion from jumps. *Journal of Financial Economics 74* (2004), 487–528.

[2] Aït-Sahalia, Y., and Jacod, J. Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics 37*, 5A (2009), 2202–2244.

[3] Aït-Sahalia, Y., and Jacod, J. Testing for jumps in a discretely observed process. *The Annals of Statistics 37*, 1 (2009), 184–222.

[4] Barndorff-Nielsen, O. E., Graversen, S., Jacod, J., Podolskij, M., and Shephard, N. A central limit theorem for realised power and bipower variations of continuous semimartingales. *In: Y. Kabanov, R. Lispter (Eds.), Stochastic Analysis to Mathematical Finance, Frestchrift for Albert Shiryaev, Springer.* (2006), 33–68.

[5] Barndorff-Nielsen, O. E., Graversen, S., Jacod, J., and Shephard, N. Limit theorems for bipower variation in financial econometrics. *Econometric Theory, 22* (2006), 677–719.

[6] Barndorff-Nielsen, O. E., and Shephard, N. Econometric analysis of realized volatility and its use in estimating volatility models. *Journal of the Royal Statistical Society, Series B 64* (2002), 253–280.
[7] Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and their Applications, 116* (2006), 796–806.
[8] Bryc, W., and Dembo, A. Large deviations for quadratic functionals of gaussian processes. *Journal of Theoretical Probability, 10* (1997), 307–322.
[9] Corsi, F., Pirino, D., and Renò, R. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics 159*, 2 (2010), 276–288.
[10] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications. Second edition.* Springer, 1998.
[11] Djellout, H., Guillin, A., and Samoura, Y. Large deviations of the realized (co-)volatility vector. *Submited* (2014), 1–36.
[12] Djellout, H., Guillin, A., and Wu, L. Large and moderate deviations for estimators of quadratic variational processes of diffusions. *Statistical Inference for Stochastic Processes*, 2 (1999), 195–225.
[13] Djellout, H., and Samoura, Y. Large and moderate deviations of realized covolatility. *Statistics & Probability Letters, 86* (2014), 30–37.
[14] Figueroa-López, J. E., and Nisen, J. Optimally thresholded realized power variations for Lévy jump diffusion models. *Stochastic Processes and their Applications 123*, 7 (2013), 2648–2677.
[15] Gobbi, F., and Mancini, C. Estimating the diffusion part of the covariation between two volatility models with jumps of Lévy type. 399–409.
[16] Jiang, H. Moderate deviations for estimators of quadratic variational process of diffusion with compound poisson. *Statistics & Probability Letters, 80*, 17-18 (2010), 1297–1305.
[17] Kanaya, S., and Otsu, T. Large deviations of realized volatility. *Stochastic Processes and their Applications, 122*, 2 (2012), 546–581.
[18] Mancini, C. Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scandinavian Actuarial Journal*, 1 (2004), 42–52.
[19] Mancini, C. Large deviation principle for an estimator of the diffusion coefficient in a jump-diffusion process. *Statistics & Probability Letters, 78*, 7 (2008), 869–879.
[20] Mancini, C. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics. Theory and Applications 36*, 2 (2009), 270–296.
[21] Mancini, C., and Gobbi, F. Identifying the brownian covariation from the co-jumps given discrete observations. *Econometric Theory, 28*, 2 (2012), 249–273.
[22] Ogihara, T., and Yoshida, N. Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes. 14*, 3 (2011), 189–229.
[23] Podolskij, M., and Vetter, M. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli 15*, 3 (2009), 634–658.
[24] Shimizu, Y. Functional estimation for Lévy measures of semimartingales with Poissonian jumps. *Journal of Multivariate Analysis 100*, 6 (2009), 1073–1092.
[25] Shimizu, Y. Threshold estimation for jump-type stochastic processes from discrete observations. *Proceedings of the Institute of Statistical Mathematics 57*, 1 (2009), 97–118.
[26] Shimizu, Y., and Yoshida, N. Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes 9*, 3 (2006), 227–277.

E-mail address: Hacene.Djellout@math.univ-bpclermont.fr

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais,BP80026, 63171 Aubière Cedex, France.

E-mail address: huijiang@nuaa.edu.cn

Department of Mathematics, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing 210016, China.

Deuxième Partie

Chaînes de Markov bifurcantes, théorèmes limites

The Annals of Applied Probability
2014, Vol. 24, No. 1, 235–291
DOI: 10.1214/13-AAP921

DEVIATION INEQUALITIES, MODERATE DEVIATIONS AND SOME LIMIT THEOREMS FOR BIFURCATING MARKOV CHAINS WITH APPLICATION

BY S. VALÈRE BITSEKI PENDA, HACÈNE DJELLOUT AND ARNAUD GUILLIN

Université Blaise Pascal

First, under a geometric ergodicity assumption, we provide some limit theorems and some probability inequalities for the bifurcating Markov chains (BMC). The BMC model was introduced by Guyon to detect cellular aging from cell lineage, and our aim is thus to complete his asymptotic results. The deviation inequalities are then applied to derive first result on the moderate deviation principle (MDP) for a functional of the BMC with a restricted range of speed, but with a function which can be unbounded. Next, under a uniform geometric ergodicity assumption, we provide deviation inequalities for the BMC and apply them to derive a second result on the MDP for a bounded functional of the BMC with a larger range of speed. As statistical applications, we provide superexponential convergence in probability and deviation inequalities (for either the Gaussian setting or the bounded setting), and the MDP for least square estimators of the parameters of a first-order bifurcating autoregressive process.

1. Introduction. Bifurcating Markov chains (BMC) are an adaptation of (usual) Markov chains to the data of a regular binary tree; see below for a more precise definition. In other terms, it is a Markov chain for which the index set is a regular binary tree. They are appropriate, for example, in the modeling of cell lineage data when each cell in one generation gives birth to two offspring in the next. Recently, they have received a great deal of attention because of the experiments of biologists on aging of Escherichia Coli; see [15, 20]. E. Coli is a rod-shaped bacterium which reproduces by dividing in the middle, thus producing two cells, one which already existed, that we call old pole progeny, and the other which is new, that we call new pole progeny. The aim of their experiments was to look for evidence of aging in E. Coli. In this section, we will introduce the model that allowed the authors of [15] to study the aging of E. Coli and we refer to their works for further motivations and insights on the data leading to the model studied here. This model is a typical example of bifurcating Markovian dynamics, and it has been the motivation for the rigorous mathematical study of BMC in [14]. This also motivates Sections 2 and 3 in the sequel, where we give a rigorous asymptotic (and

Received November 2011; revised January 2013.
MSC2010 subject classifications. Primary 60F05, 60F10, 60F15, 60E15; secondary 60G42, 60J05, 62M02, 62M05, 62P10.
Key words and phrases. Bifurcating Markov chains, limit theorems, ergodicity, deviation inequalities, moderate deviation, martingale, first-order bifurcating autoregressive process, cellular aging.

nonasymptotic) study of BMC under geometric ergodicity and uniform geometric ergodicity assumptions.

1.1. *The model.* Let $\mathbb{T}$ be a binary regular tree in which each vertex is seen as a positive integer different from 0; see Figure 1. For $r \in \mathbb{N}$, let

$$\mathbb{G}_r = \{2^r, 2^r + 1, \dots, 2^{r+1} - 1\}, \qquad \mathbb{T}_r = \bigcup_{q=0}^{r} \mathbb{G}_q,$$

which denote, respectively, the rth column and the first $(r + 1)$ columns of the tree. Then, the cardinality $|\mathbb{G}_r|$ of $\mathbb{G}_r$ is 2^r and that of $\mathbb{T}_r$ is $|\mathbb{T}_r| = 2^{r+1} - 1$. A column of a given integer n is $\mathbb{G}_{r_n}$ with $r_n = \lfloor \log_2 n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of the real number x.

The genealogy of the cells is described by this tree. In the sequel we will thus see $\mathbb{T}$ as a given population. Then the vertex n, the column $\mathbb{G}_r$ and the first $(r + 1)$ columns $\mathbb{T}_r$ designate, respectively, individual n, the rth generation and the first $(r + 1)$ generations. The initial individual is denoted 1.

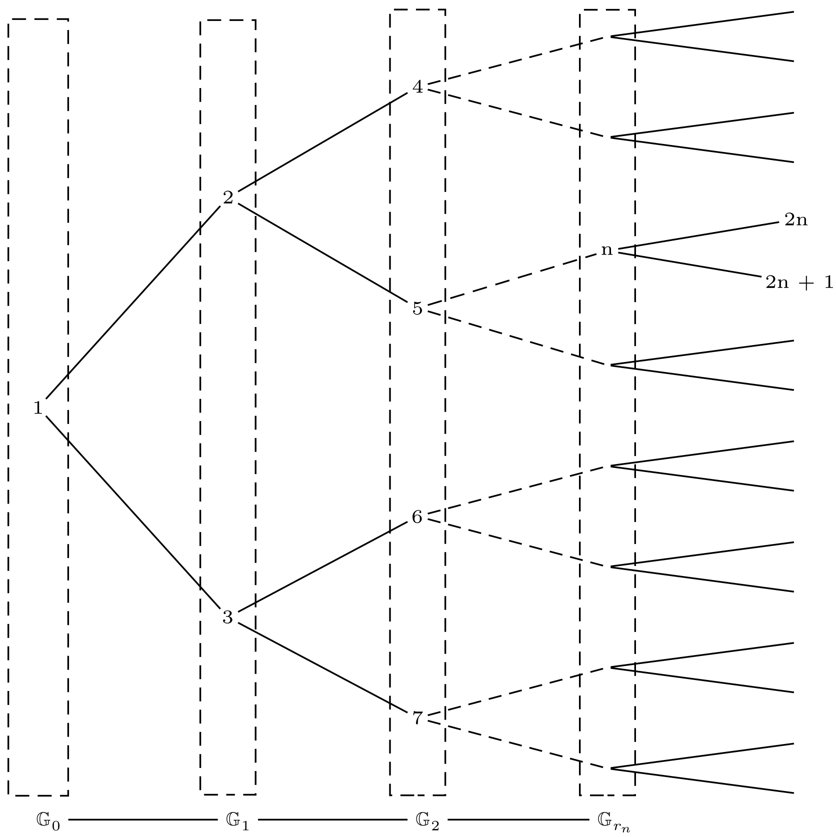

FIG. 1. *The binary tree* $\mathbb{T}$.

Guyon et al. [14, 15] proposed the following linear Gaussian model to describe the evolution of the growth rate of the population of cells derived from an initial individual:

$$\mathcal{L}(X_1) = \nu \quad \text{and} \quad \forall n \geq 1 \qquad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{1.1}$$

where X_n is the growth rate of individual n, n is the mother of $2n$ (the new pole progeny cell) and $2n+1$ (the old pole progeny cell), ν is a distribution probability on $\mathbb{R}$, $\alpha_0, \alpha_1 \in (-1, 1)$; $\beta_0, \beta_1 \in \mathbb{R}$ and $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of i.i.d. bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$, where

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad \sigma^2 > 0, \qquad \rho \in (-1, 1).$$

The processes (X_n) defined by (1.1) are typical examples of BMC which are called the first-order bifurcating autoregressive processes [BAR(1)]. The BAR(1) processes are an adaptation of autoregressive processes, when the data have a binary tree structure. They were first introduced by Cowan and Staudte [6] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation. We will not discuss here extensions to m-ary tree, which follow more or less from the same method, or Markov chains on Galton–Watson trees that are left for an other study.

In [14], Guyon, after establishing the first results on the theory of BMC, proves laws of large numbers and central limit theorem for the least-square estimators $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$ of the 4-dimensional parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$; see Section 4 for a more precise definition. He also gives some statistical tests which allow to check if the model is symmetric or not (roughly $\alpha_0 = \alpha_1$ or not), and if the new pole and the old pole populations are even distinct in mean, which allows him to conclude a statistical evidence in aging in E. Coli. Let us also mention [4], where Bercu et al., using the martingale approach, give asymptotic analysis of the least squares estimators of the unknown parameters of a general asymmetric pth-order BAR processes.

In this paper, we will give moderate deviation principle (MDP) for this estimator and the statistical tests done by Guyon. We will also give deviation inequalities for $\hat{\theta}^r - \theta$, which are important for a rigorous (nonasymptotic) statistical study. This will be done in two cases: the Gaussian case as described above and the case where the noise and the initial state X_1 are assumed to take values in a compact set. Note that the latter case implies that the BAR(1) process defined by (1.1) valued in compact set.

We are now going to give a rigorous definition of BMC. We refer to [14] for more detail.

1.2. *Definitions*. For an individual $n \in \mathbb{T}$, we are interested in the quantity X_n (it may be the weight, the growth rate, ...) with values in the metric space S endowed with its Borel σ-field $\mathcal{S}$.

DEFINITION 1.1 ($\mathbb{T}$-transition probability, see [14]). We call $\mathbb{T}$-transition probability any mapping $P : S \times \mathcal{S}^2 \to [0, 1]$ such that:

- $P(\cdot, A)$ is measurable for all $A \in \mathcal{S}^2$;
- $P(x, \cdot)$ is a probability measure on $(S^2, \mathcal{S}^2)$ for all $x \in S$.

For a $\mathbb{T}$-transition probability P on $S \times \mathcal{S}^2$, we denote by P_0, P_1 and Q, respectively, the first and the second marginal of P, and the mean of P_0 and P_1, that is, $P_0(x, B) = P(x, B \times S)$, $P_1(x, B) = P(x, S \times B)$ for all $x \in S$ and $B \in \mathcal{S}$ and $Q = \frac{P_0+P_1}{2}$.

For $p \geq 1$, we denote by $\mathcal{B}(S^p)$ [resp., $\mathcal{B}_b(S^p)$], the set of all $\mathcal{S}^p$-measurable (resp., $\mathcal{S}^p$-measurable and bounded) mappings $f : S^p \to \mathbb{R}$. For $f \in \mathcal{B}(S^3)$, we denote by $Pf \in \mathcal{B}(S)$ the function

$$x \mapsto Pf(x) = \int_{S^2} f(x, y, z) P(x, dy, dz) \qquad \text{when it is defined.}$$

DEFINITION 1.2 (Bifurcating Markov chains; see [14]). Let $(X_n, n \in \mathbb{T})$ be a family of S-valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$. Let ν be a probability on $(S, \mathcal{S})$ and P be a $\mathbb{T}$-transition probability. We say that $(X_n, n \in \mathbb{T})$ is a $(\mathcal{F}_r)$-bifurcating Markov chain with initial distribution ν and $\mathbb{T}$-transition probability P if:

- X_n is $\mathcal{F}_{r_n}$-measurable for all $n \in \mathbb{T}$;
- $\mathcal{L}(X_1) = \nu$;
- for all $r \in \mathbb{N}$ and for all family $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$

$$\mathbb{E}\Big[\prod_{n \in \mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1}) \Big/ \mathcal{F}_r\Big] = \prod_{n \in \mathbb{G}_r} Pf_n(X_n).$$

In the following, when unspecified, the filtration implicitly used will be $\mathcal{F}_r = \sigma(X_i, i \in \mathbb{T}_r)$. We denote by $(Y_r, r \in \mathbb{N})$ the Markov chain on S with $Y_0 = X_1$ and transition probability Q. The chain $(Y_r, r \in \mathbb{N})$ corresponds to a random lineage taken in the population.

We denote by $\mathfrak{G}$ the set of all permutations of $\mathbb{N}^*$ that leaves each $\mathbb{G}_r$ invariant. We draw a permutation Π uniformly on $\mathfrak{G}$, independently of $X = (X_n, n \in \mathbb{T})$. Drawing Π "uniformly" on $\mathfrak{G}$ means drawing the restriction of Π on $\mathbb{G}_r$ uniformly among the $(2^r)!$ permutations of $\mathbb{G}_r$. In particular, $(\Pi(2^r), \Pi(2^r + 1), \ldots, \Pi(2^{r+1} - 1))$ can be viewed as a random drawing of all the elements of $\mathbb{G}_r$ without replacement. Notice that Π allows one to define a random order on $\mathbb{T}$ which preserves the genealogical order. For example, $(\Pi(i), 1 \leq i \leq n)$ denotes the set of the "first" n individuals of $\mathbb{T}$. Π was introduced by Guyon in order to sample over the "first" n individuals. As mentioned in [14], this choice of Π allows one to preserve the same asymptotic behavior for the empirical means resulting from the sampling over (say) the rth generation, the first $(r + 1)$ generations or the "first"

n individuals. In general, the choice of another permutation does not preserve the asymptotic behavior of these empirical means. We refer to [14], Section 2.2, for more detail.

Throughout the paper, we will denote by:

- $f \otimes g$ the mapping $(x, y) \mapsto f(x)g(y)$.
- Q^p the pth iterated of Q recursively defined by the formulas $Q^0(x, \cdot) = \delta_x$ and $Q^{p+1}(x, B) = \int_S Q(s, dy)Q^p(y, B)$ for all $B \in \mathcal{S}$; Q^p is a transition probability in $(S, \mathcal{S})$.
- νQ the distribution on $(S, \mathcal{S})$ defined by $\nu Q(B) = \int_S \nu(dx)Q(x, B)$; νQ^p is the law of Y_p.
- $(Qf)(x) = \int_S f(y)Q(x, dy)$ when it is defined.
- (νf) or (ν, f) the integral $\int_S f\, d\nu$ when it is defined.

For all $i \in \mathbb{T}$, we set $\Delta_i = (X_i, X_{2i}, X_{2i+1})$. We introduce the following empirical quantities:

$$\begin{cases} \overline{M}_{\mathbb{G}_r}(f) = \dfrac{1}{|\mathbb{G}_r|} \displaystyle\sum_{i \in \mathbb{G}_r} f(\tilde{\Delta}_i), \\ \overline{M}_{\mathbb{T}_r}(f) = \dfrac{1}{|\mathbb{T}_r|} \displaystyle\sum_{i \in \mathbb{T}_r} f(\tilde{\Delta}_i), \\ \overline{M}_n^{\Pi}(f) = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} f(\tilde{\Delta}_{\Pi(i)}), \end{cases} \tag{1.2}$$

where $f(\tilde{\Delta}_i) = f(\Delta_i) = f(X_i, X_{2i}, X_{2i+1})$ if $f \in \mathcal{B}(S^3)$ and $f(\tilde{\Delta}_i) = f(X_i)$ if $f \in \mathcal{B}(S)$.

Guyon in [14] studied limit theorems of the empirical means (1.2), namely the law of large numbers (L^2 and almost sure versions) and the central limit theorems for (1.2) when $f \in \mathcal{B}(S^3)$, but centered by the conditional expectation rather than by the limit mean. An extension of the BMC has been proposed in [8], in which the authors studied a model of BMC with missing data. To take into account the possibility for a cell to die, the authors of [8] use Galton–Watson tree instead of a regular tree. And they give a weak law of large numbers, an invariance principle and the central limit result for the average over one generation or up to one generation. As previously mentioned, this setting will be considered in incoming works. One can also mention the work of De Saporta et al. [7] dealing with bifurcating autoregressive processes with missing data in the estimation procedure of the parameters of the asymmetric BAR process. They use a two type Galton–Watson process to model the genealogy and give convergence and asymptotic normality of their estimators. It is important to remark that the nonasymptotic study of deviation inequalities has not been considered at all in these works, despite their practical interest.

1.3. *Objectives*. Our objectives in this paper are:

- to give some limit theorems for BMC that complete those done in [14] (LLN, LIL, ...);
- to give probability inequalities and deviation inequalities for the empirical means (1.2), that is, for $f \in \mathcal{B}(S)$ and all $x > 0$
$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) - (\mu, f) \geq x\big) \leq e^{-C(x,r)},$$
where $C(x,r)$ will crucially depend on our set of assumptions on f and on the ergodic property of Q but valid for (nearly) all r;
- to study moderate deviation principle (MDP) for BMC, that is, for some range of speed $\sqrt{r} \ll b_r \ll r$ (depending on assumptions) and for $f \in \mathcal{B}_b(S^3)$ with $Pf = 0$
$$\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|} \log \mathbb{P}\bigg(\frac{1}{b_{|\mathbb{T}_r|}} M_{\mathbb{T}_r}(f) \geq x\bigg) \sim -\frac{x^2}{2\sigma^2};$$
- to obtain the MDP and deviation inequalities for the estimator of bifurcating autoregressive process, which are important for a rigorous statistical study.

All these results will be obtained under hypothesis of geometric ergodicity or uniform geometric ergodicity, meaning that Q^r converges (uniformly) exponentially fast to a limiting measure.

The limit theorems, proved in this paper, include strong law of large numbers for the empirical average $\overline{M}_n^{\Pi}(f)$ with $f \in \mathcal{B}(S)$ (this case is not studied in [14]), the law of the iterated logarithm and the almost sure functional central limit theorem. A strong law of large numbers will be obtained via control of 4th order moments. We thus generalize the computation of 2nd order moments made by Guyon in [14]. It will be noted that the technique we will use can be applied to compute the other higher-order moments, but at the price of huge and tedious computations.

Deviation inequalities will be obtained in the setting of unbounded functions, by using the classical Markov inequality and under geometric ergodicity assumption. The results are, however, at this point quite restrictive.

Exponential deviation inequalities will be shown for bounded functions and under a uniform geometric ergodicity assumption. Their proof intensively uses the Azuma–Bennett–Hoeffding inequality [1, 3, 16], which requires bounded random variables. Extension to unbounded functions and weaker ergodicity assumptions will be done in a further work, using transportation inequalities in the spirit of [12].

The MDP will be mainly deduced from these inequalities and general results on moderate deviations of martingales; see [11], recalled in the Appendix B. Their speed will depend on whether uniform geometric ergodicity or only geometric ergodicity is satisfied.

Before presenting the plan of our paper, let us recall the definition of a moderate deviation principle (MDP): let $(b_n)_{n\geq 0}$ be a positive sequence such that
$$\frac{b_n}{n} \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \frac{b_n^2}{n} \underset{n\to\infty}{\longrightarrow} \infty.$$

We say that a sequence of centered random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies a MDP with speed b_n^2/n and rate function $I : S \to \mathbb{R}_+^*$ if for each $A \in \mathcal{S}$,

$$-\inf_{x \in A^o} I(x) \leq \liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} M_n \in A\right) \leq \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} M_n \in A\right) \leq -\inf_{x\in\overline{A}} I(x);$$

here A^o and $\overline{A}$ denote the interior and closure of A, respectively.

The MDP can thus be seen as an intermediate behavior between the central limit theorem ($b_n = b\sqrt{n}$) and large deviation ($b_n = bn$). Usually, the MDP exhibits a simpler rate function inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the large deviation principle.

Our paper is organized as follows. Section 2 states the moments control inequalities and their consequences. We shall state in this section a first result on the MDP for BMC in a general framework, but with a very restricted range of speed. Section 3 deals with the exponential inequalities and their consequences. In this section, we shall generalize the MDP done in Section 2, allowing for a larger range of speed, but under more stringent assumptions. In Section 4, we will focus particularly on the first order bifurcating autoregressive processes. The proofs of some inequalities are technical so postponed in Appendix A. Appendix B is devoted to definitions and limit theorems for martingales used intensively in the paper, and are included here for completeness.

2. Moments control and consequences. Let F be a vector subspace of $\mathcal{B}(S)$ such that:

(i) F contains the constants;
(ii) $F^2 \subset F$;
(iii) $F \otimes F \subset L^1(P(x, \cdot))$ for all $x \in S$, and $P(F \otimes F) \subset F$;
(iv) there exists a probability μ on $(S, \mathcal{S})$ such that $F \subset L^1(\mu)$ and

$$\lim_{r\to\infty} \mathbb{E}_x[f(Y_r)] = (\mu, f)$$

for all $x \in S$ and $f \in F$;
(v) for all $f \in F$, there exists $g \in F$ such that for all $r \in \mathbb{N}$, $|Q^r f| \leq g$;
(vi) $F \subset L^1(\nu)$,

where we have used the notation $F^2 = \{f^2/f \in F\}$, $F \otimes F = \{f \otimes g/f, g \in F\}$ and $PE = \{Pf/f \in E\}$ whenever an operator P acts on a set E.

The following hypothesis is about the geometric ergodicity of Q:

(H1) Assume that for all $f \in F$ such that $(\mu, f) = 0$, there exists $g \in F$ such that for all $r \in \mathbb{N}$ and for all $x \in S$, $|Q^r f(x)| \leq \alpha^r g(x)$ for some $\alpha \in (0, 1)$; that is, the Markov chain $(Y_r, r \in \mathbb{N})$ is geometrically ergodic.

Recall that under this hypothesis, Guyon [14] has shown the weak law of large numbers for the three empirical average $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$ (see [14], Theorem 11 when $f \in F$ and Theorem 12 when $f \in \mathcal{B}(S^3)$) and the strong law of large numbers only for $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$; see [14], Theorem 14 and Corollary 15 when $f \in F$ and Theorem 18 when $f \in \mathcal{B}(S^3)$.

When $f \in \mathcal{B}(S^3)$ and under the additional hypothesis Pf^2 and Pf^4 exist and belong to F, he proved the central limit theorem for $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$; see [14], Theorem 19 and Corollary 21. Recall that the central limit theorem for the three empirical means (1.2) when $f \in \mathcal{B}(S)$ is still an open question; see [8] for more precision.

In this section, we complete these results by showing the strong law of large numbers for $\overline{M}_n^{\Pi}(f)$, when $f \in F$. We prove also the law of the iterated logarithm (LIL) and almost sure functional central limit theorem (ASFCLT) for $\overline{M}_n^{\Pi}(f)$ when $f \in \mathcal{B}(S^3)$.

2.1. *Control of the* 4*th order moments*. In order to establish limit theorems below, let us state the following:

THEOREM 2.1. *Let F satisfy* (i)–(vi). *Let $f \in F$ such that $(\mu, f) = 0$. We assume hypothesis* (H1). *Then for all $r \in \mathbb{N}$*,

$$\mathbb{E}\big[\big(\overline{M}_{\mathbb{G}_r}(f)\big)^4\big] \le \begin{cases} c\big(\frac{1}{4}\big)^r, & \text{if } \alpha^2 < \frac{1}{2}, \\ cr^2\big(\frac{1}{4}\big)^r, & \text{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{4r}, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.1}$$

where the positive constant c depends on α and f (and may differ line by line).

PROOF. First note that $f(X_i) \in L^4$ for all $i \in \mathbb{G}_r$. Indeed, let $(z_1, \dots, z_r) \in \{0, 1\}^r$ the unique path in the binary tree from the root 1 to i. Then,

$$\mathbb{E}\big[f^4(X_i)\big] = \nu P_{z_1} \cdots P_{z_r} f^4,$$

and from hypotheses (ii), (iii) and (vi) we conclude that $\nu P_{z_1} \cdots P_{z_r} f^4 < \infty$.

Now, the proof divides into two parts.

Part 1. *Computation of* $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$. Independently of X, let us draw four independent indices I_r, J_r, K_r and L_r uniformly from $\mathbb{G}_r$. Then

$$\mathbb{E}\big[\big(\overline{M}_{\mathbb{G}_r}(f)\big)^4\big] = \mathbb{E}\big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big].$$

For all $p \in \{0, \dots, r\}$, let us define the following events:

- E_0^p: The ancestors of I_r, J_r, K_r and L_r are different in $\mathbb{G}_p$.
- E_1^p: Exactly two of I_r, J_r, K_r and L_r have the same ancestor in $\mathbb{G}_p$.
- E_2^p: I_r, J_r, K_r and L_r have the same ancestor two by two in $\mathbb{G}_p$.

- E_3^p: Exactly three of I_r, J_r, K_r and L_r have the same ancestor in $\mathbb{G}_p$.
- E_4^p: I_r, J_r, K_r and L_r have the same ancestor in $\mathbb{G}_p$.

We also consider the following events whose for each fixed $p \leq r$, probability depend only on p.

- $E_0'^p$: Draw uniformly four independent indices from $\mathbb{G}_p$ which are different.
- $E_1'^p$: Draw uniformly four independent indices from $\mathbb{G}_p$ such that two are the same, and the others are different.
- $E_2'^p$: Draw uniformly four independent indices from $\mathbb{G}_p$ which are the same, two by two.
- $E_3'^p$: Draw uniformly four independent indices from $\mathbb{G}_p$ such that exactly three are the same.
- $E_4'^p$: Draw uniformly four independent indices from $\mathbb{G}_p$ which are all the same.

In the sequel we do the convention that E_0^{r+1} is a certain event. Then after successive conditioning by events E_i^p for $p \in \{0, \dots, r\}$ and $i \in \{0, \dots, 4\}$, we have

$$
\begin{aligned}
&\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})] \\
&\quad = \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^2] \times \mathbb{P}(E_0^2) \\
&\qquad + \sum_{p=2}^{r} \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_1^p] \times \mathbb{P}(E_1^p \cap E_0^{p+1}) \\
&\qquad + \sum_{p=2}^{r} \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_2^p] \times \mathbb{P}(E_2^p \cap E_0^{p+1}) \\
&\qquad + \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_3^r] \times \mathbb{P}(E_3^r) \\
&\qquad + \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_4^r] \times \mathbb{P}(E_4^r).
\end{aligned} \tag{2.2}
$$

Let us notice that

- for all $i \in \{1, 2, 3, 4\}$, E_i^r and $E_i'^r$ have the same probability;
- the realization of "$E_1^p \cap E_0^{p+1}$" can be seen as "draw uniformly four independent indices from $\mathbb{G}_p$ such that two are the same and others are different, and the two indices which are the same take different paths at $\mathbb{G}_{p+1}$." Thus "$E_1^p \cap E_0^{p+1}$" has the same probability that "$E_1'^p \cap A_{p,p+1}$," where "$A_{p,p+1}$" is the event, "the indices which are the same in $\mathbb{G}_p$ take different paths at $\mathbb{G}_{p+1}$";
- similarly, the realization of "$E_2^p \cap E_0^{p+1}$" may be interpreted as, "draw uniformly four independent indices from $\mathbb{G}_p$ which are the same two by two, and all the indices take different paths at $\mathbb{G}_{p+1}$." Thus "$E_2^p \cap E_0^{p+1}$" has the same probability that "$E_2'^p \cap A_{p,p+1}$," where "$A_{p,p+1}$" is the event, "the indices which are the same in $\mathbb{G}_p$ take different paths at $\mathbb{G}_{p+1}$";

- for all $p \in \{0, \dots, r\}$, we have

$$\mathbb{P}(E_1'^p) = \frac{6(2^p-1)(2^p-2)}{2^{3p}}, \qquad \mathbb{P}(E_2'^p) = \frac{3(2^p-1)}{2^{3p}},$$
$$\mathbb{P}(E_3'^p) = \frac{4(2^p-1)}{2^{3p}}, \qquad \mathbb{P}(E_4'^p) = \frac{1}{2^{3p}}.$$

We may then deduce that

$$\mathbb{P}(E_0^2) = \frac{3}{32}, \qquad \mathbb{P}(E_3^r) = \frac{4(2^r-1)}{2^{3r}}, \qquad \mathbb{P}(E_4^r) = \frac{1}{2^{3r}}$$

and for $p \in \{2, \dots, r-1\}$,

$$\mathbb{P}(E_1^p \cap E_0^{p+1}) = \mathbb{P}(E_1'^p)\mathbb{P}(A_{p,p+1}/E_1'^p) = \frac{3(2^p-1)(2^p-2)}{2^{3p}}$$

and

$$\mathbb{P}(E_2^p \cap E_0^{p+1}) = \mathbb{P}(E_2'^p)\mathbb{P}(A_{p,p+1}/E_2'^p) = \frac{3}{4}\frac{2^p-1}{2^{3p}}.$$

We are now going to compute each term which appears in (2.2). We have the following convention: $P(Q^{-1}f \otimes Q^{-1}f) = f^2$. In the sequel, we will use intensively, with a slight modification, the calculations made by Guyon [14] in order to compute conditional expectations related to the event, "draw uniformly two independent indices from $\mathbb{G}_p$," for $p \in \{0, \dots, r\}$.

(a) We have that

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_4^r] = \nu Q^r f^4.$$

(b) Conditionally on E_3^r, we may assume that the indices I_r, K_r and L_r are the same. We then have, using the calculations made by Guyon [14],

$$\begin{aligned}
&\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_3^r] \\
&\quad = \mathbb{E}[f^3(X_{I_r})f(X_{J_r})/E_3^r] \\
&\quad = \frac{2^r}{2^r-1}\Bigg\{\sum_{p=0}^{r-1} 2^{-p-2}\nu Q^p P\big(Q^{r-p-1}f^3 \otimes Q^{r-p-1}f \\
&\qquad\qquad + Q^{r-p-1}f \otimes Q^{r-p-1}f^3\big)\Bigg\}.
\end{aligned}$$

(c) Let $p \in \{2, \dots, r\}$. Conditionally on E_2^p and E_0^{p+1} we may assume that I_r and J_r have the same ancestor at $\mathbb{G}_p$, and K_r and L_r have the same ancestor at $\mathbb{G}_p$. For simplification, we will use the following notation:

$$Q_\otimes^k f := Q^k f \otimes Q^k f, \tag{2.3}$$

and we thus have

$$
\begin{aligned}
&\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_2^p] \\
&\quad = \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/\mathcal{F}_{p+1}]/\mathcal{F}_p]/E_0^{p+1}, E_2^p] \\
&\quad = \mathbb{E}[P(Q_\otimes^{r-p-1} f)(X_{I_r \wedge_p J_r}) P(Q_\otimes^{r-p-1} f)(X_{K_r \wedge_p L_r})/E_0^{p+1}, E_2^p] \\
&\quad = \frac{2^p}{2^p-1} \sum_{l=0}^{p-1} 2^{-l-1} \nu Q^l P((Q^{p-l-1} P(Q_\otimes^{r-p-1} f)) \\
&\qquad\qquad\qquad \otimes (Q^{p-l-1} P(Q_\otimes^{r-p-1} f))),
\end{aligned}
$$

where $I_r \wedge_p J_r$ (resp., $K_r \wedge_p L_r$) denotes the common ancestor of I_r and J_r which is in $\mathbb{G}_p$ (resp., the common ancestor of K_r and L_r which is in $\mathbb{G}_p$).

(d) Let $p \in \{2, \dots, r\}$. Now conditionally on E_1^p and E_0^{p+1} we may assume that it is K_r and L_r which have the same ancestor in $\mathbb{G}_p$. We denote by $p(I_r)$ and $p(J_r)$, respectively, the ancestor of I_r and J_r which are in $\mathbb{G}_p$. As before, the common ancestor of K_r and L_r, which are in $\mathbb{G}_p$, is denoted by $K_r \wedge_p L_r$. At this step, we may repeat the successive conditioning that we have done in the beginning but this time for indices $p(I_r)$, $p(J_r)$ and $K_r \wedge_p L_r$. This leads us to

$$
\begin{aligned}
&\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_1^p] \\
&\quad = \mathbb{E}[Q^{r-p} f(X_{p(I_r)}) Q^{r-p} f(X_{p(J_r)}) P(Q_\otimes^{r-p-1} f)(X_{K_r \wedge_p L_r})/E_0^{p+1}, E_1^p] \\
&\quad = \frac{2^{2p}}{(2^p-1)(2^p-2)} \sum_{l=1}^{p-1} \frac{1}{2^{l+1}} \frac{1}{2} \\
&\qquad \times \sum_{m=0}^{l-1} 2^{-m-1} \{\nu Q^m P((Q^{l-m-1} P(Q_\otimes^{r-l-1} f)) \otimes Q^{p-m-1} P(Q_\otimes^{r-p-1} f)) \\
&\qquad + \nu Q^m P((Q^{p-m-1} P(Q_\otimes^{r-p-1} f)) \otimes (Q^{l-m-1} P(Q_\otimes^{r-l-1} f))) \\
&\qquad + \nu Q^m P((Q^{l-m-1} P(Q^{r-l-1} f \otimes Q^{p-l-1} P(Q_\otimes^{r-p-1} f))) \otimes (Q^{r-m-1} f)) \\
&\qquad + \nu Q^m P(Q^{r-m-1} f \otimes (Q^{l-m-1} P(Q^{r-l-1} f \otimes Q^{p-l-1} P(Q_\otimes^{r-p-1} f)))) \\
&\qquad + \nu Q^m P((Q^{l-m-1} P(Q^{p-l-1} P(Q_\otimes^{r-p-1} f) \otimes Q^{r-l-1} f)) \otimes (Q^{r-m-1} f)) \\
&\qquad + \nu Q^m P((Q^{r-m-1} f) \otimes (Q^{l-m-1} P(Q^{p-l-1} P(Q_\otimes^{r-p-1} f) \otimes Q^{r-l-1} f)))\}.
\end{aligned}
$$

(e) Finally,

$$
\begin{aligned}
&\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^2] \\
&\quad = \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/\mathcal{F}_2]/\mathcal{F}_1]/E_0^2]
\end{aligned}
$$

$$= \mathbb{E}\big[P\big(Q_{\otimes}^{r-2} f\big)(X_2) P\big(Q_{\otimes}^{r-2} f\big)(X_3)/E_0^2\big]$$
$$= \nu P\big(P\big(Q_{\otimes}^{r-2} f\big) \otimes P\big(Q_{\otimes}^{r-2} f\big)\big).$$

Gathering together all of these terms, each multiplied by their respective probability, we obtain an explicit expression for $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$.

Part 2. *Rate*. We are now going to give some rates for the different terms that appear in the expression of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$.

Throughout this part, we will use intensively the following to bound quantities which appear in the expression of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$:

- Let $f \in F$ such that $(\mu, f) = 0$. Then from (i)–(vi) and hypothesis (H1), there exists a positive constant c such that $\forall l, m, n \in \mathbb{N}$,
$$\nu Q^l P\big(Q^m f \otimes Q^n f\big) \leq \alpha^{m+n} \nu Q^l P(g \otimes g) \leq c\alpha^{m+n},$$
where g is given in hypothesis (H1).

In the sequel, c denotes a positive constant which depends on f, and c_1 denotes a positive constant which depends on α. The constants c and c_1 may vary from one line to another and from one expression to another.

(a) For the first term appearing in (2.2), we have

$$\mathbb{E}\big[f(X_{I_r}) f(X_{J_r}) f(X_{K_r}) f(X_{L_r})/E_0^2\big] \times \mathbb{P}\big(E_0^2\big) \leq c_1 c\alpha^{4r}.$$

(b) For the fifth term appearing in (2.2), we have

$$\mathbb{E}\big[f(X_{I_r}) f(X_{J_r}) f(X_{K_r}) f(X_{L_r})/E_4^r\big] \times \mathbb{P}\big(E_4^r\big) \leq c\big(\tfrac{1}{2}\big)^{3r},$$

where, from (ii), (v) and (vi), c is such that $\nu Q^r f^4 < c$.

(c) For the fourth term appearing in (2.2), we have

$$\mathbb{E}\big[f(X_{I_r}) f(X_{J_r}) f(X_{K_r}) f(X_{L_r})/E_3^r\big] \times \mathbb{P}\big(E_3^r\big) \leq cc_1\alpha^r \left(\frac{1}{4}\right)^r \sum_{p=0}^{r-1} \left(\frac{1}{2\alpha}\right)^p,$$

where, from (ii), (iii), (v) and (vi), c is such that for all $p, q \in \mathbb{N}$

$$\max\big(\nu Q^p P\big(Q^q f^3 \otimes g\big), \nu Q^p P\big(g \otimes Q^q f^3\big)\big) < c,$$

and from hypothesis (H1), g is such that for all $p \in \{1, \dots, r-1\}$

$$Q^{r-p-1} f \leq \alpha^{r-p-1} g. \tag{2.4}$$

Now depending on the value of α, we obtain that

$$\mathbb{E}\big[f(X_{I_r}) f(X_{J_r}) f(X_{K_r}) f(X_{L_r})/E_3^r\big] \times \mathbb{P}\big(E_3^r\big)$$
$$\leq \begin{cases} c_1 c\left(\left(\dfrac{\alpha}{4}\right)^r + \left(\dfrac{1}{2^3}\right)^r\right), & \text{if } \alpha \neq \dfrac{1}{2}, \\ c_1 cr\left(\dfrac{1}{2^3}\right)^r, & \text{if } \alpha = \dfrac{1}{2}. \end{cases}$$

(d) Let us denote the third term appearing in (2.2) by

$$A_r := \sum_{p=2}^{r} \mathbb{E}\big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_2^p\big] \times \mathbb{P}\big(E_2^p \cap E_0^{p+1}\big).$$

So we have

$$A_r \le c_1 c\left(\left(\frac{1}{4}\right)^r + \alpha^{4r}\sum_{p=2}^{r-1}\left(\frac{1}{4\alpha^4}\right)^p\right),$$

where, from (ii), (iii), (v) and (vi), c is such that for all $p \in \{2, \dots, r-1\}$, $q \in \{0, \dots, r-1\}$, $l \in \{0, \dots, p-1\}$

$$\max\big(\nu Q^q P\big(Q_{\otimes}^{r-q-1} f^2\big), \nu Q^l P\big(Q_{\otimes}^{p-l-1} P(g \otimes g)\big)\big) < c,$$

and g is defined as before (2.4) and the notation $Q_{\otimes}$ is given in (2.3).

Now depending on the value of α, we obtain that:

- if $\alpha^2 \neq \frac{1}{2}$, then $A_r \le c_1 c((\frac{1}{4})^r + \alpha^{4r})$;
- if $\alpha^2 = \frac{1}{2}$, then $A_r \le c_1 c(r-1)(\frac{1}{4})^r$.

(e) For the second term appearing in (2.2), we have when $p = r$:

- if $\alpha = \frac{1}{2}$, then

$$\mathbb{E}\big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_1^r\big] \times \mathbb{P}\big(E_1^r\big) \le c_1 c\big(\tfrac{1}{4}\big)^r;$$

- if $\alpha \neq \frac{1}{2}$:
  - if $\alpha^2 = \frac{1}{2}$, then

$$\mathbb{E}\big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_1^r\big] \times \mathbb{P}\big(E_1^r\big) \le c_1(r-1)\big(\tfrac{1}{4}\big)^r;$$

  - if $\alpha^2 \neq \frac{1}{2}$, then

$$\begin{aligned}\mathbb{E}\big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_1^r\big] \times \mathbb{P}\big(E_1^r\big)\\ \le c_1 c\left(\left(\frac{\alpha^2}{2}\right)^r + \left(\frac{1}{4}\right)^r\right),\end{aligned}$$

where, from (ii), (iii), (v) and (vi), c is such that for all $l \in \{2, \dots, r-1\}$, $q \in \{0, \dots, l-1\}$

$$\begin{aligned}\max\big(&\nu Q^q P\big(Q^{l-q-1} P(g \otimes g) \otimes Q^{r-q-1} f^2\big),\\ &\nu Q^q P\big(Q^{l-q-1} P\big(g \otimes Q^{r-l-1} f^2\big) \otimes g\big)\big) < c\end{aligned}$$

and g is defined as before (2.4).

(f) For the second terms appearing in (2.2), and for the remaining term in the sum ($p \neq r$), let us denote by

$$B_r := \sum_{p=2}^{r-1} \mathbb{E}\big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_1^p\big] \times \mathbb{P}\big(E_1^p \cap E_0^{p+1}\big).$$

So we have:

- if $\alpha = \frac{1}{2}$, then $B_r \leq c_1 c(\frac{1}{4})^r$;
- if $\alpha \neq \frac{1}{2}$:
  - if $\alpha^2 = \frac{1}{2}$, then $B_r \leq c_1 c r^2 (\frac{1}{4})^r$;
  - if $\alpha^2 \neq \frac{1}{2}$, then $B_r \leq c_1 c(\alpha^{4r} + (\frac{\alpha^2}{2})^r + (\frac{1}{4})^r)$,

where c is defined in the same way as before.

Now the results of the Theorem 2.1 follow from (a)–(f) of part 2. $\square$

It leads us to an extension of Theorem 2.1 to the two empirical averages $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$.

COROLLARY 2.2. *Let F satisfy* (i)–(vi). *Let $f \in F$ such that $(\mu, f) = 0$. We assume that hypothesis* (H1) *is fulfilled. Then for all $r \in \mathbb{N}$ and $n \in \mathbb{N}$,*

$$\mathbb{E}\big[\big(\overline{M}_{\mathbb{T}_r}(f)\big)^4\big] \leq \begin{cases} c\big(\frac{1}{4}\big)^{r+1}, & \textit{if } \alpha^2 < \frac{1}{2}, \\ cr^2\big(\frac{1}{4}\big)^{r+1}, & \textit{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{4(r+1)}, & \textit{if } \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.5}$$

and

$$\mathbb{E}\big[\big(\overline{M}_n^{\Pi}(f)\big)^4\big] \leq \begin{cases} c\big(\frac{1}{4}\big)^{r_n+1}, & \textit{if } \alpha^2 < \frac{1}{2}, \\ cr_n^2\big(\frac{1}{4}\big)^{r_n+1}, & \textit{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{4(r_n+1)}, & \textit{if } \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.6}$$

where the positive constant c depends on α and f and may differ line by line.

PROOF. The proof follows the same steps as in the proof of parts 2 and 3 of Theorem 2.11, and uses the results of the proof of Theorem 2.5 to get the control of the 4th order moment in incomplete generation. See Sections 2.2 and A.1 for more detail. $\square$

REMARK 2.3. If $f \in \mathcal{B}(S^3)$ is such that Pf^2 and Pf^4 exist and belong to F, with $Pf = 0$, then we have for all $r \in \mathbb{N}$ and for some positive constant c,

$$\mathbb{E}\big[\big(\overline{M}_{\mathbb{G}_r}(f)\big)^4\big] \leq \frac{c}{|\mathbb{G}_r|^2}. \tag{2.7}$$

Indeed, let $M_{\mathbb{G}_r}(f) = \sum_{i\in\mathbb{G}_r} f(\Delta_i)$. We have

$$\begin{aligned}
\mathbb{E}\big[(M_{\mathbb{G}_r}(f))^4\big] &= \mathbb{E}[M_{\mathbb{G}_r}(f^4)] + 6\mathbb{E}\Big[\sum_{i\neq j\in\mathbb{G}_r} f^2(\Delta_i)f^2(\Delta_j)\Big] \\
&\quad + 4\mathbb{E}\Big[\sum_{i\neq j\in\mathbb{G}_r} f^3(\Delta_i)f(\Delta_j)\Big] \\
&\quad + 12\mathbb{E}\Big[\sum_{i\neq j\neq k\in\mathbb{G}_r} f^2(\Delta_i)f(\Delta_j)f(\Delta_k)\Big] \\
&\quad + 24\mathbb{E}\Big[\sum_{i\neq j\neq k\neq l\in\mathbb{G}_r} f(\Delta_i)f(\Delta_j)f(\Delta_k)f(\Delta_l)\Big] \\
&= \mathbb{E}\Big[\sum_{i\in\mathbb{G}_r} Pf^4(X_i)\Big] + 6\mathbb{E}\Big[\sum_{i\neq j\in\mathbb{G}_r} Pf^2(X_i)Pf^2(X_j)\Big],
\end{aligned}$$

where the last equality was obtained after conditioning by $\mathcal{F}_r$ and using the fact that $Pf = 0$. Now, dividing by $|\mathbb{G}_r|^4$ leads us to

$$\begin{aligned}
\mathbb{E}\big[(\overline{M}_{\mathbb{G}_r}(f))^4\big] &= \frac{6}{|\mathbb{G}_r|^2}\mathbb{E}\Big[\frac{1}{|\mathbb{G}_r|^2}\sum_{i\neq j\in\mathbb{G}_r} Pf^2(X_i)Pf^2(X_j)\Big] \\
&\quad + \frac{1}{|\mathbb{G}_r|^3}\mathbb{E}\Big[\frac{1}{|\mathbb{G}_r|}\sum_{i\in\mathbb{G}_r} Pf^4(X_i)\Big] \\
&\leq \frac{6}{|\mathbb{G}_r|^2}\mathbb{E}\big[(\overline{M}_{\mathbb{G}_r}(Pf^2))^2\big] \\
&\quad + \frac{1}{|\mathbb{G}_r|^3}\mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^4)],
\end{aligned}$$

and (2.7) then follows from the control of

$$(\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf^2))^2])_r \quad \text{and} \quad (\mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^4)])_r;$$

see [14].

REMARK 2.4. From Remark 2.3, we deduce that if $f \in \mathcal{B}(S^3)$ is such that Pf^2 and Pf^4 exist and belong to F, with $Pf = 0$, then we have for all $r \in \mathbb{N}$ and for some positive constant c,

$$\mathbb{E}\big[(\overline{M}_{\mathbb{T}_r}(f))^4\big] \leq c\big(\tfrac{1}{4}\big)^{r+1}. \tag{2.8}$$

Indeed, from the equality

$$\overline{M}_{\mathbb{T}_r}(f) = \sum_{q=0}^{r} \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \overline{M}_{\mathbb{G}_q}(f),$$

we deduce that

$$\mathbb{E}\big[\big(\overline{M}_{\mathbb{T}_r}(f)\big)^4\big] \le \left(\sum_{q=0}^{r} \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \|\overline{M}_{\mathbb{G}_q}(f)\|_4\right)^4,$$

where $\|\cdot\|_4$ stands for the L^4-norm. We then infer from (2.7) that

$$\mathbb{E}\big[\big(\overline{M}_{\mathbb{T}_r}(f)\big)^4\big] \le c\left(\sum_{q=0}^{r} \frac{(\sqrt{2})^q}{2^{r+1}}\right)^4$$

for some positive constant c. (2.8) then follows from the last inequality.

2.2. *Strong law of large numbers on incomplete subtree*. We now turn to prove the strong law of large numbers for $\overline{M}_n^{\Pi}(f)$, completing the work of Guyon [14], where the LLN was proved only for the two averages $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_{\mathbb{G}_r}(f)$.

THEOREM 2.5. *Let F satisfy* (i)–(vi). *Let $f \in F$ such that $(\mu, f) = 0$. We assume that hypothesis* (H1) *is fulfilled with $\alpha \in (0, \frac{\sqrt[4]{8}}{2})$. Then $\overline{M}_n^{\Pi}(f)$ almost surely converges to* 0 *as n goes to ∞.*

PROOF. From the decomposition

$$\overline{M}_n^{\Pi}(f) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n} \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}),$$

it is enough to check that

$$\sum_{n=1}^{\infty} \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^4\right] < \infty.$$

Indeed, since $\overline{M}_{\mathbb{G}_q}(f)$ almost surely converges to 0 (Corollary 15 in [14]), we deduce that the first term on the right-hand side of the previous decomposition almost surely converges to 0 (Lemma 13 in [14]). We have

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^4\right] \\ &= \frac{1}{n^4}\mathbb{E}\left[\sum_{i=2^{r_n}}^{n} f^4(X_{\Pi(i)})\right] + \frac{6}{n^4}\mathbb{E}\left[\sum_{i,j=2^{r_n}; i\neq j}^{n} f^2(X_{\Pi(i)}) f^2(X_{\Pi(j)})\right] \end{aligned} \tag{2.9}$$

$$+\frac{4}{n^4}\mathbb{E}\Bigg[\sum_{i,j=2^{r_n};i\neq j}^{n} f^3(X_{\Pi(i)})f(X_{\Pi(j)})\Bigg]$$
$$+\frac{12}{n^4}\mathbb{E}\Bigg[\sum_{i,j,k=2^{r_n};i\neq j\neq k}^{n} f^2(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})\Bigg]$$
$$+\frac{24}{n^4}\mathbb{E}\Bigg[\sum_{i,j,k,l=2^{r_n};i\neq j\neq k\neq l}^{n} f(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})f(X_{\Pi(l)})\Bigg].$$

We will control each term appearing in decomposition (2.9). For the first term on the right-hand side of (2.9), using (ii), (v) and (vi) we have for some positive constant c,

$$\mathbb{E}\Bigg[\sum_{i=2^{r_n}}^{n} f^4(X_{\Pi(i)})\Bigg]=\big(n-2^{r_n}+1\big)\nu Q^{r_n}f^4\leq c\big(n-2^{r_n}+1\big),$$

which implies that

$$\frac{1}{n^4}\mathbb{E}\Bigg[\sum_{i=2^{r_n}}^{n} f^4(X_{\Pi(i)})\Bigg]=O\bigg(\frac{1}{n^3}\bigg). \tag{2.10}$$

Recall the following: for i, j, k and $l\in\{2^{r_n},\dots,n\}$:

- If $i\neq j$, then $r_n\geq 1$. Independently on (X,Π), draw two independent indices I_{r_n} and J_{r_n} uniformly from $\mathbb{G}_{r_n}$. Then the law of $(\Pi(i),\Pi(j))$ is the conditional law of (I_{r_n},J_{r_n}) given $\{I_{r_n}\neq J_{r_n}\}$.
- If $i\neq j\neq k$, then $r_n\geq 2$. Independently on (X,Π), draw three independent indices I_{r_n}, J_{r_n} and K_{r_n} uniformly from $\mathbb{G}_{r_n}$. Then the law of $(\Pi(i),\Pi(j),\Pi(k))$ is the conditional law of $(I_{r_n},J_{r_n},K_{r_n})$ given $\{I_{r_n}\neq J_{r_n}\neq K_{r_n}\}$.
- If $i\neq j\neq k\neq l$, then $r_n\geq 2$. Independently on (X,Π), draw four independent indices $I_{r_n}, J_{r_n}, K_{r_n}$ and L_{r_n} uniformly from $\mathbb{G}_{r_n}$. Then the law of $(\Pi(i),\Pi(j),\Pi(k)),\Pi(l))$ is the conditional law of $(I_{r_n},J_{r_n},K_{r_n},L_{r_n})$ given $\{I_{r_n}\neq J_{r_n}\neq K_{r_n}\neq J_{r_n}\}$.

Now we have to control the second and third terms of (2.9). We have to check that

$$\frac{1}{n^4}\mathbb{E}\Bigg[\sum_{i,j=2^{r_n};i\neq j}^{n} f^2(X_{\Pi(i)})f^2(X_{\Pi(j)})\Bigg]=O\bigg(\frac{1}{n^2}\bigg) \tag{2.11}$$

and

$$\frac{1}{n^4}\mathbb{E}\Bigg[\sum_{i,j=2^{r_n};i\neq j}^{n} f^3(X_{\Pi(i)})f(X_{\Pi(j)})\Bigg]=o\bigg(\frac{1}{n^2}\bigg). \tag{2.12}$$

Indeed, from the previous reminder and (i)–(vi), we have for some positive constant c,

$$\begin{aligned}\mathbb{E}\Bigg[\sum_{i,j=2^{r_n};i\neq j}^{n} f^2(X_{\Pi(i)})f^2(X_{\Pi(j)})\Bigg] &= \frac{(n-2^{r_n})(n-2^{r_n}+1)}{(1-2^{-r_n})} \\ &\quad\times \sum_{p=0}^{r_n-1} 2^{-p-1}\nu Q^p P\big(Q_{\otimes}^{r_n-p-1} f^2\big) \\ &\leq c\big(n-2^{r_n}\big)\big(n-2^{r_n}+1\big),\end{aligned}$$

which implies (2.11). In the same way and using in addition hypothesis (H1), we obtain that

$$\begin{aligned}\mathbb{E}\Bigg[\sum_{i,j=2^{r_n};i\neq j}^{n} f^3(X_{\Pi(i)})f(X_{\Pi(j)})\Bigg] &= \frac{(n-2^{r_n})(n-2^{r_n}+1)}{(1-2^{-r_n})} \\ &\quad\times \sum_{p=0}^{r_n-1} 2^{-p-2}\nu Q^p P\big(Q^{r_n-p-1} f^3\otimes Q^{r_n-p-1} f \\ &\qquad\qquad + Q^{r_n-p-1} f\otimes Q^{r_n-p-1} f^3\big) \\ &\leq \begin{cases} c2^{-r_n}\big(n-2^{r_n}\big)\big(n-2^{r_n}+1\big), & \text{if } \alpha<\frac{1}{2}, \\ cr_n 2^{-r_n}\big(n-2^{r_n}\big)\big(n-2^{r_n}+1\big), & \text{if } \alpha=\frac{1}{2}, \\ c\alpha^{r_n}\big(n-2^{r_n}\big)\big(n-2^{r_n}+1\big), & \text{if } \alpha>\frac{1}{2},\end{cases}\end{aligned}$$

which implies (2.12).

Let us deal with the remaining term of (2.9):

$$\begin{aligned}\frac{1}{n^4}\mathbb{E}\Bigg[\sum_{i,j,k=2^{r_n};i\neq j\neq k}^{n} f^2(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})\Bigg] &= \frac{(n-2^{r_n}-1)(n-2^{r_n})(n-2^{r_n}+1)}{\mathbb{P}(I_{r_n}\neq J_{r_n}\neq K_{r_n})\times n^4} \\ &\quad\times \mathbb{E}\big[f^2(X_{I_{r_n}})f(X_{J_{r_n}})f(X_{K_{r_n}})\mathbf{1}_{\{I_{r_n}\neq J_{r_n}\neq K_{r_n}\}}\big].\end{aligned}$$

Then, we get an explicit expression for the last expectation similar to that obtained in part (d) of the calculus of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$ with a slight modification of the functions. Calculating the rate of this expression, we obtain

$$\begin{aligned}\sum_{n=4}^{\infty}\frac{1}{n^4}\mathbb{E}\Bigg[\sum_{i,j,k=2^{r_n};i\neq j\neq k}^{n} f^2(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})\Bigg]\\ \leq c\sum_{n=1}^{\infty}\frac{1}{n}\alpha^{2r_n}+c\sum_{n=1}^{\infty}\sum_{p=2}^{r_n-1}\sum_{l=0}^{p-1}\frac{1}{n}\frac{1}{2^p}\frac{1}{2^{l+1}}\alpha^{2r_n-2p}\\ +c\sum_{n=1}^{\infty}\sum_{p=2}^{r_n-1}\sum_{l=0}^{p-1}\frac{1}{n}\frac{1}{2^p}\frac{1}{2^{l+1}}\alpha^{2r_n-p-l}\end{aligned}$$

for some positive c. Now it is not hard to see that the right-hand side is finite.

Finally, to check that the series of general term

$$\frac{1}{n^4}\mathbb{E}\Bigg[\sum_{i,j,k,l=2^{r_n};i\neq j\neq k\neq l}^{n} f(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})f(X_{\Pi(l)})\Bigg]$$

is finite, it is enough, according to the calculation of rates we have done in part 2 of the proof of Theorem 2.1, to check that $\sum_{n=1}^{\infty}\alpha^{4r_n}<\infty$, which is the case if $\alpha\in(0,\frac{\sqrt[4]{8}}{2})$, and this completes the proof of Theorem 2.5. $\square$

REMARK 2.6. Note that this theorem can be improved, but the price to pay is enormous computations related to the calculation of higher moments. If f is bounded, this result is true for every $\alpha\in(0,1)$, as we will see in Section 3.

2.3. *Law of the iterated logarithm* (*LIL*). Using the LIL for martingales (see Theorem B.3 of Stout in Appendix B), we are going to prove a LIL for the BMC. This will be done when f depends on the mother-daughters triangle (Δ_i). We use the notation $M_n^{\Pi}(f)=\sum_{i=1}^{n}f(\Delta_{\Pi(i)})$ and $M_{\mathbb{T}_r}(f)=\sum_{i\in\mathbb{T}_r}f(\Delta_i)$.

THEOREM 2.7. *Let F satisfy* (i)–(vi). *Let $f\in\mathcal{B}(\mathcal{S}^3)$ such that $Pf=0$, Pf^2 and Pf^4 exist and belong to F. We assume that hypothesis* (H1) *is fulfilled. Then*

$$\limsup_{n\to\infty}\frac{M_n^{\Pi}(f)}{\sqrt{2\langle M^{\Pi}(f)\rangle_n\log\log\langle M^{\Pi}(f)\rangle_n}}=1 \qquad a.s.$$

And in particular,

$$\limsup_{r\to\infty}\frac{M_{\mathbb{T}_r}(f)}{\sqrt{2|\mathbb{T}_r|\log\log|\mathbb{T}_r|}}=\sqrt{(\mu,Pf^2)} \qquad a.s.$$

PROOF. We will check the hypothesis of Stout Theorem's B.3. Let $f\in\mathcal{B}(\mathcal{S}^3)$. We introduce the filtration $(\mathcal{H}_n)_{n\geq 0}$ defined by $\mathcal{H}_0=\sigma(X_1)$ and $\mathcal{H}_n=$

$\sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \le i \le n)$. Let $(M_n^{\Pi}(f))_{n\ge 0}$ defined by $M_0^{\Pi}(f) = 0$ and $M_n^{\Pi}(f) = \sum_{i=1}^{n} f(\Delta_{\Pi(i)})$. Then since $Pf = 0$, $(M_n^{\Pi}(f))$ is a $\mathcal{H}_n$-martingale with $\mathbb{E}[M_1^{\Pi}(f)] = 0$. The bracket of the above martingale is given by

$$\langle M^{\Pi}(f)\rangle_n = \sum_{i=0}^{n} Pf^2(X_{\Pi(i)}) = M_n^{\Pi}(Pf^2).$$

We have the following decomposition:

$$\frac{\langle M^{\Pi}(f)\rangle_n}{n} = \overline{M}_n^{\Pi}(Pf^2) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(Pf^2) + \frac{1}{n} \sum_{i=2^{r_n}}^{n} Pf^2(X_{\Pi(i)}).$$

Since

$$\forall q \le r_n - 1 \qquad \frac{2^q}{2^{r_n+1}} \le \frac{2^q}{n} \le \frac{2^q}{2^{r_n}} \quad \text{and} \quad \frac{1}{n} \sum_{i=2^{r_n}}^{n} Pf^2(X_{\Pi(i)}) \le \overline{M}_{\mathbb{G}_{r_n}}(Pf^2),$$

we deduce that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{2^{r_n+1}} \overline{M}_{\mathbb{G}_q}(Pf^2) \le \overline{M}_n^{\Pi}(Pf^2) \le \sum_{q=0}^{r_n} \frac{2^q}{2^{r_n}} \overline{M}_{\mathbb{G}_q}(Pf^2).$$

From the strong law of large numbers of $\overline{M}_{\mathbb{G}_q}(Pf^2)$ (see [14], Corollary 15) and from Lemma 5.2 of [7], we infer that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{2^{r_n+1}} \overline{M}_{\mathbb{G}_q}(Pf^2) \xrightarrow{\text{a.s.}} \frac{(\mu, Pf^2)}{2} \quad \text{and} \quad \sum_{q=0}^{r_n} \frac{2^q}{2^{r_n}} \overline{M}_{\mathbb{G}_q}(Pf^2) \xrightarrow{\text{a.s.}} 2(\mu, Pf^2).$$

Using these results, we thus deduce that $\langle M^{\Pi}(f)\rangle_n = O(n)$ and $n = O(\langle M^{\Pi}(f)\rangle_n)$ a.s. This implies in particular that $\langle M^{\Pi}(f)\rangle_n \underset{n\to\infty}{\longrightarrow} \infty$ a.s.

Now let $K_n = \frac{\sqrt{2}}{\sqrt{\log\log(n)}}$ in Theorem B.3, and we have

$$\begin{aligned} R &:= \sum_{n=1}^{\infty} \frac{2\log\log\langle M^{\Pi}(f)\rangle_n}{K_n^2 \langle M^{\Pi}(f)\rangle_n} \\ &\qquad \times \mathbb{E}\big[f^2(\Delta_{\Pi(n)}) \mathbf{1}_{\{f^2(\Delta_{\Pi(n)}) > K_n^2 \langle M^{\Pi}(f)\rangle_n/(2\log\log\langle M^{\Pi}(f)\rangle_n)\}}/\mathcal{H}_{n-1}\big] \\ &\le \sum_{n=1}^{\infty} \frac{4(\log\log\langle M^{\Pi}(f)\rangle_n)^2}{K_n^4 (\langle M^{\Pi}(f)\rangle_n)^2} Pf^4(X_{\Pi(n)}) \qquad \text{a.s.,} \end{aligned}$$

since $\langle M^{\Pi}(f)\rangle_n = O(n)$ a.s., so that for $R < \infty$ a.s., it is enough to check that

$$\sum_{n=1}^{\infty} \frac{Pf^4(X_{\Pi(n)})}{n^{\delta}} < \infty \qquad \text{a.s. with any } 1 < \delta < 2. \tag{2.13}$$

Now, according to (v) and (vi), there exists a positive constant c such that for all $n \geq 1$, $\mathbb{E}[Pf^4(X_{\Pi(n)})] = \nu Q^{r_n} Pf^4 \leq c$, and (2.13) follows. Applying Theorem B.3, we have

$$\limsup_{n\to\infty} \frac{M_n^{\Pi}(f)}{\sqrt{2\langle M^{\Pi}(f)\rangle_n \log\log\langle M^{\Pi}(f)\rangle_n}} = 1 \qquad \text{a.s.}$$

Now, for $n = |\mathbb{T}_r|$, we have the following:

$$\frac{M_{\mathbb{T}_r}(f)}{\sqrt{2\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|} \log\log\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}}$$
$$= \sqrt{\frac{|\mathbb{T}_r|\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{2\log\log\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}} \times \frac{M_{\mathbb{T}_r}(f)}{|\mathbb{T}_r|\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}/|\mathbb{T}_r|}$$

and since $\frac{\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}{|\mathbb{T}_r|} = \overline{M}_{\mathbb{T}_r}(Pf^2) \underset{r\to\infty}{\longrightarrow} (\mu, Pf^2)$ a.s. (see Theorem 18 in [14]), we get

$$\limsup_{r\to\infty} \frac{M_{\mathbb{T}_r}(f)}{\sqrt{2|\mathbb{T}_r|\log\log|\mathbb{T}_r|}} = \sqrt{(\mu, Pf^2)} \qquad \text{a.s.,}$$

which completes the proof. $\square$

REMARK 2.8. Let us note that using Theorem 2.5, we can prove that if hypothesis (H1) is fulfilled with $\alpha \in (0, \frac{\sqrt[4]{8}}{2})$, then

$$\limsup_{n\to\infty} \frac{M_n^{\Pi}(f)}{\sqrt{2n\log\log n}} = \sqrt{(\mu, Pf^2)} \qquad \text{a.s.,}$$

and via the computation of $2k$th order moments of $\overline{M}_{\mathbb{G}_r}(g)$, with $k > 2$ and $g \in \mathcal{B}(S)$, it is possible to prove the latter for all $\alpha \in (0, 1)$. But, as already emphasized, this comes at the price of enormous computations.

2.4. *Almost-sure functional central limit theorem* (*ASFCLT*). We are now going to prove an ASFCLT theorem for the BMC $(X_n, n \in \mathbb{T})$. Here again, this will be done when f depends on the mother-daughters triangle by using the ASFCLT for discrete time martingale. We refer to Chaabane, Theorem B.4, Appendix B, for the definition of an ASFCLT.

THEOREM 2.9. *Let F satisfy* (i)–(vi). *Let $f \in \mathcal{B}(\mathcal{S}^3)$ such that $Pf = 0$, Pf^2 and Pf^4 exist and belong to F. We assume that hypothesis* (H1) *is fulfilled with $\alpha \in (0, \frac{\sqrt[4]{8}}{2})$. Then $M_n^{\Pi}(f)$ verifies an ASFCLT, when n goes to ∞.*

PROOF. We use Theorem B.4. Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be the filtration defined as in Section 2.3. Then $(M_n^{\Pi}(f))$ is a $\mathcal{H}_n$ martingale. We have to check the hypotheses of Theorem B.4. For all $n \geq 1$, let $V_n = s\sqrt{n}$ where $s^2 = (\mu, Pf^2)$. Then according to Theorem 2.5,

$$\frac{\langle M^{\Pi}(f)\rangle_n}{V_n^2} = V_n^{-2} M_n^{\Pi}(Pf^2) \underset{n\to\infty}{\longrightarrow} 1 \qquad \text{a.s.}$$

Let $\varepsilon > 0$. We have

$$\begin{aligned}\sum_{n\geq 1} \frac{1}{V_n^2}\mathbb{E}\big[f^2(\Delta_{\Pi(n)})\mathbf{1}_{\{|f(\Delta_{\Pi(n)})|>\varepsilon V_n\}}/\mathcal{H}_{n-1}\big] \\ \leq \frac{1}{\varepsilon^2 s^4}\sum_{n\geq 1}\frac{Pf^4(X_{\Pi(n)})}{n^2} \qquad \text{a.s.}\end{aligned}$$

According to (v) and (vi), there exists a positive constant c such that for all $n \geq 1$, $\mathbb{E}[Pf^4(X_{\Pi(n)})] = \nu Q^{r_n} Pf^4 \leq c$, and therefore, $\forall \varepsilon > 0$

$$\sum_{n\geq 1} \frac{1}{V_n^2}\mathbb{E}\big[f^2(\Delta_{\Pi(n)})\mathbf{1}_{\{|f(\Delta_{\Pi(n)})|>\varepsilon V_n\}}/\mathcal{H}_{n-1}\big] < \infty \qquad \text{a.s.}$$

Finally, we have

$$\sum_{n\geq 1} \frac{1}{V_n^4}\mathbb{E}\big[f^4(\Delta_{\Pi(n)})\mathbf{1}_{\{|f(\Delta_{\Pi(n)})|\leq V_n\}}/\mathcal{H}_{n-1}\big] \leq \frac{1}{s^4}\sum_{n\geq 1}\frac{Pf^4(X_{\Pi(n)})}{n^2} \qquad \text{a.s.,}$$

which as before is a.s. finite, and the proof is then complete. □

REMARK 2.10. As before, let us note that this result can be extended to the general case $\alpha \in (0, 1)$, but at the price of enormous computation related to the computation of $2k$-order moments, $k > 2$, for $\overline{M}_{\mathbb{G}_r}(g)$, $g \in \mathcal{B}(S)$.

2.5. *Deviation inequalities for BMC.* We are now going to give some deviation inequalities under (i)–(vi) and (H1) for the empirical means (1.2) when $f \in \mathcal{B}(S)$ with $(\mu, f) = 0$ and when $f \in \mathcal{B}(S^3)$ with $(\mu, Pf) = 0$. This will help us in the sequel to obtain a MDP result in a general framework, that is, for functional of BMC with unbounded test functions. Let us recall that the main disadvantage of this "weak" set of assumptions is that the range of speed for the MDP is very restricted. However, we still work under geometric ergodicity assumption and general test function, which will not be the case when we would want to extend the MDP; see Section 3. Note that we postpone to Appendix A nearly all the proofs of this section, these proofs being quite long and technical.

THEOREM 2.11. *Let F satisfy conditions* (i)–(vi). *We assume that* (H1) *is fulfilled. Let $f \in F$ such that $(\mu, f) = 0$. Then we have for all $\delta > 0$ and all $r \in \mathbb{N}$ and all $n \in \mathbb{N}$,*

$$\mathbb{P}\big(\big|\overline{M}_{\mathbb{G}_r}(f)\big| > \delta\big) \leq \begin{cases} \dfrac{c}{\delta^2}\left(\dfrac{1}{2}\right)^r, & \text{if } \alpha^2 < \dfrac{1}{2}; \\ \dfrac{c}{\delta^2} r\left(\dfrac{1}{2}\right)^r, & \text{if } \alpha^2 = \dfrac{1}{2}; \\ \dfrac{c}{\delta^2}\alpha^{2r}, & \text{if } \alpha^2 > \dfrac{1}{2}; \end{cases} \tag{2.14}$$

$$\mathbb{P}\big(\big|\overline{M}_n^{\Pi}(f)\big| > \delta\big) \leq \begin{cases} \dfrac{c}{\delta^2}\left(\dfrac{1}{2}\right)^{r_n+1}, & \text{if } \alpha^2 < \dfrac{1}{2}; \\ \dfrac{c}{\delta^2} r_n\left(\dfrac{1}{2}\right)^{r_n+1}, & \text{if } \alpha^2 = \dfrac{1}{2}; \\ \dfrac{c}{\delta^2}\alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}; \end{cases} \tag{2.15}$$

and

$$\mathbb{P}\big(\big|\overline{M}_{\mathbb{T}_r}(f)\big| > \delta\big) \leq \begin{cases} \dfrac{c}{\delta^2}\left(\dfrac{1}{2}\right)^{r+1}, & \text{if } \alpha^2 < \dfrac{1}{2}; \\ \dfrac{c}{\delta^2} r\left(\dfrac{1}{2}\right)^{r+1}, & \text{if } \alpha^2 = \dfrac{1}{2}; \\ \dfrac{c}{\delta^2}\alpha^{2(r+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}; \end{cases} \tag{2.16}$$

where the positive constant c depends on f and α and may differ term by term.

PROOF. See Section A.1 in Appendix A. □

We shall also need an extension of Theorem 2.11 to the case when f does not only depend on an individual X_i, but on the mother-daughters triangle (Δ_i).

THEOREM 2.12. *Let F satisfy conditions* (i)–(vi). *We assume that* (H1) *is fulfilled. Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. Then we have the same conclusion as in Theorem 2.11 for the three empirical averages given in* (1.2): $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ *and* $\overline{M}_n^{\Pi}(f)$.

PROOF. See Section A.2 in Appendix A. □

We thus have the following first result on the superexponential convergence in probability, whose definition we present now:

DEFINITION 2.13. Let (E, d) a metric space. Let (Z_n) be a sequence of random variables valued in E, Z be a random variable valued in E and (v_n) be a rate. We say that Z_n converges v_n-superexponentially fast in probability to Z if for all

$\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mathbb{P}\big(d(Z_n, Z) > \delta\big) = -\infty.$$

This "exponential convergence" with speed v_n will be shortened as

$$Z_n \overset{\text{superexp}}{\underset{v_n}{\longrightarrow}} Z.$$

We may now set:

PROPOSITION 2.14. *Let F satisfy conditions* (i)–(vi). *Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. We assume that* (H1) *is fulfilled. Let (b_n) be a sequence of increasing positive real numbers such that*

$$\text{(2.17)} \qquad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \qquad \frac{b_n}{\sqrt{n \log n}} \longrightarrow 0, \qquad \frac{n}{b_n} \text{ is nondecreasing}.$$

Then

$$\overline{M}_n^{\Pi}(f) \overset{\text{superexp}}{\underset{b_n^2/n}{\longrightarrow}} 0.$$

PROOF. The proof is a direct consequence of Theorem 2.12. □

2.6. *Moderate deviations for BMC.* Now, using the MDP for martingale (see, e.g., [11, 24]), we are going to prove a MDP for BMC. We will use Proposition B.5, in Appendix B.

THEOREM 2.15. *Let F satisfy conditions* (i)–(vi). *We assume that* (H1) *is satisfied. Let $f \in \mathcal{B}(S^3)$ such that Pf^2 and Pf^4 exist and belong to F. Assume that $Pf = 0$. Let (b_n) be a sequence of increasing positive real numbers satisfying* (2.17). *If*

$$\text{(2.18)} \qquad \limsup_{n\to\infty} \frac{n}{b_n^2} \log\Big(n \operatorname*{ess\,sup}_{1\le k\le c^{-1}(b_{n+1})} \mathbb{P}\big(\big|f(\Delta_{\Pi(k)})\big| > b_n/\mathcal{H}_{k-1}\big)\Big) = -\infty,$$

where $c^{-1}(b_{n+1}) := \inf\{k \in \mathbb{N} : \frac{k}{b_k} \ge b_{n+1}\}$, then $(M_n^{\Pi}(f)/b_n)$ satisfies a MDP in $\mathbb{R}$ with the speed b_n^2/n and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.

PROOF. First, note that under the hypothesis, $M_n^{\Pi}(f)$ is a $\mathcal{H}_n$-martingale, with $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \le i \le n)$. From Proposition B.5 in Appendix B, we only have to check conditions (C1) and (C3).

On one hand, (2.15) applied to $Pf^4 - (\mu, Pf^4)$ implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Pf^4(X_{\Pi(i)}) > (\mu, Pf^4) + \delta\right) = -\infty,$$

and this implies the exponential Lindeberg condition (see, e.g., [24]), that is, condition (C3).

On the other hand, we have $\langle M^{\Pi}(f)\rangle_n = M_n^{\Pi}(Pf^2)$ and (2.15) applied to $Pf^2 - (\mu, Pf^2)$ implies that

$$\overline{M}_n^{\Pi}(Pf^2 - (\mu, Pf^2)) \xrightarrow[b_n^2/n]{\text{superexp}} 0,$$

that is, condition (C1). □

REMARK 2.16. One of the main difficulties in the application of this Theorem lies in the verification of (2.18). Note, however, that in the range of speed considered it is sufficient to have some uniform control in X_i of some moment of $f(X_i, X_{2i}, X_{2i+1})$ conditionally on X_i, which leads to condition of the type $P|f|^k$ bounded for some $k \geq 2$. It is, of course, the case if f is bounded.

REMARK 2.17. In the special case of model (1.1), we have (see Section 4), for f such that $Pf = 0$ and for all k,

$$\mathbb{E}\left[\exp\left(\lambda\frac{b_n}{n} f(\Delta_{\Pi(k)})\right)\Big/\mathcal{H}_{k-1}\right] = \exp\left(\frac{b_n^2}{n}\left(\frac{\lambda^2 Pf^2}{2n}\right)(X_{\Pi(k)})\right).$$

This condition implies that a MDP is satisfied for $(M_n^{\Pi}(f)/b_n)$. Indeed, if this relation is satisfied, we then have that for $\lambda \in \mathbb{R}$ the quantity

$$G_n(\lambda) = \frac{\lambda^2}{2n}\sum_{k=1}^{n} Pf^2(X_{\Pi(k)}) = \frac{\lambda^2}{2}\overline{M}_n^{\Pi}(Pf^2)$$

is an upper and lower cumulant (see, e.g., [24]), and we may apply Gärtner–Ellis-type methodology. In addition, due to (2.15) applied to $Pf^2 - (\mu, Pf^2)$, we have for $\lambda \in \mathbb{R}$,

$$G_n(\lambda) \xrightarrow[b_n^2/n]{\text{superexp}} \frac{\lambda^2(\mu, Pf^2)}{2},$$

which implies that $(M_n^{\Pi}(f)/b_n)$ satisfies a MDP in $\mathbb{R}$ with the speed b_n^2/n and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.

3. Exponential deviation inequalities for BMC and consequences. We give here stronger deviation inequalities than the one obtained in Section 2, namely exponential deviation inequalities. Of course, it requires more stringent assumptions.

3.1. *Exponential deviation inequalities.* Let us consider the following hypothesis.

(H2) There exists a probability μ on $(S, \mathcal{S})$ such that, for all $f \in \mathcal{B}_b(S)$ with $(\mu, f) = 0$, there exists a positive constant c such that

$$|Q^r f(x)| \leq c\alpha^r \qquad \text{for some } \alpha \in (0,1) \text{ and for all } x \in S.$$

One can easily check that, under hypothesis (H2), $\mathcal{B}_b(S)$ fulfills hypothesis (i)–(vi) of the previous section.

Under this assumption, we will prove exponential deviation inequalities for $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$ when $f \in \mathcal{B}_b(S)$ with $(\mu, f) = 0$ [resp., $f \in \mathcal{B}_b(S^3)$ with $(\mu, Pf) = 0$].

THEOREM 3.1. *Assume that* (H2) *is satisfied. Let* $f \in \mathcal{B}_b(S)$ *such that* $(\mu, f) = 0$. *Then we have for all* $\delta > 0$,

$$\mathbb{P}\big(\overline{M}_{\mathbb{G}_r}(f) > \delta\big) \le \begin{cases} \exp(c''\delta)\exp\big(-c'\delta^2|\mathbb{G}_r|\big), & \\ \qquad \forall r \in \mathbb{N}, & \text{if } \alpha \le \dfrac{1}{2}, \\ \exp\big(-c'\delta^2|\mathbb{G}_r|\big), & \\ \qquad \forall r \in \mathbb{N} \text{ such that } r > r_0, & \text{if } \dfrac{1}{2} < \alpha < \dfrac{\sqrt{2}}{2}, \\ \exp\Big(-c'\delta^2\dfrac{|\mathbb{G}_r|}{r}\Big), & \\ \qquad \forall r \in \mathbb{N} \text{ such that } r > r_0, & \text{if } \alpha^2 = \dfrac{1}{2}, \\ \exp\Big(-c'\delta^2\dfrac{1}{\alpha^{2r}}\Big), & \\ \qquad \forall r \in \mathbb{N} \text{ such that } r > r_0, & \text{if } \alpha^2 > \dfrac{1}{2}, \end{cases} \tag{3.1}$$

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \le \begin{cases} \exp(c''\delta)\exp\big(-c'\delta^2|\mathbb{T}_r|\big), & \\ \qquad \forall r \in \mathbb{N}, & \text{if } \alpha < \dfrac{1}{2}, \\ \exp\big(2c'\delta(r+1)\big)\exp\big(-c'\delta^2|\mathbb{T}_r|\big), & \\ \qquad \forall r \in \mathbb{N}, & \text{if } \alpha = \dfrac{1}{2}, \\ \exp\big(-c'\delta^2|\mathbb{T}_r|\big), & \\ \qquad \forall r \in \mathbb{N} \text{ such that } r > r_0 - 1, & \text{if } \dfrac{1}{2} < \alpha < \dfrac{\sqrt{2}}{2}, \\ \exp\Big(-c'\delta^2\dfrac{|\mathbb{T}_r|}{r+1}\Big), & \\ \qquad \forall r \in \mathbb{N} \text{ such that } r > r_0 - 1, & \text{if } \alpha = \dfrac{\sqrt{2}}{2}, \\ \exp\Big(-c'\delta^2\dfrac{1}{\alpha^{2(r+1)}}\Big), & \\ \qquad \forall r \in \mathbb{N}^* \text{ such that } r > r_0 - 3, & \text{if } \alpha > \dfrac{\sqrt{2}}{2}, \end{cases} \tag{3.2}$$

and

$$
\begin{aligned}
&\mathbb{P}(\overline{M}_n^{\Pi}(f) > \delta) \\
&\le \begin{cases}
\exp(c''\delta)\exp(-c'\delta^2 n), & \\
\quad \forall n \in \mathbb{N}, & \text{if } \alpha < \frac{1}{2}, \\
\exp(2c'\delta(r_n+1))\exp(-c'\delta^2 n), & \\
\quad \forall n \in \mathbb{N}, & \text{if } \alpha = \frac{1}{2}, \\
\exp(-c'\delta^2 n), & \\
\quad \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\
\exp\Bigl(-c'\delta^2 \frac{n}{r_n+1}\Bigr), & \\
\quad \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \alpha = \frac{\sqrt{2}}{2}, \\
\exp\Bigl(-c'\delta^2 \frac{1}{\alpha^{2(r_n+1)}}\Bigr), & \\
\quad \forall n \in \mathbb{N}^* \text{ such that } r_n > r_0 - 2, & \text{if } \alpha > \frac{\sqrt{2}}{2},
\end{cases}
\end{aligned}
\tag{3.3}
$$

where $r_0 := \log(\frac{\delta}{c_0})/\log(\alpha)$, *and* c_0, c' *and* c'' *are positive constants which depend on* α *and* f, *and differ line by line*; *see the proofs for the dependence*.

PROOF. The details of the proof are in Section A.3 in Appendix A. It relies mainly on successive conditioning, using carefully the uniform geometric ergodicity assumption to get rid of the conditioning. □

The condition about α less than $1/2$ or greater is of course linked to the binary structure of the tree. The extension to m-ary tree will follow from the same ideas.

THEOREM 3.2. *Assume that* (H2) *is satisfied. Let* $f \in \mathcal{B}_b(S^3)$ *such that* $(\mu, Pf) = 0$. *Then we have the same conclusions*, *for the three empirical averages* $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ *and* $\overline{M}_{\mathbb{T}_r}(f)$, *as in the Theorem* 3.1.

PROOF. See Section A.4 in Appendix A. □

Now, using the Borel–Cantelli Theorem and (3.3), we state easily the following:

COROLLARY 3.3. *Assume that* (H2) *is satisfied. Let* $f \in \mathcal{B}_b(S)$ *such that* $(\mu, f) = 0$ [*resp*., $f \in \mathcal{B}_b(S^3)$ *and* $(\mu, Pf) = 0$]. *Then* $\overline{M}_n^{\Pi}(f)$ *almost surely converges to* 0 *as* n *goes to* ∞.

REMARK 3.4. Of course uniform ergodicity and bounded test functions are surely a very strong set of assumptions, but it is not so difficult to verify if the Markov chain's daughters lie in a compact set. We are convinced that it is possible to consider the geometric ergodic case and bounded test functions, but for the price of tedious calculations that we will pursue in an other work. We will also investigate the use of transportation inequalities, leading to deviation inequality for Lipschitz test functions under some Wasserstein contraction property for the kernel P, in the spirit of the Theorems 2.5 or 2.11 in [12].

3.2. *Moderate deviation principle for BMC.* We introduce the following assumption on the speed of the MDP.

ASSUMPTION 1. Let (b_n) be an increasing sequence of positive real numbers such that

$$\frac{b_n}{\sqrt{n}} \longrightarrow +\infty$$

and:

- if $\alpha^2 < \frac{1}{2}$, the sequence (b_n) is such that $b_n/n \longrightarrow 0$;
- if $\alpha^2 = \frac{1}{2}$, the sequence (b_n) is such that $(b_n \log n)/n \longrightarrow 0$;
- if $\alpha^2 > \frac{1}{2}$, the sequence (b_n) is such that $(b_n \alpha^{r_n+1})/\sqrt{n} \longrightarrow 0$.

Using the MDP for martingale with bounded jumps (see, e.g., [9, 11]), we can now state the following:

THEOREM 3.5. *Assume that* (H2) *is satisfied. Let $f \in \mathcal{B}_b(S^3)$ such that $Pf = 0$. Let (b_n) be a sequence of real numbers satisfying the Assumption 1; then $(M_n^{\Pi}(f)/b_n)$ satisfies a MDP in S with the speed b_n^2/n and rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.*

PROOF. The proof easily follows from the previous exponential probability inequalities and the MDP for martingale with bounded jumps; see, for example, [9, 11, 24]. □

REMARK 3.6. Taking particularly $n = |\mathbb{T}_r|$ and (b_n) as a sequence of real numbers satisfying Assumption 1, we get that for all $f \in \mathcal{B}_b(S^3)$, $(M_{\mathbb{T}_r}(f)/b_{|\mathbb{T}_r|})$ satisfies a MDP in $\mathbb{R}$ with the speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.

4. Application: First order Bifurcating autoregressive processes. In this section, we seek to apply the results of the previous sections to the following bifurcating autoregressive process with memory 1 defined by

$$\mathcal{L}(X_1) = \nu \quad \text{and} \quad \forall n \geq 1 \qquad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{4.1}$$

where $\alpha_0, \alpha_1 \in (-1, 1)$; $\beta_0, \beta_1 \in \mathbb{R}$, $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of i.i.d. bivariate random variables and ν a probability measure on $\mathbb{R}$.

Several extensions of the model have been proposed and various estimators are studied in the literature for the unknown parameters; see, for instance, [2, 17–19, 25, 26]. See [4] for a relevant references.

Throughout this section, we assume that the distribution ν has finite moments of all orders.

In the sequel, we will study (4.1) in two settings:

- the Gaussian setting which corresponds to the case where $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of i.i.d. bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$ with

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad \sigma^2 > 0, \qquad \rho \in (-1, 1); \tag{4.2}$$

- the bounded setting which corresponds to the case where X_1 and $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$, which forms a sequence of centered i.i.d. bivariate random variables, take their values in a compact set. Let us note that in this case, $(X_n, n \in \mathbb{T})$ takes its values in a compact set.

Our main goal is to give deviation inequalities and MDP for the estimator of the 4-dimensional unknown parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ and for the statistical test defined in [14].

To estimate the 4-parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$, as well as σ^2 and ρ, assume we observe a complete subtree $\mathbb{T}_{r+1}$. The least square estimator $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$ of θ is given by (see [14]), for $\eta \in \{0, 1\}$,

$$\begin{cases} \hat{\alpha}_\eta^r = \dfrac{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i X_{2i+\eta} - (|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i)(|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\eta})}{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i^2 - (|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i)^2}, \\ \hat{\beta}_\eta^r = |\mathbb{T}_r|^{-1} \displaystyle\sum_{i \in \mathbb{T}_r} X_{2i+\eta} - \hat{\alpha}_\eta^r |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i. \end{cases} \tag{4.3}$$

Notice that in the Gaussian case, this least square estimator corresponds to the maximum likelihood estimator.

We also need to introduce the estimators of the conditional variance σ^2 and the conditional sister–sister correlation ρ. These estimators are naturally given by

$$\begin{cases} \hat{\sigma}_r^2 = \dfrac{1}{2\mathbb{T}_r} \displaystyle\sum_{i \in \mathbb{T}_r} (\hat{\varepsilon}_{2i}^2 + \hat{\varepsilon}_{2i+1}^2), \\ \hat{\rho}_r = \dfrac{1}{\hat{\sigma}_r^2} \displaystyle\sum_{i \in \mathbb{T}_r} \hat{\varepsilon}_{2i} \hat{\varepsilon}_{2i+1}, \end{cases} \tag{4.4}$$

where the residues are defined by $\hat{\varepsilon}_{2i+\eta} = X_{2i+\eta} - \hat{\alpha}^r_\eta X_i - \hat{\beta}^r_\eta$, with $\eta \in \{0, 1\}$.

Let us denote by $\mathcal{C}_{\text{pol}}(\mathbb{R})$ [resp., $\mathcal{C}_{\text{pol}}(\mathbb{R}^3)$] the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ (resp., $f : \mathbb{R}^3 \to \mathbb{R}$) such that $|f|$ is bounded above by a polynomial. From [14], we know that $\mathcal{C}_{\text{pol}}(\mathbb{R})$ fulfills hypotheses (i)–(vi).

We will take $F = \mathcal{C}^1_{\text{pol}}(\mathbb{R})$ the set of all $\mathcal{C}^1$ functions $f : \mathbb{R} \to \mathbb{R}$ such that $|f| + |f'|$ is bounded above by a polynomial. Then, one can check that F fulfills hypotheses (i)–(vi). Moreover, for all $f \in F$, hypothesis (H1) holds with $\alpha = \max(|\alpha_0|, |\alpha_1|)$. Let μ be the unique stationary distribution of the induced Markov chain $(Y_r, r \in \mathbb{N})$; see [14] for more details.

Let us denote by $\mathcal{C}^1_{\text{pol}}(\mathbb{R}^3)$ the set of all $\mathcal{C}^1$ functions $f : \mathbb{R}^3 \to \mathbb{R}$ such that $|f| + |f'|$ is bounded above by a polynomial. We shall denote by $\mathbf{x}$ (resp., $\mathbf{x}^2$, $\mathbf{xy}$, $\mathbf{y}, \ldots$) the element of $\mathcal{C}^1_{\text{pol}}(\mathbb{R}^3)$ defined by $(x, y, z) \mapsto x$ (resp., $x^2, xy, y, \ldots$).

We define two continuous functions $\mu_1 : \Theta \to \mathbb{R}$ and $\mu_2 : \Theta \times \mathbb{R}^*_+ \to \mathbb{R}$ by writing

$$(\mu, \mathbf{x}) = \mu_1(\theta) \quad \text{and} \quad (\mu, \mathbf{x}^2) = \mu_2(\theta, \sigma^2), \tag{4.5}$$

where $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1) \in \Theta = (-1, 1) \times \mathbb{R} \times (-1, 1) \times \mathbb{R}$.

To segregate between $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ and its alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$, we shall use the test statistic

$$\chi^{(1)}_r = \frac{|\mathbb{T}_r|}{2\hat{\sigma}^2_r(1 - \hat{\rho}_r)}\{(\hat{\alpha}^r_0 - \hat{\alpha}^r_1)^2(\hat{\mu}_{2,r} - \hat{\mu}^2_{1,r}) + ((\hat{\alpha}^r_0 - \hat{\alpha}^r_1)\hat{\mu}_{1,r} + \hat{\beta}^r_0 - \hat{\beta}^r_1)^2\},$$

where we write $\hat{\mu}_{1,r} = \mu_1(\hat{\theta}^r)$ and $\hat{\mu}_{2,r} = \mu_2(\hat{\theta}_r, \hat{\sigma}_r)$.

As usual the Gaussian setting has specific properties that allow easier calculations and more general assumptions.

4.1. *The Gaussian setting*. We introduce the following assumption on the speed of the MDP. Let (b_n) be an increasing sequence of positive real numbers such that

$$\frac{b_n}{\sqrt{n}} \longrightarrow +\infty \quad \text{and} \quad \frac{b_n}{\sqrt{n \log n}} \to 0. \tag{4.6}$$

PROPOSITION 4.1. *Let (b_n) be a sequence of real numbers satisfying (4.6). Then*

$$\hat{\theta}^r \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} \theta.$$

PROOF. We will treat the case of $\hat{\alpha}^r_0$ given in (4.3). The others, $\hat{\beta}^r_0, \hat{\alpha}^r_1$ and $\hat{\beta}^r_1$, given in (4.3), may be treated in a similar way. Note that $\hat{\alpha}^r_0 = \frac{C_r}{B_r}$, where

$$C_r = \overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})\overline{M}_{\mathbb{T}_r}(\mathbf{y}) \quad \text{and} \quad B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2.$$

Now, using Lemma B.2 and Proposition 2.14, it follows that

$$\hat{\alpha}_0^r \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} \alpha_0. \qquad \square$$

We recall that in the BAR model (4.1), we use $\alpha = \max\{|\alpha_0|, |\alpha_1|\}$, and $b := \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2$, where μ_1 and μ_2 are given in (4.5), so we have the following deviation inequality:

PROPOSITION 4.2. *For all $\delta > 0$, for all $r \in \mathbb{N}$ and for all $\gamma < \min(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}})$, where c_1 is a positive constant which depends on μ_1, we have*

$$\mathbb{P}(\|\hat{\theta}^r - \theta\| > \delta) \leq \begin{cases} \dfrac{c}{\gamma^{4q}\delta^{4-p}}\left(\dfrac{1}{4}\right)^{r+1}, & \text{if } \alpha^2 < \dfrac{1}{2}, \\ \dfrac{c}{\gamma^{4q}\delta^{4-p}} r^2 \left(\dfrac{1}{4}\right)^{r+1} & \text{if } \alpha^2 = \dfrac{1}{2}, \\ \dfrac{c}{\gamma^{4q}\delta^{4-p}} \alpha^{4(r+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}, \end{cases} \tag{4.7}$$

where the constant c depends on α, μ_1, μ_2 and differs line by line, $p = p(\delta) \in \{0, 2, 4\}$ and $q = q(\delta) \in \{0, 1\}$.

REMARK 4.3. The values of p and q in Proposition 4.2 depend on the order of δ. For example, if δ is small enough, we have $p = 0$ and $q = 0$.

PROOF. See Section A.5 in Appendix A. $\square$

REMARK 4.4. Proposition 4.2 can be improved by calculating the $2k$th order moments, with $k > 2$, as in the proof of Theorem 2.1. But, as we have said, this comes at the price of enormous computation.

PROPOSITION 4.5. *Let (b_n) be a sequence of real numbers satisfying (4.6). Then*

$$(\hat{\sigma}_r^2, \hat{\rho}_r) \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} (\sigma^2, \rho).$$

PROOF. Let us first deal with σ_r^2 given in (4.4). We have (see, e.g., [14])

$$\hat{\sigma}_r^2 = \tfrac{1}{2}\overline{M}_{\mathbb{T}_r}(f(\cdot, \theta)) + D_r,$$

where $f(x, y, z, \theta) = (y - \alpha_0 x - \beta_0)^2 + (z - \alpha_1 x - \beta_1)^2$ and

$$D_r = \frac{1}{2|\mathbb{T}_r|}\sum_{i \in \mathbb{T}_r}(f(\Delta_i, \hat{\theta}^r) - f(\Delta_i, \theta)).$$

By the Taylor–Lagrange formula, we can find $g \in \mathcal{C}_{\mathrm{pol}}(\mathbb{R}^3)$ such that (see [14])

$$|D_r| \leq \frac{1}{2}\|\hat{\theta}^r - \theta\|\big(1 + \|\theta\| + \|\hat{\theta}^r - \theta\|\big)\overline{M}_{\mathbb{T}_r}(g).$$

Now, Propositions 2.14 and 4.1 lead us to

$$\hat{\sigma}_r^2 \xrightarrow[b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|]{\text{superexp}} \sigma^2.$$

The proof for $\hat{\rho}_r$ given in (4.4) is similar. $\square$

PROPOSITION 4.6. *Let (b_n) be a sequence of real numbers satisfying (4.6). Then the sequence $(|\mathbb{T}_r|(\hat{\theta}^r - \theta)/b_{|\mathbb{T}_r|})$ satisfies the MDP on $\mathbb{R}^4$ with the speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function I given by*

$$I(x) = \frac{1}{2}x^t(\Sigma')^{-1}x,$$

where

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix}$$

with

$$K = \frac{1}{\mu_2(\theta, \sigma^2) - \mu_1(\theta)^2}\begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta, \sigma^2) \end{pmatrix}.$$

PROOF. We first observe that

$$\frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}}(\hat{\theta}^r - \theta) = M(A_r, B_r).\frac{U^r(f)}{b_{|\mathbb{T}_r|}},$$

where $f = (f_1, f_2, f_3, f_4)^t = (\mathbf{xy}, \mathbf{y}, \mathbf{xz}, \mathbf{z})^t$, $U^r(f) = M_{\mathbb{T}_r}(f - Pf)$, $A_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x})$, $B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2$ and

$$M(A_r, B_r) = \begin{pmatrix} \dfrac{1}{B_r} & \dfrac{-A_r}{B_r} & 0 & 0 \\ \dfrac{-A_r}{B_r} & \dfrac{B_r + A_r^2}{B_r} & 0 & 0 \\ 0 & 0 & \dfrac{1}{B_r} & \dfrac{-A_r}{B_r} \\ 0 & 0 & \dfrac{-A_r}{B_r} & \dfrac{B_r + A_r^2}{B_r} \end{pmatrix}.$$

For the sake of simplicity we wrote $Pf = (Pf_1, Pf_2, Pf_3, Pf_4)^t$, where P denotes the $\mathbb{T}$-transition probability associated to BAR(1) process in the Gaussian case, which is given by

$$P(x, dy, dz) = \frac{1}{2\pi\sigma^2(1-\rho^2)} \times \exp\left(-\frac{1}{2}\begin{pmatrix} y - \alpha_0 x - \beta_0 \\ z - \alpha_1 x - \beta_1 \end{pmatrix}^t \Gamma^{-1} \begin{pmatrix} y - \alpha_0 x - \beta_0 \\ z - \alpha_1 x - \beta_1 \end{pmatrix}\right) dy\, dz,$$

where Γ is the covariance matrix defined in (4.2).

