NONCONVENTIONAL LARGE DEVIATIONS THEOREMS

YURI KIFER   AND   S.R.S. VARADHAN

INSTITUTE OF MATHEMATICS   COURANT INSTITUTE
THE HEBREW UNIVERSITY   NEW YORK UNIVERSITY
JERUSALEM, ISRAEL   NEW YORK, USA

ABSTRACT. We obtain large deviations theorems for both discrete time expressions of the form $\sum_{n=1}^{N} F\big(X(q_1(n)), \ldots, X(q_\ell(n))\big)$ and similar expressions of the form $\int_0^T F\big(X(q_1(t)), \ldots, X(q_\ell(t))\big) dt$ in continuous time. Here $X(n), n \geq 0$ or $X(t), t \geq 0$ is a Markov process satisfying Doeblin's condition, F is a bounded continuous function and $q_i(n) = in$ for $i \leq k$ while for $i > k$ they are positive functions taking on integer values on integers with some growth conditions which are satisfied, for instance, when q_i's are polynomials of increasing degrees. Applications to some types of dynamical systems such as mixing subshifts of finite type and hyperbolic and expanding transformations will be obtained, as well.

1. INTRODUCTION

Nonconventional ergodic theorems which attracted substantial attention in ergodic theory (see, for instance, [2] and [13]) studied the limits of expressions having the form $1/N \sum_{n=1}^{N} T^{q_1(n)} f_1 \cdots T^{q_\ell(n)} f_\ell$ where T is a weakly mixing measure preserving transformation, f_i's are bounded measurable functions and q_i's are polynomials taking on integer values on the integers. While, for instance, [2] and [13] were interested in L^2 convergence, other papers such as [1] provided conditions for almost sure convergence in such ergodic theorems. Originally, these results were motivated by applications to multiple recurrence for dynamical systems taking functions f_i being indicators of some measurable sets.

Introducing stronger mixing or weak dependence conditions enabled us in [22] to obtain functional central limit theorems for even more general expressions of the form

$$(1.1) \qquad \frac{1}{\sqrt{N}} \sum_{n=1}^{[Nt]} \big(F(X(q_1(n)), ..., X(q_\ell(n)) - \bar{F} \big)$$

where $X(n), n \geq 0$ is a sufficiently fast mixing vector valued process with some moment conditions and stationarity properties, F is a locally Hölder continuous function with polinomial growth, $\bar{F} = \int F d(\mu \times \cdots \times \mu)$ and μ is the distribution

of $X(0)$. In order to ensure existence of limiting variances and covariances we had to impose certain assumptions concerning the functions $q_j(n)$, $j \geq 1$ saying that $q_j(n) = jn$ for $j = 1, ..., k$ while $q_j(n)$, $j \geq k$ are positive functions taking on integer values on integers with some (faster than linear) growth conditions.

The next natural step in the study of limiting behavior of nonconventional sums $S_N = \sum_{n=1}^{N} F(X(q_1(n)), ..., X(q_\ell(n)))$ is to obtain large deviations estimates. Namely, we will be interested in this paper in the asymptotical behavior as $N \to \infty$ of probabilities

$$(1.2) \qquad\qquad P\{\frac{1}{N}S_N \in \Gamma\}$$

for various (open or closed) sets $\Gamma \subset \mathbb{R}$. According to [19] under appropriate conditions $\frac{1}{N}S_N$ converges with probability one as $N \to \infty$ to $\bar{F} = \int F d\mu \times \cdots \times \mu$ where μ is the common distribution of $X(n)$'s. Thus, as usual, (1.2) describes deviations of $\frac{1}{N}S_N$ from the limit in the law of large numbers.

The study of asymptotics of probabilities in (1.2) leads to what is usually called the first level of large deviations. We will study also second level large deviations estimates which means in our setup to consider occupational measures

$$(1.3) \qquad\qquad \zeta_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{\left(X(q_1(n)), ..., X(q_\ell(n))\right)}$$

and to study the asymptotical behavior as $N \to \infty$ of probabilities $P\{\zeta_N \in \mathcal{U}\}$ where $\mathcal{U}$ is a subset in the space of probability measures on a corresponding product space. In addition, we will consider also large deviations in the averaging setup, namely, for the "slow" variable $\Xi^\varepsilon(n) = \Xi_x^\varepsilon(n)$ given by a difference equation of the form

$$(1.4)$$
$$\Xi^\varepsilon(n+1) = \Xi^\varepsilon(n) + \varepsilon F(\Xi^\varepsilon(n), X(q_1(n)), ..., X(q_\ell(n))), \ n = 0, 1, ..., \ \Xi_x^\varepsilon(0) = x$$

which is actually a generalization of the above since if $F(\xi, x_1, ..., x_\ell)$ does not depend on ξ then $\Xi_N^{\frac{1}{N}} = \frac{1}{N}S_N$. We will deal also with continuous time versions of the above results considering $S_T = \int_0^T F(X(q_1(t)), ..., X(q_\ell(t))) dt$ for some stochastic process $X(s)$, $s \geq 0$.

As for conventional sums ($\ell = k = 1$) meaningful large deviations estimates can be obtained only for some specific classes of stochastic processes and dynamical systems. In our more general situation we also assume that in the probabilistic setup $X(n)$, $n = 0, 1, ...$ is a Markov chain satisfying a (strong) Doeblin condition while in the dynamical systems setup we can consider $X(n) = X(n, \omega) = f(T^n\omega)$ where T is either a mixing subshift of finite type or a hyperbolic diffeomorphism or an expanding transformation and f is a Hölder continuous (vector) function. In the continuous time case we take the underlying process $X(t)$ to be in the probabilistic setup either an irreducible finite Markov chain with continuous time or a nondegenerate diffusion on a compact manifold while in the dynamical systems setup we can take $X(t) = X(t, \omega) = f(T^t\omega)$ where T^t, $t \geq 0$ is a hyperbolic flow on a compact manifold and f is a Hölder continuous (vector) function.

We will show that it is not difficult to reduce the problem to the case $k = \ell$ and the major problems arise only in dealing with random variables $X(n), X(2n), ..., X(kn)$. When $k = 1$ the above reduction leads to the standard (conventional) setup of large deviations. When $k > 1$ then the general case of

Markov sequences requires a quite elaborate technique and a lengthy proof and it will be treated in another paper while here when $k > 1$ we restrict ourselves to independent identically distributed (i.i.d.) sequences $X(n), n \geq 0$ which, unlike in the conventional setup, is still nontrivial.

Both probabilistic and dynamical systems setups are united by common ideas and motivations but their machineris are quite different and by this reason most of this paper deals with the probabilistic setup and only in the last Section 5 we discuss some of dynamical systems results which especially can benefit readers familiar with this field.

2. Preliminaries and main results

We start with the probabilistic discrete time setup where the underlying process $X(0)$, $X(1)$, $X(2)$, ... is a Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ and evolving on a Polish measurable space $(M, \mathcal{B})$ as its phase space. We assume a "strong" Doeblin condition saying that for some integer $n_0 > 0$, a constant $C > 0$ and a probability measure ν on M the n_0-step transition probability $P(n_0, x, \cdot)$ of the above Markov chain X satisfies

$$(2.1) \qquad C^{-1}\nu(G) \leq P(n_0, x, G) \leq C\nu(G)$$

for any $x \in M$ and every measurable set $G \subset M$. It is well known (see, for instance, [8]) that (2.1) implies existence of a unique invariant measure μ of the Markov chain X and the equality $\mu(G) = \int d\mu(x)P(n, x, G)$ yields that

$$(2.2) \qquad C^{-1} \leq \frac{d\mu}{d\nu}(x) = p(x) \leq C$$

where $d\mu/d\nu$ denotes the Radon-Nikodim derivative.

In all cases our setup includes also a bounded measurable function $F = F(x_1, x_2, ..., x_\ell)$ on the ℓ-times product space $M^\ell = M \times \cdots \times M$. The setup becomes complete with introduction of positive increasing functions q_j, $j = 1, ..., \ell$ taking on integer values on integers and such that

$$(2.3) \qquad q_j(n) = jn \quad \text{for} \quad j = 1, ..., k \quad \text{and some} \quad k \leq \ell$$

while for $j = k+1, ..., \ell$ and any $\gamma > 0$,

$$(2.4) \qquad \lim_{n\to\infty}(q_j(n) - q_j(n-1)) = \infty \quad \text{and} \quad \liminf_{n\to\infty}(q_j(\gamma n) - q_{j-1}(n)) > 0.$$

For any function W on M^ℓ we denote by $\hat{W}$ the function on M defined by

$$(2.5) \qquad \hat{W}(x) = \int \exp(W(x, x_2, ..., x_\ell))d\mu(x_2)...d\mu(x_\ell).$$

As usual we denote by P_x the probability conditioned to $X(0) = x$ and by E_x the corresponding expectation. Now, we can formulate our first result.

2.1. Theorem. *Let $W_\lambda(x_1, ..., x_\ell)$, $\lambda \in (-\infty, \infty)$ be a differentiable in λ family of bounded measurable functions on M^ℓ such that $dW_\lambda(x_1, ..., x_\ell)/d\lambda$ is bounded for each λ, as well. Assume that $k = 1$ in (2.3) and (2.4). Then for any $x \in M$ the limit*

$$(2.6) \qquad Q(W_\lambda) = \lim_{N\to\infty} \frac{1}{N} \ln E_x \exp \Big(\sum_{n=1}^N W_\lambda(X(q_1(n)), ..., X(q_\ell(n))) \Big)$$

exists, it is independent of x and it is differentiable in λ. In fact, $Q(W_\lambda) = \ln r(W_\lambda)$ where $r(W)$ is the spectral radius of the positive operator $R(W)$ acting by

$$(2.7) \qquad\qquad R(W)g(x) = \int P(x, dy)g(y)\hat{W}(y).$$

Furthermore, set $W_\lambda(x_1, ..., x_\ell) = \lambda F(x_1, ..., x_\ell)$ and

$$(2.8) \qquad\qquad J(u) = \sup_\lambda(\lambda u - r(W_\lambda)), \; u \in \mathbb{R}.$$

Then for any closed set $K \subset \mathbb{R}$,

$$(2.9) \qquad\qquad \limsup_{N\to\infty} \frac{1}{N} \ln P\{\frac{1}{N}S_N \in K\} \leq - \inf_{u\in K} J(u)$$

and for any open set $U \subset \mathbb{R}$,

$$(2.10) \qquad\qquad \liminf_{N\to\infty} \frac{1}{N} \ln P\{\frac{1}{N}S_N \in U\} \geq - \inf_{u\in U} J(u)$$

where, as before, $S_N = S_N(F) = \sum_{n=1}^N F\big(X(q_1(n)), ..., X(q_\ell(n))\big)$.

