ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

stochastic processes and their applications

www.elsevier.com/locate/spa

Large deviations for the empirical measure of a diffusion via weak convergence methods

Paul Dupuis, David Lipshutz*

Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

Received 14 May 2017; received in revised form 31 August 2017; accepted 25 September 2017
Available online xxxx

Abstract

We consider the large deviation principle for the empirical measure of a diffusion in Euclidean space, which was first established by Donsker and Varadhan. We employ a weak convergence approach and obtain a characterization for the rate function that is dual to the one obtained by Donsker and Varadhan, and which allows us to evaluate the variational form of the rate function for both reversible and nonreversible diffusions.

MSC: 60F10; 60H10; 60J60

Keywords: Diffusion; Empirical measure; Large deviations; Weak convergence method

1. Introduction

In this work we employ a weak convergence approach to prove the large deviation principle (LDP) for the empirical measure of a diffusion in Euclidean space. Given a positive integer d, let $\mathbb{R}^d$ denote d-dimensional Euclidean space, $\mathcal{B}(\mathbb{R}^d)$ be the σ-algebra of Borel subsets of $\mathbb{R}^d$ and $\mathcal{P}(\mathbb{R}^d)$ be the space of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ equipped with the topology of weak convergence. Let $\mathbb{M}^{d\times d}$ denote the set of real-valued $d \times d$ nonnegative definite symmetric matrices. Suppose a and b are continuous functions on $\mathbb{R}^d$ taking values in $\mathbb{M}^{d\times d}$ and $\mathbb{R}^d$, respectively, and for each x in $\mathbb{R}^d$ let $\sigma(x) \in \mathbb{M}^{d\times d}$ be the unique nonnegative definite square

* Corresponding author.
E-mail address: david_lipshutz@brown.edu (D. Lipshutz).

https://doi.org/10.1016/j.spa.2017.09.020

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

root of $a(x)$. Consider the stochastic differential equation (SDE)

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t), \tag{1.1}$$

where $X = \{X(t), t \geq 0\}$ is a d-dimensional continuous process, $W = \{W(t), t \geq 0\}$ is a d-dimensional Brownian motion and the integral with respect to dW is the Itô integral. Given a solution X of the SDE (1.1), define the associated empirical measure process $L = \{L_t, t > 0\}$ taking values in $\mathcal{P}(\mathbb{R}^d)$, for $t > 0$, by

$$L_t(A) \doteq \frac{1}{t}\int_0^t \delta_{X(s)}(A)ds, \qquad A \in \mathcal{B}(\mathbb{R}^d). \tag{1.2}$$

Here $\delta_x \in \mathcal{P}(\mathbb{R}^d)$ is the Dirac delta measure at $x \in \mathbb{R}^d$. In this work we establish the LDP for the family $\{L_t, t > 0\}$ on $\mathcal{P}(\mathbb{R}^d)$ as $t \to \infty$.

The LDP for the empirical measure process $\{L_t, t > 0\}$ is a classical result that follows from the works of Donsker and Varadhan [5–7], which established the LDP for the empirical measure process associated with a large class of discrete and continuous time Markov processes. Let $(\mathcal{L}, \mathcal{D})$ denote the infinitesimal generator associated with a diffusion X satisfying the SDE (1.1). Then $C_b^2(\mathbb{R}^d)$, the space of twice continuously differentiable real-valued functions on $\mathbb{R}^d$ whose first and second partial derivatives are bounded, lies in $\mathcal{D}$ and for $\phi \in C_b^2(\mathbb{R}^d)$,

$$\mathcal{L}\phi \doteq \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d a_{ij}\frac{\partial^2\phi}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial \phi}{\partial x_i}. \tag{1.3}$$

Let $\mathcal{D}^+$ denote the subset of functions $\phi \in \mathcal{D}$ that are uniformly bounded below by a positive constant on $\mathbb{R}^d$. Donsker and Varadhan [7] showed that under certain stability and regularity conditions, $\{L_t, t > 0\}$ satisfies the LDP on $\mathcal{P}(\mathbb{R}^d)$ with rate function $J : \mathcal{P}(\mathbb{R}^d) \to [0, \infty]$ defined, for $\mu \in \mathcal{P}(\mathbb{R}^d)$, by

$$J(\mu) \doteq \sup_{\phi \in \mathcal{D}^+} -\int_{\mathbb{R}^d} \frac{(\mathcal{L}\phi)(x)}{\phi(x)}\mu(dx). \tag{1.4}$$

Other approaches have been developed by Gärtner [12] and Fleming, Sheu and Soner [11] for the empirical measure of Markov processes taking values in compact manifolds; and by Feng and Kurtz [10, Chapter 12] for the empirical measure of Markov processes taking values in more general metric spaces.

We use weak convergence methods to prove the LDP for the empirical measure of a diffusion in Euclidean space. These techniques have been developed more generally for large deviation problems in the book by Dupuis and Ellis [8], and for empirical measures associated with specific classes of continuous time Markov processes in [3,4,9]. There are a couple advantages of our approach that we highlight here. First, we use standard techniques in the theory of weak convergence and SDEs, which are applied directly to the continuous time diffusion process. We find this probabilistic approach to the classical problem appealing. Second, we obtain a variational formulation of the rate function (see (3.3)) that is dual to the one defined in (1.4), and we evaluate the variational form of the rate function for the large class of both reversible and nonreversible diffusions in Euclidean space that are strong solutions of the SDE (1.1) (see Proposition 6.5). In particular, the rate function evaluated at a probability measure μ is the μ-weighted L^2-cost of a feedback control u that is reversible with respect to μ and such that μ is formally invariant under the infinitesimal generator associated with the u-controlled diffusion.

In summary, the main contributions of this work are as follows:

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

- A weak convergence approach to establish the LDP for the empirical measure of a diffusion.
- A variational form of the rate function that is the dual of the Donsker-Varadhan variational form of the rate function.
- Evaluation of the variational form of the rate function for reversible and nonreversible diffusions in Euclidean space.

To our knowledge the variational form of the rate function associated with a general diffusion in Euclidean space has not been evaluated on its domain of finiteness. Some prior results along these lines include [5, Theorem 5] which evaluates the variational form of rate function defined in (1.4) when the diffusion is reversible and takes values in a compact manifold; [16, Theorem 1.4] which provides a more tractable form of the rate function defined in (1.4) and [17, Theorem 1.4] which uses the results in [16] to evaluate the rate function defined in (1.4) for diffusions in compact regions in $\mathbb{R}^d$ with reflection along the boundary of the region; and [12, Theorem 3.2] which expresses the rate function associated with diffusions on compact manifolds evaluated at measures with sufficiently smooth densities in terms of the solution to a certain partial differential equation (PDE), and the solution can be explicitly identified when the diffusion is reversible. More recently, [18] considers a diffusion in a compact manifold whose drift can a priori be decomposed into sufficiently smooth reversible and nonreversible (with respect to the invariant measure of the diffusion) components. The authors use the results in [12] to explicitly express the rate function in terms of the rate function associated with a related reversible diffusion and the solution to a certain elliptic PDE (see [18, Theorem 2.2]).

The remainder of the paper is organized as follows. In Section 2 we present the assumptions for our main result, which is stated in Section 3. We prove the empirical measure process satisfies the Laplace principle, which is equivalent to the LDP (see the remark following Theorem 3.3). In Section 4 we state a variational representation for exponential functionals of the empirical measure that is used in Section 5 to prove the Laplace principle upper bound and in Section 7 to prove the Laplace principle lower bound. In Section 6 we state and prove properties of the rate function.

2. Preliminaries

In this section we provide conditions on the drift and diffusion coefficients that ensure existence, uniqueness and stability of solutions to the SDE (1.1). We first introduce some commonly used notation. Let $C(\mathbb{R}^d)$ denote the set of continuous functions on $\mathbb{R}^d$. For $k = 1, 2$, let $C^k(\mathbb{R}^d)$ denote the subset of functions in $C(\mathbb{R}^d)$ whose partial derivatives up to order k exist and are continuous, let $C_b^k(\mathbb{R}^d)$ denote the subset of functions in $C^k(\mathbb{R}^d)$ that are bounded and whose partial derivatives up to order k are bounded, and given $\alpha \in (0, 1)$, let $C^{k,\alpha}(\mathbb{R}^d)$ denote the subset of functions in $C^k(\mathbb{R}^d)$ whose kth partial derivatives are Hölder continuous with exponent α. Given $f \in C^1(\mathbb{R}^d)$ we let ∇f denote the gradient of f. Let $C_b^\infty(\mathbb{R}^d)$ denote the subset of functions in $C(\mathbb{R}^d)$ that are bounded and have bounded continuous derivatives of all order and let $C_c^\infty(\mathbb{R}^d)$ denote the subset of function in $C_b^\infty(\mathbb{R}^d)$ that are compactly supported.

The following local Lipschitz continuity condition ensures pathwise uniqueness of solutions. Let $\|\cdot\|$ denote a fixed norm on the set of $d \times d$ matrices.

Condition 2.1. *For each $R < \infty$, there exists $K_R < \infty$ such that for all $x, y \in \mathbb{R}^d$ satisfying $|x|, |y| \le R$,*

$$|b(x) - b(y)| + \|\sigma(x) - \sigma(y)\| \le K_R|x - y|.$$

The next condition imposes a strong stability condition on the diffusion that will be used to prove the Laplace principle upper bound. As we explain below, the first part of the condition is a version of the stability condition imposed by Donsker and Varadhan [7]. The second part of the condition is a mild technical on the growth rate of the drift $b(\cdot)$ that is imposed to ensure that the rate function (defined in (3.3)) can be explicitly evaluated.

Condition 2.2. *There is a nonnegative function $V \in C^2(\mathbb{R}^d)$ and constants $c_1, c_2 < \infty$ such that*

$$\lim_{R\to\infty} \inf\left\{g(x) : x \in \mathbb{R}^d, |x| > R\right\} = \infty, \tag{2.1}$$

and

$$|b(x)|^2 \le c_1 + c_2 g(x), \qquad x \in \mathbb{R}^d, \tag{2.2}$$

where $g : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$g(x) \doteq -(\mathcal{L}V)(x) - \frac{1}{2}|\sigma(x)\nabla V(x)|^2, \qquad x \in \mathbb{R}^d. \tag{2.3}$$

Remark 2.3. Under Condition 2.1, g is a continuous function and therefore, by (2.1), has compact level sets.

Remark 2.4. Condition 2.2 is a stronger stability condition than the one needed to ensure positive recurrence of X. For instance, a sufficient condition for X to be positive recurrent is the existence of $V \in C^2(\mathbb{R}^d)$ and a constant $k > 0$ such that $\mathcal{L}V(x) \le -k$ for all $|x|$ sufficiently large (see [14, Theorem 3.9]). In [3] the authors demonstrated that without a stronger stability condition, the empirical measure process may charge points at infinity (in the large deviations limit).

Remark 2.5. The first part of Condition 2.2 (i.e., (2.1)) is a slightly stronger stability condition than the one imposed by Donsker and Varadhan [7]. The condition in [7] requires the existence of a real-valued function g on $\mathbb{R}^d$ and a sequence $\{u_n, n \in \mathbb{N}\}$ in the domain of the generator $\mathcal{L}$ such that g has compact level sets, the sequence $\{u_n, n \in \mathbb{N}\}$ satisfies certain boundedness conditions and $g = -\lim_{n\to\infty}(\mathcal{L}u_n)/u_n$ holds pointwise. Under Condition 2.2, setting $u_n \doteq e^V$ for each $n \in \mathbb{N}$, and g as in (2.3), we see from the definition of $\mathcal{L}$ given in (1.3) and the stability condition (2.1) that the condition in [7] holds.

The third and final condition imposes regularity and uniform ellipticity conditions which are used in the proof of the Laplace principle lower bound.