On one hand, from Proposition 2.14,

$$A_r \xrightarrow[b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|]{\text{superexp}} a := \mu_1(\theta) \quad \text{and} \quad B_r \xrightarrow[b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|]{\text{superexp}} b := \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2,$$

so that by Lemma B.2, we obtain

$$M(A_r, B_r) \xrightarrow[b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|]{\text{superexp}} M(a, b) := \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}.$$

On the other hand, let $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^t \in \mathbb{R}^4$. For all $x \in \mathbb{R}$, we have that

$$\begin{aligned}
&P\exp\big(\lambda^t(f - Pf)\big)(x) \\
&\quad = \int_{\mathbb{R}^2} \exp\Bigg(\sum_{i=1}^4 \lambda_i (f_i - Pf_i)\Bigg)(x, y, z) P(x, dy, dz) \\
&\quad = \int_{\mathbb{R}^2} \exp\left(\lambda^t \begin{pmatrix} xy - x(\alpha_0 x + \beta_0) \\ y - \alpha_0 x - \beta_0 \\ xz - x(\alpha_1 x + \beta_1) \\ z - \alpha_1 x - \beta_1 \end{pmatrix}\right) P(x, dy, dz) \\
&\quad = \exp\left(-\begin{pmatrix} \alpha_0 x + \beta_0 \\ \alpha_1 x + \beta_1 \end{pmatrix}^t \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right) \\
&\qquad \times \int_{\mathbb{R}^2} \exp\left(\begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \begin{pmatrix} y \\ z \end{pmatrix}\right) P(x, dy, dz).
\end{aligned}$$

We know that

$$\begin{aligned}
&\int_{\mathbb{R}^2} \exp\left(\begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \begin{pmatrix} y \\ z \end{pmatrix}\right) P(x, dy, dz) \\
&\quad = \exp\left(\begin{pmatrix} \alpha_0 x + \beta_0 \\ \alpha_1 x + \beta_1 \end{pmatrix}^t \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right) \\
&\qquad \times \exp\left(\frac{1}{2}\begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \Gamma \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right).
\end{aligned}$$

Let $\Xi(x)$ denote the square matrix with entries $(Pf_i f_j - Pf_i Pf_j)(x)$, for $1 \leq i, j \leq 4$. So we obtain that

$$\begin{aligned}
P\exp\big(\lambda^t(f - Pf)\big)(x) &= \exp\left(\frac{1}{2}\begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \Gamma \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right) \\
&= \exp\Bigg(\frac{1}{2}\sum_{i,j=1}^4 \lambda_i \lambda_j (Pf_i f_j - Pf_i Pf_j)(x)\Bigg) \\
&= \exp\Big(\frac{1}{2}\lambda^t \Xi(x) \lambda\Big).
\end{aligned}$$

Recall that the filtration $(\mathcal{H}_n)_{n\geq 0}$ is defined by $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \leq i \leq n)$. Therefore, from the previous calculations, we deduce that for all $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}\big[\exp\big(\lambda^t (f - Pf)(\Delta_{\Pi(k)})\big)/\mathcal{H}_{k-1}\big] &= P\big(\exp\big(\lambda^t (f - Pf)\big)\big)(X_{\Pi(k)}) \\ &= \exp\big(\tfrac{1}{2}\lambda^t \,\Xi(X_{\Pi(k)})\lambda\big).\end{aligned}$$

Now, recall that $(M_n^\Pi(f - Pf))_{n\in\mathbb{N}}$ is a $(\mathcal{H}_n)$-martingale and by straightforward calculations, its increasing process is given by $\langle M^\Pi(f - Pf)\rangle_n = \sum_{k=1}^n \Xi(X_{\Pi(k)})$. From the foregoing, we infer that

$$\left(\exp\left(\lambda^t M_n^\Pi(f - Pf) - \frac{\lambda^t \langle M^\Pi(f - Pf)\rangle_n \lambda}{2}\right)\right)_{n\in\mathbb{N}}$$

is a $(\mathcal{H}_n)$-martingale. It then follows that for all $\lambda \in \mathbb{R}^4$, $G_n(\lambda) = \frac{1}{2n}\lambda^t \langle M^\Pi(f - Pf)\rangle_n \lambda$ is an upper and lower cumulant. Moreover, from Proposition 2.14 and Lemma B.2,

$$G_n(\lambda) \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} \frac{1}{2}\lambda^t \Sigma\lambda \qquad \text{where } \Sigma = \sigma^2 \begin{pmatrix} K^{-1} & \rho K^{-1} \\ \rho K^{-1} & K^{-1} \end{pmatrix}.$$

We thus deduce that (see, e.g., [24]) $(M_n^\Pi(f)/b_n)$ satisfies a MDP on $\mathbb{R}^4$ with speed b_n^2/n and the rate function

$$J(x) = \tfrac{1}{2}x^t \Sigma^{-1} x. \tag{4.8}$$

Taking $n = |\mathbb{T}_r|$, it follows that $(U^r(f)/b_{|\mathbb{T}_r|})$ satisfies a MDP with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function J given in (4.8). Finally, using the contraction principle (see, e.g., [10]) as in [23], we get the result. □

Let us now consider the test statistic.

PROPOSITION 4.7. *Let (b_n) a sequence of real numbers satisfying (4.6). Then under the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$, $\frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}}(\chi_r^{(1)})^{1/2}$ satisfies a MDP on $\mathbb{R}$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function*

$$I'(y) = \begin{cases} \dfrac{y^2}{2}, & \text{if } y \in \mathbb{R}_+, \\ +\infty, & \text{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1 of H_0, we have for all $A > 0$,

$$\limsup_{r\to\infty} \frac{|\mathbb{T}_r|}{b^2_{|\mathbb{T}_r|}} \log \mathbb{P}\big(\chi_r^{(1)} < A\big) = -\infty.$$

PROOF. We have

$$H_0 = \{g(\theta) = 0\} \qquad \text{where } g(\theta) = (\alpha_0 - \alpha_1, \beta_0 - \beta_1)^t.$$

From Proposition 4.6, $(|\mathbb{T}_r|(\hat{\theta}^r - \theta)/b_{|\mathbb{T}_r|})$ satisfies a MDP on $\mathbb{R}^4$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function $I(x) = \frac{1}{2}x^t(\Sigma')^{-1}x$. So that, using the delta method for the MDP (see, e.g., [13], Theorem 3.1) we conclude that $(|\mathbb{T}_r|(g(\hat{\theta}^r) - g(\theta))/b_{|\mathbb{T}_r|})$ satisfies a MDP on $\mathbb{R}^2$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function

$$J(y) = \inf\{I(x); y = g'(\theta)x\}.$$

Identification of this rate function by usual optimization argument leads us to

$$J(x) = \tfrac{1}{2}x^t(\Sigma'')^{-1}x \qquad \text{where } \Sigma'' = 2\sigma^2(1-\rho)K. \tag{4.9}$$

Under the null hypothesis H_0, we have $g(\theta) = 0$, so that $(|\mathbb{T}_r|g(\hat{\theta}^r)/b_{|\mathbb{T}_r|})$ satisfies a MDP on $\mathbb{R}^2$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and rate function J given in (4.9).

Now, since $K = K(\theta, \sigma)$ is a continuous function of (θ, σ) (see [14]), so that, letting $\hat{K}_r = K(\hat{\theta}^r, \hat{\sigma}_r)$, Lemma B.2, Propositions 4.6 and 4.5 entail that

$$\hat{\Sigma}''_r = 2\hat{\sigma}_r^2(1-\hat{\rho}_r)\hat{K}_r \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} \Sigma''.$$

It follows using the contraction principle (see, e.g., [23]) that

$$(|\mathbb{T}_r|\hat{\Sigma}''^{-1/2}_r g(\hat{\theta}^r)/b_{|\mathbb{T}_r|})$$

satisfies a MDP on $\mathbb{R}^2$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function $J'(y) = \frac{\|y\|^2}{2}$.

In particular,

$$\left\| \frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}} \hat{\Sigma}''^{-1/2}_r g(\hat{\theta}^r) \right\| = \frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}} \sqrt{\chi_r^{(1)}}$$

satisfies a MDP with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function I' given in the Proposition 4.7.

Now, under the alternative hypothesis H_1,

$$\frac{\chi_r^{(1)}}{|\mathbb{T}_r|} = g(\hat{\theta}^r)^t\hat{\Sigma}''^{-1}_r g(\hat{\theta}^r) \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} g(\theta)^t(\Sigma'')^{-1}g(\theta) > 0,$$

so that $\chi_r^{(1)}$ converges $\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}$-superexponentially fast to $+\infty$. This concludes the proof of the Proposition 4.7. $\square$

4.2. *Compact case*: *The uniformly ergodic setting*. We recall that the model under study in this section is the model (4.1) where we assume that the noise and initial state X_1 take their values in a compact set. The results will be given without proofs, since the proofs are similar to those done in the previous section. The novelty here is that the range of speed is improved in comparison to the previous section. However, we suppose that the process takes its values in a compact set, which is not the case in the previous section.

We take $F = \mathcal{C}_b^1(\mathbb{R})$ the set of all $\mathcal{C}^1$ functions bounded on $\mathbb{R}$. Therefore, one can easily check (as in [14], proof of Proposition 28) that hypothesis (H2) is satisfied with $\alpha = \max(|\alpha_0|, |\alpha_1|)$. We use the same notation as in the previous section.

Let us begin by the fact that the estimator of θ converges super exponentially fast to the true parameter.

PROPOSITION 4.8. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then we have*

$$\hat{\theta}^r \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} \theta.$$

We may now refine this result by proving deviation inequality.

PROPOSITION 4.9. *For all $\delta > 0$ and for all $\gamma < \min(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}})$, where c_1 is a positive constant which depends on μ_1, and for $r_0 := \frac{\log(\gamma^q \delta^{1-p/2}/c_0)}{\log \alpha}$, we have*

$$\mathbb{P}(\|\hat{\theta}^r - \theta\| > \delta) \leq \begin{cases} c_2 \exp(c''\gamma^q \delta^{1-p/2}) \exp(-c'\gamma^{2q}\delta^{2-p}|\mathbb{T}_r|), \\ \qquad \forall r \in \mathbb{N}, \qquad \text{if } \alpha < \dfrac{1}{2}, \\ c_2 \exp(c'\gamma^q \delta^{1-p/2}(r+1) - c'\gamma^{2q}\delta^{2-p}|\mathbb{T}_r|), \\ \qquad \forall r \in \mathbb{N}, \qquad \text{if } \alpha = \dfrac{1}{2}, \\ c_2 \exp(-c'\gamma^{2q}\delta^{2-p}|\mathbb{T}_r|), \\ \qquad \forall r > r_0, \qquad \text{if } \dfrac{1}{2} < \alpha < \dfrac{\sqrt{2}}{2}, \\ c_2 \exp\Big(-c'\gamma^q \delta^{2-p} \dfrac{|\mathbb{T}_r|}{r+1}\Big), \\ \qquad \forall r > r_0, \qquad \text{if } \alpha = \dfrac{\sqrt{2}}{2}, \\ c_2 \exp\Big(-c'\gamma^{2q}\delta^{2-p} \dfrac{1}{\alpha^{2(r+1)}}\Big), \\ \qquad \forall r > r_0, \qquad \text{if } \alpha > \dfrac{\sqrt{2}}{2}, \end{cases} \tag{4.10}$$

where c_2 is a positive constant, c' and c'' depend on α, and c and may differ line by line, c_0 depends on α, c and γ, and may differ line by line, $p \in \{0, 1, 3/2\}$ and $q \in \{0, 1\}$.

We have now to consider super exponential convergence of the estimators of the other parameters.

PROPOSITION 4.10. *Let (b_n) a sequence of real numbers satisfying Assumption 1. Then we have*

$$(\hat{\sigma}_r^2, \hat{\rho}_r) \underset{b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|}{\overset{\text{superexp}}{\longrightarrow}} (\sigma^2, \rho).$$

As previously we may now prove MDP for the estimator of θ.

PROPOSITION 4.11. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then $(|\mathbb{T}_r|(\hat{\theta}^r - \theta)/b_{|\mathbb{T}_r|})$ satisfies the MDP on $\mathbb{R}^4$ with the speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and rate function*

$$I(x) = \tfrac{1}{2} x^t (\Sigma')^{-1} x,$$

where

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix}$$

with

$$K = \frac{1}{\mu_2(\theta, \sigma^2) - \mu_1(\theta)^2} \begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta, \sigma^2) \end{pmatrix}.$$

REMARK 4.12. Notice that the proof of Proposition 4.11 does not need the cumulant method as in the proof of Proposition 4.6. Indeed, since we are in the bounded case, from MDP of martingale with bounded jumps (see [9]), we need only to prove the superexponential convergence of increasing process of the martingale. This convergence is easily obtained from Theorem 3.2.

Let us give us our last result by considering a MDP for the test statistic.

PROPOSITION 4.13. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then under the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$, $\frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}} (\chi_r^{(1)})^{1/2}$ satisfies a MDP on $\mathbb{R}$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function*

$$I'(y) = \begin{cases} \dfrac{y^2}{2}, & \text{if } y \in \mathbb{R}_+, \\ +\infty, & \text{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1 of H_0, we have for all $A > 0$,

$$\limsup_{r\to\infty} \frac{|\mathbb{T}_r|}{b^2_{|\mathbb{T}_r|}} \log \mathbb{P}(\chi_r^{(1)} < A) = -\infty.$$

APPENDIX A: PROOF OF THE EXPONENTIAL INEQUALITIES

This section is devoted to the proofs of Theorems 2.11, 2.12, 3.1, 3.2 and Proposition 4.2.

A.1. Proof of Theorem 2.11. Let $f \in F$ such that $(\mu, f) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. By the Markov inequality, we get, for all $\delta > 0$,

$$\mathbb{P}(|\overline{M}_{\mathbb{G}_r}(f)| > \delta) = \mathbb{P}(|\overline{M}_{\mathbb{G}_r}(f)|^2 > \delta^2) \le \frac{1}{\delta^2}\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^2].$$

By Guyon (see [14]), we have

$$\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^2] = \sum_{p=0}^{r} 2^{-p-\mathbf{1}_{p<r}} \nu Q^p P(Q^{r-p-1} f \otimes Q^{r-p-1} f).$$

Hypothesis (H1) implies that there exists $g \in F$ and $\alpha \in (0, 1)$ such that for all $p \in \{0, 1, \dots, r\}$,

$$\nu Q^p P(Q^{r-p-1} f \otimes Q^{r-p-1} f) \le \alpha^{2(r-p-1)} \nu Q^p P(g \otimes g).$$

Next, hypotheses (iii), (v) and (vi) imply that there is a positive constant c such that for all $p \in \{0, 1, \dots, r\}$,

$$\alpha^{2(r-p-1)} \nu Q^p P(g \otimes g) \le c\alpha^{2(r-p-1)}.$$

This leads us to

$$\begin{aligned} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^2] &\le c \sum_{p=0}^{r} 2^{-p-\mathbf{1}_{p<r}} \alpha^{2(r-p-1)} \\ &= \begin{cases} c\left(\dfrac{1}{2}\right)^r + c\dfrac{\alpha^{2r} - (1/2)^r}{2\alpha^2 - 1}, & \text{if } \alpha^2 \ne \dfrac{1}{2}, \\ cr\left(\dfrac{1}{2}\right)^r, & \text{if } \alpha^2 = \dfrac{1}{2}, \end{cases} \end{aligned} \tag{A.1}$$

and therefore (2.14) follows.

Part 2. Let us now consider $\overline{M}_n^{\Pi}(f)$. By the Markov inequality and the triangle inequality, we get, for all $\delta > 0$,

$$\begin{aligned} &\mathbb{P}(|\overline{M}_n^{\Pi}(f)| > \delta) \\ &\quad = \mathbb{P}(|\overline{M}_n^{\Pi}(f)|^2 > \delta^2) \le \frac{1}{\delta^2}\mathbb{E}[(\overline{M}_n^{\Pi}(f))^2] \\ &\quad \le \frac{2}{\delta^2}\mathbb{E}\left[\left(\sum_{q=0}^{r_n-1} \frac{2^q}{n}\overline{M}_{\mathbb{G}_q}(f)\right)^2\right] + \frac{2}{\delta^2}\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^2\right]. \end{aligned} \tag{A.2}$$

In the last inequality (A.2), we have used the decomposition

$$\overline{M}_n^{\Pi}(f) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n} \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}).$$

In what follows, the constant c may be slightly different from that of part 1 and may differ term by term. For the first term appearing in (A.2), we have

$$\mathbb{E}\left[\left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f)\right)^2\right] = \left\|\sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f)\right\|_2^2 \leq \left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \|\overline{M}_{\mathbb{G}_q}(f)\|_2\right)^2.$$

Using (A.1), we get that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{n} \|\overline{M}_{\mathbb{G}_q}(f)\|_2 \leq \begin{cases} \dfrac{c}{n} \displaystyle\sum_{q=0}^{r_n-1} (\sqrt{2})^q \leq c \frac{\sqrt{2}^{r_n}}{n}, & \text{if } \alpha^2 < \dfrac{1}{2}, \\ \dfrac{c}{n} \displaystyle\sum_{q=0}^{r_n} q^{1/2} \sqrt{2}^q \leq c \frac{r_n^{1/2} \sqrt{2}^{r_n}}{n}, & \text{if } \alpha^2 = \dfrac{1}{2}, \\ \dfrac{c}{n} \displaystyle\sum_{q=0}^{r_n-1} (2\alpha)^q \leq c \alpha^{r_n}, & \text{if } \alpha^2 > \dfrac{1}{2}, \end{cases}$$

which implies that

$$\text{(A.3)} \quad \mathbb{E}\left[\left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f)\right)^2\right] \leq \begin{cases} c \dfrac{2^{r_n}}{n^2} \leq c \left(\dfrac{1}{2}\right)^{r_n+1}, & \text{if } \alpha^2 < \dfrac{1}{2}, \\ c \dfrac{r_n}{2^{r_n+1}}, & \text{if } \alpha^2 = \dfrac{1}{2}, \\ c \alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}. \end{cases}$$

Now, we have to control the second term in (A.2). As in Guyon [14], we have that

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^2\right] \\ &\quad \leq \frac{n - 2^{r_n} + 1}{n^2} \nu Q^{r_n} f^2 \\ &\qquad + \frac{(n - 2^{r_n})(n - 2^{r_n} + 1)}{n^2(1 - 2^{-r_n})} \sum_{p=0}^{r_n-1} 2^{-p-1} \nu Q^p P\big(Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f\big) \\ &\quad \leq \frac{c}{n} + c \sum_{p=0}^{r_n-1} 2^{-p-1} \alpha^{2r_n-2p-2}. \end{aligned}$$

Discussing following the value of α, we obtain that

$$\text{(A.4)}\qquad \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^2\right] \le \begin{cases} c\dfrac{1}{2^{r_n+1}}, & \text{if } \alpha^2 < \dfrac{1}{2},\\ c\dfrac{r_n}{2^{r_n+1}}, & \text{if } \alpha^2 = \dfrac{1}{2},\\ c\alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}.\end{cases}$$

Inequality (2.15) then follows from (A.3) and (A.4).

Part 3. The case of $\overline{M}_{\mathbb{T}_r}(f)$ can be deduced from the previous by taking $n = |\mathbb{T}_r|$.

A.2. Proof of Theorem 2.12. Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. By the Markov inequality, we get for all $\delta > 0$,

$$\begin{aligned}\mathbb{P}(|\overline{M}_{\mathbb{G}_r}(f)| > \delta) &\le \frac{1}{\delta^2}\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^2]\\ &= \frac{1}{\delta^2}\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf))^2] + \frac{1}{\delta^2}\frac{1}{|\mathbb{G}_r|}\mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^2 - (Pf)^2)]\\ &\le \frac{1}{\delta^2}\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf))^2] + \frac{c}{\delta^2}\left(\frac{1}{2}\right)^r.\end{aligned}$$

The last inequality follows from the convergence of the sequence $(\mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^2 - (Pf)^2)])_r$ (see [14]).

Now, using part 1 of the proof of Theorem 2.11 with Pf instead of f leads us to a similar inequality (2.14) in Theorem 2.12 for $f \in \mathcal{B}(S^3)$.

Part 2. Let us now treat $\overline{M}_n^{\Pi}(f)$. Using the two equalities

$$\overline{M}_n^{\Pi}(f) = \sum_{q=0}^{r_n-1}\frac{|\mathbb{G}_q|}{n}\overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n}\sum_{i=2^{r_n}}^{n} f(\Delta_{\Pi(i)}),$$

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(\Delta_{\Pi(i)})\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} Pf(X_{\Pi(i)})\right)^2\right]\\ &\quad + \frac{1}{n}\mathbb{E}\left[\frac{1}{n}\sum_{i=2^{r_n}}^{n}(Pf^2 - (Pf)^2)(X_{\Pi(i)})\right],\end{aligned}$$

and part 2 of the proof of Theorem 2.11 with Pf instead of f leads us to a similar inequality (2.15) in Theorem 2.12 for $f \in \mathcal{B}$.

Part 3. The case of $\overline{M}_{\mathbb{T}_r}(f)$ can be deduced from the previous by taking $n = |\mathbb{T}_r|$.

A.3. Proof of Theorem 3.1. Let $f \in \mathcal{B}_b(S)$ such that $(\mu, f) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. We have for all $\lambda > 0$ and for all $\delta > 0$

$$\mathbb{P}\big(\overline{M}_{\mathbb{G}_r}(f) > \delta\big) \le \exp\big(-\lambda\delta|\mathbb{G}_r|\big)\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\Big)\Big]. \tag{A.5}$$

By subtracting and adding terms, we get

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\Big)\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\big) \\
&\qquad\qquad \times \prod_{i\in\mathbb{G}_{r-1}} \exp\big(2\lambda Qf(X_i)\big)/\mathcal{F}_{r-1}\Big]\Big].
\end{aligned}$$

Now using the fact that conditionally to the $(r-1)$ first generations the sequence $\{\Delta_i, i \in \mathbb{G}_{r-1}\}$ is a sequence of independent random variables, we have that

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\big) \\
&\qquad\qquad \times \prod_{i\in\mathbb{G}_{r-1}} \exp\big(2\lambda Qf(X_i)\big)/\mathcal{F}_{r-1}\Big]\Big] \\
&\quad = \mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(2\lambda Qf(X_i)\big) \\
&\qquad\quad \times \prod_{i\in\mathbb{G}_{r-1}} \mathbb{E}\big[\exp\big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\big)/\mathcal{F}_{r-1}\big]\Big].
\end{aligned}$$

Using the Azuma–Bennett–Hoeffding inequalities [1, 3, 16] (see Lemma B.1 for more detail), we get according to (H2), for all $i \in \mathbb{G}_{r-1}$,

$$\mathbb{E}\big[\exp\big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\big)/\mathcal{F}_{r-1}\big] \le \exp\big(2\lambda^2c^2(1+\alpha)^2\big).$$

This leads us to

$$\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\Big)\Big] \le \exp\big(\lambda^2c^2(1+\alpha)^2|\mathbb{G}_r|\big)\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(2\lambda Qf(X_i)\big)\Big].$$

Doing the same thing for $\mathbb{E}[\prod_{i\in\mathbb{G}_{r-1}}\exp(2\lambda Qf(X_i))]$ with Qf replacing f, we get

$$\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}}\exp\big(2\lambda Qf(X_i)\big)\Big]$$
$$\leq \exp\big(2\lambda^2c^2(\alpha+\alpha^2)^2|\mathbb{G}_r|\big)\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-2}}\exp\big(2^2\lambda Q^2f(X_i)\big)\Big].$$

Iterating this procedure, we get

$$\mathbb{E}\Big[\exp\Big(\lambda\sum_{i\in\mathbb{G}_r}f(X_i)\Big)\Big]\leq \mathbb{E}\big[\exp\big(2^r\lambda Q^rf(X_1)\big)\big]$$
$$\times\prod_{k=1}^{r}\exp\big(2^{k-1}\lambda^2c^2(\alpha^{k-1}+\alpha^k)^2|\mathbb{G}_r|\big).$$

Once again, according to (H2), we have

$$\mathbb{E}\Big[\exp\Big(\lambda\sum_{i\in\mathbb{G}_r}f(X_i)\Big)\Big]\leq \exp\big(\lambda c\alpha^r|\mathbb{G}_r|\big)\times\exp\Big(\lambda^2c^2(1+\alpha)^2|\mathbb{G}_r|\sum_{k=1}^{r}(2\alpha^2)^{k-1}\Big).$$

Hence:

- if $\alpha^2\neq\frac{1}{2}$, then

$$\mathbb{E}\Big[\exp\Big(\lambda\sum_{i\in\mathbb{G}_r}f(X_i)\Big)\Big]\leq \exp\Big(\lambda^2c^2(1+\alpha)^2\frac{1-(2\alpha^2)^r}{1-2\alpha^2}|\mathbb{G}_r|\Big)\times\exp\big(\lambda c\alpha^r|\mathbb{G}_r|\big);$$

- if $\alpha^2=\frac{1}{2}$, then

$$\mathbb{E}\Big[\exp\Big(\lambda\sum_{i\in\mathbb{G}_r}f(X_i)\Big)\Big]\leq \exp\big(\lambda^2c^2(1+\alpha)^2r|\mathbb{G}_r|\big)\times\exp\Big(\lambda c\Big(\frac{\sqrt{2}}{2}\Big)^r|\mathbb{G}_r|\Big).$$

We then consider three cases:

(a) If $\alpha^2<\frac{1}{2}$, then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2}<\frac{1}{1-2\alpha^2}$ for all r. Taking $\lambda=\frac{(1-2\alpha^2)\delta}{2c^2(1+\alpha)^2}$ in (A.5) leads us to

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f)>\delta)\leq\exp\Big(-\Big(\frac{(1-2\alpha^2)\delta^2}{4c^2(1+\alpha)^2}-\alpha^r\frac{(1-2\alpha^2)\delta}{2c(1+\alpha)^2}\Big)|\mathbb{G}_r|\Big).$$

- If $\alpha\leq\frac{1}{2}$, then $(2\alpha)^r\leq 1$ for all $r\in\mathbb{N}$. We then have for all $r\in\mathbb{N}$,

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f)>\delta)\leq\exp\Big(\frac{(1-2\alpha^2)\delta}{2c(1+\alpha)^2}\Big)\exp\Big(-\frac{(1-2\alpha^2)\delta^2|\mathbb{G}_r|}{4c^2(1+\alpha)^2}\Big).$$

- If $\frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}$, then for all $r \in \mathbb{N}$ such that $r > \log(\frac{\delta}{4c})/\log\alpha$, we have $(\delta - 2c\alpha^r) > \frac{\delta}{2}$, and it then follows that

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{(1-2\alpha^2)\delta^2|\mathbb{G}_r|}{8c^2(1+\alpha)^2}\right).$$

(b) If $\alpha^2 = \frac{1}{2}$, then for all $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) &\leq \exp((-\delta\lambda + c^2(1+\alpha)^2 r\lambda^2)|\mathbb{G}_r|)\\ &\quad\times \exp\left(\lambda c\left(\frac{\sqrt{2}}{2}\right)^r |\mathbb{G}_r|\right).\end{aligned}$$

Taking $\lambda = \frac{\delta}{2c^2(1+\alpha)^2 r}$, we are led to

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{\delta|\mathbb{G}_r|}{4c^2(1+\alpha)^2 r}\left(\delta - 2c\left(\frac{\sqrt{2}}{2}\right)^r\right)\right).$$

For all $r \in \mathbb{N}$ such that $r > \log(\frac{\delta}{4c})/\log(\frac{\sqrt{2}}{2})$, we have $(\delta - 2c(\frac{\sqrt{2}}{2})^r) > \frac{\delta}{2}$ and for such r, it follows that

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{\delta^2|\mathbb{G}_r|}{18c^2 r}\right).$$

(c) If $\alpha^2 > \frac{1}{2}$, then for all $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) &\leq \exp(-\lambda\delta|\mathbb{G}_r|) \times \exp\left(\lambda^2 c^2(1+\alpha)^2 \frac{(2\alpha^2)^r - 1}{2\alpha^2 - 1}|\mathbb{G}_r|\right)\\ &\quad\times \exp(\lambda c\alpha^r|\mathbb{G}_r|)\\ &\leq \exp\left(-|\mathbb{G}_r|\left(\lambda\delta - \frac{\lambda^2 c^2(1+\alpha)^2}{2\alpha^2 - 1}(2\alpha^2)^r\right)\right)\\ &\quad\times \exp(\lambda c\alpha^r|\mathbb{G}_r|).\end{aligned}$$

Taking $\lambda = \frac{(2\alpha^2-1)\delta}{2c^2(1+\alpha)^2(2\alpha^2)^r}$ leads us to

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{(2\alpha^2-1)\delta}{4c^2(1+\alpha)^2\alpha^{2r}}(\delta - 2c\alpha^r)\right).$$

Now for all $r \in \mathbb{N}$ such that $r > \log(\frac{\delta}{4c})/\log\alpha$, we have

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f)) \leq \exp\left(-\frac{(2\alpha^2-1)\delta^2}{8c^2(1+\alpha)^2\alpha^{2r}}\right).$$

Part 2. Let us now deal with $\overline{M}_{\mathbb{T}_r}(f)$. We have for all $\lambda > 0$ and all $\delta > 0$,

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp(-\lambda\delta|\mathbb{T}_r|)\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{T}_r} f(X_i)\Big)\Big]. \tag{A.6}$$

By subtracting and adding terms, we get

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{T}_r} f(X_i)\Big)\Big]\\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\big)\\
&\qquad\qquad \times \prod_{i\in\mathbb{G}_{r-1}} \exp\big(2\lambda Qf(X_i)\big) \times \prod_{i\in\mathbb{T}_{r-1}} \exp\big(\lambda f(X_i)\big)/\mathcal{F}_{r-1}\Big]\Big]\\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\big)\\
&\qquad\qquad \times \prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda(f + 2Qf)(X_i)\big) \times \prod_{i\in\mathbb{T}_{r-2}} \exp\big(\lambda f(X_i)\big)/\mathcal{F}_{r-1}\Big]\Big].
\end{aligned}$$

The fact that conditionally to the $(r-1)$ first generations the sequence $\{\Delta_i, i \in \mathbb{G}_{r-1}\}$ is a sequence of independent random variables and Azuma–Bennett–Hoeffding inequality (see Lemma B.1) lead us according to (H2) to

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{T}_r} f(X_i)\Big)\Big]\\
&\quad \leq \exp\big(2\lambda^2 c^2 (1+\alpha)^2 |\mathbb{G}_{r-1}|\big)\\
&\qquad \times \mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda(f + 2Qf)(X_i)\big) \prod_{i\in\mathbb{T}_{r-2}} \exp\big(\lambda f(X_i)\big)\Big].
\end{aligned}$$

Doing the same things for

$$\mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-1}} \exp\big(\lambda(f + 2Qf)(X_i)\big) \prod_{i\in\mathbb{T}_{r-2}} \exp\big(\lambda f(X_i)\big)\Big]$$

with $f + 2Qf$ replacing f, we get

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{T}_r} f(X_i)\Big)\Big]\\
&\quad \leq \exp\big(2\lambda^2 c^2 (1+\alpha)^2 |\mathbb{G}_{r-1}|\big) \times \exp\big(2\lambda^2 c^2 \big(1 + 3\alpha + 2\alpha^2\big)^2 |\mathbb{G}_{r-2}|\big)\\
&\qquad \times \mathbb{E}\Big[\prod_{i\in\mathbb{G}_{r-2}} \exp\big(\lambda\big(f + 2Qf + 2^2 Q^2 f\big)(X_i)\big) \prod_{i\in\mathbb{T}_{r-3}} \exp\big(\lambda f(X_i)\big)\Big].
\end{aligned}$$

Iterating this procedure leads us to

$$\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{T}_r} f(X_i)\Big)\Big]$$
$$\leq \exp\bigg(2\lambda^2 c^2(1+\alpha)^2 \sum_{q=1}^{r}\bigg(\sum_{k=0}^{q-1}(2\alpha)^k\bigg)^2 |\mathbb{G}_{r-q}|\bigg)$$
$$\times\, \mathbb{E}[\exp(\lambda(f + 2Qf + 2^2Q^2f + \cdots + 2^rQ^rf)(X_1))].$$

Using (H2) we get

$$\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{T}_r} f(X_i)\Big)\Big]$$
$$\leq \exp\bigg(\lambda c\sum_{k=0}^{r}(2\alpha)^k + 2\lambda^2c^2(1+\alpha)^2 \sum_{q=1}^{r}\bigg(\sum_{k=0}^{q-1}(2\alpha)^k\bigg)^2 |\mathbb{G}_{r-q}|\bigg).$$

Now for $\alpha \neq \frac{1}{2}$ and $\alpha^2 \neq \frac{1}{2}$ we have

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta)$$
$$\leq \exp(-\lambda\delta|\mathbb{T}_r|)\exp\bigg(2\lambda^2c^2(1+\alpha)^2\bigg(\frac{2^r-1}{(1-2\alpha)^2} - \frac{\alpha(1-\alpha^r)2^{r+1}}{(1-2\alpha)^2(1-\alpha)}$$
$$+ \frac{2\alpha^2(1-(2\alpha^2)^r)2^r}{(1-2\alpha)^2(1-2\alpha^2)}\bigg)\bigg)$$
$$\times \exp\bigg(\lambda c\frac{1-(2\alpha)^{r+1}}{1-2\alpha}\bigg)$$
$$\leq \exp\bigg(-|\mathbb{T}_r|\bigg(\lambda\delta - \frac{\lambda^2c^2(1+\alpha)^2}{(1-2\alpha)^2}\bigg(1 + \frac{4\alpha^2(1-(2\alpha^2)^r)}{1-2\alpha^2}\bigg)\bigg)\bigg)$$
$$\times \exp\bigg(\lambda c\frac{1-(2\alpha)^{r+1}}{1-2\alpha}\bigg).$$

Taking $\lambda = \frac{\delta}{(2c^2(1+\alpha)^2/(1-2\alpha)^2)(1+4\alpha^2(1-(2\alpha^2)^r)/(1-2\alpha^2))}$ leads us to

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta)$$
$$\leq \exp\bigg(-|\mathbb{T}_r|\frac{(1-2\alpha)^2\delta^2}{4c^2(1+\alpha)^2(1+4\alpha^2(1-(2\alpha^2)^r)/(1-2\alpha^2))}\bigg)$$
$$\times \exp\bigg(\frac{(1-2\alpha)^2\delta}{2c(1+\alpha)^2(1+4\alpha^2(1-(2\alpha^2)^r)/(1-2\alpha^2))}\frac{1-(2\alpha)^{r+1}}{1-2\alpha}\bigg).$$

- If $\alpha < \frac{1}{2}$, then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all $r \in \mathbb{N}$,

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\Big(\frac{1-2\alpha}{2c(1+\alpha)^2}\delta\Big) \times \exp\Big(-\frac{(1-2\alpha^2)(1-2\alpha)^2\delta^2}{4c^2(1+\alpha)^2(1+2\alpha^2)}|\mathbb{T}_r|\Big).$$

- If $\frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}$, then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all $r \in \mathbb{N}$,

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\Big(-\frac{(1-2\alpha^2)(2\alpha-1)^2\delta|\mathbb{T}_r|}{4c^2(1+\alpha)^2(1+2\alpha^2)}\Big(\delta - \frac{2c(1-2\alpha^2)\alpha^{r+1}}{(2\alpha-1)(1+2\alpha^2)}\Big)\Big).$$

Now for all $r \in \mathbb{N}$ such that $r+1 > \log(\frac{(2\alpha-1)(1+2\alpha^2)\delta}{4c(1-2\alpha^2)})/\log\alpha$, we have $\delta - \frac{2c(1-2\alpha^2)\alpha^{r+1}}{(2\alpha-1)(1+2\alpha^2)} > \frac{\delta}{2}$ so that for such r, we have

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\Big(-\frac{(1-2\alpha^2)(2\alpha-1)^2\delta^2|\mathbb{T}_r|}{8c^2(1+\alpha)^2(1+2\alpha^2)}\Big).$$

- If $\alpha^2 > \frac{1}{2}$, then for all $r \geq 1$, we have

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\Big(-\frac{(2\alpha-1)^2(2\alpha^2-1)\delta}{32c^2(1+\alpha)^2\alpha^{2(r+1)}}\Big(\delta - \frac{16\alpha^2 c\alpha^{r+1}}{(2\alpha^2-1)(2\alpha-1)}\Big)\Big).$$

For all $r \in \mathbb{N}^*$ such that $r+3 > \log(\frac{(2\alpha^2-1)(2\alpha-1)\delta}{32c})/\log\alpha$, we have $\delta - \frac{16\alpha^2 c\alpha^{r+1}}{(2\alpha^2-1)(2\alpha-1)} > \frac{\delta}{2}$ so that

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\Big(-\frac{(1-2\alpha)^2(2\alpha^2-1)\delta^2}{64c^2(1+\alpha)^2}\Big(\frac{1}{\alpha^2}\Big)^{r+1}\Big).$$

Now if $\alpha = \frac{1}{2}$, then $\sum_{q=1}^{r}\frac{q^2}{2^q} < \sum_{q=1}^{\infty}\frac{q^2}{2^q} = 6$. Then for all $\lambda > 0$,

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\big(-(\lambda\delta - 27c^2\lambda^2)|\mathbb{T}_r|\big) \times \exp\big(\lambda c(r+1)\big).$$

Taking $\lambda = \frac{\delta}{54c^2}$ leads us to

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\Big(-\frac{\delta^2}{108c^2}|\mathbb{T}_r|\Big) \times \exp\Big(\frac{\delta}{54c}(r+1)\Big).$$

Finally, if $\alpha^2 = \frac{1}{2}$, in the same way as previously, for all $r \in \mathbb{N}$ such that $r+1 > \log(\frac{(\sqrt{2}-1)\delta}{4c})/\log(\frac{\sqrt{2}}{2})$, we have

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_r}(f) > \delta\big) \leq \exp\bigg(-\frac{(\sqrt{2}-1)^2\delta^2}{4c^2(1+\sqrt{2})^2}\frac{|\mathbb{T}_r|}{r+1}\bigg).$$

Part 3. Eventually, let us look at $\overline{M}_n^{\Pi}(f)$. We have for all $\delta > 0$

$$\mathbb{P}\bigg(\frac{1}{n}M_n^{\Pi}(f) > \delta\bigg) \leq \mathbb{P}\bigg(\frac{1}{n}\sum_{i\in\mathbb{T}_{r_n-1}} f(X_i) > \frac{\delta}{2}\bigg) + \mathbb{P}\bigg(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}) > \frac{\delta}{2}\bigg).$$

On the one hand, (3.2) leads us to

$$\text{(A.7)}\quad \mathbb{P}\bigg(\frac{1}{n}\sum_{i\in\mathbb{T}_{r_n-1}} f(X_i) > \frac{\delta}{2}\bigg) \leq \begin{cases} \exp(c''\delta)\exp(-c'\delta^2 n), & \\ \qquad \forall n \in \mathbb{N}, & \text{if } \alpha < \dfrac{1}{2}, \\ \exp\big(2c'\delta(r_n+1)\big)\exp(-c'\delta^2 n), & \\ \qquad \forall n \in \mathbb{N}, & \text{if } \alpha = \dfrac{1}{2}, \\ \exp(-c'\delta^2 n), & \\ \qquad \forall r_n > r_0, & \text{if } \dfrac{1}{2} < \alpha < \dfrac{\sqrt{2}}{2}, \\ \exp\bigg(-c'\delta^2\dfrac{n}{r_n+1}\bigg), & \\ \qquad \forall r_n > r_0, & \text{if } \alpha = \dfrac{\sqrt{2}}{2}, \\ \exp\bigg(-c'\delta^2\dfrac{1}{\alpha^{2(r_n+1)}}\bigg), & \\ \qquad \forall r_n > r_0 - 2, & \text{if } \alpha > \dfrac{\sqrt{2}}{2}, \end{cases}$$

where $r_0 := \log(\frac{\delta}{c_0})/\log\alpha$ and c_0, c' and c'' are positive constants which depend on α, $\|f\|_\infty$ and c. c_0, $c'$and c'' differ line by line. On the other hand, for all $\lambda > 0$,

$$\mathbb{P}\bigg(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}) > \frac{\delta}{2}\bigg) \leq \exp\bigg(-\frac{\lambda\delta}{2}n\bigg)\mathbb{E}\bigg[\exp\bigg(\lambda\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\bigg)\bigg].$$

Now let:

- $\mathcal{O}_{r_n} = \{\Pi(2^{r_n}), \Pi(2^{r_n}+1), \ldots, \Pi(n)\}$;
- $\mathcal{O}_{r_n-1}^1$ the set of individuals of generation $\mathbb{G}_{r_n-1}$ which are ancestors of one individual in $\mathcal{O}_{r_n}$;
- $\mathcal{O}_{r_n-1}^2$ the set of individuals of generation $\mathbb{G}_{r_n-1}$ which are ancestors of two individuals in $\mathcal{O}_{r_n}$;

- $\mathcal{O}'_{r_n}$ the set of individuals of $\mathcal{O}_{r_n}$ whose parents belong to $\mathcal{O}^1_{r_n-1}$;
- $\mathcal{O}_{r_n-1} = \mathcal{O}^1_{r_n-1} \cup \mathcal{O}^2_{r_n-1}$.

We introduce the filtration $\tilde{\mathcal{F}}_r := \sigma(\mathcal{F}_r, \Pi(i), 1 \leq i \leq \mathbb{T})$. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)\right] \\
&\quad = \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} 2Qf(X_i) + \lambda \sum_{i \in \mathcal{O}^1_{r_n-1}} Qf(X_i)\right)\right. \\
&\qquad \times \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}'_{r_n}} f(X_i) - Qf(X_{[i/2]})\right) \Big/ \tilde{\mathcal{F}}_{r_n-1}\right] \\
&\qquad \left. \times \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\right) \Big/ \tilde{\mathcal{F}}_{r_n-1}\right]\right].
\end{aligned}$$

Using the Azuma–Bennett–Hoeffding inequality, as in part 1, we get

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}'_{r_n}} f(X_i) - Qf(X_{[i/2]})\right) \Big/ \tilde{\mathcal{F}}_{r_n-1}\right] \leq \exp\left(\frac{\lambda^2 c^2 (1+\alpha)^2}{2} |\mathcal{O}'_{r_n}|\right)$$

and

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\right) \Big/ \tilde{\mathcal{F}}_{r_n-1}\right] \\
&\qquad \leq \exp(2\lambda^2 c^2 (1+\alpha)^2 |\mathcal{O}^2_{r_n-1}|).
\end{aligned}$$

Now, we have

$$\begin{aligned}
&\exp\left(\frac{\lambda^2 c^2 (1+\alpha)^2}{2} |\mathcal{O}'_{r_n}|\right) + \exp(2\lambda^2 c^2 (1+\alpha)^2 |\mathcal{O}^2_{r_n-1}|) \\
&\qquad = \exp\left(\lambda^2 c^2 (1+\alpha)^2 \left(2|\mathcal{O}^2_{r_n-1}| + \frac{|\mathcal{O}'_{r_n}|}{2}\right)\right) \\
&\qquad \leq \exp(\lambda^2 c^2 (1+\alpha)^2 n).
\end{aligned}$$

This leads us to

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)\right] \\
&\qquad \leq \exp(\lambda^2 c^2 (1+\alpha)^2 n) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} 2Qf(X_i) + \lambda \sum_{i \in \mathcal{O}^1_{r_n-1}} Qf(X_i)\right)\right].
\end{aligned}$$

Now let:

- $\mathcal{O}^{1,1}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of one individual in $\mathcal{O}_{r_n-1}$ and one individual in $\mathcal{O}_{r_n}$;
- $\mathcal{O}^{1,2}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of one individual in $\mathcal{O}_{r_n-1}$ and two individuals in $\mathcal{O}_{r_n}$;
- $\mathcal{O}^{2,2}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of two individuals in $\mathcal{O}_{r_n-1}$ and two individuals in $\mathcal{O}_{r_n}$;
- $\mathcal{O}^{2,3}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of two individuals in $\mathcal{O}_{r_n-1}$ and three individuals in $\mathcal{O}_{r_n}$;
- $\mathcal{O}^{2,4}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of two individuals in $\mathcal{O}_{r_n-1}$ and four individuals in $\mathcal{O}_{r_n}$;
- $\mathcal{O}'_{r_n-1}$ the set of individuals of $\mathcal{O}_{r_n-1}$ whose parents belong to $\mathcal{O}^{1,1}_{r_n-2}$;
- $\mathcal{O}''_{r_n-1}$ the set of individuals of $\mathcal{O}_{r_n-1}$ whose parents belong to $\mathcal{O}^{1,2}_{r_n-2}$.

Then we have

$$\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathcal{O}^2_{r_n-1}} 2Qf(X_i) + \lambda \sum_{i\in\mathcal{O}^1_{r_n-1}} Qf(X_i)\Big)\Big] = \mathbb{E}[I_1 \times I_2 \times I_3 \times I_4 \times I_5 \times I_6 \times I_7],$$

where

$$I_1 = \exp\Big(\lambda \sum_{i\in\mathcal{O}^{1,1}_{r_n-2}} Q^2f(X_i) + \lambda \sum_{i\in\mathcal{O}^{1,2}_{r_n-2}} 2Q^2f(X_i) + \lambda \sum_{i\in\mathcal{O}^{2,2}_{r_n-2}} 2Q^2f(X_i) + \lambda \sum_{i\in\mathcal{O}^{2,3}_{r_n-2}} 3Q^2f(X_i) + \lambda \sum_{i\in\mathcal{O}^{2,4}_{r_n-2}} 4Q^2f(X_i)\Big),$$

$$I_2 = \mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathcal{O}'_{r_n-1}} Qf(X_i) - Q^2f(X_{[i/2]})\Big)\Big/\tilde{\mathcal{F}}_{r_n-2}\Big],$$

$$I_3 = \mathbb{E}\Big[\exp\Big(2\lambda \sum_{i\in\mathcal{O}''_{r_n-1}} Qf(X_i) - Q^2f(X_{[i/2]})\Big)\Big/\tilde{\mathcal{F}}_{r_n-2}\Big],$$

$$I_4 = \mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathcal{O}^{2,2}_{r_n-1}} Qf(X_{2i}) + Qf(X_{2i+1}) - 2Q^2f(X_i)\Big)\Big/\tilde{\mathcal{F}}_{r_n-2}\Big],$$

$$I_5 = \mathbb{E}\Big[\exp\Big(\frac{\lambda}{2} \sum_{i\in\mathcal{O}^{2,3}_{r_n-1}} 2Qf(X_{2i}) + Qf(X_{2i+1}) - 3Q^2f(X_i)\Big)\Big/\tilde{\mathcal{F}}_{r_n-2}\Big],$$

$$I_6 = \mathbb{E}\Big[\exp\Big(\frac{\lambda}{2} \sum_{i\in\mathcal{O}^{2,3}_{r_n-1}} Qf(X_{2i}) + 2Qf(X_{2i+1}) - 3Q^2f(X_i)\Big)\Big/\tilde{\mathcal{F}}_{r_n-2}\Big],$$

$$I_7 = \mathbb{E}\Bigg[\exp\Bigg(\lambda \sum_{i\in\mathcal{O}^{2,4}_{r_n-1}} 2Qf(X_{2i}) + 2Qf(X_{2i+1}) - 4Q^2 f(X_i)\Bigg)\Big/\tilde{\mathcal{F}}_{r_n-2}\Bigg].$$

Using the Azuma–Bennett–Hoeffding inequality, we get

$$\begin{aligned} I_2 \times I_3 \times I_4 \times I_5 \times I_6 \times I_7 &\\ &\leq \exp\Bigg(\lambda^2 c^2(\alpha+\alpha^2)^2\Bigg(\frac{|\mathcal{O}'_{r_n-1}|}{2} + 2|\mathcal{O}''_{r_n-1}| + 2|\mathcal{O}^{2,2}_{r_n-1}| \\ &\qquad + \frac{9|\mathcal{O}^{2,3}_{r_n-1}|}{2} + 8|\mathcal{O}^{2,4}_{r_n-1}|\Bigg)\Bigg) \\ &\leq \exp(2\lambda^2 c^2(\alpha+\alpha^2)^2 n), \end{aligned}$$

hence

$$\mathbb{E}\Bigg[\exp\Bigg(\lambda \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\Bigg)\Bigg] \leq \exp(\lambda^2 c^2(1+\alpha)^2 n)\exp(2\lambda^2 c^2(\alpha+\alpha^2)^2 n)\mathbb{E}[I_1].$$

Now, iterating this procedure we get

$$\mathbb{E}\Bigg[\exp\Bigg(\lambda \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\Bigg)\Bigg] \leq \exp\Bigg(\lambda^2 c^2(1+\alpha)^2 n \sum_{p=0}^{r_n}(2\alpha^2)^p\Bigg)\exp(\lambda c\alpha^{r_n} n).$$

Then it follows as in part 1 that

$$\mathbb{P}\Bigg(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}) > \frac{\delta}{2}\Bigg) \tag{A.8}$$
$$\leq \begin{cases} \exp(c''\delta)\exp(-c'\delta^2 n), & \\ \qquad \forall n\in\mathbb{N}, & \text{if } \alpha \leq \frac{1}{2}, \\ \exp(-c'\delta^2 n), & \\ \qquad \forall n\in\mathbb{N} \text{ such that } r_n > r_0, & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\Big(-c'\delta^2 \frac{n}{r_n}\Big), & \\ \qquad \forall n\in\mathbb{N} \text{ such that } r_n > r_0, & \text{if } \alpha^2 = \frac{1}{2}, \\ \exp\Big(-c'\delta^2\Big(\frac{1}{\alpha}\Big)^{2r_n}\Big), & \\ \qquad \forall n\in\mathbb{N} \text{ such that } r_n > r_0, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

where $r_0 := \log(\frac{\delta}{c_0})/\log(\alpha)$ and the positive constants c_0, c' and c'' depend on α, δ, c and differ line to line. Finally (A.7) and (A.8) lead us to (3.3).

A.4. Proof of Theorem 3.2. Let $f \in \mathcal{B}_b(S^3)$ such that $(\mu, Pf) = 0$.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. We have for all $\delta > 0$ and $\lambda > 0$,

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp(-\lambda\delta|\mathbb{G}_r|)\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{G}_r} f(\Delta_i)\Big)\Big].$$

Conditioning and using Bennett–Hoeffding inequality gives us

$$\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{G}_r} f(\Delta_i)\Big)\Big] \leq \exp(2\lambda^2\|f\|_\infty|\mathbb{G}_r|)\mathbb{E}\Big[\exp\Big(\lambda \sum_{i\in\mathbb{G}_r} Pf(X_i)\Big)\Big].$$

Now, applying part 1 of the proof of the Theorem 3.1 to Pf, we get (3.1) for $f \in \mathcal{B}_b(S^3)$.

Part 2. Let us now treat $\overline{M}_{\mathbb{T}_r}(f)$. We have for all $\delta > 0$,

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f - Pf) > \frac{\delta}{2}\Big) + \mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(Pf) > \frac{\delta}{2}\Big). \tag{A.9}$$

Now, since $(M_n^\Pi(f - Pf))_{n\geq 1}$ is a $\mathcal{H}_n$-martingale with bounded jumps, the Azuma inequality [1] gives us for some positive constant c',

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f - Pf) > \frac{\delta}{2}\Big) \leq \exp(-c'\delta^2|\mathbb{T}_r|).$$

For the second term on the right-hand side of (A.9), we use inequalities (3.2) with Pf instead of f. Gathering these inequalities, we get (3.2) for all r large enough.

Part 3. The proof for the case $\overline{M}_n^\Pi(f)$ follows the same lines as the proof of part 2.

A.5. Proof of Proposition 4.2. We will prove the deviation inequality for $|\hat{\alpha}_0^r - \alpha_0|$. The other deviation inequalities for $|\hat{\beta}_0^r - \beta_0|$, $|\hat{\alpha}_1^r - \alpha_1|$ and $|\hat{\beta}_1^r - \beta_1|$ may be treated in a similar way.

One easily checks that

$$\hat{\alpha}_0^r - \alpha_0 = \frac{(\overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(P(\mathbf{xy}))) - (\overline{M}_{\mathbb{T}_r}(\mathbf{x}))(\overline{M}_{\mathbb{T}_r}(\mathbf{y}) - \overline{M}_{\mathbb{T}_r}(P(\mathbf{y})))}{B_r}.$$

We then have, for all $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}(|\hat{\alpha}_0^r - \alpha_0| > \delta) \\
&\quad\leq \mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))|}{B_r} > \frac{\delta}{2}\Big) \\
&\qquad + \mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x})||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{2}\Big).
\end{aligned}$$

On one hand, for all $\gamma_1 > 0$ we have

$$\mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))|}{B_r} > \frac{\delta}{2}\Big) \leq \mathbb{P}(B_r < \gamma_1) + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))| > \frac{\delta\gamma_1}{2}\Big). \tag{A.10}$$

Now, for $b = \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2$, where μ_1 and μ_2 are given in (4.5), we have

$$\begin{aligned}\mathbb{P}(B_r < \gamma_1) \leq \mathbb{P}\Big(-\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2 - \mu_2) > \frac{b-\gamma_1}{3}\Big) &\\ + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\sqrt{b-\gamma_1}}{\sqrt{3}}\Big) &\\ + \mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1) > \frac{b-\gamma_1}{6|\mu_1|}\Big). &\end{aligned}$$

We choose $\gamma_1 < \min\{\frac{2b}{2+3\delta}, \frac{-4+\sqrt{48b\delta^2+16}}{6\delta^2}, \frac{b}{1+3\delta|\mu_1|}\}$ so that $\frac{\delta\gamma_1}{2} < \max\{\frac{b-\gamma_1}{3}, \frac{\sqrt{b-\gamma_1}}{\sqrt{3}}, \frac{b-\gamma_1}{6|\mu_1|}\}$. Then we have

$$\mathbb{P}(B_r < \gamma_1) \leq \mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(\mu_2 - \mathbf{x}^2) > \frac{\delta\gamma_1}{2}\Big) + 2\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\delta\gamma_1}{2}\Big),$$

and therefore we get

$$\begin{aligned}&\mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))|}{B_r} > \frac{\delta}{2}\Big)\\ &\quad \leq 2\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\delta\gamma_1}{2}\Big) + \mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(\mu_2 - \mathbf{x}^2) > \frac{\delta\gamma_1}{2}\Big)\\ &\qquad + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))| > \frac{\delta\gamma_1}{2}\Big).\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x})||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{2}\Big) &\leq \mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{4}\Big)\\ &\quad + \mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{4|\mu_1|}\Big).\end{aligned}$$

The last term of the previous inequality can be dealt with in the same way as inequality (A.10), using $\gamma_3 > 0$ such that

$$\gamma_3 < \min\Big\{\frac{4b|\mu_1|}{4|\mu_1| + 3\delta}, \frac{2|\mu_1|(-4 + \sqrt{24b\delta^2/|\mu_1| + 16})}{3\delta^2}, \frac{2b}{2+3\delta}\Big\}.$$

For the second term, we have

$$\mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)||\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))|}{B_r} > \frac{\delta}{4}\Big)$$
$$\leq \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\sqrt{\delta}}{2}\Big) + \mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))|}{B_r} > \frac{\sqrt{\delta}}{2}\Big).$$

Let $\gamma_2 > 0$ such that $\gamma_2 < \min\{\frac{2b}{2+3\sqrt{\delta}}, \frac{-4+\sqrt{48b\delta+16}}{b\delta}, \frac{b}{1+3\sqrt{\delta}|\mu_1|}\}$, in such a way that we obtain $\frac{\gamma_2\sqrt{\delta}}{2} < \max\{\frac{b-\gamma_2}{3}, \frac{\sqrt{b-\gamma_2}}{\sqrt{3}}, \frac{b-\gamma_2}{6|\mu_1|}\}$. We thus have

$$\mathbb{P}\Big(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)||\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))|}{B_r} > \frac{\delta}{4}\Big)$$
$$\leq \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\sqrt{\delta}}{2}\Big)$$
$$+ \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)| > \frac{\gamma_2\sqrt{\delta}}{2}\Big) + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))| > \frac{\gamma_2\sqrt{\delta}}{2}\Big)$$
$$+ 2\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\gamma_2\sqrt{\delta}}{2}\Big).$$

From the foregoing, we deduce that for all $\gamma > 0$ such that $\gamma < \min(\gamma_1, \gamma_2, \gamma_3)$,

$$\mathbb{P}(|\hat{\alpha}_0^{(r)} - \alpha_0| > \delta)$$
$$\leq 2\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\delta\gamma}{2}\Big) + \mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(\mu_2-\mathbf{x}^2) > \frac{\delta\gamma}{2}\Big)$$
$$+ \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))| > \frac{\delta\gamma}{2}\Big) + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\sqrt{\delta}}{2}\Big)$$
$$+ \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)| > \frac{\gamma\sqrt{\delta}}{2}\Big) + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))| > \frac{\gamma\sqrt{\delta}}{2}\Big)$$
$$+ 2\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\gamma\sqrt{\delta}}{2}\Big) + 2\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)| > \frac{\delta\gamma}{4|\mu_1|}\Big)$$
$$+ \mathbb{P}\Big(\Big|\overline{M}_{\mathbb{T}_r}(\mu_2-\mathbf{x}^2) > \frac{\delta\gamma}{4|\mu_1|}\Big|\Big) + \mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))| > \frac{\delta\gamma}{4|\mu_1|}\Big).$$

Now, using (2.8) and Markov's inequality we get

$$\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))| > \frac{\delta\gamma}{2}\Big) \leq \frac{c}{\delta^4\gamma^4}\Big(\frac{1}{4}\Big)^{r+1},$$
$$\mathbb{P}\Big(|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))| > \frac{\delta\gamma}{4|\mu_1|}\Big) \leq \frac{c\mu_1^4}{\delta^4\gamma^4}\Big(\frac{1}{4}\Big)^{r+1}$$

and

$$\mathbb{P}\Big(\big|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\big| > \frac{\gamma\sqrt{\delta}}{2}\Big) \le \frac{c}{\delta^2\gamma^4}\Big(\frac{1}{4}\Big)^{r+1},$$

where the constant c can be found as in Remark 2.4.

Finally, the other terms, that is, the terms related to $\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)$ and $\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)$, can be bounded as in Corollary 2.2 and this completes the proof.

APPENDIX B

Let us gather here, for the convenience of the readers, various theorems useful to establish LIL, ASFCLT, deviation inequalities and MDP.

First, let us enunciate the Azuma–Bennett–Hoeffding inequality [1, 3, 16].

LEMMA B.1. *Let X be a real-valued and centered random variable such that $a \le X \le b$ a.s., with $a < b$. Then for all $\lambda > 0$, we have*

$$\mathbb{E}\big[\exp(\lambda X)\big] \le \exp\Big(\frac{\lambda^2(b-a)^2}{8}\Big).$$

LEMMA B.2. *Let (E,d) a metric space. Let (Z_n) a sequence of random variables values in E, (v_n) a rate and $g:\mathcal{D}_E \subset E \to \mathbb{R}$ continuous. Let $z \in E$ be a deterministic value*:

$$\textit{If } Z_n \overset{\text{superexp}}{\underset{v_n}{\Longrightarrow}} z \qquad \textit{then } g(Z_n) \overset{\text{superexp}}{\underset{v_n}{\Longrightarrow}} g(z).$$

PROOF. For all $\delta > 0$, there exists (see, e.g., [22], proof of Theorem 2.3) $\alpha_0(\delta) > 0$

$$\mathbb{P}\big(\big|g(Z_n)-g(z)\big| > \delta\big) \le \mathbb{P}\big(d(Z_n,z) > \alpha_0(\delta)\big). \tag{B.1}$$

Indeed, since g is continuous, for all $\delta > 0$, there exists $\alpha_0(\delta) > 0$ such that

$$\big|g(x)-g(z)\big| \le \delta \qquad \text{whenever } d(x,z) \le \alpha_0(\delta).$$

We then have

$$\big\{\omega : d\big(Z_n(\omega),z\big) \le \alpha_0(\delta)\big\} \subset \big\{\omega : \big|g\big(Z_n(\omega)\big)-g(z)\big| \le \delta\big\}$$

and therefore inequality (B.1). Now, the result of the lemma follows since $Z_n \overset{\text{superexp}}{\underset{v_n}{\Longrightarrow}} z$. $\square$

Let $M = (M_n, \mathcal{H}_n, n \ge 0)$ be a centered square integrable martingale defined on a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and $(\langle M\rangle_n)$ its bracket. We recall some limit theorems for martingale used intensively in this paper.

We recall the following result due to W. F. Stout (Theorem 3 in [21]).

THEOREM B.3. *Let (M_n) such that $M_0 = 0$. If $\langle M\rangle_n \to \infty$ a.s. and*

$$\sum_{n=1}^{\infty} \frac{2\log\log\langle M\rangle_n}{K_n^2\langle M\rangle_n}\mathbb{E}\big[(M_n - M_{n-1})^2\mathbf{1}_{\{(M_n-M_{n-1})^2>K_n^2\langle M_n\rangle/(2\log\log\langle M\rangle_n)\}}/\mathcal{H}_{n-1}\big] < \infty \qquad a.s.,$$

where K_n are $\mathcal{H}_{n-1}$ measurable and $K_n \to 0$ a.s., then $\limsup \frac{M_n}{\sqrt{2\langle M\rangle_n \log\log\langle M\rangle_n}} = 1$ *a.s.*

We recall the following result due to Chaabane (Corollary 2.2 in [5]).

THEOREM B.4. *Let (V_n) be a $(\mathcal{H}_n)$-predictable increasing process such that*:

H-1 $V_n^{-2}\langle M\rangle_n \underset{n\to\infty}{\longrightarrow} 1$, *a.s.*;

H-2 *for all* $\varepsilon > 0$, $\sum_{n\geq 1} V_n^{-2}\mathbb{E}[(M_n - M_{n-1})^2\mathbf{1}_{|M_n-M_{n-1}|>\varepsilon V_n}/\mathcal{H}_{n-1}] < \infty$, *a.s.*;

H-3 *for some* $a > 1$, $\sum_{n\geq 1} V_n^{-2a}\mathbb{E}[(M_n - M_{n-1})^{2a}\mathbf{1}_{|M_n-M_{n-1}|\leq V_n}/\mathcal{H}_{n-1}] < \infty$, *a.s.*

Then M_n satisfies an ASFCLT; *that is*, *for almost all ω, the weighted random measures*

$$W_N(\omega, \bullet) = \big(\log V_N^2\big)^{-1}\sum_{n=1}^{N}\left(1 - \frac{V_n^2}{V_{n+1}^2}\right)\delta_{\{\psi_n(\omega)\in\bullet\}}$$

associated to the continuous processes $\Psi_n(\omega) = \{\Psi_n(\omega, t), 0 \leq t \leq 1\}$ defined by

$$\Psi_n(\omega, t) = V_n^{-1}\big\{M_k + \big(V_{k+1}^2 - V_k^2\big)^{-1}\big(tV_n^2 - V_k^2\big)(M_{k+1} - M_k)\big\},$$

when $V_k^2 \leq tV_n^2 < V_{k+1}^2$, $0 \leq k \leq n-1$, weakly converge to the Wiener measure on $\mathcal{C}([0, 1], \mathbb{R})$.

Let us enunciate the following which corresponds to the unidimensional case of Theorem 1 in [11].

PROPOSITION B.5. *Let (b_n) a sequence satisfying*

$$b_n \text{ is increasing}, \qquad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \qquad \frac{b_n}{n} \longrightarrow 0,$$

such that $c(n) := n/b_n$ is nondecreasing, *and define the reciprocal function $c^{-1}(t)$ by*

$$c^{-1}(t) := \inf\{n \in \mathbb{N} : c(n) \geq t\}.$$

Under the following conditions:

(C1) *there exists* $Q \in \mathbb{R}_+^*$ *such that* $\frac{\langle M \rangle_n}{n} \overset{\text{superexp}}{\underset{b_n^2/n}{\longrightarrow}} Q$;

(C2) $\limsup_{n\to+\infty} \frac{n}{b_n^2} \log(n \operatorname{ess\,sup}_{1\le k\le c^{-1}(b_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > b_n/\mathcal{H}_{k-1})) = -\infty$;

(C3) *for all* $a > 0$ $\frac{1}{n}\sum_{k=1}^n \mathbb{E}(|M_k - M_{k-1}|^2 \mathbf{1}_{\{|M_k - M_{k-1}|\ge an/b_n\}}/\mathcal{H}_{k-1}) \overset{\text{superexp}}{\underset{b_n^2/n}{\longrightarrow}} 0$;

$(M_n/b_n)_{n\in\mathbb{N}}$ *satisfies the MDP in* $\mathbb{R}$ *with the speed* b_n^2/n *and the rate function* $I(x) = \frac{x^2}{2Q}$.

Acknowledgments. Let us thank two anonymous referees for their very careful reading and useful suggestions, which have clearly improved both presentation and mathematical rigor of the present paper.

REFERENCES

[1] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tôhoku Math. J.* (2) **19** 357–367. MR0221571

[2] BASAWA, I. V. and ZHOU, J. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* **41A** 55–64. MR2057565

[3] BENNETT, G. (1962). Probability inequalities for sum of independant random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[4] BERCU, B., DE SAPORTA, B. and GÉGOUT-PETIT, A. (2009). Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* **14** 2492–2526. MR2563249

[5] CHAABANE, F. (1996). Version forte du théorème de la limite centrale fonctionnel pour les martingales. *C. R. Acad. Sci. Paris Sér. I Math.* **323** 195–198. MR1402542

[6] COWAN, R. and STAUDTE, R. G. (1986). The bifurcating autoregressive model in cell lineage studies. *Biometrics* **42** 769–783.

[7] DE SAPORTA, B., GÉGOUT-PETIT, A. and MARSALLE, L. (2011). Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* **5** 1313–1353. MR2842907

[8] DELMAS, J.-F. and MARSALLE, L. (2010). Detection of cellular aging in a Galton–Watson process. *Stochastic Process. Appl.* **120** 2495–2519. MR2728175

[9] DEMBO, A. (1996). Moderate deviations for martingales with bounded jumps. *Electron. Commun. Probab.* **1** 11–17 (electronic). MR1386290

[10] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* **38**. Springer, New York. MR1619036

[11] DJELLOUT, H. (2002). Moderate deviations for martingale differences and applications to ϕ-mixing sequences. *Stoch. Stoch. Rep.* **73** 37–63. MR1914978

[12] DJELLOUT, H., GUILLIN, A. and WU, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** 2702–2732. MR2078555

[13] GAO, F. and ZHAO, X. (2011). Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.* **39** 1211–1240. MR2816352

[14] GUYON, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** 1538–1569. MR2358633

[15] GUYON, J., BIZE, A., PAUL, G., STEWART, E., DELMAS, J.-F. and TADDÉI, F. (2005). Statistical study of cellular aging. In *CEMRACS* 2004—*Mathematics and Applications to Biology and Medicine*. *ESAIM Proceedings* **14** 100–114 (electronic). EDP Sci., Les Ulis. MR2226805

[16] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[17] HUGGINS, R. M. and BASAWA, I. V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36** 1225–1233. MR1746406

[18] HUGGINS, R. M. and BASAWA, I. V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42** 423–432. MR1802966

[19] HWANG, S. Y., BASAWA, I. V. and YEO, I. K. (2009). Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Stat. Methodol.* **6** 61–69. MR2655539

[20] STEWART, E. J., MADDEN, R., PAUL, G. and TADDÉI, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* **3** e45.

[21] STOUT, W. F. (1970). A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **15** 279–290. MR0293701

[22] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[23] WORMS, J. (1999). Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* **4** 28 pp. (electronic). MR1684149

[24] WORMS, J. (2001). Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.* **10** 38–72. MR1841808

[25] ZHOU, J. and BASAWA, I. V. (2005). Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74** 77–88. MR2189078

[26] ZHOU, J. and BASAWA, I. V. (2005). Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Series Anal.* **26** 825–842. MR2203513

S. V. BITSEKI PENDA
H. DJELLOUT
LABORATOIRE DE MATHÉMATIQUES
UNIVERSITÉ BLAISE PASCAL
CNRS UMR 6620
24 AVENUE DES LANDAIS
BP 80026, 63177 AUBIÈRE
FRANCE
E-MAIL: Valere.Bitsekipenda@math.univ-bpclermont.fr
Hacene.Djellout@math.univ-bpclermont.fr

A. GUILLIN
INSTITUT UNIVERSITAIRE DE FRANCE
ET LABORATOIRE DE MATHÉMATIQUES
UNIVERSITÉ BLAISE PASCAL
CNRS UMR 6620
24 AVENUE DES LANDAIS
BP 80026, 63177 AUBIÈRE
FRANCE
E-MAIL: Arnaud.Guillin@math.univ-bpclermont.fr

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 3, 806–844
DOI: 10.1214/13-AIHP545

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models

S. Valère Bitseki Penda and Hacène Djellout

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, 24 Avenue des Landais, BP80026, 63177 Aubière, France.
E-mail: Valere.Bitsekipenda@math.univ-bpclermont.fr; Hacene.Djellout@math.univ-bpclermont.fr

Received 22 August 2012; revised 10 January 2013; accepted 31 January 2013

Abstract. The purpose of this paper is to investigate the deviation inequalities and the moderate deviation principle of the least squares estimators of the unknown parameters of general pth-order asymmetric bifurcating autoregressive processes, under suitable assumptions on the driven noise of the process. Our investigation relies on the moderate deviation principle for martingales.

Résumé. L'objetcif de ce papier est d'établir des inégalités de déviations et les principes de déviations modérées pour les estimateurs des moindres carrés des paramètres inconnus d'un processus bifurcant autorégressif asymétrique d'ordre p, sous certaines conditions sur la suite des bruits. Les preuves reposent sur les principes de déviations modérées des martingales.

MSC: 60F10; 62F12; 60G42; 62M10; 62G05

Keywords: Deviation inequalities; Moderate deviation principle; Bifurcating autoregressive process; Martingale; Limit theorems; Least squares estimation

1. Motivation and context

Bifurcating autoregressive processes (BAR, for short) are an adaptation of autoregressive processes, when the data has a binary tree structure. They were first introduced by Cowan and Staudte [6] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation.

In their paper, the original BAR process is defined as follows. The initial cell is labelled 1, and the two offspring of cell k are labelled $2k$ and $2k+1$. If X_k denotes an observation of some characteristic of individual k then the first order BAR process is given, for all $k \geq 1$, by

$$\begin{cases} X_{2k} = a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} = a + bX_k + \varepsilon_{2k+1}. \end{cases}$$

The noise sequence $(\varepsilon_{2k}, \varepsilon_{2k+1})$ represents environmental effects, while numbers a and b are unknown real parameters, with $|b| < 1$, related to inherited effects. The driven noise $(\varepsilon_{2k}, \varepsilon_{2k+1})$ was originally supposed to be independent and identically distributed with normal distribution. However, since two sister cells are in the same environment at their birth, ε_{2k} and ε_{2k+1} could be correlated, inducing a correlation between sister cells, distinct from the correlation inherited from their mother.

Several extensions of the model have been proposed and various estimators for the unknown parameters have been studied in the literature, see for instance [2,19–21,28,29]. See [3] for relevant references (although [3] deals with the asymmetric case unlike the above cited papers).

Recently, there have been many studies of the asymmetric BAR process, considering cases where the quantitative characteristics of the even and odd sisters are allowed to depend on their mother's through different sets of parameters.

In [18], Guyon proposes an interpretation of the asymmetric BAR process as a bifurcating Markov chain. This enables him to derive laws of large numbers and central limit theorems for the least squares estimators of the unknown parameters of the process. This Markov chain approach was further developed by Delmas and Marsalle [10], for cells which are allowed to die. They defined the genealogy of the cells through a Galton–Watson process, studying the same model on the Galton–Watson tree instead of a binary tree.

Another approach based on martingales theory was proposed by Bercu, de Saporta and Gégout-Petit [3], to sharpen the asymptotic analysis of Guyon, under weaker assumptions. It should be pointed out that missing data is not dealt with in this work. To take it into account in the estimation procedure, de Saporta et al. [8] and [9] use a two-type Galton–Watson process to model the genealogy.

Our objective in this paper is to go a step further by

- studying the moderate deviation principle (MDP, for short) of the least squares estimators of the unknown parameters of general asymmetric pth-order bifurcating autoregressive processes (BAR(p), for short). More precisely we are interested in the asymptotic estimations of

  $$\mathbb{P}\left(\frac{\sqrt{n}}{v_n}(\Theta_n - \Theta) \in A\right),$$

  where Θ_n denotes the estimator of the unknown parameter of interest Θ, A is a given domain of deviation, $(v_n > 0)$ is some sequence denoting the scale of deviation. When $v_n = 1$ this is exactly the estimation of the central limit theorem. When $v_n = \sqrt{n}$, it becomes the *large deviation*. And when $1 \ll v_n \ll \sqrt{n}$, this is the so called *moderate deviations*. Usually, MDP has a simpler rate function inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the large deviation principle.

  To prove our result on MDP, we use

  (1) the work of Bercu et al. [3] on the almost sure convergence of the estimators with the quadratic strong law and the central limit theorem;
  (2) the work of Dembo [11], and Worms [26,27] on the one hand, and the papers of Puhalskii [24] and Djellout [13] on the other hand, on the MDP for martingales.

- giving deviation inequalities for the estimator of bifurcating autoregressive processes, which are important for a rigorous nonasymptotic statistical study. We aim at obtaining estimates such as

  $$\forall x > 0 \quad \mathbb{P}\big(\|\Theta_n - \Theta\| \geq x\big) \leq \mathrm{e}^{-C_n(x)},$$

  where $C_n(x)$ will crucially depend on our set of assumptions. The upper bound in this inequality hold for arbitrary n and x (not a limit relation, unlike the MDP results), hence they are of much more practical use (in statistics). Deviation inequalities for estimators of the parameters associated with linear regression, autoregressive and branching processes were investigated by Bercu and Touati [4]. In the martingale case, deviation inequalities for a self normalized martingale have been developed by de la Peña et al. [7]. We also refer to the work of Ledoux [22] for precise credit and references. This type of inequalities is motivated by theoretical questions as well as numerous applications in different fields including the analysis of algorithms, mathematical physics and empirical processes. For some applications in nonasymptotic model selection problems we refer to Massart [23].

Let us emphasize that to our knowledge, there are no existing studies of the above questions, that is of the MDP and deviation inequalities for the least squares estimators of the unknown parameters of the general asymmetric BAR(p) process. These questions have been adressed recently by Bitseki Penda et al. [5], but for the BAR(1) processes. Moreover, in the latter, the authors have obtained their results under stronger assumptions than those made in this paper.

The main aspect of our contribution is that our results highlight the competition between the binary division and the speed of convergence in the MDP. Our MDP holds following three regimes, depending on the value of the ergodicity parameter of the BAR(p) compared with $1/2$. This new phenomenon is not seen in the case of the previously proved limit theorems: central limit theorem and law of large numbers. However, a similar phenomenon occurs for the central limit theorem of a branching particle system: see [1].

This paper is organized as follows. First of all, in Section 2, we introduce the BAR(p) model as well as the least squares estimators for the parameters of the observed BAR(p) process and some related notation and hypotheses. In Section 3, we state our main results on the deviation inequalities and MDP for our estimators. Section 4 is dedicated to the superexponential convergence of the quadratic variation of the martingale; this section contains exponential inequalities which are crucial for the proof of the deviation inequalities. The main results are proved in Section 5.

2. Notation and hypotheses

In all the sequel, let $p \in \mathbb{N}^*$. We consider the asymmetric BAR(p) process given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} X_{2n} = a_0 + \sum_{k=1}^{p} a_k X_{[n/2^{k-1}]} + \varepsilon_{2n}, \\ X_{2n+1} = b_0 + \sum_{k=1}^{p} b_k X_{[n/2^{k-1}]} + \varepsilon_{2n+1}, \end{cases}$$

where the notation $[x]$ stands for the largest integer less than or equal to the real number x. The initial states $\{X_k, 1 \leq k \leq 2^{p-1} - 1\}$ are the ancestors while $(\varepsilon_{2n}, \varepsilon_{2n+1})$ is the driven noise of the process. The parameters $(a_0, a_1, \ldots, a_p)$ and $(b_0, b_1, \ldots, b_p)$ are unknown real vectors.

For any matrix M the notation M^t, $\|M\|$ and $\mathrm{Tr}(M)$ stand for the transpose, the Euclidean norm and the trace of M respectively.

The BAR(p) process can be rewritten in the abbreviated vector form given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} \mathbb{X}_{2n} = A\mathbb{X}_n + \eta_{2n}, \\ \mathbb{X}_{2n+1} = B\mathbb{X}_n + \eta_{2n+1}, \end{cases} \tag{2.1}$$

where $\mathbb{X}_n = (X_n, X_{[n/2]}, \ldots, X_{[n/2^{p-1}]})^t$ is the regression vector, $\eta_{2n} = (a_0 + \varepsilon_{2n})e_1$ and $\eta_{2n+1} = (b_0 + \varepsilon_{2n+1})e_1$, with $e_1 = (1, 0, \ldots, 0)^t \in \mathbb{R}^p$. Moreover, A and B are the $p \times p$ companion matrices

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & 1 & \cdot \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & 1 & \cdot \end{pmatrix}.$$

We shall assume that the matrices A and B satisfy the contraction property

$$\beta = \max\bigl(\|A\|, \|B\|\bigr) < 1. \tag{2.2}$$

One can view this BAR(p) process as a pth-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor. For all $n \geq 1$, denote the nth generation by $\mathbb{G}_n = \{2^n, 2^n + 1, \ldots, 2^{n+1} - 1\}$, see Figure 1.

In particular, $\mathbb{G}_0 = \{1\}$ is the initial generation and $\mathbb{G}_1 = \{2, 3\}$ is the first generation of offspring from the first ancestor. Let $\mathbb{G}_{r_n}$ be the generation of individual n, which means that $r_n = [\log_2(n)]$. Recall that the two offspring of individual n are labelled $2n$ and $2n + 1$, or conversely, the mother of the individual n is $[n/2]$. More generally, the ancestors of individual n are $[n/2], [n/2^2], \ldots, [n/2^{r_n}]$. Furthermore, denote by

$$\mathbb{T}_n = \bigcup_{k=0}^{n} \mathbb{G}_k$$

the subtree of all individuals from the original individual up to the nth generation. We denote by $\mathbb{T}_{n,p} = \{k \in \mathbb{T}_n, k \geq 2^p\}$ the subtree of all individuals between the pth and the nth generation ($\mathbb{T}_{p-1}$ removed). One can observe that, for all $n \geq 1$, $\mathbb{T}_{n,0} = \mathbb{T}_n$ and for all $p \geq 1$, $\mathbb{T}_{p,p} = \mathbb{G}_p$.

The BAR(p) process can be rewritten, for all $n \geq 2^{p-1}$, in the matrix form

$$Z_n = \theta^t Y_n + V_n,$$

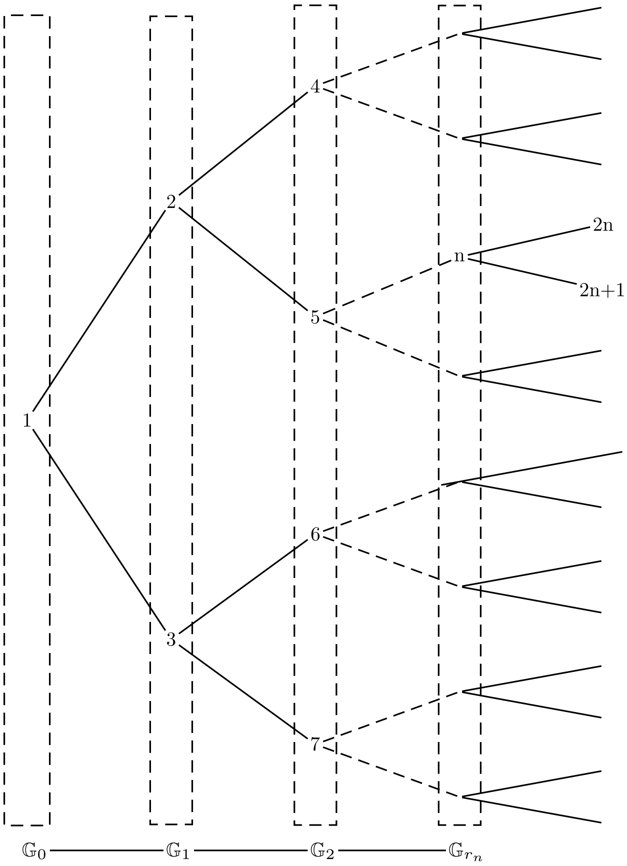

Fig. 1. The binary tree $\mathbb{T}$.

where

$$Z_n = \begin{pmatrix} X_{2n} \\ X_{2n+1} \end{pmatrix}, \qquad Y_n = \begin{pmatrix} 1 \\ \mathbb{X}_n \end{pmatrix}, \qquad V_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n+1} \end{pmatrix},$$

and the $(p+1) \times 2$ matrix parameter θ is given by

$$\theta = \begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ a_p & b_p \end{pmatrix}.$$

As in Bercu et al. [3], we introduce the least squares estimator $\hat{\theta}_n$ of θ for all $n \geq p$, from the observation of all individuals up to the nth generation (that is, the complete sub-tree $\mathbb{T}_n$)

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} Y_k Z_k^t, \tag{2.3}$$

where the $(p+1)\times(p+1)$ matrix S_n is defined as

$$S_n = \sum_{k\in\mathbb{T}_{n,p-1}} Y_k Y_k^t = \sum_{k\in\mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & \mathbb{X}_k^t \\ \mathbb{X}_k & \mathbb{X}_k\mathbb{X}_k^t \end{pmatrix}. \tag{2.4}$$

We assume, without loss of generality, that for all $n \geq p-1$, S_n is invertible. From now on, we shall make a slight abuse of notation by identifying θ and $\hat{\theta}_n$ respectively to

$$\mathrm{vec}(\theta) = \begin{pmatrix} a_0 \\ \cdot \\ \cdot \\ a_p \\ b_0 \\ \cdot \\ \cdot \\ b_p \end{pmatrix} \quad \text{and} \quad \mathrm{vec}(\hat{\theta}_n) = \begin{pmatrix} \hat{a}_{0,n} \\ \cdot \\ \cdot \\ \hat{a}_{p,n} \\ \hat{b}_{0,n} \\ \cdot \\ \cdot \\ \hat{b}_{p,n} \end{pmatrix}.$$

Let $\Sigma_n = I_2 \otimes S_n$, where $\otimes$ stands for the matrix Kronecker product. We then deduce from (2.3) that

$$\hat{\theta}_n = \Sigma_{n-1}^{-1} \sum_{k\in\mathbb{T}_{n-1,p-1}} \mathrm{vec}\big(Y_k Z_k^t\big) = \Sigma_{n-1}^{-1} \sum_{k\in\mathbb{T}_{n-1,p-1}} \begin{pmatrix} X_{2k} \\ X_k\mathbb{X}_{2k} \\ X_{2k+1} \\ X_k\mathbb{X}_{2k+1} \end{pmatrix}.$$

Consequently, (2.1) yields

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{k\in\mathbb{T}_{n-1,p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k}\mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1}\mathbb{X}_k \end{pmatrix}. \tag{2.5}$$

Denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration associated with the BAR(p) process, which means that $\mathcal{F}_n$ is the σ-algebra generated by the individuals up to the nth generation, in other words $\mathcal{F}_n = \sigma\{X_k, k\in\mathbb{T}_n\}$.

For the initial states, we set $\overline{X}_1 = \max\{\|\mathbb{X}_k\|, k \leq 2^{p-1}\}$ with the convention that $X_0 = 0$ and we introduce the following hypotheses:

(Xa) For some $a > 2$, there exists $\zeta > 0$ such that

$$\mathbb{E}\big[\exp\big(\zeta\overline{X}_1^a\big)\big] < \infty.$$

This assumption implies the weaker Gaussian integrability condition.

(X2) There is $\zeta > 0$ such that

$$\mathbb{E}\big[\exp\big(\zeta\overline{X}_1^2\big)\big] < \infty.$$

For the noise $(\varepsilon_{2n}, \varepsilon_{2n+1})$ the assumption may be of two types.

(1) In the first case we will assume the independence of the noise which allows us to impose less restrictive conditions on the exponential integrability of the noise.

Case 1: We shall assume that $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of independent and identically distributed bi-variate centered random variables with covariance matrix Γ given by

$$\Gamma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}, \quad \text{where } \sigma^2 > 0 \text{ and } |\rho| < \sigma^2. \tag{2.6}$$

For all $n \geq p-1$ and for all $k \in \mathbb{G}_n$, we set

$$\mathbb{E}\big[\varepsilon_k^2\big]=\sigma^2, \qquad \mathbb{E}\big[\varepsilon_k^4\big]=\tau^4, \qquad \mathbb{E}[\varepsilon_{2k}\varepsilon_{2k+1}]=\rho, \qquad \mathbb{E}\big[\varepsilon_{2k}^2\varepsilon_{2k+1}^2\big]=\nu^2, \quad \text{where } \tau^4>0, \nu^2<\tau^4.$$

In addition, we assume that the condition (X2) on the initial state is satisfied and that

(G2) one can find $\gamma>0$ and $c>0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_n$ and for all $|t| \leq c$

$$\mathbb{E}\big[\exp\big(t\big(\varepsilon_k^2-\sigma^2\big)\big)\big] \leq \exp\left(\frac{\gamma t^2}{2}\right).$$

In this case, we impose the following hypotheses on the scale of the deviation

(V1) (v_n) will denote an increasing sequence of positive real numbers such that

$$v_n \longrightarrow +\infty$$

and for β given by (2.2)

- if $\beta \leq \frac{1}{2}$, the sequence (v_n) is such that $\frac{v_n \log n}{\sqrt{n}} \longrightarrow 0$,
- if $\beta > \frac{1}{2}$, the sequence (v_n) is such that $(v_n\sqrt{\log n})\beta^{(r_n+1)/2} \longrightarrow 0$.

(2) In contrast with the first case, in the second case we will not assume that the sequence $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ is i.i.d. The price to pay for giving up this i.i.d. assumption is to assume higher exponential moments. Indeed we need them to make use of the MDP for martingales, especially to prove the Lindeberg condition via the Lyapunov condition.

Case 2: We shall assume that for all $n \geq p-1$ and for all $j \in \mathbb{G}_{n+1}$ $\mathbb{E}[\varepsilon_j/\mathcal{F}_n]=0$ and for all different $k, l \in \mathbb{G}_{n+1}$ with $[\frac{k}{2}] \neq [\frac{l}{2}]$, ε_k and ε_l are conditionally independent given $\mathcal{F}_n$. And we will use the same notation as in case 1: for all $n \geq p-1$ and for all $k \in \mathbb{G}_{n+1}$,

$$\mathbb{E}\big[\varepsilon_k^2/\mathcal{F}_n\big]=\sigma^2, \qquad \mathbb{E}\big[\varepsilon_k^4/\mathcal{F}_n\big]=\tau^4, \qquad \mathbb{E}[\varepsilon_{2k}\varepsilon_{2k+1}/\mathcal{F}_n]=\rho, \qquad \mathbb{E}\big[\varepsilon_{2k}^2\varepsilon_{2k+1}^2/\mathcal{F}_n\big]=\nu^2 \quad \text{a.s.}$$

where $\tau^4>0$, $\nu^2<\tau^4$ and we use also Γ for the conditional covariance matrix associated with $(\varepsilon_{2n}, \varepsilon_{2n+1})$. In this case, we assume that the condition (Xa) on the initial state is satisfied, and we shall make the following hypotheses:

(Ea) for some $a>2$, there exist $t>0$ and $E>0$ such that for all $n \geq p-1$ and for all $k \in \mathbb{G}_{n+1}$,

$$\mathbb{E}\big[\exp\big(t|\varepsilon_k|^{2a}\big)/\mathcal{F}_n\big] \leq E < \infty, \quad \text{a.s.}$$

Throughout this case, we introduce the following hypotheses on the scale of the deviation

(V2) (v_n) will denote an increasing sequence of positive real numbers such that

$$v_n \longrightarrow +\infty,$$

and for β given by (2.2)

- if $\beta^2 < \frac{1}{2}$, the sequence (v_n) is such that $\frac{v_n \log n}{\sqrt{n}} \longrightarrow 0$,
- if $\beta^2 = \frac{1}{2}$, the sequence (v_n) is such that $\frac{v_n (\log n)^{3/2}}{\sqrt{n}} \longrightarrow 0$,
- if $\beta^2 > \frac{1}{2}$, the sequence (v_n) is such that $(v_n \log n)\beta^{r_n+1} \longrightarrow 0$.

Remarks 2.1. *The condition on the scale of the deviation in case* 2, *is less restrictive than in case* 1, *since we assume a stronger integrability condition on the noise* (Ea). *This condition on the scale of the deviation naturally appears in the calculations. More precisely, the* log *term comes from the commutation of a probability and a sum.*

Remarks 2.2. *From* [14] *or* [22], *we deduce with* (Ea) *that*

(N1) *there is* $\phi > 0$ *such that for all* $n \geq p-1$, *for all* $k \in \mathbb{G}_{n+1}$ *and for all* $t \in \mathbb{R}$,

$$\mathbb{E}\big[\exp(t\varepsilon_k)/\mathcal{F}_n\big] < \exp\left(\frac{\phi t^2}{2}\right), \quad a.s.$$

We have the same conclusion in case 1, *without the conditioning*; *i.e.*

(G1) *there is* $\phi > 0$ *such that for all* $n \geq p-1$, *for all* $k \in \mathbb{G}_n$ *and for all* $t \in \mathbb{R}$,

$$\mathbb{E}\big[\exp(t\varepsilon_k)\big] < \exp\left(\frac{\phi t^2}{2}\right).$$

Remarks 2.3. *Armed with the recent development in the theory of transportation inequalities*, *exponential integrability and functional inequalities* (*see Ledoux* [22], *Gozlan* [16] *and Gozlan and Leonard* [17]), *we can prove that a sufficient condition for hypothesis* (G2) *to hold is the existence of* $t_0 > 0$ *such that for all* $n \geq p-1$ *and for all* $k \in \mathbb{G}_n$, $\mathbb{E}[\exp(t_0\varepsilon_k^2)] < \infty$.

We now turn to the estimation of the parameters σ^2 and ρ. On the one hand, we propose to estimate the conditional variance σ^2 by

$$\hat{\sigma}_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \|\hat{V}_k\|^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \big(\hat{\varepsilon}_{2k}^2 + \hat{\varepsilon}_{2k+1}^2\big),$$

where for all $n \geq p-1$ and all $k \in \mathbb{G}_n$, $\hat{V}_k^t = (\hat{\varepsilon}_{2k}, \hat{\varepsilon}_{2k+1})^t$ with

$$\begin{cases} \hat{\varepsilon}_{2k} = X_{2k} - \hat{a}_{0,n} - \sum_{i=1}^p \hat{a}_{i,n} X_{[k/2^{i-1}]}, \\ \hat{\varepsilon}_{2k+1} = X_{2k+1} - \hat{b}_{0,n} - \sum_{i=1}^p \hat{b}_{i,n} X_{[k/2^{i-1}]}. \end{cases}$$

We also introduce

$$\sigma_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} \big(\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2\big).$$

On the other hand, we estimate the conditional covariance ρ by

$$\hat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \hat{\varepsilon}_{2k}\hat{\varepsilon}_{2k+1}.$$

We also introduce

$$\rho_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} \varepsilon_{2k}\varepsilon_{2k+1}.$$

In order to establish the MDP results of our estimators, we shall make use of a martingale approach. For all $n \geq p$, set

$$M_n = \sum_{k \in \mathbb{T}_{n-1,p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k}\mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1}\mathbb{X}_k \end{pmatrix} \in \mathbb{R}^{2(p+1)}.$$

We can clearly rewrite (2.5) as

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n. \tag{2.7}$$

We know from Bercu et al. [3] that (M_n) is a square integrable martingale adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$. Its increasing process is given for all $n \geq p$ by

$$\langle M \rangle_n = \Gamma \otimes S_{n-1},$$

where S_n is given in (2.4) and Γ is given in (2.6).

Recall that for a sequence of random variables $(Z_n)_n$ on $\mathbb{R}^{d\times p}$, we say that $(Z_n)_n$ converges (v_n^2)-superexponentially fast in probability to some random variable Z if, for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\big(\|Z_n - Z\| > \delta\big) = -\infty.$$

This exponential convergence with speed v_n^2 will be abbreviated to

$$Z_n \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} Z.$$

Remarks 2.4. *Note that for a determininistic sequence that converges to some limit ℓ, it also converges (v_n^2)-superexponentially fast to ℓ for any rate v_n.*

We follow Dembo and Zeitouni [12] for the language of the large deviations, throughout this paper. Before going further, let us recall the definition of a MDP: let (v_n) be an increasing sequence of positive real numbers such that

$$v_n \longrightarrow \infty \quad \text{and} \quad \frac{v_n}{\sqrt{n}} \longrightarrow 0. \tag{2.8}$$

We say that a sequence of centered random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies a MDP with speed v_n^2 and rate function $I : S \to \mathbb{R}_+^*$ if for each $A \in \mathcal{S}$,

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{v_n} M_n \in A\right) \leq \limsup_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{v_n} M_n \in A\right) \leq -\inf_{x\in \overline{A}} I(x),$$

where A^o and $\overline{A}$ denote the interior and closure of A respectively.

Before we present the main results, let us fix some more notation. Let

$$\overline{a} = \frac{a_0 + b_0}{2}, \qquad \overline{a^2} = \frac{a_0^2 + b_0^2}{2}, \qquad \overline{A} = \frac{A + B}{2}.$$

We set

$$\Xi = \overline{a}(I_p - \overline{A})^{-1} e_1, \tag{2.9}$$

and let Λ be the unique solution of the equation (see Lemma A.4 in [3])

$$\Lambda = T + \frac{1}{2}\big(A\Lambda A^t + B\Lambda B^t\big) \tag{2.10}$$

where

$$T = \big(\sigma^2 + \overline{a^2}\big) e_1 e_1^t + \frac{1}{2}\big(a_0\big(A\Xi e_1^t + e_1 \Xi^t A^t\big) + b_0\big(B\Xi e_1^t + e_1 \Xi^t B^t\big)\big). \tag{2.11}$$

We also introduce the following matrices L and Σ given by

$$L = \begin{pmatrix} 1 & \Xi \\ \Xi & \Lambda \end{pmatrix} \quad \text{and} \quad \Sigma = I_2 \otimes L. \tag{2.12}$$

3. Main results

Let us present now the main results of this paper. In the following theorem, we give the deviation inequalities of the estimator of the parameters.

Theorem 3.1.

(i) *In case* 1, *we have for all* $\delta > 0$ *and for all* $\ell > 0$ *such that* $\ell < \|\Sigma\|/(1+\delta)$

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) \le \begin{cases} c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{2^n}{(n-1)^2}\big) & \text{if } \beta < \frac{1}{2}, \\ c_1(n-1)\exp\big(\frac{-c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{2^n}{(n-1)^2}\big) & \text{if } \beta = \frac{1}{2}, \\ c_1(n-1)\exp\big(\frac{-c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{1}{(n-1)\beta^n}\big) & \text{if } \beta > \frac{1}{2}, \end{cases} \tag{3.1}$$

where the constants c_1, c_2 *and* c_3 *depend on* σ^2, β, γ *and* ϕ, *may differ line by line and are such that* $c_1, c_2 > 0$, $c_3 \ge 0$.

(ii) *In case* 2, *we have for all* $\delta > 0$ *and for all* $\ell > 0$ *such that* $\ell < \|\Sigma\|/(1+\delta)$

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) \le \begin{cases} c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)} \frac{2^n}{(n-1)^2}\big) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)} \frac{2^n}{(n-1)^3}\big) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)} \frac{1}{(n-1)^2\beta^{2n}}\big) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.2}$$

where the constants c_1, c_2, c_3, *and* c_4 *depend on* σ^2, β, γ *and* ϕ, *may differ line by line and are such that* $c_1, c_2 > 0$, $c_3, c_4 \ge 0$, $(c_3, c_4) \ne (0,0)$.

Remarks 3.2. *Note that the estimate* (3.2) *is stronger than the estimate* (3.1). *This is due to the fact that the integrability condition* (Ea) *in case* 2 *is stronger than the integrability condition* (G2) *in case* 1.

Remarks 3.3. *Let us stress that by tedious but straightforward calculations, the constants which appear in the previous theorem can be well estimated.*

Remarks 3.4. *The upper bounds in previous theorem hold for arbitrary* $n \ge p-1$ (*not a limit relation, unlike the results below*), *hence they are very practical* (*in nonasymptotic statistics*) *when sample size does not allow the application of limit theorems.*

In the next result, we present the MDP of the estimator $\hat{\theta}_n$.

Theorem 3.5. *In case* 1 *or in case* 2, *the sequence* $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\theta}_n - \theta)/v_{|\mathbb{T}_{n-1}|})_{n\ge 1}$ *satisfies the MDP on* $\mathbb{R}^{2(p+1)}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_\theta(x) = \sup_{\lambda\in\mathbb{R}^{2(p+1)}} \big\{\lambda^t x - \lambda\big(\Gamma \otimes L^{-1}\big)\lambda^t\big\} = \frac{1}{2} x^t \big(\Gamma \otimes L^{-1}\big)^{-1} x, \tag{3.3}$$

where L *and* Γ *are given in* (2.12) *and* (2.6) *respectively.*

Remarks 3.6. *Similar results about deviation inequalities and MDP have already been obtained in* [5], *in a restrictive case of bounded or Gaussian noise and when* $p = 1$, *but results therein also hold for general Markov models. Moreover in* [5], *when the noise is Gaussian, the range of speed of MDP is very restricted in comparison to the range of speed of MDP in case* 1 *of this paper. These improvements are due to the fact that in this paper, we take advantage of the autoregressive structure of the process while in* [5], *only its Markovian nature is used.*

Let us also mention that in case 2, *the Markovian nature of* BAR(p) *processes is lost and this case is not studied in* [5]. *However in case* 2, *for* $p = 1$, *if we assume that the initial state* X_1 *and the noise take their values in a compact set*, *we can find the same results as in* [5]. *The results of this paper then allow to extend the results of the latter paper.*

Let us consider now the estimation of the parameter in the noise process.

Theorem 3.7. *Let* (v_n) *an increasing sequence of positive real numbers such that*

$$v_n \longrightarrow \infty \quad \text{and} \quad \frac{v_n}{\sqrt{n}} \longrightarrow 0.$$

In case 1 *or in case* 2,

(1) *the sequence* $(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_{\sigma^2}(x) = \frac{x^2}{\tau^4 - 2\sigma^4 + \nu^2}; \tag{3.4}$$

(2) *the sequence* $(\sqrt{|\mathbb{T}_{n-1}|}(\rho_n - \rho)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_\rho(x) = \frac{x^2}{2(\nu^2 - \rho^2)}. \tag{3.5}$$

Remarks 3.8. *Note that in this case the MDP holds for all the scales* (v_n) *verifying* (2.8) *without other restriction.*

Remarks 3.9. *It would be more interesting to prove the MDP for* $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\sigma}_n^2 - \sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$, *which will be the case if one proves for example that* $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\sigma}_n^2 - \sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *and* $(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *are exponentially equivalent in the sense of the MDP. This is described by the following convergence*

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}\big(\hat{\sigma}_n^2 - \sigma_n^2\big) \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

The proof is very technical and very restrictive with respect to the scale (v_n) *of the deviation. Actually we are only able to prove that*

$$\hat{\sigma}_n^2 - \sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

This superexponential convergence will be proved in Theorem 3.10.

In the following theorem we state the superexponential convergence.

