

LARGE DEVIATIONS IN DYNAMICAL SYSTEMS AND STOCHASTIC PROCESSES

YURI KIFER

ABSTRACT. The paper exhibits a unified approach to large deviations of dynamical systems and stochastic processes based on the existence of a pressure functional and on the uniqueness of equilibrium states for certain dense sets of functions. This enables us to generalize recent results from [OP, Y, and D] on large deviations for dynamical systems, as well, as to recover Donsker-Varadhan's [DV2] large deviation estimates for Markov processes.

1. INTRODUCTION

Suppose that X is a compact metric space, $\mathcal{P}(X)$ is the space of probability measures on X endowed with the topology of weak convergence, and $(\Omega_\lambda, \mathcal{F}_\lambda, P_\lambda)$ is a family of probability spaces indexed by λ from a directed set Λ (see, for instance [DS]). Let $\zeta^\lambda: \Omega_\lambda \rightarrow \mathcal{P}(X)$, $\zeta^\lambda: \omega \rightarrow \zeta_\omega^\lambda$, $\lambda \in \Lambda$, be a family of measurable maps with respect to the measurable structure given by \mathcal{F}_λ on Ω_λ and by the Borel σ -field on $\mathcal{P}(X)$. The theory of large deviations in this set up deals with estimates of the following form:

$$(1.1) \quad \limsup_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{ \zeta^\lambda \in K \} \leq - \inf \{ I(\nu) : \nu \in K \}$$

for any closed $K \subset \mathcal{P}(X)$ and

$$(1.2) \quad \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{ \zeta^\lambda \in G \} \geq - \inf \{ I(\nu) : \nu \in G \}$$

for any open $G \subset \mathcal{P}(X)$, where $I(\nu) \geq 0$ is a lower semicontinuous functional on $\mathcal{P}(X)$, called the rate functional, and $r(\lambda) > 0$, $\lim_{\lambda \in \Lambda} r(\lambda) = \infty$ is a scaling function on Λ , both quantities have to be identified.

The main application I have in mind concerns occupational measures for dynamical systems and stochastic processes. The use of limits over directed sets assures the applicability of the results both for one-dimensional and multi-dimensional time cases, for instance, in the framework of the thermodynamic formalism from [Ru]. In the one-dimensional time case one has a X -valued stochastic process $Y_t = Y_t(\omega)$, $\omega \in \Omega_t \equiv \Omega$ with t running over nonnegative

Received by the editors July 27, 1989.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 60F10; Secondary 58F15.

Key words and phrases. Large deviations, hyperbolic dynamical systems, Markov processes.

This work was supported by USA-Israel Binational Science Foundation.

reals \mathfrak{R}_+ (the continuous time case with $\Lambda = \mathfrak{R}_+$, $\lambda = r(\lambda) = t$) or over non-negative integers \mathbf{Z}_+ (the discrete time case with $\Lambda = \mathbf{Z}_+$, $\lambda = r(\lambda) = t$). Now the measures ζ_ω^t are defined by $\zeta_\omega^t = \frac{1}{t} \int_0^t \delta_{Y_s(\omega)} ds$ in the continuous time case and by $\zeta_\omega^t = \frac{1}{t} \sum_{l=0}^{t-1} \delta_{Y_l(\omega)}$ in the discrete time case, where δ_x denotes the unit mass at x . In particular, if $X = \Omega$ and $Y_t(x) = F^t x$ where $F^t: X \rightarrow X$ is a group or semigroup of transformations, one obtains occupational measures for a dynamical system F^t . In the n -dimensional time case one considers a group of homeomorphisms F^q , $q = (q_1, \dots, q_n) \in \mathbf{Z}^n$ of a compact $X = \Omega$ and defines $\zeta_x^a = (1/|R(a)|) \sum_{q \in R(a)} \delta_{F^q x}$ where $a = (a_1, \dots, a_n) \in \mathbf{Z}^n$, $a_i > 0$, $R(a) = \{(q_1, \dots, q_n) \in \mathbf{Z}^n, 0 \leq q_i < a_i \text{ for } i = 1, \dots, n\}$, and $|R(a)|$ is the number of points in $R(a)$. In statistical mechanics X is usually interpreted as the configuration space, $P_\lambda \equiv P$ is a distribution on X , and F^q , $q \in \mathbf{Z}^n$ are shifts.

The main feature of this paper is that large deviations estimates are derived assuming that the limit (called the pressure of V)

$$(1.3) \quad Q(V) = \lim_{\lambda \in \Lambda} (1/r(\lambda)) \log \int \exp \left(r(\lambda) \int V(x) d\zeta_\omega^\lambda(x) \right) dP_\lambda(\omega)$$

exists for any V from the space $C(X)$ of continuous functions on X . This approach was employed previously in [G, T, E, and A1]. It is known (see [A1]) that the existence of the limit (1.1) implies already upper large deviations bounds (1.1) with the rate functional $I(\nu)$ which is convex conjugate of $Q(V)$. For lower bounds (1.2) additional assumptions are needed and usually one requires Gâteaux differentiability of $Q(V)$ (see [G, E, A2]) at all $V \in C(X)$ which is rarely true in applications to dynamical systems.

It is clear that $Q(V)$ is a continuous functional and its convexity follows from Hölder’s inequality. The convex conjugate I of Q is defined by

$$(1.4) \quad I(\mu) = \sup_{V \in C(X)} \left(\int V d\mu - Q(V) \right)$$

if $\mu \in \mathcal{P}(X)$ and $I(\mu) = \infty$ for all other signed measures μ . From this definition it follows immediately that $I(\mu)$ is convex and lower semicontinuous. Thus by the well-known duality (see, for instance, [AE, p. 201]) one has

$$(1.5) \quad Q(V) = \sup_{\mu \in \mathcal{P}(X)} \left(\int V d\mu - I(\mu) \right).$$

Since $\mathcal{P}(X)$ is compact and $I(\mu)$ is lower semicontinuous one concludes that for any $V \in C(X)$ there exists $\mu_V \in \mathcal{P}(X)$, called an equilibrium state for V , such that

$$(1.6) \quad Q(V) = \int V d\mu_V - I(\mu_V).$$

In general, μ_V is not unique for many $V \in C(X)$ but it turns out that the uniqueness of μ_V for all V which are finite linear combinations of functions

from a countable dense set in $C(X)$ suffices already for lower large deviations bounds. It follows from the theorem on p. 450 of [DS] that the convex functional $Q(V)$ always has a unique subdifferential, i.e. equilibrium state, for each V from a dense in $C(X)$ set of functions. Of course, one can choose a countable dense subset from this set but the real issue is to ensure that finite linear combinations of functions from this set will also have unique equilibrium states. In the case of a smooth dynamical system in a neighborhood of a basic hyperbolic set all Hölder continuous functions have unique equilibrium states and so the above condition is satisfied. In the case of Markov processes with good transition densities the condition is satisfied, as well.

This approach enables one to derive large deviations estimates both for dynamical systems and Markov processes as simple corollaries of the main theorem which is proved in the next section. Moreover, I generalize and improve results of recent papers [OP, Y, and D] considering, in particular, the continuous time case for which methods of above papers does not seem to work. Remark that the assumption on uniqueness of equilibrium states for large classes of functions is quite natural in applications to statistical mechanics.

I would like to thank S. Varadhan for a helpful conversation we had during AMS Summer Seminar “Mathematics of Random Media” at Virginia Tech in June 1989.

2. MAIN THEOREM

It is clear that the limit (1.3) satisfies

$$(2.1) \quad |Q(V_1) - Q(V_2)| \leq \|V_1 - V_2\| \quad \text{and} \quad -\|V\| \leq Q(V) \leq \|V\|$$

where $\|V\| = \sup_{x \in X} |V(x)|$, and by Hölder’s inequality for $0 \leq p \leq 1$,

$$(2.2) \quad Q(pV_1 + (1 - p)V_2) \leq pQ(V_1) + (1 - p)Q(V_2),$$

i.e. Q is convex. The main result of this section is the following.

Theorem 2.1. *Suppose that the limit (1.3) exists for any $V \in C(X)$. Then the upper bound (1.1) holds true with the rate functional I given by (1.4). Let, furthermore, there exists a countable set of functions $V_1, V_2, \dots \in C(X)$ such that their span is dense in $C(X)$ with respect to $\|\cdot\|$, $\|V_i\| = 1$ for all $i = 1, 2, \dots$, and for each $n = 1, 2, \dots$ and all numbers $\beta_1, \beta_2, \dots, \beta_n$ the function $V = \beta_1 V_1 + \beta_2 V_2 + \dots + \beta_n V_n$ has a unique measure $\mu_V \in \mathcal{P}(X)$ satisfying (1.6). Then the lower bound (1.2) also holds true.*

Proof. I begin with the upper bound (1.1) where I follow [A1]. For any set $B \subset \mathcal{P}(X)$ put

$$(2.3) \quad J(B) = \inf\{I(\nu) : \nu \in B\}.$$

Let $K \subset \mathcal{P}(X)$ be a closed set which is therefore compact since $\mathcal{P}(X)$ is compact.