We observe that a very particular case of Theorem 2.1 when $\{X(n), \, n \geq 0\}$ are i.i.d. random variables was considered in Section 6 of [18]. Next, we describe the second level of large deviations in the nonconventional setup which deals with occupational measures ζ_N on M^ℓ given by (1.3) where M is assumed to be a compact space and δ_z is the unit mass concentrated at z. For any probability measure η on M^ℓ define

$$(2.11)$$
$$I(\eta) = - \inf_{u\in\mathbb{C}_+(M^\ell)} \int_{M^\ell} \ln \frac{E_{x_1} \int u(X(1), x_2, ..., x_\ell)d\mu(x_2)...d\mu(x_\ell)}{u(x_1, ..., x_\ell)} d\eta(x_1, ..., x_\ell)$$

where $\mathbb{C}_+(\cdot)$ denotes the space of all positive continuous functions on a space in brackets.

2.2. **Theorem.** *Let $k = 1$ in (2.3) and (2.4). Then for any continuous function $W = W(x_1, ..., x_\ell)$ on M^ℓ the limit*

$$(2.12) \qquad Q(W) = \lim_{N\to\infty} \frac{1}{N} \ln E_x \exp\big(\sum_{n=1}^N W(X(q_1(n)), ..., X(q_\ell(n)))\big)$$

is a convex lower semicontinuous functional satisfying

$$(2.13) \qquad Q(W) = \sup_{\eta\in\mathcal{P}(M^\ell)} \big(\int W(x_1, ..., x_\ell)d\eta(x_1, ..., x_\ell) - I(\eta)\big)$$

where $\mathcal{P}(\cdot)$ is the space of probability measures on a space in brackets considered with the topology of weak convergence.

Furthermore, for any closed set $K \subset \mathcal{P}(M^\ell)$,

$$(2.14) \qquad\qquad \limsup_{N\to\infty} \frac{1}{N} \ln P\{\zeta_N \in K\} \leq - \inf_{\eta\in K} I(\eta)$$

and for any open set $U \subset \mathcal{P}(M^\ell)$,

$$(2.15) \qquad\qquad \liminf_{N\to\infty} \frac{1}{N} \ln P\{\zeta_N \in U\} \geq - \inf_{\eta\in U} I(\eta).$$

Next, we exhibit continuous time versions of the above results. Here we assume that $X(t)$, $t \geq 0$ is a Markov process on a Polish measurable space $(M, \mathcal{B})$ such that for some $t_0 > 0$, a constant $C > 0$ and a probability measure ν on M the time t_0 transition probability $P(t_0, x, \cdot)$ of the above Markov process X satisfies

$$(2.16) \qquad C^{-1}\nu(G) \leq P(t_0, x, G) \leq C\nu(G)$$

for any $x \in M$ and every measurable set $G \subset M$. Again (see [8]), (2.16) implies existence of a unique invariant measure μ of the Markov process X which satisfies (2.2). Now we introduce positive increasing functions q_j, $j = 1, ..., \ell$ on $\mathbb{R}_+$ such that for some $0 < \alpha_1 < \alpha_2 < ... < \alpha_k$ and $k \leq \ell$,

$$(2.17) \qquad q_j(t) = \alpha_j t \quad \text{for} \quad j = 1, ..., k$$

while for $j = k+1, ..., \ell$ and any $\gamma > 0$,

$$(2.18) \qquad \lim_{t\to\infty} (q_j(t+\gamma) - q_j(t)) = \infty \ \ and \ \ \liminf_{t\to\infty}(q_j(\gamma t) - q_{j-1}(t)) > 0.$$

We will be interested in large deviations estimates as $T \to \infty$ for

$$S_T(F) = S_T = \int_0^T F\big(X(q_1(t)), ..., X(q_\ell(t))\big)dt.$$

2.3. Theorem. *Let $W_\lambda(x_1, ..., x_\ell)$, $\lambda \in (-\infty, \infty)$ be as in Theorem 2.1. Assume that $k = 1$ in (2.17) and (2.18). Then for any $x \in M$ the limit*

$$(2.19) \qquad Q_{cont}(W_\lambda) = \lim_{T\to\infty} \frac{1}{T} \ln E_x \exp\Big(\int_0^T W_\lambda(X(q_1(t)), ..., X(q_\ell(t)))dt \Big)$$

exists, it is independent of x and it is differentiable in λ. In fact, $Q_{cont}(W_\lambda) = \ln r_{cont}(W_\lambda)$ where $r_{cont}(W)$ is the spectral radius of the semigroup of positive operators $R_{cont}^t(W)$ acting by the formula

$$(2.20) \qquad R_{cont}^t(W)g(x) = E_x\big(g(X(t))\hat{W}_{cont}(t)\big)$$

where

$$(2.21) \qquad \hat{W}_{cont}(t) = \exp\Big(\int_0^t ds \int W_\lambda(X(\alpha_1 s), x_2, ..., x_\ell)d\mu(x_2)...d\mu(x_\ell) \Big).$$

Furthermore, set $W_\lambda(x_1, ..., x_\ell) = \lambda F(x_1, ..., x_\ell)$ and define $J(u) = J_{cont}(u)$ by (2.8) with r_{cont} in place of r. Then for any closed set $K \subset \mathbb{R}$,

$$(2.22) \qquad \limsup_{T\to\infty} \frac{1}{T} \ln P\{\frac{1}{T}S_T \in K\} \leq - \inf_{u\in K} J(u)$$

and for any open set $U \subset \mathbb{R}$,

$$(2.23) \qquad \liminf_{T\to\infty} \frac{1}{T} \ln P\{\frac{1}{T}S_T \in U\} \geq - \inf_{u\in U} J(u).$$

The second level of large deviations in the continuous time nonconventional setup deals with occupational measures

$$(2.24) \qquad \zeta_T = \frac{1}{T} \int_0^T \delta_{\big(X(q_1(t)), ..., X(q_\ell(t))\big)}dt$$

on M^ℓ. Now we assume that $X(t)$, $t \geq 0$ is a diffusion process on a compact Riemannian manifold M with the generator L which is a nondegenerate second order elliptic differential operator. For any probability measure η on M^ℓ set

$$(2.25) \quad I_{\text{cont}}(\eta) = -\inf_{u \in D_+} \int_M \frac{L_{x_1} u(x_1, x_2, ..., x_\ell) d\mu(x_2)...d\mu(x_\ell)}{u(x_1, x_2, ..., x_\ell)} d\eta(x_1, ..., x_\ell)$$

where the infimum is taken over all positive u from the domain of L.

2.4. Theorem. *Let $k = 1$ in (2.17) and (2.18). Then for any continuous function $W = W(x_1, ..., x_\ell)$ on M^ℓ the limit*
(2.26)

$$Q_{cont}(W) = \lim_{T \to \infty} \frac{1}{T} \ln E_x \exp\big(\int_0^T W(X(q_1(t)), ..., X(q_\ell(t)))dt \big) = r_{cont}(W)$$

is a convex lower semicontinuous functional satisfying

$$(2.27) \quad Q_{cont}(W) = \sup_{\eta \in \mathcal{P}(M^\ell)} \big(\int W(x_1, ..., x_\ell) d\eta(x_1, ..., x_\ell) - I_{cont}(\eta) \big).$$

Furthermore, for any closed set $K \subset \mathcal{P}(M^\ell)$,

$$(2.28) \qquad \limsup_{T \to \infty} \frac{1}{T} \ln P\{\zeta_T \in K\} \leq -\inf_{\eta \in K} I_{cont}(\eta)$$

and for any open set $\subset \mathcal{P}(M^\ell)$,

$$(2.29) \qquad \liminf_{T \to \infty} \frac{1}{T} \ln P\{\zeta_T \in U\} \geq -\inf_{\eta \in U} I_{cont}(\eta).$$

A similar result holds true when $X(t)$ is a nondegenerate continuous time Markov chain with a finite state space.

Next, we describe our large deviations estimates in a nonconventional averaging setup. Here we consider either a difference equation (1.4) for $\Xi^\varepsilon(n)$ in the discrete time case where $X(n)$, $n \geq 0$ is a Markov chain satisfying conditions of Theorem 2.1 or a differential equation for $\Xi^\varepsilon(t) = \Xi_x^\varepsilon(t) \in \mathbb{R}^d$, $t \geq 0$,

$$(2.30) \qquad \frac{d\Xi^\varepsilon(t)}{dt} = \varepsilon F\big(\Xi^\varepsilon(t), X(q_1(t)), ..., X(q_\ell(t))\big), \Xi_x^\varepsilon(0) = x$$

in the continuous time setup where $X(t)$, $t \geq 0$ is a Markov process satisfying conditions of Theorem 2.3. We assume that $F(\xi, x_1, ..., x_\ell)$ is bounded and Lipschitz continuous in ξ. The setup of (2.30) emerges considering, for instance, a time dependent small perturbation of the oscillator equation

$$(2.31) \qquad \qquad \ddot{x} + \lambda^2 x = \varepsilon g(x, \dot{x}, t)$$

where the force term g depends on time in a random way $g(x, y, t) = g(x, y, X(q_1(t)), ..., X(q_\ell(t)))$. Then passing to the polar coordinates (r, ϕ) with $x = r \sin(\lambda(t - \phi))$ and $\dot{x} = \lambda r \cos(\lambda(t - \phi))$ the equation (2.31) will be transformed into (2.30) with $\Xi^\varepsilon = (r, \phi)$. It seems reasonable that a random force may depend on versions of a same process moving with different speeds which is what we have here.

