Large Deviations in Fast-Slow Systems

Freddy Bouchet, Tobias Grafke, Tomás Tangarife, Eric Vanden-Eijnden

Large Deviations in Fast-Slow Systems

Freddy Bouchet,[1] Tobias Grafke,[2] Tomás Tangarife,[1] and Eric Vanden-Eijnden[2]

[1] *Laboratoire de Physique, ENS de Lyon, 46, allée d'Italie, F-69364 Lyon cedex 07, FRANCE*

[2] *Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012*

(Dated: October 7, 2015)

Abstract

The incidence of rare events in fast-slow systems is investigated via analysis of the large deviation principle (LDP) that characterizes the likelihood and pathway of large fluctuations of the slow variables away from their mean behavior – such fluctuations are rare on short time-scales but become ubiquitous eventually. This LDP involves an Hamilton-Jacobi equation whose Hamiltonian is related to the leading eigenvalue of the generator of the fast process, and is typically non-quadratic in the momenta – in other words, the LDP for the slow variables in fast-slow systems is different in general from that of any stochastic differential equation (SDE) one would write for the slow variables alone. It is shown here that the eigenvalue problem for the Hamiltonian can be reduced to a simpler algebraic equation for this Hamiltonian for a specific class of systems in which the fast variables satisfy a linear equation whose coefficients depend nonlinearly on the slow variables, and the fast variables enter quadratically the equation for the slow variables. These results are illustrated via examples, inspired by kinetic theories of turbulent flows and plasma, in which the quasipotential characterizing the long time behavior of the system is calculated and shown again to be different from that of an SDE.

PACS numbers:

I. INTRODUCTION

The evolution of many dynamical systems of interest involve the interplay of fast and slow variables. Examples include planetary motion, geophysical flows, climate-weather interaction models, macromolecules, etc. In such systems one is typically interested in the behavior of the slow variables on time-scales that are much longer than that over which the fast variables evolve. Under suitable conditions, the fast variables are adiabatically slaved to the slow ones, and the latter only feel the average effects of the former. When this is the case the evolution of the slow variables on their natural time-scale can be captured by a closed limiting equation for these variables alone that is obtained by averaging out the effect of the fast variables on the slow motions. This equation is valid in the limit when the scale separation between the fast and slow variables is infinitely wide, and it is an instance of the Law of Large Numbers (LLN) in the present context. Of course the scale separation is never infinite in reality and the slow variables also experience fluctuations above their average motion. Small fluctuations are captured by the Central Limit Theorem (CLT) which provides a linear stochastic differential equation (SDE) for the difference between the slow variables and their mean. Large deviations from the LLN away from the CLT scaling, on the other hand, can be characterized via a large deviation principle (LDP). On time scales that are of order one with respect to the clock of the slow variables, these large fluctuations are rare events. However, the LDP also captures the long time behavior of these variables when the effect of fluctuations is no longer negligible and large deviations from the LLN are no longer rare. This is the case, for example, if the limiting equation given by the LLN possesses multiple stable fixed points (or, more generally, multiple attractors). In such situations, fluctuations may eventually push the system from the vicinity of one such attractor to another, and the way this occurs is typically not captured by the CLT, but rather by the LDP. The aim of the present paper is to analyze in detail the structure of the LDP for a specific class of fast-slow systems and thereby provide concrete tools to characterize rare events as well as noise induced transitions in such systems. We note that this question has been investigated by many authors (see e.g. [5, 8–10, 12]), but few concrete results exist that permit to actually compute the Hamiltonian associated with the LDP: our goal here is to perform such calculations explicitly in simple examples or indicate how they could be performed numerically in more complicated ones. The examples we deal with have the peculiarity that, on the one hand the fast variables evolve linearly once the slow variable is held fixed, and on the other hand the fast variables act on the slow variable through a quadratic nonlinearity. Such a structure is very common in many examples related to the kinetic theories of both turbulent flows and plasma physics. It appears naturally for instance for the kinetic theory of the 2D Navier-Stokes or the family of quasigeostrophic models (see e.g. [1] and references therein) or the kinetic theory of plasma physics leading to either the Lenard-Balescu or the Vlasov equation (see e.g. [11]), or more generally the kinetic theory of systems with long range interactions (see e.g. [2]).

A. Set-up

We will consider fast-slow systems of the type

$$\begin{cases} \dot{X} = f(X,Y) \\ dY = \dfrac{1}{\alpha} b(X,Y)dt + \dfrac{1}{\sqrt{\alpha}}\sigma(X,Y)dW(t) \end{cases} \tag{1}$$

Here $f : \mathbb{R}^{m\times n} \to \mathbb{R}^m$ and $b : \mathbb{R}^{m\times n} \to \mathbb{R}^n$ are vector fields, $W(t) \in \mathbb{R}^p$ is a standard p-dimensional Wiener process, $\sigma : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times p}$, and $\alpha > 0$ is a parameter whose smallness measures the separation of time scale between the slow $X \in \mathbb{R}^m$ and the fast $Y \in \mathbb{R}^n$. To analyze the behavior of the slow variables X when $\alpha \ll 1$, let us introduce the virtual fast process

$$d\tilde{Y}_x = b(x, \tilde{Y}_x)d\tau + \sigma(x, \tilde{Y}_x)dW(\tau) \tag{2}$$

where x is fixed. This equation is obtained from the equation for Y in (1) by setting X to the fixed value x and rescaling time to the natural time scale of the fast Y, $\tau = t/\alpha$. Assume that the virtual fast process is ergodic at every x with respect to the invariant measure $\mu_x(dy)$ (which may depend parametrically on x) and that the following expectation exists

$$\begin{aligned} F(x) &= \int_{\mathbb{R}^n} f(x,y)\mu_x(dy) \equiv (Pf)(x) \\ &= \lim_{T\to\infty} \frac{1}{T}\int_0^T f(x, \tilde{Y}^x(\tau))d\tau \end{aligned} \tag{3}$$

where the second equality follows from ergodicity. Then for any $\varepsilon > 0$ and any fixed $T < \infty$ we have

$$\lim_{\alpha\to 0} \mathbb{P}^x \left(\sup_{0\le t\le T} |X(t) - \bar{X}(t)| < \varepsilon \right) = 1 \tag{4}$$

where $\mathbb{P}^x$ denotes the probability conditional on $X(0) = \bar{X}(0) = x$ and $\bar{X}$ satisfies the limiting equation

$$\dot{\bar{X}} = F(\bar{X}) \tag{5}$$

Property (4) is an instance of the LLN in the present context. For the reader's convenience, and also because the derivation of (5) involve formal asymptotic tools that we will need below, we recall this derivation in Appendix A.

