



HAL
open science

Almost sure invariance principle for dynamical systems by spectral methods

Sébastien Gouëzel

► **To cite this version:**

Sébastien Gouëzel. Almost sure invariance principle for dynamical systems by spectral methods. Annals of Probability, 2010, 38 (4), pp.1639-1671. 10.1214/10-AOP525 . hal-00402853v2

HAL Id: hal-00402853

<https://hal.science/hal-00402853v2>

Submitted on 13 Jan 2010 (v2), last revised 9 Feb 2011 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ALMOST SURE INVARIANCE PRINCIPLE FOR DYNAMICAL SYSTEMS BY SPECTRAL METHODS

SÉBASTIEN GOUËZEL

ABSTRACT. We prove the almost sure invariance principle for stationary \mathbb{R}^d -valued processes (with dimension-independent very precise error terms), solely under a strong assumption on 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). This kind of results has a lot of consequences, see e.g. [MN09] and references therein.

Such results are well known for one-dimensional processes, either independent or weakly dependent (see among many others [DP84, HK82]), and for independent higher dimensional processes [Ein89, Zai98]. However, for weakly dependent higher dimensional processes, difficulties arise since the techniques relying on Skorokhod representation theorem do not work efficiently. In this direction, an approximation argument introduced by [BP79] was recently generalized to a large class of weakly dependent sequences in [MN09]: their results give explicit error terms in the vector-valued almost sure invariance principle, and are applicable when the variables under study 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 study, thereby implying limit theorems. It is therefore desirable to develop an abstract argument, showing that enough control on the characteristic functions of a process implies the almost sure invariance principle, for vector-valued observables. This is our goal in this paper.

Date: January 12, 2010.

2000 Mathematics Subject Classification. 60F17, 37C30.

Key words and phrases. almost sure invariance principle, coupling, transfer operator.

[BP79, 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 paragraph, 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, distributed as $\mathcal{N}(0, \Sigma^2)$.
- (3) Almost surely in Ω ,

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

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 notations, 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 $\epsilon_0 > 0$ and $C, c > 0$ such that, for any $n, m > 0$, any $b_1 < b_2 < \dots < b_{n+m+1}$, any $k > 0$ and any $t_1, \dots, t_{n+m} \in \mathbb{R}^d$ with $|t_j| \leq \epsilon_0$, we have

$$(H) \quad \left| E \left(e^{i \sum_{j=1}^n t_j \left(\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell \right) + i \sum_{j=n+1}^{n+m} t_j \left(\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell \right)} \right) \right. \\ \left. - E \left(e^{i \sum_{j=1}^n t_j \left(\sum_{\ell=b_j}^{b_{j+1}-1} A_\ell \right)} \right) \cdot E \left(e^{i \sum_{j=n+1}^{n+m} t_j \left(\sum_{\ell=b_j+k}^{b_{j+1}+k-1} A_\ell \right)} \right) \right| \\ \leq C(1 + \max |b_{j+1} - b_j|)^{C(n+m)} e^{-ck}.$$

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, it is trivially satisfied for independent random variables. The interesting point 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 $\text{cov}(\sum_{\ell=0}^{n-1} A_\ell)/n$ 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*

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

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. [FMT03, MT04]).

If the process is not stationary, 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 $\alpha > 0$,*

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

uniformly in m, n . Then the sequence $\sum_{\ell=0}^{n-1} A_\ell/\sqrt{n}$ 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) (even more, this inequality holds with $\alpha = 0$).

Contrary to the results of [BP79], our results are dimension-independent for i.i.d. random variables (but they are not optimal in this case, see [Ein89, Zai98, Zai07]: 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 consider first a very simple example: let $T(x) = 2x \bmod 1$ on the circle $S^1 = \mathbb{R}/\mathbb{Z}$, and consider a Lipschitz function $f : S^1 \rightarrow \mathbb{R}^d$ of vanishing average for Lebesgue measure. We would like to prove an almost sure invariance principle for the process $(f(x), f(Tx), f(T^2x), \dots)$, 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, \dots, t_{n-1} in \mathbb{R}^d ,

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

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 study 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, \dots) . Let \mathcal{B} be a Banach space and let \mathcal{L}_t (for $t \in \mathbb{R}^d$, $|t| \leq \epsilon_0$) 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 $t_0, \dots, t_{n-1} \in \mathbb{R}^d$ with $|t_j| \leq \epsilon_0$,

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

In this case, we say that the characteristic function of (A_0, A_1, \dots) is coded by $(\mathcal{B}, (\mathcal{L}_t)_{|t| \leq \epsilon_0}, u_0, \xi_0)$.

We claim that the assumption (H) follows from suitable assumptions on the operators \mathcal{L}_t , that 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 $\|Q^n\|_{\mathcal{B} \rightarrow \mathcal{B}} \leq C\kappa^n$ for some $\kappa < 1$.
- (I2) There exists $C > 0$ such that $\|\mathcal{L}_t^n\|_{\mathcal{B} \rightarrow \mathcal{B}} \leq C$ 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. Then there 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, we have under the assumptions of Theorem 2.1

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

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, nor 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 explain how to check (I) in several practical situations. Let $T : X \rightarrow X$ be a dynamical system, let μ be a probability measure (invariant or not), and let $f : X \rightarrow \mathbb{R}^d$. 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), i.e., 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 *quasiconvex*).

