Rates of convergence in invariance principles for random walks on linear groups via martingale methods

Christophe Cuny, Jérôme Dedecker, Florence Merlevède

Rates of convergence in invariance principles for random walks on linear groups via martingale methods

C. Cuny[*], J. Dedecker[†], Florence Merlevède[‡]

September 16, 2019

Abstract

In this paper, we give explicit rates in the central limit theorem and in the almost sure invariance principle for general $\mathbb{R}^d$-valued cocycles that appear in the study of the left random walk on linear groups. Our method of proof lies on a suitable martingale approximation and on a careful estimation of some coupling coefficients linked with the underlying Markov structure. Concerning the martingale part, the available results in the literature are not accurate enough to give almost optimal rates whether in the central limit theorem for the Wasserstein distance, or in the strong approximation. A part of this paper is devoted to circumvent this issue. We then exhibit near optimal rates both in the central limit theorem in terms of Wasserstein distance and in the almost sure invariance principle for $\mathbb{R}^d$-valued martingales with stationary increments having moments of order $p \in]2, 3]$ (the case of sequences of reversed martingale differences is also considered). Note also that, as an application of our results for general $\mathbb{R}^d$-valued cocycles, a special attention is paid to the Iwasawa cocycle and the Cartan projection for reductive Lie groups.

Keywords: Random walks, Strong invariance principle, Wasserstein distances, Vector-valued martingales, Iwasawa cocycle, Cartan projection.

MSC: 60F17, 60G42, 60G50, 22E40

[*]Université de Brest, LMBA, UMR CNRS 6205. Email: christophe.cuny@univ-brest.fr

[†]Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 (UMR 8145). Email: jerome.dedecker@parisdescartes.fr

[‡]Université Paris-Est, LAMA (UMR 8050), UPEM, CNRS, UPEC. Email: florence.merlevede@u-pem.fr

1 Introduction

Let $G = GL_d(\mathbb{R})$ be the group of invertible real matrices of order $d \geq 2$. Denote by $\|\cdot\|$ the operator norm on G associated with the euclidean norm on $\mathbb{R}^d$. Let μ be a probability on the Borel sets of G. Let $(Y_n)_{n\geq 1}$ be independent identically distributed (iid) variables with law μ. It is well-known that, if μ admits a moment of order 1 and under some irreducibility assumption on the support of μ, then the sequence $(\frac{1}{n}\log\|Y_n\cdots Y_1\|)_{n\geq 1}$ converges almost surely to $\lambda_\mu \in \mathbb{R}$ (see for instance Furstenberg and Kesten [16]).

If μ is further assumed to admit exponential moments and to satisfy a proximality property, Le Page [25] and Guivarc'h and Raugi [18] proved a central limit theorem (CLT) with rate as well as other probabilistic results, by spectral gap methods. The CLT has been obtained by Jan [21] under a moment of order $2+\varepsilon$, by mean of martingale approximation.

Very recently, Benoist and Quint [1] managed to prove the CLT under a moment of order 2 thanks to an ingenious and explicit martingale-coboundary decomposition. Later, Cuny, Dedecker and Jan [7] gave a different proof of Benoist-Quint's result. Using precise controls of some coupling coefficients of the underlying Markov chain and an explicit martingale approximation, they derived more probabilistic results such as almost sure invariance principles with rates. Next, Cuny, Dedecker and Merlevède [8] obtained different types of deviation results.

In all the above mentioned works, the first step of the proof consists in reducing the study to a suitable cocycle on $G \times X$, where X stands for the projective space of $\mathbb{R}^d$. Recall that $\sigma : G \times X \to \mathbb{R}$ is a cocycle if $\sigma(gg', x) = \sigma(g, g'\cdot x) + \sigma(g', x)$ for every $g, g' \in G$ and $x \in X$, where $\cdot$ denotes an action of G on X. For instance to deal with $(\log\|Y_n\cdots Y_1\|)_{n\geq 1}$, one can use the cocycle $\sigma(g,x) = \log\Big(\frac{\|g\cdot x\|}{\|x\|}\Big)$.

The central limit theorem for cocycles benefited from an active research in the last years. Let us mention, beside [1], [7] and [8], the works of Björklund [5], Benoist and Quint [2] and Horbez [19].

It happens that the arguments developped in [7] are somewhat general and apply equally to any cocycle on $\sigma : G \times X \to \mathbb{R}^d$ (G a locally compact group acting on a compact metric space X) provided that the action is suitably contracting and that the cocycle is Lipschitz in the second coordinate with Lipschitz norm $\sigma_{\mathrm{Lip}}(g)$ satisfying integrability conditions with respect to μ, see Definition 3.3 and Section 3.2 for more details.

Hence a first goal of the paper is to extend results of [7] to general $\mathbb{R}^d$-valued cocycles. As a main motivation we have in mind the Iwasawa cocycle to which our results apply and for which the CLT has been obtained by Benoist and Quint [1] under a moment of order 2. Then, as in

[7] or [8] many probabilistic results follow: strong laws of large numbers with rate, the CLT (and its functional form), deviation inequalities. For those results it is not an issue to treat vector-valued processes. However, there are several results in probability theory that do not extend easily from the one-dimensional case to the multivariate one. As a matter of fact, for $\mathbb{R}^d$-valued martingales, no almost sure invariance principle (ASIP) with *explicit* rate is available in the litterature. Indeed, the only known rates are of the form $O(n^{1/2-\varepsilon})$ for some $\varepsilon > 0$, see for instance Morrow and Philipp [30], Monrad and Philipp [29] and Eberlein [15], while in the one dimensional case, martingales with stationary increments in L^p for some for $2 < p < 4$, satisfy an ASIP with rate $o(n^{1/p}\sqrt{\log n})$. A similar rate also holds for $p = 4$ but no better rate is available for $p > 4$.

A second goal of the paper is then to prove an ASIP with explicit rates for $\mathbb{R}^d$-valued martingales with stationary increments, which is the main technical result of the present work. Our result hold when the martingales are in L^p for some $2 < p \leq 3$, so that the case $3 < p \leq 4$ remains open, contrary to the one-dimensional case. The method of proof also allows to derive rates of convergence in the CLT for $\mathbb{R}^d$-valued martingales with stationary increment, in terms of the Wasserstein distance W_1. Using then a suitable martingale approximation, we derive ASIP with explicit rates for general $\mathbb{R}^d$-valued cocycles.

Let us mention that, to prove the ASIP for cocycles, a second way would be to apply a multidimensional version of the strong approximation result of Berkes, Liu and Wu [4], as given in Karmakar and Wu [22]. However, a direct application of this result would not give the good rate of convergence with respect to the moment of μ (see the introduction of [9] for more details). Hence an adaptation of the result of [22], similar to what we did in [9] for real-valued cocycles, would be necessary here, and it is not clear wether this adaptation is feasible or not. Let us also mention that the ASIP for $\mathbb{R}^d$-valued martingales is interesting in itself, and can be useful in other situations.

The paper is organized as follows. In Section 2 we state our results for $\mathbb{R}^d$-valued martingales. In section 3, we describe the type of *weakly contractive* actions that we shall deal with, and obtain several preliminary results. Then, we describe the properties that the cocycles should satisfy to implement the arguments used in [7], and we derive several probabilistic results for those cocycles, emphasizing the consequences of our martingale results. We conclude Section 3 by an application to the Iwasawa cocycle and to the Cartan projection. The fact that the Iwasawa cocycle does satisfy our conditions follows from the recent book by Benoist and Quint [3]. Finally, Section 4 is devoted to the proofs and the appendix contains a useful Fuk-Nagaev type inequality for martingales.

2 Rates of convergence in the ASIP and the CLT for $\mathbb{R}^d$-valued martingales

In this section we consider a stationary sequence of random variables with values in $\mathbb{R}^d$ and defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a stationary and non-decreasing sequence of σ-algebras (see e.g. page 10 in [27] for the definition of a stationary filtration). We suppose that $(d_n)_{n\in\mathbb{N}}$ is a stationary sequence of martingale differences with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$, i.e. for each n, d_n is integrable, $\mathcal{F}_n$-measurable and such that $\mathbb{E}(d_n|\mathcal{F}_{n-1}) = 0$ $\mathbb{P}$-a.s.

For any random variable X with values in $\mathbb{R}^d$, we shall use the notation $(X)_i$ to mean its i-th coordinate. In addition $|\cdot|_d$ means the euclidean norm on $\mathbb{R}^d$.

In this section we give, under projective conditions, rates of convergence in the almost sure invariance principle for the partial sums associated with a stationary sequence of martingale differences with values in $\mathbb{R}^d$.

Notation 2.1. For any $p > 2$, define the envelope norm $\|\,.\,\|_{1,\Phi,p}$ by

$$\|X\|_{1,\Phi,p} = \int_0^1 (1 \vee \Phi^{-1}(1-u/2))^{p-2} Q_X(u)du$$

where Φ denotes the d.f. of the $N(0,1)$ law, and Q_X denotes the quantile function of $|X|$, that is the cadlag inverse of the tail function $x \to \mathbb{P}(|X| > x)$.

Remark 2.1. Let $a > 1$ and $p > 2$. Applying Hölder's inequality, we see that there exists a positive constant $C(p,a)$ such that $\|X\|_{1,\Phi,p} \le C(p,a)\|X\|_a$.

Theorem 2.1. *Let $(d_n)_{n\in\mathbb{Z}}$ be a $\mathbb{R}^d$-valued stationary sequence of martingale differences with respect to $(\mathcal{F}_n)_{n\in\mathbb{Z}}$. Let $M_n = \sum_{k=1}^n d_k$. Let $p \in]2,3]$. Assume that $\mathbb{E}|d_0|_d^p < \infty$ and that for any i, j such that $1 \le i, j \le d$,*

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{F}_0\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_{1,\Phi,p} < \infty\,, \tag{2.1}$$

and

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+2/p}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{F}_0\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_{p/2} < \infty\,. \tag{2.2}$$

Then,

1. *For any $\varepsilon > 0$, enlarging Ω if necessary, there exists a sequence $(N_i)_{i\ge1}$ of iid $\mathbb{R}^d$-valued centered gaussian random variables with $\mathrm{Var}(N_1) = \mathbb{E}(d_0 d_0^t)$ such that*

$$M_n - \sum_{i=1}^n N_i = \begin{cases} o(n^{1/p}(\log n)^{\frac{p+1}{2p}+\varepsilon}) \text{ if } p \in]2,3[\\ o(n^{1/3}(\log n)^{1+\varepsilon}) \text{ if } p = 3 \end{cases} \text{ almost surely.}$$

2. *Enlarging Ω if necessary, there exists a sequence $(N_i)_{i\geq 1}$ of iid $\mathbb{R}^d$-valued centered gaussian random variables with $\mathrm{Var}(N_1)=\mathbb{E}(d_0 d_0^t)$ such that*

$$\Big\| \sup_{1\leq k\leq n} \big| M_k - \sum_{i=1}^k N_i \big|_d \Big\|_1 = \begin{cases} O(n^{1/p}(\log n)^{\frac{p-1}{2p}}) \text{ if } p\in]2,3[\\ O(n^{1/3}(\log n)^{2/3}) \text{ if } p=3 \end{cases}.$$

Remark 2.2. According to Remark 2.1, if $p\in]2,3]$, condition (2.1) is implied by the condition: for some $a>1$,

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{F}_0\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_a < \infty\,. \tag{2.3}$$

Hence, both (2.1) and (2.2) hold as soon as

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{F}_0\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_{p/2} < \infty\,. \tag{2.4}$$

Let $\mathrm{Lip}(|\cdot|_d)$ be the set of Lipschitz functions g from $\mathbb{R}^d$ to $\mathbb{R}$ such that $|g(x)-g(y)|\leq |x-y|_d$. For two measures μ and ν on $\mathbb{R}^d$, let

$$W_1(\mu,\nu) = \sup_{g\in \mathrm{Lip}(|\cdot|_d)} (\mu(g)-\nu(g))\,.$$

Theorem 2.2. *Let $(d_n)_{n\in\mathbb{Z}}$ be a $\mathbb{R}^d$-valued stationary sequence of martingale differences with respect to $(\mathcal{F}_n)_{n\in\mathbb{Z}}$ satisfying the assumptions of Theorem 2.1. Let G_Σ be a $\mathbb{R}^d$-valued centered gaussian random variables with $\mathrm{Var}(G_\Sigma)=\mathbb{E}(d_0 d_0^t):=\Sigma$. Then*

$$\big\| W_1\left(P_{n^{-1/2}M_n|\mathcal{F}_0}, P_{G_\Sigma}\right)\big\|_1 \leq \begin{cases} Cn^{(2-p)/2} & \text{if } p\in]2,3[\\ Cn^{-1/2}\log n & \text{if } p=3 \end{cases},$$

where C is a positive constant depending on (p,d) but not on n.

Remark 2.3. It follows from Theorem 2.2 that, for any bounded $\mathcal{F}_0$-measurable random variable Z,

$$\sup_{g\in\mathrm{Lip}(|\cdot|_d)} \left(\mathbb{E}\left(Zg\left(n^{-1/2}M_n\right)\right) - \mathbb{E}(Z)\mathbb{E}(g(G_\Sigma))\right) \leq \begin{cases} Cn^{(2-p)/2} & \text{if } p\in]2,3[\\ Cn^{-1/2}\log n & \text{if } p=3 \end{cases}.$$

In particular, taking $Z\equiv 1$, we obtain that

$$W_1\left(P_{n^{-1/2}M_n}, P_{G_\Sigma}\right) \leq \begin{cases} Cn^{(2-p)/2} & \text{if } p\in]2,3[\\ Cn^{-1/2}\log n & \text{if } p=3 \end{cases}.$$

From the above result, proceeding as in [11], one can derive rates in the central limit theorem with respect to W_1 for the normalized partial sums of a large class of stationary sequence of $\mathbb{R}^d$-valued dependent random variables satisfying some mixingale type conditions. For related results under other types of dependence conditions, let us mention the paper by Pène [31].

Remark 2.4 (Reversed martingale differences sequences)**.** Let $p \in]2,3]$. Assume that $(d_n)_{n\in\mathbb{Z}}$ is a $\mathbb{R}^d$-valued stationary sequence of *reversed* martingale differences in $\mathbb{L}^p$ with respect to a stationary and non-increasing sequence $(\mathcal{G}_n)_{n\in\mathbb{Z}}$ of σ-algebras. This means that for any integer n, d_n is $\mathcal{G}_n$-adapted and $\mathbb{E}(d_n|\mathcal{G}_{n+1}) = 0$ a.s. Let $M_n = \sum_{k=1}^n d_k$. We infer that the conclusions of Theorem 2.1 and Theorem 2.2 hold for M_n (with $\mathcal{G}_{n+1}$ in place of $\mathcal{F}_0$ in Theorem 2.2) provided that the conditions (2.1) and (2.2) are replaced by the following ones:

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{G}_{n+1}\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_{1,\Phi,p} < \infty \,, \tag{2.5}$$

and

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+2/p}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{G}_{n+1}\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_{p/2} < \infty \,. \tag{2.6}$$

See Section 4.1.5 for a short proof of this remark.

Proceeding, as in [10], this type of result for reversed martingale differences sequences allows to derive rates in the strong invariance principle for $\mathbb{R}^d$-valued observables of a large class of dynamical systems (see also [23] where reversed martingale approximations in $\mathbb{R}^d$ are provided that could be used to verify conditions such as (2.5) and (2.6)). This reversed martingale approximation method allows to derive rates up to $n^{1/3}$ (up to additional logarithm terms) which improves, in case of $\mathbb{R}^d$-valued bounded Hölder observables, the rates obtained in [26] which depend on the dimension d (in particular, when d is large, the rates in [26] are close to $1/2$). Note that, for a class of processes with exponential decay of correlations, the rate $n^{1/4+\varepsilon}$ for some $\varepsilon > 0$ can be reached for $\mathbb{R}^d$-valued observables, by using the method developed in [17].

3 Weakly contracting actions

3.1 Definitions and properties

We shall work in the general situation described in section 11 of [3]. At first reading one may assume that the group F below is trivial (i.e. reduced to its neutral element). In our applications, F will be trivial as soon as G is connected (see Remark 3.2).

Let G be a locally compact second countable group. Let $s : G \to F$ be a continuous homomorphism onto a finite group F.

Let μ be a probability on the Borel sets $\mathcal{B}(G)$ of G. Let $s(\mu) := \mu \circ s^{-1}$, and Δ_μ the support of $s(\mu)$.

Definition 3.1. We say that μ is *F-adapted* if Δ_μ spans F as a subgroup. We say that μ is F-strictly aperiodic if it is F-adapted and if the smallest normal subgroup of F a class of which contains Δ_μ is F.

Remark 3.1. Our terminology is different from [3]. Our definition of F-strict aperiodicity corresponds to their definition of *aperiodicity* (see Derriennic-Lin [14, Prop. 1.6] for a proof of the equivalence of the definitions).

Remark 3.2. When F is trivial, μ is automatically F-adapted and F-strictly aperiodic. If G is connected then $s(G) = \{e_F\}$ (e_F is the neutral element of F). In particular, in that case, if μ is F-adapted, F must be trivial.

Let X be a compact and second countable metric space on which G acts continuously (we denote that action by $g \cdot x$). Assume that X is fibered over F, meaning that there exists a G-equivariant continuous mapping $f : X \to F$. Recall that G-equivariance means that $f(g \cdot x) = s(g)f(x)$ for every $(g, x) \in G \times X$.

Notice that when X is fibered over F then the sets $(X_\mathfrak{f})_{\mathfrak{f}\in F} := (f^{-1}(\{\mathfrak{f}\}))_{\mathfrak{f}\in F}$ are compact and open and obviously form a partition of F. Those sets are called the fibers.

Before going further, to motivate our definitions, let us describe some situations to which our results will apply. Take $G = GL_d(\mathbb{R})$. Then, it is well known that we have an homeomorphism $G = KN$ where $K = O_d(\mathbb{R})$ is the group of orthogonal matrices of size d and N is the group of upper triangular matrices of size d with positive entries on the diagonal. We take $X := G/N$ which is homeomorphic to K (hence compact). Then, if $x = kN \in X$, we set $f(x) = \operatorname{sgn}\det(k)$, the sign of the determinant of k. Similarly, for $g \in G$ we set $s(g) = \operatorname{sgn}\det(g)$. The fact that the conditions imposed below may be satisfied will be explained in Section 3.4.

One may also take $G := SL_d(\mathbb{R})$. Then, $K = SO_d(\mathbb{R})$ and f and s become trivial.

When $G = GL_d(\mathbb{R})$ or $SL_d(\mathbb{R})$, one may also take for X the projective space of $\mathbb{R}^d$. This special case has been handled in [7] and [8].

