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

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

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

$$\begin{aligned} -\inf_{x\in\Gamma^o} I(x) &\le \liminf_{n\to\infty} a_n \log P(\sqrt{a}_n \hat{S}_n \in \Gamma) & (2)\\ &\le \limsup_{n\to\infty} a_n \log P(\sqrt{a}_n \hat{S}_n \in \Gamma) \le -\inf_{x\in\overline{\Gamma}} I(x) & (3)\end{aligned}$$

[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 ℓ-separated intervals $\overline{I}_i$ there exist independent random variables $\{\tilde{S}_i\}$, such that for every $\theta \in (0, q)$ and $m < \infty$

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

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 π on the unit ball of $\mathbb{R}^d$, with X_0 distributed according to μ which is an invariant measure for π. 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 $u \geq c_0/2$*

$$\sup_{y\notin N, \tilde{y}\notin N} P(\sum_{j\geq \ell_0} |Y_j - \tilde{Y}_j| \geq u | Y_0 = y, \tilde{Y}_0 = \tilde{y}) \leq \frac{1}{2} c_0 q e^{-2qu} , \tag{5}$$

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

Proof: Fix the ℓ-separated intervals $\overline{I}_i = [s_i, t_i)$ such that $s_i \geq t_{i-1} + \ell_0$ for all i setting $t_0 = 0$ without loss of generality. Choose $\tilde{Y}_{t_i}$ for $i = 0, \ldots, m-1$ independently according to law μ and let $\{\tilde{Y}_j\}$ evolve for $j \in [t_{i-1}, t_i)$ according to the kernel π. Set $X_0 = \tilde{Y}_0$ 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 $\{\tilde{Y}_j : j \in [t_{i-1}, t_i)\}$, 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 π with X_0 distributed according to μ. Moreover, this construction ensures the independence of $\tilde{S}_i = \sum_{j=s_i}^{t_i-1} \tilde{Y}_j$. Since with probability one $\{Y_{t_i}, \tilde{Y}_{t_i} : i = 0, 1, \ldots, m\}$ and N are disjoint, it follows by the uniformity of (5) with respect to $y, \tilde{y} \notin N$ that for every $u_i \in \mathbb{R}$, $i = 1, \ldots, m$,

$$P(\{|\sum_{j\in\overline{I}_i} X_j - \tilde{S}_i| \geq u_i : i = 1, \ldots, m\}) \leq \prod_{i=1}^{m} \left(\frac{1}{2}c_0 q e^{-2qu_i} 1_{u_i \geq c_0/2} + 1_{u_i < c_0/2}\right) . \tag{6}$$

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

$$E\left[e^{\theta \sum_{i=1}^m v_i}\right] = \int \theta^m e^{\theta \sum_{i=1}^m u_i} P(\{v_i \geq u_i : i = 1, \ldots, m\}) du_1 \cdots du_m \ , \tag{7}$$

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

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

using which (4) follows from (6) and (7). $\square$

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

$$\limsup_{n\to\infty} n^{-1} \log P(|n^{-1}S_n| > \eta) < 0, \tag{8}$$

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 ϕ-mixing bounded process $\{X_j\}$ (c.f. [3, Section 5] and the references therein). While (M) does not imply that $\{X_j\}$ is ϕ-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* $\lambda \in \mathbb{R}^d$

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

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 λ. Following [1] assume that

A1 $\lambda(\Gamma) \in \{0,1\}$ whenever $\lambda(\Gamma \triangle h_r(\Gamma)) = 0$ 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 μ for h which is absolutely continuous with respect to λ. Consider the stationary centered process $X_j = h_j(X_0 + \xi) - \xi$ for $\xi = \int x d\mu$ and $X_0 + \xi$ distributed according to μ. Let $\{Y_j\}$ be the corresponding time reversal Markov process. Note that the law of $\hat{S}_n$ is identical to that of $\frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j$ 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 $\{Y_j\}$, $\{\tilde{Y}_j\}$ such that for every $y \notin N$, $\tilde{y} \notin N$ and every non-negative integers ℓ, k,

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

Consequently, (5) holds for $q = \frac{\beta}{4} \log c > 0$, $c_0 = 2/(1-c^{-1}) < \infty$ and ℓ_0 such that $\frac{1}{2} c_0 q \geq K c^{\beta(1+c_0/4-\ell_0)}$. Applying Lemma 1 and Theorem 1 we conclude that $\hat{S}_n$ satisfies the MDP with rate function $I(x) = x^2/(2v)$ for $v = \lim_{n \to \infty} \mathrm{Var}(\hat{S}_n)$. 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

$$\gamma_n = K\lfloor\sqrt{n\,a_n}\rfloor, \quad m_n = \left\lceil \frac{1}{ta_n} \right\rceil$$