Condition 2.6. *The diffusion coefficient a is bounded, uniformly Lipschitz continuous and uniformly elliptic. In addition, there exists $0 < \alpha < 1$ such that b_i belongs to $C^{1,\alpha}(\mathbb{R}^d)$ for $1 \le i \le d$ and a_{ij} belongs to $C^{2,\alpha}(\mathbb{R}^d)$ for $1 \le i, j \le d$.*

Throughout this work we let $(\Omega, \mathcal{F}, P)$ denote a complete probability space, X_0 be a d-dimensional random vector on $(\Omega, \mathcal{F}, P)$ and $W = \{W(t), t \ge 0\}$ be a d-dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$ that is independent of X_0. We define the right-continuous filtration $\{\mathcal{F}_t\}$ by

$$\mathcal{F}_t \doteq \sigma(\{W(s), 0 \le s \le t\}, X_0, N), \qquad t \ge 0, \tag{2.4}$$

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

where N is the set of all P-null sets in Ω. Under Conditions 2.1 and 2.2, there exists a pathwise unique strong solution X of the SDE (1.1) with initial condition $X(0) = X_0$, and X is a time-homogeneous Feller Markov process (see, e.g., [14, Theorem 3.5]). We let P_x denote the probability measure P on $(\Omega, \mathcal{F})$ conditioned on $X_0 = x$ and E_x denote expectation under P_x. Furthermore, if Condition 2.6 also holds, X is positive recurrent, has a unique stationary distribution $\pi \in \mathcal{P}(\mathbb{R}^d)$, and for all $x \in \mathbb{R}^d$, P_x a.s. $L_t \Rightarrow \pi$ as $t \to \infty$, where we use the double arrow "$\Rightarrow$" to denote weak convergence (see, e.g., [14, Chapter 4]).

3. A large deviation principle

In order to define the rate function, we define, for $y \in \mathbb{R}^d$, the operator $\mathcal{L}^y$ on $C_b^2(\mathbb{R}^d)$ by

$$\mathcal{L}^y \phi \doteq \mathcal{L}\phi + \langle \sigma y, \nabla \phi \rangle, \qquad \phi \in C_b^2(\mathbb{R}^d), \tag{3.1}$$

where $\mathcal{L}$ is defined as in (1.3) and $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^d$. Let $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ equipped with the product topology.

Remark 3.1. Throughout this work, given $\theta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, let $[\theta]_i \in \mathcal{P}(\mathbb{R}^d)$ denote the ith marginal of θ, for $i = 1, 2$.

Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, let

$$\mathcal{S}_\mu \doteq \left\{ \nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \begin{array}{c} [\nu]_1 = \mu \text{ and } \forall\, \phi \in C_b^2(\mathbb{R}^d) \\ \displaystyle\int_{\mathbb{R}^d \times \mathbb{R}^d} (\mathcal{L}^y \phi)(x) \nu(dx\, dy) = 0 \end{array} \right\}. \tag{3.2}$$

The rate function $I : \mathcal{P}(\mathbb{R}^d) \to [0, \infty]$ evaluated at $\mu \in \mathcal{P}(\mathbb{R}^d)$ is given by

$$I(\mu) \doteq \begin{cases} \displaystyle\inf_{\nu \in \mathcal{S}_\mu} \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 \nu(dx\, dy) & \text{if } b \in L^2_\mu \text{ and } \mathcal{S}_\mu \neq \emptyset, \\ \infty & \text{otherwise.} \end{cases} \tag{3.3}$$

Remark 3.2. The condition $b \in L^2_\mu$ in (3.3) is a consequence of the mild growth condition on the drift (2.2) that we impose in Condition 2.2. The condition ensures regularity properties of measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfying $I(\mu) < \infty$ (see Lemma 6.4), which are used to obtain an explicit form for the variational form of the rate function in Proposition 6.5.

The rate function $I(\mu)$ evaluated at $\mu \in \mathcal{P}(\mathbb{R}^d)$ can equivalently be defined as the infimum over the μ-weighted L^2-cost over all feedback controls u under which μ is formally invariant under the infinitesimal generator of the associated controlled diffusion (see Lemma 6.3). The minimizing control u is identified in Proposition 6.5.

We now state our main result.

Theorem 3.3. *Suppose Conditions 2.1, 2.2 and 2.6 hold. Let I be defined in (3.3) and the empirical measure process $\{L_t, t > 0\}$ be defined as in (1.2). Then I has compact level sets and the following uniform Laplace principle holds: for any compact set K in $\mathbb{R}^d$ and bounded continuous function h mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}$, we have*

$$\lim_{t \to \infty} \sup_{x \in K} \left| \frac{1}{t} \log E_x \left[\exp\{-t h(L_t)\} \right] + \inf_{\gamma \in \mathcal{P}(\mathbb{R}^d)} \{ I(\gamma) + h(\gamma) \} \right| = 0. \tag{3.4}$$

Proof. The fact that I has compact level sets follows from Lemma 6.1. The uniform Laplace principle (3.4) follows from the uniform upper and lower bounds shown in Proposition 5.1 and Proposition 7.1, respectively. ■

Due to the equivalence between the Laplace principle and the LDP (see, e.g., [8, Theorem 1.2.3]), it follows that $\{L_t, t > 0\}$ satisfies the LDP on $\mathcal{P}(\mathbb{R}^d)$ with rate I, and the LDP is uniform for initial conditions in compact sets.

Remark 3.4. Since the rate function is unique (see, e.g., [8, Theorem 1.3.1]), it follows that our rate function I defined in (3.3) coincides with the Donsker-Varadhan rate function J defined in (1.4).

4. Variational representation

In this section we establish a variational representation for exponential functions of L_t, which will be used to prove both the Laplace principle upper and lower bounds. For $t < \infty$ let $\mathcal{V}_t$ denote the set of processes $v = \{v(s), 0 \le s \le t\}$ that are progressively measurable with respect to the filtration $\{\mathcal{F}_t\}$ defined in (2.4) and satisfy

$$E\left[\int_0^t |v(s)|^2 ds\right] < \infty.$$

The following representation was shown in [2, Theorem 4.1].

Proposition 4.1. *For $t < \infty$ let $X = \{X(s), 0 \le s \le t\}$ be the pathwise unique solution to the SDE (1.1) on $[0, t]$. Then for any bounded Borel measurable function G mapping $C([0, t], \mathbb{R}^d)$ into $\mathbb{R}$ the following representation holds:*

$$\log E\left[\exp\left(-G(X)\right)\right] = -\inf_{v\in\mathcal{V}_t} E\left[\frac{1}{2}\int_0^t |v(s)|^2 ds + G\left(X^v\right)\right],$$

where $X^v = \{X^v(s), 0 \le s \le t\}$ denotes the pathwise unique solution to the SDE

$$dX^v(s) = b(X^v(s))ds + \sigma(X^v(s))dW(s) + \sigma(X^v(s))v(s)ds, \tag{4.1}$$

for $0 \le s \le t$, with initial condition $X^v(0) = X_0$.

Using Proposition 4.1, we obtain a variational representation for functionals of the empirical measure process.

Proposition 4.2. *Let h be a bounded Borel measurable function mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}$. Then for all $t < \infty$,*

$$\frac{1}{t}\log E\left[\exp\left\{-th(L_t)\right\}\right] = -\inf_{v\in\mathcal{V}_t} E\left[\frac{1}{2t}\int_0^t |v(s)|^2 ds + h(L_t^v)\right], \tag{4.2}$$

where L_t^v denotes the empirical measure of X^v on the interval $[0, t]$ defined by

$$L_t^v(A) \doteq \frac{1}{t}\int_0^t \delta_{X^v(s)}(A)ds, \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

Proof. Let $t < \infty$. Let g be the continuous function mapping $C([0,t],\mathbb{R}^d)$ into $\mathcal{P}(\mathbb{R}^d)$ defined by

$$G(w)(\cdot) \doteq \frac{1}{t}\int_0^t \delta_{w(s)}(\cdot)ds, \qquad w \in C([0,t],\mathbb{R}^d).$$

Then $L_t = G(X)$ and the composite function $h \circ G$ mapping $C([0,t],\mathbb{R}^d)$ into $\mathbb{R}$ is bounded and Borel measurable. Thus, by Proposition 4.1 and the fact that $L_t^v = G(X^v)$, (4.2) holds. ■

5. Laplace principle upper bound

In this section we prove the uniform Laplace principle upper bound, which is stated in the following proposition.

Proposition 5.1. *Suppose b and σ satisfy Conditions* 2.1 and 2.2. *Define I as in* (3.3). *Then for each compact subset K of $\mathbb{R}^d$ and any bounded continuous function h mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}$, we have*

$$\limsup_{t\to\infty}\sup_{x\in K}\frac{1}{t}\log E_x\left[\exp\{-th(L_t)\}\right] \le -\inf_{\gamma\in\mathcal{P}(\mathbb{R}^d)}\{I(\gamma)+h(\gamma)\}. \tag{5.1}$$

The remainder of this section is devoted to the proof of Proposition 5.1. Throughout this section we assume b and σ satisfy Conditions 2.1 and 2.2. We fix a compact subset K of $\mathbb{R}^d$ and a bounded continuous function h mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}$. Let $\{t_n, n \in \mathbb{N}\}$ be an increasing sequence in $(0,\infty)$ and $\{x_n, n\in\mathbb{N}\}$ be a sequence in K such that $t_n \to \infty$ as $n \to \infty$ and

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{t_n}\log E_{x_n}\left[\exp\left\{-t_n h(L_{t_n})\right\}\right] \\ &\quad = \limsup_{t\to\infty}\sup_{x\in K}\frac{1}{t}\log E_x\left[\exp\{-th(L_t)\}\right].\end{aligned} \tag{5.2}$$

Due to the representation stated in Proposition 4.2, for each $n \in \mathbb{N}$, there is a progressively measurable process v^n in $\mathcal{V}_{t_n}$ such that

$$\begin{aligned}&\frac{1}{t_n}\log E_{x_n}\left[\exp\left\{-t_n h(L_{t_n})\right\}\right] \\ &\quad \le -E_{x_n}\left[\frac{1}{2t_n}\int_0^{t_n}|v^n(s)|^2ds + h(L_{t_n}^{v^n})\right] + \frac{1}{n}.\end{aligned} \tag{5.3}$$

Let $\|h\|_\infty \doteq \sup\{|h(\gamma)| : \gamma\in\mathcal{P}(\mathbb{R}^d)\} < \infty$. Rearranging the last display, we see that for each $n \in \mathbb{N}$,

$$E_{x_n}\left[\frac{1}{2t_n}\int_0^{t_n}|v^n(s)|^2ds\right] \le 2\|h\|_\infty + 1. \tag{5.4}$$

For each $n\in\mathbb{N}$, define $\nu_n \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ by

$$\nu_n(A\times B) \doteq \frac{1}{t_n}\int_0^{t_n}\delta_{(X^{v^n}(s),v^n(s))}(A\times B)ds \tag{5.5}$$

for Borel subsets A and B of $\mathbb{R}^d$.

Since X^{v_n} and v^n are $\{\mathcal{F}_t\}$-adapted and the mappings $(X^{v^n}, v^n) \mapsto \nu_n$ and $X^{v^n} \mapsto [\nu_n]_1$ are continuous, it follows that ν_n and $[\nu_n]_1$ are $\mathcal{F}_{t_n}$-measurable random variables taking values in $\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{R}^d)$, respectively. Our next step is to establish tightness of the family $\{\nu_n, n\in\mathbb{N}\}$. Before doing so, we recall a useful result (see, e.g., [8, Theorem A.3.12]).

Lemma 5.2. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty]$ be bounded below and lower semicontinuous. Suppose $\{\theta_n, n \in \mathbb{N}\}$ is a sequence in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ converging weakly to θ in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Then*

$$\liminf_{n\to\infty} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x, y)\theta_n(dx\, dy) \ge \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x, y)\theta(dx\, dy).$$

Given $V \in C^2(\mathbb{R}^d)$ as in Condition 2.2, let $g \in C(\mathbb{R}^d)$ be defined as in (2.3) so that g has compact level sets and by a completion of squares argument, for each $x \in \mathbb{R}^d$,

$$g(x) = \inf\left\{-(\mathcal{L}^y V)(x) + \frac{1}{2}|y|^2 : y \in \mathbb{R}^d\right\}. \tag{5.6}$$

Lemma 5.3. *The sequence $\{\nu_n, n \in \mathbb{N}\}$ of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$-valued random elements is tight.*

Proof. Let $\varepsilon > 0$. We claim there exist compact sets Γ_1 and Γ_2 in $\mathcal{P}(\mathbb{R}^d)$ such that for each $n \in \mathbb{N}$,

$$P_{x_n}([\nu_n]_1 \in \Gamma_1) \ge 1 - \varepsilon \tag{5.7}$$

$$P_{x_n}([\nu_n]_2 \in \Gamma_2) \ge 1 - \varepsilon. \tag{5.8}$$

Then $\Gamma \doteq \{\nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : [\nu]_i \in \Gamma_i, i = 1, 2\}$ is a compact subset of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and $P_{x_n}(\nu_n \in \Gamma) \ge 1 - 2\varepsilon$ for each $n \in \mathbb{N}$, which will complete the proof.

