Probab. Theory Relat. Fields 116, 125–149 (2000)

Michelle Boué · Paul Dupuis · Richard S. Ellis

Large deviations for small noise diffusions with discontinuous statistics

Received: 26 May 1998 / Revised version: 24 February 1999

Abstract. This paper proves the large deviation principle for a class of non-degenerate small noise diffusions with discontinuous drift and with state-dependent diffusion matrix. The proof is based on a variational representation for functionals of strong solutions of stochastic differential equations and on weak convergence methods.

1. Introduction

Consider the d-dimensional process $X^\varepsilon \doteq \{X^\varepsilon(t),\ 0 \le t \le 1\}$ satisfying the stochastic differential equation (SDE)

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \varepsilon^{1/2}\sigma(X^\varepsilon(t))dW(t), \quad X^\varepsilon(0) = x_0 \ , \tag{1.1}$$

where $\varepsilon > 0$ and W is a d-dimensional Brownian motion. The large deviation principle for the family $\{X^\varepsilon,\ \varepsilon > 0\}$ is a well known result under the condition that the drift vector $b(x)$ and the dispersion matrix $\sigma(x)$ are smooth functions of x [8]. However, this condition may be too restrictive for some applications, where processes which violate this smooth dependence arise naturally. We refer to these processes as having "discontinuous statistics." An example of an application where processes with discontinuous statistics arise is the modeling of communication channels incorporating a "hard limiter" in a phase locked loop, which is a form of a suboptimal nonlinear filter [12]. These communication channels can be modeled

M. Boué: Instituto Tecnologico Autonomo de Mexico, Departamento de Matematicas, Rio Hondo #1, Tizapan, San Angel, Mexico, D.F. 01000 Mexico.
e-mail: michelle@gauss.rhon.itam.mx

P. Dupuis*: Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. e-mail: dupuis@cfm.brown.edu

R.S. Ellis**: Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003, USA. e-mail: rsellis@math.umass.edu

*Research supported in part by NSF Grant DMS-9704426.

**Research supported in part by NSF Grant DMS-9700852.

Mathematics Subject Classification (1991): Primary 60F10; Secondary 60J60, 60H10, 93E99

Key words and phrases: Large deviations – Small noise diffusions – Discontinuous statistics

by a diffusion with continuous dispersion matrix but with drift that changes discontinuously as x crosses a smooth boundary in $\mathbb{R}^d$. In this paper we study the large deviation principle for precisely this kind of process.

Several previous papers have studied large deviations for diffusions with discontinuous statistics. Using continuous mapping techniques, Korostelev and Leonov [10, 11] obtained a large deviation principle for a restricted class of two-dimensional diffusions satisfying certain stability conditions. More recently, Chiang and Sheu [4] considered d-dimensional diffusions with drift that is continuous except across the $(d-1)$-dimensional hyperplane $\partial \doteq \{x \in \mathbb{R}^d : x_1 = 0\}$, where the subscript 1 denotes the 1-component of the vector. Their results assume that the dispersion matrix σ is the d-dimensional identity matrix. In this paper we extend these results significantly by allowing the dispersion matrix to depend on $x \in \mathbb{R}^d$. Although we still assume that a discontinuity occurs along the dimensional hyperplane ∂, our assumption is that (1.1) has a unique strong solution. This is certainly the case when $\sigma(\cdot)$ is the identity matrix or when it is a nondegenerate constant matrix. More general conditions for the existence and uniqueness of strong solutions can be found in [13].

A standard method for treating large deviations for diffusions is based on discretizing time. This is problematic in the context of discontinuous statistics since it is difficult to approximate the continuous-time process accurately by a discrete time analogue in a neighborhood of the discontinuity. The weak convergence approach to large deviations developed in [5] enables us to bypass this time discretization step. Although the methodology was developed there in the context of discrete-time processes, the present paper demonstrates the versatility of the approach by extending its application to the continuous-time setting. In particular, our results are a continuous-time analogue of those found in Chapter 7 of [5] concerning a random walk model with discontinuous statistics. In fact, the rate function appearing in our main theorem has the same form as the rate function appearing in Theorem 7.2.3 in [5]. The similarity is also present in the proofs of several of the preliminary results. Therefore, all those proofs which can be carried out as obvious extensions of their discrete-time counterparts will be omitted. It is important to remark that despite the similarities, the extension presented here is significant and far from immediate, and its proof requires a number of new ideas. The proof relies on general properties of SDE's and on a representation formula for functionals of strong solutions of SDE's [3].

The main result in this paper, Theorem 2.3, states the Laplace principle for the family $\{X^\varepsilon, \varepsilon > 0\}$. The term Laplace principle refers to the asymptotic analysis of normalized logarithms of expectations involving continuous functions; a precise definition is given in Section 2. Because the Laplace principle is equivalent to a large deviation principle with the same rate function, the large deviation principle for the family $\{X^\varepsilon, \varepsilon > 0\}$ is a direct consequence of Theorem 2.3.

The paper is organized as follows. The first part of Section 2 introduces the family of diffusions with discontinuous statistics considered throughout the paper. After the necessary assumptions are given, the Laplace principle for this family is stated in Theorem 2.3. In Section 3 we state a representation formula for functionals of strong solutions of SDE's which will be needed in the proof of Theorem 2.3. This section also includes general compactness and convergence results that will be used in the following section. Finally, Section 4 is devoted to the proof of the Laplace principle. The proof is split into the upper and lower bounds, each corresponding to one subsection.

A more ambitious problem than the one analyzed in the present paper concerns diffusion processes whose drifts have discontinuities along an arbitrary number of intersecting smooth $(d-1)$-dimensional manifolds. The difficulties arising in the analysis of such processes are discussed in Section 7.1 of [5].

2. Statement of the main theorem

We work with the canonical probability space $(\Omega, \mathscr{F}, P)$, where $\Omega \doteq \mathscr{C}([0,1]:\mathbb{R}^d)$ and P is d-dimensional Wiener measure. The space $\mathscr{C}([0,1]:\mathbb{R}^d)$ is endowed with the supnorm metric. For every $0 \le t \le 1$, let $W(t,\omega) \doteq \omega(t)$ and define the augmented filtration $\{\mathscr{F}_t\}$ by

$$\mathscr{F}_t \doteq \sigma\left(\mathscr{F}_t^W \cup \mathscr{N}\right), \qquad 0 \le t \le 1,$$

where $\{\mathscr{F}_t^W\} \doteq \{\sigma(W(s); 0 \le s \le t)\}$ and $\mathscr{N}$ is the collection of P-null sets. Then the process $W \doteq \{W(t), \mathscr{F}_t, 0 \le t \le 1\}$ is a d-dimensional Brownian motion.

In preparation for our main theorem, we introduce the concept of a Laplace principle. By definition, a rate function on a Polish space maps the Polish space into $[0,\infty]$ and has compact level sets.

Definition 2.1. *Let $\{Y^\epsilon, \varepsilon > 0\}$ be a family of random variables taking values in a Polish space $\mathscr{Y}$ and let I be a rate function on $\mathscr{Y}$. We say that $\{Y^\epsilon\}$ satisfies a* **Laplace principle** *with rate function I if for every bounded continuous function h mapping $\mathscr{Y}$ into $\mathbb{R}$*

$$\lim_{\epsilon\to 0} \epsilon \log E\left\{\exp\left[-\frac{h(Y^\epsilon)}{\epsilon}\right]\right\} = -\inf_{y\in\mathscr{Y}}\{h(y)+I(y)\}.$$

A Laplace principle is equivalent to a large deviation principle with the same rate function (see Theorems 2.2.1 and 2.2.3 in [5] for a proof). Thus, instead of proving a large deviation principle for diffusions with discontinuous statistics, we focus on the proof of the Laplace principle for the same family. This is the content of Theorem 2.3.

Consider the diffusion process X^ε which solves (1.1). The dispersion matrix σ is a $d \times d$ matrix of Borel measurable functions mapping $\mathbb{R}^d$ into $\mathbb{R}$. The drift b is assumed to be continuous except across the $(d-1)$-dimensional hyperplane

$$\partial \doteq \left\{x \in \mathbb{R}^d : x_1 = 0\right\}.$$

More precisely, given continuous functions $b^{(1)}$ and $b^{(2)}$ mapping $\mathbb{R}^d$ into $\mathbb{R}^d$, b is defined by

$$b(x) \doteq \begin{cases} b^{(1)}(x) & \text{if } x \in \Lambda^{(1)}, \\ b^{(2)}(x) & \text{if } x \in \Lambda^{(2)}, \end{cases}$$

where

$$\Lambda^{(1)} \doteq \left\{ x \in \mathbb{R}^d : x_1 \le 0 \right\} \text{ and } \Lambda^{(2)} \doteq \left\{ x \in \mathbb{R}^d : x_1 > 0 \right\}.$$

The inclusion of the hyperplane ∂ with the open left halfspace is arbitrary. The Laplace principle in Theorem 2.3 holds as stated if ∂ is included with the open right halfspace. We note that the more general problem where ∂ is replaced by a smooth $(d-1)$-dimensional manifold can be reduced to the one presented here by means of standard localization techniques [1].

Theorem 2.3 states the Laplace principle for the family $\{X^\epsilon, \varepsilon > 0\}$ under the following condition.

Condition 2.2. (a) *$b^{(1)}, b^{(2)}$ and σ are continuous and are bounded by a constant B_1.*
(b) *σ is uniformly nondegenerate, i.e. $\sigma(\cdot)\sigma^T(\cdot) \ge cI$ for $c > 0$.*
(c) *The SDE* (1.1) *has a unique strong solution.*

Remarks.
(i) Although in part (a) of Condition 2.2 we assume that both $b^{(1)}(x)$ and $b^{(2)}(x)$ are continuous functions of $x \in \mathbb{R}^d$, in general $b(x)$ is not continuous at ∂.
(ii) For simplicity we have assumed that $\sigma(x)$ is continuous on $\mathbb{R}^d$ (the sufficient conditions for the existence of a unique strong solution presented in [13] require this condition). However, as long as (1.1) has a unique strong solution, the result of Theorem 2.3 continues to hold if $\sigma(x)$ is also allowed to be discontinuous along the hyperplane ∂.
(iii) Part (b) of Condition 2.2 is needed for the proof of the Laplace principal lower bound. More comments on this assumption will be given below. □

For $i = 1, 2$ and x and β in $\mathbb{R}^d$, let $L^{(i)}(x, \beta)$ be the rate function associated via Cramér's Theorem with a Gaussian probability measure on $\mathbb{R}^d$ with mean vector $b^{(i)}(x)$ and covariance matrix $a(x) \doteq \sigma(x)\sigma^T(x)$. That is,

$$L^{(i)}(x, \beta) \doteq \sup_{\alpha \in \mathbb{R}^d} \left\{ \langle \alpha, \beta - b^{(i)}(x) \rangle - \tfrac{1}{2} \langle \alpha, a(x)\alpha \rangle \right\},$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{R}^d$. If $\beta - b^{(i)}(x)$ lies in the range of $a(x)$, then $L^{(i)}(x, \beta)$ is well defined by

$$L^{(i)}(x, \beta) = \tfrac{1}{2} \langle \alpha, \beta - b^{(i)}(x) \rangle,$$

where $\alpha \in \mathbb{R}^d$ is any vector satisfying $a(x)\alpha = \beta - b^{(i)}(x)$. Otherwise, $L^{(i)}(x, \beta)$ equals ∞. The effective domain of $L^{(i)}(x, \beta)$ is thus given by

$$\operatorname{dom} L^{(i)}(x, \beta) = \left\{ \beta \in \mathbb{R}^d : \beta - b^{(i)}(x) = a(x)u \text{ for some } u \in \mathbb{R}^d \right\}.$$

Therefore, part (b) of Condition 2.2 implies that $\operatorname{dom} L^{(i)}(x, \cdot) = \mathbb{R}^d$ for every $x \in \mathbb{R}^d$. It follows trivially that these sets are independent of $x \in \mathbb{R}^d$, and of $i = 1, 2$, that $0 \in \operatorname{ri}(\operatorname{dom} L^{(i)}(x, \cdot))$ and that $\operatorname{ri}(\operatorname{dom} L^{(i)}(x, \cdot))$ is not a subset of ∂. These properties are needed for the proof of the Laplace principle lower bound.