(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 $i = 1, \dots, m_n$, denote the intervals of length $\lceil tn\,a_n \rceil$ each, separated by the intervals J_i^n of length γ_n each, with $J_1^n = \{1, \dots, \gamma_n\}$ immediately followed by I_1^n, J_2^n, etc. Note that the intervals $J_1^n, I_1^n, \dots, J_{m_n}^n, I_{m_n}^n$ cover $\{1, \dots, n\}$. Lowering the value of m_n and shortening the length of $I_{m_n}^n$ and $J_{m_n}^n$ if needed, we shall assume hereafter that the union of I_i^n and J_i^n is exactly $\{1, \dots, n\}$. Let $q > 0$, $c_0 < \infty$ and $\ell_0 < \infty$ be as in condition (M). Considering the intervals I_i^n, for $\gamma_n \geq \ell_0$ the condition (M) then ensures the existence of the independent random variables $\{\tilde{S}_i^n\}$, $i = 1, \dots, m_n$, 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

$$M_n(\lambda) = E(e^{\langle \lambda, \hat{S}_n \rangle / \sqrt{a_n}})$$

(c.f. [5, Proof of Theorem 2.3.6(a)] for such an application of Chebycheff's inequality). Since $|\sum_{i=1}^{m_n} \sum_{j\in J_i^n} X_j/\sqrt{n}a_n| \leq |\lambda| K m_n$, 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

$$\tilde{M}_n(\lambda) = E\left(e^{\sum_{i=1}^{m_n} T_i^n(\lambda)}\right) \ , \qquad T_i^n = T_i^n(\lambda) = \langle \lambda, \sum_{j\in I_i^n} X_j \rangle / \sqrt{n\,a_n} \, .$$

Defining

$$V_i = |\sum_{j\in I_i^n} X_j - \tilde{S}_i^n|\,, \qquad \tilde{T}_i^n = \tilde{T}_i^n(\lambda) = \langle \lambda, \tilde{S}_i^n\rangle/\sqrt{n\,a_n}, \tag{11}$$

by (4) for $m = 1$ and $\theta = \theta_n = |\lambda|/(\epsilon(1-\epsilon)\sqrt{n\,a_n}) \to 0$, we see that

$$\begin{aligned} E\left(e^{\tilde{T}_i^n/(1-\epsilon)}\right) &\leq E\left(e^{T_i^n/(1-\epsilon)} e^{|\lambda| V_i/(1-\epsilon)\sqrt{n\,a_n}}\right) \\ &\leq E^{\epsilon}\left(e^{\theta_n V_i}\right) E^{(1-\epsilon)}\left(e^{T_i^n/(1-\epsilon)^2}\right) = (1+o(1)) E^{(1-\epsilon)}(e^{T_i^n/(1-\epsilon)^2})\,. \end{aligned} \tag{12}$$

Using (4) again, now for $m = m_n$ and $\theta = \theta_n = |\lambda|/(\epsilon\sqrt{n\,a_n}) \to 0$, the independence of $\tilde{T}_i^n$ implies that

$$\begin{aligned} \tilde{M}_n(\lambda) &\leq E\left(e^{\sum_{i=1}^{m_n}(\epsilon\theta_n V_i + \tilde{T}_i^n)}\right) \\ &\leq E^{\epsilon}\left(e^{\theta_n \sum_{i=1}^{m_n} V_i}\right) E^{(1-\epsilon)}\left(e^{(1-\epsilon)^{-1}\sum_{i=1}^{m_n}\tilde{T}_i^n}\right) \\ &\leq e^{\epsilon\theta_n c_0 m_n} E^{(1-\epsilon)}\left(e^{(1-\epsilon)^{-1}\sum_{i=1}^{m_n}\tilde{T}_i^n}\right) \\ &\leq e^{o(1)/a_n}\left[\prod_{i=1}^{m_n} E\left(e^{\tilde{T}_i^n/(1-\epsilon)}\right)\right]^{1-\epsilon}. \end{aligned} \tag{13}$$

