Moderate deviations for iterates of expanding maps

AMIR DEMBO$^1$ AND OFER ZEITOUNI$^2$

DEPARTMENT OF ELECTRICAL ENGINEERING

TECHNION, HAIFA, ISRAEL 32000

April 19, 1996.

Abstract We provide a mild mixing condition that carries the C.L.T. for normalized empirical means of centered stationary sequence of bounded random variables to the whole range of moderate deviations. It is also key for the exponential convergence of the laws of empirical means. The motivating example for this work are iterates of expanding maps, equipped with their unique invariant measure.

1 Introduction

Let $\{X_i\}$ denote a realization of a centered stationary process with values in the unit ball of $\mathbb{R}^d$. Let $S_n = \sum_{i=1}^n X_i$. Suppose that the normalized empirical means $\hat{S}_n = S_n / \sqrt{n}$ converge in distribution to a Gaussian law of zero mean and covariance matrix $V$ and define

$$I(x) = \sup_{\lambda \in \mathbb{R}^d} \{ \langle \lambda, x \rangle - \frac{1}{2} \langle \lambda, V \lambda \rangle \}. \quad (1)$$

We say that $\hat{S}_n$ satisfies the Moderate Deviation Principle (MDP) with the above rate function if for every Borel set $\Gamma$ and any $a_n \downarrow 0$ such that $na_n \to \infty$,

$$- \inf_{x \in \Gamma} I(x) \leq \lim_{n \to \infty} \inf_{x \in \Gamma} \ln \left( \frac{P(\sqrt{a_n} \hat{S}_n \in \Gamma)}{a_n} \right) \quad (2)$$

$$\leq \lim_{n \to \infty} \sup_{x \in \Gamma} \ln P(\sqrt{a_n} \hat{S}_n \in \Gamma) \leq - \inf_{x \in \Gamma} I(x) \quad (3)$$

---

$^1$On leave from the Department of Mathematics and Department of Statistics, Stanford University, Stanford, CA 94305. Partially supported by an NSF DMS-9403553 grant, a US-ISRAEL BSF grant, and by the fund for promotion of research at the Technion.

$^2$Partially supported by a US-ISRAEL BSF grant, and by the fund for promotion of research at the Technion.

Key Words: Moderate deviations, Iterates of expanding maps, discrete time Markov processes, AMS (1991) subject classifications: 60F10, 60J05.
Our goal is to provide an MDP for $\hat{S}_n$ under the following mixing condition.

(M) There exist $q > 0$ and $c_0, \ell_0 < \infty$, such that for any $\ell \geq \ell_0$ and $\ell$-separated intervals $T_i$ there exist independent random variables $\{\tilde{S}_i\}$, such that for every $\theta \in (0,q)$ and $m < \infty$

$$E(e^{\theta \sum_{i=1}^m |\sum_{j \in T_i} X_j - \tilde{S}_i|}) \leq e^\theta c_0 m.$$  \hspace{1cm} (4)

Our main result is:

Theorem 1 If (M) holds in addition to the C.L.T. then $\hat{S}_n$ satisfies the MDP with the rate function $I(\cdot)$ of (1).

Suppose $\{X_j\}$ is a realization of a discrete time Markov process of transition kernel $\pi$ on the unit ball of $\mathbb{R}^d$, with $X_0$ distributed according to $\mu$ which is an invariant measure for $\pi$. The next lemma provides a simple sufficient condition for (M) in this case.

Lemma 1 The condition (M) holds if there exist $q > 0$, $c_0, \ell_0 < \infty$ and $N \subset \mathbb{R}^d$ such that $\mu(N) = 0$ and for every $u \geq c_0/2$

$$\sup_{y \notin N \atop \bar{y} \notin N} P\left( \sum_{j \geq t_0} |Y_j - \bar{Y}_j| \geq u | Y_0 = y, \bar{Y}_0 = \bar{y} \right) \leq \frac{1}{2} c_0 q e^{-2qu},$$  \hspace{1cm} (5)

where $\{Y_j\}$ and $\{\bar{Y}_j\}$ are coupled realizations of the Markov process of transition kernel $\pi$ such that $Y_k \in \sigma(\bar{Y}_{k-1}, \bar{Y}_k, Y_{k-1}, U_k)$ for an independent sequence $\{U_k\}$ of independent variables.