Remark. We can in fact replace part (b) of Condition 2.2 with the weaker assumption that the sets ri(dom $L^{(i)}(x,\cdot)$) are independent of $x \in \mathbb{R}^d$ and of $i = 1, 2$, that $0 \in$ ri(dom $L^{(i)}(x,\cdot)$), and that ri(dom $L^{(i)}(x,\cdot)$) is not a subset of ∂. The proofs that we present here can be easily generalized to cover the cases when this weaker assumption is satisfied. ◻

Before we turn to the definition of the rate function, we indicate the form of $L^{(i)}(x,\beta)$ that will be used in the proof of the Laplace principle. Part (b) of Condition 2.2 enables us to write

$$L^{(i)}(x,\beta) = \tfrac{1}{2}\|v\|^2 \quad \text{with} \quad v \doteq \sigma^{-1}(x)(\beta - b^{(i)}(x)). \tag{2.2}$$

In order to specify the rate function for the family $\{X^\varepsilon, \varepsilon > 0\}$, for x and β in $\mathbb{R}^d$ we define

$$L^{(0)}(x,\beta) \doteq \inf\left\{\rho^{(1)}L^{(1)}(x,\beta^{(1)}) + \rho^{(2)}L^{(2)}(x,\beta^{(2)})\right\}. \tag{2.3}$$

The infimum is taken over all $\rho^{(1)} \in \mathbb{R}$, $\rho^{(2)} \in \mathbb{R}$, $\beta^{(1)} \in \mathbb{R}^d$, and $\beta^{(2)} \in \mathbb{R}^d$ satisfying

$$\rho^{(1)} \geq 0,\ \rho^{(2)} \geq 0,\ \rho^{(1)} + \rho^{(2)} = 1, \tag{2.4}$$

$$(\beta^{(1)})_1 \geq 0,\ (\beta^{(2)})_1 \leq 0, \tag{2.5}$$

$$\rho^{(1)}\beta^{(1)} + \rho^{(2)}\beta^{(2)} = \beta. \tag{2.6}$$

We then define for x and β in $\mathbb{R}^d$

$$\tilde{L}(x,\beta) \doteq \begin{cases} L^{(1)}(x,\beta) & \text{if } x_1 < 0 \\ L^{(0)}(x,\beta) & \text{if } x_1 = 0 \\ L^{(2)}(x,\beta) & \text{if } x_1 > 0. \end{cases} \tag{2.7}$$

The form of the rate function in Theorem 2.3 is the same as the rate function appearing in the Laplace principle for the random walk model with discontinuous statistics studied in Chapter 7 of [5], and it may be interpreted as follows. Our proof of the Laplace principle for the solutions to (1.1) will be based on representations for exponential integrals in terms of corresponding controlled diffusions (see Theorem 3.1). These controlled diffusions will play a role similar to that of the processes obtained via change of measure in the proof of the lower bound in Cramér's Theorem. The form of the rate function indicates costs and constraints that arise naturally when the representation is used to estimate certain events. For example, the constraints in (2.5) express a necessary and sufficient condition for the stability of the one-component of the controlled processes. In order to see the role played by this equation, let us consider, for example, the problem of estimating the probability that $\sup_{t\in[0,1]}\|X^\varepsilon(t) - \beta t\| < \delta$, where $\delta > 0$ and β is a vector in $\mathbb{R}^d$ satisfying $\beta_1 = 0$. When using the representation formula, one must consider controls that shift the drifts of the processes in the halfspaces $\Lambda^{(1)}$ and $\Lambda^{(2)}$ to $\beta^{(1)}$

and $\beta^{(2)}$, respectively, in such a way that the event of interest has a probability of order 1 as $\varepsilon \to 0$. Unless the constraints in formula (2.5) are satisfied, the probability of the event is of order 0 as $\varepsilon \to 0$. It should be stressed that these constraints do not express a stability condition on the original processes but rather on the controlled processes appearing in the representation formula. When the constraints are satisfied, $\rho^{(1)}$ and $\rho^{(2)}$ may be interpreted as the limits of the fractions of time that the controlled processes spend in $\Lambda^{(1)}$ and $\Lambda^{(2)}$, respectively; by equation (2.6) the probability of interest is indeed of order 1 as $\varepsilon \to 0$.

Theorem 2.3. *We assume Condition* 2.2. *For any $x_0 \in \mathbb{R}^d$ and $\varepsilon > 0$, let $X^\varepsilon \doteq \{X^\varepsilon(t),\ 0 \le t \le 1\}$ be the unique strong solution to the SDE*

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \varepsilon^{1/2}\sigma(X^\varepsilon(t))dW(t), \quad X^\varepsilon(0) = x_0. \tag{2.8}$$

For absolutely continuous functions $\varphi \in \mathscr{C}([0,1] : \mathbb{R}^d)$ satisfying $\varphi(0) = x_0$, we define

$$I_{x_0}(\varphi) \doteq \int_0^1 \tilde{L}(\varphi(t), \dot{\varphi}(t))\, dt,$$

where $\tilde{L}$ is defined in equations (2.3)–(2.7). *For all other $\varphi \in \mathscr{C}([0,1] : \mathbb{R}^d)$, we set $I_{x_0}(\varphi) \doteq \infty$. Then the family $\{X^\varepsilon,\ \varepsilon > 0\}$ satisfies the Laplace principle on $\mathscr{C}([0,1] : \mathbb{R}^d)$ with rate function $I_{x_0}(\cdot)$. In fact, the Laplace principle holds uniformly on compacts, in the sense that for all compact subsets K of $\mathbb{R}^d$, the Laplace principle holds uniformly for all $x_0 \in K$.*

We will give the proof of the nonuniform Laplace principle in Section 4. The proof of the uniform version uses the same arguments but with more cumbersome notation. The proof that I_{x_0} has compact level sets will be omitted since under Condition 2.2 it is identical to that of Proposition 7.6.1 in [5].

3. Preliminary results

Let P_{x_0} denote probability conditioned on $X^\varepsilon(0) = x_0$, and let E_{x_0} denote the corresponding expectation. The proof of Theorem 2.3 requires that we analyze the asymptotic behavior of $W^\varepsilon(x_0) \doteq -\varepsilon \log E_{x_0}\{\exp[-h(X^\varepsilon)/\varepsilon]\}$. A basic step in the weak convergence approach used for this analysis is the representation of $W^\varepsilon(x)$ in terms of the minimal cost function of an associated stochastic control problem. The purpose of this section is to introduce this representation and to study compactness and limit properties of certain families of controls and controlled processes arising in the representation.

We start by stating the representation formula. For a heuristic derivation of the form of the representation we refer the reader to Section 4.6 in [5]. A proof is given in [3].

Theorem 3.1. *Given $\varepsilon > 0$, let X^ε be the diffusion process that is the unique strong solution to* (2.8). *Then for any bounded Borel-measurable function h mapping $\mathscr{C}([0,1] : \mathbb{R}^d)$ into $\mathbb{R}$ the following representation holds*:

$$W^{\varepsilon}(x_0) = \inf_{v \in \mathscr{A}} E_{x_0} \left\{ \frac{1}{2} \int_0^1 \|v(t)\|^2 dt + h(X^{v,\varepsilon}) \right\},$$

where $\mathscr{A}$ is the set of all $\mathscr{F}_t$-progressively measurable d-dimensional processes $v \doteq \{v(t),\ 0 \le t \le 1\}$ satisfying

$$E\left[\int_0^1 \|v(t)\|^2 dt\right] < \infty$$

and $X^{v,\varepsilon} \doteq \{X^{v,\varepsilon}(t),\ 0 \le t \le 1\}$ is the unique strong solution to

$$dX^{v,\varepsilon}(t) = b(X^{v,\varepsilon}(t))dt + \sigma(X^{v,\varepsilon}(t))v(t)dt + \varepsilon^{1/2}\sigma(X^{v,\varepsilon}(t))dW(t),$$
$$X^{v,\varepsilon}(0) = x_0.$$

We refer to $X^{v,\varepsilon}$ as the controlled diffusion associated with the control v.

Let $\{v^{\varepsilon}, \varepsilon > 0\}$ be a family of controls in $\mathscr{A}$. Define $\bar{X}^{\varepsilon} \doteq X^{v^{\varepsilon},\varepsilon}$. Thus for every $\varepsilon > 0$ and $t \in [0, 1]$, the equation

$$\bar{X}^{\varepsilon}(t) = x_0 + \int_0^t \left[b(\bar{X}^{\varepsilon}(s)) + \sigma(\bar{X}^{\varepsilon}(s))v^{\varepsilon}(s)\right] ds + \varepsilon^{1/2} \int_0^t \sigma(\bar{X}^{\varepsilon}(s))dW(s) \tag{3.9}$$

holds w.p.1. The rest of this section is devoted to the study of the asymptotic properties of $\{v^{\varepsilon}, \varepsilon > 0\}$ and of $\{\bar{X}^{\varepsilon}, \varepsilon > 0\}$ under the assumption that the controls satisfy the following condition. The condition will be automatically satisfied by the controls that arise in the proof of the Laplace principle.

Condition 3.2. $\Delta \doteq \sup_{\varepsilon > 0} E_{x_0} \left\{ \int_0^1 \|v^{\varepsilon}(t)\|^2 dt \right\} < \infty.$

When studying the asymptotic properties of the families of controls and of controlled processes, the discontinuity of the statistics gives rise to some difficulties. First, the presence of the discontinuity implies that the family of controls may not converge in any of the usual senses. This was not the case in any of the examples treated in [5], since the admissible controls always took values in the space of probability measures on some Polish space and thus limits were guaranteed to exist in the weak sense. In order to exploit the weak convergence ideas, we represent each control v^{ε} as a measure on the Borel sets of $\mathbb{R}^d$. Under Condition 3.2, we will show below that there is a subsequence of these measures which converges weakly.

