

# Limit theorems for the left random walk on $GL_d(\mathbb{R})$

Christophe Cuny<sup>a</sup>, Jérôme Dedecker<sup>b</sup> and Christophe Jan<sup>c</sup>

<sup>a</sup>Laboratoire MICS, Centrale-Supelec, Grande Voie des Vignes, 92295 Chatenay-Malabry cedex, France. E-mail: [christophe.cuny@univ-nc.nc](mailto:christophe.cuny@univ-nc.nc)

<sup>b</sup>Laboratoire MAP5 (UMR 8145), Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75270 Paris Cedex 06, France.  
 E-mail: [jerome.dedecker@parisdescartes.fr](mailto:jerome.dedecker@parisdescartes.fr)

<sup>c</sup>Lycée Claude Fauriel, Avenue de la libération, 42000 Saint-Etienne, France. E-mail: [christophe.jan@ac-lyon.fr](mailto:christophe.jan@ac-lyon.fr)

Received 3 March 2016; accepted 9 June 2016

**Abstract.** Motivated by a recent work of Benoist and Quint and extending results from the PhD thesis of the third author, we obtain limit theorems for products of independent and identically distributed elements of  $GL_d(\mathbb{R})$ , such as the Marcinkiewicz–Zygmund strong law of large numbers, the CLT (with rates in Wasserstein's distances) and almost sure invariance principles with rates.

**Résumé.** Motivés par un travail récent de Benoist et Quint, nous étendons certains résultats issus de la thèse de doctorat du troisième auteur puis établissons des théorèmes limite pour les produits de matrices indépendantes et identiquement distribuées de  $GL_d(\mathbb{R})$ . Nous nous intéressons notamment aux lois fortes de Marcinkiewicz–Zygmund, au TLC (avec vitesses en distance de Wasserstein) et au principe d'invariance presque-sûr avec vitesse.

MSC: 60B15; 60F05; 60F17

Keywords: Central Limit theorem; Random walks on  $GL_d(\mathbb{R})$ ; Strong invariance principles

## 1. Introduction

Let  $(Y_n)_{n \geq 1}$  be independent random matrices taking values in  $G := GL_d(\mathbb{R})$ ,  $d \geq 2$  (the group of invertible  $d$ -dimensional real matrices), with common distribution  $\mu$ . Let  $\|\cdot\|$  be the euclidean norm on  $\mathbb{R}^d$ . We wish to study the asymptotic behaviour of  $(\log \|Y_n \cdots Y_1\|)_{n \geq 1}$ , where for every  $g \in GL_d(\mathbb{R})$ ,  $\|g\| := \sup_{x, \|x\|=1} \|gx\|$ .

We shall say that  $\mu$  has a (polynomial) moment of order  $p \geq 1$ , if

$$\int_G (\log N(g))^p \mu(dg) < \infty, \quad (1)$$

where  $N(g) := \max(\|g\|, \|g^{-1}\|)$ .

It follows from Furstenberg and Kesten [12] that, as soon as  $\mu$  admits a moment of order 1,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \|Y_n \cdots Y_1\| = \lambda_\mu, \quad \mathbb{P}\text{-a.s.},$$

where  $\lambda_\mu := \lim_{n \rightarrow +\infty} n^{-1} \mathbb{E}(\log \|Y_n \cdots Y_1\|)$  is the so-called first Lyapounov exponent.

If moreover, no proper subspace of  $\mathbb{R}^d$  is invariant by the closed semi-group generated by the support of  $\mu$ , then (see for instance Proposition 7.2 page 72 in [5]), for every  $x \in \mathbb{R}^d - \{0\}$ ,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \|Y_n \cdots Y_1 x\| = \lambda_\mu, \quad \mathbb{P}\text{-a.s.} \quad (2)$$

Our goal is to study the rate in the above convergences, assuming higher moments (and stronger algebraic conditions), as well as the Central Limit Theorem (CLT) or the Law of the Iterated Logarithm (LIL), and the rates of convergence in those limit theorems.

The CLT question benefited from several papers under an exponential moment, i.e.  $\int_G (N(g))^\alpha \mu(dg) < \infty$  for some  $\alpha > 0$ , and some algebraic conditions, see the next section for more details. Let us mention among others the papers by Le Page [20] and Guivarc’h and Raugi [13].

Quite recently, Benoist and Quint [2] proved the CLT under the existence of a moment of order 2. Their proof is based on Gordin’s martingale approximation method. By an elegant but somewhat tricky argument, they provide an explicit martingale-coboundary decomposition adapted to the problem. Moreover, as intermediary steps, they proved a result about *complete convergence* as well as an integrability property with respect to the invariant probability measure on  $X := P_{d-1}(\mathbb{R})$  (the projective space of  $\mathbb{R}^d$ ), see the next section for further details and definitions. Let us mention here that most of the results of [2] hold for linear groups on any local field.

Rates in the CLT under polynomial moments have been announced in Jan [16] (with proof in [17]) and the CLT has been proved in the PhD thesis of the third author [17] under a moment of order  $2 + \varepsilon$ , for any  $\varepsilon > 0$ . His method of proof is also based on martingale approximation, but relies on estimates that seem more suitable to obtain precise rates of convergence (in the CLT and the strong invariance principle) than the approach of Benoist and Quint, at least in the case of  $GL_d(\mathbb{R})$ .

In Section 2 below, we give our main results for the sequence  $(\log \|Y_n \cdots Y_1 x\|)_{n \geq 0}$  and any starting point  $x \in \mathbb{R}^d - \{0\}$ . We follow the approach described in Jan’s PhD thesis [17] (refining some of his computations), combined with recent or new results about rates in the strong invariance principle and rates in the CLT (see Section 3). At the very end of the paper (cf. Section 8), we also borrow one main argument from Benoist and Quint [2], to prove that the rates of convergence in the CLT apply to the sequence  $(\log \|Y_n \cdots Y_1\|)_{n \geq 1}$ , and to obtain some results for the sequence of matrix coefficients  $(\log |\langle Y_n \cdots Y_1 x, y \rangle|)_{n \geq 1}$ . In the same final section, we also briefly explain how to weaken the assumption of proximality (see the next section for the definition) by using another argument from [2].

## 2. Results

Let  $G := GL_d(\mathbb{R})$ ,  $d \geq 2$ , endowed with its Borel  $\sigma$ -algebra  $\mathcal{B}(G)$ . Let  $X := P_{d-1}(\mathbb{R})$  be the projective space of  $\mathbb{R}^d - \{0\}$ , and write  $\bar{x}$  as the projection of  $x \in \mathbb{R}^d - \{0\}$  to  $X$ . Then  $G$  acts continuously on  $X$  in a natural way:  $g \cdot \bar{x} = g\bar{x}$ .

Let  $\mu$  be a probability measure on  $\mathcal{B}(G)$ . Denote by  $\Gamma_\mu$  the closed semi-group generated by the support of  $\mu$ . Assume that  $\mu$  is *strongly irreducible*, i.e. that no proper finite union of subspaces of  $\mathbb{R}^d$  are invariant by  $\Gamma_\mu$  and that it is *proximal*, i.e. that there exists a matrix in  $\Gamma_\mu$  admitting a unique (with multiplicity one) eigenvalue with maximum modulus.

For such a measure  $\mu$ , it is known that there exists a unique invariant measure  $\nu$  on  $\mathcal{B}(X)$  (see for instance Theorem 3.1 of [5]) in the following sense: for any continuous and bounded function  $h$  from  $X$  to  $\mathbb{R}$

$$\int_X h(x) \nu(dx) = \int_G \int_X h(g \cdot x) \mu(dg) \nu(dx).$$

We consider the left random walk of law  $\mu$  on  $X$ . Let us recall its construction.

Let  $\Omega := X \times G^{\mathbb{N}^*}$  and  $\mathcal{F} := \mathcal{B}(X) \otimes \mathcal{B}(G)^{\otimes \mathbb{N}^*}$ , where  $\mathbb{N}^* = \{1, 2, \dots\}$ . For every probability measure  $\tau$  on  $\mathcal{B}(X)$ , we define  $\mathbb{P}_\tau := \tau \otimes \mu^{\otimes \mathbb{N}^*}$ . As usual we note  $\mathbb{P}_{\bar{x}} := \mathbb{P}_{\delta_{\bar{x}}}$ , for every  $\bar{x} \in X$ . Define the coordinate process  $(Y_n)_{n \in \mathbb{N}}$  ( $\mathbb{N} = \{0, 1, \dots\}$ ), i.e.  $Y_0((\bar{x}, g_1, g_2, \dots)) = \bar{x}$  and for every  $n \in \mathbb{N}^*$ ,  $Y_n((\bar{x}, g_1, g_2, \dots)) = g_n$ , and then  $\mathcal{F}_n$ , the  $\sigma$ -algebra generated by  $\{Y_0, \dots, Y_n\}$ .

Finally, define a measurable transformation  $\eta$  on  $\Omega$  by

$$\eta((\bar{x}, g_1, g_2, \dots)) = (g_1 \cdot \bar{x}, g_2, g_3, \dots).$$

The left random walk of law  $\mu$  is the process  $(W_n)_{n \in \mathbb{N}}$ , defined by  $W_0 := Y_0$  and for every  $n \in \mathbb{N}^*$ ,  $W_n = W_0 \circ \eta^n$ . Hence, it is a Markov chain defined by the recursive equation  $W_n = Y_n W_{n-1}$  for  $n \in \mathbb{N}^*$ .

Recall that for every probability measure  $\mathbb{P}_\tau$ ,  $Y_0$  is a random variable with law  $\tau$  independent from the sequence  $(Y_n)_{n \in \mathbb{N}^*}$  of independent and identically distributed (i.i.d.) random variables. Recall also that  $\mathbb{P}_\nu$  is  $\eta$ -invariant hence,

under  $\mathbb{P}_\nu$ ,  $(W_n)_{n \in \mathbb{N}}$  is identically distributed with common marginal distribution  $\nu$ . Moreover, since  $\nu$  is the unique  $\mu$ -invariant probability, then  $(\Omega, \mathcal{F}, \mathbb{P}_\nu, \eta)$  is ergodic (see e.g. Proposition 1.14 page 36 of [3]).

We want to study the process  $(X_n)_{n \in \mathbb{N}^*}$  given by  $X_n := \sigma(Y_n, W_{n-1})$  for every  $n \in \mathbb{N}^*$ , where for every  $g \in G$  and every  $\bar{x} \in X$ ,

$$\sigma(g, \bar{x}) = \log \left( \frac{\|g \cdot x\|}{\|x\|} \right).$$

Let us denote  $A_0 = \text{Id}$  and, for every  $n \in \mathbb{N}^*$ ,  $A_n := Y_n \cdots Y_1$ , so that  $X_n = \sigma(Y_n, A_{n-1} W_0)$ . Let  $S_n := X_1 + \cdots + X_n$ , and note that  $S_n = \log \|A_n W_0^*\|$ , where  $W_0^*$  is an element of  $\mathbb{R}^d$  such that  $\overline{W_0^*} = W_0$  and  $\|W_0^*\| = 1$ . Finally, let

$$B_n = \left\{ \frac{S_{[nt]} - [nt]\lambda_\mu}{\sqrt{n}} - \frac{(nt - [nt])}{\sqrt{n}} (X_{[nt]+1} - \lambda_\mu), t \in [0, 1] \right\}$$

be the partial sum process with values in the space  $C([0, 1])$  of continuous functions on  $[0, 1]$  equipped with the uniform metric.

As usual in the Markov chain setting, we denote by  $X_{n, \bar{x}}$  the random variable  $X_n$  for which  $W_0 = \bar{x}$ . Let also  $S_{n, \bar{x}}$  be the corresponding partial sum, and  $B_{n, \bar{x}}$  be the corresponding process. Note that  $S_{n, \bar{x}} = \log \|A_n x\|$  if  $\|x\| = 1$ .

Note that the distribution of the sequence  $(X_{n, \bar{x}})_{n \in \mathbb{N}^*}$  is the same for any probability  $\mathbb{P}_\tau$  on  $\Omega$  (in fact  $(X_{n, \bar{x}})_{n \in \mathbb{N}^*}$  is a function of  $\bar{x}$  and  $(Y_n)_{n \in \mathbb{N}^*}$ , so that its distribution depends only on  $\bar{x}$  and  $\mu$ ). Hence, we shall write “ $\mathbb{P}$ -almost surely” ( $\mathbb{P}$ -a.s.) instead of “ $\mathbb{P}_\tau$ -almost surely”, and  $\mathbb{E}(\cdot)$  instead of  $\mathbb{E}_\tau(\cdot)$ , for all the quantities involving the sequence  $(X_{n, \bar{x}})_{n \in \mathbb{N}^*}$  (and more generally for all the quantities involving only the sequence  $(Y_n)_{n \in \mathbb{N}^*}$ ). With these notations, for any positive and measurable function  $f$ ,

$$\mathbb{E}(f(X_{n, \bar{x}})) = \mathbb{E}_{\bar{x}}(f(X_n)) = \mathbb{E}(f(X_n) | W_0 = \bar{x}).$$

Our study will only require polynomial moments for  $\mu$ . As already mentioned, when  $\mu$  has a moment of order 1, the strong law of large numbers (2) holds for any starting point. Moreover, one can identify the limit  $\lambda_\mu$  via the ergodic theorem for strictly stationary sequences. It follows that, for every  $\bar{x} \in X$ ,

$$\frac{S_{n, \bar{x}}}{n} \xrightarrow[n \rightarrow +\infty]{} \lambda_\mu = \int_G \int_X \sigma(g, u) \mu(dg) \nu(du), \quad \mathbb{P}\text{-a.s.},$$

see for instance Corollary 3.4 page 54 of [5] or Theorem 3.28 of [3]. Our goal is to strengthen that strong law of large numbers when higher moments are assumed.

