The Annals of Probability
2010, Vol. 38, No. 4, 1639–1671
DOI: 10.1214/10-AOP525

ALMOST SURE INVARIANCE PRINCIPLE FOR DYNAMICAL SYSTEMS BY SPECTRAL METHODS

BY SÉBASTIEN GOUËZEL

IRMAR, Université de Rennes 1

We prove the almost sure invariance principle for stationary $\mathbb{R}^d$-valued random processes (with very precise dimension-independent error terms), solely under a strong assumption concerning the characteristic functions of these processes. This assumption is easy to check for large classes of dynamical systems or Markov chains using strong or weak spectral perturbation arguments.

The almost sure invariance principle is a very strong reinforcement of the central limit theorem: it ensures that the trajectories of a process can be matched with the trajectories of a Brownian motion in such a way that almost surely the error between the trajectories is negligible compared to the size of the trajectory (the result can be more or less precise, depending on the specific error term one can obtain). These kinds of results have a lot of consequences (see, e.g., Melbourne and Nicol [16] and references therein).

Such results are well known for one-dimensional processes, either independent or weakly dependent (see, among many others, Denker and Philipp [6], Hofbauer and Keller [13]), and for independent higher-dimensional processes [7, 25]. However, for weakly dependent higher-dimensional processes, difficulties arise since the techniques relying on the Skorokhod representation theorem do not work efficiently. In this direction, an approximation argument introduced by Berkes and Philipp [4] was recently generalized to a large class of weakly dependent sequences in Melbourne and Nicol [16]. Their results give explicit error terms in the vector-valued almost sure invariance principle and are applicable when the variables under consideration can be well approximated with respect to a suitably chosen filtration. In particular, these results apply to a large range of dynamical systems when they have some Markovian behavior and sufficient hyperbolicity.

Unfortunately, it is quite common to encounter dynamical systems for which there is no natural well-behaved filtration. It is, nevertheless, often easy to prove classical limit theorems, by using another class of arguments relying on spectral theory. These arguments automatically yield a very precise description of the characteristic functions of the process under consideration, thereby implying limit results. It is therefore desirable to develop an abstract argument, showing that sufficient control on the characteristic functions of a process implies the almost sure

Received July 2009; revised January 2010.
AMS 2000 subject classifications. 60F17, 37C30.
Key words and phrases. Almost sure invariance principle, coupling, transfer operator.

invariance principle for vector-valued observables. This is our goal in this paper. Berkes and Philipp [4], Theorem 5, gives such a result, but its assumptions are too strong for the applications we have in mind. Moreover, even when the previous approaches are applicable, our method gives much sharper error terms.

We will state our main probabilistic result, Theorem 1.2, in the next section and describe applications to dynamical systems and Markov chains in Section 2. The remaining sections are devoted to the proof of the main theorem.

1. Statement of the main result. For $d > 0$, let us consider an $\mathbb{R}^d$-valued process $(A_0, A_1, \dots)$, bounded in L^p for some $p > 2$. Under suitable assumptions to be introduced below, we wish to show that it can be almost surely approximated by a Brownian motion.

DEFINITION 1.1. For $\lambda \in (0, 1/2]$ and Σ^2 a (possibly degenerate) symmetric semi-positive-definite $d \times d$ matrix, we say that an $\mathbb{R}^d$-valued process $(A_0, A_1, \dots)$ satisfies an almost sure invariance principle with error exponent λ and limiting covariance Σ^2 if there exist a probability space Ω and two processes $(A_0^*, A_1^*, \dots)$ and $(B_0, B_1, \dots)$ on Ω such that:

1. the processes $(A_0, A_1, \dots)$ and $(A_0^*, A_1^*, \dots)$ have the same distribution;
2. the random variables $B_0, B_1, \dots$ are independent and distributed as $\mathcal{N}(0, \Sigma^2)$;
3. almost surely in Ω,

$$\left| \sum_{\ell=0}^{n-1} A_\ell^* - \sum_{\ell=0}^{n-1} B_\ell \right| = o(n^\lambda). \tag{1.1}$$

A Brownian motion at integer times coincides with a sum of i.i.d. Gaussian variables, hence this definition can also be formulated as an almost sure approximation by a Brownian motion, with error $o(n^\lambda)$.

Under some assumptions on the characteristic function of $(A_0, A_1, \dots)$, we will prove that this process satisfies an almost sure invariance principle. To simplify notation, for $t \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$, we will write e^{itx} instead of $e^{i\langle t,x\rangle}$.

Let us state our main assumption (H), ensuring that the process we consider is close enough to an independent process: there exist $\varepsilon_0 > 0$ and $C, c > 0$ such that for any $n, m > 0$, $b_1 < b_2 < \cdots < b_{n+m+1}$, $k > 0$ and $t_1, \dots, t_{n+m} \in \mathbb{R}^d$ with $|t_j| \le \varepsilon_0$, we have

$$\begin{aligned} &\bigl|E\bigl(e^{i\sum_{j=1}^n t_j(\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell)+i\sum_{j=n+1}^{n+m} t_j(\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell)}\bigr) \\ &\quad - E\bigl(e^{i\sum_{j=1}^n t_j(\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell)}\bigr)\cdot E\bigl(e^{i\sum_{j=n+1}^{n+m} t_j(\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell)}\bigr)\bigr| \\ &\quad \le C(1+\max|b_{j+1}-b_j|)^{C(n+m)}e^{-ck}. \end{aligned} \tag{H}$$

This assumption says that if one groups the random variables into $n+m$ blocks, then a gap of size k between two blocks gives characteristic functions which are exponentially close (in terms of k) to independent characteristic functions, with an error which is, for each block, polynomial in terms of the size of the block. This control is only required for Fourier parameters t_j close to 0.

Of course, the assumption is trivially satisfied for independent random variables. The interesting feature of this assumption is that it is also very easy to check for dynamical systems when the Fourier transfer operators are well understood; see Theorem 2.1 below.

Our main theorem follows.

THEOREM 1.2. *Let $(A_0, A_1, \dots)$ be a centered $\mathbb{R}^d$-valued stationary process, in L^p for some $p > 2$, satisfying (H). Then*:

1. *the covariance matrix* $\operatorname{cov}(\sum_{\ell=0}^{n-1} A_\ell)/n$ *converges to a matrix* Σ^2;
2. *the sequence* $\sum_{\ell=0}^{n-1} A_\ell/\sqrt{n}$ *converges in distribution to* $\mathcal{N}(0, \Sigma^2)$;
3. *the process* $(A_0, A_1, \dots)$ *satisfies an almost sure invariance principle with limiting covariance* Σ^2, *for any error exponent*

$$\lambda > \frac{p}{4p-4} = \frac{1}{4} + \frac{1}{(4p-4)}. \tag{1.2}$$

When $p = \infty$, the condition on the error becomes $\lambda > 1/4$, which is quite good and independent of the dimension. This condition $\lambda > 1/4$ had previously been obtained only for very specific classes of dynamical systems (in particular, closed under time reversal) for real-valued observables (see, e.g., Field, Melbourne and Török [8], Melbourne and Török [18]).

If the process is not stationary, then we need an additional assumption to ensure the (fast enough) convergence to a normal distribution.

THEOREM 1.3. *Let $(A_0, A_1, \dots)$ be an $\mathbb{R}^d$-valued process, bounded in L^p for some $p > 2$, satisfying (H). Assume, moreover, that $\sum |E(A_\ell)| < \infty$ and that there exists a matrix Σ^2 such that, for any $\alpha > 0$,*

$$\left|\operatorname{cov}\left(\sum_{\ell=m}^{m+n-1} A_\ell\right) - n\Sigma^2\right| \le Cn^\alpha, \tag{1.3}$$

uniformly in m, n. The sequence $\sum_{\ell=0}^{n-1} A_\ell/\sqrt{n}$ then converges in distribution to $\mathcal{N}(0, \Sigma^2)$. Moreover, the process $(A_0, A_1, \dots)$ satisfies an almost sure invariance principle, with limiting covariance Σ^2, for any error exponent $\lambda > p/(4p-4)$.

Theorem 1.2 is, in fact, a consequence of Theorem 1.3 since we will prove in Lemma 2.7 that a stationary process satisfying (H) always satisfies (1.3) (moreover, this inequality holds with $\alpha = 0$).

Contrary to the results of Berkes and Philipp [4], our results are dimension-independent for i.i.d. random variables (but they are not optimal in this case—see

Einmahl [7], Zaĭtsev [25, 26]—for i.i.d. sequences in L^p, $2 < p < \infty$, the almost sure invariance principle holds for any error exponent $\lambda \geq 1/p$).

In this paper, C will denote a positive constant whose precise value is irrelevant and may change from line to line.

2. Applications.

2.1. *Coding characteristic functions*. Let us first consider a very simple example: let $T(x) = 2x \bmod 1$ on the circle $S^1 = \mathbb{R}/\mathbb{Z}$ and consider a Lipschitz function $f : S^1 \to \mathbb{R}^d$ of vanishing average for Lebesgue measure. We would like to prove an almost sure invariance principle for the process $(f(x), f(Tx), f(T^2x), \ldots)$, where x is distributed on S^1 according to Lebesgue measure. Define an operator $\mathcal{L}_t$ on Lipschitz functions by $\mathcal{L}_t u(x) = \sum_{T(y)=x} e^{itf(y)} u(y)/2$. It is then easy to check that for any $t_0, \ldots, t_{n-1}$ in $\mathbb{R}^d$,

$$E(e^{i \sum_{\ell=0}^{n-1} t_\ell f \circ T^\ell}) = \int \mathcal{L}_{t_{n-1}} \cdots \mathcal{L}_{t_0} 1(x)\, dx. \tag{2.1}$$

Using the good spectral properties of the operators $\mathcal{L}_t$, it is not very hard to show that this implies (H).

In more complicated situations, it is often possible to encode in the same way the characteristic functions of the process under consideration into a family of operators. However, these operators may act on complicated Banach spaces (of distributions or measures). It is therefore desirable to introduce a more abstract setting that encompasses the essential properties of such a coding, as follows.

Consider an $\mathbb{R}^d$-valued process $(A_0, A_1, \ldots)$. Let $\mathcal{B}$ be a Banach space and let $\mathcal{L}_t$ (for $t \in \mathbb{R}^d$, $|t| \leq \varepsilon_0$) be linear operators acting continuously on $\mathcal{B}$. Assume that there exist $u_0 \in \mathcal{B}$ and $\xi_0 \in \mathcal{B}'$ (the dual of $\mathcal{B}$) such that for any $t_0, \ldots, t_{n-1} \in \mathbb{R}^d$ with $|t_j| \leq \varepsilon_0$,

$$E(e^{i \sum_{\ell=0}^{n-1} t_\ell A_\ell}) = \langle \xi_0, \mathcal{L}_{t_{n-1}} \mathcal{L}_{t_{n-2}} \cdots \mathcal{L}_{t_1} \mathcal{L}_{t_0} u_0 \rangle. \tag{2.2}$$

In this case, we say that the characteristic function of $(A_0, A_1, \ldots)$ is coded by $(\mathcal{B}, (\mathcal{L}_t)_{|t| \leq \varepsilon_0}, u_0, \xi_0)$.

We claim that the assumption (H) follows from suitable assumptions on the operators $\mathcal{L}_t$, which we now describe.

(I1) One can write $\mathcal{L}_0 = \Pi + Q$, where Π is a one-dimensional projection and Q is an operator on $\mathcal{B}$, with $Q\Pi = \Pi Q = 0$ and $\|Q^n\|_{\mathcal{B} \to \mathcal{B}} \leq C\kappa^n$ for some $\kappa < 1$.

(I2) There exists $C > 0$ such that $\|\mathcal{L}_t^n\|_{\mathcal{B} \to \mathcal{B}} \leq C$ for all $n \in \mathbb{N}$ and all small enough t.

We will denote this set of conditions by (I).

THEOREM 2.1. *Let (A_ℓ) be a process whose characteristic function is coded by a family of operators $(\mathcal{L}_t)$ and which is bounded in L^p for some $p > 2$. Assume that* (I) *holds. There then exist $a \in \mathbb{R}^d$ and a matrix Σ^2 such that $(\sum_{\ell=0}^{n-1} A_\ell - na)/\sqrt{n}$ converges to $\mathcal{N}(0, \Sigma^2)$. Moreover, the process $(A_\ell - a)_{\ell\in\mathbb{N}}$ satisfies an almost sure invariance principle with limiting covariance Σ^2 for any error exponent larger than $p/(4p-4)$.*

The proof will exhibit a as the limit of $E(A_\ell)$, give a formula for Σ^2 and derive the theorem from Theorem 1.3 since (H) and (1.3) follow from (I). Even better, under the assumptions of Theorem 2.1, we have

$$\left|\operatorname{cov}\left(\sum_{\ell=m}^{m+n-1} A_\ell\right) - n\Sigma^2\right| \le C. \tag{2.3}$$

This is proved in Lemma 2.7 below.

REMARK 2.2. Let us stress that the assumptions of this theorem are significantly weaker than those of similar results in the literature: we do not require that a perturbed eigenvalue has a good asymptotic expansion, or even that such an eigenvalue exists. In particular, to the best of the author's knowledge, the central limit theorem was not known under the assumptions of Theorem 2.1.

Before we prove Theorem 2.1 at the end of this section, let us describe some applications. We will explain how to check (I) in several practical situations. Let $T : X \to X$ be a dynamical system, let μ be a probability measure (invariant or not) and let $f : X \to \mathbb{R}^d$. We want to study the process $(f, f\circ T, f\circ T^2, \dots)$.

2.2. *Strong continuity.* Assume that the characteristic function of the process $(f, f\circ T, f\circ T^2, \dots)$ can be coded by a family of operators $\mathcal{L}_t$ on a Banach space $\mathcal{B}$ and that the operator $\mathcal{L}_0$ satisfies (I1), that is, it has a simple eigenvalue at 1, the rest of its spectrum being contained in a disk of radius $\kappa < 1$ (such an operator is said to be *quasicompact*).