For Borel subsets A of $\mathbb{R}^d$ and B of $[0, 1]$ let

$$\nu^{\varepsilon}(A|t) \doteq 1_A(v^{\varepsilon}(t)), \tag{3.10}$$

$$\nu^{\varepsilon}(A \times B) \doteq \int_B 1_A(v^{\varepsilon}(t))dt = \int_B \nu^{\varepsilon}(A|t)dt. \tag{3.11}$$

The quantities $\nu^\varepsilon(\cdot|t)$ and $\nu^\varepsilon(\cdot)$ take values in the spaces $\mathscr{P}(\mathbb{R}^d)$ and $\mathscr{P}(\mathbb{R}^d \times [0,1])$ of probability measures on $\mathbb{R}^d$ and on $\mathbb{R}^d \times [0,1]$ respectively. The quantity $\nu^\varepsilon(A \times B)$ represents the total integrated time over the Borel set $B \subset [0,1]$ that the control v^ε takes values in the Borel set $A \subset \mathbb{R}^d$. Using these measures, (3.9) can be rewritten as

$$\begin{aligned}\bar{X}^\varepsilon(t) = {} & x_0 + \int_0^t \int_{\mathbb{R}^d} \Big(b(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\nu^\varepsilon(dy|s)\,ds \\ & + \varepsilon^{1/2}\int_0^t \sigma(\bar{X}^\varepsilon(s))dW(s) \\ = {} & x_0 + \int_{\mathbb{R}^d\times[0,t]} \Big(b(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\nu^\varepsilon(dy \times ds) \\ & + \varepsilon^{1/2}\int_0^t \sigma(\bar{X}^\varepsilon(s))dW(s).\end{aligned}$$

A second consequence of the discontinuity of the statistics is that care must be exercised when analyzing the asymptotic fraction of time that the controlled processes $\bar{X}^\varepsilon$ spend in each of the halfspaces $\Lambda^{(1)}$ and $\Lambda^{(2)}$. For this we need to introduce additional measures. Define

$$\nu^{(1),\varepsilon}(A \times B) \doteq \int_B 1_{\{s\in[0,1]:(\bar{X}^\varepsilon(s))_1\le 0\}}(t)\,\nu^\varepsilon(A|t)\,dt, \tag{3.12}$$

$$\nu^{(2),\varepsilon}(A \times B) \doteq \int_B 1_{\{s\in[0,1]:(\bar{X}^\varepsilon(s))_1> 0\}}(t)\,\nu^\varepsilon(A|t)\,dt. \tag{3.13}$$

These quantities take values in the space $\mathscr{M}(\mathbb{R}^d \times [0,1])$ of subprobability measures on $\mathbb{R}^d \times [0,1]$. In terms of these measures we can further rewrite

$$\begin{aligned}\bar{X}^\varepsilon(t) = x_0 & + \int_{\mathbb{R}^d\times[0,t]} \Big(b^{(1)}(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\,\nu^{(1),\varepsilon}(dy \times ds) \\ & + \int_{\mathbb{R}^d\times[0,t]} \Big(b^{(2)}(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\,\nu^{(2),\varepsilon}(dy \times ds) \\ & + \varepsilon^{1/2}\int_0^t \sigma(\bar{X}^\varepsilon(s))dW(s).\end{aligned} \tag{3.14}$$

Finally, define $\gamma^{(1),\varepsilon}$ and $\gamma^{(2),\varepsilon}$ to be the respective second marginals of $\nu^{(1),\varepsilon}$ and $\nu^{(2),\varepsilon}$. Thus

$$\gamma^{(1),\varepsilon}(B) \doteq \nu^{(1),\varepsilon}(\mathbb{R}^d \times B) = \int_B 1_{\{s\in[0,1]:(\bar{X}^\varepsilon(s))_1\le 0\}}(t)\,dt \tag{3.15}$$

and

$$\gamma^{(2),\varepsilon}(B) \doteq \nu^{(2),\varepsilon}(\mathbb{R}^d \times B) = \int_B 1_{\{s\in[0,1]:(\bar{X}^\varepsilon(s))_1> 0\}}(t)\,dt. \tag{3.16}$$

These quantities are the Lebesgue measure of the sets of times $t \in B$ at which $\bar{X}^\varepsilon(t)$ lies in the respective halfspaces; $\gamma^{(1),\varepsilon}$ and $\gamma^{(2),\varepsilon}$ take values in $\mathscr{M}([0,1])$. For $i = 1, 2$

$$\nu^{(i),\varepsilon}(A \times B) = \int_B \nu^\varepsilon(A|t)\, \gamma^{(i),\varepsilon}(dt).$$

The following proposition gives the tightness of the family $\{(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}), \varepsilon > 0\}$ as well as a uniform integrability property of $\{\nu^\varepsilon, \varepsilon > 0\}$, $\{\nu^{(1),\varepsilon}, \varepsilon > 0\}$, and $\{\nu^{(2),\varepsilon}, \varepsilon > 0\}$.

Proposition 3.3. *Given $x_0 \in \mathbb{R}^d$, consider any family $\{v^\varepsilon, \varepsilon > 0\}$ of controls in $\mathscr{A}$ satisfying Condition* 3.2*. The following conclusions hold.*

(a) *The family $\{(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}), \varepsilon > 0\}$ is tight.*

(b) *The families $\{\nu^\varepsilon, \varepsilon > 0\}$ and $\{\nu^{(i),\varepsilon}, \varepsilon > 0\}$, $i = 1, 2$, have the uniform integrability properties*

$$\lim_{C\to\infty} \sup_{\varepsilon>0} E_{x_0}\left\{\int_{\{y\in\mathbb{R}^d:\|y\|>C\}\times[0,1]} \|y\|\, \nu^\varepsilon(dy \times dt)\right\} = 0 \tag{3.17}$$

and

$$\lim_{C\to\infty} \sup_{\varepsilon>0} E_{x_0}\left\{\int_{\{y\in\mathbb{R}^d:\|y\|>C\}\times[0,1]} \|y\|\, \nu^{(i),\varepsilon}(dy \times dt)\right\} = 0. \tag{3.18}$$

Proof. That (3.17) implies the tightness of $\{\nu^\varepsilon, \varepsilon > 0\}$ is standard. We thus focus on the proof of (3.17). For $C > 0$, Condition 3.2 gives

$$\begin{aligned}
&\sup_{\varepsilon>0} E_{x_0}\left\{\int_{\{y\in\mathbb{R}^d:\|y\|>C\}\times[0,1]} \|y\|\, \nu^\varepsilon(dy \times dt)\right\} \\
&\quad = \sup_{\varepsilon>0} E_{x_0}\left\{\int_{\mathbb{R}^d\times[0,1]} \|y\|\, 1_{\{z\in\mathbb{R}^d:\|z\|>C\}}(y)\, \nu^\varepsilon(dy \times dt)\right\} \\
&\quad = \sup_{\varepsilon>0} E_{x_0}\left\{\int_0^1 \int_{\mathbb{R}^d} \|y\|\, 1_{\{z\in\mathbb{R}^d:\|z\|>C\}}(y) \nu^\varepsilon(dy|t) dt\right\} \\
&\quad = \sup_{\varepsilon>0} E_{x_0}\left\{\int_{\{t\in[0,1]:\|v^\varepsilon(t)\|>C\}} \|v^\varepsilon(t)\| dt\right\} \\
&\quad \le \sup_{\varepsilon>0} E_{x_0}\left\{\frac{1}{C}\int_{\{t\in[0,1]:\|v^\varepsilon(t)\|>C\}} \|v^\varepsilon(t)\|^2 dt\right\} \le \frac{\Delta}{C}.
\end{aligned}$$

Letting $C \to \infty$ yields (3.17).

Since $\nu^\varepsilon = \nu^{(1),\varepsilon} + \nu^{(2),\varepsilon}$, it follows that for $i = 1, 2$ and any nonnegative measurable function g mapping $\mathbb{R}^d \times [0, 1]$ into $[0, \infty]$

$$0 \le \int_{\mathbb{R}^d\times[0,1]} g(y,t)\, \nu^{(i),\varepsilon}(dy \times dt) \le \int_{\mathbb{R}^d\times[0,1]} g(y,t)\, \nu^\varepsilon(dy \times dt).$$

As a consequence, (3.17) implies (3.18), which gives the tightness of the individual families $\{\nu^{(1),\varepsilon}, \varepsilon > 0\}$ and $\{\nu^{(2),\varepsilon}, \varepsilon > 0\}$. The random measures $\gamma^{(1),\varepsilon}$ and $\gamma^{(2),\varepsilon}$ take values in $\mathscr{M}([0,1])$, which is compact since $[0,1]$ is compact. We thus conclude that $\{(\nu^{\varepsilon}, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}), \varepsilon > 0\}$ is tight. □

Theorem 3.4 below analyzes the convergence properties of the control measures and of the associated controlled processes. Before stating it, we give decompositions of certain quantities arising in that theorem. We recall that if $(\mathscr{V}, \mathscr{A})$ is a measurable space and $\mathscr{Y}$ is a Polish space, a family $\{\tau(dy|x), x \in \mathscr{V}\}$ of probability (resp., subprobability) measures on $\mathscr{Y}$ is a stochastic kernel (resp., a substochastic kernel) on $\mathscr{Y}$ given $\mathscr{V}$ if for every Borel subset B of $\mathscr{Y}$ the function mapping $x \in \mathscr{V} \longmapsto \tau(B|x) \in [0,1]$ is measurable.

Let $(\nu, \nu^{(1)}, \nu^{(2)}, \gamma^{(1)}, \gamma^{(2)})$ denote the limits in distribution of convergent subsequences of $\{(\nu^{\varepsilon}, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}),\ \varepsilon > 0\}$. These limits can be defined on a probability space $(\Omega, \mathscr{F}, P)$ so that the following statements are valid. For $i = 1, 2$ and all Borel subsets A of $\mathbb{R}^d$ and B of $[0,1]$, P_{x_0}-a.s. for $\omega \in \Omega$

$$\nu(A \times B|\omega) = \int_B \nu(A|t,\omega)\, dt \tag{3.19}$$

for some stochastic kernel $\nu(dy|t,\omega)$;

$$\nu^{(i)}(A \times B|\omega) = \int_B \nu^{(i)}(A|t,\omega)\, dt \tag{3.20}$$

for some substochastic kernel $\nu^{(i)}(dy|t,\omega)$; and

$$\gamma^{(i)}(B|\omega) = \int_B \hat{\gamma}^{(i)}(t,\omega)\, dt \tag{3.21}$$

for some measurable function $\hat{\gamma}^{(i)}(t,\omega)$. In addition, we have with probability 1 and for a.e. $t \in [0,1]$

$$\nu^{(1)}(dy|t,\omega) + \nu^{(2)}(dy|t,\omega) = \nu(dy|t,\omega), \tag{3.22}$$

and for each $\omega \in \Omega$ and each $t \in [0,1]$ $\hat{\gamma}^{(i)}(t,\omega) = \nu^{(i)}(\mathbb{R}^d|t,\omega)$. In the sequel the displays (3.19), (3.20) and (3.21) will be summarized as

$$\nu(dy \times dt) = \nu(dy|t) \otimes dt,\ \ \nu^{(i)}(dy \times dt) = \nu^{(i)}(dy|t) \otimes dt,\ \text{ and}$$
$$\gamma^{(i)}(dt) = \hat{\gamma}^{(i)}(t)dt$$

respectively. Details for the derivation of these decompositions in an analogous situation can be found in Lemma 7.4.3 of [5]. We do remark, however, that the proof of these decompositions and the proof of Theorems 2.3 and 3.4 below make use of the Skorohod Representation Theorem [7, Theorem 1.8], which involves the introduction of a new probability space. We have retained the notation $(\Omega, \mathscr{F}, P)$ for this new space, and we will follow the same convention throughout the sequel.

The following theorem relates limit points of the family $\{(\nu^{\varepsilon}, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}), \varepsilon > 0\}$ with limits of the family $\{\bar{X}^{\varepsilon}, \varepsilon > 0\}$. In particular, it derives several key properties of the limiting quantities ν, $\nu^{(1)}$, $\nu^{(2)}$, $\gamma^{(1)}$, $\gamma^{(2)}$, and $\bar{X}$.

Theorem 3.4. *Assume Condition* 2.2. *Given* $x_0 \in \mathbb{R}^d$, *consider any family* $\{v^\varepsilon, \varepsilon > 0\}$ *of processes in* $\mathscr{A}$ *satisfying Condition* 3.2. *The following conclusions hold.*

(a) *Given any subsequence of* $\{(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}, \bar{X}^\varepsilon),\ \varepsilon > 0\}$, *there exist a subsubsequence, a probability space* $(\Omega, \mathscr{F}, P)$, *a stochastic kernel* ν *on* $\mathbb{R}^d \times [0,1]$ *given* Ω, *substochastic kernels* $\nu^{(1)}$ *and* $\nu^{(2)}$ *on* $\mathbb{R}^d \times [0,1]$ *given* Ω, *substochastic kernels* $\gamma^{(1)}$ *and* $\gamma^{(2)}$ *on* $[0,1]$ *given* Ω, *and a random variable* $\bar{X}$ *mapping* Ω *into* $\mathscr{C}([0,1] : \mathbb{R}^d)$ *such that the subsubsequence converges in distribution to* $(\nu, \nu^{(1)}, \nu^{(2)}, \gamma^{(1)}, \gamma^{(2)}, \bar{X})$. *The* (*sub*)*stochastic kernels have the decompositions given in* (3.19)–(3.22).