Theorem 3.10. *In case* 1 *or in case* 2, *we have*

$$\hat{\sigma}_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2.$$

In case 1, *if instead of* (G2), *we assume that*

(G2′) *one can find* $\gamma' > 0$ *such that for all* $n \geq p-1$, *for all* $k, l \in \mathbb{G}_{n+1}$ *with* $[\frac{k}{2}] = [\frac{l}{2}]$ *and for all* $t \in]-c, c[$ *for some* $c > 0$,

$$\mathbb{E}\big[\exp t(\varepsilon_k\varepsilon_l - \rho)\big] \leq \exp\left(\frac{\gamma' t^2}{2}\right),$$

and in case 2, *if instead of* (Ea), *we assume that*

(E2′) *one can find* $\gamma' > 0$ *such that for all* $n \geq p-1$, *for all* $k, l \in \mathbb{G}_{n+1}$ *with* $[\frac{k}{2}] = [\frac{l}{2}]$ *and for all* $t \in \mathbb{R}$

$$\mathbb{E}\big[\exp t(\varepsilon_k\varepsilon_l - \rho)/\mathcal{F}_n\big] \leq \exp\left(\frac{\gamma' t^2}{2}\right), \quad a.s.$$

Then in case 1 *or in case* 2, *we have*

$$\hat{\rho}_n \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \rho.$$

Before going into the proofs, let us gather here for the convenience of the reader two theorems useful to establish MDP for martingales and used intensively in this paper. From these two theorems, we will be able to give a strategy for the proof.

The following proposition corresponds to the unidimensional case of Theorem 1 in [13].

Proposition 3.11. *Let* $M = (M_n, \mathcal{H}_n, n \geq 0)$ *be a centered square real valued integrable martingale defined on a probability space* $(\Omega, \mathcal{H}, \mathbb{P})$ *and let* $(\langle M\rangle_n)$ *be its bracket. Let* (v_n) *be an increasing sequence of real numbers satisfying* (2.8).

Let $c(n) := \frac{\sqrt{n}}{v_n}$ *be nondecreasing, and define the reciprocal function* $c^{-1}(t)$ *by*

$$c^{-1}(t) := \inf\{n \in \mathbb{N}\colon\ c(n) \geq t\}.$$

Under the following conditions

(D1) *there exists* $Q \in \mathbb{R}^*_+$ *such that* $\frac{\langle M\rangle_n}{n} \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} Q$;

(D2) $\limsup_{n\to+\infty} \frac{n}{v_n^2} \log(n \operatorname{ess\,sup}_{1\leq k\leq c^{-1}(\sqrt{n+1}v_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > v_n\sqrt{n}/\mathcal{H}_{k-1})) = -\infty$;

(D3) *for all* $a > 0$ $\frac{1}{n}\sum_{k=1}^n \mathbb{E}(|M_k - M_{k-1}|^2 \mathbf{1}_{\{|M_k - M_{k-1}| \geq a(\sqrt{n}/v_n)\}}/\mathcal{H}_{k-1}) \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0$;

$(M_n/v_n\sqrt{n})_{n\geq 0}$ *satisfies the MDP in* $\mathbb{R}$ *with speed* v_n^2 *and rate function* $I(x) = \frac{x^2}{2Q}$.

Let us introduce a simplified version of Puhalskii's result [24] applied to a sequence of martingale differences.

Theorem 3.12. *Let* $(m_j^n)_{1\leq j\leq n}$ *be a triangular array of martingale differences with values in* $\mathbb{R}^d$, *with respect to some filtration* $(\mathcal{H}_n)_{n\geq 1}$. *Let* (v_n) *be an increasing sequence of real numbers satisfying* (2.8). *Under the following conditions*

(P1) *there exists a symmetric positive semi-definite matrix* Q *such that*

$$\frac{1}{n}\sum_{k=1}^n \mathbb{E}\big[m_k^n (m_k^n)^t|\mathcal{H}_{k-1}\big] \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} Q,$$

(P2) *there exists a constant* $c > 0$ *such that, for each* $1 \leq k \leq n$, $|m_k^n| \leq c\frac{\sqrt{n}}{v_n}$ *a.s.*,

(P3) *for all* $a > 0$, *we have the exponential Lindeberg's condition*

$$\frac{1}{n}\sum_{k=1}^n \mathbb{E}\big[|m_k^n|^2 \mathbf{1}_{\{|m_k^n|\geq a(\sqrt{n}/v_n)\}}|\mathcal{H}_{k-1}\big] \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0,$$

$(\sum_{k=1}^n m_k^n/(v_n\sqrt{n}))_{n\geq 1}$ *satisfies an MDP on* $\mathbb{R}^d$ *with speed* v_n^2 *and rate function*

$$\Lambda^*(v) = \sup_{\lambda\in\mathbb{R}^d}\left(\lambda^t v - \frac{1}{2}\lambda^t Q\lambda\right).$$

In particular, if Q is invertible, $\Lambda^*(v) = \frac{1}{2} v^t Q^{-1} v$.

As the reader can imagine naturally now, the strategy of the proof of the MDP consists in the following steps:

- the superexponential convergence of the quadratic variation of the martingale (M_n). This step is very crucial and the key for the rest of the paper. It will be realized by means of powerful exponential inequalities. This allows us to obtain the deviation inequalities for the estimator of the parameters,
- introduce a truncated martingale which satisfies the MDP, thanks to the classical Theorem 3.12,
- the truncated martingale is an exponentially good approximation of (M_n), in the sense of the moderate deviation.

4. Superexponential convergence of the quadratic variation of the martingale

First, it is necessary to establish the superexponential convergence of the quadratic variation of the martingale (M_n), properly normalized in order to prove the MDP of the estimators. Its proof is very technical, but crucial for the rest of the paper. This section contains also some deviation inequalities for some quantities needed in the proof later.

Proposition 4.1. *In case* 1 *or case* 2, *we have*

$$\frac{S_n}{|\mathbb{T}_n|} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} L, \tag{4.1}$$

where S_n is given in (2.4) *and L is given in* (2.12).

For the proof we focus on case 2. Proposition 4.1 will follow from Proposition 4.3 and Proposition 4.9 below, where we assume that the sequence (v_n) satisfies the condition (V2). Proposition 4.10 gives some ideas of the proof in case 1.

Remarks 4.2. *Using* [14], *we infer from* (Ea) *that*

(N2) *one can find $\gamma > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$ and for all $t \in \mathbb{R}$*

$$\mathbb{E}\big[\exp t\big(\varepsilon_k^2 - \sigma^2\big)/\mathcal{F}_n\big] \leq \exp\left(\frac{\gamma t^2}{2}\right) \quad a.s.$$

Proposition 4.3. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k \mathbb{X}_k^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda,$$

where Λ is given in (2.10).

Proof. Let

$$K_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \mathbb{X}_k^t \quad \text{and} \quad L_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2. \tag{4.2}$$

Then from (2.1), and after straightforward calculations (see p. 2519 in [3] for more details), we get that

$$\frac{K_n}{2^{n+1}} = \frac{1}{2^{n-p+1}} \sum_{C \in \{A;B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t + \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C T_{n-k} C^t,$$

where the notation $\{A; B\}^k$ means the set of all products of A and B with exactly k terms. The cardinality of $\{A; B\}^k$ is obviously 2^k, and

$$T_k = \frac{L_k}{2^{k+1}} e_1 e_1^t + \overline{a^2}\left(\frac{2^k - 2^{p-1}}{2^k}\right) e_1 e_1^t + I_k^{(1)} + I_k^{(2)} + \frac{1}{2^{k+1}} U_k$$

with $\overline{a^2} = (a_0^2 + b_0^2)/2$ and

$$I_k^{(1)} = \frac{1}{2}\left(a_0\left(A \frac{H_{k-1}}{2^k} e_1^t + e_1 \frac{H_{k-1}}{2^k} A^t\right) + b_0\left(B \frac{H_{k-1}}{2^k} e_1^t + e_1 \frac{H_{k-1}}{2^k} B^t\right)\right), \tag{4.3}$$

$$I_k^{(2)} = \left(\frac{1}{2^k} \sum_{l \in \mathbb{T}_{k-1,p-1}} (a_0 \varepsilon_{2l} + b_0 \varepsilon_{2l+1})\right) e_1 e_1^t, \tag{4.4}$$

$$U_k = \sum_{l \in \mathbb{T}_{k-1,p-1}} \varepsilon_{2l}\big(A \mathbb{X}_l e_1^t + e_1 \mathbb{X}_l^t A^t\big) + \varepsilon_{2l+1}\big(B \mathbb{X}_l e_1^t + e_1 \mathbb{X}_l^t B^t\big). \tag{4.5}$$

Then the proposition will follow if we prove Lemmas 4.4, 4.6, 4.7, 4.8 and 4.5. □

Lemma 4.4. *Assume that hypothesis* (Xa) *is satisfied. Then we have*

$$\frac{1}{2^{n-p+1}} \sum_{C \in \{A; B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{4.6}$$

where K_p is given in (4.2).

Proof. We get easily

$$\left\| \frac{1}{2^{n-p+1}} \sum_{C \in \{A; B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t \right\| \le c \beta^{2n} \overline{X}_1^2,$$

where β is given in (2.2), $\overline{X}_1$ is introduced in (Xa) and c is a positive constant which depends on p. Next, Chernoff inequality and hypothesis (X2) lead us easily to (4.6). □

Lemma 4.5. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A; B\}^k} C \frac{U_{n-k}}{2^{n-k+1}} C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{4.7}$$

where U_k is given by (4.5).

Proof. Let $V_n = \sum_{k \in \mathbb{T}_{n-1,p-1}} \varepsilon_{2k} X_k$. Then (V_n) is an $\mathcal{F}_n$-martingale and its increasing process satisfies, for all $n \ge p$,

$$\langle V \rangle_n = \sigma^2 \sum_{k \in \mathbb{T}_{n-1,p}} X_k^2 \le \sigma^2 \sum_{k \in \mathbb{T}_{n-1,p-1}} X_k^2 \le \sigma^2 \sum_{k \in \mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2.$$

For $\lambda > 0$, we infer from hypothesis (N1) that $(Y_k)_{p \le k \le n}$ given by

$$Y_n = \exp\left(\lambda V_n - \frac{\lambda^2 \phi}{2} \sum_{k \in \mathbb{T}_{n-1,p-1}} X_k^2\right),$$

is an $\mathcal{F}_k$-supermartingale and moreover $\mathbb{E}[Y_p] \leq 1$. For $B > 0$ and $\delta > 0$, we have

$$\begin{aligned}\mathbb{P}\left(\frac{V_n}{|\mathbb{T}_n|+1} > \delta\right) &\leq \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > B\right) + \mathbb{P}\left(Y_n > \exp\left(\lambda\delta - \frac{\lambda^2 B}{2}\right)2^{n+1}\right)\\ &\leq \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > B\right) + \exp\left(\left(-\lambda\delta + \frac{\lambda^2 B}{2}\right)2^{n+1}\right).\end{aligned}$$

Optimizing on λ, we get

$$\mathbb{P}\left(\frac{V_n}{|\mathbb{T}_n|+1} > \delta\right) \leq \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > B\right) + \exp\left(-\frac{\delta^2}{B}2^{n+1}\right).$$

Since the same thing works for $-V_n$ instead of V_n and using the following inequality,

$$\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 \leq \sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2,$$

we get

$$\mathbb{P}\left(\frac{|V_n|}{|\mathbb{T}_n|+1} > \delta\right) \leq \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2 > B\right) + \exp\left(-\frac{\delta^2}{B}2^{n+1}\right). \tag{4.8}$$

From [3], with $\alpha = \max(|a_0|, |b_0|)$, we have

$$\sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2 \leq \frac{4}{1-\beta}P_{n-1} + \frac{4\alpha^2}{1-\beta}Q_{n-1} + 2\overline{X}_1^2 R_{n-1}, \tag{4.9}$$

where

$$P_n = \sum_{k\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_k-p}\beta^i\varepsilon^2_{[k/2^i]}, \qquad Q_n = \sum_{k\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_k-p}\beta^i, \qquad R_n = \sum_{k\in\mathbb{T}_{n,p-1}}\beta^{2(r_k-p+1)}.$$

Now, to control the first term in the right hand side of (4.8), we will use the decomposition given by (4.9). From the convergence of $\frac{4\phi}{(1-\beta)(|\mathbb{T}_n|+1)}P_n$ and $\frac{4\phi\alpha^2}{(1-\beta)(|\mathbb{T}_n|+1)}Q_n$ (see [3] for more details) let l_1 and l_2 be such that

$$\frac{4\phi P_{n-1}}{(1-\beta)(|\mathbb{T}_n|+1)} \to l_1 \quad \text{and} \quad \forall n \geq p-1 \quad \frac{4\phi\alpha^2 Q_{n-1}}{(1-\beta)(|\mathbb{T}_n|+1)} < l_2.$$

For $\delta > 0$, we choose $B = \delta + l_1 + l_2$, using (4.9), we then have

$$\begin{aligned}&\mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2 > B\right)\\ &\quad\leq \mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) + \mathbb{P}\left(\frac{Q_{n-1}}{|\mathbb{T}_n|+1} - l_2' > \delta_2\right) + \mathbb{P}\left(\frac{R_{n-1}\overline{X}_1^2}{|\mathbb{T}_n|+1} > \delta_3\right),\end{aligned} \tag{4.10}$$

where

$$\delta_1 = \frac{(1-\beta)\delta}{12\phi}, \qquad l_1' = \frac{(1-\beta)l_1}{4\phi}, \qquad \delta_2 = \frac{(1-\beta)\delta}{12\alpha^2\phi}, \qquad l_2' = \frac{(1-\beta)l_2}{4\alpha^2\phi} \quad \text{and} \quad \delta_3 = \frac{\delta}{6\phi}.$$

First, by the choice of l_2, we have

$$\mathbb{P}\left(\frac{Q_{n-1}}{|\mathbb{T}_n|+1} - l_2' > \delta_2\right) = 0. \tag{4.11}$$

Next, from Chernoff inequality and hypothesis (X2) we get easily

$$\mathbb{P}\left(\frac{R_{n-1}\overline{X}_1^2}{|\mathbb{T}_n|+1} > \delta_3\right) \leq \begin{cases} c_1 \exp\left(-c_2\delta 2^{n+1}\right) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\left(-c_2\delta \frac{2^{n+1}}{n+1}\right) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\left(-c_2\delta\left(\frac{1}{\beta^2}\right)^{n+1}\right) & \text{if } \beta > \frac{\sqrt{2}}{2} \end{cases} \tag{4.12}$$

for some positive constants c_1 and c_2. Let us now control the first term of the right hand side of (4.10).

First case. If $\beta = \frac{1}{2}$, from [3]

$$P_{n-1} = \sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\varepsilon_i^2 \quad \text{and} \quad l_1' = \sigma^2.$$

We thus have

$$\frac{P_{n-1}}{|\mathbb{T}_n|+1} - \sigma^2 = \frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2 - \sigma^2\right) + \sigma^2\left(\sum_{k=p}^{n-1}\frac{n-k}{2^{n+1-k}} - 1\right).$$

In addition, we also have

$$\sigma^2\left(\sum_{k=p}^{n-1}\frac{n-k}{2^{n+1-k}} - 1\right) \leq 0.$$

We thus deduce that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2 - \sigma^2\right) > \delta_1\right).$$

On the one hand we have

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2 - \sigma^2\right) > \delta_1\right) \\ &\quad \leq \sum_{\eta=0}^{1}\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i+\eta}^2 - \sigma^2\right) > \delta_1/2\right). \end{aligned} \tag{4.13}$$

On the other hand, for all $\lambda > 0$, an application of Chernoff inequality yields

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2 - \sigma^2\right) > \delta_1/2\right) \\ &\quad \leq \exp\left(\frac{-\delta_1\lambda 2^{n+1}}{2}\right) \times \mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right]. \end{aligned}$$

From hypothesis (N2) we get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2-\sigma^2\right)\right)\right]\\
&\quad=\mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2-\sigma^2\right)\right)\Big/\mathcal{F}_n\right]\right]\\
&\quad=\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-3}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2-\sigma^2\right)\right)\prod_{i\in\mathbb{G}_{n-2}}\mathbb{E}\left[\exp\left(\lambda\left(\varepsilon_{2i}^2-\sigma^2\right)\right)/\mathcal{F}_n\right]\right]\\
&\quad\le\exp\left(\lambda^2\gamma|\mathbb{G}_{n-2}|\right)\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-3}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2-\sigma^2\right)\right)\right].
\end{aligned}$$

Iterating this procedure, we obtain

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2-\sigma^2\right)\right)\right]&\le\exp\left(\gamma\lambda^2\sum_{k=2}^{n-p+1}k^2|\mathbb{G}_{n-k}|\right)\\
&\le\exp\left(c\gamma\lambda^2 2^{n+1}\right),
\end{aligned}$$

where $c=\sum_{k=1}^{\infty}\frac{k^2}{2^{k+2}}=\frac{3}{4}$. Optimizing on λ, we are led, for some positive constant c_1 to

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2-\sigma^2\right)>\delta_1/2\right)\le\exp\left(-c_1\delta^2|\mathbb{T}_n|\right).$$

Following the same lines, we obtain the same inequality for the second term in (4.13). It then follows that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1}-l_1'>\delta_1\right)\le c_1\exp\left(-c_2\delta^2|\mathbb{T}_n|\right)\tag{4.14}$$

for some positive constants c_1 and c_2.

Second case. If $\beta\neq\frac{1}{2}$, then from [3], we have $l_1'=\frac{\sigma^2}{2(1-\beta)}$. Since

$$\sigma^2\left(\sum_{k=p}^{n-1}\frac{1-(2\beta)^{n-k}}{(1-2\beta)2^{n-k+1}}\right)\le\frac{\sigma^2}{2(1-\beta)},$$

we deduce that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1}-l_1'>\delta_1\right)\le\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}\frac{1-(2\beta)^{n-k}}{1-2\beta}\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2-\sigma^2\right)>\delta_1\right).$$

- If $\beta<\frac{1}{2}$, then for some positive constant c, we have

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1}-l_1'>\delta_1\right)\le\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2-\sigma^2\right)>c\delta_1\right).$$

Proceeding now as in the proof of (4.21), we get

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq c_1 \exp\bigl(-c_2\delta^2|\mathbb{T}_n|\bigr) \tag{4.15}$$

for some positive constants c_1 and c_2.

- If $\beta > \frac{1}{2}$, then for some positive constant c, we have

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(2\beta)^{n-k}\sum_{i\in\mathbb{G}_k}\bigl(\varepsilon_i^2-\sigma^2\bigr) > c\delta_1\right).$$

Now, from Chernoff inequality, hypothesis (N2) and after several successive conditioning, we get for all $\lambda > 0$

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(2\beta)^{n-k}\sum_{i\in\mathbb{G}_k}\bigl(\varepsilon_i^2-\sigma^2\bigr) > c\delta_1\right)\\
&\quad\leq \exp\bigl(-c\delta_1\lambda 2^{n+1}\bigr)\exp\left(\gamma\lambda^2 2^{n+1}\sum_{k=2}^{n-p+1}\bigl(2\beta^2\bigr)^k\right).
\end{aligned}$$

Next, optimizing over λ, we are led, for some positive constant c to

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \begin{cases}\exp\bigl(-c\delta^2|\mathbb{T}_n|\bigr) & \text{if } \frac{1}{2} < \beta < \frac{\sqrt{2}}{2},\\ \exp\bigl(-c\delta^2\frac{|\mathbb{T}_n|}{n}\bigr) & \text{if } \beta = \frac{\sqrt{2}}{2},\\ \exp\bigl(-c\delta^2(\frac{1}{\beta^2})^{n+1}\bigr) & \text{if } \beta > \frac{\sqrt{2}}{2}.\end{cases} \tag{4.16}$$

Now combining (4.8), (4.10), (4.11), (4.12), (4.14), (4.15) and (4.16), we have thus showed that

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}|V_n| > \delta\right)\\
&\leq \begin{cases} c_1\exp\bigl(-c_2\delta^2 2^{n+1}\bigr) + c_1\exp\bigl(-c_2\delta 2^{n+1}\bigr) + \exp\bigl(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\bigr) & \text{if } \beta < \frac{\sqrt{2}}{2},\\ c_1\exp\bigl(-c_2\delta^2\frac{2^{n+1}}{n+1}\bigr) + c_1\exp\bigl(-c_2\delta\frac{2^{n+1}}{n+1}\bigr) + \exp\bigl(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\bigr) & \text{if } \beta = \frac{\sqrt{2}}{2},\\ c_1\exp\bigl(-c_2\delta^2(\frac{1}{\beta^2})^{n+1}\bigr) + c_1\exp\bigl(-c_2\delta(\frac{1}{\beta^2})^{n+1}\bigr) + \exp\bigl(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\bigr) & \text{if } \beta > \frac{\sqrt{2}}{2},\end{cases}
\end{aligned} \tag{4.17}$$

where the positive constants c_1 and c_2 may differ term by term.

One can easily check that the coefficients of the matrix U_n are linear combinations of terms similar to V_n, so that performing calculations similar to the above for each of them, we deduce the same deviation inequalities for U_n as in (4.17).

Now we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k}C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\| > \delta\right) &\leq \mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}\frac{1}{2^{n-k+1}}\bigl\|CU_{n-k}C^t\bigr\| > \delta\right)\\
&\leq \mathbb{P}\left(\sum_{k=p}^{n}\beta^{2(n-k)}\frac{1}{|\mathbb{T}_k|+1}\|U_k\| > \delta\right)\\
&\leq \sum_{k=p}^{n}\mathbb{P}\left(\frac{\|U_k\|}{|\mathbb{T}_k|+1} > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right).
\end{aligned}$$

From (4.17), we infer the following

$$\mathbb{P}\Bigg(\sum_{k=0}^{n-p}\frac{1}{2^k}\Bigg\|\sum_{C\in\{A;B\}^k}C\frac{U_{n-k}}{2^{n-k+1}}C^t\Bigg\|>\delta\Bigg)$$

$$\leq\begin{cases} c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta^2(2\beta^4)^{k+1}}{n^2\beta^{4n}}\big)+c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta(2\beta^2)^{k+1}}{n\beta^{2n}}\big) \\ \quad +c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\big) \quad \text{if } \beta<\frac{\sqrt{2}}{2}, \\ c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta^2 4^n}{n^2(k+1)2^{k+1}}\big)+c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta 2^n}{(k+1)n}\big) \\ \quad +c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl2^{-(n-k-1)})n2^{-(n-k-1)}}\big) \quad \text{if } \beta=\frac{\sqrt{2}}{2}, \\ c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta^2(2\beta^2)^{k+1}}{n^2\beta^{4n}}\big)+c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta}{n\beta^{2n}}\big) \\ \quad +c_1\sum_{k=p}^{n}\exp\big(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\big) \quad \text{if } \beta>\frac{\sqrt{2}}{2}, \end{cases}$$

where $l=l_1+l_2$ and the positive constants c_1 and c_2 may differ term by term.

Now

- If $\beta<\frac{\sqrt{2}}{2}$, then on the one hand,

$$\sum_{k=p}^{n}\exp\Big(-c\frac{\delta^2(2\beta^4)^{k+1}}{n^2\beta^{4n}}\Big)$$

$$=\exp\Big(-c\delta^2\beta^4\frac{2^{n+1}}{n^2}\Big)\Bigg(1+\sum_{k=p}^{n-1}\Big(\exp\Big(\frac{-c\delta^2}{n^2}\Big)\Big)^{(2\beta^4)^{k+1}\beta^{-4n}(1-(2\beta^4)^{n-k})}\Bigg)$$

$$\leq\exp\Big(-c\delta^2\beta^4\frac{2^{n+1}}{n^2}\Big)\big(1+\mathrm{o}(1)\big),$$

where the last inequality follows from the fact that for some positive constant c_1,

$$\big(2\beta^4\big)^{k+1}\beta^{-4n}\big(1-\big(2\beta^4\big)^{n-k}\big)\propto c_1\big(2\beta^4\big)^{k+1}\beta^{-4n}.$$

On the other hand, following the same lines as before, we obtain

$$\sum_{k=p}^{n}\exp\Big(-\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\Big)\leq\sum_{k=p}^{n}\exp\Big(-c\delta^2\frac{2^{k+1}}{n^2\beta^{2(n-k-1)}}\Big)$$

$$\leq\exp\Big(-c\frac{\delta^2 2^{n+1}}{(\delta+l)n^2}\Big)\big(1+\mathrm{o}(1)\big),$$

and

$$\sum_{k=p}^{n}\exp\Big(-c\frac{\delta(2\beta^2)^{k+1}}{n\beta^{2n}}\Big)\leq\sum_{k=p}^{n}\exp\Big(-c\frac{\delta(2\beta^2)^{k+1}}{n^2\beta^{2n}}\Big)$$

$$\leq\exp\Big(-c\delta\frac{2^{n+1}}{n^2}\Big)\big(1+\mathrm{o}(1)\big).$$

We thus deduce that

$$\mathbb{P}\Bigg(\sum_{k=0}^{n-p}\frac{1}{2^k}\Bigg\|\sum_{C\in\{A;B\}^k}C\frac{U_{n-k}}{2^{n-k+1}}C^t\Bigg\|>\delta\Bigg)\leq c_1\exp\Big(-c_2\delta^2\frac{2^{n+1}}{n^2}\Big)+c_1\exp\Big(-c_2\delta\frac{2^{n+1}}{n^2}\Big) \tag{4.18}$$

for some positive constants c_1 and c_2.

- If $\beta = \frac{\sqrt{2}}{2}$, then following the same lines as before, we show that

$$\sum_{k=p}^{n} \exp\Big(-c\delta^2 \frac{4^n}{n^2(k+1)2^{k+1}}\Big) \leq \exp\Big(-c\delta^2 \frac{2^{n+1}}{n^3}\Big)\big(1+\mathrm{o}(1)\big),$$

$$\sum_{k=p}^{n} \exp\Big(-\frac{\delta^2 2^{k+1}}{(\delta + ln2^{-(n-k-1)})n2^{-(n-k-1)}}\Big) \leq \exp\Big(-c\frac{\delta^2 2^{n+1}}{n^2(\delta+l)}\Big)\big(1+\mathrm{o}(1)\big),$$

$$\sum_{k=p}^{n} \exp\Big(-c\delta \frac{2^n}{n(k+1)}\Big) \leq \exp\Big(-c\delta \frac{2^{n+1}}{n^3}\Big)\big(1+\mathrm{o}(1)\big).$$

It then follows that

$$\begin{aligned}&\mathbb{P}\Bigg(\sum_{k=0}^{n-p} \frac{1}{2^k}\Bigg\| \sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\Bigg\| > \delta\Bigg)\\ &\quad\leq c_1 \exp\Big(-c_2\delta^2 \frac{2^{n+1}}{n^3}\Big) + c_1 \exp\Big(-c_2 \frac{\delta^2 2^{n+1}}{n^2(\delta+l)}\Big) + c_1 \exp\Big(-c_2\delta \frac{2^{n+1}}{n^3}\Big)\end{aligned} \tag{4.19}$$

for some positive constants c_1 and c_2.

- If $\beta > \frac{\sqrt{2}}{2}$, once again following the previous lines, we get

$$\begin{aligned}&\mathbb{P}\Bigg(\sum_{k=0}^{n-p} \frac{1}{2^k}\Bigg\| \sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\Bigg\| > \delta\Bigg)\\ &\quad\leq c_1 \exp\Big(-c_2\delta^2 \frac{1}{n^2\beta^{2n}}\Big) + c_1 \exp\Big(-c_2 \frac{\delta^2}{(\delta+l)n^2\beta^{2n}}\Big) + c_1 n \exp\Big(-c_2 \frac{\delta}{n^2\beta^{2n}}\Big)\end{aligned} \tag{4.20}$$

for some positive constants c_1 and c_2.

We infer from the inequalities (4.18), (4.19) and (4.20) that

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

□

Lemma 4.6. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\frac{L_{n-k}}{2^{n-k}} e_1 e_1^t C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} l, \tag{4.21}$$

where L_k is given in the second part of (4.2) *and*

$$l = \sum_{k=0}^{+\infty} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\big(\sigma^2 e_1 e_1^t\big)C^t$$

is the unique solution of the equation

$$l = \sigma^2 e_1 e_1^t + \frac{1}{2}\big(AlA^t + BlB^t\big).$$

Proof. First, since we have for all $k \geq p$ the following decomposition on odd and even part

$$\sum_{i\in\mathbb{T}_{k,p}} \left(\varepsilon_i^2 - \sigma^2\right) = \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right) + \left(\varepsilon_{2i+1}^2 - \sigma^2\right),$$

we obtain for all $\delta > 0$ that

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\sum_{i\in\mathbb{T}_{k,p}} \left(\varepsilon_i^2 - \sigma^2\right) > \delta\right) \leq \sum_{\eta=0}^{1} \mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i+\eta}^2 - \sigma^2\right) > \frac{\delta}{2}\right).$$

We will treat only the case $\eta = 0$. Chernoff inequality gives us for all $\lambda > 0$

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right) > \frac{\delta}{2}\right) \leq \exp\left(-\lambda\frac{\delta}{2}2^{k+1}\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right].$$

We obtain from hypothesis (N2), after conditioning by $\mathcal{F}_{k-1}$

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right] \leq \exp\left(\lambda^2\gamma|\mathbb{G}_{k-1}|\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-2,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right].$$

Iterating this, we deduce that

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right] \leq \exp\left(\gamma\lambda^2\sum_{l=p-1}^{k-1}|\mathbb{G}_l|\right) \leq \exp\left(\gamma\lambda^2 2^{k+1}\right).$$

Next, optimizing on λ, we get

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right) > \frac{\delta}{2}\right) \leq \exp\left(-c\delta^2|\mathbb{T}_k|\right)$$

for some positive constant c which depends on γ. Applying the foregoing to the random variables $-(\varepsilon_i^2 - \sigma^2)$, we obtain

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\left|\sum_{i\in\mathbb{T}_{k,p}} \left(\varepsilon_i^2 - \sigma^2\right)\right| > \delta\right) \leq 4\exp\left(-c\delta^2|\mathbb{T}_k|\right). \tag{4.22}$$

Now we have

$$\begin{aligned}\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\frac{L_{n-k}}{2^{n-k}}e_1e_1^tC^t - l &= \sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\right)e_1e_1^tC^t \\ &\quad - \sum_{k=n-p+1}^{\infty}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\left(\sigma^2e_1e_1^t\right)C^t\end{aligned}$$

and since the second term of the right hand side of the last equality is deterministic and tends to 0, to prove Lemma 4.6, it suffices to show that

$$\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\right)e_1e_1^tC^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

From the following inequalities

$$\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|\le\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}\left|\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right|\left\|Ce_1e_1^tC^t\right\|$$
$$\le\sum_{k=p}^{n}\beta^{2(n-k)}\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|$$

and from (4.22) applied with $\delta/((n-p+1)\beta^{2(n-k)})$ instead of δ, we get

$$\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)\le\mathbb{P}\left(\sum_{k=p}^{n}\beta^{2(n-k)}\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|>\delta\right)$$
$$\le\sum_{k=p}^{n}\mathbb{P}\left(\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|>\frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right)$$
$$\le c_1\sum_{k=p}^{n}\exp\left(-c_2\delta^2\frac{(2\beta^4)^{k+1}}{n^2\beta^{4n}}\right).$$

Now, following the same lines as in the proof of (4.7) we obtain

$$\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)\le\begin{cases}c_1\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^2}\right) & \text{if } \beta^4<\frac{1}{2},\\ c_1n\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^2}\right) & \text{if } \beta^4=\frac{1}{2},\\ c_1\exp\left(-c_2\delta^2\frac{1}{n^2\beta^{4n}}\right) & \text{if } \beta^4>\frac{1}{2}\end{cases}\tag{4.23}$$

for some positive constants c_1 and c_2. From (4.23), we infer that (4.21) holds. □

Lemma 4.7. *Assume that hypothesis* (N1) *is satisfied. Then we have*

$$\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}CI_{n-k}^{(2)}C^t\overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}}0,\tag{4.24}$$

where $I_k^{(2)}$ *is given in* (4.4).

Proof. This proof follows the same lines as that of (4.21). □

Lemma 4.8. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}CI_{n-k}^{(1)}C^t\overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}}\Lambda',\tag{4.25}$$

where

$$\Lambda'=\sum_{k=0}^{+\infty}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\big(T-\big(\sigma^2+\overline{a^2}\big)e_1e_1^t\big)C^t,$$

is the unique solution of the equation

$$\Lambda'=T-\big(\sigma^2+\overline{a^2}\big)e_1e_1^t+\frac{1}{2}\big(A\Lambda'A^t+B\Lambda'B^t\big),$$

where T is given (2.11) and $I_k^{(1)}$ is given in (4.3).

Proof. Since in the definition of $I_n^{(1)}$ given by (4.3) there are four terms, we focus only on the first term

$$\frac{a_0}{2}A\frac{H_{k-1}}{2^k}e_1^t.$$

The other terms will be treated in the same way. Using (4.29), we obtain the following decomposition:

$$\frac{a_0}{2}\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} CA\frac{H_{n-k-1}}{2^{n-k}}e_1^tC^t = T_n^{(1)} + T_n^{(2)} + T_n^{(3)},$$

where

$$T_n^{(1)} = \frac{a_0}{2}\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} CA\left\{\overline{A}^{n-k-p}\frac{H_{p-1}}{2^p} + \sum_{l=p}^{n-k-1}\overline{A}^{n-k-l-1}\frac{H_{p-1}}{2^{l+1}}\right\}e_1^tC^t,$$

$$T_n^{(2)} = \frac{a_0}{2}\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} CA\left\{\sum_{l=p}^{n-k-1}\overline{A}^{n-k-l-1}\overline{a}\left(\frac{2^l-2^{p-1}}{2^l}\right)e_1e_1^t\right\}C^t$$

and

$$T_n^{(3)} = \frac{a_0}{2}\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} CA\sum_{l=p}^{n-k-1}\overline{A}^{n-k-l-1}\frac{P_l}{2^{l+1}}e_1e_1^tC^t, \quad \text{with } P_n = \sum_{k\in\mathbb{T}_{n,p}}\varepsilon_k.$$

On the one hand, we have

$$\|T_n^{(3)}\| \le c\sum_{k=p}^{n}\beta^{n-k}\frac{|P_k|}{2^{k+1}},$$

where c is a positive constant such that $c > |a_0|\frac{1-\beta^{n-l}}{1-\beta}$ for all $n \ge l$, so that

$$\mathbb{P}\big(\|T_n^{(3)}\| > \delta\big) \le \sum_{k=p}^{n}\mathbb{P}\left(\frac{|P_k|}{|\mathbb{T}_k|+1} > \frac{2\delta}{cn\beta^{n-k}}\right).$$

We deduce again from hypothesis (N1) and in the same way that we have obtained (4.22) that

$$\mathbb{P}\left(\frac{P_k}{|\mathbb{T}_k|+1} > \frac{2\delta}{cn\beta^{n-k}}\right) \le \exp\left(-c_1\delta^2\frac{(2\beta^2)^{k+1}}{n^2\beta^{2n}}\right) \quad \forall k \ge p$$

for some positive constant c_1. It then follows as in the proof of (4.7) that

$$\mathbb{P}\big(\|T_n^{(3)}\| > \delta\big) \le \begin{cases} \exp\big(-c_1\delta^2\frac{2^{n+1}}{n^2}\big) & \text{if } \beta^2 < \frac{1}{2}, \\ n\exp\big(-c_1\delta^2\frac{2^{n+1}}{n^2}\big) & \text{if } \beta^2 = \frac{1}{2}, \\ \exp\big(-c_1\delta^2\frac{1}{n^2\beta^{2n}}\big) & \text{if } \beta^2 > \frac{1}{2}, \end{cases}$$

so that

$$T_n^{(3)} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{4.26}$$

On the other hand, we have after tedious calculations

$$\|T_n^{(1)}\| \leq \begin{cases} c\frac{\overline{X}_1}{2^{n+1}} & \text{if } \beta < \frac{1}{2}, \\ c\frac{\overline{X}_1}{\sqrt{|\mathbb{T}_n|+1}} & \text{if } \beta = \frac{1}{2}, \\ c\beta^n\overline{X}_1 & \text{if } \beta > \frac{1}{2}, \end{cases}$$

where c is a positive constant which depends on p and $|a_0|$. Next, from hypothesis (X2) and Chernoff inequality we conclude that

$$T_n^{(1)} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{4.27}$$

Furthermore, since $(T_n^{(2)})$ is a deterministic sequence, we have (see [3], Lemma A.4)

$$T_n^{(2)} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda'', \tag{4.28}$$

where

$$\Lambda'' = \sum_{k=0}^{+\infty}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\left(\frac{1}{2}a_0 A\Xi e_1^t\right)C^t$$

is the unique solution of

$$\Lambda'' = \frac{1}{2}a_0 A\Xi e_1^t + \frac{1}{2}\left(A\Lambda'' A^t + B\Lambda'' B^t\right).$$

It then follows that

$$\frac{a_0}{2}\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} CA\frac{H_{n-k-1}}{2^{n-k}}e_1^t C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda''.$$

Doing the same for the three other terms of $I_k^{(1)}$, we end the proof of Lemma 4.8. □

Proposition 4.9. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_{n,p}}\mathbb{X}_k \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Xi,$$

where Ξ *is given in* (2.9).

Proof. Let

$$H_n = \sum_{k\in\mathbb{T}_{n,p-1}}\mathbb{X}_k \quad \text{and} \quad P_n = \sum_{k\in\mathbb{T}_{n,p}}\varepsilon_k.$$

From p. 2517 in Bercu et al. [3], we have

$$\frac{H_n}{2^{n+1}} = \sum_{k=p-1}^{n}(\overline{A})^{n-k}\frac{H_{p-1}}{2^{k+1}} + \sum_{k=p}^{n}\overline{a}(\overline{A})^{n-k}\left(\frac{2^k-2^{p-1}}{2^k}\right)e_1 + \sum_{k=p}^{n}\frac{P_k}{2^{k+1}}(\overline{A})^{n-k}e_1.$$

Since the second term in the right hand side of this equality is deterministic and converges to Ξ, this proposition will be proved if we show that

$$\sum_{k=p-1}^{n} \frac{(\overline{A})^{n-k}}{2^k} H_{p-1} \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \qquad \sum_{k=p}^{n} \frac{P_k}{2^{k+1}} (\overline{A})^{n-k} e_1 \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \tag{4.29}$$

which follows by reasoning as in the proof of Proposition 4.3 (see the proof of Proposition 4.3 for more details). □

We now explain the modification in the last proofs in case 1.

Proposition 4.10. *Within the framework of case* 1, *we have the same conclusions as Propositions* 4.9 *and* 4.3 *with the sequence* (v_n) *satisfying condition* (V1).

Proof. The proof follows exactly the same lines as the proof of Propositions 4.9 and 4.3, and uses the fact that if a superexponential convergence holds with a sequence (v_n) satisfying condition (V2), then it also holds with a sequence (v_n) satisfying condition (V1). We thus obtain the first convergence of (4.29), the convergences (4.6), (4.27), (4.28) and (4.24) within the framework of case 1 with (v_n) satisfying condition (V1). Next, following the same approach as which used to obtain (4.22), we get

$$\mathbb{P}\biggl(\frac{1}{|\mathbb{T}_k|+1}\biggl|\sum_{i\in\mathbb{T}_{k,p}} (\varepsilon_i^2 - \sigma^2)\biggr| > \delta\biggr) \le \begin{cases} c_1 \exp\bigl(-c_2\delta^2|\mathbb{T}_k|\bigr) & \text{if } \delta \text{ is small enough,} \\ c_1 \exp\bigl(-c_2\delta|\mathbb{T}_k|\bigr) & \text{if } \delta \text{ is large enough,} \end{cases} \tag{4.30}$$

where c_1 and c_2 are positive constants which do not depend on δ. On the other hand, let n_0 such that for $n > n_0\delta/(n-p+1)\gamma\beta^{2(n-n_0)}$ is large enough. We have

$$\begin{aligned} &\mathbb{P}\biggl(\biggl\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\biggl(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\biggr)e_1e_1^tC^t\biggr\| > \delta\biggr) \\ &\quad\le \sum_{k=p}^{n_0-1}\mathbb{P}\biggl(\biggl|\frac{L_k}{|\mathbb{T}_k|+1} - \sigma^2\biggr| > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\biggr) + \sum_{k=n_0}^{n}\mathbb{P}\biggl(\biggl|\frac{L_k}{|\mathbb{T}_k|+1} - \sigma^2\biggr| > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\biggr). \end{aligned}$$

Now, using (4.30) with $\delta/(n-p+1)\beta^{2(n-k)}$ instead of δ and following the same approach used to obtain (4.18)–(4.20) in the two sums of the right hand side of the above inequality, we are led to

$$\begin{aligned} &\mathbb{P}\biggl(\biggl\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\biggl(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\biggr)e_1e_1^tC^t\biggr\| > \delta\biggr) \\ &\quad\le \begin{cases} c_1\exp\bigl(-\frac{c_2\delta^2 2^{n+1}}{n^2}\bigr) + c_1\exp\bigl(-\frac{c_2\delta 2^{n+1}}{n}\bigr) & \text{if } \beta \le \frac{1}{2}, \\ c_1 n\exp\bigl(-\frac{c_2\delta^2}{n^2\beta^{4n}}\bigr) + c_1\exp\bigl(-\frac{c_2\delta}{n\beta^{2n}}\bigr) & \text{if } \beta > \frac{1}{2}, \end{cases} \end{aligned}$$

and we thus obtain convergence (4.21) with (v_n) satisfying condition (V1). In the same way we obtain

$$\mathbb{P}\bigl(\bigl\|T_n^{(3)}\bigr\| > \delta\bigr) \le \begin{cases} c_1\exp\bigl(-\frac{c_2\delta^2 2^{n+1}}{n^2}\bigr) + c_1\exp\bigl(-\frac{c_2\delta 2^{n+1}}{n}\bigr) & \text{if } \beta < \frac{1}{2}, \\ c_1 n\exp\bigl(-\frac{c_2\delta 2^{n+1}}{n}\bigr) & \text{if } \beta = \frac{1}{2}, \\ c_1\exp\bigl(-\frac{c_2\delta^2}{n^2\beta^{2n}}\bigr) + c_1\exp\bigl(-\frac{c_2\delta}{n\beta^{n}}\bigr) & \text{if } \beta > \frac{1}{2}, \end{cases}$$

so that (4.26) and then (4.25) hold for (v_n) satisfying condition (V1). To reach the convergence (4.7) and the second convergence of (4.29) with (v_n) satisfying condition (V1), we follow the same procedure as before and the proof of the proposition is then complete. □

Remark 4.11. *Let us note that we can actually prove that*

$$\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} \Xi \quad \text{and} \quad \frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\mathbb{X}_k^t \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda.$$

Indeed, let $H_n = \sum_{k=2^{p-1}}^{n}\mathbb{X}_k$ *and* $P_l^{(n)} = \sum_{k=2^{r_n-l}}^{[n/2^l]}\varepsilon_k$. *We have the following decomposition*

$$\frac{H_n}{n} - \Xi = \frac{1}{n}\sum_{k\in\mathbb{T}_{r_n-1,p-1}}(\mathbb{X}_k - \Xi) + \frac{1}{n}\sum_{k=2^{r_n}}^{n}(\mathbb{X}_k - \Xi) + \frac{2^{p-1}-1}{n}\Xi.$$

On the one hand, observing that $v_n/v_{|\mathbb{T}_{r_n-1}|} < 2$, *we infer from Proposition 4.9 that*

$$\frac{1}{n}\sum_{k\in\mathbb{T}_{r_n-1,p-1}}(\mathbb{X}_k - \Xi) \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

The sequence $(\frac{2^{p-1}-1}{n}\Xi)$ *being deterministic and converging to* 0, *we deduce that*

$$\frac{2^{p-1}-1}{n}\Xi \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

On the other hand, from (2.1) *we deduce that*

$$\begin{aligned}\sum_{k=2^{r_n}}^{n}\mathbb{X}_k &= 2^{r_n-p+1}(\overline{A})^{r_n-p+1}\sum_{k=2^{p-1}}^{[n/(2^{r_n}-p+1)]}\mathbb{X}_k + 2\overline{a}\sum_{k=0}^{r_n-p}\left(\left[\frac{n}{2^k}\right] - 2^{r_n-k}+1\right)2^k(\overline{A})^k e_1\\ &\quad + \sum_{k=0}^{r_n-p}2^k(\overline{A})^k P_k^{(n)}e_1 - \sum_{k=1}^{r_n-p+1}s_k 2^{k-1}(\overline{A})^{k-1}(B\mathbb{X}_{[n/2^k]} + \eta_{[n/2^{k-1}]+1}),\end{aligned}$$

where

$$s_k = \begin{cases} 1 & \text{if } \left[\frac{n}{2^{k-1}}\right] \text{ is even},\\ 0 & \text{if } \left[\frac{n}{2^{k-1}}\right] \text{ is odd}.\end{cases}$$

Reasoning now as in the proof of Proposition 4.9, tedious but straightforward calculations lead us to

$$\frac{1}{n}\sum_{k=2^{r_n}}^{n}(\mathbb{X}_k - \Xi) \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

It then follows that

$$\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} \Xi.$$

The term $\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\mathbb{X}_k^t$ *can be dealt with in the same way.*

The rest of the paper is dedicated to the proof of our main results. We focus on the proof in case 2, and some explanations are given on how to obtain the results in case 1.

5. Proof of the main results

We start with the proof of the deviation inequalities.

5.1. *Proof of Theorem* 3.1

We begin the proof with case 2. Let $\delta > 0$ and $\ell > 0$ such that $\ell < \|\Sigma\|/(1+\delta)$. We have from (2.7)

$$\begin{aligned}\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) &= \mathbb{P}\bigg(\frac{\|M_n\|}{\|\Sigma_{n-1}\|} > \delta, \frac{\|\Sigma_{n-1}\|}{|\mathbb{T}_{n-1}|} \geq \ell\bigg) + \mathbb{P}\bigg(\frac{\|M_n\|}{\|\Sigma_{n-1}\|} > \delta, \frac{\|\Sigma_{n-1}\|}{|\mathbb{T}_{n-1}|} < \ell\bigg) \\ &\leq \mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg) + \mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \|\Sigma\| - \ell\bigg).\end{aligned}$$

Since $\ell < \|\Sigma\|/(1+\delta)$, then

$$\mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \|\Sigma\| - \ell\bigg) \leq \mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \delta\ell\bigg).$$

It then follows that

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) \leq 2\max\bigg\{\mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg), \mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \delta\ell\bigg)\bigg\}.$$

On the one hand, we have

$$\begin{aligned}\mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg) \leq \sum_{\eta=0}^{1}\bigg\{&\mathbb{P}\bigg(\bigg|\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k+\eta}\bigg| > \frac{\delta\ell}{4}\bigg) \\ &+ \mathbb{P}\bigg(\bigg\|\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k+\eta}\mathbb{X}_k\bigg\| > \frac{\delta\ell}{4}\bigg)\bigg\}.\end{aligned}$$

Now, by carrying out the same calculations as those which have permitted us to obtain Lemma 4.7 and equation (4.17), we are led to

$$\mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg) \leq \begin{cases} c_1\exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}2^n\big) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1\exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\frac{2^n}{n}\big) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1\exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\big(\frac{1}{\beta^2}\big)^n\big) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{5.1}$$

where the positive constants c_1, c_2, c_3 and c_4 depend on σ, β, γ and ϕ and $(c_3, c_4) \neq (0,0)$.
On the other hand, noticing that $\Sigma_{n-1} = I_2 \otimes S_{n-1}$, we have

$$\mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \delta\ell\bigg) \leq 2\mathbb{P}\bigg(\bigg\|\frac{S_{n-1}}{|\mathbb{T}_{n-1}|} - L\bigg\| > \frac{\delta\ell}{2}\bigg).$$

Next, from the proofs of Propositions 4.9 and 4.3, we deduce that

$$\mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \frac{\ell}{2}\bigg) \leq \begin{cases} c_1\exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\frac{2^n}{(n-1)^2}\big) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1\exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\frac{2^n}{(n-1)^3}\big) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1\exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\big(\frac{1}{(n-1)^2\beta^{2n}}\big)\big) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{5.2}$$

where the positive constants c_1, c_2, c_3 and c_4 depend on σ, β, γ and ϕ and $(c_3, c_4) \neq (0,0)$. Now, (3.1) follows from (5.1) and (5.2).

In case 1, the proof follows exactly the same lines as before and uses the same ideas as the proof of Proposition 4.10. In particular, we have in this case

$$\mathbb{P}\left(\left\| \frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma \right\| > \frac{\ell}{2}\right) \leq \begin{cases} c_1 \exp\left(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta < \frac{1}{2}, \\ c_1(n-1)\exp\left(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta = \frac{1}{2}, \\ c_1(n-1)\exp\left(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)} \left(\frac{1}{(n-1)\beta^n}\right)\right) & \text{if } \beta > \frac{1}{2}, \end{cases}$$

where the positive constants c_1, c_2 and c_3 depend on σ, β, γ and ϕ. (3.1) then follows in this case, and this ends the proof of Theorem 3.1.

5.2. *Proof of Theorem* 3.7

First we need to prove the following

Theorem 5.1. *In case* 1 *or in case* 2, *the sequence* $(M_n/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}^{2(p+1)}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_M(x) = \sup_{\lambda\in\mathbb{R}^{2(p+1)}} \left\{\lambda^t x - \lambda^t(\Gamma\otimes L)\lambda\right\} = \frac{1}{2}x^t(\Gamma\otimes L)^{-1}x. \tag{5.3}$$

5.2.1. *Proof of Theorem* 5.1

Since the size of the data doubles at each generation, we are not able to verify the Lindeberg condition. To come over this problem, and as in Bercu et al. [3], p. 2510, we change the filtration and we will use the sister pair-wise one, that is, $(\mathcal{G}_n)_{n\geq 1}$ given by $\mathcal{G}_n = \sigma\{X_1, (X_{2k}, X_{2k+1}), 1 \leq k \leq n\}$. We introduce the following $(\mathcal{G}_n)$ martingale difference sequence (D_n), given by

$$D_n = V_n \otimes Y_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n}\mathbb{X}_n \\ \varepsilon_{2n+1} \\ \varepsilon_{2n+1}\mathbb{X}_n \end{pmatrix}.$$

We clearly have

$$D_n D_n^t = V_n V_n^t \otimes Y_n Y_n^t.$$

So we obtain that the quadratic variation of the $(\mathcal{G}_n)$ martingale $(N_n)_{n\geq 2^{p-1}}$ given by

$$N_n = \sum_{k=2^{p-1}}^{n} D_k$$

is

$$\langle N\rangle_n = \sum_{k=2^{p-1}}^{n} \mathbb{E}\left(D_k D_k^t/\mathcal{G}_{k-1}\right) = \Gamma \otimes \sum_{k=2^{p-1}}^{n} Y_k Y_k^t.$$

Now we clearly have $M_n = N_{|\mathbb{T}_{n-1}|}$ and $\langle M\rangle_n = \langle N\rangle_{|\mathbb{T}_{n-1}|} = \Gamma\otimes S_{n-1}$. From Proposition 4.1, and since $\langle M\rangle_n = \Gamma\otimes S_{n-1}$, we have

$$\frac{\langle M\rangle_n}{|\mathbb{T}_n|} \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \Gamma\otimes L. \tag{5.4}$$

Before going to the proof of the MDP results, we state the exponential Lyapounov condition for $(N_n)_{n\geq 2^{p-1}}$, which implies exponential Lindeberg condition, that is

$$\limsup \frac{1}{v_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}\mathbb{E}\big[\|D_k\|^2\mathbf{1}_{\{\|D_k\|\geq r(\sqrt{n}/v_n)\}}\big]\geq\delta\right)=-\infty$$

(see Remark 3, p. 10, in [25] for more details on this implication).

Remarks 5.2. *By* [14], *we infer from the condition* (Ea) *that*

(Na) *one can find $\gamma_a>0$ such that for all $n\geq p-1$, for all $k\in\mathbb{G}_{n+1}$ and for all $t\in\mathbb{R}$, with $\mu_a=\mathbb{E}(|\varepsilon_k|^a/\mathcal{F}_n)$ a.s.*

$$\mathbb{E}\big[\exp t\big(|\varepsilon_k|^a-\mu_a\big)/\mathcal{F}_n\big]\leq\exp\left(\frac{\gamma_a t^2}{2}\right)\quad a.s.$$

Proposition 5.3. *Let (v_n) be a sequence satisfying assumption* (V2). *Assume that hypotheses* (Na) *and* (Xa) *are satisfied. Then there exists $B>0$ such that*

$$\limsup_{n\to\infty}\frac{1}{v_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{j=2^{p-1}}^{n}\mathbb{E}\big[\|D_j\|^a/\mathcal{G}_{j-1}\big]>B\right)=-\infty.$$

Proof. We are going to prove that

$$\limsup_{n\to\infty}\frac{1}{v_{|\mathbb{T}_n|}^2}\log\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|}\sum_{j=2^p}^{|\mathbb{T}_n|}\mathbb{E}\big[\|D_j\|^a/\mathcal{G}_{j-1}\big]>B\right)=-\infty,\tag{5.5}$$

and Proposition 5.3 will follow by proceeding as in Remark 4.11. We have

$$\sum_{j\in\mathbb{T}_{n,p}}\mathbb{E}\big[\|D_j\|^a/\mathcal{G}_{j-1}\big]\leq c\mu^a\sum_{j\in\mathbb{T}_{n,p}}\big(1+\|\mathbb{X}_j\|^a\big),$$

where c is a positive constant which depends on a. From (2.1), we deduce that

$$\sum_{j\in\mathbb{T}_{n,p}}\|\mathbb{X}_j\|^a\leq\frac{c^2}{(1-\beta)^{a-1}}P_n+\frac{c^2\alpha^a Q_n}{(1-\beta)^{a-1}}+2cR_n\overline{X}_1^a,$$

where

$$P_n=\sum_{j\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_j-p}\beta^i|\varepsilon_{[j/2^i]}|^a,\qquad Q_n=\sum_{j\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_j-p}\beta^i,\qquad R_n=\sum_{j\in\mathbb{T}_{n,p}}\beta^{a(r_j-p+1)},$$

and c is a positive constant. Now, proceeding as in the proof of Proposition 4.3, using hypotheses (Na) and (Xa) instead of (N2) and (X2), we get for B large enough

$$\limsup_{n\to\infty}\frac{1}{v_{|\mathbb{T}_n|}^2}\log\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|}\sum_{j\in\mathbb{T}_{n,p}}\|\mathbb{X}_j\|^a>B\right)=-\infty.\tag{5.6}$$

Now (5.6) leads us to (5.5) and following the same approach as in Remark 4.11, we obtain Proposition 5.3. □

Remarks 5.4. *In case* 1, *we clearly have that* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$, *where*

$$\mathbb{T}_{\cdot,p-1} = \bigcup_{r=p-1}^{\infty} \mathbb{G}_r,$$

is a bifurcating Markov chain with initial state $\mathbb{X}_{2^{p-1}} = (X_{2^{p-1}}, X_{2^{p-2}}, \dots, X_1)^t$. *Let* ν *be the law of* $\mathbb{X}_{2^{p-1}}$. *From hypothesis* (X2), *we deduce that* ν *has finite moments of all orders. We denote by* P *the transition probability kernel associated to* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$. *Let* $(\mathbb{Y}_r, r \in \mathbb{N})$ *the ergodic stable Markov chain associated to* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$. *This Markov chain is defined as follows, starting from the root* $\mathbb{Y}_0 = \mathbb{X}_{2^{p-1}}$ *and if* $\mathbb{Y}_r = \mathbb{X}_n$ *then* $\mathbb{Y}_{r+1} = \mathbb{X}_{2n+\zeta_{r+1}}$ *for a sequence of independent Bernoulli r.v.* $(\zeta_q, q \in \mathbb{N}^*)$ *such that* $\mathbb{P}(\zeta_q = 0) = \mathbb{P}(\zeta_q = 1) = 1/2$.

Let μ *be the stationary distribution associated to* $(\mathbb{Y}_r, r \in \mathbb{N})$. *For more details on bifurcating Markov chain and the associated ergodic stable Markov chain, we refer to* [18] (*see also* [5]).

From [5], *we deduce that for all real bounded function* f *defined on* $(\mathbb{R}^p)^3$,

$$\frac{1}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1,p-1}} f(\mathbb{X}_k, \mathbb{X}_{2k}, \mathbb{X}_{2k+1})$$

satisfies a MDP on $\mathbb{R}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and the rate function* $I(x) = \frac{x^2}{2S^2(f)}$, *where* $S^2(f) = \langle \mu, P(f^2) - (Pf)^2 \rangle$.

Now, let f *be the function defined on* $(\mathbb{R}^p)^3$ *by* $f(x,y,z) = \|x\|^2 + \|y\|^2 + \|z\|^2$. *Then, using the relation* (4.1) *in Proposition* 4.1, *the above MDP for real bounded functionals of the bifurcating Markov chain* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$ *and the truncation of the function* f, *we prove* (*in the same manner as the proof of Lemma* 3 *in Worms* [25]) *that for all* $r > 0$

$$\limsup_{R\to\infty} \limsup_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\Bigg(\frac{1}{n} \sum_{j=2^{p-1}}^{n} \big(\|X_j\|^2 + \|X_{2j}\|^2 + \|X_{2j+1}\|^2\big) \times \mathbf{1}_{\{\|\mathbb{X}_j\| + \|\mathbb{X}_{2j}\| + \|\mathbb{X}_{2j+1}\| > R\}} > r \Bigg) = -\infty,$$

which implies the following Lindeberg condition (*for more details, we refer to Proposition* 2 *in Worms* [25])

$$\limsup_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\Bigg(\frac{1}{n} \sum_{j=2^{p-1}}^{n} \big(\|X_j\|^2 + \|X_{2j}\|^2 + \|X_{2j+1}\|^2\big) \times \mathbf{1}_{\{\|\mathbb{X}_j\| + \|\mathbb{X}_{2j}\| + \|\mathbb{X}_{2j+1}\| > r(\sqrt{n}/v_n)\}} > \delta \Bigg) = -\infty$$

for all $\delta > 0$ *and for all* $r > 0$. *Notice that the above Lindeberg condition implies in particular the Lindeberg condition on the sequence* $(\mathbb{X}_n)$.

Now, we come back to the proof of Theorem 5.1. We divide the proof into four steps. In the first one, we introduce a truncation of the martingale $(M_n)_{n\geq 0}$ and prove that the truncated martingale satisfies some MDP thanks to Puhalskii's Theorem 3.12. In the second part, we show that the truncated martingale is an exponentially good approximation of (M_n), see e.g. Definition 4.2.14 in [12]. We conclude by the identification of the rate function.

Proof in case 2. *Step 1.* From now on, in order to apply Puhalskii's result [24] (Puhalskii's Theorem 3.12) for the MDP for martingales, we introduce the following truncation of the martingale $(M_n)_{n\geq 0}$. For $r > 0$ and $R > 0$,

$$M_n^{(r,R)} = \sum_{k \in \mathbb{T}_{n-1,p-1}} D_{k,n}^{(r,R)},$$

where, for all $1 \le k \le n$, $D_{k,n}^{(r,R)} = V_k^{(R)} \otimes Y_{k,n}^{(r)}$, with

$$V_n^{(R)} = \big(\varepsilon_{2n}^{(R)}, \varepsilon_{2n+1}^{(R)}\big)^t \quad \text{and} \quad Y_{k,n}^{(r)} = \big(1, \mathbb{X}_{k,n}^{(r)}\big)^t,$$

where

$$\varepsilon_k^{(R)} = \varepsilon_k \mathbf{1}_{\{|\varepsilon_k| \le R\}} - \mathbb{E}[\varepsilon_k \mathbf{1}_{\{|\varepsilon_k| \le R\}}], \qquad \mathbb{X}_{k,n}^{(r)} = \mathbb{X}_k \mathbf{1}_{\{\|\mathbb{X}_k\| \le r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}}.$$

We introduce $\Gamma^{(R)}$ the conditional covariance matrix associated with $(\varepsilon_{2k}^{(R)}, \varepsilon_{2k+1}^{(R)})^t$ and the truncated matrix associated with S_n:

$$\Gamma^{(R)} = \begin{pmatrix} \sigma_R^2 & \rho_R \\ \rho_R & \sigma_R^2 \end{pmatrix} \quad \text{and} \quad S_n^{(r)} = \sum_{k \in \mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & (\mathbb{X}_{k,n}^{(r)})^t \\ \mathbb{X}_{k,n}^{(r)} & \mathbb{X}_{k,n}^{(r)} (\mathbb{X}_{k,n}^{(r)})^t \end{pmatrix}.$$

The condition (P2) in Puhalskii's Theorem 3.12 is verified by the construction of the truncated martingale, that is for some positive constant c, we have that for all $k \in \mathbb{T}_{n-1}$

$$\big\| D_{k,n}^{(r,R)} \big\| \le c \frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}.$$

From Proposition 5.3, we also have for all $r > 0$,

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \mathbb{X}_k \mathbf{1}_{\{\|\mathbb{X}_k\| > r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0; \tag{5.7}$$

and

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \mathbb{X}_k \mathbb{X}_k^t \mathbf{1}_{\{\|\mathbb{X}_k\| > r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.8}$$

From (5.7) and (5.8), we deduce that for all $r > 0$

$$\frac{1}{|\mathbb{T}_{n-1}|} \big(S_{n-1} - S_{n-1}^{(r)}\big) \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.9}$$

Then, we easily transfer the properties (5.4) to the truncated martingale $(M_n^{(r,R)})_{n \ge 0}$. We have for all $R > 0$ and all $r > 0$,

$$\frac{\langle M^{(r,R)} \rangle_n}{|\mathbb{T}_{n-1}|} = \Gamma^{(R)} \otimes \frac{S_{n-1}^{(r)}}{|\mathbb{T}_{n-1}|} = -\Gamma^{(R)} \otimes \left(\frac{S_{n-1} - S_{n-1}^{(r)}}{|\mathbb{T}_{n-1}|} \right) + \Gamma^{(R)} \otimes \frac{S_{n-1}}{|\mathbb{T}_{n-1}|} \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} \Gamma^{(R)} \otimes L.$$

That is condition (P1) in Puhalskii's Theorem 3.12.

Note also that Proposition 5.3 works for the truncated martingale $(M_n^{(r,R)})_{n \ge 0}$, which ensures Lindeberg's condition and thus condition (P3) for $(M_n^{(r,R)})_{n \ge 0}$. By Theorem 3.12, we deduce that $(M_n^{(r,R)}/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))_{n \ge 0}$ satisfies a MDP on $\mathbb{R}^{2(p+1)}$ with speed $v_{|\mathbb{T}_{n-1}|}^2$ and good rate function given by

$$I_R(x) = \frac{1}{2} x^t \big(\Gamma^{(R)} \otimes L\big)^{-1} x. \tag{5.10}$$

Step 2. First, we infer from the hypothesis (Ea) that:

(N1R) there is a sequence $(\kappa_R)_{R>0}$ with $\kappa_R \longrightarrow 0$ when R goes to infinity, such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$, for all $t \in \mathbb{R}$ and for R large enough

$$\mathbb{E}\big[\exp t\big(\varepsilon_k - \varepsilon_k^R\big)/\mathcal{F}_n\big] \leq \exp\left(\frac{\kappa_R t^2}{2}\right) \quad \text{a.s.}$$

Then, we have to prove that for all $r > 0$ the sequence $(M_n^{(r,R)})_n$ is an exponentially good approximation of (M_n) as R goes to infinity, see e.g. Definition 4.2.14 in [12]. This approximation in the sense of the moderate deviation, is described by the following convergence, for all $r > 0$ and all $\delta > 0$,

$$\limsup_{R\to\infty} \limsup_{n\to\infty} \frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{\|M_n - M_n^{(r,R)}\|}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} > \delta\right) = -\infty.$$

For that, we shall prove that for $\eta \in \{0, 1\}$

$$I_1 = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k+\eta} - \varepsilon_{2k+\eta}^{(R)}\big) \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{5.11}$$

$$I_2 = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k+\eta}\mathbb{X}_k - \varepsilon_{2k+\eta}^{(R)}\mathbb{X}_{k,n}^{(r)}\big) \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.12}$$

We need only prove (5.11) and (5.12) for $\eta = 0$, the same proof works for $\eta = 1$.

Proof of (5.11). We have, for all $\alpha > 0$ and R large enough

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(\alpha \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\right)\right) \\
&\quad = \mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}} \exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big) \times \mathbb{E}\left[\prod_{k\in\mathbb{G}_{n-1}} \exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big)\Big/\mathcal{F}_{n-1}\right]\right] \\
&\quad = \mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}} \exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big) \times \prod_{k\in\mathbb{G}_{n-1}} \mathbb{E}\big[\exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big)/\mathcal{F}_{n-1}\big]\right] \\
&\quad \leq \mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}} \exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big) \exp\big(|\mathbb{G}_{n-1}|\alpha^2\kappa_R\big)\right] \\
&\quad \leq \exp\big(|\mathbb{T}_{n-1}|\alpha^2\kappa_R\big),
\end{aligned}$$

where hypothesis (N1R) was used to get the first inequality, and the second was obtained by induction. By Chebyshev inequality and the previous calculation applied to $\alpha = \lambda v_{|\mathbb{T}_{n-1}|}/|\mathbb{T}_{n-1}|$, we obtain for all $\delta > 0$

$$\mathbb{P}\left(\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big) \geq \delta\right) \leq \exp\big(-v_{|\mathbb{T}_{n-1}|}^2\big(\delta\lambda - \kappa_R\lambda^2\big)\big).$$

Optimizing on λ, we obtain

$$\frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big) \geq \delta\right) \leq -\frac{\delta^2}{4\kappa_R}.$$

Letting n go to infinity and then R go to infinity, we obtain the negligibility in (5.11).

Proof of (5.12). Now, since we have the decomposition

$$\varepsilon_{2k}\mathbb{X}_k - \varepsilon_{2k}^{(R)}\mathbb{X}_{k,n}^{(r)} = \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\mathbb{X}_{k,n}^{(r)} + \varepsilon_{2k}\big(\mathbb{X}_k - \mathbb{X}_{k,n}^{(r)}\big),$$

we introduce the following notation

$$L_n^{(r)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k}\big(\mathbb{X}_k - \mathbb{X}_{k,n}^{(r)}\big) \quad \text{and} \quad F_n^{(r,R)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\mathbb{X}_{k,n}^{(r)}.$$

To prove (5.12), we will show that for all $r > 0$

$$\frac{L_n^{(r)}}{\sqrt{|\mathbb{T}_{n-1}|}\,v_{|\mathbb{T}_{n-1}|}} \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{5.13}$$

and for all $r > 0$ and all $\delta > 0$

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log\mathbb{P}\bigg(\frac{\|F_n^{(r,R)}\|}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\bigg) = -\infty. \tag{5.14}$$

Let us first deal with $(L_n^{(r)})$. Let its first component be

$$L_{n,1}^{(r)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k}\big(X_k - X_{k,n}^{(r)}\big).$$

For $\lambda\in\mathbb{R}$, we consider the random sequence $(Z_{n,1}^{(r)})_{n\geq p-1}$ defined by

$$Z_{n,1}^{(r)} = \exp\bigg(\lambda L_{n,1}^{(r)} - \frac{\lambda^2\phi}{2}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2\mathbf{1}_{\{\|\mathbb{X}_k\|>r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}}\bigg)$$

where ϕ appears in (N1). For $h > 0$, we introduce the following event

$$A_{n,1}^{(r)}(h) = \bigg\{\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2\mathbf{1}_{\{\|\mathbb{X}_k\|>r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} > h\bigg\}.$$

Using (N1), we have for all $\delta > 0$

$$\begin{aligned}
&\mathbb{P}\bigg(\frac{1}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}}L_{n,1}^{(r)} > \delta\bigg)\\
&\quad\leq \mathbb{P}\big(A_{n,1}^{(r)}(h)\big) + \mathbb{P}\bigg(Z_{n,1}^{(r)} > \exp\bigg(\delta\lambda v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} - \frac{\lambda^2\phi}{2}h|\mathbb{T}_{n-1}|\bigg)\bigg)\\
&\quad\leq \mathbb{P}\big(A_{n,1}^{(r)}(h)\big) + \exp\bigg(-v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\bigg(\delta\lambda - \frac{h\phi\sqrt{|\mathbb{T}_{n-1}|}}{2v_{|\mathbb{T}_{n-1}|}}\lambda^2\bigg)\bigg),
\end{aligned} \tag{5.15}$$

where the second term in (5.15) is obtained by conditioning successively on $(\mathcal{G}_i)_{2^{p-1}\leq i\leq|\mathbb{T}_{n-1}|-1}$ and using the fact that

$$\mathbb{E}\bigg[\exp\bigg(\lambda\varepsilon_{2^p}\big(X_{2^{p-1}} - X_{2^{p-1}}^{(r)}\big) - \frac{\lambda^2\phi}{2}X_{2^{p-1}}^2\mathbf{1}_{\{\|\mathbb{X}_{2^{p-1}}\|>r(\sqrt{2^{p-1}}/v_{2^{p-1}})\}}\bigg)\bigg] \leq 1,$$

which follows from (N1).

From Proposition 5.3, we have for all $h > 0$

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\big(A^{(r)}_{n,1}(h)\big) = -\infty,$$

so that taking $\lambda = \delta v_{|\mathbb{T}_{n-1}|}/(h\phi\sqrt{|\mathbb{T}_{n-1}|})$ in (5.15), we are led to

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\bigg(\frac{L^{(r)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\bigg) \le -\frac{\delta^2}{2h\phi}.$$

Letting $h \to 0$, we obtain that the right hand side of the last inequality goes to $-\infty$.

Proceeding in the same way for $-L^{(r)}_{n,1}$, we deduce that for all $r > 0$

$$\frac{L^{(r)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

Now, it is easy to check that the same proof works for the others components of $L^{(r)}_n$. We thus conclude the proof of (5.13).

Let us now consider the term $(F^{(r,R)}_n)$. We follow the same approach as in the proof of (5.13). Let its first component be

$$F^{(r,R)}_{n,1} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big) X^{(r)}_{k,n}.$$

For $\lambda \in \mathbb{R}$, we consider the random sequence $(W^{(r,R)}_{n,1})_{n\ge p-1}$ defined by

$$W^{(r,R)}_{n,1} = \exp\bigg(\lambda \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big) X^{(r)}_{k,n} - \frac{\lambda^2\kappa_R}{2} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(X^{(r)}_{k,n}\big)^2\bigg),$$

where κ_R appears in (N1R).

Let $h > 0$. Consider the following event $B^{(r)}_{n,1}(h) = \{\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} (X^{(r)}_{k,n})^2 > h\}$.

We have for all $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\bigg(\frac{F^{(r,R)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\bigg) \\
&\quad\le \mathbb{P}\big(B^{(r)}_{n,1}(h)\big) + \mathbb{P}\bigg(W^{(r,R)}_{n,1} > \exp\bigg(\delta\lambda v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} - \frac{\lambda^2\kappa_R}{2}|\mathbb{T}_{n-1}|h\bigg)\bigg) \\
&\quad\le \mathbb{P}\big(B^{(r)}_{n,1}(h)\big) + \exp\bigg(-v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\bigg(\delta\lambda - \frac{h\kappa_R\sqrt{|\mathbb{T}_{n-1}|}}{2v_{|\mathbb{T}_{n-1}|}}\lambda^2\bigg)\bigg),
\end{aligned} \tag{5.16}$$

where the second term in (5.16) is obtained by conditioning successively on $(\mathcal{G}_i)_{2^{p-1}\le i\le|\mathbb{T}_{n-1}|-1}$ and using the fact that

$$\mathbb{E}\bigg[\exp\bigg(\lambda\big(\varepsilon_{2^p} - \varepsilon^{(R)}_{2^p}\big)X^{(r)}_{2^{p-1}} - \frac{\lambda^2\kappa_R}{2}\big(X^{(r)}_{2^{p-1}}\big)^2\bigg)\bigg] \le 1.$$

Since $B^{(r)}_{n,1}(h) \subset \{\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} X^2_k > h\}$, from Proposition 4.3, we deduce that for h large enough

$$\limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\big(B^{(r)}_{n,1}(h)\big) = -\infty,$$

so that choosing $\lambda = \delta v_{|\mathbb{T}_{n-1}|}/(\kappa_R h\sqrt{|\mathbb{T}_{n-1}|})$, we get for all $\delta > 0$

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{F^{(r,R)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \le -\frac{\delta^2}{2\kappa_R h}.$$

Letting R go to infinity, we obtain that

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{F^{(r,R)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) = -\infty.$$

Now it is easy to check that the same works for $-F^{(r,R)}_{n,1}$ and for the others components of $F^{(r,R)}_n$. We thus conclude that (5.14) holds for all $r > 0$.

Step 3. By application of Theorem 4.2.16 in [12], we find that $(M_n/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))$ satisfies an MDP on $\mathbb{R}^{2(p+1)}$ with speed $v^2_{|\mathbb{T}_{n-1}|}$ and rate function

$$\widetilde{I}(x) = \sup_{\delta>0}\liminf_{R\to\infty}\inf_{z\in B_{x,\delta}} I_R(z),$$

where I_R is given in (5.10) and $B_{x,\delta}$ denotes the ball $\{z\colon |z-x| < \delta\}$. The identification of the rate function $\widetilde{I} = I_M$, where I_M is given in (5.3) is done easily (see for example [15]), which concludes the proof of Theorem 5.1. □

Proof in case 1. For the proof in case 1, there are no changes in Step 1, and for Step 3, instead of (5.7), (5.8), and (N1), we use Remark 5.4 and (G1). In Step 2, the negligibility in (5.11) comes from the MDP of the i.i.d. sequences $(\varepsilon_{2k} - \varepsilon^{(R)}_{2k})$ since it satisfies the condition, for $\lambda > 0$ and all $R > 0$

$$\mathbb{E}\big(\exp\big(\lambda\big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big)\big)\big) < \infty.$$

The negligibility of $(L^{(r)}_n)$ works in the same way. For $(F^{(r,R)}_n)$ we will use the MDP for martingale, see Proposition 3.11. For R large enough, we have

$$\begin{aligned}\mathbb{P}\big(\big|X^{(r)}_{k,n}\big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big)\big| > v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\big|\mathcal{F}_{k-1}\big) &\le \mathbb{P}\left(\big|\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big| > \frac{v^2_{|\mathbb{T}_{n-1}|}}{r}\right)\\ &= \mathbb{P}\left(\big|\varepsilon_{2} - \varepsilon^{(R)}_{2}\big| > \frac{v^2_{|\mathbb{T}_{n-1}|}}{r}\right) = 0.\end{aligned}$$

This implies that

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log\Big(|\mathbb{T}_{n-1}| \operatorname*{ess\,sup}_{k\ge1} \mathbb{P}\big(\big|X^{(r)}_{k,n}\big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big)\big| > v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\big|\mathcal{F}_{k-1}\big)\Big) = -\infty.$$

That is condition (D2) in Proposition 3.11.

For all $\gamma > 0$ and all $\delta > 0$, we obtain from Remark 5.4, that

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \big(X^{(r)}_{k,n}\big)^2 \mathbf{1}_{\{|X^{(r)}_{k,n}|>\gamma(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} > \delta\right)\\ &\quad\le \limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} X^2_k \mathbf{1}_{\{|X_k|>\gamma(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} > \delta\right) = -\infty.\end{aligned}$$

That is condition (D3) in Proposition 3.11. Finally, from Remark 5.4 and in the same way as in (5.9), it follows that

$$\frac{\langle F^{(r,R)}\rangle_{n,1}}{|\mathbb{T}_{n-1}|} = Q_R \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(X_{k,n}^{(r)}\big)^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} Q_R \ell$$

for some positive constant ℓ, where $Q_R = \mathbb{E}[(\varepsilon_2 - \varepsilon_2^{(R)})^2]$. That is condition (D1) in Proposition 3.11. Moreover, it is clear that Q_R converges to 0 as R goes to infinity. In light of above, we infer from Proposition 3.11 that $(F_{n,1}^{(r,R)}/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))$ satisfies an MDP on $\mathbb{R}$ of speed $v^2_{|\mathbb{T}_{n-1}|}$ and rate function $I_R(x) = x^2/(2Q_R\ell)$. In particular, this implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{|F_{n,1}^{(r,R)}|}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \le -\frac{\delta^2}{2Q_R\ell},$$

and letting R go to infinity clearly leads to the result. □

5.2.2. *Proof of Theorem* 3.5

The proof works in case 1 and in case 2. From (2.7), we have

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}(\hat{\theta}_n - \theta) = |\mathbb{T}_{n-1}|\Sigma_{n-1}^{-1} \frac{M_n}{v_{|\mathbb{T}_{n-1}|}|\mathbb{T}_{n-1}|}.$$

From Proposition 4.1, we obtain that

$$\frac{\Sigma_n}{|\mathbb{T}_n|} = I_2 \otimes \frac{S_n}{|\mathbb{T}_n|} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} I_2 \otimes L. \tag{5.17}$$

According to Lemma 4.1 of [26], together with (5.17), we deduce that

$$|\mathbb{T}_{n-1}|\Sigma_{n-1}^{-1} \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} I_2 \otimes L^{-1}. \tag{5.18}$$

From Theorem 5.1, (5.18) and the contraction principle [12], we deduce that the sequence $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\theta}_n - \theta)/v_{|\mathbb{T}_{n-1}|})_{n\ge1}$ satisfies the MDP with rate function I_θ given by (3.3).

5.3. *Proof of Theorem* 3.7

For the proof of Theorem 3.7, case 1 is an easy consequence of the classical MDP for i.i.d.r.v. applied to the sequence $(\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2)$. For case 2, we will use Proposition 3.11, rather than Puhalskii's Theorem 3.12.

We will prove that the sequence$(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/v_{|\mathbb{T}_{n-1}|})$ satisfies the MDP. For that, we will prove that conditions (D1), (D2) and (D3) of Proposition 3.11 are verified. Let us consider the $\mathcal{G}_n$-martingale $(Q_n)_{n\ge 2^{p-1}}$ given by

$$Q_n = \sum_{k=2^{p-1}}^{n} \nu_k, \quad \text{where } \nu_k = \varepsilon_{2k}^2 + \varepsilon_{2k+1}^2 - 2\sigma^2.$$

It is easy to see that its predictable quadratic variation is given by

$$\langle Q\rangle_n = \sum_{k=2^{p-1}}^{n} \mathbb{E}\big[\nu_k^2/\mathcal{G}_{k-1}\big] = \big(n - 2^{p-1} + 1\big)\big(2\tau^4 - 4\sigma^4 + 2\nu^2\big),$$

which immediately implies that

$$\frac{\langle Q\rangle_n}{n} \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 2\tau^4 - 4\sigma^4 + 2\nu^2,$$

ensuring condition (D1) in Proposition 3.11.

Next, for $B > 0$ large enough, we have for $a > 2$ (in (Ea)), and some positive constant c

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\nu_k|^a > B\right) \le 3 \max_{\eta\in\{0,1\}}\left\{\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\varepsilon_{2k+\eta}|^{2a} > \frac{B}{3c}\right)\right\}.$$

From hypothesis (Ea) and since B is large enough, we obtain for a suitable $t > 0$ via the Chernoff inequality and several successive conditionings on $(\mathcal{G}_n)$, for $\eta \in \{0, 1\}$

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\varepsilon_{2k+\eta}|^{2a} > \frac{B}{3c}\right) \le \exp\left(-tn\left(\frac{B}{3c} - \log E\right)\right) \le \exp\bigl(-tc'n\bigr),$$

where c, c' are positive generic constants. Therefore, for $B > 0$ large enough, we deduce that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\nu_k|^a > B\right) < 0,$$

and this implies (see e.g. [26]) exponential Lindeberg condition, that is for all $r > 0$

$$\frac{1}{n}\sum_{k=2^{p-1}}^{n}\nu_k^2\mathbf{1}_{\{|\nu_k|>r(\sqrt{n}/v_n)\}} \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} 0.$$

That is condition (D3) in Proposition 3.11.

Now, for all $k \in \mathbb{N}$ and a suitable $t > 0$ we have

$$\begin{aligned}
\mathbb{P}\bigl(|\nu_k| > v_n\sqrt{n}/\mathcal{G}_{k-1}\bigr) &\le \sum_{\eta=0}^{1}\mathbb{P}\left(\bigl|\varepsilon_{2k+\eta}^2 - \sigma^2\bigr| > \frac{v_n\sqrt{n}}{2}\Big/\mathcal{G}_{k-1}\right)\\
&\le \exp\left(\frac{-tv_n\sqrt{n}}{2}\right)\sum_{\eta=0}^{1}\mathbb{E}\bigl[\exp\bigl(t\bigl|\varepsilon_{2k+\eta}^2 - \sigma^2\bigr|\bigr)/\mathcal{G}_{k-1}\bigr]\\
&\le 2E'\exp\left(\frac{-tv_n\sqrt{n}}{2}\right),
\end{aligned}$$

where from hypothesis (Na), E' is finite and positive. We are thus led to

$$\frac{1}{v_n^2}\log\Bigl(n\operatorname*{ess\,sup}_{k\in\mathbb{N}^*}\mathbb{P}\bigl(|\nu_k| > v_n\sqrt{n}/\mathcal{G}_{k-1}\bigr)\Bigr) \le \frac{\log(2E'n)}{v_n^2} - \frac{t\sqrt{n}}{v_n},$$

and consequently, letting n go to infinity, we get the condition (D2) in Proposition 3.11.

Now, applying Proposition 3.11, we conclude that $(Q_n/(v_n\sqrt{n}))_{n\ge 0}$ satisfies the MDP with speed v_n^2 and rate function

$$I_Q(x) = \frac{x^2}{4(\tau^4 - 2\sigma^4 + 2\nu^2)}.$$

Applying the above to $|\mathbb{T}_{n-1}|$ and using the contraction principle (see e.g. [12]), we deduce that the sequence

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}\bigl(\sigma_n^2 - \sigma^2\bigr) = \frac{Q_{|\mathbb{T}_{n-1}|}}{2v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}}$$

satisfies a MDP with speed $v_{|\mathbb{T}_{n-1}|}^2$ and rate function I_{σ^2} given by (3.4).

We obtain as in the proof of the first part, with a slight modification, that the sequence $(|\mathbb{T}_{n-1}|(\rho_n - \rho)/v_{|\mathbb{T}_{n-1}|})$ satisfies a MDP with speed $v^2_{|\mathbb{T}_{n-1}|}$ and rate function I_ρ given by (3.5).

5.4. *Proof of Theorem* 3.10

Here also the proof works for the two cases.

Let us first deal with $\hat{\sigma}_n$. We have

$$\hat{\sigma}_n^2 - \sigma^2 = \left(\hat{\sigma}_n^2 - \sigma_n^2\right) + \left(\sigma_n^2 - \sigma^2\right).$$

From (4.22) and (4.30), we easily deduce that $\sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2$ in case 1 and in case 2. Thus, it is enough to prove that $\hat{\sigma}_n^2 - \sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0$. Let $\theta^{(0)} = (a_0, a_1, \dots, a_p)^t$, $\theta^{(1)} = (b_0, b_1, \dots, b_p)^t$, $\hat{\theta}_n^{(0)} = (\hat{a}_{0,n}, \hat{a}_{1,n}, \dots, \hat{a}_{p,n})^t$, $\hat{\theta}_n^{(1)} = (\hat{b}_{0,n}, \hat{b}_{1,n}, \dots, \hat{b}_{p,n})^t$.

Let us introduce the following function f defined for x and z in $\mathbb{R}^{p+1}$ by

$$f(x, z) = \left(x_1 - z_1 - \sum_{i=2}^{p+1} z_i x_i\right)^2,$$

where x_i and z_i denote respectively the ith component of x and z. One can observe that

$$\begin{aligned}\hat{\sigma}_n^2 - \sigma_n^2 &= \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left\{ f\left(\mathbb{X}_{2k}, \hat{\theta}_n^{(0)}\right) - f\left(\mathbb{X}_{2k}, \theta^{(0)}\right)\right\} \\ &\quad + \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left\{ f\left(\mathbb{X}_{2k+1}, \hat{\theta}_n^{(1)}\right) - f\left(\mathbb{X}_{2k+1}, \theta^{(1)}\right)\right\}.\end{aligned}$$

By the Taylor–Lagrange formula, $\forall x \in \mathbb{R}^{p+1}$ and $\forall z, z' \in \mathbb{R}^{p+1}$, one can find $\lambda \in (0, 1)$ such that

$$f\left(x, z'\right) - f(x, z) = \sum_{j=1}^{p+1} \left(z'_j - z_j\right) \partial_{z_j} f\left(x, z + \lambda\left(z' - z\right)\right).$$

Let the function g be defined by

$$g(x, z) = x_1 - z_1 - \sum_{j=2}^{p+1} z_j x_j.$$

Observing that

$$\begin{cases} \frac{\partial f}{\partial z_1}(x, z) = -2g(x, z), \\ \frac{\partial f}{\partial z_j}(x, z) = -2x_j g(x, z) \quad \forall j \geq 2, \end{cases}$$

we get easily that $|\frac{\partial f}{\partial z_j}(x, z)| \leq 4(1 + \|z\|)(1 + \|x\|^2)$ for all $j \geq 1$, and this implies

$$\left|f\left(x, z'\right) - f(x, z)\right| \leq c\left\|z' - z\right\|\left(1 + \|z\| + \left\|z' - z\right\|\right)\left(1 + \|x\|^2\right)$$

for some positive constant c. Now, applying the above to $f(\mathbb{X}_{2k},\hat{\theta}_n^{(0)}) - f(\mathbb{X}_{2k},\theta^{(0)})$ and to $f(\mathbb{X}_{2k+1},\hat{\theta}_n^{(1)}) - f(\mathbb{X}_{2k+1},\theta^{(1)})$, we deduce easily that

$$\left|\hat{\sigma}_n^2 - \sigma_n^2\right| \le c\|\hat{\theta}_n - \theta\|\left(1 + \|\theta\| + \|\hat{\theta}_n - \theta\|\right)\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}}\left(1 + \|\mathbb{X}_k\|^2\right)$$

for some positive constant c. From the MDP of $\hat{\theta}_n - \theta$, we infer that

$$\|\hat{\theta}_n - \theta\| \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.19}$$

Form Proposition 4.3 we deduce that

$$\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}}\left(1 + \|\mathbb{X}_k\|^2\right) \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 1 + \operatorname{Tr}(\Lambda). \tag{5.20}$$

We thus conclude via (5.19) and (5.20) that

$$\hat{\sigma}_n^2 - \sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

This ends the proof for $\hat{\sigma}_n$. The proof for $\hat{\rho}_n$ is very similar and uses hypotheses(G2′) and (N2′) to get inequalities similar to (4.22) and (4.30).

Acknowledgments

The authors thank Arnaud Guillin for all his advice and suggestions during the preparation of this work. We are grateful to the anonymous referee for her/his valuable comments which improve the presentation of this work.

References

[1] R. Adamczak and P. Milos. CLT for Ornstein–Uhlenbeck branching particle system. Preprint. Available at arXiv:1111.4559.
[2] I. V. Basawa and J. Zhou. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* **41** (2004) 55–64. MR2057565
[3] B. Bercu, B. de Saporta and A. Gégout-Petit. Asymtotic analysis for bifurcating autoregressive processes via martingale approach. *Electron. J. Probab.* **14** (2009) 2492–2526. MR2563249
[4] B. Bercu and A. Touati. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18** (2008) 1848–1869. MR2462551
[5] V. Bitseki Penda, H. Djellout and A. Guillin. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Ann. Appl. Probab.* **24** (2014) 235–291. MR3161647
[6] R. Cowan and R. G. Staudte. The bifurcating autoregressive model in cell lineage studies. *Biometrics* **42** (1986) 769–783.
[7] V. H. de la Peña, T. L. Lai and Q.-M. Shao. *Self-Normalized Processes. Limit Theory and Statistical Applications. Probability and Its Applications (New York)*. Springer-Verlag, Berlin, 2009. MR2488094
[8] B. de Saporta, A. Gégout-Petit and L. Marsalle. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* **5** (2011) 1313–1353. MR2842907
[9] B. de Saporta, A. Gégout-Petit and L. Marsalle. Asymmetry tests for bifurcating auto-regressive processes with missing data. *Statist. Probab. Lett.* **82** (2012) 1439–1444. MR2929798
[10] J. F. Delmas and L. Marsalle. Detection of cellular aging in a Galton–Watson process. *Stochastic Process. Appl.* **120** (2010) 2495–2519. MR2728175
[11] A. Dembo. Moderate deviations for martingales with bounded jumps. *Electron. Comm. Probab.* **1** (1996) 11–17. MR1386290
[12] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, 2nd edition. Springer, New York, 1998. MR1619036
[13] H. Djellout. Moderate deviations for martingale differences and applications to ϕ-mixing sequences. *Stoch. Stoch. Rep.* **73** (2002) 37–63. MR1914978
[14] H. Djellout, A. Guillin and L. Wu. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** (2006) 393–416. MR2242954
[15] H. Djellout and A. Guillin. Large and moderate deviations for moving average processes. *Ann. Fac. Sci. Toulouse Math.* (6) **10** (2001) 23–31. MR1928987

[16] N. Gozlan. Integral criteria for transportation-cost inequalities. *Electron. Comm. Probab.* **11** (2006) 64–77. MR2231734

[17] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** (2007) 235–283. MR2322697

[18] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** (2007) 1538–1569. MR2358633

[19] R. M. Huggins and I. V. Basawa. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36** (1999) 1225–1233. MR1746406

[20] R. M. Huggins and I. V. Basawa. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42** (2000) 423–432. MR1802966

[21] S. Y. Hwang, I. V. Basawa and I. K. Yeo. Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Stat. Methodol.* **6** (2009) 61–69. MR2655539

[22] M. Ledoux. *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, RI, 2001. MR1849347

[23] P. Massart. *Concentration Inequalities and Model Selection*. *Lecture Notes in Mathematics* **1896**. Springer, Berlin, 2007. MR2319879

[24] A. Puhalskii. Large deviations of semimartingales: A maxingale problem approach. I. Limits as solutions to a maxingale problem. *Stoch. Stoch. Rep.* **61** (1997) 141–243. MR1488137

[25] J. Worms. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* **4** (1999) 28 pp. MR1684149

[26] J. Worms. Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.* **10** (2001) 38–72. MR1841808

[27] J. Worms. Moderate deviations of some dependent variables. II. Some kernel estimators. *Math. Methods Statist.* **10** (2001) 161–193. MR1851746

[28] J. Zhou and I. V. Basawa. Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74** (2005) 77–88. MR2189078

[29] J. Zhou and I. V. Basawa. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Series Anal.* **26** (2005) 825–842. MR2203513

Troisième Partie

Inégalités du coût du transport et applications

The Annals of Probability
2004, Vol. 32, No. 3B, 2702–2732
DOI 10.1214/009117904000000531

TRANSPORTATION COST-INFORMATION INEQUALITIES AND APPLICATIONS TO RANDOM DYNAMICAL SYSTEMS AND DIFFUSIONS

BY H. DJELLOUT, A. GUILLIN AND L. WU

Université Blaise Pascal, Université Paris IX Dauphine and Université Blaise Pascal

We first give a characterization of the L^1-transportation cost-information inequality on a metric space and next find some appropriate sufficient condition to transportation cost-information inequalities for dependent sequences. Applications to random dynamical systems and diffusions are studied.

1. Introduction and questions. Let (E,d) be a metric space equipped with σ-field $\mathcal{B}$ such that $d(\cdot,\cdot)$ is $\mathcal{B}\times\mathcal{B}$-measurable. Given $p\geq 1$ and two probability measures μ and ν on E, we define the quantity

$$W_p^d(\mu,\nu)=\inf\biggl(\iint d(x,y)^p\,d\pi(x,y)\biggr)^{1/p}, \tag{1.1}$$

where the infimum is taken over all probability measures π on the product space $E\times E$ with marginal distributions μ and ν [say coupling of (μ,ν)]. This infimum is finite as soon as μ and ν have finite moments of order p. This quantity is commonly referred to as L^p-Wasserstein distance between μ and ν. When d is the trivial metric ($d(x,y)=\mathbb{1}_{x\neq y}$), $2W_1^d(\mu,\nu)=\|\mu-\nu\|_{\mathrm{TV}}$, the total variation of $\mu-\nu$.

The Kullback information (or relative entropy) of ν with respect to μ is defined as

$$H(\nu/\mu)=\begin{cases}\displaystyle\int\log\frac{d\nu}{d\mu}\,d\nu, & \text{if } \nu\ll\mu,\\ +\infty, & \text{otherwise.}\end{cases} \tag{1.2}$$

We say that the probability measure μ satisfies the L^p-transportation cost-information inequality on (E,d) if there is some constant $C>0$ such that for any probability measure ν,

$$W_p^d(\mu,\nu)\leq\sqrt{2CH(\nu/\mu)}. \tag{1.3}$$

To be short, we write $\mu\in T_p(C)$ for this relation.

Received December 2002; revised October 2003.

AMS 2000 subject classifications. 28A35, 28C20, 60E15, 60G15, 60G99, 60H10, 60J60.

Key words and phrases. Transportation cost-information inequalities, random dynamical systems, diffusions, Girsanov's transformation.

The cases "$p = 1$" and "$p = 2$" are particularly interesting. That $T_1(C)$ are related to the phenomenon of measure concentration was emphasized by Marton [10, 11], Talagrand [18], Bobkov and Götze [2] and amply explored by Ledoux [8, 9]. The $T_2(C)$, first established by Talagrand [18] for the Gaussian measure, has been brought into relation with the log-Sobolev inequality, Poincaré inequality, inf-convolution, Hamilton–Jacobi's equations by Otto and Villani [15] and Bobkov, Gentil and Ledoux [1]. Since those important works, a main trend in the field is to put on relations of $T_p(C)$ with other functional inequalities (of geometrical nature in particular). In this paper we shall study three questions around the following problem going somehow to the opposite direction: *how to establish the "$T_p(C)$" without reference to other functional inequalities in various concrete situations*?

To raise our first question, let us mention the following:

THEOREM 1.1 (Bobkov and Götze [2]). *μ satisfies the L^1-transportation cost-information inequality on (E, d) with constant $C > 0$, that is, $\mu \in T_1(C)$, if and only if for any Lipschitzian function $F : (E, d) \to \mathbb{R}$, F is μ-integrable and*

$$\int_E e^{\lambda(F-\langle F\rangle_\mu)}\,d\mu \le \exp\left(\frac{\lambda^2}{2}C\|F\|_{\mathrm{Lip}}^2\right) \qquad \forall \lambda \in \mathbb{R}, \tag{1.4}$$

where $\langle F\rangle_\mu = \int_E F\,d\mu$ and

$$\|F\|_{\mathrm{Lip}} = \sup_{x\neq y}\frac{|F(x) - F(y)|}{d(x, y)} < +\infty.$$

In that case,

$$\mu(F - \langle F\rangle_\mu > r) \le \exp\left(-\frac{r^2}{2C\|F\|_{\mathrm{Lip}}^2}\right) \qquad \forall r > 0.$$

It might be worthwhile to recall the classical Pinsker–Csizsar inequality which is the starting point of many recent works. By the coupling characterization of the total variation distance $\|\cdot\|_{\mathrm{TV}}$, the Pinsker–Csizsar inequality

$$\|\nu - \mu\|_{\mathrm{TV}} \le \sqrt{\tfrac{1}{2}H(\nu/\mu)}$$

says that w.r.t. the trivial distance $d(x, y) = \mathbb{1}_{x\neq y}$ on E, any probability measure μ satisfies the L^1-transportation cost-information inequality with the sharp constant $C = 1/4$. And by Theorem 1.1, the Pinsker–Csizsar inequality for the trivial distance follows from the classical well-known inequality: for a real bounded random variable ξ with values in $[a, b]$,

$$\mathbb{E}e^{\xi-\mathbb{E}\xi} \le \exp\left(\frac{(b-a)^2}{8}\right)$$

(and vice versa).

We now do a simple remark. Assume that $\mu \in T_1(C)$ or, equivalently, (1.4). Let $\gamma(d\lambda)$ be the standard Gaussian law $\mathcal{N}(0,1)$ on $\mathbb{R}$. We have for any Lipschitzian function F on E with $\langle F\rangle_\mu = 0$ and $\|F\|_{\mathrm{Lip}} \le 1$, and $a \in \mathbb{R}$,

$$\int_E \exp\left(\frac{a^2}{2}F^2\right) d\mu = \int_E \int_{\mathbb{R}} e^{a\lambda F}\gamma(d\lambda)\, d\mu \le \int_{\mathbb{R}} \exp\left(\frac{C}{2}a^2\lambda^2\right)\gamma(d\lambda)$$

$$= \begin{cases} \dfrac{1}{\sqrt{1-a^2C}}, & \text{if } \dfrac{a^2}{2} < \dfrac{1}{2C}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Applying it to $F(x) := d(x,x_0) - \int d(x,x_0)\, d\mu(x)$, we obtain

$$\int e^{cd^2(x,x_0)}\, d\mu(x) < +\infty \qquad \forall c \in \left(0, \frac{1}{2C}\right).$$

In particular, for all $\delta \in (0, \frac{1}{4C})$ we have,

$$\iint e^{\delta d^2(x,y)}\, d\mu(x)\, d\mu(y) < +\infty. \tag{1.5}$$

That naturally leads to the following questions:

QUESTION 1. Will the Gaussian tail (1.5) be sufficient for the L^1-transportation cost-information inequality of μ?

The second question is about dependent tensorizations of the $T_p(C)$. Let, for example, $\mathbb{P}_x^n$, the law of a homogeneous Markov chain $(X_k(x))_{1\le k\le n}$ on E^n starting from $x \in E$, with transition kernel $P(x,dy)$.

QUESTION 2. Assume that $P(x,\cdot) \in T_p(C)$ for all $x \in E$. Where is the appropriate condition under which $\mathbb{P}_x^n$ satisfies the L^p-transportation cost-information inequality w.r.t. the metric

$$d_{l_p}(x,y) := \left(\sum_{i=1}^n d(x_i,y_i)^p\right)^{1/p} ?$$

The same question can be raised for the law of an arbitrary dependent sequence $(X_k)_{1\le k\le n}$. When $(X_k)_{1\le k\le n}$ are independent, this question has a rapid and affirmative answer, see [8, 9] and references therein.

In the dependent case, when d is the trivial metric, and $p=1$ (and d_{l_1} becomes the Hamming distance on E^n), Marton [10] generalized the Pinsker–Csizsar inequality to the law of the so called "contracting" Markov chains:

$$\frac{1}{2} \sup_{(x_{i-1},y_{i-1})} \|P_i(\cdot/y_{i-1}) - P_i(\cdot/x_{i-1})\|_{\mathrm{TV}} \le r < 1. \tag{1.6}$$

Her approach is based on coupling ideas, natural by the definition of the involved Wasserstein distance. Her results have been strengthened by Marton [11, 12] and Dembo [4] and have been generalized to uniform mixing processes by Samson [17] and Rio [16].

However, the trivial distance does not reflect the natural metric structure of the state space E to which usual Markov processes such as random dynamical systems or diffusions are related and that is why the uniform mixing assumption was made in her work (and also in [17]). This is a main motivation for Question 2.

For the L^2-transportation cost-information inequality $T_2(C)$, recall that Talagrand [18] proved that the standard Gaussian law $\gamma = \mathcal{N}(0, 1)$ satisfies $T_2(C)$ on $\mathbb{R}$ w.r.t. the Euclidean distance with the sharp constant $C = 1$ and found that $T_2(C)$ is stable for product (or independent) tensorization. To our knowledge the Markovian tensorization of $T_2(C)$ has not been investigated in the literature.