Proposition 2.3. *If the family $\mathcal{L}_t : \mathcal{B} \rightarrow \mathcal{B}$ 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 $\|Q_t^n\| \leq C\kappa^n$ for some $\kappa < 1$, uniformly for small t . If $|\lambda(t)| \leq 1$ for small t , we obtain (I2).

For small t , we have

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

When $t \rightarrow 0$, 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 $|\lambda(t)| \leq 1$, concluding the proof. \square

Let us be more specific. Let T be an irreducible aperiodic subshift of finite type, let m be a Gibbs measure, and let $f : X \rightarrow \mathbb{R}^d$ be Hölder continuous with $\int f dm = 0$. Let \mathcal{L} be the transfer operator associated to 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 the Hölder functions form a Banach algebra). The condition (I) is checked e.g. in [GH88], [PP90]. Hence, Theorem 2.1 gives an almost sure invariance principle for any error exponent $> 1/4$. This result is not new, it is already given in [MN09], though with a weaker error term.

If T is an Anosov or Axiom A map and $f : X \rightarrow \mathbb{R}^d$ 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 [BT07, GL08].

Let now $T : X \rightarrow X$ be a piecewise expanding map, and assume that the expansion dominates the complexity (in the sense of [Sau00, 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 $f : X \rightarrow \mathbb{R}^d$ be β -Hölder continuous for some $\beta \in (0, 1]$. Then the perturbed transfer operator \mathcal{L}_t acts continuously on the Banach space $\mathcal{B} = V_\beta$ introduced in [Sau00], 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 $> 1/4$. This result was only known for $\dim(X) = 1$ and $d = 1$, thanks to [HK82].

This result also applies to coupled map lattices, since [BGK07] shows (I) for such maps. Let us mention that the Banach space \mathcal{B} here is not a Banach space of functions or distributions, but this is of no importance for our abstract setting.

Assume now that T is the time-one map of a contact Anosov flow. [Tsu08] 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 $> 1/4$. This result was known for real-valued observables [MT02], 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 [MT02].

Finally, assume that $T : X \rightarrow X$ is a mixing Gibbs-Markov map with invariant measure m , i.e., 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 subshift of finite type to infinite alphabets, see e.g. [MN09, Section 3.1] for precise definitions). For $f : X \rightarrow \mathbb{R}^d$ 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, it contains in particular all the weighted Lipschitz observables of [MN09, Section 3.2]).

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

This follows from [Gou08, Section 3.1], where a Banach space \mathcal{B} satisfying the assumptions of Proposition 2.3 is constructed.

Let us mention that the almost sure invariance principle is invariant under the process of *inducing*, i.e., going from a small dynamical system to a larger one. A lot

of 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 [MN09] 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, 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. [AW00]), this implies that $\langle t, f \rangle$ is an L^2 coboundary: 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 paragraph are not satisfied, while a weaker form of continuity holds. We describe such a setting in this paragraph.

Assume again that the characteristic function of a process $(f, f \circ T, f \circ T^2, \dots)$ 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), i.e., 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 $|t| \leq \epsilon_0$, we have $\|\mathcal{L}_t^n\|_{\mathcal{C} \rightarrow \mathcal{C}} \leq CM^n$.
- (2) For all $n \in \mathbb{N}$, all $|t| \leq \epsilon_0$ 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 $\|\mathcal{L}_t - \mathcal{L}_0\|_{\mathcal{B} \rightarrow \mathcal{C}}$ tends to 0 when $t \rightarrow 0$.

Then [KL99, Liv03] 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 $\tilde{\kappa} < 1$,

$$\|\Pi_t\|_{\mathcal{B} \rightarrow \mathcal{B}} \leq C, \quad \|Q_t^n\|_{\mathcal{B} \rightarrow \mathcal{B}} \leq C\tilde{\kappa}^n, \quad \|\Pi_t - \Pi\|_{\mathcal{B} \rightarrow \mathcal{C}} \rightarrow 0 \text{ when } t \rightarrow 0.$$

Therefore, (I2) follows from the arguments in the proof of Proposition 2.3 if we can prove that $\langle \xi_0, \Pi_t u_0 \rangle \rightarrow \langle \xi_0, \Pi u_0 \rangle$ when $t \rightarrow 0$. 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} . Let also $f : \mathcal{X} \rightarrow \mathbb{R}$ with $E_m(f) = 0$, we want to study the process $A_\ell = f(X_\ell)$.

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

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

By induction, we obtain

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

This is very similar to the coding property introduced in (2.2), the small 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 the integration against μ is continuous on \mathcal{B} . If the operators P_t satisfy on \mathcal{B} the condition (I), then the process $f(X_\ell)$ satisfies (H). If $f(X_\ell)$ is bounded in L^p for some $p > 2$, it follows that the process $(f(X_\ell))$ satisfies an almost sure invariant principle, for any error exponent $\lambda > p/(4p - 4)$.*

To check the condition (I), the arguments of Paragraph 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 study). We refer in particular the reader to the article [HP08], where several examples are studied, including uniformly ergodic chains, geometrically ergodic chains, and iterated random Lipschitz models. It is in particular shown in this article that the weak continuity arguments of Paragraph 2.3 are very powerful in some situations where the strong continuity of Paragraph 2.2 does not hold.