PROPOSITION 2.3. *If the family $\mathcal{L}_t : \mathcal{B} \to \mathcal{B}$ depends continuously on the parameter t at $t = 0$, then* (I2) *is satisfied.*

PROOF. By classical perturbation theory, the spectral picture for $\mathcal{L}_0$ persists for small t: we can write $\mathcal{L}_t = \lambda(t)\Pi_t + Q_t$, where $\lambda(t) \in \mathbb{C}$, Π_t is a one-dimensional projection and $\|Q_t^n\| \le C\kappa^n$ for some $\kappa < 1$, uniformly for small t. If $|\lambda(t)| \le 1$ for small t, then we obtain (I2).

For small t, we have

$$\begin{aligned} E(e^{it\sum_{\ell=0}^{n-1} f\circ T^\ell}) &= \langle \xi_0, \mathcal{L}_t^n u_0\rangle = \lambda(t)^n \langle \xi_0, \Pi_t u_0\rangle + \langle \xi_0, Q_t^n u_0\rangle \\ &= \lambda(t)^n \langle \xi_0, \Pi_t u_0\rangle + O(\kappa^n). \end{aligned} \tag{2.4}$$

When $t \to 0$, by continuity, the quantity $\langle \xi_0, \Pi_t u_0 \rangle$ converges to $\langle \xi_0, \Pi u_0 \rangle = 1$; see (2.8) below. In particular, for small enough t, $\langle \xi_0, \Pi_t u_0 \rangle \neq 0$. Since the right-hand side of (2.4) is bounded by 1, this gives $|\lambda(t)| \leq 1$, completing the proof. □

Let us be more specific. Let T be an irreducible aperiodic subshift of finite type, let m be a Gibbs measure and let $f : X \to \mathbb{R}^d$ be Hölder continuous with $\int f \, dm = 0$. Let $\mathcal{L}$ be the transfer operator associated with T, defined, by duality, by $\int u \cdot v \circ T \, dm = \int \mathcal{L}u \cdot v \, dm$ and define perturbed operators $\mathcal{L}_t$ by $\mathcal{L}_t(u) = \mathcal{L}(e^{itf} u)$. These operators code the characteristic function of the process $(f, f \circ T, \dots)$ and depend analytically on t [this follows from the series expansion $e^{ix} = \sum (ix)^n / n!$ and the fact that the Hölder functions form a Banach algebra]. The condition (I) is checked in, for example, Guivarc'h and Hardy [11], Parry and Pollicott [19]. Hence, Theorem 2.1 gives an almost sure invariance principle for any error exponent greater than $1/4$. This result is not new: it is already given in Melbourne and Nicol [16], although with a weaker error term.

If T is an Anosov or Axiom A map and $f : X \to \mathbb{R}^d$ is Hölder continuous, then the same result follows using symbolic dynamics. One can also avoid it and directly apply Theorem 2.1 to the transfer operator acting on a Banach space $\mathcal{B}$ of distributions or distribution-like objects, as in Baladi and Tsujii [2], Gouëzel and Liverani [10].

Now, let $T : X \to X$ be a piecewise expanding map and assume that the expansion dominates the complexity (in the sense of Saussol [22], Lemma 2.2). This setting includes, in particular, all piecewise expanding maps of the interval since the complexity control is automatic in one dimension. Let $f : X \to \mathbb{R}^d$ be β-Hölder continuous for some $\beta \in (0, 1]$. The perturbed transfer operator $\mathcal{L}_t$ then acts continuously on the Banach space $\mathcal{B} = V_\beta$ introduced in Saussol [22] and depends analytically on t (since $\mathcal{B}$ is a Banach algebra). With Theorem 2.1, we get an almost sure invariance principle for any error exponent greater than $1/4$. This result was only known for $\dim(X) = 1$ and $d = 1$, thanks to Hofbauer and Keller [13].

This result also applies to coupled map lattices since Bardet, Gouëzel and Keller [3] shows (I) for such maps. We should point out that the Banach space $\mathcal{B}$ here is not a Banach space of functions or distributions, but this is of no importance in our abstract setting.

Assume, now, that T is the time-one map of a contact Anosov flow. Tsujii [24] constructs a Banach space of distributions on which the transfer operator $\mathcal{L}$ acts with a spectral gap. If f is smooth enough, then $\mathcal{L}_t := \mathcal{L}(e^{itf} \cdot)$ depends analytically on t. We therefore obtain an almost sure invariance principle for any error exponent greater than $1/4$. This result was known for real-valued observables [17], but is new for $\mathbb{R}^d$-valued observables. However, our method does not apply to the whole class of rapid-mixing hyperbolic flows, contrary to the martingale arguments of Melbourne and Török [17].

Finally, assume that $T : X \to X$ is a mixing Gibbs–Markov map with invariant measure m, that is, it is Markov for a partition α with infinitely many symbols and has the big image property and Hölder distortion (this is a generalization of the notion of a subshift of finite type to infinite alphabets, see, e.g., Melbourne and Nicol [16], Section 3.1, for precise definitions). For $f : X \to \mathbb{R}^d$ and $a \in \alpha$, let $Df(a)$ denote the best Lipschitz constant of f on a. Consider f of zero average such that $\sum_{a\in\alpha} m(a) Df(a)^\rho < \infty$ for some $\rho \in (0, 1]$ (this class of observables is very large, containing, in particular, all of the weighted Lipschitz observables of Melbourne and Nicol [16], Section 3.2).

THEOREM 2.4. *If $f \in L^p$ for some $p > 2$, then the process $(f, f \circ T, \ldots)$ satisfies an almost sure invariance principle for any error exponent $> p/(4p-4)$.*

This follows from Gouëzel ([9], Section 3.1), where a Banach space $\mathcal{B}$ satisfying the assumptions of Proposition 2.3 is constructed.

It should be mentioned that the almost sure invariance principle is invariant under the process of *inducing*, that is, going from a small dynamical system to a larger one. Many hyperbolic dynamical systems can be obtained by inducing from Gibbs–Markov maps and the previous theorem implies an almost sure invariance principle for all of them (see Melbourne and Nicol [16] for several examples).

REMARK 2.5. In such dynamical contexts (when the measure is invariant and ergodic), the matrix Σ^2 is degenerate if and only if f is an L^2 coboundary in some direction. Indeed, if Σ^2 is degenerate, then it follows from (2.3) that there is a nonzero direction t such that $\langle t, S_n f\rangle$ is bounded in L^2. By Leonov's theorem (see, e.g., Aaronson and Weiss [1]), this implies that $\langle t, f\rangle$ is an L^2 coboundary, that is, there exists $u \in L^2$ such that $\langle t, f\rangle = u - u \circ T$ almost everywhere. Conversely, this condition implies that Σ^2 is degenerate.

2.3. *Weak continuity.* In several situations, the strong continuity assumptions of the previous subsection are not satisfied, while a weaker form of continuity holds. We describe such a setting in this subsection.

Again, assume that the characteristic function of a process $(f, f \circ T, f \circ T^2, \ldots)$ is coded by a family of operators $\mathcal{L}_t$ on a Banach space $\mathcal{B}$ and that the operator $\mathcal{L}_0$ satisfies (I1), that is, it is quasicompact with a simple dominating eigenvalue at 1.

We do *not* assume that the map $t \mapsto \mathcal{L}_t$ is continuous from a neighborhood of 0 to the set of linear operators on $\mathcal{B}$, hence classical perturbation theory does not apply. Let $\mathcal{C}$ be a Banach space containing $\mathcal{B}$ on which the operators $\mathcal{L}_t$ act continuously and assume that there exist $M \geq 1$, $\kappa < 1$ and $C > 0$ such that:

1. for all $n \in \mathbb{N}$ and $|t| \leq \varepsilon_0$, we have $\|\mathcal{L}_t^n\|_{\mathcal{C}\to\mathcal{C}} \leq CM^n$;
2. for all $n \in \mathbb{N}$, all $|t| \leq \varepsilon_0$ and all $u \in \mathcal{B}$, we have $\|\mathcal{L}_t^n u\|_{\mathcal{B}} \leq C\kappa^n \|u\|_{\mathcal{B}} + CM^n \|u\|_{\mathcal{C}}$;

3. the quantity $\|\mathcal{L}_t - \mathcal{L}_0\|_{\mathcal{B}\to\mathcal{C}}$ tends to 0 when $t \to 0$.

Then Keller and Liverani [14], Liverani [15] show that, for small enough t, the operator $\mathcal{L}_t$ acting on $\mathcal{B}$ has a simple eigenvalue $\lambda(t)$ close to 1 and $\mathcal{L}_t$ can be written as $\lambda(t)\Pi_t + Q_t$, where Π_t is a one-dimensional projection and, for some $C > 0$ and $\tilde{\kappa} < 1$,

$$\|\Pi_t\|_{\mathcal{B}\to\mathcal{B}} \le C, \qquad \|Q_t^n\|_{\mathcal{B}\to\mathcal{B}} \le C\tilde{\kappa}^n,$$
$$\|\Pi_t - \Pi\|_{\mathcal{B}\to\mathcal{C}} \to 0 \qquad \text{when } t \to 0.$$

Therefore, (I2) follows from the arguments in the proof of Proposition 2.3 if we can prove that $\langle \xi_0, \Pi_t u_0\rangle \to \langle \xi_0, \Pi u_0\rangle$ when $t \to 0$. By the last estimate in the previous equation, this is true if ξ_0 belongs not only to $\mathcal{B}'$, but also to $\mathcal{C}'$, which is usually the case.

2.4. *Markov chains*. Consider a Markov chain $X_0, X_1, \dots$ (with an initial measure μ and a stationary measure m, possibly different from μ) on a state space $\mathcal{X}$. Also, let $f : \mathcal{X} \to \mathbb{R}$ with $E_m(f) = 0$. We want to study the process $A_\ell = f(X_\ell)$.

Denote by P the Markov operator associated with the Markov chain and define a perturbed operator $P_t(u) = P(e^{itf}u)$. Then

$$\begin{aligned} E_\mu(e^{i\sum_{\ell=0}^{n-1} t_\ell A_\ell}) &= E_\mu\big(e^{i\sum_{\ell=0}^{n-2} t_\ell f(X_\ell)} \cdot E\big(e^{it_{n-1}f(X_{n-1})} | X_{n-2}\big)\big) \\ &= E_\mu\big(e^{i\sum_{\ell=0}^{n-2} t_\ell f(X_\ell)} P_{t_{n-1}} 1(X_{n-2})\big). \end{aligned}$$

By induction, we obtain

$$E_\mu(e^{i\sum_{\ell=0}^{n-1} t_\ell A_\ell}) = \int P_{t_0} P_{t_1} \cdots P_{t_{n-1}} 1 \, d\mu. \tag{2.5}$$

This is very similar to the coding property introduced in (2.2), the (minor) difference being that the composition is made in the reverse direction. In particular, the proof of Theorem 2.1 still works in this context. We obtain the following result.

PROPOSITION 2.6. *Let $\mathcal{B}$ be a Banach space of functions on $\mathcal{X}$ such that $1 \in \mathcal{B}$ and integration with respect to μ is continuous on $\mathcal{B}$. If the operators P_t satisfy the condition* (I) *on $\mathcal{B}$, then the process $f(X_\ell)$ satisfies* (H). *If $f(X_\ell)$ is bounded in L^p for some $p > 2$, then it follows that the process $(f(X_\ell))$ satisfies an almost sure invariant principle for any error exponent $\lambda > p/(4p-4)$.*

To check condition (I), the arguments of Sections 2.2 or 2.3 can be applied (if the Banach space $\mathcal{B}$ is carefully chosen, depending on the properties of the random walk under consideration). In particular, we refer the reader to the article [12], where several examples are studied, including uniformly ergodic chains, geometrically ergodic chains and iterated random Lipschitz models. In particular, it is shown in this article that the weak continuity arguments of Section 2.3 are very powerful in some situations where the strong continuity of Section 2.2 does not hold.

2.5. *Proof of Theorem* 2.1 *assuming Theorem* 1.3.

Step 1: there exists $u_1 \in \mathcal{B}$ such that, for $t_0, \dots, t_{n-1} \in B(0, \varepsilon_0)$,

$$\Pi(\mathcal{L}_{t_{n-1}} \cdots \mathcal{L}_{t_0} u_0) = \langle \xi_0, \mathcal{L}_{t_{n-1}} \cdots \mathcal{L}_{t_0} u_0 \rangle u_1. \tag{2.6}$$

Since Π is a rank-one projection, we can write $\Pi(u) = \langle \xi_2, u \rangle u_2$ for some $u_2 \in \mathcal{B}$ and $\xi_2 \in \mathcal{B}'$ with $\langle \xi_2, u_2 \rangle = 1$. The trivial equality

$$E(e^{i \sum_{\ell=0}^{n-1} t_\ell A_\ell}) = E(e^{i \sum_{\ell=0}^{n-1} t_\ell A_\ell + \sum_{\ell=n}^{n+N-1} 0 \cdot A_\ell})$$

gives, using the coding by the operators $\mathcal{L}_t$,

$$\langle \xi_0, \mathcal{L}_{t_{n-1}} \cdots \mathcal{L}_{t_0} u_0 \rangle = \langle \xi_0, \mathcal{L}_0^N \mathcal{L}_{t_{n-1}} \cdots \mathcal{L}_{t_0} u_0 \rangle.$$

Let $u = \mathcal{L}_{t_{n-1}} \cdots \mathcal{L}_{t_0} u_0$. When N tends to infinity, $\mathcal{L}_0^N$ tends to Π. Hence, letting N tend to infinity in the previous equality, we get

$$\langle \xi_0, u \rangle = \langle \xi_0, \Pi u \rangle = \langle \xi_0, u_2 \rangle \cdot \langle \xi_2, u \rangle. \tag{2.7}$$

Moreover,

$$\begin{aligned} \langle \xi_0, u_0 \rangle &= \langle \xi_0, \Pi u_0 \rangle = \lim \langle \xi_0, \mathcal{L}_0^N u_0 \rangle \\ &= \lim E(e^{i \sum_{\ell=0}^{N-1} 0 \cdot A_\ell}) = 1. \end{aligned} \tag{2.8}$$

Taking $u = u_0$ in (2.7), this implies, in particular, that $\langle \xi_0, u_2 \rangle \neq 0$. Finally,

$$\Pi(u) = \langle \xi_2, u \rangle u_2 = \langle \xi_0, u \rangle u_2 / \langle \xi_0, u_2 \rangle.$$

We thus obtain (2.6) for $u_1 = u_2 / \langle \xi_0, u_2 \rangle$.