Since the works of Otto and Villani [15] and Bobkov, Gentil and Ledoux [1], we know that $T_2(C)$ follows from the log-Sobolev inequality in the framework of Riemannian manifolds. Indeed, all known $T_2(C)$-inequalities up to now can be derived from the log-Sobolev inequality. An important open question in the field is whether $T_2(C)$ is strictly weaker than the log-Sobolev inequality. Hence, it would be interesting to investigate the following question:

QUESTION 3. How do we establish the $T_2(C)$-inequality in situations where the log-Sobolev inequality is unknown?

This paper is written around those three questions and it is organized as follows. The next section is the general theoretical part of this paper. After noticing the stability of $T_p(C)$ under Lipschitzian map and under weak convergence in Sections 2.1 and 2.2, in Section 2.3 we prove that condition (1.5) is, in fact, sufficient for the L^1-transportation cost-information inequality, solving Question 1. In Section 2.4 we revisit the coupling method of Marton and show that it actually works for dependent tensorization of $T_p(C)$ for $1 \leq p \leq 2$, under a contraction assumption [see (C1) in Theorem 2.5] close to Marton's (1.6). Section 2.5 is devoted to revisit the McDiarmid–Rio martingale method which allows us to obtain a much more subtle condition (C1$'$) than (C1) for tensorization of $T_1(C)$ in Theorem 2.11.

Sections 3 and 4 contain several applications of the general results in Section 2 to random dynamical systems and diffusions which are our main motivation for Question 2.

In Section 5, quite independent, we present a direct approach of $T_2(C)$ for diffusions, by means of the Girsanov transformation, with respect to the usual Cameron–Martin metric or L^2-metric.

The reader may consult the recent monograph by Villani [19] for an extended (analytical and geometrical) treatment on transportation.

2. Criteria for $T_p(C)$. Throughout this paper let (E, d) be a metric space equipped with σ-field $\mathcal{B}$ such that $d(\cdot, \cdot)$ is $\mathcal{B} \times \mathcal{B}$-measurable; and when (E, d) is separable, $\mathcal{B}$ will be the Borel σ-field.

2.1. *Stability under push-forward by Lipschitz map.* We begin with the stability of $T_p(C)$ under Lipschitzian map and under weak convergence, which will be useful later.

LEMMA 2.1. *Assume that $\mu \in T_p(C)$ on (E, d_E) and (F, d_F) is another metric space. If $\Psi : (E, d_E) \to (F, d_F)$ is Lipschitzian,*

$$d_F\big(\Psi(x), \Psi(y)\big) \le \alpha\, d_E(x, y) \qquad \forall x, y \in E,$$

then $\tilde{\mu} := \mu \circ \Psi^{-1} \in T_p(C\alpha^2)$ on (F, d_F).

PROOF. Let $\tilde{\nu}$ be a probability measure such that $H(\tilde{\nu}/\tilde{\mu}) < +\infty$. The key remark is

$$H(\tilde{\nu}/\tilde{\mu}) = \inf\{H(\nu/\mu); \nu \circ \Psi^{-1} = \tilde{\nu}\}. \tag{2.1}$$

To prove it, putting $\nu_0(dx) := \frac{d\tilde{\nu}}{d\tilde{\mu}}(\Psi(x))\mu(dx)$, we see that $\nu_0 \circ \Psi^{-1} = \tilde{\nu}$. We have for any ν so that $\nu \circ \Psi^{-1} = \tilde{\nu}$,

$$H(\nu/\mu) = H(\nu_0/\mu) + \int d\tilde{\nu}(y) H(\nu_y/\mu_y),$$

where $\nu_y := \nu(\cdot/\Psi = y)$ and $\mu_y := \mu(\cdot/\Psi = y)$ are, respectively, the regular conditional distribution of ν, μ knowing $\Psi = y$. Hence, (2.1) follows.

With (2.1) in hand, the rest of the proof is easy and is omitted. $\square$

2.2. *Stability under weak convergence.*

LEMMA 2.2. *Let (E, d) be a metric, separable and complete space (Polish, say) and $(\mu_n, \mu)_{n\in\mathbb{N}}$ a family of probability measures on E. Assume that $\mu_n \in T_p(C)$ for all $n \in \mathbb{N}$ and $\mu_n \to \mu$ weakly. Then $\mu \in T_p(C)$.*

PROOF. Recall at first two facts (see, e.g., [19]):

1. If $\mu_n \to \mu$ and $\nu_n \to \nu$ weakly, then $\liminf_{n\to\infty} W_p(\mu_n, \nu_n) \ge W_p(\mu, \nu)$.
2. If $\mu_n \to \mu$ weakly and $\{d(x, x_0)^p, \mu_n(dx)\}$ is uniformly integrable, $W_p(\mu_n, \mu) \to 0$.

What one needs to prove is

$$W_p^2(f\mu, \mu) \le 2C \int f \log f \, d\mu$$

for all f such that $f\mu$ is a probability. By approximation (and using fact 2 above), it is sufficient to prove the result for continuous f so that $1/N \le f \le N$ over E for some $N \ge 1$. Let $a_n = \int f\,d\mu_n$ and we have by "$\mu_n \in T_p(C)$,"

$$W_p^2\left(\frac{f\mu_n}{a_n}, \mu_n\right) \le 2C \int \left(\frac{f}{a_n}\right) \log\left(\frac{f}{a_n}\right) d\mu_n = \frac{2C}{a_n} \int f \log f\, d\mu_n.$$

Since μ_n converges weakly to μ, a_n converges to $\mu(f) = 1$, and one can pass to the limit in the right-hand side of this last inequality. For the convergence of the left-hand side, it is enough to apply the lower semi-continuity of W_p. $\square$

2.3. *Characterization of $T_1(C)$ by "Gaussian tail."* We present here a characterization of $T_1(C)$, based on the Bobkov and Götze [2] result, that is, some Gaussian integrability property.

THEOREM 2.3. *A given probability measure μ on (E, d) satisfies the L^1-transportation cost-information inequality with some constant C on (E, d) if and only if* (1.5) *holds. In the latter case,*

$$C \le \frac{2}{\delta} \sup_{k\ge 1}\left(\frac{(k!)^2}{(2k)!}\right)^{1/k} \cdot \left[\iint e^{\delta d^2(x,y)}\, d\mu(x)\, d\mu(y)\right]^{1/k} < +\infty. \tag{2.2}$$

PROOF. It is enough to show the sufficiency. By Bobkov–Götze's Theorem 1.1, it is enough to show that there is some constant $C = C(\delta)$ verifying (2.2) such that

$$\mathbb{E} e^{\lambda F(\xi)} \le \exp\left(\frac{C\lambda^2}{2}\right) \qquad \forall \lambda \in \mathbb{R}, \tag{2.3}$$

for all $F : E \to \mathbb{R}$ with $\|F\|_{\mathrm{Lip}} \le 1$ and $\mathbb{E}F(\xi) = 0$, where ξ is a random variable valued in E with law μ, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Let ξ' be an independent copy of ξ, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Since $\mathbb{E}F(\xi') = 0$, by the convexity of the $x \to e^x$, we have $\mathbb{E}(e^{-\lambda F(\xi')}) \ge 1$. Consequently, noting that $\mathbb{E}[F(\xi) - F(\xi')]^{2k+1} = 0$, we have

$$\begin{aligned}
\mathbb{E}\big(e^{\lambda F(\xi)}\big) &\le \mathbb{E}\big(e^{\lambda F(\xi)}\big)\mathbb{E}\big(e^{-\lambda F(\xi')}\big) \\
&= \mathbb{E} e^{\lambda(F(\xi) - F(\xi'))} \\
&= 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}\mathbb{E}[F(\xi) - F(\xi')]^{2k}}{(2k)!} \\
&\le 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}\mathbb{E}\, d(\xi, \xi')^{2k}}{(2k)!}.
\end{aligned}$$

Hence, putting

$$C := 2 \sup_{k\geq 1}\left(\frac{k! \cdot \mathbb{E}\, d(\xi,\xi')^{2k}}{(2k)!}\right)^{1/k},$$

we get

$$\mathbb{E}\big(e^{\lambda F(\xi)}\big) \leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{k!} \cdot \left(\frac{C}{2}\right)^k = \exp\left(\frac{C\lambda^2}{2}\right).$$

Thus, for (2.3), it remains to estimate C defined above. Since

$$\mathbb{E} d(\xi,\xi')^{2k} \leq k! \cdot \left(\frac{1}{\delta}\right)^k \mathbb{E} \exp\big(\delta d(\xi,\xi')^2\big),$$

we get

$$C \leq \frac{2}{\delta} \sup_{k\geq 1}\left(\frac{(k!)^2}{(2k)!}\right)^{1/k} \cdot \big[\mathbb{E} \exp\big(\delta\big(d(\xi,\xi')^2\big)\big)\big]^{1/k} < +\infty$$

the desired estimate (2.2). $\square$

REMARK 2.4. For comparison notice that the Bernoulli measure μ on $\{0,1\}$ with $\mu(1) \in (0,1)$ satisfies $T_1(1/4)$ w.r.t. the trivial metric, but does not satisfy $T_p(C)$ for any $p > 1$ (see [7]). Hence, any probability measure μ which is not a Dirac measure on E does not satisfy $T_p(C)$ for any $p > 1$ w.r.t. the trivial metric. Another example for illustrating difference of T_1 and T_2 inequalities is the following.

Let $\mu = \phi(x)^2\, dx$ on $\mathbb{R}$ with $0 \leq \phi \in C_0^\infty(\mathbb{R})$ (compact support). It satisfies always $T_1(C)$ w.r.t. the Euclidean $d(x,y) := |y - x|$ by the theorem above. But if the support of μ (or of ϕ) has two connected components I_1, I_2 with $\mathrm{dist}(I_1, I_2) > 0$, then the corresponding $T_2(C)$ fails. In fact, if contrary to $\mu \in T_2(C)$, then by [15] or [1] the following Poincaré inequality holds:

$$\mathrm{Var}_\mu(f) \leq C \int_{\mathbb{R}} f'^2\, d\mu \qquad \forall f \in C_0^\infty(\mathbb{R}).$$

Choose now f smooth enough and equal to 1 on I_1 and 0 on I_2. Then the right-hand side in the Poincaré inequality is 0, whereas the variance of f will be non zero so that the Poincaré inequality cannot hold, neither $T_2(C)$.

This example shows, moreover, that $T_1(C)$ on $\mathbb{R}$ does not imply the Poincaré inequality, unlike $T_2(C)$.

The next two sections are dedicated to the tensorization of $T_p(C)$ for dependent sequences.

2.4. *Weakly dependent sequences*: *Marton's coupling revisited.* Let $\mathbb{P}$ be a probability measure on the product space $(E^n, \mathscr{B}^n)$, $n \geq 2$. For any $x \in E^n$, $x^i := (x_1, \dots, x_i)$. Let $P_i(\cdot/x^{i-1})$ denote the regular conditional law of x_i given x^{i-1} for $i \geq 2$ (assume its existence). By convention $P_1(\cdot/x^0)$ is the law of x_1 under $\mathbb{P}$, where $x^0 = x_0$ is some fixed point. When $\mathbb{P}$ is Markov, then $P_i(\cdot/x^{i-1}) = P_i(\cdot/x_{i-1})$ is the transition kernel at step $i-1$.

Our aim in this section is to extend transportation cost-information inequalities (1.3) for a probability measure $\mathbb{P}$ on (E^n, d_{l_p}), where

$$d_{l_p}(x, y) := \left(\sum_{i=1}^n d(x_i, y_i)^p\right)^{1/p}.$$

THEOREM 2.5. *Let $\mathbb{P}$ be a probability measure on E^n, and $1 \leq p \leq 2$. Assume that $P_i(\cdot/x^{i-1}) \in T_p(C)$ on (E, d) for all $i \geq 1$, x^{i-1} in E^{i-1} ($E^0 := \{x_0\}$). If*

(C1) *there exist $a_j \geq 0$ with $r^p := \sum_{j=1}^\infty (a_j)^p < 1$ such that*

$$[W_p^d(P_i(\cdot/x^{i-1}), P_i(\cdot/\tilde{x}^{i-1}))]^p \leq \sum_{j=1}^{i-1} (a_j)^p d^p(x_{i-j}, \tilde{x}_{i-j}), \tag{2.4}$$

for all $i \geq 1$, $x^{i-1}, \tilde{x}^{i-1}$ in E^{i-1}, then for any probability measure $\mathbb{Q}$ on E^n,

$$W_p^{d_{l_p}}(\mathbb{Q}, \mathbb{P}) \leq \frac{1}{1-r}\sqrt{2Cn^{2/p-1}H(\mathbb{Q}/\mathbb{P})}.$$

PROOF. The proof is similar to the one used for the Hamming distance by Marton [10], however, we have to use the assumption $P_i(\cdot/x^{i-1}) \in T_p(C)$ instead of Pinsker's inequality. Assume that $H(\mathbb{Q}/\mathbb{P}) < +\infty$ (trivial otherwise).

Let $Q_i(\cdot/x^{i-1})$ be the regular conditional law of x_i knowing x^{i-1} for $i \geq 2$ and $Q_1(\cdot/x^0)$ the law of x_1, both under law $\mathbb{Q}$. We shall use the Kullback information between conditional distributions,

$$H_i(\tilde{x}^{i-1}) = H(Q_i(\cdot/\tilde{x}^{i-1})/P_i(\cdot/\tilde{x}^{i-1})),$$

and exploit the following important identity:

$$H(\mathbb{Q}/\mathbb{P}) = \sum_{i=1}^n \int_{E^n} H_i(\tilde{x}^{i-1})\, d\mathbb{Q}(\tilde{x}). \tag{2.5}$$

The key is to construct an appropriate coupling of $\mathbb{Q}$ and $\mathbb{P}$, that is, two random sequences $\tilde{X}^n$ and X^n distributed according to $\mathbb{Q}$ and $\mathbb{P}$, respectively, on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$.

We define a joint distribution $\mathcal{L}(\tilde{X}^n, X^n)$ by induction as follows. Add artificially time 0 and put $X_0 = \tilde{X}_0 = \tilde{x}^0 = x^0$, the fixed point. Assume that for some i, $1 \leq i \leq n$, $\mathcal{L}(\tilde{X}^{i-1}, X^{i-1})$ is already defined. We have to define the joint conditional distribution $\mathcal{L}(\tilde{X}_i, X_i/\tilde{X}^{i-1} = \tilde{x}^{i-1}, X^{i-1} = x^{i-1})$, where $(\tilde{x}^{i-1}, x^{i-1})$ is fixed (but arbitrary).

Given $\varepsilon > 0$ so small that $r(1+\varepsilon) < 1$, this distribution will have marginal laws

$$\mathcal{L}(\tilde{X}_i/\tilde{X}^{i-1} = \tilde{x}^{i-1}, X^{i-1} = x^{i-1}) = Q_i(\cdot/\tilde{x}^{i-1})$$

and

$$\mathcal{L}(X_i/\tilde{X}^{i-1} = \tilde{x}^{i-1}, X^{i-1} = x^{i-1}) = P_i(\cdot/x^{i-1})$$

so as to satisfy

$$\begin{aligned}&\mathbb{E}\big(d(\tilde{X}_i, X_i)^p/\tilde{X}^{i-1} = \tilde{x}^{i-1}, X^{i-1} = x^{i-1}\big)\\ &\qquad \le (1+\varepsilon)W_p^d\big(Q_i(\cdot/\tilde{x}^{i-1}), P_i(\cdot/x^{i-1})\big)^p\end{aligned}$$

for all $\tilde{x}^{i-1}, x^{i-1}$ in E^{i-1}. Obviously, $\tilde{X}^n, X^n$ are of law $\mathbb{Q}, \mathbb{P}$, respectively.

By the triangle inequality for the W_p^d-distance,

$$\begin{aligned}&\mathbb{E}\big(d(\tilde{X}_i, X_i)^p/\tilde{X}^{i-1} = \tilde{x}^{i-1}, X^{i-1} = x^{i-1}\big)\\ &\qquad \le (1+\varepsilon)\big[W_1^d\big(Q_i(\cdot/\tilde{x}^{i-1}), P_i(\cdot/\tilde{x}^{i-1})\big) + W_1^d\big(P_i(\cdot/\tilde{x}^{i-1}), P_i(\cdot/x^{i-1})\big)\big]^p.\end{aligned}$$

Using the elementary inequality that $(x+y)^p \le a^{p-1}x^p + b^{p-1}y^p$ (for $p \ge 1$ $\forall x, y \ge 0$) where $a, b > 1$ such that $1/a + 1/b = 1$, we have by the assumptions $P_i(\cdot/x^{i-1}) \in T_p(C)$ and (C1)

$$\begin{aligned}&\mathbb{E}\big(d^p(\tilde{X}_i, X_i)/\tilde{X}^{i-1} = \tilde{x}^{i-1}, X^{i-1} = x^{i-1}\big)\\ &\qquad \le (1+\varepsilon)\left(\sqrt{2CH_i(\tilde{x}^{i-1})} + \left[\sum_{j=1}^{i-1}(a_j)^p d^p(\tilde{x}_{i-j}, x_{i-j})\right]^{1/p}\right)^p \qquad (2.6)\\ &\qquad \le (1+\varepsilon)\left(a^{p-1}[2CH_i(\tilde{x}^{i-1})]^{p/2} + b^{p-1}\sum_{j=1}^{i-1}(a_j)^p d^p(\tilde{x}_{i-j}, x_{i-j})\right).\end{aligned}$$

By recurrence on i, this entails that $\mathbb{E}d^p(X_i, \tilde{X}_i) < +\infty$ for all $i = 1, \dots, n$. Taking the average with respect to $\mathcal{L}(\tilde{X}^{i-1}, X^{i-1})$, summing on i and using the concavity of the function $x \to x^{p/2}$ for $p \in [1, 2]$, we get by (2.5) and (2.6)

$$\begin{aligned}&\frac{1}{n(1+\varepsilon)}\sum_{i=1}^n \mathbb{E}\big(d^p(\tilde{X}_i, X_i)\big)\\ &\qquad \le a^{p-1}\left(\frac{2C}{n}\sum_{i=1}^n \mathbb{E}H_i(\tilde{X}^{i-1})\right)^{p/2} + \frac{b^{p-1}}{n}\sum_{i=1}^n\sum_{j=1}^{i-1} a_j^p \mathbb{E}\,d^p(\tilde{X}_{i-j}, X_{i-j})\\ &\qquad = a^{p-1}\left(\frac{2C}{n}H(\mathbb{Q}/\mathbb{P})\right)^{p/2} + \frac{b^{p-1}}{n}\sum_{k=1}^{n-1}\mathbb{E}\,d^p(\tilde{X}_k, X_k)\sum_{i=k+1}^n a_{i-k}^p.\end{aligned}$$

Using $\sum_{j\ge 1} a_j^p = r^p$ and letting $\varepsilon \to 0+$, the above inequality gives us, when $r^p b^{p-1} < 1$,

$$W_p^{d_{l^p}}(\mathbb{Q},\mathbb{P}) \le \left(\frac{a^{p-1}}{1-r^p b^{p-1}}\right)^{1/p} \sqrt{2Cn^{2/p-1} H(\mathbb{Q}/\mathbb{P})}.$$

Optimizing on (a,b), we get the desired inequality. $\square$

Noting that for a real function f on E^n, $\|f\|_{\mathrm{Lip}(d_{l_1})} \le \alpha$ if and only if for every $k = 1,\dots,n$,

$$|f_k(x_k) - f_k(y_k)| \le \alpha d(x_k, y_k) \qquad \forall x_k, y_k \in E, \tag{2.7}$$

where $f_k(x_k)$ is the function f w.r.t. the kth variable while the others are fixed. Then we get by Theorem 1.1,

COROLLARY 2.6. *Under the assumption of Theorem* 2.5 *for* $p = 1$, *for any real function* f *on* E^n *satisfying* (2.7),

$$\mathbb{E}_{\mathbb{P}} e^{\lambda(f-\mathbb{E}_{\mathbb{P}} f)} \le \exp\left(\frac{C\lambda^2\alpha^2 n}{2(1-r)^2}\right) \qquad \forall \lambda \in \mathbb{R}.$$

In particular, for any $t > 0$,

$$\mathbb{P}(f > \mathbb{E}_{\mathbb{P}} f + t) \le \exp\left(-\frac{t^2(1-r)^2}{2nC\alpha^2}\right).$$

REMARK 2.7. The condition $P_i(\cdot/x^{i-1}) \in T_p(C)$ is our starting point for tensorization of the $T_p(C)$ and it is verified for many interesting examples, such as the stochastic differential equation (SDE) (4.1) or random dynamical systems or Gibbs fields. Condition (C1), meaning that the dependence of the *present* on the past is very weak, is a crucial condition. Indeed, when $d(x,y) = \mathbb{1}_{x\ne y}$, $p = 1$ and $\mathbb{P}$ is Markovian, (C1) is equivalent to (1.6), and Theorem 2.5 is exactly the result of Marton mentioned in the Introduction.

REMARK 2.8. That the constant C_n for the T_1-inequality of $\mathbb{P}_x$ increases linearly on dimension n is natural in the point of view of the Hoeffding inequality in Corollary 2.6. This is completely different from the case of the T_2-inequality, for which it is hoped that the T_2-constant remains independent of dimension n, as seen for the independent tensorization of $T_2(C)$ by Talagrand [18] or its extension Theorem 2.5.

REMARK 2.9. Under $P_i(\cdot/x^{i-1}) \in T_p(C)$ and (2.4) but without the contraction condition that $r^p := \sum_j (a_j)^p < 1$, we have always $\mathbb{P}_x \in T_p(C_n)$ on E^n w.r.t. d_{l_p} for some constant C_n (but the crucial estimate of C_n in Theorem 2.5 is lost). We give only the proof of this fact for $p = 1$.

Indeed, consider the nonnegative nilpotent lower triangular matrix $A = (a_{ij})$, where $a_{ij} = a_{i-j}$ if $i > j$ and 0 otherwise. For any given $\delta \in (0, 1)$, there is always a (positive) vector $z = (z_1, \dots, z_n)$ such that $z_i > 0$, $\sum_i z_i = 1$ and

$$(zA)_k = \sum_{i=k+1}^{n} z_i a_{i-k} \le \delta z_k \qquad \forall k = 1, \dots, n.$$

Then by (2.5) for $p = 1$, we have by Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}\sum_{i=1}^{n} z_i\, d(X_i, \tilde{X}_i) \\
&\quad \le (1+\varepsilon)\left(\sum_{i=1}^{n} z_i \mathbb{E}\sqrt{2CH(\tilde{x}^{i-1})} + \sum_{i=1}^{n} z_i \sum_{j=1}^{i-1} a_j \mathbb{E}\, d(X_{i-j}, \tilde{X}_{i-j})\right) \\
&\quad \le (1+\varepsilon)\left(\sqrt{\sum_{i=1}^{n} z_i 2C\mathbb{E}H(\tilde{x}^{i-1})} + \sum_{k=1}^{n-1} \mathbb{E}\, d(X_k, \tilde{X}_k) \sum_{i=k+1}^{n} z_i a_{i-k}\right) \\
&\quad \le (1+\varepsilon)\left(\sqrt{2C \max_i z_i H(\mathbb{Q}/\mathbb{P})} + \sum_{k=1}^{n-1} \delta z_k \mathbb{E}\, d(X_k, \tilde{X}_k)\right),
\end{aligned}$$

where it follows that

$$W_1^{d_{l_1}}(\mathbb{Q}, \mathbb{P}) \le \frac{1}{(1-\delta)\min_i z_i} \sqrt{2C \max_i z_i H(\mathbb{Q}/\mathbb{P})}.$$

When $z_i = 1/n$, the best choice of δ is r, and this inequality becomes Theorem 2.5.

2.5. *$T_1(C)$ for weakly dependent sequences: McDiarmid–Rio's martingale method revisited.* The last inequality in Corollary 2.6, applied to $F(X_1, \dots, X_n) = \sum_{k=1}^{n} f(X_k)$ and the trivial metric d, where (X_k) are independent and $\|f(X_k)\|_\infty \le \alpha$, becomes exactly the sharp Hoeffding inequality (see [13]). But when it is applied to $F(X_1, \dots, X_n) = f(X_n)$, it does not furnish the good order of n for n large. As this last question is important for the $T_1(C)$ of the the invariant measure, we give now a very simple proof of the following:

PROPOSITION 2.10. *Let (E, d) be a Polish space. Let $P(x, dy)$ be a Markov kernel on E such that*:

(a) $P(x, \cdot) \in T_1(C)$ *for every* $x \in E$;
(b) $W_1^d(P(x, \cdot); P(\tilde{x}, \cdot)) \le r d(x, \tilde{x})$, *for every* $x, \tilde{x}$ *in* E *and some* $r < 1$.

Then there is a unique invariant probability measure μ of P and it satisfies $T_1(C_\infty)$ as well as $P^n(x, \cdot)$ $\forall n \ge 1$, where $C_\infty = C(1 - r^2)^{-1}$.

PROOF. When (E,d) is Polish, the space $M_1^p(E)$ of probability measures ν on E such that $\int d(x,x_0)^p\,d\nu(x) < +\infty$, equipped with the Wasserstein metric $W_p(\cdot,\cdot)$ is a metric complete separable space (see [19]). Since $\nu \in M_1^1(E) \Rightarrow \nu P \in M_1^1(E)$ by (a) and, condition (b) implies (in fact, equivalent to)

$$W_1(\nu_1 P, \nu_2 P) \le r W_1(\nu_1, \nu_2) \qquad \forall\, \nu_1, \nu_2 \in M_1^1(E),$$

hence, by the fixed point theorem, there is one and only one P-invariant measure $\mu \in M_1^p(E)$, and $P^n(x,\cdot) \to \mu$ in the metric W_1 for any initial point $x \in E$. The last point shows also that μ is the unique invariant probability measure of P [without the restriction that $\mu \in M_1^1(E)$].

Since

$$W_1^d(\nu,\mu) = \sup_{f:\,\|f\|_{\mathrm{Lip}}\le 1} \left| \int_E f\,d\nu - \int_E f\,d\mu \right|,$$

condition (b) is also equivalent to

$$\|Pf\|_{\mathrm{Lip}} \le r\|f\|_{\mathrm{Lip}} \qquad \forall\, f.$$

Thus, $\|P^N f\|_{\mathrm{Lip}} \le r^N \|f\|_{\mathrm{Lip}}$ for all $N \ge 1$. Now given a Lipschitzian function f, we have by (a) and Bobkov–Götze's Theorem 1.1,

$$\begin{aligned}
P^n(e^f) &\le P^{n-1}\left[\exp\left(Pf + \frac{C\|f\|_{\mathrm{Lip}}^2}{2}\right)\right] \\
&\le P^{n-2}\left[\exp\left(P^2 f + \frac{C\|f\|_{\mathrm{Lip}}^2}{2} + \frac{C\|Pf\|_{\mathrm{Lip}}^2}{2}\right)\right] \\
&\le \cdots \\
&\le \exp\left(P^n f + \frac{C\|f\|_{\mathrm{Lip}}^2}{2} + \frac{C\|Pf\|_{\mathrm{Lip}}^2}{2} + \cdots + \frac{C\|P^{n-1}f\|_{\mathrm{Lip}}^2}{2}\right) \\
&\le \exp\left(P^n f + \frac{C\|f\|_{\mathrm{Lip}}^2}{2(1-r^2)}\right).
\end{aligned}$$

In other words, for every $x \in E$, $P^n(x,\cdot) \in T_1(C_\infty)$, where C_∞ is given in the proposition. Letting $n \to \infty$, we obtain the desired result for μ by Lemma 2.2. $\square$

We now use the martingale method of McDiarmid [14] (in the independent case) and Rio [16] (in the uniform mixing case) for extending the argument above to the process-level law $\mathbb{P}$.

THEOREM 2.11. *Let* $\mathbb{P}$ *be a probability measure on* E^n *satisfying* $P_i(\cdot/x^{i-1}) \in T_1(C)$ $(\forall\, i, x^{i-1})$ *in Theorem* 2.5. *Assume instead of* (C1) *that*

(C1′) *there is some constant* $S > 0$ *such that for all real bounded Lipschitzian function* $f(x_{k+1},\dots,x_n)$ *with* $\|f\|_{\mathrm{Lip}(d_{l_1})} \le 1$, *for all* $x \in E^n$, $y_k \in E$,

$$\begin{aligned}&\bigl|\mathbb{E}_{\mathbb{P}}\bigl(f(X_{k+1},\dots,X_n)/X^k = x^k\bigr) - \mathbb{E}_{\mathbb{P}}\bigl(f(X_{k+1},\dots,X_n)/X^k = (x^{k-1}, y_k)\bigr)\bigr|\\ &\qquad \le Sd(x_k, y_k).\end{aligned}$$

Then for all function, F *on* E^n *satisfying* (2.7),

$$\mathbb{E}_{\mathbb{P}} e^{\lambda(F-\mathbb{E}_{\mathbb{P}}F)} \le \exp\left(\frac{C\lambda^2(1+S)^2\alpha^2 n}{2}\right) \qquad \forall \lambda \in \mathbb{R}. \tag{2.8}$$

Equivalently, $\mathbb{P} \in T_1(C_n)$ *on* (E^n, d_{l_1}) *with*

$$C_n = nC(1+S)^2.$$

PROOF. We may assume without loss of generality that $\alpha = 1$. Let $(M_k = \mathbb{E}_{\mathbb{P}}(F/X^k))_{k\ge 0}$, where $M_0 = \mathbb{E}_{\mathbb{P}}F$. It is a martingale. It is enough to show that for each k,

$$\mathbb{E}_{\mathbb{P}}\bigl(e^{\lambda(M_k - M_{k-1})}/X^{k-1}\bigr) \le \exp\left(\frac{C\lambda^2(1+S)^2}{2}\right).$$

To this end, note at first by $P_i(\cdot/x^{i-1}) \in T_1(C)$ and Theorem 1.1,

$$\mathbb{E}_{\mathbb{P}}\bigl(e^{\lambda(M_k - M_{k-1})}/X^{k-1}\bigr) \le \exp\left(\frac{C\lambda^2 b_k^2}{2}\right),$$

where

$$b_k := \sup_{x,y} \frac{|M_k(x^k) - M_k(x^{k-1}, y_k)|}{d(x_k, y_k)}.$$

But $M_k(x^k) = \int F(x^k, x_{k+1},\dots,x_n)\mathbb{P}(dx_{k+1},\dots,dx_n/x^k)$, writing $x_{k+1}^n = (x_{k+1},\dots,x_n)$ we have

$$\begin{aligned}&|M_k(x^k) - M_k(x^{k-1}, y_k)|\\ &\quad \le \left|\int \bigl(F(x^k, x_{k+1}^n) - F(x^{k-1}, y_k, x_{k+1}^n)\bigr)\mathbb{P}(dx_{k+1}^n/x^k)\right|\\ &\qquad + \left|\int F(x^{k-1}, y_k, x_{k+1}^n)\bigl(\mathbb{P}(dx_{k+1}^n/x^k) - \mathbb{P}(dx_{k+1}^n/x^{k-1}, y_k)\bigr)\right|\\ &\quad \le d(x_k, y_k) + Sd(x_k, y_k).\end{aligned}$$

Hence, $b_k \le (1+S)$, the desired result. □

REMARK 2.12. When $d(x,y) = \mathbb{1}_{x\ne y}$, $P_i(\cdot/x^{i-1}) \in T_1(1/4)$, and this result is essentially due to Rio [16]. Using a different condition than (C1′), he essentially proved that the constant S in condition (C1′) verifies $S \le 2\sum_{j=1}^{\infty}\phi_j$, where ϕ_j is the uniform mixing coefficient of the sequence (X_n). Our proof above is, in fact, inspired by his work.

REMARK 2.13. If the condition (C1) is viewed as a *backward* type, then (C1′) may be seen as a *forward* type. Indeed (C1′) is equivalent to

$$W_1^{d_{l_1}}\big(\mathbb{P}(dx_{k+1}^n/x_k, x^{k-1}), \mathbb{P}(dx_{k+1}^n/y_k, x^{k-1})\big) \le Sd(x_k, y_k).$$

It means intuitively that the present does not influence a lot the future of the process $\mathbb{P}$. In concrete situations (C1′) is often weaker than (C1) with $p = 1$. For example, let $(\mathbb{P}_x)$ be a uniformly ergodic (Doeblin recurrent, say) Markov chain with transition $P(x, dy)$ in the sense that $r_n := \sup_{x\in E} \|P^n(x, \cdot) - \mu\|_{\mathrm{TV}} \to 0$. As $2\phi_n \le r_n$, we have by Rio's estimate above,

$$S \le \sum_{n=1}^{\infty} \sup_{x\in E} \|P^n(x, \cdot) - \mu\|_{\mathrm{TV}},$$

which is finite. But Marton's condition (1.6) or (C1) means $(1/2)\sup_{x\in E} \|P^n(x, \cdot) - \mu\|_{\mathrm{TV}} \le r^n$ for all $n \ge 1$. See also Example 3.3.

It would be very interesting to generalize Theorem 2.11 to $T_2(C)$.

3. Application: study of $T_1(C)$ and $T_2(C)$ for random dynamical systems.

3.1. $T_1(C)$. Let E be a complete connected Riemannian manifold equipped with the Riemannian metric d. Consider now the nonlinear random perturbed dynamical system valued in E,

$$X_0(x) := x \in E, \qquad X_{n+1}(x) = F\big(X_n(x), W_{n+1}\big), \qquad n \ge 0, \tag{3.1}$$

where the noise $(W_n)_{n\ge 0}$ is a sequence of i.i.d. r.v. valued in some measurable space $(G, \mathcal{G})$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $F(x, w): E \times G \to E$ is measurable. Denote by $P(x, dy)$ the law of $F(x, W_1)$, and the following:

PROPOSITION 3.1. *Assume that there exists* $\delta > 0$ *such that*

$$\sup_{x\in E} \mathbb{E}\big(e^{\delta d(F(x,W_1), F(x,W_2))^2}\big) < +\infty. \tag{3.2}$$

If there exists $0 \le r < 1$ *such that*

$$\mathbb{E}\big(d\big(F(x, W_1), F(\tilde{x}, W_1)\big)\big) \le rd(x, \tilde{x}) \qquad \forall x, \tilde{x} \in E, \tag{3.3}$$

or more generally for some constant $S \ge 0$,

$$\sum_{n=1}^{\infty} \mathbb{E}\big(d\big(X_n(x), X_n(\tilde{x})\big)\big) \le S\, d(x, \tilde{x}) \qquad \forall x, \tilde{x} \in E, \tag{3.4}$$

then there is some constant $C > 0$ *such that for any* $n \ge 1$, *for every probability measure* $\mathbb{Q}^n$ *on* E^n,

$$W_1^{d_{l_1}}(\mathbb{Q}^n, \mathbb{P}_x^n) \le \sqrt{CnH(\mathbb{Q}^n/\mathbb{P}_x^n)},$$

where $\mathbb{P}_x^n$ *is the law of* $(X_k(x))_{1\le k\le n}$ *on* E^n.

PROOF. By Theorem 2.3, condition (3.2) is equivalent to "$P(x,\cdot) \in T_1(C)$ $\forall x \in E$." Notice that (3.3) is equivalent to (C1) (with $p = 1$) in Theorem 2.5, and (3.4) implies trivially (C1′) with the same constant S in Theorem 2.11. Hence, this proposition follows from Theorems 2.5 and 2.11. □

REMARK 3.2. If the largest Lyapunov exponent in L^1 given by

$$\lambda_{\max}(L^1) := \lim_{n\to\infty}\left(\sup_{x\neq\tilde{x}} \frac{\mathbb{E}d(X_n(x), X_n(\tilde{x}))}{d(x,\tilde{x})}\right)^{1/n}$$

is strictly smaller than 1, then condition (3.4) is verified.

EXAMPLE 3.3 (ARMA model). To see the difference between (C1) in Theorem 2.5 and (C1′) in Theorem 2.11, let us consider the ARMA model

$$X_0(x) = x, \qquad X_{n+1}(x) = AX_n(x) + W_{n+1}$$

in $E = \mathbb{R}^d$, where $A \in \mathcal{M}_{d\times d}$ (the space of $d \times d$ matrices) and (W_n) is a sequence of i.i.d. r.v. with values in $G = \mathbb{R}^d$. This model is a particular case of the general model above with $F(x,w) = Ax + w$. Condition (C1), equivalent to (3.3), means that $r = \|A\| := \sup\{|Ax|; |x| \le 1\} < 1$, however, (C1′) for this linear model is equivalent to

$$r_{\mathrm{sp}}(A) := \max\{|\lambda|; \lambda \text{ is an eigenvalue in } \mathbb{C} \text{ of } A\} = \lambda_{\max}(L^1) < 1,$$

which is much weaker. This last condition is a well-known sharp sufficient condition for the ergodicity of this linear ARMA model (X_n).

REMARK 3.4. For this model, the known results mentioned in the Introduction cannot be applied, for the uniform mixing condition is, in general, not satisfied when E is noncompact. For example, the ARMA model with $A \neq 0$ and W_1 unbounded is never uniformly mixing. See [22].

3.2. $T_2(C)$. Consider a particular case of the preceding model

$$X_0(x) = x, \qquad X_{n+1}(x) = f\big(X_n(x)\big) + \sigma\big(X_n(x)\big)W_{n+1}, \tag{3.5}$$

(the discrete time SDE), that is, $F(x,w) = f(x) + \sigma(x)w$, where $E = \mathbb{R}^d$, $G = \mathbb{R}^n$, $f : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathcal{M}_{d\times n}$ (the space of $d \times n$ matrices) and the noise $(W_n)_{n\in\mathbb{Z}}$ is a sequence of i.i.d. r.v. with values in $\mathbb{R}^n$ such that $\mathbb{E}W_1 = 0$. Assume that:

(i) $\mathbb{P}_W := \mathbb{P}(W_1 \in \cdot) \in T_2(C)$ on $\mathbb{R}^n$ w.r.t. the Euclidean metric;
(ii) $|\sigma(x)w| \le K|w|$ $\forall (x,w) \in \mathbb{R}^d \times \mathbb{R}^n$;
(iii) for some $r \in [0,1)$,

$$\sqrt{|f(x) - f(\tilde{x})|^2 + \mathbb{E}\big|\big(\sigma(x) - \sigma(\tilde{x})\big)W_1\big|^2} \le r|x - \tilde{x}| \qquad \forall x, \tilde{x} \in \mathbb{R}^d. \tag{3.6}$$

Notice that conditions (i) and (ii) imply that $P(x,\cdot) \in T_2(CK^2)$ for all $x \in \mathbb{R}^d$, by Lemma 2.1; and condition (iii) implies (C1) with the same r for $p = 2$. Hence, by Theorem 2.5, $\mathbb{P}_x^n \in T_2(CK^2/(1-r)^2)$. That yields, by Bobkov, Gentil and Ledoux [1], the following:

COROLLARY 3.5. *For the model* (3.5) *above assume conditions* (i)–(iii). *Then* $\mathbb{P}_x^n \in T_2(CK^2/(1-r)^2)$ *and for any measurable function* $F(x_1, \dots, x_n) \in L^1((\mathbb{R}^d)^n, \mathbb{P}_x^n)$,

$$\mathbb{E}\exp\bigl(\rho QF\bigl(X_1(x), \dots, X_n(x)\bigr)\bigr) \le \exp\bigl(\rho\mathbb{E}F\bigl(X_1(x), \dots, X_n(x)\bigr)\bigr),$$

where

$$\rho := \frac{(1-r)^2}{CK^2}, \qquad QF(x_1, \dots, x_n) := \inf_{y \in (\mathbb{R}^d)^n}\left(F(x+y) + \tfrac{1}{2}\sum_{k=1}^n |y_k|^2\right).$$

As noted in [1], several estimates of Laplace integrals are the consequence of the functional inequality version of the $T_2(C)$ above. For instance, Corollary 6.1 in [1] says that for any convex function F on $(\mathbb{R}^d)^n$,

$$\mathbb{E}_{\mathbb{P}_x^n}\exp\left(\rho\left[F - \tfrac{1}{2}\sum_{k=1}^n (\partial_k F)^2\right]\right) \le \exp\bigl(\rho\mathbb{E}_{\mathbb{P}_x^n} F\bigr).$$

REMARK 3.6. Consider the Lyapunov exponent in L^2,

$$\lambda_{\max}(L^2) := \lim_{n\to\infty}\left(\sup_{x\neq\tilde{x}} \frac{\mathbb{E}d(X_n(x), X_n(\tilde{x}))^2}{d(x,\tilde{x})^2}\right)^{1/n}.$$

Obviously, (3.6) implies $\lambda_{\max}(L^2) < 1$. It is then natural to ask whether $P(x,\cdot) \in T_2(C)$ $\forall x$ plus $\lambda_{\max}(L^2) < 1$ do imply "$\mathbb{P}_x^n \in T_2(K)$" for some constant K independent of n (for which we have no answer unlike for T_1). Notice that for the ARMA model, $\lambda_{\max}(L^2) = \lambda_{\max}(L^1) = r_{\mathrm{sp}}(A)$.

4. Application: study of $T_1(C)$ for paths of SDEs. Let us give here an application of Theorem 2.3 to SDE. Consider the SDE in $\mathbb{R}^d$,

$$dX_t = \sigma(X_t)\,dB_t + b(X_t)\,dt, \quad X_0 = x \in \mathbb{R}^d, \tag{4.1}$$

where $\sigma : \mathbb{R}^d \to \mathcal{M}_{d\times n}$, $b : \mathbb{R}^d \to \mathbb{R}^d$ and (B_t) is the standard Brownian motion valued in $\mathbb{R}^n$ defined on some well filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$.

Assume that σ, b are locally Lipschitzian and for all $x, y \in \mathbb{R}^d$,

$$\sup_{x\in\mathbb{R}^d} \|\sigma(x)\|_{\mathrm{HS}} \le A, \qquad \langle y-x, b(y)-b(x)\rangle \le B(1+|y-x|^2), \tag{4.2}$$

where $\|\sigma\|_{\mathrm{HS}} := \sqrt{\operatorname{tr}\sigma\sigma^t}$ is the Hilbert–Schmidt norm, $\langle x,y\rangle$ is the Euclidean inner product and $|x| := \sqrt{\langle x,x\rangle}$. It has a unique nonexplosive solution denoted by $(X_t(x))$ whose law on the space $C(\mathbb{R}^+, \mathbb{R}^d)$ of $\mathbb{R}^d$-valued continuous functions on $\mathbb{R}^+$ will be denoted by $\mathbb{P}_x$.

COROLLARY 4.1. *Assume the conditions above. For each $T > 0$, there exists some constant $C = C(T, A, B)$ independent of initial point x such that $\mathbb{P}_x$ satisfies the $T_1(C)$ for every $x \in \mathbb{R}^d$, on the space $C([0, T], \mathbb{R}^d)$ of $\mathbb{R}^d$-valued continuous functions on $[0, T]$ equipped with the metric*

$$d_T(\gamma_1, \gamma_2) := \sup_{t\in[0,T]} |\gamma_1(t) - \gamma_2(t)|.$$

PROOF. Let $(B_t), (\tilde{B}_t)$ be two independent Brownian motions defined on some filtered probability $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and $X_t(x), \tilde{X}_t(x)$ strong solutions of (4.1), respectively, driven by $(B_t), (\tilde{B}_t)$. Put

$$\hat{X}_t := X_t(x) - \tilde{X}_t(x), \qquad \hat{b}_t := b\big(X_t(x)\big) - b\big(\tilde{X}_t(x)\big)$$
$$a(\cdot) := \sigma\sigma^t(\cdot), \qquad \bar{a}_t := a\big(X_t(x)\big) + a\big(\tilde{X}_t(x)\big)$$
$$L_t := \int_0^t \sigma\big(X_t(x)\big)\, dB_t - \int_0^t \sigma\big(\tilde{X}_t(x)\big)\, d\tilde{B}_t.$$

Then

$$\hat{X}_t = L_t + \int_0^t \hat{b}_s\, ds.$$

By Theorem 2.3, it is enough to show that there exists some positive constant $\delta = \delta(T, A, B)$ such that

$$\mathbb{E}\exp\Big(\delta \sup_{0\le t\le T} |\hat{X}_t|^2\Big) < +\infty. \tag{4.3}$$

Let $f(x) := h(|x|)$, where $h \in C^\infty(\mathbb{R})$ is pair and such that $h(r) = r$ for $r \ge 4$ and

$$h(r) \ge r, \qquad 0 \le h'(r) \le 1 \wedge r, \qquad 0 \le h''(r) \le 1 \qquad \forall r \in [0, 4].$$

Consider $Y_t := (1 + f(\hat{X}_t))e^{-\beta t}$, where $\beta > 0$ is a constant to be determined later. By Ito's formula,

$$\begin{aligned} dY_t &= e^{-\beta t}\Bigg(\frac{1}{2}\sum_{i,j=1}^d \bar{a}_t^{ij}\partial_i\partial_j f(\hat{X}_t) + \langle\nabla f(\hat{X}_t), \hat{b}_t\rangle\Bigg)dt - \beta Y_t\, dt + dM_t \\ &= e^{-\beta t}\Bigg(\frac{1}{2}h''(|\hat{X}_t|)\frac{\langle\hat{X}_t, \bar{a}_t\hat{X}_t\rangle}{|\hat{X}_t|^2} + \frac{1}{2}h'(|\hat{X}_t|)\bigg(\frac{\operatorname{tr}\bar{a}_t}{|\hat{X}_t|} - \frac{\langle\hat{X}_t, \bar{a}_t\hat{X}_t\rangle}{|\hat{X}_t|^3}\bigg) \\ &\qquad\qquad + \frac{h'(|\hat{X}_t|)}{|\hat{X}_t|}\langle\hat{X}_t, \hat{b}_t\rangle - \beta\big(1 + h(|\hat{X}_t|)\big)\Bigg)dt + dM_t, \end{aligned}$$

where (M_t) is a local martingale (M_t) with $M_0 = 0$, whose quadratic variational process $[M]$ is given by

$$[M]_t = \int_0^t e^{-2\beta s}\langle\nabla f(\hat{X}_s), \bar{a}_s\nabla f(\hat{X}_s)\rangle\, ds \le 2A^2\int_0^t e^{-2\beta s}\, ds \le \frac{A^2}{\beta}.$$

Using our condition (4.2), we see that $Y_t \le 1 + h(0) + M_t$ once if

$$\beta > \max\{0, 2A^2 + B\}.$$

Fix such a β. For any $\lambda > 0$, using the exponential martingale,

$$\exp\left(\lambda M_t - \frac{\lambda^2}{2}[M]_t\right),$$

(Novikov's condition is satisfied) and Doob's maximal inequality [applied to the positive submartingale $\exp(\lambda M_t/2)$], we have

$$\mathbb{E}e^{\lambda(\sup_{t\le T} Y_t - 1 - h(0))} \le \mathbb{E}\sup_{t\le T} e^{\lambda M_t} \le 4(\mathbb{E}e^{\lambda M_T})^2 \le 4\exp\left(\frac{\lambda^2 A^2}{\beta}\right).$$

Hence, by Chebychev's inequality and an optimization of λ, we get

$$\mathbb{P}\left(\sup_{t\le T} Y_t > 1 + h(0) + r\right) \le 4\exp\left(-\frac{\beta r^2}{4A^2}\right) \qquad \forall r > 0.$$

Consequently,

$$\mathbb{E}\exp\left(a \sup_{t\le T} Y_t^2\right) < +\infty, \qquad \text{if } \ 0 < a < \frac{\beta}{4A^2}.$$

Hence, (4.3) is true for all $\delta \in (0, e^{-\beta T}\frac{\beta}{4A^2})$, where $\beta > \max\{0, 2A^2 + B\}$. $\square$

REMARK 4.2. If $b \in C^2$ verifies for some constant B,

$$\nabla^s b := \left(\tfrac{1}{2}(\partial_i b^j + \partial_j b^i)\right)_{1\le i,j\le d} \le B I_d \tag{4.4}$$

in the order of nonnegative definiteness where I_d is the identity matrix, then $\langle y - x, b(y) - b(x)\rangle \le B|x - y|^2$ and the condition on b in (4.2) is satisfied.

REMARK 4.3. Assume $\|\nabla b\| \le K$, $n = d$ and $\sigma(x) = \sigma = I_d$. Capitaine, Hsu and Ledoux [3] yields the log-Sobolev inequality below:

$$\int_{C([0,T],\mathbb{R}^d)} F^2 \log\frac{F^2}{\mathbb{E}_{\mathbb{P}_x} F^2}\, d\mathbb{P}_x \le 2e^{KT}\int_{C([0,T],\mathbb{R}^d)} |DF|_H^2\, d\mathbb{P}_x,$$

where DF be the Malliavin gradient and

$$H := \left\{\gamma(\cdot) := \int_0^\cdot h(s)\,ds;\ \|\gamma\|_H^2 = \int_0^T |h(s)|^2\,ds < +\infty\right\}$$

(the Cameron–Martin space). As the result of Otto and Villani [15] suggests that the log-Sobolev inequality implies the $T_2(C)$ inequality (that is proved on the smooth Riemannian manifold), we should have $\mathbb{P}_x \in T_2(C)$ on $C([0,T])$ w.r.t. the following pseudo-metric,

$$d_H(\gamma_1, \gamma_2) := \begin{cases} \|\gamma_1 - \gamma_2\|_H, & \text{if } \gamma_1 - \gamma_2 \in H, \\ +\infty, & \text{otherwise.} \end{cases}$$

This last pseudo metric is much larger than d_T used in the Corollary above. We shall give a simple proof of this last $T_2(C)$ inequality in Section 5.

Notice that as d_H above is only a pseudo-metric and $\|X.\|_H = +\infty$, a.s., Theorem 1.1 cannot be applied for $T_1(C)$ associated with d_H (since its sufficient part is no longer valid) and Theorem 2.3 (whose proof is based on Theorem 1.1) is no longer true w.r.t. d_H.

REMARK 4.4. Without essential change of proof, the same result holds if the locally Lipschitzian condition of σ, b is replaced by the well posedness of the martingale problem associated with $(\sigma\sigma^t, b)$, in the sense of Stroock–Varadhan.

REMARK 4.5. If the condition on the drift b in (4.2) is substituted by $\langle x, b(x)\rangle \le B(1+|x|^2)\ \forall x \in \mathbb{R}^d$, then with the same proof as above, we can prove that $\mathbb{E}\exp(\delta \sup_{t\in[0,T]} |X_t(x)|^2) < +\infty$ for some $\delta > 0$ depending on initial point. Hence, $\mathbb{P}_x$ satisfies the T_1-inequality with a constant $C = C_x$ depending on x.

Note the following drawback of the previous corollary: the constant C in the T_1 inequality obtained through Theorem 2.3 via inequality (2.2) is of order $e^{\beta T}$ which is not natural in regard of the results obtained via weakly dependent sequences. We now show how Theorem 2.5 enables us to get the correct order.

We know from Corollary 4.1 that the law of $(X_t(x))_{t\in[0,1]}$ satisfies the T_1-inequality with a constant C independent of x. In other words, the transition kernel of the Markov chain $Y_n := X_{[n,n+1]}$ valued in $C([0,1], \mathbb{R}^d)$ satisfies $T_1(C)$. Let us check (C1′) below.

Given two different initial points $x, \tilde{x}$, let

$$\hat{X}_t := X_t(x) - X_t(\tilde{x}),$$
$$\hat{\sigma}_t = \sigma\big(X_t(x)\big) - \sigma\big(X_t(\tilde{x})\big), \qquad \hat{b}_t = b\big(X_t(x)\big) - b\big(X_t(\tilde{x})\big).$$

By Ito's formula,

$$|\hat{X}_t|^2 = |x-\tilde{x}|^2 + \int_0^t \big(\mathrm{tr}(\hat{\sigma}_s\hat{\sigma}_s^t) + 2\langle \hat{X}_s, \hat{b}_s\rangle\big)\,ds + M_t,$$

where (M_t) is a local martingale with $M_0 = 0$, whose quadratic variational process is given by

$$[M]_t = 4\int_0^t \langle \hat{X}_s, (\hat{\sigma}_s\hat{\sigma}_s^t)\hat{X}_s\rangle\,ds.$$

Let $\hat{\tau}_n := \inf\{t \ge 0; |\hat{X}_t| \vee [M]_t = n\}$. If there is $\delta > 0$ such that

$$\begin{aligned} &\tfrac{1}{2}\,\mathrm{tr}\big[\big(\sigma(x)-\sigma(\tilde{x})\big)\big(\sigma(x)-\sigma(\tilde{x})\big)^t\big] + \langle x-\tilde{x}, b(x)-b(\tilde{x})\rangle \\ &\qquad \le -\delta|x-\tilde{x}|^2 \qquad \forall x, \tilde{x} \in \mathbb{R}^d, \end{aligned} \tag{4.5}$$

then

$$\mathbb{E}\left|\hat{X}_{t\wedge\hat{\tau}_n}\right|^2 \le |x-\tilde{x}|^2 - 2\delta\int_0^t \mathbb{E}\left|\hat{X}_{s\wedge\hat{\tau}_n}\right|^2 ds.$$

This entails by Gronwall's inequality and Fatou's lemma,

$$\mathbb{E}|X_t(x)-X_t(\tilde{x})|^2 = \mathbb{E}\left|\hat{X}_t\right|^2 \le |x-\tilde{x}|^2 e^{-2\delta t} \qquad \forall t\ge 0. \tag{4.6}$$

Moreover, if σ is globally Lipchitzian, then by Burkholder–Davis–Gundy's inequality and Gronwall's inequality, we obtain easily from the estimate above that

$$\mathbb{E}\sup_{t\le s\le t+1}|X_s(x)-X_s(\tilde{x})|^2 \le K|x-\tilde{x}|^2 e^{-2\delta t}$$

for some constant K. Thus, the Markov chain $Y_n := X_{[n,n+1]}$ valued in $C([0,1],\mathbb{R}^d)$ satisfies (C1′) too. Consequently, we obtain by Theorem 2.11, the following:

PROPOSITION 4.6. *Assume* (4.2), (4.5) *and σ is globally Lipchitzian. Then there is some constant $C>0$ such that for any $n\ge 1$ and any initial point x, the law $\mathbb{P}_x$ of $(X_t(x))_{t\in[0,n]}$ on $C([0,n],\mathbb{R}^d)$ satisfies the inequality $T_1(C\cdot n)$ w.r.t. the metric*

$$d(\gamma_1,\gamma_2) := \sum_{k=0}^{n-1}\sup_{k\le t\le k+1}|\gamma_1(t)-\gamma_2(t)|.$$

REMARK 4.7. Let (P_t) be the semigroup of transition probability kernels of our diffusion (X_t). Notice that under (4.5), we have (4.6) which entails not only the existence and uniqueness of the invariant probability measure μ of (P_t), but also

$$W_2^d\big(P_t(x,\cdot),P_t(\tilde{x},\cdot)\big) \le e^{-\delta t}|x-\tilde{x}|,$$

which gives us the exponential convergence below:

$$W_2^d\big(P_t(x,\cdot),\mu\big) \le e^{-\delta t}\left(\int |x-\tilde{x}|^2\,d\mu(\tilde{x})\right)^{1/2} \qquad \forall x\in\mathbb{R}^d, t>0.$$

Let us present a Hoeffding type inequality for

$$F(\gamma) := \int_0^n V(\gamma(t))\,dt,$$

where $V:\mathbb{R}^d\to\mathbb{R}$ satisfies $\|V\|_{\mathrm{Lip}}\le\alpha$. For such V, $\|F\|_{\mathrm{Lip}}\le\alpha$ w.r.t. the metric given in the proposition above. Hence, by Theorem 1.1, Proposition 4.6 entails

$$\mathbb{P}\left(\int_0^n \big[V\big(X_t(x)\big)-\mathbb{E}V\big(X_t(x)\big)\big]\,dt > r\right) \le \exp\left(-\frac{r^2}{2nC}\right) \qquad \forall r>0.$$

5. A direct approach to $T_2(C)$ for SDEs via stochastic calculus.

5.1. *T_2-inequality of the Wiener measure w.r.t. the Cameron–Martin metric.* Let us extend the T_2-inequality of the Gaussian measure due to Talagrand to the Wiener measure $\mathbb{P}$ on $C([0,T],\mathbb{R}^d)$, by means of Girsanov formula. Given $\mathbb{Q} \ll \mathbb{P}$ such that $H(\mathbb{Q}/\mathbb{P}) < +\infty$, then under $\mathbb{Q}$, there exist a Brownian motion (B_t) and a predictable process (β_t) such that the coordinates system (γ_t) of $C([0,T],\mathbb{R}^d)$ verifies

$$d\gamma_t = dB_t + \beta_t(\gamma)\,dt, \gamma_0 = 0.$$

Moreover, it is well known that [see the proof of (5.7) below in a much more complicated case]

$$H(\mathbb{Q}/\mathbb{P}) = \tfrac{1}{2}\mathbb{E}^{\mathbb{Q}}\int_0^T |\beta_t|^2(\gamma)\,dt. \tag{5.1}$$

Consider the Girsanov transformation $\Phi(\gamma) := \gamma(\cdot) - \int_0^{\cdot} \beta_t(\gamma)\,dt$. Then the law of $(\gamma, \Phi(\gamma))$ under $\mathbb{Q}$ is a coupling of $(\mathbb{Q},\mathbb{P})$. Hence, w.r.t. the Cameron–Martin metric d_H given in Remark 4.2,

$$\big(W_2^{d_H}(\mathbb{Q},\mathbb{P})\big)^2 \le \mathbb{E}^{\mathbb{Q}} d_H\big(\gamma,\Phi(\gamma)\big)^2 = \mathbb{E}^{\mathbb{Q}}\int_0^T |\beta_t|^2(\gamma)\,dt = 2H(\mathbb{Q}/\mathbb{P}), \tag{5.2}$$

that is, $\mathbb{P} \in T_2(1)$ on $(C([0,T],\mathbb{R}^d), d_H)$. We see now why this is sharp. Indeed, if β_t is determinist (or, equivalently, $\mathbb{Q}$ is a Gaussian measure), we claim that

$$[W_2^{d_H}(\mathbb{Q},\mathbb{P})]^2 = \int_0^T |\beta_t|^2\,dt = 2H(\mathbb{Q}/\mathbb{P}).$$

This follows by the following observation:

LEMMA 5.1. *Let X be a random variable valued in a Banach space E and H be a separable Hilbert space continuously embedded in E. Then for any element $h \in H$,*

$$W_2^{d_H}(\mathbb{P}_X, \mathbb{P}_{X+h}) = \|h\|_H,$$

where $\mathbb{P}_X$ is the law of X, $d_H(x,y) := \|x-y\|_H$ if $x-y \in H$ and $+\infty$ otherwise.

PROOF. At first $[W_2^{d_H}(\mathbb{P}_X,\mathbb{P}_{X+h})]^2 \le \mathbb{E}\|X-(X+h)\|_H^2 = \|h\|_H^2$. To show the inverse inequality, let π be a probability measure on E^2 such that its marginal laws are, respectively, laws of X and $X+h$, and $\iint \|y-x\|_H^2 \pi(dx,dy) < +\infty$. Since $y-(x+h)$ is centered in the sense that $\mathbb{E}^{\pi}\langle e_i, y-(x+h)\rangle_H = 0$ where (e_i) is an orthonormal basis of H, we have by Jensen's inequality,

$$\iint \|y-x\|_H^2 \pi(dx,dy) = \iint \big\|h + \big(y-(x+h)\big)\big\|_H^2 \pi(dx,dy) \ge \|h\|_H^2,$$

the desired result. □

Considering the mapping $\Psi(\gamma) = \gamma(T)$, which verifies

$$|\Psi(\gamma_1) - \Psi(\gamma_2)| \le \sqrt{T d_H(\gamma_1, \gamma_2)},$$

we get by Lemma 2.1 and (5.2) that $\mathcal{N}(0, T I_d) \in T_2(C)$ on $\mathbb{R}^d$ w.r.t. the Euclidean metric with the sharp constant $C = T$ (the theorem of Talagrand).

REMARK 5.2. Gentil [7] proved the dual (functional) version of the T_2-inequality of the Wiener measure w.r.t. the Cameron–Martin metric by generalizing the approach in [1]. The proof here is completely different and seems to be simpler and direct.

REMARK 5.3. Recall the method of Talagrand for proving his $T_2(C)$ for $\mathcal{N}(0, I_d)$. At first by independent tensorization, he reduces to dimension 1. And in dimension one, he uses the optimal transportation of Fréchet putting forward $\gamma = \mathcal{N}(0,1)$ to $f\, d\gamma$, and a direct integration by parts yields miraculously his $T_2(C)$. The method here is completely different, we use the Girsanov transformation which puts $\mathbb{Q}$ back to $\mathbb{P}$ instead of an (eventual) optimal transportation putting $\mathbb{P}$ forward to $\mathbb{Q}$. The approach of Talagrand is generalized recently by Feyel and Ustunel [6] who succeed to construct the optimal transportation from $\mathbb{P}$ to $\mathbb{Q}$ on an abstract Wiener space $(W, H, \mathbb{P})$.

We learned very recently (10 monthes after our first version) from Fang that the method of Girsanov transformation here has been used by Feyel and Ustunel [5] in a less elementary manner. So the result of this paragraph is due to them.

5.2. *T_2-inequality of diffusions w.r.t. the Cameron–Martin metric.* We now generalize the preceding argument to solution of the SDE

$$dX_t = dB_t + b(X_t)\, dt, \qquad X_0 = x \in \mathbb{R}^d,$$

where (B_t) is a $\mathbb{R}^d$-valued Brownian motion. We assume that $b \in C^1$ and

$$\|\nabla b\| \le K.$$

For any path $\gamma \in C([0,T], \mathbb{R}^d)$ with $\gamma(0) = 0$, let $\Phi(\gamma) = \eta$ be the solution of

$$\eta(t) = x + \gamma(t) + \int_0^t b(\eta(s))\, ds.$$

Then the solution of the SDE above is given by $X. = \Phi(B.)$. Hence, for proving the T_2-inequality of $X.$ w.r.t. the metric d_H, it is enough to show that Φ is d_H-Lipschitzian. To this end, consider

$$g(t) = \frac{d}{d\varepsilon}\Phi(\gamma + \varepsilon h)|_{\varepsilon=0},$$

where $h \in H$ is fixed. It satisfies

$$g(t) = h(t) + \int_0^t \nabla b(\eta(s))g(s)\,ds.$$

Its solution is given by

$$g(t) = \int_0^t J(s,t)h'(s)\,ds,$$

where $J(s,t)$ is the solution of the matrix differential equation

$$J(s,s) = I_d, \qquad \frac{d}{dt}J(s,t) = \nabla b(\eta(t))J(s,t). \tag{5.3}$$

Since $\nabla^s b \le BI_d$ for some $B \le K$, we have $|J(s,t)y| \le e^{B(t-s)}|y|\ \forall\, y \in \mathbb{R}^d$. Consequently,

$$|g(t)| \le \int_0^t e^{B(t-s)}|h'(s)|\,ds.$$

Thus, by Cauchy–Schwarz,

$$\begin{aligned}\|g\|_H^2 &\le 2\int_0^T |h'(t)|^2\,dt + 2\int_0^T |\nabla b(\eta(t))g(t)|^2\,dt\\ &\le 2\|h\|_H^2 + 2K^2\int_0^T \left[\int_0^t e^{B(t-s)}|h'(s)|\,ds\right]^2 dt.\end{aligned}$$

Note that

$$\begin{aligned}\int_0^T \left[\int_0^t e^{B(t-s)}|h'(s)|\,ds\right]^2 dt &= \int_0^T\int_0^T |h'(u)||h'(v)|\left[\int_{u\vee v}^T e^{2Bt-(u+v)}\,dt\right]du\,dv\\ &= \langle \Gamma|h'|, |h'|\rangle_{L^2([0,T])},\end{aligned}$$

where

$$\Gamma(u,v) = \begin{cases} e^{-B(u+v)}\dfrac{e^{2BT} - e^{2B(u\vee v)}}{2B}, & \text{if } B \ne 0,\\ T - u\vee v, & \text{if } B = 0\end{cases}$$

and $\Gamma f(u) := \int_0^T \Gamma(u,v)f(v)\,dv$. Let $\lambda_{\max}(\Gamma)$ be the largest eigenvalue of Γ in $L^2([0,T])$. We have $\lambda_{\max}(\Gamma) \le \|\Gamma\|_1$, the norm of Γ in $L^1([0,T])$. It is easy to get $\|\Gamma\|_1 \le \frac{1}{B^2}$ if $B < 0$, $\|\Gamma\|_1 \le \frac{e^{2BT}}{2B^2}$ if $B > 0$, and $\|\Gamma\|_1 = \frac{T^2}{2}$ if $B = 0$. Thus, setting

$$\alpha^2 := \alpha^2(T,K,B) = \begin{cases} 2\left(1 + \dfrac{K^2}{B^2}\right), & \text{if } B < 0,\\ 2\left(1 + K^2\dfrac{e^{2BT}}{2B^2}\right), & \text{if } B > 0,\\ 2\left(1 + \dfrac{K^2T^2}{2}\right), & \text{if } B = 0;\end{cases} \tag{5.4}$$

we get by the estimates above that $\|g\|_H^2 \le \alpha^2 \|h\|_H^2$, that is, $\|\Phi\|_{\mathrm{Lip}(d_H)} \le \alpha$. Thus, Lemma 2.1 (which remains valid for the pseudo-metric d_H) together with the T_2-inequality for the Wiener measure gives us the following:

PROPOSITION 5.4. *Assume $\nabla^s b \le BI_d$ and $\|\nabla b\| \le K$, then for every initial point x, $\mathbb{P}_x \in T_2(\alpha^2)$ on $C([0,T],\mathbb{R}^d)$ w.r.t. the metric d_H, where α^2 is given by* (5.4).

REMARK 5.5. Of course, the estimate of $\|\Phi\|_{\mathrm{Lip}(d_H)} \le \alpha$ together with the log-Sobolev inequality of Gross for the Wiener measure gives us also

$$\int_{C([0,T],\mathbb{R}^d)} F^2 \log \frac{F^2}{\mathbb{E}_{\mathbb{P}_x} F^2}\, d\mathbb{P}_x \le 2\alpha^2 \int_{C([0,T],\mathbb{R}^d)} |DF|_H^2\, d\mathbb{P}_x,$$

which is better than the Capitaine–Hsu–Ledoux's estimate in Remark 4.3 when $B < 0$.

It is interesting to investigate whether this proposition and the corresponding log-Sobolev inequality continue to hold in the case where $\nabla^s b \le BI_d$ with $B \le 0$ without condition $\|\nabla b\| \le K$.

5.3. *T_2-inequality of diffusions w.r.t. the L^2-metric.* Perhaps the most elementary metric on $C([0,T],\mathbb{R}^d)$ is the following $L^2[0,T]$-metric,

$$d_2(\gamma_1,\gamma_2) := \sqrt{\int_0^T |\gamma_1(t) - \gamma_2(t)|^2\, dt}.$$

Indeed, the argument leading to the T_2-inequality of the Wiener measure will yield the following robust T_2-inequality w.r.t. the metric above:

THEOREM 5.6. *Assume that σ, b are locally Lipschitzian and satisfy* (4.5) *for some $\delta > 0$, and $\|\sigma\|_\infty := \sup\{|\sigma(x)z|;\ x \in \mathbb{R}^d, |z| \le 1\} < +\infty$. Then $\mathbb{P}_x \in T_2(C)$ on $C([0,T],\mathbb{R}^d)$ w.r.t. the L^2-metric d_2 above for all $x \in \mathbb{R}^d$ and $T > 0$, where the constant C is given by*

$$C := \frac{\|\sigma\|_\infty^2}{\delta^2}.$$

Moreover, $P_T(x,\cdot) \in T_2(\frac{\|\sigma\|_\infty^2}{2\delta})$ on $\mathbb{R}^d$, as well as the unique invariant probability measure μ of (P_t).

REMARK 5.7. The two T_2-inequalities in this theorem are both sharp. Indeed, let $d = 1$, $\sigma(x) = 1$, $b(x) = x/2$, that is, (X_t) is the standard real Ornstein–Uhlenbeck process, whose invariant measure is $\mathcal{N}(0,1)$. By this proposition, $\mu \in T_2(C)$ with $C = \|\sigma\|_\infty^2/2\delta = 1$, which is sharp.

For the sharpness of the T_2-inequality for $\mathbb{P}_x$ w.r.t. d_2, note that any Gaussian measure $\mathcal{N}(m,\Sigma)$ on $\mathbb{R}^n$ satisfies $T_2(C)$ with the sharp constant C being the largest eigenvalue $\lambda_{\max}(\Sigma)$ of the covariance matrix Σ. This can be extended easily to any Gaussian measure $\nu=\mathcal{N}(m,\Sigma)$ on any separable Hilbert space G, where the covariance matrix Σ is a Hilbert–Schmidt operator on G. Hence, if $(X_t)_{t\ge 0}$ is a Gaussian process with paths a.s. in $L^2([0,T],dt)$, then its law $\mathbb{P}$ satisfies the $T_2(C)$ on $L^2([0,T],dt)$ with the sharp constant $C=\lambda_{\max}(\Sigma)$, the largest eigenvalue of the operator

$$\Sigma f(s):=\int_0^T \mathrm{Cov}(X_s,X_t)f(t)\,dt \qquad \forall f\in L^2([0,T],dt).$$

For the Ornstein–Uhlenbeck process law $\mathbb{P}_0$ above starting from 0, $\mathrm{Cov}(X_s,X_t)=\exp(-|t-s|/2)-\exp(-(s+t)/2)$. In that case,

$$\lambda_{\max}(\Sigma)\ge \frac{\langle \Sigma \mathbb{1}_{[0,T]},\mathbb{1}_{[0,T]}\rangle}{T}\to 4 \qquad \text{as } T\to\infty.$$

Hence, the constant $C=\|\sigma\|^2/\delta^2=4$ in the T_2-inequality for $\mathbb{P}_0$ given by our theorem becomes sharp when $T\to+\infty$.

PROOF. We shall prove that for any $\varepsilon>0$, for any probability measure $\mathbb{Q}$ on $C([0,T],\mathbb{R}^d)$,

$$\big(W_2^{d_2}(\mathbb{Q},\mathbb{P}_x)\big)^2\le 2\frac{(1-e^{(\varepsilon-2\delta)T})\|\sigma\|_\infty^2}{\varepsilon(2\delta-\varepsilon)}H(\mathbb{Q}/\mathbb{P}) \tag{5.5}$$

and for any probability measure ν on $\mathbb{R}^n$,

$$\big(W_2^{d_2}(\nu,P_T(x,\cdot))\big)^2\le 2\frac{\sup_{t\in[0,T]}e^{(\varepsilon-2\delta)t}\|\sigma\|_\infty^2}{\varepsilon}H\big(\nu/P_T(x,\cdot)\big). \tag{5.6}$$

Choosing $\varepsilon=\delta$ in (5.5), we get the first claim in the theorem; letting $\varepsilon\uparrow 2\delta$, we get $P_T(x,\cdot)\in T_2(\frac{\|\sigma\|_\infty^2}{2\delta})$ by (5.6) and then $\mu\in T_2(\frac{\|\sigma\|_\infty^2}{2\delta})$ by Lemma 2.2 and the fact that $P_T(x,\cdot)\to\mu$ as $T\to\infty$ (see Remark 4.7).

It is enough to prove (5.5) for $\mathbb{Q}\ll\mathbb{P}_x$ and $H(\mathbb{Q}/\mathbb{P}_x)<+\infty$. We divide its proof into two steps.

Step 1. We do at first some preparation of stochastic calculus. Let $(\Omega,\mathcal{F},\tilde{\mathbb{P}})$ be a complete probability space on which a n-dimensional Brownian motion $(B_t)=(B_t^j)_{j=1,\dots,n}$ is defined and let $\mathcal{F}_t=\mathcal{F}_t^B=\sigma(B_s,\ s\le t)^{\tilde{\mathbb{P}}}$ (completion by $\tilde{\mathbb{P}}$). Let $X_t(x)$ be the unique solution of (4.1) starting from x. Then the law of $X.(x)$ is $\mathbb{P}_x$. Consider

$$\tilde{\mathbb{Q}}:=\frac{d\mathbb{Q}}{d\mathbb{P}_x}(X.(x))\cdot\tilde{\mathbb{P}},\qquad M_t:=\mathbb{E}^{\tilde{\mathbb{P}}}\bigg(\frac{d\mathbb{Q}}{d\mathbb{P}}(X.(x))/\mathcal{F}_t\bigg)\qquad \forall t\in[0,T].$$

Remark that, as $\mathbb{Q}$ is a probability measure and the law of $X(x)$ under $\tilde{\mathbb{P}}$ is exactly $\mathbb{P}_x$, we have

$$\int_\Omega \frac{d\mathbb{Q}}{d\mathbb{P}_x}(X(x))\,d\tilde{\mathbb{P}} = \int_{C([0,T],R^d)} \frac{d\mathbb{Q}}{d\mathbb{P}_x}(w)\,d\mathbb{P}_x(w) = \mathbb{Q}\big(C([0,T],\mathbb{R}^d)\big) = 1.$$

(M_t) is a martingale can and will be chosen as a continuous martingale. Let $\tau := \inf\{t \in [0,T];\ M_t = 0\}$ with the convention that $\inf\varnothing := T+$, where $T+$ is an artificial added element larger than T, but smaller than any $a > T$. Then $\tilde{\mathbb{Q}}(\tau = T+) = 1$ and

$$M_t = \mathbb{1}_{t<\tau}\exp\big(L_t - \tfrac{1}{2}[L]_t\big),$$

where $L_t := \int_0^t \frac{dM_s}{M_s}$ $\forall t < \tau$. (L_t), being a $\tilde{\mathbb{P}}$-local martingale on $[0,\tau)$, can be represented in the following way: there is a predictable process $(\beta_t) = (\beta_t^j)_{0\le t<\tau}$ such that $\int_0^t |\beta_s|^2\,ds < +\infty$, $\tilde{\mathbb{P}}$-a.s. on $[t<\tau]$ and

$$L_t = \sum_{j=1}^n \int_0^t \beta_s^j\,dB_s^j = \int_0^t \langle \beta_s, dB_s\rangle \qquad \forall t < \tau.$$

Let $\tau_n = \inf\{t \in [0,\tau[;\ [L]_t = n\}$ with the same convention that $\inf\varnothing := T+$. It is elementary that $\tau_n \uparrow \tau$, $\tilde{\mathbb{P}}$-a.s. Hence, by martingale convergence,

$$\begin{aligned} H(\mathbb{Q}/\mathbb{P}) = H(\tilde{\mathbb{Q}}/\tilde{\mathbb{P}}) &= \mathbb{E}^{\tilde{\mathbb{P}}} M_T \log M_T = \lim_{n\to\infty} \mathbb{E}^{\tilde{\mathbb{P}}} M_{T\wedge\tau_n} \log M_{T\wedge\tau_n} \\ &= \lim_{n\to\infty} \mathbb{E}^{\tilde{\mathbb{Q}}}\big(L_{T\wedge\tau_n} - \tfrac{1}{2}[L]_{T\wedge\tau_n}\big). \end{aligned}$$

By Girsanov's formula, $(L_{t\wedge\tau_n} - [L]_{t\wedge\tau_n})_{t\in[0,T]}$ is a $\tilde{\mathbb{Q}}$-local martingale, then a true martingale since its quadratic variation process under $\tilde{\mathbb{Q}}$, being again $([L]_{t\wedge\tau_n})$, is bounded by n. Consequently, $\mathbb{E}^{\tilde{\mathbb{Q}}}(L_{T\wedge\tau_n} - [L]_{T\wedge\tau_n}) = 0$. Substituting it into the preceding equality and noting that $\tilde{\mathbb{Q}}(\tau_n \uparrow \tau = T+) = 1$, we get by monotone convergence,

$$H(\mathbb{Q}/\mathbb{P}) = \tfrac{1}{2}\lim_{n\to\infty} \mathbb{E}^{\tilde{\mathbb{Q}}}[L]_{T\wedge\tau_n} = \tfrac{1}{2}\mathbb{E}^{\tilde{\mathbb{Q}}}[L]_T = \tfrac{1}{2}\mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^T |\beta_t|^2\,dt. \tag{5.7}$$

Notice that this is an extension of (5.1).

Step 2. By Girsanov's theorem,

$$\tilde{B}_t := B_t - \int_0^t \beta_s\,ds$$

is a $\tilde{\mathbb{Q}}$-local martingale with $[\tilde{B}^i, \tilde{B}^j]_t = [B^i, B^j]_t = \mathbb{1}_{i=j}t$, hence, a Brownian motion under $\tilde{\mathbb{Q}}$. Under $\tilde{\mathbb{Q}}$, $X_t = X_t(x)$ verifies

$$dX_t = \sigma(X_t)\,d\tilde{B}_t + b(X_t)\,dt + \sigma(X_t)\beta_t\,dt, \qquad X_0 = x.$$

We now consider the solution Y_t (under $\tilde{\mathbb{Q}}$) of

$$dY_t = \sigma(Y_t)\, d\tilde{B}_t + b(Y_t)\, dt, \qquad Y_0 = x.$$

The law of $(Y_t)_{t\in[0,T]}$ under $\tilde{\mathbb{Q}}$ is exactly $\mathbb{P}_x$. In other words, (X, Y) under $\tilde{\mathbb{Q}}$ is a coupling of $(\mathbb{Q}, \mathbb{P}_x)$.

Setting

$$\hat{X}_t := X_t - Y_t, \qquad \hat{\sigma}_t := \sigma(X_t) - \sigma(Y_t), \qquad \hat{b}_t := b(X_t) - b(Y_t),$$

we have

$$d|\hat{X}_t|^2 = [2\langle \hat{X}_t, \hat{b}_t + \sigma(X_t)\beta_t\rangle + \mathrm{tr}(\hat{\sigma}_t\sigma_t^t)]\, dt + 2\langle \hat{X}_t, \hat{\sigma}_t\, d\tilde{B}_t\rangle. \tag{5.8}$$

Letting $\hat{\tau}_n := \inf\{t \in [0, T]; |\hat{X}_t| = n\}$, we have that for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{E}^{\tilde{\mathbb{Q}}}|\hat{X}_{t\wedge\hat{\tau}_n}|^2 &\le -2\delta \int_0^t \mathbb{E}^{\tilde{\mathbb{Q}}}|\hat{X}_{s\wedge\hat{\tau}_n}|^2\, ds + 2\mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^{t\wedge\hat{\tau}_n} \langle \hat{X}_s, \sigma(X_s)\beta_s\rangle\, ds \\ &\le (\varepsilon - 2\delta)\int_0^t \mathbb{E}^{\tilde{\mathbb{Q}}}|\hat{X}_{s\wedge\hat{\tau}_n}|^2\, ds + \frac{1}{\varepsilon}\mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^t \|\sigma\|_\infty^2|\beta_s|^2\, ds.\end{aligned}$$

Gronwall's lemma, together with Fatou's lemma, gives us

$$\mathbb{E}^{\tilde{\mathbb{Q}}}|\hat{X}_t|^2 \le \frac{\|\sigma\|_\infty^2}{\varepsilon}\mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^t e^{(\varepsilon-2\delta)(t-s)}|\beta_s|^2\, ds \qquad \forall t > 0. \tag{5.9}$$

Thus,

$$\begin{aligned}\big(W_2^{d_2}(\mathbb{Q}, \mathbb{P}_x)\big)^2 &\le \mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^T |\hat{X}_t|^2\, dt \\ &\le \frac{\|\sigma\|_\infty^2}{\varepsilon}\mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^T |\beta_s|^2\, ds \int_s^T e^{(\varepsilon-2\delta)(t-s)}\, dt \\ &\le \frac{\|\sigma\|_\infty^2}{\varepsilon}\cdot\frac{1-e^{(2\delta-\varepsilon)T}}{2\delta-\varepsilon}\mathbb{E}^{\tilde{\mathbb{Q}}}\int_0^T |\beta_s|^2\, ds,\end{aligned}$$

the desired (5.5). For (5.6), notice that by the key remark (2.1),

$$H\big(\nu/P_T(x,\cdot)\big) = \inf\big\{H\big(\mathbb{Q}|_{C([0,T],\mathbb{R}^d)}/\mathbb{P}_x|_{C([0,T],\mathbb{R}^d)}\big);\ Q_T := \mathbb{Q}(x_T \in \cdot) = \nu\big\}.$$

And for each such Q, define $\tilde{Q}$ as before, we have

$$\big[W_2^d\big(\nu, P_T(x, dy)\big)\big]^2 \le E^{\tilde{Q}}|\hat{X}_T|^2$$

and conclude using (5.9). $\square$

REMARK 5.8. After the first version was submitted, we learned from M. Ledoux the work of Wang [20] who obtained the $T_2(C)$ w.r.t. the L^2-metric for the elliptic diffusions with lower bounded Γ_2 condition of Bakry on a Riemannian manifold. His method consists of a continuous time tensorization of the $T_2(C)$

of the heat kernels (which is true by the log-Sobolev inequality due to Bakry). Hence, the method and the result here are very different from his: the volatility coefficient σ could be completely degenerated in Theorem 5.6, and our proof does not rely on the log-Sobolev inequality which is unknown in our context.

REMARK 5.9. By the proof above, we see that (5.5) and (5.6) hold under (4.5) even with $\delta \le 0$, except now the T_2-constant goes to infinity as $T \to +\infty$.

REMARK 5.10. The local Lipschitzian condition on σ, b in this theorem can be substituted by their continuity together with the well-posedness of the martingale problem associated with $(\sigma\sigma^t, b)$. Indeed, one can find (σ^n, b^n) tending locally uniformly to (σ, b), such that (σ^n, b^n) is locally Lipschitzian, $\|\sigma^n\|_\infty \le \|\sigma\|_\infty$ and verifies condition (4.5) with the same δ. Now the desired result follows from Theorem 5.6 and Lemma 2.2.

As indicated in [1], many interesting consequences can be derived from this result. For instance

COROLLARY 5.11. *Under the assumptions of Theorem* 5.6, *we have for any* $T > 0$,

(a) *for any smooth cylindrical function* F *on* $G := L^2([0,T], dt; \mathbb{R}^d) \supset C([0,T], \mathbb{R}^d)$, *that is,*

$$F \in \mathcal{F}C_b^\infty := \{f(\langle\gamma, h_1\rangle, \dots, \langle\gamma, h_n\rangle); n \ge 1, h_i \in \tilde{H}, f \in C_b^\infty(\mathbb{R}^n)\}$$

[*where* $\langle\gamma_1, \gamma_2\rangle := \int_0^T \gamma_1(t)\gamma_2(t)\,dt$], *the following Poincaré inequality holds*:

$$\operatorname{Var}_{\mathbb{P}_x}(F) \le \frac{\|\sigma\|_\infty^2}{\delta^2} \int_{C([0,T],\mathbb{R}^d)} \|\nabla F(\gamma)\|_G^2 \, d\mathbb{P}_x(\gamma), \tag{5.10}$$

where $\operatorname{Var}_{\mathbb{P}_x}(F)$ *is the variance of* F *under law* $\mathbb{P}_x$, *and* $\nabla F(\gamma) \in G$ *is the gradiant of* F *at* γ.

(b) *For any* $g \in C_b^\infty(\mathbb{R}^d)$,

$$\operatorname{Var}_{P_T(x,\cdot)}(g) \le \frac{\|\sigma\|_\infty^2}{2\delta} \int_{\mathbb{R}^d} |\nabla g(y)|^2 P_T(x, dy). \tag{5.11}$$

(c) (Inequality of Tsirel'son type.) *For any nonempty subset* K *in* G *such that* $Z(\gamma) := \sup_{h\in K}\langle\gamma, h\rangle \in L^1(\mathbb{P}_x)$, *then*

$$\int \exp\left(\frac{\delta^2}{\|\sigma\|_\infty^2} \sup_{h\in K}\left[\langle\gamma, h\rangle - \frac{|h|_G^2}{2}\right]\right) d\mathbb{P}_x \le \exp\left(\frac{\delta^2}{\|\sigma\|_\infty^2}\mathbb{E}^{\mathbb{P}_x} Z\right). \tag{5.12}$$

(d) (Inequality of Hoeffding type.) *For any* $V:\mathbb{R}^d \to \mathbb{R}$ *such that* $\|V\|_{\mathrm{Lip}} \le \alpha$,

$$\mathbb{P}\biggl(\frac{1}{T}\int_0^T V\bigl(X_t(x)\bigr)\,dt - \mathbb{E}\frac{1}{T}\int_0^T V\bigl(X_t(x)\bigr)\,dt > r\biggr)$$
$$\le \exp\biggl(-\frac{Tr^2\|\sigma\|_\infty^2}{2\alpha^2\delta^2}\biggr) \qquad \forall r > 0.$$

PROOF. For part (a), for any $F(\gamma) = f(\langle\gamma, h_1\rangle, \dots, \langle\gamma, h_n\rangle) \in \mathcal{F}C_b^\infty$, we may assume without loss of generality that $h_1, \dots, h_n$ are orthonormal. In such case,

$$\Phi:\gamma \to (\langle\gamma, h_1\rangle, \dots, \langle\gamma, h_n\rangle), \qquad G \to \mathbb{R}^n$$

is Lipschitzian with $\|\Phi\|_{\mathrm{Lip}} \le 1$. Hence, $\nu := \mathbb{P}_x \circ \Phi^{-1} \in T_2(\|\sigma\|_\infty^2/\delta^2)$ on $\mathbb{R}^n$ by Lemma 2.1. Thus, the result of [1], Section 4.1 entails

$$\mathrm{Var}_{\mathbb{P}_x}(F) = \mathrm{Var}_\nu(f) \le \frac{\|\sigma\|_\infty^2}{\delta^2}\int_{\mathbb{R}^n} |\nabla f|^2\,d\nu$$
$$= \frac{\|\sigma\|_\infty^2}{\delta^2}\int_{C([0,T],\mathbb{R}^d)} \|\nabla F(\gamma)\|_G^2\,d\mathbb{P}_x(\gamma).$$

Part (b) is a consequence of Theorem 5.6 by [1], Section 4.1. One can derive part (c) from Theorem 5.6 by the same argument as in the finite-dimensional case given in [1], Section 6.1. For part (d), note that $T_2(C) \Rightarrow T_1(C)$. Moreover, the function $F(\gamma) := (1/T)\int_0^T V(\gamma(t))\,dt$ on $C([0,T],\mathbb{R}^d)$ is Lipschitzian w.r.t. the L^2-metric and $\|F\|_{\mathrm{Lip}} \le \alpha/\sqrt{T}$. Hence, part (d) follows from Theorem 1.1. □

REMARK 5.12. Let us compare the $T_2(C)$-inequality on $C([0,T],\mathbb{R}^d)$ w.r.t. the L^2-metric d_2 or the Cameron–Martin metric d_H, denoted, respectively, by $T_2(C/d_2)$, $T_2(C/d_H)$.

(a) If $\gamma_1(0) = \gamma_2(0)$, then $d_2(\gamma_1,\gamma_2) \le \frac{2T}{\pi} d_H(\gamma_1,\gamma_2)$ by the classical Poincaré inequality. Hence, if the law $\mathbb{P}_x$ of our diffusion starting from x verifies $T_2(C/d_H)$ on $C([0,T],\mathbb{R}^d)$, then $\mathbb{P}_x \in T_2(C(4T^2/\pi^2)/d_2)$ on $C([0,T],\mathbb{R}^d)$. That order T^2 in the last T_2-inequality is of correct order. For example, for the real Wiener measure $\mathbb{P}$, we see by Section 5.1 that $\mathbb{P} \in T_2(1/d_H)$ on $C([0,T],\mathbb{R}^d)$, but the largest eigenvalue $\lambda_{\max}(\Gamma)$ of the covariance function $\Gamma(s,t) = s \wedge t$ in $L^2([0,T])$ verifies

$$\lambda_{\max}(\Gamma) \ge \frac{\langle\Gamma\mathbb{1}_{[0,T]}, \mathbb{1}_{[0,T]}\rangle}{T} = \frac{T^2}{3}.$$

Thus, by the same analysis as in Remark 5.7, $\mathbb{P} \in T_2(CT^2/d_2)$ with $4/\pi^2 \ge C = \lambda_{\max}(\Gamma) \ge 1/3$.

(b) The contribution of $|\gamma_1(t) - \gamma_2(t)|$ to the L^2-metric is homogeneous in time t, but not at all to the Cameron–Martin metric d_H. This is the principal reason for

(b.1) The $T_2(C/d_H)$ is well adapted to the small time asymptotics of the diffusions, but not for their large time asymptotics. For instance, if $\mathbb{P}_x \in T_2(C/d_H))$, since for $Z(\gamma) = \sup_{0\le t\le T} \|\gamma(t) - \gamma(0)\|$, $\|Z\|_{\mathrm{Lip}(d_H)} \le \sqrt{T}$, then by Theorem 1.1 (its necessary part remains true for d_H-Lipchitzian function F which is, moreover, integrable, by following the proof in [2]),

$$\mathbb{P}_x\left(\sup_{0\le t\le T} |X_t(x) - x| - \mathbb{E}^x \sup_{0\le t\le T} |X_t(x) - x| > r\right) \le \exp\left(-\frac{r^2}{2CT}\right)$$

which is of the correct order when $T \to 0+$, but completely meaningless for T large. See [21] for the nonadaptability of the log-Sobolev inequality w.r.t. d_H for the large time asymptotics of the diffusions.

(b.2) In contrary, we have seen that the $T_2(C/d_2)$ is very well adapted for the large time asymptotics of the diffusions.

REMARK 5.13. Theorem 5.6, together with Corollary 3.5, is our main new example for which $T_2(C)$ is true but the inequality of log-Sobolev is unknown. They are our (very partial) answer to Question 3 in the Introduction. We believe that in the situations of Theorem 5.6 and Corollary 3.5, the log-Sobolev inequality may fail without further regularity assumptions on the volatility coefficient σ.

Acknowledgments. We are grateful to P. Cattiaux, S. Fang, M. Ledoux, C. Villani and F. Y. Wang for their comments on the first version. We are particularly indebted to the referee for his very numerous and conscientious remarks which led to a complete re-organization of the paper and improved the readability of the paper and for the suggestion of the simplified proofs of Lemma 2.2 and of the example at the end of Remark 2.4.

REFERENCES

[1] BOBKOV, S., GENTIL, I. and LEDOUX, M. (2001). Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl. (9)* **80** 669–696.

[2] BOBKOV, S. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28.

[3] CAPITAINE, M., HSU, E. P. and LEDOUX, M. (1997). Martingale representation and a simple proof of logarithmic Sobolev inequality on path spaces. *Electron Comm. Probab.* **2** 71–81.

[4] DEMBO, A. (1997). Information inequalities and concentration of measure. *Ann. Probab.* **25** 927–939.

[5] FEYEL, D. and USTUNEL, A. S. (2002). Measure transport on Wiener space and Girsanov theorem. *C. R. Acad. Sci. Paris Sér. I Math.* **334** 1025–1028.

[6] FEYEL, D. and USTUNEL, A. S. (2004). The Monge–Kantorovitch problem and Monge–Ampère equation on Wiener space. *Probab. Theory Related Fields*. To appear.

[7] GENTIL, I. (2001). Inégalités de Sobolev logarithmiques et hypercontractivité en mécanique statistique et en E.D.P. Thèse de doctorat, Univ. Paul Sabatier Toulouse.

[8] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. Amer. Math. Soc., Providence, RI.

[9] LEDOUX, M. (2002). Concentration, transportation and functional inequalities. Preprint.

[10] MARTON, K. (1996). Bounding $\overline{d}$-distance by information divergence: A method to prove measure concentration. *Ann. Probab.* **24** 857–866.

[11] MARTON, K. (1997). A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** 556–571.

[12] MARTON, K. (1998). Measure concentration for a class of random processes. *Probab. Theory Related Fields* **110** 427–439.

[13] MASSART, P. (2003). Concentration inequalities and model selection. In *Saint-Flour Summer School*.

[14] MCDIARMID, C. (1989). On the method of bounded differences. *Surveys of Combinatorics* (J. Siemons, ed.). *London Math. Soc. Lecture Notes Ser.* **141** 148–188.

[15] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400.

[16] RIO, E. (2000). Inégalités de Hoeffding pour les fonctions Lipschitziennes de suites dépendantes. *C. R. Acad. Sci. Paris Sér. I Math.* **330** 905–908.

[17] SAMSON, P. M. (2000). Concentration of measure inequalities for Markov chains and ϕ-mixing process. *Ann. Probab.* **1** 416–461.

[18] TALAGRAND, M. (1996). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** 587–600.

[19] VILLANI, C. (2003). *Topics in Optimal Transportation*. Amer. Math. Soc., Providence, RI.

[20] WANG, F. Y. (2002). Transportation cost inequalities on path spaces over Riemannian manifolds. *Illinois J. Math.* **46** 1197–1206.

[21] WU, L. (2000). A deviation inequality for non-reversible Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.* **36** 435–445.

[22] WU, L. (2002). Essential spectral radius for Markov semigroups. I: Discrete time case. *Probab. Theory Related Fields* **128** 255–321.

LABORATOIRE
DE MATHÉMATIQUES APPLIQUÉES
CNRS-UMR 6620
UNIVERSITÉ BLAISE PASCAL
63177 AUBIÈRE
FRANCE
E-MAIL: djellout@math.univ-bpclermont.fr

CEREMADE
CNRS-UMR 7534
UNIVERSITÉ PARIS IX DAUPHINE
75775 PARIS
FRANCE
E-MAIL: guillin@ceremade.dauphine.fr

LABORATOIRE
DE MATHÉMATIQUES APPLIQUÉES
CNRS-UMR 6620
UNIVERSITÉ BLAISE PASCAL
63177 AUBIÈRE
FRANCE
AND
DEPARTMENT OF MATHEMATICS
WUHAN UNIVERSITY
430072
CHINA
E-MAIL: li-ming.wu@math.univ-bpclermont.fr

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2011, Vol. 47, No. 2, 450–465
DOI: 10.1214/10-AIHP360

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Lipschitzian norm estimate of one-dimensional Poisson equations and applications

Hacene Djellout[a] and Liming Wu[b]

[a]*Laboratoire de Mathématiques, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France.*
E-mail: Hacene.Djellout@math.univ-bpclermont.fr
[b]*Laboratoire de Mathématiques, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France and Institute of Applied Mathematics, Academy of China, Beijing, China. E-mail: Li-Ming.Wu@math.univ-bpclermont.fr*

Received 20 May 2009; revised 19 January 2010; accepted 1 February 2010

Abstract. By direct calculus we identify explicitly the Lipschitzian norm of the solution of the Poisson equation $-\mathcal{L}G = g$ in terms of various norms of g, where $\mathcal{L}$ is a Sturm–Liouville operator or generator of a non-singular diffusion in an interval. This allows us to obtain the best constant in the L^1-Poincaré inequality (a little stronger than the Cheeger isoperimetric inequality) and some sharp transportation–information inequalities and concentration inequalities for empirical means. We conclude with several illustrative examples.

Résumé. Par un calcul direct, on identifie explicitement la norme Lipschitzienne de la solution de l'équation de Poisson $-\mathcal{L}G = g$ en terme de différentes normes de g, où $\mathcal{L}$ est l'opérateur de Sturm–Liouville ou le générateur d'une diffusion non singulière sur un intervalle. Ainsi, nous pouvons obtenir, d'une part la meilleure constante dans l'inégalité de Poincaré L^1 (une inégalité un peu plus forte que l'inégalité isopérimétrique de Cheeger) et d'autre part certaines inégalités de transport-information et de concentration fines pour la moyenne empirique. On conclut avec des exemples illustratifs.

MSC: 47B38; 60E15; 60J60; 34L15; 35P15

Keywords: Poisson equations; Transportation–information inequalities; Concentration and isoperimetric inequalities

1. Framework and introduction

Let I be an interval of $\mathbb{R}$ so that its interior $I^0 = (x_0, y_0)$ where $-\infty \le x_0 < y_0 \le +\infty$. Consider a Sturm–Liouville operator on I:

$$\mathcal{L} = a(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + b(x)\frac{\mathrm{d}}{\mathrm{d}x}$$

with the Neumann boundary condition at $\partial I = \{x_0, y_0\} \cap \mathbb{R}$, where $a, b : I \to \mathbb{R}$ are measurable and satisfy:

(A1) a, b are locally bounded (i.e., bounded on any compact subinterval of I);
(A2) $a(x) > 0$, $\mathrm{d}x$-a.e. and $1/a$ is locally $\mathrm{d}x$-integrable on I.

Here $\mathrm{d}x$ is the Lebesgue measure. On I^0, $\mathcal{L}$ can be rewritten as the Feller's form

$$\mathcal{L} = \frac{1}{m'(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{s'(x)}\frac{\mathrm{d}}{\mathrm{d}x}\right) = \frac{\mathrm{d}}{\mathrm{d}m}\frac{\mathrm{d}}{\mathrm{d}s}, \tag{1.1}$$

where m, s are respectively the speed and scale functions of Feller, which are absolutely continuous functions on I such that $\mathrm{d}x$-a.s.

$$s'(x) = \exp\left(-\int_c^x \frac{b(u)}{a(u)}\,\mathrm{d}u\right) \quad \text{and} \quad m'(x) = \frac{1}{a(x)s'(x)}, \tag{1.2}$$

where c is some fixed point in I. Let $C_0^\infty(I)$ be the space of infinitely differentiable real functions f on I with compact support and $\mathcal{D}$ be the space of all functions f in $C_0^\infty(I)$ such that $f'|_{\partial I} = 0$ (i.e., satisfying the Neumann boundary condition). The operator $\mathcal{L}$ defined on $\mathcal{D}$ is symmetric on $L^2(I, m)$, where m denotes also the measure $m'(x)\,\mathrm{d}x$. Let $(X_t\colon\ t \geq 0)$ be the diffusion on the interval I generated by $\mathcal{L}$ (the Neumann boundary condition corresponds to the reflection at the boundary ∂I). See [17] for background and precise definitions.

We will assume that:

(A3) the diffusion is non-explosive and positively recurrent, i.e., $m(I) = \int_I m'(y)\,\mathrm{d}y < +\infty$ and

$$\int_c^{y_0} s'(x)\left(\int_c^x m'(y)\,\mathrm{d}y\right)\mathrm{d}x = +\infty \quad \text{if } y_0 \notin I,$$

$$\int_{x_0}^c s'(x)\left(\int_x^c m'(y)\,\mathrm{d}y\right)\mathrm{d}x = +\infty \quad \text{if } x_0 \notin I.$$

(A4) the generator $\mathcal{L}$, defined on $\mathcal{D} = \{f \in C_0^\infty(I);\ f'|_{\partial I} = 0\}$, is essentially self-adjoint on $L^2(I, \mathrm{d}m)$, or equivalently [11,12]:

$$s \notin L^2\big((x_0, c], \mathrm{d}m\big) \quad \text{if } x_0 \notin I; \quad \text{and} \quad s \notin L^2\big([c, y_0), \mathrm{d}m\big) \quad \text{if } y_0 \notin I.$$

Notice that when $a(x) = 1$ and $I = \mathbb{R}$, the assumptions (A3) and (A4) are automatically satisfied once if $m(I) < +\infty$ (see [17] for (A3), [12] for (A4)).