Definition 3.2. We say that a probability ν on the Borel sets $\mathcal{B}(X)$ of X is μ-invariant if

$$\int_{G\times X} h(g \cdot u)\mu(dg)\nu(du) = \int_X h(u)\nu(du)\,,$$

for every bounded Borel function h on X.

For every $g \in G$, set

$$\operatorname{Lip}(g) := \sup_{x\neq y, f(x)=f(y)} \frac{d(g \cdot x, g \cdot y)}{d(x, y)}\,, \tag{3.1}$$

where the supremum is over all $x, y \in X$ such that $x \neq y$ and $f(x) = f(y)$.

Definition 3.3. Let $p \geq 1$. We say that the action of G on (X, d) is *weakly* (μ, p)-*contracting* if

$$\int_G \left(\log^+(\mathrm{Lip}(g))\right)^p \, \mu(dg) < \infty \,, \tag{3.2}$$

where $\log^+(x) = \log(\max(1, x))$, and if there exists $n_0 \in \mathbb{N}$, such that

$$\sup_{x \neq y, f(x)=f(y)} \int_G \log\Big(\frac{d(g \cdot x, g \cdot y)}{d(x, y)}\Big) \, \mu^{*n_0}(dg) < 0 \,. \tag{3.3}$$

When $p = 1$ we shall just say that the action is weakly μ-contracting.

Remark 3.3. Notice that the left-hand side in (3.3) may be $-\infty$. Benoist and Quint [3, Definition 11.1] called an action (μ, γ)-*contracting* for some $\gamma > 0$, if

$$\int_G (\mathrm{Lip}(g))^\gamma \, \mu(dg) < \infty \,, \tag{3.4}$$

and

$$\sup_{x \neq y, f(x)=f(y)} \int_G \Big(\frac{d(g \cdot x, g \cdot y)}{d(x, y)}\Big)^\gamma \, \mu^{*n_0}(dg) < 1 \,. \tag{3.5}$$

Using Jensen's inequality, one can see that (3.4) implies (3.2) and that (3.5) implies (3.3).

Remark 3.4. Let $\alpha \in (0, 1]$, then $d_\alpha(x, y) := (d(x, y))^\alpha$ defines another metric on X. Let $p \geq 1$. For every $\alpha \in (0, 1]$, the action of G on (X, d) is weakly (μ, p)-contracting, for some $p \geq 1$ if only if the action of G on (X, d_α) is.

Benoist and Quint [3] proved (see their Lemma 11.5 p. 171) that if the action is (μ, γ)-contracting, for some $\gamma > 0$, then there exists a unique μ-invariant probability on $\mathcal{B}(X)$.

We shall prove that their result remains true under the weaker assumption that the action is weakly μ-contracting. As in [3], the proof makes use of the left (and right) random walk on X.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Assume that there exists $(Y_n)_{n \geq 1}$ iid random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in G with common law μ. Define $A_n := Y_n \cdots Y_1$ and $B_n = Y_1 \cdots Y_n$ for every $n \geq 1$ and $A_0 = \mathbf{e}$, where $\mathbf{e}$ stands for the neutral element of G.

Proposition 3.1. *Assume that the action is weakly μ-contracting and that μ is F-adapted. Then, there exists a unique μ-invariant probability on $\mathcal{B}(X)$.*

The proof is based on the following two lemmas. The proofs of the proposition and of the lemmas below are postponed to Section 4.2.

Lemma 3.2. *Assume that μ is F-adapted. Let ν be a μ-invariant probability on $\mathcal{B}(X)$. Then, $\nu(X_f) = \frac{1}{|F|}$ for every $f \in F$.*

Lemma 3.3. *Assume that the action is weakly μ-contracting. Then, there exists $\ell > 0$ such that*

$$\sum_{n\geq 0} \sup_{x\neq y, f(x)=f(y)} \mathbb{P}\left(\log\left(\frac{d(B_{2^n}\cdot x, B_{2^n}\cdot y)}{d(x,y)}\right) \geq -2^n\ell\right) < \infty\,. \tag{3.6}$$

Remark 3.5. Of course, since $(A_n)_{n\geq 1}$ has the same law as $(B_n)_{n\geq 1}$, the lemma holds with A_{2^n} in place of B_{2^n}. However, we shall need the above form in the proof of Proposition 3.1.

We shall see now that for weakly-contracting actions, the trajectories (starting from a same fiber) of the left random walk on X are almost surely arbitrary close. The next lemma is a version of Lemma 6 in [7].

Lemma 3.4. *Let $p \geq 1$. Assume that the action is weakly (μ, p)-contracting. Then, there exists $\ell > 0$, such that*

$$\sum_{k\geq 1} k^{p-2} \max_{k\leq j\leq 2k} \sup_{x\neq y, f(x)=f(y)} \mathbb{P}\left(\log\left(d(A_{j-1}\cdot x, A_{j-1}\cdot y)\right) \geq -\ell k\right) < \infty\,. \tag{3.7}$$

Moreover, there exists $\delta > 0$ such that for every $x, y \in X$, with $f(x) = f(y)$,

$$d(A_n\cdot x, A_n\cdot y) = O(\mathrm{e}^{-\delta n}) \qquad \mathbb{P}\text{-}a.s. \tag{3.8}$$

3.2 Cocycles over weakly contracting actions

Our goal is to obtain limit theorems for *cocyles* associated with a weakly μ-contracting G-action on a compact metric space X that is fibered over F, as in the previous section. We shall be concerned with cocycles taking values in a finite dimensional $\mathbb{R}$-vector space E, equipped with an euclidean norm $|\cdot|_E$.

Definition 3.4. We say that $\sigma : G\times X \to E$ is a cocycle if for every $g, g' \in G$ and every $u \in X$,

$$\sigma(gg', u) = \sigma(g, g'\cdot u) + \sigma(g', u)\,. \tag{3.9}$$

Of course we need some regularity assumptions on the considered cocycles. In order to state the needed assumptions we shall introduce some notations.

Given a cocycle $\sigma : G\times X \to E$, define, for every $g \in G$,

$$\sigma_{\sup}(g) := \sup_{x\in X} |\sigma(g,x)|_E\,, \tag{3.10}$$

$$\sigma_{\mathrm{Lip}}(g) := \sup_{x\neq y, f(x)=f(y)} \frac{|\sigma(g,x) - \sigma(g,y)|_E}{d(x,y)} \tag{3.11}$$

$$\text{and} \qquad \kappa_0(g) := \max\left(\sigma_{\sup}(g), \log(\sigma_{\mathrm{Lip}}(g)\right)\,. \tag{3.12}$$

Definition 3.5. We say that a cocycle $\sigma : G \times X \to E$ has a polynomial moment of order $p \geq 1$ if

$$\int_G \kappa_0^p(g)\, \mu(dg) < \infty\,. \tag{3.13}$$

Remark 3.6. Let $\sigma : G \times X \to E$ be a cocycle. Let $(e_i)_{1\leq i\leq d}$ be an orthonormal basis of E. Then, for every $1 \leq i \leq d$, $(g,x) \mapsto \sigma_i(g,x) := \langle e_i, \sigma(g,x)\rangle$ is an $\mathbb{R}$-valued cocycle and it is not hard to see that σ has a polynomial moment of order $p \geq 1$ if and only if σ_i has a polynomial moment of order $p \geq 1$ for every $1 \leq i \leq d$. Because of this it will be sufficient to deal with $\mathbb{R}$-valued cocycles.

Hence, due to the remark above, we assume that the dimension $d = 1$ in the remaining of this subsection.

We first recall the strong law of large numbers for cocyles. This follows here from Theorem 3.9 of [3], using Proposition 3.1.

Proposition 3.5. *Assume that the action is weakly μ-contracting and that μ is F-adapted. Assume moreover that $\sigma_{\sup} \in L^1(\mu)$. Then, for every $x \in X$,*

$$\frac{\sigma(A_n, x)}{n} \underset{n\to\infty}{\longrightarrow} \int_{G\times X} \sigma(g,x)\mu(dg)\nu(dx) =: \lambda_\mu\,,$$

$\mathbb{P}$-a.s. and in $L^1(\mathbb{P})$. Moreover, the convergence in $L^1(\mathbb{P})$ is uniform over $x \in X$.

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

The next result is a version of Lemma 5 of [7]. The proof being identical, it is therefore omitted.

Lemma 3.6. *For every $r > 1$, there exists $C_r > 0$ such that for every $g \in G$ and every $x, y \in X$, with $f(x) = f(y)$,*

$$|\sigma(g,x) - \sigma(g,y)| \leq C_r(1 + \kappa_0(g))^r\, H_{r-1}(d(x,y))\,. \tag{3.14}$$

Lemma 3.6 allows us to prove that the almost sure behaviour of the cocycle along the random walk does not depend on the starting point of a given fiber.

Lemma 3.7. *Let $p > 1$. Assume that the action is weakly (μ, p)-contracting, that μ is F-adapted and that σ admits a moment of order p. Then, for every $x, y \in X$ such that $f(x) = f(y)$,*

$$|\sigma(A_n, x) - \sigma(A_n, y)| = O(1) \qquad \mathbb{P}\text{-a.s.}$$

Proof. Since σ admits a moment of order p, it is a well-known consequence of the Borel-Cantelli lemma that $\kappa_0(Y_k) = o(k^{1/p})$ $\mathbb{P}$-a.s. Let $x, y \in X$ be such that $f(x) = f(y)$. Let $r > 2p/(p-1)$. Using the cocycle property, (3.14) and (3.8), we have, $\mathbb{P}$-a.s.

$$|\sigma(A_n, x) - \sigma(A_n, y)| \leq \sum_{k=1}^{n} |\sigma(Y_k, A_{k-1} \cdot x) - \sigma(Y_k, A_{k-1} \cdot y)|$$

$$\leq C_r(1 + \kappa_0(Y_k))^r \sum_{k=1}^{n} H_{r-1}(d(A_{k-1} \cdot x, A_{k-1} \cdot y)) \leq \tilde{C}_r \sum_{k=1}^{n} \frac{k^{r/p}}{k^{r-1}} \leq \sum_{k \geq 1} \frac{\tilde{C}_r}{k^{r(p-1)/p-1}} < \infty \,,$$

and the result follows.

Proposition 3.8. *Let $p > 1$. Assume that the action is weakly (μ, p)-contracting, that μ is F-adapted and that σ admits a moment of order p. Then, for $q \in [1, p)$,*

$$\sum_{k=1}^{\infty} k^{p-q-1} \sup_{f(x)=f(y)} \mathbb{E}\left(|\sigma(Y_k, A_{k-1} \cdot x) - \sigma(Y_k, A_{k-1} \cdot y)|^q\right) < \infty \,, \tag{3.15}$$

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

$$\sum_{k=1}^{\infty} k^{p-2} \sup_{f(x)=f(y)} \mathbb{E}\left(|\sigma(Y_k, A_{k-1} \cdot x) - \sigma(Y_k, A_{k-1} \cdot y)|^q\right) < \infty \,. \tag{3.16}$$

The proof of the proposition may be done as the proof of Proposition 3 of [7], using Lemmas 3.4 et 3.6.

Proposition 3.8 gives us a control along trajectories starting from a same fiber. To deal with trajectories starting from arbitrary fibers we shall need the following lemma (see for instance [3, Lemma 11.6b]).

Lemma 3.9. *Assume that μ is F-strictly aperiodic. There exist $C > 0$ and $0 < \rho < 1$ such that for every bounded function φ on F,*

$$\sup_{\mathfrak{f} \in F} \left| \mathbb{E}\big(\varphi(s(A_n)^{-1}\mathfrak{f}\big) - \frac{1}{|F|} \sum_{\mathfrak{f}' \in F} \varphi(\mathfrak{f}') \right| \leq C\rho^n \sup_{\mathfrak{f} \in F} |\varphi(\mathfrak{f})| \qquad \forall n \in \mathbb{N} \,.$$

Proposition 3.10. *Let $p > 1$. Assume that the action is weakly (μ, p)-contracting, that μ is F-strictly aperiodic and that σ admits a moment of order p. Then,*

$$\sum_{k \geq 1} k^{p-2} \sup_{x \in X} |\mathbb{E}(\sigma(Y_k, A_{k-1} \cdot x)) - \lambda_\mu| < \infty \,. \tag{3.17}$$

Proof. One easily sees that $\left(\sup_{x\in X}\left|\mathbb{E}(\sigma(Y_k, A_{k-1}\cdot x)) - \lambda_\mu\right|\right)_{k\geq 1}$ is non increasing. Hence, it is enough to prove that

$$\sum_{k\geq 1} k^{p-2} \sup_{x\in X} |\mathbb{E}(\sigma(Y_{2k}, A_{2k-1}\cdot x)) - \lambda_\mu| < \infty\,.$$

We shall make use of the following identity based on Lemma 3.2

$$\sigma(Y_{2k}, A_{2k-1}\cdot x) = |F| \sum_{\mathfrak{f}\in F} \int_{X_\mathfrak{f}} \nu(dy)\sigma(Y_{2k}, A_{2k-1}\cdot x)\, \mathbf{1}_{X_\mathfrak{f}}(A_{k-1}\cdot x)\,.$$

Notice that $A_{k-1}\cdot x \in X_\mathfrak{f}$ if and only if $s(A_{k-1})^{-1}\mathfrak{f} = f(x)$. Using Lemma 3.9 and independence, we infer that

$$\begin{aligned}
&|\mathbb{E}(\sigma(Y_{2k}, A_{2k-1}\cdot x)) - \lambda_\mu| \\
&\quad = \left| \mathbb{E}\left(|F| \sum_{\mathfrak{f}\in F} \int_{X_\mathfrak{f}} \Big(\sigma(Y_{2k}, A_{2k-1}\cdot x) - \sigma(Y_{2k}, Y_{2k-1}\dots Y_k\cdot y)\Big)\, \mathbf{1}_{X_\mathfrak{f}}(A_{k-1}\cdot x)\nu(dy)\right)\right. \\
&\qquad\qquad \left. + |F| \sum_{\mathfrak{f}\in F} \int_{X_\mathfrak{f}} \mathbb{E}\big(\sigma(Y_{2k}, Y_{2k-1}\dots Y_k\cdot y)\big)\nu(dy)\, \left(\mathbb{E}\left(\mathbf{1}_{\{f(x)\}}(s(A_{k-1})^{-1}\mathfrak{f})\right) - \frac{1}{|F|}\right)\right| \\
&\quad \leq \sup_{y\,:\, f(y)=f(x)} \mathbb{E}\left(|\sigma(Y_k, A_{k-1}\cdot x) - \sigma(Y_k, A_{k-1}\cdot y)|\right) + C\rho^{k-1}|F| \int_G \sigma_{\sup}(g)\mu(dg)\,.
\end{aligned}$$

Then, the result follows from (3.15) with $q = 1$. □

It is also possible to prove the following version of Proposition 4 of [7]. Define for every integer j and every $x \in X$, $\tilde{X}_j(x) := \sigma(Y_j, A_{j-1}\cdot x) - \lambda_\mu$.

Proposition 3.11. *Let $p > 2$. Assume that the action is weakly (μ, p)-contracting, that μ is F-adapted and that σ admits a moment of order p. Then*

$$\sum_{k\geq 1} k^{p-3} \sup_{f(x)=f(y)} \mathbb{E}\left(\left|\tilde{X}_n^2(x) - \tilde{X}_n^2(y)\right|\right) < \infty\,, \tag{3.18}$$

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

$$\sum_{k\geq 1} k^{\gamma} \sup_{f(x)=f(y)} \sup_{k\leq j<i\leq 2k} \mathbb{E}\left(\left|\tilde{X}_i(x)\tilde{X}_j(x) - \tilde{X}_i(y)\tilde{X}_j(y)\right|\right) < \infty\,. \tag{3.19}$$

Proposition 3.12. *Let $p > 2$. Assume that the action is weakly (μ, p)-contracting, that μ is F-strictly aperiodic and that σ admits a moment of order p. Then,*

$$\sum_{k\geq 1} k^{p-3} \sup_{x\in X} \left|\mathbb{E}\left(\tilde{X}_k^2(x)\right) - \int_X \mathbb{E}(\tilde{X}_k^2(y))\nu(dy)\right| < \infty\,, \tag{3.20}$$

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

$$\sum_{k\geq 1} k^{\gamma} \sup_{x\in X} \sup_{k\leq j<i\leq 2k} \left| \mathbb{E}\left(\tilde{X}_i(x)\tilde{X}_j(x)\right) - \int_X \mathbb{E}\left(\tilde{X}_i(y)\tilde{X}_j(y)\right)\nu(dy)\right| < \infty\,. \tag{3.21}$$

The proof is similar to that of Proposition 3.10, hence is omitted.

Finally, we state without proof the following version of Lemma 13 of [7].

Lemma 3.13. *Let* $p \geq 2$. *Assume that the action is weakly* (μ, p)*-contracting, that* μ *is* F*-adapted and that* σ *admits a moment of order* p. *Then*

$$\sup_{f(x)=f(y)} \|\sigma(A_n, x) - \sigma(A_n, y)\|_1 = O(1)$$

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

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

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

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

Remark 3.7. If F is trivial, in particular (see Remark 3.2) if G is connected, then the proposition holds with $\sup_{x,y\in X}$ rather than $\sup_{f(x)=f(y)}$.

3.3 Rates of convergence in the CLT and the ASIP

Thanks to the results of the previous section we can obtain many probabilistic results, as in [7] or [8].

Let $p \geq 1$. With the notations and definitions of the previous section, assume that the action of G on X is weakly (μ, p)-contracting, that μ is F-strictly aperiodic and that σ is a $\mathbb{R}$-valued cocycle admitting a moment of order p. Then, the conclusion of Theorems 1 and 2 of [7] hold provided that $S_{n,\overline{x}}$ is replaced with $\sigma(A_n, x)$.

If we consider an $\mathbb{R}^d$-valued cocycle, $d \geq 2$, only the items (i) and (ii) of Theorem 1 of [7] holds without any change. To extend their item (iii) to an $\mathbb{R}^d$-valued cocycle, an application of our Theorem 2.1 gives the following.