Step 2: (H) holds.

Consider $b_1 < \cdots < b_{n+m+1}$, as well as $t_1, \dots, t_{n+m} \in B(0, \varepsilon_0)$ and $k > 0$. Then

$$\begin{aligned} &E\big(e^{i \sum_{j=1}^n t_j (\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell) + i \sum_{j=n+1}^{n+m} t_j (\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell)}\big) \\ &\quad = \langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} \mathcal{L}_0^k \mathcal{L}_{t_n}^{b_{n+1}-b_n} \cdots \mathcal{L}_{t_1}^{b_2-b_1} \mathcal{L}_0^{b_1} u_0 \rangle \\ &\quad = \langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} (\mathcal{L}_0^k - \Pi) \mathcal{L}_{t_n}^{b_{n+1}-b_n} \cdots \mathcal{L}_{t_1}^{b_2-b_1} \mathcal{L}_0^{b_1} u_0 \rangle \\ &\qquad + \langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} \Pi \mathcal{L}_{t_n}^{b_{n+1}-b_n} \cdots \mathcal{L}_{t_1}^{b_2-b_1} \mathcal{L}_0^{b_1} u_0 \rangle. \end{aligned} \tag{2.9}$$

All of the operators $\mathcal{L}_{t_i}$ satisfy $\|\mathcal{L}_{t_i}^j\|_{\mathcal{B} \to \mathcal{B}} \leq C$. Since $\|\mathcal{L}_0^k - \Pi\|_{\mathcal{B} \to \mathcal{B}} \leq C\kappa^k$ for some $\kappa < 1$, it follows that the term on the penultimate line in (2.9) is bounded by $C^{n+m}\kappa^k$. Moreover, by (2.6), the term on the last line is

$$\langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} u_1 \rangle \cdot \langle \xi_0, \mathcal{L}_{t_n}^{b_{n+1}-b_n} \cdots \mathcal{L}_{t_1}^{b_2-b_1} \mathcal{L}_0^{b_1} u_0 \rangle.$$

The second factor in this equation is simply $E(e^{i\sum_{j=1}^{n} t_j(\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell)})$. Moreover,

$$
\begin{aligned}
&E\big(e^{i\sum_{j=n+1}^{n+m} t_j(\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell)}\big)\\
&\qquad = \langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} \mathcal{L}_0^{b_{n+1}+k} u_0\rangle\\
&\qquad = \langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} \Pi u_0\rangle + O(C^m \kappa^{b_{n+1}+k})\\
&\qquad = \langle \xi_0, \mathcal{L}_{t_{n+m}}^{b_{n+m+1}-b_{n+m}} \cdots \mathcal{L}_{t_{n+1}}^{b_{n+2}-b_{n+1}} u_1\rangle + O(C^m \kappa^{b_{n+1}+k}).
\end{aligned}
$$

Therefore, the last line of (2.9) is equal to

$$
E\big(e^{i\sum_{j=1}^{n} t_j(\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell)}\big) \cdot E\big(e^{i\sum_{j=n+1}^{n+m} t_j(\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell)}\big) + O(C^m \kappa^{b_{n+1}+k}).
$$

We have proven that the difference to be estimated to check (H) is bounded by $C^{n+m}\kappa^k + C^m \kappa^{b_{n+1}+k}$ for some $C > 1$ and $\kappa < 1$. If we write $C = 2^{C'}$ and $\kappa = e^{-c}$ for some $c, C' > 0$, then this is error is at most

$$
2 \cdot 2^{C'(n+m)} e^{-ck} \le 2 \cdot (1 + \max|b_{j+1} - b_j|)^{C'(n+m)} e^{-ck}.
$$

This proves (H).

Step 3: there exist $a \in \mathbb{R}^d$ and $C, \delta > 0$ such that $|E(A_\ell) - a| \le Ce^{-\delta\ell}$.

Working component by component, we can, without loss of generality, work with one-dimensional random variables.

Enriching the probability space if necessary, we can construct a centered random variable V, independent of all the A_ℓ and belonging to L^p, whose characteristic function is supported in $B(0, \varepsilon_0)$ (see Proposition 3.8 for the existence of V). Also, let $T > 0$. Then

$$
E(A_\ell) = E(A_\ell + V) = E\big((A_\ell + V)1_{|A_\ell+V|\ge T}\big) + \int_{|x|<T} x\, dP_{A_\ell+V}.
$$

The first term is bounded by $\|A_\ell + V\|_{L^2} \|1_{|A_\ell+V|\ge T}\|_{L^2} \le CP(|A_\ell + V| > T)^{1/2} \le C/T^{1/2}$. Let $\phi_\ell(t) = E(e^{itA_\ell})E(e^{itV})$ be the characteristic function of $A_\ell + V$. Let g_T be the Fourier transform of $x1_{|x|<T}$. Since the Fourier transform on $\mathbb{R}$ is an isometry up to a constant factor c_1, we have $\int_{|x|<T} x\, dP_{A_\ell+V} = c_1 \int \overline{g_T}\phi_\ell$, hence $E(A_\ell) = c_1 \int \overline{g_T}\phi_\ell + O(T^{-1/2})$.

We have

$$
\begin{aligned}
\phi_\ell(t) &= \langle \xi_0, \mathcal{L}_t \mathcal{L}_0^\ell u_0\rangle E(e^{itV})\\
&= \langle \xi_0, \mathcal{L}_t \Pi u_0\rangle E(e^{itV}) + \langle \xi_0, \mathcal{L}_t(\mathcal{L}_0^\ell - \Pi)u_0\rangle E(e^{itV}) =: \psi(t) + r_\ell(t).
\end{aligned}
$$

The function ψ is independent of ℓ, while the function $r_\ell(t)$ depends on ℓ, is bounded by $C\kappa^\ell$ and is supported in $\{|t| \le \varepsilon_0\}$. We obtain

$$
\begin{aligned}
E(A_\ell) &= c_1 \int \overline{g_T}\psi + c_1 \int \overline{g_T} r_\ell + O(T^{-1/2})\\
&= c_1 \int \overline{g_T}\psi + O(\|g_T\|_{L^2} \|r_\ell\|_{L^2}) + O(T^{-1/2}).
\end{aligned}
$$

The L^2-norm of g_T is equal to $C\|x1_{|x|<T}\|_{L^2} = CT^{3/2}$, therefore we obtain

$$E(A_\ell) = c_1 \int \overline{g_T}\psi + O(\kappa^\ell T^{3/2}) + O(T^{-1/2}).$$

Now, consider $k, \ell \in \mathbb{N}$. Taking $T = \kappa^{-\min(k,\ell)/3}$, we obtain, for some $\delta > 0$,

$$|E(A_\ell) - E(A_k)| \le Ce^{-\delta \min(k,\ell)}.$$

This shows that the sequence $E(A_\ell)$ is Cauchy, so it converges to a limit a. Moreover, letting $k \to \infty$, it also yields $|E(A_\ell) - a| \le Ce^{-\delta\ell}$, as desired.

Step 4: conclusion of the proof.

We claim that for any $m \in \mathbb{N}$, there exists a matrix s_m such that, uniformly in ℓ, m,

$$|\mathrm{cov}(A_\ell, A_{\ell+m}) - s_m| \le Ce^{-\delta\ell}. \tag{2.10}$$

Since the proof is almost identical to the third step, it will be omitted.

LEMMA 2.7. *Let (A_ℓ) be a process bounded in L^p for some $p > 2$, satisfying* (H) *and satisfying* (2.10) *for some sequence of matrices s_m. The series $\Sigma^2 = s_0 + \sum_{m=1}^\infty (s_m + s_m^*)$ then converges in norm and, uniformly in m, n,*

$$\left|\mathrm{cov}\left(\sum_{\ell=m}^{m+n-1} A_\ell\right) - n\Sigma^2\right| \le C. \tag{2.11}$$

Let us temporarily accept this lemma. The process $(A_\ell - a)$ then satisfies all the assumptions of Theorem 1.3. Theorem 2.1 follows from this theorem.

PROOF OF LEMMA 2.7. Let us first prove that for some $\delta > 0$,

$$|\mathrm{cov}(A_\ell, A_{\ell+m})| \le Ce^{-\delta m}. \tag{2.12}$$

To simplify notation, we will assume that $d = 1$. Although the estimate (2.12) follows easily from the techniques we will develop later in this paper, we will now give a direct elementary proof. Let V, V' be two independent random variables, as in the third step of the previous proof. Then

$$E(A_\ell A_{\ell+m}) = E\big((A_\ell + V)(A_{\ell+m} + V')\big) = \int xy\, dP(x, y),$$

where P is the distribution of $(A_\ell + V, A_{\ell+m} + V')$. For $T > 0$, we have

$$\begin{aligned}\int |xy| 1_{|x|>T}\, dP(x, y) &= E\big(|A_\ell + V||A_{\ell+m} + V'| 1_{|A_\ell+V|>T}\big)\\ &\le \|A_\ell + V\|_{L^p} \|A_{\ell+m} + V'\|_{L^2} \big\|1_{|A_\ell+V|>T}\big\|_{L^q},\end{aligned}$$

where $q > 1$ is chosen so that $1/p + 1/2 + 1/q = 1$. Moreover, $\|1_{|A_\ell+V|>T}\|_{L^q} = P(|A_\ell + V| > T)^{1/q} \le CT^{-1/q}$. We have proven that for some $\rho > 0$, we have

$\int |xy| 1_{|x|>T} \, dP(x, y) \leq CT^{-\rho}$. In the same way, $\int |xy| 1_{|y|>T} \, dP(x, y) \leq CT^{-\rho}$. Therefore,

$$E(A_\ell A_{\ell+m}) = \int xy 1_{|x|,|y| \leq T} \, dP(x, y) + O(T^{-\rho}).$$

The characteristic function ϕ of $(A_\ell + V, A_{\ell+m} + V')$ is given by

$$\phi(t, u) = E(e^{itA_\ell + iuA_{\ell+m}}) E(e^{itV}) E(e^{iuV'}).$$

It is therefore supported in $\{|t|, |u| \leq \varepsilon_0\}$. Denoting by h_T the Fourier transform of the function $xy1_{|x|,|y| \leq T}$ and using the fact that the Fourier transform is an isometry up to a constant factor $c_2 = c_1^2$, we get

$$E(A_\ell A_{\ell+m}) = c_2 \int \overline{h_T} \phi + O(T^{-\rho}).$$

Letting $\psi(t, u) = E(e^{itA_\ell}) E(e^{iuA_{\ell+m}}) E(e^{itV}) E(e^{iuV'})$, a similar computation shows that

$$E(A_\ell) E(A_{\ell+m}) = c_2 \int \overline{h_T} \psi + O(T^{-\rho}).$$

Therefore,

$$\begin{aligned} |E(A_\ell A_{\ell+m}) - E(A_\ell) E(A_{\ell+m})| &= c_2 \left| \int \overline{h_T} (\phi - \psi) \right| + O(T^{-\rho}) \\ &\leq C \|h_T\|_{L^2} \|\phi - \psi\|_{L^2} + O(T^{-\rho}). \end{aligned}$$

The function $\phi - \psi$ is supported in $\{|t|, |u| \leq \varepsilon_0\}$ and (H) implies that it is bounded by Ce^{-cm} for some $c > 0$. Moreover, $\|h_T\|_{L^2} = C\|xy1_{|x|,|y| \leq T}\|_{L^2} \leq CT^3$. Finally, we obtain

$$|E(A_\ell A_{\ell+m}) - E(A_\ell) E(A_{\ell+m})| \leq Ce^{-cm} T^3 + CT^{-\rho}.$$

Choosing $T = e^{cm/4}$, this gives (2.12).

When $\ell \to \infty$, $\operatorname{cov}(A_\ell, A_{\ell+m})$ tends to s_m, by assumption. Therefore, letting ℓ tend to infinity in (2.12), we get $|s_m| \leq Ce^{-\delta m}$. From (2.10), we obtain

$$|\operatorname{cov}(A_\ell, A_{\ell+m}) - s_m| \leq C \min(e^{-\delta \ell}, e^{-\delta m}). \tag{2.13}$$

We have

$$\begin{aligned} \operatorname{cov}\left(\sum_{\ell=m}^{m+n-1} A_\ell \right) &= \sum_{i=0}^{n-1} \operatorname{cov}(A_{i+m}) \\ &\quad + \sum_{0 \leq i < j \leq n-1} \big(\operatorname{cov}(A_{i+m}, A_{j+m}) + \operatorname{cov}(A_{i+m}, A_{j+m})^* \big). \end{aligned}$$

With (2.13), we get

$$\left|\operatorname{cov}\left(\sum_{\ell=m}^{m+n-1} A_\ell\right) - \sum_{i=0}^{n-1} s_0 - \sum_{0\le i<j\le n-1} (s_{j-i} + s_{j-i}^*)\right|$$
$$\le C \sum_{i=0}^{n-1} e^{-\delta(i+m)} + C \sum_{0\le i<j\le n-1} \min\bigl(e^{-\delta(i+m)}, e^{-\delta(j-i)}\bigr).$$

Up to a multiplicative constant C, this is bounded by

$$\sum_{i=0}^{\infty} e^{-\delta i} + \sum_{i=0}^{\infty} \sum_{j=i+1}^{2i} e^{-\delta i} + \sum_{i=0}^{\infty} \sum_{j=2i+1}^{\infty} e^{-\delta(j-i)} < \infty.$$

We have proven that

$$\left|\operatorname{cov}\left(\sum_{\ell=m}^{m+n-1} A_\ell\right) - ns_0 - \sum_{k=1}^{n} (n-k)(s_k + s_k^*)\right| \le C.$$

Since $\sum k|s_k + s_k^*| < \infty$, this proves (2.11). □

3. Probabilistic tools.

3.1. *Coupling*. As in Berkes and Philipp [4], the notion of coupling is central to our argument. In this subsection, we introduce this notion.

