ON LARGE DEVIATIONS FOR SMALL NOISE ITÔ PROCESSES

ALBERTO CHIARINI,* *Technische Universität Berlin*

MARKUS FISCHER,** *Università di Padova*

Abstract

The large deviation principle in the small noise limit is derived for solutions of possibly degenerate Itô stochastic differential equations with predictable coefficients, which may also depend on the large deviation parameter. The result is established under mild assumptions using the Dupuis–Ellis weak convergence approach. Applications to certain systems with memory and to positive diffusions with square-root-like dispersion coefficient are included.

Keywords: Large deviation; Itô process; stochastic differential equation; Freidlin–Wentzell estimate; time delay; CIR process; weak convergence

2010 Mathematics Subject Classification: Primary 60F10; 60H10
Secondary 60J60; 34K50

1. Introduction

Freidlin–Wentzell estimates for Itô stochastic differential equations of diffusion type are concerned with large (order-one) deviations of solutions to

$$\mathrm{d}X_t^\varepsilon = b(X_t^\varepsilon)\,\mathrm{d}t + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)\,\mathrm{d}W_t \tag{1}$$

from their small noise limit as the noise parameter $\varepsilon > 0$ tends to 0. The small noise limit here is the deterministic dynamical system given by the ordinary differential equation

$$\mathrm{d}\varphi_t = b(\varphi_t)\,\mathrm{d}t. \tag{2}$$

In (1) and (2), the solutions are $\mathbb{R}^d$-valued, b is a vector field $\mathbb{R}^d \to \mathbb{R}^d$, σ is a matrix-valued function $\mathbb{R}^d \to \mathbb{R}^{d \times m}$, and W is an m-dimensional standard Brownian motion, which serves as a model for noise. Solutions of (1) and (2) are usually considered over a finite time interval, say $[0, T]$, with the same deterministic initial condition $X_0^\varepsilon = x = \varphi_0$.

Large deviations are quantified in terms of the large deviation principle; see, for instance, Section 1.2 of Dembo and Zeitouni (1998). Let us recall the definition in the context of Polish spaces (i.e. topological spaces that are separable and compatible with a complete metric). Let $\mathcal{X}$ be a Polish space. A *rate function* on $\mathcal{X}$ is a lower-semicontinuous function $\mathcal{X} \to [0, \infty]$. A rate function is said to be *good* if its sublevel sets are compact. The *large deviation principle* is said to hold for a family $(\xi^\varepsilon)_{\varepsilon>0}$ of $\mathcal{X}$-valued random variables with rate function I if, for all $\Gamma \in \mathcal{B}(\mathcal{X})$,

$$-\inf_{x \in \Gamma^\circ} I(x) \le \liminf_{\varepsilon \to 0+} \varepsilon \log \mathbb{P}(\xi^\varepsilon \in \Gamma) \le \limsup_{\varepsilon \to 0+} \varepsilon \log \mathbb{P}(\xi^\varepsilon \in \Gamma) \le -\inf_{x \in \mathrm{cl}(\Gamma)} I(x),$$

where cl(Γ) denotes the closure and Γ° the interior of Γ. We will also need the following alternative characterization. The *Laplace principle* is said to hold for a family $(\xi^\varepsilon)_{\varepsilon>0}$ of $\mathcal{X}$-valued random variables with rate function I if, for all $F \in \mathbf{C}_b(\mathcal{X})$ (i.e. F bounded and continuous),

$$\lim_{\varepsilon \to 0+} -\varepsilon \log \mathbb{E}\left[\exp\left(-\frac{1}{\varepsilon}F(\xi^\varepsilon)\right)\right] = \inf_{x \in \mathcal{X}}\{I(x) + F(x)\}.$$

If the rate function I is good then the Laplace principle holds with rate function I if and only if the large deviation principle holds with rate function I; see, for instance, Section 1.2 of Dupuis and Ellis (1997).

Various sets of assumptions on the coefficients b and σ in (1) are known to imply that the large deviation principle holds for the family $(X^\varepsilon)_{\varepsilon>0}$ of $\mathbf{C}([0, T], \mathbb{R}^d)$-valued random variables. In the nondegenerate case, that is, if $d = m$ and the matrix-valued function $\sigma\sigma^\mathsf{T}$ is uniformly positive definite, the large deviation principle holds if, for instance, b and σ are bounded and uniformly continuous; the rate function then takes the form

$$I_x(\varphi) = \frac{1}{2}\int_0^T (\dot{\varphi}_s - b(s, \varphi_s))^\mathsf{T} (\sigma\sigma^\mathsf{T})^{-1}(s, \varphi_s)(\dot{\varphi}_s - b(s, \varphi_s)) \, \mathrm{d}s$$

whenever $\varphi \in \mathbf{C}([0, T], \mathbb{R}^d)$ is absolutely continuous with $\varphi_0 = x$, and $I_x(\varphi) = \infty$ otherwise; see Theorem 5.3.1 of Freidlin and Wentzell (1998, pp. 154–155). In the general case of a possibly degenerate diffusion matrix, the rate function I_x can be expressed as

$$I_x(\varphi) := \inf_{\{f \in L^2 : \varphi = x + \int_0^\cdot (b(\varphi_s)+\sigma(\varphi_s)f_s) \, \mathrm{d}s\}} \frac{1}{2}\int_0^T |f_t|^2 \, \mathrm{d}t, \tag{3}$$

where $\inf \varnothing = \infty$ by convention; see, for instance, Section 5.6 of Dembo and Zeitouni (1998), where b and σ are assumed to be globally Lipschitz continuous and σ is bounded. In Baldi and Caramellino (2011), building on an earlier work by Baldi and Chaleyat-Maurel (1988), which in turn improves on results obtained in Priouret (1982), the large deviation principle with rate function given by (3) is established for locally Lipschitz continuous coefficients b and σ satisfying a sublinear growth condition. The result in Baldi and Caramellino (2011) is actually more general; see our discussion in Section 4. The three works just mentioned all use a method of proof due to Azencott (1980). The idea is to show that, when $\sqrt{\varepsilon}W$ is close to a path $\psi = \int_0^\cdot f_t \, \mathrm{d}t$, where $f \in L^2([0, T], \mathbb{R}^m)$, then the probabilities that X^ε deviates from the solution φ to the integral equation

$$\varphi_t = x + \int_0^t b(\varphi_s) \, \mathrm{d}s + \int_0^t \sigma(\varphi_s) f_s \, \mathrm{d}s, \qquad t \in [0, T], \tag{4}$$

are exponentially small in ε. This can be interpreted as a quasicontinuity property of the Itô solution map associated with b and σ. To verify the quasicontinuity property, assuming that (4) is well posed given any 'control' $f \in L^2([0, T], \mathbb{R}^m)$, one first establishes, using a discretization argument applied to (1) and exponential martingale inequalities, the quasicontinuity property for the zero control and time-dependent drift coefficients; the estimate is then transferred to controls in L^2 and the original coefficients using a change of measure based on Girsanov's theorem.

In this paper we study small noise large deviations for possibly degenerate Itô stochastic differential equations with coefficients b and σ that may depend on time and the past of the solution trajectory (predictable coefficients) as well as on the large deviation parameter ε;

cf. (5) below. This general setting has also been studied in Puhalskii (2004). The proof of the large deviation principle there is based on Puhalskii's weak convergence approach to large deviations, which builds on idempotent probability theory and convergence in terms of maxingale problems, the idempotent analogues of martingale problems; see Puhalskii (2001). The assumptions needed in Puhalskii (2004) to establish the large deviation principle are very mild, the main assumption being that Luzin weak uniqueness holds for the idempotent Itô stochastic differential equation associated with the predictable coefficients b and σ; sufficient conditions in terms of regularity and growth properties of b and σ are provided.

The approach we follow here in establishing the large deviation principle, actually through the Laplace principle, is the weak convergence approach introduced in Dupuis and Ellis (1997) and adapted to the study of stochastic systems driven by finite-dimensional Brownian motion in Boué and Dupuis (1998). The approach, or more precisely the variational formula for Laplace functionals which is its starting point, has been extended to stochastic systems driven by infinite-dimensional Brownian motion and/or a Poisson random measure in Budhiraja and Dupuis (2000) and Budhiraja *et al.* (2008), (2011). Using that approach in the present situation, it is straightforward to prove the large deviation principle for solutions of (1) when the coefficients are globally Lipschitz continuous; see Section 4.2 of Boué and Dupuis (1998) or, for the case of finite-dimensional jump diffusions, Section 4.1 of Budhiraja *et al.* (2011). Here, we obtain the large deviation principle for predictable coefficients under much weaker hypotheses, which can be summarized as follows: continuity of the coefficients in the state variable; strong existence and uniqueness for the (stochastic) prelimit equations; uniqueness for a controlled version of the (deterministic) limit equation; and stability of the prelimit solutions under L^2-bounded perturbations in terms of tightness of laws. An advantage of the weak convergence method is that the large deviation principle can be derived in a unified way under mild conditions with no need for resorting to discretization arguments or exponential probability estimates. Instead, we use ordinary tightness and weak convergence for a family of controlled versions of the original processes.

The rest of this paper is organized as follows. Section 2 is dedicated to the statement and proof of the large deviation principle under general hypotheses. In Section 3 we verify the hypotheses for coefficients that are locally Lipschitz continuous with sublinear growth at ∞ but that may depend on the past as well as the large deviation parameter ε. This result yields, as an application, the large deviation principle obtained in Mohammed and Zhang (2006) for a class of systems with memory or delay; see Section 4.1. The approach used here actually allows us to easily handle more general delay models than the point delay studied in Mohammed and Zhang (2006); cf. Remark 7 below. In Section 4.2 we derive the large deviation principle for a class of positive Itô diffusions with dispersion coefficient of square-root type, such as, for example, the Cox–Ingersoll–Ross (CIR) process, which serves as a model for interest rates in mathematical finance. The result is essentially the same as the large deviation principle for positive diffusions obtained in Baldi and Caramellino (2011); it might be compared to Theorem 1.3 in Donati-Martin *et al.* (2004), which also covers degenerate cases (zero initial condition or drift vanishing in zero). Appendix A contains the variational formula for Laplace functionals of Brownian motion obtained in Boué and Dupuis (1998) as well as two related technical results.

2. General large deviation principle

Let $d, m \in \mathbb{N}$, and let $T > 0$. For $n \in \mathbb{N}$, set $\mathcal{W}^n := \mathbf{C}([0, T], \mathbb{R}^n)$ and endow $\mathcal{W}^n$ with the standard topology of uniform convergence. For $\varepsilon > 0$, let b_ε and b be functions mapping $[0, T] \times \mathcal{W}^d$ to $\mathbb{R}^d$, and let σ_ε and σ be functions mapping $[0, T] \times \mathcal{W}^d$ to $\mathbb{R}^{d \times m}$.

Let $(\mathcal{W}^m, \mathcal{B}, \theta)$ be the canonical probability space with Wiener measure θ, and let W be the coordinate process. Thus, W is an m-dimensional standard Brownian motion with respect to θ. Let $(\mathcal{G}_t)$ be the θ-augmented filtration generated by W, and let $\mathcal{M}^2[0, T]$ denote the space of $\mathbb{R}^m$-valued, square-integrable, $(\mathcal{G}_t)$-predictable processes.