As already mentioned, in the next theorem, item (ii) has been obtained by Benoist and Quint [2]. As observed in the introduction of [2], their method also allow to prove item (iii) of the next theorem when  $p = 2$ .

**Theorem 1.** *Let  $\mu$  be a proximal and strongly irreducible probability measure on  $\mathcal{B}(G)$ . Assume that  $\mu$  has a moment of order  $p \geq 1$ .*

(i) *If  $1 \leq p < 2$  then, for every  $\bar{x} \in X$ ,*

$$\frac{S_{n, \bar{x}} - n\lambda_\mu}{n^{1/p}} \xrightarrow[n \rightarrow +\infty]{} 0, \quad \mathbb{P}\text{-a.s.}$$

(ii) *If  $p = 2$  then  $n^{-1} \mathbb{E}_\nu((S_n - n\lambda_\mu)^2) \rightarrow \sigma^2$  as  $n \rightarrow \infty$ , and, for any continuous and bounded function  $\varphi$  from  $C([0, 1])$  (equipped with the sup norm) to  $\mathbb{R}$ ,*

$$\lim_{n \rightarrow \infty} \sup_{\bar{x} \in X} \left| \mathbb{E}(\varphi(B_{n, \bar{x}})) - \int \varphi(\sigma \varpi) w(d\varpi) \right| = 0,$$

where  $w$  is the distribution of a standard Wiener process.

(iii) If  $2 \leq p < 4$  then, for every (fixed)  $\bar{x} \in X$ , one can redefine  $(S_{n,\bar{x}})_{n \geq 1}$  without changing its distribution on a (richer) probability space on which there exists i.i.d. random variables  $(W_n)_{n \geq 1}$  with common distribution  $\mathcal{N}(0, \sigma^2)$ , such that,

$$\left| S_{n,\bar{x}} - n\lambda_\mu - \sum_{i=1}^n W_i \right| = o(r_n), \quad \mathbb{P}\text{-a.s.},$$

where  $r_n = \sqrt{n \log \log n}$  when  $p = 2$  and  $r_n = n^{1/p} \sqrt{\log n}$  when  $2 < p < 4$ .

(iv) If  $p = 4$  then, for every (fixed)  $\bar{x} \in X$ , one can redefine  $(S_{n,\bar{x}})_{n \geq 1}$  without changing its distribution on a (richer) probability space on which there exists i.i.d. random variables  $(W_n)_{n \geq 1}$  with common distribution  $\mathcal{N}(0, \sigma^2)$ , such that,

$$\left| S_{n,\bar{x}} - n\lambda_\mu - \sum_{i=1}^n W_i \right| = O(n^{1/4} \sqrt{\log n} (\log \log n)^{1/4}), \quad \mathbb{P}\text{-a.s.}$$

**Remark.** Let us recall the famous result by Komlós, Major and Tusnády [18]. Let  $(V_n)_{n \in \mathbb{N}}$  be i.i.d. variables in  $\mathbb{L}^p$ ,  $p > 2$ . Then, extending the probability space if necessary, it is possible to construct i.i.d. random variables  $(Z_n)_{n \geq 1}$  with common distribution  $\mathcal{N}(0, \text{Var}(X_1))$  such that

$$\left| \sum_{i=1}^n (V_i - \mathbb{E}(V_i)) - \sum_{i=1}^n Z_i \right| = o(n^{1/p}) \quad \text{a.s.}$$

Hence, for  $p \in (2, 4]$ , our results are close to the i.i.d. situation. The logarithmic loss seems to be difficult to avoid with our approach based on martingale approximation.

**Remark.** It follows from Theorem 4.11(c) of [2] that  $\sigma \neq 0$  when  $\Gamma_\mu$  has unbounded image in  $\text{PGL}(d, \mathbb{R})$  (the projective group of  $\text{GL}_d(\mathbb{R})$ ).

The proof of Theorem 1 will result from general limit theorems under projective conditions. When  $1 < p < 2$  those results are new and when  $p > 2$ , the obtained rates slightly improve previous results (see for instance [9]).

We also obtain rates of convergence for Wasserstein’s distances in the central limit theorem. Let us first recall the definition of these minimal distances. Let  $\mathcal{L}(\nu_1, \nu_2)$  be the set of the probability laws on  $\mathbb{R}^2$  with marginals  $\nu_1$  and  $\nu_2$ . The Wasserstein distances of order  $r$  between  $\nu_1$  and  $\nu_2$  are defined as follows:

$$W_r(\nu_1, \nu_2) = \begin{cases} \inf\{\int |x - y|^r P(dx, dy) : P \in \mathcal{L}(\mu, \nu)\}, & \text{if } 0 < r < 1, \\ \inf\{(\int |x - y|^r P(dx, dy))^{1/r} : P \in \mathcal{L}(\mu, \nu)\}, & \text{if } r \geq 1. \end{cases}$$

It is well known that, for  $r \in (0, 1]$ ,

$$W_r(\nu_1, \nu_2) = \sup\{\nu_1(f) - \nu_2(f) : f \in \Lambda_r\},$$

where  $\Lambda_r$  is the set of  $r$ -Hölder functions such that  $|f(x) - f(y)| \leq |x - y|^r$  for any reals  $x, y$ . For  $r \geq 1$ , one has

$$W_r(\nu_1, \nu_2) = \left( \int_0^1 |F_1^{-1}(u) - F_2^{-1}(u)|^r du \right)^{1/r},$$

where  $F_1$  and  $F_2$  are the respective distribution functions of  $\nu_1$  and  $\nu_2$ , and  $F_1^{-1}$  and  $F_2^{-1}$  are their generalized inverse.

We obtain

**Theorem 2.** Let  $\mu$  be a proximal and strongly irreducible probability measure on  $\mathcal{B}(G)$ . For any  $\bar{x} \in X$ , denote by  $\nu_{n,\bar{x}}$  the distribution of  $n^{-1/2}(S_{n,\bar{x}} - n\lambda_\mu)$ . Let also  $G_\sigma$  be the normal distribution with mean zero and variance  $\sigma^2$  given in Theorem 1(ii) (provided  $\mu$  has a moment of order 2).

(i) Assume that  $\mu$  has a moment of order  $p \in (2, 3)$ . Then, for any  $r \in [p - 2, p]$ ,

$$\sup_{\bar{x} \in X} W_r(v_{n,\bar{x}}, G_\sigma) = O(n^{-(p-2)/2 \max(1,r)}).$$

(ii) Assume that  $\mu$  has a moment of order 3. Then, for any  $r$  in  $(1, 3]$ ,

$$\sup_{\bar{x} \in X} W_r(v_{n,\bar{x}}, G_\sigma) = O(n^{-1/2r}),$$

and for  $r = 1$ ,

$$\sup_{\bar{x} \in X} W_1(v_{n,\bar{x}}, G_\sigma) = O(n^{-1/2} \log n). \tag{3}$$

**Remark.** Except for  $p = 3, r = 1$ , the rates given in Theorem 2 are consistent with the i.i.d. case, in the following sense: let  $(V_i)_{i \geq 1}$  be a sequence of i.i.d. random variables, where the  $V_i$ 's are centered and have a moment of order  $p \in (2, 3)$ . Let  $v_n$  be the distribution of  $n^{-1/2}(V_1 + \dots + V_n)$ . Then the rates given in Theorem 2 hold for  $v_n$  instead of  $v_{n,\bar{x}}$  and  $\sigma^2 = \mathbb{E}(V_1^2)$ . Moreover, these are the best known rates under the stated conditions (see the introduction of the paper [10]). For  $p = 3, r = 1$  the rate in the i.i.d. case is  $O(n^{-1/2})$ , so there is a loss of order  $\log n$  in (3).

**Remark.** Starting from Remark 2.3 of [10], we derive from Theorem 2 the following rates of convergence in the Berry–Esseen theorem: If  $\mu$  has a moment of order  $p \in (2, 3)$ , then

$$\sup_{\bar{x} \in X} \sup_{t \in \mathbb{R}} |\mathbb{P}(n^{-1/2}(S_{n,\bar{x}} - n\lambda_\mu) \leq t) - \phi_\sigma(t)| \leq O(n^{-(p-2)/2(p-1)}),$$

where  $\phi_\sigma$  is the distribution function of  $G_\sigma$ . If  $\mu$  has a moment of order 3, then

$$\sup_{\bar{x} \in X} \sup_{t \in \mathbb{R}} |\mathbb{P}(n^{-1/2}(S_{n,\bar{x}} - n\lambda_\mu) \leq t) - \phi_\sigma(t)| \leq O(n^{-1/4} \sqrt{\log n}).$$

Note that, when  $\mu$  has moments of any order, Jan [17] obtained the rate  $O(n^{-a})$  for any  $a < 1/2$  in the Berry–Esseen theorem.

### 3. Auxiliary results on the cocycle

In all this section  $\mu$  is a proximal and strongly irreducible probability measure on  $\mathcal{B}(G)$ . Let  $\tilde{X}_k = X_k - \lambda_\mu$  and  $\tilde{X}_{k,\bar{x}} = X_{k,\bar{x}} - \lambda_\mu$ . For  $p \geq 1$ , let  $\|\cdot\|_{p,\tau}$  be the  $\mathbb{L}^p$ -norm with respect to the probability  $\mathbb{P}_\tau$  on  $\Omega$ . For the quantities involving  $X_{k,\bar{x}}$ , we shall write  $\|\cdot\|_p$  instead of  $\|\cdot\|_{p,\tau}$ , in accordance with the notations of Section 2.

The proofs of Theorems 1 and 2 will make use of general results for stationary sequences under projective conditions, i.e. conditions relying on the quantities  $\|\mathbb{E}(\tilde{X}_n | W_0)\|_{p,v}$  for  $p \geq 1$  and  $\|\mathbb{E}(\tilde{X}_n \tilde{X}_k | W_0) - \mathbb{E}_v(\tilde{X}_n \tilde{X}_k)\|_{p/2,v}$  for  $p \geq 2$ .

Those quantities were already studied in [17], where polynomial rates of convergence (to 0) were obtained. By refining the arguments of [17] we obtain the following improvements.

**Proposition 3.** Assume that  $\mu$  has a moment of order  $p > 1$ . Then, for  $q \in [1, p]$ ,

$$\sum_{k=1}^{\infty} k^{p-q-1} \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(|X_{k,\bar{x}} - X_{k,\bar{y}}|^q) < \infty \tag{4}$$

and for  $q \in (0, 1]$ ,

$$\sum_{k=1}^{\infty} k^{p-2} \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(|X_{k,\bar{x}} - X_{k,\bar{y}}|^q) < \infty. \tag{5}$$

**Remark.** Since  $\mathbb{E}(X_{k,\bar{x}}) = \mathbb{E}_{\bar{x}}(X_k) = \mathbb{E}(X_k|W_0 = \bar{x})$ , and since  $\mathbb{E}_\nu(X_k) = \lambda_\mu$ , we easily infer from (4) that

$$\sum_{k \geq 1} k^{p-2} \sup_{\bar{x} \in X} |\mathbb{E}_{\bar{x}}(X_k) - \lambda_\mu| < \infty. \tag{6}$$

In particular, using that  $p + 1/p > 2$  whenever  $p > 1$ , it follows from (6) that

$$\sum_{k \geq 1} k^{-1/p} \sup_{\bar{x} \in X} |\mathbb{E}_{\bar{x}}(X_k) - \lambda_\mu| < \infty. \tag{7}$$

**Remark.** Let us notice that the third author [17] proved that for every  $p \geq 2$  and every  $\alpha \in [0, 1)$ , there exists  $C_{p,\alpha}$  such that  $\sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(|X_{k,\bar{x}} - X_{k,\bar{y}}|) \leq \frac{C_{p,\alpha}}{k^{p(\alpha-1/2)}}$ . In particular, when  $p = 2$ , using a theorem of Maxwell and Woodroffe [21], that estimate is sufficient for the CLT under a second moment and even for the invariance principle (see Peligrad and Utev [23]). Hence, the full conclusion of item (ii) in Theorem 1 follows from the latter estimate.

We shall also need the following controls.

**Proposition 4.** *Assume that  $\mu$  has a moment of order  $p > 2$ . Then*

$$\sum_{k \geq 1} k^{p-3} \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(|\tilde{X}_{k,\bar{x}}^2 - \tilde{X}_{k,\bar{y}}^2|) < \infty, \tag{8}$$

and for every  $\gamma < p - 3 + 1/p$ ,

$$\sum_{k \geq 1} k^\gamma \sup_{\bar{x}, \bar{y} \in X} \sup_{k \leq j < i \leq 2k} \mathbb{E}(|\tilde{X}_{i,\bar{x}} \tilde{X}_{j,\bar{x}} - \tilde{X}_{i,\bar{y}} \tilde{X}_{j,\bar{x}}|) < \infty. \tag{9}$$

**Remark.** As in the previous remark, we easily infer that

$$\sum_{k \geq 1} k^{p-3} \sup_{\bar{x} \in X} |\mathbb{E}_{\bar{x}}(\tilde{X}_k^2) - \mathbb{E}_\nu(\tilde{X}_k^2)| < \infty, \tag{10}$$

and for every  $\gamma < p - 3 + 1/p$ ,

$$\sum_{k \geq 1} k^\gamma \sup_{\bar{x} \in X} \sup_{k \leq j < i \leq 2k} |\mathbb{E}_{\bar{x}}(\tilde{X}_i \tilde{X}_j) - \mathbb{E}_\nu(\tilde{X}_i \tilde{X}_j)| < \infty. \tag{11}$$

The proof of Propositions 3 and 4 are based on two auxiliary lemmas. The first one gives the regularity of the cocycle  $\sigma$  with respect to a suitable metric, that we introduce right now.