The limiting equation in (5) is deterministic because we assumed that there is no explicit noise acting on the slow variables X in (1) and we let $\alpha \to 0$. In this limit the effect of the fast Y on the slow X completely averages out. For small but finite α, however, the slow variables are subject to fluctuations above their mean. To leading order, these fluctuations can be captured by the Central Limit Theorem (CLT). More precisely, denote by $C_{\tilde{f}}(x,\tau)$ the time-correlation matrix of $\tilde{f}(x,y) \equiv f(x,y) - F(x)$ along the virtual fast process $\tilde{Y}_x$,

$$C_{\tilde{f}}(x,\tau) = \int_{\mathbb{R}^n} \mathbb{E}^y \left(\tilde{f}(x,\tilde{Y}_x(\tau))\tilde{f}^T(x,y) + \tilde{f}(x,y)\tilde{f}^T(x,\tilde{Y}_x(\tau)) \right) \mu_x(dy) \tag{6}$$

where $\mathbb{E}^y$ denotes expectation over the virtual fast process conditional on $Y^x(0) = y$. Assume that the following integral exists

$$\begin{aligned} A(x) &= \int_0^\infty C_{\tilde{f}}(x,\tau)d\tau, \\ &= \lim_{T\to\infty} \frac{1}{T} \int_0^T d\tau \int_0^T d\tau' \tilde{f}(x,\tilde{Y}_x(\tau))\tilde{f}^T(x,\tilde{Y}_x(\tau')) \end{aligned} \tag{7}$$

and the following expectation exists

$$\begin{aligned} B(x) &= \int_{\mathbb{R}^n} \partial_x f(x,y)\mu_x(dy) \\ &\quad + \int_0^\infty d\tau \int_{\mathbb{R}^n} \left(\partial_y \mathbb{E}^y f(x,\tilde{Y}^x(\tau)) \right) \partial_x b(x,y)\mu_x(dy) \end{aligned} \tag{8}$$

where $(\partial_y f)_{i,j} = \partial f_i/\partial y_j$ and $(\partial_x b)_{i,j} = \partial b_i/\partial x_j$. Then on any interval $t \in [0,T]$ with $T < \infty$, the process

$$\tilde{\xi} = \frac{X - \bar{X}}{\sqrt{\alpha}} \tag{9}$$

converges in distribution towards the Gaussian process solution of

$$d\xi = B(\bar{X})\xi dt + \eta(\bar{X})dW(t) \tag{10}$$

where $\eta(x)$ is a $m \times m$ matrix such that $(\eta\eta^T)(x) = A(x)$ and $W(t)$ is a m-dimensional Wiener process – the derivation of (10) via formal asymptotic expansion techniques is recalled in Appendix B.

While the CLT indicates that typical fluctuations of the slow variables around their mean are of order $O(\sqrt{\alpha})$ on time-scales that are $O(1)$ in α, it does not permit to estimate the probability of deviations of order one away from this mean. On these time-scales, large deviations are expected to be rare, and their probability can be estimated by a LDP which takes the following form. Suppose that the following expectation over the virtual fast process exists

$$H(x,\theta) = \lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}^y \exp\left(\theta \cdot \int_0^T f(x, \tilde{Y}_x(\tau))d\tau\right) \tag{11}$$

and define the Lagrangian associated with this Hamiltonian

$$\mathcal{L}(x,y) = \sup_\theta (y \cdot x - H(x,\theta)) \tag{12}$$

as well as the action

$$I_T(x) = \int_0^T \mathcal{L}(x(t), \dot{x}(t))dt \tag{13}$$

Then this action permits to estimate the probability that the slow process wanders away from its mean behavior in the sense that for any $\Gamma \subset \{\gamma \in C([0,T], \mathbb{R}^m) : \gamma(0) = x\}$ the following LDP holds

$$\begin{aligned} -\inf_{\gamma\in\Gamma^\circ} I_T(\gamma) &\le \liminf_{\alpha\to 0} \alpha \log \mathbb{P}(X \in \Gamma) \\ &\le \limsup_{\alpha\to 0} \alpha \log \mathbb{P}(X \in \Gamma) \le -\inf_{\gamma\in\bar{\Gamma}} I_T(\gamma) \end{aligned} \tag{14}$$

The action (13) is also useful if one is interested in the slow motions on longer time-scales that can be $O(\alpha^{-1})$ or even $O(\exp(C/\alpha))$ for $C > 0$. On these time-scales, large deviations stop being rare and may, for example, lead to random transitions between the different attractors of the limiting equation in (5) if there are more than one. For example suppose that $D \subset \mathbb{R}^m$ is an open set that contains a single stable attracting point x_D of the limiting equation (5) and let us consider the first exit time from D

$$T_D = \inf\{t > 0 : X(t) \notin D\} \tag{15}$$

Then for any $x \in D$ we have

$$\lim_{\alpha\to 0} \alpha \log \mathbb{E}^x T_D = \inf_{y\in\partial D} V(x_D, y) \tag{16}$$

where $V(x,y)$ is the quasipotential

$$V(x,y) = \inf_{T>0} \inf_{\substack{\gamma(0)=x \\ \gamma(T)=y}} I_T(\gamma) \tag{17}$$

While the results above are well-known, in particular expression (11) for the Hamiltonian in the LDP (see [5, 8, 9, 12]), little attention has been given to the explicit form this LDP takes via calculation of the Hamiltonian. As stated above, our main aim here is to provide tools to perform such calculations, either analytically or numerically. Even if it goes without saying, we should stress that the Hamiltonian (11) cannot be deduced simply from the limiting equations (5) from the LLN and (10) from the CLT. Indeed, we may naively try to somehow recombine these two limiting equation and write down a nonlinear SDE whose LLN and CLT are precisely (5) and (10): since the noise in this SDE would be scaled by a factor $\sqrt{\alpha}$ one may be tempted to think that its associated LDP would also yield the LDP for the slow variables solution of (1). This program, however, is not achievable in general because there is no SDE for the slow variable alone that admits (5) as LLN and (9) as CLT, except in special cases. Indeed the drift term in this equation would have to be $F(X)$ to leading order in α, and setting $X = \bar{X} + \sqrt{\alpha}\,\xi$ and expanding in α would give a linear drift in the CLT equation equivalent to (10) involving $\partial_x F(\bar{X})$. And there lies the problem: $B(x) \neq \partial_x F(x)$ in general.

As we will see below, this problem is consistent with the fact that the LDP for the slow variables X in (1) is different in form to that of an SDE for X alone – in particular, the Hamiltonian (11) involved in this LDP is typically non-quadratic in the momenta conjugate to X, unlike that of an SDE with small noise. This essentially means that the effect of the noise induced by the fast variables on the slow ones cannot be modeled as a Gaussian white-noise in general. Below, we will re-derive the LDP stated above using formal asymptotic expansions tools. These results are complementary to the rigorous results proven e.g. in [5, 8–10, 12]. In particular we obtain an equation for the Hamiltonian which, to the best of our knowledge, is new and can be solved explicitly in some nontrivial examples.

B. Organization

The remainder of this paper is organized as follows. In Sec. II we derive the LDP for the fast-slow system (1). This derivation is formal, and to connect it with the results stated above we put it within the context of Donsker-Varadhan theory (Sec. II A) of large deviation and Gartner-Ellis theorem (Sec. II B). We also discuss in Sec. II C the link between the LLN and the LDP on the one hand, and the CLT and the LDP on the other. In Sec. III, we specialize the LPD to a class of fast-slow system for the which the equation for the Hamiltonian can be simplified. In Sec. IV we use these results to study a test-case example. Finally, some concluding remarks are given in Sec. V.

II. DERIVATION OF THE LARGE DEVIATION PRINCIPLE (LDP)

The slow-fast system in (1) defines a Markov process with generator $L = L_0 + \alpha^{-1} L_1$ where

$$\begin{aligned} L_0 &= f(x,y) \cdot \partial_x, \\ L_1 &= b(x,y) \cdot \partial_y + \tfrac{1}{2} a(x,y) : \partial_y \partial_y \end{aligned} \tag{18}$$

with $a(x,y) = (\sigma\sigma^T)(x,y)$, and $a(x,y) : \partial_y \partial_y$ is the contraction of the two tensors a and $\partial_y \partial_y$ (i.e. the trace of the product $a\partial_y\partial_y$). Given any suitable test function $\phi : \mathbb{R}^m \to \mathbb{R}$ the expectation

$$u(t,x,y) = \mathbb{E}^{x,y} \phi(X(t)) \tag{19}$$

therefore satisfies the backward Kolmogorov equation

$$\partial_t u = L_0 u + \frac{1}{\alpha} L_1 u, \qquad u(0) = \phi \tag{20}$$

Our derivation of the LDP for the slow variables X in (1) is based on formal asymptotic analysis of this equation. This approach is also at the core of the formal derivations of the LLN equation in (5) presented in Appendix A and the CLT equation in (10) presented in Appendix B.

