SIAM J. CONTROL OPTIM.
Vol. 42, No. 3, pp. 810–838

LARGE DEVIATIONS FOR A RANDOM WALK MODEL WITH STATE-DEPENDENT NOISE*

MICHELLE BOUÉ†, DANIEL HERNÁNDEZ-HERNÁNDEZ‡, AND RICHARD S. ELLIS§

Abstract. In this paper we prove the large deviation principle for a class of random walks with state-dependent noise. This type of model has important applications in queueing and communication theory and in the area of stochastic approximation.

Key words. stochastic algorithms, large deviations, Laplace principle, weak convergence

AMS subject classifications. 60F10, 60K30

DOI. 10.1137/S0363012901396618

1. Introduction. This paper is concerned with proving a large deviation principle for a certain class of random walks, where the evolution of the noise process depends on the state of the random walk. This type of model arises in a natural way in the study of recursive algorithms, which have important applications in queueing and communication theory and in the area of stochastic approximation. In fact, our main motivation for the study of these models is their application to the state-dependent stochastic approximation algorithms presented in [10]. The convergence and rate of convergence analysis of algorithms is, in general, difficult, and it is associated with the solution of a deterministic differential equation. This approach is developed in detail in [10], where a number of different models are analyzed, including classical models like Robbins–Monro and ARMAX. Further examples arising from nonlinear filtering and off-line identification can be found in [12] and from parameter tracking in [9].

A general recursive algorithm has the form

$$\theta_{k+1} = \theta_k + \gamma_k F(\theta_k, \eta_{k+1}), \quad k = 0, 1, \ldots, \tag{1.1}$$

with $\theta_k \in \mathbb{R}^d$, $\eta_k \in \mathbb{R}^m$, and $\{\gamma_k, k = 0, 1, \ldots\}$ is a sequence of positive numbers such that $\gamma_k \to 0$ as $k \to \infty$. As in the Robbins–Monro algorithm, θ_k represents an "estimate" of an object of interest, while η_k is a random variable (or observation) with distribution function possibly depending on previous estimates and observations. In our model η_k will represent the noise entering the system (1.1) and its distribution may be affected by θ_{k-1} and η_{k-1}. The sequence $\{\eta_k, k \in \mathbb{N}\}$ has its own structure, depending on the type of application at hand. It can have a linear structure, as in the identification problem described in [11], or nonlinear, as in the random direction problem in [10, p. 16]. Let us suppose that observations are given by the recursive formula

$$\eta_{k+1} = G(\theta_k, \eta_k, \nu_{k+1}), \quad k = 0, 1, \ldots, \tag{1.2}$$

*Received by the editors October 15, 2001; accepted for publication (in revised form) January 17, 2003; published electronically June 18, 2003.
http://www.siam.org/journals/sicon/42-3/39661.html
†Department of Mathematics, Trent University, Peterborough, Ontario Canada K9L 1Z6 (michelleboue@trentu.ca).
‡Centro de Investigacioń en Matemáticas, Apartado Postal 402, Guanajuato, Gto. 36000, Mexico (dher@cimat.mx). The research of this author was supported by Conacyt grant 37643-E.
§Department of Mathematics, University of Massachusetts, Amherst, MA 01003-4515 (rsellis @math.umass.edu). The research of this author was supported by National Science Foundation grant NSF-DMS-0202309.

where ν_k are $\mathbb{R}^d$ valued independent and identically distributed random variables with strictly positive density g and, for each k, ν_{k+1} is independent of $\theta_j,\ \eta_j,\ \ j \le k$. Assume that, given θ and η, $G(\theta, \eta, \cdot)$ is a diffeomorphism on $\mathbb{R}^m$ with inverse $H(\theta, \eta, \cdot)$. Then, given A, a Borel set in $\mathbb{R}^d$,

$$\begin{aligned}
&\mathrm{Prob}[\eta_{k+1} \in A|\theta_0, \ldots, \theta_k,\ \eta_0, \ldots, \eta_k] \\
&= \mathrm{Prob}[\nu_{k+1} \in H(\theta_k, \eta_k, A)|\theta_0, \ldots, \theta_k,\ \eta_0, \ldots, \eta_k] \\
&= \int_{H(\theta_k, \eta_k, A)} g(y)dy \\
&= \int_A g(H(\theta_k, \eta_k, y))|J(\theta_k, \eta_k, y)|dy,
\end{aligned}$$

where J is the Jacobian of H and $|J|$ denotes its determinant. From this argument it can be seen that, under broad general conditions on F, G, and g, the algorithm (1.1)–(1.2) satisfies Hypothesis H.1 below, so that the conclusions of our main theorem apply.

The model. Let $\mathcal{S}$ be a Polish space, and let $p(d\zeta|x, \xi)$ be a stochastic kernel on $\mathcal{S}$ given $\mathbb{R}^d \times \mathcal{S}$. For each $n \in \mathbb{N}$, we consider a sequence of random variables $\{(X_j^n, Z_j^n), j = 0, \ldots, n\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $\mathbb{R}^d \times \mathcal{S}$. For $x \in \mathbb{R}^d$, $\xi \in \mathcal{S}$, and b a function mapping $\mathbb{R}^d \times \mathcal{S}$ into $\mathbb{R}^d$, this sequence is defined by setting $X_0^n \doteq x$, $Z_0^n \doteq \xi$, and letting

$$X_{j+1}^n \doteq X_j^n + \frac{1}{n} b(X_j^n, Z_{j+1}^n),$$

where for $j \in \{0, 1, \ldots, n-1\}$ the conditional distribution of Z_{j+1}^n given the past is given by

$$P_{x,\xi}\left\{Z_{j+1}^n \in d\zeta|(X_i^n, Z_i^n), i = 0, \ldots, j\right\} = p(d\zeta|X_j^n, Z_j^n). \tag{1.3}$$

Here $P_{x,\xi}$ denotes probability conditioned on $X_0^n = x$, $Z_0^n = \xi$. We assume that the stochastic kernel p and the function b satisfy the following hypothesis.

Hypothesis H.1.

(a) $b(x, \xi)$ is bounded, continuous in ξ, and Lipschitz continuous with constant K in x, uniformly in ξ.

(b) $p(d\zeta|x, \xi)$ is weakly continuous in (x, ξ), and there exist a probability measure ϑ on $\mathcal{S}$ and a measurable function $\tilde{p}^x(\xi, \zeta)$ on $\mathcal{S} \times \mathcal{S}$ such that

$$p(d\zeta|x, \xi) = \tilde{p}^x(\xi, \zeta)\vartheta(d\zeta).$$

(c) Given any compact set $\Delta \subset \mathbb{R}^d$, there exist constants $0 < a \le A < \infty$ such that

$$a \le \tilde{p}^x(\xi, \zeta) \le A$$

for all $x \in \Delta$. Moreover, $\tilde{p}^x(\xi, \zeta)$ is continuous in x uniformly in ξ and ζ, for $x \in \Delta$.

Let $X^n = \{X^n(t), 0 \le t \le 1\}$ be the piecewise linear interpolation on $[0, 1]$ of $\{X_j^n, j = 0, \ldots, n\}$. More precisely, for $t \in [j/n, (j+1)/n]$ and $j = 0, \ldots, n-1$,

$$X^n(t) \doteq X_j^n + \left(t - \frac{j}{n}\right) b(X_j^n, Z_{j+1}^n). \tag{1.4}$$

The main result of the paper, Theorem 2.1, states the large deviation principle for the sequence $\{X^n, n \in \mathbb{N}\}$. Although our main theorem, Theorem 2.1, is closely related to the results in [3], the proof there relies on technical assumptions for the function Λ (see (2.1) and (2.3) below), which is assumed to exist. Under Hypothesis H.1, the function Λ indeed exists and satisfies the technical assumptions required there (see section 4.3 in [3]), thereby implying the large deviation principle. However, our aim is to establish a more direct connection with the applications. The proof presented here depends on assumptions made on the evolution of the process itself (the transition kernels and the function b). This has several advantages. First, for the purposes of using the results in applications, assumptions must be made on the processes, since these are the type of assumptions that can be used there. Moreover, knowledge about the process provides a lot of intuition concerning the averaging procedure required for the proof. This intuition has been heavily exploited by some of the proofs of convergence of state-dependent stochastic algorithms (see [8, 10]), and we have incorporated some of their underlying ideas into the proof. Finally, seeing where each one of the properties of the process is needed in the proof has enabled us to understand the ergodicity properties required to extend our results to more general state-dependent processes. Extensions will be dealt with elsewhere.

2. The main theorem.

THEOREM 2.1. *Let $\mathcal{S}$ be compact. Under Hypothesis* H.1 *the sequence* $\{X^n, n \in \mathbb{N}\}$ *defined in* (1.4) *satisfies a large deviation principle with rate function* $I_x(\cdot)$*, where*

$$I_x(\phi) \doteq \begin{cases} \int_0^1 L(\phi, \dot{\phi})dt & \text{if } \phi \text{ is absolutely continuous and } \phi(0) = x, \\ \infty & \text{otherwise.} \end{cases}$$

Here $L(x, \cdot)$ is the Legendre–Fenchel transform with respect to the second variable of the function $\Lambda(x, \cdot)$, which is solution to the eigenvalue problem given by

$$e^{\Lambda(x,\alpha)+\Psi(\xi)} = \int_{\mathcal{S}} e^{\langle \alpha, b(x,\zeta)\rangle + \Psi(\zeta)} p(d\zeta|x, \xi). \tag{2.1}$$

That is, for x and β in $\mathbb{R}^d$,

$$L(x, \beta) \doteq \sup_{\alpha \in \mathbb{R}^d} \{\langle \alpha, \beta \rangle - \Lambda(x, \alpha)\}. \tag{2.2}$$

Remarks.

(1) We have made the strong assumption of compactness of the state space $\mathcal{S}$ in order to guarantee tightness of the measures involved in the proof (see part (a) of Theorem C.1). If the state space is not compact, further assumptions are required. These are discussed in section 5.

(2) $\mathcal{C}([0, 1] : \mathbb{R}^d)$ maps this space into $[0, \infty]$ and has compact level sets.

(3) For a fixed $x \in \mathbb{R}^d$, define the operator T on the set of bounded and measurable functions $\psi : \mathcal{S} \to \mathbb{R}$ as

$$T\psi(\xi) = \int_{\mathcal{S}} e^{\langle \alpha, b(x,\zeta)\rangle} \psi(\zeta) p(d\zeta|x, \xi).$$

The eigenvalue problem mentioned in (2.1) consists in finding the largest eigenvalue of this operator. Under Hypothesis H.1, Theorem 10.1 in [6] guarantees the existence and uniqueness of a solution to this problem, with a bounded and uniformly positive associated eigenfunction, corresponding to e^{Ψ} in (2.1). In fact, we can identify the solution function $\Lambda(x, \alpha)$ in a very explicit manner. Given $x \in \mathbb{R}^d$ and $\xi \in \mathcal{S}$, set

$\xi_0^x = \xi$ and let $\{\xi_j^x, j \geq 0\}$ be a Markov process with transition kernel $p(\cdot|x, \xi_j^x)$. Then the function $\Lambda(x, \alpha)$ satisfies

$$\Lambda(x,\alpha) = \lim_{N\to\infty} \frac{1}{N} \log E_\xi \left\{ \exp \left\langle \alpha, \sum_{j=1}^{N} b(x, \xi_j^x) \right\rangle \right\}, \tag{2.3}$$

where E_ξ denotes expectation conditioned on $\xi_0^x = \xi$. We refer to the process $\{\xi_j^x, j \geq 0\}$ as the "fixed x" process. As can be seen, it is the Markov chain that results if the parameter X_j^n in (1.3) is held constant at value x. This process is intimately connected with the process $\{X_j^n\}$. Indeed, if n is large, then X_j^n varies slowly and thus the "local" evolution of $b(X_j^n, Z_{j+1}^n)$ is very similar to the evolution of the same quantity but with X_j^n taken to be constant (see [10, sections 2.5 and 8.4]). This idea will be exploited heavily throughout the paper; we especially refer the reader to the proof of part (e) of Theorem C.1.