For  $\bar{x}, \bar{y} \in X$ , define

$$d(\bar{x}, \bar{y}) := \frac{\|x \wedge y\|}{\|x\| \|y\|},$$

where  $\wedge$  stands for the exterior product, see e.g. [5, page 61] for the definition and some properties. Then,  $d$  is a metric on  $X$ .

For every  $q > 0$ , define a nondecreasing, concave function  $H_q$  on  $[0, 1]$  by  $H_q(0) = 0$  and for every  $x \in (0, 1]$ ,  $H_q(x) = |\log(xe^{-q-1})|^{-q}$ .

The next lemma may be seen as a version of Lemma 17 of Jan [17].

**Lemma 5.** *For every  $\kappa > 1$ , there exists  $C_\kappa > 0$  such that for every  $g \in G$  and every  $\bar{x}, \bar{y} \in X$ ,*

$$|\sigma(g, \bar{x}) - \sigma(g, \bar{y})| \leq C_\kappa (1 + \log N(g))^\kappa H_{\kappa-1}(d(\bar{x}, \bar{y})). \tag{12}$$

**Proof.** By Lemma 12.2 of [3], there exists  $C > 0$  such that for every  $\bar{x}, \bar{y} \in X$ ,

$$|\sigma(g, \bar{x}) - \sigma(g, \bar{y})| \leq CN(g)d(\bar{x}, \bar{y}). \tag{13}$$

Now, it is not hard to prove that (notice that  $\|g^{-1}\|^{-1} \leq \|x\|^{-1}\|gx\| \leq \|g\|$  for every  $g \in G$  and every  $x \in \mathbb{R}^d - \{0\}$ ), for every  $\bar{x} \in X$  and every  $g \in G$ ,

$$\sigma(g, \bar{x}) \leq \log(N(g)). \tag{14}$$

Assume that  $d(\bar{x}, \bar{y}) \leq 1/N(g)$ . Using that  $t \mapsto t(H_{\kappa-1}(t))^{-1}$  is nondecreasing on  $(0, e]$  and that  $N(g) \geq 1$ , we have

$$N(g)d(\bar{x}, \bar{y}) \leq \frac{H_{\kappa-1}(d(\bar{x}, \bar{y}))}{H_{\kappa-1}(N(g)^{-1})}.$$

Hence, by (13),

$$\begin{aligned} |\sigma(g, \bar{x}) - \sigma(g, \bar{y})| &\leq C(H_{\kappa-1}(N(g))^{-1})^{-1} H_{\kappa-1}(d(\bar{x}, \bar{y})) \\ &\leq C(\kappa + \log N(g))^{\kappa-1} H_{\kappa-1}(d(\bar{x}, \bar{y})). \end{aligned} \tag{15}$$

Assume now that  $d(\bar{x}, \bar{y}) > 1/N(g)$ . By (14),

$$|\sigma(g, \bar{x}) - \sigma(g, \bar{y})| \leq \frac{\log N(g)H_{\kappa-1}(N(g)^{-1})}{H_{\kappa-1}(N(g)^{-1})} \leq (\kappa + \log N(g))^{\kappa} H_{\kappa-1}(d(\bar{x}, \bar{y})). \tag{16}$$

Combining (15) and (16), we see that (12) holds. □

The next lemma is a result about complete convergence that may be derived from Proposition 4.1 of Benoist and Quint [2]. A different proof is given in Section 7.

**Lemma 6.** *Assume that  $\mu$  has a moment of order  $p > 1$ . Then, there exists  $\ell > 0$ , such that*

$$\sum_{k \geq 1} k^{p-2} \max_{k \leq j \leq 2k} \sup_{\bar{x}, \bar{y} \in X, \bar{x} \neq \bar{y}} \mathbb{P}(\log(d(A_{j-1} \cdot \bar{x}, A_{j-1} \cdot \bar{y})) \geq -\ell k) < \infty. \tag{17}$$

**Proof of Proposition 3.** Let  $\bar{x}, \bar{y} \in X$ . Let  $\ell > 0$  be as in Lemma 6. We start from the elementary inequality: Let  $A = \{\log N(Y_k) \geq k\}$  and  $B = \{\log d(A_{k-1} \cdot x, A_{k-1} \cdot y) \geq -\ell k\}$ ,

$$\begin{aligned} |X_{k, \bar{x}} - X_{k, \bar{y}}| &\leq |\sigma(Y_k, A_{k-1}\bar{x}) - \sigma(Y_k, A_{k-1}\bar{y})| \mathbf{1}_A + |\sigma(Y_k, A_{k-1}\bar{x}) - \sigma(Y_k, A_{k-1}\bar{y})| \mathbf{1}_B \\ &\quad + |\sigma(Y_k, A_{k-1}\bar{x}) - \sigma(Y_k, A_{k-1}\bar{y})| \mathbf{1}_{\{A^c \cap B^c\}}. \end{aligned} \tag{18}$$

Using (14) and (12) (with  $\kappa = (p + q)/q$ ), we infer from (18) that

$$\begin{aligned} \mathbb{E}(|X_{k, \bar{x}} - X_{k, \bar{y}}|^q) &\leq C\mathbb{E}(|\log N(Y_k)|^q \mathbf{1}_A) + C\mathbb{E}(|\log N(Y_k)|^q \mathbf{1}_B) \\ &\quad + C \left\| \frac{(1 + (\log N(Y_k))^{p+q}) \mathbf{1}_{A^c}}{k^p} \right\|_1, \end{aligned}$$

for some positive constant  $C$ , and consequently

$$\begin{aligned} \mathbb{E}(|X_{k, \bar{x}} - X_{k, \bar{y}}|^q) &\leq C \int_{\{\log N(g) \geq k\}} (\log N(g))^q \mu(dg) \\ &\quad + C\mathbb{P}(\log d(A_{k-1} \cdot \bar{x}, A_{k-1} \cdot \bar{y}) \geq -\ell k) \int_G (\log N(g))^q \mu(dg) \\ &\quad + C \int_{\{\log N(g) < k\}} \frac{(\log N(g))^{p+q}}{k^p} \mu(dg). \end{aligned} \tag{19}$$

Now, for  $q \in (0, p)$  there exist two positive constants  $K$  and  $L$  such that

$$\sum_{k \geq 1} k^{p-q-1} \int_{\{\log N(g) \geq k\}} (\log N(g))^q \mu(dg) \leq K \int_G (\log N)^p d\mu < \infty \tag{20}$$

and

$$\sum_{k \geq 1} k^{p-q-1} \int_{\{\log N(g) < k\}} \frac{(\log N(g))^{p+q}}{k^p} \mu(dg) \leq L \int_G (\log N)^p d\mu < \infty. \tag{21}$$

In the case where  $q \in [1, p)$ , since  $p - q - 1 \leq p - 2$ , we infer from (19), (20), (21) and (17) that (4) holds. In the case where  $q \leq 1$ , since  $p - q - 1 \geq p - 2$ , the condition (17) implies (5). This completes the proof of Proposition 3.  $\square$

**Proof of Proposition 4.** Let us first prove (8). Using (14), we see that

$$|\tilde{X}_{k,\bar{x}}^2 - \tilde{X}_{k,\bar{y}}^2| \leq 2(\log N(Y_k) + |\lambda_\mu|) |\sigma(Y_k, A_{k-1} \cdot \bar{x}) - \sigma(Y_k, A_{k-1} \cdot \bar{y})|.$$

Proceeding as in (18) and (19), we obtain that

$$\begin{aligned} \mathbb{E}(|\tilde{X}_{k,\bar{x}}^2 - \tilde{X}_{k,\bar{y}}^2|) &\leq C \int_{\{\log N(g) \geq k\}} \log N(g) (\log N(g) + |\lambda_\mu|) \mu(dg) \\ &\quad + C \mathbb{P}(\log d(A_{k-1} \cdot \bar{x}, A_{k-1} \cdot \bar{y}) \geq -\ell k) \int \log N(g) (\log N(g) + |\lambda_\mu|) \mu(dg) \\ &\quad + C \int_{\{\log N(g) < k\}} \frac{(\log N(g))^p (\log N(g) + |\lambda_\mu|)}{k^{p-1}} \mu(dg), \end{aligned}$$

for some positive constant  $C$ . We conclude as in Proposition 3 (using similar arguments as in (20) and (21)).

Let us prove (9). Let  $2k \geq i > j \geq k$ . We start from the simple decomposition

$$\begin{aligned} \tilde{X}_{i,\bar{x}} \tilde{X}_{j,\bar{x}} - \tilde{X}_{i,\bar{y}} \tilde{X}_{j,\bar{y}} &= \tilde{X}_{i,\bar{x}} (\sigma(Y_j, A_{j-1} \cdot \bar{x}) - \sigma(Y_j, A_{j-1} \cdot \bar{y})) \\ &\quad + (\sigma(Y_i, A_{i-1} \cdot \bar{x}) - \sigma(Y_i, A_{i-1} \cdot \bar{y})) \tilde{X}_{j,\bar{y}} := W_{i,j} + Z_{i,j}. \end{aligned}$$

Using (14), (12) (with  $\kappa = p$ ) and independence, and proceeding as in (18) and (19), we obtain that

$$\begin{aligned} \mathbb{E}(|W_{i,j}|) &\leq (|\lambda_\mu| + \|\log N\|_{1,\mu}) \int_{\{\log N(g) \geq j\}} \log N(g) \mu(dg) \\ &\quad + \|\log N\|_{1,\mu} (|\lambda_\mu| + \|\log N\|_{1,\mu}) \mathbb{P}(\log d(A_{j-1} \cdot \bar{x}, A_{j-1} \cdot \bar{y}) \geq -\ell j/2) \\ &\quad + C (|\lambda_\mu| + \|\log N\|_{1,\mu}) \int_{\{\log N(g) < j\}} \frac{(\log N(g))^{p+1}}{j^p} \mu(dg) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}_v(|Z_{i,j}|) &\leq (|\lambda_\mu| + \|\log N\|_{1,\mu}) \int_{\{\log N(g) \geq k\}} \log N(g) \mu(dg) \\ &\quad + \|\log N\|_{1,\mu} \mathbb{E}((|\lambda_\mu| + \log(N(Y_j))) \mathbf{1}_{\{\log d(A_{i-1} \cdot \bar{x}, A_{i-1} \cdot \bar{y}) \geq -i\ell/2\}}) \\ &\quad + C (|\lambda_\mu| + \|\log N\|_{1,\mu}) \int_{\{\log N(g) < k\}} \frac{(\log N(g))^{p+1}}{k^p} \mu(dg), \end{aligned}$$

for some positive constant  $C$ . Let  $\gamma < p - 3 + 1/p$ . It suffices to prove that

$$\sum_{k \geq 1} k^\gamma \max_{k \leq j < i \leq 2k} \mathbb{E}(\log(N(Y_j)) \mathbf{1}_{\{\log d(A_{i-1} \cdot \bar{x}, A_{i-1} \cdot \bar{y}) \geq -i\ell/2\}}) < \infty.$$



Using the Hölder inequality, it is enough to prove that

$$\sum_{k \geq 1} k^\gamma \max_{k \leq i \leq 2k} (\mathbb{P}(\log d(A_{i-1} \cdot \bar{x}, A_{i-1} \cdot \bar{y}) \geq -i\ell/2))^{(p-1)/p} < \infty.$$

Using the Hölder inequality again it suffices to find  $\delta > 1$  such that

$$\sum_{k \geq 1} k^{\delta/(p-1)} k^{\gamma p/(p-1)} \max_{k \leq i \leq 2k} (\mathbb{P}(\log d(A_{i-1} \cdot \bar{x}, A_{i-1} \cdot \bar{y}) \geq -i\ell/2)).$$

By Lemma 6, it suffices to find  $\delta > 1$ , such that  $\delta/(p-1) + \gamma p/(p-1) \leq p-2$ . in particular, it suffices that  $(p-2)(p-1) - \gamma p > 1$ , which holds by assumption.  $\square$

#### 4. General results under projective conditions

In this section, we state general results under projective conditions, that will be needed to prove versions of Theorems 1 and 2 in stationary regime. Proposition 7 is new and is somewhat optimal. Proposition 8 slightly improves previous results. Proposition 9 is taken from Dedecker, Merlevède and Rio [10]. Finally Proposition 10 is a new moment inequality, in the spirit of von Bahr and Esseen [26], that will be useful to prove that the results hold for any starting points.

The proofs of Propositions 7, 8 and 10 are given in Section 7.

We shall state Propositions 7 and 8 in presence of an invertible measure preserving transformation, since Proposition 9 has been proved in that situation. This will be enough for our purpose.

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and  $\theta$  be an invertible measure preserving transformation. Let  $\mathcal{G}_0 \subset \mathcal{F}$  be a  $\sigma$ -algebra, such that  $\mathcal{G}_0 \subset \theta^{-1}(\mathcal{G}_0)$ . For every  $n \in \mathbb{Z}$  define  $\mathcal{G}_n := \theta^{-n}(\mathcal{G}_0)$ .

For every  $Z \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ , we consider the following maximal functions

$$\mathcal{M}_p(Z, \theta) := \sup_{n \geq 1} \frac{|\sum_{k=0}^{n-1} Z \circ \theta^k|}{n^{1/p}} \quad \text{if } 1 \leq p < 2. \tag{22}$$

Write also  $T_n := Z + \dots + Z \circ \theta^{n-1}$ , and, for any real-valued random variable  $V$  and  $p \geq 1$ , let  $\|V\|_{p,\infty} = \sup_{t>0} t(\mathbb{P}(|V| > t))^{1/p}$ .