Theorem 3.14. *Let $p \in]2,3]$. Assume that the action is weakly (μ,p)-contracting and that μ is F-strictly aperiodic. Let W_0 be a random variable, with law ν, independent from $(Y_n)_{n\geq 1}$. Let σ be an $\mathbb{R}^d$-valued cocycle admitting a moment of order p. Then the series of matrices*

$$\Sigma = \mathrm{Var}(\sigma(Y_1,W_0)) + 2\sum_{k=2}^{\infty} \mathrm{Cov}(\sigma(Y_1,W_0),\sigma(Y_k,A_{k-1}W_0)) \tag{3.22}$$

converges absolutely. Moreover,

1. *For any $\varepsilon > 0$, enlarging Ω if necessary, there exists a sequence $(N_i)_{i\geq 1}$ of iid $\mathbb{R}^d$-valued centered gaussian random variables with variance Σ such that*

$$\sigma(A_n,W_0) - n\lambda_\mu - \sum_{i=1}^{n} N_i = \begin{cases} o(n^{1/p}(\log n)^{\frac{p+1}{2p}+\varepsilon}) \text{ if } p\in]2,3[\\ o(n^{1/3}(\log n)^{1+\varepsilon}) \text{ if } p=3 \end{cases} \text{ almost surely.}$$

2. *Enlarging Ω if necessary, there exists a sequence $(N_i)_{i\geq 1}$ of iid $\mathbb{R}^d$-valued centered gaussian random variables with variance Σ such that*

$$\Big\| \sup_{1\leq k\leq n} \Big|\sigma(A_k,W_0) - k\lambda_\mu - \sum_{i=1}^{k} N_i\Big|_d \Big\|_1 = \begin{cases} O(n^{1/p}(\log n)^{\frac{p-1}{2p}}) \text{ if } p\in]2,3[\\ O(n^{1/3}(\log n)^{2/3}) \text{ if } p=3 \end{cases}.$$

Moreover an application of our Theorem 2.2 (see Remark 2.3) gives the following result concerning rates in the CLT in terms of Wasserstein distance of order 1.

Theorem 3.15. *Let $p \in]2,3]$. Assume that the action is weakly (μ,p)-contracting and that μ is F-strictly aperiodic. Let W_0 be a random variable, with law ν, independent from $(Y_n)_{n\geq 1}$. Let σ be an $\mathbb{R}^d$-valued cocyle admitting a moment of order p. Let ν_n be the distribution of $n^{-1/2}(\sigma(A_n,W_0)-n\lambda_\mu)$, and let Σ be defined by (3.22). Then*

$$W_1(\nu_n, P_{G_\Sigma}) \leq \begin{cases} Cn^{(2-p)/2} & \text{if } p\in]2,3[\\ Cn^{-1/2}\log n & \text{if } p=3 \end{cases},$$

where C is a positive constant depending on (p,d) but not on n.

It is unclear, and probably not true in general, whether the above results hold for $(\sigma(A_n,x))_{n\geq 1}$ (for every $x\in X$) rather than $(\sigma(A_n,W_0))_{n\geq 1}$. However, when F is the trivial group, using Lemmas 3.7 or 3.13, one sees that the above results hold true for $(\sigma(A_n,x))_{n\geq 1}$, for every $x\in X$ (the same normal variables being used for all x).

3.4 Applications to the Iwasawa cocycle for reductive Lie groups

In this section, we shall give a general situation to which the previous sections apply.

Our presentation will borrow notations as well as results from the monograph by Benoist and Quint [3]. We refer to [3] for any complements on the topic discussed in that section. However, for the sake of clarity, we shall not cover the generality treated in [3]. In particular, we shall only condider algebraic groups over the field of real numbers, while the results presented here extend to local fields modulo several technical issues.

We give the definition of the Iwasawa cocycle in a general setting and we will make explicit all the objects in the particular case of the d-dimensional linear group.

Let G be a reductive algebraic real Lie group with lie algebra $\mathfrak{g}$. Recall that G is said to be reductive if its unipotent radical, that is the greatest connected normal subgroup of G whose elements are unipotent, is reduced to $\{e\}$, where e stands for the neutral element of G. Recall also that G is algebraic if it is the set of solutions of a (finite) system of polynomial equations (over $\mathbb{R}$).

Define the killing form on $\mathfrak{g} \times \mathfrak{g}$ by

$$\text{killing}(x, y) := \text{tr}(\text{ad}\, x \,\text{ad}\, y) \qquad \forall x, y \in \mathfrak{g}\,.$$

Let K be a maximal compact subgroup of G with Lie algebra $\mathfrak{k}$. Let $\mathfrak{s}$ be the orthogonal space of $\mathfrak{k}$ for the killing form and let $\mathfrak{a}$ be a Cartan subspace of $\mathfrak{s}$, that is $\mathfrak{a}$ is a commutative subalgebra of $\mathfrak{s}$ whose elements are diagonalizable over $\mathbb{R}$ and which is maximal with those properties. Let $A := \exp \mathfrak{a}$ be the corresponding connected lie subgroup of G.

Let U be a maximal unipotent subgroup of G that is normalized by A. Let $P := N_G(U)$ be the normalizer of U in G.

Denote by G_c, K_c and P_c the respective connected components of G, K and P.

We have the Iwasawa decomposition (the proof is sketched page 130 of [3])

$$G = KP_c = KAU\,. \tag{3.23}$$

The compact space $\mathcal{P} := G/P_c$ is known as the flag variety of G. Given $\eta = kP_c$ with $k \in K$ and $g \in G$, there exists a unique element z of $\mathfrak{a}$ such that $gk \in K \exp z\, U$. We denote by $\sigma(g, \eta)$ that unique element.

Then, σ is a continuous cocycle known as the Iwasawa cocycle. It is related to the Cartan projection that we shall introduce now.

Let Σ be the set of roots associated with $\mathfrak{a}$, that is, denoting by $\mathfrak{a}^*$ the dual vector space of $\mathfrak{a}$,

$$\Sigma := \{\alpha \in \mathfrak{a}^* - \{0\} \,:\, \mathfrak{g}^\alpha \neq \{0\}\}\,,$$

where

$$\mathfrak{g}^\alpha := \{y \in \mathfrak{g} \,:\, \forall\, x \in \mathfrak{a},\, \mathrm{ad}\, x(y) \,= \alpha(x) y\}\,.$$

Let Σ^+ be the set of positive roots associated with U, that is, if $\mathfrak{u}$ stands for the lie algebra of U, Σ^+ is characterized by

$$\mathfrak{u} = \oplus_{\alpha \in \Sigma^+} \mathfrak{g}^\alpha\,.$$

Define then $\mathfrak{a}^+ := \{x \in \mathfrak{a} \,:\, \forall \alpha \in \Sigma^+,\, \alpha(x) \geq 0\}$. We have the Cartan decomposition (see also page 130 of [3])

$$G = K \exp \mathfrak{a}^+ K_c\,.$$

Finally, define the Cartan projection as follows: for every $g \in G$, let $\kappa(g)$ be the unique $z \in \mathfrak{a}^+$ such that $g \in K \exp z\, K_c$.

For people not familiar with Lie groups we consider now the case where $G = \mathrm{Gl}_d(\mathbb{R})$, $d \geq 2$, and describe the different spaces introduced above.

In that case, one may take $K = \mathrm{O}_d(\mathbb{R})$ the orthogonal group, A the group of invertible diagonal matrices of size d with positive entries, U the group of upper triangular matrices whose diagonal terms are all 1 and P the subgroup of invertible upper triangular matrices.

Let us describe the corresponding Lie algebras are. We have $\mathfrak{g} = M_d(\mathbb{R})$, the set of matrices of size d. Then, the Killing form is given by $\mathrm{killing}(M, N) = 2d\,\mathrm{tr}(MN) - 2\mathrm{tr}M\mathrm{tr}N$ for every $M, N \in \mathfrak{g}$. Then, $\mathfrak{k} = \{M \in \mathfrak{g} \,:\, M = -M^t\}$ the set of antisymmetric matrices; $\mathfrak{s}$ is the set of symmetric matrices of size d and $\mathfrak{a}$ the set of diagonal real matrices of size d. Then, $G_c = \{M \in G \,:\, \det M > 0\}$, $K_c = \mathrm{SO}_d(\mathbb{R})$ and P_c is the group of upper triangular matrices with non-negative entries on the diagonal.

Finally, the set of positive roots is given by $\Sigma^+ = \{\alpha_{ij} \,:\, 1 \leq i < j \leq d\}$ where for $M = \mathrm{diag}(a_m)_{1 \leq m \leq d}$, $\alpha_{ij}(M) = a_i - a_j$ and $\mathfrak{u}$ is the set of upper triangular matrices with vanishing diagonal. In particular, $\mathfrak{a}^+$ consists of the set of diagonal matrices with entries in non-increasing order.

In this case, the Iwasawa decomposition may be seen as a by-product the Gram-Schmidt orthonormalization algorithm. Indeed, if $M \in G$ then M sends the canonical basis to a non necessarily orthonormal basis. Orthonormalizing the latter thanks to the Gram-Schmidt algorithm one sees that there exists an upper triangular matrix T with positive diagonal coefficients such that MT sends the canonical basis to an orthonormal basis, i.e. $MT \in O_d(\mathbb{R})$.

The Cartan projection also has a nice interpretation in this case. It may be seen as a by-product of the polar decomposition. In particular, for every $M \in G$ there exist unique matrices $K, K' \in O_d(\mathbb{R})$ and $A \in G$ such that $\det K' = 1$, A is diagonal with positive diagonal coefficients in non decreasing order. The coefficients $(\lambda_i)_{1\leq i\leq d}$ of A are the square roots of the eigenvalues of M^tM. Hence, $\kappa(M) = \operatorname{diag}(\log \lambda_i)_{1\leq i\leq d}$.

Let us come back to the general case. It follows from section 13.1 of [3] that $\mathcal{P}$ may be endowed with a metric d compatible with the quotient topology on $\mathcal{P}$. The metric is defined by (13.3) (notice that in our situation, i.e. G is a real algebraic Lie group, $\Theta = \Pi$, where Π is the set of simple roots, that is the roots of Σ^+ that are not the sum of two roots of Σ^+). Moreover, $\mathcal{P}$ is fibered over F where $F := G/G_c$ (see page 142 of [3]. The group F is finite by Lemma 6.21 of [3].

It follows from (13.4) of [3] that there exists $C, D > 0$ such that

$$d(g\cdot x, g\cdot y) \leq C\mathrm{e}^{D|\kappa(g)|_{\mathfrak{a}}} d(x,y) \qquad \forall g \in G,\, \forall x, y \in \mathcal{P}\,,$$

where $|\cdot|_{\mathfrak{a}}$ is an euclidean norm on $\mathfrak{a}$. In particular, for every $g \in G$,

$$\mathrm{Lip}(g) \leq C\mathrm{e}^{D|\kappa(g)|_{\mathfrak{a}}}\,.$$

Definition 3.6. Following [3, Section 9], we say that a Borel probability measure on G is Zariski dense if the subsemigroup Γ_μ spanned by the support of μ is G.

If μ is a Zariski dense Borel probability measure on G, then (13.10) of [3] holds, which means exactly that (3.3) is satisfied. In particular, for a Zariski dense Borel probability measure μ on G, the actions of G on $\mathcal{P}$ is weakly (μ, p)-contracting for some $p \geq 1$, as soon as

$$\int_G |\kappa(g)|^p_{\mathfrak{a}}\, \mu(dg) < \infty\,. \tag{3.24}$$

In order to apply our results, it remains to control the Iwasawa cocycle. It follows from (8.16) and (13.5) of [3], that there exist $C, D > 0$ such that

$$\begin{aligned} |\sigma(g,\eta)|_{\mathfrak{a}} &\leq C|\kappa(g)|_{\mathfrak{a}} \qquad \forall g \in G,\, \eta \in \mathcal{P} \\ \sigma_{\mathrm{Lip}}(g) &\leq C\mathrm{e}^{D|\kappa(g)|_{\mathfrak{a}}} \qquad \forall g \in G\,. \end{aligned}$$

In particular, σ admits a moment of order p, for some $p \geq 1$ as soon as (3.24) holds and our results apply.

Theorem 3.16. *Let G be a reductive algebraic Lie group. Let μ be a Zariski dense Borel probability measure on G that is F-strictly aperiodic. Let σ be the Iwasawa cocycle and κ be the Cartan projection as above. Assume that* (3.24) *holds for some $p \in (2,3]$, and let Σ be defined by* (3.22)*. Then,*

1. *For any $\varepsilon > 0$, enlarging Ω if necessary, there exists a sequence $(N_i)_{i\geq 1}$ of iid $\mathfrak{a}$-valued centered gaussian random variables with variance Σ such that*

$$\big|V_n - \sum_{i=1}^{n} N_i\big|_{\mathfrak{a}} = \begin{cases} o(n^{1/p}(\log n)^{\frac{p+1}{2p}+\varepsilon}) \text{ if } p \in]2,3[\\ o(n^{1/3}(\log n)^{1+\varepsilon}) \text{ if } p = 3 \end{cases} \text{ almost surely},$$

   where $(V_n)_{n\geq 1}$ is either of the processes $(\sigma(A_n, W_0) - n\lambda_\mu)_{n\geq 1})$, $(\kappa(A_n) - n\lambda_\mu)_{n\geq 1}$ or $(\sigma(A_n, x) - n\lambda_\mu)_{n\geq 1}$ for a given $x \in \mathcal{P}$.

2. *Enlarging Ω if necessary, there exists a sequence $(N_i)_{i\geq 1}$ of iid $\mathbb{R}^d$-valued centered gaussian random variables with variance Σ such that*

$$\Big\| \sup_{1\leq k\leq n} \big|V_k - \sum_{i=1}^{k} N_i\big|_{\mathfrak{a}} \Big\|_1 = \begin{cases} O(n^{1/p}(\log n)^{\frac{p-1}{2p}}) \text{ if } p \in]2,3[\\ O(n^{1/3}(\log n)^{2/3}) \text{ if } p = 3 \end{cases},$$

   where $(V_n)_{n\geq 1}$ is either of the processes $(\sigma(A_n, W_0) - n\lambda_\mu)_{n\geq 1})$ or $(\kappa(A_n) - n\lambda_\mu)_{n\geq 1}$. If moreover G is connected or $G = GL_d(\mathbb{R})$, then we also have

$$\sup_{x\in X} \Big\| \sup_{1\leq k\leq n} \big|\sigma(A_k, x) - k\lambda_\mu - \sum_{i=1}^{k} N_i\big|_{\mathfrak{a}} \Big\|_1 = \begin{cases} O(n^{1/p}(\log n)^{\frac{p-1}{2p}}) \text{ if } p \in]2,3[\\ O(n^{1/3}(\log n)^{2/3}) \text{ if } p = 3 \end{cases}.$$

Remark 3.8. The almost sure set in Item 1. above depends on x when $V_n = \sigma(A_n, x)$.

Theorem 3.17. *Let G be a reductive algebraic Lie group. Let μ be a Zariski dense Borel probability measure on G that is F-strictly aperiodic. Let σ be the Iwasawa cocycle and κ be the Cartan projection as above. Assume that* (3.24) *holds for some $p \in (2,3]$, and let Σ be defined by* (3.22)*. Let G_Σ be an $\mathfrak{a}$-valued centered gaussian random variables with variance Σ. Then*

$$W_1(P_{n^{-1/2}V_n}, P_{G_\Sigma}) \leq \begin{cases} Cn^{(2-p)/2} & \text{if } p \in]2,3[\\ Cn^{-1/2}\log n & \text{if } p = 3 \end{cases},$$

where C is a positive constant depending on (p, d) but not on n and $(V_n)_{n\geq 1}$ is either of the processes $(\sigma(A_n, W_0) - n\lambda_\mu)_{n\geq 1})$, $(\kappa(A_n) - n\lambda_\mu)_{n\geq 1}$. If moreover G is connected or $G = GL_d(\mathbb{R})$, then we also have

$$\sup_{x\in X} W_1(P_{n^{-1/2}(\sigma(A_n,x)-n\lambda_\mu)}, P_{G_\Sigma}) \leq \begin{cases} Cn^{(2-p)/2} & \text{if } p \in]2,3[\\ Cn^{-1/2}\log n & \text{if } p = 3 \end{cases},$$

4 Proofs

All along the proofs we denote by C a numerical constant which may vary from line to line and which may depend on d and p but not on n. We shall also denote sometimes by $\mathbb{E}_i$ the conditional expectation with respect to $\mathcal{F}_i$.

4.1 Proofs of the results of Section 2

4.1.1 Preliminaries

Let $\Sigma = \mathbb{E}(d_0 d_0^t)$. Suppose that Σ is nonnull (otherwise there is nothing to prove since the d_k's are all almost surely equal to zero). Since Σ is symmetric and positive-semidefinite, it follows that there exists a d-dimensional orthogonal matrix P such that

$$\Sigma = PDP^t\,,$$

with $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_d)$ where the λ_i's are the eigenvalues of Σ ranking in the non-increasing order. All these eigenvalues are reals and non-negative. Let $m \in \{1, \ldots, d\}$ be the number of eigenvalues that are positive and $\Delta = \mathrm{Diag}(\lambda_1, \ldots, \lambda_m, \lambda_m, \ldots, \lambda_m)$ (i.e. the d-dimensional diagonal matrix such that the first m diagonal elements are equal to the first m diagonal elements of D and the others to λ_m). Denote by J_m the d-dimensional diagonal matrix such that the first m diagonal elements are equal to 1 and the others to 0. Since $D = \Delta^{1/2}\mathrm{J}_m\Delta^{1/2}$,

$$\Sigma = P\Delta^{1/2}\mathrm{J}_m\Delta^{1/2}P^t\,.$$

Setting $\Gamma = \mathrm{Diag}(\lambda_1^{-1/2}, \ldots, \lambda_m^{-1/2}, \lambda_m^{-1/2}, \ldots, \lambda_m^{-1/2})$ (i.e. the inverse of $\Delta^{1/2}$) and $A = \Gamma P^t$, it follows that

$$A\Sigma A^t = \mathrm{J}_m\,,$$

since $PP^t = P^tP = \mathrm{I}_d$ where as usual I_d denotes the identity matrix on $\mathbb{R}^d$. Note that A is invertible and $A^{-1} = P\Delta^{1/2}$. For any integer k, let now

$$m_k = Ad_k\,.$$

Note that $(m_n)_{n\in\mathbb{N}}$ is a $\mathbb{R}^d$-valued stationary sequence of martingale differences with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $\mathbb{E}|m_0|_d^p < \infty$ and satisfying $\mathbb{E}(m_0 m_0^t) = \mathrm{J}_m$. Hence, clearly, $(m_0)_i = 0$ a.s. for any $i = m+1, \ldots, d$. For any integer k, let now d'_k be the $\mathbb{R}^m$-valued random vector whose components are equal to the m-first components of m_k, that is

$$d'_k = \big((m_k)_1, \ldots, (m_k)_m\big)^t\,.$$

Clearly, $(d'_n)_{n\in\mathbb{N}}$ is a $\mathbb{R}^m$-valued stationary sequence of martingale differences with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $\mathbb{E}|d'_0|_m^p < \infty$ and satisfying $\mathbb{E}(d'_0(d'_0)^t) = \mathrm{I}_m$. Let $M'_n = \sum_{i=1}^n d'_i$.

