AN ALMOST SURE INVARIANCE PRINCIPLE FOR SOME CLASSES OF NON-STATIONARY MIXING SEQUENCES

YEOR HAFOUTA

DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY

Abstract. In this note we (in particular) prove an almost sure invariance principle (ASIP) for non-stationary and uniformly bounded sequences of random variables which are exponentially fast \(\phi\)-mixing. The obtained rate is of order \(o(V_n^{1/2})\) for an arbitrary \(\delta > 0\), where \(V_n\) is the variance of the underlying partial sums \(S_n\). For certain classes of inhomogeneous Markov chains we also prove a vector-valued ASIP with similar rates.

1. Introduction

The central limit theorem (CLT) for partial sums \(S_n = \sum_{j=1}^n X_j\) of stationary real-valued random variables \(\{X_j\}\), exhibiting some type of “weak dependence”, is one of the main topics in probability theory, stating that \((S_n - \mathbb{E}[S_n]) / \sqrt{V_n}\), \(V_n = \text{Var}(S_n)\) converges in distribution towards a standard normal random variable. The almost sure invariance principle (ASIP) is a stronger result stating that there is a coupling between \(\{X_j\}\) and a standard Brownian motion \((W_t)_{t \geq 0}\) such that

\[ |S_n - \mathbb{E}[S_n] - W_{V_n}| = o(V_n^{1/2}), \text{ almost surely} \]

where \(W_{V_n}\) is the value of the Brownian motion at time \(t = V_n\). Both the CLT and the ASIP have corresponding versions for vector-valued sequences. The ASIP yields, for instance, the functional central limit theorem and the law of iterated logarithm (see [18]). While such results are well established for stationary sequences (see, for instance, [18], [1], [20], [19], [16] and [10] and references therein), in the non-stationary case much less is known, especially when the variance (or the covariance matrix) of \(S_n\) grows sub-linearly fast in \(n\). For instance, in [22] a vector-valued ASIP was obtained under conditions guaranteeing that the covariance matrix grows linearly fast. Similar results were obtained for random dynamical systems in [7] and [9], and the ASIP for elliptic Markov chains in random dynamical environment can be obtained similarly. For these models the variance (or the covariance matrix) of the underlying partial sums \(S_n\) grows linearly fast in \(n\) as well, while in [13] a real-valued ASIP was obtained for time-dependent hyperbolic dynamical systems under the assumption that \(\text{Var}(S_n)\) grows faster than \(n^{1/2}\).

In this paper we prove the ASIP for non-stationary, uniformly bounded, real or vector valued exponentially fast \(\phi\)-mixing sequences of random variables\(^1\). Under a certain assumption, which always holds true for real-valued sequences, we obtain the ASIP with rate \(o(s_n^{1/2+\delta})\) for an arbitrary \(\delta > 0\), where in the real-valued case \(s_n = V_n = \text{Var}(S_n)\), while in the vector-valued case\(^2\) \(s_n = \min_{|u| = 1} (\text{Cov}(S_n) u \cdot u)\). Then, in the vector-valued case, we will show that this assumption holds true for several classes of inhomogeneous contracting Markov chains.

\(^1\)We will also assume that \(\lim_{n \to \infty} \phi(n) < \frac{1}{2}\), were \(\phi(\cdot)\) are the, so-called, \(\phi\)-mixing coefficients, so the result holds true when \(\phi(n)\) decays exponentially fast.

\(^2\)Where \(|u|\) is the standard Euclidean norm of a vector and \(u \cdot v\) denotes the standard scalar product of two vectors, regardless of the underlying dimension.
The proof of the results relies on a recent modification of [10] Theorem 1.3, together with a block-partition argument, which in some sense reduces the problem to the case when the variance or the covariance matrix of $S_n$ grows linearly fast in $n$. More precisely, we show that there are “intervals” $I_j = \{a_j, a_j + 1, \ldots, b_j\}$ in the positive integers so that $a_1 = 1$ and $b_j + 1 = a_j$ (i.e. $\mathbb{N} = \bigcup_j I_j$) and the variance (covariance matrix) of each partial sum of the form $\sum_{j=1}^k \Xi_j$, $\Xi_j = \sum_{s \in I_j} X_s$ grows linearly fast in $k$. In this paper the sets $I_j$ will be referred to as “blocks”. Once the blocks $I_j$ are constructed the proof of the ASIP for $S_n$ has two steps: first, we prove the ASIP for the sequence $\tilde{S}_k = \sum_{j=1}^k \Xi_j$ using the modification of [10] Theorem 1.3 and then we approximate $S_n$ by $\tilde{S}_{k_n}$, where $k_n$ is the largest index so that $I_{k_n} \subset \{1, 2, \ldots, n\}$, and show that $k_n \approx s_n = \min_{|u|=1} (\text{Cov}(S_n) u \cdot u)$.

2. Preliminaries and main results

Let $X_1, X_2, \ldots$ be a sequence of zero-mean uniformly bounded $d$-dimensional random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $j \in \mathbb{N}$, let $\mathcal{F}_j$ denote the $\sigma$-algebra generated by $X_1, \ldots, X_j$ and let $\mathcal{F}_{j, \infty}$ denote the $\sigma$-algebra generated by $X_k$ for $k \geq j$. Recall that the $\alpha$ and $\phi$ mixing coefficients of the sequence are given by

\begin{equation}
\alpha(k) = \sup \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_j, B \in \mathcal{F}_{j+k, \infty}, j \in \mathbb{N}\}
\end{equation}

and

\begin{equation}
\phi(k) = \sup \{|\mathbb{P}(B|A) - \mathbb{P}(B)| : A \in \mathcal{F}_j, B \in \mathcal{F}_{j+k, \infty}, j \in \mathbb{N}, \mathbb{P}(A) > 0\}.
\end{equation}

Then both $\alpha(\cdot)$ and $\phi(\cdot)$ measure the long range dependence of the sequence $\{X_j\}$ in the sense that $X_j$’s are independent if and only if both sequences $\alpha(\cdot)$ and $\phi(\cdot)$ are identically zero.