(b) *With probability* 1, *for every* $t \in [0,1]$

$$\begin{aligned}\bar{X}(t) = {} & x_0 + \int_{\mathbb{R}^d \times [0,t]} \left(b^{(1)}(\bar{X}(s)) + \sigma(\bar{X}(s))y\right) \nu^{(1)}(dy \times ds) \\ & + \int_{\mathbb{R}^d \times [0,t]} \left(b^{(2)}(\bar{X}(s)) + \sigma(\bar{X}(s))y\right) \nu^{(2)}(dy \times ds) \qquad (3.23) \\ = {} & x_0 + \int_0^t \int_{\mathbb{R}^d} \left(b^{(1)}(\bar{X}(s)) + \sigma(\bar{X}(s))y\right) \nu^{(1)}(dy|s)ds \\ & + \int_0^t \int_{\mathbb{R}^d} \left(b^{(2)}(\bar{X}(s)) + \sigma(\bar{X}(s))y\right) \nu^{(2)}(dy|s)ds, \qquad (3.24)\end{aligned}$$

and $\bar{X}(t)$ *is an absolutely continuous function of* $t \in [0,1]$. *Therefore, a.s. for* $t \in [0,1]$ *the derivative of* $\bar{X}(t)$ *is given by*

$$\begin{aligned}\dot{\bar{X}}(t) = {} & \int_{\mathbb{R}^d} \left(b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))y\right) \nu^{(1)}(dy|t) \\ & + \int_{\mathbb{R}^d} \left(b^{(2)}(\bar{X}(t)) + \sigma(\bar{X}(t))y\right) \nu^{(2)}(dy|t). \qquad (3.25)\end{aligned}$$

(c) *With probability* 1, *we have a.s.* (*with respect to Lebesgue measure*) *for* $t \in [0,1]$

$$\begin{aligned}&(\bar{X}(t))_1 < 0 \text{ implies } \hat{\gamma}^{(1)}(t) = \nu^{(1)}(\mathbb{R}^d|t) = 1 \text{ and } \hat{\gamma}^{(2)}(t) = \nu^{(2)}(\mathbb{R}^d|t) = 0, \\ &(\bar{X}(t))_1 > 0 \text{ implies } \hat{\gamma}^{(2)}(t) = \nu^{(2)}(\mathbb{R}^d|t) = 1 \text{ and } \hat{\gamma}^{(1)}(t) = \nu^{(1)}(\mathbb{R}^d|t) = 0, \qquad (3.26)\end{aligned}$$

and for any value of $(\bar{X}(t))_1$

$$\hat{\gamma}^{(1)}(t) + \hat{\gamma}^{(2)}(t) = \nu^{(1)}(\mathbb{R}^d|t) + \nu^{(2)}(\mathbb{R}^d|t) = 1. \qquad (3.27)$$

(d) *With probability* 1, *we have a.s. for* $t \in [0,1]$ *that whenever* $(\bar{X}(t))_1 = 0$

$$\left(\int_{\mathbb{R}^d} \left(b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))y\right) \nu^{(1)}(dy|t)\right)_1 \geq 0 \qquad (3.28)$$

and

$$\left(\int_{\mathbb{R}^d}\left(b^{(2)}(\bar{X}(t))+\sigma(\bar{X}(t))y\right)\nu^{(2)}(dy|t)\right)_1 \leq 0. \tag{3.29}$$

Proof. (a) Tightness of $\{\bar{X}^\varepsilon, \varepsilon > 0\}$ can be easily verified under Conditions 2.2 and 3.2. The convergence in distribution asserted in part (a) is a consequence of the tightness of this family together with part (a) of Proposition 3.3. The identification of the limiting quantities is given in (3.19)–(3.21) and part (b) of the present theorem,

(b) The tightness of the family $\{\bar{X}^\varepsilon, \varepsilon > 0\}$ implies that for any subsequence of $\varepsilon > 0$ there exist a subsubsequence and a random variable $\bar{X}$ taking values in $\mathscr{C}([0,1] : \mathbb{R}^d)$ such that $\bar{X}^\varepsilon \xrightarrow{\mathscr{D}} \bar{X}$. We invoke the Skorohod Representation Theorem, which allows us to assume that $\bar{X}^\varepsilon \to \bar{X}$ w.p.1. It remains to show that w.p.1 $\bar{X}(t)$ satisfies (3.23) and (3.24) for all $t \in [0,1]$.

We start by showing that for each $i = 1, 2$, $t \in [0,1]$, and bounded continuous function g mapping $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathbb{R}$

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d\times[0,t]} g(\bar{X}^\varepsilon(s), y)\,\nu^{(i),\varepsilon}(dy\times ds) = \int_{\mathbb{R}^d\times[0,t]} g(\bar{X}(s), y)\,\nu^{(i)}(dy\times ds) \tag{3.30}$$

w.p.1. In view of [5, Theorem A.3.10], all that we need to verify is that the set of points $(y, s) \in \mathbb{R}^d \times [0,1]$ such that

$$g(\bar{X}^\varepsilon(s^\varepsilon), y^\varepsilon)1_{[0,t]}(s^\varepsilon) \longrightarrow g(X(s), y)1_{[0,t]}(s)$$

fails to hold for some sequence $\{(y^\varepsilon, s^\varepsilon), \varepsilon > 0\}$ converging to (y, s) forms a set of $\nu^{(i)}-$measure zero. Since w.p.1 $\nu^{(i),\varepsilon} \Rightarrow \nu^{(i)}$ and $\bar{X}^\varepsilon$ converges uniformly to the continuous process X, that set is a subset of $\mathbb{R}^d \times \{t\}$. Since the second marginal of $\nu^{(i)}$ equals Lebesgue measure λ on $[0,1]$, w.p.1 $\nu^{(i)}(\mathbb{R}^d \times \{t\}) = \lambda(\{t\}) = 0$, which yields (3.30).

Since $b^{(i)}$ is bounded and continuous for each $i = 1, 2$, (3.30) immediately yields

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d\times[0,t]} b^{(i)}(\bar{X}^\varepsilon(s))\nu^{(i),\varepsilon}(dy\times ds) = \int_{\mathbb{R}^d\times[0,t]} b^{(i)}(\bar{X}(s))\nu^{(i)}(dy\times ds) \tag{3.31}$$

w.p.1. For $0 < C < \infty$ let $\varphi_C(y) \doteq y$ if $\|y\| \leq C$ and $\varphi_C(y) \doteq C\frac{y}{\|y\|}$ if $\|y\| > C$. Also, for $i = 1, 2$ define

$$\psi^{(i),\varepsilon}(C) \doteq E_{x_0}\left\{\int_{\{y\in\mathbb{R}^d:\|y\|>C\}\times[0,1]} \|y\|\,\nu^{(i),\varepsilon}(dy\times ds)\right\}$$

and

$$\psi^{(i)}(C) \doteq E_{x_0}\left\{\int_{\{y\in\mathbb{R}^d:\|y\|>C\}\times[0,1]} \|y\|\,\nu^{(i)}(dy\times ds)\right\}.$$

As is shown in Theorem 5.3.5 of [5], a version of Fatou's Lemma implies

$$\psi^{(i)}(C) \leq \sup_{\varepsilon>0} \psi^{(i),\varepsilon}(C). \tag{3.32}$$

For any $\xi > 0$, we use these definitions and Chebychev's inequality to write

$$\begin{aligned}
P_{x_0} & \left\{ \left\| \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}^\varepsilon(s))y\ \nu^{(i),\varepsilon}(dy\times ds) \right.\right.\\
&\left.\left. \quad - \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}(s))y\ \nu^{(i)}(dy\times ds) \right\| \geq \xi \right\} \\
&\leq \frac{1}{\xi} E_{x_0} \left\{ \left\| \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}^\varepsilon(s))y\ \nu^{(i),\varepsilon}(dy\times ds) \right.\right.\\
&\left.\left. \quad - \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}(s))y\ \nu^{(i)}(dy\times ds) \right\| \right\} \\
&\leq \frac{1}{\xi} E_{x_0} \left\{ \left\| \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}^\varepsilon(s))\varphi_C(y)\ \nu^{(i),\varepsilon}(dy\times ds) \right.\right.\\
&\left.\left. \quad - \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}(s))\varphi_C(y)\ \nu^{(i)}(dy\times ds) \right\| \right\} \\
&\quad + \frac{2B_1}{\xi} \left[\sup_{\varepsilon>0} \psi^{(i),\varepsilon}(C) + \psi^{(i)}(C) \right].
\end{aligned}$$

We now take $\varepsilon \to 0$ followed by $C \to \infty$. The penultimate line in the last display converges to zero because of (3.30) and the Lebesgue Dominated Convergence Theorem. Combining Proposition 3.3 with (3.32), we also get convergence of the last line to zero. Therefore, as $\varepsilon \to 0$ we have

$$\int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}^\varepsilon(s))y\ \nu^{(i),\varepsilon}(dy\times ds) \to \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}(s))y\ \nu^{(i)}(dy\times ds) \tag{3.33}$$

in probability. From (3.14) we have

$$\begin{aligned}
\lim_{\varepsilon\to 0} \bar{X}^\varepsilon(t) = x_0 + \lim_{\varepsilon\to 0} & \left\{ \varepsilon^{1/2} \int_0^t \sigma(\bar{X}^\varepsilon(s))dW(s) \right.\\
&+ \sum_{i=1,2} \left[\int_{\mathbb{R}^d\times[0,t]} b^{(i)}(\bar{X}^\varepsilon(s))\nu^{(i),\varepsilon}(dy\times ds) \right.\\
&\left.\left. + \int_{\mathbb{R}^d\times[0,t]} \sigma(\bar{X}^\varepsilon(s))y\nu^{(i),\varepsilon}(dy\times ds) \right] \right\}.
\end{aligned} \tag{3.34}$$

The left-hand side of (3.34) converges w.p.1 to $\bar{X}(t)$ because of the uniform convergence of $\bar{X}^\varepsilon$ to $\bar{X}$. Taking a further subsequence if necessary, (3.31), (3.33) and the w.p.1 convergence of the stochastic integral to zero imply that the right-hand side of (3.34) equals the right-hand side of (3.23), as we wanted to show. Formula (3.24) is now a consequence of the decomposition (3.22) and the bound

$$E_{x_0} \left\{ \int_{\mathbb{R}^d\times[0,1]} \|y\| \nu^{(i)}(dy\times dt) \right\} < \infty,$$

both valid for $i = 1, 2$. It follows that $\bar{X}$ is an absolutely continuous function of $t \in [0, 1]$, and the expression for $\dot{\bar{X}}(t)$ is immediate.

For the rest of the proof we again invoke the Skorohod Representation Theorem, which allows us to assume that the convergence asserted in parts (a) and (b) occurs w.p.1.

(c) This part is a relatively straightforward consequence of weak convergence and the definition of the various measures. The proof is omitted since it is analogous to the proof of part (c) of Theorem 7.4.4 in [5].