We first define the compact set $\Gamma_1 \subset \mathcal{P}(\mathbb{R}^d)$ satisfying (5.7). Let

$$M_1 \doteq \frac{\frac{1}{t_1}\sup_{x\in K} V(x) + 2\|h\|_\infty + 1}{\varepsilon} < \infty \tag{5.9}$$

and set

$$\Gamma_1 \doteq \left\{\gamma \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} g(x)\gamma(dx) \le M_1\right\}. \tag{5.10}$$

An application of Chebyshev's inequality and the fact that g is continuous shows that Γ_1 is tight, and hence compact, in $\mathcal{P}(\mathbb{R}^d)$. Let $n \in \mathbb{N}$. By (5.6), (4.1), Itô's formula, the martingale property of the stochastic integral and the nonnegativity of V,

$$\begin{aligned} V(x_n) - E_{x_n}\left[\int_0^{t_n} g(X^{v^n}(s))ds + \frac{1}{2}\int_0^{t_n} |v^n(s)|^2 ds\right] &\ge E_{x_n}\left[V(X^{v^n}(t_n))\right] \\ &\ge 0. \end{aligned}$$

Then, by the definition of $[\nu_n]_1$, the last display and (5.4), we have

$$E_{x_n}\left[\int_{\mathbb{R}^d} g(x)[\nu_n]_1(dx)\right] \le \frac{1}{t_n}V(x_n) + 2\|h\|_\infty + 1. \tag{5.11}$$

Thus, due to (5.10), Chebyshev's inequality, (5.11) and the definition of M_1 in (5.9), we have $P_{x_n}([\nu_n]_1 \notin \Gamma_1) < \varepsilon$. This proves (5.7).

Next we construct the compact set $\Gamma_2 \subset \mathcal{P}(\mathbb{R}^d)$ satisfying (5.8). Let

$$M_2 > \frac{4\|h\|_\infty + 2}{\varepsilon} \tag{5.12}$$

and set

$$\Gamma_2 \doteq \left\{\gamma \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |y|^2\gamma(dy) \le M_2\right\}. \tag{5.13}$$

It follows from Lemma 5.2 that the function mapping γ to $\int_{\mathbb{R}^d} |y|^2\gamma(dy)$ has closed level sets and so Γ_2 is a closed set. An application of Chebyshev's inequality shows that the family Γ_2 is tight, and hence compact, in $\mathcal{P}(\mathbb{R}^d)$. Let $n \in \mathbb{N}$. By (5.13), the definition (5.5) of ν_n, Chebyshev's inequality, (5.4) and (5.12), $P_{x_n}([\nu_n]_2 \notin \Gamma_2) < \varepsilon$. This proves (5.8). ■

In light of Lemma 5.3, we can take a weakly convergent subsequence of $\{\nu_n, n \in \mathbb{N}\}$, which we also denote by $\{\nu_n, n \in \mathbb{N}\}$, and let ν denote its limit point. For the remainder of this section we fix such a convergent subsequence with limit point ν in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. According to the Skorokhod representation theorem, there is a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ and random elements $\{\bar{\nu}_n, n \in \mathbb{N}\}$ and $\bar{\nu}$ on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ taking values in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\bar{\nu}_n$ is equal in distribution to ν_n for each $n \in \mathbb{N}$, $\bar{\nu}$ is equal in distribution to ν and $\bar{P}$ a.s.

$$\bar{\nu}_n \Rightarrow \bar{\nu} \qquad \text{as } n \to \infty. \tag{5.14}$$

Let $\bar{E}$ denote expectation under $\bar{P}$. The following decomposition of $\bar{\nu}$ was shown to hold in [8, Theorem A.5.6].

Lemma 5.4. *For each $A \in \mathcal{B}(\mathbb{R}^d)$, the mapping from $(\bar{\Omega}, \bar{\mathcal{F}})$ to $([0,1], \mathcal{B}([0,1]))$ given by $\bar{\omega} \mapsto [\bar{\nu}(\bar{\omega})]_1(A)$ is measurable. Furthermore, there exists a family of probability measures $[\nu(\bar{\omega})]_{2|1}(dy|x)$ on $\mathbb{R}^d$ parameterized by $(\bar{\omega}, x) \in \bar{\Omega} \times \mathbb{R}^d$ such that for each $A \in \mathcal{B}(\mathbb{R}^d)$, the mapping from $(\bar{\Omega} \times \mathbb{R}^d, \bar{\mathcal{F}} \otimes \mathcal{B}(\mathbb{R}^d))$ to $([0,1], \mathcal{B}([0,1]))$ given by $(\bar{\omega}, x) \mapsto [\bar{\nu}(\bar{\omega})]_{2|1}(A|x)$ is measurable, and $\bar{P}$-a.s.*

$$\bar{\nu}(A \times B) = \int_A [\bar{\nu}]_{2|1}(B|x)[\bar{\nu}]_1(dx), \qquad A, B \in \mathcal{B}(\mathbb{R}^d). \tag{5.15}$$

In preparation for proving Proposition 5.1, we have the following lemma.

Lemma 5.5. *The following limit holds:*

$$\liminf_{n\to\infty} \bar{E}\left[\frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 \bar{\nu}_n(dx\,dy)\right] \geq \bar{E}\left[I([\bar{\nu}]_1)\right],$$

where I is defined in (3.3).

Proof. We have the following inequalities, which are explained below:

$$\begin{aligned}\bar{E}\left[\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\bar{\nu}(dx\,dy)\right] &\leq \liminf_{n\to\infty} \bar{E}\left[\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\bar{\nu}_n(dx\,dy)\right] \\ &\leq 2\|h\|_\infty + 1.\end{aligned} \tag{5.16}$$

The first inequality is due to (5.14), Lemma 5.2 and Fatou's lemma. The second inequality is a consequence of the Skorokhod representation theorem, the definition of ν_n in (5.5) and the bound (5.4). It follows that for $\bar{P}$-a.e. $\bar{\omega} \in \bar{\Omega}$,

$$\liminf_{n\to\infty} \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\bar{\nu}_n(\bar{\omega}; dx\,dy) < \infty.$$

By taking a subsequence $\{n_k(\bar{\omega}), k \in \mathbb{N}\}$, we can assume that

$$\sup_k \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\bar{\nu}_{n_k(\bar{\omega})}(\bar{\omega}; dx\,dy) < \infty.$$

Along this subsequence we have

$$\begin{aligned}\lim_{C\to\infty}\sup_k \int_{\mathbb{R}^d\times\mathbb{R}^d} |y| 1_{\{|y|>C\}} \bar{\nu}_{n_k(\bar{\omega})}(\bar{\omega}; dx\, dy) &\le \lim_{C\to\infty} \frac{1}{C}\sup_k \int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2 \bar{\nu}_{n_k(\bar{\omega})}(\bar{\omega}; dx\, dy)\\ &= 0.\end{aligned}$$

By the uniform integrability shown in the last display,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} (\mathcal{L}^y\phi)(x)\bar{\nu}(\bar{\omega}; dx\, dy) = \lim_{k\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} (\mathcal{L}^y\phi)(x)\bar{\nu}_{n_k(\bar{\omega})}(\bar{\omega}; dx\, dy).$$

Since ν_n and $\bar{\nu}_n$ are equal in distribution for each $n \in \mathbb{N}$, in order to show that the left hand side of the last display is equal to zero for $\bar{P}$-a.e. $\bar{\omega} \in \bar{\Omega}$, it suffices to show that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} (\mathcal{L}^y\phi)(x)\nu_n(dx\, dy) \tag{5.17}$$

converges to zero in probability as $n \to \infty$. By (5.5), (4.1) and Itô's formula, the last display is equal to

$$\frac{1}{t_n}\phi(X^{\upsilon^n}(t_n)) - \frac{1}{t_n}\phi(x) - \frac{1}{t_n}\int_0^{t_n} \langle \nabla\phi(X^{\upsilon^n}(s)), \sigma(X^{\upsilon^n}(s))dW(s)\rangle.$$

Since ϕ is bounded, the first two terms converge to zero a.s. as $n \to \infty$. For the third term, Chebyshev's inequality and the Itô isometry imply that for all $1 \le i, j \le d$ and $\varepsilon > 0$,

$$\begin{aligned}P_{x_n}\left(\left|\frac{1}{t_n}\int_0^{t_{n_k}} \sigma_{ij}(X^{\upsilon^n}(s))\frac{\partial\phi}{\partial x_i}(X^{\upsilon^n}(s))dW_j(s)\right| \ge \varepsilon\right)\\ \le \frac{1}{t_n^2\varepsilon^2} E_x\left[\int_0^{t_n}\left|\sigma_{ij}(X^{\upsilon^n}(s))\frac{\partial\phi}{\partial x_i}(X^{\upsilon^n}(s))\right|^2 ds\right].\end{aligned}$$

Since $\frac{\partial\phi}{\partial x_i}$ and σ_{ij} are bounded for $1 \le i, j \le d$, the right hand side converges to zero as $n \to \infty$. Thus, $\bar{P}$ a.s. $\bar{\nu}$ lies in $\mathcal{S}_{[\bar{\nu}]_1}$.

According to the definition of I in (3.3), we are left to show that $\bar{P}$ a.s. b lies in $L^2_{[\bar{\nu}]_1}$. Let $V \in C^2(\mathbb{R}^d)$ be as in Condition 2.2 and define $g \in C(\mathbb{R}^d)$ as in (2.3). By Condition 2.2,

$$\int_{\mathbb{R}^d} |b(x)|^2[\bar{\nu}]_1(dx) \le c_1 + c_2\int_{\mathbb{R}^d} g(x)[\bar{\nu}]_1(dx).$$

Using (5.14), Lemma 5.2, the Skorokhod representation theorem and following the exact argument carried out in the proof of Lemma 5.3 to obtain (5.11), we see that

$$\begin{aligned}\bar{E}\left[\int_{\mathbb{R}^d} g(x)[\bar{\nu}]_1(dx)\right] &\le \liminf_{n\to\infty} E_{x_n}\left[\int_{\mathbb{R}^d} g(x)[\nu_n]_1(dx)\right]\\ &\le \liminf_{n\to\infty}\frac{1}{t_n}V(x_n) + 2\|h\|_\infty + 1.\end{aligned}$$

Since $\{x_n, n \in \mathbb{N}\}$ take values in the compact set K and V is continuous, the last two displays together prove that $\bar{P}$ a.s. b lies in $L^2_{[\bar{\nu}]_1}$. ■

Proof of Proposition 5.1. Fix a compact set K in $\mathbb{R}^d$ and a bounded continuous function h mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}$. Let $\{x_n, n \in \mathbb{N}\}$ be a sequence in K and $\{t_n, n \in \mathbb{N}\}$ be an increasing

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

sequence in $(0,\infty)$ such that $t_n \to \infty$ as $n \to \infty$ and (5.2) holds. By (5.3), (5.5) and the Skorokhod representation theorem, for each $n \in \mathbb{N}$,

$$\begin{aligned}\frac{1}{t_n}\log E_{x_n}\left[\exp\left(-t_n h(L_{t_n})\right)\right] &\le -E_{x_n}\left[\frac{1}{2t_n}\int_0^{t_n}|v^n(s)|^2 ds + h(L_{t_n}^{v^n})\right]+\frac{1}{n}\\ &= -\bar{E}\left[\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\bar{\nu}_n(dx\,dy)+h([\bar{\nu}_n]_1)\right]+\frac{1}{n}.\end{aligned}$$

Then by the last display, Lemma 5.5 and (5.14), we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{t_n}\log E_{x_n}\left[\exp\left(-t_n h(L_{t_n})\right)\right] &\le -\bar{E}\left[I([\bar{\nu}]_1)+h([\bar{\nu}]_1)\right]\\ &\le -\inf_{\gamma\in\mathcal{P}(\mathbb{R}^d)}\{I(\gamma)+h(\gamma)\}.\end{aligned}$$

Along with (5.2), this proves (5.1). ■

6. Properties of the rate function

In this section we study properties of the function I defined in (3.3). In the following two lemmas we show that I has compact level sets and is convex. Let $\mathbb{R}_+ \doteq [0,\infty)$. For each $n \in \mathbb{N}$, let $\alpha_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a smooth (i.e., α_n is continuous and its restriction to $(0,\infty)$ has continuous derivatives of all orders) nondecreasing function such that

$$\begin{aligned}\alpha_n(r) &= r, && 0\le r\le n,\\ \alpha_n(r) &= 2n, && r\ge 3n,\\ \alpha_n''(r) &\le 0, && r\ge 0.\end{aligned}\tag{6.1}$$

Lemma 6.1. *Suppose Conditions 2.1 and 2.2 hold. Then the function I defined in (3.3) has compact level sets.*

Proof. As is typical in the weak convergence approach, the following proof that the rate function I has compact level sets uses arguments that are similar to those used to prove the Laplace principle upper bound in Section 5.