2.5. Proof of Theorem 2.1 assuming Theorem 1.3.

First step: there exists $u_1 \in \mathcal{B}$ such that, for $t_0, \dots, t_{n-1} \in B(0, \epsilon_0)$,

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

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

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

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 ∞ in the previous equality, we get

$$(2.7) \quad \langle \xi_0, u \rangle = \langle \xi_0, \Pi u \rangle = \langle \xi_0, u_2 \rangle \cdot \langle \xi_2, u \rangle.$$

Moreover,

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

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

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

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

Second step: (H) holds.

Consider $b_1 < \dots < b_{n+m+1}$, as well as $t_1, \dots, t_{n+m} \in B(0, \epsilon_0)$ and $k > 0$. Then

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

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

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

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

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

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

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

We have proved that a 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 $\kappa < 1$. Write $C = 2^{C'}$ and $\kappa = e^{-c}$ for some $c, C' > 0$, then this error is at most

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

This proves (H).

Third step: there exist $a \in \mathbb{R}^d$ and $C, \delta > 0$ such that $|E(A_\ell) - a| \leq C e^{-\delta \ell}$.

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 , belonging to L^p , whose characteristic

function is supported in $B(0, \epsilon_0)$ (see Proposition 3.8 for the existence of V). Let also $T > 0$. Then

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

The first term is bounded by $\|A_\ell + V\|_{L^2} \|1_{|A_\ell + V| \geq T}\|_{L^2} \leq CP(|A_\ell + V| > T)^{1/2} \leq C/T^{1/2}$. 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 $\{|t| \leq \epsilon_0\}$. 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 $C \|x1_{|x| < T}\|_{L^2} = CT^{3/2}$, we therefore obtain

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

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

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

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

Fourth step: 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 ,

$$(2.10) \quad |\text{cov}(A_\ell, A_{\ell+m}) - s_m| \leq Ce^{-\delta \ell}.$$

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 (2.10) for some sequence of matrices s_m . Then the series $\Sigma^2 = s_0 + \sum_{m=1}^{\infty} (s_m + s_m^*)$ converges in norm and, uniformly in m, n ,*

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

Let us admit this lemma for the moment. Then the process $(A_\ell - a)$ satisfies all the assumptions of Theorem 1.3. Theorem 2.1 follows from this theorem. \square

Proof of Lemma 2.7. Let us first prove that, for some $\delta > 0$,

$$(2.12) \quad |\text{cov}(A_\ell, A_{\ell+m})| \leq C e^{-\delta m}.$$

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

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

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

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

where $q > 1$ is chosen so that $1/p + 1/2 + 1/q = 1$. Moreover, $\|1_{|A_\ell + V|>T}\|_{L^q} = P(|A_\ell + V| > T)^{1/q} \leq CT^{-1/q}$. We have proved 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 $\{|t|, |u| \leq \epsilon_0\}$. Denoting by h_T the Fourier transform of the function $xy 1_{|x|, |y| \leq T}$, 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 $\{|t|, |u| \leq \epsilon_0\}$, 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 $\ell \rightarrow \infty$, $\text{cov}(A_\ell, A_{\ell+m})$ 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

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

We have

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

With (2.13), we get

$$\begin{aligned} \left| \text{cov} \left(\sum_{\ell=m}^{m+n-1} A_\ell \right) - \sum_{i=0}^{n-1} s_0 - \sum_{0 \leq i < j \leq n-1} (s_{j-i} + s_{j-i}^*) \right| \\ \leq C \sum_{i=0}^{n-1} e^{-\delta(i+m)} + C \sum_{0 \leq i < j \leq n-1} \min(e^{-\delta(i+m)}, e^{-\delta(j-i)}). \end{aligned}$$

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 proved that

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

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

3. PROBABILISTIC TOOLS

3.1. Coupling. As in [BP79], the notion of coupling is central to our argument. In this paragraph, we introduce this notion.

If $Z_1 : \Omega_1 \rightarrow E_1$ and $Z_2 : \Omega_2 \rightarrow E_2$ are two random variables on two (possibly different) probability spaces, 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 $Z'_1 : \Omega' \rightarrow E_1$ and $Z'_2 :$

$\Omega \rightarrow E_2$ 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 $\Omega = E_1 \times E_2$, with Z'_1 and Z'_2 the first and second projection.