If $J(K) \leq 0$ there is nothing to prove since the left-hand side of (1.1) is nonpositive. Let $0 < J(K) < \infty$. For each $\varepsilon > 0$ and every $W \in C(X)$ put

$$\Gamma_\varepsilon(W) = \left\{ \mu \in \mathcal{P}(X) : \int W d\mu - Q(W) > J(K) - \varepsilon \right\}.$$

Then by (1.4),

$$K \subset \{ \mu \in \mathcal{P}(X) : I(\mu) > J(K) - \varepsilon \} = \bigcup_{W \in C(X)} \Gamma_\varepsilon(W).$$

Since K is compact then there exist $W_1, \dots, W_l \in C(X)$ such that $K \subset \bigcup_{1 \leq i \leq l} \Gamma_\varepsilon(W_i)$. Then

$$\begin{aligned} (2.4) \mathcal{P}_\lambda \{ \zeta^\lambda \in K \} &\leq \sum_{1 \leq i \leq l} P_\lambda \{ \zeta^\lambda \in \Gamma_\varepsilon(W_i) \} \\ &= \sum_{1 \leq i \leq l} P_\lambda \left\{ \int W_i d\zeta^\lambda > Q(W_i) + J(K) - \varepsilon \right\} \\ &\leq \sum_{1 \leq i \leq l} \exp(-r(\lambda)(Q(W_i) + J(K) - \varepsilon)) \int \exp\left(r(\lambda) \int W_i d\zeta^\lambda\right) dP_\lambda. \end{aligned}$$

Taking here log, dividing by $r(\lambda)$, and passing to $\limsup_{\lambda \in \Lambda}$ one derives from (1.3) that

$$\limsup_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{ \zeta^\lambda \in K \} \leq -J(K) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary (1.1) follows. It remains to consider the case $J(K) = \infty$. For each $N > 0$ and $W \in C(X)$ put $\Gamma^{(N)}(W) = \{ \mu \in \mathcal{P}(X) : \int W d\mu - Q(W) > N \}$. Then $K \subset \bigcup_{W \in C(X)} \Gamma^{(N)}(W)$ and proceeding as before one obtains

$$\limsup_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{ \zeta^\lambda \in K \} \leq -N.$$

Since N is arbitrary (1.1) follows in this case, as well.

Next, I shall deal with the more difficult lower bound (1.2). Introduce a metric on $\mathcal{P}(X)$ by

$$(2.5) \quad \rho(\mu, \nu) = \sum_{k=1}^{\infty} 2^{-k} \left| \int V_k d\mu - \int V_k d\nu \right|$$

which is compatible with the topology of weak convergence since $\text{span}\{V_1, V_2, \dots\}$ is dense in $\mathcal{P}(X)$.

First I shall reduce the problem to a finite-dimensional situation.

Let $f_n: \mathcal{P}(X) \rightarrow \mathbb{R}^n$, $n = 1, 2, \dots$, be a sequence of maps acting by the formula

$$(2.6) \quad f_n(\mu) = \left(\int V_1 d\mu, \dots, \int V_n d\mu \right).$$

Clearly, the maps f_n are affine and continuous, and so if $U \subset \mathfrak{R}^n$ is an open set and $L_n = f_n(\mathcal{P}(X))$ then $f_n^{-1}U = f_n^{-1}(U \cap L_n)$ is an open subset of $\mathcal{P}(X)$. Furthermore,

$$(2.7) \quad J(f_n^{-1}U) = \inf\{I(\nu) : f_n\nu \in U\} = \inf_{\alpha \in U} I_n(\alpha)$$

where $I_n(\alpha) = \inf\{I(\nu) : f_n\nu = \alpha\}$ if $\alpha \in L_n$ and $I_n(\alpha) = \infty$ if $\alpha \notin L_n$. For each $\beta \in \mathfrak{R}^n$ denote $\langle \beta, V \rangle_n = \sum_{1 \leq k \leq n} \beta_k V_k$ and $Q_n(\beta) = Q(\langle \beta, V \rangle_n)$. From (1.5) and (2.7) it follows that

$$(2.8) \quad \begin{aligned} Q_n(\beta) &= \sup_{\mu \in \mathcal{P}(X)} \left(\sum_{1 \leq k \leq n} \beta_k \int V_k d\mu - I(\mu) \right) \\ &= \sup_{\mu \in \mathcal{P}(X)} (\langle \beta, f_n\mu \rangle - I(\mu)) = \sup_{\alpha \in L_n} (\langle \beta, \alpha \rangle - I_n(\alpha)) \\ &= \sup_{\alpha \in \mathfrak{R}^n} (\langle \beta, \alpha \rangle - I_n(\alpha)) \end{aligned}$$

where I use the brackets $\langle \cdot, \cdot \rangle$ also to denote the inner product in \mathfrak{R}^n . It follows from (2.8) that $Q_n(\beta)$ is a convex continuous function on \mathfrak{R}^n . From the definition of $I_n(\alpha)$ together with the continuity of f_n and the lower semicontinuity of $I(\nu)$ one derives the lower semicontinuity of $I_n(\alpha)$. Since $I(\nu)$ is convex I conclude that $I_n(\alpha)$ is convex, as well. Indeed, for $\alpha_1, \alpha_2 \in L_n$ and $0 \leq p \leq 1$ one has

$$(2.9) \quad \begin{aligned} I_n(p\alpha_1 + (1-p)\alpha_2) &= \inf\{I(\nu) : \nu \in \mathcal{P}(X), f_n\nu = p\alpha_1 + (1-p)\alpha_2\} \\ &\leq \inf\{I(p\nu_1 + (1-p)\nu_2) : \nu_1, \nu_2 \in \mathcal{P}(X), f_n\nu_1 = \alpha_1, f_n\nu_2 = \alpha_2\} \\ &\leq pI_n(\alpha_1) + (1-p)I_n(\alpha_2). \end{aligned}$$

If α_1 or α_2 does not belong to L_n then the right-hand side of (2.9) is infinity by the definition of I_n and there is nothing to prove. Now (2.8) together with the convexity and lower semicontinuity of $I_n(\alpha)$ imply (see [R, Theorem 12.2]) that for $\alpha \in L_n$,

$$(2.10) \quad I_n(\alpha) = \sup_{\beta \in \mathfrak{R}^n} (\langle \beta, \alpha \rangle - Q_n(\beta)).$$

To complete the proof of Theorem 2.1 I shall establish first a finite-dimensional version of the lower bound (1.2). For any set $A \subset \mathfrak{R}^n$ put $J_n(A) = \inf_{\alpha \in A} I_n(\alpha)$.

Lemma 2.1. *For any closed set $M \subset \mathfrak{R}^n$ and each $n \geq 1$,*

$$(2.11) \quad \limsup_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda\{f_n \zeta^\lambda \in M\} \leq -J_n(M).$$

For any open set $U \subset \mathfrak{R}^n$ and each $n \geq 1$,

$$(2.12) \quad \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda\{f_n \zeta^\lambda \in U\} \geq -J_n(U).$$

Proof. Since $f_n^{-1}(M) = f_n^{-1}(M \cap L_n)$ is a closed subset of $\mathcal{P}(X)$ then (2.11) follows from the upper bound (1.1) which has already been proved. To prove

(1.2) I shall follow partially [G and E]. Fix an arbitrary $\varepsilon > 0$ and choose $\alpha_\varepsilon \in U$ such that

$$(2.13) \quad I_n(\alpha_\varepsilon) < J_n(U) + \varepsilon.$$

One can assume that $\alpha_\varepsilon \in \mathcal{D}(I_n) = \{\alpha : I_n(\alpha) < \infty\}$ since for otherwise $J_n(U) = \infty$ and there is nothing to prove. Let $ri\mathcal{D}(I_n)$ be the interior of $\mathcal{D}(I_n)$ in its affine hull (see [R]). Then either $ri\mathcal{D}(I_n) \neq \emptyset$ or $\mathcal{D}(I_n)$ consists of one point. Assume, first, that $ri\mathcal{D}(I_n) \neq \emptyset$. By Corollary 7.5.1 in [R] for any $\alpha, \tilde{\alpha} \in \mathcal{D}(I_n)$ it follows from the convexity and the lower semicontinuity of I_n that

$$\lim_{p \uparrow 1} I_n(p\alpha + (1-p)\tilde{\alpha}) = I_n(\alpha),$$

and so one can choose α_ε satisfying (2.13) to belong to $ri\mathcal{D}(I_n)$. Then by Theorems 23.4 and 23.5 from [R] there exists $\beta_\varepsilon \in \mathfrak{R}^n$ such that

$$(2.14) \quad Q_n(\beta_\varepsilon) = \langle \beta_\varepsilon, \alpha_\varepsilon \rangle - I_n(\alpha_\varepsilon).$$