Let $W^n(x,\xi) \doteq -1/n \log E_{x,\xi} \{\exp[-nh(X^n)]\}$, with h in $\mathcal{C}([0,1] : \mathbb{R}^d)$. The proof of Theorem 2.1 is done in two parts. We start by proving an upper bound of the form

$$\liminf_{n\to\infty} W^n(x,\xi) \geq \inf_{\phi\in\mathcal{C}([0,1]:\mathbb{R}^d)} \{I_x(\phi) + h(\phi)\}. \tag{2.4}$$

This is the content of section 3. The lower bound

$$\limsup_{n\to\infty} W^n(x,\xi) \leq \inf_{\phi\in\mathcal{C}([0,1]:\mathbb{R}^d)} \{I_x(\phi) + h(\phi)\} \tag{2.5}$$

is then proved in section 4. These two inequalities are equivalent to a large deviation principle, as is proved in Theorems 2.2.1 and 2.2.3 in [4]. In both cases, a key step in the proof is based on studying (via weak convergence arguments) the limit properties of a sequence of associated stochastic control problems. The underlying simplicity of the basic arguments will be made clear below.

3. Proof of the upper bound. This section is devoted to the proof of (2.4). The proof can be summarized simply as follows. Based on the variational representation given in the next theorem, we associate with $W^n(x,\xi)$ an appropriate sequence of controlled processes and of control measures. The limit properties of this sequence, derived in Theorem C.1, will yield (2.4).

Let us start by introducing all the relevant quantities appearing in the representation for $W^n(x,\xi)$ (obtained in Theorem 3.1 below). The representation can be derived easily by following the same steps as those given in [4, section 4.4].

We define a discrete-time controlled process taking values in $\mathbb{R}^d \times \mathcal{S}$ denoted by $\{(\bar{X}_j^n, \bar{Z}_j^n), j = 0, \ldots, n\}$. The control at time j is the distribution of the controlled random variable $\bar{Z}_j^n$. It is given by a stochastic kernel $\nu_j^n(d\zeta|\bar{X}_0^n, \ldots, \bar{X}_j^n, \bar{Z}_j^n)$ on $\mathcal{S}$ given $(\mathbb{R}^d)^{j+1} \times \mathcal{S}$. That is, ν_j^n is a random variable mapping $(\mathbb{R}^d)^{j+1} \times \mathcal{S}$ into $\mathcal{P}(\mathcal{S})$.[1] A sequence of controls $\{\nu_j^n, j = 0, \ldots, n-1\}$ is what we refer to as an admissible control sequence. Now, setting $\bar{Z}_0^n = \xi$ and $\bar{X}_0^n = x$, the evolution of the controlled process is through the relation

$$\bar{X}_{j+1}^n = \bar{X}_j^n + \frac{1}{n} b(\bar{X}_j^n, \bar{Z}_{j+1}^n),$$

[1] For ease of presentation, the dependence on the underlying probability space of all stochastic kernels appearing in this paper is not made explicit in the notation.

where the conditional distribution of $\bar{Z}^n_{j+1}$ is given by

$$\bar{P}_{x,\xi}\left\{\bar{Z}^n_{j+1} \in d\zeta|\bar{X}^n_0,\dots,\bar{X}^n_j,\bar{Z}^n_0,\dots,\bar{Z}^n_j\right\} = \nu^n_j(d\zeta|\bar{X}^n_0,\dots,\bar{X}^n_j,\bar{Z}^n_j).$$

Finally, we let $\bar{X}^n = \{\bar{X}^n(t), t \in [0,1]\}$ be the piecewise linear interpolation of $\{\bar{X}^n_j, j = 0,\dots,n\}$.

THEOREM 3.1. *Let h be a bounded measurable function mapping $\mathcal{C}([0,1]:\mathbb{R}^d) \mapsto \mathbb{R}$. Then for all $n \in \mathbb{N}$, $x \in \mathbb{R}^d$, and $\xi \in \mathcal{S}$ we have the representation*

$$W^n(x,\xi) = \inf_{\{\nu^n_j\}} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu^n_j(\cdot)\|p(\cdot|\bar{X}^n_j,\bar{Z}^n_j)) + h(\bar{X}^n)\right\}. \tag{3.1}$$

Here R is the relative entropy function; $\nu^n_j(\cdot) = \nu^n_j(\cdot|\bar{X}^n_0,\dots,\bar{X}^n_j,\bar{Z}^n_j)$; the infimum is taken over all admissible control sequences $\{\nu^n_j, j = 0,\dots,n-1\}$; $\bar{E}_{x,\xi}$ denotes expectation conditioned on $\bar{X}^n_0 = x$ and $\bar{Z}^n_0 = \xi$; and $\{(\bar{X}^n_j,\bar{Z}^n_j), j = 0,\dots,n\}$ is the controlled process associated with a particular control sequence $\{\nu^n_j\}$.

Let $\varepsilon > 0$ be given. For each $n \in \mathbb{N}$, let $\{\nu^n_j, j = 0,\dots,n-1\}$ be a sequence of nearly optimal admissible controls for the variational problem in (3.1), so that

$$W^n(x,\xi) + \varepsilon \ge \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu^n_j(\cdot)\|p(\cdot|\bar{X}^n_j,\bar{Z}^n_j)) + h(\bar{X}^n)\right\}. \tag{3.2}$$

Here $\{(\bar{X}^n_j,\bar{Z}^n_j), j = 0,\dots,n\}$ is the controlled process associated with the nearly optimal sequence of controls.

We will obtain the limit inferior of the right-hand side of (3.2) by rewriting it in terms of a new sequence of control measures. These are defined as conveniently averaged controls in a space that is independent of n. For that purpose, let $\{m_n, n \in \mathbb{N}\}$ be a sequence of real numbers satisfying $m_n \to \infty$ as $n \to \infty$ and such that if $k_n \doteq m_n/n$, then $\lim_{n\to\infty} k_n = 0$. Also, suppose that 1 is an integer multiple of k_n. Given $\xi \in \mathcal{S}$, let δ_ξ denote the unit point measure at ξ. For $l = 0,\dots,1/k_n - 1$, and Borel subsets B_1 and B_2 of $\mathcal{S}$, let

$$\tilde{\nu}^n_l(B_1 \times B_2) \doteq \frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} \delta_{\bar{Z}^n_j}(B_1) \times \nu^n_j(B_2|\bar{X}^n_j,\bar{Z}^n_j).$$

The quantity $\tilde{\nu}^n_l$ is a stochastic kernel on $\mathcal{S}\times\mathcal{S}$ with marginals

$$(\nu^n_l)_1(B_1) = \frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} \delta_{\bar{Z}^n_j}(B_1) \quad \text{and} \quad (\nu^n_l)_2(B_2) = \frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} \nu^n_j(B_2|\bar{X}^n_j,\bar{Z}^n_j).$$

These definitions naturally result when one thinks of collecting terms of the sum appearing in (3.2) in groups of size m_n for the purposes of averaging. As was mentioned earlier, this technique is common in the proofs of convergence of state-dependent stochastic algorithms (see [8, 10]).

Now for each $n \in \mathbb{N}$ and $t \in [0,1]$ define

$$\nu^n(B_1 \times B_2|t) \doteq \begin{cases} \tilde{\nu}^n_l(B_1 \times B_2) \text{ if } t \in [lk_n,(l+1)k_n) \text{ for } l = 0,\dots,1/k_n - 2, \\ \tilde{\nu}^n_{(\frac{1}{k_n}-1)}(B_1 \times B_2) \text{ if } t \in [1-k_n, 1]. \end{cases}$$

Finally define the admissible control measure ν^n to be the random probability measure defined for Borel subsets B_1, B_2 of $\mathcal{S}$ and C of $[0,1]$ through

$$\nu^n(B_1 \times B_2 \times C) \doteq \int_C \nu^n(B_1 \times B_2|t)dt. \tag{3.3}$$

If for $B_1 \in \mathcal{B}(\mathcal{S})$ we define the first marginal $\hat{\nu}_1^n(d\zeta|t)$ of $\nu^n(d\zeta \times dy|t)$ through $\hat{\nu}_1^n(B_1|t) \doteq \nu^n(B_1 \times \mathcal{S}|t)$, then for Borel subsets B_1, B_2 of $\mathcal{S}$ and C of $[0,1]$, Theorem A.5.6 in [4] gives the decomposition

$$\nu^n(B_1 \times B_2 \times C) = \int_C \int_{B_1 \times B_2} \hat{\nu}_1^n(d\zeta|t)\hat{\nu}_2^n(dy|\zeta,t)dt = \int_C \int_{B_1} \hat{\nu}_2^n(B_2|\zeta,t)\hat{\nu}_1^n(d\zeta|t)dt, \tag{3.4}$$

where $\hat{\nu}_2^n(dy|\zeta,t)$ is a stochastic kernel on $\mathcal{S}$ given $\mathcal{S} \times [0,1]$. Following the notation in [4], we summarize this decomposition as $\nu^n(d\zeta \times dy \times dt) = \hat{\nu}_1^n(d\zeta|t) \otimes \hat{\nu}_2^n(dy|\zeta,t) \otimes \lambda$, where λ is Lebesgue measure on $[0,1]$.