Fix $M < \infty$. We first show that the family $\{\gamma \in \mathcal{P}(\mathbb{R}^d) : I(\gamma) \le M\}$ is tight in $\mathcal{P}(\mathbb{R}^d)$. Let $\varepsilon > 0$. Let $V \in C^2(\mathbb{R}^d)$ and $g \in C(\mathbb{R}^d)$ be as in Condition 2.2. Set

$$K_1 \doteq \left\{x\in\mathbb{R}^d : g(x)\le\frac{M+\varepsilon}{\varepsilon}\right\}.\tag{6.2}$$

By (2.1), K_1 is compact. Suppose $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $I(\mu) \le M$. Then there exists $\nu \in \mathcal{S}_\mu$ such that

$$\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu(dx\,dy) < M+\varepsilon.\tag{6.3}$$

For each $n \in \mathbb{N}$, define $V_n \in C_b^2(\mathbb{R}^d)$ by $V_n \doteq \alpha_n \circ V$. Then

$$(\mathcal{L}^y V_n)(x) = (\alpha_n'\circ V)(x)(\mathcal{L}^y V)(x)+\frac{1}{2}(\alpha_n''\circ V)(x)|\sigma(x)\nabla V(x)|^2$$

and $(\mathcal{L}^y V_n)(x)$ converges to $(\mathcal{L}^y V)(x)$ uniformly for (x,y) in compact subsets of $\mathbb{R}^d\times\mathbb{R}^d$ as $n\to\infty$. Since, for each $n\in\mathbb{N}$, $0\le\alpha_n'(r)\le 1$ and $\alpha_n''(r)\le 0$ for all $r>0$, the last display implies that $(\mathcal{L}^y V_n)(x) \leq \max\{(\mathcal{L}^y V)(x), 0\}$. Then by (5.6),

$$-(\mathcal{L}^y V_n)(x) + \frac{1}{2}|y|^2 \geq \min\{-(\mathcal{L}^y V)(x), 0\} + \frac{1}{2}|y|^2 \geq \min\left\{g(x), \frac{1}{2}|y|^2\right\},$$

which is bounded below by Condition 2.2. Therefore, by (5.6), (2.1), Fatou's lemma, the facts that $\nu \in \mathcal{S}_\mu$ and $V_n \in C_b^2(\mathbb{R}^d)$, and (6.3),

$$\begin{aligned}\int_{\mathbb{R}^d} g(x)\mu(dx) &\leq -\int_{\mathbb{R}^d\times\mathbb{R}^d} (\mathcal{L}^y V)(x)\nu(dx\,dy) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\nu(dx\,dy) \\ &\leq \liminf_{n\to\infty}\left[-\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathcal{L}^y V_n)(x)\mu(dx) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu(dx\,dy)\right] \\ &< M + \varepsilon.\end{aligned} \tag{6.4}$$

Thus, by (6.2), Chebyshev's inequality and the last display, $\mu(K_1^c) < \varepsilon$. This proves that the level set $\{\gamma \in \mathcal{P}(\mathbb{R}^d) : I(\gamma) \leq M\}$ is tight in $\mathcal{P}(\mathbb{R}^d)$.

We are left to show that I is lower semicontinuous. Let $\{\mu_n, n \in \mathbb{N}\}$ be a sequence in $\mathcal{P}(\mathbb{R}^d)$ that converges weakly to μ and satisfies $\liminf_{n\to\infty} I(\mu_n) < \infty$. By choosing an appropriate subsequence, also denoted $\{\mu_n, n \in \mathbb{N}\}$, we can assume $R \doteq \sup_n I(\mu_n) < \infty$ and $\lim_{n\to\infty} I(\mu_n) = \liminf_{n\to\infty} I(\mu_n)$. By the definition of I given in (3.3), for each $n \in \mathbb{N}$ we can choose $\nu_n \in \mathcal{S}_{\mu_n}$ such that

$$\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\nu_n(dx\,dy) \leq I(\mu_n) + \frac{1}{n} \leq R + 1. \tag{6.5}$$

We show that $\{\nu_n, n \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Since $[\nu_n]_1 = \mu_n$ for each $n \in \mathbb{N}$ and $\mu_n \Rightarrow \mu$ as $n \to \infty$, $\{[\nu_n]_1, n \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathbb{R}^d)$. Next, let $\varepsilon > 0$ and define the compact set K_2 by

$$K_2 \doteq \left\{y \in \mathbb{R}^d : |y|^2 \leq \frac{R+1}{\varepsilon}\right\}.$$

Then by Chebyshev's inequality and (6.5), for each $n \in \mathbb{N}$,

$$[\nu_n]_2(K_2^c) \leq \frac{\varepsilon}{R+1}\int_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\nu_n(dx\,dy) \leq 2\varepsilon.$$

This proves that $\{[\nu_n]_2, n \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathbb{R}^d)$ and hence $\{\nu_n, n \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. By possibly taking a further subsequence, also denoted $\{\nu_n, n \in \mathbb{N}\}$, we can assume there exists $\nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $[\nu]_1 = \mu$ and $\nu_n \Rightarrow \nu$ in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ as $n \to \infty$. By (6.5),

$$\begin{aligned}\lim_{C\to\infty}\sup_n \int_{\mathbb{R}^d\times\mathbb{R}^d} 1_{\{|y|\geq C\}}|y|\nu_n(dx\,dy) &\leq \lim_{C\to\infty}\frac{1}{C}\sup_n\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu_n(dx\,dy) \\ &= 0.\end{aligned}$$

Let $\phi \in C_b^2(\mathbb{R}^d)$. Due to the uniform integrability shown in the last display and the weak convergence $\nu_n \Rightarrow \nu$, we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle\sigma(x)y, \nabla\phi(x)\rangle\nu(dx\,dy) = \lim_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d}\langle\sigma(x)y, \nabla\phi(x)\rangle\nu_n(dx\,dy).$$

Consequently,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathcal{L}^y\phi)(x)\nu(dx\,dy) = \lim_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathcal{L}^y\phi)(x)\nu_n(dx\,dy) = 0.$$

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

It follows that $\nu \in \mathcal{S}_\mu$. In addition, by (2.2) of Condition 2.2 and (6.4), $b \in L^2_\mu$ holds. Therefore, we have the following relations, which are explained below:

$$\begin{aligned}\liminf_{n\to\infty} I(\mu_n) &\geq \frac{1}{2}\liminf_{n\to\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu_n(dx\,dy)\\ &\geq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu(dx\,dy)\\ &\geq I(\mu).\end{aligned}$$

The first inequality follows from (6.5). The second inequality is due to Lemma 5.2. The final inequality is due to the facts that $\nu \in \mathcal{S}_\mu$ and $b \in L^2_\mu$, and the definition of I in (3.3). This completes the proof that I is lower semicontinuous. ■

Lemma 6.2. *The function I defined in (3.3) is convex.*

Proof. Suppose $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d)$ are such that $I(\mu_1)$ and $I(\mu_2)$ are finite and $\lambda_1, \lambda_2 \in (0, 1)$ satisfy $\lambda_1 + \lambda_2 = 1$. Let $\varepsilon > 0$. For each $k = 1, 2$, by the definition of I in (3.3) we can choose $\nu_k \in \mathcal{S}_{\mu_k}$ such that

$$\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu_k(dx\,dy) \leq I(\mu_k) + \varepsilon. \tag{6.6}$$

Define $\mu \in \mathcal{P}(\mathbb{R}^d)$ by $\mu \doteq \lambda_1\mu_1 + \lambda_2\mu_2$ and $\nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ by $\nu \doteq \lambda_1\nu_1 + \lambda_2\nu_2$ so that $[\nu]_1 = \mu$. Let $\phi \in C^2_b(\mathbb{R}^d)$. By the definition of $\mathcal{L}^y$ in (3.1), the definition of ν, and the fact that $\nu_k \in \mathcal{S}_{\mu_k}$ for $k = 1, 2$, we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathcal{L}^y\phi)(x)\nu(dx\,dy) = \sum_{k=1,2}\lambda_k\int_{\mathbb{R}^d\times\mathbb{R}^d}(\mathcal{L}^y\phi)(x)\nu_k(dx\,dy) = 0.$$

Hence, $\nu \in \mathcal{S}_\mu$. The fact that $b \in L^2_{\mu_1} \cap L^2_{\mu_2}$ and the definition of μ imply that $b \in L^2_\mu$. Then, by the definition of I, the fact that $\nu = \lambda_1\nu_1 + \lambda_2\nu_2$ and (6.6),

$$I(\mu) \leq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y|^2\nu(dx\,dy) \leq \lambda_1 I(\mu_1) + \lambda_2 I(\mu_2) + 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof that I is convex. ■

We now show that the rate function $I(\mu)$ can be expressed as the infimum over feedback controls u in L^2_μ such that μ is formally invariant under the controlled infinitesimal generator. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, define

$$I^\dagger(\mu) \doteq \begin{cases}\displaystyle\inf_{u\in\mathcal{R}_\mu}\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|u(x)|^2\mu(dx) & \text{if } b \in L^2_\mu \text{ and } \mathcal{R}_\mu \neq \emptyset,\\ \infty & \text{otherwise,}\end{cases} \tag{6.7}$$

where

$$\mathcal{R}_\mu \doteq \left\{u \in L^2_\mu : \int_{\mathbb{R}^d}((\mathcal{L}\phi)(x) + \langle\sigma(x)u(x), \nabla\phi(x)\rangle)\mu(dx) = 0\ \forall\,\phi \in C^\infty_c(\mathbb{R}^d)\right\}. \tag{6.8}$$

Lemma 6.3. *$I(\mu) = I^\dagger(\mu)$ for all $\mu \in \mathcal{P}(\mathbb{R}^d)$.*

Proof. Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. We first show that $I^\dagger(\mu) \leq I(\mu)$. We can assume that $I(\mu) < \infty$. Let $\varepsilon > 0$. By (3.3), we can choose $\nu \in \mathcal{S}_\mu$ such that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 \nu(dx\,dy) < I(\mu) + \varepsilon.$$

Define

$$u(x) \doteq \int_{\mathbb{R}^d} y[\nu]_{2|1}(dy|x), \qquad x \in \mathbb{R}^d,$$

where $[\nu]_{2|1}(dy|x)$ denotes the stochastic kernel on $\mathbb{R}^d$ given $\mathbb{R}^d$ (see [8, Theorem A.5.4]). Since $x \mapsto [\nu]_{2|1}(dy|x)$ is a measurable function from $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ to $\mathcal{P}(\mathbb{R}^d)$ (equipped with the weak topology), the function $u : \mathbb{R}^d \to \mathbb{R}^d$ is Borel measurable. By Jensen's inequality,

$$\int_{\mathbb{R}^d} |u(x)|^2 \mu(dx) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 \nu(dx\,dy) < I(\mu) + \varepsilon.$$

In addition, by (3.2), for all $\phi \in C_c^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} ((\mathcal{L}\phi)(x) + \langle \sigma(x)u(x), \nabla\phi(x)\rangle)\mu(dx) = \int_{\mathbb{R}^d \times \mathbb{R}^d} (\mathcal{L}^y\phi)(x)\nu(dx\,dy) = 0.$$

Thus, $\mu \in \mathcal{R}_\mu$ and so $I^\dagger(\mu) \leq I(\mu) + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $I^\dagger(\mu) \leq I(\mu)$.