We will assume here that there are constants $C > 0$, $\delta \in (0,1)$ and $n_0 \in \mathbb{N}$ so that

\begin{equation}
\alpha(n) \leq C \delta^n, \text{ for all } n \in \mathbb{N}
\end{equation}

and

\begin{equation}
\phi(n) < \frac{1}{2}.
\end{equation}

These are the mixing (weak-dependence) assumptions discussed in Section 11.

2.1. Remark. It is clear from the definitions of $\alpha(k)$ and $\phi(k)$ that $\alpha(k) \leq \phi(k)$. Hence, both conditions (2.3) and (2.4) are in force when $\phi(n) \leq C \delta^n$ for some $C > 0$ and $\delta \in (0,1)$. Note also that for Markov chains, condition (2.4) already implies that $\phi(n)$ decays exponentially fast to 0, and so in this case (2.4) implies (2.3). In any case, all the result in this paper are new even when $\phi(n)$ decays exponentially fast.\footnote{In fact, this was the main mixing assumption in a previous version of this paper https://arxiv.org/abs/2005.02915v3}

Next, for each $n \in \mathbb{N}$ set

$$S_n = \sum_{k=1}^n X_k$$

and put $V_n = \text{Cov}(S_n)$ (which is a $d \times d$ matrix). For all $n, m \in \mathbb{N}$ so that $n \leq m$ set

$$S_{n,m} = \sum_{j=n}^m X_j, \quad V_{n,m} = \text{Cov}(S_{n,m}), \quad s_n = \min_{|u|=1} (V_n u \cdot u)$$

where $|u|$ denotes the Euclidean norm of a vector $u \in \mathbb{R}^d$ and $u \cdot v$ denotes the standard scalar product of two vectors $u, v \in \mathbb{R}^d$. Then in the scalar case $d = 1$ we have $s_n = V_n = \text{Var}(S_n)$.

Next, for a random variable $Z : \Omega \to \mathbb{R}^d$ and a number $p \in [1, \infty)$ let us denote $\|Z\|_{L^p} = (\int |Z(\omega)|^p d\mathbb{P}(\omega))^{1/p}$. We consider here the following condition.
2.2. Assumption. There are constants $C_1, C_2 \geq 1$ with the following property: for every pair of positive integers $n$ and $m$ so that $n \leq m$ and $\|S_{n,m}\|_{L^2} \geq C_1$ we have

$$\max_{|u|=1}(V_{n,m}u \cdot u) \leq C_2 \min_{|u|=1}(V_{n,m}u \cdot u).$$

This assumption trivially holds true for real-valued sequences, and in Section 5 we will verify it for certain classes of additive vector-valued functionals $X_j = f_j(\xi_j)$ of inhomogeneous “sufficiently contracting” Markov chains $\{\xi_j\}$. Note also that

$$V_{n,m}u \cdot u = \text{Var}(S_{n,m} \cdot u)$$

and so Assumption 2.2 gives us a certain type of uniform control over these variances.

Our main result here is the following:

2.3. Theorem. Under Assumption 2.2 we have the following. Suppose that (2.3) and (2.4) hold true and that $\lim_{n \to \infty} s_n = \infty$. Then for every $\varepsilon > 0$ there is a coupling between $X_1, X_2, \ldots$ and a sequence of independent zero-mean Gaussian random vectors $Z_1, Z_2, \ldots$ so that

$$\left| S_n - \sum_{j=1}^n Z_j \right| = o(s_n^{1/4+\varepsilon}), \text{ almost surely.}$$

Moreover, there is a constant $C = C_\varepsilon > 0$ so that for all $n \geq 1$ and a unit vector $u \in \mathbb{R}^d$,

$$\|S_n \cdot u\|_{L^2}^2 - C s_n^{1/2+\varepsilon} \leq \left\| \sum_{j=1}^n Z_j \cdot u \right\|_{L^2}^2 \leq \|S_n \cdot u\|_{L^2}^2 + C s_n^{1/2+\varepsilon}.$$  

2.4. Remark.

(i) In the scalar case $d = 1$, (2.6) yields that the difference between the variances is $O(V_n^{\frac{1}{2}+\delta})$. Thus, using (2.6) together with [12, Theorem 3.2 A], we conclude that in the scalar case, for every $\varepsilon > 0$ there is a coupling of $\{X_n\}$ with a standard Brownian motion $\{W_t : t \geq 0\}$ so that

$$\sum_{j=1}^n X_j - W_{V_n} = o(V_n^{\frac{1}{4}+\varepsilon}), \text{ a.s.}$$

A corresponding result in the vector-valued case seems less plausible because in the non-stationary setup the structure of the covariance matrix $V_n$ does not stabilize as $n \to \infty$, which makes it less likely that we can approximate $S_n$ by a single Gaussian process like a standard $d$-dimensional Brownian motion.

(ii) For stationary sequences $\{X_n\}$, it was shown in [20, Theorem 1.4] that if $\phi(n) \ll \ln^{-r} n$ and $E[|X_n|^{2+\delta}] < \infty$ for some $\delta > 0$ and $r > (2 + \delta)/(2 + 2\delta)$, then there is a coupling of $\{X_n\}$ with a standard Brownian motion so that the left hand side of (2.5) is of order $o(V_n^{1/2} \ln^{-\theta} V_n)$ for an arbitrary $0 < \theta < (r(1+\delta))/(2(2+2\delta)) - \frac{1}{4}$. In comparison with [20], we get better ASIP rates in the non-stationary case, but only for uniformly bounded exponentially fast $\alpha$-mixing sequences such that $\lim_{n \to \infty} \phi(n) < \frac{1}{2}$. (iii) We would like to stress that even in the scalar case $d = 1$ no growth rates on the variance (such as $V_n \geq n^s$) are required in Theorem 2.3. This is in contrast, for instance, with [13] where it was assumed that $V_n \geq n^{\frac{3}{2}+\delta}$, and [10] and [22] where a linear growth was assumed. Note that in the latter papers vector-valued variables were considered.