If $\mathcal{D}(I_n)$ consists of the single point α_ε then one derives from (2.8) that $Q_n(\beta) = \langle \beta, \alpha_\varepsilon \rangle - I(\alpha_\varepsilon)$ for all $\beta \in \mathfrak{R}^n$. So again the existence of β_ε satisfying (2.14) follows. Since U is open then for all $\delta > 0$ small enough

$$(2.15) \quad U_\delta(\alpha_\varepsilon) = \{\alpha : |\alpha - \alpha_\varepsilon| < \delta\} \subset U.$$

Put $Q_n^\lambda(\beta) = (1/r(\lambda)) \log \int \exp(r(\lambda) \int \langle \beta, V \rangle_n d\zeta^\lambda) dP_\lambda$ and introduce probability measures $P_\lambda^{(\alpha_\varepsilon)}$ on Ω_λ by

$$(2.16) \quad P_\lambda^{(\alpha_\varepsilon)}(A) = \int_A \exp\{r(\lambda)(\langle \beta_\varepsilon, f_n \zeta^\lambda \rangle - Q_n^\lambda(\beta_\varepsilon))\} dP_\lambda.$$

Then by (2.15),

$$(2.17) \quad \begin{aligned} P_\lambda\{f_n \zeta^\lambda \in U\} &\geq P_\lambda\{f_n \zeta^\lambda \in U_\delta(\alpha_\varepsilon)\} \\ &= \exp\{-r(\lambda)(\langle \beta_\varepsilon, \alpha_\varepsilon \rangle - Q_n^\lambda(\beta_\varepsilon))\} \\ &\quad \times \int_{\{\omega : f_n \zeta_\omega^\lambda \in U_\delta(\alpha_\varepsilon)\}} \exp\{-r(\lambda)\langle \beta_\varepsilon, f_n \zeta^\lambda - \alpha_\varepsilon \rangle\} dP_\lambda^{(\alpha_\varepsilon)} \end{aligned}$$

Thus by (1.3) and (2.14),

$$(2.18) \quad \begin{aligned} \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda\{f_n \zeta^\lambda \in U\} \\ \geq -I_n(\alpha_\varepsilon) - \delta |\beta_\varepsilon| + \liminf_{\lambda \in \Lambda} (1/r(\lambda)) P_\lambda^{(\alpha_\varepsilon)}\{f_n \zeta^\lambda \in U_\delta(\alpha_\varepsilon)\}. \end{aligned}$$

I claim that

$$(2.19) \quad \lim_{\lambda \in \Lambda} P_\lambda^{(\alpha_\varepsilon)}\{f_n \zeta^\lambda \in U_\delta(\alpha_\varepsilon)\} = 1,$$

i.e.

$$(2.20) \quad \lim_{\lambda \in \Lambda} P_\lambda^{(\alpha_\varepsilon)}\{f_n \zeta^\lambda \in L_n \setminus U_\delta(\alpha_\varepsilon)\} = 0.$$

To prove (2.20) I use the upper bound (2.11) for probability measures $P_\lambda^{(\alpha_\varepsilon)}$ in place of P_λ . Since $\langle \beta, f_n \zeta^\lambda \rangle = \int \langle \beta, V \rangle_n d\zeta^\lambda$ then by (1.3) and (2.16),

$$(2.21) \quad Q_n^{(\alpha_\varepsilon)}(\beta) = \lim_{\lambda \in \Lambda} (1/r(\lambda)) \log \int \exp \left(r(\lambda) \int \langle \beta, V \rangle_n d\zeta^\lambda \right) dP_\lambda^{(\alpha_\varepsilon)} \\ = Q_n(\beta + \beta_\varepsilon) - Q_n(\beta_\varepsilon).$$

Thus by (2.10) the corresponding convex conjugate functional $I_n^{(\alpha_\varepsilon)}(\alpha)$ has the form

$$(2.22) \quad I_n^{(\alpha_\varepsilon)}(\alpha) = \sup_{\beta \in \mathbb{R}^n} (\langle \beta, \alpha \rangle - Q_n(\beta + \beta_\varepsilon)) + Q_n(\beta_\varepsilon) \\ = I_n(\alpha) - (\langle \beta_\varepsilon, \alpha \rangle - Q_n(\beta_\varepsilon)) \geq 0.$$

By the upper bound (2.11) for $P_\lambda^{(\alpha_\varepsilon)}$ in place of P_λ ,

$$(2.23) \quad \limsup_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda^{(\alpha_\varepsilon)} \{f_n \zeta^\lambda \in L_n \setminus U_\delta(\alpha_\varepsilon)\} \\ \leq - \inf_{\alpha \in L_n \setminus U_\delta(\alpha_\varepsilon)} I_n^{(\alpha_\varepsilon)}(\alpha),$$

and so (2.20) will follow if the right-hand side of (2.23) is negative. By (2.14), $I_n^{(\alpha_\varepsilon)}(\alpha_\varepsilon) = 0$. I claim that α_ε is the only zero of $I_n^{(\alpha_\varepsilon)}$. Indeed, if there exists another $\tilde{\alpha} \in L_n$, $\tilde{\alpha} \neq \alpha_\varepsilon$ such that $I_n^{(\alpha_\varepsilon)}(\tilde{\alpha}) = 0$ then by the lower semicontinuity of $I(\nu)$ and the definition of $I_n(\alpha)$ there exist measures in each of the disjoint sets $f_n^{-1}\alpha_\varepsilon$ and $f_n^{-1}\tilde{\alpha}$ which are equilibrium states for the function $V = \langle \beta_\varepsilon, V \rangle_n$ in the sense of (1.6) which contradicts the assumption of Theorem 2.1. Thus by (2.10), $I_n^{(\alpha_\varepsilon)}(\alpha) > 0$ for all $\alpha \in \mathbb{R}^n \setminus \{\alpha_\varepsilon\}$. Since $L_n \setminus U_\delta(\alpha_\varepsilon)$ is a closed set and $I_n^{(\alpha_\varepsilon)}(\alpha)$ is lower semicontinuous one derives from here that

$$(2.24) \quad \inf_{\alpha \in L_n \setminus U_\delta(\alpha_\varepsilon)} I_n^{(\alpha_\varepsilon)}(\alpha) > 0$$

because, for otherwise, there would exist a sequence $\alpha_k \rightarrow \alpha$ as $k \rightarrow \infty$, $\alpha, \alpha_k \in L_n \setminus U_\delta(\alpha_\varepsilon)$, $\lim_{k \rightarrow \infty} I_n^{(\alpha_\varepsilon)}(\alpha_k) = 0$ and so $I_n^{(\alpha_\varepsilon)}(\alpha) = 0$ which is impossible. Now (2.24) together with (2.23) imply (2.20) proving (2.19). This yields that the limit in the right-hand side of (2.18) is zero. Letting $\delta \rightarrow 0$ and then $\varepsilon \rightarrow 0$ one derives the lower bound (2.12) from (2.13) and completes the proof of Lemma 2.1. \square

Next, I am able to conclude the proof of Theorem 2.1. Let $G \subset \mathcal{P}(X)$ be an open set. For any $\varepsilon > 0$ one can choose $\nu_\varepsilon \in G$ such that

$$(2.25) \quad I(\nu_\varepsilon) \leq J(G) + \varepsilon$$

where J is defined by (2.3). Since $\mathcal{P}(X) \setminus G$ is compact then

$$(2.26) \quad \delta = \inf\{\rho(\mu, \nu_\varepsilon) : \mu \in \mathcal{P}(X) \setminus G\} > 0.$$

Denote $\rho_n(\mu, \nu) = \sum_{k=1}^n 2^{-k} |\int V_k d\mu - \int V_k d\nu|$ then for any $\mu, \nu \in \mathcal{P}(X)$,

$$(2.27) \quad 0 \leq \rho(\mu, \nu) - \rho_n(\mu, \nu) \leq 2^{-(n-1)}$$

since $\|V_k\| = 1$, $k = 1, 2, \dots$. Hence if n is big enough then

$$(2.28) \quad G_{n,\delta} = \{\mu \in \mathcal{P}(X) : \rho_n(\mu, \nu_\varepsilon) < \frac{\delta}{2}\} \subset G.$$