A common key result for the proofs of Theorems 2.1 and 2.2 is the following lemma: Let $\mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)$ be the set of measurable functions $g : \mathbb{R}^m \times \Omega \to \mathbb{R}$ with respect to the σ-fields $\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{F}_0$ and $\mathcal{B}(\mathbb{R})$, such that $g(\cdot, \omega) \in \mathrm{Lip}(|\cdot|_m)$ and $g(0, \omega) = 0$ for any $\omega \in \Omega$. For the sake of brevity, we shall write $g(x)$ in place of $g(x, \omega)$.

Lemma 4.1. *Under the assumptions of Theorem 2.1, setting $T_n = \sum_{i=1}^n N_i$ where $(N_i)_{i\geq 1}$ is a sequence of iid $\mathbb{R}^m$-valued centered gaussian random variables with $\mathrm{Var}(N_1) = \mathrm{I}_m$, we have*

$$\sup_{g\in\mathrm{Lip}(|\cdot|_m,\mathcal{F}_0)} \mathbb{E}(g(M'_n)) - \mathbb{E}(g(T_n)) \leq \begin{cases} Cn^{(3-p)/2} \text{ if } p \in]2,3[\\ C\log n \text{ if } p = 3 \end{cases},$$

where C is a positive constant depending on (p, m) but not on n.

4.1.2 Proof of Theorem 2.1

The construction of the approximating sequence of Gaussian random variables uses the ideas developed in the proof of Theorem 2.1 in Merlevède and Rio [28].

Let $(m(L))_{L\in\mathbb{N}}$ be a sequence of non-negative integers that will be specified later but such that $m(L) \leq L$. Let

$$I_{k,L} =]2^L + (k-1)2^{m(L)}, 2^L + k2^{m(L)}] \cap \mathbb{N} \text{ and } U_{k,L} = \sum_{i\in I_{k,L}} d'_i\,,\ k \in \{1, \cdots, 2^{L-m(L)}\}\,.$$

Let $P_{U_{k,L}|\mathcal{F}_{2^L+(k-1)2^{m(L)}}}$ be the conditional law of $U_{k,L}$ given $\mathcal{F}_{2^L+(k-1)2^{m(L)}}$ and $\mathcal{N}_{2^{m(L)}}$ denote the $\mathcal{N}(0, 2^{m(L)}\mathrm{I}_m)$-law. The probability space is assumed to be large enough to contain a sequence $(\delta_i)_{i\in\mathbb{Z}}$ of iid random variables uniformly distributed on $[0, 1]$, independent of the sequence $(d_i)_{i\in\mathbb{Z}}$ (otherwise we enlarge it). According to Rüschendorf [32] (see also Theorem 2 in [13]), there exists a $\mathbb{R}^m$-valued random variable $V_{k,L}$ with law $\mathcal{N}_{2^{m(L)}}$, measurable with respect to $\sigma(\delta_{2^L+k2^{m(L)}}) \vee \sigma(U_{k,L}) \vee \mathcal{F}_{2^L+(k-1)2^{m(L)}}$, independent of $\mathcal{F}_{2^L+(k-1)2^{m(L)}}$ and such that

$$\begin{aligned} \mathbb{E}\big(\big|U_{k,L} - V_{k,L}\big|_m\big) &= \mathbb{E}\big(W_1(P_{U_{k,L}|\mathcal{F}_{2^L+(k-1)2^{m(L)}}}, \mathcal{N}_{2^{m(L)}})\big) \\ &= \mathbb{E}\sup_{f\in\mathrm{Lip}(|\cdot|_m)} \Big(\mathbb{E}\big(f(U_{k,L})|\mathcal{F}_{2^L+(k-1)2^{m(L)}}\big) - \mathbb{E}(f(V_{k,L}))\Big)\,, \end{aligned} \tag{4.1}$$

where we recall that $\mathrm{Lip}(|\cdot|_m)$ is the set of functions from $\mathbb{R}^m$ into $\mathbb{R}$ that are 1-Lipschitz with respect to the euclidian norm $|\cdot|_m$ on $\mathbb{R}^m$. From Point 2 of Theorem 1 in [13], the following inequality holds:

$$\mathbb{E}\big(W_1(P_{U_{k,L}|\mathcal{F}_{2^L+(k-1)2^{m(L)}}}, \mathcal{N}_{2^{m(L)}})\big) = \sup_{g\in\mathrm{Lip}(|\cdot|_m,\mathcal{F}_{2^L})} \mathbb{E}(g(U_{1,L})) - \mathbb{E}(g(V_{1,L}))\,.$$

Hence, using Lemma 4.1, we have that for any $L \in \mathbb{N}$ and any $k \in \{1, \cdots, 2^{L-m(L)}\}$,

$$\mathbb{E}\big(\big|U_{k,L} - V_{k,L}\big|_m\big) \leq \begin{cases} C2^{(3-p)m(L)/2} & \text{if } p \in]2,3[\\ Cm(L) & \text{if } p = 3 \end{cases}, \tag{4.2}$$

where C is a positive constant depending on (p, m) but not on (k, L).

By induction on k, the random variables $(V_{k,L})_{k=1,\ldots,2^{L-m(L)}}$ are mutually independent, independent of $\mathcal{F}_{2^L}$ and with law $\mathcal{N}_{2^{m(L)}}$. Hence we have constructed Gaussian random variables $(V_{k,L})_{L\in\mathbb{N},k=1,\ldots,2^{L-m(L)}}$ that are mutually independent. Now we construct a sequence $(Z_i)_{i\geq 1}$ of iid standard Gaussian random vectors in $\mathbb{R}^m$. For any $L \in \mathbb{N}$ and any $k \in \{1, \cdots, 2^{L-m(L)}\}$ the random variables

$$(Z_{2^L+(k-1)2^{m(L)}+1}, \ldots, Z_{2^L+k2^{m(L)}})$$

are defined in the following way. If $m(L) = 0$, then $Z_{2^L+k2^{m(L)}} = V_{k,L}$. If $m(L) > 0$, then by the Skorohod lemma [33], there exists a measurable function g from $\mathbb{R}^m \times [0,1]$ in $(\mathbb{R}^m)^{\otimes 2^{m(L)}}$ such that, for any pair (V, δ) of independent random variables with respective laws $\mathcal{N}_{2^{m(L)}}$ and the uniform distribution over $[0,1]$, $g(V,\delta) = (N_1^t, \ldots, N_{2^{m(L)}}^t)$ is a Gaussian random vector with iid components such that $V = N_1 + \cdots + N_{2^{m(L)}}$ a.s. Next we set

$$(Z^t_{2^L+(k-1)2^{m(L)}+1}, \ldots, Z^t_{2^L+k2^{m(L)}}) = g(V_{k,L}, \delta_{2^L+(k-1)2^{m(L)}+1})\,.$$

We have then constructed a sequence $(Z_i)_{i\geq 2}$ of iid standard Gaussian random vectors in $\mathbb{R}^m$ such that, for any $L \in \mathbb{N}$ and any $k \in \{1, \cdots, 2^{L-m(L)}\}$,

$$V_{k,L} = \sum_{i\in I_{k,L}} Z_i \quad a.s.$$

To complete the construction of the sequence $(Z_i)_{i\geq 1}$, it suffices to consider a $\mathbb{R}^m$-valued standard Gaussian random vector Z_1 independent of $(d_i, \delta_i)_{i\in\mathbb{Z}}$ which is always possible by enlarging enough the underlying probability space.

Let us now complete the proof of Theorem 2.1. First, for any $k \geq 1$, we set

$$Y_k = A^{-1}Z'_k \text{ where } Z'_k = (Z_k^t, \mathbf{0}_{d-m})^t\,,$$

$\mathbf{0}_{d-m}$ denoting the row vector of dimension $d-m$ whose all components are equal to 0. Since $\mathrm{Var}(Z'_k) = \mathrm{J}_m$, we get that

$$\mathrm{Var}(Y_k) = A^{-1}\mathrm{J}_m(A^{-1})^t = \Sigma\,.$$

So, $(Y_i)_{i\geq 1}$ is a sequence of iid centered Gaussian random vectors in $\mathbb{R}^d$ with covariance matrix

Σ. Note that

$$\sup_{k\le n}\big|M_k-\sum_{i=1}^k Y_i\big|_d=\sup_{k\le n}\big|A^{-1}\big(\sum_{i=1}^k m_i-\sum_{i=1}^k Z_i'\big)\big|_d$$
$$\le\|A^{-1}\|_2\sup_{k\le n}\big|\sum_{i=1}^k m_i-\sum_{i=1}^k Z_i'\big|_d=\sqrt{\lambda_1}\sup_{k\le n}\big|\sum_{i=1}^k d_i'-\sum_{i=1}^k Z_i\big|_m\,,$$

where for the last inequality we have used the fact that

$$\|A^{-1}\|_2=\sqrt{\rho((A^{-1})^tA^{-1})}=\sqrt{\rho(\Delta^{1/2}P^tP\Delta^{1/2})}=\sqrt{\rho(\Delta)}=\sqrt{\lambda_1}\,.$$

Above and in the rest of the paper, for any $B\in\mathcal{M}_n(\mathbb{C})$, the notation $\rho(B)$ means the spectral radius of B. To prove the theorem, it suffices then to show that

$$\sup_{k\le n}\big|\sum_{i=1}^k d_i'-\sum_{i=1}^k Z_i\big|_m=o(n^{1/p}(\log n)^a),\ \text{almost surely}\,,\tag{4.3}$$

for a suitable a. With this aim, set $S_j=\sum_{i=1}^j d_i'$ and $T_j=\sum_{i=1}^j Z_i$ and let

$$D_L:=\sup_{\ell\le 2^L}|\sum_{i=2^L+1}^{2^L+\ell}(d_i'-Z_i)|_m\,.\tag{4.4}$$

Let $N\in\mathbb{N}^*$ and $k\in]1,2^{N+1}]$. We first notice that $D_L\ge|(S_{2^{L+1}}-T_{2^{L+1}})-(S_{2^L}-T_{2^L})|_m$, so that, if K is the integer such that $2^K<k\le 2^{K+1}$, $|S_k-T_k|_m\le|d_1'-Z_1|_m+D_0+D_1+\cdots+D_K$. Consequently since $K\le N$,

$$\sup_{1\le k\le 2^{N+1}}|S_k-T_k|_m\le|d_1'-Z_1|_m+D_0+D_1+\cdots+D_N\,.\tag{4.5}$$

In addition, the following decomposition is valid:

$$D_L\le D_{L,1}+D_{L,2}\,,\tag{4.6}$$

where

$$D_{L,1}:=\sup_{k\le 2^{L-m(L)}}\Big|\sum_{\ell=1}^k(U_{\ell,L}-V_{\ell,L})\Big|_m\ \text{and}\ D_{L,2}:=\sup_{k\le 2^{L-m(L)}}\sup_{\ell\in I_{k,L}}\Big|\sum_{i=\inf I_{k,L}}^{\ell}(d_i'-Z_i)\Big|_m\,.$$

End of the proof of Item 1. Let $\varepsilon>0$ and

$$a_p=\frac12+\frac{1}{2p}+\varepsilon\ \text{if}\ p\in]2,3[\ \text{and}\ a_3=1+\varepsilon.$$

From (4.4) and (4.5), it follows that the proof of Item 1 will be complete if we can show that, for any $L \in \mathbb{N}$,

$$D_{L,1} = O(2^{L/p} L^{a_p}) \quad \text{and} \quad D_{L,2} = O(2^{L/p} L^{a_p}) \quad \text{a.s.} \tag{4.7}$$

As we shall see below, this will be achieved by selecting the sequence $(m(L))_{L\geq 0}$ as follows in the construction of the iid gaussian vectors $(Y_i)_{i\geq 1}$ as described above: set $b_p = \frac{1}{p}$ if $p \in]2,3[$ and $b_3 = 1$ and

$$m(L) = \Big[\frac{2L}{p} + b_p \log_2 L\Big]\,, \text{ so that } \frac{1}{2} 2^{2L/p} L^{b_p} \leq 2^{m(L)} \leq 2^{2L/p} L^{b_p}\,, \tag{4.8}$$

square brackets designating as usual the integer part and $\log_2(x) = (\log x)/(\log 2)$.

To prove the first part of (4.7), note that, by (4.2) and the selection of $m(L)$, we get that, for any $c > 0$,

$$\begin{aligned}\mathbb{P}\big(D_{L,1} \geq c 2^{L/p} L^{a_p}\big) \leq \frac{c}{2^{L/p} L^{a_p}} \sum_{\ell=1}^{2^{L-m(L)}} \mathbb{E}\big(\big|U_{\ell,L} - V_{\ell,L}\big|_m\big) &\leq C \frac{2^L}{2^{L/p} L^{a_p}} 2^{(1-p)m(L)/2} (1 + \mathbf{1}_{p=3} m(L)) \\ &\leq \frac{C}{L^{a_p + (p-1)b_p/2}} (1 + \mathbf{1}_{p=3} L)\,.\end{aligned}$$

Hence, for any $c > 0$,

$$\mathbb{P}\big(D_{L,1} \geq c 2^{L/p} L^{a_p}\big) \leq \frac{C}{L^{1+\varepsilon}}\,,$$

which together with the Borel-Cantelli lemma implies the first part of (4.7).

We turn now to the proof of the second part of (4.7). With this aim, we set

$$x_L = \kappa 2^{L/p} L^{a_p} \text{ for some } \kappa > 0\,,$$

and first notice that, by stationarity, for any $y > 0$,

$$\mathbb{P}(D_{L,2} \geq 2y) \leq 2^{L-m(L)} \mathbb{P}\Big(\sup_{\ell \leq 2^{m(L)}} |S_\ell|_m \geq y\Big) + 2^{L-m(L)} \mathbb{P}\Big(\sup_{\ell \leq 2^{m(L)}} |T_\ell|_m \geq y\Big)\,. \tag{4.9}$$

By Lévy's inequality (see for instance Proposition 2.3 in [24]),

$$\mathbb{P}\Big(\sup_{\ell \leq 2^{m(L)}} |T_\ell|_m \geq x_L\Big) \leq \sum_{i=1}^{m} \mathbb{P}\Big(\sup_{\ell \leq 2^{m(L)}} |(T_\ell)_i| \geq m^{-1/2} x_L\Big) \leq 2m \exp\Big(-\frac{x_L^2}{m 2^{m(L)+1}}\Big)\,. \tag{4.10}$$

On the other hand, using Proposition 5.1 of the appendix, we get that, for any integer $i \in [1,d]$, there exist two positive constants c_1 and c_2 such that, for any $x > 0$,

$$\mathbb{P}\Big(\sup_{\ell \leq n} |(M_\ell)_i| \geq x\Big) \leq c_1 \exp\Big(-\frac{x^2}{c_2 n}\Big) + c_2 n x^{-p}\,. \tag{4.11}$$

Now note that

$$\sup_{\ell\le 2^{m(L)}} |S_\ell|_m = \sup_{\ell\le 2^{m(L)}} |AM_\ell|_d \le \|A\|_2 \sup_{\ell\le 2^{m(L)}} |M_\ell|_d\,.$$

But $\|A\|_2 = \sqrt{\rho(P\Gamma^2 P^t)} = \lambda_m^{-1/2}$. Therefore, applying inequality (4.11), we get

$$\begin{aligned}\mathbb{P}\Big(\sup_{\ell\le 2^{m(L)}} |S_\ell|_m \ge x_L\Big) &\le \sum_{i=1}^m \mathbb{P}\Big(\sup_{\ell\le 2^{m(L)}} |(M_\ell)_i| \ge (m^{-1/2}\lambda_m^{1/2} x_L\Big)\Big)\\ &\le c_1 m \exp\Big(-\frac{x_L^2\lambda_m}{c_2 m 2^{m(L)}}\Big) + c_2 m^{(2+p)/2}\lambda_m^{-p/2} 2^{m(L)} x_L^{-p}\,. \quad (4.12)\end{aligned}$$

Starting from (4.9) and considering the upper bounds (4.10) and (4.12), it follows that

$$\mathbb{P}(D_{L,2}\ge 2x_L) \le m(c_1+2)2^{L-m(L)}\exp\Big(-\frac{x_L^2\lambda_m}{c(m)2^{m(L)}}\Big) + c_2 m^{(2+p)/2}\lambda_m^{-p/2}2^L x_L^{-p}\,, \quad (4.13)$$

where

$$c(m) = m\max(2, c_2\lambda_m^{-1})\,.$$

For any choice of κ, by the selection of x_L and since $pa_p > 1$, it follows that $\sum_{L\ge 1} 2^L x_L^{-p} < \infty$. On another hand,

$$\frac{x_L^2}{2^{m(L)}} \ge \kappa^2 L^{2a_p - b_p} = \kappa^2 L^{1+2\varepsilon}\,.$$

It follows that for any $\kappa > 0$,

$$\sum_{L\ge 1} 2^L \exp\Big(-\frac{x_L^2\lambda_m}{c(m)2^{m(L)}}\Big) < \infty\,.$$

So, overall, starting from (4.13) and using the Borel-Cantelli lemma, we can conclude that the second part of (4.7) holds. This ends the proof of Item 1.

End of the proof of Item 2. Let $\varepsilon > 0$ and

$$a_p = \frac{1}{2} - \frac{1}{2p} \text{ if } p \in]2,3[\text{ and } a_3 = 2/3.$$

Starting again from, (4.4) and (4.5), it follows that the proof of Item 2 will be complete if we can show that, for any $L\in\mathbb{N}$,

$$\|D_{L,1}\|_1 = O(2^{L/p}L^{a_p}) \quad \text{and} \quad \|D_{L,2}\|_1 = O(2^{L/p}L^{a_p})\,. \quad (4.14)$$

As we shall see below, this will be achieved by selecting the sequence $(m(L))_{L\ge 0}$ as follows in the construction of the iid gaussian vectors $(Y_i)_{i\ge 1}$ as described previously: set $b_p = \frac{1}{p}$ if $p\in]2,3[$ and $b_3 = -1/3$ and

$$m(L) = \Big[\frac{2L}{p} - b_p\log_2 L\Big]\,, \text{ so that } \frac{1}{2}2^{2L/p}L^{-b_p} \le 2^{m(L)} \le 2^{2L/p}L^{-b_p}\,. \quad (4.15)$$

The first part of (4.14) follows by using (4.2) together with the above selection of $m(L)$. To show the second part of (4.14), we set $y_L = \kappa_m^{-1/2}(\log 2)^{1/2} 2^{m(L)/2} L^{1/2}$ where $\kappa_m = \lambda_m / c(m)$. Hence, using (4.13), write that

$$\begin{aligned}
\|D_{L,2}\|_1 &\le y_L + \int_{y_L}^{\infty} \mathbb{P}(D_{L,2} \ge t) dt \\
&\le C\Big\{ y_L + 2^L y_L^{1-p} + 2^{L-m(L)} \int_{y_L}^{\infty} \exp\Big(-\frac{\kappa_m t^2}{2^{m(L)}}\Big) dt \Big\} \\
&\le C\Big\{ y_L + 2^L y_L^{1-p} + 2^L \exp\Big(-\frac{\kappa_m y_L^2}{2^{m(L)}}\Big) \Big\} \le C\{ y_L + 2^L y_L^{1-p} \}\,.
\end{aligned}$$

Taking into account the selection of y_L and (4.15), Item 2 follows.