(iv) Many papers about the ASIP rely on martingale approximation (e.g. [13] and [22]). However, to the best of our knowledge, the best rate in the vector-valued case that can be achieved using martingales (in the stationary case) is $o(n^{1/3}(\log n)^{1+\varepsilon}) = o(s_n^{1/3}(\log s_n)^{1+\varepsilon})$ (see [4]), and

\footnote{However, $s_n$ can still grow arbitrarily slow.}
so an attempt to use existing results for martingales seems to yield weaker rates than the ones obtained in Theorem 2.3

3. A linearization of the growth rate of the covariance matrix

The main step in the proof of Theorem 2.3 is to make a certain reduction to the case when $s_n = \min_{|u|=1} (V_n u \cdot u)$ grows linearly fast in $n$. This is the content of the following result.

3.1. Proposition. Suppose that $\sum_{m=1}^{\infty} (\alpha(m))^{-2/p} < \infty$ for some $p > 2$ and that $\lim_{n \to \infty} s_n = \infty$. Then there are constants $A_1, A_2 > 0$ and disjoint sets $I_j = \{a_j, a_j + 1, \ldots, b_j\} \subset \mathbb{N}$ whose union cover $\mathbb{N}$ (so that $a_1 = 1$ and $a_{j+1} = b_j + 1$ for all $j$) and for all $j \in \mathbb{N}$ and a unit vector $u$ we have

$$A_1 \leq \left\| \sum_{k \in I_j} X_k \cdot u \right\|_{L^2} \leq \max_{m \in I_j} \left\| \sum_{k=a_j}^{m} X_k \cdot u \right\|_{L^2} \leq A_2. \tag{3.1}$$

and so

$$\sup_{j \in \mathbb{N}} \max_{m \in I_j} \left\| \sum_{k=a_j}^{m} X_k \right\|_{L^2} \leq A_2. \tag{3.2}$$

Moreover, let $k_n = \max\{k : b_k \leq n\}$ and set $Z_j = \sum_{k \in I_j} X_k$. Then the following statement hold true.

(i) There are constants $R_1, R_2 > 0$ so that for every $n$ large enough and all unit vectors $u$,

$$R_1 k_n \leq \text{Var}(S_n \cdot u) = \text{Cov}(S_n \cdot u) \cdot u \leq R_2 k_n. \tag{3.3}$$

(ii) If also (2.4) is valid, then for every $\varepsilon > 0$ we have

$$\left| S_n - \sum_{j=1}^{k_n} Z_j \right| = o(s_n^\varepsilon), \ \mathbb{P} - a.s. \tag{3.4}$$

Proof of Proposition 3.1

First, let us fix some unit vector $u_0$, and set $\xi_j = X_j \cdot u_0$. For every finite $M \subset \mathbb{N}$ set

$$S(M) = \sum_{j \in M} X_j \cdot u_0 = \sum_{j \in M} \xi_j.$$

Next, let $A > 1$ and $r \in \mathbb{N}$ be sufficiently large constants which are yet to be determined. Let us construct a sequence $M_j$, $j \in \mathbb{N}$ of intervals (blocks) in the positive integers as follows. Let $p_1$ be the first index $p$ so that $\| \sum_{j=1}^{p} \xi_j \|_{L^2} \geq \sqrt{A}$ and set $M_1 = \{1, 2, \ldots, p_1\}$. Next, given that $M_j = \{q_j, q_j + 1, \ldots, p_j\}$ was constructed we define $q_{j+1} = p_j + r$ and $M_{j+1} = \{q_j, q_j + 1, \ldots, p_{j+1}\}$, where $p_{j+1}$ is the first index $p \geq q_{j+1}$ so that $\| S(\{q_{j+1}, \ldots, p\}) \|_{L^2} \geq \sqrt{A}$. Then the blocks $M_j = \{q_j, q_j + 1, \ldots, p_j\}$ satisfy the following properties:

1. $M_1$ contains $1$ and for each $j$ the block $M_j$ is to the left of $M_{j+1}$, and $\min M_{j+1} - \max M_j = r$;

2. For each $j$ we have $\sqrt{A} \leq \| S(M_j) \|_{L^2} \leq \sqrt{A} + L$, $L = \sup_{n}(\text{ess-sup}|X_n|)$ and

$$\max_{s \in M_j, s < p_j} \| S(\{q_j, q_j + 1, \ldots, s\}) \|_{L^2} < \sqrt{A} \leq \| S(M_j) \|_{L^2}. \tag{3.5}$$

Next, let us define $I_j = M_j + \{0, 1, \ldots, r-1\}$. Then the block $I_j$ is to the left of $I_{j+1}$ and the union of the $I_j$’s cover $\mathbb{N}$. Thus we can write $I_j = \{a_j, a_j + 1, \ldots, b_j\}$ with $a_{j+1} = b_j + 1$ and $a_1 = 1$.

We will break down the rest of the proof of Proposition 3.1 into a few steps. Between the steps we will introduce appropriate restrictions on $r$ and $A$, and the sets $I_j$ corresponding to appropriate choices of $r$ and $A$ will satisfy all the properties described in Proposition 3.1.

\footnote{Note that this series converges when (2.3) holds true.}
The first result we need is the following:

3.2. Lemma. For every \( p > 2 \) there is a constant \( C_p \geq 1 \) which does not depend on \( A \) or \( r \) so that for every \( 1 \leq i < j \) we have

\[
|\text{Cov}(S(M_i), S(M_j))| \leq C_p \left\| S(M_i) \right\|_{L^2} \left\| S(M_j) \right\|_{L^2} (\alpha(r(j - i)))^{1-2/p}.
\]

Proof. By applying [11, Corollary A.2] we get that

\[
|\text{Cov}(S(M_i), S(M_j))| \leq 8 \left\| S(M_i) \right\|_{L^p} \left\| S(M_j) \right\|_{L^p} (\alpha(r(j - i)))^{1-2/p}.
\]

On the other hand, since (2.4) holds, by applying [15, Theorem 6.17], taking into account that \( X_j \) are uniformly bounded and using (3.5) we get that

\[
\|S(M_i)\|_{L_p} \leq A_p (1 + \|S(M_i)\|_{L^2})
\]

where \( A_p \geq 1 \) is a constant that depends only on \( p, n_0 \) from (2.4) and \( \varepsilon = \frac{1}{2} - \phi(n_0) \). Now the proof is completed by recalling that \( \|S(M_j)\|_{L^2} \geq \sqrt{A} \geq 1 \) (and so we can take \( C_p = 32A_p \)). \( \square \)

Next, let \( p \) be as in Proposition 3.1. Since \( \sum_{m=1}^{\infty} (\alpha(m))^{1-2/p} < \infty \) there exists \( r_0 \in \mathbb{N} \) so that\(^6\)

\[
4C_p \sum_{m=1}^{\infty} (\alpha(r_0m))^{1-2/p} \leq 1
\]

where \( C_p \) is the constant from Lemma 3.2. Henceforth we will set \( r = r_0 \).