If $Z_1 : \Omega_1 \to E_1$ and $Z_2 : \Omega_2 \to E_2$ are two random variables on two (possibly different) probability spaces, then a *coupling* between Z_1 and Z_2 is a way to associate those random variables, usually so that this association shows that Z_1 and Z_2 are close in some suitable sense. Formally, a coupling between Z_1 and Z_2 is a probability space Ω', together with two random variables $Z_1' : \Omega \to E_1$ and $Z_2' : \Omega \to E_2$ such that Z_i' is distributed as Z_i. Considering the distribution of (Z_1', Z_2') in $E_1 \times E_2$, it follows that one may take, without loss of generality, $\Omega = E_1 \times E_2$, where Z_1' and Z_2' are the first and second projections.

The following lemma, also known as the Berkes–Philipp lemma, is Lemma A.1 of Berkes and Philipp [4]. It makes precise and justifies the intuition that, given a coupling between two random variables Z_1 and Z_2, and a coupling between Z_2 and another random variable Z_3, it is possible to ensure that those couplings live on the same probability space, giving a coupling between Z_1, Z_2 and Z_3.

LEMMA 3.1. *Let E_i, $i = 1, 2, 3$, be separable Banach spaces. Let F be a distribution on $E_1 \times E_2$ and let G be a distribution on $E_2 \times E_3$ such that the second marginal of F equals the first marginal of G. There then exist a probability space and three random variables Z_1, Z_2, Z_3 defined on this space such that the joint distribution of Z_1 and Z_2 is F and the joint distribution of Z_2 and Z_3 is G.*

As an application of this lemma, assume that two processes $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ are given and that a good coupling exists between variables X and Y distributed, respectively, like $\sum X_i$ and $\sum Y_i$. There then exists a coupling between $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ which realizes this coupling between $\sum X_i$ and $\sum Y_i$. It is sufficient to build, simultaneously:

- the trivial coupling between $(X_1, \dots, X_n)$ and X such that $X = \sum X_i$ almost surely;
- the given coupling between X and Y;
- the trivial coupling between Y and $(Y_1, \dots, Y_n)$ such that $Y = \sum Y_i$ almost surely.

This kind of argument will be used several times below, without further details.

We will need the following lemma. It ensures that, to obtain a coupling with good properties between two infinite processes $(Z_1, Z_2, \dots)$ and $(Z'_1, Z'_2, \dots)$, it is sufficient to do so for finite subsequences of these processes.

LEMMA 3.2. *Let u_n, v_n be two real sequences. Let $Z_n : \Omega \to E_n$ and $Z'_n : \Omega' \to E_n$ $(n \geq 1)$ be two sequences of random variables, taking values in separable Banach spaces. Assume that for any N there exists a coupling between $(Z_1, \dots, Z_N)$ and $(Z'_1, \dots, Z'_N)$ with*

$$P(|Z_n - Z'_n| \geq u_n) \leq v_n \tag{3.1}$$

for any $1 \leq n \leq N$. There then exists a coupling between $(Z_1, Z_2, \dots)$ and $(Z'_1, Z'_2, \dots)$ such that (3.1) holds for any $n \in \mathbb{N}$.

PROOF. For all $N \in \mathbb{N}$, there exists a probability measure P_N on $(E_1 \times \cdots \times E_N)^2$, the respective marginals of which are the distributions of $(Z_1, \dots, Z_N)$ and $(Z'_1, \dots, Z'_N)$, such that $P_N(|z_n - z'_n| \geq u_n) \leq v_n$ for $1 \leq n \leq N$, where z_n and z'_n denote the coordinates in the first and second E_n factors. Let us arbitrarily extend this measure to a probability measure $\tilde{P}_N$ on E^2, where $E = E_1 \times E_2 \times \cdots$. The sequence $\tilde{P}_N$ is tight and any of its weak limits satisfies the required property. □

3.2. *Prokhorov distance.*

DEFINITION 3.3. If P, Q are two probability distributions on a metric space, define their Prokhorov distance $\pi(P, Q)$ as the smallest $\varepsilon > 0$ such that $P(B) \leq \varepsilon + Q(B^\varepsilon)$ for any Borelian set B, where B^ε denotes the open ε-neighborhood of B.

This distance makes it possible to construct good couplings, thanks to the following result, known as the Strassen–Dudley theorem [5], Theorem 6.9.

THEOREM 3.4. *Let X, Y be two random variables taking values in a metric space with respective distributions P_X and P_Y. If $\pi(P_X, P_Y) < c$, then there exists a coupling between X and Y such that $P(d(X, Y) > c) < c$.*

We now turn to the estimation of the Prokhorov distance for processes taking values in $\mathbb{R}^d$. Let $d > 0$ and $N > 0$. We consider $\mathbb{R}^{dN}$ with the norm

$$|(x_1, \ldots, x_N)|_N = \sup_{1 \le i \le N} |x_i|,$$

where $|x|$ denotes the Euclidean norm of a point $x \in \mathbb{R}^d$.

LEMMA 3.5. *There exists a constant $C(d)$ with the following property. Let F and G be two probability distributions on $\mathbb{R}^{dN}$ with characteristic functions ϕ and γ. For any $T' > 0$,*

$$\pi(F, G) \le \sum_{j=1}^{N} F(|x_j| \ge T') + (C(d)T'^{d/2})^N \left[\int_{\mathbb{R}^{dN}} |\phi - \gamma|^2 \right]^{1/2}. \tag{3.2}$$

PROOF. After an approximation argument, we can assume, without loss of generality, that F and G have respective densities f and g. Then, for any measurable set A,

$$\begin{aligned} F(A) - G(A) &\le F(A \cap \max|x_j| \le T') + F(\max|x_j| > T') \\ &\qquad - G(A \cap \max|x_j| \le T') \\ &\le \int_{|x_1|,\ldots,|x_N| \le T'} |f - g| + \sum_{j=1}^{N} F(|x_j| > T'). \end{aligned}$$

Therefore, $\pi(F, G)$ is bounded by last line of this equation. To conclude, we have to estimate $\int_{|x_1|,\ldots,|x_N| \le T'} |f - g|$. We have

$$\int_{|x_1|,\ldots,|x_N| \le T'} |f - g| \le \|f - g\|_{L^2} \|1_{|x_1|,\ldots,|x_N| \le T'}\|_{L^2} = \|\phi - \gamma\|_{L^2} (CT')^{dN/2}$$

since the Fourier transform is an isometry on L^2 up to a factor $(2\pi)^{dN/2}$. This completes the proof. $\square$

3.3. *Classical tools.* Let us recall two classical results of probability theory that we will need later. The first is Rosenthal's inequality [21] and the second is a weak version of the Gal–Koksma strong law of large numbers [20], Theorem A1, which will be sufficient for our purposes.

PROPOSITION 3.6. *Let $X_1, \dots, X_n$ be independent centered real random variables and let $p > 2$. There exists a constant $C(p)$ such that*

$$\left\| \sum_{j=1}^{n} X_j \right\|_{L^p} \leq C(p) \left(\sum_{j=1}^{n} E(X_j^2) \right)^{1/2} + C(p) \left(\sum_{j=1}^{n} E(|X_j|^p) \right)^{1/p}. \tag{3.3}$$

PROPOSITION 3.7. *Let $X_1, X_2, \dots$ be centered real random variables and assume that for some $q \geq 1$ and some $C > 0$, for all m, n,*

$$E \left| \sum_{j=m}^{m+n-1} X_j \right|^2 \leq C n^q. \tag{3.4}$$

For any $\alpha > 0$, the sequence $\sum_{j=1}^{N} X_j / N^{q/2+\alpha}$ then tends almost surely to 0.

The following proposition will be used in several forthcoming constructions.

PROPOSITION 3.8. *There exists a symmetric random variable V on $\mathbb{R}^d$, belonging to L^q for any $q > 1$, whose characteristic function is supported in the set $\{|t| \leq \varepsilon_0\}$.*

PROOF. We start with a C^∞ function ϕ supported in $\{|t| \leq \varepsilon_0/2\}$ and consider its inverse Fourier transform $f = \mathcal{F}^{-1}(\phi)$ (which is C^∞ and rapidly decreasing). Let $g = |f|^2 = \mathcal{F}^{-1}(\phi \star \tilde{\phi})$, where $\tilde{\phi}(t) = \phi(-t)$. Finally, let $h = g / \int g$. This is nonnegative, has integral 1 and its Fourier transform is proportional to $\phi \star \tilde{\phi}$, hence it is supported in $\{|t| \leq \varepsilon_0\}$. If we let W and W' be independent random variables with density h, then $V = W - W'$ satisfies the conclusion of the proposition. □

4. L^p bounds. Our goal in this section is to show the following bound.

PROPOSITION 4.1. *Let $(A_0, A_1, \dots)$ be a centered process, bounded in L^p ($p > 2$) and satisfying* (H). *For any $\eta > 0$, there exists $C > 0$ such that, for all $m, n \geq 0$,*

$$\left\| \sum_{\ell=m}^{m+n-1} A_\ell \right\|_{L^{p-\eta}} \leq C n^{1/2}. \tag{4.1}$$

This kind of bound is classical for a large class of weakly dependent sequences. The main point of this proposition is that these bounds are established here solely under the assumption (H) on the characteristic function of the process.

For the proof, we will approximate the process $(A_0, A_1, \dots)$ by an independent process, using (H). Estimating the $L^{p-\eta}$-norm of this process via Rosenthal's inequality (Proposition 3.6), this will yield the desired estimate.

LEMMA 4.2. *Let $(A_0, A_1, \dots)$ be a centered process, bounded in L^p for some $p > 2$ and satisfying* (H). *Let $u_n = \max_{m\in\mathbb{N}} \| \sum_{\ell=m}^{m+n-1} A_\ell \|_{L^2}^2$. For any $\alpha > 0$, there exists $C > 0$ such that $u_{a+b} \le u_a + u_b + C(1 + a^\alpha + b^\alpha)(1 + u_a^{1/2} + u_b^{1/2})$ for any $a, b \ge 1$.*

PROOF. Let $m \in \mathbb{N}$ and $a \le b$. Write

$$X_1 = \sum_{\ell=m}^{m+a-1} A_\ell \quad \text{and} \quad X_2 = \sum_{\ell=m+a+\lfloor b^\alpha \rfloor}^{m+a+b-1} A_\ell.$$

Also, let $\tilde{X}_1 = X_1 + V_1$ and $\tilde{X}_2 = X_2 + V_2$, where V_1 and V_2 are independent random variables distributed like V (constructed in Proposition 3.8). Finally, let $\tilde{Y}_1$ and $\tilde{Y}_2$ be independent random variables, distributed like $\tilde{X}_1$ and $\tilde{X}_2$, respectively.

Let us prove that for some $\delta = \delta(\alpha) > 0$,

$$\pi((\tilde{X}_1, \tilde{X}_2), (\tilde{Y}_1, \tilde{Y}_2)) < Ce^{-b^\delta}. \tag{4.2}$$

Let ϕ and γ denote, respectively, the characteristic functions of (X_1, X_2) and (Y_1, Y_2), where Y_1 and Y_2 are independent copies of X_1 and X_2. Since there is a gap of size b^α between X_1 and X_2, (H) ensures that for Fourier parameters less than or equal to ε_0, $|\phi - \gamma| \le C(1 + b)^C e^{-cb^\alpha} \le Ce^{-c'b^\alpha}$. We have $\tilde{\phi} - \tilde{\gamma} = (\phi - \gamma)E(e^{i(t_1V_1+t_2V_2)})$. Since the characteristic function of V is supported in $\{|t| \le \varepsilon_0\}$, this shows that the characteristic functions $\tilde{\phi}$ and $\tilde{\gamma}$ of $(\tilde{X}_1, \tilde{X}_2)$ and $(\tilde{Y}_1, \tilde{Y}_2)$, respectively, satisfy $|\tilde{\phi} - \tilde{\gamma}| \le Ce^{-c'b^\alpha}$ and are supported in $\{|t| \le \varepsilon_0\}$. Applying Lemma 3.5 with $N = 2$ and $T' = e^{b^{\alpha/2}}$, we obtain (4.2) [since the first terms in (3.2) are bounded by $E(|\tilde{X}_i|)/T' \le Cb/e^{b^{\alpha/2}}$, while the second term is at most $CT'^d e^{-c'b^\alpha}$].

By (4.2) and Theorem 3.4, we can construct a coupling between $(\tilde{X}_1, \tilde{X}_2)$ and $(\tilde{Y}_1, \tilde{Y}_2)$ such that, outside a set O of measure at most Ce^{-b^δ}, we have $|\tilde{X}_i - \tilde{Y}_i| \le Ce^{-b^\delta}$. Then $\|\tilde{X}_1 + \tilde{X}_2\|_{L^2}$ is bounded by

$$\|1_O(\tilde{X}_1 + \tilde{X}_2)\|_{L^2} + \|1_{O^c}(\tilde{X}_1 - \tilde{Y}_1 + \tilde{X}_2 - \tilde{Y}_2)\|_{L^2} + \|\tilde{Y}_1 + \tilde{Y}_2\|_{L^2}.$$

The first term is bounded by $\|1_O\|_{L^q} \|\tilde{X}_1 + \tilde{X}_2\|_{L^p}$, where q is chosen so that $1/p + 1/q = 1/2$. Hence, it is at most $Ce^{-b^\delta/q} b \le C$. The second term is bounded by $Ce^{-b^\delta} \le C$. Finally, since $\tilde{Y}_1$ and $\tilde{Y}_2$ are independent and centered, the last term is equal to $(E(\tilde{Y}_1^2) + E(\tilde{Y}_2^2))^{1/2}$.

Since $\|V\|_{L^2}$ is finite, we finally obtain

$$\|X_1 + X_2\|_{L^2}^2 \le C + E(Y_1^2) + E(Y_2^2) = C + E(X_1^2) + E(X_2^2).$$

Taking into account the missing block $\sum_{\ell=m+a}^{m+a+\lfloor b^\alpha \rfloor - 1} A_\ell$ (whose L^2-norm is at most Cb^α) and using the trivial inequality $\|U + V\|_{L^2}^2 \le \|U\|_{L^2}^2 + \|V\|_{L^2}^2 +$

$2\|U\|_{L^2}\|V\|_{L^2}$, we finally obtain

$$\left\|\sum_{\ell=m}^{m+a+b-1} A_\ell\right\|_{L^2}^2 \leq \left\|\sum_{\ell=m}^{m+a-1} A_\ell\right\|_{L^2}^2 + \left\|\sum_{\ell=m+a}^{m+a+b-1} A_\ell\right\|_{L^2}^2 + Cb^{2\alpha} + Cb^\alpha\left(\left\|\sum_{\ell=m}^{m+a-1} A_\ell\right\|_{L^2} + \left\|\sum_{\ell=m+a}^{m+a+b-1} A_\ell\right\|_{L^2}\right).$$

This proves the lemma. □

LEMMA 4.3. *Let $u_n \geq 0$ satisfy*

$$u_{a+b} \leq u_a + u_b + C(1 + a^\alpha + b^\alpha)(1 + u_a^{1/2} + u_b^{1/2}) \tag{4.3}$$

for all $a, b \geq 1$ and some $C > 0$, $\alpha \in (0, 1/2)$. Then $u_n = O(n)$.