The following lemma, also known as the Berkes-Philipp lemma, is Lemma A.1 of [BP79]. 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 , then 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 . Then there 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 a typical 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 $\sum Y_i$. Then there exists a coupling between (X_1, \dots, X_n) and (Y_1, \dots, Y_n) realizing 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 arguments will be used several times later on, 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 $Z_n : \Omega \rightarrow E_n$ and $Z'_n : \Omega' \rightarrow E_n$ ($n \geq 1$) 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*

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

for any $1 \leq n \leq N$. Then there 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 $(E_1 \times \dots \times E_N)^2$, with respective marginals 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 the second E_n factor. Let us extend arbitrarily this measure to a probability measure \tilde{P}_N on E^2 , where $E = E_1 \times E_2 \times \dots$. The sequence \tilde{P}_N is tight, and any of its weak limits satisfies the required property. \square

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 $\epsilon > 0$ such that $P(B) \leq \epsilon + Q(B^\epsilon)$ 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 Strassen–Dudley Theorem [Bil99, 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

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

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 $T' > 0$,*

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

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

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

Therefore, $\pi(F, G)$ is bounded by the right hand side of this equation. To conclude, we have to estimate $\int_{|x_1|, \dots, |x_N| \leq T'} |f - g|$. We have

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

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

3.3. Classical tools. Let us recall two classical results of probability theory that we will need later on. The first one is Rosenthal's inequality [Ros70], and the second one is a weak version of Gal-Koksma strong law of large numbers [PS75, Theorem A1] that 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*

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

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 ,*

$$(3.4) \quad E \left| \sum_{j=m}^{m+n-1} X_j \right|^2 \leq Cn^q.$$

Then, for any $\alpha > 0$, the sequence $\sum_{j=1}^N X_j / N^{q/2+\alpha}$ 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 $\{|t| \leq \epsilon_0\}$.*

Proof. We start from a C^∞ function ϕ supported in $\{|t| \leq \epsilon_0/2\}$, 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 $\tilde{\phi}(t) = \phi(-t)$. Let finally $h = g / \int g$, it is nonnegative, has integral 1, and its Fourier transform is proportional to $\phi \star \tilde{\phi}$, hence supported in $\{|t| \leq \epsilon_0\}$. Let W and W' be independent random variables with density h , then $V = W - W'$ satisfies the conclusion of the proposition. \square

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 $m, n \geq 0$,*

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

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 thanks to 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 $u_n = \max_{m \in \mathbb{N}} \left\| \sum_{\ell=m}^{m+n-1} A_\ell \right\|_{L^2}^2$. For any $\alpha > 0$, there exists $C > 0$ such that $u_{a+b} \leq u_a + u_b + C(1 + a^\alpha + b^\alpha)(1 + u_a^{1/2} + u_b^{1/2})$ for any $a, b \geq 1$.*

Proof. Let $m \in \mathbb{N}$, and $a \leq b$. Write $X_1 = \sum_{\ell=m}^{m+a-1} A_\ell$ and $X_2 = \sum_{\ell=m+a+\lfloor b^\alpha \rfloor}^{m+a+b-1} A_\ell$. Let also $\tilde{X}_1 = X_1 + V_1$ 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). Let finally \tilde{Y}_1 and \tilde{Y}_2 be independent random variables, distributed respectively like \tilde{X}_1 and \tilde{X}_2 .

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

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

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 at most ϵ_0 , $|\phi - \gamma| \leq C(1+b)^C e^{-cb^\alpha} \leq C e^{-c'b^\alpha}$. We have $\tilde{\phi} - \tilde{\gamma} = (\phi - \gamma)E(e^{i(t_1 V_1 + t_2 V_2)})$. Since the characteristic function of V is supported in $\{|t| \leq \epsilon_0\}$, this shows that the characteristic functions $\tilde{\phi}$ and $\tilde{\gamma}$ of $(\tilde{X}_1, \tilde{X}_2)$ and $(\tilde{Y}_1, \tilde{Y}_2)$ satisfy $|\tilde{\phi} - \tilde{\gamma}| \leq C e^{-c'b^\alpha}$, and are supported in $\{|t| \leq \epsilon_0\}$. 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 $E(|\tilde{X}_i|)/T' \leq Cb/e^{b^{\alpha/2}}$, 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 of a set O of measure at most $C e^{-b^\delta}$, we have $|\tilde{X}_i - \tilde{Y}_i| \leq C e^{-b^\delta}$. We have

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

The first term is bounded by $\|1_O\|_{L^q} \left\| \tilde{X}_1 + \tilde{X}_2 \right\|_{L^p}$, where q is chosen so that $1/p + 1/q = 1/2$. Hence, it is at most $C e^{-b^\delta/q} b \leq C$. The second term is bounded by $C e^{-b^\delta} \leq C$. 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

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

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

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

This proves the lemma. \square

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

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

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

Proof. For any $\epsilon > 0$ and any $x, y \geq 0$, we have $xy \leq \epsilon x^2 + \epsilon^{-1}y^2$. We therefore obtain from the assumption

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

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

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

In particular,

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

It follows inductively that $v_k/(2 + 2\epsilon)^k \leq v_0 + C\epsilon^{-1} \sum_j \frac{2^{2\alpha j}}{(2 + 2\epsilon)^{j+1}} < \infty$. Hence, for any $\epsilon > 0$, $v_k = O((2 + 2\epsilon)^k)$, i.e., for any $\rho > 1$, $u_n = O(n^\rho)$. Choosing ρ close enough to 1, we get from (4.3) $u_{a+b} \leq u_a + u_b + Ca^\beta + Cb^\beta$, 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, i.e., $u_n = O(n)$. \square