4.1.3 Proof of Theorem 2.2

Recall the notations $A^{-1} = P\Delta^{1/2}$ and $m_k = A d_k$. Hence, we have

$$W_1(P_{M_n|\mathcal{F}_0}, P_{G_{n\Sigma}}) \le \|A^{-1}\|_2 W_1(P_{\sum_{k=1}^n m_k|\mathcal{F}_0}, P_{G_{n\mathrm{I}_m}}) = \sqrt{\lambda_1} W_1(P_{\sum_{k=1}^n m_k|\mathcal{F}_0}, P_{G_{nJ_m}}) \text{ a.s.}$$

Moreover, since $\mathbb{E}(m_0 m_0^t) = \mathrm{J}_m$, we have $(m_0)_i = 0$ a.s. for any $i = m+1, \ldots, d$. Hence, setting

$$d'_k = \big((m_k)_1, \ldots, (m_k)_m\big)^t \text{ and } M'_n = \sum_{k=1}^n d'_k\,,$$

and noticing that $\mathbb{E}(d'_0 (d'_0)^t) = \mathrm{I}_m$, we have

$$W_1(P_{\sum_{k=1}^n m_k|\mathcal{F}_0}, P_{G_{nJ_m}}) = W_1(P_{M'_n|\mathcal{F}_0}, P_{G_{nI_m}}) \text{ a.s.}$$

From Point 2 of Theorem 1 in [13], the following inequality then holds:

$$\|W_1(P_{M_n|\mathcal{F}_0}, P_{G_{n\Sigma}})\|_1 \le \sqrt{\lambda_1} \sup_{g \in \mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)} \mathbb{E}(g(M'_n)) - \mathbb{E}(g(T_n))\,,$$

where $T_n = \sum_{i=1}^n N_i$ with $(N_i)_{i\ge1}$ a sequence of iid $\mathbb{R}^m$-valued centered gaussian random variables with $\mathrm{Var}(N_1) = \mathrm{I}_m$. To end the proof of the theorem, it suffices to use Lemma 4.1.

4.1.4 Proof of Lemma 4.1

Note first that we can assume the sequence $(N_i)_{i\ge1}$ independent of $(d_i)_{i\in\mathbb{Z}}$. Let us now consider a m-dimensional standard Gaussian random vector G independent of $(N_i)_{i\ge1}$ and $(d_i)_{i\in\mathbb{Z}}$. Note that

$$\begin{aligned}
\sup_{g \in \mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)} \mathbb{E}(g(M'_n)) - \mathbb{E}(g(T_n)) &\le \sup_{g \in \mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)} \mathbb{E}(g(M'_n + G)) - \mathbb{E}(g(T_n + G)) + 2\mathbb{E}|G|_m \\
&\le \sup_{g \in \mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)} \mathbb{E}(g(M'_n + G)) - \mathbb{E}(g(T_n + G)) + 2\sqrt{m}\,.
\end{aligned}$$

The lemma is then reduced to prove that

$$\sup_{g\in \mathrm{Lip}(|\cdot|_m,\mathcal{F}_0)} \mathbb{E}(g(M'_n+G)) - \mathbb{E}(g(T_n+G)) \le \begin{cases} Cn^{(3-p)/2} & \text{if } p\in]2,3[\\ C\log n & \text{if } p=3 \end{cases} . \tag{4.16}$$

We now use the Lindeberg method to prove (4.16). With this aim, we introduce the following notation:

Notation 4.1. Let φ_a be the density of a m-dimensional centered Gaussian random vector with covariance matrix $a^2\mathrm{I}_m$ and let for $x\in\mathbb{R}^m$,

$$g*\varphi_a(x,\omega) = \int g(x+y,\omega)\varphi_a(y)dy\,.$$

For the sake of brevity, we shall write $g*\varphi_a(x)$ instead of $g*\varphi_a(x,\omega)$ (the partial derivatives will be taken with respect to x) and set $g_{i,n}(x) = g*\varphi_{i,n}$ where $\varphi_{i,n} = \varphi_{\sqrt{n-i+1}}$.

Let

$$\Delta_{i,n}(g) = g\big(M'_{i-1} + d'_i + \sum_{j=i+1}^n N_j + G\big) - g\big(M'_{i-1} + N_i + \sum_{j=i+1}^n N_j + G\big)\,.$$

We have

$$\mathbb{E}(\Delta_{i,n}(g)) = \mathbb{E}\big(g_{i,n}\big(M'_{i-1}+d'_i\big)\big) - \mathbb{E}\big(g_{i,n}\big(M'_{i-1}+N_i\big)\big)$$

Hence, noticing that $M_0=0$, the following decomposition is valid:

$$\mathbb{E}(g(M'_n+G)) - \mathbb{E}(g(T_n+G)) = \sum_{i=1}^n \mathbb{E}\big(\Delta_{i,n}(g)\big)\,. \tag{4.17}$$

Below we shall also use the following notations.

Notation 4.2. For two positive integers m and n, let $\mathcal{M}_{m,n}(\mathbb{R})$ be the set of real matrices with m lines and n columns. The Kronecker product (or Tensor product) of $A=[a_{i,j}]\in\mathcal{M}_{m,n}(\mathbb{R})$ and $B=[b_{i,j}]\in\mathcal{M}_{p,q}(\mathbb{R})$ is denoted by $A\otimes B$ and is defined to be the block matrix

$$A\otimes B = \begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{pmatrix} \in \mathcal{M}_{mp,nq}(\mathbb{R})\,.$$

For any positive integer k, the k-th Kronecker power $A^{\otimes k}$ is defined inductively by: $A^{\otimes 1}=A$ and $A^{\otimes k} = A\otimes A^{\otimes(k-1)}$.

If ∇ denotes the differentiation operator given by $\nabla = \big(\frac{\partial}{\partial x_1},\dots,\frac{\partial}{\partial x_m}\big)^t$ acting on the differentiable functions $f:\mathbb{R}^m\to\mathbb{R}$, we define

$$\nabla\otimes\nabla = \Big(\frac{\partial}{\partial x_1}\circ\nabla,\dots,\frac{\partial}{\partial x_m}\circ\nabla\Big)^t,$$

and $\nabla^{\otimes k}$ by $\nabla^{\otimes 1} = \nabla$ and $\nabla^{\otimes k} = \nabla \otimes \nabla^{\otimes(k-1)}$. If $f : \mathbb{R}^m \to \mathbb{R}$ is k-times differentiable, for any $x \in \mathbb{R}^m$, let $D^k f(x) = \nabla^{\otimes k} f(x)$, and for any vector A of $\mathbb{R}^m$, we define $D^k f(x).A^{\otimes k}$ as the usual scalar product in $\mathbb{R}^{m^k}$ between $D^k f(x)$ and $A^{\otimes k}$.

For any $i \in \{1, \dots, n\}$, let

$$\Delta_{1,i,n}(g) = g_{i,n}\big(M'_{i-1} + d'_i\big) - g_{i,n}\big(M'_{i-1}\big) - \frac{1}{2} D^2 g_{i,n}\big(M'_{i-1}\big).N_i^{\otimes 2}\,,$$

and

$$\Delta_{2,i,n}(g) = g_{i,n}\big(M'_{i-1} + N_i\big) - g_{i,n}\big(M'_{i-1}\big) - \frac{1}{2} D^2 g_{i,n}\big(M'_{i-1}\big).N_i^{\otimes 2}\,.$$

With this notation,

$$\mathbb{E}\big(\Delta_{i,n}(g)\big) = \mathbb{E}\big(\Delta_{1,i,n}(g)\big) - \mathbb{E}\big(\Delta_{2,i,n}(g)\big)\,. \tag{4.18}$$

By the Taylor integral formula, noticing that $\mathbb{E}(N_i^{\otimes 3}) = 0$, we get

$$\big|\mathbb{E}(\Delta_{2,i,n}(g))\big| \le \frac{1}{6}\Big|\mathbb{E}\int_0^1 D^4 g_{i,n}\big(M'_{i-1} + tN_i\big).N_i^{\otimes 4} dt\Big|\,.$$

But, according to Lemma 5.6 in [12], for any $y \in \mathbb{R}^m$ and any integer $k \ge 1$, there exists a positive constant c_k depending only on k such that

$$\sup_{(i_1,\dots,i_k)\in\{1,\dots,m\}^m} \Big|\frac{\partial^k g * \varphi_{i,n}}{\prod_{j=1}^k \partial x_{i_j}}(y)\Big| \le c_k (n-i+1)^{(1-k)/2}\,. \tag{4.19}$$

Therefore

$$\big|\mathbb{E}(\Delta_{2,i,n}(g))\big| \le 2^{-1} m^4 c_4 (n-i+1)^{-3/2}\,.$$

Therefore

$$\sum_{i=1}^n \big|\mathbb{E}(\Delta_{2,i,n}(g))\big| \le 2^{-1} m^4 c_4 \sum_{i=1}^n (n-i+1)^{-3/2} \le \kappa_1 m^4\,. \tag{4.20}$$

Let now

$$R_{1,i,n}(g) = g_{i,n}\big(M'_{i-1} + d'_i\big) - g_{i,n}\big(M'_{i-1}\big) - Dg_{i,n}\big(M'_{i-1}\big).d'_i - \frac{1}{2} D^2 g_{i,n}\big(M'_{i-1}\big).d_i'^{\otimes 2}\,,$$

and

$$R_{2,i,n}(g) = Dg_{i,n}\big(M'_{i-1}\big).d'_i + \frac{1}{2} D^2 g_{i,n}\big(M'_{i-1}\big).d_i'^{\otimes 2} - \frac{1}{2} D^2 g_{i,n}\big(M'_{i-1}\big).\mathbb{E}(N_i^{\otimes 2})\,.$$

With this notation,

$$\mathbb{E}(\Delta_{1,i,n}(g)) = \mathbb{E}(R_{1,i,n}(g)) + \mathbb{E}(R_{2,i,n}(g))\,. \tag{4.21}$$

By the Taylor integral formula at order two,

$$\big|\mathbb{E}(R_{1,i,n}(g))\big| \le \Big|\mathbb{E}\int_0^1 \frac{(1-t)}{2}\Big(D^2 g_{i,n}\big(M'_{i-1} + td'_i\big) - D^2 g_{i,n}\big(M'_{i-1}\big)\Big).d_i'^{\otimes 2} dt\Big|\,.$$

But, by taking into account (4.19), we infer that, for any $t \in [0,1]$,

$$\begin{aligned}\big|D^2 g_{i,n}\big(M'_{i-1}+td'_i\big) - D^2 g_{i,n}\big(M'_{i-1}\big)\Big).d_i'^{\otimes 2}\big| \\ \leq m(c_2 \vee c_3)(n-i+1)^{-1/2}\Big(\min\big(2|d'_1|_m^2, tm^{1/2}(n-i+1)^{-1/2}|d'_1|_m^3\big)\Big) \\ \leq 2m^{p/2}(c_2 \vee c_3)(n-i+1)^{-1/2}(n-i+1)^{-(p-2)/2}|d'_1|_m^p\,,\end{aligned}$$

since $p \in]2,3]$. Therefore, since $|d'_1|_m = |Ad_1|_d \leq \|A\|_2 |d_1|_d = \sqrt{\rho(P\Gamma^2 P^t)}|d_1|_d = \lambda_m^{-1/2}|d_1|_d$, it follows that

$$\big|\mathbb{E}(R_{1,i,n}(g))\big| \leq 2^{-1}\lambda_m^{-p/2} m^{p/2}(c_2 \vee c_3)(n-i+1)^{-(p-1)/2}\mathbb{E}(|d_1|_d^p)\,.$$

Hence there exists a positive constant κ_2 such that, for any positive integer n,

$$\sum_{i=1}^n \big|\mathbb{E}(R_{1,i,n}(g))\big| \leq \kappa_2 \lambda_m^{-p/2} m^{p/2}\mathbb{E}(|d_1|_d^p) n^{(3-p)/2}(1+\mathbf{1}_{p=3}\log n)\,. \tag{4.22}$$

Starting from the decomposition (4.18) and taking into account (4.20), (4.21) and (4.22), it follows that for any $g \in \mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)$,

$$\Big|\sum_{i=1}^n \mathbb{E}\big(\Delta_{i,n}(g)\big)\Big| \leq \Big|\sum_{i=1}^n (\mathbb{E}(R_{2,i,n}(g))\Big| + Cn^{(3-p)/2}(1+\mathbf{1}_{p=3}\log n)\,. \tag{4.23}$$

We handle now the term $\sum_{i=1}^n \mathbb{E}(R_{2,i,n}(g))$. With this aim, let us first write n in basis 2. Let r be the unique non-negative integer such that $2^r \leq n < 2^{r+1}$. Then, writing n in basis 2, we have

$$n = \sum_{k=0}^r b_k(n) 2^k \quad \text{where } b_r(n) = 1 \text{ and } b_k(n) \in \{0,1\} \text{ for } k = 0, \ldots, r-1\,.$$

Let also

$$n_k = \sum_{j=0}^k b_j(n) 2^j \quad \text{for } k = 0, \ldots, r-1 \text{ and } n_{-1} = 0\,.$$

It follows that

$$\sum_{i=1}^n \mathbb{E}(R_{2,i,n}(g)) = \sum_{k=0}^r b_k(n) \sum_{i=n_{k-1}+1}^{n_k} \mathbb{E}(R_{2,i,n}(g)) = \sum_{k=0}^r b_k(n) \sum_{i=1}^{2^k} \mathbb{E}(A_{i,n}(g))\,, \tag{4.24}$$

where $A_{i,n}(g) = R_{2,i+n_{k-1},n}(g)$ (note that the last inequality above holds because, when $b_k(n) = 1$, then necessarily $n_k = 2^k + n_{k-1}$). Let $i \in \{1, \ldots, 2^k\}$. Notice now that, since $(d'_i)_{i\in\mathbb{Z}}$ is a martingale differences sequence with respect to $(\mathcal{F}_i)_{i\in\mathbb{Z}}$ and such that $\mathbb{E}(d_i'^{\otimes 2}) = \mathbb{E}(N_i^{\otimes 2})$, we have, by setting $i(k) = i + n_{k-1}$,

$$\mathbb{E}(A_{i,n}(g)) = \frac{1}{2}\mathbb{E}\big(D^2 g_{i(k),n}\big(M'_{i(k)-1}\big).\big(d_{i(k)}'^{\otimes 2} - \mathbb{E}(d_{i(k)}'^{\otimes 2})\big)\,.$$

To continue the computations, as in the proof of Proposition 5.1 in [11], we introduce again a dyadic scheme. With this aim, we introduce the following notations.

Notation 4.3. Set $i_0 = i-1$ and write i_0 in basis 2 (recall that $i \in \{1,\dots,2^k\}$): $i_0 = \sum_{i=0}^k a_i 2^i$ with $a_i = 0$ or $a_i = 1$ (note that $a_k = 0$). Set $i_j = \sum_{i=j}^k a_i 2^i$, so that $i_k = 0$, and set $i_j(k) = i_j + n_{k-1}$. Let $I_{j,\ell} =]\ell 2^j, (\ell+1)2^j] \cap \mathbb{N}$ (note that $I_{k,1} =]2^k, 2^{k+1}]$), $U_j^{(\ell)} = \sum_{i\in I_{j,\ell}} d'_i$, $U_{j,k}^{(\ell)} = \sum_{i\in I_{j,\ell}} d'_{i+k}$, $\tilde{U}_j^{(\ell)} = \sum_{i\in I_{j,\ell}} N_i$ and $\tilde{U}_{j,k}^{(\ell)} = \sum_{i\in I_{j,\ell}} N_{i+k}$. For the sake of brevity, let $U_j^{(0)} = U_j$ and $\tilde{U}_j^{(0)} = \tilde{U}_j$. Set also $Z_j^{(\ell)} = \mathbb{E}_{\ell 2^j}((U_j^{(\ell)})^{\otimes 2}) - \mathbb{E}_{\ell 2^j}((U_j^{(\ell)})^{\otimes 2})$.