PROOF. For any $\varepsilon > 0$ and $x, y \geq 0$, we have $xy \leq \varepsilon x^2 + \varepsilon^{-1} y^2$. From the assumption, we therefore obtain

$$u_{a+b} \leq (1+\varepsilon)(u_a + u_b) + C\varepsilon^{-1}\max(a^{2\alpha}, b^{2\alpha}).$$

Let $v_k = \max_{0\leq n<2^{k+1}} u_n$. It follows from the previous equation that

$$v_{k+1} \leq (2+2\varepsilon)v_k + C\varepsilon^{-1}2^{2\alpha k}.$$

In particular, we have

$$\frac{v_{k+1}}{(2+2\varepsilon)^{k+1}} \leq \frac{v_k}{(2+2\varepsilon)^k} + C\varepsilon^{-1}\frac{2^{2\alpha k}}{(2+2\varepsilon)^{k+1}}.$$

It follows inductively that $v_k/(2+2\varepsilon)^k \leq v_0 + C\varepsilon^{-1}\sum_j \frac{2^{2\alpha j}}{(2+2\varepsilon)^{j+1}} < \infty$. Hence, for any $\varepsilon > 0$, $v_k = O((2+2\varepsilon)^k)$, that is, for any $\rho > 1$, $u_n = O(n^\rho)$. Choosing ρ close enough to 1, we get, from (4.3), that $u_{a+b} \leq u_a + u_b + Ca^\beta + Cb^\beta$ for some $\beta < 1$. Therefore, $v_{k+1} \leq 2v_k + C2^{\beta k}$. As above, we deduce that $v_k/2^k$ is bounded, that is, $u_n = O(n)$. □

PROOF OF PROPOSITION 4.1. Lemmas 4.2 and 4.3 show that a centered process in L^p satisfying (H) satisfies the following bound in L^2:

$$\left\|\sum_{\ell=m}^{m+n-1} A_\ell\right\|_{L^2} \leq Cn^{1/2}. \tag{4.4}$$

Let us now show that the same bound holds in $L^{p-\eta}$ for any $\eta > 0$.

Let $\alpha = 1/10$. For $n \in \mathbb{N}$, let $a = \lfloor n^{1-\alpha}\rfloor$ and $b = \lfloor n^\alpha \rfloor$. Fixing $m \in \mathbb{N}$, we decompose the interval $[m, m+n)$ as the union of the intervals $I_j = [m + ja, m +$

$(j+1)a-b^2)$ and $I'_j = [m+(j+1)a-b^2, m+(j+1)a)$ for $0 \le j < b$, and a final interval $J = [m+ba, m+n)$.

Write $X_j = \sum_{\ell \in I_j} A_\ell$ and $\tilde{X}_j = X_j + V_j$, where the V_j are independent and distributed like V constructed in Proposition 3.8. Finally, let $\tilde{Y}_0, \dots, \tilde{Y}_{b-1}$ be independent random variables such that $\tilde{Y}_j$ is distributed like $\tilde{X}_j$. We claim, for some $\delta > 0$ and any $j \le b$, that

$$\pi((\tilde{X}_0, \dots, \tilde{X}_{j-1}), (\tilde{X}_0, \dots, \tilde{X}_{j-2}, \tilde{Y}_{j-1})) < Ce^{-n^\delta}. \tag{4.5}$$

Indeed, the $\tilde{X}_j$ are blocks, each of length at most n, and there are at most n^α blocks. Since there is a gap of length $b^2 = n^{2\alpha}$ between X_{j-2} and X_{j-1}, (H) shows that the difference between the characteristic functions of the members of (4.5) is at most $Cn^{Cn^\alpha} \cdot e^{-cn^{2\alpha}} \le Ce^{-c'n^{2\alpha}}$ (the terms V_j ensure that it is sufficient to consider Fourier parameters bounded by ε_0). The estimate (4.5) then follows from Lemma 3.5 by taking $T' = e^{n^\alpha}$ and $N = j$.

Summing the estimate in (4.5) over j, we obtain

$$\pi((\tilde{X}_0, \dots, \tilde{X}_{b-1}), (\tilde{Y}_0, \dots, \tilde{Y}_{b-1})) < Ce^{-n^\delta/2}. \tag{4.6}$$

By the Strassen–Dudley Theorem 3.4, we can therefore construct a coupling between those processes such that, outside a set O of measure at most $Ce^{-n^\delta/2}$, we have $|\tilde{X}_i - \tilde{Y}_i| \le Ce^{-n^\delta/2}$ for $0 \le i \le b-1$. As in the proof of Lemma 4.2, this gives

$$\left\| \sum_{j=0}^{b-1} \tilde{X}_j \right\|_{L^{p-\eta}} \le C + \left\| \sum_{j=0}^{b-1} \tilde{Y}_j \right\|_{L^{p-\eta}}.$$

Since the $\tilde{Y}_j$ are independent and centered, Rosenthal's inequality (Proposition 3.6) applies. Let us write $v_k = \max_{t\in\mathbb{N}} \| \sum_{\ell=t}^{t+k-1} A_\ell \|_{L^{p-\eta}}$. The $\tilde{Y}_j$ are bounded in L^2 by $a^{1/2}$ [by (4.4)] and in $L^{p-\eta}$ by $C + v_{a-b^2} \le Cv_{a-b^2}$. Hence,

$$\begin{aligned} \left\| \sum_{j=0}^{b-1} \tilde{Y}_j \right\|_{L^{p-\eta}} &\le C\left(\sum_{j=0}^{b-1} a\right)^{1/2} + C\left(\sum_{j=0}^{b-1} v_{a-b^2}^{p-\eta}\right)^{1/(p-\eta)} \\ &\le Cn^{1/2} + Cv_{a-b^2} b^{1/(p-\eta)}. \end{aligned}$$

Since $\tilde{X}_j = X_j + V_j$ and V_j is bounded by C in $L^{p-\eta}$, we get, from the two previous equations, that

$$\left\| \sum_{j=0}^{b-1} X_j \right\|_{L^{p-\eta}} \le Cn^{1/2} + Cv_{a-b^2} b^{1/(p-\eta)}.$$

Finally,

$$\left\|\sum_{\ell=m}^{m+n-1} A_\ell\right\|_{L^{p-\eta}} \le \left\|\sum_{j=0}^{b-1} X_j\right\|_{L^{p-\eta}} + \sum_{j=0}^{b-1}\sum_{\ell\in I'_j} \|A_\ell\|_{L^{p-\eta}} + \left\|\sum_{\ell=m+ab}^{m+n-1} A_\ell\right\|_{L^{p-\eta}}$$

$$\le Cn^{1/2} + Cv_{a-b^2}b^{1/(p-\eta)} + Cn^{3\alpha} + v_{n-ab}.$$

Therefore, since $3\alpha < 1/2$, we have

$$v_n \le Cn^{1/2} + Cv_{a-b^2}b^{1/(p-\eta)} + v_{n-ab}.$$

Moreover, $a \le n^{1-\alpha}$, $b \le n^\alpha$ and $n-ab \le a+b+1 \le Cn^{1-\alpha}$. If $v_n = O(n^r)$, then this gives $v_n = O(n^s)$ for $s = s(r) = \max(1/2, (1-\alpha)r + \alpha/(p-\eta))$. Starting from the trivial estimate $v_n = O(n)$, we get $v_n = O(n^{s(1)})$, then $v_n = O(n^{s(s(1))})$ and so on. Since $p-\eta > 2$, this gives, in finitely many steps, that $v_n = O(n^{1/2})$. □

5. Proof of the main theorem for nondegenerate covariance matrices. In this section, we consider a process $(A_0, A_1, \dots)$ satisfying the assumptions of Theorem 1.3 and such that the matrix Σ^2 is nondegenerate. We will prove that this process satisfies the conclusions of Theorem 1.3. Replacing, without loss of generality, A_ℓ by $A_\ell - E(A_\ell)$, we can assume that A_ℓ is centered. If K is a finite subset of $\mathbb{N}$, then we denote its cardinality by $|K|$.

The strategy of the proof is very classical: we subdivide the integers into blocks with gaps between them, make the sums over the different blocks independent using the gaps and (H), use approximation results for sums of independent random variables to handle the (now independent) sums over the different blocks and, finally, show that the fluctuations inside the blocks and the terms in the gaps do not contribute much to the asymptotics.

The interesting feature of our approach is the choice of the blocks. First, we subdivide $\mathbb{N}$ into the intervals $[2^n, 2^{n+1})$ and we then cut each of these intervals following a triadic Cantor-like approach: we put a relatively large gap in the middle, then we put slightly smaller gaps in the middle of each half and we continue on in this way. This procedure gives better results than the classical arguments taking blocks along a polynomial progression: this would give an error $p/(3p-2)$ in the theorem, while we obtain the better error term $p/(4p-4)$ with the Cantor-like decomposition. The reason is that, to create n manageable blocks, the classical arguments require gaps whose union is of order n^2, while the triadic decomposition only uses gaps whose union is of order n.

We will now describe the triadic procedure more precisely. Fix $\beta \in (0,1)$ and $\varepsilon \in (0, 1-\beta)$. Let $f = f(n) = \lfloor \beta n \rfloor$. We decompose $[2^n, 2^{n+1})$ as a union of $F = 2^f$ intervals $(I_{n,j})_{0\le j<F}$ of the same length, and F gaps $(J_{n,j})_{0\le j<F}$ between them, used to ensure sufficient independence. Good intervals and gaps are placed

alternatively, and in increasing order, as follows: $[2^n, 2^{n+1}) = J_{n,0} \cup I_{n,0} \cup J_{n,1} \cup I_{n,1} \cdots \cup J_{n,F-1} \cup I_{n,F-1}$.

The lengths of the gaps $J_{n,j}$ are chosen as follows. The middle interval $J_{n,F/2}$ has length $2^{\lfloor \varepsilon n \rfloor} 2^{f-1}$. It cuts the interval $[2^n, 2^{n+1})$ into two parts. The middle intervals of each of these parts, that is, $J_{n,F/4}$ and $J_{n,3F/4}$, have length $2^{\lfloor \varepsilon n \rfloor} 2^{f-2}$. The middle intervals of the remaining four parts have length $2^{\lfloor \varepsilon n \rfloor} 2^{f-3}$, and so on. More formally, for $1 \le j < F$, we write $j = \sum_{k=0}^{f-1} \alpha_k(j) 2^k$, where $\alpha_k(j) \in \{0, 1\}$, and consider the smallest number r with $\alpha_r(j) \neq 0$. The length of $J_{n,j}$ is then $2^{\lfloor \varepsilon n \rfloor} 2^r$. We say that this interval is of rank r. This defines the length of all the intervals $J_{n,j}$, except for $j = 0$. We let $|J_{n,0}| = 2^{\lfloor \varepsilon n \rfloor} 2^f$ and say that this interval has rank f.

Since there are 2^{f-1-r} intervals of rank r for $r < f$, with length $2^{\lfloor \varepsilon n \rfloor} 2^r$, the lengths of the intervals $(J_{n,j})_{0 \le j < F}$ add up to

$$|J_{n,0}| + \sum_{r=0}^{f-1} 2^{\lfloor \varepsilon n \rfloor} 2^r \cdot 2^{f-1-r} = 2^{\lfloor \varepsilon n \rfloor} 2^{f-1}(f+2). \tag{5.1}$$

Let $|I_{n,j}| = 2^{n-f} - (f+2) 2^{\lfloor \varepsilon n \rfloor - 1}$. This is a positive integer if n is large enough and $\sum |I_{n,j}| + \sum |J_{n,j}| = 2^n$, that is, those intervals exactly fill $[2^n, 2^{n+1})$. We will denote by $i_{n,j}$ the smallest element of $I_{n,j}$.

We will use the lexicographical order $\prec$ on the set $\{(n, j) \mid n \in \mathbb{N}, 0 \le j < F(n)\}$. It can also be described as follows: $(n, j) \prec (n', j')$ if the interval $I_{n,j}$ is to the left of $I_{n',j'}$. A sequence (n_k, j_k) tends to infinity for this order if and only if $n_k \to \infty$.

