MEMOIRS
of the
American Mathematical Society

Number 944

Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging

Yuri Kifer

September 2009 • Volume 201 • Number 944 (third of 5 numbers) • ISSN 0065-9266

American Mathematical Society

MEMOIRS
of the
American Mathematical Society

Number 944

Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging

Yuri Kifer

September 2009 • Volume 201 • Number 944 (third of 5 numbers) • ISSN 0065-9266

American Mathematical Society
Providence, Rhode Island

2000 *Mathematics Subject Classification.*
Primary 34C29; Secondary 37D20, 60F10, 60J25.

Library of Congress Cataloging-in-Publication Data

Kifer, Yuri, 1948–
Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging / Yuri Kifer.
p. cm. — (Memoirs of the American Mathematical Society, ISSN 0065-9266 ; no. 944)
"Volume 201, number 944 (third of 5 numbers)."
Includes bibliographical references and index.
ISBN 978-0-8218-4425-0 (alk. paper)
1. Averaging method (Differential equations) 2. Large deviations. 3. Attractors (Mathematics) 4. Differential equations—Qualitative theory. I. Title.

QA372.K536 2009
515′.352—dc22 2009019381

Memoirs of the American Mathematical Society

This journal is devoted entirely to research in pure and applied mathematics.

Subscription information. The 2009 subscription begins with volume 197 and consists of six mailings, each containing one or more numbers. Subscription prices for 2009 are US$709 list, US$567 institutional member. A late charge of 10% of the subscription price will be imposed on orders received from nonmembers after January 1 of the subscription year. Subscribers outside the United States and India must pay a postage surcharge of US$65; subscribers in India must pay a postage surcharge of US$95. Expedited delivery to destinations in North America US$57; elsewhere US$160. Each number may be ordered separately; *please specify number* when ordering an individual number. For prices and titles of recently released numbers, see the New Publications sections of the *Notices of the American Mathematical Society.*

Back number information. For back issues see the *AMS Catalog of Publications.*

Subscriptions and orders should be addressed to the American Mathematical Society, P. O. Box 845904, Boston, MA 02284-5904 USA. *All orders must be accompanied by payment.* Other correspondence should be addressed to 201 Charles Street, Providence, RI 02904-2294 USA.

Memoirs of the American Mathematical Society (ISSN 0065-9266) is published bimonthly (each volume consisting usually of more than one number) by the American Mathematical Society at 201 Charles Street, Providence, RI 02904-2294 USA. Periodicals postage paid at Providence, RI. Postmaster: Send address changes to Memoirs, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

♾ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 14 13 12 11 10 09

Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging

Abstract

The work treats dynamical systems given by ordinary differential equations in the form $\frac{dX^\varepsilon(t)}{dt} = \varepsilon B(X^\varepsilon(t), Y^\varepsilon(t))$ where fast motions Y^ε depend on the slow motion X^ε (coupled with it) and they are either given by another differential equation $\frac{dY^\varepsilon(t)}{dt} = b(X^\varepsilon(t), Y^\varepsilon(t))$ or perturbations of an appropriate parametric family of Markov processes with freezed slow variables. In the first case we assume that the fast motions are hyperbolic for each freezed slow variable and in the second case we deal with Markov processes such as random evolutions which are combinations of diffusions and continuous time Markov chains. First, we study large deviations of the slow motion X^ε from its averaged (in fast variables Y^ε) approximation $\bar{X}^\varepsilon$. The upper large deviation bound justifies the averaging approximation on the time scale of order $1/\varepsilon$, called the averaging principle, in the sense of convergence in measure (in the first case) or in probability (in the second case) but our real goal is to obtain both the upper and the lower large deviations bounds which together with some Markov property type arguments (in the first case) or with the real Markov property (in the second case) enable us to study (adiabatic) behavior of the slow motion on the much longer exponential in $1/\varepsilon$ time scale, in particular, to describe its fluctuations in a vicinity of an attractor of the averaged motion and its rare (adiabatic) transitions between neighborhoods of such attractors. When the fast motion Y^ε does not depend on the slow one we arrive at a simpler averaging setup studied in numerous papers but the above fully coupled case, which better describes real phenomena, leads to much more complicated problems.

Received by the editor December 4, 2006.
2000 *Mathematics Subject Classification.* Primary: 34C29 Secondary: 37D20, 60F10, 60J25.
Key words and phrases. averaging, hyperbolic attractors, random evolutions,large deviations.
The author was partially supported by US–Israel BSF.

Preface

This work studies the long time behavior of slow motions in two scale fully coupled systems and it consists of two, essentially, independent parts which even have their own introductions. The first part is written having in mind readers with strong backgrounds in smooth dynamical systems and it deals with the case of Axiom A flows as fast motions. The second part is written for probabilists and it studies the case where fast motions are certain Markov processes such as random evolutions and, in particular, diffusions. As we noticed already in [**47**] principal large deviations results for Axiom A systems and Markov processes (satisfying, say, the Doeblin condition) follow from a similar scope of ideas and basic theorems though they rely on quite different machineries and backgrounds. Rate functionals of large deviations turn out to be Legendre transforms of corresponding topological pressures in the dynamical systems case while in the diffusion case they are obtained in the same way from principal eigenvalues of the corresponding infinitesimal generators. This intrinsic connection is further amplified by the fact that in the random diffusion perturbations of dynamical systems setup these principal eigenvalues converge to topological pressures when the perturbation parameter tends to zero (see [**46**]).