Since $i_k = 0$, the following elementary identity is valid

$$\begin{aligned}
&\mathbb{E}(A_{i,n}(g)) \\
&\quad = \frac{1}{2}\sum_{j=0}^{k-1} \mathbb{E}\Big(\big(D^2 g_{i_j(k)+1,n}\big(M'_{i_j(k)}\big) - D^2 g_{i_{j+1}(k)+1,n}\big(M'_{i_{j+1}(k)}\big)\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big) \\
&\qquad\qquad + \frac{1}{2}\mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big) \\
&\qquad\qquad =: \mathbb{E}(B_{i,n}(g)) + \frac{1}{2}\mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big). \quad (4.25)
\end{aligned}$$

Since $D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big)$ is a random vector which is $\mathcal{F}_{n_{k-1}}$ measurable, we have

$$\mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big) = \mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\mathbb{E}_{n_{k-1}}\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big).$$

But, by the martingale property,

$$\sum_{i=1}^{2^k} \mathbb{E}_{n_{k-1}}\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big) = \mathbb{E}_{n_{k-1}}\Big(\Big(\sum_{i=1}^{2^k} d'_{i(k)}\Big)^{\otimes 2} - \mathbb{E}\Big(\sum_{i=1}^{2^k} d'_{i(k)}\Big)^{\otimes 2}\Big).$$

Taking into account (4.19) and stationarity, we get

$$\begin{aligned}
&\Big|\sum_{i=1}^{2^k} \mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big)\Big| \\
&\qquad \le c_1 (n-n_{k-1})^{-1/2} \sum_{a=1}^m \sum_{b=1}^m \mathbb{E}\Big(\Big|\mathbb{E}_0\Big(\Big(\sum_{i=1}^{2^k} d'_i\Big)_a \Big(\sum_{i=1}^{2^k} d'_i\Big)_b\Big) - \mathbb{E}\Big(\Big(\sum_{i=1}^{2^k} d'_i\Big)_a \Big(\sum_{i=1}^{2^k} d'_i\Big)_b\Big)\Big|\Big).
\end{aligned}$$

Since $(d'_i)_a = (Ad_i)_a$ for $1 \le a \le m$ and $(Ad_i)_a = 0$ for $m+1 \le a \le d$,

$$\begin{aligned}
&\sum_{a=1}^m \sum_{b=1}^m \mathbb{E}\Big(\Big|\mathbb{E}_0\Big(\Big(\sum_{i=1}^{2^k} d'_i\Big)_a \Big(\sum_{i=1}^{2^k} d'_i\Big)_b\Big) - \mathbb{E}\Big(\Big(\sum_{i=1}^{2^k} d'_i\Big)_a \Big(\sum_{i=1}^{n} d'_i\Big)_b\Big)\Big|\Big) \\
&\qquad\qquad \le d\,\mathbb{E}\Big|A^{\otimes 2}\big(\mathbb{E}_0\big(M_{2^k}^{\otimes 2}\big) - \mathbb{E}\big(M_{2^k}^{\otimes 2}\big)\big)\Big|_d \le d\|A^{\otimes 2}\|_2 \mathbb{E}\Big|\mathbb{E}_0\big(M_{2^k}^{\otimes 2}\big) - \mathbb{E}\big(M_{2^k}^{\otimes 2}\big)\Big|_d.
\end{aligned}$$

But $\|A^{\otimes 2}\|_2 \le \|A\|_2^2 \le \lambda_m^{-1}$. So, from the above considerations,

$$\Big|\sum_{i=1}^{2^k} \mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big)\Big| \\ \le c_1 d\lambda_m^{-1}(n-n_{k-1})^{-1/2}\mathbb{E}\Big|\mathbb{E}_0\big(M_{2^k}^{\otimes 2}\big) - \mathbb{E}\big(M_{2^k}^{\otimes 2}\big)\Big|_d .$$

Since $n \ge n_{k-1} + 2^k$ if $b_k(n) = 1$, we get

$$\sum_{k=0}^{r} b_k(n)\Big|\sum_{i=1}^{2^k} \mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big)\Big| \\ \le C\sum_{k=0}^{r} b_k(n) 2^{-k/2}\mathbb{E}\Big|\mathbb{E}_0\big(M_{2^k}^{\otimes 2}\big) - \mathbb{E}\big(M_{2^k}^{\otimes 2}\big)\Big|_d \\ \le C 2^{(3-p)r/2}\sum_{k=0}^{r} 2^{-\left(2-\frac{p}{2}\right)k}\mathbb{E}\Big|\mathbb{E}_0\big(M_{2^k}^{\otimes 2}\big) - \mathbb{E}\big(M_{2^k}^{\otimes 2}\big)\Big|_d .$$

But, by the subadditivity of the sequence $\big(\mathbb{E}\big|\mathbb{E}_0\big(M_k^{\otimes 2}\big) - \mathbb{E}\big(M_k^{\otimes 2}\big)\big|_d\big)_{k\ge 1}$, condition (2.1) implies that

$$\sum_{k\ge 0} 2^{-\left(2-\frac{p}{2}\right)k}\mathbb{E}\Big|\mathbb{E}_0\big(M_{2^k}^{\otimes 2}\big) - \mathbb{E}\big(M_{2^k}^{\otimes 2}\big)\Big|_d < \infty .$$

(See for instance Remark 2.6 in [11]). Hence,

$$\sum_{k=0}^{r} b_k(n)\Big|\sum_{i=1}^{2^k} \mathbb{E}\Big(D^2 g_{n_{k-1}+1,n}\big(M'_{n_{k-1}}\big).\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big)\Big| \le C n^{(3-p)/2} . \tag{4.26}$$

Starting from (4.23) and taking into account (4.24), (4.25) and (4.26), the proof of the lemma will be complete if we can prove that for any $g \in \mathrm{Lip}(|\cdot|_m, \mathcal{F}_0)$,

$$\Big|\sum_{k=0}^{r} b_k(n)\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big| \le C n^{(3-p)/2}(1 + \mathbf{1}_{p=3}\log n) . \tag{4.27}$$

Let $i \in \{1,\dots,2^k\}$ and note that $i_j \ne i_{j+1}$ only if $a_j = 1$, then in this case $i_j = \ell 2^j$ with ℓ odd. It follows that

$$2\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g)) = \sum_{j=0}^{k-1}\sum_{\substack{\ell\in I_{k-j,0}\\ \ell \text{ odd}}} \mathbb{E}\Big(\big(D^2 g_{\ell 2^j + n_{k-1}+1,n}\big(M'_{\ell 2^j+n_{k-1}}\big) \\ - D^2 g_{(\ell-1)2^j+n_{k-1}+1,n}\big(M'_{(\ell-1)2^j+n_{k-1}}\big)\big).\sum_{\{i:i_j=\ell 2^j\}}\big(d'^{\otimes 2}_{i(k)} - \mathbb{E}(d'^{\otimes 2}_{i(k)})\big)\Big) .$$

Note that $\{i : i_j = \ell 2^j\} = I_{j,\ell}$. Now by the martingale property,

$$\mathbb{E}_{\ell 2^j+n_{k-1}}\Big(\sum_{i\in I_{j,\ell}}\big(d_{i(k)}'^{\otimes 2}-\mathbb{E}(d_{i(k)}'^{\otimes 2})\big)\Big)=\mathbb{E}_{\ell 2^j+n_{k-1}}((U_{j,n_{k-1}}^{(\ell)})^{\otimes 2})-\mathbb{E}((U_{j,n_{k-1}}^{(\ell)})^{\otimes 2}):=Z_{j,n_{k-1}}^{(\ell)}\,.$$

Consequently

$$\begin{aligned}2\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))&=\sum_{j=0}^{k-1}\sum_{\substack{\ell\in I_{k-j,0}\\ \ell\text{ odd}}}\mathbb{E}\Big(\big(D^2g_{\ell 2^j+n_{k-1}+1,n}\big(M'_{\ell 2^j+n_{k-1}}\big)\\&\qquad-D^2g_{(\ell-1)2^j+n_{k-1}+1,n}\big(M'_{(\ell-1)2^j+n_{k-1}}\big)\big).Z_{j,n_{k-1}}^{(\ell)}\Big)\\&=\sum_{j=0}^{k-1}\sum_{\substack{\ell\in I_{k-j,0}\\ \ell\text{ odd}}}\mathbb{E}\Big(\Big(D^2g_{\ell 2^j+n_{k-1}+1,n}\big(M'_{\ell 2^j+n_{k-1}}\big)\\&\qquad-D^2g_{\ell 2^j+n_{k-1}+1,n}\Big(M'_{(\ell-1)2^j+n_{k-1}}+\sum_{v=(\ell-1)2^j+1}^{\ell 2^j}N_{v+n_{k-1}}\Big)\Big).Z_{j,n_{k-1}}^{(\ell)}\Big)\,,\end{aligned}$$

since $(d_i)_{i\in\mathbb{N}}$ and $(N_i)_{i\in\mathbb{N}}$ are independent. Note that $Z_{j,n_{k-1}}^{(\ell)}$ is a m^2-dimensional random vector. Let then introduce the following notation: $(Z_{j,n_{k-1}}^{(\ell)})_{a,b}$ is the $(a-1)m+b$-th coordinate of the vector $Z_{j,n_{k-1}}^{(\ell)}$. By using this notation and (4.19), it follows that

$$\begin{aligned}2\Big|\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big|&\le\max(2c_2,c_3)\times\sum_{j=0}^{k-1}\sum_{\substack{\ell\in I_{k-j,0}\\ \ell\text{ odd}}}(n-\ell 2^j-n_{k-1})^{-1/2}\\&\qquad\times\sum_{a,b=1}^{m}\mathbb{E}\Big(|(Z_{j,n_{k-1}}^{(\ell)})_{a,b}|\min\big(1,(n-\ell 2^j-n_{k-1})^{-1/2}\sum_{c=1}^{m}\big|(U_{j,n_{k-1}}^{(\ell-1)}-\tilde U_{j,n_{k-1}}^{(\ell-1)})_c\big|\big)\Big)\,.\end{aligned}$$

But, since $p\in]2,3]$,

$$\begin{aligned}&\mathbb{E}\Big(|(Z_{j,n_{k-1}}^{(\ell)})_{a,b}|\min\big(1,(n-\ell 2^j-n_{k-1})^{-1/2}\sum_{c=1}^{m}\big|(U_{j,n_{k-1}}^{(\ell-1)}-\tilde U_{j,n_{k-1}}^{(\ell-1)})_c\big|\big)\Big)\\&\qquad\le(n-\ell 2^j-n_{k-1})^{-(p-2)/2}\mathbb{E}\Big(|(Z_{j,n_{k-1}}^{(\ell)})_{a,b}|\Big(\sum_{c=1}^{m}\big|(U_{j,n_{k-1}}^{(\ell-1)}-\tilde U_{j,n_{k-1}}^{(\ell-1)})_c\big|\Big)^{p-2}\Big)\\&\qquad\le(n-\ell 2^j-n_{k-1})^{-(p-2)/2}\sum_{c=1}^{m}\mathbb{E}\Big(|(Z_{j,n_{k-1}}^{(\ell)})_{a,b}|\big|(U_{j,n_{k-1}}^{(\ell-1)}-\tilde U_{j,n_{k-1}}^{(\ell-1)})_c\big|^{p-2}\Big)\,,\end{aligned}$$

which together with stationarity implies that

$$2b_k(n)\Big|\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big| \leq$$
$$\leq \max(2c_2,c_3)\sum_{j=0}^{k-1}\sum_{\ell=1}^{2^{k-j}-1} b_k(n)(n-\ell 2^j-n_{k-1})^{-(p-1)/2}\sum_{a,b,c=1}^{m}\mathbb{E}\Big(|(Z_j^{(1)})_{a,b}|\big|(U_j-\tilde{U}_j)_c\big|^{p-2}\Big)\,.$$

Let

$$Q_j=\sum_{a,b,c=1}^{m}\mathbb{E}\Big(|(Z_j^{(1)})_{a,b}|\big|(U_j-\tilde{U}_j)_c\big|^{p-2}\Big)\,.$$

Since $n\geq n_{k-1}+2^k$ if $b_k(n)=1$, we get, in case where $p\in]2,3[$,

$$b_k(n)\Big|\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big|\leq C\sum_{j=0}^{k-1}2^{-j(p-1)/2}2^{(3-p)(k-j)/2}Q_j\,.$$

Consider now the case $p=3$. Note that, since $b_k(n)=1$ implies that $n_k=n_{k-1}+2^k$, we have

$$b_k(n)\sum_{\ell=1}^{2^{k-j}-1}(n-\ell 2^j-n_{k-1})^{-1}=2^{-j}b_k(n)\sum_{\ell=2^{-j}(n-n_k)+1}^{2^{-j}(n-n_{k-1})-1}\ell^{-1}$$
$$\leq 2^{-j+1}b_k(n)\sum_{\ell=2^{-j}(n-n_k)+1}^{2^{-j}(n-n_{k-1})-1}(\ell+1)^{-1}\leq 2^{-j+1}b_k(n)\Big\{\log\big(\frac{n-n_{k-1}}{2^j}\big)-\log\big(\frac{n-n_k}{2^j}+1\big)\Big\}\,.$$

So, overall,

$$b_k(n)\Big|\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big|\leq C\left\{\begin{array}{ll}\sum_{j=0}^{k-1}2^{-j(p-1)/2}2^{(3-p)(k-j)/2}Q_j & \text{if } p\in]2,3[\\ b_k(n)\sum_{j=0}^{k-1}2^{-j}\Big\{\log(n-n_{k-1}+1)-\log(n-n_k+1)\Big\}Q_j & \text{if } p=3\end{array}\right.\,.$$

According to inequality (5.32) in [11], we infer that under conditions (2.1) and (2.2),

$$\mathbb{E}\Big(|(Z_j^{(1)})_{a,b}|\big|(U_j-\tilde{U}_j)_c\big|^{p-2}\Big)\leq C2^{j(1-2/p)}\|(Z_j^{(0)})_{a,b}\|_{p/2}+C2^{j(p/2-1)}\|(Z_j^{(0)})_{a,b}\|_{1,\Phi,p}\,.$$

Hence, if $p\in]2,3[$,

$$\sum_{k=0}^{r}b_k(n)\Big|\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big|$$
$$\leq C\sum_{k=0}^{r}2^{(3-p)k/2}\times\sum_{a,b=1}^{m}\Big\{\sum_{j=0}^{k-1}2^{-2j/p}\|(Z_j^{(0)})_{a,b}\|_{p/2}+\sum_{j=0}^{k-1}2^{j(p/2-2)}\|(Z_j^{(0)})_{a,b}\|_{1,\Phi,p}\Big\}\,,\quad(4.28)$$

and if $p = 3$,

$$\sum_{k=0}^{r} b_k(n)\Big|\sum_{i=1}^{2^k}\mathbb{E}(B_{i,n}(g))\Big| \le C\sum_{k=0}^{r} b_k(n)\Big\{\log(n-n_{k-1}+1)-\log(n-n_k+1)\Big\} \\ \times\sum_{a,b=1}^{m}\Big\{\sum_{j=0}^{k-1}2^{-2j/3}\|(Z_j^{(0)})_{a,b}\|_{3/2}+\sum_{j=0}^{k-1}2^{-j/2}\|(Z_j^{(0)})_{a,b}\|_{1,\Phi,3}\Big\}. \quad (4.29)$$

But, denoting by $\alpha_{2^j,d} = \max_{k,\ell\in\{1,\dots,d\}}\|\mathbb{E}_0((M_{2^j})_k((M_{2^j})_\ell - \mathbb{E}((M_{2^j})_k((M_{2^j})_\ell\|_{p/2}$, we get

$$\begin{aligned}\sum_{a,b=1}^{m}\|(Z_j^{(0)})_{a,b}\|_{p/2} &= \sum_{a,b=1}^{d}\|\mathbb{E}_0((AM_{2^j})_a((AM_{2^j})_b - \mathbb{E}((AM_{2^j})_a((AM_{2^j})_b\|_{p/2}\\ &\le \sum_{a,b=1}^{d}\sum_{k,\ell=1}^{d}\big|(A)_{a,k}(A)_{b,\ell}\big|\|\mathbb{E}_0((M_{2^j})_k((M_{2^j})_\ell - \mathbb{E}((M_{2^j})_k((M_{2^j})_\ell\|_{p/2}\\ &\le d^2\sum_{a,k=1}^{d}(A)_{a,k}^2\alpha_{2^j,d} := d^2\|A\|_{HS}^2\alpha_{2^j,d}\,.\end{aligned}$$

Now, since P is orthogonal

$$\|A\|_{HS}^2 = \|\Gamma\|_{HS}^2 = \sum_{i=1}^{m-1}\lambda_i^{-1} + (d-m+1)\lambda_m^{-1} \le d\lambda_m^{-1}\,.$$

So,

$$\sum_{a,b=1}^{m}\|(Z_j^{(0)})_{a,b}\|_{p/2} \le d^3\lambda_m^{-1}\alpha_{2^j,d}\,.$$

Similarly, setting $\beta_{2^j,d} = \max_{k,\ell\in\{1,\dots,d\}}\|\mathbb{E}_0((M_{2^j})_k((M_{2^j})_\ell - \mathbb{E}((M_{2^j})_k((M_{2^j})_\ell\|_{1,\Phi,p}$, we get

$$\sum_{a,b=1}^{m}\|(Z_j^{(0)})_{a,b}\|_{1,\Phi,p} \le d^3\lambda_m^{-1}\beta_{2^j,d}\,.$$

Note now that due to the martingale property conditions (2.1) and (2.2) are respectively equivalent to

$$\sum_{j\ge0}2^{j(p/2-2)}\beta_{2^j,d} < \infty \ \text{ and } \ \sum_{j\ge0}2^{-2j/p}\alpha_{2^j,d} < \infty\,.$$

(See Remark 2.6 in [11]). Moreover

$$\sum_{k=0}^{r} b_k(n)\Big\{\log(n-n_{k-1}+1)-\log(n-n_k+1)\Big\} = \log(n+1)$$

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

$$\sum_{k=0}^{r} 2^{(3-p)k/2} \leq C n^{(3-p)/2}$$

Starting from (4.28) and (4.29) and considering the above computations, the upper bound (4.27) follows. This ends the proof of the lemma. $\diamond$

4.1.5 Proof of Remark 2.4

To simplify, assume that $(d_i)_{i\in\mathbb{Z}}$ is a vector-valued stationary sequence of reversed martingale differences such that $\mathbb{E}(d_0 d_0^t) = \mathrm{I}$ (otherwise we transform the r.v.'s to go back to this case as done in Section 4.1.1). Let us then prove that if (2.5) and (2.6) are satisfied for some $p \in]2,3]$, the conclusions of Theorem 2.1 hold for M_n. With this aim, we need to construct the approximating Brownian motion. As in the proof of Theorem 2.1, for $L \in \mathbb{N}$, let $m(L) \in \mathbb{N}$ be such that $m(L) \leq L$, and define

$$I_{k,L} =]2^L + (k-1)2^{m(L)}, 2^L + k2^{m(L)}] \cap \mathbb{N} \text{ and } U_{k,L} = \sum_{i\in I_{k,L}} d_i \,,\ k \in \{1, \cdots, 2^{L-m(L)}\}\,.$$

Let $P_{U_{k,L}|\mathcal{G}_{2^L+k2^{m(L)}+1}}$ be the conditional law of $U_{k,L}$ given $\mathcal{G}_{2^L+k2^{m(L)}+1}$ and $\mathcal{N}_{2^{m(L)}}$ be the $\mathcal{N}(0, 2^{m(L)}\mathrm{I})$-law. Assume that the probability space is assumed to be large enough to contain a sequence $(\delta_i)_{i\in\mathbb{Z}}$ of iid random variables uniformly distributed on $[0,1]$, independent of the sequence $(d_i)_{i\in\mathbb{Z}}$. As in the proof of Theorem 2.1, we infer that there exists a $\mathbb{R}^d$-valued random variable $V_{k,L}$ with law $\mathcal{N}_{2^{m(L)}}$, measurable with respect to $\sigma(\delta_{2^L+k2^{m(L)}}) \vee \sigma(U_{k,L}) \vee \mathcal{G}_{2^L+k2^{m(L)}+1}$, independent of $\mathcal{G}_{2^L+k2^{m(L)}+1}$ and such that

$$\mathbb{E}\big(\big|U_{k,L} - V_{k,L}\big|_d\big) = \sup_{g\in \mathrm{Lip}(|\cdot|_d, \mathcal{G}_{2^L+2^{m(L)}+1})} \mathbb{E}(g(U_{1,L})) - \mathbb{E}(g(V_{1,L}))\,.$$

As, in the proof of Theorem 2.1, by induction on k, we have then constructed Gaussian random variables $(V_{k,L})_{L\in\mathbb{N},k=1,\ldots,2^{L-m(L)}}$ that are mutually independent, and using the Skorohod lemma, we can construct a sequence $(Z_i)_{i\geq 1}$ of iid standard Gaussian random vectors in $\mathbb{R}^d$ such that, for any $L \in \mathbb{N}$ and any $k \in \{1, \cdots, 2^{L-m(L)}\}$, $V_{k,L} = \sum_{i\in I_{k,L}} Z_i$ a.s. The proof can be then completed if one can prove that, in the case of reversed martingale differences, Lemma 4.1 still holds. In this case, it reads as follows:

Lemma 4.2. *Let $T_n = \sum_{i=1}^n N_i$ where $(N_i)_{i\geq 1}$ be a sequence of iid $\mathbb{R}^d$-valued centered gaussian random variables with $\mathrm{Var}(N_1) = \mathrm{I}$. Let $p \in]2,3]$ and assume that (2.5) and (2.6) are satisfied. Then, there exists a positive constant C such that, for any positive integer n,*

$$\sup_{g\in \mathrm{Lip}(|\cdot|_d, \mathcal{G}_{n+1})} \mathbb{E}(g(M_n)) - \mathbb{E}(g(T_n)) \leq \begin{cases} C n^{(3-p)/2} & \text{if } p \in]2,3[\\ C \log n & \text{if } p = 3. \end{cases},$$

The proof of this lemma can be easily deduced from the one of Lemma 4.1 by writing $M_n = \sum_{j=1}^n d_{j,n}$ where $d_{j,n} = d_{n-j+1}$ for any $j \in \mathbb{Z}$ and by noticing that, for any integer n, $(d_{j,n})_{j\in\mathbb{Z}}$ is a stationary sequence of martingale differences with respect to $(\mathcal{F}_{j,n})_{j\in\mathbb{Z}}$ where $\mathcal{F}_{j,n} = \mathcal{G}_{n-j+1}$.