We can now rewrite the right-hand side of (3.2) in terms of the control measures ν^n. We first use the fact (see [4, Lemma 1.4.3(f)]) that $R(\beta\|\gamma) = R(\alpha \times \beta\|\alpha \times \gamma)$ for any probability measures α, β, and γ on $\mathcal{S}$. This formula applied term by term enables us to write

$$\begin{aligned}&\bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu_j^n(\cdot)\|p(\cdot|\bar{X}_j^n, \bar{Z}_j^n))\right\}\\ &= \bar{E}_{x,\xi}\left\{\sum_{l=0}^{\frac{1}{k_n}-1} k_n \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n-1} R\left(\delta_{\bar{Z}_j^n}(\cdot) \times \nu_j^n(\cdot) \parallel \delta_{\bar{Z}_j^n}(d\zeta) \otimes p(\cdot|\bar{X}_j^n, \zeta)\right)\right\},\end{aligned}$$

where we have used the notation $\delta_{\bar{Z}_j^n}(\cdot) \times p(\cdot|\bar{X}_j^n, \bar{Z}_j^n) = \delta_{\bar{Z}_j^n}(d\zeta) \otimes p(\cdot|\bar{X}_j^n, \zeta)$. Applying Jensen's inequality to the convex function $R(\cdot\|\cdot)$, the right-hand side of the preceding display is no less than

$$\bar{E}_{x,\xi}\left\{\sum_{l=0}^{\frac{1}{k_n}-1} k_n R\left(\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} \delta_{\bar{Z}_j^n}(d\zeta) \times \nu_j^n(\cdot|\bar{X}_j^n, \zeta) \middle\| \frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} \delta_{\bar{Z}_j^n}(d\zeta) \otimes p\left(\cdot|\bar{X}_j^n, \zeta\right)\right)\right\},$$

which is clearly equal to

$$\bar{E}_{x,\xi}\left\{\int_0^1 R(\nu^n(\cdot|t)\|\gamma^n(\cdot|t))dt\right\}. \tag{3.5}$$

In (3.5), for each $n \in \mathbb{N}$ and $t \in [0,1]$, we have defined

$$\gamma^n(B_1 \times B_2|t) \doteq \frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} \delta_{\bar{Z}_j^n}(B_1) \otimes p(B_2|\bar{X}_j^n, \zeta) \text{ if } t \in [lk_n, (l+1)k_n),$$

for $l = 0, \dots, 1/k_n - 1$. Now define the measure γ^n on $\mathcal{S} \times \mathcal{S} \times [0,1]$ through $\gamma^n(B_1 \times B_2 \times C) \doteq \int_C \gamma^n(B_1 \times B_2|t)dt$. Since for all stochastic kernels α and β on $\mathcal{S}$ given $[0,1]$ and probability measures γ on $[0,1]$, we have (see [4, Lemma 1.4.3(f)])

$$\int_{[0,1]} R(\alpha(\cdot|x)\|\beta(\cdot|x))\gamma(dx) = R(\alpha \otimes \gamma\|\beta \otimes \gamma); \tag{3.6}$$

(3.5) can be rewritten as $\bar{E}_{x,\xi}\{R(\nu^n\|\gamma^n)\}$. (Recall the definition of ν^n given in (3.3).)

Combining this series of inequalities with (3.2), we obtain

$$W^n(x,\xi) + \varepsilon \geq \bar{E}_{x,\xi}\left\{R(\nu^n \| \gamma^n) + h(\bar{X}^n)\right\}.$$

We now wish to take the limit inferior as $n \to \infty$ of both terms in the last inequality. The asymptotic properties of the sequence $\{(\nu^n, \gamma^n, \bar{X}^n), n \in \mathbb{N}\}$ required to do this are proved in Theorem C.1. According to that theorem, there exists a probability space on which a subsequence of $\{(\nu^n, \gamma^n, \bar{X}^n), n \in \mathbb{N}\}$ converges in distribution to some limit $(\nu, \gamma, \bar{X})$. The stochastic kernels ν and γ and the random variable $\bar{X}$ satisfy all the conclusions stated in Theorem C.1. Thanks to the Skorohod representation theorem [5, p. 102] we can assume that convergence takes place with probability 1 (w.p.1). Along the convergent subsequence we thus have that

$$\begin{aligned}
&\liminf_{n\to\infty} W^n(x,\xi) + \varepsilon \\
&\geq \bar{E}_{x,\xi}\left\{R(\nu \| \gamma) + h(\bar{X})\right\} \\
&= \bar{E}_{x,\xi}\left\{R(\hat{\nu}_1(d\zeta|t) \otimes \hat{\nu}_2(dy|\zeta,t) \otimes dt \| \hat{\nu}_1(d\zeta|t) \otimes p(dy|\bar{X}(t),\zeta) \otimes dt) + h(\bar{X})\right\} \\
&= \bar{E}_{x,\xi}\left\{\int_0^1 \int_{\mathcal{S}} R(\hat{\nu}_2(dy|\zeta,t) \| p(dy|\bar{X}(t),\zeta))\hat{\nu}_1(d\zeta|t)dt + h(\bar{X})\right\} \\
&\geq \bar{E}_{x,\xi}\left\{\int_0^1 L\left(\bar{X}(t), \int_{\mathcal{S}} b(\bar{X}(t),\zeta)\hat{\nu}_1(d\zeta|t)\right)dt + h(\bar{X})\right\} \\
&= \bar{E}_{x,\xi}\left\{I_x(\bar{X}) + h(\bar{X})\right\} \geq \inf_{\phi \in \mathcal{C}([0,1]:\mathbb{R}^d)}\left\{I_x(\phi) + h(\phi)\right\}.
\end{aligned}$$

Lower semicontinuity of $R(\cdot\|\cdot)$, Fatou's lemma, and continuity of h yield the second line of the above display. The third line uses parts (b) and (f) of Theorem C.1 and the fourth uses (3.6). Finally, part (b) of Lemma B.2 and part (e) of Theorem C.1 give the fifth and sixth lines, respectively. Since the above inequality is valid for all $\varepsilon > 0$, (2.4) follows, concluding the proof of the upper bound. □

4. Proof of the lower bound. This section is devoted to showing that (2.5) holds for all h in $\mathcal{C}([0,1] : \mathbb{R}^d)$ that are Lipschitz continuous. Thanks to [4, Corollary 1.2.5(b)], this is enough to show that (2.5) holds for all h in $\mathcal{C}([0,1] : \mathbb{R}^d)$. As in the proof of Proposition 6.6.1 in [4], the proof of (2.5) is done by introducing a perturbation to the original random walk by means of a random walk with Gaussian noise. This allows one to obtain necessary smoothness properties for a function L_σ, which is the analogue of the function L defined in (2.2) but for the perturbed process. Weak convergence arguments make use of these continuity properties, implying the desired lower bound when taking the perturbation to be sufficiently small.

Let us first focus on the perturbed problem; the connection with (2.5) will be clear after (4.3). Given $\sigma > 0$, let $\{G_{j,\sigma}, j \in \mathbb{N}_0\}$ be a sequence of independent and identically distributed random variables on $\mathbb{R}^d$ with common Gaussian distribution ρ_σ, with mean zero and variance σI. We assume them to be independent of $\{\xi_j^x, x \in \mathbb{R}^d, j \in \mathbb{N}_0\}$, where ξ_j^x is the "fixed x" Markov process with transition kernel $p(\cdot|x,\xi_j^x)$. Given $n \in \mathbb{N}$ and $j \in \{0,1,\ldots,n-1\}$, let X_j^n and Z_j^n be as before, and define

$$U_{0,\sigma}^n \doteq 0, \qquad U_{j+1,\sigma}^n \doteq U_{j,\sigma}^n + \frac{1}{n}\, G_{j,\sigma}.$$

Denote by $X^n(t)$ and $U_\sigma^n(t)$ the piecewise linear interpolations of $\{X_j^n, j = 1,\ldots,n\}$ and $\{U_{j,\sigma}^n, j = 0,\ldots,n\}$ on $[0,1]$, respectively (see (1.4)). Also, define

$$Y_\sigma^n(t) \doteq X^n(t) + U_\sigma^n(t), \tag{4.1}$$

which is the piecewise linear interpolation of $\{X_j^n + U_{j,\sigma}^n\}$. As was mentioned earlier, the point of introducing a perturbation is to replace the function L by a continuous function L_σ. This latter function is defined as the Legendre–Fenchel transform of some convex function Λ_σ. Once again, the function Λ_σ is identified via an eigenvalue problem, which we now describe.

For fixed $x \in \mathcal{S}$, we can identify an additive component of the process (see [7, p. 376]), namely, $b(x, \xi_j^x) + G_{j,\sigma}$. Here ξ_j^x is the "fixed x" Markov process described earlier. Let Q_σ^x be the stochastic kernel on $\mathcal{S} \times \mathbb{R}^d$ given $\xi \in \mathcal{S}$ defined by

$$Q_\sigma^x(B_1 \times B_2|\xi) = \int_{B_1} \int_{\mathbb{R}^d} 1_{B_2}(b(x,\zeta) + y)\rho_\sigma(dy)p(d\zeta|x,\xi),$$

where $B_1 \in \mathcal{B}(\mathcal{S})$ and $B_2 \in \mathcal{B}(\mathbb{R}^d)$. Then, letting

$$\upsilon_\sigma(B_1 \times B_2|x) \doteq \int_{B_1} \int_{\mathbb{R}^d} 1_{B_2}(b(x,\zeta) + y)\rho_\sigma(dy)\vartheta(d\zeta),$$

with ϑ as in Hypothesis H.1, $B_1 \in \mathcal{B}(\mathcal{S})$, and $B_2 \in \mathcal{B}(\mathbb{R}^d)$, we have

$$a\upsilon_\sigma(B_1 \times B_2|x) \le Q_\sigma^x(B_1 \times B_2|\xi) \le A\upsilon_\sigma(B_1 \times B_2|x).$$

These bounds on $Q_\sigma^x(\cdot,\cdot|\xi)$ and the fact that the convex hull of the support of $\upsilon_\sigma(\mathcal{S} \times \cdot|x)$ is $\mathbb{R}^d$ guarantee the existence of a solution to the eigenvalue problem for each $x, \alpha \in \mathbb{R}^d$ [7, Lemma 3.1]. That is, for each $x, \alpha \in \mathbb{R}^d$, there exist a unique $\Lambda_\sigma(x,\alpha) \in \mathbb{R}$ and a bounded function $\Psi_\sigma(x;\alpha,\cdot) : \mathcal{S} \mapsto \mathbb{R}$ such that

$$e^{\Lambda_\sigma(x,\alpha)+\Psi_\sigma(x;\alpha,\xi)} = \int_{\mathcal{S}} \int_{\mathbb{R}^d} e^{\langle \alpha, b(x,\zeta)+y\rangle + \Psi_\sigma(x;\alpha,\zeta)} \rho_\sigma(dy)p(d\zeta|x,\xi).$$

Furthermore, $\Lambda_\sigma(x,\alpha) = \Lambda(x,\alpha) + \frac{\sigma^2}{2}||\alpha||^2$, and $\Psi_\sigma(x;\alpha,\xi) = \Psi(x;\alpha,\xi)$, where $\Psi(x;\alpha,\xi)$ is the eigenfunction associated with $\Lambda(x,\alpha)$ (see (2.1)). The Legendre–Fenchel transform of Λ_σ is given by

$$L_\sigma(x,\beta) \doteq \sup_{\alpha \in \mathbb{R}^d} \{\langle \alpha, \beta\rangle - \Lambda_\sigma(x,\alpha)\}. \tag{4.2}$$

Having introduced the necessary definitions, we now proceed to relate the original and the perturbed processes. Let K_1 be the Lipschitz constant of h and define $B \doteq 2||h||_\infty$. Then

$$h(Y_\sigma^n) = h(X^n + U_\sigma^n) \ge h(X^n) - (K_1||U_\sigma^n||_\infty \wedge B),$$

and, because of independence,

$$\begin{aligned} \frac{1}{n} \log E_{x,\xi}\{\exp[-nh(Y_\sigma^n)]\} &\le \frac{1}{n} \log E_{x,\xi}\{\exp[-nh(X^n)] \cdot \exp[n(K_1||U_\sigma^n||_\infty \wedge B)]\} \\ &= -W^n(x,\xi) + \frac{1}{n} \log E_{x,\xi}\{\exp[n(K_1||U_\sigma^n||_\infty \wedge B)]\}. \end{aligned}$$