Proof: Fix the $\ell$-separated intervals $T_i = [s_i, t_i)$ such that $s_i \geq t_{i-1} + \ell_0$ for all $i$ setting $t_0 = 0$ without loss of generality. Choose $\bar{Y}_{t_i}$ for $i = 0, \ldots, m - 1$ independently according to law $\mu$ and let $\{\bar{Y}_j\}$ evolve for $j \in [t_{i-1}, t_i)$ according to the kernel $\pi$. Set $X_0 = \bar{Y}_0$ and for $i = 1, \ldots, m$ let $X_j = Y_j$, $j \in (t_{i-1}, t_i)$ be the Markov process realization coupled to $\{\bar{Y}_j : \bar{Y}_j \in [t_{i-1}, t_i)\}$, with $X_{t_i}$ arbitrarily.
chosen according to \( \pi (\cdot | X_{t_{i-1}}) \). Then, \( \{ X_j \} \) is a realization of the Markov process of transition kernel \( \pi \) with \( X_0 \) distributed according to \( \mu \). Moreover, this construction ensures the independence of \( \bar{S}_i = \sum_{j=s_i}^{t_i-1} Y_j \). Since with probability one \( \{ Y_i, \bar{Y}_i : i = 0, 1, \ldots, m \} \) and \( N \) are disjoint, it follows by the uniformity of (5) with respect to \( y, \bar{y} \notin N \) that for every \( u_i \in \mathbb{R}, i = 1, \ldots, m, \)

\[
P(\{| \sum_{j \in I_i} X_j - \bar{S}_i | \geq u_i : i = 1, \ldots, m \}) \leq \prod_{i=1}^{m} \left( \frac{1}{2} c_0 q e^{-2q u_i} 1_{u_i \geq \epsilon_0/2} + 1_{u_i < \epsilon_0/2} \right).
\]

(6)

By Fubini’s theorem, for any \( \theta > 0 \) and any random vector \( v = (v_1, \ldots, v_m) \in \mathbb{R}^m, \)

\[
E[ e^{\theta \sum_{i=1}^{m} v_i} ] = \int \theta^m e^{\theta \sum_{i=1}^{m} u_i} P(\{ v_i \geq u_i : i = 1, \ldots, m \}) du_1 \cdots du_m ,
\]

(7)

Moreover, for any \( \theta \in (0, q), \)

\[
\int_{-\infty}^{\infty} \theta e^{\theta u} \left( \frac{1}{2} c_0 q e^{-2q u} 1_{u \geq \epsilon_0/2} + 1_{u < \epsilon_0/2} \right) du \leq e^{\theta \epsilon_0/2} (1 + \theta \epsilon_0/2) \leq e^{\theta \epsilon_0},
\]

using which (4) follows from (6) and (7).

We say that \( n^{-1} S_n \to 0 \) with exponential tails if for every \( \eta > 0 \)

\[
\limsup_{n \to \infty} n^{-1} \log P(|n^{-1} S_n| > \eta) < 0,
\]

(8)

that is, if the law of \( n^{-1} S_n \) converges exponentially rapidly to its limit. This property holds for the empirical means \( n^{-1} S_n \) of a \( \phi \)-mixing bounded process \( \{ X_j \} \) (c.f. [3, Section 5] and the references therein). While (M) does not imply that \( \{ X_j \} \) is \( \phi \)-mixing, we show that