(d) Let us temporarily fix $m > 0$. The first step of the proof is to build an approximation to the function G mapping $\mathbb{R}$ into $\mathbb{R}$ defined by

$$G(z) \doteq \begin{cases} |z| & \text{if } |z| \le m \\ m & \text{if } |z| > m. \end{cases}$$

For each $\kappa > 0$ let $G_\kappa(z)$ be a twice continuously differentiable function of $z \in \mathbb{R}\backslash\{0\}$ such that

$$\begin{aligned} |G_\kappa(z)| &\le 2m && \text{for all } z \in \mathbb{R}, \\ G_\kappa(z) &= G(z) && \text{for } |z| \le m, \\ |d^2 G_\kappa(z)/dz^2| &< B_0 && \text{for } |z| > m/4, \end{aligned}$$

where $B_0 < \infty$ depends on $\kappa > 0$. Define

$$g_\kappa(z) \doteq \begin{cases} dG_\kappa(z)/dz & \text{if } z \ne 0 \\ -1 & \text{if } z = 0. \end{cases}$$

By suitably defining $G_\kappa(z)$, we can assume that for $|z| > m$ $g_\kappa(z) \to 0$ as $\kappa \to 0$ and that $g_\kappa(z)$ is uniformly bounded for $\kappa > 0$ and $z \in \mathbb{R}$. Although g_κ is not continuous on all of $\mathbb{R}$, the restrictions of g_κ to $(-\infty, 0]$ and to $(0, \infty)$ are bounded and continuous, and $g_\kappa(z)$ is Lipschitz continuous for $|z| \ge m/4$ with constant B_0.

For any two points x and y in $\mathbb{R}$, $4m \ge G_\kappa(x) - G_\kappa(y)$. Substituting $x = (\bar{X}^\varepsilon(1))_1$ and $y = (\bar{X}^\varepsilon(0))_1$ in this relation we get $4m \ge G_\kappa\left((\bar{X}^\varepsilon(1))_1\right) - G_\kappa\left((\bar{X}^\varepsilon(0))_1\right)$. Since G_κ can be written as the sum of a twice continuously differentiable function and a convex function, we can apply the generalized Itô's rule [9, Theorem 7.1] to the semimartingale $G_\kappa((\bar{X}^\varepsilon(t))_1)$ to obtain

$$\begin{aligned} 4m &\ge G_\kappa\left((\bar{X}^\varepsilon(1))_1\right) - G_\kappa\left((\bar{X}^\varepsilon(0))_1\right) \\ &= \int_0^1 g_\kappa\left((\bar{X}^\varepsilon(s))_1\right)\left(b(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))v^\varepsilon(s)\right)_1 ds \\ &\quad + \varepsilon^{1/2}\sum_{i=1}^d \int_0^1 g_\kappa\left((\bar{X}^\varepsilon(s))_1\right)\sigma_{1i}(\bar{X}^\varepsilon(s))dW^{(i)}(s) \\ &\quad + \frac{\varepsilon}{2}\int_0^1 \frac{d^2G_\kappa}{dx^2}\left((\bar{X}^\varepsilon(s))_1\right)a_{11}(\bar{X}^\varepsilon(s))ds + 2\Lambda_1(0). \end{aligned}$$

Here $a(x)$ is the diffusion matrix $a(x) \doteq \sigma(x)\sigma^T(x)$ and $\Lambda_t(0)$ denotes the semimartingale local time of $(\bar{X}^\varepsilon(\cdot))_1$ at the origin.The definition of G^κ, Condition 2.2, and the nonnegativity of the local time imply that

$$4m \geq \int_{\mathbb{R}^d\times[0,1]} g_\kappa\left((\bar{X}^\varepsilon(s))_1\right)\left(b(\bar{X}^\varepsilon(s))+\sigma(\bar{X}^\varepsilon(s))y\right)_1 \nu^\varepsilon(dy\times ds) + \varepsilon^{1/2}\sum_{i=1}^{d}\int_0^1 g_\kappa\left((\bar{X}^\varepsilon(s))_1\right)\sigma_{1i}(\bar{X}^\varepsilon(s))dW^{(i)}(s) - \frac{\varepsilon B_0 B_1}{2}. \quad (3.35)$$

Let $g_\kappa^{(1)}$ be a bounded continuous extension to $\mathbb{R}$ of the restriction of g_κ to $(-\infty,0]$ and let $g_\kappa^{(2)}$ be a bounded continuous extension to $\mathbb{R}$ of the restriction of g_κ to $(0,\infty)$. W.p.1 $\nu^{(i),\varepsilon} \Longrightarrow \nu^{(i)}$ and $\bar{X}^\varepsilon$ converges uniformly on $[0,1]$ to the continuous process $\bar{X}$. Therefore Theorem 5.5 in [2] and the uniform integrability given in Proposition 3.3 yield

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d\times[0,1]} g_\kappa\left((\bar{X}^\varepsilon(t))_1\right)(b(\bar{X}^\varepsilon(t))+\sigma(\bar{X}^\varepsilon(t))\,y)_1\,\nu^\varepsilon(dy\times dt)\\
&\quad=\lim_{\varepsilon\to 0}\sum_{i=1,2}\int_{\mathbb{R}^d\times[0,1]} g_\kappa^{(i)}\left((\bar{X}^\varepsilon(t))_1\right)(b^{(i)}(\bar{X}^\varepsilon(t))+\sigma(\bar{X}^\varepsilon(t))\,y)_1\,\nu^{(i),\varepsilon}(dy\times dt)\\
&\quad=\sum_{i=1,2}\int_{\mathbb{R}^d\times[0,1]} g_\kappa^{(i)}\left((\bar{X}(t))_1(b^{(i)}(\bar{X}(t))+\sigma(\bar{X}(t))\,y)_1\right)\nu^{(i)}(dy\times dt).
\end{aligned}$$

We now combine this expression with (3.35) and use (3.22) together with the fact that if $(\bar{X}(t))_1 < 0$, then $\nu^{(2)}(dy|t) = 0$ and if $(\bar{X}(t))_1 > 0$, then $\nu^{(1)}(dy|t) = 0$ [see (3.26)]. Since along a suitable subsequence the second term on the right-hand side of (3.35) converges to zero w.p.1, we obtain

$$\begin{aligned}
4m \geq &\int_0^1\int_{\mathbb{R}^d}(b^{(1)}(\bar{X}(t))+\sigma(\bar{X}(t))\,y)_1\, g_\kappa^{(1)}\left((\bar{X}(t))_1\right)1_{(-\infty,0]}\left((\bar{X}(t))_1\right)\nu^{(1)}(dy|t)\,dt\\
&+\int_0^1\int_{\mathbb{R}^d}(b^{(2)}(\bar{X}(t))+\sigma(\bar{X}(t))\,y)_1\, g_\kappa^{(2)}\left((\bar{X}(t))_1\right)\\
&\qquad\times 1_{[0,\infty)}\left((\bar{X}(t))_1\right)\nu^{(2)}(dy|t)\,dt. \qquad (3.36)
\end{aligned}$$

The functions $\{g_\kappa^{(1)}, \kappa > 0\}$ and $\{g_\kappa^{(2)}, \kappa > 0\}$ are uniformly bounded and for $z \in \mathbb{R}$

$$\lim_{\kappa\to 0} g_\kappa^{(1)}(z)\,1_{(-\infty,0]}(z) = -1_{[-m,0]}(z) \quad\text{and}\quad \lim_{\kappa\to 0} g_\kappa^{(2)}(z)\,1_{[0,\infty)}(z) = 1_{[0,m]}(z).$$

By the Lebesgue Dominated Convergence Theorem, sending $\kappa \to 0$ in formula (3.36) gives w.p.1

$$\begin{aligned}
4m \geq &\int_0^1\int_{\mathbb{R}^d} 1_{[0,m]}\left((\bar{X}(t))_1\right)(b^{(2)}(\bar{X}^\varepsilon(t))+\sigma(\bar{X}^\varepsilon(t))\,y)_1\,\nu^{(2)}(dy|t)\,dt\\
&-\int_0^1\int_{\mathbb{R}^d} 1_{[-m,0]}\left((\bar{X}(t))_1\right)(b^{(1)}(\bar{X}^\varepsilon(t))+\sigma(\bar{X}^\varepsilon(t))\,y)_1\,\nu^{(1)}(dy|t)\,dt.
\end{aligned}$$

Similar proofs that are based on approximating

$$G(z) \doteq \begin{cases} z & \text{if } |z| \le m \\ m & \text{if } z > m \\ -m & \text{if } z < -m \end{cases}$$

and $-G(z)$ show that w.p.1

$$4m \ge \left| \int_0^1 \int_{\mathbb{R}^d} 1_{[0,m]}\big((\bar{X}(t))_1\big)\, (b^{(2)}(\bar{X}^\varepsilon(t)) + \sigma(\bar{X}^\varepsilon(t))\, y)_1\, \nu^{(2)}(dy|t)\, dt \right.$$
$$\left. + \int_0^1 \int_{\mathbb{R}^d} 1_{[-m,0]}\big((\bar{X}(t))_1\big)\, (b^{(1)}(\bar{X}^\varepsilon(t)) + \sigma(\bar{X}^\varepsilon(t))\, y)_1\, \nu^{(1)}(dy|t)\, dt \right|.$$

Combining these equations gives w.p.1

$$\int_0^1 \int_{\mathbb{R}^d} 1_{[-m,0]}\big((\bar{X}(t))_1\big)\, (b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y)_1\, \nu^{(1)}(dy|t)\, dt \ge -4m$$

and

$$\int_0^1 \int_{\mathbb{R}^d} 1_{[0,m]}\big((\bar{X}(t))_1\big)\, (b^{(2)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y)_1\, \nu^{(2)}(dy|t)\, dt \le 4m.$$

By taking $m \to 0$, we conclude that w.p.1

$$\int_0^1 \int_{\mathbb{R}^d} 1_{\{0\}}\big((\bar{X}(t))_1\big)\, (b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y)_1\, \nu^{(1)}(dy|t)\, dt \ge 0$$

and

$$\int_0^1 \int_{\mathbb{R}^d} 1_{\{0\}}\big((\bar{X}(t))_1\big)\, (b^{(2)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y)_1\, \nu^{(2)}(dy|t)\, dt \le 0.$$

Let $[\alpha, \beta]$ be any closed interval in $[0, 1]$. Repeating the argument leading to the last two displays, we obtain

$$\theta^{(1)}(\alpha, \beta) \doteq \int_\alpha^\beta \int_{\mathbb{R}^d} 1_{\{0\}}\big((\bar{X}(t))_1\big)\, (b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y)_1\, \nu^{(1)}(dy|t)\, dt \ge 0$$

and

$$\theta^{(2)}(\alpha, \beta) \doteq \int_\alpha^\beta \int_{\mathbb{R}^d} 1_{\{0\}}\big((\bar{X}(t))_1\big)\, (b^{(2)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y)_1\, \nu^{(2)}(dy|t)\, dt \le 0.$$

With probability 1 these inequalities hold simultaneously for all intervals $[\alpha, \beta] \subset [0, 1]$ with rational endpoints, and thus by continuity they hold simultaneously for all intervals $[\alpha, \beta] \subset [0, 1]$. This implies that w.p.1 $\theta^{(1)}(0, \beta)$ is nondecreasing and $\theta^{(2)}(0, \beta)$ is nonincreasing for $\beta \in [0, 1]$. Since a nondecreasing (resp., nonincreasing) function has a nonnegative (resp., nonpositive) derivative a.s., it follows that w.p.1 we have a.s. for $t \in [0, 1]$, whenever $(\bar{X}(t))_1 = 0$

$$\int_{\mathbb{R}^d} \left(b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y \right)_1 \nu^{(1)}(dy|t) \geq 0$$

and

$$\int_{\mathbb{R}^d} \left(b^{(2)}(\bar{X}(t)) + \sigma(\bar{X}(t))\, y \right)_1 \nu^{(2)}(dy|t) \leq 0.$$

This proves part (d), completing the proof of Theorem 3.4. □

4. Proof of Theorem 2.3

We split the proof of the Laplace principle in Theorem 2.3 into two parts: the Laplace principle upper bound and the Laplace principle lower bound.