The second result we need is as follows.

3.3. Lemma. If the sets \( \{M_j\} \) are constructed with \( r = r_0 \) so that (3.8) holds true, then for every \( k \in \mathbb{N} \) we have

\[
\frac{1}{2} \sum_{i=1}^{k} \text{Var}(S(M_i)) \leq \text{Var}(S(M_1 \cup M_2 \cup \cdots \cup M_k)) \leq \frac{3}{2} \sum_{i=1}^{k} \text{Var}(S(M_i)).
\]

Proof. First,

\[
\text{Var}(S(M_1 \cup M_2 \cup \cdots \cup M_k)) = \sum_{i=1}^{k} \|S(M_i)\|_{L^2}^2 + 2 \sum_{1 \leq i < j \leq k} \text{Cov}(S(M_i), S(M_j)).
\]

Next, set \( \gamma(k) = (\alpha(k))^{1-2/p} \). Then by (3.6),

\[
2 \sum_{1 \leq i < j \leq k} |\text{Cov}(S(M_i), S(M_j))| \leq 2C_p \sum_{1 \leq i < j \leq k} \gamma(r(j - i))\|S(M_i)\|_{L^2}\|S(M_j)\|_{L^2}
\]

\[
\leq C_p \sum_{1 \leq i < j \leq k} \gamma(r(j - i))\|S(M_i)\|_{L^2}^2 + \|S(M_j)\|_{L^2}^2 = C_p \sum_{j=2}^{k} \|S(M_j)\|_{L^2}^2 \sum_{i=1}^{j-1} \gamma(r(j - i)) +
\]

\[
C_p \sum_{i=1}^{k} \|S(M_i)\|_{L^2}^2 \sum_{j=i+1}^{k} \gamma(r(j - i)) \leq \left( 2C_p \sum_{m \geq 1} \gamma(r_m) \right) \sum_{j=1}^{k} \|S(M_j)\|_{L^2}^2.
\]

The proof is completed using that \( 2C_p \sum_{m \geq 1} \gamma(r_m) \leq \frac{1}{2} \).

Next, let \( r_0 \) satisfy (3.9) and set \( Q_0 = 2C_p r_0 d^2 L^2 \sum_{m \geq 1} (\alpha(m))^{1-2/p} + (r_0 d L)^2 \), where \( d \) is the dimension of the random vectors \( X_j \). For each \( A \) set

\[
Q(A) = Q(A, r_0, p, L) = Q_0 + 2 \sqrt{3AQ_0}.
\]

\(^6\)Indeed \( \sum_{m=1}^{\infty} (\alpha(r_m))^{1-2/p} \leq \sum_{m=r}^{\infty} (\alpha(m))^{1-2/p} \rightarrow 0 \) as \( r \rightarrow \infty \).
Then \(Q(A)/A \to 0\) as \(A \to \infty\). Let \(A_0 > 1\) be so that for all \(A \geq A_0\) we have
\[
\sqrt{A} \geq 2r_0 dL, \ A \geq 4Q(A) \quad \text{and} \quad (\sqrt{A} + L)^2 \leq 2A.
\]
Note that the second restriction on \(A\) guarantees that \(A \leq \text{Var}(S(M_j)) \leq 2A\) for each \(j\).

The last auxiliary result we need before completing the proof of Proposition \ref{thm:3.1} is as follows.

3.4. **Lemma.** Suppose that the sets \(M_j\) are constructed with \(r = r_0\) so that \((3.9)\) holds true and with \(A \geq A_0\). Fix some \(k \in \mathbb{N}\) and set \(A_1 = M_1 \cup M_2 \cup \cdots \cup M_k\) and \(A_2 = I_1 \cup I_2 \cup \cdots \cup I_k\). Then,
\[
(3.11) \quad \left| \frac{\text{Var}(S(A_2))}{\text{Var}(S(A_1))} - 1 \right| \leq \frac{2Q(A)}{A} \leq \frac{1}{2}.
\]

**Proof.** Let \(X = S(A_1)\) and \(Y = S(A_2) - X\). Then
\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)
\]
and so by the Cauchy-Schwarz inequality,
\[
(3.12) \quad |\text{Var}(X + Y) - \text{Var}(X)| \leq \text{Var}(Y) + 2(\text{Var}(X)\text{Var}(Y))^{1/2}.
\]

Now, by Lemma \ref{lem:3.3}
\[
(3.13) \quad \frac{A}{2}k \leq \sum_{j=1}^{k} \text{Var}(S(M_j)) \leq \text{Var}(X) \leq \sum_{j=1}^{k} \frac{2}{3} \text{Var}(S(M_j)) \leq 3Ak
\]
where we have used that \(A \leq \text{Var}(S(M_j)) \leq 2A\). On the other hand, let \(D_j = I_j \setminus M_j\). Then \(Y = \sum_{j=1}^{k} S(D_j)\) and so
\[
\text{Var}(Y) = \text{Cov}(Y, Y) \leq \sum_{j=1}^{k} |\text{Cov}(S(D_j), Y)|.
\]

Now, fix some \(j\) and write \(D_j = \{d_j + 1, \ldots, d_j + r - 1\}\). Then
\[
|\text{Cov}(S(D_j), Y)| \leq \sum_{m \leq d_j} |\text{Cov}(S(D_j), X_m)| + \sum_{m \geq d_j + r} |\text{Cov}(S(D_j), X_m)| + \text{Var}(S(D_j)).
\]