Fix $x \in \mathbb{R}^d$. For $\varepsilon > 0$, we consider the Itô stochastic differential equation

$$dX_t^\varepsilon = b_\varepsilon(t, X^\varepsilon)\, dt + \sqrt{\varepsilon}\sigma_\varepsilon(t, X^\varepsilon)\, dW_t, \tag{5}$$

and, with $v \in \mathcal{M}^2[0, T]$ its controlled counterpart

$$dX_t^{\varepsilon, v} = b_\varepsilon(t, X^{\varepsilon, v})\, dt + \sigma_\varepsilon(t, X^{\varepsilon, v})v_t\, dt + \sqrt{\varepsilon}\sigma_\varepsilon(t, X^{\varepsilon, v})\, dW_t, \tag{6}$$

both over the time interval $[0, T]$ and with initial condition $X_0^{\varepsilon, v} = X_0^\varepsilon = x$. Observe that if $\varepsilon = 0$ then (5) becomes a deterministic functional equation, namely,

$$\varphi_t = x + \int_0^t b(s, \varphi)\, ds.$$

Similarly, if $\varepsilon = 0$ and we pick $v = f \in L^2([0, T]; \mathbb{R}^m)$, then (6) reduces to

$$\varphi_t = x + \int_0^t b(s, \varphi)\, ds + \int_0^t \sigma(s, \varphi) f_s\, ds. \tag{7}$$

Let us introduce the following hypotheses.

(H1) The coefficients b and σ are predictable. Moreover, $b(t, \cdot)$ and $\sigma(t, \cdot)$ are uniformly continuous on compact subsets of $\mathcal{W}^d$, uniformly in $t \in [0, T]$, and $t \mapsto \sigma(t, \varphi)$ is in $L^2([0, T]; \mathbb{R}^d)$ for any $\varphi \in \mathcal{W}^d$.

(H2) The coefficients b_ε and σ_ε are predictable maps such that $b_\varepsilon \to b$ and $\sigma_\varepsilon \to \sigma$ as $\varepsilon \to 0$ uniformly on $[0, T] \times \mathcal{W}^d$.

(H3) For all sufficiently small $\varepsilon > 0$, pathwise uniqueness and existence in the strong sense hold for (5).

(H4) For any $f \in L^2([0, T]; \mathbb{R}^m)$, (7) has a unique solution so that the map

$$\Gamma_x : L^2([0, T]; \mathbb{R}^m) \to \mathcal{W}^d$$

which takes $f \in L^2[0, T]$ to the solution of (7) is well defined.

(H5) For all $N \in \mathbb{N}$, the map Γ_x is continuous when restricted to

$$S_N := \left\{ f \in L^2([0, T], \mathbb{R}^m) : \int_0^T |f_s|^2 ds \le N \right\}$$

endowed with the weak topology of $L^2[0, T]$.

(H6) If $\{\varepsilon_n\} \subset (0, 1]$ is such that $\varepsilon_n \to 0$ as $n \to \infty$ and $\{v_n\}_{n \in \mathbb{N}} \subset \mathcal{M}^2[0, T]$ is such that, for some constant $N > 0$,

$$\sup_{n \in \mathbb{N}} \int_0^T |v_s^n(\omega)|^2\, ds \le N \quad \text{for } \theta\text{-almost all } \omega \in \mathcal{W}^m,$$

then $\{X^{\varepsilon_n, v_n}\}_{n \in \mathbb{N}}$ is tight as a family of $\mathcal{W}^d$-valued random variables and

$$\sup_{n \in \mathbb{N}} \int_0^T \mathbb{E}[|\sigma(s, X^{\varepsilon_n, v_n})|^2]\, ds < \infty.$$

Remark 1. We shall see in Section 3 that hypothesis (H2) can be weakened. Specifically, we shall require uniform convergence of b_ε and σ_ε to b and σ, respectively, only on bounded subsets of $\mathcal{W}^d$.

Remark 2. As will be clear from the proof of Theorem 1, existence of solutions to (7) is a consequence of hypotheses (H1)–(H3) and (H6). Thus, hypothesis (H4) reduces to the requirement of uniqueness of solutions for the deterministic integral equation (7).

Remark 3. The spaces $S_N := \{f \in L^2([0, T], \mathbb{R}^m) \colon \int_0^T |f_s|^2 \, ds \leq N\}$, $N \in \mathbb{N}$, introduced in hypothesis (H5) are compact Polish spaces when endowed with the weak topology of $L^2[0, T]$. Continuity of the restriction of Γ_x to S_N as required by (H5) is only needed to guarantee that the rate function has compact sublevel sets, and accordingly is good.

Theorem 1. *Assume that (H1)–(H6) hold. Then the family $\{X^\varepsilon\}_{\varepsilon>0}$ of solutions to the stochastic differential equation (5) with initial condition $X_0^\varepsilon = x$ satisfies the Laplace principle with good rate function $I_x \colon \mathcal{W}^d \to [0, \infty]$ given by*

$$I_x(\varphi) = \inf_{\{f \in L^2([0,T];\mathbb{R}^m) \colon \Gamma_x(f)=\varphi\}} \frac{1}{2} \int_0^T |f_t|^2 \, dt$$

whenever $\{f \in L^2([0, T]; \mathbb{R}^m) \colon \Gamma_x(f) = \varphi\} \neq \varnothing$, and $I_x(\varphi) = \infty$ otherwise.

Proof. Lower bound. The first step in proving Theorem 1 is the Laplace principle lower bound. We have to show that, for any bounded and continuous function $F \colon \mathcal{W}^d \to \mathbb{R}$,

$$\liminf_{\varepsilon \to 0+} -\varepsilon \log \mathbb{E}[\mathrm{e}^{-F(X^\varepsilon)/\varepsilon}] \geq \inf_{\varphi \in \mathcal{W}^d} \{F(\varphi) + I_x(\varphi)\}.$$

It suffices to prove that any sequence $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset (0, 1]$ such that $\varepsilon_n \to 0$ as $n \to \infty$ has a subsequence for which the above limit relation holds.

Let $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset (0, 1]$ be such that $\varepsilon_n \to 0$. By hypothesis (H3), for any $n \in \mathbb{N}$, $X^n := X^{\varepsilon_n}$ is a strong solution of (5). Hence, there exists a measurable map $h^n \colon \mathcal{W}^m \to \mathcal{W}^d$ such that $X^n = h^n(W)$, θ-almost surely. Representation formula (27) in Appendix A applies and yields

$$-\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F(X^n)/\varepsilon_n}] = -\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F\circ h^n(W)/\varepsilon_n}]$$

$$= \varepsilon_n \inf_{v\in\mathcal{M}^2[0,T]} \mathbb{E}\left[\frac{1}{2} \int_0^T |v_s|^2 \, ds + \frac{1}{\varepsilon_n} F \circ h^n\left(W + \int_0^\cdot v_s \, ds\right)\right]$$

$$= \inf_{v\in\mathcal{M}^2[0,T]} \mathbb{E}\left[\frac{1}{2} \int_0^T |v_s|^2 \, ds + F \circ h^n\left(W + \frac{1}{\sqrt{\varepsilon_n}} \int_0^\cdot v_s \, ds\right)\right]. \quad (8)$$

Fix $\delta > 0$. We claim that there exists a constant $N > 0$ such that, for every $n \in \mathbb{N}$, there exists $v^n \in \mathcal{M}^2[0, T]$ such that $\int_0^T |v_s^n|^2 \, ds \leq N$ and

$$-\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F(X^n)/\varepsilon_n}] \geq \mathbb{E}\left[\frac{1}{2} \int_0^T |v_s^n|^2 \, ds + F \circ h^n\left(W + \frac{1}{\sqrt{\varepsilon_n}} \int_0^\cdot v_s^n \, ds\right)\right] - \delta. \quad (9)$$

Indeed, by the definition of the infimum, for any $n \in \mathbb{N}$, there exists $u^n \in \mathcal{M}^2[0, T]$ such that

$$-\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F(X^n)/\varepsilon_n}] \geq \mathbb{E}\left[\frac{1}{2} \int_0^T |u_s^n|^2 \, ds + F \circ h^n\left(W + \frac{1}{\sqrt{\varepsilon_n}} \int_0^\cdot u_s^n \, ds\right)\right] - \frac{\delta}{2}.$$

Setting $M := \|F\|_\infty$, it follows that

$$\sup_{n \in \mathbb{N}} \mathbb{E}\left[\frac{1}{2} \int_0^T |u_s^n|^2 \, \mathrm{d}s\right] \le 2M + \frac{\delta}{2} < \infty. \tag{10}$$

For $N \in \mathbb{N}$, define the stopping time

$$\tau_N^n := \inf\left\{t \in [0, T]: \int_0^t |u_s^n|^2 \, \mathrm{d}s \ge N\right\} \wedge T.$$

The processes $u_s^{n,N} := u_s^n \mathbf{1}_{[0, \tau_N^n]}(s)$ belong to $\mathcal{M}^2[0, T]$ with $\int_0^T |u_s^{n,N}|^2 \, \mathrm{d}s \le N$. By the Chebychev inequality and (10),

$$\theta(u^n \ne u^{n,N}) \le \theta\left(\int_0^T |u_s^n|^2 \, \mathrm{d}s \ge N\right) \le \frac{4M + \delta}{N}.$$

This observation implies that

$$-\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F(X^n)/\varepsilon_n}]$$
$$\ge \mathbb{E}\left[\frac{1}{2} \int_0^T |u_s^{n,N}|^2 \, \mathrm{d}s + F \circ h^n\left(W + \frac{1}{\sqrt{\varepsilon_n}} \int_0^\cdot u_s^{n,N} \, \mathrm{d}s\right)\right] - \frac{2M(4M + \delta)}{N} - \frac{\delta}{2}. \tag{11}$$

In view of (11), to verify the claim, we take N large enough so that

$$\frac{2M(4M + \delta)}{N} < \frac{\delta}{2}$$

and set, for $n \in \mathbb{N}$, $v^n := u^{n,N}$.

Choose N and $\{v^n\} \subset \mathcal{M}^2[0, T]$ according to the claim, $\delta > 0$ being fixed. Thanks to hypothesis (H3) and Lemma 1 in Appendix A, the controlled stochastic equation

$$\mathrm{d}X_t^{n,v^n} = b_{\varepsilon_n}(t, X^{n,v^n}) \, \mathrm{d}t + \sigma_{\varepsilon_n}(t, X^{n,v^n}) v_t^n \, \mathrm{d}t + \sqrt{\varepsilon_n} \sigma_{\varepsilon_n}(t, X^{n,v^n}) \, \mathrm{d}W_t$$

possesses a unique strong solution with $X_0^{n,v^n} = x$, and

$$h^n\left(W + \frac{1}{\sqrt{\varepsilon_n}} \int_0^\cdot v_s^n \, \mathrm{d}s\right) = X^{n,v^n} \quad \theta \text{ almost everywhere.}$$

It follows that, for any $n \in \mathbb{N}$, we can rewrite (9) to obtain

$$-\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F(X^n)/\varepsilon_n}] \ge \mathbb{E}\left[\frac{1}{2} \int_0^T |v_s^n|^2 \, \mathrm{d}s + F(X^{n,v^n})\right] - \delta,$$

where X^{n,v^n} is the unique strong solution of (6) with $\varepsilon = \varepsilon_n$ and control $v = v_n$.