Hence

$$\limsup_{n\to\infty} W^n(x,\xi) \le \limsup_{n\to\infty} \left(-\frac{1}{n} \log E_{x,\xi}\{\exp[-nh(Y_\sigma^n)]\}\right) + \frac{K_1^2\sigma^2}{2},$$

where the second term of the inequality follows from [4, p. 189]. This implies that (2.5) holds as long as we can show that

$$\limsup_{n\to\infty} W_\sigma^n(x,\xi) \le \inf_{\varphi\in C([0,1]:\mathbb{R}^d)} \{I_x(\varphi)+h(\varphi)\} + \theta(\sigma), \tag{4.3}$$

with $W_\sigma^n \doteq -\frac{1}{n}\log E_{x,\xi}\exp[-nh(Y_\sigma^n)]$ and $\theta(\sigma)\to 0$ when $\sigma\to 0$. What we will show in fact is that, given $\varepsilon>0$ and $\psi\in\mathcal{C}([0,1]:\mathbb{R}^d)$ satisfying

$$I_x(\psi)+h(\psi) \le \inf_{\varphi\in C([0,1]:\mathbb{R}^d)} \{I_x(\varphi)+h(\varphi)\} + \varepsilon < \infty, \tag{4.4}$$

we have

$$\limsup_{n\to\infty} W_\sigma^n(x,\xi) \le \int_0^1 L_\sigma(\psi(t),\dot\psi(t))dt + h(\psi) + \theta(\sigma). \tag{4.5}$$

Since $L_\sigma(x,\beta)\le L(x,\beta)$ for all x and $\beta\in\mathbb{R}^d$ (part (a) of Lemma B.3), (4.3) will follow after that.

The steps in the proof of (4.5) can be described in simple terms. Starting with the nearly optimal function ψ in (4.4), we construct a sequence of nearly optimal admissible controls for the stochastic control problem that is associated with $W_\sigma^n(x,\xi)$ through the representation in Theorem A.1. The limit properties of this sequence, as well as estimates on the associated sequence of running costs (where continuity of L_σ is required), will lead directly to (4.5).

Let ψ satisfy (4.4), and let ψ^* be as in part (e) of Lemma B.3. The admissible control sequence that we define based on ψ^* (see (4.20) below) has the following properties: the running costs are nearly optimal in (A.3), and, with probability converging to 1, the associated controlled process $\bar Y^n \doteq \bar X^n + \bar U^n$ (see (A.1)) enters a small neighborhood of ψ^* as $n\to\infty$. The construction is given in the following paragraphs.

Define the compact set

$$\Delta \equiv \cup_{t\in[0,1]}\{y\in\mathbb{R}^d : ||y-\psi^*(t)||\le 1\}.$$

Let $\eta=\eta(\Delta,\sigma)\in(0,1)$ satisfy the conclusions of part (d) of Lemma B.3 when taking $\varepsilon=\sigma$. Also, let $\{x_j, j=1,\dots,n\}$ be a sequence in Δ satisfying $\|\psi^*(j/n)-x_j\|<\eta$. For every $n\in\mathbb{N}$, $j=1,\dots,n$, and with $x=\psi^*(j/n)$, $y=x_j$, and $\beta=\dot\psi^*(j/n)$, part (d) of that lemma implies that there exists $\bar\beta_j^n\in\mathbb{R}^d$ such that

$$L_\sigma(x_j,\bar\beta_j^n) - L_\sigma(\psi^*(j/n),\dot\psi^*(j/n)) \le \sigma \tag{4.6}$$

and

$$||\bar\beta_j^n - \dot\psi^*(j/n)|| \le K\|\psi^*(j/n)-x_j\|.$$

Further, $\bar\beta_j^n = \bar\beta_j^{1,n}+\bar\beta_j^{2,n}$, with

$$\bar\beta_j^{1,n} = \int_{\mathcal{S}} b(x_j,\xi)\mu_{j,n}^*(d\xi) \quad \text{ and } \quad \bar\beta_j^{2,n} = \int_{\mathbb{R}^d} y\nu_{j,n}^*(dy).$$

Here $\mu_{j,n}^*$ is the invariant measure corresponding to the kernel $\gamma_{j,n}^*$ defined for $B_1\in\mathcal{B}(\mathcal{S})$ as

$$\begin{aligned}\gamma_{j,n}^*(B_1|\psi^*(j/n),\xi) = \int_{B_1} &\exp\{\langle\alpha, b(\psi^*(j/n),\zeta)\rangle - \Lambda(\psi^*(j/n),\alpha)\\ &+\Psi_\sigma(\psi^*(j/n);\alpha,\zeta) - \Psi_\sigma(\psi^*(j/n);\alpha,\xi)\}p(d\zeta|\psi^*(j/n),\xi),\end{aligned}$$

and for $B_2 \in \mathcal{B}(\mathbb{R}^d)$ as

$$\nu^*_{j,n}(B_2) = \int_{B_2} \exp\left\{ \langle \alpha, y \rangle - \frac{\sigma^2 ||\alpha||^2}{2} \right\} \rho_\sigma(dy).$$

Note that $\alpha = \alpha(\psi^*(j/n), \dot{\psi}^*(j/n))$ and $\dot{\psi}^*(j/n) = \int_{\mathcal{S}} b(\psi^*(j/n), \xi)\mu^*_{j,n}(d\xi) + \int_{\mathbb{R}^d} y\nu^*_{j,n}(dy)$. We observe that, from part (b) of Lemma B.2 in Appendix B,

$$\begin{aligned} L(x_j, \bar{\beta}^{1,n}_j) &\leq \int_{\mathcal{S}} R(\gamma^*_{j,n}(\cdot|\psi^*(j/n), \xi) \| p(\cdot|x_j, \xi))\mu^*_{j,n}(d\xi) \\ &= \langle \alpha(\psi^*(j/n), \dot{\psi}^*(j/n)), \bar{\beta}^n_j - \bar{\beta}^{2,n}_j \rangle - \Lambda(\psi^*(j/n), \alpha(\psi^*(j/n), \dot{\psi}^*(j/n))) \\ &\leq L(\psi^*(j/n), \bar{\beta}^n_j - \bar{\beta}^{2,n}_j). \end{aligned}$$

Now, from part (c) of Lemma B.3 and for $\bar{\beta}^n_j$ as in (4.6), the stochastic kernel $\gamma^{1,n}_j(\cdot|x_j, \xi)$ on $\mathcal{S}$ given $\mathcal{S} \times \mathbb{R}^d$ (with invariant measure μ^n_j) and the measure $\gamma^{2,n}_j$ on $\mathbb{R}^d$ given by

$$\gamma^{1,n}_j(B_1|x_j, \xi) = \int_{B_1} e^{\langle \alpha, b(x_j,\zeta) \rangle + \Psi_\sigma(x_j;\alpha,\zeta) - \Psi_\sigma(x_j;\alpha,\xi) - \Lambda(x_j,\alpha)} p(d\zeta|x_j, \xi) \tag{4.7}$$

and

$$\gamma^{2,n}_j(B_2) = \int_{B_2} e^{\langle \alpha, y \rangle - \frac{\sigma^2}{2}\|\alpha\|^2} \rho_\sigma(dy)$$

achieve the infimum in the representation for L_σ. That is,

$$L_\sigma(x_j, \bar{\beta}^n_j) = \int_{\mathcal{S}} R(\gamma^{1,n}_j(\cdot|x_j, \xi)||p(\cdot|x_j, \xi))\mu^n_j(d\xi) + R(\gamma^{2,n}_j(\cdot)||\rho_\sigma(\cdot)), \tag{4.8}$$

and

$$\bar{\beta}^n_j = \int_{\mathcal{S}} b(x_j, \xi)\mu^n_j(d\xi) + \int_{\mathbb{R}^d} y\gamma^{2,n}_j(dy) = \bar{\beta}^{1,n}_j + \bar{\beta}^{2,n}_j,$$

where we have used the fact that [4, Corollary C.3.3] for any probability measures γ and θ on $\mathcal{S}$, and λ and μ on $\mathbb{R}^d$,

$$R(\gamma \times \lambda \| \theta \times \mu) = R(\gamma\|\theta) + R(\lambda\|\mu). \tag{4.9}$$

Note that $\alpha = \alpha(x_j, \bar{\beta}^n_j) = \alpha(x_j, \psi^*(j/n), \dot{\psi}^*(j/n))$ in both $\gamma^{1,n}_j$ and $\gamma^{2,n}_j$. Hence $\gamma^{2,n}_j$ depends implicitly on x_j (through α), but we do not write this dependence explicitly for ease of notation.

We now use the kernels $\gamma^{1,n}_j$ and $\gamma^{2,n}_j$ to finish the definition of the required sequence of admissible controls. As was the case in section 3, grouping for the purposes of averaging motivates part of the construction. Let $\{m_n, n \in \mathbb{N}\}$ be a sequence as the one used there, so that $m_n \to \infty$ and $k_n = m_n/n \to 0$ as $n \to \infty$. Let $l \in \{0, \ldots, \frac{1}{k_n} - 1\}$. Then for $lm_n \leq j < (l+1)m_n - 1$ we define

$$\nu^{1,n}_j(d\zeta|x_0, \ldots, x_j, u_0, \ldots, u_j, \xi_j) \doteq \begin{cases} \gamma^{1,n}_{lm_n}(d\zeta|x_{lm_n}, \xi_j) & \text{if } \max\limits_{0\leq i\leq j}\|x_i - \psi^*(i/n)\| \leq \eta, \\ p(d\zeta|x_j, \xi_j) & \text{if } \max\limits_{0\leq i\leq j}\|x_i - \psi^*(i/n)\| > \eta \end{cases} \tag{4.10}$$

and

$$\nu_j^{2,n}(dy|x_0,\ldots,x_j,u_0,\ldots,u_j,\xi_j) \doteq \begin{cases} \gamma_{lm_n}^{2,n}(dy) & \text{if} \quad \max_{0\le i\le j}\|x_i-\psi^*(i/n)\|\le\eta, \\ \rho_\sigma(dy) & \text{if} \quad \max_{0\le i\le j}|x_i-\psi^*(i/n)|>\eta. \end{cases} \tag{4.11}$$

To simplify notation we have not made explicit the dependence on σ of $\nu_j^{1,n}$ and $\nu_j^{2,n}$. Finally, we define the required admissible control sequence $\{\nu_{j,prod}^n, j=0,\ldots,n-1\}$ on $\mathcal{S}\times\mathbb{R}^d$ as

$$\nu_{j,prod}^n(d\zeta\times dy)=\nu_j^{1,n}(d\zeta)\times\nu_j^{2,n}(dy). \tag{4.12}$$

To show that the control sequence just constructed is nearly optimal in (A.3), we compute the associated running cost directly. In what follows, $\bar{X}_j^n$, $\bar{Z}_j^n$, $j=0,\ldots,n$, are controlled random variables associated with the sequence $\{\nu_j^n\}$ through definitions (A.1) and (A.2).