Next, by applying \cite{11} Corollary A.2 and using \ref{lem:3.8} we see that if \(m \notin D_j\) then
\[
|\text{Cov}(S(D_j), X_m)| \leq C_p \|S(D_j)\|_{L^p} \|X_m\|_{L^p} (\alpha(\rho_{m,j}))^{1 - 2/p}, \quad \rho_{m,j} = \min_{s \in D_j} |m - s|.
\]

Using also that \(\|S(D_j)\|_{L^p} \leq rdL\) and \(\|X_m\|_{L^p} \leq dL\) for every \(p > 1\) we see that
\[
|\text{Cov}(S(D_j), Y)| \leq 2C_p (rdL)(dL) \sum_{m \geq 1} (\alpha(m))^{1 - 2/p} + (rdL)^2 = Q_0.
\]
Thus,
\[
\text{Var}(Y) \leq Q_0 k.
\]

Finally, using \ref{lem:3.11} and \ref{lem:3.9} we conclude that
\[
|\text{Var}(X + Y) - \text{Var}(X)| \leq (Q_0 + 2\sqrt{3A Q_0}) k = Q(A) k.
\]

The proof is completed by dividing the above left hand side by \(\text{Var}(X)\) and using \ref{lem:3.13}. \(\Box\)

**Completion of the proof of Proposition \ref{thm:3.1}**. Let us construct the blocks \(\{I_j\}\) with constants \(A \geq A_0\) and \(r = r_0\) with the same restrictions described before. First, since \(\sqrt{A} \geq 2r_0 dL\), using the second property of \(M_j\) and that \(I_j \setminus M_j\) is of cardinality \(r_0 - 1\) we obtain \ref{thm:3.1} with the specific unit vector \(u = u_0\) and the constants \(A_1 = \frac{1}{2} \sqrt{A}\) and \(A_2 = \frac{3}{4} \sqrt{A}\). By using Assumption \ref{assump:2.2} we see that if \(A\) is large enough then \ref{thm:3.1} holds true all unit vectors \(u_i\), possibly with different constants. The estimate \ref{thm:3.2} follows by taking the supremum over all unit vectors \(u\) in the third inequality from the left in \ref{thm:3.1}. Next, by applying Lemmas \ref{lem:3.3} and \ref{thm:3.1} we see that \ref{lem:3.3} holds true with
the specific unit vector $u = u_0$. Thus, by Assumption 2.2, if $A$ is large enough then (3.3) holds for an arbitrary unit vector (possibly with different constants).

In order to prove (3.3), let us assume (2.4). For each $q \geq 1$ set

$$D_q := \max_{b_q \leq n \leq b_{q+1}} |S_n - S_{b_q}| = \max_{m \in I_{q+1}} \left| \sum_{j=a_q+1}^{m} X_j \right|$$

where in the second inequality we used that $b_q + 1 = a_{q+1}$. Then with $\Xi_j = \sum_{k \in I_j} X_k$ and $k_n = \max\{k : b_k \leq n\}$ we have

$$(3.14) \quad \left| S_n - \sum_{j=1}^{k_n} \Xi_j \right| \leq D_{k_n}.$$ 

By applying [15, Theorem 6.17] with the random variables $\{X_n : n \in I_{q+1}\}$ (which is possible due to (2.4)) we see that for every $p > 2$ there are constants $c_p$ and $R_p$ so that for all $q \in \mathbb{N}$ we have

$$\|D_q\|_{L^p} \leq R_p \left( \|\max\{|X_n| : n \in I_{q+1}\}\|_{L^p} + \max\{|S_n - S_{b_q}| : n \in I_{q+1}\} \right) \leq c_p$$

where in the second inequality we have used that $\sup_n (\text{ess-sup} |X_n|) < \infty$ and (3.2). Thus, by applying the Markov inequality we see that for every $\varepsilon > 0$ and $p > 2$ we have

$$P(\|D_q\| \geq q^\varepsilon) = P(\|D_q\|^p \geq q^{p\varepsilon}) \leq c_p q^{-p\varepsilon}.$$

Taking $p > 1/\varepsilon$ we get from the Borel-Cantelli lemma that

$$(3.15) \quad \|D_q\| = O(q^\varepsilon), \text{ a.s.}$$

The desired estimate (3.4) follows by plugging in $q = k_n$ in (3.15) and using (3.14) and (3.3). □

4. ASP: proof Theorem 2.3

The proof of Theorem 2.3 is based on an application of [10, Theorem 2.1] with an arbitrary $p > 4$. The latter theorem is a modification of [10, Theorem 1.3] suited for more general non-stationary sequences of random vectors. The standing assumption in both theorems can be described as follows. Let $(A_1, A_2, \ldots)$ be an $\mathbb{R}^d$-valued process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists $\varepsilon_0 > 0$ and $C, c > 0$ such that for all $n, m \in \mathbb{N}$, $a_1 < a_2 < \ldots < a_{n+m+k}$, $k \in \mathbb{N}$ and $t_1, \ldots, t_{n+m} \in \mathbb{R}^d$ with $|t_j| \leq \varepsilon_0$, we have that

$$(4.1) \quad \mathbb{E} \left( e^{i \sum_{j=1}^n t_j \cdot (\sum_{\ell=a_j}^{a_{j+1}-1} A_\ell)} \right) \leq C(1 + \max\{|a_{j+1} - a_j|\})^{C(n+m)} e^{-ck}.$$