**Proposition 1** If (M) holds then \( n^{-1} S_n \to 0 \) with exponential tails.

In contrast with the strong mixing condition (S) of [2] and the mixing assumptions of [8], the condition (M) involves approximation by independent random variables instead of approximation
of laws by the product measure. This is apparent in our motivation for the assumption (M) which comes from the study of iterates of expanding maps. The latter result with discrete time Markov processes on the compact space $[0, 1]$ which fail to satisfy the uniform mixing assumptions of [4], needed for a Donsker-Varadhan type LDP. Indeed, for some atomic initial measures they satisfy neither LDP nor MDP. Nevertheless, as we show in Section 2 they satisfy the condition (5) hence the MDP for the normalized empirical means of such processes is a direct corollary of Theorem 1 and the convergence with exponential tails of the empirical means follows from Proposition 1.

The regeneration structure of iterative expanding maps is discussed in [7] and sufficient conditions for MDP to hold for Markov processes are provided in [9]. However, we do not see how these works may provide the MDP in the case dealt with here.

In Section 3 we provide the proof of Theorem 1. Its proof relies on the next lemma which uses (M) to control exponential moments in the convergence in distribution of $\hat{S}_n$.

**Lemma 2** The condition (M) implies that for any $\lambda \in \mathbb{R}^d$

$$\limsup_{n \to \infty} E(\langle \lambda, \hat{S}_n \rangle^2) < \infty \implies \limsup_{n \to \infty} E(e^{\langle \lambda, S_n \rangle}) < \infty.$$ 

The proof of Lemma 2 is provided in Section 4, and that of Proposition 1 in Section 5.

### 2 Application to iterates of expanding maps

Let $h : I = [0, 1] \to I$ be piecewise monotone, $C^1$ and uniformly expanding: that is, there is a finite set $U = U(h)$ of points

$$0 = u_0 < u_1 < \cdots < u_m = 1$$
in $I$ such that, for each interval $J_i = J_i(h) = (u_{i-1}, u_i)$, both $h$ restricted to $J_i$, and its continuous extension to $[u_{i-1}, u_i]$ are $C^1$ and monotone, satisfying

\[1 < \inf_{x \in J_i} |h'(x)|, \quad \sup_{x \in J_i} |h'(x)| < \infty.\]

The iterates of $h$ are denoted by $h_r$ and Lebesgue measure on $I$ by $\lambda$. Following [1] assume that

**A1** $\lambda(\Gamma) \in \{0, 1\}$ whenever $\lambda(\Gamma \Delta h_r(\Gamma)) = 0$ for some $r \geq 1$;

**A2** for some $r \geq 1$, $|h_r^{-1}(x)| \geq 4$ for all $x \notin h_r(U(h_r))$;

**A3** $h'$ is piecewise Hölder continuous with some positive exponent.

Then, there is exactly one invariant measure $\mu$ for $h$ which is absolutely continuous with respect to $\lambda$. Consider the stationary centered process $X_j = h_j(X_0 + \xi) - \xi$ for $\xi = \int xd\mu$ and $X_0 + \xi$ distributed according to $\mu$. Let $\{Y_j\}$ be the corresponding time reversal Markov process. Note that the law of $\hat{S}_n$ is identical to that of $\frac{1}{n} \sum_{j=1}^n Y_j$ and the C.L.T. for $\hat{S}_n$ follows as in [6, Theorem 5]. Moreover, assuming **A1–A3**, by [1, Theorem 3.4] there exist $c > 1$, $\beta > 0$, $K < \infty$, a set $N \subset I$ with $\mu(N) = 0$ and coupled realizations of the reversal Markov process $\{Y_j\}$, $\{\vec{Y}_j\}$ such that for every $y \notin N$, $\vec{y} \notin N$ and every non-negative integers $\ell$, $k$,

\[P(\sum_{j \geq \ell} |Y_j - \vec{Y}_j| \geq 2k + (1 - c^{-1})^{-1}|Y_0 = y, \vec{Y}_0 = \vec{y}) \leq P(\sup_{j \geq 0} c^j|Y_j - \vec{Y}_j| \geq c^{\ell+k}|Y_0 = y, \vec{Y}_0 = \vec{y}) \leq Kc^{-\beta(\ell+k)}.\]