4.1. Proof of the Laplace principle upper bound

For each $\varepsilon > 0$ let X^ε be the unique strong solution to (2.8). To prove the Laplace principle upper bound we must show that for all bounded continuous functions h mapping $\mathscr{C}([0,1] : \mathbb{R}^d)$ into $\mathbb{R}$

$$\limsup_{\varepsilon \to 0} \varepsilon \log E_{x_0} \left\{ \exp \left[-\frac{h(X^\varepsilon)}{\varepsilon} \right] \right\} \leq - \inf_{\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)} \{ I_{x_0}(\varphi) + h(\varphi) \}.$$

We will show the equivalent lower limit

$$\liminf_{\varepsilon \to 0} W^\varepsilon(x_0) \geq \inf_{\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)} \{ I_{x_0}(\varphi) + h(\varphi) \}, \tag{4.37}$$

where

$$W^\varepsilon(x_0) \doteq -\varepsilon \log E_{x_0} \left\{ \exp \left[-\frac{h(X^\varepsilon)}{\varepsilon} \right] \right\}.$$

It is sufficient to prove the lower limit (4.37) when ε is replaced by any subsequence along which $W^\varepsilon(x_0)$ converges. Such a subsequence exists since $|W^\varepsilon(x_0)| \leq \|h\|_\infty$. We will work with a fixed such subsequence for the remainder of the proof, and for convenience we relabel the indices with $\varepsilon > 0$.

The key to the proof is the use of the representation formula for $W^\varepsilon(x_0)$ given in Theorem 3.1. Thanks to this theorem, we can construct a family $\{v^\varepsilon, \varepsilon > 0\}$ of controls in $\mathscr{A}$ so that for each $\varepsilon > 0$

$$W^\varepsilon(x_0) \geq E_{x_0} \left\{ \frac{1}{2} \int_0^1 \|v^\varepsilon(t)\|^2 dt + h(\bar{X}^\varepsilon) \right\} - \varepsilon, \tag{4.38}$$

where $\bar{X}^\varepsilon$ is the controlled diffusion associated with v^ε via (3.9). Since by definition $|W^\varepsilon(x_0)| \leq \|h\|_\infty$, the family $\{v^\varepsilon, \varepsilon > 0\}$ satisfies Condition 3.2. Therefore, if we use the family of controls $\{v^\varepsilon, \varepsilon > 0\}$ to define the measures ν^ε, $\nu^{(1),\varepsilon}$, $\nu^{(2),\varepsilon}$, $\gamma^{(1),\varepsilon}$ and $\gamma^{(2),\varepsilon}$ as in Section 3, then along some subsubsequence of $\varepsilon > 0$

$$(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}, \bar{X}^\varepsilon) \xrightarrow{\mathscr{D}} (\nu, \nu^{(1)}, \nu^{(2)}, \gamma^{(1)}, \gamma^{(2)}, \bar{X}).$$

The limit quantities $\nu, \nu^{(1)}, \nu^{(2)}, \gamma^{(1)}, \gamma^{(2)}$ and $\bar{X}$ satisfy the conclusions of Theorem 3.4. By the Skorohod Representation Theorem we can assume that the convergence in the last display occurs w.p.1.

We can now evaluate the limit inferior of $W^\varepsilon(x_0)$. Each step of the process is explained after the display. We have

$$\begin{aligned}
&\liminf_{\varepsilon\to 0} W^\varepsilon(x_0) \qquad (4.39)\\
&\geq \liminf_{\varepsilon\to 0} \left\{ E_{x_0} \left\{ \frac{1}{2}\int_0^1 \|v^\varepsilon(t)\|^2 dt + h(\bar{X}^\varepsilon) \right\} - \varepsilon \right\}\\
&\geq \liminf_{\varepsilon\to 0} E_{x_0} \left\{ \frac{1}{2}\int_{\mathbb{R}^d\times[0,1]} \|y\|^2 \nu^\varepsilon(dy\times dt) + h(\bar{X}^\varepsilon) \right\}\\
&\geq E_{x_0} \left\{ \frac{1}{2}\int_{\mathbb{R}^d\times[0,1]} \|y\|^2 \nu(dy\times dt) + h(\bar{X}) \right\}\\
&= E_{x_0}\left\{ \frac{1}{2}\int_{\mathbb{R}^d\times[0,1]} \|y\|^2 \nu^{(1)}(dy\times dt) + \frac{1}{2}\int_{\mathbb{R}^d\times[0,1]} \|y\|^2 \nu^{(2)}(dy\times dt) + h(\bar{X}) \right\}\\
&= E_{x_0} \left\{ \frac{1}{2}\int_0^1\int_{\mathbb{R}^d} \|y\|^2 \nu^{(1)}(dy|t)dt + \frac{1}{2}\int_0^1\int_{\mathbb{R}^d} \|y\|^2 \nu^{(2)}(dy|t)dt + h(\bar{X}) \right\}\\
&\geq E_{x_0} \Bigg\{ \frac{1}{2}\int_0^1 \left\| \frac{1}{\hat{\gamma}^{(1)}(t)}\int_{\mathbb{R}^d} y\nu^{(1)}(dy|t) \right\|^2 \hat{\gamma}^{(1)}(t)\, dt\\
&\qquad + \frac{1}{2}\int_0^1 \left\| \frac{1}{\hat{\gamma}^{(2)}(t)}\int_{\mathbb{R}^d} y\nu^{(2)}(dy|t) \right\|^2 \hat{\gamma}^{(2)}(t)dt + h(\bar{X}) \Bigg\}\\
&\geq E_{x_0} \Bigg\{ \int_0^1 \hat{\gamma}^{(1)}(t) L^{(1)}\Big(\bar{X}(t), b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t))\frac{1}{\hat{\gamma}^{(1)}(t)}\int_{\mathbb{R}^d} y\nu^{(1)}(dy|t) \Big)\, dt\\
&\qquad + \int_0^1 \hat{\gamma}^{(2)}(t) L^{(2)}\Big(\bar{X}(t), b^{(2)}(\bar{X}(t))\\
&\qquad + \sigma(\bar{X}(t))\frac{1}{\hat{\gamma}^{(2)}(t)}\int_{\mathbb{R}^d} y\nu^{(2)}(dy|t) \Big)\, dt + h(\bar{X}) \Bigg\}\\
&\geq E_{x_0} \left\{ \int_0^1 \tilde{L}\Big(\bar{X}(t), \dot{\bar{X}}(t)\Big)\, dt + h(\bar{X}) \right\}\\
&\geq \inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)} \{I_x(\varphi) + h(\varphi)\}.
\end{aligned}$$

The first two lines of this display are consequences of (4.38). Line three uses the control measure representations ν^ε for the controls v^ε given in (3.11). Since h is continuous on $\mathscr{C}([0,1]:\mathbb{R}^d)$ and w.p.1 $\{\bar{X}^\varepsilon\}$ converges uniformly on $[0,1]$ to $\bar{X}$,

$h(\bar{X}^\varepsilon) \to h(\bar{X})$ w.p.1. Line four in (4.39) now follows from Fatou's Lemma and the w.p.1 convergence $\nu^\varepsilon \Rightarrow \nu$, and line five from the probability–1 equality $\nu = \nu^{(1)} + \nu^{(2)}$. Line six is obtained from the probability–1 decompositions $\nu^{(i)}(dy \times dt) = \nu^{(i)}(dt|y) \otimes dt$, and after normalization, lines seven and eight follow from Jensen's inequality. The normalization is well defined if we adopt the convention $0 \cdot \infty = 0$. Lines nine and ten follow from the expressions for $L^{(i)}(x, \beta)$ given in (2.2), and line eleven from the fact that w.p.1 and a.s. for $t \in [0, 1]$

$$\begin{aligned}\tilde{L}(\bar{X}(t), \dot{\bar{X}}(t)) \le \hat{\gamma}^{(1)}(t) L^{(1)}\bigg(\bar{X}(t), b^{(1)}(\bar{X}(t)) + \sigma(\bar{X}(t)) \int_{\mathbb{R}^d} y \nu^{(1)}(dy|t)\bigg) \\ + \hat{\gamma}^{(2)}(t) L^{(2)}\bigg(\bar{X}(t), b^{(2)}(\bar{X}(t)) + \sigma(\bar{X}(t)) \int_{\mathbb{R}^d} y \nu^{(2)}(dy|t)\bigg).\end{aligned}$$

This formula follows from the definition of $\tilde{L}(\cdot, \cdot)$ (details can be found in the proof of Proposition 7.4.1 in [5]). Finally, the last line of (4.39) is a consequence of the definition of the rate function.

We have shown that every convergent subsequence of $\{W^\varepsilon(x_0), \varepsilon > 0\}$ has a subsequence satisfying

$$\liminf_{\varepsilon \to 0} W^\varepsilon(x_0) \ge \inf_{\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)} \{I_{x_0}(\varphi) + h(\varphi)\}.$$

An argument by contradiction establishes this lower limit for the entire family $\{W^\varepsilon(x_0), \varepsilon > 0\}$. Thus the proof of the Laplace principle upper bound is complete.

4.2. Proof of the Laplace principle lower bound

The proof of the Laplace principle lower bound requires some properties of $L^{(i)}(x, \beta)$, $i = 0, 1, 2$, and of $\tilde{L}(x, \beta)$. We state them in the following lemma.

Lemma 4.1. *Under Condition* 2.2, *the functions* $L^{(0)}(x, \beta)$, $L^{(1)}(x, \beta)$, $L^{(2)}(x, \beta)$, *and* $\tilde{L}(x, \beta)$ *have the following properties.*

(a) *For x and β in* $\mathbb{R}^d$

$$L^{(0)}(x, \beta) \le L^{(1)}(x, \beta) \text{ if } \beta_1 \ge 0 \text{ and } L^{(0)}(x, \beta) \le L^{(2)}(x, \beta) \text{ if } \beta_1 \le 0.$$

(b) *For $i = 1, 2$, the functions $L^{(i)}(x, \beta)$ are continuous functions of* $(x, \beta) \in \mathbb{R}^d \times \mathbb{R}^d$.

(c) $L^{(0)}(x, \beta)$ *is a continuous function of* $(x, \beta) \in \mathbb{R}^d \times \partial$.

(d) *For each $x \in \mathbb{R}^d$* dom $L^{(0)}(x, \cdot)$ *equals* $\mathbb{R}^d$ *and thus* dom $\tilde{L}(x, \cdot)$ *equals* $\mathbb{R}^d$.

Remarks. The proof of part (b) follows directly from Condition 2.2. The other parts of the lemma can be proved as in Lemma 7.5.3 in [5]. □

In order to prove the Laplace principle lower bound, we must verify that

$$\liminf_{\varepsilon \to 0} \varepsilon \log E_{x_0}\{\exp[-h(X^\varepsilon)/\varepsilon]\} \ge -\inf_{\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)} \{I_{x_0}(\varphi) + h(\varphi)\}.$$

To prove this statement, we adapt an approximation procedure introduced in Chapter 7 of [5]. Let $\mathscr{N}_0$ be the class of functions $\psi^* \in \mathscr{C}([0,1]:\mathbb{R}^d)$ which satisfy the following conditions:

(a) $\dot{\psi}^*(t)$ is piecewise constant with only finitely many jumps in the interval $(0,1)$.
(b) Either $(\psi^*(t))_1 \neq 0$ or $(\psi^*(t))_1 = 0$ on each interval of constancy of $\dot{\psi}^*$.