Usually, Markov processes are easier to deal with since we can use the Markov property there for free while in the dynamical systems case we have to look for some substitute. We felt that the first part of this work would be quite difficult to follow for most of probabilists in view of its heavy dynamical systems machinery. By this reason the second part is written in the way that it can be read independently of the first one and it relies only on the standard probabilistic background though the strategies of the proof in both parts are similar with the Markov property making arguments easier in the second part which also does not require to deal with geometric pecularities of the hyperbolic deterministic dynamics of the first part. In order to ensure a convenient independent reading of the second part we give full arguments there except for very few references to some general proofs in the first part which do not rely on the specific dynamical systems setup there. Still, the readers having sufficient background both in dynamical systems and Markov processes will certainly benefit from having proofs for both cases in one place and such exposition demonstrates boldly unifying features of these two quite different objects. We observe, that it could be possible to start with some very general (though quite unwieldy) assumptions which would enable us to prove similar results and then verify these assumptions for both cases we are dealing with but we believe that such exposition would make the paper quite difficult to read for both groups of mathematicians this work is addressed to.

Contents

Part 1

Hyperbolic Fast Motions

1.1. Introduction

Many real systems can be viewed as a combination of slow and fast motions which leads to complicated double scale equations. Already in the 19th century in applications to celestial mechanics it was well understood (though without rigorous justification) that a good approximation of the slow motion can be obtained by averaging its parameters in fast variables. Later, averaging methods were applied in signal processing and, rather recently, to model climate–weather interactions (see **[36]**, **[18]**, **[37]** and **[52]**). The classical setup of averaging justified rigorously in **[12]** presumes that the fast motion does not depend on the slow one and most of the work on averaging treats this case only. On the other hand, in real systems both slow and fast motions depend on each other which leads to the more difficult fully coupled case which we study here. This setup emerges, in particular, in perturbations of Hamiltonian systems which leads to fast motions on manifolds of constant energy and slow motions across them.

In this work we consider a system of differential equations for $X^\varepsilon = X^\varepsilon_{x,y}$ and $Y^\varepsilon = Y^\varepsilon_{x,y}$,

$$\frac{dX^\varepsilon(t)}{dt} = \varepsilon B(X^\varepsilon(t), Y^\varepsilon(t)), \quad \frac{dY^\varepsilon(t)}{dt} = b(X^\varepsilon(t), Y^\varepsilon(t)) \tag{1.1.1}$$

with initial conditions $X^\varepsilon(0) = x$, $Y^\varepsilon(0) = y$ on the product $\mathbb{R}^d \times \mathbf{M}$ where $\mathbf{M}$ is a compact $n_{\mathbf{M}}$-dimensional C^2 Riemannian manifold and $B(x,y)$, $b(x,y)$ are smooth in x, y families of bounded vector fields on $\mathbb{R}^d$ and on $\mathbf{M}$, respectively, so that y serves as a parameter for B and x for b. The solutions of (1.1.1) determine the flow of diffeomorphisms Φ^t_ε on $\mathbb{R}^d \times \mathbf{M}$ acting by $\Phi^t_\varepsilon(x,y) = (X^\varepsilon_{x,y}(t), Y^\varepsilon_{x,y}(t))$. Taking $\varepsilon = 0$ we arrive at the flow $\Phi^t = \Phi^t_0$ acting by $\Phi^t(x,y) = (x, F^t_x y)$ where F^t_x is another family of flows given by $F^t_x y = Y_{x,y}(t)$ with $Y = Y_{x,y} = Y^0_{x,y}$ being the solution of

$$\frac{dY(t)}{dt} = b(x, Y(t)), \quad Y(0) = y. \tag{1.1.2}$$

It is natural to view the flow Φ^t as describing an idealised physical system where parameters $x = (x_1, ..., x_d)$ are assumed to be constants (integrals) of motion while the perturbed flow Φ^t_ε is regarded as describing a real system where evolution of these parameters is also taken into consideration. Essentially, the proofs of this paper work also in the slightly more general case when B and b in (1.1.1) together with their derivatives depend Lipschitz continuously on ε (cf. **[55]**) but in order to simplify notations and estimates we do not consider this generalisation here.

Assume that the limit

$$\bar{B}(x) = \bar{B}_y(x) = \lim_{T\to\infty} T^{-1} \int_0^T B(x, F^t_x y)dt \tag{1.1.3}$$

exists and it is the same for "many" y's. For instance, suppose that μ_x is an ergodic invariant measure of the flow F^t_x then the limit (1.1.3) exists for μ_x–almost all y and is equal to

$$\bar{B}(x) = \bar{B}_{\mu_x}(x) = \int B(x,y)d\mu_x(y).$$

If $b(x,y)$ does not, in fact, depend on x then $F^t_x = F^t$ and $\mu_x = \mu$ are also independent of x and we arrive at the classical uncoupled setup. Here the Lipschitz continuity of B implies already that $\bar{B}(x)$ is also Lipshitz continuous in x, and so there exists a unique solution $\bar{X} = \bar{X}_x$ of the averaged equation

$$\frac{d\bar{X}^\varepsilon(t)}{dt} = \varepsilon\bar{B}(\bar{X}^\varepsilon(t)), \quad \bar{X}^\varepsilon(0) = x. \tag{1.1.4}$$

In this case the standard averaging principle says (see [**73**]) that for μ-almost all y,

$$\lim_{\varepsilon\to 0}\sup_{0\le t\le T/\varepsilon}|X^\varepsilon_{x,y}(t) - \bar{X}^\varepsilon_x(t)| = 0. \tag{1.1.5}$$