Next we check that $\{(X^{n,v^n}, v^n)\}_{n \in \mathbb{N}}$ is tight as a family of random variables with values in $\mathcal{W}^d \times S_N$. Since both S_N and $\mathcal{W}^d$ are Polish spaces, it suffices to show that $\{X^{n,v^n}\}_{n \in \mathbb{N}}$ is tight as a family of $\mathcal{W}^d$-valued random variables and that $\{v^n\}_{n \in \mathbb{N}}$ is tight as a family of S_N-valued random variables. But tightness of $\{X^{n,v^n}\}_{n \in \mathbb{N}}$ follows by hypothesis (H6), while tightness of $\{v^n\}$ is automatic since S_N is compact. Therefore, possibly taking a subsequence,

(X^{n,v^n}, v^n) converges in distribution to a $(\mathcal{W}^d \times S_N)$-valued random variable (X, v) defined on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Let us denote by $\mathbb{E}_\mathrm{P}$ the expectation with respect to the measure P. We will show that X satisfies

$$X_t = x + \int_0^t b(s, X)\,\mathrm{d}s + \int_0^t \sigma(s, X)v_s\,\mathrm{d}s \quad \mathrm{P}\text{-almost surely.} \tag{12}$$

To this end, for $t \in [0, T]$, consider the map $\Psi_t \colon \mathcal{W}^d \times S_N \to \mathbb{R}$ defined by

$$\Psi_t(\varphi, f) := \left| \varphi(t) - x - \int_0^t b(s, \varphi(s))\,\mathrm{d}s - \int_0^t \sigma(s, \varphi(s))f_s\,\mathrm{d}s \right| \wedge 1.$$

Clearly, Ψ_t is bounded. Moreover, Ψ_t is continuous. Indeed, let $\varphi^n \to \varphi$ in $\mathcal{W}^d$ and $f^n \to f$ in S_N with respect to the weak topology of L^2. The set $\mathcal{C} := \{\varphi^n \colon n \in \mathbb{N}\} \cup \{\varphi\}$ is a compact subset of $\mathcal{W}^d$. Therefore, by hypothesis (H1), there exist moduli of continuity ρ_b and ρ_σ mapping $[0, \infty)$ into $[0, \infty)$ such that $|b(s, \varphi) - b(s, \psi)| \le \rho_b(\|\varphi - \psi\|_\infty)$ and $|\sigma(s, \varphi) - \sigma(s, \psi)| \le \rho_\sigma(\|\varphi - \psi\|_\infty)$ for all $s \in [0, T]$ and all $\varphi, \psi \in \mathcal{C}$. Using Hölder's inequality and the fact that $\|f^n\|_{L^2} \le \sqrt{N}$, we find that

$$\begin{aligned}
|\Psi_t(\varphi^n, f^n) &- \Psi_t(\varphi, f)| \\
&\le |\varphi_t^n - \varphi_t| + \int_0^t |b(s, \varphi^n) - b(s, \varphi)|\,\mathrm{d}s + \int_0^t |\sigma(s, \varphi) - \sigma(s, \varphi^n)||f_s^n|\,\mathrm{d}s \\
&\quad + \left| \int_0^t \sigma(s, \varphi)(f - f_s^n)\,\mathrm{d}s \right| \\
&\le \|\varphi^n - \varphi\|_\infty + T\rho_b(\|\varphi^n - \varphi\|_\infty) + \sqrt{NT}\rho_\sigma(\|\varphi^n - \varphi\|_\infty) \\
&\quad + \left| \int_0^t \sigma(s, \varphi)(f_s - f_s^n)\,\mathrm{d}s \right|.
\end{aligned}$$

The terms involving $\|\varphi - \varphi^n\|_\infty$ in the above display go to 0 as $n \to \infty$. Thanks to hypothesis (H1), the function $\sigma(\cdot, \varphi)$ is in $L^2[0, T]$; since f^n converges weakly to f, the rightmost term of the previous display goes to 0 as well. This shows that Ψ_t is continuous. Since (X^{n,v^n}, v^n) converges in distribution to (X, v) and Ψ_t is bounded and continuous, the continuous mapping theorem for weak convergence implies that

$$\lim_{n \to \infty} \mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] = \mathbb{E}_\mathrm{P}[\Psi_t(X, v)]. \tag{13}$$

If we show that the limit in (13) is actually 0 then, by the definition of Ψ_t, X will satisfy (12) P-almost surely for all $t \in [0, T]$. Since X has continuous paths, it follows that X satisfies (12) for all $t \in [0, T]$ P-almost surely. Observe that

$$\begin{aligned}
\mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] &\le \mathbb{E}\left[\int_0^t |b_{\varepsilon_n}(s, X^{n,v^n}) - b(s, X^{n,v^n})|\,\mathrm{d}s \right] \\
&\quad + \mathbb{E}\left[\int_0^t |\sigma_{\varepsilon_n}(s, X^{n,v^n}) - \sigma(s, X^{n,v^n})||v_s^n|\,\mathrm{d}s \right] \\
&\quad + \sqrt{\varepsilon_n}\,\mathbb{E}\left[\left| \int_0^t \sigma_{\varepsilon_n}(s, X^{n,v^n})\,\mathrm{d}W_s \right| \right].
\end{aligned}$$

Using the uniform convergence of σ_ε to σ and the uniform convergence of b_ε to b on $[0, T] \times \mathcal{W}^d$ according to (H2), we obtain

$$\mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] \leq t\|b_{\varepsilon_n} - b\|_\infty + \|\sigma_{\varepsilon_n} - \sigma\|_\infty \mathbb{E}\left[\int_0^T |v_s^n|\, \mathrm{d}s\right]$$

$$+ \sqrt{\varepsilon_n}\sqrt{\int_0^t \mathbb{E}[|\sigma_{\varepsilon_n}(s, X^{n,v^n})|^2]\, \mathrm{d}s},$$

which goes to 0 as $n \to \infty$. The last term in the above display tends to 0 since

$$\sup_{n\in\mathbb{N}} \int_0^t \mathbb{E}[|\sigma_{\varepsilon_n}(s, X^{n,v^n})|^2]\, \mathrm{d}s \leq 2\sup_{n\in\mathbb{N}} \int_0^T \mathbb{E}[|\sigma(s, X^{n,v^n})|^2]\, \mathrm{d}s$$

$$+ 2\sup_{n\in\mathbb{N}} \int_0^T \mathbb{E}[|\sigma_{\varepsilon_n}(s, X^{n,v^n}) - \sigma(s, X^{n,v^n})|^2]\, \mathrm{d}s$$

$$\leq 2T\sup_{n\in\mathbb{N}}\|\sigma_{\varepsilon_n} - \sigma\|_\infty^2 + 2\sup_{n\in\mathbb{N}} \int_0^T \mathbb{E}[|\sigma(s, X^{n,v^n})|^2]\, \mathrm{d}s,$$

which is finite thanks to hypothesis (H6) (and (H2)). Recalling (13), we have shown that

$$\lim_{n\to\infty} \mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] = \mathbb{E}_\mathrm{P}[\Psi_t(X, v)] = 0.$$

Thus, X satisfies (12) for all $t \in [0, T]$ P-almost surely. If $f \in L^2([0, T]; \mathbb{R}^m)$ then applying the same argument to the (constant) sequence of deterministic control processes $v^n = f$, we find that (7) possesses a solution. The existence part of hypothesis (H4) is therefore a consequence of hypotheses (H1), (H2), (H3), and (H6).

The mapping $S_N \ni f \to \int_0^T |f_s|^2\, \mathrm{d}s \in \mathbb{R}$ is nonnegative and lower semicontinuous (with respect to the weak L^2-topology on S_N). Since the trajectories of v^n are in S_N for all $n \in \mathbb{N}$ and v^n converges in distribution to v, a version of Fatou's lemma (Theorem A.3.12 of Dupuis and Ellis (1997, p. 307)) entails that

$$\liminf_{n\to\infty} \mathbb{E}\left[\int_0^T |v_s^n|^2\, \mathrm{d}s\right] \geq \mathbb{E}_\mathrm{P}\left[\int_0^T |v_s|^2\, \mathrm{d}s\right].$$

Using this inequality and the continuous mapping theorem (recalling that F is bounded and continuous), we find that

$$\liminf_{n\to\infty} -\varepsilon_n \log \mathbb{E}[\mathrm{e}^{-F(X^n)/\varepsilon_n}] \geq \liminf_{n\to\infty} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_s^n|^2\, \mathrm{d}s + F(X^{n,v^n})\right] - \delta$$

$$\geq \mathbb{E}_\mathrm{P}\left[\frac{1}{2}\int_0^T |v_s|^2\, \mathrm{d}s + F(X^v)\right] - \delta$$

$$\geq \inf_{\{(f,\varphi)\in L^2\times\mathcal{W}^d:\, \varphi=\Gamma_x(f)\}}\left\{\frac{1}{2}\int_0^T |f_s|^2\, \mathrm{d}s + F(\varphi)\right\} - \delta$$

$$\geq \inf_{\varphi\in\mathcal{W}^d}\{I_x(\varphi) + F(\varphi)\} - \delta.$$

The second but last inequality is obtained by evaluating the random variable inside the expectation ω by ω. Since δ was arbitrary, the lower bound follows.

Upper bound. We now prove the Laplace principle upper bound,

$$\limsup_{\varepsilon \to 0} -\varepsilon \log \mathbb{E}[e^{-F(X^\varepsilon)/\varepsilon}] \le \inf_{\varphi \in \mathcal{W}^d} \{I_x(\varphi) + F(\varphi)\} \tag{14}$$

for $F \colon \mathcal{W}^d \to \mathbb{R}$ bounded and continuous. As for the lower bound, it suffices to show that any sequence $\{\varepsilon_n\}_{n \in \mathbb{N}} \subset (0, 1]$ such that $\varepsilon_n \to 0$ has a subsequence for which the limit in (14) holds.

Fix $\delta > 0$. If the infimum in (14) is not finite, the inequality is trivially satisfied; hence, we may assume that the infimum is finite. Since F is bounded, there exists $\varphi \in \mathcal{W}^d$ such that

$$I_x(\varphi) + F(\varphi) \le \inf_{\psi \in \mathcal{W}^d} \{I_x(\psi) + F(\psi)\} + \frac{\delta}{2} < \infty. \tag{15}$$

For such φ, choose $\tilde{v} \in L^2([0, T]; \mathbb{R}^m)$ such that

$$\frac{1}{2} \int_0^T |\tilde{v}_s|^2 \, ds \le I_x(\varphi) + \frac{\delta}{2},$$

and $\varphi = \Gamma_x(\tilde{v})$. This choice is possible by the definition of I_x and since $I_x(\varphi) < \infty$. Let $\{\varepsilon_n\}_{n \in \mathbb{N}} \subset (0, 1]$ be such that $\varepsilon_n \to 0$ as $n \to \infty$. For $n \in \mathbb{N}$, let $X^{n,\tilde{v}}$ be the unique strong solution of (6) with $\varepsilon = \varepsilon_n$ and (deterministic) control $v = \tilde{v}$. Then the family $\{(X^{n,\tilde{v}}, \tilde{v})\}_{n \in \mathbb{N}}$ is tight. Therefore, possibly taking a subsequence, $(X^{n,\tilde{v}}, \tilde{v})$ converges in distribution to a random variable $(X, \tilde{v})$ defined on some probability space $(\Omega, \mathcal{F}, P)$. As in the proof of the lower bound, it follows that, P-almost surely,

$$X_t = x + \int_0^t b(s, X) \, ds + \int_0^t \sigma(s, X)\tilde{v}_s \, ds \quad \text{for all } t \in [0, T].$$