Since the LDP for the slow variables X in (1) is concerned with estimating the probability of having a fluctuation of X away from $\bar{X}$ that is $O(1)$ in α, and since such probability is expected to be exponentially small in α^{-1} on time-scales that are $O(1)$ in α, let us consider the expectation

$$u(t,x,y) = \mathbb{E}^{x,y} \exp\left(\frac{1}{\alpha} h(X(t))\right), \tag{21}$$

so that u satisfies (20) for the initial condition

$$u(0,x,y) = \exp\left(\frac{1}{\alpha} h(x)\right), \tag{22}$$

and look for a solution of the type

$$u(t,x,y) = \Big(w(t,x,y) + O(\alpha)\Big) \exp\left(\frac{1}{\alpha} S(t,x,y)\right) \tag{23}$$

where both S and w are assumed to be independent of α. Inserting this ansatz in (20) and collecting terms of increasing power in α, we obtain at leading order, $O(\alpha^{-2})$:

$$\tfrac{1}{2} a : \partial_y S \partial_y S = 0 \tag{24}$$

This equation indicates that S depends on x but not y, $S(t,x,y) \equiv S(t,x)$, or, equivalently,

$$PS = S \tag{25}$$

where P is the expectation operator with respect to the invariant measure of the fast virtual process (2), see (3). The function S plays the role of the action of the LDP, since combining (21) and (23) with $S(t,x,y) \equiv S(t,x)$ implies that

$$S(x,t) = \lim_{\alpha\to 0} \alpha \log \mathbb{E}^{x,y} \exp\left(\frac{1}{\alpha} h(X(t))\right) \tag{26}$$

which is a variant of Varadhan's Lemma [3]. The question which remains to be addressed is how to estimate $S(x,t)$?

To this end, go to next order, $O(\alpha^{-1})$, where the ansatz (23) used in (20) gives

$$w\partial_t S = wL_0 S + L_1 w. \tag{27}$$

where we used the properties that L_0 is a first order operator, i.e. $L_0(fg) = fL_0 g + gL_0 f$, and $\partial_y S = L_1 S = 0$. Since $S = PS$ from (25), we can get an equation for S from (27) by dividing this equation by w (note that $w > 0$ since $u > 0$ by definition, see (21)) and applying P. This gives

$$\begin{aligned} \partial_t S &= L_0 S + w^{-1} L_1 w \\ &= PL_0 PS + P(w^{-1} L_1 w) \end{aligned} \tag{28}$$

Inserting (28) back in (27) gives

$$L_1 w + L_0 S w = \big(PL_0 PS + P(w^{-1} L_1 w)\big)\, w \tag{29}$$

This equation can be written explicitly as

$$b(x,y)\cdot\partial_y w + \tfrac{1}{2} a(x,y) : \partial_y\partial_y w + f(x,y)\cdot\partial_x S\; w = H(x,\partial_x S) w \tag{30}$$

where we defined

$$H(x,\partial_x S) = F(x)\cdot\partial_x S + \tfrac{1}{2}\int_{\mathbb{R}^n} \mu_x(dy)\, a(x,y) : \partial_y \log w\, \partial_y \log w \tag{31}$$

Here we used $w^{-1}L_1 w = L_1 \log w + \frac{1}{2} a(x,y) : \partial_y \log w\, \partial_y \log w$ and $PL_1 = 0$, to express the last term at the right hand side of (28) as

$$P(w^{-1}L_1 w) = \tfrac{1}{2} P(a(x,y) : \partial_y \log w\, \partial_y \log w) \tag{32}$$

Since x and $\partial_x S$ only enter as parameters in (30), this equation can be viewed an eigenvalue problem for the operator L_1, where $H(x,\partial_x S) - f(x,y)\cdot\partial_x S$ plays the role of (leading) eigenvalue. The function $H(x,\partial_x S)$ also is the Hamiltonian of the LDP we are trying to derive. Indeed, if we go back to equation (28) for S, we see that it can be expressed in terms of $H(x, \partial_x S)$ as the following Hamilton-Jacobi equation:

$$\partial_t S = H(x, \partial_x S), \qquad S(x, 0) = h(x) \tag{33}$$

This also means that if we introduce the Lagrangian associated with H via

$$\mathcal{L}(x, y) = \sup_\theta (y \cdot \theta - H(x, \theta)) \tag{34}$$

then the process X will satisfy a large deviation principle with respect to the action

$$I_T(x) = \int_0^T \mathcal{L}(x(t), \dot{x}(t))dt \tag{35}$$

and the solution to (33) can be expressed as

$$S(t, x) = \inf\{h + I_t(\tilde{x})\} \tag{36}$$

where the infimum is taken over all paths $\tilde{x} : [0, t) \to \mathbb{R}^m$ such that $\tilde{x}(t) = x$. Since $S(t, x)$ permits via (26) to evaluate the limit of α time the logarithm of the expectation in (26), this is again Varadhan's Lemma.

To make these results concrete, it remains to see whether we can analyze (30) and get a more convenient equation for $H(x, \theta)$. This will be done in Sec. III for a specific class of systems. Before going there, however, we show that our results are consistent with those of Donsker-Varadhan large deviations theory (Sec. II A) as well as with Ellis-Gartner theorem (Sec. II B). We also discuss the link between the LDP we just derived and the LLN on the one hand and the CLT on the other (Sec. II C).

A. Connection with Donsker-Varadhan Theory of Large Deviations

Denoting the solution of (30) for x and $\partial_x S = \theta$ fixed by $w(y, x, \theta)$, this solution can be expressed as the expectation

$$\begin{aligned} w(y, x, \theta) &= \lim_{T\to\infty} \mathbb{E}^y \exp\Big(-TH(x, \theta) + \theta \cdot \int_0^T f(x, \tilde{Y}_x(\tau))d\tau\Big) \\ &= \lim_{T\to\infty} e^{-TH(x,\theta)} \mathbb{E}^y \exp\Big(\theta \cdot \int_0^T f(x, \tilde{Y}_x(\tau))d\tau\Big) \end{aligned} \tag{37}$$

where $\tilde{Y}_x(t)$ is the virtual fast process, solution to (2), and $H(x, \theta)$ needs to be adjusted to make the limit converge to a finite, nonzero value. The requirement that such an adjustment be possible also gives the condition for the solution to (30) to exist and be uniquely given by (37). By taking the logarithm of the factor under the limit in (37), it is easy to see that this requirement imposes as a necessary condition that the following limit exists

$$H(x,\theta) = \lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}^y \exp\Big(\theta \cdot \int_0^T f(x, \tilde{Y}_x(\tau))d\tau\Big) \tag{38}$$

This is the alternative expression for H already given in (11) that is consistent with the one in Donsker-Varadhan large deviations theory.