As the main motivation for the study of averaging is the setup of perturbations described above we have to deal in real problems with the fully coupled system (1.1.1) which only in very special situations can be reduced by some change of variables to a much easier uncoupled case where the fast motion does not depend on the slow one. Observe that in the general case (1.1.1) the averaged vector field $\bar{B}(x)$ in (1.1.3) may even not be continuous in x, let alone Lipschitz, and so (1.1.4) may have many solutions or none at all. Moreover, there may exist no natural well dependent on $x \in \mathbb{R}^d$ family of invariant measures μ_x since dynamical systems F^t_x may have rather different properties for different x's. Even when all measures μ_x are the same the averaging principle often does not hold true in the form (1.1.5), for instance, in the presence of resonances (see [**63**] and [**56**]). Thus even basic results on approximation of the slow motion by the averaged one in the fully coupled case cannot be taken for granted and they should be formulated in a different way requiring usually stronger and more specific assumptions.

If convergence in (1.1.3) is uniform in x and y then (see, for instance, [**52**]) any limit point $\bar{Z}(t) = \bar{Z}_x(t)$ as $\varepsilon \to 0$ of $Z^\varepsilon_{x,y}(t) = X^\varepsilon_{x,y}(t/\varepsilon)$ is a solution of the averaged equation

$$\frac{d\bar{Z}(t)}{dt} = \bar{B}(\bar{Z}(t)), \quad \bar{Z}(0) = x. \tag{1.1.6}$$

It is known that the limit in (1.1.3) is uniform in y if and only if the flow F^t_x on $\mathbf{M}$ is uniquely ergodic, i.e. it possesses a unique invariant measure, which occurs rather rarely. Thus, the uniform convergence in (1.1.3) assumption is too restrictive and excludes many interesting cases. Probably, the first relatively general result on fully coupled averaging is due to Anosov [**1**] (see also [**63**] and [**52**]). Relying on the Liouville theorem he showed that if each flow F^t_x preserves a probability measure μ_x on $\mathbf{M}$ having a C^1 dependent on x density with respect to the Riemannian volume m on $\mathbf{M}$ and μ_x is ergodic for Lebesgue almost all x then for any $\delta > 0$,

$$\text{mes}\{(x,y) : \sup_{0\le t\le T/\varepsilon}|X^\varepsilon_{x,y}(t) - \bar{X}^\varepsilon_x(t)| > \delta\} \to 0 \text{ as } \varepsilon \to 0, \tag{1.1.7}$$

where mes is the product of m and the Lebesgue measure in a relatively compact domain $\mathcal{X} \subset \mathbb{R}^d$. An example in Appendix to [**56**] shows that, in general, this convergence in measure cannot be strengthened to the convergence for almost all initial conditions and, moreover, in this example the convergence (1.1.5) does not hold true for any initial condition from a large open domain. Such examples exist due to the presence of resonances , more specifically to the "capture into resonance" phenomenon, which is rather well understood in perturbations of integrable Hamiltonian systems. Resonances lead there to the wealth of ergodic invariant measures and to different time and space averaging. It turns out (see [**11**]) that wealth of ergodic invariant measures with nice properties (such as Gibbs measures) for Axiom A and expanding dynamical systems also yields in the fully coupled averaging setup with the latter fast motions examples of nonconvergence as $\varepsilon \to 0$ for large sets of initial conditions (see Remark 1.2.12).

In Hamiltonian systems, which are a classical object for applications of averaging methods, the whole space is fibered into manifolds of constant energy. For some mechanical systems these manifolds have negative curvature with respect to the natural metric and their motion is described by geodesic flows there. Hyperbolic Hamiltonian systems were discussed, for instance, in [**64**] and a specific example of a particle in a magnetic field leading to such systems was considered recently in [**74**]. Of course, these lead to Hamiltonian systems which are far from integrable. Such situations fall in our framework and they are among main motivations for this work. This suggests to consider the equation (1.1.1) on a (locally trivial) fiber bundle $\mathcal{M} = \{(x,y) : x \in U,\, y \in M_x\}$ with a base U being an open subset in a Riemannian manifold N and fibers M_x being diffeomorphic compact Riemannian manifolds (see [**75**]). On the other hand, $\mathcal{M}$ has a local product structure and if $\|B\|$ is bounded then the slow motion stays in one chart during time intervals of order Δ/ε with Δ small enough. Hence, studying behavior of solutions of (1.1.1) on each such time interval separately we come back to the product space $\mathbb{R}^d \times \mathbf{M}$ setup and will only have to piece results together to see the picture on a larger time interval of length T/ε.

We assume in the first part of this work that $b(x,y)$ is C^2 in x and y and that for each x in a closure of a relatively compact domain $\mathcal{X}$ the flow F_x^t is Anosov or, more generally, Axiom A in a neighborhood of an attractor Λ_x. Let μ_x^{SRB} be the Sinai-Ruelle-Bowen (SRB) invariant measure of F_x^t on Λ_x and set $\bar{B}(x) = \int B(x,y)d\mu_x^{\mathrm{SRB}}(y)$. It is known (see [**16**]) that the vector field $\bar{B}(x)$ is Lipschitz continuous in x, and so the averaged equations (1.1.4) and (1.1.6) have unique solutions $\bar{X}^\varepsilon(t)$ and $\bar{Z}(t) = \bar{X}^\varepsilon(t/\varepsilon)$. Still, in general, the measures μ_x^{SRB} are singular with respect to the Riemannian volume on $\mathbf{M}$, and so the method of [**1**] cannot be applied here. We proved in [**55**] that, nevertheless, (1.1.7) still holds true in this case, as well, and, moreover, the measure in (1.1.7) can be estimated by $e^{-c/\varepsilon}$ with some $c = c(\delta) > 0$. The convergence (1.1.7) itself without an exponential estimate can be proved by another method (see [**57**]) which can be applied also to some partially hyperbolic fast motions . An extension of the averaging principle in the sense of convergence of Young measures is discussed in Section 1.11.