Throughout this paper we assume that (A1)–(A4) are satisfied. In that case $(X_t)_{t\geq 0}$ is reversible w.r.t. the probability measure $\mu(\mathrm{d}x) = \frac{1}{m(I)} m'(x)\,\mathrm{d}x$. Let $(P_t)_{t\geq 0}$ be the transition semigroup of $(X_t)_{t\geq 0}$, $\mathcal{L}_2$ the generator of (P_t) on $L^2(I, \mu)$ with domain $\mathbb{D}(\mathcal{L}_2)$, which is an extension of $(\mathcal{L}, \mathcal{D})$.

Consider the Poisson equation

$$-\mathcal{L}_2 G = g, \tag{1.3}$$

where $g \in L^2(I, \mu)$ such that $\mu(g) := \int_I g\,\mathrm{d}\mu = 0$. By the ergodicity of the diffusion, the solution G of the Poisson equation, if exists, is unique in $L^2(I, \mu)$ up to the difference of some constant. In the physical interpretation of the heat diffusion, g represents the heat source, G is the equilibrium heat distribution.

The objective of this paper is to estimate

$$\|G\|_{\mathrm{Lip}(\rho)} := \sup_{x,y\in I, x<y} \frac{|G(y) - G(x)|}{|\rho(y) - \rho(x)|} \tag{1.4}$$

in terms of various norms on the heat source g. Here ρ is some absolutely continuous function on I such that $\rho'(x) > 0$, $\mathrm{d}x$-a.e.

Let λ_1 be the spectral gap of $\mathcal{L}_2$, i.e. the lowest eigenvalue or spectral point above zero of $-\mathcal{L}_2$. Then $c_P := \lambda_1^{-1}$ is the best constant in the following Poincaré inequality

$$\mathrm{Var}_\mu(f) \leq c_P \int_I a(x) f'(x)^2\,\mathrm{d}\mu(x), \quad f \in \mathcal{D}, \tag{1.5}$$

where $\mathrm{Var}_\mu(f) := \mu(f^2) - (\mu(f))^2$ is the variance of f w.r.t. μ and $\mu(f) := \int_I f\,\mathrm{d}\mu$. The importance of the spectral gap is that it describes the exponential convergence rate:

$$\|P_t f - \mu(f)\|_2 \leq \mathrm{e}^{-\lambda_1 t}\|f - \mu(f)\|_2 \quad \forall t \geq 0,$$

where $\|\cdot\|_2$ is the $L^2(I,\mu)$-norm. The constant λ_1 can be also interpreted by means of the Poisson equation:

$$\|G-\mu(G)\|_2 \le c_P\|g\|_2 \quad \text{or} \quad \int_I a(x)G'(x)^2\,\mathrm{d}\mu(x) \le c_P\|g\|_2^2.$$

Those physical interpretations explain why the study of λ_1 or c_P is of fundamental importance. Since the study on λ_1 is of a very long history, it is not possible for us to describe even the main line, the reader is referred to the books [9,24] for bibliographies. For the stronger log-Sobolev inequality, the first characterization was due to Bobkov–Götze [3], see [2,9] for further improvements of constant.

Our initial motivation was to understand Chen's variational formula for λ_1 [8]:

$$c_P = \inf_\rho \sup_{x\in I} \frac{s'(x)}{\rho'(x)} \int_x^{y_0} \left[\rho(t)-\mu(\rho)\right] m'(t)\,\mathrm{d}t, \tag{1.6}$$

where ρ runs over all $C^1(I)$ functions with $\rho'>0$, in $L^2(I,\mu)$. Notice that no variational formula is known for the best log-Sobolev constant on the real line. But our main motivation comes from some concentration inequalities for the empirical mean $(1/t)\int_0^t g(X_s)\,\mathrm{d}s$, which are immediate consequences of the estimate on $\|G\|_{\mathrm{Lip}(\rho)}$ via the forward–backward martingale decomposition or transportation–information inequalities developed in [15], see also [18].

Our method for estimate of $\|G\|_{\mathrm{Lip}(\rho)}$ is direct: the solution of the Poisson equation (1.3) can be solved explicitly (unlike the corresponding heat equation), only some further (easy) control is needed for completing the job. Besides those motivations, the estimation of G' is physically meaningful: in the heat diffusion problem, in presence of the heat source g with $\mu(g)=0$, G represents the equilibrium heat distribution; an estimate on $|G'|$ allows us to control the variation of the equilibrium heat distribution.

This paper is organized as follows. In the next section, we state the main results and present several applications in concentration inequalities and transportation–information inequalities, L^1-Poincaré inéquality (a little stronger than the Cheeger isoperimetric inequality), and provide several examples to illustrate the results. In Section 3 the proof of the main result is given.

2. Main results and applications

2.1. *Main results*

Given an absolutely continuous function $\rho: I \longrightarrow \mathbb{R}$ such that $\rho'>0$, $\mathrm{d}x$-a.e., let $d_\rho(x,y)=|\rho(x)-\rho(y)|$ be the metric on I associated with ρ. If the Lipschitzian norm $\|f\|_{\mathrm{Lip}(\rho)}$ of f w.r.t. d_ρ defined in (1.4) is finite, we say that f is ρ-Lipschitzian. Let $L_0^2(I,\mu) := \{f\in L^2(I,\mu);\ \mu(f)=0\}$.

Now, we can state the main result in this paper.

Theorem 2.1. *Assume* (A1)–(A4) *and let ρ, ρ_1, ρ_2 be absolutely continuous functions on I such that $\rho, \rho_k \in L^2(I,\mu)$, $\rho', \rho_k' > 0$, $\mathrm{d}x$-a.e.*

(i) *If*

$$c_{\mathrm{Lip}}(\rho_1,\rho_2) := \operatorname*{ess\,sup}_{x\in I} \frac{s'(x)}{\rho_2'(x)} \int_x^{y_0} \left[\rho_1(t)-\mu(\rho_1)\right] m'(t)\,\mathrm{d}t < +\infty, \tag{2.1}$$

then for any ρ_1-Lipschitzian function $g\in L_0^2(I,\mu)$, there is a unique solution G with $\mu(G)=0$ belonging to the domain $\mathbb{D}(\mathcal{L}_2)$ of the Poisson equation (1.3). *Moreover, G (or one $\mathrm{d}x$-version of it) is ρ_2-Lipschitzian and satisfies*

$$\|G\|_{\mathrm{Lip}(\rho_2)} \le c_{\mathrm{Lip}}(\rho_1,\rho_2)\|g\|_{\mathrm{Lip}(\rho_1)}. \tag{2.2}$$

Furthermore, this inequality (2.2) *becomes equality for $g=\rho_1-\mu(\rho_1)$.*

(ii) *Let $\varphi : I \to \mathbb{R}^+$ be a nonnegative function in $L^2(I,\mu)$. If*

$$c(\varphi,\rho) := \operatorname*{ess\,sup}_{x\in I} \frac{s'(x)}{\rho'(x)} m(I)\left(\mu\big(I_x^+\big)\int_{I_x^-}\varphi\,\mathrm{d}\mu + \mu\big(I_x^-\big)\int_{I_x^+}\varphi\,\mathrm{d}\mu\right) < +\infty, \tag{2.3}$$

where $I_x^+ = \{y \in I;\ y \ge x\}$, $I_x^- = \{y\in I;\ y < x\}$, then for any function $g \in L^2(I,\mu)$ such that $|g| \le \varphi$, there is a unique solution G with $\mu(G) = 0$ to the Poisson equation $-\mathcal{L}_2 G = g - \mu(g)$. Moreover, G (or one $\mathrm{d}x$-version of it) is ρ-Lipschitzian and satisfies

$$\sup_{g:|g|\le\varphi} \|G\|_{\mathrm{Lip}(\rho)} = c(\varphi,\rho). \tag{2.4}$$

Its proof is postponed to Section 3.

Remark 2.2. *Let $C_{\mathrm{Lip}(\rho),0}$ be the Banach space of all ρ-Lipschitzian functions g with $\mu(g) = 0$ equipped with norm $\|\cdot\|_{\mathrm{Lip}(\rho)}$. Part* (i) *above says that the Poisson operator $(-\mathcal{L}_2)^{-1} : C_{\mathrm{Lip}(\rho_1),0} \to C_{\mathrm{Lip}(\rho_2),0}$ is bounded and*

$$\big\|(-\mathcal{L}_2)^{-1}\big\|_{C_{\mathrm{Lip}(\rho_1),0}\to C_{\mathrm{Lip}(\rho_2),0}} = c_{\mathrm{Lip}}(\rho_1,\rho_2). \tag{2.5}$$

Since $\mathcal{L}_2$ is self-adjoint on $L_0^2(I,\mu)$, a general functional analysis result (*see* [25], *Proposition* 2.9) *says that*

$$\big\|(-\mathcal{L}_2)^{-1}\big\|_{L_0^2(I,\mu)} \le \big\|(-\mathcal{L}_2)^{-1}\big\|_{C_{\mathrm{Lip}(\rho),0}\to C_{\mathrm{Lip}(\rho),0}}.$$

But the left-hand side is exactly the Poincaré constant c_P, so we get

$$c_P \le \big\|(-\mathcal{L}_2)^{-1}\big\|_{C_{\mathrm{Lip}(\rho),0}\to C_{\mathrm{Lip}(\rho),0}} = c_{\mathrm{Lip}}(\rho,\rho)$$

which is exactly the '$\le$' part in (1.6). *We now outline the idea of Chen for the converse inequality in* (1.6). *If the eigenfunction ρ associated with $\lambda_1 = 1/c_P$ exists, i.e. $-\mathcal{L}_2\rho = \lambda_1\rho$, it must be strictly monotone* (*see* [9]) *and then could be assumed to be increasing, and ρ' is given by* (3.2) *with $C = 0$ and $g = \lambda_1\rho$* (*see Section* 3 *for the reason why* $C = 0$), *i.e.*

$$\rho'(x) = \lambda_1 s'(x)\int_x^{y_0}\big[\rho(t) - \mu(\rho)\big]m'(t)\,\mathrm{d}t, \quad \mathrm{d}x\text{-}a.s.,$$

where the '$\ge$' part in (1.6) *follows. When λ_1 has no eigenfunction, Chen proved the converse inequality by using a sequence of increasing functions $\rho \in L_0^2(I,\mu)$ approximating this virtual eigenfunction.*

That is our interpretation to Chen's variational formula (1.6).

Remark 2.3. *Let $\|g\|_\varphi$ be the largest constant c such that $|g(x)| \le c\varphi(x)$ over I and $b_\varphi\mathcal{B}$ be the Banach space of those measurable functions g such that its norm $\|g\|_\varphi$ is finite. Let $Pg = g - \mu(g) : L^2(I,\mu) \to L_0^2(I,\mu)$, the orthogonal projection. Part* (ii) *above means that $(-\mathcal{L})^{-1}P$ is bounded from $b_\varphi\mathcal{B}$ to $C_{\mathrm{Lip}(\rho),0}$ and its norm is exactly $c(\varphi,\rho)$.*

2.2. *Applications to transportation–information inequalities and concentration inequalities*

For any probability measure ν on I, say $\nu \in \mathcal{M}_1(I)$, the Wasserstein distance between ν and μ w.r.t. a given metric d on I is defined by

$$W_{1,d}(\nu,\mu) = \inf_\pi \iint_{I^2} d(x,y)\pi(\mathrm{d}x,\mathrm{d}y),$$

where π runs over all couplings of ν,μ, i.e. all probability measures π on I^2 with the first and second marginal distributions ν,μ, respectively. When d is the trivial metric ($d(x,y) = 1_{x\ne y}$), $2W_{1,d}(\mu,\nu) = \|\mu-\nu\|_{\mathrm{TV}} := \sup_{|f|\le 1}|(\mu-\nu)(f)|$, the total variation of $\mu - \nu$.

Under (A1)–(A4), the Dirichlet form $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ associated with the transition semigroup (P_t) of (X_t) is given by

$$\mathbb{D}(\mathcal{E}) = \mathbb{D}\big(\sqrt{-\mathcal{L}_2}\big) = \left\{ f \in L^2(I,\mu) \cap \mathcal{AC}(I), \int_I a(x) f'(x)^2 \,\mathrm{d}\mu(x) < +\infty \right\},$$
$$\mathcal{E}(f,f) := \int_I a(x) f'(x)^2 \,\mathrm{d}\mu(x), \quad f \in \mathbb{D}(\mathcal{E}).$$

For $f, g \in \mathbb{D}(\mathcal{E})$, let $\Gamma(f,g) = af'g'$ be the *carré-du-champs* operator. The Fisher–Donsker–Varadhan information of ν w.r.t. μ is defined by

$$I(\nu|\mu) = \begin{cases} \mathcal{E}\big(\sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}}, \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}}\big), & \text{if } \nu \ll \mu \text{ and } \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}} \in \mathbb{D}(\mathcal{E}), \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.6}$$

Recall that for $\rho_0(x) = \int_c^x \frac{1}{\sqrt{a(y)}} \,\mathrm{d}y$ the associated metric $d_{\rho_0}(x,y) = |\rho_0(y) - \rho_0(x)|$ is the intrinsic metric of the diffusion (X_t).

Corollary 2.4. *Assume* (A1)–(A4). *Let $\rho \in \mathcal{AC}(I) \cap L^2(I,\mu)$ so that $\rho'(x) > 0$, $\mathrm{d}x$-a.e. and*

$$c_\rho = \operatorname*{ess\,sup}_{x\in I} s'(x)\sqrt{a(x)} \int_x^{y_0} \big[\rho(y) - \mu(\rho)\big] m'(y) \,\mathrm{d}y < +\infty. \tag{2.7}$$

Then for all $\nu \in \mathcal{M}_1(I)$

$$\big(W_{1,d_\rho}(\nu,\mu)\big)^2 \le 4c_\rho^2 I(\nu|\mu), \tag{2.8}$$

or equivalently for every ρ-Lipschitzian function g on I, we have for any initial measure $\nu \ll \mu$ and $t, r > 0$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t g(X_s)\,\mathrm{d}s > \mu(g) + r\right) \le \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|_{L^2(I,\mu)} \exp\left(-\frac{tr^2}{4c_\rho^2 \|g\|_{\mathrm{Lip}(\rho)}^2}\right). \tag{2.9}$$

Proof. Remark that $c_\rho = c_{\mathrm{Lip}}(\rho, \rho_0)$, the constant given in (2.1). The equivalence between the transportation–information inequality (2.8) and the Gaussian concentration inequality (2.9) is due to Guillin et al. [15], Theorem 2.4.

By Kantorovitch–Rubinstein dual equality, (2.8) is equivalent to: if $\|g\|_{\mathrm{Lip}(\rho)} \le 1$,

$$\left(\int_I g \,\mathrm{d}(\nu - \mu)\right)^2 \le 4c_\rho^2 I(\nu|\mu) \quad \forall \nu \in \mathcal{M}_1(I).$$

We may assume that $I(\nu|\mu) < +\infty$, i.e., $\nu = h^2\mu$ with $h \in \mathbb{D}(\mathcal{E})$. Let G be the solution of $-\mathcal{L}_2 G = g - \mu(g)$ with $\mu(G) = 0$ (its existence and uniqueness is assured by Theorem 2.1(i)). Notice that with $f = h^2$ $(h \ge 0)$,

$$\begin{aligned} \int_I g \,\mathrm{d}(\nu - \mu) &= \langle -\mathcal{L}_2 G, f\rangle = \mathcal{E}(G,f) = \int_I a(x) G'(x) f'(x) \,\mathrm{d}\mu(x) \\ &\le \operatorname*{ess\,sup}_{x\in I}\big[\sqrt{a(x)}\big|G'(x)\big|\big] \int_I \sqrt{a(x)}\big|f'\big|(x)\,\mathrm{d}\mu(x) \\ &\le 2c_\rho\sqrt{\mu\big(h^2\big)\mu\big[ah'^2\big]} = 2c_\rho\sqrt{I(\nu|\mu)}, \end{aligned} \tag{2.10}$$

where the last inequality follows by Theorem 2.1(i) and Cauchy–Schwarz inequality, for $\operatorname{ess\,sup}_{x\in I}\sqrt{a(x)}|G'(x)| = \|G\|_{\mathrm{Lip}(\rho_0)}$. □

Remark 2.5 (Proposed by the referee). *If the observable g is fixed and absolutely continuous, the best choice of ρ for the Gaussian concentration inequality* (2.9) *is $\tilde\rho$ such that $\tilde\rho' = |g'|$ by Lemma* 3.2 *in Section* 3 (*though such $\tilde\rho$ is not strictly increasing, but Theorem* 2.1 *is still valid as seen for its proof*).

Remark 2.6. *The second inequality in* (2.10) *can be read as*

$$W_{1,\rho}(f\mu, \mu) \le c_\rho \int_I \sqrt{\Gamma(f,f)}\,\mathrm{d}\mu \le 2c_\rho \sqrt{I(f\mu|\mu)}.$$

Repeating the argument above but using part (ii) of Theorem 2.1, we get (2.11) below.

Corollary 2.7. *Assume* (A1)–(A4). *Let* $0 \le \varphi \in L^2(I,\mu)$ *such that* $c(\varphi,\rho_0) < +\infty$. *Then for all* $\nu = f\mu \in \mathcal{M}_1(I)$,

$$\left\|\varphi(\nu-\mu)\right\|_{\mathrm{TV}} \le c(\varphi,\rho_0) \int_I \sqrt{\Gamma(f,f)}\,\mathrm{d}\mu \le 2c(\varphi,\rho_0)\sqrt{I(\nu|\mu)}. \tag{2.11}$$

Or equivalently for every $g: I \to \mathbb{R}$ *such that* $|g(x)-g(y)| \le \beta_\varphi(x,y) := [\varphi(x)+\varphi(y)]1_{x\neq y}$ (*i.e.*, $\|g\|_{\mathrm{Lip}(\beta_\varphi)} \le 1$), *we have for any initial measure* $\nu \ll \mu$ *and* $t, r > 0$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t g(X_s)\,\mathrm{d}s > \mu(g) + r\right) \le \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(I,\mu)} \exp\left(-\frac{tr^2}{4c(\varphi,\rho_0)^2}\right). \tag{2.12}$$

The Gaussian concentration inequality (2.12) follows from (2.11) by [15], Theorem 2.4, and the fact that $\|\varphi(\nu-\mu)\|_{\mathrm{TV}} = \sup_{g:\|g\|_{\mathrm{Lip}(\beta_\varphi)}\le 1} \int_I g\,\mathrm{d}(\nu-\mu)$ (cf. [14]). Notice that β_φ is a metric once φ is positive (and a pseudo-metric satisfying the triangular inequality in the general case).

Remark 2.8. *When* $\varphi = 1$ *in* (2.11), *the constant* $c(\varphi,\rho_0)$ *becomes*

$$c_\delta := 2 \operatorname*{ess\,sup}_{x\in I} \sqrt{a(x)}s'(x)m(I)\mu\left(I_x^+\right)\mu\left(I_x^-\right), \tag{2.13}$$

where I_x^+, I_x^- *are given in Theorem* 2.1; *and the inequality* (2.11) *becomes*: *for every* μ*-probability density* $f \in \mathcal{AC}(I)$

$$\int_I |f-1|\,\mathrm{d}\mu \le c_\delta \int_I \sqrt{\Gamma(f,f)}\,\mathrm{d}\mu \le 2c_\delta\sqrt{I(f\mu|\mu)}. \tag{2.14}$$

It was proved by Guillin et al. [15], *Theorem* 3.1, *that if the Poincaré inequality holds*, *then*

$$\int_I |f-1|\,\mathrm{d}\mu \le \sqrt{2c_G I(\nu|\mu)}$$

with the best constant $c_G \le 2c_P$ (*the index G is referred to the equivalent Gaussian concentration inequality*); *and conversely if the last inequality holds*, *then* $c_P \le 2c_G$.

Remark 2.9. *Gozlan* [13] *has established some connections between Talagrand's transportation–entropy inequalities and weighted Poincaré inequalities*, *see also* [23].

The concentration inequalities (2.9) and (2.12) do not contain the asymptotic variance of g:

$$\sigma^2(g) := \lim_{t\to\infty} \frac{1}{t} \mathrm{Var}_{\mathbb{P}_\mu}\left(\int_0^t g(X_s)\,\mathrm{d}s\right)$$

which plays a fundamental role in the central limit theorem (then in statistical applications). This is provided in the following Bernstein's type concentration inequality.

Corollary 2.10. *Assume* (A1)–(A4). *Suppose that the constant* c_δ *in* (2.13) *is finite.*

(i) *If the constant* c_ρ *in* (2.7) *is finite, then for every* ρ*-Lipschitzian function* g *with* $\|g\|_{\mathrm{Lip}(\rho)} \le 1$, *we have for any initial measure* $\nu \ll \mu$ *and* $t, x > 0$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t g(X_s)\,\mathrm{d}s > \mu(g) + \sqrt{\left(2\sigma^2(g) + 4c_\rho^2 \min\{1, c_\delta\sqrt{x}\}\right)x}\right) \le \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(I,\mu)} \mathrm{e}^{-tx}.$$

(ii) *If the constant* $c(\varphi, \rho_0)$ *in* (2.3) *is finite, then for every measurable function* g *such that* $|g(x) - g(y)| \le \varphi(x) + \varphi(y)$, *the inequality in* (i) *holds with* c_ρ *replaced by* $c(\varphi, \rho_0)$.

Proof. Our proof below follows [16].

(i) We may and will assume that $\|g\|_{\mathrm{Lip}(\rho)} \le 1$. Let G be the solution of $-\mathcal{L}G = g - \mu(g)$. Notice that $\sigma^2(g) = 2\langle G, g\rangle_\mu = 2\mathcal{E}(G, G)$.

We have for $\nu = h^2\mu$ with $I(\nu|\mu) < +\infty$,

$$\int_I g\,\mathrm{d}(\nu - \mu) = 2\int_I aG'hh'\,\mathrm{d}\mu(x) \le 2\sqrt{\int_I aG'^2h^2\,\mathrm{d}\mu \cdot I(\nu|\mu)}.$$

Since $0 \le aG'^2 \le \|G\|^2_{\mathrm{Lip}(\rho_0)} \le c_\rho^2$ by Theorem 2.1(i), using the fact that $\int_I F\,\mathrm{d}(\nu - \mu) \le \frac{1}{2}\|\nu - \mu\|_{\mathrm{TV}}$ for F verifying $|F(x) - F(y)| \le 1$, we have by (2.14),

$$\int_I aG'^2h^2\,\mathrm{d}\mu \le \int_I aG'^2\,\mathrm{d}\mu + \frac{c_\rho^2}{2}\int_I |h^2 - 1|\,\mathrm{d}\mu \le \frac{\sigma^2(g)}{2} + c_\rho^2 \min\{1, c_\delta\sqrt{I(\nu|\mu)}\}.$$

Plugging it into the previous inequality (for $\pm g$), we obtain

$$\left(\int_I g\,\mathrm{d}(\nu - \mu)\right)^2 \le \left(2\sigma^2(g) + 4c_\rho^2 \min\{1, c_\delta\sqrt{I(\nu|\mu)}\}\right)I(\nu|\mu) \quad \forall\nu.$$

This is equivalent to the desired concentration inequality by [15], Theorem 2.4.

(ii) The same argument as above (but using part (ii) of Theorem 2.1 instead of part (i)), we have $\forall g$ such that $|g| \le \varphi$

$$\left(\int_I g\,\mathrm{d}(\nu - \mu)\right)^2 \le \left(2\sigma^2(g) + 4c(\varphi, \rho_0)^2 \min\{1, c_\delta\sqrt{I(\nu|\mu)}\}\right)I(\nu|\mu) \quad \forall\nu.$$

This leads to the desired concentration inequality again by [15], Theorem 2.4. □

2.3. L^1-*Poincaré inequality and Cheeger's isoperimetric inequality*

The Poincaré inequality has a L^1 counterpart related to Cheeger's isoperimetric inequality. Namely, let $c_{P,1}$ be the best constant such that the following L^1-Poincaré inequality holds: for any $f \in \mathcal{AC}(I) \cap L^1(I, \mu)$

$$\int_I |f - \mu(f)|\,\mathrm{d}\mu \le c_{P,1}\int_I \sqrt{a(x)}\,|f'|\,\mathrm{d}\mu, \tag{2.15}$$

where $\mathcal{AC}(I)$ is the space of all absolutely continuous functions on I. Theorem 2.1 allows us to identify the best constant $c_{P,1}$ in the L^1-Poincaré inequality (2.15).

Theorem 2.11. *Assume* (A1)–(A4). *The best constant* $c_{P,1}$ *in the* L^1*-Poincaré inequality* (2.15) *is finite if and only if* c_δ *given in* (2.13) *is finite. In this case* $c_{P,1} = c_\delta$.

Proof. At first $c_{P,1} \le c_\delta$, by (2.14) (the passage from μ-density f to general f in (2.15) is easy). For the converse inequality, we may assume that $c_{P,1} < +\infty$. In that case for any $g \in b\mathcal{B}$ such that $|g| \le 1$, $G = (-\mathcal{L}_2)^{-1}(g - \mu(g))$ exists (because the Poincaré inequality holds by Cheeger's inequality). We have for any μ-probability density $f \in \mathcal{AC}(I)$,

$$\int_I a(x)G'(x)f'(x)\,\mathrm{d}\mu(x) = \langle -\mathcal{L}_2 G, f\rangle_\mu = \langle g, f-1\rangle_\mu$$
$$\le \int_I |f-1|\,\mathrm{d}\mu \le c_{P,1}\int_I \sqrt{a(x)}\bigl|f'(x)\bigr|\,\mathrm{d}\mu(x).$$

That implies $\|G\|_{\mathrm{Lip}(\rho_0)} = \operatorname{ess\,sup}_{x\in I}\sqrt{a(x)}|G'(x)| \le c_{P,1}$ (however this elementary fact do no longer work in the multi-dimensional case). Hence $c_\delta \le c_{P,1}$ by Theorem 2.1(ii). □

Let us discuss now some connections between (2.15) and isoperimetric inequalities. Consider the intrinsic metric d_{ρ_0} associated with the diffusion where $\rho_0(x) = \int_c^x \frac{1}{\sqrt{a(y)}}\,\mathrm{d}y$, and the corresponding isoperimetric function

$$I_\mu(p) := \inf\bigl\{\mu_\partial(\partial A); \mu(A) = p\bigr\}, \quad p \in (0,1).$$

Here ∂A is the boundary of A and the surface measure μ_s of A is defined by $\mu_s(\partial A) = \liminf_{\varepsilon\to 0_+}\frac{\mu(A_\varepsilon)-\mu(A)}{\varepsilon}$ and $A_\varepsilon = \{x \in I,\ \text{such that } d_{\rho_0}(x, A) \le \varepsilon\}$, the ε-neighborhood of A.

Remark 2.12. *Let $c_{cheeger}$ be Cheeger's isoperimetric constant of μ w.r.t. the intrinsic metric d_{ρ_0}, i.e. the best constant in the following Cheeger isoperimetric inequality*

$$\min\bigl(\mu(A), 1-\mu(A)\bigr) \le c_{cheeger}\mu_s(\partial A)$$

for all measurable subsets $A \subset I$, or equivalently $I_\mu(p) \ge \frac{1}{c_{cheeger}}\min\{p, 1-p\}$. It is well known (cf. [4,20]) that $c_{cheeger}$ is also the best constant in the functional version of Cheeger's isoperimetric inequality below: for any $f \in \mathcal{AC}(I) \cap L^1(I,\mu)$

$$\int_I \bigl|f - m_\mu(f)\bigr|\,\mathrm{d}\mu \le c_{cheeger}\int_I \sqrt{a(x)}\bigl|f'\bigr|\,\mathrm{d}\mu, \tag{2.16}$$

where $m_\mu(f)$ is a median of f w.r.t. μ (via Co-Area formula). Since

$$\frac{1}{2}\mu\bigl(|f-\mu(f)|\bigr) \le \mu\bigl(|f - m_\mu(f)|\bigr) \le \mu\bigl(|f-\mu(f)|\bigr)$$

we have

$$\frac{1}{2}c_{P,1} \le c_{cheeger} \le c_{P,1}. \tag{2.17}$$

The two inequalities above are both sharp as seen for the examples later. An important result of Bobkov–Houdré [5], Theorem 1.3, says that

$$c_{cheeger} = \operatorname*{ess\,sup}_{x\in I}\frac{m(I)\min\{\mu(I_x^+), \mu(I_x^-)\}}{m'(x)\sqrt{a(x)}}$$
$$= \operatorname*{ess\,sup}_{x\in I} m(I)\sqrt{a(x)}s'(x)\min\bigl\{\mu\bigl(I_x^+\bigr), \mu\bigl(I_x^-\bigr)\bigr\} \tag{2.18}$$

or say roughly, the extreme set for $c_{cheeger}$ is a semi-interval I_x^+. In recent years, the best constant $c_{cheeger}$ (in multi-dimensional case) has been extensively investigated, see [1,5,7,20,22,26] and relevant references therein.

Remark 2.13 (Proposed by the referee). *By Bobkov–Houdré* [4], *Theorem* 1.2, *the* L^1*-Poincaré inequality* (2.15) *is equivalent to the following isoperimetric inequality associated with* d_{ρ_0}:

$$2\mu(A)\mu\bigl(A^c\bigr) \le c_{P,1}\mu_s(\partial A) \tag{2.19}$$

for any measurable subset A of I, *or equivalently* $I_\mu(p) \ge \frac{2}{c_{P,1}} p(1-p)$, $p \in (0,1)$. *That equivalence holds on a general metric space.*

Notice that if $a(x)$ *is continuous and positive,* c_δ *is just the best constant in* (2.19) (*in place of* $c_{P,1}$) *for A varying over* I_x^+, $x \in I$.

When $a(x) = 1$ *and* μ *is log-concave* (*i.e.,* $\mu = f\,\mathrm{d}x$ *with* $\log f$ *concave*), *Bobkov–Houdré* [4], *Corollary* 13.8, *showed that the optimal set for* $I_\mu(p)$ *is* I_x^+ *with* $\mu(I_x^+) = p$ *for every* $p \in (0,1)$, *and then* $c_{P,1} = c_\delta$.

The referee indicates another approach for Theorem 2.11 *even for general* μ *not necessarily log-concave, when* $a(x) = 1$. *The idea goes as follows. At first notice that* $p(1-p)$ *is the isoperimetric function* $I_\nu(p)$ *of the logistic distribution* ν: $\nu(-\infty, x] = (1 + \mathrm{e}^{-x})^{-1}$. *Following the proof of Bobkov–Houdré* [5], *proof of Theorem* 1.3, *if* $c_\delta < +\infty$, *the increasing mapping* $U : \mathbb{R} \to I$ *pushing forward* ν *to* μ *must be Lipschitzian and* $\|U\|_{\mathrm{Lip}} = 2c_\delta$. *Then one sees that the best constant* $c_{P,1}$ *in* (2.19) *for* μ *is just* $\|U\|_{\mathrm{Lip}}/2 = c_\delta$.

Let us remark finally that the L^1*-Poincaré inequality* (2.15) *is equivalent to the following concentration inequality* ([4], *Theorem* 2.1, *pp.* 20–21):

$$\mu(A_\varepsilon) \ge \frac{p}{p + (1-p)\exp(-2\varepsilon/c_{P,1})}, \quad \mu(A) = p \in (0,1), \varepsilon > 0.$$

2.4. *A qualitative description for the boundedness of the Poisson operator*

For $g \in L_0^2(I, \mu)$, the solution G with $\mu(G) = 0$ of the Poisson equation $-\mathcal{L}_2 G = g$, if exists, will be denoted by $(-\mathcal{L})^{-1}g$. One may think naturally that when φ is bounded but tends to zero at the boundary ∂I, the Lipschitzian norm $c(\varphi, \rho_0)$ may be finite even if $c_\delta = +\infty$. The same picture might appear in one's mind for $c_{\mathrm{Lip}}(\rho, \rho_0)$ when ρ' tends to 0 at the boundary ∂I. However this is not the case.

Proposition 2.14. *Assume* (A1)–(A4). *Let* ρ, φ *be as in Theorem* 2.1, *but moreover bounded and* $\varphi > 0$. *Let* $\rho_0(x) = \int_c^x \frac{1}{\sqrt{a(y)}}\,\mathrm{d}y$. *Consider the following properties*:

(i) $c_\rho = c_{\mathrm{Lip}}(\rho, \rho_0) = \|(-\mathcal{L}_2)^{-1}\|_{C_{\mathrm{Lip}(\rho),0} \to C_{\mathrm{Lip}(\rho_0),0}} < +\infty$.

(ii) $c(\varphi, \rho_0) = \sup_{g:|g| \le \varphi} \|(-\mathcal{L}_2)^{-1}(g - \mu(g))\|_{\mathrm{Lip}(\rho_0)} < +\infty$.

(iii) $c_\delta = \sup_{|g| \le 1} \|(-\mathcal{L}_2)^{-1}(g - \mu(g))\|_{\mathrm{Lip}(\rho_0)} < +\infty$.

(iv) *The* L^1*-Poincaré inequality* (2.15) *holds, i.e.,* $c_{P,1} < +\infty$.

(v) *The transportation–information inequality below holds*: *there is some finite best constant* $c_G > 0$ *such that for all* $\nu = f\mu \in \mathcal{M}_1(I)$,

$$\int_I |f - 1|\,\mathrm{d}\mu \le \sqrt{2c_G I(\nu|\mu)}.$$

(vi) *The Poincaré inequality* (1.5) *holds, i.e.,* $c_P < +\infty$.

Then

(a) *the properties* (i)–(iv) *are equivalent.*

(b) (iv) $\Rightarrow$ (v) $\Leftrightarrow$ (vi).

(c) *If* $a(x) = 1$ *and* $b' \le K$ (*i.e., the Bakry–Emery curvature is bounded from below by* $-K$), (vi) $\Rightarrow$ (iv) *and then* (i)–(vi) *are all equivalent.*

Proof. (a) Equivalence between (i), (ii) and (iii). It is enough to regard the behavior at the boundary of the functions appearing in the definitions of $c_\rho = c_{\mathrm{Lip}}(\rho, \rho_0)$, c_δ and $c(\varphi, \rho_0)$. For instance, if $y_0 \notin I$, for x close to y_0, say $x \ge z > c$,

we have

$$\big(\rho(z)-\mu(\rho)\big)\mu\big(I_x^+\big)\le\int_x^{y_0}\big(\rho(y)-\mu(\rho)\big)\,\mathrm{d}\mu(y)\le\big(\rho(y_0)-\mu(\rho)\big)\mu\big(I_x^+\big),$$
$$\mu\big(I_z^-\big)\mu\big(I_x^+\big)\le\mu\big(I_x^+\big)\mu\big(I_x^-\big)\le\mu\big(I_x^+\big),$$
$$\mu(1_{I_z^-}\varphi)\mu\big(I_x^+\big)\le\mu\big(I_x^+\big)\mu(1_{I_x^-}\varphi)+\mu\big(I_x^-\big)\mu(1_{I_x^+}\varphi)\le 2\|\varphi\|_\infty\mu\big(I_x^+\big).$$

Hence the supremums over $[c,y_0)$ of the functions appearing in the definitions of c_ρ, c_δ and $c(\varphi,\rho_0)$ are simultaneously finite or infinite. The same argument works when $x_0\notin I$. That completes the proof of the equivalence between (i), (ii) and (iii).

(iii) ⇔ (iv). That is contained in Theorem 2.11: $c_\delta=c_{P,1}$.

(b) (iv) ⇒ (v). Since $\int_I\sqrt{a(x)}|f'|\,\mathrm{d}\mu\le 2\sqrt{I(f\mu|\mu)}$, we have $c_G\le 2c_{P,1}^2$.

(v) ⇔ (vi). This is noticed in Remark 2.8: $c_P/2\le c_G\le 2c_P$.

(c) (vi) ⇒ (iv). This converse of the Cheeger's inequality is known in the actual lower bounded Bakry–Emery's curvature case see Buser [7] and Ledoux [21], Theorem 5.2 (otherwise there are counter-examples). □

2.5. *Several examples*

Example 2.15 (Gaussian measure). *Let $I=\mathbb{R}$, $a(x)=1$ and $b(x)=-x/\sigma^2$ where $\sigma>0$. Then $m'(x)=\mathrm{e}^{-x^2/2\sigma^2}$ and $\mu=\mathcal{N}(0,\sigma^2)$, the centered Gaussian law with variance σ^2. For $\rho_0(x)=x$, we see that*

$$c_{\mathrm{Lip}}(\rho_0,\rho_0)=c_{\rho_0}=\sup_{x\in\mathbb{R}}\mathrm{e}^{x^2/2\sigma^2}\int_x^\infty y\mathrm{e}^{-y^2/2\sigma^2}\,\mathrm{d}y=\sigma^2.$$

By Remark 2.2, $c_P\le c_{\rho_0}=\sigma^2$ which is in reality an equality as well known [20]. The transportation inequality (2.8) becomes equality for $\nu=\mathcal{N}(m,\sigma^2)$.

By calculus we identify the constant c_δ in (2.13) as

$$c_\delta=2\sup_{x\in\mathbb{R}}\mathrm{e}^{x^2/2\sigma^2}\sqrt{2\pi}\sigma\mu\big([x,+\infty)\big)\mu\big((-\infty,x)\big)=\sqrt{\frac{\pi}{2}}\sigma.$$

On the other side, $c_{cheeger}\ge\sqrt{\frac{\pi}{2}}\sigma$ as seen for $A=\mathbb{R}^+$. Then by (2.17) and Theorem 2.11, $c_{cheeger}=c_\delta=c_{P,1}$.

Example 2.16 (Uniform distribution). *Let $I=[-D/2,D/2]$ where $D>0$, $a(x)=1$ and $b(x)=0$. The unique invariant probability measure μ is the uniform measure on I. Since $m'(x)=1=s'(x)$, we have*

$$c_{\rho_0}=c_{\mathrm{Lip}(\rho_0,\rho_0)}=\sup_{x\in[-D/2,D/2]}\int_x^{D/2}y\,\mathrm{d}y=\frac{D^2}{8}$$

and the constant $c_\delta=c(\varphi,\rho_0)$ with $\varphi=1$ is given by

$$c_\delta=\sup_{x\in[-D/2,D/2]}2D\mu\big([-D/2,x]\big)\mu\big([x,D/2]\big)=\frac{D}{2}.$$

As $c_{cheeger}\ge D/2$ (as seen for $A=[0,D/2]$), we have $c_{cheeger}=D/2=c_\delta=c_{P,1}$ by (2.17) and Theorem 2.11.

Example 2.17 (Exponential measure on $\mathbb{R}^+$). *Let $I=\mathbb{R}^+=[0,+\infty)$, $a(x)=1$ and $b(x)=-\lambda$ where $\lambda>0$. Then $m'(x)=\mathrm{e}^{-\lambda x}=1/s'(x)$, $\rho_0(x)=x$ and μ is the exponential distribution with parameter λ. It is easy to see that $c_{\rho_0}=c_{\mathrm{Lip}(\rho_0,\rho_0)}=+\infty$: no spectral gap in the ρ_0-Lipschitzian norm. In fact the transportation–information inequality (2.8) is false for $\rho=\rho_0$. By Theorem 2.11*

$$c_{P,1}=c_\delta=2\sup_{x\ge 0}\frac{1}{\lambda}\mathrm{e}^{\lambda x}\mu(0,x)\mu(x,+\infty)=2\sup_{x\ge 0}\frac{1}{\lambda}\mu(0,x)=\frac{2}{\lambda}.$$

However $c_{cheeger} = \frac{1}{\lambda}$ *by Bobkov–Houdré* [5], *which together with the Gaussian measure above shows that the two inequalities in* (2.17) *are both sharp* (*as promised*). *We have also the transportation–information inequality* (2.14), *which is read as*

$$\|\nu - \mu\|_{\mathrm{TV}} \le \frac{4}{\lambda}\sqrt{I(\nu|\mu)} \quad \forall \nu.$$

It is sharp. Indeed let ν *be the exponential law with parameter* $\tilde{\lambda} \in (0, \lambda)$. *We have* $I(\nu|\mu) = (\lambda - \tilde{\lambda})^2/4$, *and the right-hand side above is given by* $2(1-x)$ *where* $x = \tilde{\lambda}/\lambda$. *The left-hand side above is given by* $2(x^{x/(1-x)} - x^{1/(1-x)})$. *Then the inequality above for such* ν *says*

$$x^{x/(1-x)} - x^{1/(1-x)} \le 1 - x, \quad 0 < x < 1,$$

which is sharp as $x \to 0$.

For this model it is well known that $c_P = 4/\lambda^2$ [20]. *The inequality above is same as provided by* [15], *Theorem* 3.1 (*from the Poincaré inequality*).

Example 2.18 (Log-concave measure on $\mathbb{R}$***).*** *Let* $I = \mathbb{R}$, $a(x) = 1$ *and* $b(x) = -V'(x)$ *where* V *is* C^2, *strictly convex on* $\mathbb{R}$ *such that* $V(0) = 0$ *and* $\int_{\mathbb{R}} \mathrm{e}^{-V}\,\mathrm{d}x < +\infty$. *Then* $m'(x) = \mathrm{e}^{-V(x)}$ *and* $s'(x) = \mathrm{e}^{V(x)}$ *and* $\rho_0(x) = x$. *Let* $\rho(x) = V'(x)$, *which is* μ*-integrable and* $\mu(\rho) = 0$. *We have*

$$c_\rho = c_{\mathrm{Lip}}(\rho, \rho_0) = \sup_{x\in\mathbb{R}} \mathrm{e}^{V(x)} \int_x^{+\infty} V'(y)\mathrm{e}^{-V(y)}\,\mathrm{d}y = \sup_{x\ge 0} \mathrm{e}^{V(x)}\mathrm{e}^{-V(x)} = 1.$$

Thus assuming $\int V'^2 \mathrm{e}^{-V}\,\mathrm{d}x < +\infty$, *we have the transportation–information inequality* (2.8) *and the Gaussian concentration inequality* (2.9). *For instance, for any* $g \in C^1(\mathbb{R})$ *such that* $|g'| \le V''$ *we have for any initial measure* $\nu \ll \mu$ *and* $t, r > 0$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t g(X_s)\,\mathrm{d}s > \mu(g) + r\right) \le \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(I,\mu)} \exp\left(-\frac{tr^2}{4}\right). \tag{2.20}$$

Furthermore, for any nonnegative $\varphi \le M(1 + |V'|)$, *it is easy to see that* $c(\varphi, \rho_0) < +\infty$, *then the transportation–information inequality* (2.11) *holds.*

In comparison recall the Lyapunov function criterion in [15], *Theorem* 5.1, *for* (2.11): *for some* $0 \le U \in C^2$, $-U'' + V'U' + |U'|^2 \ge c\varphi^2 - K$ *for some two positive constants* c, K (*which does not require the convexity of* V).

It will be very interesting to generalize it to log-concave measures on multi-dimensional spaces $\mathbb{R}^d$. *See Bobkov–Ledoux* [6] *for some results in this direction.*

Example 2.19 (Jacobi diffusion). *Let* $I =]0, 1[$, $a(x) = x(1-x)$ *and* $b(x) = -x + 1/2$, *then* $\mu(x) = 1/(\pi\sqrt{x(1-x)})$ [10]. *For* $\rho_0(x) = \frac{\pi}{2} + \mathrm{Arcsin}(2x-1)$, *we see that*

$$c_{\mathrm{Lip}}(\rho_0, \rho_0) = c_{\rho_0} = \sup_{x\in]0,1[}\left(\frac{\pi^2}{8} - \frac{1}{2}\mathrm{Arcsin}^2(2x-1)\right) = \frac{\pi^2}{8}.$$

By calculus we identify the constant c_δ *in* (2.13) *as*

$$c_\delta = \frac{2}{\pi}\sup_{x\in]0,1[}\left(\frac{\pi^2}{4} - \mathrm{Arcsin}^2(2x-1)\right) = \frac{\pi}{2}.$$

Using (2.18), *see Bobkov–Houdré* [5], *we obtain* $c_{cheeger} = \frac{\pi}{2}$, *so we have* $c_{cheeger} = c_{P,1} = c_\delta = \frac{\pi}{2}$ *by Theorem* 2.11.

Example 2.20 (Continuous branching process). *Let* $I =]0, +\infty[$, $a(x) = 2x$ *and* $b(x) = -2x + 1$, *then* $\mu(x) = \frac{1}{\sqrt{\pi}} \frac{e^{-x}}{\sqrt{x}}$. *This process arise as diffusion limits of discrete space branching process, see* [19]. *For* $\rho_0(x) = \sqrt{2x}$, *we see that*

$$c_{\mathrm{Lip}}(\rho_0, \rho_0) = c_{\rho_0} = \sup_{x \in \mathbb{R}^+} \left(1 - \frac{e^x}{\sqrt{\pi}} \int_x^\infty \frac{e^{-y}}{\sqrt{y}} \, dy\right) = 1.$$

Example 2.21. *See Example* 1.4.2 *in* [24]. *Let* $I = \mathbb{R}^+$, $a(x) = (1+x)^\alpha$ *with* $\alpha > 1$, *and* $b(x) = 0$, *then* $\mu(x) = \frac{\alpha-1}{(1+x)^\alpha}$. *For* $\alpha > 2$ *and* $\rho_0(x) = \frac{2}{\alpha-2}(1 - (1+x)^{-(\alpha-2)/2})$, *we see that*

$$c_{\mathrm{Lip}}(\rho_0, \rho_0) = c_{\rho_0} = \frac{4}{(\alpha-2)(3\alpha-4)} \sup_{x \in \mathbb{R}^+} (1+x)^{-(\alpha-2)/2}\bigl(1 - (1+x)^{-(\alpha-2)/2}\bigr) = \frac{1}{(\alpha-2)(3\alpha-4)}.$$

By calculus we identify the constant c_δ *in* (2.13) *as*

$$c_\delta = \frac{2}{\alpha-1} \sup_{x \in \mathbb{R}^+} (1+x)^{-(\alpha-2)/2}\bigl(1 - (1+x)^{-(\alpha-1)}\bigr) = \frac{4}{3\alpha-4}\left(\frac{\alpha-2}{3\alpha-4}\right)^{(\alpha-2)/(2(\alpha-1))}.$$

By Theorem 2.11, *we have* $c_{P,1} = c_\delta$. *However, using* (2.18), *we obtain* $c_{cheeger} = \frac{1}{\alpha-1}(\frac{1}{2})^{(\alpha-2)/(2(\alpha-1))}$.

3. Proof of Theorem 2.1

3.1. *Several lemmas*

Let $\mathcal{L}^*$ be the adjoint operator of $(\mathcal{L}, \mathcal{D})$ in $L^2(I, m)$, more precisely a function f in $L^2(I, m)$ belongs to the domain of definition $\mathbb{D}_2(\mathcal{L}^*)$ of $\mathcal{L}^*$ if there is $g \in L^2(I, m)$ such that $\langle f, \mathcal{L}h\rangle_m = \langle g, h\rangle_m$ for all $h \in \mathcal{D}$, in such case $\mathcal{L}^* f = g$. Here $\langle \cdot, \cdot\rangle_m$ is the inner product on $L^2(I, m)$.

We want to understand the Poisson equation (1.3) as an ordinary differential equation. That is the purpose of the following lemma.

Lemma 3.1. *Assume* (A1) *and* (A2). *For a given* $f \in L^2(I, m)$, $f \in \mathbb{D}_2(\mathcal{L}^*)$ *if and only if*

(i) f *admits a* dx*-version* $\tilde f$ *such that* $\tilde f \in C^1(I)$, $\tilde f'|_{\partial I} = 0$, *and* $\tilde f' \in \mathcal{AC}(I)$;
(ii) $a\tilde f'' + b\tilde f' \in L^2(I, m)$.

In that case $\mathcal{L}^* f = a\tilde f'' + b\tilde f'$.

Proof. This follows by integration by parts argument and the distribution theory, as in [12], Appendix C, Theorem 2.7, or [27], Lemma 4.5. So we omit the details. □

Since $\mathcal{L}_2$ is an extension of $(\mathcal{L}, \mathcal{D})$, then $\mathcal{L}^*$ is an extension of $\mathcal{L}_2^* = \mathcal{L}_2$ (because the generator of a symmetric strongly continuous semigroup is always self-adjoint). Of course under (A4), $\mathcal{L}^* = \mathcal{L}_2$. Then in our framework (i.e., (A1)–(A4) are satisfied), solving the Poisson equation (1.3) is equivalent to check $G \in C^1(I) \cap L_0^2(I, \mu)$ such that $G' \in \mathcal{AC}(I)$ and $G'|_{\partial I} = 0$ and

$$-\bigl(aG'' + bG'\bigr) = g. \tag{3.1}$$

This is a first-order differential equation for G'. It can be easily solved as

$$G'(x) = s'(x)\left[C + \int_x^{y_0} g(y)m'(y)\, dy\right] \tag{3.2}$$

for some constant C (to be determined).

Lemma 3.2. *Let ρ be as in Theorem 2.1 and $g : I \to \mathbb{R}$ be ρ-Lipschitzian with $\mu(g) = 0$. Then for all $x \in I$,*

$$\int_x^{y_0} g(t)m'(t)\,\mathrm{d}t \le \|g\|_{\mathrm{Lip}(\rho)} \int_x^{y_0} \big[\rho(t) - \mu(\rho)\big]m'(t)\,\mathrm{d}t.$$

Proof. Without loss of generality, we may suppose that $\|g\|_{\mathrm{Lip}(\rho)} = 1$ and $m(I) = 1$ and $\mu(\rho) = 0$. Letting $m(x) := \int_{x_0}^x m'(t)\,\mathrm{d}t$ and $\tilde{g} = g \circ m^{-1}$, $\tilde{\rho} = \rho \circ m^{-1}$, we have

$$\int_x^{y_0} g(t)m'(t)\,\mathrm{d}t = \int_{m(x)}^1 \tilde{g}(u)\,\mathrm{d}u \quad \text{and} \quad \int_x^{y_0} \rho(t)m'(t)\,\mathrm{d}t = \int_{m(x)}^1 \tilde{\rho}(u)\,\mathrm{d}u.$$

As $\|\tilde{g}\|_{\mathrm{Lip}(\tilde{\rho})} = \|g\|_{\mathrm{Lip}(\rho)}$, we have only to prove that for all $x \in [0, 1]$

$$h(x) = \int_x^1 \tilde{\rho}(s)\,\mathrm{d}s - \int_x^1 \tilde{g}(s)\,\mathrm{d}s \ge 0.$$

Since $h(0) = h(1) = 0$ and h' is absolutely continuous on [0, 1] and for dx-a.s. $x \in [0, 1]$

$$h''(x) = -\tilde{\rho}'(x) + \tilde{g}'(x) \le \|\tilde{g}\|_{\mathrm{Lip}(\tilde{\rho})}\tilde{\rho}'(x) - \tilde{\rho}'(x) \le 0.$$

So h is concave on [0, 1]. Consequently $h(x) \ge 0$ for all $x \in [0, 1]$. □

Lemma 3.3. *Let $0 \le \varphi \in L^2(I, \mu)$. Then for every $x \in I$,*

$$\sup_{g:\,|g|\le\varphi} \int_x^{y_0} \big[g(t) - \mu(g)\big]m'(t)\,\mathrm{d}t = m(I)\bigg(\mu\big(I_x^+\big)\int_{I_x^-} \varphi\,\mathrm{d}\mu + \mu\big(I_x^-\big)\int_{I_x^+} \varphi\,\mathrm{d}\mu\bigg),$$

where $I_x^+ = [x, y_0] \cap I$, $I_x^- = [x_0, x) \cap I$. The supremum is attained for $g = 1_{I_x^+}\varphi - 1_{I_x^-}\varphi$.

Proof. We may assume that $m(I) = 1$ and then $\mu = m$. Fix $x \in I$. The functional $\Phi(g) = \int_x^{y_0} [g(t) - \mu(g)]m'(t)\,\mathrm{d}t = \mathrm{Cov}_\mu(g, 1_{I_x^+})$ (the covariance of g and $1_{I_x^+}$ under μ) is a linear functional of g. Since the closed convex hull of $\{1_A\varphi, A \in \mathcal{B}\}$ ($\mathcal{B}$ is the Borel σ-field on I, and the closure is to be understood in $L^2(I, \mu)$) is $\{g;\ 0 \le g \le \varphi\}$, and $\{g;\ |g| \le \varphi\} = \{h_1 - h_2; 0 \le h_1, h_2 \le \varphi\}$, then

$$\sup_{g:\,|g|\le\varphi} \Phi(g) = \sup_{h_1:\,0\le h_1\le\varphi} \Phi(h_1) - \inf_{0\le h_2\le\varphi} \Phi(h_2) = \sup_{A\in\mathcal{B}} \mathrm{Cov}_\mu(1_A\varphi, 1_{I_x^+}) + \sup_{A\in\mathcal{B}} \mathrm{Cov}_\mu(-1_A\varphi, 1_{I_x^+}).$$

We examine the first supremum at the right-hand side. Note that

$$\mathrm{Cov}_\mu(1_A\varphi, 1_{I_x^+}) = \int_{A\cap I_x^+} \varphi\,\mathrm{d}\mu - \mu\big(I_x^+\big)\int_A \varphi\,\mathrm{d}\mu.$$

With $A \cap I_x^+ = B$ fixed, this functional of A attaint the maximum when A becomes the smallest B. Next for $A = B$ or equivalently $A \subset I_x^+$, the right-hand side above equals to

$$\int_A \varphi\,\mathrm{d}\mu\big(1 - \mu\big(I_x^+\big)\big)$$

which attaint the maximum if $A = I_x^+$. So we have proven that

$$\max_{A\in\mathcal{B}} \mathrm{Cov}_\mu(1_A\varphi, 1_{I_x^+}) = \int_{I_x^+} \varphi\,\mathrm{d}\mu\big(1 - \mu\big(I_x^+\big)\big) = \mu\big(I_x^-\big)\int_{I_x^+} \varphi\,\mathrm{d}\mu.$$

Now we turn to the last supremum. Note $\mathrm{Cov}_\mu(-1_A\varphi, 1_{I_x^+}) = -\int_{A\cap I_x^+}\varphi\,\mathrm{d}\mu + \mu(I_x^+)\int_A \varphi\,\mathrm{d}\mu$ and

$$\begin{aligned}\max_{A\in\mathcal{B}}\left(-\int_{A\cap I_x^+}\varphi\,\mathrm{d}\mu + \mu\big(I_x^+\big)\int_A\varphi\,\mathrm{d}\mu\right) &= \max_{B\subset I_x^+}\ \max_{A:\ A\cap I_x^+=B}\left(-\int_{A\cap I_x^+}\varphi\,\mathrm{d}\mu + \mu\big(I_x^+\big)\int_A\varphi\,\mathrm{d}\mu\right)\\ &= \max_{B\subset I_x^+}\left(-\int_B\varphi\,\mathrm{d}\mu + \mu\big(I_x^+\big)\int_{B\cup I_x^-}\varphi\,\mathrm{d}\mu\right)\\ &= \max_{B\subset I_x^+}\left(-\mu\big(I_x^-\big)\int_B\varphi\,\mathrm{d}\mu + \mu\big(I_x^+\big)\int_{I_x^-}\varphi\,\mathrm{d}\mu\right).\end{aligned}$$

The last functional in $B\subset I_x^+$ attaints the maximum if B is the smallest empty set. Thus

$$\max_{A\in\mathcal{B}}\left(-\int_{A\cap I_x^+}\varphi\,\mathrm{d}\mu + \mu\big(I_x^+\big)\int_A\varphi\,\mathrm{d}\mu\right) = \mu\big(I_x^+\big)\int_{I_x^-}\varphi\,\mathrm{d}\mu.$$

Summarizing the conclusions in two cases, we obtain the desired result. □

3.2. *Proof of Theorem* 2.1(i)

We separate its proof into three cases: $y_0\in I$, $x_0\in I$ or $I=(x_0,y_0)$:

Case 1. $y_0\in I$. Let g be ρ_1-Lipschitzian such that $\mu(g)=0$. By Lemma 3.1, if G is a solution of the Poisson equation (1.3), $G\in C^1(I)$, $G'\in\mathcal{AC}(I)$, and G' is given by (3.2). Since $G'(y_0)=0$, the constant C there must be zero. Now applying Lemma 3.2, we get

$$\big|G'(x)\big| \le \|g\|_{\mathrm{Lip}(\rho_1)} s'(x)\int_x^{y_0}\big[\rho_1(t)-\mu(\rho_1)\big]m'(t)\,\mathrm{d}t \le \|g\|_{\mathrm{Lip}(\rho_1)} c_{\mathrm{Lip}}(\rho_1,\rho_2)\rho_2'(x)$$

for $\mathrm{d}x$-a.e. $x\in I$. This yields to (2.2).

We turn to prove the existence of solution to the Poisson equation (1.3). Let G be a primitive of

$$G'(x) = s'(x)\int_x^{y_0} g(y)m'(y)\,\mathrm{d}y.$$

By what shown above, $\|G\|_{\mathrm{Lip}(\rho_2)}\le c_{\mathrm{Lip}}(\rho_1,\rho_2)\|g\|_{\mathrm{Lip}(\rho_1)}<\infty$, then $G\in L^2(I,\mu)$ (for $\rho_2\in L^2(I,\mu)$). By Lemma 3.1 and (A4), $G\in\mathbb{D}_2(\mathcal{L}^*)=\mathbb{D}(\mathcal{L}_2)$. Hence G is a solution of (1.3).

Finally, for $g=\rho_1-\mu(\rho_1)$, we see that $G'(x)=s'(x)\int_x^{y_0}[\rho_1(y)-\mu(\rho_1)]m'(y)\,\mathrm{d}y$. Then (2.2) becomes equality for that g.

Case 2. $x_0\in I$. Parallel to the case 1, for $G'(x)$ is again given by (3.2) with $C=0$.

Case 3. $I=(x_0,y_0)$. By the proof in case 1, we have only to show that for any solution G of (1.3), G' is given by (3.2) with $C=0$.

Assume in contrary that $C\ne 0$ in (3.2). Let G_0 be a fixed primitive of $s'(x)\int_x^{y_0} g(y)m'(y)\,\mathrm{d}y$. As shown above $\|G_0\|_{\mathrm{Lip}(\rho_2)}\le c_{\mathrm{Lip}}(\rho_1,\rho_2)\|g\|_{\mathrm{Lip}(\rho_1)}<+\infty$, then $G_0\in L^2(I,\mu)$ (for $\rho_2\in L^2(I,\mu)$). Therefore for some constant K,

$$G = Cs + G_0 + K.$$

But $s\notin L^2(I,\mu)$ by (A4), then $G\notin L^2(I,\mu)$, contrary to the assumption that $G\in\mathbb{D}(\mathcal{L}_2)\subset L^2(I,\mu)$. Thus $C=0$ as desired.

3.3. *Proof of Theorem* 2.1(ii)

At first notice that by (3.2), if $-\mathcal{L}_2G=g-\mu(g)$, then $G\in C^1(I)$, $G'\in\mathcal{AC}(I)$ and

$$G'(x) = s'(x)\left[C + \int_x^{y_0}\big[g(y)-\mu(g)\big]m'(y)\,\mathrm{d}y\right]. \tag{3.3}$$

We denote by G_0' the function above when $C=0$. We separate its proof into the three cases as in the proof of part (i):

Case 1. $y_0\in I$. Fix the measurable function g on I such that $|g|\le\varphi$. If G is a solution of $-\mathcal{L}_2G=g-\mu(g)$, as $G'(y_0)=0$, $C=0$ in (3.3), i.e., $G'=G_0'$. By Lemma 3.3, we have for dx-a.e. $x\in I$,

$$\left|G'(x)\right|=\left|G_0'(x)\right|\le s'(x)m(I)\left(\mu\left(I_x^+\right)\int_{I_x^-}\varphi\,\mathrm{d}\mu+\mu\left(I_x^-\right)\int_{I_x^+}\varphi\,\mathrm{d}\mu\right)\le c(\varphi,\rho)\rho'(x)$$

which gives us $\|G\|_{\mathrm{Lip}(\rho)}\le c(\varphi,\rho)$. Moreover, any primitive G_0 of G_0' satisfies $\|G_0\|_{\mathrm{Lip}(\rho)}\le c(\varphi,\rho)$, then $G_0\in L^2(I,\mu)$ (for $\rho\in L^2(I,\mu)$). By Lemma 3.1 and (A4), G_0 is a solution of $-\mathcal{L}_2G=g-\mu(g)$.

Finally the supremum of $\|G\|_{\mathrm{Lip}(\rho)}$ over $\{g;|g|\le\varphi\}$ equals to $c(\varphi,\rho)$, by Lemma 3.3.

Case 2. $x_0\in I$. Same as the proof of case 1.

Case 3. $x_0,y_0\notin I$. As in the proof of case 3 in part (i), we have G' is given by (3.3) with $C=0$. Now one can repeat the proof of case 1 to conclude.

Remark 3.4. *For some partial extensions of the results here to multi-dimensional Riemannian manifolds case, see the second named author* [26].

Acknowledgements

We are grateful to the referee for the two constructive reports which clarify and improve the presentation of the paper. The second named author thanks Prof. M. F. Chen for his interests in this work and kind invitation to Beijing Normal University.

References

[1] F. Barthe and A. V. Kolesnikov. Mass transport and variants of the logarithmic Sobolev inequality. *J. Geom. Anal.* **18** (2008) 921–979. MR2438906
[2] F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Math.* **159** (2003) 481–497. MR2052235
[3] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999) 1–28. MR1682772
[4] S. G. Bobkov and C. Houdre. Some connections between isoperimetric and Sobolev-type inequalities. *Mem. Amer. Math. Soc.* **129** (1997) No. 616. MR1396954
[5] S. G. Bobkov and C. Houdre. Isoperimetric constants for product probability measures. *Ann. Probab.* **25** (1997) 184–205. MR1428505
[6] S. G. Bobkov and M. Ledoux. Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab.* **37** (2009) 403–427. MR2510011
[7] P. Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup.* **15** (1982) 213–230. MR0683635
[8] M. F. Chen. Analytic proof of dual variational formula for the first eigenvalue in dimension one. *Sci. China Ser. A* **42** (1999) 805–815. MR1738551
[9] M. F. Chen. *Eigenvalues, Inequalities, and Ergodic Theory*. Springer, London, 2005. MR2105651
[10] N. Demni and M. Zani. Large deviations for statistics of the Jacobi process. *Stochastic Process. Appl.* **119** (2009) 518–533. MR2494002
[11] H. Djellout. L^p-uniqueness for one-dimensional diffusions. In *Mémoire de D.E.A*. Université Blaise Pascal, Clermont-Ferrand, 1997.
[12] A. Eberle. *Uniqueness and Non-Uniqueness of Semigroups Generated by Singular Diffusion Operators. Lecture Notes in Mathematics* **1718**. Springer, Berlin, 1999. MR1734956
[13] N. Gozlan, Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. H. Poincaré Probab. Statist.* To appear.
[14] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** (2007) 235–283. MR2322697
[15] A. Guillin, C. Léonard, L. Wu and N. Yao. Transport–information inequalities for Markov processes (I). *Probab. Theory Related Fields* **144** (2009) 669–695. MR2496446
[16] A. Guillin, C. Léonard, F. Y. Wang and L. Wu. Transportation–information inequalities for Markov processes (II): Relations with other functional inequalities. Preprint. Available at http://arxiv.org/abs/0902.2101 or http://hal.archives-ouvertes.fr/hal-00360854/fr/.
[17] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*, 2nd edition. *North-Holland Mathematical Library* **24**. North-Holland, Amsterdam, 1989. MR1011252
[18] T. Klein, Y. Ma and N. Privault. Convex concentration inequality and forward/backward martingale stochastic calculus. *Electron. J. Probab.* **11** (2006) 486–512. MR2242653
[19] O. Ludger. Estimation for continuous branching processes. *Scand. J. Statist.* **25** (1998) 111–126. MR1614256
[20] M. Ledoux. *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI, 2001. MR1849347

[21] M. Ledoux. Spectral gap, logarithmic Sobolev constant, and geometric bounds. *Surv. Differ. Geom.* **IX** (2004) 219–240. MR2195409

[22] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** (2009) 1–43. MR2507637

[23] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000) 361–400. MR1760620

[24] F. Y. Wang. *Functional Inequalities, Markov Semigroup and Spectral Theory.* Chinese Sciences Press, Beijing, 2005.

[25] L. M. Wu. Moderate deviations of dependent random variables related to CLT. *Ann. Probab.* **23** (1995) 420–445. MR1330777

[26] L. Wu, Gradient estimates of Poisson equations on Riemannian manifolds and applications. *J. Funct. Anal.* **29** (2009) 1008–1022. MR2557733

[27] L. Wu and Y. P. Zhang. A new topological approach to the L^∞-uniqueness of operators and the L^1-uniqueness of Fokker–Planck equations. *J. Funct. Anal.* **241** (2006) 557–610. MR2271930

Quatrième Partie

Déviations modérées pour des variables aléatoires dépendantes

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Ann. I. H. Poincaré – PR 42 (2006) 393–416

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES

www.elsevier.com/locate/anihpb

Moderate deviations of empirical periodogram and non-linear functionals of moving average processes

H. Djellout [a,*], A. Guillin [b], L. Wu [a,c]

[a] *Laboratoire de Mathématiques, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France*
[b] *Ceremade, CNRS-UMR 7534, Université Paris IX Dauphine, 75775 Paris, France*
[c] *Department of Mathematics, Wuhan University, 430072, China*

Received 28 April 2004; received in revised form 17 February 2005; accepted 8 April 2005

Available online 17 November 2005

Abstract

A moderate deviation principle for non-linear functionals, with at most quadratic growth, of moving average processes (or linear processes) is established. The main assumptions on the moving average process are a Logarithmic Sobolev Inequality for the driving random variables and the continuity, or some (weaker) integrability condition on the spectral density (covering some cases of long range dependence). We also obtain the moderate deviation estimate for the empirical periodogram, exhibiting an interesting new form of the rate function, i.e. with a correction term compared to the Gaussian rate functional. As statistical applications we provide the moderate deviation estimates of the least square and the Yule–Walker estimators of the parameter of a stationary autoregressive process and of the Neyman–Pearson likelihood ratio test in the Gaussian case.

Résumé

Un principe de déviations modérées pour des fonctionnelles non linéaires, à croissances quadratiques, des processus de moyennes mobiles (ou processus linéaire) est établi. Les conditions imposées sur le processus de moyennes mobiles sont une inégalité de Sobolev Logarithmique sur les variables aléatoires d'innovation et la continuité, ou une condition (plus faible) d'intégrabilité sur la densité spectrale (couvrant certains cas de longue mémoire). On obtient aussi une estimation des déviations modérées pour le périodogramme empirique, faisant apparaître une nouvelle forme de la fonction de taux, avec un terme correctif comparé à la fonction de taux gaussienne. Comme applications statistiques, on donne des estimations de déviations modérées pour les estimateurs de Yule–Walker et des moindres carrés du paramètre de processus autoregressif stationnaire, ainsi que pour le test de Neyman–Pearson pour le rapport de vraisemblance dans le cadre gaussien.

MSC: 60F10; 60G10; 60G15

Keywords: Moderate deviations; Moving average processes; Logarithmic Sobolev inequalities; Toeplitz matrices

* Corresponding author.
E-mail addresses: djellout@math.univ-bpclermont.fr (H. Djellout), guillin@ceremade.dauphine.fr (A. Guillin), Li-Ming.Wu@math.univ-bpclermont.fr (L. Wu).

0246-0203/$ – see front matter
doi:10.1016/j.anihpb.2005.04.006

1. Introduction

Consider the moving average process (or the linear process)

$$X_n := \sum_{j=-\infty}^{+\infty} a_{j-n}\xi_j = \sum_{j=-\infty}^{+\infty} a_j\xi_{n+j}, \quad \forall n \in \mathbb{Z} \tag{1.1}$$

where $(\xi_n)_{n\in\mathbb{Z}}$ is a sequence of $\mathbb{R}$-valued centered square integrable i.i.d.r.v., with common law $\mathcal{L}(\xi_0) = \mu$, and $(a_n)_{n\in\mathbb{Z}}$ be a sequence of real numbers such that

$$\sum_{n\in\mathbb{Z}} |a_n|^2 < +\infty. \tag{1.2}$$

This last condition (1.2) is necessary and sufficient for the a.s. convergence or convergence in law of the series (1.1) (see [17, Chapter 2]). The sequence (X_k) is strictly stationary with the spectral density given by

$$f(\theta) := \mathrm{Var}(\xi_0)\big|g(\theta)\big|^2$$

where

$$g(\theta) := \sum_{n=-\infty}^{+\infty} a_n \,\mathrm{e}^{\mathrm{i}n\theta}. \tag{1.3}$$

Moving average processes (or linear processes) are of special importance in time series analysis, filtrage of noise and they arise in a wide variety of contexts. Applications to economics, engineering and physical sciences are very broad and a vast amount of literature is devoted to the study of the limit theorems for moving average processes under various conditions (e.g. Brockwell and Davis [4] and references therein). A most important class of moving average processes is the real stationary Gaussian processes (X_n) with a square integrable spectral density function f (which can be represented as (1.1) with $\xi = \mathcal{N}(0,1)$ in law).

Let

$$\mathcal{I}_n(\theta) := \frac{1}{n}\left|\sum_{k=1}^{n} X_k \,\mathrm{e}^{\mathrm{i}k\theta}\right|^2, \tag{1.4}$$

be the so-called empirical periodogram of order n of the process (X_k). It is one of the main tools in the study of non-parametric statistical estimation of the unknown spectral density f on the basis of the sample $(X_1,\ldots,X_n)$ from the process (X_n). And for an observable $F(x) = F(x_0,\ldots,x_l)$ valued in $\mathbb{R}^m$, let

$$\frac{1}{n}S_n(F) := \frac{1}{n}\sum_{k=1}^{n} F(X_k, X_{k+1},\ldots,X_{k+l})$$

be the empirical mean of F. We begin with reviewing some known results which motivate our investigation.

(I) *Linear observables* $F(x) = x_0$.

  (a) The *minimal* condition for the central limit theorem (CLT in short) for $\frac{1}{n}\sum_{k=1}^n X_k$ is the continuity of g at $\theta = 0$ (see [17, Corollary 5.2, p. 135]).

  (b) Large deviations for $\frac{1}{n}\sum_{k=1}^n X_k$. See Burton and Dehling [7], Jiang, Rao and Wang [18,19], Djellout and Guillin [12] etc.

  For non-linear observables F, the limit theorems for $\frac{1}{n}S_n(F)$ becomes much more difficult, even in the particular Gaussian case.

(II) *Quadratic observables* $F(x) = (x_0^2, x_0x_1,\ldots,x_0x_l)$ *and* $\mathcal{I}_n(\theta)$.

  By Fourier series, one can often reduce the limit theorems of the empirical periodogram $\mathcal{I}_n(\theta)$ to those for $\frac{1}{n}S_n(F)$ where $F(x) = (x_0^2, x_0x_1,\ldots,x_0x_l)$. There exists an abundant literature on limit theorems of $\mathcal{I}_n(\theta)$ and of $\frac{1}{n}S_n(F)$ because of their importance in practice, especially in Gaussian case.

  (a) CLT. Avram [1] and Fox and Taqqu [15] proved the CLT for $\mathcal{I}_n(\cdot)$ and $\frac{1}{n}S_n(F)$ in the Gaussian case. This CLT was generalized by Giraitis and Surgailis [16] to non-Gaussian case.

(b) Large and moderate deviations. Bryc and Dembo [6] have considered quadratic functional $F(x) = x_0^2$ of Gaussian processes both at the level of large and moderate deviations, under the boundedness of f or the L^q-integrability of f respectively. But for $F(x) = x_0 x_l$ with $l \geqslant 1$, they have assumed that $f = 1$ (i.e., (X_k) are i.i.d.), an assumption excluding the dependent case.
Their result on large deviations (LDP in short) was generalized for general quadratic F by Bercu, Gamboa and Rouault [2] under some condition on the distribution of the eigenvalues of the involved Toeplitz matrix, always in the Gaussian case. This last "technical" condition, (wrongly) omitted in the precedent works, is optimal but quite difficult to check in practice. In [2], they provided several concrete important statistical examples for which their condition is fulfilled.
In [24], the third author proved the LDP of $\mathcal{I}_n$ and $\frac{1}{n} S_n(F)$ for general quadratic F, without the technical condition in [2], but under the following integrability condition $\mathbb{E}\,\mathrm{e}^{\lambda\xi^2} < +\infty$, $\forall \lambda > 0$, which excludes unfortunately the Gaussian case.

(III) *General non-linear observables F.*

(a) CLT. The literature is again abundant, we refer the reader to Rosenblatt [21] and the references therein.

(b) Large deviations. The seminal work of Donsker and Varadhan [14] established the LDP of the empirical process $R_n := \frac{1}{n}\sum_{k=1}^n \delta_{(X_k, X_{k+1}, \ldots)}$ for the stationary Gaussian processes such that $f \in C_b(\mathbb{T})$ and $\log f \in L^1(\mathbb{T})$. This implies the LDP of $\frac{1}{n} S_n(F)$ once if F is continuous and bounded. Bryc and Dembo [5] showed that the continuity of the spectral density f cannot be weakened but the condition $\log f \in L^1(\mathbb{T})$ can be removed, for the LDP result of Donsker and Varadhan. More recently, the third author [24] generalized this last result to all moving average processes such that $\mathbb{E}\,\mathrm{e}^{\delta\xi^2} < +\infty$ for some $\delta > 0$.

The main purpose of this paper consists to investigate the moderate deviation principle (MDP in short) for the so-called empirical periodogram $\mathcal{I}_n(\theta)$ of order n of the process (X_k) defined by (1.4) in the space $L^p(\mathbb{T}, \mathrm{d}\theta)$ of p-integrable function on the torus $\mathbb{T}$ identified with $[-\pi, \pi[$ equipped with the weak convergence topology. We establish the MDP for $\mathcal{I}_n(\theta)$ under some conditions such as the $L^q(\mathbb{T}, d\theta)$-integrability of the spectral density of (X_k) and a Logarithmic Sobolev Inequality (in short LSI) for the law μ of the driven random variable ξ. Moreover our approach allows us to obtain the MDP of $\frac{1}{n} S_n(F)$ for non-linear $\mathbb{R}^m$-valued observables F of at most quadratic growth.

To our knowledge, it is the first time that a MDP for a general class of non-linear observables of moving average processes is established (not only in the Gaussian case). Our investigation is a natural continuation of the known works [14,6,2,24] etc. We also consider statistical applications such as

(1) the MDP of the least square and Yule–Walker estimators of the autoregression parameter in a stationary autoregressive process, complementing known CLT results and the LDP (limited to the Gaussian case) obtained by Bercu et al. [2];
(2) the MDP in the Neyman–Pearson likelihood ratio test (largely inspired by Bercu and al. [2]) in the Gaussian case.

Besides the standard techniques in large deviations (such as approximation lemmas, projective limit etc.), our method is mainly based on the LSI technique, as developed by Ledoux [20] and al.

This paper is structured as follows. The MDP for the empirical spectral density and non-linear functionals are stated in next section. In Section 3, we provide statistical applications. We establish the key a priori estimations in Section 4. The last section is devoted to the proofs of the main results.

2. Main results

2.1. MDP for the empirical periodogram

For the sake of completeness, we recall the definition of the LDP [10] and [11]. A sequence of random variables (Y_n) with values in a regular Hausdorff topological space E is said to satisfy the LDP with speed $\lambda_n \to \infty$ and good rate function $I(\cdot): E \to \mathbb{R}^+$ if: I has compact level sets and for all measurable sets A of X:

$$-\inf_{x\in\mathring{A}} I(x) \leqslant \liminf_{n\to+\infty} \frac{1}{\lambda_n}\log\mathbb{P}(Y_n\in A) \leqslant \limsup_{n\to+\infty}\frac{1}{\lambda_n}\log\mathbb{P}(Y_n\in A) \leqslant -\inf_{x\in\bar{A}} I(x)$$

where $\mathring{A}$, $\bar{A}$ denote the interior and closure of A, respectively.

In the whole paper we shall study only a special type of LDP, called usually moderate deviation principle (MDP in short, cf. [10]).

Let $(\xi_n)_{n\in\mathbb{Z}}$ is a sequence of $\mathbb{R}$-valued centered i.i.d.r.v., with common law $\mathcal{L}(\xi_0)=\mu$, and let $a:=(a_n)_{n\in\mathbb{Z}}$ be a sequence of real numbers satisfying (1.2), and define (X_n) by (1.1). Our basic assumption, supposed throughout this paper, is that μ satisfies a LSI, i.e. there exists $C>0$ such that

$$\mathrm{Ent}_\mu(h^2)\leqslant 2C\mathbb{E}_\mu\big(|\nabla h|^2\big) \tag{2.1}$$

for every smooth h such that $\mathbb{E}_\mu(h^2\log^+ h^2)<\infty$, where

$$\mathrm{Ent}_\mu(h^2)=\mathbb{E}_\mu(h^2\log h^2)-\mathbb{E}_\mu(h^2)\log\mathbb{E}_\mu(h^2).$$

See Ledoux [20] for further details on LSI. Note that it implies in particular that there exists some positive δ such that

$$\mathbb{E}_\mu\big(\mathrm{e}^{\delta|x|^2}\big)<\infty. \tag{2.2}$$

Remark 2.1. First note that there exists some practical criteria ensuring the LSI. For example, consider a C^2 function W on $\mathbb{R}^d$ such that e^{-W} is integrable with respect to Lebesgue measure and let

$$\mathrm{d}\mu(x)=Z^{-1}\,\mathrm{e}^{-W(x)}\,\mathrm{d}x$$

where Z is the normalization constant, and suppose that for some $c\in\mathbb{R}$, $W''(x)\geqslant cI$ for every x and that for some $\epsilon>0$,

$$\iint \mathrm{e}^{(c^-+\epsilon)|x-y|^2}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)<\infty$$

where $c^-=-\min(c,0)$. Then μ satisfies (2.1) by the criterion of Wang [20]. Obviously Gaussian variables fulfill this criterion. See Bobkov and Götze [3] for a necessary and sufficient condition in the actual one-dimensional case, relying on generalized Hardy's inequalities.

We are interested in the moderate deviation principle (MDP in short) of the empirical spectral density (or periodogram) of (X_n) defined by

$$\mathcal{I}_n(\theta):=\frac{1}{n}\left|\sum_{k=1}^{n}X_k\,\mathrm{e}^{\mathrm{i}k\theta}\right|^2$$

which are random elements in the space $L^p(\mathbb{T},\mathrm{d}\theta)$ equipped with the weak convergence topology, where $\mathbb{T}$ is the torus identified with $[-\pi,\pi[$ in the usual way.

We first present here the MDP for the empirical autocorrelation vector which will be our main tool for the MDP of the empirical spectral density, and has its own interest in statistics. Let

$$\kappa_4=\frac{\mathbb{E}(\xi^4)-3[\mathbb{E}(\xi^2)]^2}{\mathbb{E}(\xi^2)^2},$$

the cumulant of order 4 of the driven random variable ξ.

Theorem 2.1. *Assume that μ satisfies the LSI* (2.1). *Suppose moreover that*

(H1) *the spectral density function f is in $L^q(\mathbb{T},\mathrm{d}\theta)$, where $2<q\leqslant+\infty$; and*
(H2) *the moderate deviation scale (b_n) is a sequence of positive numbers satisfying $1\ll b_n\ll\sqrt{n}$ (i.e. $b_n\to+\infty$ and $b_nn^{-1/2}\to 0$, the moderate deviation scale) and*

$$b_nn^{1/q-1/2}\to 0,$$

then for every $\lambda \in \mathbb{R}^{m+1}$,

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E} \exp\left(\frac{b_n}{\sqrt{n}} \sum_{\ell=0}^{m} \lambda_l \sum_{k=1}^{n} (X_k X_{k+\ell} - \mathbb{E} X_k X_{k+\ell})\right) = \frac{1}{2} \lambda^* \Sigma^2 \lambda \tag{2.3}$$

where $\Sigma^2 = (\Sigma^2_{k,\ell})_{0\leqslant k,\ell\leqslant m}$ *is given by*

$$\begin{aligned}\Sigma^2_{k,\ell} &= \frac{1}{2\pi} \int_{\mathbb{T}} \left(e^{i(k-\ell)\theta} + e^{i(k+\ell)\theta}\right) f^2(\theta)\, d\theta + \kappa_4 \left(\frac{1}{2\pi} \int_{\mathbb{T}} f(\theta) e^{ik\theta}\, d\theta\right)\left(\frac{1}{2\pi} \int_{\mathbb{T}} f(\theta) e^{i\ell\theta}\, d\theta\right) \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} 2\cos(k\theta)\cos(\ell\theta) f^2(\theta)\, d\theta + \kappa_4 \left(\frac{1}{2\pi} \int_{\mathbb{T}} f(\theta)\cos(k\theta)\, d\theta\right)\left(\frac{1}{2\pi} \int_{\mathbb{T}} f(\theta)\cos(l\theta)\, d\theta\right).\end{aligned} \tag{2.4}$$

In particular

$$\left(\frac{1}{b_n\sqrt{n}} \sum_{k=1}^{n} (X_k X_{k+\ell} - \mathbb{E} X_k X_{k+\ell})\right)_{0\leqslant \ell\leqslant m}$$

satisfies the LDP on $\mathbb{R}^{m+1}$ *with speed* b_n^2 *and with the rate function given by*

$$I(z) = \sup_{\lambda\in\mathbb{R}^{m+1}} \left\{\langle \lambda, z\rangle - \frac{1}{2}\langle \lambda, \Sigma^2 \lambda\rangle\right\}.$$

Remark 2.2. By Cauchy–Schwartz inequality we have $[\mathbb{E}(\xi^2)]^2 \leqslant \mathbb{E}(\xi^4)$, so $\kappa_4 \geqslant -2$ and $\kappa_4 = -2$ iff $\xi^2 = C$, a.s. Under the assumption (2.1), ξ^2 cannot be constant by [13, Remark 2.4], so $\kappa_4 > -2$. Consequently the matrix Σ^2 is symmetric and non-negative definite. Notice that the rate function I given above can be calculated explicitly as

$$I(z) = \begin{cases} \frac{1}{2}\langle z, \Sigma^{-2} z\rangle, & \text{if } z \in \operatorname{Ran}(\Sigma^2); \\ +\infty, & \text{otherwise,}\end{cases}$$

where Σ^{-2} is the inverse of Σ^2 restricted to the range $\operatorname{Ran}(\Sigma^2)$ of Σ^2.

Remark 2.3. The assumptions (H1) and (H2) on f and the scale b_n are exactly the ones imposed in Bryc and Dembo [6, Theorem 2.3] for the MDP of $\frac{1}{n}\sum_{k=1}^{n} X_k^2$ in the Gaussian case. Their large deviations result (namely Proposition 2.5 in [6]) for the empirical autocorrelation is further restricted to the i.i.d. case.

Remark 2.4. Notice that the condition (H1) on the dependence is indeed quite weak and general. It covers not only the short-range case (i.e. $\sum |\operatorname{Cov}(X_0, X_n)| < +\infty$), but also some cases of long range. To illustrate this case, consider the following example: let $\{B_H(t),\ t \in \mathbb{R}\}$ be the fractional Brownian motion with Hurst parameter $0 < H < 1$. Consider its increments

$$Y_j = B_H(j+1) - B_H(j), \quad j \in \mathbb{Z},$$

which form a stationary Gaussian sequence with mean zero and variance $\mathbb{E}(B_H^2(1)) = \sigma_0^2$. The sequence $\{Y_j,\ j \in \mathbb{Z}\}$ has the covariance function

$$\alpha(j) = \mathbb{E}(Y_1 Y_{j+1}) = \frac{\sigma_0^2}{2}\left(|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right),$$

and the spectral density

$$f(\lambda) = \frac{\sigma_0^2}{C^2}\left|e^{i\lambda} - 1\right|^2 \sum_{k=-\infty}^{+\infty} \frac{1}{|\lambda + 2\pi k|^{2H+1}}, \quad -\pi \leqslant \lambda \leqslant \pi,$$

where C is a constant depending only on H. It is known that (see [22])

$$\alpha(j) \sim \sigma_0^2 H(2H-1) j^{2H-2}, \quad \text{as } j \to \infty, \text{ for } H \neq 1/2,$$

and f is continuous on $\mathbb{T}\setminus\{0\}$ and

$$f(\lambda)\sim\sigma_0^2C^{-2}(H)|\lambda|^{1-2H},\quad \text{as } \lambda\to 0.$$

When $0<H\leqslant 1/2$, f is continuous (and then bounded) on $\mathbb{T}$. So the MDP in Theorem 2.1 holds for every moderate deviation scale (b_n).

When $1/2<H<1$, the series $\sum\alpha(j)$ diverges. In this case $\{Y_j, j\in\mathbb{Z}\}$ exhibits long range dependence. The condition (H1) is satisfied if $1/2<H<1/2+1/(2q)$, (so $1/2<H<3/4$ for $q>2$), and (1.2) is thus easily verified. In this case, we obtain the MDP of Theorem 2.1 for the sequence $\{Y_j,\ j\in\mathbb{Z}\}$ with $\kappa_4=0$ for the moderate deviation scale (b_n) verifying (H2).

The following corollary follows from Theorem 2.1 by the contraction principle

Corollary 2.2. *Under the assumptions of Theorem* 2.1*, we have for all* $\ell\geqslant 0$, $(1/(\sqrt{n}b_n)\sum_{k=1}^n(X_kX_{k+\ell}-\mathbb{E}X_kX_{k+\ell}))$ *satisfies the LDP on* $\mathbb{R}$ *with speed* b_n^2 *and rate function given by*

$$I^\ell(z)=\frac{1}{2}\frac{z^2}{1/(2\pi)\int_{\mathbb{T}}2\cos^2(\ell\theta))f^2(\theta)\,\mathrm{d}\theta+\kappa_4(1/(2\pi)\int_{\mathbb{T}}f(\theta)\cos(\ell\theta)\,\mathrm{d}\theta)^2}$$

with the convention that $a/0=+\infty$ *for* $a>0$ *and* $0/0:=0$.

Let us present now the main result of this paper. From Theorem 2.1 (and its proof) together with the projective limit method, we yield the functional type's MDP below, for

$$\mathcal{L}_n(\theta)=\frac{\sqrt{n}}{b_n}\big(\mathcal{I}_n(\theta)-\mathbb{E}\mathcal{I}_n(\theta)\big).$$

Theorem 2.3. *Suppose that* μ *satisfies the LSI* (2.1) *and* (H1), (H2). *Let* $1\leqslant p<2$ *and* $p'\in[2,+\infty]$ *the conjugated number, i.e.,* $1/p+1/p'=1$. *Assume moreover*

(H3) *the moderate deviation scale* b_n *satisfies*

$$b_nn^{1/q+1/p'-1/2}\to 0,\quad \frac{1}{p'}+\frac{1}{q}<\frac{1}{2}.$$

Then $\mathcal{I}_n(\theta)$ *satisfies the MDP, i.e.,* $(\mathcal{L}_n)_{n\geqslant 0}$ *satisfies the LDP on* $(L^p(\mathbb{T},\mathrm{d}\theta),\sigma(L^p(\mathbb{T},\mathrm{d}\theta),L^{p'}(\mathbb{T},\mathrm{d}\theta)))$ *with speed* b_n^2 *and with the rate function given by*

$$J(\eta)=\begin{cases}\dfrac{1}{2\pi}\displaystyle\int_{\mathbb{T}}\frac{\eta^2(\theta)}{4f^2(\theta)}\,\mathrm{d}\theta-\frac{\kappa_4}{2+\kappa_4}\left(\frac{1}{2\pi}\int_{\mathbb{T}}\frac{\eta(\theta)}{2f(\theta)}\,\mathrm{d}\theta\right)^2,\\ \qquad \textit{if } \kappa_4>-2,\ \eta \textit{ is even},\ \eta\,\mathrm{d}\theta\ll f\,\mathrm{d}\theta \textit{ and } \dfrac{\eta}{f}\in L^2(\mathbb{T},\mathrm{d}\theta);\\ +\infty,\quad \textit{otherwise.}\end{cases}$$

As a consequence of Theorem 2.3 we have the following marginal MDP:

Corollary 2.4. *Under the assumptions of Theorem* 2.3*, we have that for all* $h\in L^{p'}(\mathbb{T},\mathrm{d}\theta)$,

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\big(\mathrm{e}^{b_n^2\frac{1}{2\pi}\int_{\mathbb{T}}h(\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta}\big)=\frac{1}{2}\sigma^2(h),$$

where

$$\sigma^2(h):=\frac{1}{2\pi}\int_{\mathbb{T}}2\tilde{h}^2(\theta)f^2(\theta)\,\mathrm{d}\theta+\kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}}h(\theta)f(\theta)\,\mathrm{d}\theta\right)^2$$

and $\tilde{h}(\theta)=(h(\theta)+h(-\theta))/2$. *In particular* $\frac{1}{2\pi}\int_{\mathbb{T}}h(\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta$ *satisfies the LDP on* $\mathbb{R}$ *with speed* b_n^2 *and with the rate function given by* $I_h(z):=(1/2)z^2/(\sigma^2(h))$.

Remark 2.5. One cannot hope that the MDP in Theorem 2.3 holds w.r.t. the strong topology of $L^p(\mathbb{T}, \mathrm{d}\theta)$, because the rate function $I(\eta)$ is not inf-compact w.r.t. this topology.

The assumption (H3) is stronger than (H2). When $p = 1$ (and then $p' = +\infty$), (H3) becomes (H2) and thus under the LSI for ξ and (H1) and (H2), $\mathcal{L}_n(\theta)$ satisfies the LDP on $L^1(\mathbb{T})$ w.r.t. the weak convergence topology $\sigma(L^1, L^\infty)$ in Theorem 2.3, and $\frac{1}{2\pi}\int_{\mathbb{T}} h(\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta$ satisfies the LDP in Corollary 2.4 for every $h \in L^\infty(\mathbb{T})$.

Remark 2.6. Now assume that (ξ_n) is a sequence of real i.i.d. normal random variables, so (X_n) is a stationary Gaussian process and inversely any real Gaussian stationary process (X_n) with a square integrable spectral density function f can be represented as (1.1). In this case, we have $\mathbb{E}(\xi^4) = 3\mathbb{E}(\xi^2)^2$ and thus $\kappa_4 = 0$, so under the assumptions of Theorem 2.3 we obtain that $(\mathcal{L}_n)_{n\geqslant 0}$ satisfies the LDP on $L^p(\mathbb{T}, d\theta)$ with speed b_n^2 and with the rate function given by

$$J(\eta) = \begin{cases} \dfrac{1}{2\pi}\displaystyle\int_{\mathbb{T}} \dfrac{\eta^2(\theta)}{4f^2(\theta)}\,\mathrm{d}\theta, & \text{if } \eta \text{ is even, } \eta\,\mathrm{d}\theta \ll f\,\mathrm{d}\theta \text{ and } \dfrac{\eta}{f} \in L^2(\mathbb{T}, \mathrm{d}\theta), \\ +\infty, & \text{otherwise.} \end{cases}$$

We thus give the MDP for the spectral empirical measure in the setting of Bercu and al. [2]. Note however that they only consider the marginal LDP, i.e. LDP for $\mathcal{I}_n(h)$ for some bounded h on the torus with an extra assumption on the eigenvalues of the Toeplitz matrix, where $\mathcal{I}_n(h) = \frac{1}{2\pi}\int_{\mathbb{T}} \mathcal{I}_n(\theta)h(\theta)\,\mathrm{d}\theta$.

Remark 2.7. For any real and symmetric function $h \in L^1(\mathbb{T}, \mathrm{d}\theta)$, let $T_n(h)$ be the Toeplitz matrix of order n associated with h i.e. $T_n(h) = (\hat{r}_{k-l}(h))_{1\leqslant k,l\leqslant n}$ where $\hat{r}_k(h)$ is the kth Fourier coefficient of h given by

$$\hat{r}_k(h) = \frac{1}{2\pi}\int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}k\theta} h(\theta)\,\mathrm{d}\theta, \quad \forall k \in \mathbb{Z}. \tag{2.5}$$

The matrix $T_n(h)$ is obviously real and symmetric, is positive definite whenever $h \geqslant 0$.

Notice that the extra term with respect to the Gaussian case in the evaluation of the asymptotic variance has been known for a long time (see [21]). The result of [16] about CLT for $\mathcal{I}_n$ can be summarized as below: if

$$\lim_{n\to\infty} \frac{1}{n}\operatorname{tr}\bigl(\bigl(T_n(f)T_n(h)\bigr)^2\bigr) = \frac{1}{2\pi}\int_{\mathbb{T}} f^2(\theta)h^2(\theta)\,\mathrm{d}\theta; \tag{2.6}$$

(where $T_n(h)$ is the Toeplitz matrix of h) then $\sqrt{n}(\mathcal{I}_n(h) - \mathbb{E}\mathcal{I}_n(h))$ converges in law (as $n \to \infty$) to the normal distribution $\mathcal{N}(0, \sigma^2(h))$ with $\sigma^2(h)$ given in Corollary 2.4. In Gaussian case this result was already proved by Avram [1] and Fox and Taqqu [15].

In the next corollaries of Theorem 2.3, we replace $\mathbb{E}\mathcal{I}_n(\theta)$ by $f(\theta)$ in the definition of $\mathcal{L}_n(\theta)$, more useful in practice, but need more assumptions. More precisely we are interested in the MDP of

$$\tilde{\mathcal{L}}_n(\theta) = \frac{\sqrt{n}}{b_n}\bigl(\mathcal{I}_n(\theta) - f(\theta)\bigr).$$

Corollary 2.5. *Suppose that μ satisfies the LSI* (2.1) *and the spectral density f verifies*

$$f \in L^\infty(\mathbb{T}) \quad \text{and} \quad \bigl\|f(t+\cdot) - f(\cdot)\bigr\|_{L^p(\mathbb{T})} = \mathrm{O}(\sqrt{t}) \tag{2.7}$$

then for every scale $1 \ll b_n \ll n^{1/2-1/p'}$, $(\tilde{\mathcal{L}}_n)_{n\geqslant 0}$ satisfies the LDP on $L^p(\mathbb{T}, \mathrm{d}\theta)$ w.r.t. the weak topology $\sigma(L^p(\mathbb{T}), L^{p'}(\mathbb{T})))$, with speed b_n^2 and with the rate function J given in Theorem 2.3.

We have also the following consequence of Corollary 2.4 for the marginals of the empirical spectral measures

Corollary 2.6. *Assume* (2.1) *and* (H1), (H2). *Suppose that*

$$h \in L^\infty(\mathbb{T}) \quad \text{and} \quad \bigl\|h(t+\cdot) - h(\cdot)\bigr\|_{L^{q'}(\mathbb{T})} = \mathrm{O}(\sqrt{t}) \tag{2.8}$$

then the conclusion of Corollary 2.4 *holds for $\int_\pi h(\theta)\tilde{\mathcal{L}}_n(\theta)\,\mathrm{d}\theta$ instead of $\int_\pi h(\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta$.*

2.2. MDP for non-linear functionals

We now present the MDP of $\frac{1}{n}S_n(F)$, i.e., the LDP of

$$M_n(F) = \frac{1}{\sqrt{n}b_n}\sum_{k=1}^{n}\big(F(X_k, \dots, X_{k+l}) - \mathbb{E}\big(F(X_k, \dots, X_{k+l})\big)\big),$$

where the observable $F:\mathbb{R}^{l+1} \to \mathbb{R}^m$ is a general non-linear differentiable function.

Theorem 2.7. *Suppose that μ satisfies the LSI* (2.1)*, and g given in* (1.3) *is continuous on $\mathbb{T}$. Assume moreover that $\partial_{x_i}F$ is Lipschitz for $i = 0, \dots, l$. Then*

$$\Sigma_F^2 := \lim_{n\to+\infty}\frac{1}{n}\Gamma\left(\sum_{k=1}^{n}F(X_k, \dots, X_{k+l})\right) \tag{2.9}$$

exists where $\Gamma(\cdot)$ is the covariance matrix of the random vector $\cdot$, and for every *moderate deviation scale $1 \ll b_n \ll \sqrt{n}$, $M_n(F)$ satisfies the LDP on $\mathbb{R}^m$ with speed b_n^2 and good rate function I_F given by*

$$I_F(z) = \sup_{\lambda\in\mathbb{R}^m}\left\{\langle\lambda, z\rangle - \frac{1}{2}\langle\lambda, \Sigma_F^2\lambda\rangle\right\} = \begin{cases}\frac{1}{2}\langle z, \Sigma_F^{-2}z\rangle, & \text{if } z \in \mathrm{Ran}(\Sigma_F^2),\\ +\infty, & \text{otherwise,}\end{cases}$$

where $\Sigma_F^{-2}:\mathrm{Ran}(\Sigma_F^2) \to \mathrm{Ran}(\Sigma_F^2)$ is the inverse of the limit covariance matrix Σ_F^2 restricted to $\mathrm{Ran}(\Sigma_F^2)$.

Note also the following corollary in the linear case $F(x_0, \dots, x_l) = x_0$ in which the assumption on g can be largely weakened.

Corollary 2.8. *Suppose that μ satisfies the integrability condition* (2.2)*, if*

$$f^N(0) = \sum_{|k|\leqslant N}\left(1 - \frac{|k|}{N}\right)\hat{r}_k(f) \to \sigma^2,$$

then for every moderate deviation scale $1 \ll b_n \ll \sqrt{n}$, $\frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n}X_k$ satisfies the LDP on $\mathbb{R}$ with speed b_n^2 and rate $I(z) = \frac{1}{2}\frac{z^2}{\sigma^2}$.

Remark 2.8. When f admits a version which is continuous at 0, then $f^N(0) \to f(0) = \sigma^2$. This corollary generalizes Theorem 3.1 of Djellout and Guillin [12] to the case of unbounded r.v.

3. Statistical applications

We now provide two statistical applications. The first deals with the least square estimator of the parameter of the autoregressive linear process and the second about the likelihood ratio test on spectral densities in the Gaussian case.

3.1. Autoregressive stationary process

Consider the autoregressive process (not necessarily Gaussian)

$$X_{n+1} = \theta X_n + \sigma\xi_{n+1},$$

where the noises sequence $((\xi_n)_{n\in\mathbb{Z}})$ is i.i.d. with common law μ, satisfying a LSI, and $\mathbb{E}(\xi_n) = 0$, $\mathbb{E}(\xi^2) = 1$, $\sigma > 0$ and $\theta \in (-1, 1)$ is the unknown parameter. Assume that X_0 is independent of $(\xi_n)_{n\geqslant1}$ and has the same law as $\sum_{k=0}^{\infty}\theta^k\sigma\xi_{-k}$. (X_n) is thus a centered stationary process of the form (1.1), with spectral density given by

$$f(t) = \frac{\sigma^2}{1 + \theta^2 - 2\theta\cos t}, \quad \forall t \in \mathbb{T}.$$

Let $\hat{\theta}_n$ be the least square estimator of θ, given by:

$$\hat{\theta}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}.$$

It is well-known that $\hat{\theta}_n \to \theta$ a.s. and $\sqrt{n}(\hat{\theta}_n - \theta)$ satisfies the CLT. We show in the next proposition the MDP of the least square estimator.

Proposition 3.1. *For every moderate deviation scale $1 \ll b_n \ll \sqrt{n}$, $\frac{\sqrt{n}}{b_n}(\hat{\theta}_n - \theta)$ satisfies a LDP on $\mathbb{R}$ with speed b_n^2 and with the rate function given by*

$$I(x) = \frac{x^2}{2(1-\theta^2)}.$$

Remark 3.1. Let $\tilde{\theta}_n$ be the Yule–Walker estimator of θ:

$$\tilde{\theta}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=0}^n X_i^2}.$$

It is well-known that the Yule–Walker estimator share the same almost sure property and the same CLT. Bercu et al. [2] showed however that the LDP of the Yule–Walker estimator is better than the one of the least-squares.