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 :

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

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 $b = \lfloor n^\alpha \rfloor$. Fix $m \in \mathbb{N}$, 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 $0 \leq j < b$, and the last 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 that, for some $\delta > 0$, for any $j \leq b$

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

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 $Cn^{Cn^\alpha} \cdot e^{-cn^{2\alpha}} \leq Ce^{-c'n^{2\alpha}}$ (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 over j the estimate in (4.5), we obtain

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

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

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

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

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

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

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

Finally,

$$\begin{aligned} \left\| \sum_{\ell=m}^{m+n-1} A_\ell \right\|_{L^{p-\eta}} &\leq \left\| \sum_{j=0}^{b-1} X_j \right\|_{L^{p-\eta}} + \sum_{j=0}^{b-1} \sum_{\ell \in I'_j} \|A_\ell\|_{L^{p-\eta}} + \left\| \sum_{\ell=m+ab}^{m+n-1} A_\ell \right\|_{L^{p-\eta}} \\ &\leq Cn^{1/2} + Cv_{a-b^2} b^{1/(p-\eta)} + Cn^{3\alpha} + v_{n-ab}. \end{aligned}$$

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

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

Moreover, $a \leq n^{1-\alpha}$, $b \leq n^\alpha$ and $n - ab \leq a + b + 1 \leq Cn^{1-\alpha}$. If $v_n = O(n^r)$, 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 $v_n = O(n^{1/2})$. \square

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} , 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 then we 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 go 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 requires 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 $\epsilon \in (0, 1-\beta)$. Let $f = f(n) = \lfloor \beta n \rfloor$. We decompose $[2^n, 2^{n+1})$ as a union of $F = 2^f$ intervals $(I_{n,j})_{0 \leq j < F}$ of the same length, and F gaps $(J_{n,j})_{0 \leq j < F}$ between them, used to give enough independence. Good intervals and gaps are laid 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 $2^{\lfloor \epsilon n \rfloor} 2^{f-1}$. It cuts the interval $[2^n, 2^{n+1})$ into two parts. The middle interval of each of these parts, i.e., $J_{n,F/4}$ and $J_{n,3F/4}$, have length $2^{\lfloor \epsilon n \rfloor} 2^{f-2}$. The middle intervals of the remaining four parts have length $2^{\lfloor \epsilon n \rfloor} 2^{f-3}$, and so on. More formally, for $1 \leq j < F$, 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 $\alpha_r(j) \neq 0$, then the length of $J_{n,j}$ is $2^{\lfloor \epsilon n \rfloor} 2^r$. 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 $|J_{n,0}| = 2^{\lfloor \epsilon n \rfloor} 2^f$, and say that this interval has rank f .

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

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

Let $|I_{n,j}| = 2^{n-f} - (f+2)2^{\lfloor \epsilon n \rfloor - 1}$, this is a positive integer if n is large enough, and $\sum |I_{n,j}| + \sum |J_{n,j}| = 2^n$, i.e., 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 $\{(n, j) \mid n \in \mathbb{N}, 0 \leq j < F(n)\}$. 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 $n_k \rightarrow \infty$.

Let $X_{n,j} = \sum_{\ell \in I_{n,j}} A_\ell$, for $n \in \mathbb{N}$ and $0 \leq j < F(n)$. Write finally $\mathcal{I} = \bigcup_{n,j} I_{n,j}$ 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 $(n, j) \rightarrow \infty$,

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

- (2) There exists a coupling between $(Y_{n,j})$ and a sequence of independent gaussian random variables $Z_{n,j}$, with $\text{cov}(Z_{n,j}) = |I_{n,j}| \Sigma^2$, such that, almost surely, when $(n, j) \rightarrow \infty$,

$$\left| \sum_{(n',j') \prec (n,j)} Y_{n',j'} - Z_{n',j'} \right| = o(2^{(\beta+\epsilon)n/2} + 2^{((1-\beta)/2+\beta/p+\epsilon)n}).$$

- (3) Coupling the $X_{n,j}$ with the $Z_{n,j}$ thanks to 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,

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

- (4) We check that, almost surely, when $(n, j) \rightarrow \infty$,

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

Moreover, a similar estimate holds for the B_ℓ s.

- (5) Combining the last two steps, we get when k tends to infinity

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

- (6) Finally, we prove that the gaps can be neglected: almost surely,

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

and a similar estimate holds for the B_ℓ s.

Altogether, this gives a coupling for which, almost surely

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

Let us choose β so that the two error terms are equal, i.e., $\beta = p/(2p-2)$. We obtain an almost sure invariance principle with error term $p/(4p-4) + \epsilon$, for any $\epsilon > 0$. 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 the steps (1), (2), (4) and (6), since the steps (3) and (5) are trivial. This is done in the next paragraphs.