The first part of the proof is to show that $A_j = \Xi_j = \sum_{k \in I_j} X_k$ satisfies (4.1), which follows directly from the exponential $\alpha$-mixing rates (2.3). Next, let us verify the rest of the conditions of [10, Theorem 2.1]. Set

$$A_n = \sum_{j=1}^n A_j.$$ 

Then, by applying (3.3) with $b_n$ instead of $n$ we see that for all $n$ large enough we have

$$\min_{|u| = 1} (\text{Cov}(A_n) u \cdot u) \geq Cn$$

where $C > 0$ is a constant. This shows that the first additional condition in [10, Theorem 2.1] is satisfied. To show that $A_j$ are uniformly bounded in $L^p$, combining our assumption (2.4) with [15, Theorem 6.17] and taking into account (3.2), we see that for every $p > 2$,

$$(4.2) \quad B_p := \sup_j \|A_j\|_{L^p} < \infty.$$
The last condition we need to verify is that
\begin{equation}
(4.3) \quad |\text{Cov}(A_n \cdot u, A_{n+k} \cdot u)| \leq C_0 \eta^k
\end{equation}
for some $C_0 > 0$, $\eta \in (0, 1)$, all $k, n \in \mathbb{N}$ and all unit vectors $u \in \mathbb{R}^d$. To establish that, let us fix some $p > 2$. Then by [11 Corollary A.2] we have
\[ |\text{Cov}(A_n \cdot u, A_{n+k} \cdot u)| \leq \|A_n \cdot u\|_{L^p} \|A_{n+k} \cdot u\|_{L^p} (\alpha(k))^{1-2/p} \]
and so by (2.3) and (4.2) we see that (4.3) holds true with $C_0 = B_p^2 C^{1-2/p}$ and $\eta = \delta^{1-2/p}$ (where $C$ and $\delta$ come from (2.3)).

Next, by applying [9 Theorem 2.1] with the sequence $A_j = \Xi_j = \sum_{k \in I_j} X_k$ we conclude that there is a coupling between the sequence $A_1, A_2, ...$ and a sequence $Z_1, Z_2, ...$ of independent centered Gaussian random vectors so that for every $\varepsilon > 0$,
\begin{equation}
(4.4) \quad \left| \sum_{i=1}^k A_i - \sum_{j=1}^k Z_j \right| = o(k^{\delta + \varepsilon}), \text{ a.s.}
\end{equation}
and all the properties specified in Theorem 2.3 hold true for the new sequence $A_j = \Xi_j$. Now Theorem 2.3 follows by plugging in $k = k_n$ in (4.4), using (3.3), and then approximating $S_n$ by $A_{k_n} = \sum_{j=1}^{k_n} \Xi_j$, relying on (3.3) and using the, so-called, Berkes-Philipp lemma (which allows us to further couple $(X_j)$ with the Gaussian sequence).

5. Verification of the additional conditions in the non-scalar case: Markov chains

Assumption 2.2 trivially holds true for real-valued random variables $X_j$. In this section we discuss natural sufficient conditions for Assumption 2.2 for certain additive functionals of contracting Markov chains.

Dobrushin’s contracting chains. Let us recall the definition of Dobrushin’s contraction coefficients $\pi(\cdot)$ (see [6]). If $Q(x, \cdot)$ is a regular family of Markov transition operators between two spaces $\mathcal{X}$ and $\mathcal{Y}$, then
\[ \pi(Q) = \sup\{|Q(x_1, E) - Q(x_2, E)| : x_1, x_2 \in \mathcal{X}, E \in \mathcal{B}(\mathcal{Y})\} \]
where $\mathcal{B}(\mathcal{Y})$ is the underlying $\sigma$-algebra on $\mathcal{Y}$.

Let $\{\xi_j\}$ be a Markov chain with corresponding state spaces $\mathcal{X}_j$. Let $Q_j(x, \Gamma) = \mathbb{P}(\xi_{j+1} \in \Gamma | \xi_j = x)$ and suppose that
\begin{equation}
(5.1) \quad \delta := \sup_j \pi(Q_j) < 1.
\end{equation}

Then, as proven in [21], the chain $\{\xi_j\}$ is exponentially fast $\phi$-mixing. Let us take a sequence $f_j$ of bounded measurable functions on $\mathcal{X}_j$ and set $X_j = f_j(\xi_j) - \mathbb{E}[f_j(\xi_j)]$. Then by the results\footnote{In [21] only the lower bound was derived, however in this setup the upper bound is easier to obtain.} in [21] (see also [17 Proposition 13]), there are positive constants $A = A_\delta$ and $B = B_\delta$ so that for every $n, m$ with $n \leq m$ and each unit vector $u$,
\[ A \sum_{j=n}^m \text{Var}(X_j \cdot u) \leq \text{Var}(S_{n,m} \cdot u) \leq B \sum_{j=n}^m \text{Var}(X_j \cdot u). \]

We thus get the following result.

5.1. Proposition. Assumption 2.2 (and hence Theorem 2.3) holds true if $\delta < 1$ and there is a constant $C \geq 1$ so that for every $j \in \mathbb{N}$ we have
\[ \max_{|u| = 1} (\text{Cov}(X_j u \cdot u) \leq C \min_{|u| = 1} (\text{Cov}(X_j) u \cdot u). \]
5.0.1. **Uniformly elliptic chains.** In this section we consider a (somewhat) less general class of Markov chains \( \{\xi_j\} \), but more general functionals. Let \( \{\xi_j\} \) be a Markov chain with transition densities
\[
\mathbb{P}(\xi_{j+1} \in \Gamma|\xi_j = x) = \int p_j(x, y) d\mu_{j+1}(y)
\]
where \( \mu_{j+1} \) is a measure on the space \( \mathcal{X}_{j+1} \) of \( \xi_{j+1} \) and \( \Gamma \subset \mathcal{X}_{j+1} \) is a measurable set. We assume that there exists \( \varepsilon_0 > 0 \) so that for any \( i \) we have \( \sup_{x,y} p_i(x,y) \leq 1/\varepsilon_0 \), and the second step transition densities of \( \xi_{i+2} \) given \( \xi_i \) are bounded below by \( \varepsilon_0 \) (this is the uniform ellipticity condition):
\[
\inf_{i \geq 1} \inf_{x, y} \int p_i(x,y) p_{i+1}(y,z) d\mu_{i+1}(y) \geq \varepsilon_0.
\]
Then the resulting Markov chain \( \{\xi_j\} \) is exponentially fast \( \phi \)-mixing (see [5 Proposition 1.22]). Note that if the first step transition densities \( p_i \) were bounded below then we would get (5.1), but the assumption about the second step transition densities does necessary yield (5.1).