Let $X_{n,j} = \sum_{\ell \in I_{n,j}} A_\ell$ for $n \in \mathbb{N}$ and $0 \le j < F(n)$. Finally, write $\mathcal{I} = \bigcup_{n,j} I_{n,j}$ and $\mathcal{J} = \bigcup_{n,j} J_{n,j}$. The main steps of the proof are the following:

1. there exists a coupling between $(X_{n,j})$ and a sequence of independent random variables $(Y_{n,j})$, with $Y_{n,j}$ distributed like $X_{n,j}$, such that almost surely when $(n, j) \to \infty$,
$$\left| \sum_{(n',j') \prec (n,j)} X_{n',j'} - Y_{n',j'} \right| = o\big(2^{(\beta+\varepsilon)n/2}\big);$$
2. there exists a coupling between $(Y_{n,j})$ and a sequence of independent Gaussian random variables $Z_{n,j}$, with $\operatorname{cov}(Z_{n,j}) = |I_{n,j}| \Sigma^2$, such that almost surely when $(n, j) \to \infty$,
$$\left| \sum_{(n',j') \prec (n,j)} Y_{n',j'} - Z_{n',j'} \right| = o\big(2^{(\beta+\varepsilon)n/2} + 2^{((1-\beta)/2+\beta/p+\varepsilon)n}\big);$$
3. coupling the $X_{n,j}$ with the $Z_{n,j}$, by virtue of the first two steps, and writing $Z_{n,j}$ as the sum of $|I_{n,j}|$ Gaussian random variables $\mathcal{N}(0, \Sigma^2)$, we obtain (using Lemma 3.1 and the example that follows it) a coupling between $(A_\ell)_{\ell \in \mathcal{I}}$ and

$(B_\ell)_{\ell\in\mathcal{I}}$, where the B_ℓ are i.i.d. and distributed like $\mathcal{N}(0, \Sigma^2)$, such that, when (n, j) tends to infinity, we have

$$\left|\sum_{\ell<i_{n,j},\ell\in\mathcal{I}} A_\ell - B_\ell\right| = o\big(2^{(\beta+\varepsilon)n/2} + 2^{((1-\beta)/2+\beta/p+\varepsilon)n}\big);$$

4. we check that almost surely when $(n, j) \to \infty$,

$$\max_{m<|I_{n,j}|}\left|\sum_{\ell=i_{n,j}}^{i_{n,j}+m} A_\ell\right| = o\big(2^{((1-\beta)/2+\beta/p+\varepsilon)n}\big)$$

and, moreover, that a similar estimate holds for the B_ℓ's;
5. combining the last two steps, we get that when k tends to infinity,

$$\left|\sum_{\ell<k,\ell\in\mathcal{I}} A_\ell - B_\ell\right| = o\big(k^{(\beta+\varepsilon)/2} + k^{(1-\beta)/2+\beta/p+\varepsilon}\big);$$

6. finally, we prove that the gaps can be neglected, that is, almost surely

$$\sum_{\ell<k,\ell\in\mathcal{J}} A_\ell = o(k^{\beta/2+\varepsilon}) \tag{5.2}$$

and a similar estimate holds for the B_ℓ's.

Altogether, this gives a coupling for which almost surely

$$\left|\sum_{\ell<k} A_\ell - B_\ell\right| = o\big(k^{\beta/2+\varepsilon} + k^{(1-\beta)/2+\beta/p+\varepsilon}\big).$$

Let us choose β such that the two error terms are equal, that is, $\beta = p/(2p-2)$. We obtain an almost sure invariance principle with error term $p/(4p-4)+\varepsilon$ for any $\varepsilon > 0$. Since the almost sure invariance principle implies the central limit theorem, this proves Theorem 1.3, under the assumption that Σ^2 is nondegenerate.

It remains to justify Steps 1, 2, 4 and 6 since Steps 3 and 5 are trivial. This is done in the following subsections.

5.1. *Step* 1: *Coupling with independent random variables*. In this subsection, we justify the first step of the proof of Theorem 1.3 with the following proposition.

PROPOSITION 5.1. *There exists a coupling between* $(X_{n,j})$ *and* $(Y_{n,j})$ *such that*, *almost surely*, *when* (n, j) *tends to infinity*,

$$\left|\sum_{(n',j')\prec(n,j)} X_{n',j'} - Y_{n',j'}\right| = o\big(2^{(\beta+\varepsilon)n/2}\big).$$

The rest of this subsection is devoted to the proof of this proposition.

Let $V_{n,j}$, for $n, j \in \mathbb{N}$, be independent copies of V (constructed in Proposition 3.8), independent of everything else (we may need to enlarge the space to ensure their existence). Let $\tilde{X}_{n,j} = X_{n,j} + V_{n,j}$.

We will write $X_n = (X_{n,j})_{0 \le j < F(n)}$ and $\tilde{X}_n = (\tilde{X}_{n,j})_{0 \le j < F(n)}$. The proof of Proposition 5.1 has two parts: first, we make the different $\tilde{X}_n$ independent of each other, using the gaps $J_{n,0}$; then, inside each block $\tilde{X}_n$, we make the variables $\tilde{X}_{n,j}$ independent by using the smaller gaps $J_{n,j}$.

LEMMA 5.2. *Let $\tilde{Q}_n$ be a random variable distributed like $\tilde{X}_n$, but independent of $(\tilde{X}_1, \dots, \tilde{X}_{n-1})$. We have*

$$\pi((\tilde{X}_1, \dots, \tilde{X}_{n-1}, \tilde{X}_n), (\tilde{X}_1, \dots, \tilde{X}_{n-1}, \tilde{Q}_n)) < C4^{-n}. \tag{5.3}$$

PROOF. The random process $(X_1, \dots, X_n)$ takes its values in $\mathbb{R}^{dD}$ for $D = \sum_{m=1}^{n} F(m) \le \sum_{m=1}^{n} 2^{\beta m} \le C2^{\beta n}$. Moreover, each component in $\mathbb{R}^d$ of this process is one of the $X_{n,j}$, hence it is a sum of at most 2^n consecutive variables A_ℓ. On the other hand, the interval $J_{n,0}$ is a gap between $(X_j)_{j<n}$ and X_n, and its length k is $C^{\pm 1} 2^{\varepsilon n + \beta n}$. Let ϕ and $\gamma : \mathbb{R}^{dD} \to \mathbb{C}$ denote the respective characteristic functions of $(X_1, \dots, X_{n-1}, X_n)$ and $(X_1, \dots, X_{n-1}, Q_n)$, where Q_n is distributed like X_n and is independent of $(X_1, \dots, X_{n-1})$. The assumption (H) ensures that for Fourier parameters $t_{m,j}$ all bounded by ε_0, we have

$$|\phi - \gamma| \le C(1 + 2^n)^{CD} e^{-ck} \le C2^{nC2^{\beta n}} e^{-c2^{\beta n + \varepsilon n}} \le Ce^{-c' 2^{\beta n + \varepsilon n}},$$

if n is large enough.

Let $\tilde{\phi}$ and $\tilde{\gamma}$ be the characteristic functions of, respectively, $(\tilde{X}_1, \dots, \tilde{X}_n)$ and $(\tilde{X}_1, \dots, \tilde{Q}_n)$: they are obtained by multiplying ϕ and γ by the characteristic function of V in each variable. Since this function is supported in $\{|t| \le \varepsilon_0\}$, we obtain, in particular, that

$$|\tilde{\phi} - \tilde{\gamma}| \le Ce^{-c2^{\beta n + \varepsilon n}}. \tag{5.4}$$

We then use Lemma 3.5 with $N = D$ and $T' = e^{2^{\varepsilon n/2}}$ to get

$$\begin{aligned} &\pi((\tilde{X}_1, \dots, \tilde{X}_n), (\tilde{X}_1, \dots, \tilde{X}_{n-1}, \tilde{Q}_n)) \\ &\qquad \le \sum_{m \le n} \sum_{j < F(m)} P(|\tilde{X}_{m,j}| \ge e^{2^{\varepsilon n/2}}) + e^{CD2^{\varepsilon n/2}} e^{-c2^{\beta n + \varepsilon n}}. \end{aligned}$$

The second term is, again, bounded by $e^{-c' 2^{\beta n + \varepsilon n}}$, while each term in the first sum is bounded by $e^{-2^{\varepsilon n/2}} E(|\tilde{X}_{m,j}|) \le e^{-2^{\varepsilon n/2}} \cdot C2^n$. Summing over m and j, we obtain a bound of the form $Ce^{-2^{\delta n}}$, which is stronger than (5.3). □

COROLLARY 5.3. *Let $\tilde{R}_n = (\tilde{R}_{n,j})_{j<F(n)}$ be distributed like $\tilde{X}_n$ and such that the $\tilde{R}_n$ are independent of each other. There then exist $C > 0$ and a coupling between $(\tilde{X}_1, \tilde{X}_2, \dots)$ and $(\tilde{R}_1, \tilde{R}_2, \dots)$ such that for all (n, j),*

$$P(|\tilde{X}_{n,j} - \tilde{R}_{n,j}| \geq C4^{-n}) \leq C4^{-n}. \tag{5.5}$$

PROOF. By Lemma 3.2, it is enough to build such a coupling between $(\tilde{X}_1, \dots, \tilde{X}_N)$ and $(\tilde{R}_1, \dots, \tilde{R}_N)$ for fixed N (we just have to ensure that the constant C we obtain is independent of N, of course).

We use Lemma 5.2 to get a good coupling that makes $\tilde{X}_N$ independent of the other variables, then again use this lemma to make $\tilde{X}_{N-1}$ independent of the other ones and so on. In the end, this yields the desired coupling between $\tilde{X}$ and $\tilde{R}$.

Let us be more formal. For $n \leq N$, we denote by $(\tilde{R}_1^n, \dots, \tilde{R}_n^n)$ a process distributed like $(\tilde{X}_1, \dots, \tilde{X}_n)$. Also, let $\tilde{R}_n$ be distributed like $\tilde{X}_n$, independent of everything else. Lemma 5.2 and the Strassen–Dudley Theorem 3.4 give a good coupling between $(\tilde{R}_1^n, \dots, \tilde{R}_n^n)$ and $(\tilde{R}_1^{n-1}, \dots, \tilde{R}_{n-1}^{n-1}, \tilde{R}_n)$. Putting all those couplings together on a single space (by Lemma 3.1), we obtain a space on which live, in particular, $(\tilde{R}_1^N, \dots, \tilde{R}_N^N)$ and $(\tilde{R}_1, \dots, \tilde{R}_N)$, which are the processes we are trying to couple. Moreover,

$$|\tilde{R}_n^N - \tilde{R}_n| \leq \sum_{j=n+1}^{N} |\tilde{R}_n^j - \tilde{R}_n^{j-1}| + |\tilde{R}_n^n - \tilde{R}_n|.$$

If $|\tilde{R}_n^j - \tilde{R}_n^{j-1}| \leq C4^{-j}$ for $j \in [n+1, N]$ and $|\tilde{R}_n^n - \tilde{R}_n| \leq C4^{-n}$, then we get $|\tilde{R}_n^N - \tilde{R}_n| \leq C'4^{-n}$ for some constant C' independent of n and N. In particular, $P(|\tilde{R}_n^N - \tilde{R}_n| > C'4^{-n})$ is bounded by

$$\begin{aligned}
&\sum_{j=n+1}^{N} P(|\tilde{R}_n^j - \tilde{R}_n^{j-1}| > C4^{-j}) + P(|\tilde{R}_n^n - \tilde{R}_n| > C4^{-n}) \\
&\qquad \leq \sum_{j=n}^{N} C4^{-j} \leq C'4^{-n}.
\end{aligned}$$

□

LEMMA 5.4. *For any $n \in \mathbb{N}$, we have*

$$\pi\big((\tilde{R}_{n,j})_{0 \leq j < F(n)}, (\tilde{Y}_{n,j})_{0 \leq j < F(n)}\big) < C4^{-n}, \tag{5.6}$$

where $\tilde{Y}_{n,j} = Y_{n,j} + V_{n,j}$.

PROOF. Let $f = f(n) = \lfloor \beta n \rfloor$ and $F = 2^f$. We will first make the variables $(\tilde{R}_{n,j})_{j<F/2}$ independent of the variables $(\tilde{R}_{n,j})_{F/2 \leq j < F}$ by using the large gap $J_{n,F/2}$, then proceed in each remaining half using the gap in the middle of this half and so on.

We define $\tilde{Y}^i_{n,j}$ for $0 \le i \le f$ as follows: for $0 \le k < 2^{f-i}$, the random variable $\tilde{\mathcal{Y}}^i_{n,k} := (\tilde{Y}^i_{n,j})_{k2^i \le j < (k+1)2^i}$ is distributed like $(\tilde{X}_{n,j})_{k2^i \le j<(k+1)2^i}$, and $\tilde{\mathcal{Y}}^i_{n,k}$ is independent of $\tilde{\mathcal{Y}}^i_{n,k'}$ if $k \neq k'$. Hence, the process $(\tilde{Y}^f_{n,j})_{0 \le j < F}$ coincides with $(\tilde{R}_{n,j})_{0\le j<F}$, while $(\tilde{Y}^0_{n,j})_{0\le j<F}$ coincides with $(\tilde{Y}_{n,j})_{0\le j<F}$.

Writing $\tilde{Y}^i = (\tilde{Y}^i_{n,j})_{j<F}$, let us estimate $\pi(\tilde{Y}^i, \tilde{Y}^{i-1})$ for $1 \le i \le f$. Since the variables $\tilde{\mathcal{Y}}^i_{n,k}$ are already independent of one another for $0 \le k < 2^{f-i}$, we have

$$\pi(\tilde{Y}^i, \tilde{Y}^{i-1}) \le \sum_{k=0}^{2^{f-i}-1} \pi(\tilde{\mathcal{Y}}^i_{n,k}, (\tilde{\mathcal{Y}}^{i-1}_{n,2k}, \tilde{\mathcal{Y}}^{i-1}_{n,2k+1})). \tag{5.7}$$

Moreover, $\tilde{\mathcal{Y}}^i_{n,k}$ is made of 2^i sums of variables A_ℓ along blocks, each of these blocks has length at most 2^{n-f} and there is a gap $J_{n,k2^i+2^{i-1}}$ of size $C^{\pm 1}2^{\varepsilon n+i}$ in the middle. Therefore, (H) ensures that the difference between the characteristic functions of $\tilde{\mathcal{Y}}^i_{n,k}$ and $(\tilde{\mathcal{Y}}^{i-1}_{n,2k}, \tilde{\mathcal{Y}}^{i-1}_{n,2k+1})$ is at most

$$C(1+2^{n-f})^{C2^i} e^{-c2^{\varepsilon n+i}} \le Ce^{Cn2^i - c2^{\varepsilon n+i}} \le Ce^{-c'2^{\varepsilon n+i}},$$

if n is large enough. Taking $N = 2^i$ and $T' = e^{2^{\varepsilon n/2}}$ in Lemma 3.5, we obtain (with computations very similar to those in the proof of Lemma 5.2)

$$\pi(\tilde{\mathcal{Y}}^i_{n,k}, (\tilde{\mathcal{Y}}^{i-1}_{n,2k}, \tilde{\mathcal{Y}}^{i-1}_{n,2k+1})) \le Ce^{-2^{\delta n}}$$

for some $\delta > 0$. Summing over k in (5.7) and then over i, we obtain

$$\pi(\tilde{Y}^f, \tilde{Y}^0) \le \sum_{i=1}^{f} \pi(\tilde{Y}^i, \tilde{Y}^{i-1}) \le f2^f Ce^{-2^{\delta n}} \le Ce^{-2^{\delta n}/2}.$$

This gives, in particular, (5.6). □

PROOF OF PROPOSITION 5.1. We combine the coupling constructed in Corollary 5.3 with the couplings constructed in Lemma 5.4 for each n, using the Strassen–Dudley Theorem 3.4. We obtain a coupling between $(\tilde{X}_{n,j})$ and $(\tilde{Y}_{n,j})$ such that $P(|\tilde{X}_{n,j} - \tilde{Y}_{n,j}| \ge C4^{-n}) \le C4^{-n}$. Since $\sum_{n,j} 4^{-n} < \infty$, the Borel–Cantelli lemma ensures that almost surely

$$\sup_{(n,j)} \left| \sum_{(n',j') \prec (n,j)} \tilde{X}_{n',j'} - \tilde{Y}_{n',j'} \right| < \infty. \tag{5.8}$$

Moreover, $\tilde{X}_{n,j} = X_{n,j} + V_{n,j}$, where the random variables $V_{n,j}$ are centered, independent and in L^2. By the law of the iterated logarithm, almost surely, for any $\alpha > 0$,

$$\left| \sum_{(n',j') \prec (n,j)} V_{n',j'} \right| = o\big(\mathrm{Card}\{(n',j') \mid (n',j') \prec (n,j)\}^{1/2+\alpha}\big).$$

Moreover, $\mathrm{Card}\{(n', j') \mid (n', j') \prec (n, j)\} \leq \sum_{n'=1}^{n} \sum_{j'<F(n')} 1 \leq C2^{\beta n}$. We obtain almost surely

$$\left| \sum_{(n',j') \prec (n,j)} X_{n',j'} - \tilde{X}_{n',j'} \right| = o\big(2^{\beta n(1/2+\alpha)}\big).$$

A similar estimate holds for $Y_{n,j} - \tilde{Y}_{n,j}$. With (5.8), this proves the proposition. □

5.2. *Step* 2: *Coupling with Gaussian random vectors*. We are going to use Corollary 3 of Zaĭtsev [26]. Let us recall it here, for the convenience of the reader, in a form that is suitable for us (it is obtained from the statement of Zaĭtsev by taking $r = 10/e$, $\gamma = q$, $L_\gamma = M^q$, $n = b$ and $z' = Mz/5$).