5.1. Step (1): Coupling with independent random variables. In this paragraph, 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,*

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

The rest of this paragraph 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 from 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 $X_n = (X_{n,j})_{0 \leq j < F(n)}$ and $\tilde{X}_n = (\tilde{X}_{n,j})_{0 \leq j < F(n)}$. The proof of Proposition 5.1 has two parts: first, we make the different \tilde{X}_n independent from 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 distributed like \tilde{X}_n but independent of $(\tilde{X}_1, \dots, \tilde{X}_{n-1})$. We have*

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

Proof. The process (X_1, \dots, X_n) takes its values in \mathbb{R}^{dD} for $D = \sum_{m=1}^n F(m) \leq \sum_{m=1}^n 2^{\beta m} \leq C2^{\beta n}$. Moreover, each component in \mathbb{R}^d of this process is one of the $\tilde{X}_{n,j}$, 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 $C^{\pm 1}2^{\epsilon n + \beta n}$. Let ϕ and $\gamma : \mathbb{R}^{dD} \rightarrow \mathbb{C}$ denote respectively the characteristic functions of $(X_1, \dots, X_{n-1}, X_n)$ and $(X_1, \dots, X_{n-1}, Q_n)$, where Q_n is a copy of X_n , independent of (X_1, \dots, X_{n-1}) . The assumption (H) ensures that, for Fourier parameters $t_{n,j}$ all bounded by ϵ_0 , we have

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

if n is large enough.

Let $\tilde{\phi}$ and $\tilde{\gamma}$ be the characteristic functions of $(\tilde{X}_1, \dots, \tilde{X}_n)$ 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 $\{|t| \leq \epsilon_0\}$, we obtain in particular

$$(5.4) \quad |\tilde{\phi} - \tilde{\gamma}| \leq C e^{-c2^{\beta n + \epsilon n}}.$$

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

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

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

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 from each other. There 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) ,*

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

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 use again 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 $(\tilde{X}_1, \dots, \tilde{X}_n)$. Let also \tilde{R}_n be distributed like \tilde{X}_n , independent of everything else. For $1 \leq n \leq N$, Lemma 5.2 and 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 $|\tilde{R}_n^n - \tilde{R}_n| \leq C4^{-n}$, 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

$$\sum_{j=n+1}^N P(|\tilde{R}_n^j - \tilde{R}_n^{j-1}| > C4^{-j}) + P(|\tilde{R}_n^n - \tilde{R}_n| > C4^{-n}) \leq \sum_{j=n}^N C4^{-j} \leq C'4^{-n}. \quad \square$$

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

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

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.

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

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

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

Moreover, $\tilde{\mathcal{Y}}_{n,k}^i$ is made of 2^i sums of variables A_ℓ s 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 $C^{\pm 1}2^{\epsilon n+i}$ in the middle. Therefore, (H) ensures that the difference between the characteristic functions of $\tilde{\mathcal{Y}}_{n,k}^i$ and $(\tilde{\mathcal{Y}}_{n,2k}^{i-1}, \tilde{\mathcal{Y}}_{n,2k+1}^{i-1})$ is at most

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

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

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

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

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

This gives in particular (5.6). \square

Proof of Proposition 5.1. Let us put together the coupling constructed in Corollary 5.3 with the couplings constructed in Lemma 5.4 for each n , thanks to Strassen-Dudley Theorem 3.4. We get a coupling between $(\tilde{X}_{n,j})$ and $(\tilde{Y}_{n,j})$ such that

$P(|\tilde{X}_{n,j} - \tilde{Y}_{n,j}| \geq C4^{-n}) \leq C4^{-n}$. Since $\sum_{n,j} 4^{-n} < \infty$, Borel-Cantelli ensures that, almost surely,

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

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 $\alpha > 0$,

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

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

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

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

5.2. Step (2): Coupling with Gaussian random vectors. We will use Corollary 3 of [Zai07], let us recall it here for the convenience of the reader in a form that is convenient for us (it is obtained from the statement of Zaitsev 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 let $M = \left(\sum_{j=0}^{b-1} E|Y_j|^q \right)^{1/q}$. Assume that there exists a sequence $0 = m_0 < m_1 < \dots < m_s = b$ satisfying the following condition. Let $\zeta_k = Y_{m_k} + \dots + Y_{m_{k+1}-1}$ and let $B_k = \text{cov } \zeta_k$, we assume that*

$$(5.9) \quad 100M^2|v|^2 \leq \langle B_k v, v \rangle \leq 100CM^2|v|^2,$$

for all $v \in \mathbb{R}^d$ and all $0 \leq k < s$, and for some constant $C \geq 1$. Then there 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 $\text{cov } S_j = \text{cov } Y_j$, and moreover

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

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 $\text{cov } S_{n,j} = \text{cov } Y_{n,j}$, such that*

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

Proof. Let $q \in (2, p)$ and let $n \in \mathbb{N}$. 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 $\|Y_{n,j}\|_{L^q} \leq C 2^{(1-\beta)n/2}$. This implies that $M := \left(\sum_{j=0}^{F-1} E|Y_{n,j}|^q \right)^{1/q}$ satisfies

$$(5.12) \quad M \leq C 2^{\beta n/q} \cdot 2^{(1-\beta)n/2}.$$

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

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

Moreover, we assume that matrix Σ^2 is nondegenerate. Therefore, there exists a constant C_0 such that, for any $0 \leq m < m' \leq F$ with $m' - m$ large enough, and for any vector v ,