In the regime of the MDP, following the same proof as for the least square estimator we see that the Yule–Walker estimator share the same MDP.

3.2. Likelihood ratio test in the Gaussian case

Let f_0 and f_1 be two spectral densities which differ on a positive Lebesgue measure subset of $\mathbb{T}$. If we wish to test H_0: $f = f_0$ against H_1: $f = f_1$, on the basis of the stationary centered Gaussian observation $X_1, \ldots, X_n$, the Neyman–Pearson theorem tells us that the optimal strategy is the likelihood ratio test:

$$L_n = \frac{1}{2n}\left(\log \frac{\det T_n(f_0)}{\det T_n(f_1)} + \left\langle X^{(n)}, \left[T_n(f_0)^{-1} - T_n(f_1)^{-1}\right] X^{(n)}\right\rangle\right).$$

The study of the MDP properties of (L_n) under hypothesis H_0 or H_1 is useful to control asymptotically the threshold or the power of the test. We now make the two following assumptions:

(A_1) the spectral density f_0 is in the Szegö class, i.e. $\log(f_0) \in L^1(\mathbb{T})$;
(A_2) the ratio $f_0/f_1 \in L^\infty(\mathbb{T})$.

Under those assumption, Bercu and al. [2] proved that L_n converges a.s. to

$$\frac{1}{4\pi}\left(\int_{\mathbb{T}} \log f_0(t)\,\mathrm{d}t - \int_{\mathbb{T}} \log f_1(t)\,\mathrm{d}t + \int_{\mathbb{T}} \left(1 - \frac{f_0(t)}{f_1(t)}\right)\mathrm{d}t\right)$$

and satisfies the LDP. Inspired by their work we have furthermore

Proposition 3.2. *Assume that* (A_1) *and* (A_2) *are satisfied. Then, under the null hypothesis* H_0, *for every moderate deviation scale $1 \ll b_n \ll \sqrt{n}$, the sequence $\frac{\sqrt{n}}{b_n}(L_n - \mathbb{E}(L_n))$ satisfies a LDP on $\mathbb{R}$ with speed b_n^2 and good rate function*

$$G(x) = \frac{x^2}{(1/(2\pi))\int_{\mathbb{T}}(1 - f_0/f_1)^2(\theta)\,\mathrm{d}\theta}.$$

4. Several lemmas

In this section we first establish the a priori estimate, next recall several facts concerning the Toeplitz matrix and the Fejèr approximation and the MDP of m-dependent stationary sequences.

4.1. A priori estimation

We recall the following well known elementary result

Lemma 4.1. *Suppose that $Y^{(n)} = (Y_1, \ldots, Y_n)^*$ is a standard $\mathcal{N}(0, I)$ centered Gaussian vector valued in $\mathbb{R}^n$ and let A be a symmetric real valued $n \times n$-matrix. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the matrix A. Then for every $z \in \mathbb{R}$*

$$\log \mathbb{E} \exp\big(z\langle Y^{(n)}, AY^{(n)}\rangle\big) = \begin{cases} -\dfrac{1}{2}\displaystyle\sum_{j=1}^{n} \log(1 - 2z\lambda_j) & \text{if } \max\limits_{1 \leqslant j \leqslant n} (z\lambda_j) < \dfrac{1}{2}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.1}$$

We give a crucial lemma which was first proved in Wu [24], and reproduced here for completeness.

Lemma 4.2. *If the centered r.v. ξ_0 satisfies* (2.2)*, then there is some constant $K > 0$ such that*

$$L(y) := \mathbb{E} \exp(y\xi_0) \leqslant \exp\left(\frac{K^2}{2} y^2\right), \quad \forall y \in \mathbb{R}. \tag{4.2}$$

Proof. Let $\delta > 0$ be given in (2.2). Since

$$2y\xi_0 \leqslant 2\delta\xi_0^2 + \frac{1}{2\delta} y^2,$$

there is $C_1 > 0$ such that (4.2) holds for all $|y| > 1$.

For $|y| \leqslant 1$, notice that $\log L(y) \in C^\infty(\mathbb{R})$, and

$$\log L(0) = 0, \qquad \frac{\mathrm{d}}{\mathrm{d}y} \log L(y)\bigg|_{y=0} = \mathbb{E}\xi_0 = 0.$$

By Taylor's formula of order 2, we have for all y with $|y| \leqslant 1$,

$$\log L(y) \leqslant \frac{1}{2} C_2^2 y^2,$$

where

$$C_2 := \sup_{|y| \leqslant 1} \left| \frac{\mathrm{d}^2}{\mathrm{d}y^2} \log L(y) \right|^{1/2}.$$

Thus (4.2) follows with $K := C_1 \vee C_2$. □

We now extend (4.1) from Gaussian distribution to general law μ satisfying (2.2), which is a generalization of the preceding lemma.

Lemma 4.3. *Let (X_k) be the moving average process given by* (1.1) *and $T_n(f)$ the Toeplitz matrix associated with the spectral density function f of (X_k), given in Remark* 2.7*. Assume the integrability condition* (2.2) (*but not the stronger LSI*).

Let $X^{(n)} = (X_1, \ldots, X_n)^*$, B be a real non-negative definite symmetric $n \times n$-matrix, and $\mu_1^n, \ldots, \mu_n^n$ the eigenvalues of the matrix $\sqrt{B}\, T_n(f)\sqrt{B}$. Then for all $\lambda \geqslant 0$ satisfying $\lambda \max_{1 \leqslant j \leqslant n} \mu_j^n < 1/(2K^2)$, we have

$$\log \mathbb{E} \exp\big(\lambda\langle X^{(n)}, BX^{(n)}\rangle\big) \leqslant -\frac{1}{2}\sum_{j=1}^{n} \log(1 - 2K^2\lambda\mu_j^n),$$

where $K > 0$ is given in Lemma 4.2.

Proof. The main difficulty resides in the non-linear property of $\langle x, Bx\rangle$. The trick consists to reduce it to an estimation of linear type in the following way:

$$\mathbb{E}\{\mathrm{e}^{\frac{1}{2}t^2\langle X^{(n)},BX^{(n)}\rangle}\} = \mathbb{E}\{\mathrm{e}^{\frac{1}{2}t^2|\sqrt{B}X^{(n)}|^2}\} = \int_{\mathbb{R}^n} \mathbb{E}\{\mathrm{e}^{t\langle\sqrt{B}X^{(n)},y^{(n)}\rangle}\}\gamma(\mathrm{d}y^{(n)})$$

where γ is the standard Gaussian law $\mathcal{N}(0, I)$ on $\mathbb{R}^n$.

Since

$$\langle\sqrt{B}X^{(n)}, y^{(n)}\rangle = \langle X^{(n)}, \sqrt{B}y^{(n)}\rangle = \sum_{k=1}^{n} X_k(\sqrt{B}y^{(n)})_k = \sum_{j\in\mathbb{Z}} \xi_j \sum_{k=1}^{n} a_{j-k}(\sqrt{B}y^{(n)})_k.$$

We get by Lemma 4.2 and the i.i.d. property of (ξ_j),

$$\mathbb{E}\{\exp[t\langle\sqrt{B}X^{(n)}, y^{(n)}\rangle]\} \leqslant \exp\left[\frac{K^2t^2}{2}\sum_{j\in\mathbb{Z}}\left|\sum_{k=1}^{n} a_{j-k}(\sqrt{B}y^{(n)})_k\right|^2\right].$$

Now observe that

$$\begin{aligned}\sum_{j\in\mathbb{Z}}\left|\sum_{k=1}^{n} a_{j-k}(\sqrt{B}y^{(n)})_k\right|^2 &= \sum_{k,l=1}^{n}\sum_{j\in\mathbb{Z}} a_{j-k}a_{j-l}(\sqrt{B}y^{(n)})_k(\sqrt{B}y^{(n)})_l \\ &= \sum_{k,l=1}^{n} \big(T_n(f)\big)_{k,l}(\sqrt{B}y^{(n)})_k(\sqrt{B}y^{(n)})_l \\ &= \big\langle y^{(n)}, \sqrt{B}T_n(f)\sqrt{B}y^{(n)}\big\rangle.\end{aligned}$$

Then letting $\mu_1^n, \dots, \mu_n^n$ be the eigenvalues of the matrix $\sqrt{B}T_n(f)\sqrt{B}$ (which are also the eigenvalues of $T_n(f)B$), we get for all t such that $K^2t^2\max_{1\leqslant j\leqslant n}\mu_j^n < 1$,

$$\begin{aligned}\mathbb{E}\left\{\exp\left[\frac{1}{2}t^2\langle X^{(n)}, BX^{(n)}\rangle\right]\right\} &\leqslant \int_{\mathbb{R}^n}\left\{\exp\left[\frac{K^2t^2}{2}\big\langle y^{(n)}, \sqrt{B}T_n(f)\sqrt{B}y^{(n)}\big\rangle\right]\right\}\gamma(\mathrm{d}y^{(n)}) \\ &= -\frac{1}{2}\sum_{j=1}^{n}\log(1 - K^2t^2\mu_j^n)\end{aligned}$$

where the last equality follows by Lemma 4.1. Finally the desired result follows with $\lambda = t^2/2$. □

Remark 4.1. If we assume $\|g\|_\infty = \|g(\theta)\|_{L^\infty(\mathbb{T},\mathrm{d}\theta)} < +\infty$, and $B = I$ we obtain exactly the result in Wu [24]. In fact in this case, we have for any $\lambda > 0$ such that $2\lambda K^2\|g\|_\infty^2 < 1$,

$$\log\mathbb{E}\,\mathrm{e}^{\lambda\langle X^{(n)},X^{(n)}\rangle} \leqslant -\frac{1}{2}\log\big(1 - 2\lambda K^2\|g\|_\infty^2\big)^n, \tag{4.3}$$

because the eigenvalues of $T_n(f)$ are bounded by $\|f\|_\infty = \|g\|_\infty^2$.

Remark 4.2. Instead of Lemma 4.2, we can use the consequence of the LSI (5.3) below to prove Lemma 4.3, but (5.3) is stronger than (2.2).

4.2. Preparating lemmas

For an $n\times n$ matrix A, we consider the usual operator norm $\|A\| = \sup_{x\in\mathbb{R}^n}(|Ax|/|x|)$. Recall (cf. Remarks 2.7) that for any real and even function $h \in L^1(\mathbb{T}, \mathrm{d}\theta)$, $T_n(h)$ is the Toeplitz matrix of order n associated with h i.e. $T_n(h) = (\hat{r}_{k-l}(h))_{1\leqslant k,l\leqslant n}$ where $\hat{r}_k(h)$ is the kth Fourier coefficient of h given by

$$\hat{r}_k(h) = \frac{1}{2\pi}\int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}k\theta}h(\theta)\,\mathrm{d}\theta, \quad \forall k \in \mathbb{Z}.$$

Lemma 4.4 *((Avram [1], Lemma 1)). If $f \in L^q(\mathbb{T})$ where $1 \leqslant q \leqslant \infty$, then for all $n > 1$ we have $\|T_n(f)\| \leqslant n^{1/q}\|f\|_q$.*

Lemma 4.5 *((Avram [1], Theorem 1)). Let $f_k \in L^{q_k}(\mathbb{T}, \mathrm{d}\theta)$ with $q_k \geqslant 1$ for $k = 1, \ldots, p$ and $\sum_{k=1}^{p}(1/q_k) \leqslant 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{tr}\left(\prod_{k=1}^{p} T_n(f_k)\right) = \hat{r}_0\left(\prod_{k=1}^{p} f_k\right).$$

Introduce now the Fejèr approximation of g:

$$g^N(\theta) = \sum_{j\in\mathbb{Z}} a_j^N \,\mathrm{e}^{\mathrm{i}j\theta}, \quad \forall\theta \in \mathbb{R}, \text{ where } a_j^N = a_j\left(1 - \frac{|j|}{N}\right)1_{|j|\leqslant N}.$$

We recall the following (see [8])

Lemma 4.6. *$g^N(\theta) = \int_{-\pi}^{\pi} g(\theta - t)K_N(t)\,\mathrm{d}t$ where K_N is the Fejèr kernel of order N given by*

$$K_N(t) = \frac{1}{2\pi N}\left(\frac{\sin(Nt/2)}{\sin(t/2)}\right)^2, \quad t \in \mathbb{T}.$$

Furthermore for $g \in L^p(\mathbb{T})$ where $1 \leqslant p < \infty$, $g^N \to g$ in $L^p(\mathbb{T})$ and $g^N \to g$ uniformly on $\mathbb{T}$ if g is continuous. Moreover, K_n is even, non-negative and possesses the following properties for small δ:

$$(a) \quad \int_{\mathbb{T}} K_n(t)\,\mathrm{d}t = 1,$$

$$(b) \quad \int_{|t|\geqslant\delta} K_n(t)\,\mathrm{d}t \leqslant \frac{C}{n},$$

$$(c) \quad \int_{|t|\leqslant\delta} K_n(t)t^\alpha\,\mathrm{d}t \leqslant \begin{cases} Cn^{-\alpha}, & \alpha < 1,\\ Cn^{-1}\ln n, & \alpha = 1,\\ Cn^{-1}, & \alpha > 1. \end{cases}$$

Let m be a given positive integer, a sequence $(Z_n)_{n\geqslant 1}$ of strictly stationary random variables is called m-dependent if for every $k \geqslant 1$ the two collections $\{Z_1, \ldots, Z_k\}$ and $\{Z_{k+m}, Z_{k+m+1}, \ldots\}$ are independent. We have the following

Lemma 4.7 *((Chen X. [9])). Let $(Z_n)_{n\geqslant 1}$ be a stationary sequence of m-dependent random variables taking values in $\mathbb{R}^m$, such that*

$$\mathbb{E}\big(\mathrm{e}^{\alpha|Z_1|}\big) < +\infty, \quad \textit{for some } \alpha > 0.$$

Then for all $\lambda \in \mathbb{R}^m$,

$$\begin{aligned}\lim_{n\to+\infty} \frac{1}{b_n^2} \log \mathbb{E}\big(\mathrm{e}^{b_n^2\langle\lambda, \frac{1}{\sqrt{n}b_n}\sum_{k=1}^{n}(Z_k - \mathbb{E}Z_k)\rangle}\big) &= \frac{1}{2}\lim_{n\to\infty} \mathbb{E}\left\langle\lambda, \sum_{k=1}^{n}(Z_k - \mathbb{E}Z_k)\right\rangle^2\\ &= \frac{1}{2}\left(\mathbb{E}\langle\lambda, Z_1\rangle^2 + 2\sum_{k=2}^{m+1}\mathbb{E}\langle\lambda, Z_1\rangle\mathbb{E}\langle\lambda, Z_k\rangle\right).\end{aligned}$$

5. Proofs of the main results in Section 2

5.1. Proof of Theorem 2.1

The proof is divided into three steps. In the first one, we approximate the moving average process by a bilateral moving average process of finite range $2N$ which satisfies the MDP. Then we will show that this approximation is a good one in the sense of the MDP. In third step, we will finally establish the convergence of the rate function and the subsequent existence of the limiting variance.

Step 1 (Approximation by bilateral moving average process of finite range $2N$). Let $X_k^N = \sum_{j\in\mathbb{Z}} a_j^N \xi_{k+j}$, where $a_j^N = a_j(1-\frac{|j|}{N})1_{|j|\leqslant N}$, be the Fejèr approximation of X_k.

Put

$$Q_n^N = (Q_n^{N,l}) = \left(\frac{1}{\sqrt{n}b_n} Z_n^{N,l}\right)_{l=0,\ldots,m} \quad \text{and} \quad Q_n = (Q_n^l)_{l=0,\ldots,m} = \left(\frac{1}{\sqrt{n}b_n} Z_n^l\right),$$

where

$$Z_n^{N,l} = \sum_{k=1}^n (X_k^N X_{k+l}^N - \mathbb{E}X_k^N X_{k+l}^N) \quad \text{and} \quad Z_n^l = \sum_{k=1}^n (X_k X_{k+l} - \mathbb{E}X_k X_{k+l}).$$

The crucial remark is that the sequence $\{(X_k^N X_{k+l}^N)_{l=0,\ldots,m} \in \mathbb{R}^{m+1},\ k\in\mathbb{Z}\}$ is a $2N$-dependent stationary sequence. By (2.2), we get for all N there is $\eta > 0$ such that $\mathbb{E}(\mathrm{e}^{\eta|X_k^N X_{k+l}^N|}) < \infty$.

Then applying Lemma 4.7, we get that for each N fixed, for all $\lambda \in \mathbb{R}^{m+1}$,

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\left(\mathrm{e}^{b_n^2\langle\lambda, Q_n^N\rangle}\right) = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\langle\lambda, Z_n^{N,\cdot}\rangle^2 := \frac{1}{2}\langle\lambda, \Sigma^{2,N}\lambda\rangle \in \mathbb{R} \tag{5.1}$$

where $\Sigma^{2,N}$ is the limit covariance matrix given in Lemma 4.7, and that Q_n^N satisfies the MDP on $\mathbb{R}^{m+1}$ with the good rate function

$$I^N(x) = \sup_{\lambda\in\mathbb{R}^{m+1}} \left\{\langle\lambda, x\rangle - \frac{1}{2}\langle\lambda, \Sigma^{2,N}\lambda\rangle\right\}.$$

Furthermore, by [21], $\Sigma_{k,l}^{2,N}$ can be expressed as (2.4) with f replaced by f^N.

Step 2 (Exponential contiguity, see Section 4.2 in [10]). The purpose of this step will be to prove the asymptotic negligibility of $Q_n - Q_n^N$ with respect to the MDP as N goes to ∞, i.e. we will establish that for all $\lambda \in \mathbb{R}^{m+1}$,

$$\limsup_{N\to\infty} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\left(\mathrm{e}^{b_n^2\langle\lambda, Q_n - Q_n^N\rangle}\right) = 0.$$

As our functional $Q_n - Q_n^N$ are centered, by Jensen inequality we only have to establish the upper inequality in the equality above. By Jensen's inequality again,

$$\mathbb{E}\left(\mathrm{e}^{b_n^2\langle\lambda, Q_n - Q_n^N\rangle}\right) \leqslant \frac{1}{m+1} \sum_{l=0}^m \mathbb{E}\left(\mathrm{e}^{(m+1)b_n^2\lambda_l(Q_n^l - Q_n^{N,l})}\right),$$

we need only to show that for each $l = 0,\ldots,m$ fixed and for every $\lambda \in \mathbb{R}$,

$$\lim_{N\to\infty} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\left(\mathrm{e}^{b_n^2\lambda(Q_n^l - Q_n^{N,l})}\right) \leqslant 0. \tag{5.2}$$

To this end, our main tool is the following consequence of the LSI (2.1), see Ledoux [20, Theorem 2.7] (after having extended (2.1) by tensorization to the product measure of μ): for any integrable C^1 functional G of $\xi = (\xi_k)_{|k|\leqslant m}$,

$$\mathbb{E}\left(\mathrm{e}^{\lambda(b_n/\sqrt{n})(G-\mathbb{E}G)}\right) \leqslant \mathbb{E}\left(\mathrm{e}^{\lambda^2(b_n^2/n)C|\nabla_\xi G|^2}\right), \tag{5.3}$$

with C given in (2.1), where $|\nabla_\xi G|^2 := \sum_k |\partial_{\xi_k} G|^2$. This inequality can be extended to all integrable functionals $G = F(X_1,\ldots,X_n)$ where $F \in C^1(\mathbb{R}^n)$ by dominated convergence (even now X_k depends on the infinite sequence $(\xi_k)_{k\in\mathbb{Z}}$, the detail is left to the reader).

Let apply it to

$$G_n^{N,l}\big((\xi_i)_{i\in\mathbb{Z}}\big) = \sum_{k=1}^n (X_k X_{k+l} - X_k^N X_{k+l}^N),$$

so that our main estimations are now transferred to the gradient of $G_n^{N,l}$.

Clearly

$$\partial_{\xi_i} G_n^{N,l} = \sum_{k=1}^{n} (a_{i-k} X_{k+l} + a_{i-k-l} X_k - a_{i-k}^N X_{k+l}^N - a_{i-k-l}^N X_k^N);$$

so

$$\begin{aligned}|\nabla G_n^{N,l}|^2 &\leqslant 4 \sum_{i\in\mathbb{Z}} \left(\left(\sum_{k=1}^{n} (a_{i-k} - a_{i-k}^N) X_{k+l} \right)^2 + \left(\sum_{k=1}^{n} (a_{i-k-l} - a_{i-k-l}^N) X_k \right)^2 \right. \\ &\quad \left. + \left(\sum_{k=1}^{n} a_{i-k}^N (X_{k+l} - X_{k+l}^N) \right)^2 + \left(\sum_{k=1}^{n} a_{i-k-l}^N (X_k - X_k^N) \right)^2 \right) \\ &= (I) + (II) + (III) + (IV).\end{aligned}$$

By Hölder inequality,

$$\begin{aligned}\log \mathbb{E}\left(\mathrm{e}^{\lambda (b_n/\sqrt{n})(G_n^{N,l} - \mathbb{E} G_n^{N,l})}\right) &\leqslant \log \mathbb{E}\left(\mathrm{e}^{C\lambda^2 (b_n^2/n) \|\nabla_\xi G_n^{N,l}\|^2}\right) \\ &\leqslant \frac{1}{4} \log \mathbb{E}\left(\mathrm{e}^{4C\lambda^2 (b_n^2/n)(I)}\right) + \frac{1}{4} \log \mathbb{E}\left(\mathrm{e}^{4C\lambda^2 (b_n^2/n)(II)}\right) \\ &\quad + \frac{1}{4} \log \mathbb{E}\left(\mathrm{e}^{4C\lambda^2 (b_n^2/n)(III)}\right) + \frac{1}{4} \log \mathbb{E}\left(\mathrm{e}^{4C\lambda^2 (b_n^2/n)(IV)}\right). \end{aligned} \tag{5.4}$$

Let us deal with the first term of this inequality. Using the definition of $\hat{r}_l$ given in (2.5) and the fact that the spectral density of $(X_k - X_k^N)$ is $|g - g^N|^2$, we rewrite the expression of (I) as

$$\begin{aligned}(I) &= 4 \sum_{i\in\mathbb{Z}} \sum_{k,k'=1}^{n} (a_{i-k} - a_{i-k}^N)(a_{i-k'} - a_{i-k'}^N) X_{k+l} X_{k'+l} = 4 \sum_{k,k'=1}^{n} \hat{r}_{k'-k}\left(|g - g^N|^2\right) X_{k+l} X_{k'+l} \\ &= 4 \left\langle X_{\cdot+l}^{(n)}, T_n\left(|g - g^N|^2\right) X_{\cdot+l}^{(n)} \right\rangle,\end{aligned}$$

where $X_{\cdot+l}^{(n)} = (X_{l+1}, \ldots, X_{l+n})^*$. Let $\mu_1^{n,N}, \ldots, \mu_n^{n,N}$ be the eigenvalues of the matrix

$$\sqrt{T_n\left(|g - g^N|^2\right)} T_n(f) \sqrt{T_n\left(|g - g^N|^2\right)}.$$

Its operator norm is bounded from above by (using Lemma 4.4)

$$\left\| T_n(f) \right\| \cdot \left\| T_n\left(|g - g^N|^2\right) \right\| \leqslant n^{1/q} \|f\|_q n^{1/q} \left\| |g - g^N|^2 \right\|_q.$$

Since $(b_n/\sqrt{n}) n^{1/q} \to 0$ by (H2) and $f \in L^q(\mathbb{T}, \mathrm{d}\theta)$ by (H1), we have for all n sufficiently large, $32CK^2\lambda^2 b_n^2/n \max_{1\leqslant j\leqslant n} \mu_j^{n,N} < 1$. Applying the crucial Lemma 4.3, we get

$$\log \mathbb{E}\left(\mathrm{e}^{4C\lambda^2 (b_n^2/n)(I)}\right) \leqslant -\frac{1}{2} \sum_{j=1}^{n} \log\left(1 - 32CK^2\lambda^2 \frac{b_n^2}{n} \mu_j^{n,N}\right). \tag{5.5}$$

Similarly, for all n sufficiently large such that $32CK^2\lambda^2(b_n^2/n) \max\limits_{1\leqslant j\leqslant n} \mu_j^{n,N} < 1$, we have

$$\log \mathbb{E}\left(\mathrm{e}^{4C\lambda^2 (b_n^2/n)(II)}\right) = \log \mathbb{E}\, \mathrm{e}^{16C(b_n^2/n)\lambda^2 \langle X_\cdot^{(n)}, T_n(|g-g^N|^2) X_\cdot^{(n)} \rangle} \leqslant -\frac{1}{2} \sum_{j=1}^{n} \log\left(1 - 32CK^2\lambda^2 \frac{b_n^2}{n} \mu_j^{n,N}\right). \tag{5.6}$$

Let us deal with the third term. We rewrite the expression of (III) as

$$(III) = 4 \sum_{i\in\mathbb{Z}} \sum_{k,k'=1}^{n} a_{i-k}^N a_{i-k'}^N (X_{k+l} - X_{k+l}^N)(X_{k'+l} - X_{k'+l}^N)$$

$$= 4 \sum_{k,k'=1}^{n} \hat{r}_{k'-k}\big(|g^N|^2\big)(X_{k+l} - X^N_{k+l})(X_{k'+l} - X^N_{k'+l})$$

$$= 4\big\langle X^{(n)}_{\cdot+l} - (X^N)^{(n)}_{\cdot+l}, T_n\big(|g^N|^2\big)\big(X^{(n)}_{\cdot+l} - (X^N)^{(n)}_{\cdot+l}\big)\big\rangle,$$

where $(X^N)^{(n)}_{\cdot+l} = (X^N_{1+l}, \dots, X^N_{n+l})^*$. Let $\nu^{n,N}_1, \dots, \nu^{n,N}_n$ the eigenvalues of the matrix

$$\sqrt{T_n\big(|g^N|^2\big)}\, T_n\big(|g - g^N|^2\big)\sqrt{T_n\big(|g^N|^2\big)}.$$

Its operator norm is bounded from above by (using Lemma 4.4)

$$\big\|T_n\big(|g^N|^2\big)\big\| \cdot \big\|T_n\big(|g - g^N|^2\big)\big\| \leqslant n^{1/q}\big\||g^N|^2\big\|_q n^{1/q}\big\||g - g^N|^2\big\|_q.$$

By our assumptions (H1) and (H2) on b_n and f, we have for all n sufficiently large, $32CK^2\lambda^2(b_n^2/n)\max_{1\leqslant j\leqslant n}\nu^{n,N}_j < 1$. Applying the crucial Lemma 4.3, we get

$$\log\mathbb{E}\big(e^{4C\lambda^2(b_n^2/n)(III)}\big) \leqslant -\frac{1}{2}\sum_{j=1}^{n}\log\left(1 - 32CK^2\lambda^2\frac{b_n^2}{n}\nu^{n,N}_j\right). \tag{5.7}$$

Similarly for all n sufficiently large such that $32CK^2\lambda^2(b_n^2/n)\max_{1\leqslant j\leqslant n}\nu^{n,N}_j < 1$ and we have

$$\log\mathbb{E}\big(e^{4C\lambda^2(b_n^2/n)(IV)}\big) \leqslant -\frac{1}{2}\sum_{j=1}^{n}\log\left(1 - 32CK^2\lambda^2\frac{b_n^2}{n}\nu^{n,N}_j\right). \tag{5.8}$$

By (5.4) and the previous estimations (5.5)–(5.8), we obtain

$$\log\mathbb{E}\big(e^{\lambda b_n^2(Q^l_n - Q^{N,l}_n)}\big) \leqslant -\frac{1}{4}\sum_{j=1}^{n}\left(\log\left(1 - 32CK^2\lambda^2\frac{b_n^2}{n}\mu^{n,N}_j\right) + \log\left(1 - 32CK^2\lambda^2\frac{b_n^2}{n}\nu^{n,N}_j\right)\right). \tag{5.9}$$

Notice that by the Taylor's expansion of order 1, we have for $|z| < 1$

$$\log(1 - z) = -z(1 - tz)^{-1}$$

where $t = t(z) \in [0, 1]$. This applied here to $z^{n,N}_j = 32CK^2\lambda^2(b_n^2/n)\lambda^{n,N}_j$, where $\lambda^{n,N}_j = \nu^{n,N}_j$ or $\lambda^{n,N}_j = \mu^{n,N}_j$ which satisfies $\sup_{1\leqslant j\leqslant n}|z^{n,N}_j| \to 0$ as $n \to \infty$, yields by (5.9),

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\big(e^{\lambda b_n^2(Q^l_n - Q^{N,l}_n)}\big) \leqslant 16C^2\lambda^2\lim_{n\to\infty}\left(\frac{1}{n}\sum_{j=1}^{n}(\mu^{n,N}_j + \nu^{n,N}_j)\right).$$

Thanks to Lemma 4.5, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mu^{n,N}_j = \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}\big(T_n(f)T_n\big(|g - g^N|^2\big)\big) = \hat{r}_0\big(|g - g^N|^2 f\big).$$

Similarly

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\nu^{n,N}_j = \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}\big(T_n\big(|g^N|^2\big)T_n\big(|g - g^N|^2\big)\big) = \hat{r}_0\big(|g^N|^2|g - g^N|^2\big).$$

So we get

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\big(e^{\lambda b_n^2(Q^l_n - Q^{N,l}_n)}\big) \leqslant 16C^2\lambda^2\big[\hat{r}_0\big(|g - g_N|^2 f\big) + \hat{r}_0\big(|g - g^N|^2|g_N|^2\big)\big]$$

where the desired negligibility (5.2) follows.

Step 3. Now we establish (2.3), i.e., for all $\lambda \in \mathbb{R}^{m+1}$,

$$\lim_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\,e^{b_n^2\langle\lambda, Q_n\rangle} = \frac{1}{2}\langle\lambda, \Sigma^2\lambda\rangle. \tag{5.10}$$

At first we have $\Sigma^{2,N}_{k,l} \to \Sigma^2_{k,l}$ for all k,l as N goes to infinity, by (H1) and the expression (2.4). Next for any fixed $\alpha, \beta > 1$ with $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, by Hölder inequality we have that

$$\log \mathbb{E}\,\mathrm{e}^{b_n^2\langle\lambda, Q_n\rangle} \leqslant \frac{1}{\alpha}\log \mathbb{E}\,\mathrm{e}^{\alpha b_n^2\langle\lambda, Q_n^N\rangle} + \frac{1}{\beta}\log \mathbb{E}\,\mathrm{e}^{\beta b_n^2\langle\lambda, Q_n - Q_n^N\rangle}$$

for all $\lambda \in \mathbb{R}^{m+1}$. From (5.1) and (5.2) it follows that

$$\limsup_{n\to\infty} \frac{1}{b_n^2}\log \mathbb{E}\,\mathrm{e}^{b_n^2\langle\lambda, Q_n\rangle} \leqslant \frac{\alpha}{2}\langle\lambda, \Sigma^{2,N}\lambda\rangle + \delta_N$$

where $\delta_N := \limsup_{n\to\infty}(1/\beta b_n^2)\log \mathbb{E}\,\mathrm{e}^{\beta b_n^2\langle\lambda, Q_n - Q_n^N\rangle} \to 0$. Letting $N \to \infty$, we get

$$\limsup_{n\to\infty} \frac{1}{b_n^2}\log \mathbb{E}\,\mathrm{e}^{b_n^2\langle\lambda, Q_n\rangle} \leqslant \frac{\alpha}{2}\langle\lambda, \Sigma^{2}\lambda\rangle. \tag{5.11}$$

Similarly, by Hölder inequality, we have for every λ,

$$\frac{1}{b_n^2}\log \mathbb{E}\,\mathrm{e}^{\alpha^{-1}b_n^2\langle\lambda, Q_n^N\rangle} \leqslant \frac{1}{b_n^2}\left(\frac{1}{\alpha}\log \mathbb{E}\,\mathrm{e}^{b_n^2\langle\lambda, Q_n\rangle} + \frac{1}{\beta}\log \mathbb{E}\,\mathrm{e}^{(\beta b_n^2/\alpha)\langle\lambda, Q_n^N - Q_n\rangle}\right).$$

Taking first $\liminf_{n\to\infty}$ and next $\lim_{N\to\infty}$ we get from (5.1) and (5.2)

$$\frac{1}{2\alpha^2}\langle\lambda, \Sigma^2\lambda\rangle \leqslant \liminf_{n\to\infty} \frac{1}{b_n^2}\log \mathbb{E}\,\mathrm{e}^{b_n^2\langle\lambda, Q_n\rangle}. \tag{5.12}$$

Letting $\alpha \to 1$ in (5.11) and (5.12) yields (5.10). Finally the desired MDP follows from (5.10) by Ellis–Gärtner's theorem ([10], Section 2.3).

5.2. Proof of Theorem 2.3

We begin with the following

Lemma 5.1. *Under the hypothesis Theorem* 2.3*, we have that for all $h \in L^{p'}(\mathbb{T}, \mathrm{d}\theta)$,*

$$\Lambda(h) := \limsup_{n\to\infty} \frac{1}{b_n^2}\log \mathbb{E}\left(\mathrm{e}^{b_n^2(1/2\pi)\int_{\mathbb{T}} h(\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta}\right) \leqslant \frac{4CK^2}{2\pi}\int_{\mathbb{T}} f^2(t)h^2(t)\,\mathrm{d}t. \tag{5.13}$$

*In particular $\mathbb{P}(\mathcal{L}_n \in \cdot)$ is exponentially *-tight in $(L^p(\mathbb{T}, \mathrm{d}\theta), \sigma(L^p(\mathbb{T}, \mathrm{d}\theta), L^{p'}(\mathbb{T}, \mathrm{d}\theta)))$, where $1/p' + 1/p = 1$.*

Proof. The last claim follows from (5.13) by [23, Chapter 2, Proposition 2.5] when $1 < p < 2$ and by [23, Chapter 2, Theorem 2.1] when $p = 1$. So it is enough to prove (5.13). For every function $h \in L^{p'}(\mathbb{T}, \mathrm{d}\theta)$, the function $\tilde{h}(\theta) = \frac{1}{2}[h(\theta) + h(-\theta)]$ is even and

$$\frac{1}{2\pi}\int_{\mathbb{T}} h(\theta)\mathcal{I}_n(\theta)\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{\mathbb{T}} \tilde{h}(\theta)\mathcal{I}_n(\theta)\,\mathrm{d}\theta,$$

we shall hence restrict ourselves to the case where h is even. Since

$$\frac{1}{2\pi}\int_{\mathbb{T}} h(\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta = \frac{1}{b_n\sqrt{n}}\left(\langle X^{(n)}_{\cdot}, T_n(h)X^{(n)}_{\cdot}\rangle - \mathbb{E}\langle X^{(n)}_{\cdot}, T_n(h)X^{(n)}_{\cdot}\rangle\right).$$

Applying (2.1) to $H((\xi_l)_{l\in\mathbb{Z}}) = \langle X^{(n)}_{\cdot}, T_n(h)X^{(n)}_{\cdot}\rangle$, we have

$$\mathbb{E}\left(\mathrm{e}^{b_n^2(1/(2\pi))\int_{\mathbb{T}} h(\theta)\mathcal{L}_n(\mathrm{d}\theta)}\right) = \mathbb{E}\left(\mathrm{e}^{(b_n/\sqrt{n})(H - \mathbb{E}H)}\right) \leqslant \mathbb{E}\left(\mathrm{e}^{(b_n^2/n)C|\nabla_\xi H|^2}\right).$$

Since $T_n(h)$ is symmetric, we have

$$\partial_{\xi_i} H\left((\xi_l)_{l\in\mathbb{Z}}\right) = \sum_{l,k=1}^{n} a_{i-k}X_l T_n(h)_{k,l} + a_{i-l}X_k T_n(h)_{k,l} = 2\sum_{l,k=1}^{n} a_{i-k}X_l T_n(h)_{k,l}.$$

Clearly

$$|\nabla_\xi H|^2 = \sum_{i\in\mathbb{Z}} (\partial_{\xi_i} H)^2 = \sum_{i\in\mathbb{Z}} \left(2\sum_{l,k=1}^{n} a_{i-k} X_l T_n(h)_{k,l}\right)^2 = 4\sum_{l,l'=1}^{n}\left(\sum_{k,k'=1}^{n} T_n(h)_{k,l} T_n(f)_{k,k'} T_n(h)_{k',l'}\right) X_l X_{l'}$$

$$= 4\sum_{l,l'=1}^{n} \big(T_n(h)T_n(f)T_n(h)\big)_{l,l'} X_l X_{l'} = 4\big\langle X_\cdot^{(n)}, T_n(h)T_n(f)T_n(h) X_\cdot^{(n)}\big\rangle.$$

Let $\alpha_1^n, \ldots, \alpha_n^n$ the eigenvalues of the matrix

$$\sqrt{T_n(h)T_n(f)T_n(h)}\, T_n(f) \sqrt{T_n(h)T_n(f)T_n(h)}.$$

Its operator norm is bounded from above by (using Lemma 4.4)

$$\|T_n(f)\| \cdot \|T_n(h)T_n(f)T_n(h)\| \leqslant \big(n^{1/q}\|f\|_q\big)^2 \big(n^{1/p'}\|h\|_{p'}\big)^2.$$

Since $b_n n^{1/q+1/p'-1/2} \to 0$, $f \in L^q(\mathbb{T}, \mathrm{d}\theta)$ and $h \in L^{p'}(\mathbb{T}, \mathrm{d}\theta)$, we have $8CK^2(b_n^2/n)\max_{1\leqslant j\leqslant n}\alpha_j^n < 1$ for n large enough. Applying Lemma 4.3, we get

$$\log \mathbb{E}\big(\mathrm{e}^{b_n^2(1/(2\pi))\int_{\mathbb{T}} h(\theta)\mathcal{L}_n(\mathrm{d}\theta)}\big) \leqslant -\frac{1}{2}\sum_{j=1}^{n} \log\left(1 - 8CK^2\frac{b_n^2}{n}\alpha_j^n\right).$$

Thus

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\big(\mathrm{e}^{b_n^2(1/(2\pi))\int_{\mathbb{T}} h(\theta)\mathcal{L}_n(\mathrm{d}\theta)}\big) \leqslant 4CK^2 \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n}\alpha_j^n.$$

Since $f \in L^q(\mathbb{T}, \mathrm{d}\theta)$ and $h \in L^{p'}(\mathbb{T}, \mathrm{d}\theta)$ with $1/p' + 1/q \leqslant \frac{1}{2}$, applying Lemma 4.5, we obtain

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{j=1}^{n}\alpha_j^n = \lim_{n\to+\infty}\frac{1}{n}\operatorname{tr}\big(\big(T_n(f)T_n(h)\big)^2\big) = \hat{r}_0(f^2h^2) < +\infty.$$

Hence (5.13) follows. □

We now turn to the

Proof of Theorem 2.3. *Step 1*. By Lemma 4.3, $\mathbb{E}\,\mathrm{e}^{\lambda|X_0|^2} < +\infty$ for some $\lambda > 0$. Then by Chebychev inequality

$$\left(\frac{1}{b_n\sqrt{n}}\sum_{k=n-\ell+1}^{n}(X_kX_{k+\ell} - \mathbb{E}X_kX_{k+\ell})\right)_{0\leqslant\ell\leqslant m}$$

is negligible with respect to the MDP. Using Theorem 2.1, we get the finite dimensional MDP on $\mathbb{R}^{m+1}$ of

$$\left(\frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n-\ell}(X_kX_{k+\ell} - \mathbb{E}X_kX_{k+\ell})\right)_{0\leqslant\ell\leqslant m}$$

with the rate function given by

$$I(z) = \sup_{\lambda\in\mathbb{R}^{m+1}}\left\{\langle\lambda, z\rangle - \frac{1}{2}\langle\lambda, \Sigma^2\lambda\rangle\right\},$$

where

$$\langle\lambda, \Sigma^2\lambda\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} 2\left(\sum_{k=0}^{m}\lambda_k\cos(k\theta)\right)^2 f^2(\theta)\,\mathrm{d}\theta + \kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}}\left(\sum_{k=0}^{m}\lambda_k\cos(k\theta)\right) f(\theta)\,\mathrm{d}\theta\right)^2.$$

Now notice that $\mathcal{L}_n(\theta)$ is even and

$$\hat{\mathcal{L}}_n(\ell) := \frac{1}{2\pi}\int_{\mathbb{T}} \cos(\ell\theta)\mathcal{L}_n(\theta)\,\mathrm{d}\theta = \frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n-\ell}(X_kX_{k+\ell} - \mathbb{E}X_kX_{k+\ell}), \quad \ell \geqslant 0.$$

Thus $(\hat{\mathcal{L}}_n(\ell))_{0\leqslant\ell\leqslant m}$ satisfies the MDP on $\mathbb{R}^{m+1}$ with the same rate function. By Lemma 4.3 and the projective limit Theorem [10, Theorem 4.6.9], we deduce that $(\mathcal{L}_n)_{n\geqslant 0}$ satisfies the MDP on $(L^p(\mathbb{T},\mathrm{d}\theta), \sigma(L^p(\mathbb{T},\mathrm{d}\theta), L^{p'}(\mathbb{T},\mathrm{d}\theta)))$ with the rate function given by

$$I(\eta) = \begin{cases} \sup\limits_{m\geqslant 0}\sup\limits_{\lambda_0,\ldots,\lambda_m\in\mathbb{R}}\left\{\frac{1}{2\pi}\int_{\mathbb{T}}\left(\sum_{k=0}^m \lambda_k\cos(k\theta)\right)\eta(\theta)\,\mathrm{d}\theta - \frac{1}{2}\Lambda\left(\sum_{k=0}^m\lambda_k\cos(k\theta)\right)\right\}, & \text{if } \eta \text{ is even,} \\ +\infty, \quad \text{otherwise} \end{cases} \tag{5.14}$$

where

$$\Lambda\left(\sum_{k=0}^m \lambda_k\cos(k\theta)\right) = \langle\lambda, \Sigma^2\lambda\rangle.$$

Step 2. Identification of the rate function. Introduce $L^p_{\text{even}}(\mathbb{T},\nu) = \{h \in L^p(\mathbb{T},\nu),\ h \text{ even}\}$. Remark as trigonometric polynomials are dense in $L^2(\mathbb{T}, f^2\,\mathrm{d}\theta)$, one can find for every $h \in L^2_{\text{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta)$, an approximation by some cosine polynomials sequence h_n, such that

$$\lim_{n\to\infty}\int_{\mathbb{T}}(h_n - h)^2(\theta)f^2(\theta)\,\mathrm{d}\theta = 0. \tag{5.15}$$

So we can extend continuously the definition of Λ to all functions $h \in L^2_{\text{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta)$,

$$\Lambda(h) = \frac{1}{2\pi}\int_{\mathbb{T}} 2h^2(\theta)f^2(\theta)\,\mathrm{d}\theta + \kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}} h(\theta)f(\theta)\,\mathrm{d}\theta\right)^2. \tag{5.16}$$

(a) Suppose that η is even, $\eta\,\mathrm{d}\theta$ is absolutely continuous w.r.t. $f^2\,\mathrm{d}\theta$, and $\eta/f \in L^2(\mathbb{T},\mathrm{d}\theta)$. For any $h \in L^2_{\text{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta)$, let h_n the sequence defined in (5.15), by Cauchy–Schwartz inequality, we get

$$\left(\int_{\mathbb{T}}\left|(h_n - h)(\theta)\eta(\theta)\right|\mathrm{d}\theta\right)^2 \leqslant \int_{\mathbb{T}}\left|h_n(\theta) - h(\theta)\right|^2 f^2(\theta)\,\mathrm{d}\theta\int_{\mathbb{T}}\left(\frac{\eta}{f}\right)^2(\theta)\,\mathrm{d}\theta \underset{n\to\infty}{\longrightarrow} 0.$$

So $I(\eta)$ defined in (5.14) coincides with

$$I(\eta) = \sup_{h\in L^2_{\text{even}}(\mathbb{T},f^2\,\mathrm{d}\theta)}\left\{\frac{1}{2\pi}\int_{\mathbb{T}} h(\theta)\eta(\theta)\,\mathrm{d}\theta - \frac{1}{2}\Lambda(h)\right\} := \sup_{h\in L^2_{\text{even}}(\mathbb{T},f^2\,\mathrm{d}\theta)} D(h).$$

Let us find explicitly the maximizer h_0 of $D(h)$. Let $k \in L^2_{\text{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta)$ and $\epsilon > 0$,

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{D(h+\epsilon k) - D(h)}{\epsilon} = \frac{1}{2\pi}\int_{\mathbb{T}} k(\theta)\eta(\theta)\,\mathrm{d}\theta - \frac{1}{2}\Bigg(&\frac{2}{2\pi}\int_{\mathbb{T}} 2f^2(\theta)h(\theta)k(\theta)\,\mathrm{d}\theta \\ &+ 2\kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(\theta)h(\theta)\,\mathrm{d}\theta\right)\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(\theta)k(\theta)\,\mathrm{d}\theta\right)\Bigg).\end{aligned}$$

So

$$\lim_{\epsilon\to 0}\frac{D(h+\epsilon k) - D(h)}{\epsilon} = 0, \quad \forall k \in L^2_{\text{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta) \tag{5.17}$$

iff

$$\eta(\theta) = 2f(\theta)^2 h(\theta) + \kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(\theta)h(\theta)\,\mathrm{d}\theta\right) f(\theta). \tag{5.18}$$

Dividing (5.18) by f and integrating over $\mathbb{T}$, we obtain

$$\int_{\mathbb{T}} f(\theta)h(\theta)\,\mathrm{d}\theta = \frac{1}{2+\kappa_4}\int_{\mathbb{T}} \frac{\eta(\theta)}{f(\theta)}\,\mathrm{d}\theta.$$

Plugging this last expression in (5.18), it is then easy to verify that the only functional $h_0 \in L^2_{\mathrm{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta)$ realizing (5.17) is given by

$$h_0(\theta)f(\theta) = \frac{\eta(\theta)}{2f(\theta)} - \frac{\kappa_4}{2+\kappa_4}\left(\frac{1}{2\pi}\int_{\mathbb{T}} \frac{\eta(u)}{2f(u)}\,\mathrm{d}u\right).$$

Calculating $D(h_0)$ gives finally the announced rate function.

(b) Now we have to treat the case where $\eta\,\mathrm{d}\theta$ is absolutely continuous w.r.t. $f^2\,\mathrm{d}\theta$ but $\frac{\eta}{f} \notin L^2(\mathbb{T}, \mathrm{d}\theta)$. So there exists $g \in L^2_{\mathrm{even}}(\mathbb{T}, \mathrm{d}\theta)$ such that $\int_{\mathbb{T}} g(\theta)\frac{\eta}{f}(\theta)\,\mathrm{d}\theta = +\infty$, and $g\frac{\eta}{f} \geqslant 0$. Let $h := \frac{g}{f}$, so $h \in L^2_{\mathrm{even}}(\mathbb{T}, f^2\,\mathrm{d}\theta)$, we choose $h_n = (h \vee (-n)) \wedge n$. We get by dominated convergence

$$\lim_{n\to\infty}\int_{\mathbb{T}} \big(h_n(\theta) - h(\theta)\big)^2 f(\theta)^2\,\mathrm{d}\theta = 0,$$

so it follows that

$$\lim_{n\to+\infty} \Lambda(h_n) = \Lambda(h).$$

By Fatou's lemma we get

$$\liminf_{n\to\infty}\int_{\mathbb{T}} h_n(\theta)\eta(\theta)\,\mathrm{d}\theta \geqslant \int_{\mathbb{T}} \liminf_{n\to\infty} h_n(\theta)\eta(\theta)\,\mathrm{d}\theta = +\infty.$$

Since by approximation,

$$I(\eta) \geqslant \frac{1}{2\pi}\int_{\mathbb{T}} h_n(\theta)\eta(\theta)\,\mathrm{d}\theta - \frac{1}{2}\Lambda(h_n),$$

letting n to ∞, we obtain $I(\eta) = \infty$.

(c) Now we have to treat the case where $\eta\,\mathrm{d}\theta$ is not absolutely continuous w.r.t. $f^2\,\mathrm{d}\theta$, i.e. there exists a measurable and symmetric set $K \subset \mathbb{T}$ such that $\int_K f^2(\theta)\,\mathrm{d}\theta = 0$ while $\int_K \eta(\theta)\,\mathrm{d}\theta > 0$. For any $t > 0$, we approximate the function $t1_K$ by a sequence of cosine polynomials h_n in $L^2(\mathbb{T}, (f^2 + |\eta|)\,\mathrm{d}\theta)$ and get

$$I(\eta) \geqslant \lim_{n\to+\infty} D(h_n) \geqslant t\int_K \eta(\theta)\,\mathrm{d}\theta.$$

Letting t to infinity, we get $I(\eta) = +\infty$. □

5.3. Proof of the corollaries of Theorem 2.3

Proof of Corollary 2.4. It is enough to prove it for h even. When h is a cosine polynomial, this was established in the proof of Theorem 2.3. For general $h \in L^{p'}(\mathbb{T}, \mathrm{d}\theta)$, let (h_N) be a sequence of cosine polynomials such that $h_N \to h$ in $L^{p'}(\mathbb{T}, \mathrm{d}\theta)$. To get the desired result, it remains to show

$$\lim_{N\to\infty}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\exp\left(\lambda b_n^2\frac{1}{2\pi}\int_{\mathbb{T}}(h_N - h)\mathcal{L}_n(\theta)\,\mathrm{d}\theta\right) = 0, \quad \forall\lambda \in \mathbb{R}.$$

This follows by Lemma 5.1. □

Proof of Corollary 2.5. To deduce MDP for $(\tilde{\mathcal{L}}_n)$ from Theorem 2.3 (with $q=+\infty$), we only need to prove that for all $h \in L^{p'}(\mathbb{T})$,

$$\frac{\sqrt{n}}{b_n}\left(\int_{\mathbb{T}} h(t)\mathbb{E}\mathcal{I}_n(t)\,\mathrm{d}t - \int_{\mathbb{T}} f(t)h(t)\,\mathrm{d}t\right) \xrightarrow[n\to 0]{} 0. \tag{5.19}$$

It is easy to see that

$$\int_{\mathbb{T}} h(t)\mathbb{E}\mathcal{I}_n(t)\,\mathrm{d}t = \int_{\mathbb{T}}\int_{\mathbb{T}} K_n(u-t)f(u)h(t)\,\mathrm{d}t\,\mathrm{d}u = \langle K_n * f, h\rangle$$

where K_n is the Fejèr kernel function given in Lemma 4.6. Since the function K_n is even, we have

$$\int_{\mathbb{T}} h(t)\mathbb{E}\mathcal{I}_n(t)\,\mathrm{d}t = \frac{1}{2}\int_{\mathbb{T}}\int_{\mathbb{T}} K_n(t)f(u+t)h(t)\,\mathrm{d}t\,\mathrm{d}u + \frac{1}{2}\int_{\mathbb{T}}\int_{\mathbb{T}} K_n(t)f(u)h(t+u)\,\mathrm{d}t\,\mathrm{d}u.$$

Taking into account the equalities $\int_{\mathbb{T}} f(u)h(u)\,\mathrm{d}u = \int_{\mathbb{T}} f(u+t)h(u+t)\,\mathrm{d}u$, $\int_{\mathbb{T}} K_n(t)\,\mathrm{d}t = 1$ we get

$$\begin{aligned}\left|\int_{\mathbb{T}} h(t)\mathbb{E}\mathcal{I}_n(t)\,\mathrm{d}t - \int_{\mathbb{T}} f(t)h(t)\,\mathrm{d}t\right| &= \frac{1}{2}\left|\int_{\mathbb{T}} K_n(t)\int_{\mathbb{T}}\big(f(u)-f(u+t)\big)\big(h(t+u)-h(u)\big)\,\mathrm{d}u\,\mathrm{d}t\right| \\ &\leqslant \frac{1}{2}\int_{\mathbb{T}} K_n(t)\big\|f(\cdot)-f(\cdot+t)\big\|_p\big\|h(t+\cdot)-h(\cdot)\big\|_{p'}\,\mathrm{d}t.\end{aligned}$$

By our assumption (2.7) on f, for $\delta>0$ small and $|t|\leqslant\delta$ we have

$$\big\|f(\cdot)-f(\cdot+t)\big\|_p \leqslant C\sqrt{|t|} \quad \text{and} \quad \big\|h(t+\cdot)-h(\cdot)\big\|_{p'} \leqslant 2\|h\|_{p'}.$$

By Lemma 4.6, the last quantity above is smaller than

$$C\|h\|_{p'}\int_{|t|\leqslant\delta} K_n(t)\sqrt{|t|}\,\mathrm{d}u + 2\|f\|_p\|h\|_{p'}\int_{|t|\geqslant\delta} K_n(t)\,\mathrm{d}t = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right).$$

Hence (5.19) follows. □

Proof of Corollary 2.6. By Corollary 2.4 with $p=1$, we only need to prove (5.19) for all h satisfies (2.8), and the proof of (5.19) is completely similar to that of Corollary 2.5. □

5.4. Proof of Theorem 2.7

Let us describe briefly how the preceding proof of Theorem 2.1 can be easily extended to the general non-linear functional F. We only consider $F(x_0,\ldots,x_l)=F(x_0)$ and it is real-valued (for simplicity).

Since F' is Lipschitz continuous, we get for some positive L, and for all N

$$\big|F(X_k^N)\big| \leqslant L\big(1+|X_k^N|^2\big) \leqslant 2L(N+1)\left(1+\sum_{j=-N}^{N} a_j^2\xi_{k+j}^2\right)$$

so that, setting $\delta' = \delta/(2L(N+1)^2\sup_j a_j^2)$ where δ is given in (2.2), by the assumption on the validity of the LSI, we get

$$\mathbb{E}\big(\mathrm{e}^{\delta'|F(X_k^N)|}\big) \leqslant \mathrm{e}^{\delta' L(N+1)}\mathbb{E}\big(\mathrm{e}^{\delta\xi_0^2}\big) < \infty.$$

Hence for every N fixed, by Lemma 4.7, $(1/n)\sum_{k=1}^n F(X_k^N)$ satisfies the MDP as in Step 1 in the proof of Theorem 2.1.

Thus by the argument in Step 3 of Theorem 2.1, it remains to prove that $\forall\lambda \in \mathbb{R}$,

$$\limsup_{N\to\infty}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\exp\left(\lambda\frac{b_n}{\sqrt{n}}\sum_{k=1}^{n}\left[F(X_k)-F(X_k^N)-\mathbb{E}\left(F(X_k)-F(X_k^N)\right)\right]\right)=0.$$

We apply again (5.3) to

$$G_n^N\left((\xi_l)_{l\in\mathbb{Z}}\right)=\sum_{k=1}^{n}\left(F(X_k)-F(X_k^N)\right).$$

Writing $F'(X.) := (F'(X_1), \ldots, F'(X_n))^*$ and similarly $F'(X.)$, we have

$$\begin{aligned}|\nabla_\xi G_n^N|^2 &= \sum_{i\in\mathbb{Z}}\left(\sum_{k=1}^{n}a_{i-k}F'(X_k)-a_{i-k}^N F'(X_k^N)\right)^2\\ &\leqslant 2\sum_{i\in\mathbb{Z}}\left(\sum_{k=1}^{n}(a_{i-k}-a_{i-k}^N)F'(X_k)\right)^2+2\sum_{i\in\mathbb{Z}}\left(\sum_{k=1}^{n}a_{i-k}^N\left(F'(X_k^N)-F'(X_k)\right)\right)^2\\ &= 2\left|\sqrt{T_n\left(|g-g^N|^2\right)}F'(X.)\right|^2+2\left|\sqrt{T_n\left(|g^N|^2\right)}\left(F'(X.)-F'(X^N_.)\right)\right|^2.\end{aligned}$$

By the fact that the derivative of F is Lipschitz and the spectral density is bounded, the last term above is bounded by

$$2L\|g-g^N\|_\infty^2\left(n+\sum_{k=1}^{n}X_k^2\right)+2\|g^N\|_\infty^2\sum_{k=1}^{n}(X_k^N-X_k)^2.$$

Finally as $\lambda^2(b_n^2/n)\|g-g^N\|_\infty^2$ can be chosen arbitrary small for large n, we have by Lemma 4.3

$$\begin{aligned}\frac{1}{b_n^2}\log\mathbb{E}\exp\left(\lambda\frac{b_n}{\sqrt{n}}(G_n^N-\mathbb{E}G_n^N)\right)&\leqslant LC\lambda^2\|g-g^N\|_\infty^2-\frac{n}{4b_n^2}\log\left(1-4CLK^2\lambda^2\frac{b_n^2}{n}\|g-g^N\|_\infty^2\|g\|_\infty^2\right)\\ &\quad-\frac{n}{4b_n^2}\log\left(1-4CLK^2\lambda^2\frac{b_n^2}{n}\|g^N\|_\infty^2\|g-g^N\|_\infty^2\right)\end{aligned}$$

and the r.h.s. of this last inequality is easily seen to behave as $n \to \infty$ as

$$\|g-g^N\|_\infty^2\left(LC\lambda^2+2CLK^2\lambda^2\|g\|_\infty^2\right).$$

By the famous Fejèr Theorem (Lemma 4.6), under the assumption of continuity of g, we get that

$$\lim_{N\to\infty}\|g-g^N\|_\infty^2=0,$$

which yields to the desired negligibility.

5.5. *Proof of Corollary 2.8*

Under assumption (2.2), the crucial inequality (5.3), as a consequence of the LSI, may not be used. However, we may encompass this difficulty by noting that integrability (2.2) is, by Djellout and al. [13, Theorem 2.3], equivalent to a Transport inequality in L_1-Wasserstein distance which is itself equivalent to the inequality (5.3) with the Lipschitz norm instead of the gradient in the right hand side, but for this particular linear case, the gradient and Lipschitz norm are equal so that the same proof works.

5.6. *Proof of statistical results in Section 3*

5.6.1. *Proof of Proposition 3.1*

Considering X_n/σ if necessary, we can assume without loss of generality that $\sigma = 1$. Let us introduce

$$r_n := \frac{\sqrt{n}}{b_n}(\hat{\theta}_n-\theta)\quad\text{and}\quad R_n=\frac{1-\theta^2}{\sqrt{n}b_n}\sum_{i=1}^{n}(X_iX_{i-1}-\theta X_{i-1}^2).$$

By Theorem 2.1, R_n satisfies the MDP. Before identifying its rate function let us first show that $r_n - R_n$ is negligible w.r.t. the MDP. To that end, note

$$r_n = \frac{\sqrt{n}}{b_n}\frac{\sum_{i=1}^n (X_i X_{i-1} - \theta X_{i-1}^2)}{\sum_{i=1}^n X_{i-1}^2} = \frac{1}{\sqrt{n}b_n}\sum_{i=1}^n (X_i X_{i-1} - \theta X_{i-1}^2) \times \frac{n}{\sum_{i=1}^n X_{i-1}^2}.$$

So

$$r_n - R_n = \frac{1}{\sqrt{n}b_n}\sum_{i=1}^n (X_i X_{i-1} - \theta X_{i-1}^2) \times \frac{\frac{1}{n}\sum_{i=1}^n X_{i-1}^2 - (1-\theta^2)^{-1}}{\frac{1}{n}\sum_{i=1}^n X_{i-1}^2} \times (1-\theta^2).$$

For $\epsilon > 0,\ L > 0$ and $\delta > 0$, we have

$$\begin{aligned}\mathbb{P}\big(|r_n - R_n| \geqslant \epsilon\big) = {} & \mathbb{P}\left(\left|\frac{\sqrt{n}}{b_n}\sum_{i=1}^n (X_i X_{i-1} - \theta X_{i-1}^2)\right| \geqslant \frac{1}{1-\theta^2} L\sqrt{\delta\epsilon}\right) \\ & + \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_{i-1}^2 - \frac{1}{1-\theta^2}\right| \geqslant \frac{\sqrt{\delta\epsilon}}{L}\right) + \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_{i-1}^2 < \delta\right).\end{aligned}$$

For $\delta,\ \varepsilon$ sufficiently small but fixed, the two last terms are bounded by (for n large enough)

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}b_n}\sum_{i=1}^n \left(X_{i-1}^2 - \frac{1}{1-\theta^2}\right)\right| \geqslant \frac{\sqrt{n}}{b_n}\right)$$

which is clearly negligible as $n \to +\infty$ by the MDP of

$$\frac{1}{\sqrt{n}b_n}\sum_{i=1}^n \left(X_{i-1}^2 - \frac{1}{1-\theta^2}\right).$$

The first one is negligible by the MDP of R_n by letting $L \to \infty$. So r_n satisfies the same MDP as R_n. It remains to identify the rate function governing the MDP of R_n. By Theorem 2.1, the rate function governing the MDP of R_n is given by

$$I(x) = \frac{x^2}{2(1-\theta^2)^2 A^2} \tag{5.20}$$

with $A^2 := \theta^2\Sigma_{00}^2 - 2\theta\Sigma_{01}^2 + \Sigma_{11}^2$, where

$$\begin{aligned}\Sigma_{00}^2 &= \frac{1}{2\pi}\int_{\mathbb{T}} 2f^2(u)\,\mathrm{d}u + \kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(u)\,\mathrm{d}u\right)^2, \\ \Sigma_{01}^2 &= \frac{1}{2\pi}\int_{\mathbb{T}} 2\cos(u)f^2(u)\,\mathrm{d}u + \kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(u)\,\mathrm{d}u\right)\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(u)\cos(u)\,\mathrm{d}u\right), \\ \Sigma_{11}^2 &= \frac{1}{2\pi}\int_{\mathbb{T}} (1+\cos 2u)f^2(u)\,\mathrm{d}u + \kappa_4\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(u)\cos(u)\,\mathrm{d}u\right)^2.\end{aligned}$$

Hence

$$A^2 = 2\theta^2\hat{r}_0(f^2) - 2\theta\hat{r}_1(f^2) + \hat{r}_0(f^2) + \hat{r}_2(f^2) + \kappa_4\left(\frac{\theta}{2\pi}\int_{\mathbb{T}} f(u)\,\mathrm{d}u - \frac{1}{2\pi}\int_{\mathbb{T}} f(u)\cos u\,\mathrm{d}u\right)^2$$

where $\hat{r}_k(f^2) = \frac{1}{2\pi}\int_{\mathbb{T}} f^2(u)\,\mathrm{e}^{-\mathrm{i}ku}\,\mathrm{d}u = \frac{1}{2\pi}\int_{\mathbb{T}} f^2(u)\cos(ku)\,\mathrm{d}u$ is the kth Fourier coefficient. The last term with coefficient κ_4 is zero for

$$\frac{\theta}{2\pi}\int_{\mathbb{T}} f(u)\,\mathrm{d}u - \frac{1}{2\pi}\int_{\mathbb{T}} f(u)\cos u\,\mathrm{d}u = \theta\,\mathrm{Var}(X_0) - \mathrm{Cov}(X_0, X_1) = 0.$$

Furthermore (recalling that $\sigma = 1$ and $\mathbb{E}X_0X_n = \theta^n(1-\theta^2)^{-1}$ for $n \geqslant 0$),

$$f^2(u) = \frac{1}{(1-\theta^2)^2}\left(\sum_{n\in\mathbb{Z}}\theta^{|n|}\,\mathrm{e}^{\mathrm{i}nu}\right)^2 = \frac{1}{(1-\theta^2)^2}\sum_{k\in\mathbb{Z}}\mathrm{e}^{\mathrm{i}ku}\sum_{n\in\mathbb{Z}}\theta^{|n|+|k-n|}$$

where it follows that $\hat{r}_k(f^2) = \frac{1}{(1-\theta^2)^2}\sum_{n\in\mathbb{Z}}\theta^{|n|+|k-n|}$. From this last relation we deduce easily

$$\hat{r}_0(f^2) = \frac{1+\theta^2}{(1-\theta^2)^3}, \quad \hat{r}_1(f^2) = \frac{2\theta}{(1-\theta^2)^3}, \quad \hat{r}_2(f^2) = \frac{3\theta^2-\theta^4}{(1-\theta^2)^3}.$$

Substituting to the expression of A^2, we get

$$A^2 = \frac{1}{1-\theta^2}.$$

Substituting in (5.20), we obtain the claimed rate function.

5.6.2. Proof of Proposition 3.2

We need the following stronger result in the centered Gaussian case, inspired by [2]

Lemma 5.2. *Assume that* (ξ_i) *are Gaussian* $\mathcal{N}(0,1)$. *Let* $X^{(n)} = (X_1,\ldots,X_n)^*$ *and* M_n *be a* $n\times n$ *order symmetric matrix. Denote by* $(\lambda_j^n)_{1\leqslant j\leqslant n}$ *the eigenvalues (counting up to the multiplicity) of* $M_nT_n(f)$. *Assume that* $\sup_n \max_j |\lambda_j^n| < +\infty$ *and for some measurable function* m *on* $\mathbb{T}$ *such that* $fm \in L^\infty(\mathbb{T})$,

$$\frac{1}{n}\sum_{j=1}^n(\lambda_j^n)^2 \to \frac{1}{2\pi}\int_{\mathbb{T}}\big(f(\theta)m(\theta)\big)^2\,\mathrm{d}\theta. \tag{5.21}$$

Then for every moderate deviation scale $1 \ll b_n \ll \sqrt{n}$, *we have for all* λ,

$$\lim_{n\to+\infty}\frac{1}{b_n^2}\log\mathbb{E}\exp\left(\frac{\lambda b_n}{\sqrt{n}}\big(\langle X^{(n)}, M_nX^{(n)}\rangle - \mathbb{E}\langle X^{(n)}, M_nX^{(n)}\rangle\big)\right) = \frac{\lambda^2}{2\pi}\int_{\mathbb{T}}f^2(\theta)m^2(\theta)\,\mathrm{d}\theta.$$

Proof. Denote $T_n = \frac{1}{\sqrt{n}b_n}(\langle X^{(n)}, M_nX^{(n)}\rangle - \mathbb{E}\langle X^{(n)}, M_nX^{(n)}\rangle)$ we have

$$\log\mathbb{E}\big(\mathrm{e}^{b_n^2\lambda T_n}\big) = -\lambda\frac{b_n}{\sqrt{n}}\mathbb{E}\langle X^{(n)}, M_nX^{(n)}\rangle - \frac{1}{2}\sum_{j=1}^n\log\left(1-2\lambda\frac{b_n}{\sqrt{n}}\lambda_j^n\right).$$

Notice that by Taylor's Theorem for $|z|<1$

$$\log(1-z) = -z - \frac{1}{2}z^2(1-tz)^{-2},$$

where $t = t(z) \in [0,1]$. This applied here to $z_j^n = 2\lambda(b_n/\sqrt{n})\lambda_j^n$, which satisfies $\sup_{1\leqslant j\leqslant n}|z_j^n| \to 0$ as $n\to\infty$, and hence $|1 - t(z_j^n)z_j^n| \to 1$ uniformly in $1\leqslant j\leqslant n$. Since the Gaussian process (X_k) is assumed centered we have

$$\sum_{j=1}^n\lambda_j^n = \mathrm{tr}\big(M_nT_n(f)\big) = \mathbb{E}\langle X^{(n)}, M_nX^{(n)}\rangle.$$

Thus

$$\lim_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\big(\mathrm{e}^{b_n^2\lambda T_n}\big) = \lambda^2\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n(\lambda_j^n)^2.$$

The conclusion follows by our hypothesis. $\square$

Proof of Proposition 3.2. We have

$$\frac{\sqrt{n}}{b_n}\big(L_n - \mathbb{E}(L_n)\big) = \frac{1}{2\sqrt{n}b_n}\Big(\big\langle X^{(n)}, \big[T_n(f_0)^{-1} - T_n(f_1)^{-1}\big]X^{(n)}\big\rangle - \mathbb{E}\big(\big\langle X^{(n)}\big[T_n(f_0)^{-1} - T_n(f_1)^{-1}\big]X^{(n)}\big\rangle\big)\Big).$$

We want to apply Lemma 5.2 with $f = f_0$, $M_n = \frac{1}{2}(T_n(f_0)^{-1} - T_n(f_1)^{-1})$ and $m = \frac{1}{2}(f_0^{-1} - f^{-1})$. The boundness of the eigenvalues (λ_j^n) of $M_n T_n(f_0)$, $n \geqslant 1$ is given by Lemma 10 in [2], and it is proved in [2, Proof of Proposition 7] that $(1/n)\sum_{j=1}^n \delta_{\lambda_j^n}$ converges weakly to the image measure of the normalized Lebesgue measure $\mathrm{d}t/(2\pi)$ by $fm = \frac{1}{2}(1 - f_0/f_1)$ (the factor 1/2 is missed in [2]). Thus condition (5.21) is satisfied for $(\lambda_j^n)_{n,j}$ is bounded. Now the desired MDP follows by Lemma 5.2 and Ellis–Gärtner's theorem. □

Acknowledgement

We are grateful to M. Bouaziz and W. Bryc for fruitful discussions and valuable comments. We also thank the referees whose remarks and comments lead to significant improvement both in the presentation of the paper and in statistical applications.

References

[1] F. Avram, On bilinear forms in Gaussian random variables and Toeplitz matrices, Probab. Theory Related Fields 79 (1988) 37–45.
[2] B. Bercu, F. Gamboa, A. Rouault, Large deviations for quadratic forms of Gaussian stationary processes, Stochastic Process. Appl. 71 (1997) 75–90.
[3] S. Bobkov, F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, J. Funct. Anal. 163 (1999) 1–28.
[4] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Springer-Verlag, New York, 1991.
[5] W. Bryc, A. Dembo, On large deviations of empirical measures for stationary Gaussian processes, Stochastic Process. Appl. 58 (1995) 23–34.
[6] W. Bryc, A. Dembo, Large deviations for quadratic functionals of Gaussian functionals, J. Theoret. Probab. 10 (1997) 307–332.
[7] R.M. Burton, H. Dehling, Large deviations for some weakly dependent random processes, Statist. Probab. Lett. 9 (1990) 397–401.
[8] P.L. Butzer, R.J. Nessel, Fourier Analysis and Approximation, vol. I, Birkhäuser, 1971.
[9] X. Chen, Moderate deviations for m-dependent random variables with Banach space values, Statist. Probab. Lett. 35 (1997) 123–134.
[10] A. Dembo, O. Zeitouni, Large Deviations Techniques and their Applications, Jones and Bartlett, Boston, MA, 1993.
[11] J.D. Deuschel, D.W. Stroock, Large Deviations, Academic Press, Boston, 1989.
[12] H. Djellout, A. Guillin, Large deviations and moderate deviations for moving average processes, Ann. Math. Fac. Toulouse 10 (2001) 23–31.
[13] H. Djellout, A. Guillin, L. Wu, Transportation cost-information inequalities for random dynamical systems and diffusions, Ann. Probab. 32 (2004) 2702–2732.
[14] M.D. Donsker, S.R.S. Varadhan, Large deviations for stationary Gaussian processes, Commun. Math. Phys. 97 (1985) 187–210.
[15] R. Fox, M. Taqqu, Central limit theorems for quadratic forms in random variables having long-range dependence, Probab. Theory Related Fields 74 (1987) 213–240.
[16] L. Giraitis, D. Surgailis, A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate, Probab. Theory Related Fields 86 (1990) 87–104.
[17] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, New York, 1980.
[18] T. Jiang, M.B. Rao, X. Wang, Moderate deviations for some weakly dependent random processes, Statist. Probab. Lett. 15 (1992) 71–76.
[19] T. Jiang, M.B. Rao, X. Wang, Large deviations for moving average processes, Stochastic Process. Appl. 59 (1995) 309–320.
[20] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, in: Séminaire de probabilités XXXIII, in: Lecture Notes in Math., vol. 1709, Springer, 1999, pp. 120–216.
[21] M. Rosenblatt, Gaussian and Non-Gaussian Linear Time Series and Random Fields, Springer-Verlag, New York, 2000.
[22] G. Samorodnitsky, M.S. Taqqu, Stable Non-Gaussian Random Processes, Chapman and Hall, New York, 1994.
[23] L. Wu, An introduction to large deviations, in: J.A. Yan, S. Peng, S. Fang, L. Wu (Eds.), Several Topics in Stochastic Analysis, Academic Press of China, Beijing, 1997, pp. 225–336 (in Chinese).
[24] L. Wu, On large deviations for moving average processes, in: T.L. Lai, H.L. Yang, S.P. Yung (Eds.), Probability, Finance and Insurance, Proceeding of a Workshop at the University of Hong-Kong, 15–17 July 2002, World Scientific, Singapore, 2004, pp. 15–49.

ESAIM: PS 18 (2014) 308–331
DOI: 10.1051/ps/2013038

ESAIM: Probability and Statistics
www.esaim-ps.org

MODERATE DEVIATIONS FOR THE DURBIN–WATSON STATISTIC RELATED TO THE FIRST-ORDER AUTOREGRESSIVE PROCESS

S. Valère Bitseki Penda[1], Hacène Djellout[2] and Frédéric Proïa[3]

Abstract. The purpose of this paper is to investigate moderate deviations for the Durbin–Watson statistic associated with the stable first-order autoregressive process where the driven noise is also given by a first-order autoregressive process. We first establish a moderate deviation principle for both the least squares estimator of the unknown parameter of the autoregressive process as well as for the serial correlation estimator associated with the driven noise. It enables us to provide a moderate deviation principle for the Durbin–Watson statistic in the case where the driven noise is normally distributed and in the more general case where the driven noise satisfies a less restrictive Chen–Ledoux type condition.

Mathematics Subject Classification. 60F10, 60G42, 62M10, 62G05.

Received August 31, 2012. Revised February 26, 2013.

1. Introduction

This paper is focused on the stable first-order autoregressive process where the driven noise is also given by a first-order autoregressive process. The purpose is to investigate moderate deviations for both the least squares estimator of the unknown parameter of the autoregressive process as well as for the serial correlation estimator associated with the driven noise. Our goal is to establish moderate deviations for the Durbin–Watson statistic [11–13], in a lagged dependent random variables framework. First of all, we shall assume that the driven noise is normally distributed. Then, we will extend our investigation to the more general framework where the driven noise satisfies a less restrictive Chen–Ledoux type condition [5, 17]. We are inspired by the recent paper of Bercu and Proïa [3], where the almost sure convergence and the central limit theorem are established for both the least squares estimators and the Durbin–Watson statistic. Our results are proved *via* an extensive use of the results of Dembo [6], Dembo and Zeitouni [7] and Worms [24, 25] on the one hand, and of the paper of Puhalskii [21] and Djellout [8] on the other hand, about moderate deviations for martingales. In order to introduce the Durbin–Watson statistic, the first-order autoregressive process of interest is as follows, for all

Keywords and phrases. Durbin–Watson statistic, moderate deviation principle, first-order autoregressive process, serial correlation.

[1] Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais, 63177 Aubière, France. Valere.Bitsekipenda@math.univ-bpclermont.fr

[2] Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, Avenue des Landais, 63177 Aubière, France. Hacene.Djellout@math.univ-bpclermont.fr

[3] Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, and INRIA Bordeaux, team ALEA, 200 Avenue de la Vieille Tour, 33405 Talence cedex, France. Frederic.Proia@inria.fr

Article published by EDP Sciences

$n \geq 1$,

$$\begin{cases} X_n = \theta X_{n-1} + \varepsilon_n \\ \varepsilon_n = \rho \varepsilon_{n-1} + V_n \end{cases} \tag{1.1}$$

where we shall assume that the unknown parameters $|\theta| < 1$ and $|\rho| < 1$ to ensure the stability of the model.

In all the sequel, we also assume that (V_n) is a sequence of independent and identically distributed random variables with zero mean and positive variance σ^2. The square-integrable initial values X_0 and ε_0 may be arbitrarily chosen. We have decided to estimate θ by the least squares estimator

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2}. \tag{1.2}$$

Then, we also define a set of least squares residuals given, for all $1 \leq k \leq n$, by

$$\widehat{\varepsilon}_k = X_k - \widehat{\theta}_n X_{k-1}, \tag{1.3}$$

which leads to the estimator of ρ,

$$\widehat{\rho}_n = \frac{\sum_{k=1}^n \widehat{\varepsilon}_k \widehat{\varepsilon}_{k-1}}{\sum_{k=1}^n \widehat{\varepsilon}_{k-1}^2}. \tag{1.4}$$

Finally, the Durbin–Watson statistic is defined, for $n \geq 1$, as

$$\widehat{D}_n = \frac{\sum_{k=1}^n (\widehat{\varepsilon}_k - \widehat{\varepsilon}_{k-1})^2}{\sum_{k=0}^n \widehat{\varepsilon}_k^2}. \tag{1.5}$$

This well-known statistic was introduced by the pioneer work of Durbin and Watson [11–13], in the middle of last century, to test the presence of a significative first order serial correlation in the residuals of a regression analysis. A wide range of literature is available on the asymptotic behavior of the Durbin–Watson statistic, frequently used in Econometry. While it appeared to work pretty well in the classical independent framework, Malinvaud [18] and Nerlove and Wallis [19] observed that, for linear regression models containing lagged dependent random variables, the Durbin–Watson statistic may be asymptotically biased, potentially leading to inadequate conclusions. Durbin [10] proposed alternative tests to prevent this misuse, such as the *h-test* and the *t-test*, then substantial contributions were brought by Inder [15], King and Wu [16] and more recently Stocker [22]. Lately, a set of results have been established by Bercu and Proïa in [3] for the first-order autoregressive process, and by Proïa [20] for the autoregressive process of any order, in particular a test procedure as powerful as the *h-test* and more accurate than the usual *portmanteau* tests, and they will be summarized thereafter as a basis for this paper in the one-dimensional case. This work can be seen as an extension of [3] in the sense that more powerful convergences are reached and that a better precision than the central limit theorem is provided for the same random sequences. Hence, the establishment of moderate deviations is the natural continuation following the proof of central limit theorems and laws of iterated logarithm. We are now interested in the asymptotic estimation of

$$\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(\Theta_n - \Theta) \in A\right)$$

where Θ_n denotes the estimator of the unknown parameter of interest Θ, A is a given domain of deviations and (b_n) denotes the scale of deviations. When $b_n = 1$, this is exactly the estimation of the *central limit theorem* (CLT). When $b_n = \sqrt{n}$, it becomes a *large deviation principle* (LDP). And when $1 \ll b_n \ll \sqrt{n}$, this is the so-called *moderate deviation principle* (MDP). Usually, an MDP has a simpler rate function inherited from the approximated gaussian process which does not necessarily depend on the parameters under investigation and holds for a larger class of dependent random variables than the LDP. Furthermore, an MDP can be seen as a refinement of the CLT in the sense that the MDP tells us that the gaussian estimation still holds up to the scale of large deviations. For the sake of clarity, all useful definitions will be given later.

The paper is organized as follows. First of all, we recall the results recently established by Bercu and Proïa [3]. In Section 2, we propose moderate deviation principles for the estimators of θ and ρ and for the Durbin–Watson statistic, given by (1.2), (1.4) and (1.5), under the normality assumption on the driven noise. Section 3 deals with the generalization of the latter results under a less restrictive Chen–Ledoux type condition on (V_n). Finally, all technical proofs are postponed to Section 4.

Lemma 1.1. *Assume that* (V_n) *is independent and identically distributed with positive finite variance. Then, we have the almost sure convergence of the autoregressive estimator,*

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta^* \qquad \text{a.s.}$$

where the limiting value

$$\theta^* = \frac{\theta+\rho}{1+\theta\rho}\cdot \tag{1.6}$$

In addition, as soon as $\mathbb{E}[V_1^4] < \infty$, *we also have the asymptotic normality,*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\theta^2)$$

where the asymptotic variance

$$\sigma_\theta^2 = \frac{(1-\theta^2)(1-\theta\rho)(1-\rho^2)}{(1+\theta\rho)^3}\cdot \tag{1.7}$$

Lemma 1.2. *Assume that* (V_n) *is independent and identically distributed with positive finite variance. Then, we have the almost sure convergence of the serial correlation estimator,*

$$\lim_{n\to\infty} \widehat{\rho}_n = \rho^* \qquad \text{a.s.}$$

where the limiting value

$$\rho^* = \theta\rho\,\theta^*. \tag{1.8}$$

Moreover, as soon as $\mathbb{E}[V_1^4] < \infty$, *we have the asymptotic normality,*

$$\sqrt{n}\left(\widehat{\rho}_n - \rho^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\rho^2)$$

with the asymptotic variance

$$\sigma_\rho^2 = \frac{(1-\theta\rho)}{(1+\theta\rho)^3}\left((\theta+\rho)^2(1+\theta\rho)^2 + (\theta\rho)^2(1-\theta^2)(1-\rho^2)\right). \tag{1.9}$$

In addition, we have the joint asymptotic normality,

$$\sqrt{n}\begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

where the covariance matrix

$$\Gamma = \begin{pmatrix} \sigma_\theta^2 & \theta\rho\sigma_\theta^2 \\ \theta\rho\sigma_\theta^2 & \sigma_\rho^2 \end{pmatrix}. \tag{1.10}$$

Lemma 1.3. *Assume that (V_n) is independent and identically distributed with positive finite variance. Then, we have the almost sure convergence of the Durbin–Watson statistic,*

$$\lim_{n\to\infty} \widehat{D}_n = D^* \qquad a.s.$$

where the limiting value

$$D^* = 2(1-\rho^*). \tag{1.11}$$

In addition, as soon as $\mathbb{E}[V_1^4] < \infty$, we have the asymptotic normality,

$$\sqrt{n}\left(\widehat{D}_n - D^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_D^2)$$

where the asymptotic variance

$$\sigma_D^2 = 4\sigma_\rho^2. \tag{1.12}$$

Proof. The proofs of Lemma 1.1, Lemmas 1.2 and 1.3 may be found in [3]. $\square$

Our objective is now to establish a set of moderate deviation principles on these estimates in order to get a better asymptotic accuracy than the central limit theorem.

In the whole paper, for any matrix M, M' and $\|M\|$ stand for the transpose and the euclidean norm of M, respectively. In addition, for a sequence of random variables $(Z_n)_n$ on $\mathbb{R}^{d\times p}$, we say that $(Z_n)_n$ converges $(a_n)-$superexponentially fast in probability to some random variable Z with $a_n \to \infty$ if, for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}\Big(\|Z_n - Z\| > \delta\Big) = -\infty.$$

This *exponential convergence* with speed a_n will be shortened as

$$Z_n \underset{a_n}{\overset{\text{superexp}}{\longrightarrow}} Z.$$

The *exponential equivalence* with speed a_n between two sequences of random variables $(Y_n)_n$ and $(Z_n)_n$, whose precise definition is given in Definition 4.2.10 of [7], will be shortened as

$$Y_n \underset{a_n}{\overset{\text{superexp}}{\sim}} Z_n.$$

We start by recalling some useful definitions.

Definition 1.4 (Large Deviation Principle)**.** We say that a sequence of random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies an LDP with speed a_n and rate function $I : S \to \mathbb{R}^+$ if $a_n \to \infty$ and, for each $A \in \mathcal{S}$,

$$-\inf_{x\in A^o} I(x) \le \liminf_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}\Big(M_n \in A\Big) \le \limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}\Big(M_n \in A\Big) \le -\inf_{x\in \bar{A}} I(x)$$

where A^o and $\bar{A}$ denote the interior and the closure of A, respectively. The rate function I is lower semicontinuous, *i.e.* all the sub-level sets $\{x \in S \mid I(x) \le c\}$ are closed, for $c \ge 0$.

Let (b_n) be a sequence of increasing positive numbers satisfying $1 = o(b_n^2)$ and $b_n^2 = o(n)$,

$$b_n \longrightarrow \infty, \qquad\qquad \frac{b_n}{\sqrt{n}} \longrightarrow 0. \tag{1.13}$$

Definition 1.5 (Moderate Deviation Principle)**.** We say that a sequence of random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies an MDP with speed b_n^2 such that (1.13) holds, and rate function $I : S \to \mathbb{R}^+$ if the sequence $(\sqrt{n}M_n/b_n)_n$ satisfies an LDP with speed b_n^2 and rate function I.

Formally, our main results about the MDP for a sequence of random variables $(M_n)_n$ will be stated as the LDP for the sequence $(\sqrt{n}M_n/b_n)_n$.

2. On moderate deviations under the Gaussian condition

In this first part, we focus our attention on moderate deviations for the Durbin–Watson statistic in the easy case where the driven noise (V_n) is normally distributed. This restrictive assumption allows us to reduce the set of hypotheses to the existence of $t > 0$ such that

G1

$$\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big] < \infty,$$

G2

$$\mathbb{E}\Big[\exp(tX_0^2)\Big] < \infty.$$

Theorem 2.1. *Assume that there exists $t > 0$ such that* **G1** *and* **G2** *are satisfied and that (V_n) follows the $\mathcal{N}(0, \sigma^2)$ distribution. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and rate function

$$I_\theta(x) = \frac{x^2}{2\sigma_\theta^2} \tag{2.1}$$

where σ_θ^2 is given by (1.7).

Theorem 2.2. *Assume that there exists $t > 0$ such that* **G1** *and* **G2** *are satisfied and that (V_n) follows the $\mathcal{N}(0, \sigma^2)$ distribution. Then, as soon as $\theta \neq -\rho$, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\begin{pmatrix}\widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^*\end{pmatrix}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and rate function

$$K(x) = \frac{1}{2}x'\varGamma^{-1}x \tag{2.2}$$

where $\varGamma$ is given by (1.10). In particular, the sequence

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n - \rho^*\right)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and rate function

$$I_\rho(x) = \frac{x^2}{2\sigma_\rho^2} \tag{2.3}$$

where σ_ρ^2 is given by (1.9).

Remark 2.3. The covariance matrix $\varGamma$ is invertible if and only if $\theta \neq -\rho$ since one can see by a straightforward calculation that its determinant is given by

$$\det(\varGamma) = \frac{\sigma_\theta^2(\theta+\rho)^2(1-\theta\rho)}{(1+\rho^2)}\cdot$$

Moreover, in the particular case where $\theta = -\rho$, the sequences

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right)\right)_{n\geq 1} \qquad \text{and} \qquad \left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n - \rho^*\right)\right)_{n\geq 1}$$

satisfy LDP on $\mathbb{R}$ with speed b_n^2 and rate functions respectively given by

$$I_\theta(x) = \frac{x^2(1-\theta^2)}{2(1+\theta^2)} \qquad \text{and} \qquad I_\rho(x) = \frac{x^2(1-\theta^2)}{2\theta^4(1+\theta^2)}.$$

Theorem 2.4. *Assume that there exists $t > 0$ such that* **G1** *and* **G2** *are satisfied and that (V_n) follows the $\mathcal{N}(0, \sigma^2)$ distribution. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{D}_n - D^*\right)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and rate function

$$I_D(x) = \frac{x^2}{2\sigma_D^2} \tag{2.4}$$

where σ_D^2 is given by (1.12).

Proof. Theorem 2.1, Theorems 2.2 and 2.4 are proved in Section 4. □

3. On moderate deviations under the Chen–Ledoux type condition

Via an extensive use of Puhalskii's result, we will now focus our attention on the more general framework where the driven noise (V_n) is assumed to satisfy the Chen–Ledoux type condition. Accordingly, one shall introduce the following hypothesis, for any $a > 0$.

CL1(a) Chen–Ledoux.

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log n\mathbb{P}\left(|V_1|^a > b_n\sqrt{n}\right) = -\infty.$$

CL2(a)

$$\frac{|\varepsilon_0|^a}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0.$$

CL3(a)

$$\frac{|X_0|^a}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0.$$

Remark 3.1. If the random variable V_1 satisfies **CL1**(2), then

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log n\mathbb{P}\left(\left|V_1^2 - \mathbb{E}[V_1^2]\right| > b_n\sqrt{n}\right) = -\infty, \tag{3.1}$$

which implies in particular that $\mathrm{Var}(V_1^2) < \infty$. Moreover, if the random variable V_1^2 has exponential moments, *i.e.* if there exists $t > 0$ such that

$$\mathbb{E}\left[\exp\left(tV_1^2\right)\right] < \infty,$$

then **CL1**(2) is satisfied for every increasing sequence (b_n). From [1, 2, 14], condition (3.1) is equivalent to say that the sequence

$$\left(\frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n}\Big(V_k^2-\mathbb{E}[V_k^2]\Big)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and rate function

$$I(x)=\frac{x^2}{2\mathrm{Var}(V_1^2)}\cdot$$

Remark 3.2. If we choose $b_n = n^\alpha$ with $0 < \alpha < 1/2$, **CL1**(2) is immediately satisfied if there exists $t > 0$ and $0 < \beta < 1$ such that

$$\mathbb{E}\Big[\exp\big(tV_1^{2\beta}\big)\Big]<\infty,$$

which is clearly a weaker assumption than the existence of $t > 0$ such that

$$\mathbb{E}\Big[\exp\big(tV_1^{2}\big)\Big]<\infty,$$

imposed in the previous section.

Remark 3.3. If **CL1**(a) is satisfied, then **CL1**(b) is also satisfied for all $0 < b < a$.

Remark 3.4. In the technical proofs that will follow, rather than **CL1**(4), the weakest assumption really needed is summarized by the existence of a large constant C such that

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}V_k^4>C\right)=-\infty.$$

Theorem 3.5. *Assume that* **CL1**(4), **CL2**(4) *and* **CL3**(4) *are satisfied. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n-\theta^*\right)\right)_{n\geq 1}$$

satisfies the LDP on $\mathbb{R}$ *stated in Theorem 2.1.*

Theorem 3.6. *Assume that* **CL1**(4), **CL2**(4) *and* **CL3**(4) *are satisfied. Then, as soon as* $\theta \neq -\rho$, *the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\begin{pmatrix}\widehat{\theta}_n-\theta^*\\ \widehat{\rho}_n-\rho^*\end{pmatrix}\right)_{n\geq 1}$$

satisfies the LDP on $\mathbb{R}^2$ *stated in Theorem 2.2. In particular, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n-\rho^*\right)\right)_{n\geq 1}$$

satisfies the LDP on $\mathbb{R}$ *also stated in Theorem 2.2.*

Remark 3.7. We have already seen in Remark 2.3 that the covariance matrix Γ is invertible if and only if $\theta \neq -\rho$. In the particular case where $\theta = -\rho$, the sequences

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n-\theta^*\right)\right)_{n\geq 1}\qquad\text{and}\qquad\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n-\rho^*\right)\right)_{n\geq 1}$$

satisfy the LDP on $\mathbb{R}$ stated in Remark 2.3.

Theorem 3.8. *Assume that* **CL1**(4), **CL2**(4) *and* **CL3**(4) *are satisfied. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{D}_n - D^*\right)\right)_{n\geq 1}$$

satisfies the LDP on $\mathbb{R}$ *stated in Theorem 2.4.*

Proof. Theorem 3.5, Theorems 3.6 and 3.8 are proved in Section 4. □

4. Proof of the main results

For a matter of readability, some notations commonly used in the following proofs have to be introduced. First, for all $n \geq 1$, let

$$L_n = \sum_{k=1}^{n} V_k^2. \tag{4.1}$$

Then, let us define M_n, for all $n \geq 1$, as

$$M_n = \sum_{k=1}^{n} X_{k-1} V_k \tag{4.2}$$

where $M_0 = 0$. For all $n \geq 1$, denote by $\mathcal{F}_n$ the σ-algebra of the events occurring up to time n, $\mathcal{F}_n = \sigma(X_0, \varepsilon_0, V_1, \cdots, V_n)$. We infer from (4.2) that $(M_n)_{n\geq 0}$ is a locally square-integrable real martingale with respect to the filtration $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ with predictable quadratic variation given by $\langle M\rangle_0 = 0$ and for all $n \geq 1$, $\langle M\rangle_n = \sigma^2 S_{n-1}$, where

$$S_n = \sum_{k=0}^{n} X_k^2. \tag{4.3}$$

Moreover, $(N_n)_{n\geq 0}$ is defined, for all $n \geq 2$, as

$$N_n = \sum_{k=2}^{n} X_{k-2} V_k \tag{4.4}$$

and $N_0 = N_1 = 0$. It is not hard to see that $(N_n)_{n\geq 0}$ is also a locally square-integrable real martingale sharing the same properties than $(M_n)_{n\geq 0}$. More precisely, its predictable quadratic variation is given by $\langle N\rangle_n = \sigma^2 S_{n-2}$. To conclude, let $P_0 = 0$ and, for all $n \geq 1$,

$$P_n = \sum_{k=1}^{n} X_{k-1} X_k. \tag{4.5}$$

To smooth the reading of the following proofs, we introduce some relations.

Lemma 4.1. *For any* $\eta > 0$,

$$\sum_{k=0}^{n} |X_k|^{\eta} \leq (1+\alpha(\eta))|X_0|^{\eta} + \alpha(\eta)\beta(\eta)|\varepsilon_0|^{\eta} + \alpha(\eta)\beta(\eta)\sum_{k=1}^{n} |V_k|^{\eta}$$

where

$$\alpha(\eta) = (1-|\theta|)^{-\eta} \qquad \text{and} \qquad \beta(\eta) = (1-|\rho|)^{-\eta}.$$

In addition,

$$\max_{1\le k\le n} X_k^2 \le \alpha(1)\, X_0^2 + \alpha(2)\beta(1)\, \varepsilon_0^2 + \alpha(2)\beta(2) \max_{1\le k\le n} V_k^2.$$

Proof. The proof follows from (1.1). Details are given in the proof of Lemma A.2 in [3]. □

Lemma 4.2. *For all $n \geq 2$,*

$$\frac{S_n}{n} - \ell = \frac{\ell}{\sigma^2}\left[\left(\frac{L_n}{n} - \sigma^2\right) + 2\theta^* \frac{M_n}{n} - 2\theta\rho \frac{N_n}{n} + \frac{R_n}{n}\right] \tag{4.6}$$

where L_n, M_n, S_n and N_n are respectively given by (4.1), (4.2), (4.3) and (4.4),

$$R_n = [2(\theta+\rho)\rho^* - (\theta+\rho)^2 - (\theta\rho)^2]X_n^2 - (\theta\rho)^2 X_{n-1}^2 + 2\rho^* X_n X_{n-1} + \xi_1,$$

and where the remainder term

$$\xi_1 = (1 - 2\theta\rho - \rho^2)X_0^2 + \rho^2\varepsilon_0^2 + 2\theta\rho X_0\varepsilon_0 - 2\rho\rho^*(\varepsilon_0 - X_0)X_0 + 2\rho(\varepsilon_0 - X_0)V_1.$$

In addition, for all $n \geq 1$,

$$\frac{P_n}{n} - \theta^* \frac{S_n}{n} = \frac{1}{1+\theta\rho}\frac{M_n}{n} + \frac{1}{1+\theta\rho}\frac{R_n(\theta)}{n} - \theta^* \frac{X_n^2}{n} \tag{4.7}$$

with

$$R_n(\theta) = \theta\rho X_n X_{n-1} + \rho X_0(\varepsilon_0 - X_0).$$

Proof. The results follow from direct calculation. □

4.1. Proof of Theorem 2.1

Before starting the Proof of Theorem 2.1, we need to introduce some technical tools. Denote by ℓ the almost sure limit of S_n/n [3], given by

$$\ell = \frac{\sigma^2(1+\theta\rho)}{(1-\theta^2)(1-\theta\rho)(1-\rho^2)}. \tag{4.8}$$

Lemma 4.3. *Under the assumptions of Theorem 2.1, we have the exponential convergence*

$$\frac{S_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \ell \tag{4.9}$$

where ℓ is given by (4.8).

Proof. First of all, (V_n) is a sequence of independent and identically distributed gaussian random variables with zero mean and variance $\sigma^2 > 0$. It immediately follows from Cramér–Chernoff's Theorem, expounded *e.g.* in [7], that for all $\delta' > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left|\frac{L_n}{n} - \sigma^2\right| > \delta'\right) < 0. \tag{4.10}$$

Since $b_n^2 = o(n)$, the latter convergence leads to

$$\frac{L_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \sigma^2, \tag{4.11}$$

ensuring the exponential convergence of L_n/n to σ^2 with speed b_n^2. Moreover, for all $\delta > 0$ and a suitable $t > 0$, we clearly obtain from Markov's inequality that

$$\mathbb{P}\left(\frac{X_0^2}{n} > \delta\right) \leq \exp\left(-tn\delta\right) \mathbb{E}\Big[\exp(tX_0^2)\Big],$$

which immediately implies *via* **G2**,

$$\frac{X_0^2}{n} \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.12}$$

and we get the exponential convergence of X_0^2/n to 0 with speed b_n^2. The same is true for V_1^2/n, ε_0^2/n and more generally for any isolated term in ξ_1 given after (4.6). Let us now focus our attention on X_n^2/n. The model (1.1) can be rewritten in the vectorial form,

$$\Phi_n = A\Phi_{n-1} + W_n \tag{4.13}$$

where $\Phi_n = \big(X_n \ X_{n-1}\big)'$ stands for the lag vector of order 2, $W_n = \big(V_n \ 0\big)'$ and

$$A = \begin{pmatrix} \theta + \rho & -\theta\rho \\ 1 & 0 \end{pmatrix}. \tag{4.14}$$

It is easy to show that the spectral radius of A is given by $\rho(A) = \max(|\theta|, |\rho|) < 1$ under the stability conditions. Then,

$$\frac{\|\Phi_n\|^2}{n} \xrightarrow[b_n^2]{\text{superexp}} 0,$$

according to [23], which is clearly sufficient to deduce that

$$\frac{X_n^2}{n} \xrightarrow[b_n^2]{\text{superexp}} 0. \tag{4.15}$$

The exponential convergence of R_n/n to 0 with speed b_n^2 is achieved following exactly the same lines. To conclude the proof of Lemma 4.3, it remains to study the exponential asymptotic behavior of M_n/n. For all $\delta > 0$ and a suitable $y > 0$,

$$\begin{aligned}\mathbb{P}\left(\frac{M_n}{n} > \delta\right) &= \mathbb{P}\left(\frac{M_n}{n} > \delta, \langle M\rangle_n \leq y\right) + \mathbb{P}\left(\frac{M_n}{n} > \delta, \langle M\rangle_n > y\right), \\ &\leq \exp\left(-\frac{n^2\delta^2}{2y}\right) + \mathbb{P}\Big(\langle M\rangle_n > y\Big),\end{aligned} \tag{4.16}$$

by application of Theorem 4.1 of [4] in the case of a gaussian martingale, and Remark 4.2 that follows. From Lemma 4.1, one can find α and β such that, for a suitable $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(\langle M\rangle_n > y\Big) &\leq \mathbb{P}\Big(X_0^2 > \frac{y}{3\alpha\sigma^2}\Big) + \mathbb{P}\left(\varepsilon_0^2 > \frac{y}{3\beta\sigma^2}\right) + \mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right), \\ &\leq 3\max\left(\exp\left(\frac{-yt}{3\alpha\sigma^2}\right)\mathbb{E}\Big[\exp(tX_0^2)\Big], \exp\left(\frac{-yt}{3\beta\sigma^2}\right)\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big],\right. \\ &\qquad \left.\mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right)\right).\end{aligned}$$

Let us choose $y = nx$, assuming $x > 3\beta\sigma^4$. It follows that

$$\begin{aligned}\frac{1}{b_n^2}\log\mathbb{P}\Big(\langle M\rangle_n > nx\Big) \leq \frac{\log 3}{b_n^2} + \frac{1}{b_n^2}\max\left(\frac{-nxt}{3\alpha\sigma^2} + \log\mathbb{E}\Big[\exp(tX_0^2)\Big],\right. \\ \left.\frac{-nxt}{3\beta\sigma^2} + \log\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big], \log\mathbb{P}\left(L_{n-1} > \frac{nx}{3\beta\sigma^2}\right)\right).\end{aligned}$$

Since $b_n^2 = o(n)$ and by virtue of (4.10) with $\delta' = x/(3\beta\sigma^2) - \sigma^2 > 0$, we obtain that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\langle M\rangle_n > nx\Big) = -\infty. \tag{4.17}$$

It enables us by (4.16) to deduce that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{M_n}{n} > \delta\right) = -\infty. \tag{4.18}$$

The same result is also true replacing M_n by $-M_n$ in (4.18) since M_n and $-M_n$ share the same distribution. Therefore, we find that

$$\frac{M_n}{n} \underset{b_n^2}{\overset{\text{superexp}}{\longrightarrow}} 0. \tag{4.19}$$

A similar reasoning leads to the exponential convergence of N_n/n to 0, with speed b_n^2. Finally, we obtain (4.9) from Lemma 4.2 together with (4.11), (4.12), (4.15) and (4.19) which achieves the proof of Lemma 4.3. □

Corollary 4.4. *By virtue of Lemma 4.3 and under the same assumptions, we have the exponential convergence*

$$\frac{P_n}{n} \underset{b_n^2}{\overset{\text{superexp}}{\longrightarrow}} \ell_1 \tag{4.20}$$

where $\ell_1 = \theta^*\ell$.