Let $\tau^n \doteq \frac{1}{n}(\min\{j\in\{0,1,\ldots,n\}:\|\bar{X}_j^n-\psi^*(j/n)\|>\eta\}\wedge n)$. Then (4.9) and the definition of τ^n give

$$\begin{aligned}
&\bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1}R(\nu_{j,prod}^n(\cdot)||(p\times\rho_\sigma)(\cdot|\bar{X}_j^n,\bar{Z}_j^n))\right\} \\
&=\bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n\tau^n-1}\left[R(\nu_j^{1,n}(\cdot)\|p(\cdot|\bar{X}_j^n,\bar{Z}_j^n))+R(\nu_j^{2,n}(\cdot)\|\rho_\sigma(\cdot))\right]\right\} \\
&=\bar{E}_{x,\xi}\left\{\sum_{l=0}^{q_n-1}k_n\left[\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1}R(\gamma_{lm_n}^{1,n}(\cdot|\bar{X}_{lm_n}^n,\bar{Z}_j^n)\|p(\cdot|\bar{X}_j^n,\bar{Z}_j^n))+R(\gamma_{lm_n}^{2,n}(\cdot)\|\rho_\sigma(\cdot))\right]\right. \\
&\qquad\left.+\frac{1}{n}\sum_{j=q_nm_n}^{n\tau^n-1}\left[R(\gamma_{q_nm_n}^{1,n}(\cdot|\bar{X}_{q_nm_n}^n,\bar{Z}_j^n)\|p(\cdot|\bar{X}_j^n,\bar{Z}_j^n))+R(\gamma_{q_nm_n}^{2,n}(\cdot)||\rho_\sigma(\cdot))\right]\right\},
\end{aligned} \tag{4.13}$$

where q_n is such that $n\tau^n=q_nm_n+r_n$, with $0\le r_n<m_n$ and $q_n,r_n\in\mathbb{N}_0$.

To continue our estimates on the running costs, we must prove the following claim: for each $j\le n\tau^n-1$, $lm_n\le j\le(l+1)m_n-1$ for some $l\in\{0,\ldots,q_n\}$, and n large enough,

$$R(\gamma_{lm_n}^{1,n}(\cdot|\bar{X}_{lm_n}^n,\bar{Z}_j^n)\|p(\cdot|\bar{X}_j^n,\bar{Z}_j^n))\le R(\gamma_{lm_n}^{1,n}(\cdot|\bar{X}_{lm_n}^n,\bar{Z}_j^n)\|p(\cdot|\bar{X}_{lm_n}^n,\bar{Z}_j^n))+\sigma. \tag{4.14}$$

We first note that part (c) of Hypothesis H.1 implies that for any $x,y\in\Delta$, there exists $\delta>0$ such that for $\|x-y\|<\delta$

$$\frac{\tilde{p}^y(\xi,\zeta)}{\tilde{p}^x(\xi,\zeta)}=1+\frac{\tilde{p}^y(\xi,\zeta)-\tilde{p}^x(\xi,\zeta)}{\tilde{p}^x(\xi,\zeta)}\le1+\frac{\tilde{p}^y(\xi,\zeta)-\tilde{p}^x(\xi,\zeta)}{a}\le e^\sigma. \tag{4.15}$$

Then taking n large enough so that $m_n\|b\|_\infty/n<\delta$, we have $\|\bar{X}_{lm_n+i}^n-\bar{X}_{lm_n}^n\|\le\frac{i}{n}\|b\|_\infty<\delta$ for $0\le i<m_n$, and hence

$$\begin{aligned}
&\gamma_{lm_n}^{1,n}(B_1|\bar{X}_{lm_n}^n,\bar{Z}_j^n) \\
&=\int_{B_1}e^{\langle\alpha,b(\bar{X}_{lm_n}^n,\zeta)\rangle+\Psi_\sigma(\bar{X}_{lm_n}^n;\alpha,\zeta)-\Psi_\sigma(\bar{X}_{lm_n}^n;\alpha,\xi)-\Lambda(\bar{X}_{lm_n}^n,\alpha)}p(d\zeta|\bar{X}_{lm_n}^n,\bar{Z}_j^n) \\
&\le e^\sigma\int_{B_1}e^{\langle\alpha,b(\bar{X}_{lm_n}^n,\zeta)\rangle+\Psi_\sigma(\bar{X}_{lm_n}^n;\alpha,\zeta)-\Psi_\sigma(\bar{X}_{lm_n}^n;\alpha,\xi)-\Lambda(\bar{X}_{lm_n}^n,\alpha)}\tilde{p}^{\bar{X}_j^n}(\bar{Z}_j^n,\zeta)\vartheta(d\zeta)
\end{aligned}$$

for any $B_1 \in \mathcal{B}(\mathcal{S})$. From the above we get that $\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)$ is absolutely continuous with respect to $p(\cdot|\bar{X}^n_j, \bar{Z}^n_j)$ and that

$$\frac{d\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)}{dp(\cdot|\bar{X}^n_j, \bar{Z}^n_j)} \le e^{\sigma} \cdot \frac{d\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)}{dp(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)},$$

which implies (4.14).

Now fix $\bar{\xi} \in \mathcal{S}$ and normalize Ψ_σ in such a way that $\Psi_\sigma(x;\alpha,\bar{\xi}) = 0$. Then, observing that

$$\frac{a}{A} e^{\Psi_\sigma(x;\alpha,\xi_1)} \le e^{\Psi_\sigma(x;\alpha,\xi_2)} \le \frac{A}{a} e^{\Psi_\sigma(x;\alpha,\xi_1)}$$

for all $\xi_1, \xi_2 \in \mathcal{S}$, and taking $\xi_1 = \bar{\xi}$, we get that

$$\frac{a}{A} \le e^{\Psi_\sigma(x;,\alpha,\xi)} \le \frac{A}{a} \quad \forall \xi \in \mathcal{S},\ x,\ \alpha \in \mathbb{R}^d.$$

Then, from (4.7),

$$\gamma^{1,n}_{lm_n}(d\zeta|x,\xi) = e^{\langle \alpha, b(x,\zeta)\rangle + \Psi_\sigma(x;\alpha,\zeta) - \Psi_\sigma(x;\alpha,\xi) - \Lambda(x,\alpha)} \tilde{p}^x(\xi,\zeta)\vartheta(d\zeta),$$

with $\alpha = \alpha(x, \bar{\beta}^n_{lm_n})$, $x \in \Delta$, and $\bar{\beta}^n_{lm_n}$ to satisfy (4.6), and hence

$$\gamma^{1,n}_{lm_n}(d\zeta|x,\xi) \ge \frac{a^3}{A^2} e^{\langle \alpha, b(x,\zeta)\rangle - \Lambda(x,\alpha)} \vartheta(d\zeta) \ge \frac{a^3}{A^2} e^{-2||b||_\infty \max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}} \vartheta(d\zeta),$$

where $\Theta \doteq \cup_{t\in[0,1]}\{\beta \in \mathbb{R}^d : ||\beta - \dot{\psi}^*(t)|| \le K\}$. Using the fact that $(x,\beta) \mapsto \alpha(x,\beta)$ is continuous (part (f) of Lemma B.3), we get that for $x \in \Delta$, $\bar{\beta}^n_{\ell m_n}$ belongs to Θ and, moreover, that $\max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}$ is bounded. Denoting the jth iteration of the kernel $\gamma^{1,n}_{lm_n}$ by $\gamma^{1,n,j}_{lm_n}$, we observe that [13, Theorem 16.0.2]

$$\|\gamma^{1,n,j}_{lm_n}(\cdot|x,\zeta) - \mu^n_{lm_n}(\cdot)\| \le \left(1 - \frac{a^3}{A^2} e^{-2\max_{x,\zeta}|\langle \alpha, b(x,\zeta)\rangle|}\right)^j. \tag{4.16}$$

We complete the estimate on the running cost for our admissible control sequence in the inequalities that follow, using (4.14), standard properties of conditional expectation, and (4.7). We have that (4.13) is less than or equal to

$$\begin{aligned}
&\bar{E}_{x,\xi}\Bigg\{\sum_{l=0}^{q_n-1} k_n \left[\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} R(\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)||p(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)) + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))\right] \\
&\quad + \frac{1}{n}\sum_{j=q_nm_n}^{n\tau^n-1} \left[R(\gamma^{1,n}_{q_nm_n}(\cdot|\bar{X}^n_{q_nm_n}, \bar{Z}^n_{q_nm_n})||p(\cdot|\bar{X}^n_{q_nm_n}, \bar{Z}^n_j)) + R(\gamma^{2,n}_{q_nm_n}(\cdot)||\rho_\sigma(\cdot))\right]\Bigg\} \\
&\quad + \sigma \\
&\le \sum_{r=1}^{1/k_n} k_n \bar{E}_{x,\xi}\left\{\sum_{l=0}^{r-1}\left[\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} R(\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)||p(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j))\right.\right. \\
&\qquad \left.\left. + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))\right] 1_{[(r-1)m_n < n\tau^n \le rm_n]}\right\} + \sigma
\end{aligned} \tag{4.17}$$

$$
\begin{aligned}
= \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \bar{E}_{x,\xi} \Bigg\{ & 1_{[(r-1)m_n < n\tau^n \le rm_n]} \\
& \cdot \bar{E}_{x,\xi} \Bigg\{ \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n - 1} R(\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)||p(\cdot|\bar{X}^n_{lm_n}, \bar{Z}^n_j)) \\
& \quad + \; R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))|\bar{Z}^n_0, \ldots, \bar{Z}^n_{lm_n}, \bar{X}^n_0, \ldots, \bar{X}^n_{rm_n} \Bigg\}\Bigg\} \\
+ \sigma & \\
= \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \bar{E}_{x,\xi} \Bigg\{ & 1_{[(r-1)m_n < n\tau^n \le rm_n]} \\
& \cdot \bar{E}_{x,\xi} \Bigg\{ \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n - 1} \Bigg[\Big\langle \alpha_{lm_n}, \int_{\mathcal{S}} b(\bar{X}^n_{lm_n}, \zeta) \gamma^{1,n}_{lm_n}(d\zeta|\bar{X}^n_{lm_n}, \bar{Z}^n_j) \Big\rangle \\
& \quad + \int_{\mathcal{S}} \Psi_\sigma(\bar{X}^n_{lm_n}; \alpha_{lm_n}, \zeta) \gamma^{1,n}_{lm_n}(d\zeta|\bar{X}^n_{lm_n}, \bar{Z}^n_j) \\
& \quad - \Psi_\sigma(\bar{X}^n_{lm_n}; \alpha_{lm_n}, \bar{Z}^n_j) \\
& \quad - \; \Lambda(\bar{X}^n_{lm_n}, \alpha_{lm_n}) \Bigg] \\
& \quad + \; R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))|\bar{Z}^n_0, \ldots, \bar{Z}^n_{lm_n}, \bar{X}^n_0, \ldots, \bar{X}^n_{lm_n} \Bigg\}\Bigg\} \\
+ \sigma. &
\end{aligned}
$$