B. Connection with Gartner-Ellis Theorem

Gartner-Ellis Theorem tells us that the Hamiltonian in (31) can also be defined by taking the limit as $\alpha \to 0$ of

$$\alpha \log \mathbb{E} \exp\left(\frac{1}{\alpha}\int_0^T \theta(t)\cdot f(x(t), Y_x(t))dt\right) \tag{39}$$

where $Y_x(t)$ is the solution of the second equation in (1) for a given $x(\cdot)$, i.e.

$$dY_x = \frac{1}{\alpha} b(x(t), Y_x)dt + \frac{1}{\sqrt{\alpha}}\sigma(x(t), Y_x)dW(t) \tag{40}$$

Letting

$$\hat{u}(t,y) = \mathbb{E}^y \exp\left(\frac{1}{\alpha}\int_0^t \theta(t')\cdot F(x(t'), Y_x(t'))dt'\right) \tag{41}$$

this function satisfies

$$\partial_t \hat{u} = \frac{1}{\alpha} L_1 \hat{u} + \frac{1}{\alpha}\theta(t)\cdot f(x(t), y)\,\hat{u} \tag{42}$$

Look for a solution of the type

$$\hat{u}(t,y) = \hat{w}(t,y)\exp\left(\frac{1}{\alpha}\hat{\phi}(t)\right) \tag{43}$$

Proceeding similarly as above then lead to the following system of equations for $\hat{\phi}$ and $\hat{w}$ (compare (30) and (33)):

$$b(x,y)\cdot\partial_y \hat{w} + \tfrac{1}{2} a(x,y) : \partial_y\partial_y \hat{w} + f(x,y)\cdot\theta\, \hat{w} = H(x,\theta)\hat{w} \tag{44}$$

and

$$\partial_t \hat{\phi} = H(x,\theta) \tag{45}$$

for the Hamiltonian H defined in (31). This also means that

$$\lim_{\alpha\to 0} \alpha \log \mathbb{E}\exp\left(\frac{1}{\alpha}\int_0^T \theta(t)\cdot f(x(t), Y_x(t))dt\right) = \int_0^T H(x(t), \theta(t))dt \tag{46}$$

which leads again to the action in (35).

C. Link between the LDP, the LLN, and the CLT

Suppose that we expand the exponential in (38) to second order in θ, take the expectation, then expand the logarithm to second order in θ as well, and finally take the limit as $T \to \infty$. This sequence of operations corresponds to making a cumulant expansion of the variable $\theta \cdot \int_0^T f(x, \tilde{Y}_x(\tau)) d\tau$ truncated to second order, and it gives the following quadratic approximation for H:

$$H_{\text{quad}}(x, \theta) = \theta \cdot F(x) + \tfrac{1}{2}\theta^T A(x)\theta \tag{47}$$

where $F(x)$ is defined in (3) and $A(x)$ in (7). This is the Hamiltonian for the LDP associated with the SDE

$$dX = F(X)dt + \sqrt{\alpha}\, \eta(X) dW(t) \tag{48}$$

in the limit as $\alpha \to 0$. The process defined by (48) satisfies a LLN principle with limiting equation (5), meaning that the LDP contains the information about the LLN. Yet, it also highlights the subtle (and well-know) differences between the CLT and the LDP. Indeed, the process defined by (48) satisfies a CLT with respect to

$$d\xi = \partial_x F(\bar{X})\xi dt + \eta(\bar{X}) dW(t) \tag{49}$$

This equation is not identical with (10) – their drift terms are different. In fact, it is easy to see that we would have to add deterministic terms of order $O(\sqrt{\alpha})$ in (48) in order that the CLT associated with this modified equation coincide with (10) – as already mentioned in the introduction, the addition of these terms would render the resulting equation unclosed since they depend on ξ rather than X. These additional terms do not affect the Hamiltonian of the LDP which would still be (47). This goes to show that the CLT cannot be recovered from the LDP: in the range of values for X where it applies, it contains finer information than that in the LDP.

Conversely, the LDP cannot be deduced from the CLT, as the actual H in (38) is different in general from its quadratic approximation H_{quad}. More precisely, $H = H_{\text{quad}}$ *iff* $\theta \cdot \int_0^T f(x, \tilde{Y}_x(\tau)) d\tau$ is Gaussian, and Marcinkiewicz's theorem states that in all other cases H is not a polynomial of θ of any order (i.e. its expansion in θ involves infinitely many terms). As we will see below in Sec. III, $H_{\text{quad}} = H$ if the fast Y enter linearly the equation for the slow X. When this is not the case and $H_{\text{quad}} \neq H$, H_{quad} can be used to describe moderate fluctuations of order $O(\alpha^\nu)$ with $\frac{1}{2} \leq \nu < 1$, that is, outside the range of validity of the CLT but only moderately so. To describe large fluctuations of order $O(1)$, however, we need to use the actual H in (38).

III. A SPECIFIC CLASS OF SYSTEMS

Next let us specialize (1) to systems in which the dynamics of the fast Y is linear and they enter quadratically the equation for the slow X

$$\begin{cases} \dot{X} = r(X) + Y^T s(X) + Y^T M(X) Y \\ dY = -\frac{1}{\alpha} L(X) Y + \frac{1}{\sqrt{\alpha}} \sigma(X) dW \end{cases} \tag{50}$$

where $X \in \mathbb{R}$ (the generalization to the vectorial case is straightforward but it makes notations more cumbersome, so we will stick to the scalar case here – see however Sec. IV C for an illustration with $X \in \mathbb{R}^2$), $Y \in \mathbb{R}^n$, $r : \mathbb{R} \to \mathbb{R}$ (e.g. $r(x) = -\nu x$ for some $\nu > 0$), $s : \mathbb{R} \to \mathbb{R}^n$, $L : \mathbb{R} \to \mathbb{R}^{n\times n}$ is a positive-definite matrix, $M : \mathbb{R} \to \mathbb{R}^{n\times n}$ is a symmetric matrix, and $\sigma : \mathbb{R} \to \mathbb{R}^{n\times n}$. For such systems, the equation (30) leading to the Hamiltonian of the LDP can be written down more explicitly.

To see how, notice first that the virtual fast process defined in (2) is given explicitly by

$$\tilde{Y}^x(\tau) = e^{-L(x)\tau} y + \int_0^\tau e^{-L(x)(\tau-\tau')} \sigma(x) dW(\tau') \tag{51}$$

As a result

$$\mu_x(dy) = (2\pi)^{-n/2} (\det C(x))^{-1/2} \exp\left(-\frac{1}{2} y^T C^{-1}(x) y\right) dy \tag{52}$$

where $C(x)$ is the equilibrium covariance matrix of $\tilde{Y}^x(\tau)$ satisfying the Lyapunov equation

$$L(x)C(x) + C(x)L^T(x) = a(x) \tag{53}$$

where $a(x) = (\sigma\sigma^T)(x)$. This means that the limiting equation (5) from the LLN reads

$$\dot{\bar{X}} = r(\bar{X}) + \text{tr}\left(C(\bar{X})M(\bar{X})\right) \tag{54}$$

and the equation (10) from the CLT reads

$$d\xi = r'(\bar{X})\xi dt + \text{tr}\left(C(\bar{X})M'(\bar{X})\right)\xi dt + g(\bar{X})\xi dt + \eta(\bar{X}) dW(t) \tag{55}$$

Here

$$g(x) = -\int_0^\infty \text{tr}\left(C(x)(L'(x) + [L'(x)]^T) e^{-L^T(x)\tau} M(x) e^{-L(x)\tau}\right) d\tau \tag{56}$$

and

$$\begin{aligned}\eta^2(x) &= \int_0^\infty s^T(x)\left(C(x)e^{-L^T(x)\tau} + e^{-L(x)\tau}C(x)\right)s(x)d\tau \\ &\quad + 4\int_0^\infty \text{tr}\left(C(x)e^{-L(x)\tau}M(x)e^{-L^T(x)\tau}C(x)M(x)\right)d\tau \\ &= s^T(x)L^{-1}(x)a(x)L^{-T}(x)s(x) \\ &\quad + 4\int_0^\infty \text{tr}\left(C(x)e^{-L(x)\tau}M(x)e^{-L^T(x)\tau}C(x)M(x)\right)d\tau\end{aligned} \tag{57}$$

where we used the Lyapunov equation (53).