PROPOSITION 5.5. *Let $Y_0, \dots, Y_{b-1}$ be independent centered $\mathbb{R}^d$-valued random vectors. Let $q \geq 2$ and $M = (\sum_{j=0}^{b-1} E|Y_j|^q)^{1/q}$. Assume that there exists a sequence $0 = m_0 < m_1 < \cdots < m_s = b$ satisfying the following condition. Letting $\zeta_k = Y_{m_k} + \cdots + Y_{m_{k+1}-1}$ and $B_k = \operatorname{cov} \zeta_k$, we assume that*

$$100M^2|v|^2 \leq \langle B_k v, v\rangle \leq 100CM^2|v|^2 \tag{5.9}$$

for all $v \in \mathbb{R}^d$, all $0 \leq k < s$ and some constant $C \geq 1$. There then exists a coupling between $(Y_0, \dots, Y_{b-1})$ and a sequence of independent Gaussian random vectors $(S_0, \dots, S_{b-1})$ such that $\operatorname{cov} S_j = \operatorname{cov} Y_j$ and, moreover,

$$P\left(\max_{1\leq i\leq b}\left|\sum_{j=0}^{i-1} Y_j - S_j\right| \geq Mz\right) \leq C'z^{-q} + \exp(-C'z) \tag{5.10}$$

for all $z \geq C' \log n$. Here, C' is a positive quantity depending only on C, the dimension d and the integrability exponent q.

The following lemma easily follows from the previous proposition.

LEMMA 5.6. *For $n \in \mathbb{N}$, there exists a coupling between $(Y_{n,0}, \dots, Y_{n,F(n)-1})$ and $(S_{n,0}, \dots, S_{n,F(n)-1})$, where the $S_{n,j}$ are independent centered Gaussian vectors with $\operatorname{cov} S_{n,j} = \operatorname{cov} Y_{n,j}$, such that*

$$\sum_n P\left(\max_{1\leq i\leq F(n)}\left|\sum_{j=0}^{i-1} Y_{n,j} - S_{n,j}\right| \geq 2^{((1-\beta)/2+\beta/p+\varepsilon/2)n}\right) < \infty. \tag{5.11}$$

PROOF. Let $q \in (2, p)$ and $n \in \mathbb{N}$. We want to apply Proposition 5.5 to the independent vectors $(Y_{n,j})_{0\leq j<F}$, where $F = F(n) = 2^{\lfloor \beta n\rfloor}$.

By Proposition 4.1, we have $\|Y_{n,j}\|_{L^q} \le C2^{(1-\beta)n/2}$. This implies that $M := (\sum_{j=0}^{F-1} E|Y_{n,j}|^q)^{1/q}$ satisfies

$$M \le C2^{\beta n/q} \cdot 2^{(1-\beta)n/2}. \tag{5.12}$$

By the assumptions of Theorem 1.3, $\operatorname{cov} Y_{n,j} = |I_{n,j}|\Sigma^2 + o(|I_{n,j}|^\alpha)$ for any $\alpha > 0$. In particular,

$$\operatorname{cov} Y_{n,j} = 2^{(1-\beta)n}\Sigma^2\big(1 + o(1)\big). \tag{5.13}$$

Moreover, we assume that the matrix Σ^2 is nondegenerate. Therefore, there exists a constant C_0 such that for any large enough n, any $0 \le m < m' < F(n)$ and any vector v, we have

$$C_0^{-1}(m'-m)2^{(1-\beta)n}|v|^2 \le \left\langle \sum_{j=m}^{m'-1} \operatorname{cov} Y_{n,j} v, v \right\rangle \le C_0(m'-m)2^{(1-\beta)n}|v|^2.$$

For $m = 0$ and $m' = F$, the quantity $(m'-m)2^{(1-\beta)n} = 2^{\lfloor \beta n \rfloor} \cdot 2^{(1-\beta)n}$ is much larger than M^2, by (5.12). On the other hand, each individual term (for $m' = m+1$) is bounded by

$$|\operatorname{cov} Y_{n,j}||v|^2 \le \|Y_{n,j}\|_{L^2}^2|v|^2 \le \|Y_{n,j}\|_{L^q}^2|v|^2 \le M^2|v|^2.$$

Therefore, we can group the $Y_{n,j}$ into consecutive blocks so that (5.9) is satisfied for some constant C.

Applying Proposition 5.5, we get a coupling between $(Y_{n,0}, \dots, Y_{n,F-1})$ and $(S_{n,0}, \dots, S_{n,F-1})$ such that

$$P\left(\max_{1 \le i \le F} \left| \sum_{j=0}^{i-1} Y_{n,j} - S_{n,j} \right| \ge 2^{\varepsilon n/3} M \right) \le C2^{-q\varepsilon n/3} \tag{5.14}$$

by (5.10), for $z = 2^{\varepsilon n/3}$. This quantity is summable in n. Since $2^{\varepsilon n/3}M \le 2^{((1-\beta)/2+\beta/p+\varepsilon/2)n}$ if q is close enough to p and n is large enough, this completes the proof of the lemma. □

LEMMA 5.7. *Let $Z_{n,j}$ be independent Gaussian random vectors such that $\operatorname{cov} Z_{n,j} = |I_{n,j}|\Sigma^2$. There then exists a coupling between $(S_{n,j})$ and $(Z_{n,j})$ such that almost surely*

$$\sum_{(n',j') \prec (n,j)} S_{n',j'} - Z_{n',j'} = o\big(2^{(\beta+\varepsilon)n/2}\big). \tag{5.15}$$

PROOF. Let $\alpha > 0$. Let $E_{n,j} = \mathcal{N}(0, |I_{n,j}|\Sigma^2 + 2^{\alpha n} I_d)$, where I_d is the identity matrix of dimension d. By assumption, $\operatorname{cov} S_{n,j} = |I_{n,j}|\Sigma^2 + o(2^{\alpha n})$. In particular, if n is large enough, we can write $|I_{n,j}|\Sigma^2 + 2^{\alpha n} I_d = \operatorname{cov} S_{n,j} + M_{n,j}$, where the matrix $M_{n,j}$ is positive definite and $|M_{n,j}| \le C2^{\alpha n}$. Therefore, $E_{n,j}$ is the sum of $S_{n,j}$ and an independent random variable distributed like $\mathcal{N}(0, M_{n,j})$. In the same way, $E_{n,j}$ is the sum of $Z_{n,j}$ and of an independent Gaussian $\mathcal{N}(0, 2^{\alpha n} I_d)$. Putting those decompositions on a single space, using Lemma 3.1, we obtain a coupling between $(S_{n,j})$ and $(Z_{n,j})$ such that the difference $D_{n,j} = S_{n,j} - Z_{n,j}$ is centered and where $\|D_{n,j}\|_{L^2} \leq C 2^{\alpha n/2}$.

We claim that this coupling satisfies the conclusion of the lemma if $\alpha < \varepsilon/2$. Indeed, by Etemadi's inequality ([5], Paragraph M19), we have, for any n,

$$
\begin{aligned}
P\left(\max_{1\leq i\leq F(n)}\left|\sum_{j=0}^{i-1} D_{n,j}\right| > 2^{(\beta+\varepsilon/2)n/2}\right)
&\leq C \max_{1\leq i\leq F(n)} P\left(\left|\sum_{j=0}^{i-1} D_{n,j}\right| > 2^{(\beta+\varepsilon/2)n/2}/3\right)\\
&\leq C \max_{1\leq i\leq F(n)} E\left(\left|\sum_{j=0}^{i-1} D_{n,j}\right|^2\right)\Big/ 2^{(\beta+\varepsilon/2)n}\\
&\leq C \sum_{j=0}^{F(n)-1} E(|D_{n,j}|^2)/2^{(\beta+\varepsilon/2)n} \leq C 2^{\beta n} 2^{\alpha n}/2^{(\beta+\varepsilon/2)n}.
\end{aligned}
$$

This is summable. Therefore, almost surely for large enough n and for $1 \leq i \leq F(n)$, we have $|\sum_{j=0}^{i-1} D_{n,j}| \leq 2^{(\beta+\varepsilon/2)n/2}$. The estimate (5.15) follows. □

Putting together the couplings constructed in Lemmas 5.6 and 5.7, we obtain a coupling satisfying the conclusions of Step 2.

5.3. *Step* 4: *Handling the maxima.* We recall that $i_{n,j}$ is the smallest element of the interval $I_{n,j}$.

LEMMA 5.8. *Almost surely when* $(n, j) \to \infty$,

$$
\max_{m<|I_{n,j}|}\left|\sum_{\ell=i_{n,j}}^{i_{n,j}+m} A_\ell\right| = o\big(2^{((1-\beta)/2+\beta/p+\varepsilon)n}\big).
$$

PROOF. Let $q \in (2, p)$. In L^q, the partial sums $\sum_{\ell=a}^{b-1} A_\ell$ are bounded by $C(b-a)^{1/2}$, by Proposition 4.1. Let $M_a^b = \max_{a\leq n\leq b}|\sum_{\ell=a}^{n-1} A_\ell|$. Corollary B1 in Serfling [23] then also shows that

$$
\|M_a^b\|_{L^q} \leq C(b-a)^{1/2} \tag{5.16}
$$

for a different constant C. In particular, if $\nu = (1-\beta)/2 + \beta/p + \varepsilon/2$, then

$$
\begin{aligned}
P\big(M_{i_{n,j}}^{i_{n,j}+|I_{n,j}|} \geq 2^{\nu n}\big) &\leq E\big(\big(M_{i_{n,j}}^{i_{n,j}+|I_{n,j}|}\big)^q\big)/2^{\nu n q}\\
&\leq C|I_{n,j}|^{q/2}/2^{\nu n q}.
\end{aligned}
$$

Moreover,

$$\sum_{n,j} |I_{n,j}|^{q/2}/2^{\nu nq} \leq \sum_n 2^{\beta n} \cdot 2^{(1-\beta)nq/2-\nu nq}.$$

This sum is finite if q is close enough to p. The Borel–Cantelli lemma gives the desired result. $\square$

5.4. *Step* 6: *The gaps.* Recall that $\mathcal{J}$ is the union of the gaps $J_{n,j}$. In this subsection, we prove the following lemma.

LEMMA 5.9. *For any $\alpha > 0$, there exists $C > 0$ such that for any interval $J \subset \mathbb{N}$,*

$$E\left|\sum_{\ell \in J \cap \mathcal{J}} A_\ell\right|^2 \leq C|J \cap \mathcal{J}|^{1+\alpha}.$$

Together with the Gal–Koksma strong law of large numbers (Proposition 3.7) applied with $q = 1 + \alpha$, this shows that for every $\alpha > 0$, almost surely

$$\sum_{\ell < k, \ell \in \mathcal{J}} A_\ell = o(|\mathcal{J} \cap [0,k]|^{1/2+\alpha}).$$

Moreover, for $k \in [2^n, 2^{n+1})$, we have [by (5.1)]

$$|\mathcal{J} \cap [0,k]| \leq \sum_{m=1}^{n} \sum_{j=0}^{F(m)-1} |J_{m,j}| \leq C \sum_{m=1}^{n} m 2^{\varepsilon m + \beta m} \leq C n 2^{\varepsilon n + \beta n} \leq C k^{\beta + 3\varepsilon/2}.$$

With the previous equation, we obtain (if α is small enough)

$$\sum_{\ell < k, \ell \in \mathcal{J}} A_\ell = o(k^{\beta/2+\varepsilon}).$$

This is (5.2), as desired.