To end the proof of Remark 2.4, it remains to notice that, as in the proof of Theorem 2.2, Lemma 4.2 immediately leads to the conclusion of Theorem 2.2 for the reversed martingale M_n with $\mathcal{G}_{n+1}$ replacing $\mathcal{F}_0$. $\diamond$

4.2 Proof of the results of Section 3

4.2.1 Proof of Proposition 3.1.

The existence of a μ-invariant probability on $\mathcal{B}(X)$ is a well-known fact based on the compactness of X, see for instance Lemma 2.16 of [3]. Hence, we shall only prove uniqueness.

By Lemma 3.3 and the Borel-Cantelli lemma we see that for every $x, y \in X$ such that $f(x) = f(y)$,

$$d(B_{2^n} \cdot x, B_{2^n} \cdot y) \underset{n\to\infty}{\longrightarrow} 0 \qquad \mathbb{P}\text{-a.s.} \tag{4.30}$$

Let ν be a μ-invariant probability on X. By Lemma 2.17 of [3], for $\mathbb{P}$-almost every $\omega \in \Omega$, there exists a probability ν_ω on X such that for every continuous function φ on X, we have

$$\int_X \varphi(B_n(\omega) \cdot x)\nu(dx) \underset{n\to\infty}{\longrightarrow} \int_X \varphi(x)\nu_\omega(dx)\,. \tag{4.31}$$

Let φ be a continuous function on X. For every $\mathfrak{f} \in F$, pick $y_\mathfrak{f} \in X_\mathfrak{f}$. Using (4.30), combined with Fubini's theorem for $\mathbb{P}\otimes\nu$ and the uniform continuity of φ, we infer that for $\mathbb{P}$-almost every $\omega \in \Omega$ we have

$$|\varphi(B_{2^n}(\omega) \cdot y_\mathfrak{f}) - \varphi(B_{2^n}(\omega) \cdot x)| \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{for } \nu\text{-almost every } x \in X_\mathfrak{f}\,.$$

Recall that $(X_\mathfrak{f})_{\mathfrak{f}\in F}$ forms a partition of X and that, by Lemma 3.2, $\nu(X_\mathfrak{f}) = \frac{1}{|F|}$ for every $\mathfrak{f} \in F$ Hence, by Lebesgue dominated convergence theorem, using (4.31), we infer that, for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\frac{1}{|F|}\sum_{\mathfrak{f}\in F} \varphi(B_{2^n}(\omega) \cdot y_\mathfrak{f}) \underset{n\to\infty}{\longrightarrow} \int_X \varphi(x)\nu_\omega(dx)\,. \tag{4.32}$$

Let ω be such that (4.32) holds. Since X is compact, there exist $(n_k)_{k\geq 1}$ and $(Z_\mathfrak{f}(\omega))_{\mathfrak{f}\in F}$ such that $B_{2^{n_k}}(\omega) \cdot y_\mathfrak{f} \underset{k\to\infty}{\longrightarrow} Z_\mathfrak{f}(\omega)$.

From the previous considerations, we conclude that for every continuous φ on X,

$$\int_X \varphi(x)\nu_\omega(dx) = \frac{1}{|F|}\sum_{\mathfrak{f}\in F} \varphi(Z_\mathfrak{f}(\omega))\,.$$

Then, we infer that $\nu_\omega = \frac{1}{|F|}\sum_{\mathfrak{f}\in F}\delta_{Z_{\mathfrak{f}}(\omega)}$, hence does not depend on ν.

The uniqueness of ν then follows from the fact that for every continuous φ on X,

$$\int_X \varphi(x)\nu(dx) = \int_\Omega \Big(\int_X \varphi(x)\nu_\omega(dx)\Big) d\mathbb{P}(\omega)\,,$$

see Lemma 2.19 of [3]. □

4.2.2 Proof of Lemma 3.2.

Let ν be a μ-invariant probability on $\mathcal{B}(X)$. Let $\psi : F \to \mathbb{R}$. We have

$$\int_{G\times X} \psi(f(g\cdot x))\mu(dg)\nu(dx) = \int_X \psi(f(x))\nu(dx)\,.$$

Hence

$$\int_G \left(\sum_{\mathfrak{f}\in F} \psi(s(g)\mathfrak{f})\nu(X_{\mathfrak{f}})\right)\mu(dg) = \sum_{\mathfrak{f}\in F}\psi(\mathfrak{f})\nu(X_{\mathfrak{f}})\,.$$

Hence, $\sum_{\mathfrak{f}\in F}\nu(X_{\mathfrak{f}})\delta_{\mathfrak{f}}$ is an $s(\mu)$-invariant probability measure on F. Since $s(\mu)$ is adapted it follows from the Choquet-Deny theorem for compact groups, see for instance [20], (notice that for finite groups the proof is quite direct) that the latter measure is the Haar measure on F. □

4.2.3 Proof of Lemma 3.3.

Of course, it is enough to prove (3.6) with A_{2^n} in place of B_{2^n}. Let n_0 be the integer appearing in Definition 3.3.

Set $\eta_n := \sup_{x\neq y, f(x)=f(y)} \int_G \log\left(\frac{d(g\cdot x, g\cdot y)}{d(x,y)}\right)\mu^{*n}(dg)$. Then for every integers $k \geq 1$ and $\ell \in [0, n_0-1]$, $\eta_{kn_0+\ell} \leq k\eta_{n_0} + \eta_\ell \leq k\eta_{n_0} + \max_{0\leq m\leq n_0}\eta_m$. Since $\eta_{n_0} < 0$, it follows that $\eta_n < 0$ for every n large enough. Hence we may assume that $n_0 = 2^{r_0}$.

Using that $A_0 = e$, we have

$$\log\Big(\frac{d(A_{2^{n+r_0}}\cdot x, A_{2^{n+r_0}}\cdot y)}{d(x,y)}\Big) = \sum_{k=1}^{2^n}\log\Big(\frac{d(A_{2^{r_0}k}\cdot x, A_{2^{r_0}k}\cdot y)}{d(A_{2^{r_0}(k-1)}\cdot x, A_{2^{r_0}(k-1)}\cdot y)}\Big)$$

As mentionned in Remark 3.3 the right-hand side of (3.3) may be $-\infty$. To take care of that technical matter let us introduce some $R > 0$ (that clearly exists) such that

$$\sup_{x\neq y, f(x)=f(y)} \int_G \max\left(-R, \log\Big(\frac{d(g\cdot x, g\cdot y)}{d(x,y)}\Big)\right)\mu^{*2^{r_0}}(dg) < 0\,.$$

Let $\mu_0 := \mu^{*2^{r_0}}$. For every $x, y \in X$ and $g \in G$, set

$$\varphi_R(g,x,y) := \max\left(-R, \log\Big(\frac{d(g\cdot x, g\cdot y)}{d(x,y)}\Big)\right). \tag{4.33}$$

Set also $V_k := Y_{2^{r_0}k} \cdots Y_{2^{r_0}(k-1)+1}$.

Then, we have

$$\log\Big(\frac{d(A_{2^{n+r_0}}\cdot x, A_{2^{n+r_0}}\cdot y)}{d(x,y)}\Big) \le \sum_{k=1}^{2^n} \varphi_R(V_k, A_{2^{r_0}(k-1)}\cdot x, A_{2^{r_0}(k-1)}\cdot y)$$

$$= \sum_{k=1}^{2^n}\Big(\varphi_R(V_k, A_{2^{r_0}(k-1)}\cdot x, A_{2^{r_0}(k-1)}\cdot y) - \int_G \varphi_R\big(g, A_{2^{r_0}(k-1)}\cdot x, A_{2^{r_0}(k-1)}\cdot y\big)\,\mu_0(dg)\Big)$$

$$+ \sum_{k=1}^{2^n}\int_G \varphi_R\big(g, A_{2^{r_0}(k-1)}\cdot x, A_{2^{r_0}(k-1)}\cdot y\big)\,\mu_0(dg)$$

$$:= \sum_{k=1}^{2^n} D_{R,k}(x,y) + \sum_{k=1}^{2^n} E_{R,k}(x,y)$$

$$:= M_{R,2^n}(x,y) + Q_{R,2^n}(x,y)\,.$$

Notice that $(M_{R,n}(x,y))_{n\in\mathbb{N}}$ is a $(\mathcal{F}_{2^{r_0}n})_{n\in\mathbb{N}}$-martingale. In order to prove (3.6) it is enough to prove that for every $\varepsilon > 0$,

$$\sum_{n\ge 0}\sup_{x\neq y, f(x)=f(y)} \mathbb{P}(|M_{R,2^n}(x,y)| \ge 2^n\varepsilon) < \infty\,, \tag{4.34}$$

and that there exists $\tilde{\ell} > 0$ such that

$$\sum_{n\ge 0}\sup_{x\neq y, f(x)=f(y)} \mathbb{P}\left(Q_{R,2^n}(x,y) \ge -2^n\tilde{\ell}\right) < \infty\,. \tag{4.35}$$

We first prove (4.35). Set

$$-2\tilde{\ell} := \sup_{x\neq y, f(x)=f(y)} \int_G \max\left(-R, \log\Big(\frac{d(g\cdot x, g\cdot y)}{d(x,y)}\Big)\right)\,\mu^{*2^{r_0}}(dg) < 0\,.$$

Since $Q_{R,2^n}(x,y) \le -2^{n+1}\tilde{\ell}$, $\mathbb{P}(\left(Q_{R,2^n}(x,y) \ge -2^n\tilde{\ell}\right) = 0$ and the result follows.

Let us prove (4.34) (for every $\varepsilon > 0$). Define a function Λ on G by setting $\Lambda(g) := \max(R, \log^+ \mathrm{Lip}(g))$ for every $g \in G$.

Let $\Gamma_n := \cup_{k=1}^{2^n}\{\Lambda(V_k) > \varepsilon 2^n\}$ and

$$D'_{R,k}(x,y) := D_{R,k}(x,y)\mathbf{1}_{\{\Lambda(V_k)\le\varepsilon 2^n\}}$$
$$D''_{R,k}(x,y) := D'_{R,k}(x,y) - \mathbb{E}(D'_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)})\,.$$

Notice that $|D_{R,k}| \le \Lambda(V_k)+\mathbb{E}(\Lambda(V_0))$ and since $D_{R,k}$ is such that $\mathbb{E}(D_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)}) = 0$,

$$\mathbb{E}(D'_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)}) = -\mathbb{E}(D_{R,k}(x,y)\mathbf{1}_{\{\Lambda(V_k)>\varepsilon 2^n\}}|\mathcal{F}_{2^{r_0}(k-1)})\,.$$

Hence, using independence,

$$\begin{aligned}|\mathbb{E}(D'_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)})| &\le \mathbb{E}(\Lambda(V_k)\mathbf{1}_{\{\Lambda(V_k)>\varepsilon 2^n\}}|\mathcal{F}_{2^{r_0}(k-1)}) + \mathbb{E}(\Lambda(V_0))\mathbb{E}(\mathbf{1}_{\{\Lambda(V_k)>\varepsilon 2^n\}}|\mathcal{F}_{2^{r_0}(k-1)})\\ &= \mathbb{E}(\Lambda(V_0)\mathbf{1}_{\{\Lambda(V_0)>\varepsilon 2^n\}}) + \mathbb{E}(\Lambda(V_0))\mathbb{P}(\Lambda(V_0)>\varepsilon 2^n) = 3\mathbb{E}(\Lambda(V_0)\mathbf{1}_{\{\Lambda(V_0)>\varepsilon 2^n\}})\,.\end{aligned}$$

Since $\Lambda \in L^1(\mu_0)$, it follows

$$|\mathbb{E}(D'_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)})| \le 3\int_G \Lambda(g)\mathbf{1}_{\{\Lambda(g)>\varepsilon 2^n\}}\,\mu_0(dg) \underset{n\to\infty}{\longrightarrow} 0\,. \tag{4.36}$$

Now we have

$$\begin{aligned}\mathbb{P}(|M_{R,2^n}(x,y)| \ge 2^n\varepsilon) &\le \mathbb{P}(\Gamma_n) + \mathbb{P}(|\sum_{k=1}^{2^n} D'_{R,k}(x,y)| \ge 2^n\varepsilon)\\ &\le 2^n\mathbb{P}(\Lambda(V_0)\ge 2^n\varepsilon) + \mathbb{P}\big(\big|\sum_{k=1}^{2^n} D''_{R,k}(x,y)\big| \ge 2^{n-1}\varepsilon\big)\\ &\quad + \mathbb{P}\big(\big|\sum_{k=1}^{2^n}\mathbb{E}(D'_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)})\big| \ge 2^{n-1}\varepsilon\big)\,.\end{aligned}$$

Using (4.36), we see that for n large enough, $\mathbb{P}(|\sum_{k=1}^{2^n}\mathbb{E}(D'_{R,k}(x,y)|\mathcal{F}_{2^{r_0}(k-1)})| \ge 2^{n-1}\varepsilon) = 0$. Hence, for n large enough

$$\mathbb{P}(|M_{R,2^n}(x,y)| \ge 2^n\varepsilon) \le 2^n\mathbb{P}(\Lambda(V_0)\ge 2^n\varepsilon) + \frac{1}{\varepsilon^2 2^{n-2}}\mathbb{E}(\Lambda(V_0)^2\mathbf{1}_{\{\Lambda(V_0)\le 2^n\varepsilon\}})\,.$$

Then, (4.34) follows from standard computations and the fact that $\mathbb{E}(\Lambda(V_0)) < \infty$. □

4.2.4 Proof of Lemma 3.4.

Let n_0 be as in Definition 3.3.

Using that $(Y_n)_{n\ge 1}$ is iid we see that for every $k \le j \le 2k$

$$\sup_{x,y\in X, x\neq y}\mathbb{P}\left(\log\left(d(A_{j-1}\cdot x, A_{j-1}\cdot y)\right) \ge -\ell k\right) \le \sup_{x,y\in X, x\neq y}\mathbb{P}\left(\log\left(d(A_{k-1}\cdot x, A_{k-1}\cdot y)\right) \ge -\ell k\right)\,.$$

Similarly, for every $mn_0 \le k \le (m+1)n_0$, for some $m \ge 3$, we have (notice that $mn_0 - 1 \ge (m-1)n_0$ and $m+1 \le 2(m-1)$)

$$\begin{aligned}&\sup_{x,y\in X, x\neq y}\mathbb{P}\left(\log\left(d(A_{k-1}\cdot x, A_{k-1}\cdot y)\right) \ge -\ell k\right)\\ &\le \sup_{x,y\in X, x\neq y}\mathbb{P}\left(\log\left(d(A_{(m-1)n_0}\cdot x, A_{(m-1)n_0}\cdot y)\right) \ge -2\ell(m-1)n_0\right)\end{aligned}$$

Let $k \geq 1$. Proceeding as in the proof of Lemma 3.3, we see that, setting $W_k := Y_{kn_0} \cdots Y_{(k-1)n_0+1}$ and using the notation (4.33),

$$\begin{aligned}
&\log\left(d(A_{kn_0}\cdot x, A_{kn_0}\cdot y)\right) - \log\left(d(x,y)\right) \\
&\leq \sum_{j=1}^{k} \Big(\varphi_R(W_j, A_{(j-1)n_0}\cdot x, A_{(j-1)n_0}\cdot y) - \int_G \varphi_R(g, A_{(j-1)n_0}\cdot x, A_{(j-1)n_0}\cdot y)\mu^{*n_0}(dg) \\
&\qquad + \sum_{j=1}^{k} \int_G \varphi_R(g, A_{(j-1)n_0}\cdot x, A_{(j-1)n_0}\cdot y)\mu^{*n_0}(dg) \\
&\qquad := M_k + R_k\,.
\end{aligned}$$

Clearly, $R_k \leq -\gamma k$ where $\gamma := -\sup_{x,y\in X, x\neq y} \int_G \varphi_R(g, A_{(j-1)n_0}\cdot x, A_{(j-1)n_0}\cdot y)\mu^{*n_0}(dg) > 0$. Hence, (3.7) will hold, say with $\ell = \gamma/(4n_0)$, if we can prove prove that for every $\varepsilon > 0$,

$$\sum_{k\geq 1} k^{p-2}\mathbb{P}\Big(\max_{1\leq j\leq k} |M_j| > \varepsilon k\Big) < \infty\,. \tag{4.37}$$

Now, the fact that (4.37) holds for every $\varepsilon > 0$ may be proved as Theorem 4.1 of [8], where a more general result is obtained.

Hence, the proof of (3.7) is completed and it remains to prove (3.8).