Turning ourselves to the LDP next, (30) is explicitly

$$\begin{aligned}&-L(x)y\cdot\partial_y w + \tfrac{1}{2}a(x):\partial_y\partial_y w \\ &\quad + \left(r(x) + y^T\cdot s(x) + y^T M(x)y\right)\theta\, w = H(x,\theta)w\end{aligned} \tag{58}$$

where $\theta = \partial_x S$ and

$$\begin{aligned}H(x,\theta) &= (r(x) + \text{tr}(C(x)M(x)))\,\theta \\ &\quad + \tfrac{1}{2}\int \mu_x(dy)\, a(x):\partial_y \log w\, \partial_y \log w\end{aligned} \tag{59}$$

Look for a solution of the form

$$w = \exp(y^T m(x,\theta) + y^T N(x,\theta)y) \tag{60}$$

for some unknown $m(x,\theta)$ and $N(x,\theta)$. Then (dropping the dependencies in x for simplicity of notation)

$$H(\theta) = (r + \text{tr}(CM))\,\theta + \tfrac{1}{2}m^T(\theta)am(\theta) + 2\,\text{tr}\,(CN(\theta)aN(\theta)) \tag{61}$$

and (58) becomes

$$\begin{aligned}&-y^T N(\theta)Ly - y^T L^T N(\theta)y - y^T L^T m(\theta) + \text{tr}(aN(\theta)) \\ &+ 2y^T N(\theta)am(\theta) + 2y^T N(\theta)aN(\theta)y \\ &+ (y^T s + y^T My)\theta = \text{tr}(CM)\theta + 2\,\text{tr}\,(aN(\theta)CN(\theta))\end{aligned} \tag{62}$$

Collecting the terms that are of order 0, 1, and 2 in y, respectively, gives the equations

$$\text{tr}(aN(\theta)) = \text{tr}(CM)\theta + 2\,\text{tr}\,(CN(\theta)aN(\theta)) \tag{63}$$

$$(L^T - 2N(\theta)a)m(\theta) = s\,\theta \tag{64}$$

$$N(\theta)L + L^T N(\theta) = 2N(\theta)aN(\theta) + M\theta \tag{65}$$

By right multiplying (65) by C, taking the trace, and using the Lyapunov equation (53), it is easy to see that the result is (63), i.e. if (65) is satisfied then (63) automatically holds. We can also solve (64) in $m(\theta)$ to get

$$m(\theta) = (L^T - 2N(\theta)a)^{-1} s\,\theta \tag{66}$$

and be left with solving (65) in $N(\theta)$ – note that this equation may have more than one solution, and we should take the one such that $N(0) = 0$, so that $H(0) = 0$. Inserting (63) and (66) in (61) we then obtain the Hamiltonian of the LDP in terms of $N(\theta)$ alone

$$H(\theta) = r\theta + \text{tr}\,(aN(\theta)) + \tfrac{1}{2} s^T (L^T - 2N(\theta)a)^{-T} a (L^T - 2N(\theta)a)^{-1} s\,\theta^2 \tag{67}$$

The solution to (65) is not available explicitly in general. There is one trivial case, however, namely when $M = 0$. In this case it is easy to see that $N(\theta) = 0$, $m(\theta) = L^{-T} s\,\theta$ and the Hamiltonian is quadratic

$$H(\theta) = r\theta + \tfrac{1}{2} s^T L^{-1} a L^{-T} s\,\theta^2 \tag{68}$$

This Hamiltonian is that of the LDP associated with SDE

$$dX = r(X)dt + \sqrt{\alpha}\eta_0(X)dW(t) \tag{69}$$

where $\eta_0(x)$ is the factor defined in (57) evaluated at $M = 0$. It is easy to see that this the limiting equation from the LLN for this equation is (54) (with $M = 0$) and the equation from the CLT is (55) (again with $M = 0$). Thus, if the fast variables are Gaussian, and their action on the slow one is linear, the LDP contains all the information about the CLT, and the SDE (69) can be used to investigate large deviations. Notice that this includes nontrivial situations with metastability, when $\dot{x} = r(x)$ has more than one stable fixed point and one is interested in the rate and mechanism of transitions between these points.

Another case where the Hamiltonian can be computed explicitly is the following one: assume that a is invertible, and that the following conditions hold:

$$La = aL^T \quad \text{and} \quad L^T M a = M a L^T. \tag{70}$$

Then, it is straightforward to check that whenever $B(\theta) = \left(L^T\right)^2 - 2\theta M a$ admits a square root, the matrix

$$N(\theta) = \frac{1}{2}\left[L^T - \sqrt{B(\theta)}\right] a^{-1} \tag{71}$$

satisfies $N(\theta)L = L^T N(\theta)$, and is a solution of (65). Using (70), we also have that $aB(\theta)$ is symmetric, so inverting (71) and using (65), we prove that $N(\theta)$ is symmetric, which is consistent with its definition (60). It is then also easy to prove that $a\sqrt{B(\theta)}^{-1} = \sqrt{B^T(\theta)}^{-1} a$, so the Hamiltonian (67) reads

$$H(\theta) = r\theta + \tfrac{1}{2}\mathrm{tr}\left(L - \sqrt{L^2 - 2\theta a M}\right) + \tfrac{1}{2} s^T \left[L^2 - 2\theta a M\right]^{-1} a s\, \theta^2\,, \tag{72}$$

whenever the square roots and inverses appearing in this equation exist (which is the case for $\theta = 0$). Note that the square root in the trace should be chosen such that $H(\theta = 0) = 0$.

IV. A CASE STUDY

In this section, we illustrate our results on the following test case example:

$$\begin{cases} \dot{X} = \dfrac{1}{K}\displaystyle\sum_{k=1}^{K} Y_k^2 - \nu X \\ dY_k = -\dfrac{1}{\alpha}\gamma(X) Y_k dt + \dfrac{\sigma}{\sqrt{\alpha}}\, dW_k, \qquad k = 1, \dots, K \end{cases} \tag{73}$$

where W_k are independent Wiener processes, $\alpha > 0$, $\nu > 0$ and σ are parameters and $\gamma(X) > 0$ is a function to be specified later. (73) consists of a scalar X (the vectorial case is discussed below in Sec. IV C) coupled to K Ornstein-Uhlenbeck processes Y_k which each feel an independent noise and whose common decay rate depends on X. (73) is in the class of (50) with

$$\begin{aligned} &r(x) = -\nu x, \quad s(x) = 0, \quad M(x) = K^{-1}\, \mathbf{\mathit{Id}}, \\ &L(x) = \gamma(x)\, \mathbf{\mathit{Id}}, \quad \sigma(x) = \sigma\, \mathbf{\mathit{Id}} \end{aligned} \tag{74}$$

We will be interested in studying (73) in the limits $\alpha \to 0$ and $K \to \infty$ – the former limit is in the realm of the formalism developed here, whereas the latter can be estimated by direct calculation. As we will below these two limits commute.