Consequently, (5) holds for $q = \frac{\beta}{4} \log c > 0$, $c_0 = 2/(1 - c^{-1}) < \infty$ and $\ell_0$ such that $\frac{1}{2}c_0q \geq Kc^{\beta(1+c_0/4-\ell_0)}$. Applying Lemma 1 and Theorem 1 we conclude that $\hat{S}_n$ satisfies the MDP with rate function $I(x) = x^2/(2v)$ for $v = \lim_{n \to \infty} \text{Var}(\hat{S}_n)$. Similarly, by Lemma 1 and Proposition 1 we conclude that $n^{-1}S_n \to 0$ with exponential tails.
3 Proof of Theorem 1

Fix $a_n \downarrow 0$ such that $na_n \to \infty$. Fix $\epsilon > 0$ and a large integer $K$, let $t = 2K/\epsilon$ and

$$
\gamma_n = K \sqrt{na_n}, \quad m_n = \left\lfloor \frac{1}{ta_n} \right\rfloor
$$

(whereas any $K$ shall do for the upper bound (3) we need $K \to \infty$ for the lower bound (2)). Let $I_i^n$ for $i = 1, \ldots, m_n$, denote the intervals of length $[tna_n]$ each, separated by the intervals $J_i^n$ of length $\gamma_n$ each, with $J_1^n = \{1, \ldots, \gamma_n\}$ immediately followed by $I_1^n, J_2^n$, etc. Note that the intervals $J_1^n, J_2^n, \ldots, J_m^n, I_m^n$ cover $\{1, \ldots, n\}$. Lowering the value of $m_n$ and shortening the length of $I_m^n$ and $J_m^n$ if needed, we shall assume hereafter that the union of $I_i^n$ and $J_i^n$ is exactly $\{1, \ldots, n\}$. Let $q > 0$, $\epsilon_0 < \infty$ and $\epsilon_0 < \infty$ be as in condition (M). Considering the intervals $I_i^n$, for $\gamma_n \geq \epsilon_0$ the condition (M) then ensures the existence of the independent random variables $\{\tilde{S}_i^n\}, i = 1, \ldots, m_n$, for which (4) holds.

**Upper bound.** By Chebycheff's inequality, suffices for the upper bound (3) to show that for every $\lambda \in \mathbb{R}^d$

$$
\limsup_{n \to \infty} a_n \log M_n(\lambda) \leq \frac{1}{2} \langle \lambda, V \lambda \rangle \tag{9}
$$

where

$$
M_n(\lambda) = E(\exp(\lambda \tilde{S}_n)/\sqrt{\sigma_n})
$$

(c.f. [5, Proof of Theorem 2.3.6(a)] for such an application of Chebycheff's inequality). Since

$$
|\sum_{i=1}^{m_n} \sum_{j \in I_i^n} X_j / \sqrt{n a_n}| \leq |\lambda| K m_n,
$$

it follows by the choice of $t = 2K/\epsilon$ that

$$
\limsup_{n \to \infty} a_n \log M_n(\lambda) \leq \limsup_{n \to \infty} a_n \log \tilde{M}_n(\lambda) + |\lambda| \epsilon, \tag{10}
$$