As it is well known (see, for instance, [25]), if $F(\xi, x_1, ..., x_\ell)$ is bounded and Lipschitz continuous in ξ and for each ξ the limit

$$\bar{F}(\xi) = \lim_{\mathcal{T} \to \infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} F(\xi, X(q_1(t)), ..., X(q_\ell(t)))dt$$

exists then for any $T \geq 0$,

$$\lim_{\varepsilon \to 0} \sup_{0 \leq t \leq T/\varepsilon} |\Xi^\varepsilon(t) - \bar{\Xi}^\varepsilon(t)| = 0$$

where

$$\frac{d\bar{\Xi}^\varepsilon(t)}{dt} = \varepsilon \bar{F}(\bar{\Xi}^\varepsilon(t)).$$

In the discrete time case we have to take

$$\bar{F}(\xi) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N} F(\xi, X(q_1(n)), ..., X(q_\ell(n))).$$

Almost everywhere limits of the averages above can be obtained by nonconventional pointwise ergodic theorems from [4] and [1], respectively, in rather general circumstances in the dynamical systems case and under another set of conditions existence of such limits follows from [19]. The next natural step here is to obtain large deviations estimates for the above approximation of the slow motion Ξ^ε by the averaged one $\bar{\Xi}^\varepsilon$.

For any $\eta \in \mathcal{P}(M^\ell)$ set

(2.32) $$\bar{B}_\eta(\xi) = \int B(\xi, x_1, ..., x_\ell) d\eta(x_1, ..., x_\ell).$$

For each absolutely continuous curve γ_t, $t \in [0, \mathcal{T}]$ set

$$\mathcal{S}_{0T}(\gamma) = \int_0^{\mathcal{T}} \inf\{I(\eta) : \dot{\gamma}_t = \bar{B}_\eta(\gamma_t)\} dt$$

where $I(\eta)$ is given by (2.11) or $I(\eta) = I_{\text{cont}}(\eta)$ given by (2.25) in the discrete or continuous time cases, respectively. If γ_t, $t \in [0, T]$ is not absolutely continuous we set $S_{0T}(\gamma) = \infty$.

2.5. **Theorem.** *Let $k = 1$ in (2.3) and (2.4) or in (2.17) and (2.18) and set $\Psi^\varepsilon(t) = \Xi^\varepsilon([t/\varepsilon])$ or $\Psi^\varepsilon(t) = \Xi^\varepsilon(t/\varepsilon)$ in the discrete or continuous time cases, respectively. Then for any continuous function $W_t(x_1, ..., x_\ell)$ on $\mathbb{R}_+ \times M^\ell$,*
(2.33)

$$\lim_{\varepsilon \to 0} \varepsilon \ln E_x \exp\left(\varepsilon^{-1} \int_0^T W_t\big(X(q_1(t/\varepsilon)), ..., X(q_\ell(t/\varepsilon))\big) dt\right) = \int_0^T r_{cont}(W_t) dt$$

where r_{cont} is the same as in Theorem 2.3 with W_t considered as a function on M^ℓ and in the discrete time case we either extend $q_j(t) = q_j([t])$ to all $t \geq 0$ in order to write the integral in exponent in (2.33) or replace this integral by the corresponding sum.

Furthermore, for any $a, \delta, \lambda > 0$ and every continuous γ_t, $t \in [0, \mathcal{T}]$, $\gamma_0 = x$ there exist $\varepsilon_0 > 0$ such that for all positive $\varepsilon < \varepsilon_0$,

(2.34) $$P\{\rho_{0,\mathcal{T}}(\Psi_x^\varepsilon, \gamma) < \delta\} \geq \exp\{-\frac{1}{\varepsilon}(\mathcal{S}_{0,\mathcal{T}}(\gamma) + \lambda)\} \quad and$$

(2.35) $$P\{\rho_{0,\mathcal{T}}(\Psi_x^\varepsilon, \Phi_{0,\mathcal{T}}^a(x)) \geq \delta\} \leq \exp\{-\frac{1}{\varepsilon}(a - \lambda)\}$$

where $\Psi_x^\varepsilon(0) = x$, $\rho_{0,\mathcal{T}}$ is the uniform distance and $\Phi_{0,\mathcal{T}}^a(x) = \{\gamma : \gamma_0 = x, \mathcal{S}_{0,T}(\gamma) \leq a\}$.

2.6. **Remark.** Suppose that the averaged motion $\bar{\bar{\Xi}}^\varepsilon$ has several attracting fixed points and limit circles. Then similarly to [12] (Markov chains case) and [17] (dynamical systems case) we can study rare transitions of the slow motion Ξ^ε between these attractors. However, in the nonconventional setup the situation is more complicated and this problem will not be dealt with in this paper.

Certain versions of Theorems 2.2–2.5 can be obtained for some classes of dynamical systems such as mixing subshifts of finite type and C^2 hyperbolic and expanding transformations but in order not to interrupt probabilistic exposition here we discuss some of these results in the last Section 5.

In the next section we will show that the study of large deviations in our non-conventional setup can be always reduced to the case $k = \ell$, i.e. we have to deal only with $q_j(n) = jn$, $j = 1, ..., k$. So we discuss next this situation allowing any $k \geq 1$ while assuming that $X(n)$, $n \geq 0$, q_j and F are the same as in Theorem 2.1. It turns out that the treatment of the general case when $X(0), X(1), X(2), ...$ is a Markov chain requires a quite complicated and technical proof whose exposition here would make this paper too long, and so it will be discussed in another paper. Thus, we will restrict ourselves here to a particular case when $X(n)$, $n \geq 0$ are independent identically distributed (i.i.d.) random variables (or vectors). Namely, we are interested in large deviations estimates for $S_N(F) = \sum_{n=1}^N F(X(n), X(2n), ..., X(kn))$ where $X(n) \in M$, $n \geq 1$ are i.i.d. random variables (vectors) with a compact support M. Let $r_1, ..., r_m \geq 2$ be all primes not exceeding k. Set $A_n = \{a \leq n : a \text{ is relatively prime with } r_1, ..., r_m\}$ and $B_\eta(a) = \{b \leq \eta : b = ar_1^{d_1} r_2^{d_2} \cdots r_m^{d_m} \text{ for some nonnegative integers } d_1, ..., d_m\}$. Now for any bounded measurable function V on M^k we write

(2.36)
$$S_N(V) = \sum_{a \in A_N} S_{N,a}(V) \text{ with } S_{N,a}(V) = \sum_{b \in B_N(a)} V(X(b), X(2b), ..., X(kb)).$$

Observe that $S_{N,a}(V)$, $a \in A_V$ is a collection of independent random variables.

2.7. **Theorem.** *For any continuous function V on M^k the limit*

(2.37) $Q(V) = \lim_{N \to \infty} \frac{1}{N} \ln E \exp\left(\sum_{n=1}^N V(X(n), X(2n), ..., X(kn))\right)$

$\qquad\qquad = \lim_{N \to \infty} \frac{1}{N} \sum_{a \in A_N} \ln E \exp S_{N,a}(V)$

exists and the functional $Q(V)$ is convex and lower semicontinuous. If $V = V_\lambda$ depends on a parameter λ and has a bounded derivative in λ then $Q(V_\lambda)$ is also differentiable in λ. Thus taking $V_\lambda = \lambda F$ we obtain that also for $k \geq 2$ in the above i.i.d. setup both upper and lower large deviations bounds (2.9) and (2.10) hold true with the rate functional J being the Fenchel-Legendre transform $J(u) = \sup_\lambda(\lambda u - Q(\lambda F))$ of Q.

In Section 4 we will provide a rather explicit computation of the limit (2.37). As a model application of Theorem 2.7 we can consider digits $X(n) = X(n, \omega)$, $n \geq 1$ of base M expansions $\omega = \sum_{n=1}^\infty \frac{X(n,\omega)}{M^n}$, $X(n, \omega) \in \{0, 1, ..., M-1\}$ of numbers $\omega \in [0, 1)$ which are i.i.d. random variables on the probability space $([0, 1), \mathcal{B}, P)$ where $\mathcal{B}$ is the Borel σ-algebra and P is the Lebesgue measure. Take, for instance, $V(x_1, ..., x_k) = \delta_{\alpha_1 x_1} \delta_{\alpha_2 x_2} \cdots \delta_{\alpha_k x_k}$ for some $\alpha_1, ..., \alpha_k \in \{0, 1, ..., M-1\}$ with $\delta_{ij} = 1$ if $i = j$ and $= 0$, otherwise. Then Theorem 2.7 provides large deviations estimates

for the number

$$(2.38) \qquad n_{\alpha_1,\ldots,\alpha_k}(N,\omega) = \#\{n \le N : X(n,\omega) = \alpha_1, X(2n,\omega) = \alpha_2,$$
$$\ldots, X(kn,\omega) = \alpha_k\} = \sum_{n=1}^{N} V(X(n,\omega),\ldots,X(kn,\omega)).$$

The same setup can be reformulated in the following way. Consider infinite sequences of letters (colors, spins, etc.) taken out of an alphabet of size M. Let $n_{\alpha_1,\ldots,\alpha_k}(N)$ be the number of arithmetic progressions of length k with both the first term and the difference equal $n \le N$ and having the letter (color, spin, etc.) α_i on the place $i = 1, 2, \ldots, k$. Then Theorem 2.7 yields large deviations bounds for $n_{\alpha_1,\ldots,\alpha_k}(N)$ as $N \to \infty$ considered as a random variable on the space of sequences of letters with any product probability measure, in particular, with uniform probability measure which assigns the same weight to each combination of n consecutive letters (i.e. to each cylinder set of length n) for all $n = 1, 2, \ldots$. We observe that another statistical physics interpretation of a particular case of the above i.i.d. setup appeared independently in a recent paper [6] though large deviations bounds were obtained there only for the case $k = M = 2$.

3. LARGE DEVIATIONS FOR MARKOV PROCESSES: $k = 1$ CASE

3.1. **Reduction to the $k = \ell$ case.** First, we will show that the study of the limit (2.6) for any $k \le \ell$ can be reduced to the case $k = \ell$. In order to apply this result not only to Markov chains but also to other fast mixing processes, in particular to dynamical systems considered in Section 5, we will deal here with a somewhat more general setup.