Once the convergence of $Z_{x,y}^\varepsilon(t) = X_{x,y}^\varepsilon(t/\varepsilon)$ to $\bar{Z}_x(t) = \bar{X}_x^\varepsilon(t/\varepsilon)$ as $\varepsilon \to 0$ is established it is interesting to study the asymptotic behavior of the normalized error

$$V_{x,y}^{\varepsilon,\theta}(t) = \varepsilon^{\theta-1}(Z_{x,y}^\varepsilon(t) - \bar{Z}_x(t)), \quad \theta \in [\frac{1}{2}, 1]. \tag{1.1.8}$$

Namely, in our situation it is natural to study the distributions $m\{y : V_{x,y}^{\varepsilon,\theta}(\cdot) \in A\}$ as $\varepsilon \to 0$ where m is the normalized Riemannian volume on $\mathbf{M}$ and A is a Borel subset in the space C_{0T} of continuous paths $\varphi(t)$, $t \in [0,T]$ on $\mathbb{R}^d$. We will obtain in this work large deviations bounds for $V_{x,y}^\varepsilon = V_{x,y}^{\varepsilon,1}$ which will give, in particular, the result from [**55**] saying that

$$m\{y : \|V_{x,y}^\varepsilon\|_{0,T} > \delta\} \to 0 \quad \text{as} \quad \varepsilon \to 0 \tag{1.1.9}$$

exponentially fast in $1/\varepsilon$ where $\|\cdot\|_{0,T}$ is the uniform norm on C_{0T}. However, the main goal of this work is not to provide another derivation of (1.1.9) but to obtain precise upper and lower large deviations bounds which not only estimate measure of sets of initial conditions for which the slow motion Z^ε exhibits substantially different behavior than the averaged one $\bar{Z}$ but also enable us to go further and to investigate much longer exponential in $1/\varepsilon$ time behavior of Z^ε. Namely, we will be able to study exits of the slow motion from a neighborhood of an attractor of the averaged one and transitions of Z^ε between basins of attractors of $\bar{Z}$. Such evolution, which becomes visible only on much longer than $1/\varepsilon$ time scales, is usually called adiabatic in the framework of averaging. In the simpler case when the fast motion does not depend on the slow one such results were discussed in [**49**]. Still, even in this uncoupled situation descriptions of transitions of the slow motion between attractors of the averaged one were not justified rigorously both in the Markov processes case of [**29**] and in the dynamical systems case of [**49**]. Extending these technique to three scale equations may exhibit stochastic resonance type phenomena producing a nearly periodic motion of the slowest motion which is described in Section 1.10 below. These problems seem to be important in the study of climate–weather interactions and they were discussed in [**18**] and [**37**] in the framework of a model describing transitions between steady climatic states with weather evolving as a fast chaotic system and climate playing the role of the slow motion. Such "very long" time description of the slow motion is usually impossible in the traditional averaging setup which deals with perturbations of integrable Hamiltonian systems. In the fully coupled situation we cannot work just with one hyperbolic flow but have to consider continuously changing fast motions which requires a special technique. In particular, the full flow Φ^t_ε on $\mathbb{R}^d \times \mathbf{M}$ defined above and viewed as a small perturbation of the partially hyperbolic system Φ^t plays an important role in our considerations. It is somewhat surprising that the "very long time" behavior of the slow motion which requires certain "Markov property type" arguments still can be described in the fully coupled setup which involves continuously changing fast hyperbolic motions. It turns out that the perturbed system still possesses semi-invariant expanding cones and foliations and a certain volume lemma type result on expanding leaves plays an important role in our argument for transition from small time were perturbation techniques still works to the long and "very long" time estimates.

It is plausible that moderate deviations type results can be proved for $V^{\varepsilon,\theta}_{x,y}$ when $1/2 < \theta < 1$ and that the distribution of $V^{\varepsilon,1/2}_{x,y}(\cdot)$ in y converges to the distribution of a Gaussian diffusion process in $\mathbb{R}^d$. Still, this requires somewhat different methods and it will not be discussed here. In this regard we can mention limit theorems obtained in [**14**] for a system of two heavy and light particles which leads to an averaging setup for a billiard flow. For the simpler case when b does not depend on x, i.e. when all flows F^t_x are the same, the moderate deviations and Gaussian approximations results were obtained previously in [**50**]. Related results in this uncoupled situation concerning Hasselmann's nonlinear (strong) diffusion approximation of the slow motion X^ε were obtained in [**56**].