Now, adding and subtracting $\int_{\mathcal{S}} R(\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \xi)||p(\cdot|\bar{X}^n_{lm_n}, \xi))\mu^n_{lm_n}(d\xi)$ inside the expectation and collecting terms, we get that the above expression is less than or equal to

$$
\begin{aligned}
\sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \bar{E}_{x,\xi} \Bigg\{ \Bigg[& \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n - 1} \Big\langle \alpha_{lm_n}, \int_{\mathcal{S}} b(\bar{X}^n_{lm_n}, \zeta) \gamma^{1,n,j+1}_{lm_n}(d\zeta|\bar{X}^n_{lm_n}, \bar{Z}^n_{lm_n}) \\
& \qquad - \int_{\mathcal{S}} b(\bar{X}^n_{lm_n}, \zeta) \mu^n_{lm_n}(d\zeta) \Big\rangle \\
& + \int_{\mathcal{S}} R(\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n}, \xi)||p(\cdot|\bar{X}^n_{lm_n}, \xi))\mu_{lm_n}(d\xi) \\
& + \; R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot)) \Bigg] 1_{[(r-1)m_n < n\tau^n \le rm_n]} \Bigg\} + \frac{4}{n} \ln \frac{A}{a} + \sigma \qquad (4.18)
\end{aligned}
$$

$$
\begin{aligned}
\le \frac{4}{n} \ln \frac{A}{a} + \sigma + \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \bar{E}_{x,\xi} \Big\{ & 1_{[(r-1)m_n < n\tau^n \le rm_n]} \\
\cdot \Bigg[\frac{||b||_\infty A^2}{m_n a^3} \max_{x \in \Delta, \beta \in \Theta} \{\|\alpha(x,\beta)\|\} e^{2||b||_\infty \max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}} &
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_{\mathcal{S}} R(\gamma^{1,n}_{lm_n}(\cdot|\bar{X}^n_{lm_n},\xi)||p(\cdot|\bar{X}^n_{lm_n},\xi))\mu^n_{lm_n}(d\xi) + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))\Big]\Big]\Big\} \\
&\le \frac{4}{n}\ln\frac{A}{a} + \sigma + \frac{||b||_\infty A^2}{na^3}\max_{x\in\Delta,\beta\in\Theta}\{||\alpha(x,\beta)\|\}e^{2||b||_\infty \max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}} \\
&\quad + \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \bar{E}_{x,\xi}\Big\{1_{[(r-1)m_n<n\tau^n\le rm_n]}\Big[L_\sigma(\bar{X}^n_{lm_n},\bar{\beta}^n_{lm_n})\Big]\Big\} \\
&\le \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \bar{E}_{x,\xi}\Big\{1_{[(r-1)m_n<n\tau^n\le rm_n]}L_\sigma(\psi^*(lm_n/n),\dot{\psi}^*(lm_n/n))\Big\} \\
&\quad + \frac{4}{n}\ln\frac{A}{a} + 2\sigma + \frac{||b||_\infty A^2}{na^3}\max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}e^{2||b||_\infty \max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}} \\
&\le \bar{E}_{x,\xi}\left\{\sum_{l=0}^{[\frac{\tau^n}{k_n}]} k_n L_\sigma(\psi^*(lm_n/n),\dot{\psi}^*(lm_n/n))\right\} + 3\sigma \quad \text{for } n \text{ large enough}, \\
&\le k_n \sum_{l=0}^{[1/k_n]} L_\sigma(\psi^*(lm_n/n),\dot{\psi}^*(lm_n/n)) + 3\sigma.
\end{aligned}
$$

In the first, second, and third inequalities we have used (4.16), (4.8), and (4.6), respectively. We conclude that the admissible control sequence that we constructed has a running cost which is nearly optimal, as we had claimed.

We can now return to the proof of (4.5). Using (A.3), (4.18), and Lemma B.3(e) (with $\varepsilon = \sigma$), we get that

$$
\begin{aligned}
&\limsup_{n\to\infty} W^n_\sigma(x,\xi) \\
&\le \limsup_{n\to\infty} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=1}^{n-1}[R(\nu^{1,n}_j(\cdot)||p(\cdot|\bar{X}^n_j,\bar{Z}^n_j)) + R(\nu^{2,n}_j(\cdot)||\rho(\cdot))] + h(\bar{Y}^n)\right\} \\
&\le \int_0^1 L_\sigma(\psi^*(t),\dot{\psi}^*(t))dt + 3\sigma + \limsup_{n\to\infty}\bar{E}_{x,\xi}\{h(\bar{Y}^n)\} \\
&\le \int_0^1 L_\sigma(\psi(t),\dot{\psi}(t))dt + 4\sigma + \limsup_{n\to\infty}\bar{E}_{x,\xi}\{h(\bar{Y}^n)\}.
\end{aligned}
$$

Thus, the proof of (4.5) will be complete once we prove that

$$\limsup_{n\to\infty}\bar{E}_{x,\xi}\{h(\bar{Y}^n)\} \le h(\psi) + \tilde{\theta}(\sigma),$$

with $\tilde{\theta}(\sigma) \to 0$ when $\sigma \to 0$. This in turn will be implied by

$$\lim_{\sigma\to 0}\limsup_{n\to\infty}\bar{P}_{x,\xi}\left\{\sup_{t\in[0,1]}||\bar{Y}^n(t) - \psi^*(t)|| \ge \sigma\right\} = 0, \tag{4.19}$$

because of the Lipschitz property of h and part (e) of Lemma B.3.

To show (4.19), it is convenient to define a sequence of control measures associated with the controls $\nu^{1,n}_j$ and $\nu^{2,n}_j$ given in (4.10) and (4.11). For $B_1, B_2 \in \mathcal{B}(S)$, $B \in$

$\mathcal{B}(\mathbb{R}^d)$, define

$$\tilde{\nu}_l^{1,n}(B_1 \times B_2) \doteq \frac{1}{m_n} \sum_{j=l_n}^{(l+1)m_n - 1} \delta_{\bar{Z}_j^n}(B_1) \times \nu_j^{1,n}(B_2 | \bar{X}_0^n, \ldots, \bar{X}_j^n, \bar{Z}_j^n),$$

$$\tilde{\nu}_l^{2,n}(B) \doteq \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n - 1} \nu_j^{2,n}(B | \bar{X}_0^n, \ldots, \bar{X}_j^n),$$

$$\tilde{\nu}^{1,n}(\cdot|t) \doteq \begin{cases} \tilde{\nu}_l^{1,n}(\cdot) & \text{if} \quad t \in [lk_n, (l+1)k_n) \quad \text{for } l = 0, \ldots, 1/k_n - 2, \\ \tilde{\nu}_{(\frac{1}{k_n}-1)}^{1,n}(\cdot) & \text{if} \quad t \in [1-k_n, 1], \end{cases}$$

and

$$\tilde{\nu}^{2,n}(\cdot|t) \doteq \begin{cases} \tilde{\nu}_l^{2,n}(\cdot) & \text{if} \quad t \in [lk_n, (l+1)k_n) \quad \text{for } l = 0, \ldots, 1/k_n - 2, \\ \tilde{\nu}_{(\frac{1}{k_n}-1)}^{2,n}(\cdot) & \text{if} \quad t \in [1-k_n, 1]. \end{cases}$$

With B_1, B_2, and B as before, and with $C \in \mathcal{B}([0,1])$, now define the random measures $\nu^{1,n}$ and $\nu^{2,n}$ on $\mathcal{S} \times \mathcal{S} \times [0,1]$ and $\mathbb{R}^d \times [0,1]$, respectively, by

$$\nu^{1,n}(B_1 \times B_2 \times C) \doteq \int_C \nu^{1,n}(B_1 \times B_2|t)dt \quad \text{and} \quad \nu^{2,n}(B \times C) \doteq \int_C \nu^{2,n}(B|t)dt.$$

Finally, let ν_{prod}^n be the random measure on $\mathcal{S} \times \mathcal{S} \times \mathbb{R}^d \times [0,1]$ defined as

$$\nu_{prod}^n(B_1 \times B_2 \times B \times C) = \int_C \nu^{1,n}(B_1 \times B_2|t)\nu^{2,n}(B|t)dt. \tag{4.20}$$

Let us also define

$$S^{1,n}(t) \doteq x + \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(S^{1,n}(s), y)\nu^{1,n}(d\zeta \times dy \times ds)$$

and

$$S^{2,n}(t) \doteq \int_{\mathbb{R}^d\times[0,t]} y\nu^{2,n}(dy \times ds).$$

Since L_σ is continuous, ψ^* is continuous and $\dot{\psi}^*$ has only a finite number of discontinuities,

$$\sup_n \bar{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^{1,n}(\cdot) \times \nu_j^{2,n}(\cdot) || p(\cdot|\bar{X}_j^n, \bar{Z}_j^n) \times \rho_\sigma(\cdot)) \right\} < \infty.$$

Theorem 5.3.5 in [4], the fact that $\mathcal{S}$ and $[0,1]$ are compact, and arguments analogous to the proof of Theorem C.1(e) then imply that given any subsequence of $\{(\nu^{1,n}, \nu^{2,n}, \bar{X}^n, \bar{U}^n, \tau^n, S^{1,n}, S^{2,n}), n \in \mathbb{N}\}$ there exists a subsubsequence such that $(\nu^{1,n}, \nu^{2,n}, \bar{X}^n, \bar{U}^n, \tau^n, S^{1,n}, S^{2,n})$ converges in distribution to $(\nu^1, \nu^2, \bar{X}, \bar{U}, \tau, \bar{X}, \bar{U})$ when $n \to \infty$. We define $\bar{Y}(t) \doteq \lim_{n\to\infty}(\bar{X}^n(t) + \bar{U}^n(t))$.