Next, we show that $I(\mu) \leq I^\dagger(\mu)$, which will complete the proof. We can assume that $I^\dagger(\mu) < \infty$. Let $\varepsilon > 0$ and choose $u \in \mathcal{R}_\mu$ such that

$$\int_{\mathbb{R}^d} |u(x)|^2 \mu(dx) < I^\dagger(\mu) + \varepsilon.$$

Define $\nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ by

$$\nu(A \times B) \doteq \int_A \delta_{u(x)}(B)\mu(dx), \qquad A, B \in \mathcal{B}(\mathbb{R}^d).$$

Then

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 \nu(dx\,dy) = \int_{\mathbb{R}^d} |u(x)|^2 \nu(dx) < I^\dagger(x) + \varepsilon.$$

In addition, by (3.1) and (6.8), for all $\phi \in C_c^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\mathcal{L}^y\phi)(x)\nu(dx\,dy) = \int_{\mathbb{R}^d} (\mathcal{L}(x) + \langle u(x)\sigma(x), \nabla\phi(x)\rangle)\mu(dx) = 0.$$

Thus, $\nu \in \mathcal{S}_\mu$ and so $I(\mu) \leq I^\dagger(\mu) + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this proves $I(\mu) \leq I^\dagger(\mu)$. ■

Next we obtain an explicit characterization of the rate function. Given $\mu \in \mathcal{P}(\mathbb{R}^d)$ we equip L^2_μ with the inner product

$$\langle f, g\rangle_{L^2_\mu} \doteq \int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \mu(dx). \tag{6.9}$$

Then L^2_μ is a separable Hilbert space. Let G_μ be the linear subspace of L^2_μ defined as the L^2_μ-closure of the set

$$G \doteq \left\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^d)\right\},$$

where $C_c^\infty(\mathbb{R}^d)$ denotes the space of real-valued functions on $\mathbb{R}^d$ with compact support and continuous derivatives of all orders. We let Π_μ denote the orthogonal projection of L^2_μ onto

G_μ. Suppose $b \in L^2_\mu$ and Condition 2.6 holds. Define the continuous function $\beta : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\beta_i \doteq b_i - \frac{1}{2}\sum_{j=1}^{d} \frac{\partial a_{ij}}{\partial x_j}, \qquad 1 \le i \le d. \tag{6.10}$$

Since $b \in L^2_\mu$ and Condition 2.6 holds, $a^{-1}\beta \in L^2_\mu$. For the following lemma, let $H^1(\mathbb{R}^d)$ denote the Sobolev space of real-valued functions f on $\mathbb{R}^d$ that are weakly differentiable and such that f and its weak first partial derivatives are L^2-integrable with respect to Lebesgue measure.

Lemma 6.4. *Suppose Condition 2.6 holds and $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $I(\mu) < \infty$. Then μ is absolutely continuous with respect to Lebesgue measure, $\varphi \doteq \sqrt{\rho}$ lies in $H^1(\mathbb{R}^d)$ and $(\nabla\rho/\rho)1_{\{\rho>0\}}$ coincides μ-a.e. with $\Pi_\mu(a^{-1}\beta + \sigma^{-1}u)$.*

Proof. Suppose $I(\mu) < \infty$. By Lemma 6.3 and (6.8), we can choose $u \in L^2_\mu$ such that, for all $\phi \in C_c^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} ((\mathcal{L}\phi)(x) + \langle \sigma(x)u(x), \nabla\phi(x)\rangle)\mu(dx) = 0. \tag{6.11}$$

Since σ is bounded and $b \in L^2_\mu$ by (3.3), it follows that $b + \sigma u \in L^2_\mu$. The conclusion of the lemma then follows from (6.11) and [1, Theorem 1.1] (with $A = a$ and $B = b + \sigma u$). ■

Proposition 6.5. *Suppose Conditions 2.2 and 2.6 hold and $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $I(\mu) < \infty$. Then*

$$I(\mu) = \frac{1}{2}\int_{\mathbb{R}^d} |u(x)|^2 \mu(dx), \tag{6.12}$$

where $u : \mathbb{R}^d \to \mathbb{R}^d$ is the Borel measurable function given by

$$u = \sigma\left\{\frac{1}{2}\frac{\nabla\rho}{\rho}1_{\{\rho>0\}} - \Pi_\mu(a^{-1}\beta)\right\}, \tag{6.13}$$

where $\rho \doteq d\mu/dx$ is the density of μ with respect to Lebesgue measure. Moreover, if $\rho \in C_b^2(\mathbb{R}^d)$ and is strictly positive on $\mathbb{R}^d$, then u_i lies in $C^{1,\alpha}(\mathbb{R}^d)$ for $1 \le i \le d$.

Remark 6.6. This result can be compared to the characterization of the variational form of the Donsker-Varadhan rate function for diffusions in compact regions (with reflection along the boundary) that was obtained in equations (1.7) and (1.8) of [17]. When the reflection vector in [17] is normal to the boundary of the region (i.e., when $T = 0$), the characterization of the rate function in (1.7) and (1.8) of [17] takes a similar form as our characterization in (6.12) and (6.13). The main difference being that in [17] the projection term $\Pi_\mu(a^{-1}\beta)$ is characterized as ∇h, where h is the solution to the variational problem (1.8) of [17].

Remark 6.7. A useful interpretation of the feedback control u in (6.13) is as the product of σ and the difference between half the logarithmic gradient of ρ and the component of $a^{-1}\beta$ that is reversible with respect to μ. This interpretation can be compared to [15, Corollary 1.5], which provides a probabilistic interpretation of the Donsker-Varadhan rate function for diffusions in compact regions (with reflection along the boundary of the region) in terms of invariant measures for such diffusions.

Proof. We first show that $u \in \mathcal{R}_\mu$. Due to the definition of u given in (6.13), the definition of $\langle \cdot, \cdot \rangle_{L^2_\mu}$ given in (6.9), the fact that $\nabla \phi \in G_\mu$ and the definition of β given in (6.10), we have

$$\begin{aligned}\int_{\mathbb{R}^d} \langle \sigma(x)u(x), \nabla\phi(x)\rangle \mu(dx) &= \frac{1}{2}\langle \rho^{-1}\nabla\rho, \nabla\phi\rangle_{L^2_\mu} - \langle a^{-1}\beta, \nabla\phi\rangle_{L^2_\mu} \\ &= \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d \int_{\mathbb{R}^d} \frac{\partial(a_{ij}\rho)}{\partial x_i}(x)\frac{\partial\phi}{\partial x_j}(x)dx \\ &\quad - \int_{\mathbb{R}^d} \langle b(x), \nabla\phi(x)\rangle \mu(dx).\end{aligned}$$

By the last display and integration by parts, we obtain

$$\begin{aligned}&\int_{\mathbb{R}^d} ((\mathcal{L}\phi)(x) + \langle \sigma(x)u(x), \nabla\phi(x)\rangle)\mu(dx) \\ &\quad = \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d \int_{\mathbb{R}^d} a_{ij}(x)\frac{\partial^2\phi}{\partial x_i \partial x_j}(x)\rho(x)dx \\ &\qquad + \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d \int_{\mathbb{R}^d} \frac{\partial(a_{ij}\rho)}{\partial x_i}(x)\frac{\partial\phi}{\partial x_j}(x)dx \\ &\quad = 0.\end{aligned}$$

This proves $u \in \mathcal{R}_\mu$.

Now suppose $\tilde{u}$ also lies in $\mathcal{R}_\mu$. Then by the definition of $\langle \cdot, \cdot \rangle_{L^2_\mu}$,

$$\begin{aligned}\int_{\mathbb{R}^d} |\tilde{u}(x)|^2\mu(dx) - \int_{\mathbb{R}^d} |u(x)|^2\mu(dx) \geq \int_{\mathbb{R}^d} |\tilde{u}(x) - u(x)|^2\mu(dx) \\ + 2\langle \sigma^{-1}(\tilde{u} - u), \sigma^{-1}u\rangle_{L^2_\mu}.\end{aligned}$$

According to Lemma 6.3, once we show $\langle \sigma^{-1}(\tilde{u} - u), \sigma^{-1}u\rangle_{L^2_\mu} = 0$, the proof of (6.12) will be complete. By Lemma 6.4, $(\nabla\rho/\rho)1_{\{\rho>0\}} \in G_\mu$. Then (6.13) and the definition of Π_μ imply $\sigma^{-1}u \in G_\mu$. Therefore, there exists $\{\phi_n, n \in \mathbb{N}\}$ in $C_c^\infty(\mathbb{R}^d)$ such that

$$\langle \sigma^{-1}(\tilde{u} - u), \sigma^{-1}u\rangle_{L^2_\mu} = \lim_{n\to\infty} \langle \sigma^{-1}(\tilde{u} - u), \nabla\phi_n\rangle_{L^2_\mu}.$$

For each $n \in \mathbb{N}$, since $u, \tilde{u} \in \mathcal{R}_\mu$, we have

$$\begin{aligned}\langle \sigma^{-1}(\tilde{u} - u), \nabla\phi_n\rangle_{L^2_\mu} &= \int_{\mathbb{R}^d} \langle \sigma(x)(\tilde{u}(x) - u(x)), \nabla\phi_n(x)\rangle \mu(dx\, dy) \\ &= \int_{\mathbb{R}^d} ((\mathcal{L}\phi_n)(x) + \langle \sigma(x)\tilde{u}(x), \nabla\phi(x)\rangle)\mu(dx) \\ &\quad - \int_{\mathbb{R}^d} ((\mathcal{L}\phi_n)(x) + \langle \sigma(x)u(x), \nabla\phi(x)\rangle)\mu(dx) \\ &= 0.\end{aligned}$$

Thus, (6.13) holds.

Lastly, suppose ρ lies in $C_b^2(\mathbb{R}^d)$ and is strictly positive on $\mathbb{R}^d$. Since $\Pi_\mu(a^{-1}\beta) \in G_\mu$, there exists a weakly differentiable function $\psi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\psi = \Pi_\mu(a^{-1}\beta)$. Thus, $\langle \nabla\psi, \nabla\phi\rangle_{L^2_\mu} = \langle a^{-1}\beta, \nabla\phi\rangle_{L^2_\mu}$ for all $\phi \in C_c^\infty(\mathbb{R}^d)$. Using the definition of $\langle \cdot, \cdot \rangle_{L^2_\mu}$ in (6.9), integrating by parts and dividing both sides by ρ, we see that ψ is a weak solution to the PDE

$$\nabla \cdot (a\nabla\psi) + \langle \nabla \log \rho, \nabla\psi\rangle = \nabla \cdot (\rho\beta). \tag{6.14}$$

Then according to [13, Theorem 9.19] and because Conditions 2.1 and 2.6 hold, ψ belongs to $C^{2,\alpha}(\mathbb{R}^d)$. Since $\psi \in C^{2,\alpha}(\mathbb{R}^d)$, $\rho \in C_b^2(\mathbb{R}^d)$ and ρ is strictly positive on $\mathbb{R}^d$, it follows from (6.13) that $u_i \in C^{1,\alpha}(\mathbb{R}^d)$ for $1 \le i \le d$. ■

Lemma 6.8. *Suppose Condition 2.6 holds and $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $I(\mu) < \infty$. Then there exists a sequence $\{\mu_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{R}^d)$ such that for each $n \in \mathbb{N}$, μ_n is absolutely continuous with respect to Lebesgue measure, its density $\rho_n \doteq d\mu_n/dx$ belongs in $C_b^\infty(\mathbb{R}^d)$ and is strictly positive on $\mathbb{R}^d$, $\mu_n \Rightarrow \mu$ as $n \to \infty$ and $I(\mu_n) \to I(\mu)$ as $n \to \infty$.*

The proof of Lemma 6.8 is given in the Appendix.