where

$$
\tilde{M}_n(\lambda) = E\left( \exp\left( \sum_{i=1}^{m_n} T_i^n(\lambda) \right) \right), \quad T_i^n = T_i^n(\lambda) = \langle \lambda, \sum_{j \in I_i^n} X_j \rangle / \sqrt{n a_n}.
$$
Defining
\[ V_i = | \sum_{j \in I^n_i} X_j - \bar{S}_i^n |, \quad \bar{T}_i^n = \bar{T}_i^n(\lambda) = \langle \lambda, \bar{S}_i^n \rangle / \sqrt{n a_n}, \]
(11)
by (4) for \( m = 1 \) and \( \theta = \theta_n = |\lambda|/(\epsilon(1 - \epsilon)\sqrt{n a_n}) \to 0 \), we see that
\[
E \left( e^{\bar{T}_i^n / (1 - \epsilon)} \right) \leq E \left( e^{T_i^n / (1 - \epsilon)} e^{\lambda |V_i| / (1 - \epsilon) \sqrt{n a_n}} \right)
\leq E^c \left( e^{\theta_n V_i} \right) E^{(1 - \epsilon)} \left( e^{\bar{T}_i^n / (1 - \epsilon)^2} \right) = (1 + o(1)) E^{(1 - \epsilon)} \left( e^{T_i^n / (1 - \epsilon)^2} \right). \tag{12}
\]
Using (4) again, now for \( m = m_n \) and \( \theta = \theta_n = |\lambda|/(\epsilon \sqrt{n a_n}) \to 0 \), the independence of \( \bar{T}_i^n \) implies that
\[
M_n(\lambda) \leq E \left( e^{\sum_{i=1}^{m_n}(e^{\theta_n V_i} + \bar{T}_i^n)} \right)
\leq E^c \left( e^{\theta_n \sum_{i=1}^{m_n} V_i} \right) E^{(1 - \epsilon)} \left( e^{(1 - \epsilon)^{-1} \sum_{i=1}^{m_n} \bar{T}_i^n} \right)
\leq e^{e^{\theta_n \sum_{i=1}^{m_n} V_i}} E^{(1 - \epsilon)} \left( e^{(1 - \epsilon)^{-1} \sum_{i=1}^{m_n} \bar{T}_i^n} \right)
\leq e^{o(1)/a_n} \prod_{i=1}^{m_n} E \left( e^{\bar{T}_i^n / (1 - \epsilon)} \right)^{1 - \epsilon}. \tag{13}
\]
Since \( a_n m_n \to t^{-1} \), combining (12), (13), by the stationarity of \( \{X_j\} \) we have that
\[
\limsup_{n \to \infty} a_n \log M_n(\lambda) \leq t^{-1}(1 - \epsilon)^2 \limsup_{n \to \infty} m_n^{-1} \sum_{i=1}^{m_n} \log E (e^{T_i^n(\lambda) / (1 - \epsilon)^2})
= t^{-1}(1 - \epsilon)^2 \limsup_{n \to \infty} \log E (e^{\sqrt{T(\lambda, \bar{S}[\lfloor t a_n \rfloor]) / (1 - \epsilon)^2}}).
\]
By the C.L.T. for \( \bar{S}[\lfloor t a_n \rfloor] \) and Lemma 2 we thus have
\[
\limsup_{n \to \infty} a_n \log M_n(\lambda) \leq \frac{1}{2(1 - \epsilon)^2} \langle \lambda, V \lambda \rangle,
\]
which in view of (10) establishes (9) by taking \( \epsilon \to 0 \).
Lower bound. Recall that \( I(x) < \infty \) only if \( x = V \lambda \) for some \( \lambda \in \mathbb{R}^d \), in which case \( I(x) = \langle \lambda, V \lambda \rangle / 2 \). Hence, fixing \( x = V \lambda \), \( \lambda \in \mathbb{R}^d \) and \( \epsilon > 0 \), suffices for the lower bound (2) to show that

\[
\liminf_{n \to \infty} a_n \log P(|\sqrt{a_n} \hat{S}_n - x| < 4\epsilon) \geq -I(x)
\]