From the definition of $\bar{\beta}_j^{1,n}$ and Lemma B.3(d), for each $l \in \{0, \ldots, [\frac{\tau^n}{k_n}]\}$,

$$\begin{aligned} ||\bar{\beta}_{lm_n}^n - \dot{\psi}^*(lm_n/n)|| &= ||\bar{\beta}_{lm_n}^{1,n} + \bar{\beta}_{lm_n}^{2,n} - \dot{\psi}^*(lm_n/n)|| \\ &= \left\| \int_{\mathcal{S}} b(\bar{X}_{lm_n}^n, \xi)\mu_{lm_n}^n(d\xi) + \int_{\mathbb{R}^d} y\tilde{\nu}_{lm_n}^{2,n}(dy) - \dot{\psi}^*(lm_n/n) \right\| \\ &\leq K||\bar{X}_{lm_n}^n - \psi^*(lm_n/n)||. \end{aligned}$$

Then for $t \in [0, \tau]$

$$
\begin{aligned}
||\bar{Y}(t) - \psi^*(t)|| = \lim_{n\to\infty} \Bigg\| \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} & b(S^{1,n}(s), y)\nu^{1,n}(d\zeta \times dy \times ds) \\
& - \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\tilde{X}^n(s), y)\nu^{1,n}(d\zeta \times dy \times ds) \\
& + \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\tilde{X}^n(s), y)\nu^{1,n}(d\zeta \times dy \times ds) \\
& + \sum_{l=0}^{[\frac{t}{k_n}]} k_n \int_{\mathbb{R}^d} y\tilde{\nu}^{2,n}_{lm_n}(dy) - \sum_{l=0}^{[\frac{t}{k_n}]} k_n \int_{\mathcal{S}} b(\tilde{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi) \\
& + \sum_{l=0}^{[\frac{t}{k_n}]} k_n \int_{\mathcal{S}} b(\tilde{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi) - \int_0^t \dot{\psi}^*(s)ds \Bigg\| \\
\leq \limsup_{n\to\infty} \int_0^1 & (K||S^{1,n}(s) - \tilde{X}^n(s)|| \wedge 2\|b\|_\infty)ds \\
& + \limsup_{n\to\infty} \sum_{l=0}^{[\frac{t}{k_n}]} k_n \Bigg\| \frac{1}{m_n} \sum_{j=l}^{(l+1)m_n} \int_{\mathcal{S}} b(\bar{X}_{lm_n}, \zeta)\gamma^{1,n,j}_{lm_n}(d\zeta|\bar{X}^n_{lm_n}, \bar{Z}_{lm_n}) \\
& \qquad\qquad - \int_{\mathcal{S}} b(\bar{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi) \Bigg\| \\
& + \limsup_{n\to\infty} K \int_0^t ||\bar{X}^n(s) - \psi^*(s)||ds \\
\leq \limsup_{n\to\infty} & \sum_{l=0}^{[\frac{t}{k_n}]} k_n ||b||_\infty \frac{A^2}{na^3} e^{2||b||_\infty \max_{x\in\Delta,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}} \\
& + K \int_0^t ||\bar{X}(s) - \psi^*(s)||ds \\
\leq K \sup_{s\in[0,\tau]} & ||\bar{X}(s) - \bar{Y}(s)|| + K \int_0^t ||\bar{Y}(s) - \psi^*(s)||ds.
\end{aligned}
$$

Gronwall's inequality with $\bar{K} \doteq Ke^K$ gives

$$
\sup_{s\in[0,\tau]} ||\bar{Y}(s) - \psi^*(s)|| \leq \bar{K} \sup_{s\in[0,\tau]} ||\bar{X}(s) - \bar{Y}(s)|| = \bar{K} \sup_{s\in[0,\tau]} ||\bar{U}(s)||,
$$

which together with Lemma C.2 implies that

$$
\lim_{\sigma\to 0} \bar{P}_{x,\xi} \left\{ \sup_{s\in[0,\tau]} ||\bar{Y}(s) - \psi^*(s)|| \geq \frac{\sigma}{2} \right\} \leq \lim_{\sigma\to 0} \bar{P}_{x,\xi} \left\{ \sup_{s\in[0,\tau]} ||\bar{U}(s)|| \geq \frac{\sigma}{2\bar{K}} \right\} = 0.
$$

Finally, writing $\tau = \tau_\sigma$ and following the same arguments given in [4, pp. 205–206], it is proved that $\lim_{\sigma\to 0} \bar{P}_{x,\xi}\{\tau_\sigma < 1\} = 0$. Since $\bar{Y}^n \to \bar{Y}$ w.p.1 uniformly on $[0, 1]$, we obtain (4.19), which completes the proof of the lower bound. □

5. The case of noncompact $\mathcal{S}$. Without the assumption of compactness of $\mathcal{S}$, a strong positive recurrence hypothesis on $p(d\zeta|x,\xi)$ is required to guarantee tightness of the measures appearing in the proof of Theorem 2.1. This hypothesis is analogous to Condition 8.2.2 in [4].

Hypothesis H.2. There exists a measurable function $U : \mathcal{S} \to [0,\infty)$ with the following properties:

(a) $\inf_{\zeta\in\mathcal{S}} \left\{U(\zeta) - \log \int_{\mathcal{S}} e^{U(\zeta)} p(d\zeta|x,\xi)\right\} > -\infty$.

(b) For each $M < \infty$ and compact set $\Delta \subset \mathbb{R}^d$, the set

$$Z(M,\Delta) = \left\{(\xi,y) \in \mathcal{S}\times\Delta \,:\, c(\xi,y) := U(\xi) - \log \int_{\mathcal{S}} e^{U(\zeta)} p(d\zeta|y,\xi) \le M\right\}$$

is a compact subset of $\mathcal{S}\times\mathbb{R}^d$.

(c) U is bounded above on every compact subset of $\mathcal{S}$.

Under Hypothesis H.2, part (a) of Theorem C.1 remains valid for $\mathcal{S}$ noncompact, which implies Theorem 2.1 as well. Because the proof requires only small changes in the proofs of Lemma 8.2.4 and Proposition 8.2.5 in [4], we omit the details.

Appendix A. A representation formula. In this appendix we state the variational representation formula required in the proof of the lower bound. It can be derived easily by following the same steps as those given in [4, section 4.4].

Let $p\times\rho_\sigma$ be the stochastic kernel on $S\times\mathbb{R}^d$ given $\xi\in\mathcal{S}$, $x\in\mathbb{R}^d$ defined by $(p\times\rho_\sigma)(d\zeta\times dy|x,\xi) \doteq p(d\zeta|x,\xi)\times\rho_\sigma(dy)$. We consider admissible control sequences consisting of stochastic kernels $\nu_j^n(d\zeta\times dy|\bar{X}_0^n,\ldots,\bar{X}_j^n,\bar{U}_0^n,\ldots,\bar{U}_j^n,\bar{Z}_j^n)$ on $\mathcal{S}\times\mathbb{R}^d$ given $(\mathbb{R}^d)^{j+1}\times(\mathbb{R}^d)^{j+1}\times\mathcal{S}$. For each admissible control sequence $\{\nu_j^n, j=0,\ldots,n-1\}$, the controlled system is defined by setting $\bar{X}_0^n \doteq x$, $\bar{U}_0^n \doteq 0$ and for $j=0,\ldots,n-1$ through

$$\bar{X}_{j+1}^n \doteq \bar{X}_j^n + \frac{1}{n}\, b(\bar{X}_j^n,\bar{Z}_{j+1}^n),\ \bar{U}_{j+1}^n \doteq \bar{U}_j^n + \frac{1}{n}\,\bar{G}_j^n,\ \text{and } \bar{Y}_j^n = \bar{X}_j^n + \bar{U}_j^n, \tag{A.1}$$

where the conditional distribution of $(\bar{Z}_{j+1}^n,\bar{G}_j^n)$ is given by

$$\begin{aligned}\bar{P}_{x,\xi}&\left\{(\bar{Z}_{j+1}^n,\bar{G}_j^n)\in(d\zeta\times dy)|\bar{X}_0^n,\ldots,\bar{X}_j^n,\bar{U}_0^n,\ldots,\bar{U}_j^n,\bar{Z}_j^n\right\}\\ &= \nu_j^n(d\zeta\times dy|\bar{X}_0^n,\ldots,\bar{X}_j^n,\bar{U}_0^n,\ldots,\bar{U}_j^n,\bar{Z}_j^n).\end{aligned} \tag{A.2}$$

We define the processes $\bar{X}^n \doteq \{\bar{X}^n(t), t\in[0,1]\}$, $\bar{U}^n = \{\bar{U}^n(t), t\in[0,1]\}$, and $\bar{Y}^n \doteq \{\bar{Y}^n(t), t\in[0,1]\}$ as the linear interpolations of $\{\bar{X}_j^n\}$, $\{\bar{U}_j^n\}$, and $\{\bar{Y}_j^n\}$, respectively.

THEOREM A.1. *Let $W_\sigma^n(x,\xi) \doteq -1/n \log E_{x,\xi}\{\exp[-nh(Y_\sigma^n)]\}$, where Y_σ^n is defined by (4.1), $E_{x,\xi}$ denotes expectation conditioned on $X_0^n = x$ and $Z_0^n=\xi$, and h is a bounded measurable function mapping $\mathcal{C}([0,1]:\mathbb{R}^d)\mapsto\mathbb{R}$. Then for all $n\in\mathbb{N}$, $x\in\mathbb{R}^d$, $\xi\in\mathcal{S}$, and $\sigma>0$, we have the representation*

$$W_\sigma^n(x,\xi) = \inf_{\{\nu_j^n\}} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu_j^n(\cdot)\|(p\times\rho_\sigma)(\cdot|\bar{X}_j^n,\bar{Z}_j^n)) + h(\bar{Y}^n)\right\}. \tag{A.3}$$

Appendix B. Properties of the functions $\boldsymbol{\Lambda}$, $\boldsymbol{L}$, and $\boldsymbol{L_\sigma}$. In this appendix we establish properties of the functions $\Lambda(x,\alpha)$ and $L(x,\beta)$ defined in (2.3) and (2.2), respectively, and of the function L_σ defined in (4.2).

LEMMA B.1. *Under Hypothesis* H.1*, the function* $\Lambda(x,\alpha)$ *defined in* (2.3) *satisfies the following properties. For each* $x \in \mathbb{R}^d$, $\Lambda(x,\alpha)$ *is a finite strictly convex function of* $\alpha \in \mathbb{R}^d$ *which is differentiable for all* α. *In addition,* $\Lambda(x,\alpha)$ *is a continuous function of* $(x,\alpha) \in \mathbb{R}^d \times \mathbb{R}^d$.

These properties follow from Lemmas 3.1 and 3.4 in [7] given the relation between the function Λ and the solution to the eigenvalue problem given in (2.1).

Lemma 2.1 in [3] gives a list of properties that are satisfied by the function L, the Legendre–Fenchel transform of Λ. These include convexity and lowersemicontinuity in β, positivity, and uniqueness. Part (a) of the following lemma is also among those properties, and we state it here for use in the proof of part (b), which provides an important variational representation for the function L.

LEMMA B.2. *Under Hypothesis* H.1*, the function* $L(x,\beta)$ *defined in* (2.2) *satisfies the following properties.*

(a) *If* $L(x,\beta)$ *is finite in a neighborhood of* β', *then* $\nabla L(x,\beta')$ *exists and* $L(x,\beta') = \langle \alpha, \beta' \rangle - \Lambda(x,\alpha)$ *if and only if* $\alpha = \nabla L(x,\beta')$.

(b) *For each* x *and* β *in* $\mathbb{R}^d$,

$$L(x,\beta) = \inf\left\{ \int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\|p(\cdot|x,\xi))\mu(d\xi) : \gamma\mu = \mu, \int_{\mathcal{S}} b(x,\xi)d\mu = \beta \right\}.$$

If L *is finite, then the infimum is attained uniquely.*

Proof. (a) Since $\Lambda(x,\cdot)$ is strictly convex on $\mathbb{R}^d$, $L(x,\cdot)$ is differentiable on $\mathrm{int}(\mathrm{dom}L(x,\cdot))$. See Theorem D.2.8 in [4]. The last part follows from standard results.