A. The limit as $\alpha \to 0$

Using the formulas given in section III, it is easy to see that the LLN equation (54) becomes

$$\dot{\bar{X}} = \frac{\sigma^2}{2\gamma(\bar{X})} - \nu \bar{X} \tag{75}$$

and the CLT equation (55) is

$$d\xi = -\left(\nu + \frac{\sigma^2\gamma'(\bar{X})}{2\gamma^2(\bar{X})}\right)\xi dt + \frac{\sigma^2}{\sqrt{2K\gamma^3(\bar{X})}} dW(t) \tag{76}$$

As far as the LDP is concerned, note that (73) is of the form

$$\dot{X} = \frac{1}{K}\sum_{k=1}^{K} f_0\left(X, Y_k\right) + f_1(X) \tag{77}$$

where Y_k are *i.i.d.* random processes. From (38), we see that the Hamiltonian reads

$$H(x,\theta) = \theta f_1(x) + K H_0\left(x, \frac{\theta}{K}\right) \tag{78}$$

with H_0 the one-particle Hamiltonian

$$H_0(x,\theta) = \lim_{T\to\infty}\frac{1}{T}\log\mathbb{E}^{y_1}\exp\Big(\theta\cdot\int_0^T f_0(x,\tilde{Y}_{1,x}(\tau))d\tau\Big). \tag{79}$$

To get the Lagrangian associated with this Hamiltonian, we solve

$$\dot{x} = \frac{\partial H}{\partial\theta} = f_1(x) + \frac{\partial H_0}{\partial\theta}\left(x, \frac{\theta}{K}\right) \tag{80}$$

and get θ/K as a function of $(x,\dot{x})$. Denoting this solution as $\theta = K\vartheta(x,\dot{x})$ and using (78) and (80), we then deduce that the Lagrangian is proportional to K and reads

$$\begin{aligned}\mathcal{L}(x,\dot{x}) &= \dot{x}\theta - H(x,\theta)\\ &= K\left(\vartheta(x,\dot{x})\frac{\partial H_0}{\partial\theta}\left(x,\vartheta(x,\dot{x})\right) - H_0\left(x,\vartheta(x,\dot{x})\right)\right).\end{aligned} \tag{81}$$

As a result, the path that minimizes the action (35) (instanton) does not depend on K, but the probabilistic weight of this path decreases exponentially with K. The quasi-potential $V(x)$, which by definition is the solution of

$$H(x,\partial_x V) = 0 \tag{82}$$

with H given by (78), is then also proportional to K.

Going back to the particular case (73), the one-particle Hamiltonian (79) can be computed using (72). Indeed, we are now solving a one-dimensional problem, so the conditions (70) are fulfilled and (72) is simply

$$H_0(x,\theta) = \frac{1}{2}\left[\gamma(x) - \sqrt{\gamma^2(x) - 2\sigma^2\theta}\right], \tag{83}$$

defined whenever $\theta \leqslant \gamma^2(x)/2\sigma^2$. To get the Lagrangian associated with this Hamiltonian, we solve

$$\dot{x} + \nu x = \frac{\partial H_0}{\partial\theta}\left(x, \frac{\theta}{K}\right) = \frac{1}{2}\frac{\sigma^2}{\sqrt{\gamma^2(x) - 2\sigma^2\theta/K}}. \tag{84}$$

A solution requires $\dot{x} + \nu x > 0$, as should be expected from (73). We then obtain

$$\vartheta(x, \dot{x}) = \frac{\theta}{K} = \frac{\gamma^2(x)}{2\sigma^2} - \frac{\sigma^2}{8(\dot{x} + \nu x)^2} \tag{85}$$

and get

$$\mathcal{L}(x, \dot{x}) = \frac{K}{8\sigma^2} \frac{|2\gamma(x)(\dot{x} + \nu x) - \sigma^2|^2}{\dot{x} + \nu x} \tag{86}$$

whenever $\dot{x} + \nu x > 0$.

The quasi-potential is given by (82), using (78) and (83) it is easy to see that this implies that either $\partial_x V \equiv V' = 0$, or

$$\frac{V'(x)}{K} = \frac{\nu x \gamma(x) - \frac{1}{2}\sigma^2}{\nu^2 x^2} \tag{87}$$

This result should be compared with the one obtained from the quadratic approximation to H,

$$H_{\text{quad}} = -\nu x \theta + K \left[\frac{\sigma^2 \theta / K}{2\gamma(x)} + \frac{\sigma^4 (\theta/K)^2}{4\gamma^3(x)} \right], \tag{88}$$

for which we deduce

$$\frac{V'_{\text{quad}}(x)}{K} = \frac{4\gamma^3(x)}{\sigma^4} \left(\nu x - \frac{\sigma^2}{2\gamma(x)} \right) \tag{89}$$

The potentials $V(x)$ and $V_{\text{quad}}(x)$ are different in general. To give a concrete example, consider the case $\gamma(x) = x^4/10 - x^2 + 3$, $\nu = 1$, $\sigma = \sqrt{3}$, which leads to bistability of the slow process $x(t)$. The potentials V and V_{quad} are represented in figure 1. The extrema of V_{quad} are also extrema of V, and the second derivatives of these two potentials are the same at these extrema, as should be expected. Figure 2 illustrates this last point for the potential minima. However, we see that the global shape of the potentials are very different, and in particular the energy barrier between the attractors of V_{quad} is almost twice the one between the attractors of V, which means that the probability of rare transitions obtained from the quadratic approximation in this case will be much lower than the actual one.

B. The limit as $K \to \infty$

Interestingly, we can corroborate the results obtained in Sec. IV A by taking the limit as $K \to \infty$ first. If we define

$$E = \frac{1}{K} \sum_{k=1}^{K} Y_k^2 \tag{90}$$

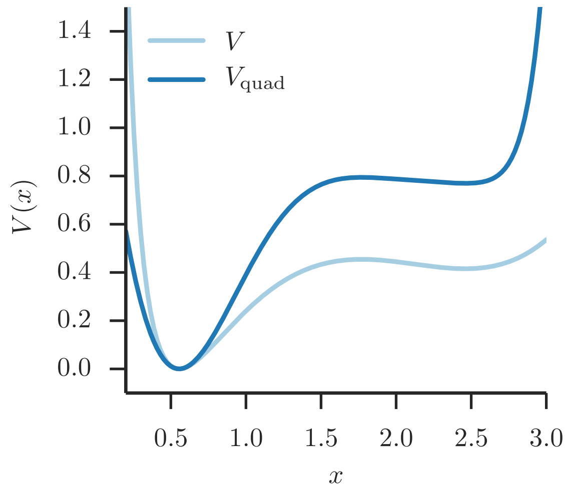

FIG. 1: The potential $V(x)/K$ and the one obtained from a quadratic approximation of the Hamiltonian $V_{\text{quad}}(x)/K$, see (89), for $\gamma(x) = x^4/10 - x^2 + 3$, $\nu = 1$, $\sigma = \sqrt{3}$. The quadratic potential obtained from the quadratic approximation is quite different from the actual potential.

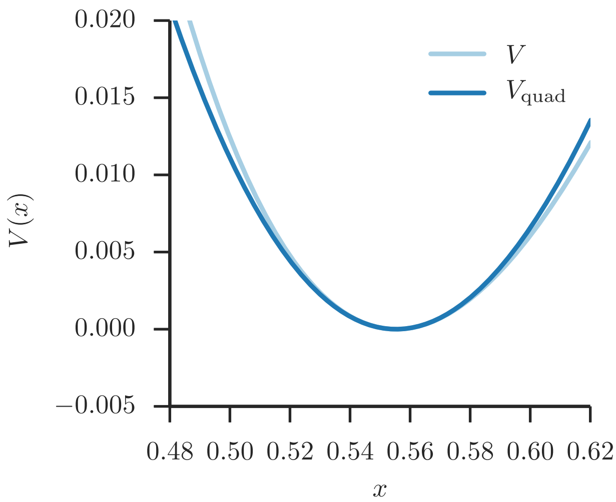

FIG. 2: The potential $V(x)/K$ and the one obtained from a quadratic approximation of the Hamiltonian $V_{\text{quad}}(x)/K$ in the vicinity of the main attractor.

it is easy to see that (73) can be rewritten as

$$\begin{cases} \dot{X} = E - \nu X \\ dE = -\dfrac{2}{\alpha}\gamma(X)Edt + \dfrac{1}{\alpha}\sigma^2 dt + 2\dfrac{\sigma}{\sqrt{\alpha K}}\sqrt{E}\,dW \end{cases} \tag{91}$$

where we used the identity

$$\frac{1}{K}\sum_{k=1}^{K} Y_k dW_k = K^{-1/2}\sqrt{E}\, dW \qquad \text{(in law)} \tag{92}$$

Since the noise term in (91) is small when K is large, we can use large deviation theory to analyze the behavior of the system in the limit as $K \to \infty$. The Freidlin-Wentzell action associated with (91) reads

$$I_T^\alpha(E,x) = \frac{\alpha}{8\sigma^2}\int_0^T \frac{|\dot{E} + 2\gamma(x)E/\alpha - \sigma^2/\alpha|^2}{E} dt \tag{93}$$

if $E = \dot{x} + \nu x$ and $I_T(E,x) = \infty$ otherwise. Letting $\alpha \to 0$ and keeping only the leading order term gives

$$I_T^\alpha(E,x) \sim \frac{1}{8\sigma^2\alpha}\int_0^T \frac{|2\gamma(x)(\dot{x}+\nu x) - \sigma^2|^2}{\dot{x}+\nu x} dt \tag{94}$$

which is consistent with (86) since they both imply that the probability weight on paths is roughly

$$\exp\left(-\frac{K}{8\sigma^2\alpha}\int_0^T \frac{|2\gamma(x)(\dot{x}+\nu x) - \sigma^2|^2}{\dot{x}+\nu x} dt\right) \tag{95}$$

when α is small and/or K is large.