(c.f. [5, inequality (1.2.8)]). For all \( n \) large enough, \(|\sqrt{a_n} \sum_{i=1}^{m_n} \sum_{j \in J_i} X_j / \sqrt{n}| \leq \epsilon\), implying for \( V_i \) as in (11) that

\[
P(|\sqrt{a_n} \hat{S}_n - x| < 4\epsilon) \geq P \left( |\sqrt{a_n/n} \sum_{i=1}^{m_n} \sum_{j \in J_i} X_j - x| < 3\epsilon \right)
\]

\[
\geq P \left( |\sqrt{a_n/n} \hat{S}_{i}^{n} - x| < 2\epsilon \right) - P \left( \sqrt{a_n/n} \sum_{i=1}^{m_n} V_i > \epsilon \right)
\]

\[
\geq \prod_{i=1}^{m_n} P \left( |m_n \sqrt{\gamma_{n}} a_n/n \hat{S}_{i}^{n} - x| < 2\epsilon \right) - P \left( \sqrt{a_n/n} \sum_{i=1}^{m_n} V_i > \epsilon \right) ,
\]

where the last inequality follows by convexity of \( \{ y : |y - x| < 2\epsilon \} \) and the independence of \( \hat{S}_{i}^{n} \). By the stationarity of \( \{X_j\} \) and since \( m_n \sqrt{|tn a_n|} a_n/n \to 1/\sqrt{t} \) while \( \gamma_{n} m_n \sqrt{a_n/n} \to \epsilon / 2 \), for all \( n \) large enough,

\[
P(|m_n \sqrt{\gamma_{n}} a_n/n \hat{S}_{i}^{n} - x| < 2\epsilon) \geq P \left( |\frac{1}{\sqrt{t}} \hat{S}_{[n a_n]} - x| < \epsilon \right) - P(V_i > \gamma_{n}) .
\]

Applying (4) for \( m = 1 \) and \( \theta = 1/\sqrt{\gamma_{n}} \to 0 \) we see that \( \sup_i P(V_i > \gamma_{n}) \to 0 \). Thus,

\[
\liminf_{l \to \infty} \liminf_{n \to \infty} a_n \log \prod_{i=1}^{m_n} P(|m_n \sqrt{\gamma_{n}} a_n/n \hat{S}_{i}^{n} - x| < 2\epsilon) \geq \liminf_{l \to \infty} \frac{1}{l} \liminf_{n \to \infty} \log P \left( |\frac{1}{\sqrt{t}} \hat{S}_{[n a_n]} - x| < \epsilon \right)
\]

\[
\geq - \frac{1}{2} \langle \lambda, V \lambda \rangle = -I(x) ,
\]

where the last inequality follows by the C.L.T. for \( \hat{S}_{[n a_n]} \) and the large deviations of the Gaussian law of zero mean and covariance matrix \( V \). By Chebycheff’s inequality and (4) for \( m = m_n \) and
\[ \theta = \theta_n = 2t/\sqrt{n}a_n \to 0, \]

\[ P(\sqrt{a_n/n} \sum_{i=1}^{m_n} V_i > \epsilon) \leq e^{-K/a_n} E(e^{\theta_n \sum_{i=1}^{m_n} V_i}) \leq e^{-K/a_n} e^{\theta_n a_m n} = e^{-(K+o(1))/a_n}. \]  \hfill (17)

Combining (15), (16) and (17) we establish (14) by first taking \( n \to \infty \) followed by \( K \to \infty \) (hence \( t = 2K/\epsilon \to \infty \) as well).

\[ \square \]

4 \hspace{1em} Proof of Lemma 2

Let \( q > 0, c_0, \ell_0 < \infty \) be as in condition (M) and fix \( \lambda \in \mathbb{R}^d \) for which \( C(\lambda) = \sup_m E(\langle \lambda, \hat{S}_m \rangle^2) \) < \( \infty \). Let \( J_i^m = \{1, \ldots, \ell_0\} \) followed by the interval \( I_i^m \) of length \( m \) followed by the interval \( J_{2i}^m \) of length \( \ell_0 \) etc. defining

\[ \bar{M}_m(\lambda) = E\left(e^{\sum_{i=1}^m T_i^m(\lambda)}\right), \quad T_i^m = T_i^m(\lambda) = \langle \lambda, \sum_{j \in I_i^m} X_j \rangle/m. \]