PROOF OF LEMMA 5.9. We will freely use the convexity inequality

$$(a_1 + \cdots + a_k)^2 \leq k(a_1^2 + \cdots + a_k^2). \tag{5.17}$$

Let $J \subset \mathbb{N}$ be an interval. We decompose $J \cap \mathcal{J}$ as $J_0 \cup J_1 \cup J_2$, where J_0 and J_2 are, respectively, the first and the last interval of $J \cap \mathcal{J}$, and J_1 is the remaining part (it is therefore a union of full intervals of $\mathcal{J}$). Then

$$\left|\sum_{\ell \in J \cap \mathcal{J}} A_\ell\right|^2 \leq 3\left|\sum_{\ell \in J_0} A_\ell\right|^2 + 3\left|\sum_{\ell \in J_1} A_\ell\right|^2 + 3\left|\sum_{\ell \in J_2} A_\ell\right|^2.$$

The set J_0 is an interval, hence Proposition 4.1 gives $E|\sum_{\ell\in J_0} A_\ell|^2 \leq C|J_0|$. A similar inequality holds for J_2. To conclude the proof, it is therefore sufficient to prove that

$$E\left|\sum_{\ell\in J_1} A_\ell\right|^2 \leq C|J_1|^{1+\alpha}. \tag{5.18}$$

Since J_1 is not always an interval, this does not follow directly from Proposition 4.1. However, this is trivial if J_1 is empty. Otherwise, let N be such that $\max J_1 \in [2^N, 2^{N+1})$. Since the last interval in J_1 is contained in $[2^N, 2^{N+1})$, its length is $2^{\lfloor \varepsilon N\rfloor + r}$ for some $r \in [0, f(N)]$. In particular, $N \leq C\log|J_1|$.

We defined the notion of rank of an interval $J_{n,j}$ in the paragraph before equation (5.1): such an interval has rank $r \in [0, f(n)]$ if its length is $2^{\lfloor \varepsilon n\rfloor + r}$. There are $2^{f(n)-1-r}$ intervals of rank r in $[2^n, 2^{n+1})$ for $r < f(n)$ and one interval of rank $f(n)$.

For $n \in \mathbb{N}$ and $0 \leq r \leq f(n)$, let $\mathcal{J}^{(n,r)}$ denote the union of the intervals $J_{n,j}$ which are of rank r. The number of sets $\mathcal{J}^{(n,r)}$ intersecting J_1 is at most $\sum_{n=0}^{N}(f(n)+1) \leq CN^2$. Hence, by the convexity inequality (5.17),

$$\left|\sum_{\ell\in J_1} A_\ell\right|^2 \leq CN^2 \sum_{n,r}\left|\sum_{\ell\in J_1\cap\mathcal{J}^{(n,r)}} A_\ell\right|^2. \tag{5.19}$$

Let us fix some (n,r) and enumerate the intervals of $\mathcal{J}^{(n,r)}$ as $K_1,\dots,K_t$ for $t = 2^{f(n)-1-r}$ if $r < f(n)$ [or $t = 1$ if $r = f(n)$]. Let $T_s = \sum_{\ell\in K_s} A_\ell$. We claim that for any subset S of $\{1,\dots,t\}$,

$$E\left|\sum_{s\in S} T_s\right|^2 \leq C\sum_{s\in S} E|T_s|^2 + C|S|. \tag{5.20}$$

Let us show how this completes the proof. By Proposition 4.1, we have $E|T_s|^2 \leq C|K_s|$. Therefore, for any set K which is a union of intervals in $\mathcal{J}^{(n,r)}$, we obtain $E|\sum_{\ell\in K} A_\ell|^2 \leq C|K|$. This applies, in particular, to $K = J_1 \cap \mathcal{J}^{(n,r)}$. Therefore, (5.19) gives

$$E\left|\sum_{\ell\in J_1} A_\ell\right|^2 \leq CN^2\sum_{n,r}|J_1\cap\mathcal{J}^{(n,r)}| = CN^2|J_1|.$$

Together with the inequality $N \leq C\log|J_1|$, this proves (5.18), as desired.

It remains to prove (5.20). We first make the T_s independent, as follows. Let $(U_1,\dots,U_t)$ be independent random variables such that U_s is distributed like T_s. Also, let $V_1,\dots,V_t,V_1',\dots,V_t'$ be independent random variables distributed like V (constructed in Proposition 3.8) and write $\tilde{T}_s = T_s + V_s$, $\tilde{U}_s = U_s + V_s'$. We claim that for some $\delta > 0$ and $C > 0$,

$$\pi((\tilde{T}_1,\dots,\tilde{T}_t),(\tilde{U}_1,\dots,\tilde{U}_t)) < Ce^{-2^{\delta n}}. \tag{5.21}$$

To prove this estimate, we use the intervals of rank greater than r as gaps: we first make $\tilde{T}_1, \dots, \tilde{T}_{t/2}$ independent of $\tilde{T}_{t/2+1}, \dots, \tilde{T}_t$ using the gap $J_{n,F/2}$, then proceed in each half using the central gaps $J_{n,F/4}$ and $J_{n,3F/4}$, and so on. The details of the argument are exactly the same as in the proof of Lemma 5.4.

Thanks to this estimate and the Strassen–Dudley Theorem 3.4, we can construct a coupling between $(\tilde{T}_j)_{1\le j\le t}$ and $(\tilde{U}_j)_{1\le j\le t}$ such that, outside a set O of measure at most $Ce^{-2^{\delta n}}$, we have $|\tilde{T}_j - \tilde{U}_j| \le Ce^{-2^{\delta n}}$ for $1 \le j \le t$. For any subset S of $\{1, \dots, t\}$, we obtain (as in the proof of Lemma 4.2)

$$\Big\|\sum_{s\in S}\tilde{T}_s\Big\|_{L^2} \le \Big\|1_O\sum_{s\in S}\tilde{T}_s\Big\|_{L^2} + \Big\|1_{O^c}\sum_{s\in S}\tilde{T}_s - \tilde{U}_s\Big\|_{L^2} + \Big\|\sum_{s\in S}\tilde{U}_s\Big\|_{L^2}.$$

The first term is bounded by $\|1_O\|_{L^q}\|\sum_{s\in S}\tilde{T}_s\|_{L^p}$, where q is chosen such that $1/p + 1/q = 1/2$. Hence, it is at most $Ce^{-2^{\delta n}/q}2^n \le C$. The second term is bounded by $Cte^{-2^{\delta n}} \le C$. Therefore, $\|\sum_{s\in S} T_s\|_{L^2}$ is bounded by

$$\Big\|\sum_{s\in S}\tilde{T}_s\Big\|_{L^2} + \Big\|\sum_{s\in S}V_s\Big\|_{L^2} \le C + \Big\|\sum_{s\in S}U_s\Big\|_{L^2} + \Big\|\sum_{s\in S}V_s\Big\|_{L^2} + \Big\|\sum_{s\in S}V'_s\Big\|_{L^2}.$$

Since the U_s are centered independent random variables, $\|\sum_{s\in S} U_s\|_{L^2} = (\sum E(U_s^2))^{1/2} = (\sum E(T_s^2))^{1/2}$. In the same way, we have $\|\sum_{s\in S} V_s\|_{L^2} = \|\sum_{s\in S} V'_s\|_{L^2} = C|S|^{1/2}$. We get $\|\sum_{s\in S} T_s\|_{L^2} \le C + (\sum E(T_s^2))^{1/2} + C|S|^{1/2}$, which implies (5.20). □

6. Completing the proof of the main theorems. In this section, we first finish the proof of Theorem 1.3 when the matrix Σ^2 is degenerate and then derive Theorem 1.2 from Theorem 1.3.

LEMMA 6.1. *Let $(A_0, A_1, \dots)$ be a process satisfying the assumptions of Theorem 1.3 for $\Sigma^2 = 0$. Then almost surely $\sum_{\ell=0}^{n-1} A_\ell = o(n^\lambda)$ for any $\lambda > p/(4p-4)$.*

PROOF. Let $\beta > 0$ and $\varepsilon > 0$. Define a sequence of intervals $I_n = [n^{\beta+1}, (n+1)^{\beta+1}) \cap \mathbb{N}$ and denote by $i_n = \lceil n^{\beta+1}\rceil$ the smallest element of I_n. We claim that almost surely

$$\Big|\sum_{\ell=0}^{i_n-1} A_\ell\Big| = O(n^{1/2+\varepsilon}) \tag{6.1}$$

and

$$\max_{i\in I_n}\Big|\sum_{\ell=i_n}^{i} A_\ell\Big| = O(n^{\beta/2+1/p+\varepsilon}). \tag{6.2}$$

Taking $\beta = (p-2)/p$ to equate the error terms, we get $|\sum_{\ell \le k} A_\ell| = O(n^{1/2+\varepsilon})$, where $n = n(k)$ is the index of the interval I_n containing k. Since $n \le Ck^{1/(1+\beta)}$, we finally obtain an error term $O(k^{\lambda+2\lambda\varepsilon})$ for

$$\lambda = \frac{1}{2} \cdot \frac{1}{1+(p-2)/p} = \frac{p}{4p-4}.$$

This concludes the proof. It remains to establish (6.1) and (6.2).

By (1.3), $\|\sum_{\ell=0}^{i_n-1} A_\ell\|_{L^2} = O(n^{\alpha/2})$ for any $\alpha > 0$. Therefore,

$$P\left(\sum_{\ell=0}^{i_n-1} A_\ell \ge n^{1/2+\varepsilon}\right) \le \left\|\sum_{\ell=0}^{i_n-1} A_\ell\right\|_{L^2}^2 \Big/ n^{1+2\varepsilon} \le Cn^\alpha / n^{1+2\varepsilon}.$$

Taking $\alpha = \varepsilon$, this quantity is summable. Equation (6.1) follows.

Let $M_a^b = \max_{a \le n \le b} |\sum_{\ell=a}^{n-1} A_\ell|$. For $q < p$, we have

$$\begin{aligned} P\left(\max_{i \in I_n} \left|\sum_{\ell=i_n}^{i} A_\ell\right| \ge n^{\beta/2+1/p+\varepsilon}\right) &= P(M_{i_n}^{i_{n+1}} \ge n^{\beta/2+1/p+\varepsilon}) \\ &\le \|M_{i_n}^{i_{n+1}}\|_{L^q}^q / n^{q(\beta/2+1/p+\varepsilon)}. \end{aligned}$$

By (5.16), $\|M_{i_n}^{i_{n+1}}\|_{L^q} \le C(i_{n+1} - i_n)^{1/2} \le Cn^{\beta/2}$. Therefore, the last equation is bounded by $C/n^{q(1/p+\varepsilon)}$. This is summable if q is close enough to p. The estimate (6.2) follows. $\square$

Let $(A_0, A_1, \ldots)$ be a process satisfying the assumptions of Theorem 1.3 for some matrix Σ^2. Replacing A_ℓ by $A_\ell - E(A_\ell)$, we can assume that this process is centered. We decompose $\mathbb{R}^d$ as an orthogonal sum $E \oplus F$, where Σ^2 is nondegenerate on E and vanishes on F. The almost sure invariance principle along E is proved in Section 5, while Lemma 6.1 handles F. This proves Theorem 1.3.

Finally, Theorem 1.2 follows directly from Lemma 2.7 and Theorem 1.3.

REFERENCES

[1] AARONSON, J. and WEISS, B. (2000). Remarks on the tightness of cocycles. *Colloq. Math.* **84/85** 363–376. MR1784202

[2] BALADI, V. and TSUJII, M. (2007). Anisotropic Hölder and Sobolev spaces for hyperbolic diffeomorphisms. *Ann. Inst. Fourier* (*Grenoble*) **57** 127–154. MR2313087

[3] BARDET, J.-B., GOUËZEL, S. and KELLER, G. (2007). Limit theorems for coupled interval maps. *Stoch. Dyn.* **7** 17–36. MR2293068

[4] BERKES, I. and PHILIPP, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54. MR515811

[5] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[6] DENKER, M. and PHILIPP, W. (1984). Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam. Systems* **4** 541–552. MR779712

[7] EINMAHL, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR996984
[8] FIELD, M., MELBOURNE, I. and TÖRÖK, A. (2003). Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions. *Ergodic Theory Dynam. Systems* **23** 87–110. MR1971198
[9] GOUËZEL, S. (2008). Characterization of weak convergence of Birkhoff sums for Gibbs–Markov maps. Preprint.
[10] GOUËZEL, S. and LIVERANI, C. (2008). Compact locally maximal hyperbolic sets for smooth maps: Fine statistical properties. *J. Differential Geom.* **79** 433–477. MR2433929
[11] GUIVARC'H, Y. and HARDY, J. (1988). Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov. *Ann. Inst. H. Poincaré Probab. Statist.* **24** 73–98. MR937957
[12] HERVÉ, L. and PÈNE, F. (2008). Nagaev method via Keller–Liverani theorem. Preprint.
[13] HOFBAUER, F. and KELLER, G. (1982). Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.* **180** 119–140. MR656227
[14] KELLER, G. and LIVERANI, C. (1999). Stability of the spectrum for transfer operators. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (4) **28** 141–152. MR1679080
[15] LIVERANI, C. (2003). Invariant measures and their properties. A functional analytic point of view. In *Dynamical Systems. Part II. Pubbl. Cent. Ric. Mat. Ennio Giorgi* 185–237. Scuola Norm. Sup., Pisa. MR2071241
[16] MELBOURNE, I. and NICOL, M. (2009). A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab.* **37** 478–505.
[17] MELBOURNE, I. and TÖRÖK, A. (2002). Central limit theorems and invariance principles for time-one maps of hyperbolic flows. *Comm. Math. Phys.* **229** 57–71. MR1917674
[18] MELBOURNE, I. and TÖRÖK, A. (2004). Statistical limit theorems for suspension flows. *Israel J. Math.* **144** 191–209. MR2121540
[19] PARRY, W. and POLLICOTT, M. (1990). Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque* **187-188** 268. MR1085356
[20] PHILIPP, W. and STOUT, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* 2 **161** iv+140. MR0433597
[21] ROSENTHAL, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. MR0271721
[22] SAUSSOL, B. (2000). Absolutely continuous invariant measures for multidimensional expanding maps. *Israel J. Math.* **116** 223–248. MR1759406
[23] SERFLING, R. J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Statist.* **41** 1227–1234. MR0268938
[24] TSUJII, M. (2008). Quasi-compactness of transfer operators for contact Anosov flows. Preprint.
[25] ZAĬTSEV, A. Y. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM Probab. Stat.* **2** 41–108. MR1616527
[26] ZAĬTSEV, A. Y. (2007). Estimates for the rate of strong approximation in the multidimensional invariance principle. *J. Math. Sci.* (*N. Y.*) **145** 4856–4865. MR2355400

IRMAR
CNRS UMR 6625
UNIVERSITÉ DE RENNES 1
35042 RENNES
FRANCE
E-MAIL: sebastien.gouezel@univ-rennes1.fr