7. Laplace principle lower bound

The following proposition is the main result of this section.

Proposition 7.1. *Suppose Conditions 2.1, 2.2 and 2.6 hold. For each compact subset K of $\mathbb{R}^d$ and bounded continuous function h mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}$ we have the following uniform Laplace principle lower bound*

$$\liminf_{t\to\infty} \inf_{x\in K} \frac{1}{t} \log E_x \left[\exp\{-th(L_t)\}\right] \ge - \inf_{\gamma\in\mathcal{P}(\mathbb{R}^d)} \{I(\gamma) + h(\gamma)\}. \tag{7.1}$$

Proof. Fix a bounded continuous function h mapping $\mathcal{P}(\mathbb{R}^d)$ into $\mathbb{R}^d$. Let $\varepsilon > 0$ be arbitrary. According to Lemma 6.8 and the continuity of h, there exists $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that μ is absolutely continuous with respect to Lebesgue measure, its density with respect to Lebesgue measure $\rho \doteq d\mu/dx$ lies in $C_b^\infty(\mathbb{R}^d)$ and is strictly positive on $\mathbb{R}^d$, and

$$I(\mu) + h(\mu) \le \inf_{\gamma\in\mathcal{P}(\mathbb{R}^d)} \{I(\gamma) + h(\gamma)\} + \varepsilon. \tag{7.2}$$

Define $u \in L_\mu^2$ as in (6.13). Since $\rho \in C_b^\infty(\mathbb{R}^d)$ it follows from Proposition 6.5 that u is continuously differentiable and hence locally Lipschitz continuous. Therefore, along with Conditions 2.1 and 2.2, this ensures there exists a pathwise unique solution $X^u = \{X^u(t), t \ge 0\}$ of the controlled SDE

$$dX^u(t) = b(X^u(t))dt + \sigma(X^u(t))u(X^u(t))dt + \sigma(X^u(t))dW(t), \qquad t \ge 0,$$

with $X^u(0) = X_0$. We let L_t^u denote the associated empirical measure process. Since u and X^u are continuous, the process $\{u(X^u(t)), t \ge 0\}$ is continuous and hence progressively measurable with respect to the filtration $\{\mathcal{F}_t\}$.

Let $\{t_n, n \in \mathbb{N}\}$ and $\{x_n, n \in \mathbb{N}\}$ be sequences in $(0, \infty)$ and K, respectively, such that $t_n \to \infty$ as $n \to \infty$. By Proposition 4.2, for each $n \in \mathbb{N}$, we have

$$\begin{aligned}\frac{1}{t_n} \log E_{x_n} \left[\exp\left\{-t_n h(L_{t_n})\right\}\right] &\ge -E_{x_n} \left[\frac{1}{2t_n} \int_0^{t_n} |u(X^u(s))|^2 ds + h(L_{t_n}^u)\right] \\ &= -E_{x_n} \left[\frac{1}{2} \int_{\mathbb{R}^d} |u(x)|^2 L_{t_n}^u(dx) + h(L_{t_n}^u)\right].\end{aligned}$$

We claim that

$$\lim_{n\to\infty} E_{x_n}\left[\frac{1}{2}\int_{\mathbb{R}^d}|u(x)|^2 L^u_{t_n}(dx)+h(L^u_{t_n})\right]=\frac{1}{2}\int_{\mathbb{R}^d}|u(x)|^2\mu(dx)+h(\mu) \tag{7.3}$$
$$=I(\mu)+h(\mu),$$

which along with (7.2) will complete the proof. Since the proof of the claim is analogous to a portion of the proof of [8, Proposition 8.6.1], which establishes the uniform Laplace principle lower bound for discrete time Markov chains, we provide a sketch of the argument here and refer the reader to [8] for the details.

Let P_μ denote the probability measure on $(\Omega,\mathcal{F})$ given by $P_\mu(A)\doteq\int_{\mathbb{R}^d}P_x(A)\mu(dx)$ for all $A\in\mathcal{F}$, and let E_μ denote expectation under μ. Since μ is a stationary distribution of the controlled diffusion X^u, an application of the L^1 ergodic theorem implies that

$$\lim_{n\to\infty} E_\mu\left[\int_{\mathbb{R}^d}|u(x)|^2 L^u_{t_n}(dx)\right]=\int_{\mathbb{R}^d}|u(x)|^2\mu(dx).$$

Then, using the definition of P_μ, Chebyshev's inequality, the Borel–Cantelli lemma and by possibly taking a subsequence, which we also denote by $\{t_n, n\in\mathbb{N}\}$, we have for μ-a.e. $x\in\mathbb{R}^d$,

$$\lim_{n\to\infty} E_x\left[\int_{\mathbb{R}^d}|u(x)|^2 L^u_{t_n}(dx)\right]=\int_{\mathbb{R}^d}|u(y)|^2\mu(dy).$$

Next, given a bounded, uniformly continuous function mapping $\mathbb{R}^d$ to $\mathbb{R}$, the pointwise ergodic theorem implies that P_μ a.s.

$$\lim_{n\to\infty}\int_{\mathbb{R}^d}g(y)L^u_{t_n}(dy)=\int_{\mathbb{R}^d}g(y)\mu(dy). \tag{7.4}$$

Since there is a metric on $\mathbb{R}^d$ under which the space of bounded, uniformly continuous functions on $\mathbb{R}^d$ is separable (see [8, Theorem A.6.1]), the definition of P_μ and last display implies that for μ-a.e. $x\in\mathbb{R}^d$, P_x a.s. $L^u_{t_n}\Rightarrow\mu$. Combining our results so far, we have (7.3) holds for μ-a.e. $x\in\mathbb{R}^d$. The extension to all $x\in\mathbb{R}^d$ then follows from an argument using the Feller continuity of the diffusion X. This completes the proof of our claim. ■

Acknowledgment

We thank Vivek Borkar and Ari Arapostathis for a useful discussion about the existence and regularity of solutions to the PDE (6.14). We also thank Thomas Kurtz for his suggestion to use the Gram-Schmidt decomposition of the orthogonal projection operators in Appendix A and Johnny Guzmán for a useful discussion related to the construction of the approximations $\varphi_{n,m}$ introduced in Appendix B. Finally, we thank the reviewer for a careful reading of this paper and for their helpful comments. The first author's research was supported in part by the National Science Foundation (DMS-1317199) and the Defense Advanced Research Projects Agency (W911NF-15-2-0122). The second author's research was supported in part by the National Science Foundation (DMS-1148284) and the Air Force Office of Scientific Research (FA9550-12-1-0399).

Appendix A. Properties of weighted L^2-projection operators

Let $\mu\in\mathcal{P}(\mathbb{R}^d)$ and $\{\mu_n, n\in\mathbb{N}\}$ be a sequence in $\mathcal{P}(\mathbb{R}^d)$ satisfying the following condition.

Condition A.1. *The following hold:*

- $L^2_\mu \subset L^2_{\mu_n}$ *for each* $n \in \mathbb{N}$,
- $\lim_{n\to\infty}\langle f, g\rangle_{L^2_{\mu_n}} = \langle f, g\rangle_{L^2_\mu}$ *for all* $f, g \in L^2_\mu$.

Let G_μ be an infinite-dimensional linear subspace of L^2_μ and let Π_μ denote the orthogonal projection from L^2_μ onto G_μ. For each $n \in \mathbb{N}$, let G_{μ_n} denote the closure of G_μ in $L^2_{\mu_n}$ and let Π_{μ_n} denote the orthogonal projection from $L^2_{\mu_n}$ onto G_{μ_n}. The following proposition is the main result of this section.

Proposition A.2. *Suppose Condition A.1 holds. Then for all* $f \in L^2_\mu$, $\langle \Pi_{\mu_n} f, \Pi_{\mu_n} f\rangle_{L^2_{\mu_n}} \to \langle \Pi_\mu f, \Pi_\mu f\rangle_{L^2_\mu}$ *as* $n \to \infty$.

The remainder of this section is devoted to proving Proposition A.2. Throughout the section we assume Condition A.1 holds and, given $\gamma \in \mathcal{P}(\mathbb{R}^d)$ and $f \in L^2_\gamma$, we let $\|f\|_{L^2_\gamma} \doteq \sqrt{\langle f, f\rangle_{L^2_\gamma}}$.

Since L^2_μ is separable, there is a sequence $\{g_i, i \in \mathbb{N}\}$ of linearly independent vector fields in G_μ such that G_μ is equal to the closure of $G \doteq \text{span}\{g_i, i \in \mathbb{N}\}$ in L^2_μ. For each $m \in \mathbb{N}$, define the finite dimensional linear subspace G^m by

$$G^m \doteq \text{span}\{g_i, i = 1, \ldots, m\}.$$

Let Π^m_μ denote the orthogonal projection of L^2_μ onto G^m. We can recursively construct these projections via Gram–Schmidt as follows: Set

$$\Pi^1_\mu f \doteq \frac{\langle f, g_1\rangle_{L^2_\mu}}{\|g_1\|^2_{L^2_\mu}} g_1, \tag{A.1}$$

and for $m \in \mathbb{N}$, set

$$\Pi^{m+1}_\mu f \doteq \Pi^m_\mu f + \frac{\langle f, g_{m+1} - \Pi^m_\mu g_{m+1}\rangle_{L^2_\mu}}{\|g_{m+1} - \Pi^m_\mu g_{m+1}\|^2_{L^2_\mu}}(g_{m+1} - \Pi^m_\mu g_{m+1}). \tag{A.2}$$

Here the linear independence of the vector fields G ensures that (A.1) and (A.2) are well defined. For each $n, m \in \mathbb{N}$, let $\Pi^m_{\mu_n}$ denote the orthogonal projection from $L^2_{\mu_n}$ onto G^m. Given $n \in \mathbb{N}$, we can similarly recursively construct the projections $\Pi^m_{\mu_n}$, $m \in \mathbb{N}$, via Gram–Schmidt. Since the procedure is analogous to the one explained above, we omit the details.