Proof. The proof is immediately derived from previous statements and Lemma 4.2. □

We are now in the position to prove Theorem 2.1. We shall make use of the following MDP for martingales established by Worms [23].

Theorem 4.5 (Worms)**.** *Let (Y_n) be an adapted sequence with values in $\mathbb{R}^p$, and (V_n) a gaussian noise with variance $\sigma^2 > 0$. We suppose that (Y_n) satisfies, for some invertible square matrix C of order p and a speed sequence (b_n^2) such that $b_n^2 = o(n)$, the exponential convergence for any $\delta > 0$,*

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=0}^{n-1} Y_k Y_k' - C\right\| > \delta\right) = -\infty. \tag{4.21}$$

Then, the sequence

$$\left(\frac{1}{b_n\sqrt{n}} \sum_{k=1}^{n} Y_{k-1}V_k\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}^p$ of speed b_n^2 and rate function

$$I(x) = \frac{1}{2\sigma^2} x' C^{-1} x. \tag{4.22}$$

Proof. The proof of Theorem 4.5 is contained in the one of Theorem 5 of [23] with $d = 1$. □

Proof of Theorem 2.1. Let us consider the decomposition

$$\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right) = \frac{n}{\langle M\rangle_n} A_n + B_n, \tag{4.23}$$

with

$$A_n = \left(\frac{\sigma^2}{1+\theta\rho}\right) \frac{M_n}{b_n\sqrt{n}} \quad \text{and} \quad B_n = \frac{\sqrt{n}}{b_n}\left(\frac{1}{1+\theta\rho}\right)\frac{R_n(\theta)}{S_{n-1}},$$

that can be obtained by a straighforward calculation, where the remainder term $R_n(\theta)$ is defined after (4.7). First, by using the same methodology as in convergence (4.12), we obtain that for all $\delta > 0$ and for a suitable $t > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{X_0^2}{b_n\sqrt{n}} > \delta\right) \leq \lim_{n\to\infty}\left(-t\delta\frac{\sqrt{n}}{b_n}\right) + \lim_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\Big[\exp(tX_0^2)\Big], \\ = -\infty, \tag{4.24}$$

since $b_n = o(\sqrt{n})$, and the same is true for any isolated term in (4.23) of order 2 whose numerator does not depend on n. Moreover, under the gaussian assumption on the driven noise (V_n), it is not hard to see that

$$\frac{1}{b_n\sqrt{n}} \max_{1\leq k\leq n} V_k^2 \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0. \tag{4.25}$$

As a matter of fact, for all $\delta > 0$ and for all $t > 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq n} V_k^2 \geq \delta b_n\sqrt{n}\right) = \mathbb{P}\left(\bigcup_{k=1}^{n}\left\{V_k^2 \geq \delta b_n\sqrt{n}\right\}\right) \leq \sum_{k=1}^{n}\mathbb{P}\Big(V_k^2 \geq \delta b_n\sqrt{n}\Big), \\ \leq n\exp\left(-t\delta b_n\sqrt{n}\right)\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big].$$

In addition, as soon as $0 < t < 1/(2\sigma^2)$, $\mathbb{E}\big[\exp(tV_1^2)\big] < \infty$. Consequently,

$$\frac{1}{b_n^2}\log\mathbb{P}\left(\max_{1\leq k\leq n} V_k^2 \geq \delta b_n\sqrt{n}\right) \leq \frac{\sqrt{n}}{b_n}\left(\frac{\log n}{b_n\sqrt{n}} - t\delta + \frac{\log\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big]}{b_n\sqrt{n}}\right)$$

which clearly leads to (4.25). Then, we deduce from (4.24), (4.25) and Lemma 4.1 that

$$\frac{1}{b_n\sqrt{n}} \max_{1\leq k\leq n} X_k^2 \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0, \tag{4.26}$$

which of course imply the exponential convergence of $X_n^2/(b_n\sqrt{n})$ to 0, with speed b_n^2. Therefore, we obtain that

$$\frac{R_n(\theta)}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0. \tag{4.27}$$

We infer from Lemma 4.3 and Lemma 2 of [23] that the following convergence is satisfied,

$$\frac{n}{S_n} \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} \frac{1}{\ell} \tag{4.28}$$

where $\ell > 0$ is given by (4.8). According to (4.27), the latter convergence and again Lemma 2 of [23], we deduce that

$$B_n \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0. \tag{4.29}$$

Hence, we obtain from (4.28) that the same is true for

$$A_n\left(\frac{n}{\langle M\rangle_n} - \frac{1}{\sigma^2\ell}\right) \overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0, \tag{4.30}$$

since Lemma 4.3 together with Theorem 4.5 with $p = 1$ directly show that $(M_n/(b_n\sqrt{n}))$ satisfies an LDP with speed b_n^2 and rate function given, for all $x \in \mathbb{R}$, by

$$J(x) = \frac{x^2}{2\ell\sigma^2}\cdot \tag{4.31}$$

As a consequence,

$$\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right) \underset{b_n^2}{\overset{\text{superexp}}{\sim}} \frac{1}{\ell(1+\theta\rho)}\frac{M_n}{b_n\sqrt{n}}, \tag{4.32}$$

and this implies that both of them share the same LDP, see Theorem 4.2.13 in [7]. One shall now take advantage of the contraction principle ([7], Thm. 4.2.1), to establish that $(\sqrt{n}(\widehat{\theta}_n - \theta^*)/b_n)$ satisfies an LDP with speed b_n^2 and rate function $I_\theta(x) = J(\ell(1+\theta\rho)x)$ given by (2.1), that is

$$I_\theta(x) = \frac{x^2}{2\sigma_\theta^2},$$

which achieves the Proof of Theorem 2.1. □

4.2. Proof of Theorem 2.2

We need to introduce some more notations. For all $n \geq 2$, let

$$Q_n = \sum_{k=2}^{n} X_{k-2}X_k. \tag{4.33}$$

In addition, for all $n \geq 1$, denote

$$T_n = 1 + \theta^*\rho^* - \left(1 + \rho^*(\widehat{\theta}_n + \theta^*)\right)\frac{S_n}{S_{n-1}} + \left(2\rho^* + \widehat{\theta}_n + \theta^*\right)\frac{P_n}{S_{n-1}} - \frac{Q_n}{S_{n-1}}, \tag{4.34}$$

where S_n and P_n are respectively given by (4.3) and (4.5). Finally, for all $n \geq 0$, let

$$J_n = \sum_{k=0}^{n} \widehat{\varepsilon}_k^{\,2} \tag{4.35}$$

where the residual sequence $(\widehat{\varepsilon}_n)$ is given in (1.3). A set of additional technical tools has to be expounded to make the Proof of Theorem 2.2 more tractable.

Corollary 4.6. *By virtue of Lemma 4.3 and under the same assumptions, we have the exponential convergence*

$$\frac{Q_n}{n} \underset{b_n^2}{\overset{\text{superexp}}{\longrightarrow}} \ell_2$$

where $\ell_2 = ((\theta+\rho)\theta^* - \theta\rho)\ell$.

Proof. The Proof of Corollary 4.6 immediately follows from the relation

$$\frac{Q_n}{n} - ((\theta+\rho)\theta^* - \theta\rho)\frac{S_n}{n} = \theta^*\frac{M_n}{n} + \frac{N_n}{n} + \frac{\xi_n^Q}{n} \tag{4.36}$$

where ξ_n^Q is a residual term made of isolated terms such that

$$\frac{\xi_n^Q}{n} \underset{b_n^2}{\overset{\text{superexp}}{\longrightarrow}} 0,$$

see *e.g.* the proof of Theorem 3.2 in [3] where more details are given on ξ_n^Q. □

Lemma 4.7. *Under the assumptions of Theorem 2.2, we have the exponential convergence*

$$A_n \xrightarrow[b_n^2]{\text{superexp}} A$$

where

$$A_n = \frac{n}{1+\theta\rho}\begin{pmatrix} \dfrac{1}{S_{n-1}} & 0 \\ \dfrac{T_n}{J_{n-1}} & -\dfrac{(\theta+\rho)}{J_{n-1}} \end{pmatrix}, \tag{4.37}$$

and

$$A = \frac{1}{\ell(1+\theta\rho)(1-(\theta^*)^2)}\begin{pmatrix} 1-(\theta^*)^2 & 0 \\ \theta\rho+(\theta^*)^2 & -(\theta+\rho) \end{pmatrix}. \tag{4.38}$$

Proof. Via (4.28), we directly obtain the exponential convergence,

$$\frac{1}{(1+\theta\rho)}\frac{n}{S_{n-1}} \xrightarrow[b_n^2]{\text{superexp}} \frac{1}{\ell(1+\theta\rho)}. \tag{4.39}$$

The combination of Lemma 4.3, Corollary 4.4, Corollary 4.6 and Lemma 2 of [23] shows, after a simple calculation, that

$$T_n \xrightarrow[b_n^2]{\text{superexp}} (\theta^*)^2 + \theta\rho. \tag{4.40}$$

Moreover, J_n given by (4.35) can be rewritten as

$$J_n = S_n - 2\widehat{\theta}_n P_n + \widehat{\theta}_n^{\,2} S_{n-1},$$

which leads, *via* Lemma 2 of [23], to

$$\frac{J_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \ell(1-(\theta^*)^2). \tag{4.41}$$

Convergences (4.40) and (4.41) imply

$$\left(\frac{n}{1+\theta\rho}\right)\frac{T_n}{J_{n-1}} \xrightarrow[b_n^2]{\text{superexp}} \frac{(\theta^*)^2+\theta\rho}{\ell(1+\theta\rho)(1-(\theta^*)^2)}, \tag{4.42}$$

and consequently,

$$\left(\frac{n}{1+\theta\rho}\right)\frac{\theta+\rho}{J_{n-1}} \xrightarrow[b_n^2]{\text{superexp}} \frac{\theta+\rho}{\ell(1+\theta\rho)(1-(\theta^*)^2)}. \tag{4.43}$$

Finally, (4.39) together with (4.42) and (4.43) achieve the proof of Lemma 4.7. □

Proof of Theorem 2.2. We shall make use of the decomposition

$$\frac{\sqrt{n}}{b_n}\begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} = \frac{1}{b_n\sqrt{n}} A_n Z_n + B_n, \tag{4.44}$$

where A_n is given by (4.37), $(Z_n)_{n\geq 0}$ is the 2-dimensional vector martingale given by

$$Z_n = \begin{pmatrix} M_n \\ N_n \end{pmatrix}, \tag{4.45}$$

and where the remainder term

$$B_n = \frac{1}{(1+\theta\rho)}\frac{\sqrt{n}}{b_n}\begin{pmatrix} \dfrac{R_n(\theta)}{S_{n-1}} \\ \dfrac{R_n(\rho)}{J_{n-1}} \end{pmatrix}. \tag{4.46}$$

The first component $R_n(\theta)$ is given in (4.7) while $R_n(\rho)$, whose definition may be found in the proof of Theorem 3.2 in [3], is made of isolated terms. Consequently, (4.24) and (4.27) are sufficient to ensure that

$$\frac{R_n(\theta)}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0 \qquad \text{and} \qquad \frac{R_n(\rho)}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0.$$

Therefore, we obtain that

$$B_n \xrightarrow[b_n^2]{\text{superexp}} 0. \tag{4.47}$$

In addition, it follows from Lemma 4.7 and Theorem 4.5 with $p=2$ that $(Z_n/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and rate function given, for all $x \in \mathbb{R}^2$, by

$$J(x) = \frac{1}{2\sigma^2} x' \Lambda^{-1} x, \tag{4.48}$$

where

$$\Lambda = \ell \begin{pmatrix} 1 & \theta^* \\ \theta^* & 1 \end{pmatrix}, \tag{4.49}$$

since we have the exponential convergence

$$\frac{\langle Z\rangle_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \sigma^2 \Lambda \tag{4.50}$$

by application of Lemma 4.3 and Corollary 4.4. One observes that Λ is invertible. As a consequence,

$$\frac{1}{b_n\sqrt{n}}(A_n - A)Z_n \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.51}$$

and we deduce from (4.44) that

$$\frac{\sqrt{n}}{b_n}\begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \underset{b_n^2}{\overset{\text{superexp}}{\sim}} \frac{1}{b_n\sqrt{n}} A Z_n. \tag{4.52}$$

This of course implies that both of them share the same LDP, see Theorem 4.2.13 in [7]. The contraction principle ([7], Thm. 4.2.1) enables us to conclude that the rate function of the LDP on $\mathbb{R}^2$ with speed b_n^2 associated with equivalence (4.52) is given, for all $x \in \mathbb{R}^2$, by $K(x) = J(A^{-1}x)$, that is

$$K(x) = \frac{1}{2} x' \Gamma^{-1} x,$$

where $\Gamma = \sigma^2 A \Lambda A'$ is given by (1.10), and where we shall suppose that $\theta \neq -\rho$ to ensure that A is invertible. In particular, the latter result also implies that the rate function of the LDP on $\mathbb{R}$ with speed b_n^2 associated with $(\sqrt{n}(\widehat{\rho}_n - \rho^*)/b_n)$ is given, for all $x \in \mathbb{R}$, by

$$I_\rho(x) = \frac{x^2}{2\sigma_\rho^2},$$

where σ_ρ^2 is the last element of the matrix Γ. This achieves the Proof of Theorem 2.2. □

4.3. Proof of Theorem 2.4

For all $n \geq 1$, denote by f_n the explosion coefficient associated with J_n given by (4.35), that is

$$f_n = \frac{J_n - J_{n-1}}{J_n} = \frac{\widehat{\varepsilon}_n^{\,2}}{J_n}. \tag{4.53}$$

It follows from decomposition (C.4) in [3] that

$$\frac{\sqrt{n}}{b_n}\Big(\widehat{D}_n - D^*\Big) = -2\frac{\sqrt{n}}{b_n}\Big(1 - f_n\Big)\Big(\widehat{\rho}_n - \rho^*\Big) + \frac{\sqrt{n}}{b_n}\zeta_n, \tag{4.54}$$

where the remainder term ζ_n is made of negligible terms, that is

$$\zeta_n = 2(\rho^* - 1)f_n + \frac{\widehat{\varepsilon}_n^{\,2} - \widehat{\varepsilon}_0^{\,2}}{J_n}.$$

From the definition of $(\widehat{\varepsilon}_n)$ in (1.3), from (4.26), (4.41) and considering that $\widehat{\varepsilon}_0 = X_0$, we clearly have that

$$\frac{\sqrt{n}}{b_n}\zeta_n \xrightarrow[b_n^2]{\text{superexp}} 0 \qquad \text{and} \qquad f_n \xrightarrow[b_n^2]{\text{superexp}} 0.$$

As a consequence,

$$\frac{\sqrt{n}}{b_n}\Big(\widehat{D}_n - D^*\Big) \underset{b_n^2}{\overset{\text{superexp}}{\sim}} -2\frac{\sqrt{n}}{b_n}\Big(\widehat{\rho}_n - \rho^*\Big), \tag{4.55}$$

and this implies that both of them share the same LDP. The contraction principle [7] enables us to conclude that the rate function of the LDP on $\mathbb{R}$ with speed b_n^2 associated with equivalence (4.55) is given, for all $x \in \mathbb{R}$, by $I_D(x) = I_\rho(-x/2)$, that is

$$I_D(x) = \frac{x^2}{2\sigma_D^2},$$

which achieves the Proof of Theorem 2.4.

4.4. Proofs of Theorem 3.5, Theorems 3.6 and 3.8

We shall now propose a technical lemma ensuring that all results already proved under the gaussian assumption still hold under the Chen–Ledoux type condition.

Lemma 4.8. *Under* **CL1**(4), **CL2**(4) *and* **CL3**(4), *all exponential convergences of Lemma 4.3, Corollary 4.4, Corollary 4.6 and Lemma 4.7 still hold.*

Proof. Following the same methodology as the one used to establish (4.27), we get

$$\mathbb{P}\left(\max_{1\le k\le n} V_k^2 \ge \delta b_n\sqrt{n}\right) \;\le\; \sum_{k=1}^{n}\mathbb{P}\Big(V_k^2 \ge \delta b_n\sqrt{n}\Big) \;=\; n\,\mathbb{P}\Big(V_1^2 \ge \delta b_n\sqrt{n}\Big).$$

Via **CL1**(2), **CL2**(2), **CL3**(2) and the same reasoning,

$$\frac{X_n^2}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.56}$$

and Cauchy–Schwarz inequality implies that this is also the case for any isolated term of order 2, such as $X_nX_{n-1}/(b_n\sqrt{n})$. This allows us to control each remainder term. Note that under **CL2**(4) and **CL3**(4) and using (4.56), ε_0^4/n, X_0^4/n, ε_0^2/n, X_0^2/n and X_n^2/n also exponentially converge to 0, since $b_n\sqrt{n} = o(n)$. Moreover, it follows from Theorem 2.2 of [14] under **CL1**(2), that

$$\frac{L_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \sigma^2. \tag{4.57}$$

Furthermore, since (M_n) is a locally square integrable martingale, we infer from Theorem 2.1 of [4] that for all $x, y > 0$,

$$\mathbb{P}\Big(|M_n| > x, \langle M\rangle_n + [M]_n \leq y\Big) \leq 2\exp\left(-\frac{x^2}{2y}\right), \tag{4.58}$$

where the predictable quadratic variation $\langle M\rangle_n = \sigma^2 S_{n-1}$ is described in (4.3) and the total quadratic variation is given by $[M]_0 = 0$ and, for all $n \geq 1$, by

$$[M]_n = \sum_{k=1}^{n} X_{k-1}^2 V_k^2. \tag{4.59}$$

According to (4.58), we have for all $\delta > 0$ and a suitable $b > 0$,

$$\begin{aligned}\mathbb{P}\left(\frac{|M_n|}{n} > \delta\right) &\leq \mathbb{P}\Big(|M_n| > \delta n, \langle M\rangle_n + [M]_n \leq nb\Big) + \mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big),\\ &\leq 2\exp\left(-\frac{n\delta^2}{2b}\right) + \mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big).\end{aligned}$$

Consequently,

$$\limsup_{n\to\infty} \frac{1}{b_n^2}\log\mathbb{P}\left(\frac{|M_n|}{n} > \delta\right) \leq \limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big). \tag{4.60}$$

Moreover, for all $n \geq 1$, let us define

$$T_n = \sum_{k=0}^{n} X_k^4 \qquad \text{and} \qquad \varGamma_n = \sum_{k=1}^{n} V_k^4.$$

From Lemma 4.1 and for n large enough, one can find $\gamma > 0$ such that

$$T_n \;\leq\; \gamma\,\varGamma_n$$

under **CL2**(4) and **CL3**(4). According to Theorem 2.2 of [14] under **CL1**(4), we also have the exponential convergence,

$$\frac{\varGamma_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \tau^4, \tag{4.61}$$

where $\tau^4 = \mathbb{E}[V_1^4]$, leading, *via* Cauchy–Schwarz inequality, to

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{[M]_n}{n} > \delta\right) &\leq \limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\varGamma_n}{n} > \frac{\delta}{\sqrt{\gamma}}\right),\\ &= -\infty,\end{aligned} \tag{4.62}$$

where $\delta > \tau^4\sqrt{\gamma}$. Exploiting (4.57) and again Lemma 4.1, the same result can be achieved for $\langle M\rangle_n/n$ under **CL1**(2) and $\delta > \sigma^4\gamma$. As a consequence, it follows from (4.62) that

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{\langle M\rangle_n + [M]_n}{n} > b\right) = -\infty, \tag{4.63}$$

as soon as $b > \sigma^4\gamma + \tau^4\sqrt{\gamma}$. Therefore, the exponential convergence of M_n/n to 0 with speed b_n^2 is obtained *via* (4.60) and (4.63), that is, for all $\delta > 0$,

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{|M_n|}{n} > \delta\right) = -\infty. \tag{4.64}$$

Explicitly, (4.64) is equivalent of (4.19) which was the main element for the proof of Lemma 4.3, and the same obviously holds for N_n/n. In consequence, one can proceed similarly to establish Corollary 4.4, Corollary 4.6 and Lemma 4.7. Indeed, hypotheses **CL2**(4) and **CL3**(4) together with exponential convergences (4.56), (4.57) and (4.64) are sufficient to achieve the proof of Lemma 4.8. □

Let us introduce a simplified version of Puhalskii's result [21] applied to a sequence of martingale differences, and two technical lemmas that shall help us to prove our results.

Theorem 4.9 (Puhalskii)**.** *Let $(m_j^n)_{1\le j\le n}$ be a triangular array of martingale differences with values in $\mathbb{R}^d$, with respect to a filtration $(\mathcal{F}_n)_{n\ge 1}$. Let (b_n) be a sequence of real numbers satisfying (1.13). Suppose that there exists a symmetric positive-semidefinite matrix Q such that*

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big[m_k^n(m_k^n)'\,\big|\,\mathcal{F}_{k-1}\Big]\overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} Q. \tag{4.65}$$

Suppose that there exists a constant $c>0$ such that, for each $1\le k\le n$,

$$|m_k^n|\le c\frac{\sqrt{n}}{b_n}\qquad \text{a.s.} \tag{4.66}$$

Suppose also that, for all $a>0$, we have the exponential Lindeberg's condition

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big[|m_k^n|^2\,\mathrm{I}_{\left\{|m_k^n|\ge a\frac{\sqrt{n}}{b_n}\right\}}\,\big|\,\mathcal{F}_{k-1}\Big]\overset{\text{superexp}}{\underset{b_n^2}{\longrightarrow}} 0. \tag{4.67}$$

Then, the sequence

$$\left(\frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n}m_k^n\right)_{n\ge 1}$$

satisfies an LDP on $\mathbb{R}^d$ with speed b_n^2 and rate function

$$\Lambda^*(v)=\sup_{\lambda\in\mathbb{R}^d}\left(\lambda' v-\frac{1}{2}\lambda' Q\lambda\right)\cdot$$

In particular, if Q is invertible,

$$\Lambda^*(v)=\frac{1}{2}\,v'Q^{-1}v. \tag{4.68}$$

Proof. The proof of Theorem 4.9 is contained *e.g.* in the proof of Theorem 3.1 in [21]. □

Lemma 4.10. *Under* **CL1**(a), **CL2**(a) *and* **CL3**(a) *for any $a>2$, we have for all $\delta>0$,*

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}X_k^2\,\mathrm{I}_{\{|X_k|>R\}}>\delta\right)=-\infty.$$

Remark 4.11. Lemma 4.10 implies that the exponential Lindeberg's condition given by (4.67) is satisfied.

Proof. From Lemma 4.1, for any $\eta>0$ and n large enough, one can find $\gamma>0$ such that

$$\sum_{k=0}^{n}|X_k|^{2+\eta}\;\le\;\gamma\sum_{k=1}^{n}|V_k|^{2+\eta} \tag{4.69}$$

under **CL2**$(2+\eta)$ and **CL3**$(2+\eta)$. If we suppose that **CL1**$(2+\eta)$ holds, then it follows that, for $R>0$,

$$R^{\eta}\sum_{k=1}^{n}X_{k-1}^{2}\,\mathrm{I}_{\{|X_{k-1}|>R\}}\;\leq\;\sum_{k=1}^{n}|X_{k-1}|^{2+\eta}\;\leq\;\gamma\sum_{k=1}^{n}|V_k|^{2+\eta},$$

for n large enough and $\eta>0$, leading to

$$\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}X_{k-1}^{2}\,\mathrm{I}_{\{|X_{k-1}|>R\}}>\delta\right)\;\leq\;\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}|V_k|^{2+\eta}>\frac{\delta}{\gamma}R^{\eta}\right).$$

Using Theorem 2.2 of [14] and letting R go to infinity, we immediately reach the end of the proof of Lemma 4.10. □

Remark 4.12. The same result can be achieved under the less restrictive **CL1**(2) condition, *via* a technical proof using the empirical measure associated with the geometric ergodic Markov chain $(X_n)_{n\geq 0}$. A same reasoning can be found in [9].

Lemma 4.13. *Under* **CL1**(4)*,* **CL2**(4) *and* **CL3**(4)*, the sequence*

$$\left(\frac{M_n}{b_n\sqrt{n}}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ *with speed* b_n^2 *and rate function*

$$J(x)=\frac{x^2}{2\ell\sigma^2} \tag{4.70}$$

where ℓ *is given by* (4.8).

Proof. From now on, in order to apply Puhalskii's result concerning MDP for martingales, we introduce the following modification of the martingale $(M_n)_{n\geq 0}$, for $r>0$ and $R>0$,

$$M_n^{(r,R)}=\sum_{k=1}^{n}X_{k-1}^{(r)}V_k^{(R)} \tag{4.71}$$

where, for all $1\leq k\leq n$,

$$X_k^{(r)}=X_k\,\mathrm{I}_{\left\{|X_k|\leq r\frac{\sqrt{n}}{b_n}\right\}}\qquad\text{and}\qquad V_k^{(R)}=V_k\,\mathrm{I}_{\left\{|V_k|\leq R\right\}}-\mathbb{E}\Big[V_k\,\mathrm{I}_{\left\{|V_k|\leq R\right\}}\Big]. \tag{4.72}$$

Then, we have to prove that for all $r>0$ the sequence $(M_n^{(r,R)})$ is an exponentially good approximation of (M_n) as R goes to infinity, see *e.g.* Definition 4.2.14 in [7]. This approximation, in the sense of the large deviations, is described by the following convergence, for all $r>0$ and all $\delta>0$,

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{|M_n-M_n^{(r,R)}|}{b_n\sqrt{n}}>\delta\right)=-\infty. \tag{4.73}$$

From Lemma 4.8, and since $\langle M\rangle_n=\sigma^2S_{n-1}$, we have

$$\frac{\langle M\rangle_n}{n}\xrightarrow[b_n^2]{\text{superexp}}\sigma^2\ell. \tag{4.74}$$

From Lemma 4.10 and Remark 4.11, we also have for all $r > 0$,

$$\frac{1}{n}\sum_{k=0}^{n} X_k^2 \mathrm{I}_{\left\{|X_k| > r\frac{\sqrt{n}}{b_n}\right\}} \xrightarrow[b_n^2]{\text{superexp}} 0. \tag{4.75}$$

We introduce the following notations,

$$\sigma_R^2 = \mathbb{E}\left[(V_1^{(R)})^2\right] \qquad \text{and} \qquad S_n^{(r)} = \sum_{k=0}^{n} (X_k^{(r)})^2.$$

Then, we easily transfer properties (4.74) and (4.75) to the truncated martingale $(M_n^{(r,R)})_{n\geq 0}$. We have for all $R > 0$ and all $r > 0$,

$$\frac{\langle M^{(r,R)}\rangle_n}{n} = \sigma_R^2 \frac{S_{n-1}^{(r)}}{n} = -\sigma_R^2\left(\frac{S_{n-1}}{n} - \frac{S_{n-1}^{(r)}}{n}\right) + \sigma_R^2 \frac{S_{n-1}}{n} \xrightarrow[b_n^2]{\text{superexp}} \sigma_R^2 \ell$$

which ensures that (4.65) is satisfied for the martingale $(M_n^{(r,R)})_{n\geq 0}$. Note also that Lemma 4.8 and Remark 4.11 work for the martingale $(M_n^{(r,R)})_{n\geq 0}$. So, for all $r > 0$, the exponential Lindeberg's condition and thus (4.67) are satisfied for $(M_n^{(r,R)})_{n\geq 0}$. By Theorem 4.9, we deduce that $(M_n^{(r,R)}/b_n\sqrt{n})$ satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and rate function

$$J_R(x) = \frac{x^2}{2\sigma_R^2 \ell}. \tag{4.76}$$

We intend to transfer the MDP result for the martingale $(M_n)_{n\geq 0}$ by proving relation (4.73). For that purpose, let us now introduce the following decomposition,

$$M_n - M_n^{(r,R)} = L_n^{(r)} + F_n^{(r,R)}$$

where

$$L_n^{(r)} = \sum_{k=1}^{n}\left(X_{k-1} - X_{k-1}^{(r)}\right)V_k \qquad \text{and} \qquad F_n^{(r,R)} = \sum_{k=1}^{n}\left(V_k - V_k^{(R)}\right)X_{k-1}^{(r)}.$$

One has to show that for all $r > 0$,

$$\frac{L_n^{(r)}}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.77}$$

and, for all $r > 0$ and all $\delta > 0$, that

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{b_n^2}\log\mathbb{P}\left(\frac{|F_n^{(r,R)}|}{b_n\sqrt{n}} > \delta\right) = -\infty. \tag{4.78}$$

Via inequality (4.69), for n large enough,

$$\begin{aligned}
\frac{|L_n^{(r)}|}{b_n\sqrt{n}} &= \frac{1}{b_n\sqrt{n}}\left|\sum_{k=1}^{n} X_{k-1}\mathrm{I}_{\left\{|X_{k-1}| > r\frac{\sqrt{n}}{b_n}\right\}} V_k\right|, \\
&\leq \frac{1}{b_n\sqrt{n}}\left(r\frac{\sqrt{n}}{b_n}\right)^{-\eta}\left(\sum_{k=1}^{n}|X_{k-1}|^{2+\eta}\right)^{1/2}\left(\sum_{k=1}^{n} V_k^2\,|X_{k-1}|^{\eta}\right)^{1/2}, \\
&\leq \lambda(r,\eta,\gamma)\left(\frac{b_n}{\sqrt{n}}\right)^{\eta-1}\frac{1}{n}\sum_{k=1}^{n}|V_k|^{2+\eta}
\end{aligned} \tag{4.79}$$

by virtue of Hölder's inequality, where $\lambda(r,\eta,\gamma) > 0$ can be easily evaluated. As a consequence, for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|L_n^{(r)}|}{b_n\sqrt{n}} > \delta\right) \le \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n |V_k|^{2+\eta} > \frac{\delta}{\lambda(r,\eta,\gamma)}\left(\frac{\sqrt{n}}{b_n}\right)^{\eta-1}\right), \\ = -\infty, \tag{4.80}$$

as soon as $\eta > 1$, by application of Theorem 2.2 of [14] under **CL1**$(2+\eta)$, since

$$\lim_{n\to\infty}\left(\frac{\sqrt{n}}{b_n}\right)^{\eta-1} = \infty.$$

We deduce that

$$\frac{L_n^{(r)}}{b_n\sqrt{n}} \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.81}$$

which achieves the proof of (4.77), under **CL1**$(2+\eta)$, **CL2**$(2+\eta)$ and **CL3**$(2+\eta)$ for $\eta > 1$. On the other hand, $(F_n^{(r,R)})_{n\ge 0}$ is a locally square-integrable real martingale whose predictable quadratic variation is given by $\langle F^{(r,R)}\rangle_0 = 0$ and, for all $n \ge 1$, by

$$\langle F^{(r,R)}\rangle_n = \mathbb{E}\left[\left(V_1 - V_1^{(R)}\right)^2\right] S_{n-1}^{(r)}.$$

To prove (4.78), we will use Theorem 1 of [8]. For R large enough and all $k \ge 1$, we have

$$\mathbb{P}\left(\left|X_{k-1}^{(r)}\left(V_k - V_k^{(R)}\right)\right| > b_n\sqrt{n}\;\middle|\;\mathcal{F}_{k-1}\right) \le \mathbb{P}\left(\left|V_k - V_k^{(R)}\right| > \frac{b_n^2}{r}\right), \\ = \mathbb{P}\left(\left|V_1 - V_1^{(R)}\right| > \frac{b_n^2}{r}\right) = 0.$$

This implies that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log\left(n \operatorname*{ess\,sup}_{k\ge 1} \mathbb{P}\left(\left|X_{k-1}^{(r)}\left(V_k - V_k^{(R)}\right)\right| > b_n\sqrt{n}\;\middle|\;\mathcal{F}_{k-1}\right)\right) = -\infty. \tag{4.82}$$

For all $\nu > 0$ and all $\delta > 0$, we obtain from Lemma 4.10 and Remark 4.11, that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n \left(X_{k-1}^{(r)}\right)^2 \mathrm{I}_{\left\{|X_{k-1}^{(r)}| > \nu\frac{\sqrt{n}}{b_n}\right\}} > \delta\right) \le \\ \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n X_{k-1}^2 \mathrm{I}_{\left\{|X_{k-1}| > \nu\frac{\sqrt{n}}{b_n}\right\}} > \delta\right) = -\infty.$$

Finally, from Lemma 4.8, Lemma 4.10 and Remark 4.11, it follows that

$$\frac{\langle F^{(r,R)}\rangle_n}{n} = Q_R\frac{S_{n-1}^{(r)}}{n} = -Q_R\left(\frac{S_{n-1}}{n} - \frac{S_{n-1}^{(r)}}{n}\right) + Q_R\frac{S_{n-1}}{n} \xrightarrow[b_n^2]{\text{superexp}} Q_R\ell$$

where

$$Q_R = \mathbb{E}\left[\left(V_1 - V_1^{(R)}\right)^2\right],$$

and ℓ is given by (4.8). Moreover, it is clear that Q_R converges to 0 as R goes to infinity. Consequently, we infer from Theorem 1 of [8] that $(F_n^{(r,R)}/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}$ of speed b_n^2 and rate function

$$I_R(x) = \frac{x^2}{2Q_R\ell}.$$

In particular, this implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|F_n^{(r,R)}|}{b_n\sqrt{n}} > \delta\right) = -\frac{\delta^2}{2Q_R\ell}, \tag{4.83}$$

and letting R go to infinity clearly leads to the end of the proof of (4.78). We are able to conclude now that $(M_n^{(r,R)}/(b_n\sqrt{n}))$ is an exponentially good approximation of $(M_n/(b_n\sqrt{n}))$. By application of Theorem 4.2.16 in [7], we find that $(M_n/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and rate function

$$\widetilde{J}(x) = \sup_{\delta>0} \liminf_{R\to\infty} \inf_{z\in B_{x,\delta}} J_R(z),$$

where J_R is given in (4.76) and $B_{x,\delta}$ denotes the ball $\{z : |z-x| < \delta\}$. The identification of the rate function $\widetilde{J} = J$, where J is given in (4.70) is done easily, which concludes the proof of Lemma 4.13. $\square$

Lemma 4.14. *Under* **CL1**(4), **CL2**(4) *and* **CL3**(4), *the sequence*

$$\left(\frac{1}{b_n\sqrt{n}}\begin{pmatrix} M_n \\ N_n \end{pmatrix}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and rate function

$$J(x) = \frac{1}{2\sigma^2} x' \Lambda^{-1} x \tag{4.84}$$

where Λ is given by (4.49).

Proof. We follow the same approach as in the proof of Lemma 4.13. We shall consider the 2-dimensional vector martingale $(Z_n)_{n\geq 0}$ defined in (4.45). In order to apply Theorem 4.9, we introduce the following truncation of the martingale $(Z_n)_{n\geq 0}$, for $r > 0$ and $R > 0$,

$$Z_n^{(r,R)} = \begin{pmatrix} M_n^{(r,R)} \\ N_n^{(r,R)} \end{pmatrix}$$

where $M_n^{(r,R)}$ is given in (4.71) and where $N_n^{(r,R)}$ is defined in the same manner, that is, for all $n \geq 2$,

$$N_n^{(r,R)} = \sum_{k=2}^{n} X_{k-2}^{(r)} V_k^{(R)} \tag{4.85}$$

with $X_n^{(r)}$ and $V_n^{(R)}$ given by (4.72). The exponential convergence (4.50) still holds, by virtue of Lemma 4.8, which immediately implies hypothesis (4.65). In addition, Lemma 4.10 ensures that, for all $r > 0$,

$$\frac{1}{n}\sum_{k=0}^{n} X_k^2 \mathrm{I}_{\left\{|X_k| > r\frac{\sqrt{n}}{b_n}\right\}} \xrightarrow[b_n^2]{\text{superexp}} 0, \tag{4.86}$$

justifying hypothesis (4.67). *Via* Theorem 4.9, $(Z_n^{(r,R)}/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and rate function J_R given by

$$J_R(x) = \frac{1}{2\sigma_R^2} x' \Lambda^{-1} x. \tag{4.87}$$

Finally, it is straightforward to prove that $(Z_n^{(r,R)}/(b_n\sqrt{n}))$ is an exponentially good approximation of $(Z_n/(b_n\sqrt{n}))$. By application of Theorem 4.2.16 in [7], we deduce that $(Z_n/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and rate function given by

$$\widetilde{J}(x) = \sup_{\delta>0} \liminf_{R\to\infty} \inf_{z\in B_{x,\delta}} J_R(z),$$

where J_R is given in (4.87) and $B_{x,\delta}$ denotes the ball $\{z : |z-x| < \delta\}$. The identification of the rate function $\widetilde{J} = J$ is done easily, which concludes the proof of Lemma 4.14. □

Proofs of Theorem 3.5, Theorems 3.6 and 3.8. The residuals appearing in the decompositions (4.23), (4.44) and (4.54) still converge exponentially to zero under **CL1**(4), **CL2**(4) and **CL3**(4), with speed b_n^2, as it was already proved. Therefore, for a better readability, we may skip the most accessible parts of these proofs whose development merely consists in following the same lines as those in the proofs of Theorem 2.1, Theorem 2.2 and Theorem 2.4, taking advantage of Lemmas 4.13 and 4.14, and applying the contraction principle given *e.g.* in [7]. □

Acknowledgements. The authors thank Bernard Bercu and Arnaud Guillin for all their advices and suggestions during the preparation of this work. The authors are also very grateful to the Associate Editor and the Reviewer for spending time to evaluate this manuscript and for providing comments and suggestions that improve the paper substantially.

References

[1] M.A. Arcones, The large deviation principle for stochastic processes I. *Theory Probab. Appl.* **47** (2003) 567–583.
[2] M.A. Arcones, The large deviation principle for stochastic processes II. *Theory Probab. Appl.* **48** (2003) 19–44.
[3] B. Bercu and F. Proïa, A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. *ESAIM: PS* **17** (2013) 500–530.
[4] B. Bercu and A. Touati, Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18** (2008) 1848–1869.
[5] X. Chen, Moderate deviations for m-dependent random variables with Banach space value. *Stat. Probab. Lett.* **35** (1998) 123–134.
[6] A. Dembo, Moderate deviations for martingales with bounded jumps. *Electron. Commun. Probab.* **1** (1996) 11–17.
[7] A. Dembo and O. Zeitouni, Large deviations techniques and applications, 2nd edition, vol. 38 of *Appl. Math.* Springer (1998).
[8] H. Djellout, Moderate deviations for martingale differences and applications to ϕ-mixing sequences. *Stoch. Stoch. Rep.* **73** (2002) 37–63.
[9] H. Djellout and A. Guillin, Moderate deviations for Markov chains with atom. *Stochastic Process. Appl.* **95** (2001) 203–217.
[10] J. Durbin, Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* **38** (1970) 410–421.
[11] J. Durbin and G.S. Watson, Testing for serial correlation in least squares regression I. *Biometrika* **37** (1950) 409–428.
[12] J. Durbin and G.S. Watson, Testing for serial correlation in least squares regression II. *Biometrika* **38** (1951) 159–178.
[13] J. Durbin and G.S. Watson, Testing for serial correlation in least squares regession III. *Biometrika* **58** (1971) 1–19.
[14] P. Eichelsbacher and M. Löwe, Moderate deviations for i.i.d. random variables. *ESAIM: PS* **7** (2003) 209–218.
[15] B.A. Inder, An approximation to the null distribution of the Durbin-Watson statistic in models containing lagged dependent variables. *Econometric Theory* **2** (1986) 413–428.
[16] M.L. King and P.X. Wu, Small-disturbance asymptotics and the Durbin-Watson and related tests in the dynamic regression model. *J. Econometrics* **47** (1991) 145–152.
[17] M. Ledoux, Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. Henri-Poincaré* **35** (1992) 123–134.
[18] E. Malinvaud, Estimation et prévision dans les modèles économiques autorégressifs. *Rev. Int. Inst. Statis.* **29** (1961) 1–32.
[19] M. Nerlove and K.F. Wallis, Use of the Durbin-Watson statistic in inappropriate situations. *Econometrica* **34** (1966) 235–238.
[20] F. Proïa, Further results on the H-Test of Durbin for stable autoregressive processes. *J. Multivariate. Anal.* **118** (2013) 77–101.

[21] A. Puhalskii, Large deviations of semimartingales: a maxingale problem approach I. Limits as solutions to a maxingale problem. *Stoch. Stoch. Rep.* **61** (1997) 141–243.

[22] T. Stocker, On the asymptotic bias of OLS in dynamic regression models with autocorrelated errors. *Statist. Papers* **48** (2007) 81–93.

[23] J. Worms, Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* **4** (1999) 1–28.

[24] J. Worms, Moderate deviations of some dependent variables I. Martingales. *Math. Methods Statist.* **10** (2001) 38–72.

[25] J. Worms, Moderate deviations of some dependent variables II. Some kernel estimators. *Math. Methods Statist.* **10** (2001) 161–193.

Stochastics and Stochastics Reports, 2002 Vol. 73 (1–2), pp. 37–63

MODERATE DEVIATIONS FOR MARTINGALE DIFFERENCES AND APPLICATIONS TO ϕ-MIXING SEQUENCES

HACÈNE DJELLOUT

Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620, Université Blaise Pascal, 24 avenue des Landais, 63177 Aubière, France

(Received 6 November 2000; In final form 25 July 2001)

For a $\mathbb{R}^d$-valued sequence of martingale differences $\{m_k\}_{k\geq 1}$, we obtain a moderate deviation principle for the sequence of partial sums $\{Z_n(t) := \sum_{k=1}^{[nt]} m_k/b_n,\ t \in [0,1]\}$, in the space of càdlàg functions equipped with the Skorohod topology, under the following conditions: a Chen–Ledoux type condition, an exponential convergence in probability of the associated quadratic variation process of the martingale, and a condition of "Lindeberg" type. For the small jumps of $Z_n(\cdot)$, we apply the general result of Puhalskii [Puhalskii, A. (1994). "Large deviations of semimartingales via convergence of the predictable characteristics". *Stoch. Stoch. Rep.*, **49**, pp. 27–85]. Following the method of Ledoux [Ledoux, M. (1992). "Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi". *Ann. Inst. H. Poincaré*, **28**, pp. 267–280] and Arcones [Arcones, A. (1999). "The large deviation principle for stochastic processes", Submitted for publication], we prove that the large jumps part of $Z_n(\cdot)$ is negligible in the sense of the moderate deviations. One can regard our result as an extension to martingale differences, of the beautiful characterization of moderate deviations for i.i.d.r.v. case due to Chen [Chen, X. (1991). "The moderate deviations of independent vectors in Banach space". *Chin. J. Appl. Probab. Stat.*, **7**, pp. 124–32] and Ledoux [Ledoux, M. (1992). "Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi". *Ann. Inst. H. Poincaré*, **28**, pp. 267–280]. Using the Gordin [Gordin, M.I. (1969). "The central limit theorem for stationary processes". *Soviet Math. Dokl.*, **10**, pp. 1174–1176] decomposition, the martingale result is applied to prove the moderate deviation principle for a wide class of stationary ϕ-mixing sequences of random variables.

Keywords: Moderate deviations; Martingale differences; Quadratic variation process; CLT; ϕ-mixing sequence; Döeblin recurrence

AMS classifications: Primary 60F10

*Tel.: +33-4-73-40-78-95. E-mail: djellout@math.univ-bpclermont.fr.

ISSN 1045-1129 print/ISSN 1029-0346 online
DOI: 10.1080/1045112021000004207

INTRODUCTION

Let $(m_k)_{k\geq 1}$ be a sequence of $\mathbb{R}^d$-valued martingale differences defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with respect to the filtration $(\mathscr{F}_n)_{n\in\mathbb{N}}$, $(\mathscr{F}_n \subseteq \mathscr{F}_{n+1} \subseteq \mathscr{F}, \forall n \in \mathbb{N})$. All elements of $\mathbb{R}^d$ are assumed to be column vectors. For $x \in \mathbb{R}^d$, x^* denotes the transposed vector, $|x|$ is the Euclidean norm. We denote by $\mathscr{S}(d^2)$ (resp. $\mathscr{S}^+(d^2)$) the space of symmetric (resp. symmetric positive semi-definite) $d \times d$ matrix equipped with the following norm $\|A\| = \sup_{|x|=1} |x^*Ax|$ for $A \in \mathscr{S}(d^2)$. $\mathbb{I}_B$ denotes the indicator function of a set B.

Let

$$M_0 = 0, \quad M_n = \sum_{k=1}^{n} m_k, \quad \forall n \in \mathbb{N}^*.$$

We denote by $\langle M\rangle_n$ the quadratic variation process of the martingale (M_n) given by

$$\langle M\rangle_n = \sum_{k=1}^{n} \mathbb{E}(m_k m_k^* | \mathscr{F}_{k-1}) = \sum_{k=1}^{n} \mathbb{E}_{k-1}(m_k m_k^*).$$

Here and throughout, $\mathbb{E}_{k-1}$ and $\mathbb{P}_{k-1}$ denote the conditional expectation and conditional probability knowing $\mathscr{F}_{k-1}$.

For many purposes in statistics one needs to estimate the limit behavior (when $n \to \infty$), of

$$\mathbb{P}\left(\frac{1}{b_n} M_n \in \cdot\right), \tag{1.1}$$

where b_n is a sequence of positive numbers tending to infinity.

When $b_n = \sqrt{n}$, the estimation (1.1) becomes the central limit theorem (in short CLT) for martingale, a traditional subject in probability. It is known that under the following conditions:

(a) $\displaystyle \frac{1}{n}\sum_{k=1}^{n} m_k m_k^* \to Q,$ in probability, for some $Q \in \mathscr{S}^+(d^2)$ or

(a)′ $\displaystyle \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{k-1}(m_k m_k^*) \to Q,$ in probability, for some $Q \in \mathscr{S}^+(d^2)$ and

$$\text{(b)} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_{k-1}(|m_k|^2 \mathbb{1}_{\{|m_k| \geq \delta\sqrt{n}\}}) \longrightarrow 0,$$

in probability, $\forall \delta > 0$; (condition of Lindeberg)

then $M_n/\sqrt{n}$ converges in distribution to a Gaussian law of zero mean and covariance matrix Q. Uniform and nonuniform bounds on the rate of the last convergence have been obtained by many authors, see Refs. [18,19,31,32]. For the proof of the CLT above as well as for general information about the martingale CLT, we refer to Refs. [21,22,26].

If $b_n = n$, the estimation (1.1) becomes the large deviation principle (in short LDP) extensively studied by Puhalskii [27], in general case (sequence of semimartingale).

Now assume,

$$b_n \quad \text{increasing}, \ \frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \quad \frac{b_n}{n} \longrightarrow 0. \tag{1.2}$$

The estimation of the probabilities (1.1) is usually called the moderate deviation principle (in short: MDP). So the MDP is an intermediate estimation between CLT and LDP, it is often used to give further estimations related to CLT and the law of iterated logarithm. The moderate deviation estimation arise from the requirements of statistics. Our aim in this work is to give the asymptotic behavior of the functionals associated with the Donsker invariance principle:

$$Z_n(t) = \frac{M_{[nt]}}{b_n}, \quad t \in [0,1],$$

in $(\mathbb{D}[0,1], \mathbb{R}^d)$, the space of all $\mathbb{R}^d$-valued *càdlàg* functions on [0, 1], equipped with the Skorohod topology and with the Borel σ-field $\mathbb{B}$.

Let us begin with some few bibliographical notes on the MDP (the reader will find the detail in the references quoted below).

The Case of Banach Space Valued I.i.d.r.v.'s

Borovkov and Mogulskii [3,4] considered the MDP for Banach valued i.i.d.r.v. sequences. Under the condition that $\mathbb{E}\, e^{\delta|m_1|} < +\infty$, for some $\delta > 0$, they proved the MDP for $Z_n(1)$. Baldi [2] obtained the MDP for $Z_n(\cdot)$ under the same condition.

For $b_n = n^\alpha$ with $(1/2) < \alpha < 1$, Chen [6] found the necessary and sufficient condition for the MDP in a Banach space, and he obtained the lower bound for general b_n under very weak conditions. Using the isoperimetry techniques, Ledoux [25] obtained the necessary and sufficient condition in the general case (i.e. for any sequence b_n satisfying Eq. (1.2) and a technical condition roughly meaning that b_n cannot be too near to n), extending the works of Chen [6]. Indeed, he proved the equivalence between the MDP and the following exponential tail condition:

$$\exists R > 0, \quad \text{such that} \, \forall u > 0,$$
$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log (n\mathbb{P}(|m_1| > ub_n)) \leq -\frac{u^2}{R}. \tag{1.3}$$

The results of Ledoux [25] is extended to functional empirical processes (in the setting of non parametrical statistics) by Wu [35]. The further developments are given by Dembo and Zajic [11]. Arcones [1] obtains the MDP of functional type, without the technical assumption on b_n of Ledoux [25] but with Eq. (1.3) substituted by

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log (n\mathbb{P}(|m_1| > b_n)) = -\infty. \tag{1.4}$$

On Markov Processes

How to extend the MDP to the dependent situations has recently attracted much attention and remarkable works. The Markovian case has been studied under successively less restrictive conditions (see Refs. [7,30,36], for the relevant references) and very recently under weak conditions by de Acosta [8] and Chen [7] for the lower bound (under different and non comparable conditions) and by de Acosta–Chen [9] and Chen [7] (under the same condition but different proof) for the upper bound. Guillin [20] obtained uniform (in time) MDP for functional empirical processes. Using regeneration split chain method, Djellout and Guillin [13,14] extend the characterization of MDP for i.i.d.r.v. case of Ledoux [25] to Markov chains. The geometric ergodicity is substituted by a tail on the first time of return to the atom. Their conditions are weaker than Theorem 1 in Ref. [9], which allow them to obtain MDP for empirical measures and functional empirical processes.

Those works motivate directly the studies here.

On Martingale Case

However, the studies on the MDP of martingale are more recent: see Refs. [10,16,24,27,29,32,34]. Ra$\check{c}$kauskas [32] obtained the upper bound of the MDP for a real sequence of **bounded** martingale differences $Z_n(1)$ under strong conditions:

$$\frac{\langle M \rangle_n}{n} = \sigma^2 \quad a.s., \, \sigma \in \mathbb{R}, \quad \frac{b_n}{n^{\frac{2}{3}}} \xrightarrow[n\to\infty]{} 0. \tag{1.5}$$

His proof is based on sharp estimations about the rate of the convergence of $M_n/\sqrt{n}$ to the normal law. Though he has not given the corresponding lower bound, but it is a rather easy exercise to deduce it from his Theorem 1, under the same conditions, as for the upper bound. His estimations depend heavily on his conditions (1.5), and it is difficult to obtain MDP for the functional $Z_n(\cdot)$ through his method. For a **bounded** $\mathbb{R}^d$-valued sequence of martingale differences (m_k), Dembo [10], using the cumulant method of Puhalskii [28], gives the MDP for $Z_n(\cdot)$ in $\mathbb{D}[0, \infty)$, equipped with the locally uniform topology, under the following condition

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\left\| \frac{\langle M \rangle_t}{t} - Q \right\| > \delta \right) < 0, \quad \forall \delta > 0; \tag{1.6}$$

for some $Q \in \mathscr{S}^+(d^2)$.

Puhalskii [27] established conditions for the LDP in the Skorohod topology to hold for a sequence of semi martingales in terms of the convergence of their predictable characteristics. Gao [16] discussed the MDP for martingale differences sequence under conditional exponential integrability of the martingale, with some applications to mixing sequence. Worms [34] presents some criterion for the MDP of vectorial martingale (and of some class of regression sequences) with a deterministic normalization or autonormalization. Some of his results are based on the cumulant method developed by Puhalskii [28].

Our main aim is to prove the Chen–Ledoux type theorem for the MDP of a sequence of martingale differences. Our method is the following:

(1) for small jumps of $Z_n(\cdot)$, we apply the general result of Puhalskii [27];
(2) by following the method of Ledoux [25] and Arcones [1], we prove that the large jumps part of $Z_n(\cdot)$ is negligible in the sense of the MDP.

Several technical difficulties arise in the second step (2), when one passes from the i.i.d. case of Ledoux–Arcones to the general martingale case here.

Let us present now the structure of this paper. In the second section, we give our main result, the MDP for martingale differences sequence. Applications to ϕ-mixing sequence are discussed in the third section. The fourth section is devoted to the proof of the MDP.

MAIN RESULT

Firstly, we recall the result of Puhalskii [27, Theorem 3.1], applied to the sequence of martingale differences, which gives the MDP below in the case of very small jumps (for $Z_n(\cdot)$).

PROPOSITION 1 (PUHALSKII [27]) Let $\{m_j^n : 1 \le j \le n\}$ be a triangular array of martingale differences, with values in $\mathbb{R}^d$. Let b_n be a sequence of real numbers satisfying Eq. (1.2). Suppose that:

i) there exists $Q \in \mathscr{S}^+(d^2)$ such that for all $\delta > 0$,

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\sup_{t\in[0,1]} \left\| \frac{1}{n} \sum_{k=1}^{[nt]} \mathbb{E}_{k-1}(m_k^n (m_k^n)^*) - tQ \right\| \ge \delta \right) = -\infty;$$

ii) there exists a constant c such that for each $1 \le k \le n$, $|m_k^n| \le c(n/b_n)$;

iii) $\forall a > 0$ and $\forall \delta > 0$,

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_{k-1}\left(|m_k^n|^2 \mathbb{1}_{\{|m_k^n| \ge a\frac{n}{b_n}\}} \right) \ge \delta \right) = -\infty;$$

then $Z_n(\cdot)$ satisfies the MDP in $\mathbb{D}([0,1], \mathbb{R}^d)$ (equipped with the Skorohod topology), with speed b_n^2/n and the good rate function

$$I(\phi) = \begin{cases} \int_0^1 \Lambda^*(\phi'(t))\mathrm{d}t & \text{if } \phi \in \mathscr{AC}_0([0,1]) \\ +\infty & \text{otherwise}, \end{cases} \tag{2.1}$$

where Λ^* is given by

$$\Lambda^*(v) = \sup_{\lambda\in\mathbb{R}^d} \left(\lambda^* v - \frac{1}{2}\lambda^* Q \lambda \right)$$
$$\left(= \frac{1}{2} v^* Q^{-1} v, \quad \text{if } Q \text{ is invertible} \right), \tag{2.2}$$

and

$$\mathscr{AC}_0([0,1]) = \{\phi : [0,1] \to \mathbb{R}^d \quad \text{is absolutely continuous with } \phi(0) = 0\}.$$

Remark 2.1 This result is a consequence of Theorem 3.1 in Puhalskii [27]. We present here the parallel with his conditions: (sup B), (K + L) and (L), in his notations are implied by (iii); (C) by (i); v^F is (ii), and (0) is verified since $M_0 = 0$.

Our main result is the following.

THEOREM 1 Let $(b_n = b(n))$ be a sequence satisfying Eq. (1.2), such that $c(n) := n/b_n$ is non-decreasing, and define the reciprocal function $c^{-1}(t)$ by

$$c^{-1}(t) := \inf\{n \in \mathbb{N} : c(n) \geq t\}.$$

Under the following conditions:

(H1) there exists $Q \in \mathcal{S}^+(d^2)$ such that $\forall \delta > 0$,

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\left\| \frac{\langle M \rangle_n}{n} - Q \right\| > \delta \right) = -\infty;$$

(H2) $$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \left(n \operatorname*{ess\,sup}_{1\leq k\leq c^{-1}(b(n+1))} \mathbb{P}_{k-1}(|m_k| > b_n) \right) = -\infty;$$

(H3) $\forall a > 0$, and $\forall \delta > 0$

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_{k-1}\left(|m_k|^2 \mathbb{1}_{\{|m_k| \geq a\frac{n}{b_n}\}} \right) \geq \delta \right) = -\infty;$$

$Z_n(\cdot)$ satisfies the MDP in $\mathbb{D}([0,1], \mathbb{R}^d)$ (equipped with the Skorohod topology), with speed b_n^2/n and the good rate function given by Eq. (2.1). More precisely, for any Borel-measurable subset $A \subset \mathbb{D}([0,1], \mathbb{R}^d)$, we have that

$$-\inf_{\phi\in A^\circ} I(\phi) \leq \liminf_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}(Z_n(\cdot) \in A) \leq \limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}(Z_n(\cdot) \in A)$$
$$\leq -\inf_{\phi\in \bar{A}} I(\phi);$$

where A° and $\bar{A}$ denote the interior and the closure of A, respectively.

We present some remarks and comments on the conditions (**H1**), (**H2**) and (**H3**).

Remark 2.2 In the i.i.d.r.v.'s case, conditions (**H1**) and (**H3**) are automatically satisfied, and condition (**H2**) is exactly the tail condition in the work of Arcones [1]. Assumption (**H2**) is the so called Chen–Ledoux condition. Moreover, for the processes level MDP, the condition (**H2**) is necessary in the i.i.d.r.v.'s case, as indicated by Arcones [1].

Ledoux [25] assumes that for some $A > 1$ and some $0 < \delta < 1$, $b_{nk} \leq Ak^{1-\delta} b_n$, $\forall n, k \in \mathbb{N}$. This condition is not satisfied by sequences very close to n, for example by $b_n = n(\log n)^{-\alpha}$ with $\alpha > 0$. This technical condition on (b_n) is removed as in Arcones [1] for the i.i.d.r.v. case.

Remark 2.3 If $\max_{1\leq k\leq n}|m_k| \leq cn/b_n$, for some constant c, this theorem is proved, under (**H1**) and (**H3**) by Puhalskii [27, Theorem 3.1], (in the bounded case and under (**H1**) by Dembo [10, Proposition 1]). In the general case, if one imposes Eq. (4.4) (see the "Proof of theorem 1" section), (**H1**) and (**H3**), then it is not difficult to see that by approximation technique, the MDP in Theorem 1 can be reduced to the previous case. This was realized by Worms [34, Théorème 2.1]. Worms [34, Chapitre 1 and 2] also discussed some conditions which are sufficient for Eq. (4.4) and compared them with Puhalskii's criterion [27, Corollary 6.3]. The key point of this paper is to substitute condition (4.4), which is not explicit and quite difficult to check, by the much easier condition (**H2**). Finally our conditions are weaker than those of Gao [16, Theorem 1.1].

Remark 2.4 The condition (**H3**) is of "Lindeberg" type, compared with condition (b) of the CLT.

Remark 2.5 Under conditions (**H1**), (**H2**) and (**H3**), $Z_n(1)$ satisfies the MDP in $\mathbb{R}^d$ with speed b_n^2/n and the good rate function given by Eq. (2.2), by the contraction principle [12, Theorem 4.2.1].

Remark 2.6 [37] Even for a martingale (M_n) of stationary ergodic bounded jumps with $\mathbb{E}(m_1)^2 = 1$, the condition (**H1**) is indispensable to Theorem 1, shown by the following counter-example: let $(\xi_k)_{k\geq 1}$ be a sequence of real bounded i.i.d.r.v. with $\mathbb{E}(\xi_k) = 0$, $\mathbb{E}(\xi_k)^2 = 1$. Let us construct another sequence of $\mathbb{N}$-valued stationary ergodic bounded r.v. $(\tau_i)_{i\geq 1}$, independent of $(\xi_k)_{k\geq 1}$, satisfying $1 < \mathbb{E}\tau_i < +\infty$ and for some $\delta > 0$,

$$C(\delta) := \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{\sigma_n}{n} - \mathbb{E}\tau_1 > \delta\right) > -\infty,$$

where $\sigma_n = \sum_{k=1}^n \tau_k$, $b_n \geq \sqrt{n\log(n)}$ is fixed.

In fact let $(X_n)_{n\geq 0}$ be a stationary ergodic Markov chain on $\mathbb{Z}$ with transition probability $P(i,j)$, satisfying

i) it is independent of $(\xi_k)_{k\geq 1}$;
ii) the probability P is transitive in $\mathbb{Z}$, and $\mathbb{E}T^N = +\infty$ for some $N \geq 2$, where $T = \inf\{n \geq 0; X_n = 0\}$.

So, P is not recurrent of degree N. Let $\tau_i = \mathbb{I}_{\mathbb{Z}^*}(X_{i-1})$. Take $0 < \delta < 1 - \mathbb{P}(X_0 \in \mathbb{Z}^*)$. We claim that $C(\delta) > -\infty$. If in contrary $C(\delta) = -\infty$, then for all $p > 0$, and for all n large enough

$$\mathbb{P}(T > n-1) = \mathbb{P}\left(\frac{\sigma_n}{n} = 1\right) \leq \mathbb{P}\left(\frac{\sigma_n}{n} - \mathbb{E}\tau_1 > \delta\right) \leq \mathrm{e}^{-p\frac{b_n^2}{n}} \leq \frac{1}{n^p},$$

which is in contradiction with (ii).

The counter-example is then given by

$$M_n = \sum_{k=1}^{\sigma_n} \frac{\xi_k}{h}, \quad \text{where } h = \sqrt{\mathbb{E}\tau_1},$$

which is a ($\mathscr{F}_n = \sigma(\xi_k, k \leq n, \tau_i, i \geq 1)$)-martingale with $\langle M \rangle_n = \sigma_n/h^2$ and with bounded differences. So the condition **(H1)** is violated.

But by the MDP of $S_n = \sum_{k=1}^n \xi_k$ (well known, see Ref. [25]), we have for $r > 0$,

$$\begin{aligned}
\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}(Z_n(1) > r) &= \limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{S_{\sigma_n}}{h} > rb_n\right) \\
&\geq \limsup_{n\to+\infty} \frac{n}{b_n^2} \log \inf_{m \geq n(h^2+\delta)} \mathbb{P}(S_m > hrb_n)\mathbb{P}(\sigma_n > n(h^2+\delta)) \\
&\geq -\frac{1}{2}\frac{r^2h^2}{h^2+\delta} + C(\delta).
\end{aligned}$$

Consequently the upper bound of the MDP of $Z_n(1)$ fails for $\bar{A} = F = [r, +\infty)$ for r large enough.

Remark 2.7 If (M_t) is continuous, then **(H1)** (without **(H2)**, **(H3)**) alone is sufficient to the MDP of $Z_n(\cdot)$ on the space of all continuous maps with speed b_n^2/n and good rate function given by Eq. (2.1). It is a consequence of the approximation lemma [12, Theorem 4.2.13], the Skorohod representation of martingale [33, Theorem 1.6, Chapter V] and Schilder's theorem [12, Theorem 5.2.3].

Remark 2.8 In contrast with the MDP, note that the LDP with speed n may fail for M_n/n even when (M_n) is a real valued discrete parameter martingale with bounded independent jumps such that $\mathbb{E}(m_1^2) = 1$, see Remark 6 in Ref. [10].

APPLICATIONS

We present now some interesting applications of Theorem 1.

Let $(X_n,\ n \in \mathbb{Z})$ be a sequence of random variables with values in a Polish space $(E, \mathscr{E})$ on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Denote by σ-fields $\mathscr{F}_n^m = \sigma(X_k, m \le k \le n)$, $\mathscr{F}_n = \sigma(X_k, k \le n)$ and $\mathscr{F}^m = \sigma(X_k, k \ge m)$. Set

$$\phi(n) = \sup(|\mathbb{P}(B/A) - \mathbb{P}(B)|; A \in \mathscr{F}_k \quad \text{with}\, \mathbb{P}(A) > 0,\ B \in \mathscr{F}^{k+n}, k \in \mathbb{Z}).$$

The sequence $(X_n, n \in \mathbb{Z})$ is said to be ϕ-mixing if $\phi(n) \to 0$ as $n \to +\infty$.

If $(X_k, k \in \mathbb{Z})$ is a stationary process, we have the following well known inequality, which is a direct consequence of Ref. [21, Theorem A.6], $\forall i \ge 1$

$$\|\mathbb{E}_0 g(X_i) - \mu(g)\|_\infty \le 2\phi(i)\|g\|_\infty, \tag{3.1}$$

for every bounded g, where μ is the law of X_0.

Stationary Martingale Case

We begin with the stationary martingale differences. The following corollary is a quite pleasant extension of the Chen–Ledoux characterization of the MDP in the i.i.d. case.

COROLLARY 1 Let $(X_k,\ k \in \mathbb{Z})$ be a square integrable $\mathbb{R}^d$-valued stationary martingale differences, ϕ-mixing such that:

$$\exists C \text{ a positive constant such that } \mathbb{P}_0(X_1 \in \cdot) \le C\mu(\cdot), \tag{3.2}$$

where μ is the law of X_0. If

$$\limsup_{n \to +\infty} \frac{n}{b_n^2} \log(n\mathbb{P}(|X_0| > ab_n)) = -\infty, \quad \forall a > 0, \tag{3.3}$$

then the MDP of Theorem 1 holds.

Proof In view of Eq. (3.3) above and Theorem 1, Corollary 1 follows provided we can verify **(H1)** and **(H3)**. In order to check **(H1)**, we first recall the following result taken from [38, Lemma 2] □

LEMMA 1 Let $(\xi_k)_{k\in\mathbb{N}}$ be a sequence of real random variables adapted to the filtration $(\mathscr{F}_k)_{k\in\mathbb{N}}$ on $(\Omega, \mathscr{F}, \mathbb{P})$, such that $|\xi_k| \le C$, a.s. for all k. Let

$$L_{j,n} := \frac{1}{n} \sum_{k=j+1}^{j+n} \xi_k.$$

If there is some $z \in \mathbb{R}$ so that

$$\sup_{j\ge 0} \|\mathbb{E}_j(L_{j,n}) - z\|_{L^\infty(\mathbb{P})} \to 0,$$

as $n \to \infty$, then $L_{0,n}$ converges to z exponentially in probability, i.e.

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|L_{0,n}(\xi) - z\| > \delta) < 0, \quad \forall \delta > 0.$$

Set $\xi_k = E_{k-1}(X_k^2)$ (bounded, by condition (3.2)). In the notation of Lemma 1, we have that $\langle M\rangle_n/n = L_{0,n}$, so by the stationarity, for (**H1**) we have only to show that

$$\mathbb{E}_{-1}\left(\frac{\langle M\rangle_n}{n}\right) \to \mathbb{E}(X_0^2), \quad \text{as } n \to +\infty, \quad \text{in } L^\infty(\mathbb{P}). \tag{3.4}$$

It is not hard to see that for Eq. (3.4), it is enough to prove

$$\|\mathbb{E}_{-1}X_n^2 - \mu(X_0^2)\|_\infty \to 0, \quad \text{as } n \to \infty. \tag{3.5}$$

Let $A \geq 1$ given. By Eqs. (3.1) and (3.2), we have

$$\begin{aligned}
\|\mathbb{E}_{-1}X_n^2 - \mu(X_0^2)\|_\infty &\leq \|\mathbb{E}_{-1}X_n^2 \mathbb{I}_{\{|X_n|>A\}}\|_\infty + \|\mathbb{E}_{-1}X_n^2\mathbb{I}_{\{|X_n|\leq A\}} - \mu(X_0^2\mathbb{I}_{\{|X_0|\leq A\}})\|_\infty \\
&\quad + \mu(X_0^2\mathbb{I}_{\{|X_0|>A\}}) \\
&\leq (C+1)\mu(X_0^2\mathbb{I}_{\{|X_0|>A\}} + \|\mathbb{E}_{-1}X_n^2\mathbb{I}_{\{|X_n|\leq A\}} \\
&\quad - \mu(X_0^2\mathbb{I}_{\{|X_0|\leq A\}})\|_\infty \\
&\leq (C+1)\mu(X_0^2\mathbb{I}_{\{|X_0|>A\}}) + 2A^2\phi(n+1).
\end{aligned}$$

Since $(X_n)_{n\in\mathbb{Z}}$ is square integrable, we can, given ϵ, choose A so the first term in the right hand side (RHS) of the last inequality is less than $\epsilon/2$. By the ϕ-mixingness, we deduce that the second term in the RHS is also less than $\epsilon/2$ for all n large enough. So Eq. (3.5) follows.

Now it remains to verify (**H3**). Again using Eq. (3.2), we have $\forall a > 0$

$$\mathbb{E}_{k-1}(X_k^2 1_{\{|X_k|\geq \frac{an}{b_n}\}}) \leq C\mu(X_0^2\mathbb{I}_{\{|X_0|\geq\frac{an}{b_n}\}}),$$

and condition (**H3**) follows from the square integrability of X_0.

Stationary ϕ-mixing Sequences

Now we make use of Theorem 1 to establish a MDP for a certain class of stationary ϕ-mixing sequences. The basic idea is a representation for the increments of the stationary process in terms of the increments of a stationary martingale differences plus other terms whose sum is negligible in the sense of large deviation. This idea is due to Gordin [17], who has used it to prove a central limit theorem for stationary ergodic sequences and it has been widely developed, see Ref. [21].

Suppose that the process $(X_n, n \in \mathbb{Z})$ is stationary and ϕ-mixing such that

$$\sum_{n=1}^{+\infty} \sqrt{\phi(n)} < +\infty. \tag{3.6}$$

Given a real measurable function f such that $\mu(f) := \int f \mathrm{d}\mu = 0$, define for every $n \geq 1$, $t \in [0,1]$

$$S_n(t,f) = \sum_{k=0}^{[nt]} f(X_k), \quad S_n(f) := S_n(1,f).$$

COROLLARY 2 Let $f \in L^4(\mu)$. Assume that Eqs. (3.2) and (3.6) hold. If

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log(n\mathbb{P}(|f(X_0)| > ab_n)) = -\infty, \quad \forall a > 0, \tag{3.7}$$

then the processes $S_n(\cdot,f)/b_n$, as n goes to infinity, satisfies the moderate deviation principle on $\mathbb{D}[0, 1]$ with speed b_n^2/n and with the good rate function J^f : $\mathbb{D}[0,1] \to [0,+\infty]$ given by

$$J^f(\gamma) = \begin{cases} \frac{1}{2\sigma^2(f)} \int_0^1 \gamma'(t)^2 \mathrm{d}t & \text{if } \gamma \in AC_0([0,1]) \text{ and } \sigma^2(f) > 0 \\ +\infty & \text{otherwise,} \end{cases}$$

where $\sigma^2(f)$ is given by

$$\sigma^2(f) = \mathbb{E}(f(X_0)^2) + 2\sum_{k=1}^{+\infty} \mathbb{E}(f(X_0)f(X_k)). \tag{3.8}$$

Remarks

(1) Compared with the hyper-exponential convergence of $\phi(n)$ to 0 (i.e. $\phi(n)\mathrm{e}^{\lambda n} \to 0$, $\forall \lambda > 0$) required for the large deviations in the words of Bryc [5], the condition (3.6) for the MDP here is much more weaker, it is an old sufficient condition for the CLT see Corollary 5.5 in Ref. [21].
(2) The condition (3.7), being the best possible as is seen for the i.i.d. case, is weaker than the boundedness of f imposed in Gao [16], but our conditions (3.2) and (3.6) are stronger.
(3) We can get the vectorial version of the corollary: if $F = (f_1, \ldots, f_n) \in (L^4(\mu))^d$ such that $\mu(f_i) = 0$, then $S_n(\cdot, F)/b_n$ satisfies the moderate

deviation principle in $\mathbb{D}([0,1], \mathbb{R}^d)$ with speed b_n^2/n and with good rate function $J^F : \mathbb{D}([0,1], \mathbb{R}^d) \to [0, +\infty]$ given by

$$J^F(\gamma) = \begin{cases} \frac{1}{2}\int_0^1 \langle \sigma^{-2}(F)\gamma'(t), \gamma'(t)\rangle \mathrm{d}t & \text{if } \gamma \in AC_0([0,1], \mathbb{R}^d) \\ +\infty & \text{otherwise,} \end{cases}$$

where the matrix $\sigma^2(F) := (\sigma(f_i, f_j))$ is given by

$$\sigma^2(f_i, f_j) = \langle f_i, f_j\rangle + \sum_{k=1}^{+\infty} \mathbb{E}(f_i(X_0)f_j(X_k) + f_j(X_0)f_i(X_k)).$$

and $\sigma^{-2}(F)$ is the generalized inverse of $\sigma^2(F)$.

PROOF Without loss of generality, we assume $(X_n, n \in \mathbb{Z})$ is defined on $(\Omega, \mathscr{F}) = (E, \mathscr{E})^{\mathbb{Z}}$ as the system of coordinates and $\mathbb{P}$ is the law of $(X_n, n \in \mathbb{Z})$. Let θ the shift on Ω (i.e. $\theta\omega(i) = \omega(i+1)$), we assume also that $\mu(f) = 0$.

We recall that Eq. (3.1) holds for every bounded g. We have also, that for every integrable function g and $i \geq 1$,

$$\|\mathbb{E}_0 g(X_i) - \mu(g)\|_\infty \leq (C+1)\|g\|_1. \tag{3.9}$$

By the Riesz–Thorin interpolation theorem, we deduce that for every $g \in L^2(\mu)$ and $i \geq 1$,

$$\|\mathbb{E}_0 g(X_i) - \mu(g)\|_\infty \leq \sqrt{(C+1)}\sqrt{2\phi(i)}\|g\|_2. \tag{3.10}$$

Thus for every $f \in L^2(\mu)$ such that $\mu(f) = 0$, by Eqs. (3.6) and (3.10), we deduce that

$$\sum_{i=1}^{+\infty} \|\mathbb{E}_0 f(X_i)\|_\infty \leq \sqrt{C+1}\sqrt{2}\|f\|_2 \sum_{i=1}^{+\infty} \sqrt{\phi(i)} < +\infty.$$

Hence we can define the potential

$$Gf(\omega) := \sum_{k=0}^{+\infty} \mathbb{E}_0(f(X_k)),$$

which is absolutely convergent in $L^2(\mu)$. The series (Eq. 3.8) is then absolutely convergent. This last claim implies

$$\mathbb{E}\left(\frac{S_n^2(f)}{n}\right) \to \sigma^2(f), \quad \text{as } n \to +\infty. \tag{3.11}$$

The key to the proof is the following martingale decomposition (see Refs. [17,21])

$$S_n(f) = \sum_{k=0}^{n-1} f(X_k) = Gf(\omega) - Gf(\theta^n\omega) + M_n(f), \tag{3.12}$$

where

$$M_n(f) = \sum_{k=1}^{n} [Gf(\theta^k\omega) - \mathbb{E}_{k-1}(Gf(\theta^k\omega))] := \sum_{k=1}^{n} m_k, \tag{3.13}$$

is a $(\mathscr{F}_n)$-martingale with stationary differences (which is not ϕ-mixing in general, unlike the underlying process $(X_n, n \in \mathbb{Z})$). We have

$$\sigma^2(f) = \lim_{n\to+\infty} \mathbb{E}\left(\frac{S_n(f)^2}{n}\right) = \lim_{n\to\infty} \mathbb{E}\left(\frac{M_n(f)^2}{n}\right) = \mathbb{E}(M_1(f)^2).$$

Notice that

$$Gf(\omega) = f(X_0) + \sum_{j=1}^{+\infty} \mathbb{E}_0(f(X_j)) := f(X_0) + G_1(f), \tag{3.14}$$

where $G_1 f \in L^\infty(\mu)$ as proved precedently. By conditions (3.7) and (3.12), the moderate deviation of $S_n(f)/b_n$ is equivalent to that of $M_n(f)/b_n$, by means of the approximation Lemma. By Theorem 1, it remains to verify (**H1**) and (**H3**) for $(M_n(f))$, since (**H2**) is entailed by Eq. (3.7).

To verify (**H1**), we need to use again Lemma 1. Set $\xi_k = E_{k-1}(m_k^2)$. We have

$$m_k^2 = [Gf(\theta^k\omega) - \mathbb{E}_{k-1}(Gf(\theta^k\omega))]^2 \le 2Gf(\theta^k\omega)^2 + 2\|\mathbb{E}_{k-1}Gf(\theta^k\omega)\|_\infty^2.$$

Since there exist two positive constants C_1 and C_2 depending only on C and f such that

$$\|\mathbb{E}_{k-1}Gf(\theta^k\omega)\|_\infty^2 \le 2C_1^2;$$

and

$$Gf(\theta^k\omega)^2 \le 2f(X_k)^2 + 2C_2^2,\ a.s.$$

we deduce that for $\eta = \sqrt{C_1^2 + C_2^2}$, $m_k^2 \le 4f^2(X_k) + 4\eta^2$, and then

$$\mathbb{E}_{k-1}(m_k^2) \le 4\mathbb{E}_{k-1}(f^2(X_k) + \eta^2) \le 4C\mu((f^2 + \eta^2)).$$

So ξ_k is bounded.

In the notation of Lemma 1, we have that $\langle M(f)\rangle_n/n = L_{0,n}$, so by the stationarity, for (**H1**) we have only to show that

$$\mathbb{E}_{-1}\left(\frac{\langle M(f)\rangle_n}{n}\right) \to \sigma^2(f), \quad \text{as } n \to +\infty, \quad \text{in } L^\infty(\mathbb{P}). \tag{3.15}$$

By Eq. (3.12), we have

$$\mathbb{E}_{-1}(\langle M(f)\rangle_n) = \mathbb{E}_{-1}M_n(f)^2 = \mathbb{E}_{-1}(S_n(f) + Gf(\theta^n\omega) - Gf(\omega))^2.$$

Since $\sup_n \|\mathbb{E}_{-1}Gf(\omega)^2 + \mathbb{E}_{-1}(Gf(\theta^n\omega))^2\|_\infty < +\infty$, we deduce that Eq. (3.15) is equivalent to

$$\frac{\mathbb{E}_{-1}S_n(f)^2}{n} \to \sigma^2(f), \quad \text{in } L^\infty(\mathbb{P}).$$

By Eq. (3.11), we have to show that

$$\frac{\mathbb{E}_{-1}S_n(f)^2 - \mathbb{E}S_n(f)^2}{n} \to 0, \quad \text{in } L^\infty(\mathbb{P}). \tag{3.16}$$

Since

$$\begin{aligned}
&|\mathbb{E}_{-1}S_n(f)^2 - \mathbb{E}S_n(f)^2| \\
&\quad = \left|\sum_{k=0}^{n-1}\mathbb{E}_{-1}f(X_k)^2 - \mu(f^2) + 2\sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1}\mathbb{E}_{-1}(f(X_i)f(X_j)) - \mathbb{E}(f(X_i)f(X_j))\right| \\
&\quad \le \sum_{k=0}^{n-1}\|\mathbb{E}_{-1}f(X_k)^2 - \mu(f^2)\|_\infty \\
&\qquad + 2\sum_{i=0}^{n-1}\sum_{j=i+1}^{\infty}\|\mathbb{E}_{-1}(f(X_i)f(X_j)) - \mathbb{E}(f(X_i)f(X_j))\|_\infty.
\end{aligned}$$

For Eq. (3.16), it is enough to show

$$\|\mathbb{E}_{-1}f(X_n)^2 - \mu(f^2)\|_\infty \to 0, \quad \text{as } n \to \infty, \tag{3.17}$$

and

$$a_i := \sum_{j=i+1}^{\infty}\|\mathbb{E}_{-1}(f(X_i)f(X_j)) - \mathbb{E}(f(X_i)f(X_j))\|_\infty \to 0, \quad \text{as } i \to \infty. \tag{3.18}$$

The claim (3.17) follows obviously from the square integrability of f, condition (3.2) and the ϕ-mixingness, see the proof of Corollary 1.

It remains to show Eq. (3.18). To do that, we need the fact that $f \in L^4(\mu)$ rather then $f \in L^2(\mu)$.

$$\begin{aligned}
a_i &\leq \sum_{j=i+1}^{\infty} \sqrt{1+C}\sqrt{2\phi(i)}\|f(X_i)f(X_j)\|_2 \\
&= \sum_{j=i+1}^{\infty} \sqrt{1+C}\sqrt{2\phi(i)}\sqrt{\mathbb{E}f^2(X_i)\mathbb{E}_i f^2(X_j)} \\
&\leq \sum_{j=i+1}^{\infty} \sqrt{1+C}\sqrt{2\phi(i)}\|f(X_i)\|_2\|\mathbb{E}_i f^2(X_j)\|_\infty \\
&\leq 2(1+C)\|f\|_2\|f^2\|_2\sqrt{\phi(i)}\sum_{j=i+1}^{\infty}\sqrt{\phi(j-i)} \\
&\leq 2(1+C)\|f\|_2\|f^2\|_2\left(\sum_{j=1}^{\infty}\sqrt{\phi(j)}\right)\sqrt{\phi(i)} \longrightarrow 0,
\end{aligned}$$

as $i \longrightarrow +\infty$

So, we get the desired result. □

We turn now to verify the Lindeberg type condition (**H3**). Using the previous notations, we have

$$\begin{aligned}
\mathbb{E}_{k-1}\left(m_k^2 \mathbb{1}_{\{|m_k| \geq \frac{an}{b_n}\}}\right) &\leq 4\mathbb{E}_{k-1}\left((f^2(X_k)+\eta^2)\mathbb{1}_{\{|f(X_k)| \geq \frac{an}{2b_n}-\eta\}}\right) \\
&\leq 4C\mu\left((f^2+\eta^2)\mathbb{1}_{\{|f| \geq \frac{an}{2b_n}-\eta\}}\right),
\end{aligned}$$

where (**H3**) follows by the square integrability of f.

Stationary Markov Case

Now we apply Theorem 1 to establish the MDP for functionals of a Markov process satisfying Döeblin recurrence condition.

Let $(X_k, k \in \mathbb{Z})$ be a Markov process satisfying the Döeblin recurrence condition, (condition D_0 in Doob [15] pages 192 and 221). With this in mind, the n-step transition function $P^n(x,B)$ satisfies for some $\gamma > 0$ and $0 < \rho < 1$,

$$|P^n(x,B) - \mu(B)| \leq \gamma\rho^n,$$

uniformly in $x \in E$ and $B \in \mathscr{E}$, where μ is the unique stationary probability distribution of the Markov transition function. It is well known that Döeblin recurrence condition implies the ϕ-mixing condition with $\phi(n) = c\rho^n$ for some $c > 0$, see Ref. [23].

COROLLARY 3 Assume that Eqs. (3.2) and (3.7) hold. Suppose that $f \in L^2(\mu)$ in this context. Then the MDP in Corollary 2 holds.

The proof of this result is identical to that of Corollary 2, except that condition **(H1)** is more easier to check with the help of the following

Remark In the Gordin representation it is not clear that the martingale differences sequence $(m_k, k \in \mathbb{Z})$ (see Eqs. (3.12) and (3.13)) inherits the ϕ-mixing property of the original sequence $(X_n, n \in \mathbb{Z})$. In the case of the Markov process satisfying Döeblin recurrence condition, the Markov property helped in preserving ϕ-mixingness for $(m_k, k \in \mathbb{Z})$.

PROOF OF THEOREM 1

We now go to prove the main result of this paper. We need the following well known lemma stated in the real case in Puhalskii [27, Lemma 3.1]. We make the following convention: let $C(t)$ be an $\mathscr{S}(d^2)$-valued function, we say that $C(t)$ is a non-decreasing function if for $s \le t$, we have $u^* C(s)u \le u^* C(t)u$, for all $u \in \mathbb{R}^d$.

LEMMA 2 Let $A_n = (A_n(t), t \in [0,1])$, $A_n(0) = 0$ be an $\mathscr{S}(d^2)$-valued non-decreasing process on $(\Omega, \mathscr{F}, \mathbb{P})$ and $A = (A(t), t \in [0,1])$, $A(0) = 0$, be a deterministic $\mathscr{S}(d^2)$-valued non-decreasing continuous function. If for all $t \in [0,1]$ and for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(\|A_n(t) - A(t)\| > \delta) = -\infty,$$

then this convergence is uniform:

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\sup_{t\in[0,1]} \|A_n(t) - A(t)\| > \delta\right) = -\infty, \quad \forall \delta > 0.$$

Proof of Lemma 2 For $N > 0$, choose $t_i^N = \frac{i}{N}$, $0 \le i \le N$, a partition of [0, 1]. Since A_n and A are non-decreasing, for $t \in [t_{i-1}^N, t_i^N]$,

$$\|A_n(t) - A(t)\| \le \|A_n(t_i^N) - A(t_{i-1}^N)\| \vee \|A_n(t_{i-1}^N) - A(t_i^N)\|.$$

Hence, using that $A_n(0) = A(0) = 0$,

$$\sup_{t\in[0,1]}\|A_n(t) - A(t)\| \leq \sup_{1\leq i\leq N}\|A_n(t_i^N) - A(t_i^N)\| + \sup_{|u-v|\leq\frac{1}{N}u,v\leq 1}\|A(u) - A(v)\|.$$

By the continuity of A,

$$\sup_{|u-v|\leq\frac{1}{N}u,v\leq 1}\|A(u) - A(v)\| \underset{N\to\infty}{\longrightarrow} 0.$$

So for each $\delta > 0$ and for N large enough

$$\frac{n}{b_n^2}\log\mathbb{P}\left(\sup_{t\in[0,1]}\|A_n(t) - A(t)\| > \delta\right) \leq \frac{n}{b_n^2}\log\sum_{i=1}^{N}\mathbb{P}\left(\|A_n(t_i^N) - A(t_i^N)\| > \frac{\delta}{2}\right).$$

Hence

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\sup_{t\in[0,1]}\|A_n(t) - A(t)\| > \delta\right)$$

$$\leq \max_{i=1}^{N}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\|A_n(t_i^N) - A(t_i^N)\| > \frac{\delta}{2}\right).$$

The latter goes to $-\infty$ as $n\to\infty$, by the assumptions. The lemma is proved. □

Remark 4.1 This lemma gives the equivalence between (**H1**) and **sup** (**H1**), where

$$\sup(\mathbf{H1})\,\forall\delta > 0,\ \limsup_{n\to+\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\sup_{t\in[0,1]}\left\|\frac{\langle M\rangle_{[nt]}}{n} - tQ\right\| > \delta\right) = -\infty.$$

Proof of Theorem 1 Let (m_k) be a sequence of martingale differences satisfying conditions of Theorem 1. For $n = 1, 2, \ldots$ and $1 \leq k \leq n$, define:

$$x_{n,k} = m_k\mathbb{1}_{\{|m_k|\leq\frac{n}{b_n}\}} - \mathbb{E}_{k-1}\left(m_k\mathbb{1}_{\{|m_k|\leq\frac{n}{b_n}\}}\right),$$

$$y_{n,k} = m_k\mathbb{1}_{\{\frac{n}{b_n}<|m_k|\leq b_n\}} - \mathbb{E}_{k-1}\left(m_k\mathbb{1}_{\{\frac{n}{b_n}<|m_k|\leq b_n\}}\right),$$

$$w_{n,k} = m_k\mathbb{1}_{\{|m_k|>b_n\}} - \mathbb{E}_{k-1}\left(m_k\mathbb{1}_{\{|m_k|>b_n\}}\right),\quad X_n(t) = \sum_{k=1}^{[nt]}x_{n,k},$$

$$Y_n(t) = \sum_{k=1}^{[nt]}y_{n,k},\quad W_n(t) = \sum_{k=1}^{[nt]}w_{n,k}.$$

$(x_{n,k})$, $(y_{n,k})$, $(w_{n,k})$ are martingale differences arrays. Because (m_k) is a sequence of martingale differences,

$$m_k = x_{n,k} + y_{n,k} + w_{n,k}.$$

We control now each term of this decomposition, showing that only the first term contributes to the MDP, and the two other terms are negligible in the sense of moderate deviation. The proof is then divided into three steps.

Step 1: We shall show here that for any $\delta > 0$,

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}|W_n(t)| \geq \delta\right) = -\infty. \tag{4.1}$$

Since

$$\mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}|W_n(t)| \geq \delta\right) \leq \mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]} m_k \mathbb{1}_{\{|m_k|\geq b_n\}}\right| \geq \frac{\delta}{2}\right)$$
$$+ \mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]} \mathbb{E}_{k-1}(m_k \mathbb{1}_{\{|m_k|\geq b_n\}})\right| \geq \frac{\delta}{2}\right);$$

for Eq. (4.1), it is enough to show that for all $\delta > 0$,

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]} m_k \mathbb{1}_{\{|m_k|\geq b_n\}}\right| \geq \delta\right) = -\infty; \tag{4.2}$$

and

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]} \mathbb{E}_{k-1}(m_k \mathbb{1}_{\{|m_k|\geq b_n\}})\right| \geq \delta\right) = -\infty. \tag{4.3}$$

Noting

$$\left\{\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]} m_k \mathbb{1}_{\{|m_k|\geq b_n\}}\right| \geq \delta\right\} \subset \bigcup_{k=1}^{n} \{|m_k| \geq b_n\},$$

we see that

$$\mathbb{P}\left\{\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]} m_k \mathbb{1}_{\{|m_k|\geq b_n\}}\right| \geq \delta\right\} \leq \sum_{k=1}^{n} \mathbb{E}(\mathbb{1}_{\{|m_k|\geq b_n\}}) \leq \sum_{k=1}^{n} \mathbb{E}(\mathbb{P}_{k-1}(|m_k| \geq b_n))$$
$$\leq n \operatorname*{ess\,sup}_{1\leq k\leq n} \mathbb{P}_{k-1}(|m_k| \geq b_n).$$

Since $n \leq c^{-1}(b(n+1))$ (because $c(n) < b(n+1)$) for sufficiently large n, then Eq. (4.2) follows from condition (**H2**).

On the other hand, we have for all sufficiently large n,

$$\frac{1}{b_n}\sup_{t\in[0,1]}\left|\sum_{k=1}^{[nt]}\mathbb{E}_{k-1}(m_k\mathbb{1}_{\{|m_k|\geq b_n\}})\right| \leq \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{k-1}\left(|m_k|^2\mathbb{1}_{\{|m_k|\geq\frac{n}{b_n}\}}\right);$$

and then using condition (**H3**), we obtain Eq. (4.3).

Step 2: We will show that for each $\delta > 0$,

$$\limsup_{n\to+\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}|Y_n(t)|\geq\delta\right) = -\infty. \tag{4.4}$$

This step requires more efforts, and is the main new point of this paper.

Since $Y_n(t)$ is a bounded $\mathscr{F}_{[nt]}$-martingale, we deduce that $\exp(\lambda b_n Y_n(t)/n)$ is a sub-martingale. So by the maximal inequality, we have for any $\lambda > 0$,

$$\begin{aligned}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{1}{b_n}\sup_{t\in[0,1]}|Y_n(t)|\geq\delta\right) &= \frac{n}{b_n^2}\log\mathbb{P}\left(\exp\left(\lambda\frac{b_n}{n}\sup_{t\in[0,1]}|Y_n(t)|\right)\geq e^{\lambda\delta\frac{b_n^2}{n}}\right)\\ &\leq -\lambda\delta + \frac{n}{b_n^2}\log\mathbb{E}\left(\exp\left(\lambda\frac{b_n}{n}|Y_n(1)|\right)\right),\end{aligned}$$

and the last term above is

$$\begin{aligned}&\leq \frac{n}{b_n^2}\log\mathbb{E}\left(\exp\left(\lambda\frac{b_n}{n}\sum_{k=1}^{n}\left(|m_k|\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}} + \mathbb{E}_{k-1}|m_k|\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right)\right)\right)\\ &\leq \frac{n^2}{b_n^2}\log\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{E}_{k-1}\left(\exp\left(\lambda\frac{b_n}{n}|m_k|\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right)\right)\\ &\quad+\lambda\frac{n}{b_n}\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{E}_{k-1}\left(|m_k|\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right).\end{aligned}$$

Therefore for Eq. (4.4), it is enough to show that, for each $\lambda > 0$,

$$\limsup_{n\to\infty}\frac{n^2}{b_n^2}\log\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{E}_{k-1}\left(\exp\left(\lambda\frac{b_n}{n}|m_k|\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right)\right) = 0, \tag{4.5}$$

and

$$\limsup_{n\to\infty}\frac{n}{b_n}\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{E}_{k-1}\left(|m_k|\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right) = 0. \tag{4.6}$$

If we denote $J_n = [n/b_n, b_n]$, we have for all $\lambda > 0$

$$\begin{aligned}
&\mathbb{E}_{k-1}\left(\exp\left(\lambda\frac{b_n}{n}|m_k|\mathbb{1}_{\{|m_k|\in J_n\}}\right)\right)\\
&= \mathbb{E}_{k-1}\left(\exp\left(\lambda\frac{b_n}{n}|m_k|\mathbb{1}_{\{|m_k|\in J_n\}}\right)\mathbb{1}_{\{|m_k|\in J_n\}}\right.\\
&\qquad \left.+\exp\left(\lambda\frac{b_n}{n}|m_k|\mathbb{1}_{\{|m_k|\in J_n\}}\right)\mathbb{1}_{\{|m_k|\notin J_n\}}\right)\\
&= \mathbb{E}_{k-1}\left(\mathrm{e}^{\lambda\frac{b_n}{n}}\mathbb{1}_{\{|m_k|\in J_n\}} + 1 - \mathbb{1}_{\{|m_k|\in J_n\}}\right)\\
&= 1 + \mathbb{E}_{k-1}\left(\left(\mathrm{e}^{\lambda\frac{b_n}{n}|m_k|} - 1\right)\mathbb{1}_{\{|m_k|\in J_n\}}\right).
\end{aligned}$$

Since $\forall x \geq 0$, $\log(1+x) \leq x$, to prove Eq. (4.5), it is enough to show that

$$\limsup_{n\to\infty}\frac{n^2}{b_n^2}\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{E}_{k-1}\left(\left(\mathrm{e}^{\lambda\frac{b_n}{n}|m_k|} - 1\right)\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right) = 0. \tag{4.7}$$

Given $0 < R < \infty$, and letting $\mu_{k,\omega}(\cdot) = \mathbb{P}_{k-1}(|m_k| \in \cdot)$,

$$\begin{aligned}
&\frac{n^2}{b_n^2}\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{E}_{k-1}\left(\left(\mathrm{e}^{\lambda\frac{b_n}{n}|m_k|} - 1\right)\mathbb{1}_{\{\frac{n}{b_n}\leq|m_k|\leq b_n\}}\right)\\
&= \frac{n^2}{b_n^2}\operatorname*{ess\,sup}_{1\leq k\leq n}\int_0^\infty\left(\mathrm{e}^{\lambda\frac{b_n}{n}x} - 1\right)\mathbb{1}_{\{\frac{n}{b_n}\leq x\leq b_n\}}\mathrm{d}\mu_{k,\omega}(x)\\
&= \frac{n^2}{b_n^2}\operatorname*{ess\,sup}_{1\leq k\leq n}\int_0^\infty\left(\int_0^{\frac{\lambda b_n x}{n}}\mathrm{e}^u\mathrm{d}u\right)\mathbb{1}_{\{\frac{n}{b_n}\leq x\leq b_n\}}\mathrm{d}\mu_{k,\omega}(x)\\
&= \frac{n^2}{b_n^2}\operatorname*{ess\,sup}_{1\leq k\leq n}\int_0^\infty\int_0^\infty\mathrm{e}^u\mathbb{1}_{\{u\leq\frac{\lambda b_n x}{n}\}}\mathbb{1}_{\{\frac{n}{b_n}\leq x\leq b_n\}}\mathrm{d}\mu_{k,\omega}(x)\mathrm{d}u\\
&\leq \frac{n^2}{b_n^2}\int_0^\infty\mathrm{e}^u\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{P}_{k-1}\left(u\leq\frac{\lambda b_n}{n}|m_k|, \frac{n}{b_n}\leq|m_k|\leq b_n\right)\mathrm{d}u\\
&= \frac{n^2}{b_n^2}\int_0^{\frac{\lambda b_n^2}{n}}\mathrm{e}^u\operatorname*{ess\,sup}_{1\leq k\leq n}\mathbb{P}_{k-1}\left(u\leq\frac{\lambda b_n}{n}|m_k|, \frac{n}{b_n}\leq|m_k|\leq b_n\right)\mathrm{d}u
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{n^2}{b_n^2}\int_0^R \mathrm{e}^u \operatorname*{ess\,sup}_{1\leq k\leq n} \mathbb{P}_{k-1}\left(|m_k| \geq \frac{n}{b_n}\right)\mathrm{d}u \\
&\quad + \frac{n^2}{b_n^2}\int_R^{\frac{\lambda b_n^2}{n}} \mathrm{e}^u \operatorname*{ess\,sup}_{1\leq k\leq n} \mathbb{P}_{k-1}\left(|m_k| \geq \frac{nu}{b_n\lambda}\right)\mathrm{d}u \\
&= (\mathrm{e}^R - 1)\frac{n^2}{b_n^2} \operatorname*{ess\,sup}_{1\leq k\leq n} \mathbb{P}_{k-1}\left(|m_k| \geq \frac{n}{b_n}\right) \\
&\quad + \int_R^{\frac{\lambda b_n^2}{n}} \mathrm{e}^u \frac{n^2}{b_n^2} \operatorname*{ess\,sup}_{1\leq k\leq n} \mathbb{P}_{k-1}\left(|m_k| \geq \frac{nu}{b_n\lambda}\right)\mathrm{d}u \\
&:= A(R,n) + B(R,n).
\end{aligned}
$$

Let us first establish that $A(R,n) \to 0$, $\forall R > 1$, when n goes to infinity.

Let $L = L(n)$ such that $b_L \leq n/b_n < b_{L+1}$, we have that $L \to +\infty$, when $n \to +\infty$. We also have for n large $b_{L+1}^2 \leq b_L^2((L+1)/L)^2 \leq 4b_L^2$, by the non-decreasingness of n/b_n. Given $\beta > 1$, using condition (**H2**), and the fact that $n \leq c^{-1}(b(L+1))$, we have for all sufficiently large n,

$$
L \operatorname*{ess\,sup}_{1\leq k\leq n} \mathbb{P}_{k-1}(|m_k| > b_L) \leq L \operatorname*{ess\,sup}_{1\leq k\leq c^{-1}(b(L+1))} \mathbb{P}_{k-1}(|m_k| > b_L) \leq \mathrm{e}^{-\beta\frac{b_L^2}{L}}.
$$

Hence

$$
\begin{aligned}
A(R,n) &\leq (\mathrm{e}^R - 1)4b_L^2 \operatorname*{ess\,sup}_{1\leq k\leq c^{-1}(b(L+1))} \mathbb{P}_{k-1}(|m_k| \geq b_L) \leq (\mathrm{e}^R - 1)4\frac{b_L^2}{L}\,\mathrm{e}^{-\beta\frac{b_L^2}{L}} \\
&\leq (\mathrm{e}^R - 1)4\,\mathrm{e}^{-(\beta-1)\frac{b_L^2}{L}} \to 0,
\end{aligned}
$$

as n goes to infinity.