The above integral equation, which is deterministic since $\tilde{v} \in L^2([0, T]; \mathbb{R}^m)$ is deterministic, coincides with (7). The solution to that equation is unique by hypothesis (H4); hence, $X = \Gamma_x(\tilde{v}) = \varphi$, P-almost surely. Using representation (8), we obtain

$$\limsup_{n \to \infty} -\varepsilon_n \log \mathbb{E}[e^{-F(X^{\varepsilon_n})/\varepsilon_n}]$$

$$= \limsup_{n \to \infty} \inf_{v \in \mathcal{M}^2[0,T]} \mathbb{E}\left[\frac{1}{2} \int_0^T |v_s|^2 \, ds + F \circ h^n\left(W + \frac{1}{\sqrt{\varepsilon_n}} \int_0^\cdot v_s \, ds\right) \right]$$

$$\le \limsup_{n \to \infty} \mathbb{E}\left[\frac{1}{2} \int_0^T |\tilde{v}_s|^2 \, ds + F(X^{n,\tilde{v}}) \right]$$

$$= \frac{1}{2} \int_0^T |\tilde{v}_s|^2 \, ds + \lim_{n \to \infty} \mathbb{E}[F(X^{n,\tilde{v}})]$$

$$\le I_x(\varphi) + \frac{\delta}{2} + \lim_{n \to \infty} \mathbb{E}[F(X^{n,\tilde{v}})].$$

Since F is bounded and continuous, and $X^{n,\tilde{v}}$ converges in distribution to $X = \varphi$, we have $\lim_{n \to \infty} \mathbb{E}[F(X^{n,\tilde{v}})] = F(\varphi)$. Thanks to (15), we can end the above chain of inequalities by

$$I_x(\varphi) + \frac{\delta}{2} + F(\varphi) \le \inf_{\psi \in \mathcal{W}^d} \{I_x(\psi) + F(\psi)\} + \delta.$$

Since $\delta > 0$ is arbitrary, the proof of the Laplace principle upper bound is complete.

Goodness of the rate function. To prove that I_x is actually a good rate function, it remains to check that I_x has compact sublevel sets. This follows from the compactness of S_N for any $N > 0$, and by the continuity on these sets of the map Γ_x, which takes v to the unique solution of (7), according to hypothesis (H5). Indeed, $\{\varphi \in \mathcal{W}^d : I_x(\varphi) \leq N\} = \bigcap_{\varepsilon > 0} \Gamma_x(S_{N+\varepsilon})$ is the intersection of compact sets, and is hence compact.

3. Locally Lipschitz continuous coefficients

In this section we show that hypotheses (H1)–(H6) hold in the important case of locally Lipschitz continuous coefficients which are predictable and satisfy a sublinear growth condition. With the notation of Section 2, let us introduce the following assumptions.

(A1) The functions b and σ satisfy a sublinear growth condition. Specifically, there exists $M > 0$ such that, for all $t \in [0, T]$ and all $\varphi \in \mathcal{W}^d$,

$$|b(t, \varphi)| \vee |\sigma(t, \varphi)| \leq M\left(1 + \sup_{s \in [0,t]} |\varphi_s|\right).$$

(A2) The functions b and σ are locally Lipschitz continuous. Specifically, for any $R > 0$, there exists $L_R > 0$ such that, for all $t \in [0, T]$ and all $\varphi, \tilde{\varphi} \in \mathcal{W}^d$ with $\sup_{s \in [0,t]} |\varphi_s| \vee |\tilde{\varphi}_s| \leq R$,

$$|b(t, \varphi) - b(t, \tilde{\varphi})| \vee |\sigma(t, \varphi) - \sigma(t, \tilde{\varphi})| \leq L_R \sup_{s \in [0,t]} |\varphi_s - \tilde{\varphi}_s|.$$

(A3) The functions b_ε and σ_ε enjoy property (A1) (with the same constant M as b and σ) as well as property (A2).

(A4) The coefficients b_ε and σ_ε converge as $\varepsilon \to 0$ to b and σ, respectively, uniformly on bounded subsets of $[0, T] \times \mathcal{W}^d$.

Remark 4. We distinguish between hypotheses (A1)–(A2) and (A3) since (A3) is not needed to verify (H4) and (H5). Observe that (A4) is not exactly (H2); indeed, the convergence is not on the whole $\mathcal{W}^d$, but on the bounded subsets of $\mathcal{W}^d$.

Remark 5. Assumption (A2) implies that if $0 \leq t \leq T$ and $\varphi, \psi \in \mathcal{W}^d$ are such that $\varphi_s = \psi_s$ for all $s \in [0, t]$, then $b(t, \varphi) = b(t, \psi)$ and the process $\{b(t, \cdot)\}_{t \geq 0}$ is adapted to the canonical filtration. In particular, if $\{b(t, \cdot)\}_{t \geq 0}$ is càdlàg then b is also predictable. The same remark is also true for σ.

Theorem 2. *Assume that (A1)–(A4) hold. Then the family $\{X^\varepsilon\}_{\varepsilon > 0}$ of solutions to the stochastic differential equation (5) with initial condition $X_0^\varepsilon = x$ satisfies the Laplace principle with good rate function $I_x \colon \mathcal{W}^d \to [0, \infty]$ given by*

$$I_x(\varphi) = \inf_{\{f \in L^2([0,T];\mathbb{R}^m) : \varphi = \Gamma_x(f)\}} \frac{1}{2} \int_0^T |f_t|^2 \, \mathrm{d}t$$

whenever $\{f \in L^2([0, T]; \mathbb{R}^m) : \varphi = \Gamma_x(f)\} \neq \varnothing$, and $I_x(\varphi) = \infty$ otherwise.

Proof. It is enough to show that hypotheses (H1)–(H6) of Theorem 1 are entailed by assumptions (A1)–(A4). As mentioned above, we will not be able to prove (H2). Instead, we show that in this special setting (H2) is not really needed; this discussion is postponed to the end of the section.

Hypotheses (H1) and (H3). Hypothesis (H1) is satisfied; in fact, $b(t, \cdot)$ and $\sigma(t, \cdot)$ are uniformly continuous on bounded subsets of $\mathcal{W}^d$, uniformly in $t \in [0, T]$ because of assumption (A2). Moreover, $\sigma(\cdot, \varphi)$ belongs to $L^2[0, T]$ for any $\varphi \in \mathcal{W}^d$ since

$$\sup_{t \in [0,T]} |\sigma(t, \varphi)|^2 \le 2M^2(1 + \|\varphi\|_\infty^2)$$

as a consequence of (A1). Assumption (A3) implies that pathwise uniqueness and existence of strong solutions hold for (5); see, for instance, Theorem 12.1 of Rogers and Williams (2000, p. 132).

Hypotheses (H4) and (H5). In view of Remark 2, to verify (H4), it suffices to show that, given any $f \in L^2([0, T], \mathbb{R}^m)$, there is a unique solution $\varphi \in \mathbf{C}([0, T], \mathbb{R}^d)$ of (7). Moreover, for φ, we have the growth estimate

$$\sup_{0 \le s \le t} |\varphi_t|^2 \le (3|x|^2 + 6M^2t^2 + 6M^2t\|f\|^2)e^{6M^2t(t+\|f\|^2)}, \qquad t \in [0, T]. \tag{16}$$

To verify that uniqueness holds, let $\varphi, \psi \in \mathbf{C}([0, T], \mathbb{R}^d)$ be solutions of (7). Then, for $t \in [0, T]$,

$$|\varphi_t - \psi_t| \le \int_0^t |b(s, \varphi) - b(s, \psi)|\, \mathrm{d}s + \int_0^t |\sigma(s, \varphi) - \sigma(s, \psi)||f_s|\, \mathrm{d}s.$$

By taking the square, using Hölder's inequality and the local Lipschitz continuity according to (A2), we obtain, for large enough $R > 0$ (since φ and ψ are bounded),

$$|\varphi_t - \psi_t|^2 \le 2L_R^2(T + \|f\|^2) \int_0^t \sup_{u \in [0,s]} |\varphi_u - \psi_u|^2\, \mathrm{d}s.$$

Gronwall's inequality now entails that $\|\varphi - \psi\|_\infty = 0$, which yields uniqueness. Similarly, also using the sublinear growth condition (A1), we find that

$$|\varphi_t|^2 \le 3|x|^2 + 3t \int_0^t |b(s, \varphi)|^2\, \mathrm{d}s + 3\left(\int_0^t |\sigma(s, \varphi)||f_s|\, \mathrm{d}s\right)^2$$

$$\le 3|x|^2 + 6M^2(t + \|f\|^2) \int_0^t \left(1 + \sup_{0 \le u \le s} |\varphi_u|^2\right) \mathrm{d}s$$

$$\le 3|x|^2 + 6M^2t^2 + 6M^2t\|f\|^2 + 6M^2(t + \|f\|^2) \int_0^t \sup_{0 \le u \le s} |\varphi_u|^2\, \mathrm{d}s.$$

An application of Gronwall's inequality now yields the growth estimate (16).

In order to establish (H5), we have to show that, given any $N \in \mathbb{N}$, the map Γ_x defined in (H4) is continuous when restricted to S_N. Recall that S_N is a compact Polish space. Take $\{f^n\} \subset S_N$ such that $f^n \to f$ weakly, and define $\varphi^n := \Gamma_x(f^n)$ and $\varphi := \Gamma_x(f)$. Then, for $t \in [0, T]$,

$$\varphi_t^n - \varphi_t = \int_0^t (b(s, \varphi^n) - b(s, \varphi))\, \mathrm{d}s + \int_0^t (\sigma(s, \varphi^n) - \sigma(s, \varphi)) f_s^n\, \mathrm{d}s$$

$$+ \int_0^t \sigma(s, \varphi)(f_s^n - f_s)\, \mathrm{d}s.$$

Since $\|f^n\|^2 \le N$, it follows from estimate (16) that $R := \sup_{n\in\mathbb{N}} \|\varphi^n\|_\infty \vee \|\varphi\|_\infty$ is finite. Therefore, using (A2),

$$\sup_{u\in[0,t]} |\varphi_u^n - \varphi_u| \le L_R \int_0^t \sup_{u\in[0,s]} |\varphi_u^n - \varphi_u| \, \mathrm{d}s$$

$$+ L_R \int_0^t \sup_{u\in[0,s]} |\varphi_u^n - \varphi_u||f_s^n| \, \mathrm{d}s + \sup_{u\in[0,T]} \left| \int_0^u \sigma(s,\varphi)(f_s^n - f_s) \, \mathrm{d}s \right|.$$

Set $\Delta_\sigma^n := \sup_{u\in[0,T]} |\int_0^u \sigma(s,\varphi)(f_s^n - f_s) \, \mathrm{d}s|$. By Hölder's inequality and since $\|f^n\|^2 \le N$ for all $n \in \mathbb{N}$, it follows that

$$\sup_{u\in[0,t]} |\varphi_u^n - \varphi_u|^2 \le 3L_R^2(t+N) \int_0^t \sup_{u\in[0,s]} |\varphi_u^n - \varphi_u|^2 \, \mathrm{d}s + 3(\Delta_\sigma^n)^2.$$

An application of Gronwall's lemma yields

$$\|\Gamma_x(f^n) - \Gamma_x(f)\|_\infty = \sup_{t\in[0,T]} |\varphi_t^n - \varphi_t|^2 \le 3(\Delta_\sigma^n)^2 \mathrm{e}^{3L^2T(T+N)}.$$

In order to establish continuity of Γ_x on S_N, it remains to check that Δ_σ^n goes to 0 as $n \to \infty$. Thanks to assumption (A1), the function $\sigma(\cdot,\varphi)$ is in $L^\infty[0,T]$. It follows that $\sigma(\cdot,\varphi)f^n$ converges weakly to $\sigma(\cdot,\varphi)f$ in L^2. Moreover, the family $\{\sigma(\cdot,\varphi)f^n\}_{n\in\mathbb{N}}$ is bounded in L^2 with respect to the L^2-norm. Hence,

$$\int_0^t \sigma(s,\varphi)f_s^n \, \mathrm{d}s \to \int_0^t \sigma(s,\varphi)f_s \, \mathrm{d}s \quad \text{as } n \to \infty,$$

uniformly in $t \in [0,T]$, which implies that $\Delta_\sigma^n \to 0$ as $n \to \infty$.