Since $a_n m_n \to t^{-1}$, combining (12), (13), by the stationarity of $\{X_j\}$ we have that

$$\begin{aligned} \limsup_{n\to\infty} a_n \log \tilde{M}_n(\lambda) &\leq t^{-1}(1-\epsilon)^2 \limsup_{n\to\infty} m_n^{-1}\sum_{i=1}^{m_n} \log E(e^{T_i^n(\lambda)/(1-\epsilon)^2}) \\ &= t^{-1}(1-\epsilon)^2 \limsup_{n\to\infty} \log E(e^{\sqrt{t}\langle\lambda, \hat{S}_{\lceil tn\,a_n\rceil}\rangle/(1-\epsilon)^2})\,. \end{aligned}$$

By the C.L.T. for $\hat{S}_{\lceil tn\,a_n\rceil}$ and Lemma 2 we thus have

$$\limsup_{n\to\infty} a_n \log \tilde{M}_n(\lambda) \leq \frac{1}{2(1-\epsilon)^2}\langle \lambda, V\lambda\rangle\,,$$

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

$$\liminf_{n\to\infty} a_n \log P(|\sqrt{a_n}\hat{S}_n - x| < 4\epsilon) \geq -I(x) \tag{14}$$

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

$$\begin{aligned}
P(|\sqrt{a_n}\hat{S}_n - x| < 4\epsilon) &\geq P\left(|\sqrt{a_n/n}\sum_{i=1}^{m_n}\sum_{j\in I_i^n} X_j - x| < 3\epsilon\right) \\
&\geq P(|\sum_{i=1}^{m_n}\sqrt{a_n/n}\tilde{S}_i^n - x| < 2\epsilon) - P(\sqrt{a_n/n}\sum_{i=1}^{m_n} V_i > \epsilon) \\
&\geq \prod_{i=1}^{m_n} P(|m_n\sqrt{a_n/n}\tilde{S}_i^n - x| < 2\epsilon) - P(\sqrt{a_n/n}\sum_{i=1}^{m_n} V_i > \epsilon) \ , \qquad (15)
\end{aligned}$$

where the last inequality follows by convexity of $\{y : |y - x| < 2\epsilon\}$ and the independence of $\tilde{S}_i^n$. By the stationarity of $\{X_j\}$ and since $m_n\sqrt{\lceil tn\, a_n\rceil a_n/n} \to 1/\sqrt{t}$ while $\gamma_n m_n \sqrt{a_n/n} \to \epsilon/2$, for all n large enough,

$$P(|m_n\sqrt{a_n/n}\tilde{S}_i^n - x| < 2\epsilon) \geq P(|\frac{1}{\sqrt{t}}\hat{S}_{\lceil tn\, a_n\rceil} - x| < \epsilon) - P(V_i > \gamma_n) \ .$$

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,

$$\begin{aligned}
\liminf_{t\to\infty}\liminf_{n\to\infty} a_n \log \prod_{i=1}^{m_n} P(|m_n\sqrt{a_n/n}\tilde{S}_i^n - x| < 2\epsilon) &\geq \liminf_{t\to\infty}\frac{1}{t}\liminf_{n\to\infty}\log P(|\frac{1}{\sqrt{t}}\hat{S}_{\lceil tn\, a_n\rceil} - x| < \epsilon) \\
&\geq -\frac{1}{2}\langle\lambda, V\lambda\rangle = -I(x) \ , \qquad (16)
\end{aligned}$$

where the last inequality follows by the C.L.T. for $\hat{S}_{\lceil tn\, a_n\rceil}$ 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

$\theta = \theta_n = 2t/\sqrt{na_n} \to 0,$

$$P(\sqrt{a_n/n}\sum_{i=1}^{m_n} V_i > \epsilon) \leq e^{-K/a_n} E(e^{\theta_n \sum_{i=1}^{m_n} V_i}) \leq e^{-K/a_n} e^{\theta_n c_0 m_n} = e^{-(K+o(1))/a_n} \,. \tag{17}$$

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 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 $J_1^m = \{1, \ldots, \ell_0\}$ followed by the interval I_1^m of length m followed by the interval J_2^m of length ℓ_0 etc. defining

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

Since $|X_j| \leq 1$ and $\tilde{M}_m(\lambda) \geq 1$, 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

$$E(e^{\langle \lambda, \hat{S}_n \rangle}) \leq e^{|\lambda|(\ell_0+3)} \tilde{M}_m(\lambda) \,.$$

Hence, suffices to show that

$$\limsup_{m \to \infty} \tilde{M}_m(\lambda) < \infty \tag{18}$$

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

$$V_i = |\sum_{j \in I_i^m} X_j - \tilde{S}_i^m| \,, \qquad \tilde{T}_i^m = \tilde{T}_i^m(\lambda) = \langle \lambda, \tilde{S}_i^m \rangle / m.$$