For each $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ denote $\|\alpha\|_n = \sum_{k=1}^n 2^{-k} |\alpha_k|$ and put $\alpha_\varepsilon = f_n \nu_\varepsilon$. If $U_{n,\delta} = \{\alpha : \|\alpha_\varepsilon - \alpha\|_n < \frac{\delta}{2}\}$ then clearly $f_n G_{n,\delta} = L_n \cap U_{n,\delta}$. Thus by (2.7), (2.12), (2.25), and (2.28) one obtains

$$(2.29) \quad \begin{aligned} \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{\zeta^\lambda \in G\} &\geq \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{\zeta^\lambda \in G_{n,\delta}\} \\ &= \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{f_n \zeta^\lambda \in U_{n,\delta}\} \geq -J_n(U_{n,\delta}) \\ &\geq -I_n(\alpha_\varepsilon) \geq -I(\nu_\varepsilon) \geq -J(G) - \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary (1.2) follows and the proof of Theorem 2.1 is complete. \square

Let W_1, \dots, W_n be arbitrary continuous functions on X . Similarly to the construction in the proof of Theorem 2.1 define $\tilde{f}_n: \mathcal{P}(X) \rightarrow \mathbb{R}^n$ by $\tilde{f}_n(\mu) = (\int W_1 d\mu, \dots, \int W_n d\mu)$ and put $\tilde{I}_n(\alpha) = \inf\{I(\nu) : \tilde{f}_n \nu = \alpha\}$ if $\alpha \in \tilde{f}_n(\mathcal{P}(X))$ and $\tilde{I}_n(\alpha) = \infty$ if $\alpha \notin \tilde{f}_n(\mathcal{P}(X))$. In the same way as in (2.8) and (2.9) it follows for $\tilde{Q}_n(\beta) = Q(\sum_{1 \leq k \leq n} \beta_k W_k)$, $\beta = (\beta_1, \dots, \beta_n)$ that $\tilde{Q}_n(\beta) = \sup_{\alpha \in \mathbb{R}^n} (\langle \beta, \alpha \rangle - \tilde{I}_n(\beta))$ and $\tilde{I}_n(\alpha) = \sup_{\beta \in \mathbb{R}^n} (\langle \beta, \alpha \rangle - \tilde{Q}_n(\beta))$ for $\alpha \in \tilde{f}_n(\mathcal{P}(X))$. One obtains easily from Theorem 2.1

Corollary 2.1. *Let $\tilde{J}_n(A) = \inf_{\alpha \in A} \tilde{I}_n(\alpha)$. Then for any closed set $M \subset \mathbb{R}^n$,*

$$(2.30) \quad \limsup_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{\tilde{f}_n \zeta^\lambda \in M\} \leq -\tilde{J}_n(M)$$

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

$$(2.31) \quad \liminf_{\lambda \in \Lambda} (1/r(\lambda)) \log P_\lambda \{\tilde{f}_n \zeta^\lambda \in U\} \geq -\tilde{J}_n(U).$$

Remark also that the derivation of Corollary 2.1 from Theorem 2.1 is a partial case of the general contraction principle.

Remark 2.1. According to Theorems 23.5 and 25.1 from [R] my assumption that the functions $\langle \beta, V \rangle_n$ have unique equilibrium state is equivalent to the differentiability of the functions $Q_n(\beta)$ in β .

Remark 2.2. One can generalize Theorem 2.1 to a noncompact (but locally compact) X assuming that except for events whose P_λ -probabilities decrease fast in λ the measures ζ_ω^λ form a tight family (cf. [A1]).

3. APPLICATIONS TO DYNAMICAL SYSTEMS

Let $F^t: M \rightarrow M$, $t \in \mathbb{Z}_+$ or $t \in \mathbb{R}_+$ be a semigroup of continuous maps of a locally compact metric space M . Suppose that $X \subset M$ is a compact

set and put $X_t = \{x : F^u x \in X \text{ for all } u \in [0, t]\}$. If $F^t X \subset X$ for all $t \geq 0$ then, clearly, $X_t \equiv X$. I am going to employ the general set up of the previous sections when $r(\lambda) = \lambda = t$, $\Omega_t = X_t$, \mathcal{F}_t is the Borel σ -field in X_t , $P_t = m_t = \frac{m}{m(X_t)}$ where m is a probability measure on X such that $m(X_t) > 0$ for all $t \geq 0$, and, finally, for $x \in X_t$ one defines $\zeta_x^t = \frac{1}{t} \int_0^t \delta_{F^s x} ds$ in the continuous time case and $\zeta_x^t = \frac{1}{t} \sum_{k=0}^{t-1} \delta_{F^k x}$ in the discrete time case.

The study of large deviations for the measures ζ_x^t involves the topological pressure whose definition is based on the notion of (δ, t) -separated sets. A set $S \subset X_t$ will be called (δ, t) -separated if $y, z \in S$, $y \neq z$ imply $\text{dist}(F^s y, F^s z) > \delta$ for some $s \in [0, t]$. Define

$$(3.1) \quad Z_X(\psi, \delta, t) = \sup \left\{ \sum_{x \in S} \exp \left(t \int_X \psi(y) d\zeta_x^t(y) \right) : S \subset X_t \text{ is } (\delta, t)\text{-separated} \right\},$$

$$Z_X(\psi, \delta, t) = 0 \quad \text{if } X_t = \emptyset, \text{ and}$$

$$(3.2) \quad Q_X(\psi, \delta) = \limsup_{t \rightarrow \infty} \frac{1}{t} \log Z_X(\psi, \delta, t)$$

where ψ is a continuous function on X .

Then

$$(3.3) \quad Q_X(\psi) = \limsup_{\delta \rightarrow 0} Q_X(\psi, \delta)$$

is called the topological pressure of a function ψ . Denote by \mathcal{M}_X^F the space of F^t -invariant measures $\mu \in \mathcal{P}(X)$, i.e. $\mu(F^{-t}U) = \mu(U)$ for all $t \geq 0$ and any Borel $U \subset X$. The following statement is called the variational principle.

Proposition 3.1. *If $\mathcal{M}_X^F \neq \emptyset$ then for any continuous function ψ ,*

$$(3.4) \quad Q_X(\psi) = \sup_{\mu \in \mathcal{M}_X^F} \left(\int \psi d\mu + h_\mu(F^1) \right)$$

where $h_\mu(F^1)$ is the Kolmogorov-Sinai entropy of the map F^1 with respect to a F^t -invariant measure μ . If $\mathcal{M}_X^F = \emptyset$ then

$$(3.5) \quad Q_X(\psi) = -\infty.$$

Proof. If $\mathcal{M}_X^F = \emptyset$ then $X_\infty = \bigcap_{t \geq 0} X_t = \emptyset$ since X_∞ is compact as an intersection of the compact sets $X_t = \bigcap_{0 \leq u \leq t} F^{-u} X$, X_∞ is forward F^t -invariant, and so if $X_\infty \neq \emptyset$ then it must support a F^t -invariant probability measure. The intersection X_∞ of the decreasing sequence of compact sets X_t being empty implies that $X_t = \emptyset$ for all $t \geq t_0$ and (3.5) follows. Next, I assume that $\mathcal{M}_X^F \neq \emptyset$. The standard form of the variational principle concerns the case where X is a forward F^t -invariant set, i.e. $F^t X \subset X$, and so I shall indicate how to modify slightly the proof of Theorem 9.10 in [W] to obtain this

more general version. It seems that, L.-S. Young knew a version of this result which she needed to establish Theorem 4 in [Y]. Let Φ be the closure of the set $\bigcup\{\text{supp } \mu: \mu \in \mathcal{M}_X^F\}$.