We consider also the discrete time case where (1.1.1) is replaced by difference equations for sequences $X^\varepsilon(n) = X^\varepsilon_{x,y}(n)$ and $Y^\varepsilon(n) = Y^\varepsilon_{x,y}(n)$, $n = 0, 1, ...$ so that

$$X^\varepsilon(n+1) - X^\varepsilon(n) = \varepsilon B(X^\varepsilon(n), Y^\varepsilon(n)),$$
$$Y^\varepsilon(n+1) = F_{X^\varepsilon(n)} Y^\varepsilon(n),\ \ X^\varepsilon(0) = x, Y^\varepsilon(0) = y \tag{1.1.10}$$

where $B:\mathcal{X}\times\mathbf{M}\to\mathbb{R}^d$ is Lipschitz in both variables and the maps $F_x:\mathbf{M}\to\mathbf{M}$ are smooth and depend smoothly on the parameter $x\in\mathbb{R}^d$. Introducing the map

$$\Phi_\varepsilon(x,y)=(X^\varepsilon_{x,y}(1),Y^\varepsilon_{x,y}(1))=(x+\varepsilon B(x,y),F_xy)$$

we can also view this setup as a perturbation of the map $\Phi(x,y)=(x,F_xy)$ describing an ideal system where parameters $x\in\mathbb{R}^d$ do not change. Assuming that F_x, $x\in\mathbb{R}^d$ are C^2 depending on x families of either C^2 expanding transformations or C^2 Axiom A diffeomorphisms in a neighborhood of an attractor Λ_x we will derive large deviations estimates for the difference $X^\varepsilon_{x,y}(n)-\bar{X}^\varepsilon_x(n)$ where $\bar{X}^\varepsilon=\bar{X}^\varepsilon_x$ solves the equation

$$\frac{d\bar{X}^\varepsilon(t)}{dt}=\varepsilon\bar{B}(\bar{X}^\varepsilon(t)),\ \ \bar{X}^\varepsilon(0)=x \tag{1.1.11}$$

where $\bar{B}(x)=\int B(x,y)d\mu^{\mathrm{SRB}}_x(y)$ and μ^{SRB}_x is the corresponding SRB invariant measure of F_x on Λ_x. The discrete time results are obtained, essentially, by simplifications of the corresponding arguments in the continuous time case which enable us to describe "very long" time behavior of the slow motion also in the discrete time case. Since our methods work not only for fast motions being Axiom A diffeomorphisms but also when they are expanding transformations we can construct simple examples satisfying conditions of our theorems and exhibiting corresponding effects. In particular, we produce in Section 1.9 computational examples which demonstrate transitions of the slow motion between neighborhoods of attractors of the averaged system.

A series of related results for the case when ordinary differential equations in (1.1.1) are replaced by fully coupled stochastic differential equations appeared in [**45**], [**77**]–[**79**], [**66**], and [**5**]. Hasselmann's nonlinear (strong) diffusion approximation of the slow motion in the fully coupled stochastic differential equations setup was justified in [**10**]. When the fast process does not depend on the slow one such results were obtained in [**44**], [**29**], and [**54**]. Especially relevant for our results here is [**78**] and we employ some elements of the probabilistic strategy from this paper. Still, the methods there are quite different from ours and they are based heavily, first, on the Markov property of processes emerging there and, secondly, on uniformity and nondegeneracy of the fast diffusion term assumptions which cannot be satisfied in our circumstances as our deterministic fast motions are very degenerate from this point of view. Note that the proof in [**78**] contains a vicious cycle and substantial gaps which recently were essentially fixed in [**79**]. Some of the dynamical systems technique here resembles [**49**] but the dependence of the fast motion on the slow one complicates the analysis substantially and requires additional machinery. A series of results on Cramer's type asymptotics for fully coupled averaging with Axiom A diffeomorphisms as fast motions appeared recently in [**4**]–[**7**]. Observe that the methods there do not work for continuous time Axiom A dynamical systems considered here, they cannot lead, in principle, to the standard large deviations estimates of our work and they deal with deviations of X^ε from the averaged motion only at the last moment and not of its whole path. Various limit theorems for the difference equations setup (1.1.10) with partially hyperbolic fast motions were obtained recently in [**20**] and [**21**].

The study of deviations from the averaged motion in the fully coupled case seems to be quite important for applications, especially, from phenomenological point of view. In addition to perturbations of Hamiltonian systems mentioned above there are many non Hamiltonian systems which are naturally to consider from the beginning as a combination of fast and slow motions. For instance, Hasselmann [**36**] based his model of weather–climate interaction on the assumption that weather is a fast chaotic motion depending on climate as a slow motion which differs from the corresponding averaged motion mainly by a diffusion term. Though, as shown in [**54**], [**10**] and [**56**], this diffusion error term does not help in the study of large deviations which are responsible for rare transitions of the slow motion between attractors of the averaged one, the latter phenomenon can be described in our framework and it seems to be important in certain models of climate fluctuations (see [**18**] and [**37**]). Very slow nearly periodic motions appearing in the stochastic resonance framework considered in Section 1.10 may also fit into this subject in the discussion on "ice ages". Of course, it is hard to believe that real world chaotic systems can be described precisely by an Anosov or Axiom A flow but one may take comfort in the Chaotic Hypothesis [**32**]: " A chaotic mechanical system can be regarded for practical purposes as a topologically mixing Anosov system".

1.2. Main results

Let F^t be a C^2 flow on a compact Riemannian manifold $\mathbf{M}$ given by a differential equation

$$\frac{dF^ty}{dt} = b(F^ty), \ \ F^0y = y. \tag{1.2.1}$$

A compact F^t–invariant set $\Lambda \subset \mathbf{M}$ is called hyperbolic if there exists $\kappa > 0$ and the splitting $T_\Lambda \mathbf{M} = \Gamma^s \oplus \Gamma^0 \oplus \Gamma^u$ into the continuous subbundles $\Gamma^s, \Gamma^0, \Gamma^u$ of the tangent bundle $T\mathbf{M}$ restricted to Λ, the splitting is invariant with respect to the differential DF^t of F^t, Γ^0 is the one dimensional subbundle generated by the vector field b, and there is $t_0 > 0$ such that for all $\xi \in \Gamma^s$, $\eta \in \Gamma^u$, and $t \geq t_0$,

$$\|DF^t\xi\| \leq e^{-\kappa t}\|\xi\| \quad \text{and} \quad \|DF^{-t}\eta\| \leq e^{-\kappa t}\|\eta\|. \tag{1.2.2}$$