Lemma A.3. *For each* $m \in \mathbb{N}$ *and all* $f \in L^2_\mu$,

$$\lim_{n\to\infty} \|\Pi^m_{\mu_n} f\|_{L^2_{\mu_n}} = \|\Pi^m_\mu f\|_{L^2_\mu}.$$

Proof. We proceed with a proof by induction. Let $f \in L^2_\mu$. By (A.1) and Condition A.1,

$$\lim_{n\to\infty} \|\Pi^1_{\mu_n} f\|_{L^2_{\mu_n}} = \lim_{n\to\infty} \frac{\langle f, g_1\rangle_{L^2_{\mu_n}}}{\|g_1\|_{L^2_{\mu_n}}} = \frac{\langle f, g_1\rangle_{L^2_\mu}}{\|g_1\|_{L^2_\mu}} = \|\Pi^1_\mu f\|_{L^2_\mu}.$$

This establishes the base case. Now let $m \geq 1$ and assume the following induction hypothesis: $\lim_{n\to\infty}\|\Pi^m_{\mu_n} f\|_{L^2_{\mu_n}} = \|\Pi^m_\mu f\|_{L^2_\mu}$ for all $f \in L^2_\mu$. Since the orthogonal projection operators are idempotent and self-adjoint,

$$\langle f, \Pi^m_{\mu_n} g\rangle_{L^2_{\mu_n}} = \langle \Pi^m_{\mu_n} f, \Pi^m_{\mu_n} g\rangle_{L^2_{\mu_n}} \to \langle \Pi^m_\mu f, \Pi^m_\mu g\rangle_{L^2_\mu} = \langle f, \Pi^m_\mu g\rangle_{L^2_\mu} \tag{A.3}$$

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

as $n \to \infty$ for all $f, g \in L^2_\mu$, where the convergence follows from the identity

$$4\langle \Pi^m_{\mu_n} f, \Pi^m_{\mu_n} g\rangle_{L^2_{\mu_n}} = \|\Pi^m_{\mu_n}(f+g)\|_{L^2_{\mu_n}} - \|\Pi^m_{\mu_n}(f-g)\|_{L^2_{\mu_n}}$$

and the induction hypothesis. Then by orthogonality, the induction hypothesis, Condition A.1 and (A.3), given $f \in L^2_\mu$,

$$\begin{aligned}\lim_{n\to\infty} \|\Pi^{m+1}_{\mu_n} f\|_{L^2_{\mu_n}} &= \lim_{n\to\infty} \|\Pi^m_{\mu_n} f\|_{L^2_{\mu_n}} + \lim_{n\to\infty} \frac{\langle f, g_{m+1} - \Pi^m_{\mu_n} g_{m+1}\rangle_{L^2_{\mu_n}}}{\|g_{m+1} - \Pi^m_{\mu_n} g_{m+1}\|_{L^2_{\mu_n}}} \\ &= \|\Pi^m_\mu f\|_{L^2_\mu} + \frac{\langle f, g_{m+1} - \Pi^m_\mu g_{m+1}\rangle_{L^2_\mu}}{\|g_{m+1} - \Pi^m_\mu g_{m+1}\|_{L^2_\mu}} \\ &= \|\Pi^{m+1}_\mu f\|_{L^2_\mu}.\end{aligned}$$

This proves the induction step. The lemma now follows from the principle of mathematical induction. ■

Lemma A.4. *For all* $f \in L^2_\mu$,

$$\liminf_{n\to\infty} \|\Pi_{\mu_n} f\|^2_{L^2_{\mu_n}} \ge \|\Pi_\mu f\|^2_{L^2_\mu}.$$

Proof. By Lemma A.3, for each $m \in \mathbb{N}$, we have

$$\liminf_{n\to\infty} \|\Pi_{\mu_n} f\|^2_{L^2_{\mu_n}} \ge \liminf_{n\to\infty} \|\Pi^m_{\mu_n} f\|^2_{L^2_{\mu_n}} = \|\Pi^m_\mu f\|^2_{L^2_\mu}.$$

Letting $m \to \infty$ yields the conclusions of the lemma. ■

Let $G^\perp_\mu$ denote the orthogonal complement of G_μ in L^2_μ and let $\Pi^\perp_\mu$ denote the orthogonal projection operator from L^2_μ onto $G^\perp_\mu$. Similarly, for each $n \ge 1$, let $G^\perp_{\mu_n}$ denote the orthogonal complement of G_{μ_n} in $L^2_{\mu_n}$ and let $\Pi^\perp_{\mu_n}$ denote the orthogonal projection operator from $L^2_{\mu_n}$ onto $G^\perp_{\mu_n}$. The next lemma follows immediately from Lemma A.4, but with $\Pi^\perp_{\mu_n}$ and $\Pi^\perp_\mu$ in place of Π_{μ_n} and Π_μ, respectively. (Lemma A.4 assumes that G_μ is infinite-dimensional; however, it is readily checked that the result still holds if G_μ is finite-dimensional.)

Lemma A.5. *For all* $f \in L^2_\mu$,

$$\liminf_{n\to\infty} \|\Pi^\perp_{\mu_n} f\|^2_{L^2_{\mu_n}} \ge \|\Pi^\perp_\mu f\|^2_{L^2_\mu}.$$

Proof of Proposition A.2. Let $f \in L^2_\mu$. By orthogonality, Condition A.1 and Lemma A.5,

$$\begin{aligned}\limsup_{n\to\infty} \|\Pi_{\mu_n} f\|^2_{L^2_{\mu_n}} &\le \limsup_{n\to\infty} \|f\|^2_{L^2_{\mu_n}} - \liminf_{n\to\infty} \|\Pi^\perp_{\mu_n} f\|^2_{L^2_{\mu_n}} \\ &\le \|f\|^2_{L^2_\mu} - \|\Pi^\perp_\mu f\|^2_{L^2_\mu} \\ &= \|\Pi_\mu f\|^2_{L^2_\mu}.\end{aligned}$$

Along with Lemma A.4, this completes the proof. ■

Appendix B. Proof of Lemma 6.8

Throughout this section we fix $\mu \in \mathcal{P}(\mathbb{R}^d)$ with $I(\mu) < \infty$. By Lemma 6.4, μ is absolutely continuous with respect to Lebesgue measure and if $\varphi \doteq \sqrt{d\mu/dx}$, then φ lies in $H^1(\mathbb{R}^d)$.

For each $n \in \mathbb{N}$, let $\alpha_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a smooth function such that $|\alpha_n'(r)| \leq 1$ for all $r \in \mathbb{R}_+$ and (6.1) holds. In addition, define $\zeta_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a nonincreasing smooth function that satisfies $\zeta_n(r) = 1$ for all $r \leq n$, $\zeta_n(r) = 0$ for all $r \geq n+2$ and $-1 \leq \zeta_n'(r) \leq 0$ for all $n \leq r \leq n+2$. Define $\varphi_n \in H^1(\mathbb{R}^d)$ by

$$\varphi_n(x) \doteq (\alpha_n \circ \varphi(x))\zeta_n(|x|), \qquad x \in \mathbb{R}^d.$$

Then, for all $n \in \mathbb{N}$ sufficiently large, $\varphi_n(x)$ is positive on a set of positive measure. Without loss of generality, we assume that for all $n \in \mathbb{N}$, φ_n is positive on a set of positive measure. Since $\zeta_n(r) = 1$ and $\alpha_n(r) = r$ for all $r \leq n$ and φ lies in $H^1(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} |\varphi(x) - \varphi_n(x)|^2 dx \leq 4 \int_{\max\{|x|,|\varphi(x)|\} \geq n} |\varphi(x)|^2 dx \to 0$$

as $n \to \infty$, and

$$\begin{aligned}\int_{\mathbb{R}^d} |\nabla\varphi(x) - \nabla\varphi_n(x)|^2 dx &\leq 2 \int_{\mathbb{R}^d} |\nabla\varphi(x)|^2 |1 - (\alpha_n' \circ \varphi(x))\zeta_n(|x|)|^2 dx \\ &\quad + 2 \int_{\mathbb{R}^d} |(\alpha_n \circ \varphi(x))\zeta_n'(|x|)|^2 dx \\ &\leq 4 \int_{\max\{|x|,|\varphi(x)|\} \geq n} |\nabla\varphi(x)|^2 dx \\ &\quad + 2 \int_{|x| \geq n} |\varphi(x)|^2 dx \\ &\to 0\end{aligned}$$

as $n \to \infty$. Thus, $\varphi_n \to \varphi$ in $H^1(\mathbb{R}^d)$ as $n \to \infty$.

Let $\eta \in C_c^\infty(\mathbb{R}^d)$ be given by $\eta(x) \doteq c\exp(-(1-|x|^2)^{-1})$ for $|x| < 1$ and $\eta(x) \doteq 0$ for all $|x| \geq 1$, where the normalization constant $c > 0$ is chosen so that $\int_{\mathbb{R}^d} \eta(x)dx = 1$. For each $m \in \mathbb{N}$, define $\eta_m(x) \doteq m^d \eta(mx)$ for all $x \in \mathbb{R}^d$. Given $n, m \in \mathbb{N}$, define $\varphi_{n,m} \in C_c^\infty(\mathbb{R}^d) \cap H^1(\mathbb{R}^d)$ by

$$\varphi_{n,m}(x) \doteq (\alpha_n \circ (\varphi * \eta_m)(x))\zeta_n(|x|), \qquad x \in \mathbb{R}^d \tag{B.1}$$

where $\varphi * \eta_m$ denotes the convolution of φ and η_m. Since φ and η_m both take values in $\mathbb{R}_+$, their convolution also takes values in $\mathbb{R}_+$, so the above expression is well defined. It is well known that $\varphi * \eta_m \to \varphi$ in $H^1(\mathbb{R}^d)$ and a.e. as $m \to \infty$. Since $|\alpha_n'(r)| \leq 1$, $|\zeta_n(r)| \leq 1$ and $|\zeta_n'(r)| \leq 1$ for all $r \geq 0$, we have, for each $n \in \mathbb{N}$,

$$\begin{aligned}\int_{\mathbb{R}^d} |\varphi_{n,m}(x) - \varphi_n(x)|^2 dx &\leq \int_{\mathbb{R}^d} |(\varphi * \eta_m)(x) - \varphi(x)|^2 dx \\ &\to 0\end{aligned}$$

as $m \to \infty$. In addition, because $\varphi * \eta_m \to \varphi$ in $H^1(\mathbb{R}^d)$ and a.e. as $m \to \infty$, and due to the bounds $|\zeta_n(r)| \leq 1$, $|\zeta_n'(r)| \leq 1$ and $|\alpha_n'(r)| \leq 1$ for all $r \in \mathbb{R}_+$, we also have

$$\begin{aligned}&\int_{\mathbb{R}^d} |\nabla\varphi_{n,m}(x) - \nabla\varphi_n(x)|^2 dx \\ &\quad \leq 2 \int_{\mathbb{R}^d} |(\alpha_n' \circ (\varphi * \eta_m)(x))(\nabla\varphi * \eta_m)(x) - (\alpha_n' \circ \varphi(x))\nabla\varphi(x)|^2 dx \\ &\qquad + 2 \int_{\mathbb{R}^d} |\alpha_n \circ (\varphi * \eta_m)(x) - \alpha_n \circ \varphi(x)|^2 dx\end{aligned}$$

$$
\begin{aligned}
&\leq 2\int_{\mathbb{R}^d} |\alpha_n' \circ (\varphi * \eta_m)(x)|^2 |(\nabla\varphi * \eta_m)(x) - \nabla\varphi(x)|^2 dx \\
&\qquad + 2\int_{\mathbb{R}^d} |\alpha_n' \circ (\varphi * \eta_m)(x) - \alpha_n' \circ \varphi(x)|^2 |\nabla\varphi(x)|^2 dx \\
&\qquad + 2\int_{\mathbb{R}^d} |(\varphi * \eta_m)(x) - \varphi(x)|^2 dx \\
&\to 0
\end{aligned}
$$

as $m \to \infty$. Hence, $\varphi_{n,m} \to \varphi_n$ in $H^1(\mathbb{R}^d)$ as $m \to \infty$. For $n, m \in \mathbb{N}$, set

$$
Z_{n,m} \doteq \int_{\mathbb{R}^d} \varphi_{n,m}^2(x)dx \in (0, \infty), \tag{B.2}
$$

and define $\gamma_{n,m} \in \mathcal{P}(\mathbb{R}^d)$ to be absolutely continuous with respect to Lebesgue measure with density

$$
\frac{d\gamma_{n,m}}{dx} \doteq \frac{1}{Z_{n,m}} \varphi_{n,m}^2. \tag{B.3}
$$

Lemma B.1. *Suppose that on each compact set K in $\mathbb{R}^d$, φ is strictly bounded from below by a positive constant. Then there exists a subsequence $\{m_n, n \in \mathbb{N}\}$ such that $m_n \to \infty$ as $n \to \infty$ and, upon setting $\mu_n \doteq \gamma_{n,m_n}$ for each $n \in \mathbb{N}$, Condition A.1 holds, $\mu_n \Rightarrow \mu$ as $n \to \infty$ and $\sqrt{d\mu_n/dx} \to \sqrt{d\mu/dx}$ in $H^1(\mathbb{R}^d)$ as $n \to \infty$.*

Proof. Let $f, g \in L^2_\mu$. Given $n, m \in \mathbb{N}$, due to (B.1) and the facts that α_n is bounded above by $2n$, the support of ζ_n is contained in $K_n \doteq \{x \in \mathbb{R}^d : |x| \leq n + 2\}$ and φ is bounded below by a positive constant on K_n, we have

$$
\begin{aligned}
\int_{\mathbb{R}^d} |\langle a(x)f(x), g(x)\rangle| \gamma_{n,m}(dx) &= \frac{1}{Z_{n,m}} \int_{\mathbb{R}^d} |\langle a(x)f(x), g(x)\rangle| \frac{\varphi_{n,m}^2(x)}{\varphi^2(x)} \mu(dx) \\
&\leq \frac{4n^2}{Z_{n,m}} \frac{\langle f, g\rangle_{L^2_\mu}}{\inf_{x\in K_n} \varphi^2(x)} \\
&< \infty.
\end{aligned}
$$