Hypothesis (H6). Let $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset (0,1]$ be such that $\varepsilon_n \to 0$ as $n \to \infty$, and let $\{v^n\}_{n\in\mathbb{N}} \subset \mathcal{M}^2[0,T]$ be such that, for some constant $N > 0$,

$$\sup_{n\in\mathbb{N}} \int_0^T |v_s^n(\omega)|^2 \, \mathrm{d}s < N \quad \text{for } \theta\text{-almost all } \omega \in \mathcal{W}^m.$$

For $n \in \mathbb{N}$, let X^{n,v_n} be the solution of (6) with $\varepsilon = \varepsilon_n$, control $v = v^n$, and initial condition x. Observe that if $\varepsilon \le 1$ then $\sqrt{\varepsilon}\sigma_\varepsilon$ has sublinear growth at ∞ with constant M, thanks to (A3). By Lemma 2 in Appendix A, it follows that, for all $p \ge 2$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{n,v_n}|^p \right] \le C(1 + |x|^p) \tag{17}$$

for some finite constant $C = C_p(T,N,M)$. Estimate (17), together with the sublinear growth at ∞ of σ (according to (A1)), implies in particular that

$$\sup_{n\in\mathbb{N}} \int_0^T \mathbb{E}[|\sigma(s, X^{n,v_n})|^2] \, \mathrm{d}s < \infty.$$

It remains to verify that the family $\{X^{n,v_n}\}_{n\in\mathbb{N}}$ is tight. In view of the Kolmogorov tightness criterion (see, for instance, Theorem 13.1.8 of Revuz and Yor (1999, pp. 517–518)), it suffices to show that there exist strictly positive constants α, β, and γ such that for all $t, s \in [0,T]$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_t^{n,v_n} - X_s^{n,v_n}|^\alpha] \le \beta |t-s|^{\gamma+1}.$$

Without loss of generality, let $s < t$. Set $K := 2^{p-1} M^p (T + C(1 + |x|^p))$. Exploiting the sublinear growth, we obtain, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
\mathbb{E}[|X_t^{n,v_n} - X_s^{n,v_n}|^p] &\leq 3^{p-1} (t-s)^{p-1} \mathbb{E}\left[\int_s^t |b_{\varepsilon_n}(u, X^{n,v_n})|^p \, du\right] \\
&\quad + 3^{p-1} \mathbb{E}\left[\left(\int_s^t |\sigma_{\varepsilon_n}(u, X^{n,v_n})||v^n| \, du\right)^p\right] \\
&\quad + 3^{p-1} (\varepsilon_n)^{p/2} \mathbb{E}\left[\left|\int_s^t \sigma_{\varepsilon_n}(u, X^{n,v_n}) \, dW_u\right|^p\right] \\
&\leq 3^{p-1} K((t-s)^p + N^{p/2}(t-s)^{p/2} + c_p(\varepsilon_n)^{p/2}(t-s)^{p/2}) \\
&\leq (t-s)^{p/2} 3^{p-1} K(T^{p/2} + N^{p/2} + (\varepsilon_n)^{p/2}).
\end{aligned}
$$

The hypotheses of Kolmogorov's criterion are therefore satisfied if we choose $p > 2$, and set $\alpha := p$, $\beta := 3^{p-1} K(T^{p/2} + N^{p/2} + 1)$, and $\gamma := p/2 - 1 > 0$.

Hypothesis (H2) modified. In the proof of Theorem 1, hypothesis (H2) is only needed to show that for all $t \in [0, T]$,

$$
\lim_{n \to \infty} \mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] = 0,
$$

where $\Psi_t \colon \mathcal{W}^d \times S_N \to \mathbb{R}$ is defined by

$$
\Psi_t(\varphi, f) := \left|\varphi_t - x - \int_0^t b(s, \varphi) \, ds - \int_0^t \sigma(s, \varphi) f_s \, ds\right| \wedge 1.
$$

We show that the same conclusion holds if we assume that (A3) and (A4) hold. Define $b_\varepsilon^R \colon [0, T] \times \mathcal{W}^d \to \mathbb{R}^d$ by

$$
b_\varepsilon^R(s, \varphi) := \begin{cases} b_\varepsilon(s, \varphi) & \text{if } \sup_{u \in [0,s]} |\varphi_u| \leq R, \\ b_\varepsilon\left(s, \dfrac{R}{\|\varphi\|_\infty} \varphi\right) & \text{otherwise.} \end{cases}
$$

In the same way, define σ_ε^R, σ^R, and b^R. It is clear that the functions just defined are globally Lipschitz and bounded. Thanks to assumption (A4), $b_\varepsilon^R \to b^R$ and $\sigma_\varepsilon^R \to \sigma^R$ uniformly on $[0, T] \times \mathcal{W}^d$. In analogy with Ψ_t, set

$$
\Psi_t^R(\varphi, f) := \left|\varphi_t - x - \int_0^t b^R(s, \varphi) \, ds - \int_0^t \sigma^R(s, \varphi) f_s \, ds\right| \wedge 1.
$$

Observe that if $\sup_{u \in [0,t]} |\varphi_u| \leq R$ then $\Psi_t^R(\varphi, f) = \Psi_t(\varphi, f)$. Now consider the family $\{X^{R,n}\}$ of solutions to the equation

$$
dX_t^{R,n} = b_{\varepsilon_n}^R(t, X^{R,n}) \, dt + \sigma_{\varepsilon_n}^R(t, X^{R,n}) v_t^n \, dt + \sqrt{\varepsilon_n} \sigma_{\varepsilon_n}^R(t, X^{R,n}) \, dW_t,
$$

with $X_0^{R,n} = x$. The same argument as used in the proof of Theorem 1 yields

$$
\lim_{n \to \infty} \mathbb{E}[\Psi_t^R(X^{R,n}, v^n)] = 0.
$$

For $R > 0$ and $n \in \mathbb{N}$, let τ_R^n denote the first exit time of X^{n,v^n} from the open ball of radius R centered at the origin. By the locality of the stochastic integral,

$$
\mathrm{P}(X_t^{R,n} = X_t^{n,v^n} \text{ for all } t \leq \tau_R^n) = 1.
$$

On the event $\{t < \tau_R^n\}$, we have $\Psi_t(X^{n,v^n}, v^n) = \Psi_t^R(X^{R,n}, v^n)$. It follows that

$$\mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] = \mathbb{E}[\mathbf{1}_{\{t < \tau_R^n\}}\Psi_t(X^{n,v^n}, v^n)] + \mathbb{E}[\mathbf{1}_{\{t \geq \tau_R^n\}}\Psi_t(X^{n,v^n}, v^n)]$$

$$\leq \mathbb{E}[\Psi_t^R(X^{R,n}, v^n)] + \mathrm{P}(t \geq \tau_R^n). \tag{18}$$

Using the sublinear growth condition and the estimate of Lemma 2, we find that, for all $n \in \mathbb{N}$,

$$\mathrm{P}(t \geq \tau_R^n) = \mathrm{P}\left(\sup_{0 \leq s \leq t} |X_t^{R,n}| \geq R\right) \leq \frac{C_2(T, N, M)(1 + |x|^2)}{R^2} =: \frac{C}{R^2}.$$

Taking upper limits on both sides of (18), we obtain

$$\limsup_{n \to \infty} \mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] \leq \limsup_{n \to \infty} \mathrm{P}(t \geq \tau_R^n) \leq \frac{C}{R^2}.$$

Since $R > 0$ has been chosen arbitrarily, it follows that

$$\lim_{n \to \infty} \mathbb{E}[\Psi_t(X^{n,v^n}, v^n)] = 0.$$

The job of assumption (H2) is therefore carried out by (A3) and (A4).

Example 1. (*Freidlin–Wentzell estimates.*) Let $\bar{b}$ and $\bar{\sigma}$ be measurable functions from $[0, T] \times \mathbb{R}^d$ to $\mathbb{R}^d$ and $\mathbb{R}^{d \times m}$, respectively. Assume that $\bar{b}$ and $\bar{\sigma}$ are locally Lipschitz continuous and satisfy a sublinear growth condition, uniformly in the time variable; that is, for every $R > 0$, there exists $L_R > 0$ such that, for all $t \in [0, T]$ and all $y, z \in \mathbb{R}^d$ with $|y|, |z| \leq R$,

$$|\bar{b}(t, y) - \bar{b}(t, z)| \leq L_R|y - z|, \qquad |\bar{\sigma}(t, y) - \bar{\sigma}(t, z)| \leq L_R|y - z|,$$

and there exists a constant $M > 0$ such that, for all $t \in [0, T]$, all $y \in \mathbb{R}^d$,

$$|\bar{b}(t, x)| \leq M(1 + |x|), \qquad |\bar{\sigma}(t, x)| \leq M(1 + |x|).$$

Let X^ε be the unique strong solution of the stochastic differential equation

$$\mathrm{d}X_t^\varepsilon = \bar{b}(t, X_t^\varepsilon)\,\mathrm{d}t + \sqrt{\varepsilon}\bar{\sigma}(t, X_t^\varepsilon)\,\mathrm{d}W_t$$

over the time interval $[0, T]$ with initial condition $X_0^\varepsilon = x$. Set $b(t, \varphi) := \bar{b}(t, \varphi_t)$ and $\sigma(t, \varphi) := \bar{\sigma}(t, \varphi_t)$. Then b and σ satisfy assumptions (A1)–(A4). By Theorem 2, the family $\{X^\varepsilon\}_{\varepsilon > 0}$ satisfies the large deviation principle with rate function $I_x \colon \mathcal{W}^d \to [0, \infty]$ given by

$$I_x(\varphi) = \inf_{\{f \in L^2([0,T];\mathbb{R}^m)\colon \varphi_t = x + \int_0^t \bar{b}(s, \varphi_s)\,\mathrm{d}s + \int_0^t \bar{\sigma}(s, \varphi_s)f_s\,\mathrm{d}s\}} \frac{1}{2}\int_0^T |f_t|^2\,\mathrm{d}t \tag{19}$$

whenever $\{f \in L^2([0, T]; \mathbb{R}^m)\colon \varphi_t = x + \int_0^t \bar{b}(s, \varphi_s)\,\mathrm{d}s + \int_0^t \bar{\sigma}(s, \varphi_s)f_s\,\mathrm{d}s\} \neq \varnothing$, and $I_x(\varphi) = \infty$ otherwise.