**Proposition 7.** *Let  $1 < p < 2$ . Let  $Z \in L^p(\Omega, \mathcal{G}_0, \mathbb{P})$  be such that*

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(T_n | \mathcal{G}_0)\|_p}{n^{1+1/p}} < \infty. \tag{23}$$

*There exists a constant  $C_p > 0$ , depending only on  $p$  such that*

$$\|\mathcal{M}_p(Z)\|_{p,\infty} \leq C_p \left( \|Z\|_p + \sum_{n \geq 1} \frac{\|\mathbb{E}(T_n | \mathcal{G}_0)\|_p}{n^{1+1/p}} \right). \tag{24}$$

Moreover,

$$T_n = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.} \tag{25}$$

and, there exists  $K > 0$  such that for every positive integer  $d$ ,

$$\left\| \max_{1 \leq i \leq 2^d} |T_i| \right\|_p \leq \frac{K}{p-1} 2^{d/p} \left( \|Z\|_p + \sum_{k=0}^d 2^{-k/p} \|\mathbb{E}(T_{2^k} | \mathcal{G}_{-2^k})\|_p \right). \tag{26}$$

**Remarks.** An inequality similar to (26) is given in Theorem 3 of [27]. It is not hard to prove that (23) holds as soon as

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_p}{n^{1/p}} < \infty. \tag{27}$$

Condition (23) may be seen as an  $L^p$ -analogue of the so-called Maxwell–Woodroffe condition [21]. As in the papers [23,24] or [7, Section D3], it can be shown that (23) is somewhat optimal for (25).

**Proposition 8.** *Let  $2 \leq p \leq 4$  and assume that  $\theta$  is ergodic if  $p = 2$ . Let  $Z \in L^p(\Omega, \mathcal{G}_0, \mathbb{P})$  be such that*

$$\sum_{n \geq 1} \|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_p < \infty \quad \text{for } p \in [2, 4) \tag{28}$$

and

$$\sum_{n \geq 1} \log(n) \|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_4 < \infty \quad \text{for } p = 4. \tag{29}$$

If  $p \in (2, 4]$ , assume also that

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(T_n^2 | \mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2}}{n^{1+2/p}} < \infty.$$

Then  $\mathbb{E}(T_n^2)/n \rightarrow \sigma^2$  as  $n \rightarrow \infty$ , and

(i) *If  $2 \leq p < 4$ , one can redefine  $(T_n)_{n \geq 1}$  without changing its distribution on a (richer) probability space on which there exists i.i.d. random variables  $(W_n)_{n \geq 1}$  with common distribution  $\mathcal{N}(0, \sigma^2)$ , such that*

$$\left| T_n - \sum_{i=1}^n W_i \right| = o(r_n), \quad \mathbb{P}\text{-a.s.}, \tag{30}$$

where  $r_n = \sqrt{n \log \log n}$  when  $p = 2$  and  $r_n = n^{1/p} \sqrt{\log n}$  when  $2 < p < 4$ .

(ii) *If  $p = 4$ , one can redefine  $(T_n)_{n \geq 1}$  without changing its distribution on a (richer) probability space on which there exists i.i.d. random variables  $(W_n)_{n \geq 1}$  with common distribution  $\mathcal{N}(0, \sigma^2)$ , such that*

$$\left| T_n - \sum_{i=1}^n W_i \right| = O(n^{1/4} \sqrt{\log n} (\log \log n)^{1/4}), \quad \mathbb{P}\text{-a.s.} \tag{31}$$

**Remark.** The condition  $\sum_{n \geq 1} \|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_p < \infty$  ensures a martingale-coboundary decomposition. It is possible to weaken this condition as done for instance in [9]. Since in our application the condition  $\sum_{n \geq 1} \|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_p < \infty$  is satisfied, we do not state those refinements.

**Proposition 9.** *Let  $2 < p \leq 3$ . Let  $Z \in L^p(\Omega, \mathcal{G}_0, \mathbb{P})$  be such that*

$$\sum_{n \geq 1} \|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_p < \infty \quad \text{for } p \in (2, 3)$$

and

$$\sum_{n \geq 1} \log(n) \|\mathbb{E}(Z \circ \theta^n | \mathcal{G}_0)\|_3 < \infty \quad \text{for } p = 3.$$

Assume also that

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(T_n^2 | \mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2}}{n^{3-p/2}} < \infty.$$

Then  $n^{-1}\mathbb{E}(T_n^2) \rightarrow \sigma^2$  as  $n \rightarrow \infty$ , and, denoting by  $L_n$  the distribution of  $n^{-1/2}T_n$  and by  $G_\sigma$  the normal distribution with mean zero and variance  $\sigma^2$ , one has:

(i) If  $p \in (2, 3)$ , then, for any  $r \in [p - 2, p]$ ,

$$W_r(L_n, G_\sigma) = O(n^{-(p-2)/2 \max(1,r)}).$$

(ii) If  $p = 3$ , then, for any  $r \in (1, 3]$ ,

$$W_r(L_n, G_\sigma) = O(n^{-1/2r}),$$

and for  $r = 1$ ,

$$W_1(L_n, G_\sigma) = O(n^{-1/2} \log n).$$

To prove Theorem 2 we shall also need the following von Bahr–Esseen type inequality. This inequality is stated in the nonstationary case: the  $Z_i$ 's are real-valued random variables adapted to an increasing filtration  $(\mathcal{F}_i)_{i \geq 0}$ , and  $T_n = Z_1 + \dots + Z_n$ .

**Proposition 10.** *Let  $r \in (1, 2]$ . The following inequality holds:*

$$\|T_n\|_r^r \leq 2^{2-r} \left( \sum_{i=1}^n \|Z_i\|_r^r + r \sum_{i=1}^{n-1} \mathbb{E}(|Z_i|^{r-1} |\mathbb{E}(T_n - T_i | \mathcal{F}_i)|) \right).$$

Moreover, letting  $T_n^* = \max(0, T_1, \dots, T_n)$ ,

$$\|T_n^*\|_r^r \leq \frac{4}{r-1} \sum_{i=1}^n \|Z_i\|_r^r + \frac{6r}{r-1} \sum_{i=1}^{n-1} \mathbb{E}(|Z_i|^{r-1} |\mathbb{E}(T_n - T_i | \mathcal{F}_i)|).$$

### 5. On the convergence of series $\sum_n n^{-(1+\beta)} \|\mathbb{E}(T_n^2 | \mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2}$

We keep the same notations as in previous section. For simplicity, if  $Z$  belongs to  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ , we shall write  $Z_n := Z \circ \theta^n$ .

We want to find conditions relying on series of the type considered in Proposition 3 such that the above series converges for a given  $p > 2$  and a given  $\beta \in [1/2, 1)$ . To do so we shall use computations as well as notations from Dedecker, Doukhan and Merlevède [9].

For every  $k, m \in \mathbb{N}$ , define

$$\begin{aligned} \gamma_p(m, k) &:= \|\mathbb{E}(Z_m Z_{m+k} | \mathcal{G}_0) - \mathbb{E}(Z_m Z_{m+k})\|_{p/2}, \\ \tilde{\gamma}_p(m) &:= \sup_{m \leq j < i \leq 2m} \|\mathbb{E}(Z_i Z_j | \mathcal{G}_0) - \mathbb{E}(Z_i Z_j)\|_{p/2}. \end{aligned}$$

Notice that in our definition of  $\tilde{\gamma}_p(m)$  we take the supremum  $\sup_{m \leq j < i \leq 2m}$  while in [9] they use  $\sup_{i \geq j \geq m}$ .

Let  $\gamma \in (0, 1)$ , be fixed for the moment. Proceeding as in (4.18) in [9], we see that (notice that  $[m^\gamma] + 1 \leq 2[m^\gamma]$ , for  $m \geq 1$ )

$$\|\mathbb{E}(T_n^2 | \mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2} \leq \sum_{k=1}^n \gamma_p(k, 0) + 4 \sum_{m=1}^n [m^\gamma] \tilde{\gamma}_p(m) + 2 \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \gamma_p(m, k),$$

with the usual convention that an empty sum equals 0. We derive that the sum  $\sum_n n^{-(1+\beta)} \|\mathbb{E}(T_n^2|\mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2}$  is finite provided that the following conditions hold (recall that  $\gamma - \beta > -1$ ):

$$\sum_{m \geq 1} m^{-\beta} \gamma_p(m, 0) < \infty, \tag{32}$$

$$\sum_{n \geq 1} n^{\gamma-\beta} \tilde{\gamma}_p(n) < \infty, \tag{33}$$

$$\sum_{n \geq 1} \frac{1}{n^{1+\beta}} \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \gamma_p(m, k) < \infty. \tag{34}$$

For every  $m, k \in \mathbb{N}$ , using the notation  $Z_m^{(0)} := Z_m - \mathbb{E}(Z_m|\mathcal{G}_0)$ , define

$$\gamma_p^*(m, k) := \|\mathbb{E}(Z_m^{(0)} Z_{m+k}^{(0)}|\mathcal{G}_0) - \mathbb{E}(Z_m^{(0)} Z_{m+k}^{(0)})\|_{p/2}.$$

Writing  $P_1(\cdot) := \mathbb{E}(\cdot|\mathcal{G}_1) - \mathbb{E}_v(\cdot|\mathcal{G}_0)$ , and combining (4.20), (4.23) and (4.24) of [9], we infer that,

$$\begin{aligned} \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \gamma_p(m, k) &\leq \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \sum_{\ell=1}^m \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p \\ &\quad + \sum_{m=1}^n \sum_{k=0}^n \|\mathbb{E}(Z_m|\mathcal{G}_0)\|_p \|\mathbb{E}(Z_{m+k}|\mathcal{G}_0)\|_p \\ &\leq \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \sum_{\ell=1}^m \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p + \left( \sum_{m=1}^{2n} \|\mathbb{E}(Z_m|\mathcal{G}_0)\|_p \right)^2. \end{aligned}$$

Hence (34) holds provided that the following conditions are satisfied

$$\sum_{n \geq 1} \frac{1}{n^{1+\beta}} \left( \sum_{m=1}^{2n} \|\mathbb{E}(Z_m|\mathcal{G}_0)\|_p \right)^2 < \infty, \tag{35}$$

$$\sum_{n \geq 1} \frac{1}{n^{1+\beta}} \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \sum_{\ell=1}^m \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p < \infty. \tag{36}$$

Now, using that  $\sum_{k=[m^\gamma]+1}^n \leq \sum_{k \geq [m^\gamma]+1}$  in the second equation, we see that

$$\begin{aligned} &\sum_{n \geq 1} \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \sum_{\ell=1}^m n^{-1-\beta} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p \\ &\leq C \sum_{m \geq 1} \sum_{k=[m^\gamma]+1}^m \sum_{\ell=1}^m m^{-\beta} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p \\ &\quad + C \sum_{m \geq 1} \sum_{k \geq m+1}^m \sum_{\ell=1}^m k^{-\beta} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p, \end{aligned}$$

so that

$$\sum_{n \geq 1} \sum_{m=1}^n \sum_{k=[m^\gamma]+1}^n \sum_{\ell=1}^m n^{-1-\beta} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p$$

$$\begin{aligned} &\leq C \sum_{k \geq 1} \sum_{\ell \geq 1} \sum_{m=1}^{\lfloor k^{1/\gamma} \rfloor} m^{-\beta} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p + C \sum_{k \geq 1} \sum_{\ell \geq 1} \sum_{m=1}^k k^{-\beta} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p \\ &\leq \tilde{C} \sum_{k, \ell \geq 1} k^{(1-\beta)/\gamma} \|P_1 Z_\ell\|_p \|P_1 Z_{\ell+k}\|_p \leq \sum_{\ell \geq 1} \|P_1 Z_\ell\|_p \sum_{k \geq 1} k^{(1-\beta)/\gamma} \|P_1 Z_k\|_p. \end{aligned}$$

Hence, (36) holds as soon as

$$\begin{aligned} \sum_{k \geq 1} k^{(1-\beta)/\gamma} \|P_1 Z_k\|_p &\leq C \sum_{\ell \geq 0} 2^{\ell((1-\beta)/\gamma)} \sum_{k=2^\ell}^{2^{\ell+1}-1} \|P_1 Z_\ell\|_p \\ &\leq C \sum_{\ell \geq 0} 2^{\ell((1-\beta)/\gamma+1-1/p)} \left( \sum_{k=2^\ell}^{2^{\ell+1}-1} \|P_1 Z_\ell\|_p^p \right)^{1/p} < \infty, \end{aligned}$$

where we used Hölder for the last inequality. Applying Lemma 5.2 of [9] with  $q = p$  we infer that (36) holds as soon as

$$\sum_{n \geq 1} n^{(1-\beta)/\gamma-1/p} \left( \sum_{k \geq n} \frac{\|\mathbb{E}(Z_k | \mathcal{G}_0)\|_p^p}{k} \right)^{1/p}. \tag{37}$$

By stationarity, the sequence  $(\|\mathbb{E}(Z_k | \mathcal{G}_0)\|_p)_{k \geq 1}$  is nonincreasing. Hence, using that  $\|\cdot\|_{\ell^p} \leq \|\cdot\|_{\ell^1}$ , we see that (37) holds provided that

$$\begin{aligned} &\sum_{\ell \geq 0} 2^{\ell((1-\beta)/\gamma-1/p+1)} \left( \sum_{k \geq \ell} \|\mathbb{E}(Z_{2^k} | \mathcal{G}_0)\|_p^p \right)^{1/p} \\ &\leq \sum_{\ell \geq 0} 2^{\ell((1-\beta)/\gamma-1/p+1)} \sum_{k \geq \ell} \|\mathbb{E}(Z_{2^k} | \mathcal{G}_0)\|_p < \infty. \end{aligned}$$