A hyperbolic set Λ is said to be basic hyperbolic if the periodic orbits of $F^t|_\Lambda$ are dense in Λ, $F^t|_\Lambda$ is topologically transitive, and there exists an open set $U \supset \Lambda$ with $\Lambda = \cap_{-\infty<t<\infty} F^tU$. Such a Λ is called a basic hyperbolic attractor if for some open set U and $t_0 > 0$,

$$F^{t_0}\bar{U} \subset U \quad \text{and} \quad \cap_{t>0} F^tU = \Lambda$$

where $\bar{U}$ denotes the closure of U. If $\Lambda = \mathbf{M}$ then F^t is called an Anosov flow.

1.2.1. Assumption. *The family $b(x,\cdot)$ in (1.1.2) consists of C^2 vector fields on a compact $n_{\mathbf{M}}$-dimensional Riemannian manifold $\mathbf{M}$ with uniform C^2 dependence on the parameter x belonging to a neighborhood of the closure $\bar{\mathcal{X}}$ of a relatively compact open connected set $\mathcal{X} \subset \mathbb{R}^d$. Each flow F^t_x, $x \in \bar{\mathcal{X}}$ on $\mathbf{M}$ given by*

$$\frac{dF^t_xy}{dt} = b(x, F^t_xy), \quad F^0_xy = y \tag{1.2.3}$$

possesses a basic hyperbolic attractor Λ_x with a splitting $T_{\Lambda_x}\mathbf{M} = \Gamma^s_x \oplus \Gamma^0_x \oplus \Gamma^u_x$ satisfying (1.2.2) with the same $\kappa > 0$ and there exists an open set $\mathcal{W} \subset \mathbf{M}$ and $t_0 > 0$ such that

$$\Lambda_x \subset \mathcal{W}, \ F^t_x\bar{\mathcal{W}} \subset \mathcal{W} \, \forall t \geq t_0, \ \text{and} \ \cap_{t>0} F^t_x\mathcal{W} = \Lambda_x \ \forall x \in \bar{\mathcal{X}}. \tag{1.2.4}$$

Let $J^u_x(t,y)$ be the absolute value of the Jacobian of the linear map $DF^t_x(y)$: $\Gamma^u_x(y) \to \Gamma^u_x(F^t_x y)$ with respect to the Riemannian inner products and set

$$\varphi^u_x(y) = -\frac{dJ^u_x(t,y)}{dt}\Big|_{t=0}. \tag{1.2.5}$$

The function $\varphi^u_x(y)$ is known to be Hölder continuous in y, since the subbundles Γ^u_x are Hölder continuous (see **[13]** and **[60]**), and $\varphi^u_x(y)$ is C^1 in x (see **[16]**).

Let $\mathcal{W}$ satisfy (1.2.4) and set $\mathcal{W}^t_x = \{y \in \mathcal{W} : F^s_x y \in \bar{\mathcal{W}} \ \forall s \in [0,t]\}$. A set $E \subset \mathcal{W}^t_x$ is called $(\delta,t)-$separated for the flow F_x if $y, z \in E$, $y \neq z$ imply $d(F^s_x y, F^s_x z) > \delta$ for some $s \in [0,t]$, where $d(\cdot,\cdot)$ is the distance function on $\mathbf{M}$. For each continuous function ψ on $\mathcal{W}$ set $P_x(\psi,\delta,t) = \sup\{\sum_{y\in E}\exp\int_0^t \psi(F^s_x y)ds$: $E \subset \mathcal{W}^t_x$ is $(\delta,t)-$ separated for $F_x\}$, $P_x(\psi,\delta,t) = 0$ if $\mathcal{W}^t_x = \emptyset$, and

$$P_x(\psi,\delta) = \limsup_{t\to\infty}\frac{1}{t}\log P_x(\psi,\delta,t).$$

The latter is monotone in δ, and so the limit

$$P_x(\psi) = \lim_{\delta\to 0} P_x(\psi,\delta)$$

exists and it is called the topological pressure of ψ for the flow F^t_x. Let $\mathcal{M}_x$ denotes the space of F^t_x-invariant probability measures on Λ_x then (see, for instance, **[60]**) the following variational principle

$$P_x(\psi) = \sup_{\mu\in\mathcal{M}_x}\Big(\int \psi d\mu + h_\mu(F^1_x)\Big) \tag{1.2.6}$$

holds true where $h_\mu(F^1_x)$ is the Kolmogorov–Sinai entropy of the time-one map F^1_x with respect to μ. If q is a Hölder continuous function on Λ_x then there exists a unique F^t_x-invariant measure μ^q_x on Λ_x, called the equilibrium state for $\varphi^u_x + q$, such that

$$P_x(\varphi^u_x + q) = \int(\varphi^u_x + q)d\mu^q_x + h_{\mu^q_x}(F^1_x). \tag{1.2.7}$$

We denote μ^0_x by μ^{SRB}_x since it is usually called the Sinai–Ruelle– Bowen (SRB) measure . Since Λ_x are attractors we have that $P_x(\varphi^u_x) = 0$ (see **[13]**).