Remark 6. If $\bar{\sigma}$ is a square matrix such that $a(t, y) := \bar{\sigma}(t, y)\bar{\sigma}(t, y)^\mathsf{T}$ is uniformly positive definite, then (19) simplifies to

$$I_x(\varphi) = \frac{1}{2}\int_0^T (\dot{\varphi}_s - \bar{b}(s, \varphi_s))^\mathsf{T} a^{-1}(s, \varphi_s)(\dot{\varphi}_s - \bar{b}(s, \varphi_s))\,\mathrm{d}s$$

whenever $\varphi \in \mathcal{W}^d$ is absolutely continuous on $[0, T]$ with $\varphi_0 = x$, and $I_x(\varphi) = \infty$ otherwise.

4. Two applications

In Subsection 4.1 we apply Theorem 2 to derive the large deviation principle for stochastic systems with memory or delay established in Mohammed and Zhang (2006). They considered systems with point delay. Their proof is based on a discretization argument analogous to the method of steps for proving properties (including existence of solutions) of delay differential equations. This allows us to derive the large deviation principle for Itô processes with delay from the (well-established) large deviation principle for Itô diffusions with time-dependent coefficients. The coefficients are assumed to be globally Lipschitz.

In Subsection 4.2 we go back to Theorem 1 to derive the large deviation principle obtained in Baldi and Caramellino (2011) for a class of positive Itô diffusions with dispersion coefficient σ of square-root type. In that work, as mentioned in the introduction, a general large deviation principle is established for the diffusion case (the coefficients may actually depend on the parameter ε). The assumptions can be summarized as follows (c.f. Baldi and Caramellino (2011, Assumption A.2.3 and Theorem 2.4)): assumptions on b and σ in terms of (7) equivalent to our hypotheses (H4) and (H5), including existence of solutions; local Lipschitz continuity of b_ε and σ_ε for $\varepsilon > 0$ as well as strong existence (and uniqueness) of solutions for the corresponding prelimit equations; the quasicontinuity property (assumption A.2.3(c) there), which relates the prelimit solutions to solutions of the limit equation (7). All assumptions are verified for locally Lipschitz continuous coefficients with sublinear growth at ∞ when b_ε and σ_ε converge to b and σ uniformly on compacts. Although the diffusion coefficient σ in the case of positive diffusions is locally Lipschitz only on $\mathbb{R} \setminus \{0\}$, Proposition 1 below, which is Proposition 3.11 of Baldi and Caramellino (2011), allows us to invoke the large deviation principle for locally Lipschitz coefficients. Here, we use their result only to check that uniqueness holds for the controlled deterministic limit equation ((24) below).

4.1. Systems with memory

Let $\bar{b} \colon [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $\bar{\sigma} \colon [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ be Borel measurable functions. Let us make the following assumptions, which are those of Mohammed and Zhang (2006).

(Q1) The functions $\bar{b}$ and $\bar{\sigma}$ satisfy a global Lipschitz condition, that is, there exists a constant $L > 0$ such that, for all $x_1, x_2, y_1, y_2 \in \mathbb{R}^d$ and all $t \in [0, T]$,

$$|\bar{b}(s, x_1, y_1) - \bar{b}(s, x_2, y_2)| \leq L(|x_1 - x_2| + |y_1 - y_2|),$$
$$|\bar{\sigma}(s, x_1, y_1) - \bar{\sigma}(s, x_2, y_2)| \leq L(|x_1 - x_2| + |y_1 - y_2|).$$

(Q2) The functions $\bar{b}(\cdot, x, y)$ and $\bar{\sigma}(\cdot, x, y)$ are continuous on $[0, T]$, uniformly in $x, y \in \mathbb{R}^d$.

Let $\tau \in (0, T)$ and $\psi \in \mathbf{C}([-\tau, 0], \mathbb{R}^d)$; τ will be the length of the (fixed) point delay and ψ the initial segment. For $\varepsilon > 0$, consider the stochastic delay differential equation

$$\mathrm{d}X_t^\varepsilon = \bar{b}(t, X_t^\varepsilon, X_{t-\tau}^\varepsilon)\,\mathrm{d}t + \sqrt{\varepsilon}\bar{\sigma}(t, X_t^\varepsilon, X_{t-\tau}^\varepsilon)\,\mathrm{d}W_t \qquad (20)$$

over $t \in [0, T]$ and with initial condition $X_s^\varepsilon = \psi_s$ for all $s \in [-\tau, 0]$. Denote by $\mathcal{C}_\psi$ the set of all continuous functions $\varphi \colon [-\tau, T] \to \mathbb{R}^d$ such that $\varphi_s = \psi_s$ for all $s \in [-\tau, 0]$. Let G_ψ be the map $L^2([0, T], \mathbb{R}^m) \to \mathcal{C}_\psi$ which takes $f \in L^2([0, T], \mathbb{R}^m)$ to the unique solution of

the integral equation

$$\varphi_t = \begin{cases} \psi_0 + \int_0^t \bar{b}(s, \varphi_s, \varphi_{s-\tau}) \, \mathrm{d}s + \int_0^t \bar{\sigma}(s, \varphi_s, \varphi_{s-\tau}) f_s \, \mathrm{d}s & \text{if } t \in (0, T], \\ \psi_t & \text{if } t \in [-\tau, 0]. \end{cases} \quad (21)$$

Theorem 3. *Assume that (Q1) and (Q2) hold. Then the map G_ψ is well defined and the family $\{X^\varepsilon\}_{\varepsilon > 0}$ of solutions to the stochastic delay differential equation (20) with initial condition $X_s^\varepsilon = \psi_s$ for $s \in [-\tau, 0]$ satisfies the large deviation principle with good rate function $I_\psi \colon \mathcal{C}_\psi \to [0, \infty]$ given by*

$$I_\psi(\varphi) = \inf_{\{f \in L^2([0,T];\mathbb{R}^m) \colon \varphi = G_\psi(f)\}} \frac{1}{2} \int_0^T |f_t|^2 \, \mathrm{d}t$$

whenever $\{f \in L^2([0, T]; \mathbb{R}^m) \colon \varphi = G_\psi(f)\} \neq \varnothing$, and $I_\psi(\varphi) = \infty$ otherwise.

Proof. Define a function $\Phi \colon \mathcal{W}^d \to \mathcal{C}_\psi$ according to

$$\Phi[\varphi](s) \doteq \begin{cases} \psi_s \mathbf{1}_{[-\tau,0]}(s) + \varphi_s \mathbf{1}_{(0,T]}(s) & \text{if } \varphi_0 = \psi_0, \\ \psi_s \mathbf{1}_{[-\tau,0]}(s) + \psi_0 \mathbf{1}_{(0,T]}(s) & \text{otherwise.} \end{cases}$$

Define mappings b and σ from $[0, T] \times \mathcal{W}^d$ to $\mathbb{R}^d$ and to $\mathbb{R}^{d \times m}$, respectively, according to

$$b(s, \varphi) := \bar{b}(s, \varphi_s, \psi_{s-\tau}) \mathbf{1}_{[0,\tau)}(s) + \bar{b}(s, \varphi_s, \varphi_{s-\tau}) \mathbf{1}_{[\tau,T]}(s),$$
$$\sigma(s, \varphi) := \bar{\sigma}(s, \varphi_s, \psi_{s-\tau}) \mathbf{1}_{[0,\tau)}(s) + \bar{\sigma}(s, \varphi_s, \varphi_{s-\tau}) \mathbf{1}_{[\tau,T]}(s),$$

and consider the stochastic differential equation

$$\mathrm{d}Y_t^\varepsilon = b(t, Y^\varepsilon) \, \mathrm{d}t + \sqrt{\varepsilon} \sigma(t, Y^\varepsilon) \, \mathrm{d}W_t \quad (22)$$

over $[0, T]$ with initial condition $Y_0^\varepsilon = \psi_0$. We show that the functions b and σ enjoy assumptions (A1)–(A4). Since the coefficients do not depend on ε, it suffices to verify (A1) and (A2). We check the assumptions only for b, the work for σ being completely analogous. Let us start with (A1). Thanks to (Q1) we have

$$|\bar{b}(t, x, y)| \leq L(|x| + |y|) + |\bar{b}(t, 0, 0)|.$$

By (Q2), it follows that $\sup_{t \in [0,T]} |\bar{b}(t, 0, 0)| < \infty$. Let $\varphi \in \mathcal{W}^d$. Then

$$|b(s, \varphi)| \leq \begin{cases} L(|\varphi_s| + |\psi_{s-\tau}|) + |\bar{b}(s, 0, 0)| & \text{if } s \in [0, \tau), \\ L(|\varphi_s| + |\varphi_{s-\tau}|) + |\bar{b}(s, 0, 0)| & \text{if } s \in [\tau, T]. \end{cases}$$

Set $M := 2L \vee (\sup_{t \in [0,T]} |\bar{b}(t, 0, 0)| + \sup_{s \in [-\tau,0]} L|\psi_s|)$. Then $|b(s, \varphi)| \leq M(1 + \sup_{t \in [0,s]} |\varphi_t|)$, which yields (A1). Next we verify (A2). Let $\varphi, \tilde{\varphi} \in \mathcal{W}^d$. Then, thanks to (Q1),

$$|b(s, \varphi) - b(s, \tilde{\varphi})| \leq \begin{cases} L|\varphi_s - \tilde{\varphi}_s| & \text{if } s \in [0, \tau), \\ L(|\varphi_s - \tilde{\varphi}_s| + |\varphi_{s-\tau} - \tilde{\varphi}_{s-\tau}|) & \text{if } s \in [\tau, T]. \end{cases}$$

Thus, $b(t, \cdot)$ is globally Lipschitz continuous with constant $2L$, uniformly in $t \in [0, T]$.

Since b and σ satisfy both (A1) and (A2), Theorem 2 applies and reveals that the family $\{Y^\varepsilon\}_{\varepsilon>0}$ of solutions to (22) with initial condition $Y_0^\varepsilon = \psi_0$ satisfies the large deviation principle with good rate function $J\colon \mathcal{W}^d \to [0, \infty]$ given by

$$J(\varphi) = \inf_{\{f\in L^2([0,T];\mathbb{R}^m)\colon \varphi=\Gamma(f)\}} \frac{1}{2} \int_0^T |f_t|^2 \, dt,$$

where $\inf \varnothing = \infty$ by convention and $\Gamma := \Gamma_{\psi_0}$ as in (H4). In particular, Γ is well defined as the mapping $L^2([0, T], \mathbb{R}^m) \to \mathcal{W}^d$ that takes $f \in L^2([0, T], \mathbb{R}^m)$ to the unique solution of (7), that is, to the unique solution $\varphi \in \mathcal{W}^d$ of the integral equation

$$\varphi_t = x + \int_0^t b(s, \varphi) \, ds + \int_0^t \sigma(s, \varphi) f_s \, ds, \qquad t \in [0, T].$$

Now let $\varphi \in \mathcal{C}_\psi$. Then φ solves the integral equation (21) with $f \in L^2([0, T], \mathbb{R}^m)$ if and only if $\varphi_{|[0,T]} = \Gamma(f)$. Recalling the definitions of b and σ, it follows that (21) has a unique solution and that the mapping G is well defined. Moreover, for every $\varepsilon > 0$, (20) possesses a unique strong solution X^ε with initial segment ψ, and $X^\varepsilon = \Phi[Y^\varepsilon]$, θ-almost surely.