Since \( |X_j| \leq 1 \) and \( \bar{M}_m(\lambda) \geq 1 \), for any \( m \geq \ell_0 + 1 \) and any \( n \in [m^2 + m\ell_0, (m + 1)^2 + (m + 1)\ell_0] \) we have

\[ E(e^{\langle \lambda, \hat{S}_n \rangle}) \leq e^{(1-q)(\ell_0+3)} \bar{M}_m(\lambda). \]

Hence, suffices to show that

\[ \limsup_{m \to \infty} \bar{M}_m(\lambda) < \infty \]  \hfill (18)

Fixing \( \epsilon > 0 \) and \( m > |\lambda|/(\epsilon(1-\epsilon)q) \), the condition (M) ensures the existence of independent variables \( \tilde{S}_i^m \) for which (4) holds. Let

\[ V_i = | \sum_{j \in I_i^m} X_j - \tilde{S}_i^m |, \quad \bar{T}_i^m = T_i^m(\lambda) = \langle \lambda, \tilde{S}_i^m \rangle/m. \]
Then, by (4) for $\theta_m = |\lambda|/(me) < q$ and the independence of $\{T_i^m\}$,

$$M_m(\lambda) \leq E \left( e^{\sum_{i=1}^m \hat{T}_i^m} e^{\theta_m \sum_{i=1}^m V_i} \right) \leq E^{(1-\epsilon)} \left( e^{(1-\epsilon)^{-1} \sum_{i=1}^m \hat{T}_i^m} \right) E^{\epsilon} \left( e^{\theta_m \sum_{i=1}^m V_i} \right) \leq \left( \prod_{i=1}^m E \left( e^{(1-\epsilon)^{-1} \hat{T}_i} \right) \right)^{1-\epsilon} e^{\epsilon |\lambda| c_0} \cdot (19)$$

By stationarity of $\{X_j\}$ the law of $T_i^m$ is independent of $i$. Hence, similar to the derivation of (19) we get,

$$\left( \prod_{i=1}^m E \left( e^{(1-\epsilon)^{-1} \hat{T}_i} \right) \right)^{1-\epsilon} \leq \left( \prod_{i=1}^m E \left( e^{(1-\epsilon)^{-2} \hat{T}_i} \right) \right)^{(1-\epsilon)^2} \left( \prod_{i=1}^m E \left( e^{\theta_m \sum_{i=1}^m V_i} \right) \right) \leq \left[ E \left( e^{(1-\epsilon)^{-2} \hat{T}} \right) \right] m^{(1-\epsilon)^2} e^{\epsilon |\lambda| c_0} \cdot (20)$$

Note that $|T^m| \leq |\lambda|$, while $E(T^m) = \langle \lambda, EX \rangle = 0$ and

$$m \sup_{\xi \in [0,t]} E \left( (T^m)^2 e^{\xi T^m} \right) \leq e^{t|\lambda|} E(\langle \lambda, \hat{S}_m \rangle)^2 \leq e^{t|\lambda|} C(\lambda).$$

Consequently, for all $m$ and $t > 0$,

$$E \left( e^{tT^m} \right) \leq 1 + t E(T^m) + \frac{t^2}{2} \sup_{\xi \in [0,t]} E \left( (T^m)^2 e^{\xi T^m} \right) \leq 1 + \frac{t^2}{2m} e^{t|\lambda|} C(\lambda),$$

implying that

$$\lim_{m \to \infty} \sup_{\xi \in [0,t]} \left[ E \left( e^{(1-\epsilon)^{-2} \hat{T}} \right) \right]^{m(1-\epsilon)^2} < \infty \cdot (21)$$