Then $F^t\Phi \subset \Phi \subset X$, and so by the standard version of the variational principle

$$(3.6) \quad Q_X(\psi) \geq Q_\Phi(\psi) = \sup_{\mu \in \mathcal{M}_X^F} \left(\int \psi d\mu + h_\mu(F^1) \right).$$

To prove the inequality in the other direction it suffices to produce for any $\varepsilon > 0$ a measure $\mu_\varepsilon \in \mathcal{M}_X^F$ such that

$$(3.7) \quad h_{\mu_\varepsilon}(F^t) + \int \psi d\mu_\varepsilon \geq Q_X(\psi, \varepsilon).$$

For each $t \geq 0$ let S_t be (ε, t) -separated sets in X_t with

$$\log \sum_{x \in S_t} \exp \left(t \int_X \psi d\zeta_x^t \right) \geq \log Z_X(\psi, \varepsilon, t) - 1.$$

Define atomic measures $\sigma_t \in \mathcal{P}(X)$ concentrated on S_t by the formula

$$\sigma_t = \left(\sum_{z \in S_t} \exp \left(t \int_X \psi d\zeta_z^t \right) \right)^{-1} \sum_{y \in S_t} \delta_y \exp \left(t \int_X \psi d\zeta_y^t \right)$$

and put

$$\mu_t = \frac{1}{t} \sum_{k=0}^{t-1} \sigma_k \circ F^{-k} \quad \text{or} \quad \mu_t = \frac{1}{t} \int_0^t \sigma_u \circ F^{-u} du$$

in the discrete and continuous time cases, respectively, where $\nu \circ F^{-t}(U) = \nu(F^{-t}U)$ for $\nu \in \mathcal{P}(X)$ and $U \in \mathcal{B}(X)$. The measures μ_t are probability measures on the compact X , and so one can choose a sequence $t_i \rightarrow \infty$ such that

$$\lim_{i \rightarrow \infty} \frac{1}{t_i} \log Z_X(\psi, \varepsilon, t_i) = Q_X(\psi, \varepsilon)$$

and μ_{t_i} weakly converges to some $\mu \in \mathcal{P}(X)$ which according to Theorem 6.9 of [W] satisfies $\mu \in \mathcal{M}_X^F$. Next one shows that $\mu = \mu_\varepsilon$ satisfies (3.7) taking a partition $\zeta = \{A_1, \dots, A_k\}$ of $(X, \mathcal{B}(X))$ such that $\text{diam}(A_i) < \varepsilon$ and $\mu(\partial A_i) = 0$ for $1 \leq i \leq k$ and proceeding verbatim as on pp. 220–221 in [W]. Finally, (3.7) being true for all $\varepsilon > 0$ implies the inequality opposite to (3.6). \square

For any $\delta > 0$, $t \geq 0$, and $x \in X_t$ put

$$U_\delta(x, t, X) = \{y \in X_t: \text{dist}(F^u x, F^u y) \leq \delta \text{ for all } u \in [0, t]\}.$$

Proposition 3.2 (cf. Theorem 6.2 in [K2]). *Suppose that $m \in \mathcal{P}(X)$, the support of m is the whole X , and for some continuous function φ on X and for all*

$t, \delta > 0, x \in X_t$ one has

$$(3.8) \quad (A_\delta(t))^{-1} \leq m(U_\delta(x, t, X)) \exp \left(-t \int_X \varphi d\zeta_x^t \right) \leq A_\delta(t)$$

where $A_\delta(t) > 0$ satisfies

$$(3.9) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log A_\delta(t) = 0.$$

Then for any $V \in C(X)$,

$$(3.10) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log \int_{X_t} \exp \left(t \int_X V d\zeta_x^t \right) dm(x) = Q_X(\varphi + V) = Q_\Phi(\varphi + V)$$

where, recall, $\Phi =$ closure of $\bigcup_{\mu \in \mathcal{M}_X^F} \text{supp } \mu$.

Proof. For each $\delta, t > 0$ denote by $S(\delta, t)$ a maximal (δ, t) -separated set in X_t . Then $\bigcup_{x \in S(\delta, t)} U_\delta(x, t, X) = X_t \supset \bigcup_{x \in S(\delta, t)} U_{\delta/2}(x, t, X)$ and for any $y, z \in S(\delta, t), y \neq z$ the sets $U_{\delta/2}(y, t, X)$ and $U_{\delta/2}(z, t, X)$ are disjoint. Thus

$$(3.11) \quad \begin{aligned} & \sum_{x \in S(\delta, t)} m(U_\delta(x, t, X)) \exp \left(t \int_X (V + \gamma_\delta(V)) d\zeta_x^t \right) \\ & \geq \int_{X_t} \exp \left(t \int_X V d\zeta_x^t \right) dm(x) \\ & \geq \sum_{x \in S(\delta, t)} m(U_{\delta/2}(x, t, X)) \exp \left(t \int_X (V - \gamma_\delta(V)) d\zeta_x^t \right) \end{aligned}$$

where $\gamma_\delta(V) = \sup\{|V(y) - V(z)| : y, z \in X, \text{dist}(y, z) \leq \delta\}$ which tends to zero as $\delta \rightarrow 0$ since $V \in C(X)$. Taking log in (3.11), dividing by t , employing (3.8) and (3.9), and letting first $t \rightarrow \infty$ and then $\delta \rightarrow 0$ we derive (3.10) taking into account the definition (3.1)–(3.3) and Proposition 3.1. \square

Taking in (3.10), $V \equiv 0$ one obtains the escape rates from X with respect to $m \in \mathcal{P}(X)$.

Proposition 3.3. *Let $m \in \mathcal{P}(X)$ satisfies the conditions of Proposition 3.2. Then*

$$(3.12) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log m(X_t) = Q_\Phi(\varphi)$$

If $m \in \mathcal{M}_X^F$ then both the left- and the right-hand sides of (3.12) is zero.

Now I am able to derive a general result on large deviations in dynamical systems.

Theorem 3.4. *Let m satisfy the conditions of Proposition 3.2. Suppose that the entropy $h_\mu(F^1)$ is upper semicontinuous at all $\mu \in \mathcal{M}_X^F$. Then for any closed $K \subset \mathcal{P}(X)$,*

$$(3.13) \quad \limsup_{t \rightarrow \infty} \frac{1}{t} \log m\{x : \zeta_x^t \in K\} \leq -\inf\{I(\nu) : \nu \in K\} \leq Q_X(\varphi)$$

where

$$(3.14) \quad I(\nu) = \begin{cases} -\int \varphi d\nu - h_\nu(F^1) & \text{if } \nu \in \mathcal{M}_X^F, \\ \infty & \text{if } \nu \notin \mathcal{M}_X^F. \end{cases}$$

If, in addition, there exists a countable set of functions, $V_1, V_2, \dots \in C(X)$ such that their span is dense in $C(X)$ and for each linear combination $V = \beta_1 V_1 + \dots + \beta_n V_n$ there exists a unique measure $\mu_V \in \mathcal{P}(X)$ such that

$$(3.15) \quad Q_X(\varphi + V) = \int V d\mu_V - I(\mu_V)$$

then for any open $G \subset \mathcal{P}(X)$,

$$(3.16) \quad \liminf_{t \rightarrow \infty} \frac{1}{t} \log m\{x : \zeta_x^t \in G\} \geq -\inf\{I(\nu) : \nu \in G\}.$$

Proof. The entropy $h_\mu(F^1)$ is affine as a function of μ . Thus if it is also upper semicontinuous then by (3.4) and the duality theorem (see [AE, p. 201] and Theorem 9.12 in [W]) the functional $I(\nu)$ defined in Theorem 3.4 is convex conjugate to $Q_X(\varphi + V)$. Taking into account (3.10) and (3.12) one derives (3.13) from (1.1) applied to the measures $P_t = m_t = m/m(X_t)$ as explained at the beginning of this section. Indeed, in view of (3.10) and (3.12) the pressure of V given by (1.3) for $P_\lambda = P_t = m_t$ equals $Q_X(\varphi + V) - Q_\Phi(\varphi)$, and so the corresponding I -functional defined by (1.4) equals $I(\nu) + Q_\Phi(\varphi)$. Thus by (1.1),

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log m_t\{x : \zeta_x^t \in K\} \leq -\inf_{\nu \in K} I(\nu) - Q_\Phi(\varphi)$$

which together with (3.12) yield (3.13). The lower bound (3.16) follows from (1.2) by the second part of Theorem 2.1. \square

Next, I shall describe situations in smooth dynamical systems where conditions of Theorem 3.4 are satisfied. Relevant examples are expanding transformations and uniformly partially hyperbolic and hyperbolic dynamical systems. A C^2 -endomorphism $F = F^1$ of a compact Riemannian manifold M is called expanding if there exists $\gamma > 0$ such that

$$(3.17) \quad \|DF^n \xi\| \geq \gamma e^{\gamma n} \|\xi\| \quad \text{for all } \xi \in T_x M, \quad x \in M, \quad n \geq 0,$$

where DF is the differential of the map F , TM is the tangent bundle over M and $T_x M$ is the tangent space at x , and $\|\cdot\|$ is the Riemannian norm on TM . Let $J(x)$ denote the Jacobian of the linear map $DF: T_x M \rightarrow T_{Fx} M$ with respect to inner products induced by the Riemannian metric. Define

$$(3.18) \quad \varphi(x) = -\log J(x).$$

Clearly, φ is a smooth function. It is well know that the entropy $h_\nu(F)$ is upper semicontinuous for expanding maps since they are expansive (see [W and Ru]). Let m denote the normalized Riemannian volume on M . Then it is easy to see (cf. p. 136 in [K1]) that (3.8) holds true with φ given by (3.18) and with $A_\delta(t) > 0$ independent of t . This is a simple case of the volume lemma. It is

also well known (see [Ru, Chapter 7]) that for any Hölder continuous function ψ on M there exists a unique measure ν_ψ at which the supremum in (3.4) with $X = M$ is attained, i.e. $Q_M(\psi) = \int \psi d\nu_\psi + h_{\nu_\psi}(F)$. If V_1, V_2, \dots is a countable dense in $C(X)$ set of Hölder continuous functions on M then all functions $\psi = \varphi + \sum_{i=1}^n \beta_i V_i$ are also Hölder continuous and so (3.15) will be satisfied for a unique probability measure $\mu_\nu = \nu_\psi$. Thus both upper and lower bounds (3.13) and (3.16) hold true in this case.