Let $\{X(n), n = 0, 1, \ldots\}$ be a sequence of measurable mappings of a measurable space $(\Omega, \mathcal{F})$ to a Polish space M considered with its Borel σ-algebra $\mathcal{B}$. Since $(M, \mathcal{B})$ is isomorphic to a Borel subset Υ of an interval we can and do identify M with Υ and assume that each $X(n)$ is real (or vector) valued. Then $\{X(n), n = 0, 1, \ldots\}$ becomes a real (or vector) valued stochastic process under each probability measure on $(\Omega, \mathcal{F})$. Our setup includes two such measures P and Π while we assume that $X(n)\Pi = \mu$ does not depend on n, i.e. that the one dimensional distribution μ of $X(n)$ on the probability space $(\Omega, \mathcal{F}, \Pi)$ is the same for all n. In order to state our conditions we introduce also a family of σ-algebras $\mathcal{F}_{ml} \subset \mathcal{F}$, $-\infty \le m \le l \le \infty$ satisfying $\mathcal{F}_{-\infty,\infty} = \mathcal{F}$ and $\mathcal{F}_{ml} \subset \mathcal{F}_{m'l'}$ if $m' \le m$ and $l' \ge l$. Next, we define a modified ψ-mixing (dependence) coefficient by

$$\psi(n) = \psi_{P,\Pi}(n) = \sup_{l \ge 0, g}\{\|E_P(g|\mathcal{F}_{-\infty,l}) - E_\Pi g\|_\infty :$$
$$g \text{ is } \mathcal{F}_{l+n,\infty} - \text{measurable and } E_\Pi|g| \le 1\}$$

where E_Q is the expectation with respect to a probability measure Q and $\|\cdot\|_\infty$ is the $L^\infty(\Omega, P)$ norm. The rational behind introduction of two probability measures P and Π above is to allow $X(n), n \ge 0$ to be a Markov chain with an arbitrary initial distribution (in particular, starting at a point) under P while $X(n)$ is stationary under Π and the distribution of $X(n)$ under P converges to $\mu = X(0)\Pi$. Furthermore, we will not assume measurability of $X(n)$'s with respect to some of σ-algebras $\mathcal{F}_{m,l}$ but instead will rely on approximation coefficients defined for each bounded continuous function $V = V(x_1, \ldots, x_\ell)$ on M^ℓ by

$$\beta_V(n) = \max_{1 \le j \le \ell} \sup_{x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_\ell \in M} \sup_{m \ge 0} \|V(x_1,\ldots,x_{j-1},$$
$$X(m), x_{j+1}, \ldots, x_\ell) - V(x_1, \ldots, x_{j-1}, E_P(X(m)|\mathcal{F}_{m-n,m+n}), x_{j+1}, \ldots, x_\ell)\|_\infty.$$

Since V is continuous we can take here the supremum over a countable dense set in $M^{\ell-1}$, and so outside of one P-measure zero set $\beta_V(n)$ gives a uniform bound of the difference above.

3.1. Proposition. *Let $V(x_1, ..., x_\ell)$ be a bounded continuous function on M^ℓ and assume that*

$$
(3.1) \qquad \lim_{n \to \infty} (\psi(n) + \beta_V(n)) = 0
$$

together with the conditions (2.3) and (2.4) on functions q_j, $j = 1, ..., \ell$. Then,

$$
(3.2) \qquad \lim_{N \to \infty} \tfrac{1}{N} \Big(\ln E_P \exp \big(\sum_{n=1}^N V(X(q_1(n)), ..., X(q_\ell(n))) \big)
$$
$$
- \ln E_P \exp \big(\sum_{n=1}^N V^{(k)}(X(n), X(2n), ..., X(kn)) \big) \Big) = 0
$$

where for each $m < \ell$,

$$
(3.3) \qquad V^{(m)}(x_1, ..., x_m) = \ln \int_M ... \int_M \exp(V(x_1, ..., x_m, x_{m+1}, ..., x_\ell))
$$
$$
d\mu(x_{m+1})...d\mu(x_\ell) \ and \ V^{(\ell)} = V.
$$

If, in fact, $X(n)$ is $\mathcal{F}_{n,n}$-measurable then (3.2) holds true for any bounded measurable function V assuming only that $\psi(n) \to 0$ as $n \to \infty$.

Proof. Observe that (2.4) yields

$$
(3.4) \qquad \lim_{n \to \infty} (q_j(\gamma n) - q_{j-1}(n)) = \infty \ \text{for any} \ j > k \ \text{and} \ \gamma > 0.
$$

Set

$$
d_\gamma(n) = \min_{k+1 \le j \le \ell} \min \big(q_j(\gamma n) - q_{j-1}(n), \min_{l \ge \gamma n}(q_j(l) - q_j(l-1)) \big)
$$

and observe that $d_\gamma(n) \to \infty$ as $n \to \infty$ in view of (2.4) and (3.4). For any $l = 0, 1, ...$ and $0 \le r \le \infty$ set

$$
X_r(l) = E_P(X(l)|\mathcal{F}_{l-r,l+r}).
$$

Next, for $m = 1, 2, ..., \ell$, $a \le b \le c$ and $0 \le r \le \infty$ denote

$$
Z_r^{(m)}(a, b, c) = E_P \exp \big(\sum_{a < l \le b} V^{(m)}(X_r(q_1(l)), ..., X_r(q_m(l)))
$$
$$
+ \sum_{b < l \le c} V^{(m-1)}(X_r(q_1(l)), ..., X_r(q_{m-1}(l))) \big).
$$

If $b = c$, i.e. we have only the first sum above, we set $Z_r^{(m)}(a, b, c) = Z_r^{(m)}(a, b)$. If $r = \infty$ we drop the index r and write just $Z^{(n)}(a, b, c)$ or $Z^{(m)}(a, b)$. Observe that

$$
(3.5) \qquad e^{-C(V)\gamma N} Z_r^{(m)}(\gamma N, N) \le Z_r^{(m)}(0, N) \le e^{C(V)\gamma N} Z_r^{(m)}(\gamma N, N)
$$

where $C(V) = \sup_{(x_1, ..., x_\ell) \in M^\ell} |V(x_1, ..., x_\ell)|$. By the definition of $\beta_V(n)$ (and the remark after it) we obtain also that for any $m = 1, 2, ..., \ell$, $a \le b \le c$ and $0 \le r \le \infty$,

$$
(3.6) \qquad Z_r^{(m)}(a, b, c)e^{-(c-a)\ell\beta_V(r)} \le Z^{(m)}(a, b, c) \le Z_r^{(m)}(a, b, c)e^{(c-a)\ell\beta_V(r)}.
$$

Let $g = g(x, y)$ be a bounded measurable function on a product $M_1 \times M_2$ (for some measurable spaces $(M_1, \mathcal{B}_1)$ and $(M_2, \mathcal{B}_2)$) and $X : \Omega \to M_1$ and $Y : \Omega \to M_2$ be $\mathcal{F}_{-\infty,l}-$ and $\mathcal{F}_{l+n,\infty}-$measurable random variables (maps), respectively. Then it follows from the definition of $\psi(n) = \psi_{P,\Pi}(n)$ that

$$
(3.7) \qquad |E_P(g(X, Y)|\mathcal{F}_{-\infty,l}) - g_\Pi(X)| \le \psi(n)|g|_\Pi(X)
$$

where $g_\Pi(x) = E_\Pi g(x, Y)$ and $|g|_\Pi(x) = E_\Pi |g(x, Y)|$. Now take $r = r_\gamma(N) = [\frac{1}{3}d_\gamma(N)]$ where $[\cdot]$ denotes the integral part. Then for all $N \geq n \geq \gamma N + 1$, $m = k+1, ..., \ell$ and N large enough

$$(3.8)\ Z_r^{(m)}(\gamma N, n, N) = E_P\big(J \exp\big(\sum_{\gamma N < l \leq n-1} V^{(m)}(X_r(q_1(l)), ..., X_r(q_m(l)))$$
$$+ \sum_{n < l \leq N} V^{(m-1)}(X_r(q_1(l)), ..., X_r(q_{m-1}(l))))\big)$$

where

$$J = J_r(n) = E_P\big(\exp\big(V^{(m)}(X_r(q_1(n)), ..., X_r(q_m(n)))\big)\big|\mathcal{F}_{-\infty, q_m(n-1)+r}\big).$$

By (3.7) and the definition of β_V we conclude that

$$(3.9) \qquad \big|J - \int \exp\big(V^{(m)}(X_r(q_1(n)), ..., X_r(q_{m-1}(n)), y)\big) d\mu(y)\big|$$
$$\leq \eta(r) \int \exp\big(V^{(m)}(X_r(q_1(n)), ..., X_r(q_{m-1}(n)), y)\big) d\mu(y)$$

where $\eta(n) = (\psi(n) + 2\beta_V(n) + 2\beta_V(n)\psi(n))e^{C(V)} \to 0$ as $n \to \infty$. Employing (3.8) and (3.9) for $n = N, N-1, ..., [\gamma N] + 1$ we obtain that

$$(3.10)\ (1 - \eta(r))^N Z_r^{(m-1)}(\gamma N, N) \leq Z_r^{(m)}(\gamma N, N) \leq (1 + \eta(r))^N Z_r^{(m-1)}(\gamma N, N).$$

Next, we use (3.10) for $m = \ell, \ell - 1, ..., k+1$ which together with (3.5) and (3.6) yields that

$$(3.11) \qquad (1 - \eta(r))^{\ell N} e^{-2N(\ell \beta_V(r) + C(V)\gamma)} Z^{(k)}(0, N) \leq Z^{(\ell)}(0, N)$$
$$\leq (1 + \eta(r))^{\ell N} e^{2N(\ell \beta_V(r) + C(V)\gamma)} Z^{(k)}(0, N).$$

Taking ln in (3.11), dividing by N, letting $N \to \infty$ and taking into account that then $r = r(N) \to \infty$, we obtain that

$$\limsup_{N \to \infty} \frac{1}{N}\big|\ln Z^{(\ell)}(0, N) - \ln Z^{(k)}(0, N)\big| \leq 2C(V)\gamma$$

and (3.2) follows since $\gamma > 0$ is arbitrary.