Then, by (4) for $\theta_m = |\lambda|/(m\epsilon) < q$ and the independence of $\{\tilde{T}_i^m\}$,

$$\begin{aligned}
\tilde{M}_m(\lambda) &\leq E\left(e^{\sum_{i=1}^m \tilde{T}_i^m} e^{\epsilon\theta_m \sum_{i=1}^m V_i}\right) \\
&\leq E^{(1-\epsilon)}\left(e^{(1-\epsilon)^{-1}\sum_{i=1}^m \tilde{T}_i^m}\right) E^{\epsilon}\left(e^{\theta_m \sum_{i=1}^m V_i}\right) \\
&\leq \left(\prod_{i=1}^m E(e^{(1-\epsilon)^{-1}\tilde{T}_i^m})\right)^{1-\epsilon} e^{|\lambda| c_0} . \qquad (19)
\end{aligned}$$

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

$$\begin{aligned}
\left(\prod_{i=1}^m E(e^{(1-\epsilon)^{-1}\tilde{T}_i^m})\right)^{1-\epsilon} &\leq \left(\prod_{i=1}^m E(e^{(1-\epsilon)^{-2}T_i^m})\right)^{(1-\epsilon)^2} \left(\prod_{i=1}^m E(e^{\frac{|\lambda|}{m\epsilon(1-\epsilon)}V_i})\right)^{\epsilon(1-\epsilon)} \\
&\leq \left[E\left(e^{(1-\epsilon)^{-2}T^m}\right)\right]^{m(1-\epsilon)^2} e^{|\lambda| c_0} . \qquad (20)
\end{aligned}$$

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 $t > 0$,

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

impling that

$$\limsup_{m \to \infty} \left[E\left(e^{(1-\epsilon)^{-2}T^m}\right)\right]^{m(1-\epsilon)^2} < \infty . \qquad (21)$$

Combining (19), (20) and (21) we establish (18). □

5 Proof of Proposition 1

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

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

With $q > 0$, $c_0, \ell_0 < \infty$ as in (M) fix $k \geq 1$ integer, setting $m_n = \lfloor n/(k+\ell_0) \rfloor$. Let $\{I_i^n : i = 1, \ldots, m_n\}$ be ℓ_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 $\tilde{S}_i^k$ and the stationarity of $\{X_j\}$, for $|\lambda| < q/4$

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

Applying yet again (4) and Cauchy-Schwarz,

$$\Lambda_n(\lambda) \leq n^{-1}(\ell_0 m_n + k + \ell_0 + c_0 m_n)|\lambda| + (2n)^{-1} \log E\left(e^{\sum_{i=1}^{m_n} 2\langle \lambda, \tilde{S}_i^n \rangle}\right) .$$

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

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

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

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

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 $\theta > 0$,

$$P(|n^{-1}S_n| > \eta) \leq 2d \max_{\ell=1}^{2d} P(d^{1/2}\langle u_\ell, S_n \rangle > \eta n) \leq 2d \max_{\ell=1}^{2d} e^{n(\Lambda_n(\theta d^{1/2} u_\ell) - \theta\eta)} .$$

By (23), $\max_\ell \Lambda(\theta d^{1/2} u_\ell) \le \theta\eta/2$ for all $\theta > 0$ small enough. Using such a value of θ,

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

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

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

[1] A. D. Barbour, R. M. Gerrard and G. Reinert, Iterates of expanding maps, preprint (1995).

[2] W. Bryc and A. Dembo, Large deviations and strong mixing, to appear in Annals Inst. Henri. Poincare. (1996).

[3] W. Bryc and W. Smolenski, On the convergence of averages of mixing sequences, Journal of Theoretical Probability 6 (1993) 473–483.

[4] A. De Acosta, Large deviations for empirical measures of Markov chains, Journal of Theoretical Probability 3 (1990) 395–431.

[5] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Jones and Bartlett, Boston, MA, 1993.

[6] F. Hofbauer and G. Keller. Ergodic properties of invariant measures for piecewise monotone transformations, Math. Z. 180 (1982) 119–140.

[7] P. Ney and E. Nummelin, Regeneration for chains of infinite order and random maps, Contemporary Math. 169 (1993) 503–520.

[8] E. Nummelin, Large deviations for functionals of stationary processes, Probab. Th. and Rel. Fields 86 (1990) 387–401.

[9] L. Wu, Moderate deviations of dependent random variables related to CLT, Ann. Probab. 23 (1995) 420-445.