Thus, $f, g \in L^2_{\gamma_{n,m}}$. By dominated convergence, given $n \in \mathbb{N}$,

$$
\int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \varphi_{n,m}^2(x)dx \to \int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \varphi_n^2(x)dx
$$

as $m \to \infty$. Now for each $n \in \mathbb{N}$,

$$
\int_{\mathbb{R}^d} |\langle a(x)f(x), g(x)\rangle| \varphi_n^2(x)dx \leq \int_{\mathbb{R}^d} |\langle a(x)f(x), g(x)\rangle| \varphi^2(x)dx < \infty.
$$

Therefore, again by dominated convergence,

$$
\int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \varphi_n^2(x)dx \to \int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \varphi^2(x)dx
$$

as $n \to \infty$. Hence, we can choose a subsequence $\{m_n, n \in \mathbb{N}\}$ such that $m_n \to \infty$ as $n \to \infty$ and

$$
\int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \varphi_{n,m_n}^2(x)dx \to \int_{\mathbb{R}^d} \langle a(x)f(x), g(x)\rangle \varphi^2(x)dx
$$

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

as $n \to \infty$. Since $\varphi_{n,m} \to \varphi_n$ in $H^1(\mathbb{R}^d)$ as $m \to \infty$ and $\varphi_n \to \varphi$ in $H^1(\mathbb{R}^d)$ as $n \to \infty$, by possibly selecting $m_n \in \mathbb{N}$ larger for each $n \in \mathbb{N}$, we can assume that $\varphi_{n,m_n} \to \varphi$ in $H^1(\mathbb{R}^d)$ as $n \to \infty$, and thus $Z_{n,m_n} \to \int_{\mathbb{R}^d} |\varphi(x)|^2 dx = 1$ as $n \to \infty$. It follows that Condition A.1 holds and $\sqrt{d\mu_n/dx} \to \sqrt{d\mu/dx}$ in $H^1(\mathbb{R}^d)$ as $n \to \infty$, which implies the weak convergence $\mu_n \Rightarrow \mu$ as $n \to \infty$. This completes the proof. ■

Proposition B.2. *Suppose $\mu \in \mathcal{P}(\mathbb{R}^d)$ is such that $I(\mu) < \infty$ and μ admits a density with respect to Lebesgue measure, denoted $\rho \doteq d\mu/dx$, such that on each compact set K its density ρ is bounded below by a positive constant. Then there is a sequence $\{\mu_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{R}^d)$ such that for each $n \in \mathbb{N}$, μ_n is absolutely continuous with respect to Lebesgue measure and its density $\rho_n \doteq d\mu_n/dx$ lies in $C_c^\infty(\mathbb{R}^d)$, $\mu_n \Rightarrow \mu$ as $n \to \infty$ and $I(\mu_n) \to I(\mu)$ as $n \to \infty$.*

Proof. Define the sequence $\{\mu_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{R}^d)$ as in Lemma B.1. Then we are left to show that $I(\mu_n) \to I(\mu)$ as $n \to \infty$. Set $\rho \doteq d\mu/dx$ and, for each $n \in \mathbb{N}$, set $\rho_n \doteq d\mu_n/dx$. By Proposition 6.5, in order to show that $I(\mu_n) \to I(\mu)$ as $n \to \infty$, it suffices to show that

$$\lim_{n\to\infty} \langle \rho_n^{-1}\nabla\rho_n, \rho_n^{-1}\nabla\rho_n\rangle_{L^2_{\mu_n}} = \langle \rho^{-1}\nabla\rho, \rho^{-1}\nabla\rho\rangle_{L^2_\mu} \tag{B.4}$$

and

$$\lim_{n\to\infty} \langle \Pi_{\mu_n}(a^{-1}\beta), \Pi_{\mu_n}(a^{-1}\beta)\rangle_{L^2_{\mu_n}} = \langle \Pi_{\mu}(a^{-1}\beta), \Pi_{\mu}(a^{-1}\beta)\rangle_{L^2_\mu}. \tag{B.5}$$

By Lemma B.1, $\sqrt{\rho_n} \to \sqrt{\rho}$ in $H^1(\mathbb{R}^d)$ and Condition A.1 holds. Since σ is bounded, we have

$$\begin{aligned}\lim_{n\to\infty} \langle \rho_n^{-1}\nabla\rho_n, \rho_n^{-1}\nabla\rho_n\rangle_{L^2_{\mu_n}} &= 4\lim_{n\to\infty}\int_{\mathbb{R}^d} |\sigma(x)(\nabla\sqrt{\rho_n})(x)|^2 dx \\ &= 4\int_{\mathbb{R}^d} |\sigma(x)(\nabla\sqrt{\rho})(x)|^2 dx \\ &= \langle \rho^{-1}\nabla\rho, \rho^{-1}\nabla\rho\rangle_{L^2_\mu}.\end{aligned}$$

This proves (B.4). Since $a^{-1}\beta \in L^2_\mu$ due to the facts that $b \in L^2_\mu$, a^{-1} is bounded and a has bounded derivatives, it follows from Proposition A.2 that

$$\lim_{n\to\infty} \langle \Pi_{\mu_n}(a^{-1}\beta), \Pi_{\mu_n}(a^{-1}\beta)\rangle_{L^2_{\mu_n}} = \langle \Pi_{\mu}(a^{-1}\beta), \Pi_{\mu}(a^{-1}\beta)\rangle_{L^2_\mu}.$$

This establishes (B.5), thus completing the proof. ■

Proof of Lemma 6.8. Suppose $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $I(\mu) < \infty$. Choose $\mu^* \in \mathcal{P}(\mathbb{R}^d)$ such that $I(\mu^*) < \infty$, μ^* is absolutely continuous with respect to Lebesgue measure with density $d\mu^*/dx$ that lies in $C_b^\infty(\mathbb{R}^d)$ and is strictly positive on $\mathbb{R}^d$. (One is tempted to use the stationary distribution π here for μ^*; however, in this way we avoid having to verify that π satisfies the desired regularity properties. The existence of μ^* with the stated regularity and satisfying $I(\mu^*) < \infty$ can be readily inferred from (6.12) and (6.13).) For each $\delta \in (0, 1)$, define $\mu^\delta \in \mathcal{P}(\mathbb{R}^d)$ by

$$\mu^\delta \doteq (1-\delta)\,\mu + \delta\mu^*. \tag{B.6}$$

By the convexity of the rate function shown in Lemma 6.2, $I(\mu^\delta) \le (1-\delta)I(\mu) + \delta I(\mu^*) < \infty$. Then by Lemma 6.4 and the fact that $d\mu^*/dx$ is continuous and strictly positive on $\mathbb{R}^d$, μ^δ admits a density with respect to Lebesgue measure such that for each compact subset K of $\mathbb{R}^d$, $d\mu^\delta/dx$ is bounded below by a positive constant on K. Thus, by Proposition B.2, there is a sequence

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.

$\{\mu^\delta_m, m \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{R}^d)$ such that for each $m \in \mathbb{N}$, μ^δ_m is absolutely continuous with respect to Lebesgue measure and its density $\rho^\delta_m \doteq \frac{d\mu^\delta_m}{dx}$ lies in $C^2_c(\mathbb{R}^d)$, $\mu^\delta_m \Rightarrow \mu^\delta$ as $m \to \infty$ and $I(\mu^\delta_m) \to I(\mu^\delta)$ as $m \to \infty$. For each $\delta > 0$, choose $m_\delta \in \mathbb{N}$ sufficiently large so that

$$|I(\mu^\delta) - I(\mu^\delta_{m_\delta})| < \delta, \tag{B.7}$$

and $d_w(\mu^\delta, \mu^\delta_{m_\delta}) < \delta$, where $d_w(\cdot, \cdot)$ is a metric on $\mathcal{P}(\mathbb{R}^d)$ compatible with the topology of weak convergence. For each $n \in \mathbb{N}$, set $\delta_n \doteq \frac{1}{n}$ and define

$$\mu_n \doteq (1 - \delta_n)\mu^{\delta_n}_{m_{\delta_n}} + \delta_n \mu^*. \tag{B.8}$$

Then μ_n admits a strictly positive continuous density with respect to Lebesgue measure and

$$d_w(\mu_n, \mu) \le d_w(\mu_n, \mu^{\delta_n}_{m_{\delta_n}}) + d_w(\mu^{\delta_n}_{m_{\delta_n}}, \mu^{\delta_n}) + d_w(\mu^{\delta_n}, \mu).$$

By (B.6)–(B.8), each term on the right converges to zero as $n \to \infty$, and so $\mu_n \Rightarrow \mu$ as $n \to \infty$. By (B.6)–(B.8) and the convexity of the rate function,

$$I(\mu_n) \le I(\mu) + \delta_n(1 + 2I(\mu^*)).$$

This, along with the fact thats $I(\mu^*) < \infty$, $\mu_n \Rightarrow \mu$ as $n \to \infty$ and the lower semicontinuity of the rate function shown in Lemma 5.2, implies that $I(\mu_n) \to I(\mu)$ as $n \to \infty$. ■

References

[1] V.I. Bogachev, N. Krylov, M. Röckner, Regularity of invariant measures: the case of non-constant diffusion part, J. Funct. Anal. 138 (1996) 223–242.
[2] M. Boué, P. Dupuis, A variational representation for certain functions of Brownian motion, Ann. Probab. 26 (4) (1998) 1641–1659.
[3] A. Budhiraja, P. Dupuis, Large deviations for the empirical measures of reflecting Brownian motion and related constrained processes in $\mathbb{R}_+$, Electron. J. Probab. 8 (16) (2003) 1–46.
[4] A. Budhiraja, P. Dupuis, M. Fischer, Large deviation properties of weakly interacting processes via weak convergence methods, Ann. Probab. 40 (1) (2012) 74–102.
[5] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, I, Comm. Pure Appl. Math. 28 (1975) 1–47.
[6] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, II, Comm. Pure Appl. Math. 28 (1975) 279–301.
[7] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, III, Comm. Pure Appl. Math. 29 (1976) 389–461.
[8] P. Dupuis, R.S. Ellis, A Weak Convergence Approach to Large Deviations, Wiley, New York, 1997.
[9] P. Dupuis, Y. Liu, On the large deviation rate function for the empirical measures of reversible jump Markov processes, Ann. Probab. 43 (3) (2015) 1121–1156.
[10] J. Feng, T.G. Kurtz, Large Deviations for Stochastic Processes, American Mathematical Society, Providence, 2006.
[11] W.H. Fleming, S.-J. Sheu, H.M. Soner, A remark on the large deviations of an ergodic Markov process, Stochastics 22 (1987) 187–199.
[12] J. Gärtner, On large deviations from the invariant measure, Theory Probab. Appl. 22 (1) (1977) 24–40.
[13] D. Gilbarg, N.S. Trudinger, Elliptic Partial Differential Equations of Second Order, second ed., Springer-Verlag, Berlin Heidelberg, 2001.
[14] R. Khasminskii, Stochastic Stability of Differential Equations, Springer, Heidelberg, 2012.
[15] R. Pinsky, A classification of diffusion processes with boundaries by their invariant measures, Ann. Probab. 13 (3) (1985) 693–697.
[16] R. Pinsky, On evaluating the Donsker-Varadhan I-function, Ann. Probab. 13 (2) (1985) 342–362.
[17] R. Pinsky, The I-function for diffusion processes with boundaries, Ann. Appl. Probab. 13 (3) (1985) 676–692.
[18] L. Rey-Bellet, K. Spiliopoulos, Irreversible Langevin samplers and variance reduction: a large deviations approach, Nonlinearity 28 (2015) 2081–2103.

Please cite this article in press as: P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods, Stochastic Processes and their Applications (2017), https://doi.org/10.1016/j.spa.2017.09.020.