If $X(n)$ is $\mathcal{F}_{n,n}$-measurable for each n then we do not have to deal with the approximation coefficient $\beta_V(r)$ and $X_r = X$, $Z_r^{(m)} = Z^{(m)}$ above. Hence all above arguments remain true with $\beta_V(r) = 0$ for any bounded measurable V and we obtain (3.2) provided $\psi(n) \to 0$ as $n \to \infty$. $\qquad\square$

It is easy to check the conditions of Proposition 3.1 for Markov chains $X(n)$, $n \geq 0$ satisfying the "strong" Doeblin condition (2.1). Indeed, denote by $\mathcal{F}_{l,m}$, $l \leq m$ the σ-algebra generated by $X(l), ..., X(m)$ with $\mathcal{F}_{l,\infty}$ being the minimal σ-algebra containing all $\mathcal{F}_{l,m}$, $m \geq l$ and we set $\mathcal{F}_{l,m} = \mathcal{F}_{0,m}$ for $l < 0$ and $m \geq 0$. If g is $\mathcal{F}_{l+n,\infty}$−measurable then by the Markov property

$$(3.12) \qquad\qquad E_P(g|\mathcal{F}_{-\infty,l}) = \int P(n, X(l), dy) E_{P_y} g$$

where P_y is the probability measure on the path space of the Markov chain $X(n)$ starting at y. The Chapman-Kolmogorov equation sais that for any $n \geq n_0$,

$$P(n, x, G) = \int P(n - n_0, x, dy) P(n_0, y, G),$$

and so by (2.1) for all such n,

$$C^{-1}\nu(G) \leq P(n, x, G) \leq C\nu(G).$$

This together with the Radon-Nikodim theorem yields existence for ν-almost all y and $n \geq n_0$ of the transition density $p(n, x, y)$ satisfying

$$C^{-1} \leq p(n, x, y) = \frac{dP(n, x, \cdot)}{d\nu}(y) \leq C.$$

It is well known (see, for instance, [8]) that (2.1) and (2.2) imply that

(3.13) $$(1 - Ke^{-\kappa n})p(y) \leq p(n, x, y) \leq (1 + Ke^{-\kappa n})p(y)$$

for some $K, \kappa > 0$ independent of $n \geq n_0$. If Π is the stationary probability of the Markov chain on the path space then

$$E_\Pi g = \int p(y)E_{P_y}g d\nu(y).$$

Hence, by (3.12) and (3.13),

$$\|E_P(g|\mathcal{F}_{-\infty,l}) - E_\Pi g\|_\infty \leq Ke^{-\kappa n}E_\Pi|g|.$$

Thus the condition (3.1) with $\beta_V(n) = 0$ is satisfied in our Markov chains case.

3.2. **Corollary.** *Assume that conditions of Proposition 3.1 hold true. Suppose that for any bounded measurable function $V_\lambda(x_1, ..., x_k)$ on $\mathbb{R} \times M^k$ having a bounded in $x_1, ..., x_k$ derivative in a parameter $\lambda \in (-\infty, \infty)$ the limit*

$$Q(V_\lambda) = \lim_{N \to \infty} \frac{1}{N} \ln E_x \exp\left(\sum_{n=1}^N V_\lambda(X(n), X(2n), ..., X(kn))\right)$$

exists, it is a lower semicontinuous convex functional and it is differentiable in the parameter λ. Then for any bounded measurable function $W_\lambda(x_1, ..., x_\ell)$ on $\mathbb{R} \times M^\ell$ having a bounded in $x_1, ..., x_\ell$ derivative in a parameter $\lambda \in (-\infty, \infty)$ the limit

$$Q(W_\lambda) = \lim_{N \to \infty} \frac{1}{N} \ln E_x \exp\left(\sum_{n=1}^N W_\lambda(X(q_1(n)), , ..., X(q_\ell(n)))\right) = Q(W_\lambda^{(k)})$$

exists, it is a lower semicontinuous convex functional and it is differentiable in the parameter λ. In particular, the large deviations estimates in the form (2.9) and (2.10) hold true then with the rate functional J given by (2.8) with $W_\lambda = \lambda F$.

Proof. By Proposition 3.1, $Q(W_\lambda) = Q(W_\lambda^{(k)})$ and we see from (3.3) that if W_λ is bounded and has a bounded derivative in λ then so does $W_\lambda^{(k)}$. Hence, by the assumption $Q(W_\lambda^{(k)})$ is a lower semicontinuous convex functional and it is differentiable in λ which implies the same for $Q(W_\lambda)$ and the result follows. $\square$

Now let $k = 1$ and $V = W_\lambda$ as in Theorem 2.1. Then $\hat{W}_\lambda = V^{(1)}$ and by Proposition 3.1,

(3.14) $$Q(W_\lambda) = \lim_{N \to \infty} \frac{1}{N} \ln E_x \exp\left(\sum_{n=1}^N \hat{W}_\lambda(X(n))\right).$$

Thus we arrive at the standard limit appearing in "conventional" large deviations results which is well known for Markov chains $X(n)$, $n \geq 0$ satisfying our conditions as it is described in Theorem 2.1. Differentiability of $Q(W_\lambda)$ in λ follows from standard results on positive operatos (see, for instance, [20]) and we derive now Theorem 2.1 from well known "conventional" large deviations results (see, for instance, [9], [16] and Section 2.3 in [11]). $\square$

3.2. **2nd level of large deviations.** Recall that in the setup of Theorem 2.2 we have $k = 1$, M being a compact space and the result is about large deviations for occupational measures ζ_N appearing there. Let W be a continuous function on M^ℓ with $\hat{W}$ defined by (2.5). By Proposition 3.1 together with the well known facts (see, for instance, [9], [11] and [16]),

$$(3.15) \quad Q(W) = \lim_{N \to \infty} \frac{1}{N} \ln E_x \exp \Big(\sum_{n=1}^{N} W(X(q_1(n)), ..., X(q_\ell(n))) \Big) = \ln(r(W))$$

where $r(W)$ is the spectral radius of the operator

$$(3.16) \qquad R(W)g(x) = E_x\big(g(X(1))\hat{W}(X(1))\big) = E_x\big(g(X(1))e^{\ln \hat{W}(X(1))}\big).$$

Observe, that by the Donsker–Varadhan variational formula (see [9] and [10]),

$$(3.17) \qquad\qquad Q(W) = \sup_{\nu \in \mathcal{P}(M)} \Big(\int_M \ln \hat{W}(x) d\nu(x) - \hat{I}(\nu) \Big)$$

where $\hat{I}(\nu) = -\inf_{u \in C_+(M)} \int \ln \frac{E_x u(X(1))}{u(x)} d\nu(x)$ and the infimum is taken over positive continuous functions on M.

Next, let $Y^{(i)}(n)$, $i = 2, ..., \ell$; $n = 0, 1, 2, ...$ be i.i.d. M-valued random variables with the distribution μ, all of them independent of the Markov chain $X(n)$, $n \geq 0$. Then it is easy to see that

$$(3.18) \qquad \lim_{N \to \infty} \frac{1}{N} \ln E_x \exp \Big(\sum_{n=1}^{N} W(X_n, Y^{(2)}(n), ..., Y^{(\ell)}(n)) \Big) = Q(W).$$

Indeed, let $\mathcal{F}_X$ be the σ-algebra generated by the Markov chain $X(n), n \geq 0$. Then

$$(3.19) \qquad E_x \exp \Big(\sum_{n=1}^{N} W(X(n), Y^{(2)}(n), ..., Y^{(\ell)}(n)) \Big)$$
$$= E_x\Big(E_x(\exp(\sum_{n=1}^{N} W(X(n), Y^{(2)}(n), ..., Y^{(\ell)}(n)))|\mathcal{F}_X)\Big)$$
$$= E_x \exp(\sum_{n=1}^{N} \ln \hat{W}(X(n)))$$

and (3.18) follows. But now we have the standard situation for the Markov chain $(X(n), Y^{(2)}(n), ..., Y^{(\ell)}(n))$, $n \geq 0$, and so by the Donsker–Varadhan variational formula (see [9] and [10]),

$$(3.20) \qquad Q(W) = \sup_{\nu \in \mathcal{P}(M \times \cdots \times M)} \Big(\int W(x_1, x_2, ..., x_\ell) d\nu(x_1, ..., x_\ell) - I(\nu) \Big)$$

where

$$(3.21) \qquad\qquad I(\nu) = -\inf_{u \in C_+(M \times \cdots \times M)} \int_{M \times \cdots \times M}$$
$$\ln \frac{E_{x_1} \int u(X(1), x_1, ..., x_\ell) d\mu(x_2)...d\mu(x_\ell)}{u(x_1, ..., x_\ell)} d\nu(x_1, ..., x_\ell).$$

It is known here (see, for instance, Proposition 5.1 in [15]) that there exists a unique $\nu = \nu_W$ on which the supremum in (3.20) is attained and it follows from the standard theory (see, for instance, [16]) that $I(\nu)$ is the rate functional for the second level large deviations both for the auxiliary occupational measures

$$\frac{1}{N} \sum_{n=1}^{N} \delta_{(X_n, Y_n^{(2)}, ..., Y_n^{(\ell)})}$$

and for our nonconventional occupational measures ζ_N. $\qquad\qquad\qquad\square$

3.3. Continuous time case. Similarly to the discrete time case, the main step in the proof of Theorem 2.3 is to establish (2.19) and to identify the limit there as the spectral radius of the semigroup (2.20).