Next, let $F^t : M \rightarrow M$, $t \in \mathbf{Z}$ or $t \in \mathfrak{R}$ be a group of diffeomorphisms of a locally compact Riemannian manifold M .

A compact F^t -invariant set $\Gamma \subset M$ (i.e. $F^t\Gamma = \Gamma$ for all t) is called uniformly partially hyperbolic (abbreviated uph) if there exist $\gamma > 0$ and DF^t -invariant splitting $T_\Gamma M = E^u \oplus E^{cs}$ into continuous subbundles E^u and E^{cs} of the tangent bundle over Γ such that the dimensions of E^u and E^{cs} are constant on Γ and given $\varepsilon > 0$ there is $N > 0$ so that for any $x \in \Gamma$ and all $t > N$,

$$(3.19) \quad \|DF_x^t \xi\| \geq e^{\gamma t} \|\xi\| \quad \text{provided } \xi \in E_x^{cs} \subset T_x M,$$

and

$$(3.20) \quad \|DF_x^t \eta\| \leq e^{-\varepsilon t} \|\eta\| \quad \text{provided } \eta \in E_x^{cs} \subset T_x M.$$

The set Γ is called hyperbolic in the discrete time case, i.e. when $t \in \mathbf{Z}$, if $E^{cs} = E^s$ and in place of (3.20) one has

$$(3.21) \quad \|DF_x^t \eta\| \leq e^{-\gamma t} \|\eta\| \quad \text{provided } \eta \in E^s.$$

The set Γ is called hyperbolic in the continuous time case if $E^{cs} = E^\circ \oplus E^s$ where E^s satisfies (3.21) and E° is the direction of the flow F^t , i.e. it is the one-dimensional subbundle generated by the vector field B satisfying

$$(3.22) \quad \frac{dF^t x}{dt} = B(F^t x).$$

A hyperbolic set Γ is said to be basic hyperbolic if the periodic orbits of $F^t|_\Gamma$ are dense in Γ , $F^t|_\Gamma$ is topologically transitive, and there exists an open set $U \supset \Gamma$ with $\Gamma = \bigcap_{-\infty < t < \infty} F^t U$. In the discrete time case let $J(x)$ be the Jacobian with respect to the Riemannian inner products of the linear map $DF : E_x^u \rightarrow E_{F^1 x}^u$ where $F = F^1$ and put $\varphi^u(x) = -\log J(x)$. In the continuous time case let $J_t(x)$ be the Jacobian of the linear map $DF^t : E_x^u \rightarrow E_{F^t x}^u$ and put $\varphi^u(x) = -\frac{dJ_t(x)}{dt}|_{t=0}$.

The function φ^u is known to be Hölder continuous even in uph cases since the subbundle E^u is Hölder continuous (see Appendix in [BK]). I shall need this fact only in hyperbolic cases where it is proved in [B and BR].

Let Γ be a locally maximal uph set meaning that there exists an open set U satisfying $\bigcap_{-\infty < t < \infty} F^t \bar{U} = \Gamma$. If m is the normalized Riemannian volume in \bar{U} then the relations (3.8)–(3.9) with $\varphi = \varphi^u$ follows from the volume

estimates in [Y]. If Γ is a basic hyperbolic set then (3.8) with $\varphi = \varphi^u$ follows with $A_\delta(t)$ independent of t from the volume lemma in [BR]. Recently I proved together with S. Newhouse that a global volume lemma holds true when F^t is an Axiom-A dynamical system on a compact manifold M with strongly transversal intersections of stable and unstable manifolds in the sense that (3.8) holds true with φ which is a continuous extension of φ^u from basic hyperbolic sets to the whole M . The upper semicontinuity of the entropy in the uph case and in the hyperbolic continuous time case holds true since these dynamical systems are entropy expansive (h -expansive: see [DGS]). In the hyperbolic discrete time case the upper semicontinuity of the entropy is simpler since the corresponding diffeomorphism is expansive (see [W]). Thus in the above cases one has the upper bound (3.13). The uniqueness of equilibrium states corresponding to Hölder continuous functions in the case of a basic hyperbolic set Γ is well known (see [B and BR]), and so if $X = \overline{U}$, $\bigcap_{-\infty < t < \infty} F^t \overline{U} = \Gamma$, and V_1, V_2, \dots is a countable dense in $C(X)$ set of Hölder continuous functions then for each $V = \varphi^u + \sum_{i=1}^n \beta_i V_i$ the equality (3.15) will be satisfied for a unique probability measure. In the last case one has the lower bound (3.16), as well. Thus for the above classes of dynamical systems the following result is the corollary of Theorem 3.4.

Theorem 3.5. *Let m denote the Riemannian volume on a C^2 locally compact manifold M .*

(i) *If $F^t: M \rightarrow M$ is a semigroup of C^2 expanding transformations and M is compact then (3.13) and (3.16) hold true with $I(\nu)$ given by (3.14), $X = M$, and φ defined by (3.18);*

(ii) *Let $F^t: M \rightarrow M$ be a group of C^2 diffeomorphisms. Suppose that an uph Γ is locally maximal in a neighborhood U with a compact closure \overline{U} . Then the upper bound (3.13) holds true with $X = \overline{U}$ and $\varphi = \varphi^u$;*

(iii) *If $F^t: M \rightarrow M$ is C^2 Axiom-A dynamical system with strong transversality condition then (3.13) is satisfied with $X = M$ and φ which is a continuous extension of the functions φ^u from the basic hyperbolic sets to the whole M ;*

(iv) *Let $F^t: M \rightarrow M$ be a group of C^2 diffeomorphisms having a basic hyperbolic set Γ in an open set U with a compact closure \overline{U} such that $\bigcap_{-\infty < t < \infty} F^t \overline{U} = \Gamma$. Then both the upper bound (3.13) and the lower bound (3.16) hold true with $X = \overline{U}$ and $\varphi = \varphi^u$.*

Remark 3.1. If Γ is a basic hyperbolic set and $m = \nu_\psi$ is an equilibrium state corresponding to a Hölder continuous function ψ on Γ then (3.8) will be satisfied with $\varphi = \psi - Q_X(\psi)$ (see [B and BR]), and so (3.13) and (3.16) remain true with such m and φ , and $X = \Gamma$ which generalizes [V].

Remark 3.2. If Γ is an uph set of F^t (in particular, if Γ is hyperbolic) then $-\int \varphi^u d\nu$ with $\nu \in \mathcal{M}_\Gamma^F$ is the integral of the sum of positive Lyapunov characteristic exponents corresponding to ν .

Remark 3.3. One should not expect reasonable lower bounds of the type (3.16) if the uniqueness of equilibrium states for a large set of functions does not take place. Let, for instance, F^t be an Axiom-A dynamical system having at least two attractors Γ_1 and Γ_2 . Then for certain Hölder continuous functions the equilibrium states will be all convex linear combinations of certain equilibrium states on Γ_1 and Γ_2 , i.e. the uniqueness will not hold true. Now take any $\nu_1 \in \mathcal{M}_{\Gamma_1}^F$ and $\nu_2 \in \mathcal{M}_{\Gamma_2}^F$. Let G be small neighborhood of $\nu = p_1\nu_1 + p_2\nu_2$ with $p_1, p_2 > 0$ and $p_1 + p_2 = 1$. Then by (3.14), $I(\nu) = p_1I(\nu_1) + p_2I(\nu_2)$. On the other hand, since there is no orbit of F^t which visits both small neighborhoods of Γ_1 and Γ_2 then $m\{x : \zeta_x^t \in G\} = 0$. Thus the left-hand side of (3.16) will always be $-\infty$.