Changing the order of summation and using again that  $(\|\mathbb{E}(Z_k | \mathcal{G}_0)\|_p)_{k \geq 1}$  is nonincreasing, we infer that (37) holds provided that

$$\sum_{k \geq 1} k^{(1-\beta)/\gamma-1/p} \|\mathbb{E}(Z_{2^k} | \mathcal{G}_0)\|_p < \infty. \tag{38}$$

Collecting all the above estimates and taking care of Proposition 3, we obtain the following result:

**Proposition 11.** *Let  $p > 2$  and  $\beta \in [1/2, 1)$ . Assume that (32) and (35) hold and that there exists  $\gamma \in (0, 1)$  such that (33) and (38) hold. Then,*

$$\sum_{n > 0} \frac{\|\mathbb{E}(T_n^2 | \mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2}}{n^{1+\beta}} < \infty. \tag{39}$$

## 6. Proofs of Theorems 1 and 2

### 6.1. Proof of Theorem 1

We first prove a version in stationary regime, i.e. under  $\mathbb{P}_v$ . The proof makes use of Propositions 7 and 8. Those results are stated in the context of an invertible dynamical system. Let us explain how to circumvent that technical matter. Theorem 1 is a limit theorem for the process  $(X_n)_{n \geq 1}$ , which is a functional of the Markov chain  $((Y_n, W_{n-1}))_{n \geq 1}$  with

state space  $G \times X$  and stationary distribution  $\mu \otimes \nu$ . Since that Markov chain is stationary, it is well-known that, by Kolmogorov's theorem, there exists a probability  $\hat{\mathbb{P}}$  on the measurable space  $(\hat{\Omega}, \hat{\mathcal{F}}) = ((G \times X)^{\mathbb{Z}}, (\mathcal{B}(G) \otimes \mathcal{B}(X))^{\otimes \mathbb{Z}})$ , invariant by the shift  $\hat{\eta}$  on  $\hat{\Omega}$ , and such that the law of the coordinate process  $(\hat{V}_n)_{n \in \mathbb{Z}}$  (with values in  $G \times X$ ) under  $\hat{\mathbb{P}}$  is the same as the one of the process  $((Y_n, W_{n-1}))_{n \geq 1}$  under  $\mathbb{P}_\nu$ . In particular they both are Markov chains. Moreover,  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}}, \hat{\eta})$  is ergodic, which is not difficult to prove.

For every  $n \in \mathbb{Z}$ , define  $\hat{X}_n := \sigma(\hat{V}_0) \circ \hat{\eta}^n - \hat{\mathbb{E}}(\sigma(\hat{V}_0))$  and  $\hat{\mathcal{G}}_n := \sigma\{\hat{X}_k : k \leq n\}$ . Then, using the Markov property one can prove easily that for every  $p \geq 1$ , and every  $m \geq n \geq 1$ ,

$$\|\hat{\mathbb{E}}(\hat{X}_n | \hat{\mathcal{G}}_0)\|_p = \|\mathbb{E}_\nu(\tilde{X}_n | \tilde{X}_0)\|_p \leq \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(|X_{n, \bar{x}} - X_{n, \bar{y}}|), \tag{40}$$

$$\begin{aligned} \|\hat{\mathbb{E}}(\hat{X}_n \hat{X}_m | \hat{\mathcal{G}}_0) - \hat{\mathbb{E}}(\hat{X}_n \hat{X}_m)\|_p &= \|\mathbb{E}_\nu(\tilde{X}_n \tilde{X}_m | \tilde{X}_0) - \mathbb{E}_\nu(\tilde{X}_n \tilde{X}_m)\|_p \\ &\leq \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(|\tilde{X}_{n, \bar{x}} \tilde{X}_{m, \bar{x}} - \tilde{X}_{n, \bar{y}} \tilde{X}_{m, \bar{y}}|). \end{aligned} \tag{41}$$

Let us prove (i). Let us apply Proposition 7 with  $Z := \hat{X}_1$ . Notice that by (40) and Proposition 3 (see the remark after it), (27) holds. It follows that

$$\hat{X}_1 + \dots + \hat{X}_n = o(n^{1/p}), \quad \hat{\mathbb{P}}\text{-a.s.}$$

Then, we infer that

$$S_n - n\lambda_\mu = o(n^{1/p}), \quad \mathbb{P}_\nu\text{-a.s.}$$

or equivalently that for  $\nu$ -almost every  $\bar{x} \in X$ ,

$$S_{n, \bar{x}} - n\lambda_\mu = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.}$$

In particular, there exists  $y \in \mathbb{R}^d$  with  $\|y\| = 1$  such that

$$\log \|A_n y\| - n\lambda_\mu = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.}$$

Let  $x \in \mathbb{R}^d$  be such that  $\|x\| = 1$ . By Proposition 3.2 page 52 of [5], there exists a random variable  $C$  satisfying  $C(\omega) > 0$  for  $\mathbb{P}$ -almost every  $\omega \in \Omega$ , and such that, for every  $n \in \mathbb{N}$ ,

$$C \leq \frac{\|A_n x\|}{\|A_n\|} \leq 1. \tag{42}$$

Applying this inequality with  $x = y$ , we infer that

$$\log \|A_n\| - n\lambda_\mu = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.}$$

and then that for every  $x \in \mathbb{R}^d$  such that  $\|x\| = 1$ ,

$$\log \|A_n x\| - n\lambda_\mu = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.}$$

Let us prove items (iii) and (iv). Let us apply Proposition 8 with  $Z = \hat{X}_1$ . Then clearly, the conclusion of Proposition 8 will hold for  $Z = \hat{X}_1$  and, arguing as above, items (iii) and (iv) of Theorem 1 will follow from Lemma 4.1 of Berkes, Liu and Wu [4].

Notice first that by (40) and (6), (28) (or (29)) holds. Hence, it remains to check (39) with  $\beta = 2/p$ , which follows from Proposition 12 below.

**Proposition 12.** *Let  $p > 2$  and  $\beta \in [1/2, 1)$ . Take  $Z := \hat{X}_1$ . Then, (39) holds if  $\beta > 3 - p$ . For instance, one may take  $\beta = 2 - p/2$  when  $2 < p \leq 3$  and  $\beta = 2/p$  when  $2 < p \leq 4$ .*

**Proof.** Let  $\beta > 3 - p$  (with  $\beta \geq 1/2$ ). Using (6) and (40), we see that (35) is satisfied, since  $\beta > 0$ . Using, (6), (8) and (9) combined with (40) and (41), we infer that (39) holds if

$$-\beta \leq p - 3; \tag{43}$$

$$\gamma - \beta \leq p - 3 + 1/p; \tag{44}$$

$$(1 - \beta)/\gamma - 1/p \leq p - 2. \tag{45}$$

Now, (43) holds by assumption and then (44) holds with  $\gamma = 1/p$ . It is then not difficult to prove that (45) also holds with  $\gamma = 1/p$ .  $\square$

Let us prove item (ii). By Proposition 3, we have

$$\sum_{n \geq 1} \left( \int_X |\mathbb{E}_u(X_n) - \lambda_\mu|^2 \nu(du) \right)^{1/2} < \infty.$$

It is well-known then that Gordin’s method applies, i.e. that we have a martingale-coboundary decomposition (with respect to  $\mathbb{P}_\nu$ ), and the martingale has stationary and ergodic increments. Hence we have the weak invariance principle under  $\mathbb{P}_\nu$  (see [15]) meaning that, for any continuous and bounded function  $\varphi$  from  $C([0, 1])$  to  $\mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \left| \mathbb{E}_\nu(\varphi(B_n)) - \int \varphi(\sigma \varpi) w(d\varpi) \right| = 0, \tag{46}$$

where  $w$  is the distribution of a standard Wiener process.

Assume now that (ii) does not hold. Then, there exists a continuous and bounded function  $\varphi_0$  from  $C([0, 1])$  to  $\mathbb{R}$ , and a sequence  $\bar{x}_n$  of elements of  $X$  such that

$$\left| \mathbb{E}(\varphi_0(B_{n, \bar{x}_n})) - \int \varphi_0(\sigma \varpi) w(d\varpi) \right| \text{ does not converge to } 0 \text{ as } n \rightarrow \infty. \tag{47}$$

Now, if  $\psi$  is any bounded and Lipschitz function from  $C([0, 1])$  to  $\mathbb{R}$ , it follows from the first assertion of Lemma 13 below that

$$\lim_{n \rightarrow \infty} \left| \mathbb{E}(\psi(B_{n, \bar{x}_n})) - \mathbb{E}_\nu(\psi(B_n)) \right| = 0. \tag{48}$$

Putting together (46) and (48), we infer that  $B_{n, \bar{x}_n}$  converges in distribution to  $\sigma W$ , where  $W$  is a standard Wiener process. This is in contradiction with (47), which completes the proof of (ii).  $\square$

It remains to prove the following lemma (note that the first assertion has already been proved in [17] when  $p > 2$ ).

**Lemma 13.** *Assume that  $\mu$  has a moment of order  $p \geq 2$ . Then*

$$\sup_{x, y, \|x\|=\|y\|=1} \left\| \log \|A_n x\| - \log \|A_n y\| \right\|_1 < \infty,$$

for  $r \in (1, 2]$ ,

$$\sup_{x, y, \|x\|=\|y\|=1} \left\| \log \|A_n x\| - \log \|A_n y\| \right\|_r = \begin{cases} O(1) & \text{if } r \leq p - 1, \\ O(n^{(r+1-p)/r}), & \text{if } r > p - 1, \end{cases}$$

and for  $p \in [2, 3]$ ,

$$\sup_{x, y, \|x\|=\|y\|=1} \left\| \log \|A_n x\| - \log \|A_n y\| \right\|_p = O(n^{1/p}).$$

**Proof.** For any  $x, y \in \mathbb{R}^d$  such that  $\|x\| = \|y\| = 1$ , one has

$$\log \|A_n x\| - \log \|A_n y\| = \sum_{k=1}^n X_{k,\bar{x}} - X_{k,\bar{y}}.$$

Hence

$$\|\log \|A_n x\| - \log \|A_n y\|\|_1 \leq \sum_{k=1}^n \|X_{k,\bar{x}} - X_{k,\bar{y}}\|_1. \tag{49}$$

Using (49) and (4) (with  $q = 1$  and  $p \geq 2$ ), the first assertion of Lemma 13 follows.

For the case  $r \in (1, 2]$ , we apply Proposition 10. Let  $s_n(\bar{x}, \bar{y}) = \sum_{k=1}^n X_{k,\bar{x}} - X_{k,\bar{y}}$ . Then

$$\begin{aligned} \|\log \|A_n x\| - \log \|A_n y\|\|_r^r &\leq 2 \sum_{k=1}^n \|X_{k,\bar{x}} - X_{k,\bar{y}}\|_r^r \\ &\quad + 4 \sum_{k=1}^{n-1} \| |X_{k,\bar{x}} - X_{k,\bar{y}}|^{r-1} \mathbb{E}(s_n(\bar{x}, \bar{y}) - s_k(\bar{x}, \bar{y}) | \mathcal{F}_k) \|_1. \end{aligned} \tag{50}$$

From equality (3.9) in [2] (which can also be deduced from (6)) we infer that

$$X_{k,\bar{x}} - X_{k,\bar{y}} = d_k(\bar{x}, \bar{y}) + \psi(A_{k-1}\bar{x}, A_{k-1}\bar{y}) - \psi(A_k\bar{x}, A_k\bar{y}), \tag{51}$$

where  $d_k(\bar{x}, \bar{y})$  is  $\mathcal{F}_k$ -measurable and such that  $\mathbb{E}(d_k(\bar{x}, \bar{y}) | \mathcal{F}_{k-1}) = 0$ , and  $\psi$  is a bounded function (with  $|\psi| < M$ ). In particular, it follows from (51) that

$$\| |X_{k,\bar{x}} - X_{k,\bar{y}}|^{r-1} \mathbb{E}(s_n(\bar{x}, \bar{y}) - s_k(\bar{x}, \bar{y}) | \mathcal{F}_k) \|_1 \leq 2M \| |X_{k,\bar{x}} - X_{k,\bar{y}}|^{r-1} \|_1,$$

so that, by (50),

$$\|\log \|A_n x\| - \log \|A_n y\|\|_r^r \leq D \sum_{k=1}^n (\|X_{k,\bar{x}} - X_{k,\bar{y}}\|_r^r + \| |X_{k,\bar{x}} - X_{k,\bar{y}}|^{r-1} \|_1), \tag{52}$$

for some positive constant  $D$ . Applying (4) (with  $p \geq 2$  and  $q = r$ ) and (5) (with  $p \geq 2$  and  $q = r - 1$ ), we infer that

$$\sum_{k=1}^n (\|X_{k,\bar{x}} - X_{k,\bar{y}}\|_r^r + \| |X_{k,\bar{x}} - X_{k,\bar{y}}|^{r-1} \|_1) = O(\max(1, n^{(r+1-p)})),$$

and the second assertion of Lemma 13 follows from (52).