In order to have $\dot{\psi}^*(t)$ defined for all $t \in [0,1)$, the almost everywhere defined function $\dot{\psi}$ is replaced by its right continuous regularization. We will show that given $x_0 \in \mathbb{R}^d$ the inequality

$$\limsup_{\varepsilon\to 0} W^\varepsilon(x_0) \doteq \limsup_{\varepsilon\to 0} \big(-\varepsilon \log E_{x_0}\{\exp[-h(X^\varepsilon)/\varepsilon]\}\big) \leq I_{x_0}(\psi^*) + h(\psi^*) \tag{4.40}$$

holds for all $\psi^* \in \mathscr{N}_0$. Thanks to the following lemma, whose proof can be found in [5, Theorem 7.5.4], this property can then be extended to the whole space $\mathscr{C}([0,1]:\mathbb{R}^d)$.

Lemma 4.2. *Assume Condition* 2.2. *Given* $x_0 \in \mathbb{R}^d$, *let* $\psi \in \mathscr{C}([0,1]:\mathbb{R}^d)$ *satisfy* $I_{x_0}(\psi) < \infty$. *Then for each* $\eta > 0$, *there exists* $\psi^* \in \mathscr{N}_0$ *such that*

$$\|\psi^* - \psi\|_\infty \leq \eta \ \textit{ and } \ I_{x_0}(\psi^*) \leq I_{x_0}(\psi) + \eta$$

and for each $k \in \{1,2,\ldots,r\}$ *either* $(\psi^*(t))_1 \neq 0$ *for all* $t \in (t_k, t_{k+1})$ *or* $(\psi^*(t))_1 = 0$ *for all* $t \in (t_k, t_{k+1})$. *The intervals* (t_k, t_{k+1}), $k = 1,2,\ldots,r$, *denote the interiors of the successive intervals on which* $\dot{\psi}^*$ *is constant.*

Assume that (4.40) holds for $\psi^* \in \mathscr{N}_0$. We now describe how to complete the proof of the Laplace principle lower bound. Given $\eta > 0$, let $\psi \in \mathscr{C}([0,1]:\mathbb{R}^d)$ satisfy

$$I_{x_0}(\psi) + h(\psi) \leq \inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)} \{I_{x_0}(\varphi) + h(\varphi)\} + \eta < \infty.$$

Since h is bounded, this implies that $I_{x_0}(\psi) < \infty$. Since h is continuous, there exists $\psi^* \in \mathscr{N}_0$ so that

$$h(\psi^*) \leq h(\psi) + \eta \ \text{ and } \ I_{x_0}(\psi^*) \leq I_{x_0}(\psi) + \eta.$$

It follows that

$$\begin{aligned}
\liminf_{\varepsilon\to 0} \ & \varepsilon \log E_{x_0}\left\{\exp\left[-\frac{h(X^\varepsilon)}{\varepsilon}\right]\right\} \\
& \geq -I_{x_0}(\psi^*) - h(\psi^*) \\
& \geq -I_{x_0}(\psi) - h(\psi) - 2\eta \\
& \geq -\inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)} \{I_{x_0}(\varphi) + h(\varphi)\} - 3\eta.
\end{aligned}$$

Since $\eta > 0$ is arbitrary, this inequality yields the Laplace principle lower bound.

We now proceed to show that (4.40) holds for all $\psi^* \in \mathcal{N}_0$. To this end, let $\psi^* \in \mathcal{N}_0$ satisfy $I_{x_0}(\psi^*) < \infty$. For each $k \in \{1, \ldots, r\}$ let $\beta_k \doteq \dot{\psi}^*(t)$, where t is any point in the interior of the interval of constancy (t_k, t_{k+1}) of this derivative. Let $\beta_k^{(1)}$ and $\beta_k^{(2)}$ be defined as follows. If $(\psi^*(t))_1 \neq 0$ for all $t \in (t_k, t_{k+1})$, then $\beta_k^{(1)} = \beta_k^{(2)} \doteq \beta_k$. Otherwise (that is, if $(\psi^*(t))_1 = 0$ for all $t \in (t_k, t_{k+1})$), given $\eta > 0$ $\beta_k^{(1)}$ and $\beta_k^{(2)}$ are chosen together with constants $\rho_k^{(1)}$ and $\rho_k^{(2)}$ so that

$$\rho_k^{(1)} > 0,\ \rho_k^{(2)} > 0,\ (\beta_k^{(1)})_1 > 0,\ (\beta_k^{(2)})_1 < 0,$$

$$\rho_k^{(1)} + \rho_k^{(2)} = 1,\ \rho_k^{(1)}\beta_k^{(1)} + \rho_k^{(2)}\beta_k^{(2)} = \beta_k, \tag{4.41}$$

and

$$\rho_k^{(1)} L^{(1)}(\psi^*(t), \beta_k^{(1)}) + \rho_k^{(2)} L^{(2)}(\psi^*(t), \beta_k^{(2)}) \leq L^{(0)}(\psi^*(t), \beta_k) + \eta \tag{4.42}$$

for all $t \in (t_k, t_k + \lambda)$, where $\lambda > 0$. The existence of $\beta_k^{(1)}$, $\beta_k^{(2)}$, $\rho_k^{(1)}$ and $\rho_k^{(2)}$ satisfying (4.41) and (4.42) follows from the continuity of $L^{(1)}(\cdot,\cdot)$ and $L^{(2)}(\cdot,\cdot)$ on $\mathbb{R}^d \times \mathbb{R}^d$ (Lemma 4.1), as is shown in Lemma 7.5.3 in [5]. The continuity properties of $L^{(0)}$, $L^{(1)}$, $L^{(2)}$ and ψ^* imply that if necessary we can add points to the original subdivision $0 = t_1 < t_s < \ldots < t_{r+1} = 1$ of $(0, 1)$ to obtain a finer subdivision $0 = \tilde{t}_1 < \tilde{t}_2 < \ldots < \tilde{t}_{s+1} = 1$ of $(0, 1)$ for which $(\psi^*(t))_1 = 0$ for all $t \in (\tilde{t}_k, \tilde{t}_{k+1})$ implies that (4.42) holds for all $t \in (\tilde{t}_k, \tilde{t}_{k+1})$. For simplicity we will retain the same notation for the original subdivision and the refined subdivision.

We next describe how to use the vectors $\beta_k^{(i)}, k \in \{1, \ldots, r\}, i = 1, 2$, to design a family of controls $\{v^\varepsilon, \varepsilon > 0\}$ in $\mathscr{A}$ and a corresponding family of controlled processes $\{\bar{X}^\varepsilon, \varepsilon > 0\}$. These families will be used to prove (4.40) through the representation formula given in Theorem 3.1. For every $x \in \mathbb{R}^d$, $i = 1, 2$ and $t \in [0, 1]$ such that $t \in [t_k, t_{k+1})$ for some $k \in \{1, \ldots, r\}$, define

$$v^{(i)}(x, t) \doteq \sigma^{-1}(x)(\beta_k^{(i)} - b^{(i)}(x)).$$

According to (2.2), these vectors satisfy

$$\tfrac{1}{2}\|v^{(i)}(x, t)\|^2 = L^{(i)}(x, \beta_k^{(i)}).$$

Now define

$$f(x, t) \doteq v^{(1)}(x, t)1_{\{x_1 \leq 0\}}(x) + v^{(2)}(x, t)1_{\{x_1 > 0\}}(x)$$

and let $\bar{X}^\varepsilon$ be the solution to the SDE

$$\bar{X}^\varepsilon(t) = x_0 + \int_0^t b(\bar{X}^\varepsilon(s))ds + \int_0^t \sigma(\bar{X}^\varepsilon(s)) f(\bar{X}^\varepsilon(s), s)ds + \varepsilon^{1/2}\int_0^t \sigma(\bar{X}^\varepsilon(s))dW(s).$$

Finally, let

$$v^\varepsilon(t) \doteq f(\bar{X}^\varepsilon(t), t).$$

The controls v^ε have the following properties. First, for each $\varepsilon > 0$, the control v^ε is an element of $\mathscr{A}$. Indeed, since $\bar{X}^\varepsilon$ and t are progressively measurable, v^ε is also progressively measurable. Since v^ε is also bounded, $v^\varepsilon \in \mathscr{A}$. Next, if $t \in [t_k, t_{k+1})$ and $(\bar{X}^\varepsilon(t))_1 \leq 0$, then

$$\tfrac{1}{2}\|v^\varepsilon(t)\|^2 = L^{(1)}(\bar{X}^\varepsilon(t), \beta_k^{(1)}) \quad \text{and} \quad \beta_k^{(1)} = b^{(1)}(\bar{X}^\varepsilon(t)) + \sigma(\bar{X}^\varepsilon(t))v^\varepsilon(t), \tag{4.43}$$

and if $(\bar{X}^\varepsilon(t))_1 > 0$, then

$$\tfrac{1}{2}\|v^\varepsilon(t)\|^2 = L^{(2)}(\bar{X}^\varepsilon(t), \beta_k^{(2)}) \quad \text{and} \quad \beta_k^{(1)} = b^{(2)}(\bar{X}^\varepsilon(t)) + \sigma(\bar{X}^\varepsilon(t))v^\varepsilon(t). \tag{4.44}$$

Finally, we claim that the family $\{v^\varepsilon, \varepsilon > 0\}$ satisfies Condition 3.2. For each $\varepsilon > 0$ (4.43) and (4.44) give

$$\begin{aligned} E_{x_0}\left\{\frac{1}{2}\int_0^1 \|v^\varepsilon(t)\|^2 dt\right\} &= E_{x_0}\left\{\sum_{k=1}^{r}\int_{t_k}^{t_{k+1}} \left[L^{(1)}\left(\bar{X}^\varepsilon(t), \beta_k^{(1)}\right)\cdot 1_{\{(\bar{X}^\varepsilon(t))_1\leq 0\}}\right.\right. \\ &\qquad \left.\left. +\ L^{(2)}\left(\bar{X}^\varepsilon(t), \beta_k^{(2)}\right)\cdot 1_{\{(\bar{X}^\varepsilon(t))_1> 0\}}\right]\right\} \\ &\leq \sup_{x\in\mathbb{R}^d}\left\{\|a^{-1}(x)\|\sum_{k=1}^{r}\left[\|\beta_k^{(1)} - b^{(1)}(x)\|^2\right.\right. \\ &\qquad \left.\left. +\ \|\beta_k^{(2)} - b^{(2)}(x)\|^2\right]\right\}. \end{aligned} \tag{4.45}$$

Condition 2.2 thus implies that the sequence of expected values in the last display is bounded for $\varepsilon > 0$, as claimed.

Since Condition 3.2 is satisfied, we can apply the compactness and convergence results obtained in Section 3. Using the family $\{v^\varepsilon,\ \varepsilon > 0\}$ constructed above, we define the measures ν^ε, $\nu^{(1),\varepsilon}$, $\nu^{(2),\varepsilon}$, $\gamma^{(1),\varepsilon}$, and $\gamma^{(2),\varepsilon}$ as in (3.10)–(3.16). Proposition 3.3 implies that the family $\left\{\left(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}, \bar{X}^\varepsilon\right), \varepsilon > 0\right\}$ is tight. Theorem 3.4 and Prohorov's Theorem then imply that given any subsequence there exists a subsubsequence that satisfies

$$\left(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}, \bar{X}^\varepsilon\right) \xrightarrow{\mathscr{D}} \left(\nu, \nu^{(1)}, \nu^{(2)}, \gamma^{(1)}, \gamma^{(2)}, \bar{X}\right).$$

The limit quantities ν, $\nu^{(1)}$, $\nu^{(2)}$, $\gamma^{(1)}$, $\gamma^{(2)}$ and $\bar{X}$ satisfy (3.19)–(3.22) and the conclusions of Theorem 3.4. By the Skorohod Representation Theorem, we can assume that convergence takes place w.p.1.