Next, we take a uniformly bounded sequence of measurable functions \( f_j : \mathcal{X}_j \times \mathcal{X}_{j+1} \to \mathbb{R}^d \) and set \( X_j = f_j(\xi_j, \xi_{j+1}) - \mathbb{E}[f_j(\xi_j, \xi_{j+1})] \). Let us fix some unit vector \( u \). Then, by applying [5 Theorem 2.1] with the real-valued functions \( f_j \cdot u \) (which are uniformly bounded in both \( j \) and \( u \)) we see that there are non-negative numbers \( u_i(f; u) = u_i(f_{i-2} \cdot u, f_{i-1} \cdot u, f_i \cdot u) \) and constants \( A, B, C, D > 0 \) which depend only on \( \varepsilon_0 \) and \( K := \sup_j \sup_i |f_j| \) so that for all \( m, n \) with \( m - n \geq 3 \) we have
\[
(5.2) \quad A \sum_{j=n+3}^{m} u_j^2(f; u) - B \leq \text{Var}(S_{n,m} \cdot u) \leq C \sum_{j=n+3}^{m} u_j^2(f; u) + D
\]
where we recall that \( S_{n,m} = \sum_{j=n}^{m} X_j \). The numbers \( u_i(f; u) \) are given in [5 Definition 1.14]:
\[
u_j^2(f; u) = (u_i(f; u))^2
\]
is the variance of the balance (in the terminology of [5]) function \( \Gamma_i = \Gamma_i \cdot u \) given by
\[
\Gamma_i(x_{i-2}, x_{i-1}, x_i, y_{i-1}, y_i, y_{i+1}) = f_{i-2}(x_{i-2}, x_{i-1}) \cdot u + f_{i-1}(x_{i-1}, x_i) \cdot u + f_i(x_i, y_{i+1}) \cdot u \\
- f_{i-2}(x_{i-2}, y_{i-1}) \cdot u - f_{i-1}(y_{i-1}, y_i) \cdot u - f_i(y_i, y_{i+1}) \cdot u
\]
corresponding to the hexagon generated by \( (x_{i-1}, x_i, y_{i+1}; y_{i-1}, y_i, y_{i+1}) \), with respect to the probability measure on the space of hexagons positioned at “time” \( i \), as introduced in [5 Section 1.3]. We thus have the following result.

5.2. **Proposition.** **Assumption 2.2** (and hence Theorem 2.3) holds true if there is a constant \( C \geq 1 \) so that for each \( j \) the matrix \( B_j \) defined by \( (B_j)_{k,l} = \frac{1}{2} (u_j^2(f, e_k) + u_j^2(f, e_l)) \) (where \( e_m \) is the \( m \)-th standard unit vector), satisfies
\[
\max_{|u|=1} (B_j u \cdot u) \leq C \min_{|u|=1} (B_j u \cdot u).
\]

**Weaker results for uniformly contracting Markov chains.** Let \( \{\xi_j\} \) be a Markov chain. Let us consider the transition operators \( Q_j \) given by \( Q_j g(x) = \mathbb{E}[g(\xi_{j+1})|\xi_j = x] \). For each \( j \geq 1 \) let \( \rho_j \) be the \( L^2 \)-operator norm of the restriction of \( Q_j \) to the space of zero-mean square-integrable functions \( g(\xi_{j+1}) \) (see [17]). We assume here that
\[
\rho := \sup_j \rho_j < 1.
\]
In these circumstances the Markov chain \( \{\xi_j\} \) is exponentially fast \( \rho \)-mixing (see [17]), and so by [2 (1.22)] we get (2.3). Note also that by [21 Lemma 4.1] we have,
\[
\rho_j \leq \sqrt{\pi(Q_j)}
\]
and so this is a weaker assumption than (5.1).
Let $f_j : X_j \to \mathbb{R}^d$ be a sequence of measurable uniformly bounded functions and set $X_j = f_j(\xi_j)$. We prove here the following result.

5.3. **Theorem.** Suppose that $s_n = \min_{|u|=1} (V_n u \cdot u) \geq c_0 n^{\delta_0}$ for some constants $c_0, \delta_0 > 0$. Assume also that there exists $C \geq 1$ so that for each $j$ we have
\[
\max_{|u|=1} (\text{Cov}(X_j) u \cdot u) \leq C \min_{|u|=1} (\text{Cov}(X_j) u \cdot u).
\]

Then there is a coupling of $X_1, X_2, \ldots$ with a sequence of independent centered Gaussian vectors $Z_1, Z_2, \ldots$ with the properties described in Theorem 2.3.

5.4. **Remark.** Relying on (5.4) below, the condition $s_n \geq c_0 n^{\delta_0}$ is satisfied if $\sum_{j=1}^n c_j \geq c_0 C^{-1} n^{\delta_0}$ where $c_j = \min_{|u|=1} (\text{Cov}(X_j) u \cdot u) = \min_{|u|=1} \text{Var}(X_j \cdot u)$.

**Proof of Theorem 5.3.** First, by [17, Proposition 13], there are constants $C_1, C_2 > 0$ so that for all $n, m$ with $n \leq m$ and every unit vector $u$ we have
\[
C_1 \sum_{j=n}^m \text{Var}(X_j \cdot u) \leq \text{Var}(S_{n,m} \cdot u) \leq C_2 \sum_{j=n}^m \text{Var}(X_j \cdot u)
\]
(5.3)

By using (5.4) and (5.3) we see that Assumption 2.2 is valid.