C. Two-dimensional generalization

To illustrate the impact that the non-quadratic nature of the Hamiltonian has on the pathway of the transition, let us now consider the following generalization of (73):

$$\begin{cases} \dot{X}_i = -\beta_i X_i + \kappa\sum_{j=1}^2 D_{ij}X_j + Y_i^2, & i = 1,2 \\ dY_i = -\dfrac{1}{\alpha}\gamma(X_i)Y_i dt + \dfrac{1}{\sqrt{\alpha}}\sigma dW_i, & i = 1,2 \end{cases} \tag{96}$$

with

$$D_{11} = D_{22} = -1, \qquad D_{12} = D_{21} = 1 \tag{97}$$

The LLN equations for the system (96) are given by

$$\dot{\bar{X}}_i = \frac{\sigma^2}{2\gamma(\bar{X}_i)} - \beta_i\bar{X}_i + \kappa\sum_{j=1}^2 D_{ij}X_j, \quad i = 1,2. \tag{98}$$

For the specific choice $\gamma(x) = (x-5)^2 + 1$, and $\beta_1 = 0.6$, $\beta_2 = 0.3$ and $\sigma = \sqrt{10}$, the flow field associated with (98) is shown in figure 3: it has the two stable fixed points (shown as red circles in the figure) with one unstable critical point (shown as a red square) in between.

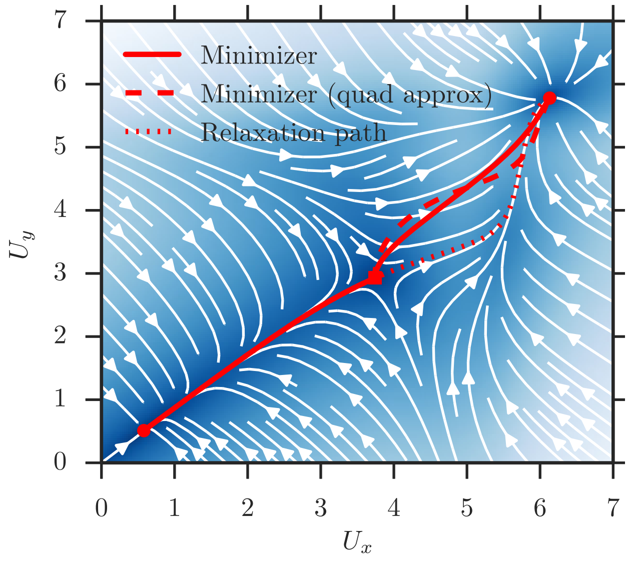

FIG. 3: Dynamics of the coupled slow-fast system ODE model for $\kappa = 0.2$. The arrows denote the direction of the deterministic flow, the color its magnitude. The red solid line depicts the minimizer of the actual action associated with the Hamiltonian in (99), the red dashed line the minimizer of the quadratic approximation of this Hamiltonian, and the red dotted line the relaxation paths from the saddle via the limiting equation (98). Red markers are located at the fixed points (circle: stable; square: saddle).

The Hamiltonian associated with (96) can be written as

$$H(x,\theta) = \sum_{i=1}^{2} h(x_i,\theta_i) - \sum_{i=1}^{2} \beta_i x_i \theta_i + \kappa \sum_{i,j=1}^{2} \theta_j D_{ij} x_j \tag{99}$$

with

$$h(x,\theta) = \tfrac{1}{2}\left(\gamma(x) - \sqrt{\gamma^2(x) - 2\sigma^2\theta}\right) . \tag{100}$$

A numerical computation of the transition trajectories between the two stable fixed points was performed using the geometric minimum action method GMAM [6, 7] (building on the method introduced in [4]) and is shown in figure 3: The red solid line depicts the minimizing trajectory for the full Hamiltonian (99), while the dashed line represents the minimizing trajectory for a quadratic approximation of $H(x,\theta)$, with clear differences between the two. The respective probability of the minimizers can be seen by evaluating the action along the trajectories, as shown in Fig. 4: The action evaluated along its minimizer is lower than evaluated along the minimizer of the quadratic action. Also showed in dashed lines are the relaxation paths from the unstable to the stable equilibrium points from the limiting equation (98): These paths are followed by the mini-

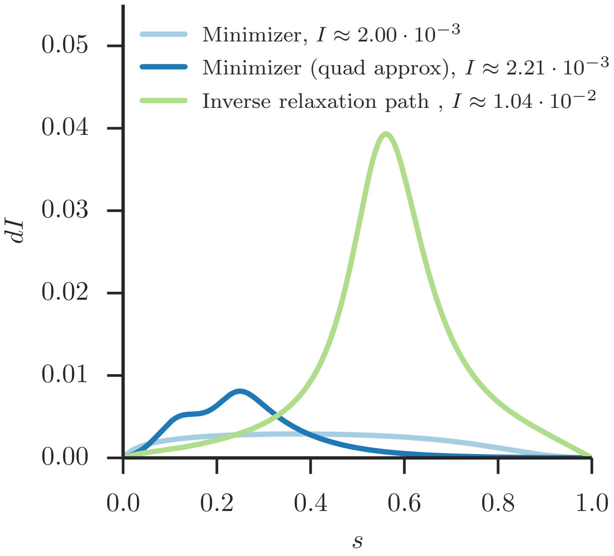

FIG. 4: Action density dI (variation of the action per unit length) versus path length s, for the coupled slow-fast system ODE model in (96) for $\kappa = 0.2$, for paths up to the saddle point. The action density is computed with respect to the full Hamiltonian (99) for the three trajectories depicted in figure 3, i.e. the minimizer of the actual action, the minimizer of the quadratic approximation of this action, and the relaxation pathway from the limiting equation (98).

mizing trajectories on the way down from the unstable critical points (as they should: no noise is necessary for this part of the transition paths) but not on the way uphill.

V. CONCLUDING REMARKS

We have investigated how large deviations affect the slow variables in fast-slow systems via analysis of the large deviation principle (LDP) that characterize their likelihood and pathways. For a specific class of systems, we derived an algebraic equation for the Hamiltonian involved in this LDP, and we discussed several situations in which this equation can be solved explicitly. These results show that the way rare events or infrequent transitions arise in fast-slow systems is intrinsically different from the way they would arise if the dynamics of the slow variables was approximated by an SDE – these difference stem from the fact that the Hamiltonian is non-quadratic in the momenta in general. The examples treated in the present paper were simple enough to allow for analytic treatment. However, we believe that our results will be useful in more complicated situations, in which the algebraic equation for the Hamiltonian will have to be solved numerically.

Acknowledgments

E. V.-E. thank David Kelly for interesting discussions. The research leading to these results has received funding from the European Research Council under the European Union's seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 616811) (F. Bouchet, and T. Tangarife).