Combining (19), (20) and (21) we establish (18).
5 Proof of Proposition 1

For every $\lambda \in \mathbb{R}^d$ let

$$
\Lambda_n(\lambda) = n^{-1} \log E(e^{(\lambda,S_n)}), \quad \Lambda(\lambda) = \limsup_{n \to \infty} \Lambda_n(\lambda), \quad D\Lambda(\lambda) = \lim_{\theta \to 0} \theta^{-1} \Lambda(\theta \lambda) .
$$

With $q > 0$, $c_0, \ell_0 < \infty$ as in (M) fix $k \geq 1$ integer, setting $m_n = [n/(k + \ell_0)]$. Let $\{I^n_i : i = 1, \ldots, m_n\}$ be $\ell_0$-separated intervals of size $k$ within $\{1, \ldots, n\}$ and $\tilde{S}_i^n$ the corresponding independent variables. Applying (4) and Cauchy-Schwarz, by the independence of $\tilde{S}_i^n$ and the stationarity of $\{X_j\}$, for $|\lambda| < q/4$

$$
\log E \left(e^{\sum_{i=1}^{m_n} 2(\lambda, \tilde{S}_i^n)}\right) = \sum_{i=1}^{m_n} \log E \left(e^{2(\lambda, \tilde{S}_i^n)}\right) \leq 2c_0|\lambda|m_n + \frac{1}{2}m_n \log E \left(e^{4(\lambda, S_k)}\right).
$$

Applying yet again (4) and Cauchy-Schwarz,

$$
\Lambda_n(\lambda) \leq n^{-1}(\ell_0m_n + k + \ell_0 + c_0m_n)|\lambda| + (2n)^{-1} \log E \left(e^{\sum_{i=1}^{m_n} 2(\lambda, \tilde{S}_i^n)}\right).
$$

With $m_n/n \to k^{-1}$, it follows that

$$
\Lambda(\lambda) \leq k^{-1}(\ell_0 + 2c_0)|\lambda| + (4k)^{-1} \log E \left(e^{4(\lambda, S_k)}\right) . \quad (22)
$$

Since $|S_k| \leq k$ and $ES_k = 0$, (22) implies that for every $\lambda \in \mathbb{R}^d$,

$$
D\Lambda(\lambda) \leq k^{-1}(\ell_0 + 2c_0)|\lambda| + k^{-1} \lim_{t \to 0} t^{-1} \log E \left(e^{t(\lambda, S_k)}\right) = k^{-1}(\ell_0 + 2c_0)|\lambda| \to_{k \to \infty} 0 . \quad (23)
$$

Let $\{u_\ell : \ell = 1, \ldots, 2d\}$ be vectors of unit norm such that $|x| \leq d^{1/2} \max_\ell \langle u_\ell, x \rangle$ for every $x \in \mathbb{R}^d$.

By Chebycheff’s inequality, for any $\theta > 0$,

$$
P(|n^{-1}S_n| > \eta) \leq 2d \max_{\ell=1}^{2d} P(d^{1/2}\langle u_\ell, S_n \rangle > \eta n) \leq 2d \max_{\ell=1}^{2d} e^{n(\Lambda_n(\theta d^{1/2}u_\ell) - \theta n)} .
$$
By (23), \( \max_{\ell} \Lambda(\theta d^{1/2} u_{\ell}) \leq \theta \eta / 2 \) for all \( \theta > 0 \) small enough. Using such a value of \( \theta \),

\[
\limsup_{n \to \infty} n^{-1} \log P(|n^{-1} S_n| > \eta) \leq \max_{\ell=1}^{2d} \Lambda(\theta d^{1/2} u_{\ell}) - \theta \eta \leq -\theta \eta / 2 < 0 ,
\]

and \( n^{-1} S_n \to 0 \) with exponential tails since \( \eta > 0 \) is arbitrary. \( \square \)

**Acknowledgement** We thank Andrew Barbour for asking us about the moderate and large deviations of iterates of expanding maps and for providing us with a preliminary draft of [1].

**References**