The proof of Theorem 5.3 proceeds now similarly to the proof of Theorem 2.3 with the following exception: we cannot use [15, Theorem 6.17] in order to obtain (3.14), since it requires (2.4). In order to overcome this difficulty, consider first the scalar case. Let $\rho = \sum_{j=1}^n \text{Var}(X_j) \cdot u \cdot u$, so that for all $n, m$ with $n \leq m$ and $\sum_{j=1}^m \text{Var}(X_j) \geq \rho$, we have
\[
\|S_{m,n}\|_{L^p} \leq E_{p,K} \left( \sum_{j=n}^m \text{Var}(X_j) \right)^{1/2}.
\]
(5.5)

Now, by (5.4) we have that
\[
\sum_{j=n}^m \text{Var}(f_j(X_j)) \leq C_1^{-1} \text{Var}(S_{n,m})
\]
and so there are constants $R_p, U_p > 0$ so that for all $n, m$ with $\|S_{m,n}\|_{L^2} \geq U_p$, we have
\[
\|S_{n,m}\|_{L^p} \leq R_p \|S_{n,m}\|_{L^2}.
\]
(5.6)

By replacing $X_j$ with $X_j \cdot u$ for an arbitrary unit vector $u$ and then taking the supremum over $u$, we see that (5.6) holds true also in the vector-valued case (i.e. when $d > 1$).

Finally, let us obtain (3.14). Set $B_n = \sum_{j=1}^n \xi_j$. Then by the Markov inequality for every $\varepsilon > 0$ and $q > 1$ we have
\[
\mathbb{P}(\|S_n - B_n\| \geq n^\varepsilon) = \mathbb{P}(\|S_n - B_n\|^q \geq n^{\varepsilon q}) \leq n^{-\varepsilon q} \|S_n - B_n\|_{L^q}^q \leq R_{q,K}(1+c)n^{-\varepsilon q}
\]
where in the last inequality we have also used (5.6) and that $\|S_n - B_n\|_{L^q} \leq c$ is bounded in $n$. Taking $q > 1/\varepsilon$ and applying the Borel-Cantelli lemma we get that
\[
|S_n - B_n| = o(n^\varepsilon) = o(s_n^\varepsilon), \text{ a.s.}
\]
Since $\varepsilon$ is arbitrary small we get that for every $\varepsilon > 0$ we have
\[
|S_n - B_n| = o(s_n^\varepsilon), \text{ a.s.}
\]
Now the proof of Theorem 5.3 is completed similarly to the end of the proof of Theorem 2.3. □
Acknowledgment. The original rates obtained in previous versions of this paper were $o(n^{\delta}) + o(V_n^{1/4+\delta})$, for any $\delta > 0$. I would like to thank D. Dolgopyat for several discussions which helped improving these rates to the current rates $o(V_n^{1/4+\delta})$ in Theorem 2.3.

References

[1] J. Berkes and W. Philipp, Approximation theorems for independent and weakly dependent random vectors, Ann. Probab. 29-54 (1979).
[2] R. Bradley, Basic properties of strong mixing conditions. A survey and some open questions, Probability Surveys, Vol. 2 (2005) 107–144.
[3] R.C. Bradley, Introduction to Strong Mixing Conditions, Volume 1, Kendrick Press, Heber City, 2007.
[4] C. Cuny, J. Dedecker, F. Merlevède, Rates of convergence in invariance principles for random walks on linear groups via martingale methods, Trans. Amer. Math. Soc. 374 (2021), 137-174.
[5] P. Doukhan, Mixing: Properties and Examples, Lecture Notes in Statistics, Vol. 85, Springer, Berlin (1994).
[6] R. Dobrushin, R. Central limit theorems for non-stationary Markov chains I, II. Theory Probab. Appl.1, 65-80, 329-383 (1956).
[7] D. Dragičević, G. Froyland, C. Gonzalez-Tokman and S. Vaienti, Almost Sure Invariance Principle for random piecewise expanding maps, Nonlinearity 31 (2018), 2252-2280.
[8] D. Dolgopyat, O. Sarig, Local limit theorems for inhomogeneous Markov chains, https://arxiv.org/abs/2109.05560
[9] D. Dragičević, Y. Hafouta, Almost sure invariance principle for random dynamical systems via Gouëzel’s approach, Nonlinearity, 34 6773.
[10] S. Gouëzel, Almost sure invariance principle for dynamical systems by spectral methods, Annals of Probability 38 (2010), 1639–1671.
[11] P.G. Hall and C.C. Hyde, Martingale central limit theory and its application, Academic Press, New York, 1980.
[12] D. L. Hanson and R. P. Russo, Some Results on Increments of the Wiener Process with Applications to Lag Sums of I.I.D. Random Variables, Ann. Probab. 11 (1983), 609–623.
[13] Nicolai Haydn, Matthew Nicol, Andrew Török and Sandro Vaienti, Almost sure invariance principle for sequential and non-stationary dynamical systems, Trans. Amer. Math. Soc. 369 (2017), 5293-5316.
[14] M. Iosifescu and R. Theodorescu, Random processes and learning, Die Grundlehren der mathematischen Wissenschaften, Band 150. Springer-Verlag, New York (1969).
[15] F. Merlevède, M. Peligrad, M. and S. Utev, S, Functional Gaussian Approximation for Dependent Structures, Oxford University Press (2019).
[16] M. Peligrad and S. Utev, A new maximal inequality and invariance principle for stationary sequences. Ann. Probab. 33, 798-815 (2005).
[17] M. Peligrad, Central limit theorem for triangular arrays of non-homogeneous Markov chains, Probab. Theory Relat. Fields (2012) 154:409-428.
[18] W. Philipp and W.F. Stout, Almost sure invariance principles for partial sums of weakly dependent random variables, Mem. Amer. Math. Soc. 161 (1975).
[19] E. Rio, Théorie asymptotique des processus aléatoires faiblement dépendants. Mathématiques et Applications 31, Springer-Verlag, Berlin, 2000
[20] Q.M. Shao, Almost sure invariance principles for mixing sequences of random variables, Stochastic Processes and their Applications 48, 319-334 (1993).
[21] S. Sethuraman and S.R.S Varadhan, A martingale proof of Dobrushin’s theorem for non-homogeneous Markov chains, Electron. J. Probab. 10, 1221–1235 (2005).
[22] W. Wu and Z. Zou, Gaussian approximations for non-stationary multiple time series Statistica Sinica, Vol. 21, No. 3 , pp. 1397-1413 (2011).

Email address: yeor.hafouts@mail.huji.ac.il, hafuta.1@osu.edu