For any probability measure ν on $\bar{\mathcal{W}}$ define

$$I_x(\nu) = \begin{cases} -\int\varphi^u_x d\nu - h_\nu(F^1_x) & \text{if } \nu\in\mathcal{M}_x \\ \infty & \text{otherwise.}\end{cases} \tag{1.2.8}$$

Then

$$P_x(\varphi^u_x + q) = \sup_\nu\Big(\int q d\nu - I_x(\nu)\Big).$$

Observe that by the Ruelle inequality (see, for instance, **[60]**, Theorem S.2.13), $I_x(\nu) \geq 0$, and so in view of Assumption 1.2.1 for any $\nu \in \mathcal{M}_x$,

$$I_x(\nu) \leq \sup_{y\in\Lambda_x}|\varphi^u_x(y)| \leq \sup_{x\in\bar{\mathcal{X}},y\in\Lambda_x}|\varphi^u_x(y)| < \infty. \tag{1.2.9}$$

It is known that $h_\nu(F^1_x)$ is upper semicontinuous in ν since hyperbolic flows are entropy expansive (see **[8]**). Thus $I_x(\nu)$ is a lower semicontinuous functional in ν and it is also convex (and affine on $\mathcal{M}_x$) since entropy h_ν is affine in ν (see, for instance, **[80]**). Hence, by the duality theorem (see **[2]**, p.201),

$$I_x(\nu) = \sup_{q\in\mathcal{C}(\mathbf{M})}\Big(\int q d\nu - P_x(\varphi^u_x + q)\Big).$$

Observe that this formula can be proved more directly. Namely, if we define $I_x(\nu)$ by it in place of (1.2.8) then (1.2.8) follows for $\nu \in \mathcal{M}_x$ from Theorem 9.12 in [**80**] and it is easy to show directly that $I_x(\nu)$ defined in this way equals ∞ for any finite signed measure ν which is not F_x-invariant.

Since we assume that the vector field B is C^1 in both arguments (here only continuity in y is needed) then for any $x, x' \in \mathcal{X}$ and $\alpha, \beta \in \mathbb{R}^d$ we can define $H(x, x', \beta) = P_x(< \beta, B(x', \cdot) > +\varphi_x^u)$ and $H(x, \beta) = H(x, x, \beta)$. Then

$$(1.2.10) \qquad H(x, x', \beta) = \sup_\nu \left(\int < \beta, B(x', y) > d\nu(y) - I_x(\nu)\right) \\ = \sup_{\alpha \in \mathbb{R}^d} \left(< \alpha, \beta > -L(x, x', \alpha)\right)$$

where

$$(1.2.11) \qquad L(x, x', \alpha) = \inf\{I_x(\nu) : \int B(x', y) d\nu(y) = \alpha\}$$

if $\nu \in \mathcal{M}_x$ satisfying the condition in brackets exists and $L(x, x', \alpha) = \infty$, otherwise. Since, $H(x, x', \beta)$ is convex and continuous the duality theorem (see [**2**], p.201) yields that

$$(1.2.12) \qquad L(x, x', \alpha) = \sup_{\beta \in \mathbb{R}^d} \left(< \alpha, \beta > -H(x, x', \beta)\right)$$

provided there exists a probability measure $\nu \in \mathcal{M}_x$ such that $\int B(x', y) d\nu(y) = \alpha$ and $L(x, x', \alpha) = \infty$, otherwise. Clearly, $L(x, x', \alpha)$ is convex and lower semi-continuous in all arguments and, in particular, it is measurable. We set also $L(x, \alpha) = L(x, x, \alpha)$.

Denote by C_{0T} the space of continuous curves $\gamma_t = \gamma(t)$, $t \in [0, T]$ in $\mathcal{X}$ which is the space of continuous maps of $[0, T]$ into $\mathcal{X}$. For each absolutely continuous $\gamma \in C_{0T}$ its velocity $\dot{\gamma}_t$ can be obtained as the almost everywhere limit of continuous functions $n(\gamma_{t+n^{-1}} - \gamma_t)$ when $n \to \infty$. Hence $\dot{\gamma}_t$ is measurable in t, and so we can set

$$(1.2.13) \quad S_{0T}(\gamma) = \int_0^T L(\gamma_t, \dot{\gamma}_t) dt = \int_0^T \inf\{I_{\gamma_t}(\nu) : \dot{\gamma}_t = \bar{B}_\nu(\gamma_t),\ \nu \in \mathcal{M}_{\gamma_t}\} dt,$$

where $\bar{B}_\nu(x) = \int B(x, y) d\nu(y)$, provided for Lebesgue almost all $t \in [0, T]$ there exists $\nu_t \in \mathcal{M}_{\gamma_t}$ for which $\dot{\gamma}_t = \bar{B}_{\nu_t}(\gamma_t)$, and $S_{0T}(\gamma) = \infty$ otherwise. It follows from [**13**] and [**16**] that

$$S_{0T}(\gamma) \geq S_{0T}(\gamma^u) = -\int_0^T P_{\gamma_t^u}(\varphi^u_{\gamma_t^u}) dt = 0$$

where γ_t^u is the unique solution of the equation

$$(1.2.14) \qquad \dot{\gamma}_t^u = \bar{B}(\gamma_t^u), \quad \gamma_0^u = x,$$

where $\bar{B}(z) = \bar{B}_{\mu_z^{\mathrm{SRB}}}(z)$, and the equality $S_{0T}(\gamma) = 0$ holds true if and only if $\gamma = \gamma^u$.