Remark 3.4. In view of Corollary 2.1 the results of [OP, Y, and D] concerning large deviations for $m\{x : \frac{1}{n} \sum_{k=0}^{n-1} V(F^k x) \in U\}$, $U \subset \mathbb{R}^1$, $V \in C(X)$ are partial cases of Theorem 3.5.

Remark 3.5. The machinery of this paper works perfectly in the symbolic dynamic set up, in particular, for shifts of finite type and for Gibbs measures of general shifts, and so the corresponding results from [OP and Y] will follow from Theorem 3.4. Indeed, let $P = (p_{ij})$ be a $k \times k$ irreducible stochastic matrix and let m be the stationary measure of the corresponding Markov shift $F: X \rightarrow X$ where $X = \{x : x = (x_0, x_1, \dots), x_i \in \{1, \dots, k\}, p_{x_i x_{i+1}} > 0 \text{ for all } i \geq 0\}$. Set $\varphi(x) = \log p_{x_0 x_1}$ then $m\{x : x_0 = i_0, \dots, x_{n-1} = i_{n-1}\} = p_{i_0} \exp(\sum_{l=0}^{n-1} \varphi(F^l x))$, where $p = (p_1, \dots, p_k)$ is the stationary vector of P , and so the conditions of Proposition 3.2 are satisfied. The uniqueness of equilibrium states for Hölder continuous functions on X is well known (see [Ru, Chapter 5]) and so Theorem 3.5 is applicable. Moreover one can treat dynamical systems with the multidimensional time in the framework of the thermodynamic formalism from [Ru]. In the one-dimensional time case one derives directly assertions of Theorem 3.5 for homeomorphisms of Smale spaces (see [Ru, Chapter 7]). In the multidimensional time case the uniqueness of equilibrium states is more difficult to obtain which is the main obstacle for application of the above methods. Remark that the results of this paper yield large deviation estimates for hyperbolic rational maps considered in [L].

4. APPLICATIONS TO MARKOV PROCESSES

In this section the large deviations estimates from [DV2] will be derived as the corollary of Theorem 2.1. Suppose that Y_t is a Markov process on a locally compact metric space M with a discrete $t = 0, 1, 2, \dots$ or continuous time $t \in [0, \infty)$. I shall consider two cases. In the discrete time case $Y_t, t = 0, 1, \dots$ will be a Markov chain on M whose transition probabilities $P(x, \Gamma) = P\{Y_{t+1} \in \Gamma | Y_t = x\}$ have continuous positive densities $p(x, y)$ with respect to certain probability measure m on M such that $m(U) > 0$ where $U \subset M$ is an open set with a compact closure \bar{U} where Y_t will only be considered, in particular, U may coincide with the whole M which is supposed

then to be compact. In the continuous time case Y_t , $t \in [0, \infty)$ will be a diffusion process in a connected open subset U of a locally compact Riemannian manifold M with absorption on the boundary ∂U so that ∂U is smooth and $\bar{U} = U \cup \partial U$ is compact. In particular, U may coincide with M which is compact then. Put $\tau_U = \inf\{t \geq 0: Y_t \notin U\}$ and for each $V \in C(\bar{U})$ introduce the semigroups of operators

$$(4.1) \quad T_V(t)g(x) = E_x \chi_{\tau_U > t} g(Y_t) \exp\left(\sum_{k=1}^t V(Y_k)\right)$$

in the discrete time case, and

$$(4.2) \quad T_V(t)g(x) = E_x \chi_{\tau_U > t} g(Y_t) \exp\left(\int_0^t V(Y_s) ds\right)$$

in the continuous time case, where E_x denotes the expectation for the process Y_t starting at x and $\chi_A = 1$ if A occurs and $\chi_A = 0$ for otherwise. The operators $T_V(t)$ map $C(\bar{U})$ into itself and in the continuous time case when U is a proper subset of M then one considers $T_V(t)$ acting on the subspace $C_0(U) \subset C(U)$ of functions with zero data on ∂U . By the submultiplicative property of the norm the limit

$$(4.3) \quad Q(V) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \|T_V(t)\| = \inf_{t > 0} \frac{1}{t} \log \|T_V(t)\|$$

exists and is finite. It follows from the theory of positive operators (see [Kr]) both in the discrete time case and in the continuous time case with $U = M$ compact that the spectrum of the operator $T_V(1)$ is pure point and $e^{Q(V)}$ is its principal eigenvalue, i.e. the eigenvalue with the maximal absolute value. If U is a proper subset of M and Y_t is the diffusion in U with the absorption on ∂U then for any $V \in C(\bar{U})$ the number $e^{Q(V)}$ belongs to the spectrum of $T_V(1)$ (see [DV3]) but one can claim that $e^{Q(V)}$ is an eigenvalue only if V is Hölder continuous (see [Kr]). Let L be the generator of the diffusion Y_t which is a second order elliptic differential operator. Put $L_V = L + V$ then $Q(V)$ belongs to the spectrum of L_V for all $V \in C(\bar{U})$ and $Q(V)$ is the principal eigenvalue of L_V , i.e. the eigenvalue with the maximal real part, if V is Hölder continuous (see [Kr]).

By [DV1–DV2] for any $V \in C(\bar{U})$,

$$(4.4) \quad Q(V) = \sup_{\mu \in P(\bar{U})} \left(\int_{\bar{U}} V d\mu - I(\mu) \right)$$

where in the discrete time case

$$(4.5) \quad I(\mu) = - \inf_{u \in C_+(\bar{U})} \int_{\bar{U}} \log\left(\frac{Pu}{u}\right) d\mu,$$

with $Pu(x) = \int_{\bar{U}} p(x, y)u(y) dm(y)$, $C_+(\bar{U}) = \{u \in C(\bar{U}): u > 0\}$, and in the continuous time case

$$(4.6) \quad I(\mu) = - \inf_{u \in D_+} \int_{\bar{U}} \frac{Lu}{u} d\mu$$

with D_+ being the set of functions from the domain of L having positive upper and lower bounds. A simple proof of (4.4) goes as follows. The lower semicontinuity of $I(\mu)$ is built in its definitions (4.5) and (4.6), and so for any $V \in C(\bar{U})$ there exists $\mu_V \in \mathcal{P}(\bar{U})$ such that in the discrete time case,

$$\begin{aligned}
 (4.7) \quad & \sup_{\mu} \left(\int V d\mu - I(\mu) \right) = \int V d\mu_V - I(\mu_V) \\
 & = \int_{\bar{U}} V d\mu_V + \inf_{u \in C_+(\bar{U})} \int_{\bar{U}} \log \left(\frac{Pu}{u} \right) d\mu_V \\
 & \leq \int_{\bar{U}} \log \left(\frac{T_V(1)r_V}{r_V} \right) d\mu_V = Q(V)
 \end{aligned}$$

where r_V is the positive eigenfunction corresponding to the eigenvalue $Q(V)$ (see [Kr]). It turns out (see [K2]) that μ_V satisfies (4.7) if and only if it is an invariant measure of a Markov process with Markov transition operators.

$$(4.8) \quad \mathcal{T}_V(t)g(x) = e^{-Q(V)t}(r_V(x))^{-1}T_V(t)(gr_V)(x).$$

Then for any $u \in C(\bar{U})$,

$$(4.9) \quad \int_{\bar{U}} \mathcal{T}_V(t)u d\mu_V = \int_{\bar{U}} u d\mu_V.$$

Take $u = e^h$ then by Jensen's inequality

$$\log \mathcal{T}_V(t)u \geq \mathcal{T}_V(t) \log u = \mathcal{T}_V h,$$

and so by (4.9),

$$\int_{\bar{U}} \log \mathcal{T}_V(t)u d\mu_V \geq \int_{\bar{U}} \mathcal{T}_V(t)h d\mu_V = \int_{\bar{U}} h d\mu_V = \int_{\bar{U}} \log u d\mu,$$

i.e.

$$(4.10) \quad \int_{\bar{U}} \log \left(\frac{\mathcal{T}_V(t)u}{u} \right) d\mu_V \geq 0.$$

Since (4.10) holds true for any $u \in C_+(\bar{U})$ then by (4.5) and (4.8) it implies that

$$\int_{\bar{U}} V d\mu - I(\mu) \geq Q(V)$$

which together with (4.7) give (4.4) for $I(\mu)$ given by (4.5). In the continuous time case one remarks that for small $t > 0$,

$$(4.11) \quad \log \left(\frac{\mathcal{T}_V(t)u}{u} \right) \sim t \left(\frac{\mathcal{L}_V u}{u} \right)$$

where $\mathcal{L}_V = -Q(V) + V + (r_V)^{-1}L(r_V)$ is the generator of the semigroup $\mathcal{T}_V(t)$. In the continuous time case with $U = M$ this gives (4.4) immediately. If $U \neq M$, the time is continuous, and one deals with functions having zero data on ∂U then (4.7) and (4.10) need some additional justification. First, in this

case the eigenfunctions r_V exist only for Hölder continuous V , and so the above arguments may go through only for such V . But since by (4.3),

$$(4.12) \quad |Q(V_1) - Q(V_2)| \leq \|V_1 - V_2\|$$

then (4.4) proved for all Hölder continuous V will be true for all $V \in C(\bar{U})$. Secondly, one will have to deal only with functions with zero data on ∂U which behave near ∂U as r_V which does not cause any problems since the measures μ_V (for V Hölder continuous) have densities with respect to the volume which near ∂U behave as $(r_V)^2$ (for details, see Proposition 3.1 in [K2]).