From (2.16) it follows that for any $t \geq t_0$ and every measurable set $G \subset M$,

$$(3.22) \qquad P(t,x,G) = \int_G p(t,x,y)d\nu(y) \text{ with } C^{-1} \leq p(t,x,y) \leq C.$$

Furthermore, similarly to (3.5) (see [8]),

$$(3.23) \qquad\qquad (1 - Ke^{-\kappa t})p(y) \leq p(t,x,y) \leq (1 + Ke^{-\kappa t})p(y)$$

where $p(y) = \frac{d\mu}{d\nu}(y)$ is the density of the unique invariant measure μ of the Markov process X. Observe that (2.18) implies also that for any $j \geq k+1$ and $\gamma > 0$,

$$(3.24) \qquad\qquad \lim_{t \to \infty}(q_j(\gamma t) - q_{j-1}(t)) = \infty.$$

Let $V = V(x_1, ..., x_\ell)$ be a bounded measurable function on M^ℓ and for $m = 1, 2, ..., \ell$ set

$$V_{\text{cont}}^{(m)}(x_1, ..., x_m) = \int_M ... \int_M V(x_1, ..., x_m, x_{m+1}, ..., x_\ell)d\mu(x_{m+1})...d\mu(x_\ell)$$

with $V_{\text{cont}}^{(\ell)} = V$. Set $t_n(\gamma, T) = \gamma T + n(\gamma + \gamma^2)$ for $n = 0, 1, 2, ..., M(\gamma, T) - 1$ where $M(\gamma, T) = [(T(1-\gamma)/(\gamma + \gamma^2)]$. Next, for $a \leq b \leq c$ and $m = 1, 2, ..., \ell$ we denote

$$Z_x^{(m)}(a,b,c) = E_x \exp\big(\sum_{a \leq n < b} \int_{t_n(\gamma,T)}^{t_n(\gamma,T)+\gamma} V_{\text{cont}}^{(m)}(X(q_1(t)), ..., X(q_m(t)))dt$$

$$+ \sum_{b \leq n < c} \int_{t_n(\gamma,T)}^{t_n(\gamma,T)+\gamma} V_{\text{cont}}^{(m-1)}(X(q_1(t)), ..., X(q_{m-1}(t)))dt \big).$$

It is easy to see that

$$(3.25) \qquad \exp(-2C(V)\gamma T)Z^{(\ell)}(0, M(\gamma,T)) \leq E_x \exp\big(\int_0^T V(X(q_1(t)),$$

$$..., X(q_\ell(t)))dt\big) \leq \exp(2C(V)\gamma T)Z^{(\ell)}(0, M(\gamma,T))$$

where $Z_x^{(m)}(a,b) = Z_x^{(m)}(a,b,b)$ and $C(V) = \sup_{(x_1,...,x_\ell)} |V(x_1, ..., x_\ell)|$.

Denote by $\mathcal{F}_t$ the σ-algebra generated by $X(s), s \leq t$. Then by (2.18) and (3.24) for all T large enough if $n \geq 1, T \geq t \geq t_n(\gamma, T)$ and $s \leq t_{n-1}(\gamma, T) + \gamma$ then $X(q_1(s)), ..., X(q_m(s))$ and $X(q_1(t)), ..., X(q_{m-1}(t))$ are $\mathcal{F}_{q_m(t_n(\gamma,T)-\gamma^2)}$–measurable. Hence,

$$(3.26) \qquad Z_x^{(m)}(0, n, M(\gamma,T)) = E_x\big(J_{m,n} \exp\big(\sum_{0 \leq l < n} \int_{t_l(\gamma,T)}^{t_l(\gamma,T)+\gamma}$$

$$V_{\text{cont}}^{(m)}(X(q_1(s)), ..., X(q_m(s)))ds + \sum_{n+1 \leq l < M(\gamma,T)} \int_{t_l(\gamma,T)}^{t_l(\gamma,T)+\gamma}$$

$$V_{\text{cont}}^{(m-1)}(X(q_1(s)), ..., X(q_{m-1}(s)))ds\big)\big)$$

where

$$J_{m,n} = E_x\big(\exp\big(\int_{t_n(\gamma,T)}^{t_n(\gamma,T)+\gamma} V_{\text{cont}}^{(m)}(X(q_1(s)), ..., X(q_m(s)))ds\big) \big| \mathcal{F}_{q_m(t_n(\gamma,T)-\gamma^2)}\big).$$

Let

$$\tilde{J}_{m,n} = \exp\big(\int_{t_n(\gamma,T)}^{t_n(\gamma,T)+\gamma} E_x\big(V_{\text{cont}}^{(m)}(X(q_1(s)), ..., X(q_m(s)))\big| \mathcal{F}_{q_m(t_n(\gamma,T)-\gamma^2)}\big)ds\big).$$

Since $|e^\alpha - 1 - \alpha| \leq \alpha^2$ if $|\alpha| \leq 1$ then

$$(3.27) \qquad\qquad |J_{m,n} - \tilde{J}_{m,n}| \leq 2\gamma^2(C(V))^2.$$

On the other hand, by the Markov property

$$(3.28) \qquad \tilde{J}_{m,n} = \exp\Big(\int_{t_n(\gamma,T)}^{t_n(\gamma,T)+\gamma} ds \int_M p\big(q_m(s) - q_m(t_n(\gamma,T) - \gamma^2),$$

$$X(q_m(t_n(\gamma,T) - \gamma^2)), y\big) V_{\text{cont}}^{(m)}(X(q_1(s)), ..., X(q_{m-1}(s)), y) d\nu(y)\Big).$$

Set

$$d_\gamma(t) = \inf_{s \geq \gamma t} \min_{k+1 \leq j \leq \ell} \min(q_j(s) - q_{j-1}(s\gamma^{-1}), \, q_j(s) - q_j(s - \gamma^2))$$

and observe that $d_\gamma(t) \to \infty$ as $t \to \infty$ for each fixed $\gamma > 0$ in view of the assumption (2.18). Now, by (3.23) and (3.28),

$$(3.29) \; \exp(-Ke^{-\kappa d_\gamma(T)}C(V)\gamma) \leq J_{m,n}\exp\Big(-\int_{t_n(\gamma,T)}^{t_n(\gamma,T)+\gamma} V_{\text{cont}}^{(m-1)}(X(q_1(s)),$$

$$..., X(q_{m-1}(s)))ds\Big) \leq \exp(Ke^{-\kappa d_\gamma(T)}C(V)\gamma).$$

Employing (3.26)–(3.29) for $n = M(\gamma,T), M(\gamma,T) - 1, ..., 1$ with each $m = \ell, \ell - 1, ..., k+1$ we obtain that

$$(3.30) \qquad \limsup_{T\to\infty} \frac{1}{T}\Big|\ln\big(Z_x^{(\ell)}(0, M(\gamma,T))\big) - \ln\big(Z_x^{(k)}(0, M(\gamma,T))\big)\Big| = 0.$$

Now taking ln in (3.25) and letting first $T \to \infty$ and then $\gamma \to 0$ we obtain from (3.30) and the definition of $Z_x^{(k)}$ that

$$(3.31) \qquad \lim_{T\to\infty} \tfrac{1}{T}\Big(\ln E_x \exp\big(\int_0^T V(X(q_1(t)), ..., X(q_\ell(t)))dt\big)$$

$$- \ln E_x \exp\big(\int_0^T V_{\text{cont}}^{(k)}(X(\alpha_1 t), ..., X(\alpha_k t))dt\big)\Big) = 0.$$

If $k = 1$ then $\frac{1}{T}$ of the second expression in brackets in (3.31) converges as $T \to \infty$ to the logarithm of the spectral radius of the semigroup of operators $R_{\text{cont}}^t(V)$ defined in (2.20). Thus, the assertions of Theorems 2.3 and 2.4 follow from the well known results on large deviations (see [9], [10], [16] and [11]) in the same way as in the discrete time case. $\qquad\square$

3.4. **Nonconventional averaging.** According to [12] the large deviations estimates (2.34) and (2.35) follow once we establish (2.33) for all continuous functions $W_t(x_1, ..., x_\ell)$ on $\mathbb{R}_+ \times M^\ell$. First, we claim that even without the assumption $k = 1$,

$$(3.32) \qquad \lim_{\varepsilon\to 0} \varepsilon\big(\ln E_x \exp\big(\varepsilon^{-1}\int_0^T W_t(X(q_1(t/\varepsilon)), ..., X(q_\ell(t/\varepsilon)))dt\big)$$

$$- \ln E_x \exp\big(\varepsilon^{-1}\int_0^T W_t^{(k)}(X(q_1(t/\varepsilon)), ..., X(q_k(t/\varepsilon)))dt\big)\big) = 0$$

where in the discrete time case q_j's are extended to all $s \geq 0$ by writing $q_j(s) = q_j([s])$ and we set

$$W_t^{(k)}(x_1, ..., x_k) = \ln \int_M ... \int_M \exp(W_t(x_1, ..., x_\ell))d\mu(x_{k+1})...d\mu(x_\ell)$$

while in the continuous time case we set

$$W_t^{(k)}(x_1, ..., x_k) = \int_M ... \int_M W_t(x_1, ..., x_\ell)d\mu(x_{k+1})...d\mu(x_\ell).$$

The proof of (3.32) is the same as the proofs of (3.1) in the discrete time case and of (3.31) in the continuous time case while the dependence of W_t on t does not play any role in the arguments employed there.

Next, when $k = 1$ we arrive at the "conventional" setup and (2.33) follows in the same way as in [12] (see also [17]). $\qquad\square$

4. Large deviations for any $k \geq 1$: i.i.d. case

Here we assume that $X(n)$, $n \geq 1$ are i.i.d. random variables (vectors) and rely on the decomposition (2.36). In view of independency of $S_{N,a}(V)$ for different $a \in A_N$ we can write

$$(4.1) \qquad Z_N(V) = E \exp \Big(\sum_{n=1}^{N} V(X(n), X(2n), ..., X(kn)) \Big) = \prod_{a \in A_N} Z_{N,a}(V)$$

where

$$Z_{\eta,a}(V) = E \exp \Big(\sum_{b \in B_\eta(a)} V(X(b), X(2b), ..., X(kb)) \Big)$$

with A_N and $B_\eta(a)$ defined in Section 2.