Set $\mathfrak{C}_{\psi_0} := \{\varphi \in \mathcal{W}^d\colon \varphi_0 = \psi_0\}$. Observe that the effective domain of J, namely, $\mathcal{D}_J := \{\varphi \in \mathcal{W}^d\colon J(\varphi) < \infty\}$, is contained in $\mathfrak{C}_{\psi_0}$. The map Φ is continuous on $\mathfrak{C}_{\psi_0}$ (in fact, a continuous bijection $\mathfrak{C}_{\psi_0} \to \mathcal{C}_\psi$). Since the processes Y^ε take values in $\mathfrak{C}_{\psi_0}$ and $X^\varepsilon = \Phi[Y^\varepsilon]$, it follows by the contraction principle (see, for instance, Theorem 4.2.1 with Remark (c) of Dembo and Zeitouni (1998, pp. 126–127)) that the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies the large deviation principle with good rate function $I\colon \mathcal{C}_\psi \to [0, \infty]$ given by

$$\begin{aligned}
I(\bar{\varphi}) &= \inf\{J(\varphi)\colon \varphi \in \mathcal{W}^d \text{ such that } \Phi[\varphi] = \bar{\varphi}\} \\
&= J(\bar{\varphi}_{|[0,T]}) \\
&= \inf_{\{f\in L^2([0,T];\mathbb{R}^m)\colon \bar{\varphi}_{|[0,T]}=\Gamma(f)\}} \frac{1}{2} \int_0^T |f_t|^2 \, dt \\
&= \inf_{\{f\in L^2([0,T];\mathbb{R}^m)\colon \bar{\varphi}=G(f)\}} \frac{1}{2} \int_0^T |f_t|^2 \, dt.
\end{aligned}$$

Remark 7. A closer look at the proof above shows that we can generalize without any effort the result of Mohammed and Zhang (2006). We can assume that the coefficients are locally Lipschitz continuous and depend on ε as well, provided that a sublinear growth condition is satisfied and that $\bar{b}_\varepsilon \to b$ and $\bar{\sigma}_\varepsilon \to \sigma$. In particular, the uniform continuity condition (Q2) is no longer needed, and it suffices to assume predictability of the coefficients. With the approach used here, the large deviation analysis can be performed in the same way also for other delay models, such as distributed delay or dependence on the running maximum; in those cases the coefficients could be (locally) Lipschitz functions of expressions like

$$\int_{-\tau}^0 g(\psi_s) \, ds, \qquad \sum_{s\in J} g_s(\psi_s), \qquad \max_{s\in[-\tau,0]} g(\psi_s),$$

where g and g_s are suitable functions, and $J \subset [-\tau, 0]$ is a countable set. These generalizations would be difficult to obtain with a method-of-steps approach.

4.2. Positive diffusions with Hölder dispersion coefficient

In this subsection we derive the large deviation principle for a class of scalar Itô diffusions where the dispersion coefficient σ is positive away from 0 and Hölder continuous with exponent $\gamma \geq \frac{1}{2}$. We can rely on the work by Baldi and Caramellino (2011) in proving uniqueness for the deterministic limit system (7) as required by hypothesis (H4) (see Proposition 1 below); then we invoke Theorem 1.

Let W denote a one-dimensional Brownian motion. Slightly changing notation, let $x_0 > 0$ be the initial condition and consider, for $\varepsilon > 0$, the scalar stochastic differential equation

$$\mathrm{d}X_t^\varepsilon = \bar{b}(X_t^\varepsilon)\,\mathrm{d}t + \sqrt{\varepsilon}\bar{\sigma}(X_t^\varepsilon)\,\mathrm{d}W_t \tag{23}$$

with $X_0^\varepsilon = x_0$. We make the following assumptions on the coefficients $\bar{b}$ and $\bar{\sigma}$, which we take independent of ε for the sake of simplicity.

(R1) The dispersion coefficient $\bar{\sigma} : \mathbb{R} \to [0, \infty)$ is locally Lipschitz continuous on $\mathbb{R} \setminus \{0\}$, has sublinear growth at ∞, and $\bar{\sigma}(0) = 0$, while $\bar{\sigma}(x) > 0$ for all $x \neq 0$. Moreover, there exists a continuous increasing function $\rho : (0, \infty) \to (0, \infty)$ such that $\int_{0+}^\infty \rho^{-2}(u)\,\mathrm{d}u = +\infty$ and

$$|\bar{\sigma}(x) - \bar{\sigma}(y)| \leq \rho(|x - y|) \quad \text{for all } x, y \in \mathbb{R}, \ x \neq y.$$

(R2) The drift coefficient $\bar{b} : \mathbb{R} \to \mathbb{R}$ is locally Lipschitz continuous, has sublinear growth at ∞, and $\bar{b}(0) > 0$.

Condition (R1) is satisfied, in particular, if $\bar{\sigma}(x) = \sqrt{|x|}$. The large deviation principle will be derived from Theorem 1. To this end, set

$$\sigma(s, \varphi) := \bar{\sigma}(\varphi_s), \qquad b(s, \varphi) = \bar{b}(\varphi_s), \qquad (s, \varphi) \in [0, T] \times \mathcal{W}^1.$$

Let us check that hypotheses (H1)–(H6) hold for b and σ. Since $\bar{b}$ and $\bar{\sigma}$ are continuous, b and σ are predictable with $b(t, \cdot)$ and $\sigma(t, \cdot)$ uniformly continuous on all bounded subsets of $\mathcal{W}^1$, uniformly in $t \in [0, T]$. Moreover, given any $\varphi \in \mathcal{W}^1$, $\sigma(\cdot, \varphi)$ is bounded by $M(1 + \|\varphi\|_\infty)$ for some M independent of φ thanks to the sublinear growth condition, and is hence square integrable. Thus, (H1) holds. Hypothesis (H2) is clearly satisfied as $b_\varepsilon \equiv b$ and $\sigma_\varepsilon \equiv \sigma$. Under (R1) and (R2), pathwise uniqueness holds for (23) (or (5) with b and σ as above); this follows from Theorem 1 of Yamada and Watanabe (1971). Continuity and sublinear growth of the coefficients implies existence of a weak solution (for instance, Theorems 2.3 and 2.4 of Ikeda and Watanabe (1989)), which together with pathwise uniqueness actually implies that any solution is strong (see Corollary 3 of Yamada and Watanabe (1971) or Theorem IX.1.7 of Revuz and Yor (1999, p. 368)). Accordingly, hypothesis (H3) holds. The fact that hypothesis (H4) holds is a consequence of Remark 2 and Proposition 1 stated next, which can be proved exactly as Proposition 3.11 of Baldi and Caramellino (2011).

Proposition 1. *Assume that (R1) and (R2) hold. Let $f \in L^2([0, T])$. Then uniqueness of solutions holds for the integral equation*

$$\varphi_t = x_0 + \int_0^t \bar{b}(\varphi_s)\,\mathrm{d}s + \int_0^t \bar{\sigma}(\varphi_s) f_s\,\mathrm{d}s. \tag{24}$$

Moreover, for every $N > 0$, there exists $\eta > 0$ such that $\inf_{t \in [0, T]} \varphi_t \geq \eta$ whenever φ is a solution of (24) and $\|f\|_{L^2} < N$.

Proposition 1 also implies that hypothesis (H5) is satisfied. The map Γ_x which takes $f \in S_N$ to the unique solution of the integral equation

$$\varphi_t = x + \int_0^t b(\varphi_s)\,\mathrm{d}s + \int_0^t \sigma(\varphi_s) f_s\,\mathrm{d}s$$

coincides with the map defined by replacing σ with a function which is locally Lipschitz on the whole $\mathbb{R}$ and equals σ outside a sufficiently small neighborhood of 0. Indeed, there exists $\xi > 0$ such that, for all $f \in S_N$, $\Gamma_x(f) \geq \xi$. Therefore, Γ_x is continuous on S_N endowed with the weak topology of L^2, as a consequence of what we have shown in Section 3 in the case of locally Lipschitz continuous coefficients.

Finally, by assumptions (R1) and (R2), the coefficients b and σ have sublinear growth at ∞. Based on this property, we can argue exactly as in Section 3 to show that (H6) holds.

Theorem 4. *Assume that (R1) and (R2) hold. Then the family $\{X^\varepsilon\}_{\varepsilon>0}$ of solutions to the stochastic differential equation (23) with initial condition x_0 satisfies the large deviation principle with good rate function $I : \mathbb{C}([0, T], \mathbb{R}) \to [0, \infty]$ given by*

$$I(\varphi) = \frac{1}{2} \int_0^T \frac{(\dot{\varphi}_t - \bar{b}(\varphi_t))^2}{\bar{\sigma}^2(\varphi_t)}\,\mathrm{d}t$$

whenever φ is absolutely continuous on $[0, T]$ such that $\varphi_0 = x_0$ and $(\dot{\varphi} - \bar{b})/\bar{\sigma}(\varphi) \in L^2([0, T], \mathbb{R})$, and $I(\varphi) = \infty$ otherwise.

Proof. We have already checked that (R1) and (R2) imply (H1)–(H6). Theorem 1 therefore yields the large deviation principle for the family $\{X^\varepsilon\}_{\varepsilon>0}$ with good rate function $J = J_{x_0}$ given by

$$J(\varphi) = \inf_{\{f \in L^2([0,T],\mathbb{R})\,:\,\varphi_t = x_0 + \int_0^t \bar{b}(\varphi_s)\,\mathrm{d}s + \int_0^t \bar{\sigma}(\varphi_s) f_s\,\mathrm{d}s\}} \frac{1}{2} \int_0^T |f_t|^2\,\mathrm{d}t$$

whenever $\{f \in L^2([0, T], \mathbb{R}) : \varphi_t = x_0 + \int_0^t \bar{b}(\varphi_s)\,\mathrm{d}s + \int_0^t \bar{\sigma}(\varphi_s) f_s\,\mathrm{d}s\} \neq \varnothing$, and $I(\varphi) = \infty$ otherwise. In particular, $J(\varphi) < \infty$ if and only if φ solves (24) for some $f \in L^2([0, T], \mathbb{R})$. Let $\varphi \in \mathcal{W}^1$ be such that $J(\varphi) < \infty$. Then φ solves (24) for some $f \in L^2([0, T], \mathbb{R})$; hence,

$$\dot{\varphi}_t = b(\varphi_t) + \sigma(\varphi_t) f_t \quad \text{for almost every } t \in [0, T],$$

and φ is absolutely continuous on $[0, T]$ with $\varphi_0 = x_0$. By Proposition 1, $\varphi_t > 0$ for all $t \in [0, T]$; thus, $\bar{\sigma}(\varphi_t) \neq 0$; hence,

$$\frac{\dot{\varphi}_t - \bar{b}(\varphi_t)}{\bar{\sigma}(\varphi_t)} = f_t$$

for almost every $t \in [0, T]$. It follows that

$$\frac{1}{2} \int_0^T |f_t|^2\,\mathrm{d}t = \frac{1}{2} \int_0^T \frac{(\dot{\varphi}_t - \bar{b}(\varphi_t))^2}{\bar{\sigma}^2(\varphi_t)}\,\mathrm{d}t,$$

which implies that $J(\varphi) = I(\varphi)$. On the other hand, if $\varphi \in \mathcal{W}^1$ is absolutely continuous on $[0, T]$ with $\varphi_0 = x_0$ such that $\int_0^T ((\dot{\varphi}_t - \bar{b}(\varphi_t))^2/\bar{\sigma}^2(\varphi_t))\,\mathrm{d}t < \infty$, then

$$f_t := \frac{\dot{\varphi}_t - \bar{b}(\varphi_t)}{\bar{\sigma}(\varphi_t)}$$

is well defined as an element of $L^2([0, T], \mathbb{R})$ and φ solves (24) with control f. It follows also in this case that $J(\varphi) = I(\varphi)$.