Since the transition densities of Y_t are positive in U then it is easy to see that for any $x \in U$,

$$(4.13) \quad Q(V) = \lim_{t \rightarrow \infty} \frac{1}{t} \log(T_V(t)1)(x)$$

where $1(x) \equiv 1$. Considering the occupational measures $\zeta_\omega^t = \frac{1}{t} \sum_{l=0}^{t-1} \delta_{Y_l(\omega)}$ in the discrete time case and $\zeta_\omega^t = \frac{1}{t} \int_0^t \delta_{Y_s(\omega)} ds$ in the continuous time case one can rewrite (4.13) as

$$(4.14) \quad Q(V) = \lim_{t \rightarrow \infty} \frac{1}{t} \log E_x \chi_{\tau_U > t} \exp \left(\int_{\bar{U}} V d\zeta_\omega^t \right)$$

which has the form (1.3). I proved in Propositions 3.1 and 5.1 from [K2] that the measure μ_V giving the supremum in (4.4) is unique, at least, for all Hölder continuous functions V (in fact, in the discrete time case and in the continuous time case with $U = M$, μ_V is unique for all $V \in C(\bar{U})$).

Thus Theorem 2.1 with $X = \bar{U}$ yields

Theorem 4.1. *For any $x \in U$ and every closed set $K \subset \mathcal{P}(\bar{U})$,*

$$(4.15) \quad \limsup_{t \rightarrow \infty} \frac{1}{t} \log P_x \{\zeta^t \in K\} \leq -\inf\{I(\nu) : \nu \in K\},$$

and for every open set $G \subset \mathcal{P}(\bar{U})$,

$$(4.16) \quad \liminf_{t \rightarrow \infty} \frac{1}{t} \log P_x \{\zeta^t \in G\} \geq -\inf\{I(\nu) : \nu \in G\}$$

where $P_x\{A\} = E_x \chi_A$ and $I(\nu)$ is defined by (4.5) or (4.6) accordingly to the case under the consideration.

Remark 4.1. Employing the theory of Harris-recurrent Markov chains as in Lemmas 2.3 and 4.2 of [A2] in place of Doeblin's condition in Proposition 5.1 of [K2] one can generalize the above theory in the spirit of [A2] to wider classes of Markov chains since the formula (4.4) itself is proved in [DV1] under rather general circumstances.

Remark 4.2. The method works perfectly also for Markov chains $Y(t)$ with continuous time on a finite state space $\{1, \dots, N\}$. Let $P(t) = (p_{kl}(t))$ be the matrix of transition probabilities of $Y(t)$ for the time t and put $L = \frac{d}{dt} P(t)|_{t=0}$. Suppose that all nondiagonal elements of L are positive. Consider

the semigroup of operators $T_V(t)g(k) = E_k g(Y_t) \exp(\int_0^t V(Y_s) ds)$ where both $g = (g(1), \dots, g(N))$ and $V = (V(1), \dots, V(N))$ are N -vectors. This is a semigroup of positive matrices, and so by the Perron-Frobenius theorem $T_V(t)$ has a simple eigenvalue $\exp(tQ(V))$ with maximal absolute value where $Q(V)$ is an eigenvalue of the matrix L . Moreover, it follows that for each k ,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log E_k \exp \left(\int_0^t V(Y_s) ds \right) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \|T_V(t)\| = Q(V).$$

Again by the Perron-Frobenius theorem the eigenvector r_V corresponding to $Q(V)$ is positive, and so one can define a Markov operator $\mathcal{T}_V(t)$ by means of (4.8). It is easy to see that this operator has a unique invariant probability vector (measure on $\{1, \dots, N\}$) and proceeding as in (4.7)–(4.11) one concludes that μ_V is the only measure on which the supremum in (4.4) is attained with $I(u)$ given by (4.6) where $u = (u(1), \dots, u(N))$ and $\int \frac{Lu}{u} d\mu = \sum_k (u(k))^{-1} \mu(k) (Lu)(k)$. Thus the large deviation bounds (4.15) and (4.16) follow.

REFERENCES

- [A1] A. de Acosta, *Upper bounds for large deviations of dependent random vectors*, Z. Wahrsch. Verw. Gebiete **69** (1985), 551–565.
- [A2] ———, *Large deviations for vector-valued functionals of a Markov chain: lower bounds*, Ann. Probab. **16** (1988), 925–960.
- [AE] J.-P. Aubin and I. Ekeland, *Applied nonlinear analysis*, Wiley, New York, 1984.
- [B] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Math., vol. 470, Springer-Verlag, Berlin, 1975.
- [BK] M. Brin and Y. Kifer, *Dynamics of Markov chains and stable manifolds for random diffeomorphisms*, Ergodic Theory Dynamical Systems **7** (1987), 351–374.
- [BR] R. Bowen and D. Ruelle, *The ergodic theory of Axiom-A flows*, Invent. Math. **29** (1975), 181–202.
- [D] M. Denker, *Large deviations and the pressure function*, Preprint.
- [DGS] M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Math., vol. 257, Springer-Verlag, Berlin, 1976.
- [DS] N. Dunford and J. T. Schwartz, *Linear operators*, Part I, Interscience, New York, 1958.
- [DV1] M. D. Donsker and S. R. S. Varadhan, *On a variational formula for the principal eigenvalue for operators with maximum principle*, Proc. Nat. Acad. Sci. U.S.A. **72** (1975), 780–783.
- [DV2] ———, *Asymptotic evaluation of certain Markov processes expectations for large time. I*, Comm. Pure Appl. Math. **28** (1975), 1–47.
- [DV3] ———, *On the principal eigenvalue of second order elliptic differential operators*, Comm. Pure Appl. Math. **29** (1976), 595–621.
- [E] R. S. Ellis, *Large deviations for a general class of random vectors*, Ann. Probab. **12** (1984), 1–12.
- [G] J. Gärtner, *On large deviations from the invariant measure*, Theor. Probab. Appl. **22** (1977), 24–39.
- [K1] Y. Kifer, *Random perturbations of dynamical systems*, Birkhäuser, Boston, Mass., 1988.
- [K2] ———, *Principal eigenvalues, topological pressure, and stochastic stability of equilibrium states*, Israel J. Math. (to appear).
- [Kr] M. A. Krasnoselskii, *Positive solutions of operator equations*, Noordhoff, Groningen, 1964.

- [L] A. O. Lopes, *Entropy and large deviations*, Preprint.
- [OP] S. Orey and S. Pelikan, *Deviations of trajectory averages and the defect in Pesin's formula for Anosov diffeomorphisms*, Trans. Amer. Math. Soc. **315** (1989), 741–753.
- [R] R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, N.J., 1970.
- [Ru] D. Ruelle, *Thermodynamic formalism*, Encyclopedia of Math. and its Appl., vol. 5, Addison-Wesley, Reading, Mass., 1978.
- [T] Y. Takahashi, *Entropy functional (free energy) for dynamical systems and their random perturbations*, Taniguchi Sympos. SA, Katata, 1982, pp. 437–467.
- [V] S. Vaienti, *Computing the pressure for Axiom-A attractors by time series and large deviations for the Lyapunov exponent*, J. Statist. Phys. **56** (1989), 403–413.
- [W] P. Walters, *An introduction to ergodic theory*, Springer-Verlag, New York, 1982.
- [Y] L.-S. Young, *Large deviations in dynamical systems*, Preprint.

INSTITUTE OF MATHEMATICS, THE HEBREW UNIVERSITY OF JERUSALEM, JERUSALEM, ISRAEL