In order to study $Z_{N,a}(V)$ we introduce also

$$B(a) = \{b \geq 1 : b = ar_1^{d_1} r_2^{d_2} \cdots r_m^{d_m} \text{ for some nonnegative integers } d_1, ..., d_m\}.$$

Observe that each $l = 1, 2, ..., k$ can be written uniquely in the form $l = r_1^{d_1(l)} r_2^{d_2(l)} \cdots r_m^{d_m(l)}$ for some nonnegative integers $d_1(l), ..., d_m(l)$. Now, if $b = ar_1^{d_1} \cdots r_m^{d_m} \in B(a)$ and $l = 1, 2, ..., k$ then $lb = ar_1^{d_1 + d_1(l)} \cdots r_m^{d_m + d_m(l)} \in B(a)$. Next, consider the lattice $\mathbb{Z}^m$ and set

$$\mathbb{Z}_+^m = \{n = (n_1, ..., n_m), \, n_i \geq 0 \text{ for all } i = 1, ..., m\}.$$

Then the formula $\varphi_a(n_1, ..., n_m) = ar_1^{n_1} \cdots r_m^{n_m}$ provides a one-to-one correspondence

$$\varphi_a : \mathbb{Z}_+^m \to B(a)$$

where, recall, a is relatively prime with $r_1, ..., r_m$. Set

$$D(\rho) = \{n = (n_1, ..., n_m) \in \mathbb{Z}^m : n_i \geq 0, \, i = 1, ..., m \text{ and } \sum_{i=1}^{m} n_i \ln r_i \leq \rho\}.$$

Then, clearly,

$$(4.2) \qquad\qquad\qquad \varphi_a D(\ln(N/a)) = B_N(a).$$

It follows that

$$(4.3) \qquad\qquad |B_N(a)| \leq \prod_{i=1}^{m} \Big(1 + \frac{1}{\ln r_i} \ln \frac{N}{a}\Big) \leq \Big(1 + \frac{1}{\ln 2} \ln \frac{N}{a}\Big)^m$$

where $|\Gamma|$ denotes the cardinality of a set Γ. Hence

$$(4.4) \qquad\qquad\qquad a \leq N 2^{-(|B_N(a)|^{1/m} - 1)}.$$

Next, we claim that $Z_{N,a}(N)$ is determined only by $|B_N(a)|$ and not by N and a themselves. Indeed, since $|D(\rho)|$ is nondecreasing in ρ then it determines the set $D(\rho)$ itself, and so $|D(\ln(N/a))| = |B_N(a)| = |B_{N/a}(1)|$ determines the set $B_{N/a}(1)$ in view of (4.2). Set $\hat{B}_\eta(a) = B_\eta(a) \cup \{n : n = ln' \text{ for some } n' \in B_\eta(a) \text{ and } l = 2, 3, ..., k\}$. Then we can write

$$Z_{\eta,a}(V) = \int ... \int \exp \Big(\sum_{b \in B_\eta(a)} V(x_b, x_{2b}, ..., x_{kb}) \Big) \prod_{b' \in \hat{B}_\eta(a)} d\mu(x_{b'}).$$

It is easy to see from here that $Z_{\eta,a}(V) = Z_{\eta/a,1}(V)$ for any $\eta > 0$ and an integer $a \geq 2$ relatively prime with $r_1, ..., r_m$. Indeed, $Z_{\eta,a}(V)$ is determined by the labeled

directed graph $\Gamma_\eta(a)$ having $B_\eta(a)$ as its vertices and having arrows of $k-1$ types so that an arrow with a label $l = 2, 3, ..., k$ is drawn from $n \in B_\eta(a)$ to $n' \in B_\eta(a)$ if $n' = ln$. Clearly, the graphs $\Gamma_\eta(a)$ and $\Gamma_{\eta/a}(1)$ are isomorphic in the sense that there exists a one-to-one map $\varphi : B_\eta(a) \to B_{\eta/a}(1)$ such that if $n, n' \in B_\eta(a)$ and $n' = ln$ then $\varphi n, \varphi n' \in B_{\eta/a}(1)$ and $\varphi n' = l\varphi n$. Since $X(n)$, $n \geq 1$ are i.i.d., $Z_{\eta,a}(V)$ is determined, in fact, by the isomorphism class of $\Gamma_\eta(a)$ and not by $\Gamma_\eta(a)$ itself, and so $Z_{\eta,a}(V) = Z_{\eta/a,1}(V)$. Since $|B_N(a)|$ determines the set $B_{N/a}(1)$ we conclude that it determines $Z_{N,a}(V)$, as well, proving the claim.

Let $l = |B_N(a)|$ and set $R_l(V) = Z_{N,a}(V)$ since the latter depends only on l (and, of course, on V). Observe that

$$(4.5) \qquad \ln R_l(V) \leq lC(V)$$

where $C(V) = \sup_{x_1,...,x_k \in M} |V(x_1, ..., x_k)|$. Set $A_N^{(l)} = \{a \in A_N : |B_N(a)| = l\}$. By (4.4),

$$(4.6) \qquad |A_N^{(l)}| \leq N2^{-(l^{1/m}-1)}.$$

Observe that $|D(\rho)|$ is a nondecreasing right continuous piecewise constant function and since $r_1, r_2, ..., r_m$ are primes the jumps of $|D_N(\rho)|$ can only be of size 1, i.e. for all $\tilde{\rho} > 0$,

$$|D(\tilde{\rho})| - \lim_{\rho \uparrow \tilde{\rho}} |D(\rho)| \leq 1.$$

It follows that

$$\rho_{\min}(l) = \inf\{\rho \geq 0 : |D(\rho)| = l\} \text{ and } \rho_{\max}(l) = \sup\{\rho \geq 0 : |D(\rho)| = l\}$$

are well defined for each integer $l \geq 1$ and $\rho_{\min}(l) < \rho_{\max}(l)$. Denote $\hat{A}_N^{(l)} = \{a \in \mathbb{N} : Ne^{-\rho_{\max}(l)} \leq a \leq Ne^{-\rho_{\min}(l)}, a \text{ is relatively prime with } r_1, r_2, ..., r_m\}$. Then by (4.2) and the above,

$$(4.7) \qquad \frac{1}{N}\big||A_N^{(l)}| - |\hat{A}_N^{(l)}|\big| \leq \frac{1}{N} \to 0 \text{ as } N \to \infty.$$

We will show next that the limit

$$(4.8) \qquad \lim_{N\to\infty} \frac{1}{N}|\hat{A}_N^{(l)}| = (e^{-\rho_{\min}(l)} - e^{-\rho_{\max}(l)})r$$

exists with

$$(4.9) \quad r = 1 - \frac{1}{2} - \frac{1}{3} + \frac{1}{2 \cdot 3} - \frac{1}{5} + \frac{1}{2 \cdot 5} + \frac{1}{3 \cdot 5} - \frac{1}{2 \cdot 3 \cdot 5} + \cdots + (-1)^m \frac{1}{r_1 \cdot r_2 \cdots r_m}.$$

Indeed, for each integer $n \geq 1$ set $G(n) = \{in : i \in \mathbb{Z}_+\}$ and $G_N^{(l)}(n) = \{j \in G(n) : Ne^{-\rho_{\max}(l)} \leq j \leq Ne^{-\rho_{\min}(l)}\}$. Then (by the inclusion-exclusion principle),
(4.10)
$$|\hat{A}_N^{(l)}| = |G_N^{(l)}(1)| - |G_N^{(l)}(2)| - |G_N^{(l)}(3)| + |G_N^{(l)}(2 \cdot 3)| + \cdots + (-1)^m |G_N^{(l)}(r_1 \cdot r_2 \cdots r_m)|.$$

Since each $G(n)$ is an arithmetic progression with the difference n we obtain that

$$(4.11) \qquad \lim_{N\to\infty} \frac{1}{N}|G_N^{(l)}(n)| = \frac{1}{n}(e^{-\rho_{\min}(l)} - e^{-\rho_{\max}(l)})$$

and (4.8)–(4.9) follows from (4.10)–(4.11).

Observe that by (4.2), (4.3) and the definition of $\rho_{\min}$ and $\rho_{\max}$,

$$(4.12) \qquad \rho_{\max}(l) \geq \rho_{\min}(l) \geq (l^{1/m} - 1)\ln 2,$$

and so we obtain from (4.1) and (4.5)–(4.9) that

(4.13) $\frac{1}{N} \ln Z_N(V) = \frac{1}{N} \sum_{a \in A_N} \ln Z_{N,a}(V)$

$$= \frac{1}{N} \sum_{1 \le l \le (1+\frac{1}{\ln 2} \ln \frac{N}{a})^m} |A_N^{(l)}| \ln R_l(V)$$

$$\longrightarrow r \sum_{l=1}^{\infty} (e^{-\rho_{\min}(l)} - e^{-\rho_{\max}(l)}) \ln R_l(V) \text{ as } N \to \infty$$

while the last series converges absolutely in view of (4.5) and (4.12). Furthermore, if $V = V_\lambda$ depends on a parameter λ in a differentiable way with a derivative bounded by $\tilde{C}$ then each $\ln R_l(V_\lambda)$ is also differentiable in λ with a derivative bounded by $\tilde{C}l$. Hence, in this case we can differentiate in λ the series in the right hand side of (4.13) and the assertion of Theorem 2.7 follows. $\qquad\square$

5. Nonconventional large deviations for dynamical systems

In this section we discuss nonconventional large deviations results in the dynamical systems case and a reader which is not familiar with hyperbolic dynamical systems and is interested only in the probabilistic setup may skip this section altogether. We assume now that $T : M \to M$ is either a subshift of finite type or a C^2 expanding endomorphism or a hyperbolic diffeomorphism on a compact Riemannian manifold (see [3] and [21]). By the latter we mean a C^2 Anosov diffeomorphism or, more generally, a C^2 diffeomorphism defined in a neighborhood of a hyperbolic attractor. We identify now the probability space $(\Omega, \mathcal{F}, P)$ with $(M, \mathcal{B}, \mu)$ where $\mathcal{B}$ is the Borel σ-algebra on M and μ is a Gibbs T-invariant measure constructed by a Hölder continuous potential g (see [3] and [21]). Let $k = 1$ in (2.3), (2.4) and (2.17), (2.18).

5.1. **Theorem.** *Let $X(n) = X(n,\omega) = X(n,x) = f(T^n x)$, $n \ge 0$, where f is a Hölder continuous (vector) function, and we take also q_j's as in Theorem 2.1. Let $k = 1$ then for any $W_\lambda = W_\lambda(x_1, ..., x_\ell)$ continuous in $x_1, ..., x_\ell$,*

(5.1) $Q(W_\lambda) = \lim_{N \to \infty} \frac{1}{N} \ln \int_M \exp \left(\sum_{n=1}^N W_\lambda(T^{q_1(n)}x, ..., T^{q_\ell(n)}x) \right) d\mu(x)$

$$= \mathfrak{P}(\ln \hat{W}_\lambda + g)$$

with $\hat{W}$ defined by (2.5), g being the potential of μ and $\mathfrak{P}(\cdot)$ being the topological pressure of a function in brackets for the transformation T (see [3] and [21]). If the derivative $dW_\lambda/d\lambda$ exists and is bounded in $x_1, ..., x_\ell$ for each λ then $Q(W_\lambda)$ is differentiable in λ, as well. In the expanding and hyperbolic cases the limit in (5.1) remains the same if we integrate in (5.1) either with respect to the normalized Riemannian volume or with respect to the Sinai-Ruelle-Bowen (SRB) measure $\mu = \mu^{SRB}$ which is the Gibbs measure corresponding to the potential $g = -\ln\varphi$ where φ is the differential expanding coefficient on unstable leaves (see [3] and [21]). The large deviations estimates (2.9) and (2.10) hold true with the rate functional J given by (2.8) with $Q(W_\lambda)$ for $W_\lambda = \lambda F$ given by (5.1) in place of $r(W_\lambda)$ in (2.8).