Taking $p = 1$ in (4.37) and using that, for $2^\ell \leq k \leq 2^{\ell+1} - 1$, $\mathbb{P}(\max_{1\leq j\leq 2^\ell} |M_j| > \varepsilon 2^\ell) \leq \mathbb{P}(\max_{1\leq j\leq k} |M_j| > \varepsilon k/2)$, we infer that

$$\sum_{k\geq 1} \mathbb{P}\Big(\max_{1\leq j\leq 2^k} |M_j| > \varepsilon 2^k\Big) < \infty\,,$$

for every $\varepsilon > 0$ and (3.8) follows from the Borel-Cantelli lemma, since $R_k \leq -\gamma k$. □

4.2.5 Proof of Theorems 3.14 and 3.15.

Let $n \geq 1$. We have

$$T_n := \sigma(A_n, W_0) - n\lambda_\mu = \sum_{k=1}^{n} (\sigma(Y_k, A_{k-1}\cdot W_0) - \lambda_\mu)\,, \tag{4.38}$$

where all the summands have the same law.

We start by noticing that, by Proposition 3.10 (and Remark 3.6), using that $p \geq 2$,

$$\begin{aligned}
\sum_{n\geq 1} \|\mathbb{E}\big(\sigma(Y_n, A_{n-1}\cdot W_0\,|\sigma\{W_0\}\big) - \lambda_\mu\|_p &\leq \sum_{n\geq 1} n^{p-2}\|\mathbb{E}\big(\sigma(Y_n, A_{n-1}\cdot W_0\,|\sigma\{W_0\}\big) - \lambda_\mu\|_p \\
&\leq \sum_{n\geq 1} n^{p-2} \sup_{x\in X} |\mathbb{E}\big(\sigma(Y_n, A_{n-1}x\big) - \lambda_\mu| < \infty\,. \quad (4.39)
\end{aligned}$$

Hence, by Gordin's L^p-criteria, we infer that $\sigma(Y_n, A_{n-1}\cdot W_0) - \lambda_\mu = D_n + R_n - R_{n-1}$, where $(D_n)_{n\geq 1}$ is a stationary sequence of martingale differences in L^p and (R_n) is a stationary process in L^p. Write $M_n := \sum_{k=1}^n D_k$ and $(M_n)_i := \langle M_n, e_i\rangle$ for $1 \leq i \leq d$.

Then, it is clearly sufficient to prove the theorems for $(M_n)_{n\geq 1}$ instead of $(T_n)_{n\geq 1}$. Since we are back to the study of a martingale with stationary increments, we only have to check the conditions of Theorems 2.1 and 2.2. In particular, it suffices to check that

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big((M_n)_i (M_n)_j \big| \mathcal{F}_0\big) - \mathbb{E}\big((M_n)_i (M_n)_j\big) \big\|_{p/2} < \infty\,. \tag{4.40}$$

Clearly, using the basic equality $(a+b)^2 - a^2 - b^2 = 2ab$, the latter holds provided that one can prove that

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big(\widetilde{M}_n^2 \big| \mathcal{F}_0\big) - \mathbb{E}\big(\widetilde{M}_n^2\big) \big\|_{p/2} < \infty\,,$$

for any $\widetilde{M}_n \in \{(M_n)_i \,:\, 1 \leq i \leq d\} \cup \{(M_n)_i + (M_n)_j \,:\, 1 \leq i < j \leq d\}$. In particular, we are back to prove that (4.40) holds in the case of an $\mathbb{R}$-valued cocycle. As in the proof of Propositions 8 and 9 of [7], taking into account (4.39), it is enough to prove that

$$\sum_{n=1}^{\infty} \frac{1}{n^{3-p/2}} \big\| \mathbb{E}\big(T_n^2 \big| \mathcal{F}_0\big) - \mathbb{E}\big(T_n^2\big) \big\|_{p/2} < \infty\,. \tag{4.41}$$

Then, proceeding as in the proof of Proposition 12 of [7] (based on Proposition 11 there), making use of (3.17), (3.18) and (3.19), we infer that (4.41) holds. □

4.2.6 Proof of Theorems 3.16 and 3.17.

The case where $V_n = \sigma(A_n, W_0)$ follows directly from Theorems 3.14 and 3.15. It remains to prove them for the other possible values of V_n. We shall only complete the proof of Theorem 3.16 since the proof of Theorem 3.17 may be done as the proof of item 2. of Theorem 3.16. The proofs make use of results from [3].

Let us start by proving Item 1 of Theorem 3.16. It follows from item (d) of Theorem 10.9 of [3] that $\mathbb{P}$-almost surely

$$\sup_{n\geq 1} |\sigma(A_n, W_0) - \kappa(A_n)|_{\mathfrak{a}} < \infty\,.$$

Hence Item 1 for the Cartan projection follows from the result for $(\sigma(A_n, W_0) - n\lambda_\mu)_{n\geq 1}$. Similary, Theorem 10.9 of [3] implies that for every $x \in \mathcal{P}$,

$$\sup_{n\geq 1} |\sigma(A_n, x) - \kappa(A_n)|_{\mathfrak{a}} < \infty\,.$$

Hence Item 1 for $(\sigma(A_n,x)-n\lambda_\mu)_{n\geq 1}$, for every $x\in\mathcal{P}$, follows from the result for $(\kappa(A_n)-n\lambda_\mu)_{n\geq 1}$.

Let us prove Item 2 of Theorem 3.16. By (13.31) of [3], there exist $M>0$ and $\eta_1,\dots,\eta_r\in\mathcal{P}$, such that, for every $g\in G$ there exists $i\in\{1,\dots,r\}$ such that

$$|\sigma(g,\eta_i)-\kappa(g)|_{\mathfrak{a}}\leq M\,. \tag{4.42}$$

Let $(\Gamma_i)_{1\leq i\leq r}$ be a partition of Γ_μ such that for every $g\in\Gamma_i$ (4.42) holds. Let $(N_n)_{n\geq 1}$ be the sequence of iid normal variables appearing in Item 2 and approximating $(\sigma(A_n,W_0))_{n\geq 1}$. Set $W_n:=\sum_{k=1}^n N_k$. We have

$$\begin{aligned}\big|\kappa(A_n)-W_n\big|_{\mathfrak{a}}&=\sum_{\mathfrak{f}\in F}\sum_{i\in\{1,\dots,r\},:\,\eta_i\in X_{\mathfrak{f}}}\big|\kappa(A_n)-W_n\big|_{\mathfrak{a}}\mathbf{1}_{\Gamma_i}(A_n)\\&\leq M+\sum_{\mathfrak{f}\in F}\sum_{i\in\{1,\dots,r\},:\,\eta_i\in X_{\mathfrak{f}}}\big|\sigma(A_n,\eta_i)-W_n|_{\mathfrak{a}}\mathbf{1}_{\Gamma_i}(A_n)\,.\end{aligned}$$

For $\eta_i\in X_{\mathfrak{f}}$ write

$$\sigma(A_n,\eta_i)-W_n=\frac{1}{|F|}\Big[\int_{X_{\mathfrak{f}}}\big(\sigma(A_n,\eta_i)-\sigma(A_n,x)\big)\nu(dx)+\int_{X_{\mathfrak{f}}}\big(\sigma(A_n,x)-W_n\big)\nu(dx)\Big]\,.$$

Combining the above estimates and using Proposition 3.13, we infer that there exists $K>0$

$$\big\|\,\big|\kappa(A_n)-W_n\big|_{\mathfrak{a}}\,\big\|_1\leq rK+r\|\,|\sigma(A_n,W_0)-W_n|_{\mathfrak{a}}\,\|_1\,,$$

and the desired result follows.

The case where G is connected follows from Lemma 3.13 (see the remark after the lemma). The case where $G=GL_d(\mathbb{R})$ follows also from Lemma 3.13 once one has noticed that if

$$\varepsilon=\begin{pmatrix}-1&0&\dots&\dots&0\\0&1&0&\dots&0\\\vdots&\ddots&\ddots&\ddots&\vdots\\\vdots&&\ddots&\ddots&0\\0&\dots&\dots&0&1\end{pmatrix}$$

then $f(kP_c)=-f(k\varepsilon P_c)$ and $\sigma(g,kP_c)=\sigma(g,k\varepsilon P_c)$ for every $g\in GL_d(\mathbb{R})$ and every $kP_c\in G/P_c$. □

5 Appendix

The following result is a Fuk-Nagaev type inequality for martingales with moments of order $p \in]2,3]$.

Proposition 5.1. *Let $(d_i)_{i\in\mathbb{Z}}$ be a real stationary martingale differences sequence with respect to $(\mathcal{F}_i)_{i\in\mathbb{Z}}$. Let σ denote the standard deviation of d_0. Let $p \in]2,3]$. Assume that $\mathbb{E}|d_0|^p < \infty$ and that the conditions (2.1) and (2.2) are fulfilled. Let $M_n^* = \max_{1\le k\le n} M_k$. Then for any positive real x,*

$$\mathbb{P}(M_n^* \ge x) \le 7 \times 2^{2p-1/2}\Big(\frac{n\sigma^2}{x^2}\Big)^{p+1/2} \exp\Big(-\frac{x^2}{8n\sigma^2}\Big) + C_p n x^{-p}\,,$$

where C_p is a positive constant depending only on p, $\mathbb{E}|d_0|^p$ and on the two series (2.1) and (2.2) but not on n.

Proof of Proposition 5.1. The proof uses similar arguments as in the proof of Proposition 5.2 in [28].

Let x be a positive real and φ be the function from $\mathbb{R}$ to $\mathbb{R}^+$ defined by

$$\varphi(t) = \frac{(t-x/2)_+^p}{p(p-1)}\,.$$

Since φ is nonnegative and convex, $\varphi(M_n)$ is a submartingale. Consequently Doob's maximal inequality entails that

$$\mathbb{P}(M_n^* \ge x) = \mathbb{P}\big(\varphi(M_n^*) \ge \varphi(x)\big) \le \frac{\mathbb{E}(\varphi(M_n))}{\varphi(x)}\,.$$

Let now Y be a standard gaussian random variable independent of $(d_i)_{i\in\mathbb{Z}}$. Then

$$\mathbb{P}(M_n^* \ge x) \le \frac{2^p p(p-1)}{x^p}\big(\mathbb{E}(\varphi(\sigma\sqrt{n}Y)) + \mathbb{E}(\varphi(M_n)) - \mathbb{E}(\varphi(\sigma\sqrt{n}Y))\big)\,. \tag{5.1}$$

Since $\varphi''(t) = (t-x/2)_+^{p-2}$, it follows that φ'' is $(p-2)$ Hölder. Then,

$$\mathbb{E}(\varphi(M_n)) - \mathbb{E}(\varphi(\sigma\sqrt{n}Y)) \le \sup_{f\in\Lambda_p} \mathbb{E}(f(M_n) - f(\sigma\sqrt{n}Y))\,,$$

where (recall that $p \in]2,3]$) Λ_p is the class of real functions f which are 2-times continuously differentiable and such that

$$|f''(x) - f''(y)| \le |x-y|^{p-2} \quad \text{for any } (x,y) \in \mathbb{R}\times\mathbb{R}\,.$$

Now denoting by P_{M_n} the law of M_n and by $G_{n\sigma^2}$ the normal distribution $N(0, n\sigma^2)$ we have

$$\sup_{f\in\Lambda_p} \mathbb{E}(f(M_n) - f(\sigma\sqrt{n}Y)) := \zeta_p(P_{M_n}, G_{n\sigma^2})\,,$$

that is the Zolotarev distance of order p between P_{M_n} and $G_{n\sigma^2}$. Applying Theorem 2.1 in [11], we derive that, under (2.1) and (2.2), there exists a positive constant K_p depending only on p, $\mathbb{E}|d_0|^p$ and on the two series (2.1) and (2.2), but not on n, such that

$$\mathbb{E}(\varphi(M_n)) - \mathbb{E}(\varphi(\sigma\sqrt{n}Y)) \leq K_p n\,. \tag{5.2}$$

On an other hand, using the fact that $(u + x/2)^2 \geq ux + x^2/4$, we get that

$$\mathbb{E}(\varphi(\sigma\sqrt{n}Y)) \leq \frac{e^{-x^2/(8n\sigma^2)}}{p(p-1)} \int_0^\infty e^{-xu/(2n\sigma^2)} \frac{u^p}{\sigma\sqrt{2n\pi}} du\,.$$

With the change of variables $v = xu/(2n\sigma^2)$ and since $p \in]2, 3]$, we derive that

$$\begin{aligned} p(p-1)\mathbb{E}(\varphi(\sigma\sqrt{n}Y)) &\leq e^{-x^2/(8n\sigma^2)} \Big(\frac{2n\sigma^2}{x}\Big)^{p+1} \frac{1}{\sigma\sqrt{2n\pi}} \int_0^\infty v^p e^{-v} dv \\ &\leq \frac{7}{2} \frac{(2n\sigma^2)^{p+1/2}}{x^{p+1}} e^{-x^2/(8n\sigma^2)}\,. \end{aligned} \tag{5.3}$$

Starting from (5.1) and using (5.2) and (5.3) we get the result.

Acknowledgements We would like to thank Çağri Sert for useful discussions and references on the Iwasawa cocycle.

References

[1] Y. Benoist and J.-F. Quint, *Central limit theorem for linear groups*, Ann. Probab. **44** (2016) 1308-1340.

[2] Y. Benoist and J.-F. Quint, *Central limit theorem on hyperbolic groups*, Izvestiya Mathematics **80** (2016) p.5-26.

[3] Y. Benoist and J.-F. Quint, *Random walks on reductive groups*, manuscript, http://www.math.u-psud.fr/∼benoist/prepubli/prepublication.html

[4] I. Berkes, W. Liu, W. B. Wu, *Komlós-Major-Tusnády approximation under dependence*, Ann. Probab. **42** (2014) 794-817.

[5] M. Björklund, *Central limit theorems for Gromov hyperbolic groups* J. Theoret. Probab. **23** (2010), no. 3, 871-887.

[6] P. Bougerol and J. Lacroix, Products of random matrices with applications to Schrödinger operators. Progress in Probability and Statistics, 8. Birkhäuser Boston, Inc., Boston, MA, 1985.

[7] C. Cuny, J. Dedecker and C. Jan, *Limit theorems for the left random walk on $GL_d(\mathbb{R})$*, Ann. Inst. Henri Poincaré Probab. Stat. **53** (2017) 1839-1865.

[8] C. Cuny, J. Dedecker and F. Merlevède, *Large and moderate deviations for the left random walk on $GL_d(\mathbb{R})$*, ALEA Lat. Am. J. Probab. Math. Stat. **14** (2017) 503-527

[9] C. Cuny, J. Dedecker and F. Merlevède, *On the Komlós, Major and Tusnády strong approximation for some classes of random iterates.* Stochastic Process. Appl. **128** (2018) 1347-1385.

[10] C. Cuny and F. Merlevède, *Strong invariance principles with rate for reverse martingales and applications.* J. Theoret. Probab. **28** (2015), no. 1, 137–183.

[11] 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** (2009), no. 35, 978–1011.

[12] J. Dedecker J., F. Merlevède and E. Rio, *On strong approximation for the empirical process of stationary sequences.* Ann. Probab. **41** (2013), no. 5, 3051-3696.

[13] J. Dedecker, C. Prieur and P. Raynaud De Fitte, *Parametrized Kantorovich-Rubinstein theorem and application to the coupling of random variables.* Dependence in Probability and Statistics, Lectures Notes in Statistics. **187** (2006), 105-121.

[14] Y. Derriennic and M. Lin, *Convergence of iterates of averages of certain operator representations and of convolution powers*, J. Funct. Anal. **85** (1989), no. 1, 86-102.

[15] E. Eberlein, *On strong invariance principles under dependence assumptions* Ann. Probab. **14** (1986), no. 1, 260-270.

[16] H. Furstenberg and H. Kesten. Products of random matrices. Ann. Math. Stat. **31** (1960) 457469.

[17] Gouëzel, S. *Almost sure invariance principle for dynamical systems by spectral methods.* Ann. Probab. **38** (2010), no. 4, 1639–1671.

[18] Y. Guivarc'h and A. Raugi, *Frontire de Furstenberg, propriétés de contraction et théorèmes de convergence*, Z. Wahrsch. Verw. Gebiete **69** (1985), no. 2, 187-242.

[19] C. Horbez, *Central limit theorems for mapping class groups and* $\mathrm{OUT}(F_N)$, Geom. Topol. **22** (2018), no. 1, 105-156.

[20] K. Ito and Y. Kawada, *On the probability distribution on a compact group*, I. Proc. Phys.-Math. Soc. Japan (3) 22, (1940). 977-998.

[21] 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 (2001), thesis number 01REN10073.

[22] S. Karmakar and W. B. Wu, *Optimal Gaussian approximation for multiple times series*, to appear in Statistica Sinica.

[23] A. Korepanov, Z. Kosloff and I. Melbourne, *Martingale-coboundary decomposition for families of dynamical systems.* Ann. Inst. H. Poincaré Anal. Non Linéaire **35** (2018), no. 4, 859–885.

[24] M. Ledoux and M. Talagrand, M., *Probability in Banach spaces. Isoperimetry and processes.* Springer-Verlag, Berlin, 1991.

[25] E. Le Page, *Théorèmes limites pour les produits de matrices aléatoires*, Probability measures on groups (Oberwolfach, 1981), pp. 258-303, Lecture Notes in Math., 928, Springer, Berlin-New York, 1982.

[26] I. Melbourne and M. Nicol. *A vector-valued almost sure invariance principle for hyperbolic dynamical systems.* Ann. Probab. **37** (2009), no. 2, 478–505.

[27] F. Merlevède, M. Peligrad, and S. Utev, *Functional Gaussian Approximation for Dependent Structures*. (2019). Oxford Studies in Probability. **6**, Oxford University Press.

[28] F. Merlevède and E. Rio (2012). Strong approximation of partial sums under dependence conditions with application to dynamical systems. *Stochastic Process. Appl.* **122**, 386-417.

[29] D. Monrad and W. Philipp, *The problem of embedding vector-valued martingales in a Gaussian process*, Teor. Veroyatnost. i Primenen. **35** (1990), no. 2, 384–387; translation in Theory Probab. Appl. 35 (1990), no. 2, 374-377 (1991)

[30] G. Morrow and W. Philipp, *An almost sure invariance principle for Hilbert space valued martingales*, Trans. Amer. Math. Soc. **273** (1982), no. 1, 231-251.

[31] F. Pène, *Rate of convergence in the multidimensional central limit theorem for stationary processes. Application to the Knudsen gas and to the Sinai billiard.* Ann. Appl. Probab. **15** (2005), no. 4, 2331–2392.

[32] L. Rüschendorf, *The Wasserstein Distance and Approximation Theorems*, Z. Wahrsch. Verw. Gebiete **70** (1985), 117-129.

[33] A. V. Skorohod, *On a representation of random variables*, Teor. Verojatnost. i Primenen. **21**, (1976), 645-648.