Appendix A: Derivation of the limiting equation (5) from the LLN

Here we derive the limiting equation (5) of the LLN by formally taking the limit as $\alpha \to 0$ on the backward Kolmogorov equation (20). To this end expand u as $u = u_0 + \alpha u_1 + O(\alpha^2)$, insert this ansatz in (20), and collect term of increasing power in α. This gives the hierarchy

$$\begin{aligned} &L_1 u_0 = 0 \\ &L_1 u_1 = \partial_t u_0 - L_0 u_0 \\ &\quad\vdots \end{aligned} \tag{A1}$$

The first implies that u_0 is a only a function of x and not of y, or equivalently

$$P u_0 = u_0. \tag{A2}$$

Since L_1 is not invertible ($PL_1 = 0$), the second equation requires a solvability condition, which reads

$$0 = \partial_t P u_0 - P L_0 u_0 = \partial_t u_0 - P L_0 P u_0 \tag{A3}$$

It is easy to see that $PL_0P = F(x) \cdot \partial_x$, i.e. (A3) is the backward Kolmogorov equation of the limiting ODE (5).

Appendix B: Derivation of the CLT equation (10)

To derive the linear SDE (10) of the CLT, notice that, using (5), (1) can be rewritten as

$$\begin{cases} \dot{\tilde{\xi}} = \dfrac{1}{\sqrt{\alpha}} \tilde{f}(\bar{X}, Y) + \partial_x f(\bar{X}, Y)\tilde{\xi} + O(\sqrt{\alpha}) \\ dY = \dfrac{1}{\alpha} b(\bar{X}, Y)dt + \dfrac{1}{\sqrt{\alpha}} \partial_x b(\bar{X}, Y)\tilde{\xi} + \dfrac{1}{\sqrt{\alpha}} \sigma(\bar{X}, Y) dW(t) + O(1) \end{cases} \tag{B1}$$

This means that the joint process $(\bar{X}, \tilde{\xi}, Y)$ is Markov with generator $L' = L_0' + \alpha^{-1/2} L_{\frac{1}{2}} + \alpha^{-1} L_1 + O(\alpha^{3/2})$ where L_1 is defined in (18) and

$$\begin{aligned} L_0' &= F(\bar{x}) \cdot \partial_{\bar{x}} + \partial_x f(\bar{x}, y)\xi \cdot \partial_\xi + \text{operator in } y \\ L_{\frac{1}{2}} &= \tilde{f}(\bar{x}, y) \cdot \partial_\xi + \partial_x b(\bar{x}, y)\xi \cdot \partial_y \end{aligned} \tag{B2}$$

Letting

$$v(t, \bar{x}, \xi, y) = \mathbb{E}^{\bar{x},\xi,y} g(\bar{X}(t), \tilde{\xi}(t)) \tag{B3}$$

this function satisfies the backward Kolmogorov equation

$$\partial_t v = L_0' v + \frac{1}{\sqrt{\alpha}} L_{\frac{1}{2}} v + \frac{1}{\alpha} L_1 v + \text{higher order terms}, \qquad v(0) = g \tag{B4}$$

Formally expand v as $v = v_0 + \sqrt{\alpha} v_{\frac{1}{2}} + \alpha v_1 + O(\alpha^{3/2})$, insert this ansatz in (B4), and collect term of increasing power in α:

$$\begin{aligned} L_1 v_0 &= 0 \\ L_1 v_{\frac{1}{2}} &= -L_{\frac{1}{2}} v_0 \\ L_1 v_1 &= \partial_t v_0 - L_0' v_0 - L_{\frac{1}{2}} v_{\frac{1}{2}} \\ &\dots \end{aligned} \tag{B5}$$

The first equation implies that $v_0 = P v_0$, i.e. v_0 is a function of $\bar{x}$ and ξ only. The solvability condition for the second equation is automatically satisfied since $P\tilde{f} = 0$ implies that $PL_{\frac{1}{2}}P = 0$. Therefore, the solution to this equation is

$$v_{\frac{1}{2}} = -L_1^{-1} L_{\frac{1}{2}} P v_0 \tag{B6}$$

where L_1^{-1} denotes the pseudo-inverse of L_1. Alternatively, this solution can also be expressed as

$$v_{\frac{1}{2}} = \int_0^\infty d\tau \, e^{\tau L_1} L_{\frac{1}{2}} P v_0 \tag{B7}$$

Using this expression in the solvability condition for the third equation in (B5) finally gives the evolution equation for v_0:

$$\partial_t v_0 = P L_0' P v_0 + P L_{\frac{1}{2}} \int_0^\infty d\tau \, e^{\tau L_1} L_{\frac{1}{2}} P v_0 \tag{B8}$$

The first term at the right hand side is explicitly

$$P L_0' P v_0 = \bar{F}(\bar{x}) \cdot \partial_{\bar{x}} v_0 + B_1(\bar{x})\xi \cdot \partial_\xi v_0 \tag{B9}$$

where $B_1(x)$ is

$$B_1(x) = \int_{\mathbb{R}^n} \partial_x f(x,y)\mu_x(dy) \tag{B10}$$

The second term at the right hand side of (B8) is

$$PL_{\frac{1}{2}} \int_0^\infty d\tau\, e^{\tau L_1} L_{\frac{1}{2}} P v_0 = B_2(\bar{x})\xi \cdot \partial_\xi v_0 + A(\bar{x}) : \partial_\xi \partial_\xi v_0 \tag{B11}$$

where $A(x)$ is the matrix defined in (7) and

$$B_2(x) = \int_0^\infty d\tau \int_{\mathbb{R}^n} \left(\partial_y \mathbb{E}^y f(x, \tilde{Y}^x(\tau))\right) \partial_x b(x,y)\mu_x(dy) \tag{B12}$$

Inserting (B9) and (B11) in (B8) shows that this equation is indeed the backward Kolmogorov equation of the joint process governed by (5) and (10).

---

[1] F. Bouchet, C. Nardini, and T. Tangarife. Kinetic theory of jet dynamics in the stochastic barotropic and 2d navier-stokes equations. *J. Stat. Phys.*, 153(4):572–625, 2013.

[2] A. Campa, T. Dauxois, D. Fanelli, and S. Ruffo. *Physics of long-range interacting systems*. Oxford University Press, 2014.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

[4] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Minimum action method for the study of rare events. *Comm. Pure Appl. Math.*, 57(5):637–656, 2004.

[5] Mark Iosifovich Freidlin. The averaging principle and theorems on large deviations. *Russian Mathematical Surveys*, 33(5):117–176, 1978.

[6] Matthias Heymann and Eric Vanden-Eijnden. The geometric minimum action method: A least action principle on the space of curves. *Comm. Pure App. Math.*, 61(8):1052–1117, 2008.

[7] Matthias Heymann and Eric Vanden-Eijnden. Pathways of maximum likelihood for rare events in nonequilibrium systems: Application to nucleation in the presence of shear. *Phys Rev Lett*, 100(14):140601, 2008.

[8] Yuri Kifer. Averaging in dynamical systems and large deviations. *Inventiones Mathematicae*, 110(1):337–370, 1992.

[9] Yuri Kifer. Averaging principle for fully coupled dynamical systems and large deviations. *Ergodic Theory and Dynamical Systems*, 24(03):847–871, 2004.

[10] Robert Liptser. Large deviations for two scaled diffusions. *Probab. Theory Related Fields*, 106(1):71–104, September 1996.

[11] D. R. Nicholson. *Introduction to plasma theory*. Cambridge Univ Press, 1983.

[12] A Yu Veretennikov. On large deviations for sdes with small diffusion and averaging. *Stochastic Processes and their Applications*, 89(1):69–79, 2000.