Proof. For T being a C^2 Axiom A diffeomorphism (in particular, Anosov) in a neighborhood of an attractor or T being an expanding C^2 endomorphism of a Riemannian manifold M (see [3]) let ζ be a finite Markov partition for T. Then we can take $\mathcal{F}_{kl}$ to be the finite σ-algebra generated by the partition $\cap_{i=k}^l T^i \zeta$. Another case for the above theorem is when T is a topologically mixing subshift of finite type, i.e. T is the left shift on a subspace Ξ of the space of one-sided sequences $\varsigma = (\varsigma_i, i \ge 0), \varsigma_i = 1, ..., l_0$ such that $\varsigma \in \Xi$ if $\xi_{\varsigma_i \varsigma_{i+1}} = 1$ for all $i \ge 0$ where

$\Xi = (\xi_{ij})$ is an $l_0 \times l_0$ matrix with 0 and 1 entries and such that Ξ^n for some n is a matrix with positive entries. Again, we take μ to be a Gibbs invariant measure corresponding to some Hölder continuous function and to define $\mathcal{F}_{kl}$ as the finite σ-algebra generated by cylinder sets with fixed coordinates having numbers from k to l. The exponentially fast ψ-mixing is well known in the above cases (see [3]). In fact, convergence to zero of the modified $\psi-$mixing coefficient $\psi_{P,\Pi}(n)$ holds true, as well, in the hyperbolic and expanding case when P is the normalized Riemannian volume and $\Pi = \mu^{\mathrm{SRB}}$.

If the function $W_\lambda = W_\lambda(x_1, ..., x_\ell)$ is continuous in $x_1, ..., x_\ell$ then $\beta_{W_\lambda}(n)$ from Proposition 3.1 tends to zero as $n \to \infty$, and so the condition (3.1) will be satisfied here. It follows from [16] that

$$\lim_{N \to \infty} \frac{1}{N} \ln \int \exp \big(\sum_{n=1}^{N} \hat{W}_\lambda(f(T^n x)) \big) d\mu(x) = \mathfrak{P}(\ln \hat{W}_\lambda + g)$$

and Theorem 5.1 follows from Proposition 3.1 and Corollary 3.2 considered with $k = 1$ since in our circumstances differentiability of the topological pressure in parameters of the potential is well known (see, for instance, [24] and [23]). $\qquad \square$

5.2. **Remark.** (i) A version of Theorem 2.2 can also be obtained in the present dynamical systems setup where the limit $Q(W) = \mathfrak{P}(\ln \hat{W}(x) + g)$ is obtained in the same way as in Theorem 5.1. Since $\mathfrak{P}(q)$ is Gateaux differentiable at any Hölder continuous q (see [26] and [23]) then $Q(W)$ is also Gateaux differentiable at any Hölder continuous W and the large deviations for occupational measures

$$\zeta_N = \zeta_{N,x} = \frac{1}{N} \sum_{n=1}^{N} \delta_{\big(T^{q_1(n)} x, ..., T^{q_\ell(n)} x \big)}$$

follow from Section 4.5.3 in [11] with a rate function which is the Fenchel–Legendre transform of Q.

(ii) In the continuous time dynamical systems case of C^2 hyperbolic flows (Axiom A flows in a neghborhood of an attractor, in particular, Anosov flows) the limits of the type (2.19) for $k = 1$ are obtained again relying on a result similar to (3.31), namely that,

$$(5.2) \qquad \lim_{\mathcal{T} \to \infty} \tfrac{1}{\mathcal{T}} \ln \int_M \exp \big(\int_0^{\mathcal{T}} V(T^{q_1(t)} x, ..., T^{q_\ell(t)} x) dt \big) d\mu(x)$$
$$= \lim_{\mathcal{T} \to \infty} \tfrac{1}{\mathcal{T}} \ln \int_M \exp \big(\int_0^{\mathcal{T}} V_{\mathrm{cont}}^{(1)}(T^{\alpha_1 t} x) dt \big) d\mu(x)$$
$$= \mathfrak{P}(V_{\mathrm{cont}}^{(1)} + g)$$

where μ is the SRB measure having g as its potential, V is continuous and $V_{\mathrm{cont}}^{(1)}$ is defined in Section 3.3. The second equality in (5.2) follows from [16] while the first equality there cannot be obtained via a general result like Proposition 3.1 (in view of lack of appropriate $\psi-$mixing for hyperbolic flows) and it requires a quite technical proof relying on some arguments which replace the Markov property employed in Section 3.3. Another approach which can work here is to rely on exponentially fast decay of correlations obtained for Markov families on unstable leafs in [7] together with arguments simular to Proposition 3.1.

(iii) In order to obtain a dynamical systems version of Theorem 2.5 we have to show that for any continuous function $V_t(x_1, ..., x_\ell)$ on $\mathbb{R}_+ \times M^\ell$,

(5.3) $\qquad \lim_{\mathcal{T} \to \infty} \frac{1}{\mathcal{T}} \ln \int_M \exp \big(\int_0^{\mathcal{T}} V_t(T^{q_1(t)}x, ..., T^{q_\ell(t)}x) dt \big) d\mu(x)$
$$= \int_0^{\mathcal{T}} Q_{\text{cont}}(V_t) dt$$

where $Q_{\text{cont}}(V_t) = \mathfrak{P}(V_{t,\text{cont}}^{(1)} + g)$. This is done by combining methods of Section 3 from [17] and of (ii) above.

(iv) Theorem 2.6 provides a direct application to the dynamical systems case when T is a full shift (on a finite alphabet sequence space) considered with a Bernoulli invariant measure taking $X(n) = f \circ T^n$ with a function f on the sequence space depending only on zero coordinate. Nonconventional large deviations when $k > 1$ for more general cases (e.g. subshifts of finite type with Gibbs invariant measures, hyperbolic and expanding transformations etc.) require more elaborate technique and they will not be treated in this paper.

References

[1] I. Assani, *Multiple recurrence and almost sure convergence for weakly mixing dynamical systems*, Israel J. Math. **103**, 111–124 (1998).

[2] V. Bergelson, *Weakly mixing PET*, Ergod. Th.& Dynam. Sys. **7**, 337–349 (1987).

[3] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math. 470, Springer–Verlag, 2nd ed., 2008.

[4] V. Bergelson, A. Leibman and C.G. Moreira, *From discrete-to continuous time ergodic theorems*, Ergod. Th.& Dyn. Sys., to appear.

[5] G. Contreras, *Regularity of topological and metric entropy of hyperbolic flows*, Math. Z. 210 (1992), 97–111.

[6] G. Carinci, J-R Chazottes, C. Giardina and F. Redig, *Nonconventional averages along arithmetic progressions and lattice spin systems*, Indag. Math. (2012).

[7] D. Dolgopyat, *Limit theorems for partially hyperbolic systems*, Trans. Amer. Math. Soc. 356 (2003), 1637–1689.

[8] J. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[9] M.D. Donsker and S.R.S. Varadhan, *Asymptotic evaluation of certain Markov processes expectations for large time. I,* Comm. Pure Appl. Math. 28 (1975), 1–47.

[10] M.D. Donsker and S.R.S. Varadhan, *On the variational formula for the principal eigenvalue for operators with maximum principle*, Proc. Nat. Acad. Sci. U.S.A. 72 (1975), 780–783.

[11] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications, 2nd. ed.*, Springer, Heidelberg, 1998.

[12] M.I. Freidlin, *The averaging principle and theorems on large deviations,* Russ. Math. Surv., **33,** No.5 (1978), 107–160.

[13] H. Furstenberg, *Nonconventional ergodic averages*, Proc. Symp. Pure Math. 50, 43–56 (1990).

[14] I.A. Ibragimov and Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters–Noordhoff, Groningen (1971).

[15] Yu. Kifer, *Principal eigenvalues, topological pressure, and stochastic stability of equilibrium states,* Israel J. Math. 70 (1990), 1–47.

[16] Yu. Kifer, *Large deviations in dynamical systems and stochastic processes,* Trans. Amer. Math. Soc., 321 (1990), 505–524.

[17] Yu. Kifer, *Averaging in dynamical systems and large deviations,* Invent. Math., 110 (1992), 337–370.

[18] Yu. Kifer, *Nonconventional limit theorems*, Prob. Th. Rel. Fields 148 (2010), 71–106.

[19] Yu. Kifer, *A nonconventional strong law of large numbers and fractal dimensions of some multiple recurrence sets*, Stoch. Dynam. 12 (2012)

[20] M.A. Krasnoselskii, *Positive Solutions of Operator Equations*, Noordhoff, Groningen, 1964.

[21] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems,* (1995), Cambridge Univ. Press, Cambridge.

[22] Yu. Kifer and S.R.S. Varadhan, *Nonconventional limit theorems in discrete and continuous time via martingales*, Preprint, arXiv:1012.2223.

[23] W. Parry and M. Pollicott, *Zeta Functions and the Periodic Structure of Hyperbolic Dynamics*, Astérisque 187-188, Soc. Math. de France, 1990.

[24] D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, Reading, 1978.

[25] J.A. Sanders, F.Verhurst and J. Murdock *Averaging Methods in Nonlinear Dynamical Systems,* 2nd ed. (2007), Springer, New York.

[26] P. Walters, *Differentiability properties of the pressure of a continuous transformation on a compact metric space*, J. London Math. Soc. (2) 46 (1992), 471–481.

INSTITUTE OF MATHEMATICS, THE HEBREW UNIVERSITY, JERUSALEM 91904, ISRAEL
E-mail address: kifer@math.huji.ac.il

COURANT INSTITUTE FOR MATHEMATICAL STUDIES, NEW YORK UNIVERSITY, 251 MERCER ST, NEW YORK, NY 10012, USA
E-mail address: varadhan@cims.nyu.edu