Let us prove the last assertion. Let

$$Z_{n,x,y} = X_{n,\bar{x}} - X_{n,\bar{y}} \quad \text{and} \quad T_n(x, y) = \sum_{k=1}^n Z_{k,x,y} := g_n(x, y, Y_1, \dots, Y_n).$$

With these notations, let  $\hat{T}_n(x, y) = g_n(x, y, Y_2, \dots, Y_{n+1})$ . Now, it is easy to see that  $T_n(x, y) = Z_{1,x,y} + \hat{T}_{n-1}(Y_1 x, Y_1 y)$ . Letting  $\psi_p(t) = |t|^p$ , we have

$$|T_n(x, y)|^p = |\hat{T}_{n-1}(Y_1 x, Y_1 y)|^p + Z_{1,x,y} \int_0^1 \psi_p'(\hat{T}_{n-1}(Y_1 x, Y_1 y) + t Z_{1,x,y}) dt.$$

Hence

$$|T_n(x, y)|^p \leq |\hat{T}_{n-1}(Y_1 x, Y_1 y)|^p + 2^{p-2} |Z_{1,x,y}|^p + p 2^{p-2} |Z_{1,x,y}| |\hat{T}_{n-1}(Y_1 x, Y_1 y)|^{p-1}.$$



Let  $G_{n,p}(x, y) = \mathbb{E}(|T_n(x, y)|^p)$ . Taking the conditional expectation with respect to  $Y_1$ , we get

$$\mathbb{E}(|T_n(x, y)|^p | Y_1) \leq G_{n-1,p}(Y_1x, Y_1y) + 2^{p-2}|Z_{1,x,y}|^p + p2^{p-2}|Z_{1,x,y}|G_{n-1,p-1}(Y_1x, Y_1y).$$

Let  $u_n = \sup_{x,y,x \neq 0,y \neq 0} G_{n,p}(x, y)$  and  $v_n = \sup_{x,y,x \neq 0,y \neq 0} G_{n,p-1}(x, y)$ . It follows that

$$\mathbb{E}(|T_n(x, y)|^p | Y_1) \leq u_{n-1} + 2^{p-2}|Z_{1,x,y}|^p + p2^{p-2}|Z_{1,x,y}|v_{n-1}.$$

Taking first the expectation, and then the maximum, we get

$$u_n \leq u_{n-1} + 2^{p-2} \sup_{x,y,x \neq 0,y \neq 0} \mathbb{E}(|Z_{1,x,y}|^p) + p2^{p-2} \sup_{x,y,x \neq 0,y \neq 0} \mathbb{E}(|Z_{1,x,y}|)v_{n-1}.$$

Since  $p - 1 \in [1, 2]$ , we know from the second assertion of the lemma that  $v_n = O(1)$ . Consequently, there exists a positive constant  $C$  such that

$$u_n \leq u_{n-1} + C.$$

It follows that  $u_n = O(n)$ , which is the desired result, since

$$u_n = \sup_{x,y,x \neq 0,y \neq 0} \left\| \sum_{k=1}^n X_{k,\bar{x}} - X_{k,\bar{y}} \right\|_p^p = \sup_{x,y,\|x\|=\|y\|=1} \|\log \|A_n x\| - \log \|A_n y\|\|_p^p. \quad \square$$

### 6.2. Proof of Theorem 2

Let  $\nu_n$  be the distribution of  $n^{-1/2}(S_n - n\lambda_\mu)$  under  $\mathbb{P}_\nu$ . As in the proof of Theorem 1, it is enough to apply Proposition 9 with  $Y = \hat{X}_1$ . From Proposition 12 (with  $\beta = 2 - p/2$ ) combined with (40) and (41), we see that the assumptions of Proposition 9 are satisfied. It follows that:

(i) If  $\mu$  has a moment of order  $p \in (2, 3)$ , then, for any  $r \in [p - 2, p]$ ,

$$W_r(\nu_n, G_\sigma) = O(n^{-(p-2)/2 \max(1,r)}).$$

(ii) If  $\mu$  has a moment of order 3, then, for any  $r \in (1, 3]$ ,

$$W_r(\nu_n, G_\sigma) = O(n^{-1/2r}),$$

and for  $r = 1$ ,

$$W_1(\nu_n, G_\sigma) = O(n^{-1/2} \log n).$$

Recall that  $W_0^*$  is an element of  $\mathbb{R}^d$  such that  $\overline{W_0^*} = W_0$  and  $\|W_0^*\| = 1$ . To prove the results for any starting point, we use the following elementary inequalities:

For  $r \leq 1$ ,

$$\sup_{\bar{x} \in X} W_r(\nu_n, \nu_{n,\bar{x}}) \leq n^{-r/2} \sup_{x,\|x\|=1} \|\log \|A_n x\| - \log \|A_n W_0^*\|\|_{1,\nu}^r.$$

For  $r > 1$ ,

$$\sup_{\bar{x} \in X} W_r(\nu_n, \nu_{n,\bar{x}}) \leq n^{-1/2} \sup_{x,\|x\|=1} \|\log \|A_n x\| - \log \|A_n W_0^*\|\|_{r,\nu}.$$

From the first assertion of Lemma 13, we infer that: for  $r \leq 1$ ,

$$\sup_{\bar{x} \in X} W_r(\nu_n, \nu_{n,\bar{x}}) = O(n^{-r/2}).$$

This proves Theorem 2 for  $r \in [p - 2, 1]$ , since in that case

$$\sup_{\bar{x} \in X} W_r(v_n, v_{n,\bar{x}}) = O(n^{-(p-2)/2}).$$

From the last assertion of Lemma 13, we infer that: for  $p \in (2, 3]$ ,

$$\sup_{\bar{x} \in X} W_p(v_n, v_{n,\bar{x}}) = O(n^{-(p-2)/2p}).$$

This proves the result for  $r = p$ .

It remains to consider the case  $r \in (1, p)$ . We use the elementary inequality

$$(W_r(v_n, v_{n,\bar{x}}))^r \leq (W_1(v_n, v_{n,\bar{x}}))^{(p-r)/(p-1)} (W_p(v_n, v_{n,\bar{x}}))^{p(r-1)/(p-1)}.$$

It follows from the preceding upper bounds for  $W_1(v_n, v_{n,\bar{x}})$  and  $W_p(v_n, v_{n,\bar{x}})$  that

$$\sup_{\bar{x} \in X} (W_r(v_n, v_{n,\bar{x}}))^r = O(n^{-(p-2)(p-r)/2(p-1)} n^{-(p-2)(r-1)/2(p-1)}) = O(n^{-(p-2)/2}),$$

which concludes the proof.

### 7. Proofs of the intermediate results

#### 7.1. Proof of Lemma 6

We first recall the following notation: for any  $x \in \mathbb{R}^d - \{0\}$  and  $g$  in  $G$ ,  $g \cdot \bar{x} = \overline{g \cdot x}$ .

Since  $\int_G \log N(g)\mu(dg) < \infty$ , we may define a bounded function  $F_1$ , by setting

$$F_1(\bar{x}, \bar{y}) = \int_G \log(d(g \cdot \bar{x}, g \cdot \bar{y}) / (d(\bar{x}, \bar{y}))) \mu(dg) \quad \forall \bar{x}, \bar{y} \in X, \bar{x} \neq \bar{y}.$$

Then, we define a cocycle as follows. For every  $g \in G$  and every  $\bar{x}, \bar{y} \in X$  with  $\bar{x} \neq \bar{y}$ , set  $\sigma_1(g, (\bar{x}, \bar{y})) := \log(d(g \cdot \bar{x}, g \cdot \bar{y}) / (d(\bar{x}, \bar{y}))) - F_1(\bar{x}, \bar{y})$ .

Finally, write

$$\log(d(A_n \bar{x}, A_n \bar{y}) / d(\bar{x}, \bar{y})) = M_n + R_n,$$

with

$$R_n = R_n(\bar{x}, \bar{y}) := \sum_{k=1}^n F_1(A_{k-1} \bar{x}, A_{k-1} \bar{y})$$

and

$$M_n := \sum_{k=1}^n \sigma_1(Y_k, (A_{k-1} \bar{x}, A_{k-1} \bar{y})),$$

and notice that  $(M_n)_{n \geq 1}$  is a martingale in  $L^p$ , since  $\mu$  has a moment of order  $p$ .

Using that  $d(\bar{x}, \bar{y}) \leq 1$ , the proposition will be proved if we can prove that there exists  $\ell > 0$ , such that

$$\sum_{k \geq 1} k^{p-2} \max_{k \leq j \leq 2k} \sup_{\bar{x}, \bar{y} \in X, \bar{x} \neq \bar{y}} \mathbb{P}(R_j(\bar{x}, \bar{y}) \geq -2\ell k) < \infty, \tag{53}$$

and

$$\sum_{k \geq 1} k^{p-2} \max_{k \leq j \leq 2k} \sup_{\bar{x}, \bar{y} \in X, \bar{x} \neq \bar{y}} \mathbb{P}(|M_j(\bar{x}, \bar{y})| \geq \ell k) < \infty. \tag{54}$$

*Proof of (53).* Let  $K > 0$  be such that  $|F_1| \leq K$ . Let  $n \geq 1$  be an integer. Then  $|R_n| \leq 2nK$  and using that  $|e^x - 1 - x| \leq x^2 e^{|x|}$  for every  $x \in \mathbb{R}$ , we see that, for every  $a > 0$ ,

$$|\mathbb{E}(e^{aR_n}) - 1 - a\mathbb{E}(R_n)| \leq a^2 K^2 e^{aK}.$$

By Proposition 6.4(ii) in [5], there exists  $n_0 \in \mathbb{N}$  and  $\delta > 0$ , such that

$$\sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(R_{n_0}(\bar{x}, \bar{y})) \leq -\delta.$$

For this  $n_0$ , we can find  $a_0 > 0$  small enough such that

$$\sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(e^{a_0 R_{n_0}(\bar{x}, \bar{y})}) \leq 1 - a_0 \delta / 2 := \rho < 1.$$

Using that  $R_{(k+1)n_0} = R_{kn_0} + R_{n_0} \circ \eta^{kn_0}$  and conditioning with respect to  $\mathcal{F}_{kn_0}$ , we infer that

$$\sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(e^{a_0 R_{(k+1)n_0}(\bar{x}, \bar{y})}) \leq \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(e^{a_0 R_{kn_0}(\bar{x}, \bar{y})}) \sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(e^{a_0 R_{n_0}(\bar{x}, \bar{y})}) \leq \rho^{k+1}.$$

Hence, there exists  $C > 0$ , such that for every  $n \in \mathbb{N}$ ,

$$\sup_{\bar{x}, \bar{y} \in X} \mathbb{E}(e^{a_0 R_n(\bar{x}, \bar{y})}) \leq C \rho^{n/n_0}.$$

Let  $k \geq 1$  and  $k \leq j \leq 2k$  and let  $\alpha := |\log \rho| / (2a_0 n_0)$ . Then

$$\mathbb{P}(R_j(\bar{x}, \bar{y}) \geq -\alpha k) \leq e^{a_0 \alpha k} \mathbb{E}(e^{a_0 R_j(\bar{x}, \bar{y})}) \leq C e^{a_0 \alpha k} \rho^{k/n_0} \leq C \rho^{k/(2n_0)},$$

and (53) holds with  $\ell = \alpha/2$ .

*Proof of (54).* The proof makes use of a result about complete convergence for martingales that we recall below. This result represents a very small sample of the general situations treated by Alsmeyer [1], and later generalized by Hao and Liu [14].

Recall that a sequence of random variables  $(D_n)_{n \geq 1}$  is said to be dominated by a (nonnegative) random variable  $X$ , if there exists  $C > 0$  such that for every  $x > 0$ ,  $\mathbb{P}(|D_n| > x) \leq C \mathbb{P}(X > x)$ .

The next theorem follows directly from Theorem 2.2 of [14].

**Theorem 14 (Alsmeyer [1], Hao and Liu [14]).** *Let  $(D_n)_{n \in \mathbb{N}}$  be a sequence of  $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -martingale differences dominated by a variable  $X$ . For every  $q > 1$ , every  $\gamma \in (1, 2]$  and every  $L \in \mathbb{N}$ , there exists  $C > 0$ , such that for every  $n \geq 1$  and every  $\varepsilon > 0$ ,*

$$\begin{aligned} & \mathbb{P}\left(\max_{1 \leq k \leq n} |D_1 + \dots + D_k| \geq \varepsilon n\right) \\ & \leq n \mathbb{P}\left(X > \frac{\varepsilon n}{4(L+1)}\right) \\ & \quad + \frac{C}{(\varepsilon n)^{q\gamma(L+1)/(q+L)}} \left\| \mathbb{E}(|D_1|^\gamma | \mathcal{F}_0) + \dots + \mathbb{E}(|D_n|^\gamma | \mathcal{F}_{n-1}) \right\|_q^{(L+1)/(q+L)}. \end{aligned} \tag{55}$$

We apply Theorem 14 with  $D_k := \sigma_1(Y_k, (A_{k-1}\bar{x}, A_{k-1}\bar{y}))$ ,  $X := 2 \log N(Y_1)$ ,  $\gamma = \min(p, 2)$  and  $q = L$  (to be chosen later). Notice that  $(D_n)_{n \in \mathbb{N}}$  is dominated by  $X$ , see for instance Lemma 5.3 page 62 of [5].

Since  $\mathbb{E}(X^p) < \infty$ , it is easy to check that for every  $\delta > 0$ ,

$$\sum_{n \geq 1} n^{p-2} \mathbb{P}(X > \delta n) < \infty.$$

Moreover,

$$\|\mathbb{E}(|D_1|^\gamma | \mathcal{F}_0) + \dots + \mathbb{E}(|D_n|^\gamma | \mathcal{F}_{n-1})\|_q \leq 2n \|X\|_p.$$

Hence, the series

$$\sum_{n \geq 1} n^{p-2} \mathbb{P}\left(\max_{1 \leq k \leq n} |D_1 + \dots + D_k| \geq \varepsilon n\right)$$

converges for every  $\varepsilon > 0$ , as soon as

$$\sum_{n \geq 1} \frac{n^{p-2}}{n^{(\gamma-1)(q+1)/2}} < \infty,$$

which holds provided that  $q > 2(p - 1)/(\gamma - 1) - 1$ . In particular, we infer that (54) holds by taking  $\varepsilon = \ell$ .

### 7.2. Proof of Proposition 7

We first give a maximal inequality in the spirit of Proposition 2 of [22]. The present form is just Proposition 4.1 of [7].