Define the uniform metric on C_{0T} by

$$\mathbf{r}_{0T}(\gamma, \eta) = \sup_{0 \leq t \leq T} |\gamma_t - \eta_t|$$

for any $\gamma, \eta \in C_{0T}$. Set

$$\Psi_{0T}^a(x) = \{\gamma \in C_{0T} : \gamma_0 = x,\ S_{0T}(\gamma) \leq a\}.$$

Since $L(x,\alpha)$ is lower semicontinuous and convex in α and, in addition, $L(x,\alpha)=\infty$ if $|\alpha|>\sup_{y\in\mathbf{M}}|B(x,y)|$ we conclude that the conditions of Theorem 3 in Ch.9 of [**41**] are satisfied as we can choose a fast growing minorant of $L(x,\alpha)$ required there to be zero in a sufficiently large ball and to be equal, say, $|\alpha|^2$ outside of it. As a result, it follows that S_{0T} is lower semicontinuous functional on C_{0T} with respect to the metric $\mathbf{r}_{0T}$, and so $\Psi^a_{0T}(x)$ is a closed set which plays a crucial role in the large deviations arguments below.

We suppose that the coefficients of (1.1.1) satisfy the following

1.2.2. ASSUMPTION. *There exists $K>0$ such that*

$$\|B(x,y)\|_{C^1(\mathcal{X}\times\mathbf{M})}+\|b(x,y)\|_{C^2(\mathcal{X}\times\mathbf{M})}\le K \tag{1.2.15}$$

where $\|\cdot\|_{C^i(\mathcal{X}\times\mathbf{M})}$ is the C^i norm of the corresponding vector fields on $\mathcal{X}\times\mathbf{M}$.

Set $\mathcal{X}_t=\{x\in\mathcal{X}:\,X^\varepsilon_{x,y}(s)\in\mathcal{X}$ and $\bar X^\varepsilon_x(s)\in\mathcal{X}$ for all $y\in\bar{\mathcal{W}},\ s\in[0,t/\varepsilon],\ \varepsilon>0\}$. Clearly, $\mathcal{X}_t\supset\{x\in\mathcal{X}:\,\inf_{z\in\partial\mathcal{X}}|x-z|\ge 2Kt\}$. The following is one of the main results of this paper.

1.2.3. THEOREM. *Suppose that $x\in\mathcal{X}_T$ and $X^\varepsilon_{x,y}$, $Y^\varepsilon_{x,y}$ are solutions of (1.1.1) with coefficients satisfying Assumptions 1.2.1 and 1.2.2. Set $Z^\varepsilon_{x,y}(t)=X^\varepsilon_{x,y}(t/\varepsilon)$ then for any $a,\delta,\lambda>0$ and every $\gamma\in C_{0T},\,\gamma_0=x$ there exists $\varepsilon_0=\varepsilon_0(x,\gamma,a,\delta,\lambda)>0$ such that for $\varepsilon<\varepsilon_0$,*

$$m\left\{y\in\mathcal{W}:\,\mathbf{r}_{0T}(Z^\varepsilon_{x,y},\gamma)<\delta\right\}\ge\exp\left\{-\frac{1}{\varepsilon}(S_{0T}(\gamma)+\lambda)\right\} \tag{1.2.16}$$

and

$$m\left\{y\in\mathcal{W}:\,\mathbf{r}_{0T}(Z^\varepsilon_{x,y},\Psi^a_{0T}(x))\ge\delta\right\}\le\exp\left\{-\frac{1}{\varepsilon}(a-\lambda)\right\} \tag{1.2.17}$$

where, recall, m is the normalized Riemannian volume on $\mathbf{M}$. The functional $S_{0T}(\gamma)$ for $\gamma\in C_{0T}$ is finite if and only if $\dot\gamma_t=\bar B_{\nu_t}(\gamma_t)$ for $\nu_t\in\mathcal{M}_{\gamma_t}$ and Lebesgue almost all $t\in[0,T]$. Furthermore, $S_{0T}(\gamma)$ achieves its minimum 0 only on γ^u satisfying (1.2.14) for all $t\in[0,T]$. Finally, for any $\delta>0$ there exist $c(\delta)>0$ and $\varepsilon_0>0$ such that for all $\varepsilon<\varepsilon_0$,

$$m\left\{y\in\mathcal{W}:\,\mathbf{r}_{0T}(Z^\varepsilon_{x,y},\bar Z_x)\ge\delta\right\}\le\exp\left(-\frac{c(\delta)}{\varepsilon}\right) \tag{1.2.18}$$

where $\bar Z_x=\gamma^u$ is the unique solution of (1.2.14).

Observe that (1.2.18) (which was proved already in [**55**] by a less precise large deviations argument) follows from (1.2.17) and the lower semicontinuity of the functional S_{0T} and it says, in particular, that $Z^\varepsilon_{x,\cdot}$ converges to $\bar Z_x$ in measure on the space $(\mathcal{W},m)$ with respect to the metric $\mathbf{r}_{0T}$. It is naturally to ask whether we have here also the convergence for m-almost all $y\in\mathcal{W}$. An example due to A.Neishtadt discussed in [**56**] shows that in the classical situation of perturbations of integrable Hamiltonian systems, in general, the averaging principle holds true only in the sense of convergence in measure on the space of intitial conditions but not in the sense of the almost everywhere convergence. This example concerns the simple system $\dot I=\varepsilon(4+8\sin\varphi-I)$, $\dot\varphi=I$ with the one dimensional slow motion I and the fast motion φ evolving on the circle while the corresponding averaged motion $J=\bar I$ satisfies the equation $\dot J=\varepsilon(4-J)$. The resonance occurs here only when $I=0$ but it suffices already to create troubles in the averaging principle.