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

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

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

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

$$(5.14) \quad P \left(\max_{1 \leq i \leq F} \left| \sum_{j=0}^{i-1} Y_{n,j} - S_{n,j} \right| \geq 2^{\epsilon n/3} M \right) \leq C 2^{-q \epsilon n/3},$$

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

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

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

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, $\text{cov } S_{n,j} = |I_{n,j}|\Sigma^2 + o(2^{\alpha n})$. In particular, if n is large enough, we can write $|I_{n,j}|\Sigma^2 + 2^{\alpha n}I_d = \text{cov } Z_{n,j} + M_{n,j}$ where the matrix $M_{n,j}$ is positive definite and $|M_{n,j}| \leq C2^{\alpha n}$. 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 thanks to 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 with $\|D_{n,j}\|_{L^2} \leq C2^{\alpha n/2}$.

We claim that this coupling satisfies the conclusion of the lemma if $\alpha < \epsilon/2$. Indeed, by Etemadi's inequality [Bil99, Paragraph M19], we have for any n

$$\begin{aligned} P \left(\max_{1 \leq i \leq F(n)} \left| \sum_{j=0}^{i-1} D_{n,j} \right| > 2^{(\beta+\epsilon/2)n/2} \right) \\ \leq C \max_{1 \leq i \leq F(n)} P \left(\left| \sum_{j=0}^{i-1} D_{n,j} \right| > 2^{(\beta+\epsilon/2)n/2} / 3 \right) \\ \leq C \max_{1 \leq i \leq F(n)} E \left(\left| \sum_{j=0}^{i-1} D_{n,j} \right|^2 \right) / 2^{(\beta+\epsilon/2)n} \\ \leq C \sum_{j=0}^{F(n)-1} E(|D_{n,j}|^2) / 2^{(\beta+\epsilon/2)n} \leq C2^{\beta n} 2^{\alpha n} / 2^{(\beta+\epsilon/2)n}. \end{aligned}$$

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

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 $(n, j) \rightarrow \infty$,*

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

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. Denote $M_a^b = \max_{a \leq n \leq b} |\sum_{\ell=a}^{n-1} A_\ell|$. Corollary B1 in [Ser70] then also shows that

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

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

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

Moreover,

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

This sum is finite if q is close enough to p . Borel-Cantelli 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 paragraph, 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 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))

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

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

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

This is (5.2), as desired.

Proof of Lemma 5.9. We will freely use the convexity inequality

$$(5.17) \quad (a_1 + \cdots + a_k)^2 \leq k(a_1^2 + \cdots + a_k^2).$$

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 $E \left| \sum_{\ell \in J_0} A_\ell \right|^2 \leq C|J_0|$. A similar inequality holds for J_2 . To conclude the proof, it is therefore sufficient to prove that

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

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 $2^{\lfloor \epsilon N \rfloor + r}$ 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 $2^{\lfloor \epsilon n \rfloor + r}$. 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),

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

Let us fix some (n, r) , and let us 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 $\{1, \dots, t\}$,

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

Let us show how this concludes 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 $E \left| \sum_{\ell \in K} A_\ell \right|^2 \leq C|K|$. 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 . Let also $V_1, \dots, V_t, V'_1, \dots, V'_t$ 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 $C > 0$,

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

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

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

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

$$\left\| \sum_{s \in S} T_s \right\|_{L^2} \leq \left\| \sum_{s \in S} \tilde{T}_s \right\|_{L^2} + \left\| \sum_{s \in S} V_s \right\|_{L^2} \leq C + \left\| \sum_{s \in S} U_s \right\|_{L^2} + \left\| \sum_{s \in S} V_s \right\|_{L^2} + \left\| \sum_{s \in S} V'_s \right\|_{L^2}.$$

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

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 we 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 $\epsilon > 0$. 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,

$$(6.1) \quad \left| \sum_{\ell=0}^{i_n-1} A_\ell \right| = O(n^{1/2+\epsilon})$$

and

$$(6.2) \quad \max_{i \in I_n} \left| \sum_{\ell=i_n}^i A_\ell \right| = O(n^{\beta/2+1/p+\epsilon}).$$

Taking $\beta = (p-2)/p$ to equate the error terms, we get $|\sum_{\ell \leq k} A_\ell| = O(n^{1/2+\epsilon})$, where $n = n(k)$ is the index of the interval I_n containing k . Since $n \leq Ck^{1/(1+\beta)}$, we finally obtain an error term $O(k^{\lambda+2\lambda\epsilon})$ 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), $\left\| \sum_{\ell=0}^{i_n-1} A_\ell \right\|_{L^2} = O(n^{\alpha/2})$ for any $\alpha > 0$. Therefore,

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

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

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

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

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

Let (A_0, A_1, \dots) 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 directly follows from Lemma 2.7 and Theorem 1.3.