**Proposition 15.** *Let  $X \in L^1(\Omega, \mathcal{G}_0, \mathbb{P})$ . For every  $k \geq 0$ , write  $u_k := |\mathbb{E}(T_{2^k} | \mathcal{G}_{-2^k})|$  and  $d_k := \mathbb{E}(T_{2^k} | \mathcal{G}_{-2^k}) + (\mathbb{E}(T_{2^k} | \mathcal{G}_{-2^k})) \circ \theta^{2^k} - \mathbb{E}(T_{2^{k+1}} | \mathcal{G}_{-2^{k+1}})$ . Then, for every integer  $d \geq 0$ , we have (with the convention  $\sum_{k=0}^{-1} = 0$ )*

$$\begin{aligned} \max_{1 \leq i \leq 2^d} |T_i| &\leq \max_{1 \leq i \leq 2^d} \left| \sum_{\ell=0}^{i-1} (Z - \mathbb{E}(Z | \mathcal{G}_{-1})) \circ \theta^\ell \right| + \sum_{k=0}^{d-1} \max_{1 \leq i \leq 2^{d-k-1}} \left| \sum_{\ell=0}^{i-1} d_k \circ \theta^{2^{k+1}\ell} \right| \\ &\quad + u_d + \sum_{k=0}^{d-1} \max_{0 \leq \ell \leq 2^{d-1-k}-1} u_k \circ \theta^{2^{k+1}\ell}. \end{aligned} \tag{56}$$

In particular, there exists  $C > 0$ , such that for every  $p \geq 1$ ,

$$\begin{aligned} \mathcal{M}_p(X, \theta) &\leq C \left( \sum_{k \geq 0} \frac{u_k}{2^{k/p}} + \sum_{k \geq 0} \frac{(\mathcal{M}_1(u_k^p, \theta^{2^{k+1}}))^{1/p}}{2^{k/p}} \right) \\ &\quad + \mathcal{M}_p(X - \mathbb{E}_{-1}(X), \theta) + \sum_{k \geq 0} \frac{\mathcal{M}_p(d_k, \theta^{2^{k+1}})}{2^{k/p}}. \end{aligned} \tag{57}$$

By Hopf's dominated ergodic theorem (see Corollary 2.2 page 6 of [19]), for every  $f \in L^1(\Omega, \mathbb{P})$  and every  $k \in \mathbb{N}$ ,

$$\|\mathcal{M}_1(f, \theta^{2^k})\|_{1, \infty} \leq \|f\|_1.$$

Then, (24) follows from Proposition 15 combined with Proposition 2.1 of [6].

Let us prove (25). Define  $MW_p := \{Z \in L^p(\Omega, \mathcal{G}_0, \mathbb{P}) : \|Z\|_{MW_p} < \infty\}$ . Then,  $(MW_p, \|\cdot\|_{MW_p})$  is a Banach space.

For every  $Z \in L^1(\Omega, \mathcal{G}_0, \mathbb{P})$  define  $QZ = \mathbb{E}_0(Z \circ \theta)$ . Notice that  $Q^n(Z) = \mathbb{E}_0(Z \circ \theta^n)$ . Then, clearly  $Q$  is a contraction of  $L^p(\Omega, \mathcal{G}_0)$ . Now, we see that

$$\|Z\|_{MW_p} = \sum_{n \geq 0} \frac{\|\sum_{k=0}^{2^n-1} Q^k Z\|_p}{2^{n/p}} \quad \text{if } 1 < p < 2.$$

Hence, in any case,  $Q$  is a contraction on  $MW_p$ .

Writing  $V_n := I + \dots + Q^{n-1}$  and using that  $\|V_n V_k Z\|_p \leq C \min(k \|V_n\|_p, n \|V_k Z\|_p)$ , we see that, for every  $Z \in MW_p$ ,

$$\frac{\|V_{2^n} Z\|_{MW_p}}{2^n} \leq C_p \left( \frac{\|V_{2^n} Z\|_p}{2^{n/p}} + \sum_{k \geq n+1} \frac{\|V_{2^k} Z\|_p}{2^{k/p}} \right) \xrightarrow{n \rightarrow +\infty} 0. \quad (58)$$

Now, for every  $n \geq 1$ , taking  $m$  such that  $2^m \leq n < 2^{m+1}$ , we have  $\|V_n Z\|_{MW_p} \leq C \sum_{k=0}^m \|V_{2^k} Z\|_{MW_p} = o(n)$ .

In particular, we see that  $Q$  is mean ergodic on  $MW_p$  and has no nontrivial fixed point (see e.g. Theorem 1.3 page 73 of [19]), i.e.,

$$MW_p = \overline{(I - Q)MW_p}^{MW_p}. \quad (59)$$

Now, by (24) and the Banach principle (see Proposition C.1 of [6]) it is enough to prove (25) for a set of elements of  $MW_p$  that is dense, in particular on  $(I - Q)MW_p$ . So let  $Z = (I - Q)Y$ , with  $Y \in MW_p$ . Then,  $Z = Y \circ \theta - QY + Y - Y \circ \theta$ , is a martingale-coboundary decomposition in  $L^p(\Omega, \mathbb{P})$ . Hence (25) holds since  $Y \circ \theta^n = o(n^{1/p})$ ,  $\mathbb{P}$ -a.s. (by the Borel–Cantelli lemma) and by the Marcinkiewicz–Zygmund strong law of large numbers for martingales with stationary differences in  $L^p$ .

It remains to prove (26). We shall apply once more (56). The  $L_p$ -norm of the first two terms may be estimated thanks to Proposition 10. To estimate the  $L_p$ -norm of the last term in (56), we just notice that

$$\max_{0 \leq \ell \leq 2^{d-1-k-1}} u_k \circ \theta^{2^{k+1}\ell} \leq \left( \sum_{0 \leq \ell \leq 2^{d-1-k-1}} u_k^p \circ \theta^{2^{k+1}\ell} \right)^{1/p},$$

and (26) follows.

### 7.3. Proof of Proposition 8

Since  $\sum_{n \geq 1} \|\mathbb{E}(Z_n | \mathcal{G}_1)\|_p < \infty$ , we define a variable  $R$  in  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  by setting

$$R := \sum_{n \geq 1} \mathbb{E}(Z_n | \mathcal{G}_1).$$

Then, we have

$$Z_1 = R \circ \theta - \mathbb{E}(R \circ \theta | \mathcal{G}_1) + R - R \circ \theta := D + R - R \circ \theta.$$

Since  $R \in L^p(\Omega, \mathcal{F}, \mathbb{P})$  it is a standard consequence of the Borel–Cantelli lemma that  $n^{-1/p} R \circ \theta^n \rightarrow 0$ ,  $\mathbb{P}$ -a.s. as  $n$  tends to infinity. Hence, it suffices to prove (30) with  $M_n := D + \dots + D \circ \theta^{n-1}$  in place of  $S_n$ .

Since  $D \in L^p$ , it follows from Theorem 2.1 of Shao [25] (see the proofs of Corollaries 2.5, 2.7 and 2.8 in Cuny and Merlevède [8]) that we only have to prove that:

$$\sum_{i=1}^n (\mathbb{E}(D^2 | \mathcal{G}_1) - \mathbb{E}(D^2)) \circ \theta^{i-1} = o(n^{2/p}), \quad \mathbb{P}\text{-a.s.}, \quad (60)$$

if  $2 \leq p < 4$ , and that

$$\sum_{i=1}^n (\mathbb{E}(D^2|\mathcal{G}_1) - \mathbb{E}(D^2)) \circ \theta^{i-1} = O((n \log \log n)^{1/2}), \quad \mathbb{P}\text{-a.s.}, \tag{61}$$

if  $p = 4$ .

When  $p = 2$ , (60) follows from the ergodic theorem. Now, by Proposition 7 (the ergodicity of  $\theta$  is not required) and using orthogonality of martingale increments, (60) holds provided that

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(M_n^2|\mathcal{G}_0) - \mathbb{E}(M_n^2)\|_{p/2}}{n^{1+2/p}} < \infty.$$

Similarly, using Theorem 5.2 of [7], (61) holds provided that

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(M_n^2|\mathcal{G}_0) - \mathbb{E}(M_n^2)\|_2}{n^{3/2}} < \infty.$$

Since by assumption, for  $p \in (2, 4]$ ,

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(T_n^2|\mathcal{G}_0) - \mathbb{E}(T_n^2)\|_{p/2}}{n^{1+2/p}} < \infty,$$

it suffices to prove that

$$\sum_{n \geq 1} \frac{\|\mathbb{E}(M_n^2|\mathcal{G}_0) - \mathbb{E}(T_n^2|\mathcal{G}_0)\|_{p/2}}{n^{1+2/p}} < \infty.$$

In case  $p \in (2, 4)$ , this can be done as to prove (5.38) in [10] (see the proof of Theorem 3.1 in [10]). In case  $p = 4$ , this can be done as to prove (5.43) in [10] (see the proof of Theorem 3.2 in [10]).

#### 7.4. Proof of Proposition 10

We proceed as in the proof of Proposition 1 of [11]. For  $r \in (1, 2]$ , let  $\psi_r$  be the function from  $\mathbb{R}$  to  $\mathbb{R}^+$  defined by  $\psi_r(x) = |x|^r$ . We start from the following elementary decomposition (using the convention  $T_0 = 0$ ):

$$\begin{aligned} |T_n|^r &= \psi_r(T_n) = \sum_{i=1}^n \psi_r(T_i) - \psi_r(T_{i-1}) = \sum_{i=1}^n Z_i \int_0^1 \psi_r'(T_{i-1} + tZ_i) dt \\ &= \sum_{i=1}^n Z_i \int_0^1 (\psi_r'(T_{i-1} + tZ_i) - \psi_r'(T_{i-1})) dt + \sum_{i=1}^n Z_i \psi_r'(T_{i-1}). \end{aligned}$$

Consequently,

$$\begin{aligned} |T_n|^r &= \sum_{i=1}^n Z_i \int_0^1 (\psi_r'(T_{i-1} + tZ_i) - \psi_r'(T_{i-1})) dt + \sum_{i=1}^n Z_i \left( \sum_{j=1}^{i-1} \psi_r'(T_j) - \psi_r'(T_{j-1}) \right) \\ &= \sum_{i=1}^n Z_i \int_0^1 (\psi_r'(T_{i-1} + tZ_i) - \psi_r'(T_{i-1})) dt + \sum_{i=1}^{n-1} (\psi_r'(T_i) - \psi_r'(T_{i-1}))(T_n - T_i). \end{aligned}$$

Now, it is easy to check that  $|\psi'_r(x) - \psi'_r(y)| \leq r2^{2-r}|x - y|^{r-1}$ . Using this simple fact and taking the conditional expectation, we obtain

$$\mathbb{E}(|T_n|^r) \leq 2^{2-r} \sum_{i=1}^n \mathbb{E}(|Z_i|^r) \int_0^1 rt^{r-1} dt + r2^{2-r} \sum_{i=1}^{n-1} \mathbb{E}(|Z_i|^{r-1} |\mathbb{E}(T_n - T_i | \mathcal{F}_i)|),$$

and the first inequality is proved.

Let us prove the second inequality. We first write that

$$(T_n^*)^r = \sum_{i=1}^n (T_i^*)^r - (T_{i-1}^*)^r.$$

Note that for  $a \geq b \geq 0$ ,  $(r - 1)(a^r - b^r) \leq ra(a^{r-1} - b^{r-1})$ . Hence,

$$(T_n^*)^r \leq \frac{r}{r-1} \sum_{i=1}^n T_i^* ((T_i^*)^{r-1} - (T_{i-1}^*)^{r-1}) = \frac{r}{r-1} \sum_{i=1}^n T_i ((T_i^*)^{r-1} - (T_{i-1}^*)^{r-1}), \tag{62}$$

the last equality being true because  $(T_i^*)^{r-1} - (T_{i-1}^*)^{r-1}$  is nonzero iff  $T_i^* = T_i$ . Now

$$\begin{aligned} \sum_{i=1}^n T_i ((T_i^*)^{r-1} - (T_{i-1}^*)^{r-1}) &= \sum_{i=1}^n T_i (T_i^*)^{r-1} - T_{i-1} (T_{i-1}^*)^{r-1} - \sum_{i=1}^n Z_i (T_{i-1}^*)^{r-1} \\ &= T_n (T_n^*)^{r-1} - \sum_{i=1}^n Z_i (T_{i-1}^*)^{r-1}. \end{aligned} \tag{63}$$

Recall Young's inequality (based on the concavity of the logarithm): for any  $a, b \geq 0$  and any  $p, q > 1$  such that  $1/p + 1/q = 1$ ,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Hence, for  $x, y \geq 0$

$$xy^{r-1} \leq \frac{2^{r-1}}{r} x^r + \frac{r-1}{2r} y^r.$$

We infer that

$$\frac{r}{r-1} |T_n| (T_n^*)^{r-1} \leq \frac{2^{r-1}}{r-1} |T_n|^r + \frac{1}{2} (T_n^*)^r. \tag{64}$$

Combining (62), (63) and (64), we get that

$$(T_n^*)^r \leq \frac{2^r}{r-1} |T_n|^r - \frac{2r}{r-1} \sum_{i=1}^n Z_i (T_{i-1}^*)^{r-1}.$$

Proceeding as for the first inequality, we get that

$$(T_n^*)^r \leq \frac{2^r}{r-1} |T_n|^r - \frac{2r}{r-1} \sum_{i=1}^{n-1} ((T_i^*)^{r-1} - (T_{i-1}^*)^{r-1}) (T_n - T_i).$$

Since, for  $x, y \geq 0$ ,  $|x^{r-1} - y^{r-1}| \leq |x - y|^{r-1}$ , we finally get that

$$\mathbb{E}((T_n^*)^r) \leq \frac{2^r}{r-1} \mathbb{E}(|T_n|^r) + \frac{2^r}{r-1} \sum_{i=1}^{n-1} \mathbb{E}(|Z_i|^{r-1} |\mathbb{E}(T_n - T_i | \mathcal{F}_i)|).$$

Combining this inequality with the first inequality of the proposition, the result follows.