Appendix A

As above, let $(\mathcal{W}^m, \mathcal{B}, \theta)$ be the canonical probability space for m-dimensional Brownian motion over the time interval $[0, T]$, and let $(\mathcal{G}_t)$ be the θ-augmented filtration generated by the coordinate process W. Let $\mathcal{M}^2[0, T]$ denote the space of all $\mathbb{R}^m$-valued, square-integrable, $(\mathcal{G}_t)$-predictable processes. Theorem 3.1 of Boué and Dupuis (1998) provides the following representation for Laplace functionals of the Brownian motion W. For all $F \colon \mathcal{W}^m \to \mathbb{R}$ bounded and measurable,

$$-\log \mathbb{E}[\mathrm{e}^{-F(W)}] = \inf_{v \in \mathcal{M}^2[0,T]} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_s|^2 \,\mathrm{d}s + F\left(W + \int_0^{\cdot} v_s \,\mathrm{d}s\right)\right], \qquad (25)$$

where $\mathbb{E}$ denotes the expectation with respect to the Wiener measure θ.

Let $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ be predictable functions from $[0, T] \times \mathcal{W}^d$ to $\mathbb{R}^d$ and to $\mathbb{R}^{d \times m}$, respectively. Fix $x \in \mathbb{R}^d$, and consider the stochastic differential equation

$$\mathrm{d}X_t = b(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}W_t \qquad (26)$$

for $t \in [0, T]$ and with initial condition $X_0 = x$. Suppose that (26) has a strong solution. Then there exists a $\mathcal{B}(\mathcal{W}^m) \setminus \mathcal{B}(\mathcal{W}^d)$-measurable function $h \colon \mathcal{W}^m \to \mathcal{W}^d$ such that $X = h[W]$, θ-almost surely; see, for instance, Theorem 10.4 of Rogers and Williams (2000, p. 126). Hence, for any $F \colon \mathcal{W}^d \to \mathbb{R}$ bounded and measurable, $F \circ h$ is a bounded and measurable map from $\mathcal{W}^m$ into $\mathbb{R}$. By representation formula (25) for Brownian motion, it follows that

$$-\log \mathbb{E}[\mathrm{e}^{-F(X)}] = -\log \mathbb{E}[\mathrm{e}^{-F \circ h(W)}]$$
$$= \inf_{v \in \mathcal{M}^2[0,T]} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_s|^2 \,\mathrm{d}s + F \circ h\left(W + \int_0^{\cdot} v_s \,\mathrm{d}s\right)\right]. \qquad (27)$$

For $v \in \mathcal{M}^2[0, T]$, consider the controlled stochastic differential equation

$$\mathrm{d}X_t^v = b(t, X^v)\,\mathrm{d}t + \sigma(t, X^v)v_t \,\mathrm{d}t + \sigma(t, X^v)\,\mathrm{d}W_t \qquad (28)$$

for $t \in [0, T]$ and with initial condition $X_0^v = x$. If strong existence and pathwise uniqueness hold for (26) then the term $F \circ h(W + \int_0^{\cdot} v_s \,\mathrm{d}s)$ in (27) can be rewritten in terms of solutions to (28). We only need that identity for control processes v with deterministically bounded L^2-norm. Lemma 1 below should be compared to Theorem 4.1 of Boué and Dupuis (1998).

Lemma 1. *Let $v \in \mathcal{M}^2[0, T]$ be such that $\int_0^T |v_s|^2 \,\mathrm{d}s \leq N$, θ-almost surely for some $N > 0$. Suppose that strong existence and pathwise uniqueness hold for (26) with initial condition $X_0 = x$. Then (28) has a unique strong solution X^v with $X_0^v = x$ and*

$$h\left(W + \int_0^{\cdot} v_s \,\mathrm{d}s\right) = X^v \quad \theta\text{-almost surely.}$$

Proof. Define the process

$$\tilde{W}_t := W_t + \int_0^t v_s \,\mathrm{d}s, \qquad t \in [0, T].$$

Since $\int_0^t |v_s|^2 \,\mathrm{d}s \leq N$, θ-almost surely, Girsanov's theorem is applicable; accordingly, there exists a measure γ over $\mathcal{W}^m$ equivalent to θ such that $\tilde{W}$ is a $(\mathcal{G}_t)$-Brownian motion on $[0, T]$

(see, for instance, Theorem 5.2 of Karatzas and Shreve (1991, p. 191)). With respect to the measure γ, the controlled equation (28) becomes

$$dX_t^v = b(t, X^v)\, dt + \sigma(t, X^v)\, d\tilde{W}_t. \tag{29}$$

Uniqueness of solutions to (28) follows by assumption of pathwise uniqueness for (26). Indeed, if X and Y are two solutions of (28) with respect to W and θ, then they are solutions of (29) with respect to γ and $\tilde{W}$. By pathwise uniqueness, X and Y are indistinguishable.

We now prove existence of solutions. For continuous and $(\mathcal{G}_t)$-adapted processes Z, define the map $\Psi(Z)\colon \mathcal{W}^m \to \mathcal{W}^d$ according to

$$\Psi(Z)(\omega) := x + \int_0^\cdot b(s, h[Z(\omega)])\, ds + \left(\int_0^\cdot \sigma(s, h[Z(\omega)])\, dZ_s\right)(\omega).$$

The map $\Psi(Z)$ is certainly well defined when Z is given by

$$Z_t(\omega) := \tilde{W}_t(\omega) = \omega(t) + \int_0^t v_s(\omega)\, ds$$

with $v \in \mathcal{M}^2[0, T]$. In this situation, for θ-almost all $\omega \in \mathcal{W}^m$,

$$\begin{aligned}
\Psi(\tilde{W})(\omega) = &\, x + \int_0^\cdot b(s, h[\tilde{W}(\omega)])\, ds + \int_0^\cdot \sigma(s, h[\tilde{W}(\omega)])v_s(\omega)\, ds \\
&+ \left(\int_0^\cdot \sigma(s, h[\tilde{W}])\, dW_s\right)(\omega),
\end{aligned} \tag{30}$$

where W is the coordinate process on $\mathcal{W}^m$. Since $h[W]$ is a solution of (26), by construction we have

$$h[W(\omega)] = \Psi(W)(\omega) \quad \text{for } \theta\text{-almost all } \omega \in \mathcal{W}^m.$$

By Theorem 10.4 of Rogers and Williams (2000, p. 126), $h(\tilde{W})$ satisfies

$$h[\tilde{W}] = x + \int_0^\cdot b(s, h[\tilde{W}])\, ds + \int_0^\cdot \sigma(s, h[\tilde{W}])\, d\tilde{W}_s \qquad \gamma\text{-almost surely.}$$

Since γ is equivalent to θ, it follows that

$$h[\tilde{W}] = \Psi(\tilde{W}) \quad \theta\text{-almost surely.}$$

Thanks to (30), this implies that, θ-almost surely,

$$h[\tilde{W}]_t = \Psi(\tilde{W})_t = x + \int_0^t b(s, h[\tilde{W}])\, ds + \int_0^t \sigma(s, h[\tilde{W}])v_s\, ds + \int_0^t \sigma(s, h[\tilde{W}])\, dW_s,$$

showing that $h[\tilde{W}]$ is a strong solution of (28) with respect to W and θ. We have already seen that pathwise uniqueness holds for (28). It follows that

$$h\left(W + \int_0^\cdot v_s\, ds\right) = X^v \quad \theta\text{-almost surely.}$$

for any solution X^v of (28) with $X_0^v = x$.

The following lemma provides a growth estimate if the coefficients b and σ satisfy a sublinear growth condition. The proof uses only standard arguments, including localization along times of first exit, the Burkholder–Davis–Gundy inequality, and Gronwall's lemma.

Lemma 2. *Let $v \in \mathcal{M}^2[0, T]$ be such that $\int_0^T |v_s|^2 \, ds \leq N$, θ-almost surely for some $N > 0$. Assume that b and σ are such that, for some $M > 0$,*

$$|b(t, \varphi)| \vee |\sigma(t, \varphi)| \leq M \left(1 + \sup_{s \in [0,t]} |\varphi_s| \right)$$

for all $t \in [0, T]$, all $\varphi \in \mathcal{W}^d$. If X^v is a solution of (28) with $X_0^v = x$ then, for all $p \geq 2$,

$$\mathbb{E}\left[\sup_{t \in [0,T]} |X_t^v|^p \right] \leq C_p(T, N, M)(1 + |x|^p),$$

where $C_p(T, N, M)$ is nondecreasing in each of its three arguments.

Acknowledgement

References

AZENCOTT, R. (1980). Grandes déviations et applications. In *Eighth Saint Flour Probability Summer School—1978* (Lecture Notes Math. **774**), Springer, Berlin, pp. 1–176.

BALDI, P. AND CARAMELLINO, L. (2011). General Freidlin-Wentzell large deviations and positive diffusions. *Statist. Prob. Lett.* **81,** 1218–1229.

BALDI, P. AND CHALEYAT-MAUREL, M. (1988). An extension of Ventsel'–Freidlin estimates. In *Stochastic Analysis and Related Topics* (Lecture Notes Math. **1316**), Springer, Berlin, pp. 305–327.

BOUÉ, M. AND DUPUIS, P. (1998). A variational representation for certain functionals of Brownian motion. *Ann. Prob.* **26,** 1641–1659.

BUDHIRAJA, A. AND DUPUIS, P. (2000). A variational representation for positive functionals of infinite dimensional Brownian motion. *Prob. Math. Statist.* **20,** 39–61.

BUDHIRAJA, A., DUPUIS, P. AND MAROULAS, V. (2008). Large deviations for infinite dimensional stochastic dynamical systems. *Ann. Prob.* **36,** 1390–1420.

BUDHIRAJA, A., DUPUIS, P. AND MAROULAS, V. (2011). Variational representations for continuous time processes. *Ann. Inst. H. Poincaré Prob. Statist.* **47,** 725–747.

DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd edn. Springer, New York.

DONATI-MARTIN, C., ROUAULT, A., YOR, M. AND ZANI, M. (2004). Large deviations for squares of Bessel and Ornstein–Uhlenbeck processes. *Prob. Theory Relat. Fields* **129,** 261–289.

DUPUIS, P. AND ELLIS, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley, New York.

FREIDLIN, M. I. AND WENTZELL, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd edn. Springer, New York.

IKEDA, N. AND WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd edn. North-Holland, Amsterdam.

KARATZAS, I. AND SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus* (Graduate Texts Math. **113**), 2nd edn. Springer, New York.

MOHAMMED, S-E. A. AND ZHANG, T. (2006). Large deviations for stochastic systems with memory. *Discrete Contin. Dyn. Syst. B* **6,** 881–893.

PRIOURET, P. (1982). Remarques sur les petites perturbations de systèmes dynamiques. In *Seminar on Probability, XVI* (Lecture Notes Math. **920**), Springer, Berlin, pp. 184–200.

PUHALSKII, A. A. (2001). *Large Deviations and Idempotent Probability* (Monogr. Surveys Pure Appl. Math. **119**). Chapman & Hall/CRC, Boca Raton, FL.

PUHALSKII, A. A. (2004). On some degenerate large deviation problems. *Electron. J. Prob.* **9,** 862–886.

REVUZ, D. AND YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd edn. Springer, Berlin.

ROGERS, L. C. G. AND WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales*, Vol. 2, *Itô Calculus*. Cambridge University Press.

YAMADA, T. AND WATANABE, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11,** 155–167.