We now go to control $B(R, n)$. Let $N = N(n,u)$ be such that

$$
b_N \leq nu\lambda^{-1}b_n^{-1} < b_{N+1}.
$$

Observe that $N(n,u) \geq b^{-1}(uc(n)/\lambda) - 1 \to +\infty$, when $n \to \infty$, uniformly in $u \geq R$. So, given $\beta > 1 + 2\lambda$ and $R > \lambda$, for n large enough, using assumption (**H2**), and the fact that

$$
n \leq c^{-1}\left(\frac{\lambda}{u}b(N+1)\right) \leq c^{-1}(b(N+1)),
$$

we have

$$N \operatorname*{ess\,sup}_{1\le k\le n} \mathbb{P}_{k-1}(|m_k| > b_N) \le \mathrm{e}^{-\beta \frac{b_N^2}{N}}, \quad \forall u \ge R.$$

Hence, for all $u \ge R$,

$$\frac{n^2}{b_n^2} \operatorname*{ess\,sup}_{1\le k\le n} \mathbb{P}_{k-1}(|m_k| \ge nu\lambda^{-1}b_n^{-1})\mathrm{e}^u \le \lambda^2 u^{-2} b_{N+1}^2 \operatorname*{ess\,sup}_{1\le k\le n} \mathbb{P}_{k-1}(|m_k| \ge b_N)\mathrm{e}^u$$

$$\le \lambda^2 u^{-2} b_{N+1}^2 N^{-1}\, \mathrm{e}^{-\beta\frac{b_N^2}{N}+u}.$$

For n large enough, $b_{N+1}^2 \le 4b_N^2$, as noted before. Hence,

$$\lambda^2 u^{-2} N^{-1} b_{N+1}^2 \le \lambda^2 u^{-2} N^{-1} 4 b_N^2; \quad u \le b_{N+1}\lambda \frac{b_n}{n} \le 2 b_N \lambda \frac{b_n}{n} \le 2\lambda \frac{b_N^2}{N}.$$

So, we have for all sufficiently large n and for all $u \ge R > 1 \vee \lambda$,

$$\frac{n^2}{b_n^2} \operatorname*{ess\,sup}_{1\le k\le n} \mathbb{P}_{k-1}(|m_k| \ge nu\lambda^{-1}b_n^{-1})\mathrm{e}^u \le 4\lambda^2 \frac{b_N^2}{N}\, \mathrm{e}^{-\beta\frac{b_N^2}{N}+u} \le 4\lambda^2\, \mathrm{e}^{-(\beta-1)\frac{b_N^2}{N}+u}$$

$$\le 4\lambda^2\, \mathrm{e}^{-\left(\left(\frac{\beta-1}{2\lambda}\right)-1\right)u}.$$

Substituting that estimation in $B(R,n)$, we obtain

$$\limsup_{n\to\infty} \frac{n^2}{b_n^2} \int_R^{\lambda\frac{b_n^2}{n}} \operatorname*{ess\,sup}_{1\le k\le n} \mathbb{P}_{k-1}(|m_k| \ge nu\lambda^{-1}b_n^{-1})\mathrm{e}^u\, \mathrm{d}u \le \frac{4\lambda^2}{\frac{\beta-1}{2\lambda}-1}\, \mathrm{e}^{-\left(\frac{\beta-1}{2\lambda}-1\right)R}.$$

Since β is chosen so that $\beta > 2\lambda + 1$, we have

$$\lim_{R\to\infty} \limsup_{n\to\infty} B(R,n) = 0.$$

Hence Eq. (4.5) follows.

Since $x \le \mathrm{e}^x - 1$, $\forall x \ge 0$, Eq. (4.6) follows from Eq. (4.7).

Step 3: By the previous estimations, it suffices to show that

$$X_n(t) = \sum_{k=1}^{[nt]} x_{n,k} = \sum_{k=1}^{[nt]} \left(m_k \mathbb{I}_{\{|m_k| \le \frac{n}{b_n}\}} - \mathbb{E}_{k-1}\left(m_k \mathbb{I}_{\{|m_k| \le \frac{n}{b_n}\}} \right) \right),$$

satisfies the assumptions of the Proposition 1, to conclude the MDP. Using the fact that

$$\mathbb{E}_{k-1}\left(m_k \mathbb{I}_{\{|m_k| \le \frac{n}{b_n}\}} \right) = -\mathbb{E}_{k-1}\left(m_k \mathbb{I}_{\{|m_k| \ge \frac{n}{b_n}\}} \right), \tag{4.8}$$

we have

$$\mathbb{E}_{k-1}(x_{n,k}x_{n,k}^*) = \mathbb{E}_{k-1}(m_k m_k^*) - \mathbb{E}_{k-1}\left(m_k m_k^* \mathbb{I}_{\{|m_k|\geq \frac{n}{b_n}\}}\right) - \mathbb{E}_{k-1}\left(m_k \mathbb{I}_{\{|m_k|\geq \frac{n}{b_n}\}}\right)$$
$$\times \left(\mathbb{E}_{k-1}\left(m_k \mathbb{I}_{\{|m_k|\geq \frac{n}{b_n}\}}\right)\right)^*.$$

Hence

$$\sup_{t\in[0,1]}\left\|\frac{1}{n}\sum_{k=1}^{[nt]}\mathbb{E}_{k-1}(x_{n,k}x_{n,k}^*) - tQ\right\| \leq \sup_{t\in[0,1]}\left\|\frac{1}{n}\sum_{k=1}^{[nt]}\mathbb{E}_{k-1}(m_k m_k^*) - tQ\right\|$$
$$+\frac{2}{n}\sum_{k=1}^{n}\mathbb{E}_{k-1}\left(|m_k|^2\mathbb{I}_{\{|m_k|\geq \frac{n}{b_n}\}}\right).$$

By condition (**H1**) and (**H3**), we deduce that

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\sup_{t\in[0,1]}\left\|\frac{1}{n}\sum_{k=1}^{[nt]}\mathbb{E}_{k-1}(x_{n,k}x_{n,k}^*) - tQ\right\| \geq \delta\right) = -\infty, \quad \forall\delta > 0.$$

Then condition (i) of Proposition 1 is checked.

Condition (ii) is obviously satisfied since $|x_{n,k}| \leq n/b_n$, (with $c = 1$).

To check condition (iii) in Proposition 1, we need to prove that, for all $a > 0$ and for all $\delta > 0$

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{k-1}\left(|x_{n,k}|^2\mathbb{I}_{\{|x_{n,k}|\geq a\frac{n}{b_n}\}}\right) > \delta\right) = -\infty. \tag{4.9}$$

Let $a > 0$, using Eq. (4.8), we have

$$|x_{n,k}|^2\mathbb{I}_{\{|x_{n,k}|\geq a\frac{n}{b_n}\}} \leq 2|m_k|^2\mathbb{I}_{\{|m_k|\leq\frac{n}{b_n}\}}\mathbb{I}_{\{|x_{n,k}|\geq a\frac{n}{b_n}\}} + 2\left|\mathbb{E}_{k-1}\left(m_k\mathbb{I}_{\{|m_k|\geq\frac{n}{b_n}\}}\right)\right|^2$$
$$\leq 2|m_k|^2\mathbb{I}_{\{|m_k|\leq\frac{n}{b_n}\}}\mathbb{I}_{\{|x_{n,k}|\geq a\frac{n}{b_n}\}} + 2\mathbb{E}_{k-1}\left(|m_k|^2\mathbb{I}_{\{|m_k|\geq\frac{n}{b_n}\}}\right), \tag{4.10}$$

and denoting $V_{n,k}$ the first term at the right hand side of Eq. (4.10), we have

$$V_{n,k} \leq 2|m_k|^2\mathbb{I}_{\{|m_k|\geq a\frac{n}{2b_n}\}} + 2|m_k|^2\mathbb{I}_{\{|m_k|\leq\frac{n}{b_n}\}}\mathbb{I}_{\left\{\left|\mathbb{E}_{k-1}\left(m_k\mathbb{I}_{|m_k|\geq\frac{n}{b_n}}\right)\right|\geq a\frac{n}{2b_n}\right\}}$$
$$\leq 2|m_k|^2\mathbb{I}_{\{|m_k|\geq a\frac{n}{2b_n}\}} + \frac{8}{a^2}\left|\mathbb{E}_{k-1}\left(m_k\mathbb{I}_{|m_k|\geq\frac{n}{b_n}}\right)\right|^2$$
$$\leq 2|m_k|^2\mathbb{I}_{\{|m_k|\geq a\frac{n}{2b_n}\}} + \frac{8}{a^2}\mathbb{E}_{k-1}\left(|m_k|^2\mathbb{I}_{\{|m_k|\geq\frac{n}{b_n}\}}\right). \tag{4.11}$$

Combining the estimations (4.10) and (4.11), we obtain that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{k-1}\left(|x_{n,k}|^2 \mathbb{1}_{\{|x_{n,k}|\geq a\frac{n}{b_n}\}}\right) > \delta\right)$$

$$\leq \max\left\{ \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{k-1}\left(|m_k|^2 \mathbb{1}_{\{|m_k|\geq a\frac{n}{2b_n}\}}\right) > \frac{\delta}{4}\right),\right.$$

$$\left.\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{k-1}\left(|m_k|^2 \mathbb{1}_{\{|m_k|\geq \frac{n}{b_n}\}}\right) > \frac{1}{2(\frac{8}{a^2}+2)}\delta\right)\right\}.$$

Then Eq. (4.9) follows from condition (**H3**) and the estimations above.

Hence, we deduce from Proposition 1 that $X_n(\cdot)$ satisfies the MDP in $\mathbb{D}([0,1],\mathbb{R}^d)$ (equipped with the Skorohod topology), with speed b_n^2/n and good rate function given by Eq. (2.1). That finishes the proof of Theorem 1. □

REMARK 4.2 As it was kindly pointed out to us by Puhalskii, we can invoke Corollary 6.3 in Ref. [27] rather than Theorem 3.1. This Corollary requires his conditions: (0), (A) + (a), (MD), (sup A) and (C_0). Our proof can be seen as a verification of his conditions, which has some technical difficulties. But the essential motivation behind our work, is to attain Chen–Ledoux and Arcones [1] condition (**H2**), which is necessary and sufficient in the i.i.d. case.

Acknowledgements

I want to express my sincere thanks to my advisor Prof Liming WU for all his advices during the preparation of this work. I am indebted to Prof Puhalskii for his encouragements and careful attention to my paper. I am grateful to the anonymous referee for his valuable comments which improve the presentation of this work.

References

[1] Arcones, A. (1999). "The large deviation principle for stochastic processes", Submitted for publication.

[2] Baldi, P. (1988) "Large deviations and stochastic homogenization", *Ann. Mat. Pura Appl.* **151**, 161–177.

[3] Borovkov, A.A. and Mogulskii, A.A. (1978) "Probabilities of large deviations in topological vector space I", *Sib. Math. J.* **19**, 697–709.

[4] Borovkov, A.A. and Mogulskii, A.A. (1980) "Probabilities of large deviations in topological vector space II", *Sib. Math. J.* **21**, 12–26.

[5] Bryc, W. (1992) "On large deviations for uniformly strong mixing sequences", *Stoch. Proc. Appl.* **41**, 191–202.

[6] Chen, X. (1991) "The moderate deviations of independent vectors in Banach space", *Chin. J. Appl. Probab. Stat.* **7**, 124–132.

[7] Chen, X. (1999). "Limit theorems for functionals of ergodic Markov chains with general state space". *Memoirs of the AMS*, p. 139.

[8] de Acosta, A. (1997) "Moderate deviations for empirical measures of Markov chains: lower bound", *Ann. Probab.* **25**, 259–284.

[9] de Acosta, A. and Chen, X. (1998) "Moderate deviations for empirical measure of Markov chains: upper bound", *J. Theor. Probab.* **4**(11), 75–110.

[10] Dembo, A. (1996) "Moderate deviations for martingales with bounded jumps", *Elect. Comm. Probab.* **1**, 11–17.

[11] Dembo, A. and Zajic, T. (1997) "Uniform large and moderate deviations for functional empirical process", *Stoch. Proc. Appl.* **67**, 195–211.

[12] Dembo, A. and Zeitouni, O. (1998) Large Deviations Techniques and Applications, 2nd Ed. (Springer, New York).

[13] Djellout, H. and Guillin, A. (2001) "Moderate deviations for Markov chains with atom", *Stoch. Proc. Appl.*, In press.

[14] Djellout, H. and Guillin, A. (2000) "Principe de déviations modérées pour le processus empirique fonctionnel d'une chaîne de Markov", *CRAS* **330**, 377–380.

[15] Doob, J.L. (1953) Stochastic Processes (Wiley, New York).

[16] Gao, F.Q. (1996) "Moderate deviations for martingales and mixing random processes", *Stoch. Proc. Appl.* **61**, 263–275.

[17] Gordin, M.I. (1969) "The central limit theorem for stationary processes", *Sov. Math. Dokl.* **10**, 1174–1176.

[18] Grama, I. (1997) "On moderate deviation for martingales", *Ann. Probab.* **25**, 152–183.

[19] Grama, I. and Haeusler, E. (2000) "Large deviations for martingales via Cramér's method", *Stoch. Proc. Appl.* **85**, 279–293.

[20] Guillin, A. (2000) "Uniform moderate deviations of functional empirical processes of Markov chains", *Probab. Math. Statist.* **20**, 237–260.

[21] Hall, P. and Heyde, C.C. (1980) Martingale Limit Theory and its Application (Academic Press, New York).

[22] Jacod, J. and Shiryaev, A.N. (1987) Limit Theorems for Stochastic Processes (Springer, Berlin).

[23] Jain, N., Jogdeo, K. and Stout, W. (1975) "Upper and lower functions for martingales and mixing processes", *Ann. Probab.* **3**, 119–145.

[24] Khoshnevisan, D. (1996) "Deviation inequalities for continuous martingales", *Stoch. Proc. Appl.* **65**, 17–30.

[25] Ledoux, M. (1992) "Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi", *Ann. Inst. H. Poincaré* **28**, 267–280.

[26] Liptser, R.Sh. and Shiryaev, A.N. (1989) Theory of Martingales (Kluwer Academic Publishers, Dordrecht).

[27] Puhalskii, A. (1994) "Large deviations of semimartingales via convergence of the predictable characteristics", *Stoch. Stoch. Rep.* **49**, 27–85.

[28] Puhalskii, A. (1994) "The method of stochastic exponentials for large deviations", *Stoch. Proc. Appl.* **54**, 45–70.

[29] Puhalskii, A. (1997) "Large deviations of semimartingales: a maxingale problem approach. I. Limits as solutions to a maxingale problem", *Stoch. Stoch. Rep.* **61**, 141–243.

[30] Puhalskii, A. (1999) "Large deviations of semimartingales a maxingale problem approach. II. Uniqueness of the maxingale problem", *Appl. Stoch. Stoch. Rep.* **68**, 65–143.

[31] Rackauskas, A. (1990) "On probabilities of large deviations for martingales", *Leituvos Mathematikos Rinkinys*, 784–794.

[32] Rackauskas, A. (1995) "Large deviations for martingales with some applications", *Acta Appl. Math.* **38**, 109–129.

[33] Revuz, D. and Yor, M. (1994) Continuous Martingales and Brownian Motion (Springer, Berlin).

[34] Worms, J., "Principes de déviations modérées pour des martingales et applications statistiques" Thése de Doctoratde l'Université Marne-la-Vallée.

[35] Wu, L. (1994) "Large deviations, moderate deviations and LIL for empirical-processes", *Ann. Probab.* **22**, 17–27.
[36] Wu, L. (1995) "Moderate deviations for dependent random variables related to CLT", *Ann. Probab.* **23**, 420–445.
[37] Wu, L. (1996) "Moderate deviations for martingales and an application to uniformly mixing sequences", Preprint Laboratoire de Mathématiques Appliquées (Université Blaise Pascal.
[38] Wu, L. (1999) "Exponential convergence in probability for empirical means of Brownian motion and of random walks", *J. Theor. Probab.* **12**(3), 661–673.

ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE

H. DJELLOUT
A. GUILLIN

Large and moderate deviations for moving average processes

Annales de la faculté des sciences de Toulouse 6e série, tome 10, nº 1 (2001), p. 23-31.

<http://www.numdam.org/item?id=AFST_2001_6_10_1_23_0>

Annales de la Faculté des Sciences de Toulouse Vol. X, n° 1, 2001
pp. 23–31

Large and moderate deviations for moving average processes (*)

H. Djellout and A. Guillin [(1)]

Résumé. — Soit $X_n = \sum_{i\in\mathbb{Z}} a_{i+n}\omega_i$, $n \geqslant 1$, le processus à moyenne mobile, $(\omega_i)_{i\in\mathbb{Z}}$ étant une suite de v.a.i.i.d. réelles, et soit $S_n = (X_1 + \dots + X_n)$. Dans ce papier, nous établissons un principe de grandes déviations et de déviations modérées pour S_n/n, sous les conditions suivantes : ω_i sont bornées pour tout $i \in \mathbb{Z}$ et $\sum_{n\in\mathbb{Z}} a_n^2 < \infty$.

Abstract. — Let $X_n = \sum_{i\in\mathbb{Z}} a_{i+n}\omega_i$, $n \geqslant 1$, the moving average process, $(\omega_i)_{i\in\mathbb{Z}}$ i.i.d. real random values, and $S_n = (X_1 + \dots + X_n)$. In this note, we prove large and moderate deviations principle for S_n/n, under the boundedness of ω_i and $\sum_{n\in\mathbb{Z}} a_n^2 < \infty$.

1. Introduction.

Let $\{\omega_i,\ i \in \mathbb{Z}\}$ be a doubly infinite sequence of independent and identically distributed square integrable real random variables with $\mathbb{E}(\omega_i) = 0$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{a_n, n \in \mathbb{Z}\}$ be a doubly infinite sequence of real numbers such that $\sum_{i\in\mathbb{Z}} a_i^2 < \infty$.

The moving average process X_k, $k \geqslant 1$, is defined by

$$X_k = \sum_{i\in\mathbb{Z}} a_{i+k}\ \omega_i,$$

(*) Reçu le 8 avril 1999, accepté le 8 juin 2001
(1) Laboratoire de Mathématiques Appliquées, Université Blaise Pascal, 24 Avenue des Landais, 63177 Aubière.
email: {djellout, guillin}@math.univ-bpclermont.fr

H. Djellout and A. Guillin

and let

$$S_n = \sum_{k=1}^{n} X_k.$$

Numerous works have been made on the problem of large and moderate deviations of S_n/n under the strong condition $\sum_{i\in\mathbb{Z}} |a_i| < \infty$. For example, Burton and Dehling [BD90] have proved a large deviation principle for $\{S_n/n; n \geqslant 1\}$ with speed $\{n; n \geqslant 1\}$ and a good rate function depending only on the moment generating function. Their proof, like many of the referenced papers, relies on the powerful Ellis Theorem. The moderate deviations of $\{S_n/n; n \geqslant 1\}$ are obtained by Jiang and al. [JWR92] under the condition of exponential integrability of ω_0. Jiang and al. [JWR95] proved that the upper bound of large deviations for S_n in a Banach space B holds if and only if the condition $\mathbb{E}\left(e^{q_K(\omega)}\right) < \infty$ is fulfilled for some compact K of B, where q_K is the Minkowski functional of the set K. And the lower bound of large deviations is obtained in [JWR95] without any condition, with a rate function which may not have compact level sets, and which can be different from the rate function of the upper bound.

Remember also the famous work of Donsker and Varadhan [DV85], on large deviations of level-3 for stationary Gaussian processes, under moving average form, which has motivated our study.

In this note, we prove a large deviation principle and a moderate deviation principle for moving average processes, substituting the absolute convergence condition by the continuity of $g(\theta) = \sum_{n\in\mathbb{Z}} a_n e^{in\theta}$ at 0, a well known condition for the Central Limit Theorem of $\{S_n/\sqrt{n}\}$, see ([HH80], Corollary 5.2. pp 135). But we need the boundedness of ω_i instead of the exponential integrability in the works cited above.

2. A large deviation principle for the moving average processes.

About the language of large deviations, see Dembo and Zeitouni [DZ93], Deuschel and Stroock [DS89]. The main result of this paper is

THEOREM 2.1.— *Let $(\omega_i)_{i\in\mathbb{Z}}$ be a family of $\mathbb{P}$-i.i.d. real valued random variables. Suppose the following conditions*

(H1) $\sum_{i\in\mathbb{Z}} a_i^2 < \infty$, $\mathbb{E}(\omega_i) = 0$, $\mathbb{E}(\omega_i^2) = 1$ *and* $|\omega_i| \leqslant C$ *for all* $i \in \mathbb{Z}$.

(H2) *The function g given by $|g(\theta)|^2 := \left|\sum_n a_n e^{in\theta}\right|^2 = f(\theta)$ (the spectral density function of X_k) is continuous at 0 and belongs to $\mathrm{L}^2([-\pi,\pi], d\theta)$.*

Then $\mathbb{P}\left(\frac{S_n}{n} \in \cdot\right)$ satisfies a large deviation principle with speed n and the good rate function I given by

$$I(x) = \Lambda^*\left(\frac{x}{g(0)}\right), \quad \forall x \in \mathbb{R}, \tag{2.1}$$

where Λ^* is the Fenchel-Legendre transform of the logarithmic moment generating function $\Lambda(\lambda) := \log \mathbb{E}e^{\lambda\omega_0}$ of the common law of ω_i, $i \in \mathbb{Z}$.

Remark 2.i. — Except the boundedness of ω_i, the assumption **(H1)** is minimal to define X_k.

Remark 2.ii. — Condition **(H2)** is the usual condition for the Central Limit Theorem for S_n, see [HH80]. Notice also that it is much weaker than the condition $f(\theta) \in \mathrm{C}([-\pi,\pi])$ used in the pioneering work of Donsker-Varadhan [DV85] for the level-3 large deviations of stationary gaussian processes.

To prove Theorem 2.1 we need the following concentration inequality for S_n, which is a translation of the well known Hoeffding inequality [Hoe63] in our context.

LEMMA 2.2. — *Under condition* **(H1)**, *we have*

$$\mathbb{P}\left(|S_n| \geqslant t\right) \leqslant 2\, e^{-\frac{t^2}{2C^2\mathbb{E}(S_n^2)}}. \tag{2.2}$$

Proof of Lemma 2.2. — By Hoeffding inequality [Hoe63] (or more exactly its proof), applied to

$$\tilde{S}_n^K = \sum_{|i| \leqslant K} \tilde{X}_i^n$$

where $\tilde{X}_i^n = \sum_{k=1}^n a_{i+k}\omega_i$, we have for all $\lambda \geqslant 0$

$$\begin{aligned}
\mathbb{E}\left(e^{\lambda \tilde{S}_n^K}\right) &\leqslant \exp\left(\frac{\lambda^2 C^2}{2} \sum_{|i| \leqslant K} \left(\sum_{k=1}^n a_{i+k}\right)^2\right) \\
&= \exp\left(\frac{\lambda^2 C^2}{2} \mathbb{E}(\tilde{S}_n^K)^2\right).
\end{aligned}$$

Letting $K \to \infty$, $\tilde{S}_n^K \longrightarrow S_n$ in $L^2(\mathbb{P})$ and we get by Fatou's lemma:

$$\mathbb{E}(e^{\lambda S_n}) \leqslant e^{\frac{\lambda^2 C^2}{2} \mathbf{E} S_n^2}.$$

Then by Markov inequality, we have that for all $t > 0$

$$\mathbb{P}(S_n \geqslant t) \leqslant e^{-t\lambda + \frac{\lambda^2 C^2}{2} \mathbf{E} S_n^2}, \ \forall \lambda \geqslant 0,$$

and optimizing in λ, we obtain

$$\forall t \geqslant 0, \quad \mathbb{P}\left(S_n \geqslant t\right) \leqslant e^{-\frac{t^2}{2C^2 \mathbf{E} S_n^2}}.$$

We have obviously the same inequality for $-S_n$. Thus (2.2) follows.

Remark 2.iii. — Inequality (2.2) can be proved by means of logarithmic Sobolev inequality for convex functions [Led96] (with less better constant), in a way which is also valid for $|S_n|$ and thus give an alternate way to establish *Step 2* in the following proof.

Proof of Theorem 2.1 : we separate its proof into three steps.

Step 1. Let $S_n^K = \sum_{k=1}^{n} X_k^K$ where we have for some fixed K in $\mathbb{N}$

$$X_k^K = \sum_{|j|<K} a_{j+k} \left(1 - \frac{|j|}{K}\right) \omega_j.$$

Then $\mathbb{P}\left(\frac{S_n^K}{n} \in \cdot\right)$ satisfies the large deviation principle with speed n and some good rate function I^K, by Sanov's theorem and the contraction principle, or using results of [BD90] which give I^K with the same notations as in Theorem 2.1:

$$I^K(x) = \Lambda^* \left(\frac{x}{\sum_{|j|<K} a_j \left(1 - \frac{|j|}{K}\right)} \right).$$

Step 2. We show that for all $\delta > 0$

$$\limsup_{K\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left| \frac{S_n}{n} - \frac{S_n^K}{n} \right| > \delta \right) = -\infty. \tag{2.3}$$

By our hypothesis, we can apply Hoeffding inequality for $S_n - S_n^K$, and noting that $\mathbb{E}(S_n - S_n^K) = 0$ we get by lemma 2.2

$$\frac{1}{n}\log \mathbb{P}\left(\frac{1}{n}\left|S_n - S_n^K\right| > \delta\right) \leqslant \frac{\log 2}{n} - \frac{n\delta^2}{2C^2\mathbb{E}\left((S_n - S_n^K)^2\right)}$$

We have now to control the right term of this inequality. Let f_K denote the spectral density of X_n^K, i.e. $f_K(\theta) = |g_K(\theta)|^2 := \left|\sum_{|j|<K} a_j \left(1 - \frac{|j|}{K}\right) e^{ij\theta}\right|^2$.

Let introduce Fejer's kernel F_K

$$F_K(\theta) = \frac{1}{2\pi K}\left(\frac{\sin \frac{K}{2}\theta}{\sin \frac{1}{2}\theta}\right)^2.$$

An obvious property is that $\int_{-\pi}^{\pi} F_K(\theta)d\theta = 1$. Moreover, we have $g_K = F_K * g$, where "*" denotes the usual convolution product. By the well known theorem ([IL71], Theorem 18.2.1. pp322), we have

$$\frac{\mathbb{E}\left((S_n - S_n^K)^2\right)}{n} = 2\pi\ \left[|g - g_K|^2 * F_n\right](0). \tag{2.4}$$

For any $\varepsilon > 0$, let $[-\delta, \delta]$ be such that $|g(\theta) - g(0)| < \frac{\varepsilon}{2}$ for $|\theta| < \delta$. For $|\theta| < \frac{\delta}{2}$

$$\begin{aligned}
|g_K(\theta) - g(\theta)| &= |g * F_K(\theta) - g(\theta)| \\
&= \left|\int_{-\pi}^{\pi} F_K(\theta')g(\theta - \theta')d\theta' - g(\theta)\right| \\
&\leqslant \left|\int_{|\theta'|<\delta/2} F_K(\theta')\left(g(\theta - \theta') - g(\theta)\right)d\theta'\right| \\
&\qquad + \left|\int_{|\theta'|\geqslant\delta/2} F_K(\theta')\left(g(\theta - \theta') - g(\theta)\right)d\theta'\right| \\
&\leqslant \varepsilon\int_{|\theta'|<\delta/2} F_K(\theta')d\theta' + C(K,\delta)2\pi\left(\|g\|_{L^2(d\theta')} + |g(\theta)|\right) \\
&\leqslant \varepsilon\int_{|\theta'|<\delta/2} F_K(\theta')d\theta' + C(K,\delta)2\pi\left(\|g\|_{L^2(d\theta')} + |g(0)| + \frac{\varepsilon}{2}\right),
\end{aligned}$$

where $C(K, \delta) = \sup_{|\theta'|\geqslant\delta/2} |F_K(\theta')| \leqslant (2\pi K \sin^2(\delta/4))^{-1} \to 0$ as $K \to \infty$. We can now control the right hand side of (2.4)

$$\limsup_{K\to\infty}\left[|g - g_K|^2 * F_n\right](0) = \limsup_{K\to\infty}\int_{-\pi}^{\pi} |g - g_K|^2(\theta)F_n(\theta)d\theta$$

$$\leqslant \limsup_{K\to\infty}\left(\sup_{|\theta|<\delta/2}|g-g_K|^2(\theta)+C(n,\delta)\int_{|\theta'|\geqslant\delta/2}|g-g_K|^2(\theta)d\theta\right)$$
$$\leqslant \varepsilon.$$

We further conclude that

$$\limsup_{K\to\infty}\limsup_{n\to\infty}\frac{\mathbb{E}\left(\left(S_n-S_n^K\right)^2\right)}{n}=0.$$

We then deduce (2.3). Applying the approximation lemma ([DZ93], Th. 4.2.16.), we obtain that S_n satisfies the large deviations principle with speed n and the rate function

$$\tilde{I}(x):=\sup_{\delta>0}\liminf_{K\to\infty}\inf_{y\in B(x,\delta)}I^K(y)=\sup_{\delta>0}\limsup_{K\to\infty}\inf_{y\in B(x,\delta)}I^K(y)$$

where $B(x,\delta)$ is the ball of radius δ centered at x.

Step 3. It remains to show that $\tilde{I}(x)=I(x)$, where I is given by (2.1). We will first prove that $\tilde{I}(x)\leqslant I(x)$. Assume that $I(x)<\infty$ (trivial otherwise). This inequality is obvious for $x=0$ (as $I^K(0)=0$). Now for $x\neq 0$, the finiteness of $I(x)$ implies that $g(0)\neq 0$. For each $\delta\geqslant 0$, we have by the convergence of $g_K(0)$ to $g(0)$ that $yg(0)\in B(x,\delta)$ implies $yg_K(0)\in B(x,2\delta)$ for sufficiently large K, so that we have

$$\inf_{z\in B(x,\delta)}\{\Lambda^*(y),\ yg(0)=z\}\geqslant\inf_{z\in B(x,2\delta)}\{\Lambda^*(y),\ yg_K(0)=z\},$$

which yields $\tilde{I}(x)\leqslant I(x)$.

We now have to prove the converse inequality. Assume at first $g(0)\neq 0$, by the lower semi-continuity of I (inf-compact in reality), we have

$$\tilde{I}(x)=\sup_{\delta>0}\limsup_{K\to\infty}\inf_{y\in B(x,\delta)}\Lambda^*\left(\frac{y}{g_K(0)}\right)\geqslant\liminf_{z\to\frac{x}{g(0)}}\Lambda^*(z)\geqslant\Lambda^*\left(\frac{x}{g(0)}\right).$$

Now assume $g(0)=0$. $\tilde{I}(0)\geqslant I(0)$ (trivial). For $x\neq 0$,

$$\tilde{I}(x)=\sup_{\delta>0}\limsup_{K\to\infty}\inf_{y\in B(x,\delta)}\Lambda^*\left(\frac{y}{g_K(0)}\right)\geqslant\liminf_{|z|\to+\infty}\Lambda^*(z)=+\infty=I(x).$$

So we have that $\tilde{I}(x)=I(x)$, which ends the proof of theorem 2.1.

Remarks. — Under the boundedness "$|\omega_i|\leqslant C$" and the strong condition $\sum_{i\in\mathbb{Z}}|a_i|<\infty$, the level-3 large deviations principle for $(X_n)_{n\in\mathbb{Z}}$ holds.

Indeed, assume without loss of generality that (ω_i) is the coordinates system on the product space $\Omega_C = [-C, C]^{\mathbb{Z}}$ equipped with the product measure $\mathbb{P} = \mu^{\otimes \mathbb{Z}}$, where μ is the common law of ω_i. Let θ_k be the shift operator acting on Ω_C, defined by $(\theta_k \omega)_l = \omega_{l-k}$, $\forall k, l \in \mathbb{Z}$, and let E_n be the empirical process of the i.i.d. sequence, defined on the space $\mathcal{P}(\Omega_C)$ of all probability measures on Ω_C:

$$E_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{\theta_k \omega}.$$

By the results of Donsker-Vardahan [DV85], E_n satisfies a level-3 large deviation principle on $\mathcal{P}(\Omega_C)$ equipped with the weak convergence topology, with speed n and the good rate function given by the Donsker-Varadhan level-3 entropy $H(Q)$, see [DV85] for some details on $H(Q)$. Let ϕ be the map given by

$$\begin{aligned} \phi \quad : \quad & \Omega \rightarrow \mathbb{R}^{\mathbb{Z}} \\ & \omega \rightarrow \left(\sum_{i \in \mathbb{Z}} a_{i+k} \omega_i \right)_{k \in \mathbb{Z}} . \end{aligned}$$

By the absolute summability $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$, ϕ is continuous from Ω_C to $\mathbb{R}^{\mathbb{Z}}$ both equipped with product topology. Let $\mathcal{P}(\mathbb{R}^{\mathbb{Z}})$ be the space of all probability measures on $\mathbb{R}^{\mathbb{Z}}$ equipped with the weak convergence topology. Define on $\mathcal{P}(\mathbb{R}^{\mathbb{Z}})$ the empirical measure

$$R_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_{\cdot + k}}.$$

We obviously have $R_n = E_n \circ \phi^{-1}$. By the contraction principle, we conclude that R_n satisfies a level-3 large deviation principle on $\mathcal{P}(\mathbb{R}^{\mathbb{Z}})$ with speed n and the good rate function $I(Q) = \inf \left\{ H(\tilde{Q}); Q = \tilde{Q} \circ \phi^{-1} \right\}$.

3. Moderate deviations.

We are now studying moderate deviations for S_n in the same way as we have proved large deviations in the preceding section, we keep the same conditions on a_i and ω_i. To this purpose, let $(b_n)_{n \geqslant 1}$ be a sequence of positive numbers such that

$$b_n \rightarrow +\infty \quad , \quad \frac{b_n}{\sqrt{n}} \rightarrow 0. \tag{3.1}$$

H. Djellout and A. Guillin

THEOREM 3.1. — *Under the conditions* **(H1)** *and* **(H2)**, $\mathbb{P}\left(\frac{S_n}{b_n\sqrt{n}} \in \cdot\right)$ *satisfies a large deviation principle with the speed* b_n^2 *and the good rate function* I_M *given by*

$$\forall x \in \mathbb{R}, \qquad I_M(x) = \frac{x^2}{2f(0)}. \tag{3.2}$$

Proof. — We separate its proof into two steps.

Step 1. Let $S_n^K = \sum_{k=1}^{n} X_k^K$ as before.

Then, by [JWR92], $\mathbb{P}\left(\frac{S_n^K}{b_n\sqrt{n}} \in \cdot\right)$ satisfies a large deviation principle with speed b_n^2 and the good rate function I_M^K given by

$$I_M^K(x) = \frac{x^2}{2\left(\sum_{|j|<K} a_j\left(1 - \frac{|j|}{K}\right)\right)^2}.$$

Step 2. Since $S_n - S_n^K$ satisfies assumptions of lemma 2.2, we apply the concentration inequality (2.2) to $\frac{S_n - S_n^K}{b_n\sqrt{n}}$: $\forall t \geqslant 0$,

$$\frac{1}{b_n^2}\log \mathbb{P}\left(\left|\frac{S_n - S_n^K}{b_n\sqrt{n}}\right| \geqslant t\right) \leqslant -\frac{n\, t^2}{2C^2\mathbb{E}\left((S_n - S_n^K)^2\right)} + \frac{\log 2}{b_n^2}.$$

But by the proof of Theorem 2.1 we have

$$\limsup_{K\to\infty}\limsup_{n\to\infty}\frac{\mathbb{E}\left((S_n - S_n^K)^2\right)}{n} = 0$$

where it follows

$$\limsup_{K\to\infty}\limsup_{n\to\infty}\frac{1}{b_n^2}\log \mathbb{P}\left(\left|\frac{S_n - S_n^K}{b_n\sqrt{n}}\right| \geqslant t\right) = -\infty.$$

According to the approximation lemma ([DZ93], Th. 4.2.16.), we deduce that S_n satisfies the moderate deviations principle with speed b_n^2 and the rate function

$$\tilde{I}_M(x) := \sup_{\delta>0}\liminf_{K\to\infty}\inf_{y\in B(x,\delta)} I_M^K(y) = \sup_{\delta>0}\limsup_{K\to\infty}\inf_{y\in B(x,\delta)} I_M^K(y).$$

The identification of the rate function is done like in *Step3* of the proof of theorem 2.1.

Acknowledgment. — The authors thank Pr. Liming Wu for fruitful discussions and advices. We are grateful to the anonymous referee for his valuable comments which improve the presentation of this work.

Bibliography

[BD90] BURTON (R.M.) and DEHLING (H.). — Large deviations for some weakly dependent random processes. *Statistics and Probability Letters*, 9:397–401, 1990.

[DS89] DEUSCHEL (J.D.) and STROOCK (D.W.). — *Large deviations.* Academic Press, Boston, 1989.

[DV85] DONSKER (M.D.) and VARADHAN (S.R.S.). — Large deviations for stationary gaussian processes. *Communications in Mathematical Physics*, 97:187–210, 1985.

[DZ93] DEMBO (A.) and ZEITOUNI (O.). — *Large deviations techniques and their applications.* Jones and Bartlett, Boston, MA, 1993.

[HH80] HALL (P.) and HEYDE (C.C.). — *Martingale limit theory and its application.* Academic Press, New York, 1980.

[Hoe63] HOEFFDING (W.). — Probability inequalities for sums of bounded random variables. *American Statistical Association Journal*, pages 13–30, 1963.

[IL71] IBRAGIMOV (I.A.) and LINNIK (Y.V.). — *Independent and stationnary sequences of random variables.* Wolters-Noordhoff Publishing, 1971.

[JWR92] JIANG (T.), WANG (X.) and RAO (M.B.). — Moderate deviations for some weakly dependent random processes. *Statistics and Probability Letters*, 15:71–76, 1992.

[JWR95] JIANG (T.), WANG (X.) and RAO (M.B.). — Large deviations for moving average processes. *Stochastic Processes and Their Applications*, 59:309–320, 1995.

[Led96] LEDOUX (M.). — On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1996.

[Led99] LEDOUX (M.). — Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probab. XXXIII, LNM Springer*, 1709:120–216, 1999.

ELSEVIER Stochastic Processes and their Applications 95 (2001) 203–217

stochastic processes and their applications

www.elsevier.com/locate/spa

Moderate deviations for Markov chains with atom

H. Djellout, A. Guillin *

Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620, Université Blaise Pascal, 24 Avenue des Landais, 63177 Aubière, France

Received 1 February 2001; received in revised form 10 April 2001; accepted 11 April 2001

Abstract

We obtain in this paper moderate deviations for functional empirical processes of general state space valued Markov chains with atom under weak conditions: a tail condition on the first time of return to the atom, and usual conditions on the class of functions. Our proofs rely on the regeneration method and sharp conditions issued of moderate deviations of independent random variables. We prove our result in the nonseparable case for additive and unbounded functionals of Markov chains, extending the work of de Acosta and Chen (J. Theoret. Probab. (1998) 75–110) and Wu (Ann. Probab. (1995) 420–445). One may regard it as the analog for the Markov chains of the beautiful characterization of moderate deviations for i.i.d. case of Ledoux 1992. Some applications to Markov chains with a countable state space are considered. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary 60F10

Keywords: Moderate deviations; Markov chains; Regeneration chain method; Functional empirical processes; Countable state space

1. Introduction and main result

Let $(E,\mathscr{E})$ be a measurable space and $\mathscr{M}(E)$ be the space of all finite signed measures on $(E,\mathscr{E})$ equipped with the total variation norm $\|\cdot\|_{\mathrm{var}}$. Let $\{X_j\}_{j\geqslant 0}$ be an E-valued irreducible ergodic Markov chain with transition probability P and invariant probability measure π. Throughout the paper, we assume that the chain $\{X_j\}$ has an atom, i.e. $\exists\alpha\subset E$ with $\pi(\alpha)>0$, v a probability measure such that

$$\forall x\in\alpha,\qquad P(x,\cdot)=v(\cdot), \tag{1.1}$$

α is then called a atom. Note that, when the state space is discrete, every state charged by π is an atom. We introduce the first time of entrance of the chain in this atom

* Corresponding author. Tel.: +33-4-73-40-70-50; fax: +33-4-73-40-70-64.

E-mail addresses: djellout@math.univ-bpclermont.fr (H. Djellout), arnaud.guillin@math.univ-bpclermont.fr (A. Guillin).

which will play an important role in the study:

$$\tau = \inf\{n \geqslant 0;\ X_n \in \alpha\},$$

and we will always assume that $\mathbb{E}_\nu \tau^2 < \infty$.

Given a probability measure μ on $(E,\mathscr{E})$, $\mathbb{P}_\mu$ will be the Markovian probability measure on $(E^{\mathbb{N}}, \mathscr{E}^{\otimes\mathbb{N}})$ determined by the transition probability P and the initial law μ. $\{X_j\}_{j\geqslant 0}$ will be then the sequence of coordinates on $E^{\mathbb{N}}$.

Let M_n, $n \geqslant 1$, be random elements of $\mathscr{M}(E)$ defined by

$$M_n := \frac{1}{b_n}\sum_{j=0}^{n-1}(\delta_{X_j} - \pi), \tag{1.2}$$

where b_n is a sequence of positive numbers tending to infinity. We are interested in this paper in the asymptotic behaviour of $\mathbb{P}_\mu(M_n \in \cdot)$.

When $b_n = \sqrt{n}$, it is the Central Limit Theorem obtained first by Nummelin (1978) and Chen (1997) under various conditions. If $b_n = n$, it is the large deviations case extensively studied since the pioneering works of Donsker–Varadhan (see for instance Deuschel and Stroock, 1989; Wu, 1993 for a survey on this topic).

Now assume,

$$\frac{b_n}{\sqrt{n}} \uparrow +\infty, \qquad \frac{b_n}{n} \to 0. \tag{1.3}$$

The estimation of the probabilities $\mathbb{P}_\mu(M_n \in \cdot)$ is usually called the moderate deviation problem. We will suppose moreover the following: $\exists A > 1,\ 0 < \delta < 1$ such that

$$\forall n,k \geqslant 1, \qquad b_{nk} \leqslant A k^{1-\delta} b_n. \tag{1.4}$$

It is the usual condition on the speed of moderate deviations in the i.i.d. case (Ledoux, 1992), it means that b_n cannot be too near of n (the scale of large deviations). Sharp results on moderate deviation are quite recent, even for the i.i.d. case: the works of Ledoux (1992) for the upper bound in Banach space (which are largely used in this paper) and results of Wu (1994) for the functional empirical process (nonseparable Banach space case). See also Djellout (2000) for the extension to the martingale differences case and applications to mixing sequences.

The Markovian case has been studied under successively less restrictive conditions (Mogulskii (1984), Gao (2000), Wu (1994)) and recently under weak conditions by de Acosta (1988a,b) and Chen (1997) for the lower bound (under different and non-comparable conditions) and by de Acosta and Chen (1998), and Chen (1997) (under same conditions but different proof) for the upper bound. de Acosta and Chen (1998) have established their results under the assumptions of geometric ergodicity and a regularity condition (de Acosta and Chen, 1998, assumption (1.5)). Very recently, Guillin (2000) extends their results to the uniform trajectorial case, and Guillin, 2001 for Markov processes (continuously indexed).

We will be interested here by the asymptotic behaviour of M_n uniformly over a class of function (context of Wu, 1994).

Given a class of real measurable functions $\mathscr{F}$ such that $\forall f \in \mathscr{F}$, $\pi(f) = 0$, $f \in L_2(\pi)$ and $\mathbb{E}_\nu(\sum_{j=1}^{\tau}\|f(X_j)\|)^2 < \infty$, let $l_\infty(\mathscr{F})$ be the space of all bounded real functions on $\mathscr{F}$ with norm $\|F\|_{\mathscr{F}} = \sup_{f\in\mathscr{F}}|F(f)|$.

If $\mathscr{F}$ is infinite, $l_\infty(\mathscr{F})$ is a nonseparable Banach space. Every $\beta \in \mathscr{M}(E)$ can be regarded as an element $\beta^{\mathscr{F}} \in l_\infty(\mathscr{F})$ given by $\beta^{\mathscr{F}}(f) = \beta(f) = \int_E f \, d\beta$. We will now establish the moderate deviations estimations of $(M_n)^{\mathscr{F}}$ in $l_\infty(\mathscr{F})$.

In the sequel, we will suppose that $\mathscr{F}$ is countable, or that the processes $\{M_n(f); f \in \mathscr{F}\}$ are separable in the sense of Doob, to avoid measurability problems. Let d_2 be the following metric for $\mathscr{F}$: $\forall f, g \in \mathscr{F}$:

$$d_2(f,g) = \sigma(f-g),$$

where

$$\sigma^2(f) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\sum_{k=0}^{n-1} f(X_k)\right)^2 = \pi(\alpha)\mathbb{E}_\nu\left(\sum_{j=1}^{\tau} f(X_j)\right)^2$$

is the associated variance.

For an irreducible Markov chain taking integer values which has a finite second moment for the first return time from some integer to itself, Levental (1990) find necessary and sufficient conditions for the uniform CLT over all subsets of the integers. Tsai (1997) generalized this result to unbounded classes, $\mathscr{F} = \{f\colon |f| \leqslant F\}$, where F is a non-negative function, say the envelope function, on the countable state space. Tsai (2000) gives sufficient and nearly necessary conditions (weaker than condition of the uniform CLT) for the compact and bounded law of the iterated logarithm for Markov chains with a countable state space.

We will first give the moderate deviation principle in the general framework where an atom is present, and then present some applications on a countable state space, where some conditions can be more explicit.

Here is our main result:

Theorem 1. *Suppose that $(\mathscr{F}, d_2)$ is totally bounded and $(M_n)^{\mathscr{F}} \to 0$ in probability in $l_\infty(\mathscr{F})$. Assume*
(H1) $\limsup_{n\to+\infty} n/b_n^2 \log(n\mathbb{P}_\nu(\tau \geqslant b_n)) = -\infty$,
(H2) $\limsup_{n\to+\infty} n/b_n^2 \log(n\mathbb{P}_\nu(\sum_{k=0}^{\tau} \|f(X_k)\|_{\mathscr{F}} \geqslant b_n)) = -\infty$.
Then for every probability measure μ on $(E, \mathscr{E})$ verifying

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu\left(\sum_{k=0}^{\tau} \|f(X_k)\|_{\mathscr{F}} \geqslant b_n\right) = -\infty. \tag{1.5}$$

$\mathbb{P}_\mu((M_n)^{\mathscr{F}} \in \cdot)$ satisfies a moderate deviation principle on $l_\infty(\mathscr{F})$ with speed b_n^2/n and good rate function $J_{\mathscr{F}}$ given by

$$J_{\mathscr{F}}(F) = \sup\{J_{(f_1,\ldots,f_m)}(F(f_1),\ldots,F(f_m)); f_1,\ldots,f_m \in \mathscr{F}, m \geqslant 1\},$$

where $J_{(f_1,\ldots,f_m)}$ is given by

$$J_f(x) = \sup_{\xi\in\mathbb{R}^m}\left[\langle x,\xi\rangle - \frac{1}{2}\sigma^2(\langle f,\xi\rangle)\right]. \tag{1.6}$$

Suppose moreover $\sum_{k=1}^{\infty}\langle\xi, f\rangle P^k\langle f,\xi\rangle \in L_1(\pi)$ *for all* $\xi\in\mathbb{R}^m$, *then*

$$\sigma^2(\langle f,\xi\rangle) = \int \langle f,\xi\rangle^2 \,\mathrm{d}\pi + 2\int \sum_{k=1}^{\infty} \langle f,\xi\rangle P^k\langle f,\xi\rangle \,\mathrm{d}\pi.$$

Remarks. (i) Note that when an atom is present, the geometric ergodicity condition is equivalent to

$$\exists\delta > 0 \text{ such that } \mathbb{E}_\nu(\mathrm{e}^{\delta\tau}) < \infty.$$

Condition (H1) is then strictly weaker than the geometric ergodicity imposed in the work de Acosta and Chen (1998). Moreover (H1) can be more explicitly given. For example, in the particular case $b_n = n^{1/p}$ with $1 < p < 2$, for which conditions (1.3) and (1.4) are then obviously verified, then (H1) is easily seen to be implied by

$$\exists\delta > 0 \text{ such that } \mathbb{E}_\nu(\mathrm{e}^{\delta\tau^{2-p}}) < \infty. \tag{1.7}$$

Remark also that we consider here the nonseparable case of the functional empirical process and unbounded functions, cases which are not studied by de Acosta and Chen. To their credit, note however that they suppose neither the existence of an atom nor the condition (1.4) on (b_n) and their sole assumption is the well known geometric ergodicity.

(ii) Still in the context $b_n = n^{1/p}$ with $1 < p < 2$, following Nummelin and Tweedie (1978) and Nummelin and Tuominen (1982) (or Meyn and Tweedie, 1993 for a complete review) one can see that condition (1.7) is equivalent to the following *sub-geometric ergodicity*: there exists $r > 1$ such that for π-a.e. x (with $\|\cdot\|_V$ denoting the total variation norm)

$$\sum_{n=1}^{\infty} r^{n^{2-p}} \|P^n(x,\cdot) - \pi\|_V < \infty, \tag{1.8}$$

which is stronger than ergodicity of degree 2 (see Chen, 1999) but weaker than geometric ergodicity. Such an assertion implies in particular that (1.7) is valid independently of the choice of the atom and so (H1) in this context.

We have not been able to derive the independence of the recurrence condition (H1) on atom nor its characterization by means of some type of ergodicity for general b_n, but fortunately our results are proved if (H1) and (H2) are satisfied by some and then any atom.

(iii) Under (H2), condition (1.5) is verified, for instance, by the invariant measure π of the Markov chain and then by the Dirac measure δ_x for π-a.e. $x\in E$, see the appendix.

2. Applications to Markov chains with a countable state space

We will give in this section some applications where some conditions can be given explicitly, more precisely when the total boundedness of $\mathscr{F}$ with respect to the pseudo-metric d_2 or $(M_n)^{\mathscr{F}} \to 0$ in probability can be proved under satisfying hypotheses. We are much inspired here by the works of Levental (1990) and Tsai (1997).

We then consider the MDP for Markov chains with a countable state space $E = \{1, 2, 3 \dots\}$. Here τ_i will be the ith hitting time of state 1, i.e.

$$\tau := \tau_1 = \min\{n\colon n \geqslant 1, X_n = 1\} \text{ and for } i > 1, \quad \tau_i = \min\{n\colon n > \tau_{i-1}, X_n = 1\},$$

and $m_{i,j}$ be the expected minimal number of steps from state i to state j, i.e.

$$m_{i,j} = \mathbb{E}(\min\{n\colon n \geqslant 1, X_n = j\} \mid X_0 = i).$$

Let us first consider the case where $\mathscr{F}$ is the family of all indicator functions, i.e. $\mathscr{F} = \{1_A - \pi(A)\colon A \subset E\}$, related with Kolmogorov–Smirnov nonparametrical statistics.

Corollary 2. *Assume that* (H1) *is satisfied and*

$$\sum_{k=1}^{+\infty} \pi(k)\sqrt{m_{1,k}} < \infty, \tag{2.1}$$

for all orderings of E. Then for every probability measure μ satisfying (1.5), *the MDP of Theorem* 1 *holds for the family of all indicator functions on E.*

Proof. Condition (2.1) is the necessary and sufficient condition for the uniform CLT over all subsets of the integers for Markov chains satisfying $\mathbb{E}(\tau_2 - \tau_1)^2 < \infty$ by Levental (1990). The uniform CLT implies in particular $M_n^{\mathscr{F}} \to 0$ in probability and $(\mathscr{F}, d_2)$ totally bounded. For this family of indicator functions $\mathscr{F}$, (H2) is identical to (H1). The proof is completed by Theorem 1.

In the particular case of the law of the iterated logarithm ($b_n = \sqrt{2n \log\log n}$), we have the following:

Corollary 3. *Assume that*

$$\mathbb{E}_\nu(\tau^2(\log \tau)^a) < \infty, \quad \forall a > 0. \tag{2.2}$$

Suppose moreover

$$\frac{1}{\sqrt{\log\log n}} \sum_{k=1}^{n} \pi(k)\sqrt{m_{1,k}} \to 0, \tag{2.3}$$

for all orderings of E. Then for every probability measure μ satisfying (1.5), *the MDP of Theorem* 1 *holds for the family of all indicator functions on E and for $b_n = \sqrt{2n \log\log n}$.*

Proof. Remark that by Theorem 1, and by the fact that Ledoux (1992, Corollaire 2) shows that (H1) is implied by (2.2), we only have to prove that $(M_n)^{\mathscr{F}} \to 0$ in probability in $l_\infty(\mathscr{F})$ and that $(\mathscr{F}, d_2)$ is totally bounded. But, Tsai (2000) proves that under the square integrability of τ under ν (obvious by (2.2)), the compact LIL is implied by (2.3) and that the compact LIL is equivalent to the needed convergence in probability. Note also that the compact LIL implies that $(\mathscr{F}, d_2)$ is totally bounded, so ends our proof. □

We can extend Corollaries 2 and 3 to unbounded classes of functions $\mathscr{F} = \{f\colon |f| \leqslant F\}$ centred with finite variance, where F is a nonnegative function on E (called envelope of $\mathscr{F}$).

Corollary 4. *Suppose that* (H1) *and* (H2) *hold and that* $(\mathscr{F}, d_2)$ *is totally bounded. Assume either*

(a) $b_n = n^{\alpha+1/2}$ *with* $0 < \alpha \leqslant 1/2$, *and*

$$\exists \alpha' < \alpha \text{ such that } \frac{1}{n^{\alpha'}} \sum_{k=1}^{n} F(k)\pi(k)\sqrt{m_{1,k}} \to 0 \quad \text{as } n \to \infty;$$

or

(b) b_n *general and*

$$\frac{\sqrt{n}}{b_n} \sum_{k=1}^{[\sqrt{n}b_n]} F(k)\pi(k)\sqrt{m_{1,k}} \to 0 \quad \text{as } n \to \infty, \tag{2.4}$$

for all orderings of E. *Then for every probability measure* μ *on* $(E, \mathscr{E})$, *verifying* (1.5), *the MDP of Theorem* 1 *holds.*

First remark that the particular condition of part (a) is slightly weaker than in (b). By trivial facts, one can show that, for $b_n = \sqrt{2n \log\log n}$, condition (2.4) and (2.3) are equivalent.

Proof. Once again, by Theorem 1, we only have to prove that $(M_n)^{\mathscr{F}} \to 0$ in probability in $l_\infty(\mathscr{F})$. The conditions for cases (a) and (b) are obtained through a rewriting of the proof of Tsai (2000) in our context.

3. Proof

3.1. The separable case

We first need following lemma which gives us the moderate deviation principle when $\mathscr{F}$ is finite, i.e. in the separable case.

Let f be a measurable mapping from E to $\mathbb{R}^d$, suppose moreover that $\pi(f) = 0$ and $\sigma^2(f) < \infty$.

Lemma 5. *Assume that* (H1) *is satisfied and*
(H2′) $\limsup_{n\to+\infty} n/b_n^2 \log(n\mathbb{P}_\nu(\sum_{k=0}^{\tau} \|f(X_k)\| \geqslant b_n)) = -\infty$.
Then for every probability measure μ *on* $(E, \mathscr{E})$ *verifying*

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sum_{k=0}^{\tau} \|f(X_k)\| \geqslant b_n \right) = -\infty, \tag{3.1}$$

$\mathbb{P}_\mu(M_n(f) \in \cdot)$ *verifies a moderate deviation principle with speed* b_n^2/n *and good rate function* J_f *given by* (1.6).

Remark. By assumptions (1.3) and (1.4) on the speed b_n, for each $\varepsilon > 0$, we may choose some $l(\varepsilon) > 0$ such that $\varepsilon b_n > b_{[l(\varepsilon)n]}$. Then, using (1.4), it is not hard to conclude that (H1) implies $\forall \varepsilon > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log(n\mathbb{P}_\mu(\tau > \varepsilon b_n)) = -\infty.$$

The same extension can be made for (H2), (H2′), (3.1) and (1.5), which will be referred in the sequel, with little abuse, again as (H1), (H2), (H2′), (3.1) and (1.5).

Proof of Lemma 5. The proof relies principally on a decomposition into blocks of return to the atom, which by a regeneration argument enables us to reduce the problem to the case of i.i.d. random variables.

We divide the proof of the Lemma into 4 steps: *Step 1* is dedicated to the key decomposition of $M_n(f)$. We give an extended version of Ledoux (1992) moderate deviations of i.i.d.r.v. and we apply it to our setting in *Step 2*. The negligibilities in the decomposition are established in the third and fourth steps.

Step 1: First, introduce by induction the following successive times of return to α:

$$\tau(0) = \tau = \inf\{n \geqslant 0;\ X_n \in \alpha\},$$

$$\tau(k+1) = \inf\{n > \tau(k);\ X_n \in \alpha\}. \tag{3.2}$$

Obviously, $\{\tau(k)\}$ are stopping times w.r.t. $\{X_n\}$, and are almost surely finite. Note that $\mathbb{E}_\nu \tau = \pi(\alpha)^{-1}$.

Here is the classical decomposition of the sum $M_n(f)$; which is again crucial here,

$$M_n(f) = \frac{1}{b_n}\sum_{i=0}^{n-1} f(X_i)$$

$$= M_{\tau\wedge(n-1)}(f) + \frac{1}{b_n}\sum_{k=1}^{i(n)-1}\xi_k(f) + \frac{1}{b_n}\sum_{l(n)+1\leqslant j\leqslant n-1} f(X_j) \quad \text{p.s.}, \tag{3.3}$$

where the random $\xi_k(f)$ are defined by

$$\xi_k(f) = \sum_{j=\tau(k-1)+1}^{\tau(k)} f(X_j), \tag{3.4}$$

and

$$i(n) = \sum_{k=0}^{n-1} I_\alpha(X_k),$$

and $l(n) = \tau((i(n)-1)\vee 0)$. Note that, by Nummelin (1984), $\{\xi_k\}(f)$ is a sequence of independent random variables with common law $\mathscr{L}_{\mathbb{P}_\nu}(\sum_{j=0}^{\tau} f(X_j))$.

Let us introduce for all n, $e(n) = [\pi(\alpha)n]$, (3.3) becomes

$$M_n(f) = \frac{1}{b_n}\sum_{k=1}^{e(n)}\xi_k(f) + M_{\tau\wedge(n-1)}(f) + \frac{1}{b_n}\left(\sum_{k=1}^{i(n)-1}\xi_k(f) - \sum_{k=1}^{e(n)}\xi_k(f)\right)$$

$$+ \frac{1}{b_n}\sum_{l(n)+1\leqslant j\leqslant n-1} f(X_j). \tag{3.5}$$

We control now each term of this decomposition, showing that only the first term contributes to the moderate deviations. It is the decomposition Nummelin used to establish the Central Limit Theorem (Nummelin, 1978, Theorem 7.6).

Step 2: We deal here with the moderate deviations of the first term of (3.5). First note the following.

Lemma 6. *Let* (ω_k) *be a centred and square integrable i.i.d. sequence of* $\mathbb{R}^d$ *such that* $\exists M$, $\forall u \in \mathbb{R}$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log(n\mathbb{P}(\|\omega_0\| > ub_n)) \leqslant -\frac{u^2}{M}, \tag{3.6}$$

then, if $a(n)$ *is some positive increasing sequence such that* $a(n)/n \to a < \infty$, $1/b_n \sum_{k=1}^{a(n)} \omega_k$ *satisfies a moderate deviation principle with rate function* $a^{-1}I$ *where*

$$I(x) = \sup_{y\in\mathbb{R}^d} \left\{ \langle x, y\rangle - \frac{1}{2}\mathbb{E}\langle \omega_0, y\rangle^2 \right\}.$$

Ledoux (1992) proves only this result with $a(n)$ substituted by n, but his proof works in this context. So its proof is omitted.

Obviously, $(\xi_k(f))$ verifies condition (3.6) by (H2$'$) (regarded as its extension see remark after Lemma 5). Then $\mathbb{P}_\mu(1/b_n \sum_{k=1}^{e(n)} \xi_k(f) \in \cdot)$ satisfies a moderate deviation principle with speed b_n^2/n and rate function J_f given by

$$J_f(x) = \sup_{\phi\in\mathbb{R}^m} \left\{ \langle \phi, x\rangle - \frac{1}{2}\sigma^2(\langle \phi, f\rangle) \right\}.$$

Let us deal now with the other terms in the summation (3.5).

Step 3: We will prove that $\forall \varepsilon > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{k=1}^{\tau\wedge(n-1)} f(X_k) \right\| \geqslant \varepsilon b_n \right) = -\infty, \tag{3.7}$$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{l(n)+1\leqslant j\leqslant n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right) = -\infty, \tag{3.8}$$

In fact,

$$\left\| \sum_{k=0}^{\tau\wedge(n-1)} f(X_k) \right\| \leqslant \sum_{j=0}^{\tau\wedge(n-1)} \left\| f(X_j) \right\|$$

and then (3.7) follows exactly from condition (3.1):

$$\begin{aligned}
\mathbb{P}_\mu \left(\left\| \sum_{l(n)+1\leqslant j\leqslant n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right) &\leqslant \mathbb{P}_\mu \left(\sum_{l(n)+1\leqslant j\leqslant n-1} \|f(X_j)\| \geqslant \varepsilon b_n \right) \\
&\leqslant \mathbb{P}_\mu \left(\sum_{\tau(i(n)-1)+1\leqslant j\leqslant \tau(i(n))} \|f(X_j)\| \geqslant \varepsilon b_n \right) \\
&\leqslant \mathbb{P}_\mu \left(\max_{0\leqslant k\leqslant n-1} \sum_{j=\tau(k)+1}^{\tau(k+1)} \|f(X_j)\| \geqslant \varepsilon b_n \right).
\end{aligned}$$

By Nummelin (1984), $\{\sum_{j=\tau(k)+1}^{\tau(k+1)} \|f(X_j)\|\}$ are i.i.d. random variables under $\mathbb{P}_\mu$ with common law $\mathscr{L}_{\mathbb{P}_\nu}(\sum_{k=0}^{\tau} \|f(X_k)\|)$, so we get

$$\mathbb{P}_\mu\left(\left\|\sum_{l(n)+1\leqslant j\leqslant n-1} f(X_j)\right\| \geqslant \varepsilon b_n\right) \leqslant n\mathbb{P}_\nu\left(\sum_{k=0}^{\tau}\|f(X_k)\| \geqslant \varepsilon b_n\right),$$

and (3.8) is a straightforward consequence of condition (H2$'$).

Step 4: We shall prove here

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu\left(\left\|\sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)}\xi_k(f)\right\| \geqslant \varepsilon b_n\right) = -\infty. \tag{3.9}$$

This limit needs more effort than the previous negligibilities: let $0 < \delta < \pi(\alpha)$ be fixed but arbitrary, and n be sufficiently large in order that $e(n) \geqslant \delta n$. We have by stationarity

$$\begin{aligned}
&\mathbb{P}_\mu\left(\left\|\sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)}\xi_k(f)\right\| \geqslant \varepsilon b_n\right)\\
&\quad= \mathbb{P}_\mu\left(\left\|\sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)}\xi_k(f)\right\| \geqslant \varepsilon b_n; |i(n)-1-e(n)| > \delta n\right)\\
&\qquad+ \mathbb{P}_\mu\left(\left\|\sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)}\xi_k(f)\right\| \geqslant \varepsilon b_n; |i(n)-1-e(n)| \leqslant \delta n\right)\\
&\quad\leqslant \mathbb{P}_\mu\left(\max_{e(n)-[\delta n]\leqslant k\leqslant e(n)+[\delta n]}\left\|\sum_{i=e(n)-[\delta n]}^{k}\xi_i(f)\right\| \geqslant \frac{\varepsilon}{2}b_n\right)\\
&\qquad+ \mathbb{P}_\mu(|i(n)-1-e(n)| > \delta)\\
&\quad\leqslant \mathbb{P}_\mu\left(\max_{1\leqslant k\leqslant 2[\delta n]}\left\|\sum_{i=1}^{k}\xi_i(f)\right\| \geqslant \frac{\varepsilon}{2}b_n\right)\\
&\qquad+ \mathbb{P}_\mu(i(n)-1-e(n) > \delta n)\\
&\qquad+ \mathbb{P}_\mu(i(n) < e(n)-\delta n+1).
\end{aligned}\tag{3.10}$$

Let us begin with the last two terms of the right side of this last inequality.

$$\begin{aligned}
&\mathbb{P}_\mu(i(n)-1-e(n) > n\delta) \leqslant \mathbb{P}_\mu(\tau(e(n)+[\delta n]) \leqslant n-1)\\
&\quad\leqslant \mathbb{P}_\mu(\tau(e(n)+[\delta n]) - \tau(0) \leqslant n-1)\\
&\quad\leqslant \mathbb{P}_\mu\left(\sum_{k=1}^{e(n)+[\delta n]}(\tau(k)-\tau(k-1)) \leqslant n-1\right)\\
&\quad\leqslant \mathbb{P}_\mu\left(\frac{1}{b_{k(n)}}\sum_{k=1}^{k(n)}\left(\tau(k)-\tau(k-1)-\frac{1}{\pi(\alpha)}\right) \leqslant \frac{n-1-k(n)\pi(\alpha)^{-1}}{b_{k(n)}}\right)
\end{aligned}$$

with $k(n) = e(n) + [\delta n]$. We have

$$\frac{n-1-k(n)\pi(\alpha)^{-1}}{b_{k(n)}} = \frac{k(n)}{b_{k(n)}}\left(\frac{n-1}{k(n)} - \frac{1}{\pi(\alpha)}\right).$$

Note now that for sufficiently large n, $(n-1)/k(n) \simeq (\pi(\alpha)+\delta)^{-1}$, and $k(n)/b_{k(n)} \to \infty$. By Ledoux (1992), condition (H1) implies the upper bound of the moderate deviations for the i.i.d. sequence $\{\tau(k)-\tau(k-1)-\pi(\alpha)^{-1}\}$ with rate function I_1 such that $I_1(x) \to \infty$ when $|x| \to \infty$. Therefore, we have $\forall L > 0$,

$$\limsup_{n\to\infty} \frac{k(n)}{b_{k(n)}^2} \log \mathbb{P}_\mu(i(n) - 1 - e(n) > \delta) \leqslant -\inf_{t \leqslant -L} I_1(t).$$

Letting L tend to infinity and noting that $k(n)/n \to \pi(\alpha) + \delta$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu(i(n) \geqslant e(n) + [\delta n] - 1) = -\infty.$$

Using the same argument, we obtain also

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu(i(n) \leqslant e(n) - [\delta n] - 1) = -\infty.$$

We have the control of the last two terms of (3.9).

Because $1/b_n \sum_{k=1}^n \xi_k(f) \to 0$ in probability (by CLT for example), we have for sufficiently large n,

$$\max_{k \leqslant 2\delta n} \mathbb{P}_\mu\left(\left\|\sum_{j=k+1}^n \xi_j(f)\right\| \geqslant \frac{\varepsilon}{6} b_n\right) \leqslant \frac{1}{2}.$$

Then, by the Ottavianii's inequality for independent random variables, we get

$$\mathbb{P}_\mu\left(\max_{1\leqslant k\leqslant 2[\delta n]}\left\|\sum_{j=e(n)-\delta n}^{k} \xi_j(f)\right\| \geqslant \frac{\varepsilon}{2} b_n\right) \leqslant 2\mathbb{P}_\mu\left(\left\|\sum_{j=1}^{2[\delta n]}\xi_j(f)\right\| \geqslant \frac{\varepsilon}{6} b_n\right) + 2\mathbb{P}_\mu\left(\max_{k\leqslant 2\delta n}\|\xi_k(f)\| \geqslant \frac{\varepsilon}{6} b_n\right).$$

Obviously by the same approach as in (3.8), condition (H2′) implies

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu\left(\max_{k\leqslant 2\delta n}\|\xi_k(f)\| \geqslant \frac{\varepsilon}{6} b_n\right) = -\infty.$$

Taking $F = \{x; \|x\| \geqslant \varepsilon/6\}$, we have by the results of *Step 2*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu\left(\left\|\sum_{j=1}^{2[\delta n]}\xi_j(f)\right\| \geqslant \frac{\varepsilon}{6} b_n\right) \leqslant -\pi(\alpha)(2\delta)^{-1} \inf_{\|x\|\geqslant \varepsilon/6} J_f(x).$$

Combining these last results, we obtain

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu\left(\left\|\sum_{j=1}^{i(n)-1}\xi_j(f) - \sum_{k=1}^{e(n)}\xi_k(f)\right\| \geqslant \varepsilon b_n\right) \leqslant -\pi(\alpha)(2\delta)^{-1} \inf_{\|x\|\geqslant \varepsilon/6} J_f(x).$$

As $\sigma^2(\langle x,\xi\rangle)$ is differentiable on ξ and $\partial_\xi \sigma^2(\langle x,\xi\rangle)|_{\xi=0}=0$, $J_f(x)=0 \Leftrightarrow x=0$. Thus, by the inf-compactness of J_f, we get

$$\inf_{\|x\|\geqslant \varepsilon/6} J_f(x) > 0.$$

As δ is arbitrary, letting $\delta\to 0^+$, we have then the negligibility (3.9).

Using estimates (3.7)–(3.9) and the moderate deviations of *Step 2*, we get the result by Dembo and Zeitouni (1993, Theorem 4.2.13).

3.2. *Proof of Theorem 1*

Theorem 1 is established using the line of the proof of Wu (1994) for i.i.d. case. In fact, we reduce our proof to the use of Lemma 5 and to an exponential asymptotic equicontinuity with respect to the pseudometric d_2 associated with $\mathscr{F}$.

Under our hypothesis, by Lemma 5, we have the finite dimensional moderate deviation principle, i.e. for each $f_1,\dots,f_m\in\mathscr{F}$, $M_n((f_1,\dots,f_m))$ satisfies the moderate deviation principle with speed b_n^2/n and the good rate function $J_{(f_1,\dots,f_m)}$. We introduce the following notation $\forall\eta>0$:

$$\mathscr{F}_\eta=\{f-g;\, f,g\in\mathscr{F} \text{ and } d_2(f,g)\leqslant\eta\}.$$

We have obviously that $(\mathscr{F}_\eta,d_2)$ is totally bounded. Moreover $(M_n)^{\mathscr{F}_\eta}\to 0$ in probability in $l_\infty(\mathscr{F}_\eta)$ by our assumption. By Wu (1994), for the MDP we have only to verify the following condition: $\forall\varepsilon>0$,

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}_\mu(\|M_n\|_{\mathscr{F}_\eta}\geqslant\varepsilon)=-\infty. \tag{3.11}$$

Or equivalently, $\forall\varepsilon>0$

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{j=0}^{n-1}f(X_j)\right\|\geqslant\varepsilon b_n\right)=-\infty. \tag{3.12}$$

We use the same decomposition as in the proof of the preceding theorem:

$$\begin{aligned}
&\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{j=0}^{n-1}f(X_j)\right\|\geqslant\varepsilon b_n\right)\\
&\quad=\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{j=0}^{\tau\wedge(n-1)}f(X_j)+\sum_{i=1}^{i(n)-1}\xi_i(f)+\sum_{l(n)+1\leqslant i\leqslant n-1}f(X_i)\right\|\geqslant\varepsilon b_n\right)\\
&\quad\leqslant\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{j=0}^{\tau\wedge(n-1)}f(X_j)\right\|\geqslant\frac{\varepsilon}{3}b_n\right)+\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{i=1}^{i(n)-1}\xi_i(f)\right\|\geqslant\frac{\varepsilon}{3}b_n\right)\\
&\qquad+\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{l(n)+1\leqslant i\leqslant n-1}f(X_i)\right\|\geqslant\frac{\varepsilon}{3}b_n\right).
\end{aligned}$$

We then have to prove the negligibility of all the terms in the right side of this inequality.

Using conditions (1.5) and (H2), we get as for (3.7) and (3.8)

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{j=0}^{\tau\wedge(n-1)}f(X_j)\right\|\geqslant\frac{\varepsilon}{3}b_n\right)=-\infty,$$

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{l(n)+1\leqslant m\leqslant n-1}f(X_m)\right\|\geqslant\frac{\varepsilon}{3}b_n\right)=-\infty. \tag{3.13}$$

For the middle term, first note

$$\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{i(n)-1}\xi_k(f)\right\|\geqslant\frac{\varepsilon}{3}b_n\right)\leqslant\mathbb{P}_\mu\left(\max_{1\leqslant i\leqslant n}\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{i}\xi_k(f)\right\|\geqslant\frac{\varepsilon}{3}b_n\right).$$

We obviously have that $(1/b_n\sum_{k=1}^n\xi_k(f))^{\mathscr{F}_\eta}\to 0$ in probability in $l_\infty(\mathscr{F}_\eta)$ by the assumption that $(M_n)^{\mathscr{F}}\to 0$ in probability. Therefore, for sufficiently large n, we have

$$\max_{k\leqslant n}\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{n}\xi_k(f)\right\|\geqslant\frac{\varepsilon}{9}b_n\right)\leqslant\frac{1}{2},$$

and we use again the Ottavianii's inequality in Banach space for independent random variables, then for sufficiently large n,

$$\mathbb{P}_\mu\left(\max_{1\leqslant i\leqslant n}\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{i}\xi_k(f)\right\|\geqslant\frac{\varepsilon}{3}b_n\right)\leqslant 2\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{n}\xi_k(f)\right\|\geqslant\frac{\varepsilon b_n}{9}\right)$$
$$+2\mathbb{P}_\mu\left(\max_{k\leqslant n}\sup_{f\in\mathscr{F}_\eta}\|\xi_k(f)\|\geqslant\frac{\varepsilon b_n}{9}\right).$$

The negligibility of the last term is done as in the proof of Lemma 5. For the first term, we then use Lemma 3 of Wu (1994), an extension of Ledoux moderate deviations in the nonseparable case, we get identifying $B=l_\infty(\mathscr{F}_\eta)$ (in the notations of Wu, 1994, Lemma 3)

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{n}\xi_k(f)\right\|\geqslant\frac{\varepsilon b_n}{9}\right)\leqslant-\frac{\varepsilon^2}{C_0\sigma^2},$$

where C_0 is some universal positive constant and $\sigma^2=\sup_{f\in\mathscr{F}_\eta}\mathbb{E}(\xi_1(f))^2$. Remark that $\sigma^2\leqslant\eta^2$ and consequently

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}_\mu\left(\sup_{f\in\mathscr{F}_\eta}\left\|\sum_{k=1}^{n}\xi_k(f)\right\|\geqslant\frac{\varepsilon b_n}{9}\right)=-\infty. \tag{3.14}$$

Combining (3.3), (3.4), and preceding inequalities, we get (3.2) and then our theorem.

Acknowledgements

We deeply thank Liming Wu for all his advice during the preparation of this work, and the anonymous referee for well pointed remarks leading in particular to remark (ii).

Appendix

This section is devoted to the proof of the following lemma:

Lemma 7. *Under hypothesis* (H2), *the invariant measure* π *satisfies condition* (1.5).

Proof. In fact, we will prove the following strongest assertion:
(i) $\limsup_{n\to\infty} n/b_n^2 \log \mathbb{P}_\nu(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n) = -\infty$ is equivalent to
(ii) $\limsup_{n\to\infty} n/b_n^2 \log \mathbb{P}_\pi(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n) = -\infty$.

This assertion being proved, the conclusion of the lemma follows.
(ii) $\Rightarrow$ (i): It follows simply from

$$\begin{aligned}\pi(\cdot) &= \int_E \pi(\mathrm{d}x)P(x,\cdot)\\ &\geqslant \int_\alpha \pi(\mathrm{d}x)P(x,\cdot)\\ &\geqslant \int_\alpha \pi(\mathrm{d}x)\nu(\cdot)\\ &= \pi(\alpha)\nu(\cdot).\end{aligned}$$

(i) $\Rightarrow$ (ii): By Nummelin (1984), letting $g(x) = \mathbb{P}_x(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n)$, we have

$$\begin{aligned}\mathbb{P}_\pi\left(\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n\right) &= \int_E \pi(\mathrm{d}x)\mathbb{P}_x\left(\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n\right)\\ &= \int_E g(x)\pi(\mathrm{d}x)\\ &= \pi(\alpha)\mathbb{E}_\nu\left(\sum_{k=1}^{\tau} g(X_k)\right).\end{aligned}$$

Let $\sigma_k = \{n \geqslant k; X_n \in \alpha\}$, using the strong Markov property, we get

$$\begin{aligned}&\mathbb{P}_\pi\left(\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n\right)\\ &\quad= \pi(\alpha)\mathbb{E}_\nu\left(\sum_{k=1}^{\tau}\mathbb{P}_\nu\left(\sum_{j=k}^{\sigma_k}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n/\mathscr{F}_k\right)\right)\\ &\quad= \pi(\alpha)\mathbb{E}_\nu\left(\sum_{k=1}^{\infty}1_{\{\tau\geqslant k\}}\mathbb{P}_\nu\left(\sum_{j=k}^{\sigma_k}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n/\mathscr{F}_k\right)\right).\end{aligned}$$

Since $\{\tau \geqslant k\}$ is $\mathscr{F}_k$ measurable, and on $\{\tau \geqslant k\}, \sigma_k = \tau$, we have

$$\begin{aligned}
\mathbb{P}_\pi\left(\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n\right) &= \pi(\alpha)\mathbb{E}_v\left(\sum_{k=1}^{\infty}1_{\{\tau\geqslant k\}}1_{\{\sum_{j=k}^{\sigma_k}\|f(X_j)\|_{\mathscr{F}}\geqslant b_n\}}\right)\\
&\leqslant \pi(\alpha)\mathbb{E}_v\left(\sum_{k=1}^{\tau}1_{\{\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}}\geqslant b_n\}}\right)\\
&= \pi(\alpha)\mathbb{E}_v(\tau 1_{\{\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}}\geqslant b_n\}})\\
&\leqslant \pi(\alpha)\sqrt{\mathbb{E}_v(\tau^2)}\sqrt{\mathbb{P}_v\left(\sum_{j=0}^{\tau}\|f(X_j)\|_{\mathscr{F}} \geqslant b_n\right)},
\end{aligned}$$

where the last step is obtained by Cauchy–Schwartz inequality. We now easily derive (ii) from (i).

4. For further reading

The following references are also of interest to the reader: Athreya and Ney, 1978; Chen, 1991; de Acosta, 1990; de Acosta, 1997; Gao, 1994; Wu, 1995.

References

Athreya, K.B., Ney, P., 1978. A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. 245, 493–501.

Chen, X., 1991. Probabilities of moderate deviations for independent random vectors in Banach space. Chinese J. Appl. Probab. Statist. 7, 24–32.

Chen, X., 1997. Moderate deviation for m-dependent random variables with Banach space value. Statist. Probab. Lett. 35, 123–134.

Chen, X., 1999. Limit theorems for functionals of ergodic Markov chains with Banach space value. Memoirs AMS, pp. 139.

de Acosta, A., 1988a. Large deviations for vector-valued functionals of a Markov chain: lower bounds. Ann. Probab. 16, 925–960.

de Acosta, A., 1988b. Moderate deviation for empirical measures of Markov chains: lower bound. Ann. Probab. 25, 259–284.

de Acosta, A., 1990. Large deviations for empirical measures of Markov chain. J. Theoret. Probab. 3, 395–431.

de Acosta, A., 1997. Exponential tightness and projective systems in large deviation theory. Festschrift for Lucien Le Cam. Springer, Berlin, pp. 143–156.

de Acosta, A., Chen, X., 1998. Moderate deviation for empirical measure of Markov chains: Upper bound. J. Theoret. Probab. 4.11, 75–110.

Dembo, A., Zeitouni, O., 1993. Large Deviations Techniques and Their Applications. Jones and Bartlett, Boston, MA.

Deuschel, J.D., Stroock, D.W., 1989. Large Deviations. Academic Press, Boston.

Djellout, H., 2000. Moderate deviations for martingale differences and applications to ϕ-mixing sequences. Submitted for publication.

Gao, F.Q., 1994. Uniform moderate deviations for Markov processes. Research Announcements, Advances in Mathematics, China.

Guillin, A., 2000. Uniform moderate deviations of functional empirical processes of Markov chains. Probab. Math. Stat. 20 (2), 237–260.

Guillin, A., 2001. Moderate deviations of inhomogeneous functional of Markov processes and application to averaging. Stochast. Proc. Appl. 92 (2), 287–313.

Ledoux, M., 1992. Sur les grandes déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. Ann. Inst. H. Poincaré 35, 123–134.

Levental, S., 1990. Uniform CLT for Markov chains with a countable state space. Stochast. Proc. Appl. 34, 245–253.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer-Verlag, Berlin.

Mogulskii, A.A., 1984. On moderatly large deviation from the invariant measure. In: Borokov, A.A. (Ed.), Advances in Probability Theory: Limit Theorems and Related Problems. Optimization Software, New York.

Nummelin, E., 1978. A splitting technique for Harris recurrent chains. Z. Wahrs. nerw Gebiete 43, 309–318.

Nummelin, E., 1984. General Irreducible Markov Chains and Non-negative Operators. Cambridge University Press, Cambridge, England.

Nummelin, E., Tuominen, P., 1982. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. Stochast. Proc. Appl. 12, 187–202.

Nummelin, E., Tweedie, R.L., 1978. Geometric ergodicity and R-positivity for general Markov chains. Ann. Probab. 6, 404–420.

Tsai, T., 1997. Uniform CLT for Markov chains with a countable state space. Taiwanese J. Math. 1 (4), 481–498.

Tsai, T., 2000. Empirical law of the iterated logarithm for Markov chains with a countable state space. Stochast. Proc. Appl. 89, 175–191.

Wu, L.M., 1993. Habilitation à diriger des recherches. Université PARIS 6.

Wu, L.M., 1994. Large deviations, moderate deviations and LIL for empirical processes. Ann. Probab. 22, 17–27.

Wu, L.M., 1995. Moderate deviations of dependent random variables related to CLT. Ann. Probab. 23, 420–445.

C. R. Acad. Sci. Paris, t. 330, Série I, p. 377–380, 2000
Probabilités/*Probability Theory*

Principe de déviations modérées pour le processus empirique fonctionnel d'une chaîne de Markov

Hacène DJELLOUT, Arnaud GUILLIN

Laboratoire de mathématiques appliquées, Université Blaise-Pascal, 21, avenue des Landais, 63177 Aubière, France
Courriel : djellout@ucfma.univ-bpclermont.fr, guillin@ucfma.univ-bpclermont.fr

(Reçu le 24 novembre 1999, accepté le 10 janvier 2000)

Résumé. Nous nous intéressons dans cette Note aux déviations modérées pour le processus empirique fonctionnel d'une chaîne de Markov à valeurs dans un espace d'état général, possédant un atome. Ce principe est établi sous les hypothèses suivantes : une condition sur le premier temps d'entrée dans l'atome, et des conditions sur la classe de fonctions. Le cas général est également présenté sous des conditions sur la « split chain » associée. Ces résultats peuvent être considérés comme l'extension aux chaînes de Markov de la caractérisation des déviations modérées de v.a. i.i.d. de Ledoux. © 2000 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Moderate deviation principle for the functional empirical process of Markov Chains

Abstract. *We obtain in this paper moderate deviations for the functional empirical processes of general state space valued Markov chains with atom under the following conditions: an exponential tail for the first time of entrance in the atom of the Markov chain, and usual conditions on the class of functions for the second case. Our proofs relie on sharp conditions issued of moderate deviations of independent random variables. One can see our results as an extension to the Markov case of the beautiful characterization of moderate deviations for i.i.d. case of M. Ledoux.* © 2000 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

1. Introduction

Soit $\{X_j\}_{j\geqslant 0}$ une chaîne de Markov irréductible sur un espace d'état mesurable $(E,\mathcal{E})$, de probabilité de transition P, de mesure de probabilité invariante π. Nous supposerons de plus que cette chaîne possède un *atome*, i.e. $\exists\,\alpha\in\mathcal{E}$, ν une mesure de probabilité sur E tels que :

$$\forall x\in\alpha,\quad P(x,\cdot)=\nu(\cdot),$$

Note présentée par Marc YOR.

S0764-4442(00)00188-9/FLA

et α est ainsi appelé *atome*. On remarquera que, dans le cas où E est discret, chaque état est un *atome*. On note par $\mathbb{P}_\mu$ la loi markovienne sur $\left(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}\right)$ determinée uniquement par la transition P et la loi initiale μ. La chaîne $\{X_j\}$ représente alors la suite des coordonnées sur $E^{\mathbb{N}}$.

On définit les mesures empiriques M_n, $n \geqslant 1$, par :

$$M_n := \frac{1}{b_n} \sum_{j=0}^{n-1} (\delta_{X_j} - \pi),$$

qui sont des éléments aléatoires de $\mathcal{M}_b(E)$, l'espace des mesures finies signées sur $(E, \mathcal{E})$ muni de la norme de la variation totale $\|\cdot\|_{\mathrm{var}}$, et où b_n est une suite positive telle que $b_n/\sqrt{n} \uparrow +\infty$, $b_n/n \to 0$. On supposera en outre que b_n est strictement croissante et qu'il existe $A > 1$, $0 < \delta < 1$ tels que :

$$\forall n, k \geqslant 1, \quad b_{nk} \leqslant A k^{1-\delta} b_n.$$

Cette condition technique est la condition usuelle pour la vitesse des déviations modérées dans le cas i.i.d. [8]. Elle signifie que b_n ne peut être trop près de n (l'échelle des grandes déviations).

En statistique non paramétrique, on a besoin des estimations uniformes de $M_n(f) := \int f \, \mathrm{d}M_n = \frac{1}{b_n} \sum_{j=0}^{n-1} (f(X_j) - \pi(f))$, sur une classe de fonctions $\mathcal{F}$. C'est l'objectif de cette Note.

Notre volonté est d'étendre les résultats sur les déviations modérées de Ledoux [8] et Wu [12] établis pour le cas i.i.d. dans un espace de Banach séparable et non séparable respectivement, aux chaînes de Markov. Le cas markovien a été étudié de façon extensive sous des hypothèses de plus en plus générales (*voir* [9, 13],...), et très récemment par de Acosta [1] pour la borne inférieure sous la seule condition de l'ergodicité de degré 2, et de Acosta–Chen [2] pour la borne supérieure sous l'ergodicité géométrique et une condition de régularité de la mesure initiale, dont l'objectif principal est le principe de déviations modérées pour les fonctionnelles bornées à valeurs dans un espace de Banach séparable de M_n et pour la mesure empirique M_n. Nous améliorons et étendons ces résultats dans cette Note.

2. Résultat principal

Précisons quelques notations. Étant donné une classe de fonctions réelles et mesurables $\mathcal{F}$ telle que $\forall f \in \mathcal{F}$ on a $\pi(f) = 0$, $f \in \mathrm{L}_2(\pi)$, soit $\ell_\infty(\mathcal{F})$ l'espace de toutes les fonctions réelles bornées sur $\mathcal{F}$, muni de la norme $\|F\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |F(f)|$, qui est un espace de Banach non séparable quand $\mathcal{F}$ est infini. À toute mesure $\nu \in \mathcal{M}_b(E)$ correspond un élément $\nu^{\mathcal{F}}$ dans $\ell_\infty(\mathcal{F})$ donné par $\nu^{\mathcal{F}} = \int f \, \mathrm{d}\nu$, $\forall f \in \mathcal{F}$.

Afin d'éviter les problèmes de mesurabilité, on suppose que $\mathcal{F}$ est dénombrable, ou que le processus $\{M_n(f),\ f \in \mathcal{F}\}$ est séparable au sens de Doob pour tout n. Nous nous intéressons ici au comportement asymptotique de $\mathbb{P}_\mu\big((M_n)^{\mathcal{F}} \in \cdot\big)$.

THÉORÈME 1. – *Supposons que* $M_n^{\mathcal{F}} \to 0$ *en probabilité dans* $\ell_\infty(\mathcal{F})$. *Soit* $\tau = \inf\{n \geqslant 0,\ X_n \in \alpha\}$. *Supposons de plus que* :

(A1) $\limsup_{n \to +\infty} \dfrac{n}{b_n^2} \log\big(n \mathbb{P}_\nu(\tau \geqslant b_n)\big) = -\infty$,

(A2) $\limsup_{n \to +\infty} \dfrac{n}{b_n^2} \log\Big(n \mathbb{P}_\nu\Big(\sum_{k=0}^{\tau} \|f(X_k)\|_{\mathcal{F}} \geq b_n\Big)\Big) = -\infty$,

(A3) $\sigma^2(f) = \pi(\alpha) \mathbb{E}_\nu\Big(\sum_{k=0}^{\tau} f(X_k)\Big)^2 < \infty$, *for all* $f \in \mathcal{F}$,

et que $(\mathcal{F}, d_2)$ *est totalement bornée avec* $d_2(f, g) = \sigma(f - g)$. *Alors, pour toute mesure initiale de probabilité* μ *sur* $(E, \mathcal{E})$ *vérifiant*

$$\limsup_{n \to +\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu\left(\sum_{k=0}^{\tau} \big\|f(X_k)\big\|_{\mathcal{F}} \geqslant b_n\right) = -\infty, \tag{1}$$

$\mathbb{P}_\mu\big((M_n)^{\mathcal{F}} \in \cdot\big)$ *satisfait un principe de déviations modérées dans* $\ell_\infty(\mathcal{F})$ *de vitesse* b_n^2/n *et de bonne fonction de taux* $J_{\mathcal{F}}$*, définie par :* $J_{\mathcal{F}}(F) = \sup\big\{J_{(f_1,\dots,f_m)}\big(F(f_1),\dots,F(f_m)\big)\,;\, f_1,\dots,f_m \in \mathcal{F},\, m \geqslant 1\big\}$*, où*

$$J_f(x) = \sup_{\xi\in\mathbb{R}^m}\Big[\langle x,\xi\rangle - \frac{1}{2}\,\sigma^2\big(\langle f,\xi\rangle\big)\Big]. \tag{2}$$

Remarques. – (i) Ce cadre contient le cas de Banach séparable au sens suivant : lorsque B est un espace de Banach séparable, il existe $(y_n) \in \mathrm{B}'(0,1)$ (boule unité du dual de B), tel que $\|x\| = \sup_n |\langle x, Y_n\rangle|$. Et donc pour $\mathcal{F} = \{\langle\cdot, y_n\rangle\,;\, n\in\mathbb{N}\}$, B est un espace fermé de $\ell_\infty(\mathcal{F})$.

(ii) Même dans le cadre d'espace de Banach séparable et pour des classes de fonctions telles que $\|f\|_{\mathcal{F}} < C$, où (A2) équivaut à (A1), notre condition (A1) est plus faible que la récurrence géométrique imposée par de Acosta–Chen [2]. Dans ce dernier cas, notre condition (1) est satisfaite par toutes les mesures de probabilité vérifiant : $\exists r > 1$, $\sum_{k=1}^{\infty} r^n \|\mu P^k - \pi\|_{\mathrm{var}} < \infty$.

(iii) Rappelons que les déviations modérées pour la mesure empirique (que nous obtenons également) ne permettent pas d'obtenir de résultats sur les classes infinies (ni dans le cadre du (i)).

3. Esquisse de la démonstration

Notre preuve repose sur les travaux de Ledoux [8] et Wu [12], dans le cas i.i.d., récemment formalisé dans un cadre général par Arcones [3]. Ainsi, on réduit la preuve du théorème au cas de dimension finie et une équicontinuité exponentielle asymptotique par rapport à la pseudo-métrique d_2 associée à $\mathcal{F}$. Nous devons donc démontrer que pour tout $\varepsilon > 0$,

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\,\log\mathbb{P}_\mu\big(\|M_n\|_{\mathcal{F}_\eta} > \varepsilon\big) = -\infty,$$

où $\mathcal{F}_\eta = \{f - g\,;\, f,\, g\in\mathcal{F},\, d_2(f,g) < \eta\}$ et le résultat en dimension finie suivant :

LEMME 2. – *Sous les conditions* (A1), (A2) (*où* $\mathcal{F}$ *est finie*), *pour toute mesure* μ *vérifiant* (1), *pour toute fonction* $f : B \to \mathbb{R}^m$ *telle que* $\pi(f) = 0$, $f \in \mathrm{L}^2(\pi)$ *et* $\sigma^2(\langle f,\xi\rangle) < \infty$ *pour tout* $\xi\in\mathbb{R}^m$, $\mathbb{P}_\mu(M_n(f)\in\cdot)$ *satisfait un principe de déviations modérées de vitesse* b_n^2/n *et de bonne fonction de taux* J_f *donnée par* (2).

Pour démontrer ces deux résultats, on décompose tout d'abord $M_n(f)$ en blocs de retour à l'atome α, puis on utilise les résultats de Ledoux [8] sur les déviations modérées de v.a. i.i.d. sur un espace de Banach comme estimation a priori pour affaiblir l'hypothèse de l'ergodicité géométrique de la chaîne de Markov, et de la bornitude de f.

4. Extension au cas général

D'après Nummelin ([11], théorème 2.1), dans le cas général, la chaîne de Markov $\{X_j\}$ possède des ensembles petits, i.e. $\exists\, m\geqslant 1$, $b < 1$, $C\subset E$, ν une mesure de probabilité sur $(E,\mathcal{E})$ tels que :

$$\forall x\in E,\ A\subset E,\quad P^m(x,A)\geqslant bI_C(x)\nu(A).$$

On utilise ensuite la technique dite de « *regeneration split chain method* » systématiquement développée dans Nummelin [10,11] (pour le théorème limite centrale dans notre contexte) et Athreya–Ney [4] : on crée la chaîne $\Phi_n = \{(X_{nm}, Y_n)\}$, à partir de la chaîne initiale et d'une suite de variables aléatoires $\{Y_k\}$ à valeur dans $\{0,1\}$, avec probabilité de transition $\widetilde{P}$ (*voir* [11] pour de plus amples détails sur cette construction). $C\times\{1\}$ est alors un *atome* de Φ_n.

Soit une mesure de probabilité $\mu\in\mathcal{M}_b(E)$, on définit la mesure de probabilité $\tilde\mu$ sur $(E\times I, \mathcal{E}\otimes\mathcal{I})$ par $\tilde\mu = \mu((1-bI_C)I_{(\cdot)})\otimes\delta_0 + \mu(bI_CI_{(\cdot)})\otimes\delta_1$. On note ensuite $\tilde\tau_C = \inf\big\{n\geqslant 0\,;\,\Phi_n\in C\times\{1\}\big\}$, et remplaçons σ^2 par $\sigma_m^2(f) = \tilde\pi(C\times\{1\})\mathbb{E}_{\tilde\nu}\big(\sum_{k=0}^{m\tilde\tau_C+m-1} f(X_k)\big)^2$. On a alors :

H. Djellout, A. Guillin

THÉORÈME 3. – *Supposons que* $(\mathcal{F}, d_2)$ *est totalement bornée et que* $M_n^{\mathcal{F}} \to 0$ *en probabilité dans* $\ell_\infty(\mathcal{F})$. *Supposons de plus* :

(A1′) $\limsup_{n\to+\infty} \dfrac{n}{b_n^2} \log\left(n\widetilde{\mathbb{P}}_{\tilde\nu}(\tilde\tau_C \geqslant b_n)\right) = -\infty$,

(A2′) $\limsup_{n\to+\infty} \dfrac{n}{b_n^2} \log\Big(n\widetilde{\mathbb{P}}_{\tilde\nu}\Big(\sum_{k=0}^{m\tilde\tau_C+m-1} \|f(X_k)\|_{\mathcal{F}} \geqslant b_n\Big)\Big) = -\infty$,

et (A3) *avec* σ_m^2. *Alors pour toute mesure initiale de probabilité* μ *sur* $(E, \mathcal{E})$ *vérifiant*

$$\limsup_{n\to+\infty} \frac{n}{b_n^2} \log \widetilde{\mathbb{P}}_{\tilde\mu}\left(\sum_{k=0}^{m\tilde\tau_C+m-1} \big\|f(X_k)\big\|_{\mathcal{F}} \geqslant b_n\right) = -\infty,$$

$\mathbb{P}_\mu\big((M_n)^{\mathcal{F}} \in \cdot\big)$ *satisfait un principe de déviations modérées dans* $\ell_\infty(\mathcal{F})$ *de vitesse* b_n^2/n *et de bonne fonction de taux* $J_{\mathcal{F}}$.

Pour la preuve, on décompose la somme $M_n(f)$ en blocs de retour à $C \times \{1\}$, de manière à revenir à un problème de déviations modérées de variables aléatoires 1-dépendantes. On démontre ainsi une amélioration du théorème sur les déviations modérées de v.a. 1-dépendantes dans le cas Banach [6] en utilisant les résultats de Ledoux [8]. L'équicontinuité (2) provient de l'extension de l'inégalité d'Ottavianii au v.a. 1-dépendantes [5] et des résultats de Ledoux [8].

Remerciements. Nous remercions Liming Wu pour nous avoir dirigé vers ce problème et pour tous ses précieux conseils.

Références bibliographiques

[1] de Acosta A., Moderate deviation for empirical measures of Markov chains: Lower bound, Ann. Probab. 25 (1988) 259–284.
[2] de Acosta A., Chen X., Moderate deviation for empirical measure of Markov chains : Upper bound, J. Theor. Probab. 11 (4) (1998) 75–110.
[3] Arcones M., The large deviation principle for stochastic processes in $\ell_\infty(\mathcal{F})$, Preprint, 1999.
[4] Athreya K.B., Ney P., A new approach to the limit theory of recurrent Markov chains, Trans. Amer. Math. Soc. 245 (1978) 493–501.
[5] Chen X., Phd dissertation, 1996, Memoirs of the Amer. Math. Soc. (à apparaître).
[6] Chen X., Moderate deviation for m-dependent random variables with Banach space value, Statis. and Probab. Letters. 35 (1998) 123–134.
[7] Djellout H., Guillin A., Moderate deviations for Markov chains, (en préparation).
[8] Ledoux M., Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi, Ann. Inst. Henri-Poincaré 35 (1992) 123–134.
[9] Mogulskii A.A., On moderatly large deviations from the invariant measure, in: Advances in Probability Theory : Limit Theorems and Related Problems, Borokov A.A. (Ed.), Optimization Software, New York, 1984.
[10] Nummelin E., A splitting technique for Harris recurrent chains, Z. Wahrs. nerw Gebiete. 43 (1978) 309–318.
[11] Nummelin E., General Irreducible Markov Chains and Non-negative Operators, Cambridge University Press, Cambridge, England, 1984.
[12] Wu L., Large deviations, moderate deviations and LIL for empirical processes, Ann. Probab. 22 (1994) 17–27.
[13] Wu L., Moderate deviations of dependent random variables related to CLT, Ann. Probab. 23 (1995) 420–445.
[14] Wu L., On large deviations for moving average processes, Preprint, 1999.