Our next step is to show that $\bar{X}(s) = \psi^*(s)$ for all $s \in [0, 1]$. It suffices to show that w.p.1 $\bar{X}(s) = \psi^*(s)$ for each $k \in \{1, 2, \ldots, r+1\}$ and all $s \in [0, t_k]$. The proof of this last statement is by induction on k. For $k = 1$ the equality holds because $t_1 = 0$ and w.p.1 $\bar{X}(0) = x_0 = \phi^*(0)$. Assuming that w.p.1 $\bar{X}(s) = \psi^*(s)$ for some $k \in \{1, 2, \ldots, r\}$ and all $s \in [0, t_k]$, we must show that w.p.1 $\bar{X}(s) = \psi^*(s)$ for all $s \in [0, t_{k+1}]$. First consider the case where $(\psi^*(s))_1 \neq 0$ for all $s \in (t_k, t_{k+1})$.

Using (4.43) and (4.44) and the fact that, by definition, β_k equals the constant value $\dot{\psi}^*(s)$ for any $s \in (t_k, t_{k+1})$, we obtain

$$\begin{aligned}\bar{X}(t) - \bar{X}(t_k) &= \lim_{\varepsilon\to 0} \int_{\mathbb{R}^d\times(t_k,t]} \Big(b^{(1)}(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\, \nu^{(1),\varepsilon}(dy \times ds) \\ &\quad + \int_{\mathbb{R}^d\times(t_k,t]} \Big(b^{(2)}(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\, \nu^{(2),\varepsilon}(dy \times ds) \\ &= \lim_{\varepsilon\to 0} \int_{\mathbb{R}^d\times(t_k,t]} \beta_k \nu^\varepsilon(dy \times ds) \\ &= \int_{t_k}^{t} \dot{\psi}^*(s)\, ds = \psi^*(t) - \psi^*(t_k),\end{aligned}$$

valid w.p.1 for any $t \in [t_k, t_{k+1})$. The equality $\bar{X}(t) - \bar{X}(t_k) = \psi^*(t) - \psi^*(t_k)$ extends by continuity to $t = t_{k+1}$. Thanks to the inductive hypothesis, this implies that w.p.1 $\bar{X}(t) = \psi^*(t)$ for all $t \in [t_k, t_{k+1}]$ and thus for all $t \in [0, t_{k+1}]$.

Now assume $(\psi^*(s))_1 = 0$ for all $s \in (t_k, t_{k+1})$. Given numbers α and t satisfying $t_k < \alpha < t < t_{k+1}$,

$$\begin{aligned}&\int_{\mathbb{R}^d\times(\alpha,t]} \Big(b^{(1)}(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\, \nu^{(1),\varepsilon}(dy \times ds) \\ &\quad + \int_{\mathbb{R}^d\times(\alpha,t]} \Big(b^{(2)}(\bar{X}^\varepsilon(s)) + \sigma(\bar{X}^\varepsilon(s))y\Big)\, \nu^{(2),\varepsilon}(dy \times ds) \\ &\quad = \beta_k^{(1)}\gamma^{(1),\varepsilon}((\alpha, t]) + \beta_k^{(2)}\gamma^{(2),\varepsilon}((\alpha, t]).\end{aligned}$$

According to (3.21), w.p.1 $\gamma^{(i)}$ is absolutely continuous with respect to Lebesgue measure on $[0, 1]$. Hence, after normalizing, part (e) of Portmanteau's Theorem [2, Theorem 2.1] applied to the weakly convergent sequences $\{\gamma^{(i),\varepsilon}\}$ implies that w.p.1, for all α and t satisfying $t_k < \alpha < t < t_{k+1}$,

$$\begin{aligned}\bar{X}(t) - \bar{X}(\alpha) &= \int_{\mathbb{R}^d\times(\alpha,t]} \Big(b^{(1)}(\bar{X}(s)) + \sigma(\bar{X}(s))y\Big)\, \nu^{(1)}(dy \times ds) \\ &\quad + \int_{\mathbb{R}^d\times(\alpha,t]} \Big(b^{(2)}(\bar{X}(s)) + \sigma(\bar{X}(s))y\Big)\, \nu^{(2)}(dy \times ds) \\ &= \beta_k^{(1)}\gamma^{(1)}((\alpha, t]) + \beta_k^{(2)}\gamma^{(2)}((\alpha, t]).\end{aligned}$$

As in Proposition 7.5.1 in [5], this relation implies that w.p.1

$$\hat{\gamma}^{(1)}(s) = \rho_k^{(1)} \text{ and } \hat{\gamma}^{(2)}(s) = \rho_k^{(2)} \tag{4.46}$$

and that $\bar{X}(s) = \psi^*(s)$ for all $s \in [t_k, t_{k+1}]$ and thus for all $s \in [0, t_{k+1}]$, as we wanted to show.

To evaluate the limit superior of $W^\varepsilon(x_0)$, we need to introduce additional notation. For $k \in \{1, 2, \ldots, r\}$ and $t \in [0, 1]$, we define

$$g_k(t) \doteq \begin{cases} L^{(1)}(\psi^*(t), \beta_k) & \text{if } (\psi^*(t))_1 < 0 \\ L^{(2)}(\psi^*(t), \beta_k) & \text{if } (\psi^*(t))_1 > 0 \\ \rho_k^{(1)}\, L^{(1)}(\psi^*(t), \beta_k^{(1)}) + \rho_k^{(2)}\, L^{(2)}(\psi^*(t), \beta_k^{(2)}) & \text{if } (\psi^*(t))_1 = 0. \end{cases}$$

It follows easily that

$$g_k(t) \le \tilde{L}\big(\psi^*(t), \dot{\psi}^*(t)\big) + \eta. \tag{4.47}$$

We use this property in the following display, where the limit superior of $W^\varepsilon(x_0)$ is evaluated along a subsubsequence of $\varepsilon > 0$ for which w.p.1

$$\Big(\nu^\varepsilon, \nu^{(1),\varepsilon}, \nu^{(2),\varepsilon}, \gamma^{(1),\varepsilon}, \gamma^{(2),\varepsilon}, \bar{X}^\varepsilon\Big) \to \Big(\nu, \nu^{(1)}, \nu^{(2)}, \gamma^{(1)}, \gamma^{(2)}, \bar{X}\Big).$$

We have

$$\begin{aligned}
\limsup_{\varepsilon\to 0} W^\varepsilon(x_0) &= \limsup_{\varepsilon\to 0} \inf_{v\in\mathscr{A}} E_{x_0}\left\{\frac{1}{2}\int_0^1 \|v(t)\|^2 dt + h(X^{v,\varepsilon})\right\} \\
&\le \limsup_{\varepsilon\to 0} E_{x_0}\left\{\frac{1}{2}\int_0^1 \|v^\varepsilon(t)\|^2 dt + h(\bar{X}^\varepsilon)\right\} \\
&\le \lim_{\varepsilon\to 0} E_{x_0}\Bigg\{\sum_{k=1}^r \int_{t_k}^{t_{k+1}} L^{(1)}\Big(\bar{X}^\varepsilon(t), \beta_k^{(1)}\Big)\, \gamma^{(1),\varepsilon}(dt) \\
&\qquad + \sum_{k=1}^r \int_{t_k}^{t_{k+1}} L^{(2)}\Big(\bar{X}^\varepsilon(t), \beta_k^{(2)}\Big)\, \gamma^{(2),\varepsilon}(dt) + h(\bar{X}^\varepsilon)\Bigg\} \\
&= E_{x_0}\Bigg\{\sum_{k=1}^r \int_{t_k}^{t_{k+1}} \Bigg[L^{(1)}\Big(\bar{X}(t), \beta_k^{(1)}\Big)\, \hat{\gamma}^{(1)}(t) \\
&\qquad + L^{(2)}\Big(\bar{X}(t), \beta_k^{(2)}\Big)\, \hat{\gamma}^{(2)}(t)\Bigg] dt + h(\bar{X})\Bigg\} \\
&= \sum_{k=1}^r \int_{t_k}^{t_{k+1}} g_k(t)\, dt + E_{x_0}\{h(\bar{X})\} \\
&\le \int_0^1 \tilde{L}\big(\psi^*(t), \dot{\psi}^*(t)\big)\, ds + h(\psi^*) + \eta \\
&= I_{x_0}(\psi^*) + h(\psi^*) + \eta.
\end{aligned}$$

The first line of the display is a consequence of the representation formula given in Theorem 3.1. In line two we introduce the family of controls $\{v^\varepsilon, \varepsilon > 0\}$. Equation (4.45) and the definitions of the measures $\gamma^{(i),\varepsilon}$ and of the controlled processes $\bar{X}^\varepsilon$ give lines three and four. W.p.1, for $i = 1, 2$ $\gamma^{(i),\varepsilon} \Longrightarrow \gamma^{(i)}$, which is absolutely continuous with respect to Lebesgue measure and has the decomposition $\gamma^{(i)}(dt) = \hat{\gamma}^{(i)}(t)\, dt$. Therefore the uniform convergence of $\bar{X}^\varepsilon$ to $\bar{X}$ w.p.1 on $[0, 1]$, the continuity on $\mathbb{R}^d$ of $L^{(i)}(\cdot, \beta)$ for each $\beta \in \mathbb{R}^d$, Theorem 5.5 in [2], the continuity of h on $\mathscr{C}([0, 1] : \mathbb{R}^d)$, and the Lebesgue Dominated Convergence Theorem give lines five and six. W.p.1 $\bar{X}(t) = \psi^*(t)$ for all $t \in [0, 1]$. The last three lines of the display now follow from the properties of $\hat{\gamma}^{(i)}$ given in (3.26) and (4.46), the inequality (4.47) relating g_k and $\tilde{L}(\cdot, \cdot)$, and the definition of $I_{x_0}(\psi^*)$.

Since $\eta > 0$ is arbitrary, we have shown that the upper limit

$$\limsup_{\varepsilon \to 0} W^{\varepsilon}(x_0) \leq I_{x_0}(\psi^*) + h(\psi^*)$$

holds for the convergent subsubsequence. An argument by contradiction applied to an arbitrary subsequence of the original family $\{W^{\varepsilon}(x_0), \varepsilon > 0\}$ yields the same upper limit for the entire family. This completes the proof of the Laplace principle lower bound. Since $I_{x_0}(\cdot)$ has compact level sets [5, Proposition 7.6.1], this completes the proof of the Laplace principle stated in Theorem 2.3. □

References

1. Anderson, P., Orey, S.: Small random perturbations of dynamical systems with reflecting boundary. Nagoya Mathematical Journal. 60, 189–216, (1976)
2. Billingsley, P.: Convergence of Probability Measures. John Wiley & Sons, New York (1968)
3. Boué, M., Dupuis, P.: A variational representation for certain functionals of Brownian motion. Ann. Probab. 26, 1641–1659, (1998)
4. Chiang, T.S., Sheu, S.J.: Large deviation of diffusion processes and occupation times with discontinuous drift. Preprint (1997)
5. Dupuis, P., Ellis, R.S.: A Weak Convergence Approach to the Theory of Large Deviations. John Wiley & Sons (1997)
6. Dupuis, P., Ellis, R.S.: The large deviation principle for a general class of queueing systems, I. Trans. Amer. Math. Soc. 347, 2689–2751, (1995)
7. Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence. John Wiley & Sons New York (1986)
8. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Springer-Verlag Berlin (1984)
9. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Second edition. Springer-Verlag, New York (1991)
10. Korostelev, A.P., Leonov, S.L.: Action functional for diffusion processes with discontinuous drift (in Russian). Theory Probab. Appl., 37, 570–576, (1992)
11. Korostelev, A.P., Leonov, S.L.: Action functional for diffusions in discontinuous media, Probab. Theory and Related Fields. 94, 317–333, (1993)
12. Kushner, H.J.: Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory. MIT Press, Cambridge (1984)
13. Veretennikov, A.Ju.: On strong solutions and explicit formulas for solutions of stochastic integral equations. Math. USSR Sbornik. 39, 387–403, (1981)