REFERENCES

- [AW00] Jon Aaronson and Benjamin Weiss, *Remarks on the tightness of cocycles*, Colloq. Math. **84/85** (2000), 363–376, Dedicated to the memory of Anzelm Iwanik. MR1784202. Cited page 7.
- [BGK07] Jean-Baptiste Bardet, Sébastien Gouëzel, and Gerhard Keller, *Limit theorems for coupled interval maps*, Stoch. Dyn. **7** (2007), no. 1, 17–36. MR2293068. Cited page 6.
- [Bil99] Patrick Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. MR1700749. Cited pages 14 and 26.

- [BP79] István Berkes and Walter Philipp, *Approximation theorems for independent and weakly dependent random vectors*, Ann. Probab. **7** (1979), no. 1, 29–54. MR515811. Cited pages 1, 2, 3, 12, and 13.
- [BT07] Viviane Baladi and Masato Tsujii, *Anisotropic Hölder and Sobolev spaces for hyperbolic diffeomorphisms*, Ann. Inst. Fourier (Grenoble) **57** (2007), no. 1, 127–154. MR2313087. Cited page 6.
- [DP84] Manfred Denker and Walter Philipp, *Approximation by Brownian motion for Gibbs measures and flows under a function*, Ergodic Theory Dynam. Systems **4** (1984), no. 4, 541–552. MR779712. Cited page 1.
- [Ein89] Uwe Einmahl, *Extensions of results of Komlós, Major, and Tusnády to the multivariate case*, J. Multivariate Anal. **28** (1989), no. 1, 20–68. MR996984. Cited pages 1 and 3.
- [FMT03] Michael Field, Ian Melbourne, and Andrew Török, *Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions*, Ergodic Theory Dynam. Systems **23** (2003), no. 1, 87–110. MR1971198. Cited page 3.
- [GH88] Yves Guivarc’h and Jean Hardy, *Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d’Anosov*, Ann. Inst. H. Poincaré Probab. Statist. **24** (1988), no. 1, 73–98. MR937957. Cited page 6.
- [GL08] Sébastien Gouëzel and Carlangelo Liverani, *Compact locally maximal hyperbolic sets for smooth maps: fine statistical properties*, J. Differential Geom. **79** (2008), no. 3, 433–477. MR2433929. Cited page 6.
- [Gou08] Sébastien Gouëzel, *Characterization of weak convergence of Birkhoff sums for Gibbs-Markov maps*, Preprint, 2008. Cited page 6.
- [HK82] Franz Hofbauer and Gerhard Keller, *Ergodic properties of invariant measures for piecewise monotonic transformations*, Math. Z. **180** (1982), no. 1, 119–140. MR656227. Cited pages 1 and 6.
- [HP08] Loïc Hervé and Françoise Pène, *Nagaev method via Keller-Liverani theorem*, Preprint, 2008. Cited page 8.
- [KL99] Gerhard Keller and Carlangelo Liverani, *Stability of the spectrum for transfer operators*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **28** (1999), no. 1, 141–152. MR1679080. Cited page 7.
- [Liv03] Carlangelo Liverani, *Invariant measures and their properties. A functional analytic point of view*, Dynamical systems. Part II, Pubbl. Cent. Ric. Mat. Ennio Giorgi, Scuola Norm. Sup., Pisa, 2003, pp. 185–237. MR2071241. Cited page 7.
- [MN09] Ian Melbourne and Matthew Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Ann. Probab. **37** (2009), 478–505. Cited pages 1, 6, and 7.
- [MT02] Ian Melbourne and Andrew Török, *Central limit theorems and invariance principles for time-one maps of hyperbolic flows*, Comm. Math. Phys. **229** (2002), no. 1, 57–71. MR1917674. Cited page 6.
- [MT04] Ian Melbourne and Andrei Török, *Statistical limit theorems for suspension flows*, Israel J. Math. **144** (2004), 191–209. MR2121540. Cited page 3.
- [PP90] William Parry and Mark Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque (1990), no. 187-188, 268. MR1085356. Cited page 6.
- [PS75] Walter Philipp and William Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Mem. Amer. Math. Soc. **2** (1975), no. issue 2, 161, iv+140. MR0433597. Cited page 14.
- [Ros70] Haskell P. Rosenthal, *On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables*, Israel J. Math. **8** (1970), 273–303. MR0271721. Cited page 14.

- [Sau00] Benoît Saussol, *Absolutely continuous invariant measures for multidimensional expanding maps*, Israel J. Math. **116** (2000), 223–248. MR1759406. Cited page 6.
- [Ser70] Robert J. Serfling, *Moment inequalities for the maximum cumulative sum*, Ann. Math. Statist. **41** (1970), 1227–1234. MR0268938. Cited page 27.
- [Tsu08] Masato Tsujii, *Quasi-compactness of transfer operators for contact Anosov flows*, Preprint, 2008. Cited page 6.
- [Zai98] Andrei Yu. Zaitsev, *Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments*, ESAIM Probab. Statist. **2** (1998), 41–108 (electronic). MR1616527. Cited pages 1 and 3.
- [Zai07] ———, *Estimates for the rate of strong approximation in the multidimensional invariance principle*, J. Math. Sci. (N. Y.) **145** (2007), 4856–4865. MR2355400. Cited pages 3 and 24.

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