**8. Extension of the results**

In this section, we shall first explain why the results of Section 2 still hold for the sequence  $(\log \|A_n\|)_{n \geq 1}$  (with obvious changes in the statements). Next, we shall briefly explain how to deal with  $(\log |\langle A_n x, y \rangle|)_{n \geq 1}$ .

8.1. *Matrix norm*

The fact that the statements of Theorem 1 hold for  $\log \|A_n\|$  instead of  $S_{n,\bar{x}}$  is clear from the proof Theorem 1 (cf. Section 6.1). The crucial point here is Inequality (42).

Let  $\mu_n$  be the distribution of  $\log \|A_n\|$ . The fact that the statements of Theorem 2 hold for  $\mu_n$  instead of  $\nu_{n,\bar{x}}$  requires some explanations.

Let  $\tilde{\mu}_n$  be the distribution of

$$\int_X \log S_{n,u} \nu(du).$$

The first point to notice is that the statements of Theorem 2 are valid for  $\tilde{\mu}_n$  instead  $\nu_{n,\bar{x}}$ . This can be proved exactly as for the proof of Theorem 2, by using some easy consequences of Lemma 13, such as

$$\sup_{x, \|x\|=1} \left\| \log \|A_n x\| - \int_X S_{n,u} \nu(du) \right\|_p = O(n^{1/p}),$$

if  $\mu$  has a moment of order  $p \in [2, 3]$ .

The next step is to replace  $\tilde{\mu}_n$  by  $\mu_n$ . To do this, we need to introduce

$$\delta(\bar{x}, \bar{y}) := \frac{|\langle x, y \rangle|}{\|x\| \|y\|}. \tag{65}$$

It follows from Proposition 4.5 of [2] that if  $\mu$  has a moment of order  $p > 1$ , then

$$\sup_{v \in X} \int_X |\log(\delta(u, v))|^{p-1} \nu(du) < \infty. \tag{66}$$

Now, from [5] pages 52–53, we know that there exists a random variable  $V(\omega)$  with values in  $X$ , such that, for any  $x \in \mathbb{R}^d$  such that  $\|x\| = 1$ ,

$$0 \leq \log \|A_n\| - \log \|A_n x\| \leq |\log \delta(\bar{x}, V)|.$$

Integrating this inequality, we get

$$\left\| \log \|A_n\| - \int_X S_{n,u} \nu(du) \right\|_\infty \leq \sup_{v \in X} \int_X |\log(\delta(u, v))| \nu(du) < \infty$$

the term on right-hand being finite because  $\mu$  has a moment of order 2. The result easily follows.



### 8.2. Results without proximality

Proceeding as in the proof of Theorem 4.11 of [2] (using their Lemma 4.13) we infer that the results for matrix norm hold without proximality. Then, we see that the results of Theorems 1 and 2 hold also without proximality, since (42) does not require proximality but only strong irreducibility.

### 8.3. Matrix coefficients

We shall now explain how to derive results for matrix coefficients, i.e. for any given  $x, y \in \mathbb{R}^d$  with  $\|x\| = \|y\| = 1$ , we study the behaviour of  $(\log |\langle A_n x, y \rangle|)_{n \geq 1}$ .

We were not able to extend Theorem 2 to the matrix coefficients. We only succeeded to extend Theorem 1, but under a stronger moment assumption (and it does not seem possible to get rid of the proximality assumption here). Our argument is inspired by [17]. We shall use the distance  $\delta$  defined in (65) and the upper bound (66).

Let  $1 < p \leq 4$ . Assume that  $\mu$  has a moment of order  $p + 1$ . Let us explain why the results from Theorem 1 may be extended to the matrix coefficients. Actually, using similar arguments as below, one may see that a moment of order 2 is enough to derive item (ii) of Theorem 1 for the matrix coefficients.

Let  $x, y \in \mathbb{R}^d$  such that  $\|x\| = \|y\| = 1$ . We have

$$\log |\langle A_n x, y \rangle| = \log \|A_n x\| + \log \|y\| + \log \frac{|\langle A_n x, y \rangle|}{\|A_n x\| \|y\|}.$$

The behaviour of  $(\log \|A_n x\|)_{n \geq 1}$  is described in Theorem 1.

It is obvious (using Lemma 4 of [4] to deal with items (iii) and (iv)) that the results of Theorem 1 will hold for the matrix coefficients if we can prove that

$$\left| \log \frac{|\langle A_n x, y \rangle|}{\|A_n x\| \|y\|} \right| = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.}$$

or, equivalently, that

$$|\log \delta(A_n \bar{x}, \bar{y})| = o(n^{1/p}), \quad \mathbb{P}\text{-a.s.}$$

(recall that  $A_n \bar{x} = \overline{A_n x}$ ). Since  $\delta \leq 1$ , we are back to prove that for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s., we have

$$\delta(A_n \bar{x}, \bar{y}) \geq e^{-\varepsilon n^{1/p}} \quad \text{for all } n \text{ large enough.}$$

Now, it is well-known (see e.g. Definition 4.1 page 55 and (9) page 61 of [5]), that, for any  $x', y'$  in  $\mathbb{R}^d - \{0\}$ ,

$$\frac{|\langle x', y' \rangle|}{\|x'\| \|y'\|} = \sqrt{1 - (d(\bar{x}', \bar{y}'))^2}. \tag{67}$$

Hence

$$\delta^2(A_n \bar{x}, \bar{y}) = 1 - d^2(A_n \bar{x}, \bar{y}).$$

Now,

$$\begin{aligned} 1 - d^2(A_n \bar{x}, \bar{y}) &\geq 1 - (d(A_n \bar{x}, W_n) + d(W_n, \bar{y}))^2 \\ &\geq 1 - d^2(A_n \bar{x}, W_n) - 2d(A_n \bar{x}, W_n)d(W_n, \bar{y}) - d^2(W_n, \bar{y}). \end{aligned}$$

Since for every  $n \in \mathbb{N}$ ,  $W_n$  has law  $\nu$ , the variables  $(\log \delta(W_n, \bar{y}))_{n \in \mathbb{N}}$  are identically distributed in  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ . In particular, it follows from the Borel–Cantelli lemma that

$$|\log \delta(W_n, \bar{y})| = o(n^{1/p}), \quad \mathbb{P}_\nu\text{-a.s.} \tag{68}$$

Hence, for every  $\varepsilon > 0$ ,  $\mathbb{P}_\nu$ -a.s., we have

$$\delta(W_n, \bar{y}) \geq e^{-\varepsilon n^{1/p}} \quad \text{for all } n \text{ large enough,}$$

and using (67) again, for every  $\varepsilon > 0$ ,  $\mathbb{P}_\nu$ -a.s., we have

$$d(W_n, \bar{y}) \leq \sqrt{1 - e^{-2\varepsilon n^{1/p}}} \quad \text{for all } n \text{ large enough.} \quad (69)$$

As in the proof of Lemma 6, we may write

$$\log\left(\frac{d(A_n \bar{x}, W_n)}{d(\bar{x}, W_0)}\right) := M_n + R_n,$$

where  $(M_n)_{n \geq 1}$  is a (centered) martingale with increments dominated by a variable in  $L^{p+1}$  and  $(R_n)_{n \geq 1}$  is such that there exists  $\ell > 0$  such that

$$\sum_{n \geq 1} \mathbb{P}(R_n \geq -\ell n) < \infty.$$

It is well-known, since  $p + 1 \geq 2$ , that  $(M_n)_{n \geq 1}$  satisfies the strong law of large numbers (actually even the law of the iterated logarithm). Hence, we have,  $\mathbb{P}_\nu$ -a.s.

$$\log d(A_n \bar{x}, W_n) \leq -\ell n/2 \quad \text{for every } n \text{ large enough.}$$

Finally, we infer that, for every  $\varepsilon$ ,  $\mathbb{P}_\nu$ -a.s., we have

$$1 - d^2(A_n \bar{x}, \bar{y}) \geq 1 - e^{-\ell n} - e^{-\ell n} \sqrt{1 - e^{-2\varepsilon n^{1/p}}} - (1 - e^{-2\varepsilon n^{1/p}}) \geq C e^{-2\varepsilon n^{1/p}},$$

which is exactly what we wanted to prove.

## References

- [1] G. Alsmeyer. Convergence rates in the law of large numbers for martingales. *Stochastic Process. Appl.* **36** (2) (1990) 181–194. [MR1084974](#)
- [2] Y. Benoist and J.-F. Quint. Central limit theorem for linear groups. *Ann. Probab.* **44** (2) (2016) 1308–1340. [MR3474473](#)
- [3] Y. Benoist and J.-F. Quint. Random walks on reductive groups manuscript. Available at <http://www.math.u-psud.fr/~benoist/prepubli/prepublication.html>.
- [4] I. Berkes, W. Liu and W. B. Wu. Komlós–Major–Tusnády approximation under dependence. *Ann. Probab.* **42** (2) (2014) 794–817. [MR3178474](#)
- [5] P. Bougerol and J. Lacroix. *Products of Random Matrices with Applications to Schrödinger Operators. Progress in Probability and Statistics 8*. Birkhäuser, Boston, MA, 1985. [MR0886674](#)
- [6] C. Cuny. ASIP for martingales in 2-smooth Banach spaces. Applications to stationary processes. *Bernoulli* **21** (1) (2015) 374–400. [MR3322323](#)
- [7] C. Cuny. Invariance principles under the Maxwell–Woodroffe condition in Banach spaces. *Ann. Probab.*, 2016. To appear. Available at [arXiv:1403.0772](https://arxiv.org/abs/1403.0772).
- [8] C. Cuny and F. Merlevède. Strong invariance principles with rate for “reverse” martingale differences and applications. *J. Theoret. Probab.* **28** (1) (2015) 137–183. [MR3320963](#)
- [9] J. Dedecker, P. Doukhan and F. Merlevède. Rates of convergence in the strong invariance principle under projective criteria. *Electron. J. Probab.* **17** (2012) 16. DOI: [10.1214/EJP.v17-1849](https://doi.org/10.1214/EJP.v17-1849). [MR2900457](#)
- [10] J. Dedecker, F. Merlevède and E. Rio. Rates of convergence for minimal distances in the central limit theorem under projective criteria. *Electron. J. Probab.* **14** (35) (2009) 978–1011. [MR2506123](#)
- [11] J. Dedecker and E. Rio. On the functional central limit theorem for stationary processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **36** (2000) 1–34. [MR1743095](#)
- [12] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Stat.* **31** (1960) 457–469. [MR0121828](#)
- [13] Y. Guivarc’h and A. Raugi. Frontière de furstenberg, propriétés de contraction et théorèmes de convergence. *Z. Wahrsch. Verw. Gebiete* **69** (2) (1985) 187–242. [MR0779457](#)
- [14] S. Hao and Q. Liu. Convergence rates in the law of large numbers for arrays of martingale differences. *J. Math. Anal. Appl.* **417** (2) (2014) 733–773. [MR3194512](#)

- [15] C. C. Heyde. On the central limit theorem and iterated logarithm law for stationary processes. *Bull. Aust. Math. Soc.* **12** (1975) 1–8. [MR0372954](#)
- [16] C. Jan. Vitesse de convergence dans le TCL pour des chaînes de Markov et certains processus associés à des systèmes dynamiques. *C. R. Acad. Sci. Paris Sér. I Math.* **331** (5) (2000) 395–398. [MR1784921](#)
- [17] C. Jan. Vitesse de convergence dans le TCL pour des processus associés à des systèmes dynamiques ou des produits de matrices aléatoires. Thèse de l'université de Rennes, 1. Thesis number 01REN10073, 2001.
- [18] J. Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. I. *Z. Warsch. Verw. Gebiete* **32** (1975) 111–131. [MR0375412](#)
- [19] U. Krengel. *Ergodic Theorems. De Gruyter Studies in Mathematics* **6**. Walter de Gruyter, Berlin, 1985. [MR0797411](#)
- [20] E. Le Page. Théorèmes limites pour les produits de matrices aléatoires. In *Probability Measures on Groups* 258–303. *Oberwolfach, 1981. Lecture Notes in Math.* **928**. Springer, Berlin, 1982. [MR0669072](#)
- [21] M. Maxwell and M. Woodroffe. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28** (2) (2000) 713–724. [MR1782272](#)
- [22] F. Merlevède and M. Peligrad. Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. *Ann. Probab.* **41** (2) (2013) 914–960. [MR3077530](#)
- [23] M. Peligrad and S. Utev. A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.* **33** (2) (2005) 798–815. [MR2123210](#)
- [24] M. Peligrad, S. Utev and W. B. Wu. A maximal  $l_p$ -inequality for stationary sequences and its applications. *Proc. Amer. Math. Soc.* **135** (2) (2007) 541–550. [MR2255301](#)
- [25] Q. M. Shao. Almost sure invariance principles for mixing sequences of random variables. *Stochastic Process. Appl.* **48** (1993) 319–334. [MR1244549](#)
- [26] B. von Bahr and C.-G. Esseen. Inequalities for the  $r$ th absolute moment of a sum of random variables,  $1 \leq r \leq 2$ . *Ann. Math. Stat.* **36** (1965) 299–303. [MR0170407](#)
- [27] W. B. Wu and Z. Zhao. Moderate deviations for stationary processes. *Statist. Sinica* **18** (2) (2008) 769–782. [MR2411619](#)