(b) First we consider the case when $\beta \in \mathrm{ri}(\mathrm{dom}L(x,\cdot))$. For $\alpha \in \mathbb{R}^d$ let γ_α be the stochastic kernel defined by

$$\frac{d\gamma_\alpha(\cdot|x,\xi)}{dp(\cdot|x,\xi)}(\zeta) = \frac{e^{\langle \alpha, b(x,\zeta)\rangle + \psi(\zeta)}}{\int_{\mathcal{S}} e^{\langle \alpha, b(x,\zeta)\rangle + \psi(\zeta)} p(d\zeta|x,\xi)}.$$

In terms of the function Λ defined in (2.1) we can write

$$\frac{d\gamma_\alpha(\cdot|x,\xi)}{dp(\cdot|x,\xi)}(\zeta) = e^{-\Lambda(x,\alpha) - \psi(\xi) + \langle \alpha, b(x,\zeta)\rangle + \psi(\zeta)}.$$

Let μ^α be the unique invariant measure of γ_α. (Proposition 4.1 in [7] guarantees that such a measure exists.) Part (a) of the present lemma and the fact that $\Lambda(x,\cdot)$ is strictly convex and differentiable imply that there exists a unique $\alpha = \alpha(x,\beta)$ such that

$$L(x,\beta) = \langle \alpha(x,\beta), \beta \rangle - \Lambda(x, \alpha(x,\beta)) \tag{B.1}$$

with $\alpha(x,\beta) \in \partial L(\beta)$ if and only if $\beta = \nabla\Lambda(x,\alpha(x,\beta))$ (see Corollary 26.3.1 in [14]). Then, Proposition 4.1 in [7] gives

$$E_{\mu^\alpha}^{\gamma_\alpha} b(x,\xi) = \int_{\mathcal{S}} b(x,\xi)\mu^\alpha(d\xi) = \beta. \tag{B.2}$$

Now let γ be any kernel (with corresponding invariant measure μ^γ) satisfying

$$\int_{\mathcal{S}} b(x,\xi)\mu^\gamma(d\xi) = \beta \quad \text{and} \quad \int_{\mathcal{S}} R(\gamma(\cdot|\xi)\|p(\cdot|x,\xi))\mu^\gamma(\xi) < \infty.$$

Then $\gamma(\cdot|x,\xi) \ll p(\cdot|x,\xi)$ for μ^γ-almost all ξ. Since $\frac{d\gamma_\alpha}{dp}$ is strictly positive, $\gamma(\cdot|x,\xi) \ll \gamma_\alpha(\cdot|x,\xi)$ for almost all ξ (with respect to μ^γ) and

$$\begin{aligned}
&\int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\|p(\cdot|x,\xi))\mu^\gamma(d\xi)\\
&= \int_{\mathcal{S}}\int_{\mathcal{S}} \log\frac{d\gamma(\cdot|x,\xi)}{dp(\cdot|x,\xi)}(\zeta)\gamma(d\zeta|x,\xi)\mu^\gamma(d\xi)\\
&= \int_{\mathcal{S}}\left[\int_{\mathcal{S}} \log\frac{d\gamma(\cdot|x,\xi)}{d\gamma_\alpha(\cdot|x,\xi)}(\zeta)\gamma(d\zeta|x,\xi) + \int_{\mathcal{S}} \log\frac{d\gamma_\alpha(\cdot|x,\xi)}{dp(\cdot|x,\xi)}(\zeta)\gamma(d\zeta|x,\xi)\right]\mu^\gamma(d\xi)\\
&= \int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\|\gamma_\alpha(\cdot|x,\xi))\mu^\gamma(d\xi)\\
&\quad + \int_{\mathcal{S}}\int_{\mathcal{S}} \log\Big[\exp\left[\langle\alpha, b(x,\zeta)\rangle + \psi(\zeta) - \psi(\xi) - \Lambda(x,\alpha)\right]\Big]\gamma(d\zeta|x,\xi)\mu^\gamma(d\xi)\\
&= \int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\|\gamma_\alpha(\cdot|x,\xi))\mu^\gamma(d\xi) - \Lambda(x,\alpha)\\
&\quad - \int_{\mathcal{S}} \psi(\xi)\mu^\gamma(d\xi) + \int_{\mathcal{S}}\int_{\mathcal{S}} \left[\langle\alpha, b(x,\zeta)\rangle + \psi(\zeta)\right]\gamma(d\zeta|x,\xi)\mu^\gamma(d\xi)\\
&= \int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\|\gamma_\alpha(\cdot|x,\xi))\mu^\gamma(d\xi) - \Lambda(x,\alpha) + \int_{\mathcal{S}} \langle\alpha, b(x,\xi)\rangle\mu^\gamma(d\xi)\\
&= \int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\|\gamma_\alpha(\cdot|x,\xi))\mu^\gamma(d\xi) + L(x,\beta) \geq L(x,\beta).
\end{aligned}$$

Equality is obtained if and only if $\gamma \equiv \gamma_\alpha$. If β does not belong to $\mathrm{ri}(\mathrm{dom} L(x,\cdot))$, analogous arguments to those given in Appendix C.5 in [4] can be adapted. ∎

The next result establishes properties of the function L_σ defined in (4.2) that are needed in the proof of the lower bound.

LEMMA B.3. *Given $\sigma > 0$, the function $L_\sigma(x,\beta)$ satisfies the following properties:*

(a) $L_\sigma(x,\beta) = \inf_{z\in\mathbb{R}^d}\{L(x,\beta - z) + \frac{||z||^2}{2\sigma^2}\}$ *and* $L_\sigma(x,\beta) \leq L(x,\beta)$.

(b) $L_\sigma(x,\beta)$ *is a finite, nonnegative, continuous function of* $(x,\beta) \in \mathbb{R}^d \times \mathbb{R}^d$. *Moreover,* $L_\sigma(x,\cdot)$ *is differentiable on* $\mathbb{R}^d$.

(c)

$$\begin{aligned}
L_\sigma(x,\beta) = \inf\bigg\{ &\int_{\mathcal{S}} R(\gamma(\cdot|x,\xi)\times \upsilon(\cdot)||p(\cdot|x,\xi)\times\rho_\sigma(\cdot))\mu(d\xi) :\\
&\mu\gamma = \mu, \int_{\mathcal{S}} b(x,\xi)\mu(d\xi) + \int_{\mathbb{R}^d} y\upsilon(dy) = \beta\bigg\}.
\end{aligned}$$

Further, for each $x, \beta \in \mathbb{R}^d$ there exist a stochastic kernel γ^* and a measure υ^* such that the infimum on the right-hand side is achieved. For Borel sets B_1 of $\mathcal{S}$ and B_2 of $\mathbb{R}^d$, these are given by

$$\gamma^*(B_1|x,\xi) \doteq \int_{B_1} e^{\langle\alpha, b(x,\zeta)\rangle - \Lambda(x,\alpha) - \Psi_\sigma(x;\alpha,\xi) + \Psi_\sigma(x;\alpha,\zeta)} p(d\zeta|x,\xi)$$

and

$$\upsilon^*(B_2) \doteq \int_{B_2} e^{\langle\alpha,y\rangle - \frac{\sigma^2||\alpha||^2}{2}}\rho_\sigma(dy),$$

with $\alpha = \alpha(x,\beta) \in \mathrm{argmax}\{\langle\alpha,\beta\rangle - \Lambda_\sigma(x,\alpha) : \alpha \in \mathbb{R}^d\}$.

(d) *Given any compact set* $\Delta \subset \mathbb{R}^d$ *and* $\varepsilon \in (0,1)$, *there exists* $\eta \in (0,1)$ *such that, whenever* $x, y \in \Delta$, $\beta \in \mathbb{R}^d$, *and* $||x-y|| \leq \eta$, *there exists* $\bar{\beta} \in \mathbb{R}^d$ *such that*

$$L_\sigma(y,\bar{\beta}) - L_\sigma(x,\beta) \leq \varepsilon \quad \text{and} \quad ||\bar{\beta}-\beta|| \leq K||x-y||,$$

where K *is the Lipschitz constant of* b.

(e) *Given* $\psi \in \mathcal{C}([0,1] : \mathbb{R}^d)$ *satisfying* $I_x(\psi) < \infty$ *and* $\varepsilon > 0$, *there exists* $\psi^* \in \mathcal{C}([0,1] : \mathbb{R}^d)$, *with* $\dot{\psi}^*$ *piecewise constant with only finitely many jumps in the interval* $(0,1)$, *such that* $||\psi - \psi^*||_\infty < \varepsilon$ *and*

$$\int_0^1 L_\sigma(\psi^*(t),\dot{\psi}^*(t))dt \leq \int_0^1 L_\sigma(\psi(t),\dot{\psi}(t))dt + \varepsilon \leq I_x(\psi) + \varepsilon.$$

(f) *The function* $(x,\beta) \to \alpha(x,\beta) \in \operatorname{argmax}\{\langle\alpha,\beta\rangle - \Lambda_\sigma(x,\alpha) : \alpha \in \mathbb{R}^d\}$ *is continuous.*

Proof. (a) The first statement follows from Corollary D.4.2 in [4], while for the second part we take $z = 0$.

(b) From Theorem 26.4 in [14] and Lemma 3.4(iv) in [7], we have that $\operatorname{int}(\operatorname{Dom} L_\sigma(x,\cdot)) = \operatorname{Range}(\nabla\Lambda_\sigma(x,\cdot)) = \mathbb{R}^d$. So $L_\sigma(x,\beta) < \infty$ for all $(x,\beta) \in \mathbb{R}^d \times \mathbb{R}^d$. The nonnegativity of $L_\sigma(x,\beta)$ follows from the nonnegativity of $L(x,\beta)$. The continuity follows from Lemma C.8.1 in [4] and the continuity of $\Lambda(x,\alpha)$ in both variables. Finally, the differentiability follows from the strict convexity of Λ and Theorem D.2.8 in [4].

(c) Let $x, \beta \in \mathbb{R}^d$. From part (b) there exists $\alpha \in \mathbb{R}^d$, with $\alpha = \alpha(x,\beta)$, such that $\beta = \nabla_\alpha \Lambda_\sigma(x,\alpha)$ and $L_\sigma(x,\beta) = \langle\alpha,\beta\rangle - \Lambda_\sigma(x,\alpha)$. Let

$$\gamma_\alpha(d\zeta \times dy|x,\xi) = \exp\Big\{\langle\alpha, b(x,\zeta)\rangle + \Psi(x;\alpha,\zeta) - \Psi(x;\alpha,\xi) - \Lambda(x,\alpha) - \frac{\sigma^2}{2}||\alpha||^2\Big\} p(d\zeta|x,\xi)\rho_\sigma(dy).$$

From Proposition 4.1 in [7],

$$\beta = \int_{\mathcal{S}} b(x,\xi)\mu(d\xi) + \int_{\mathcal{S}} y e^{\langle\alpha,y\rangle - \frac{\sigma^2}{2}||\alpha||^2}\rho_\sigma(dy),$$

where μ is the unique invariant measure of the first marginal of γ_α given by

$$\gamma_{\alpha,1}(d\zeta|x,\xi;\alpha) = \exp\{\langle\alpha, b(x,\zeta)\rangle + \Psi(x;\alpha,\zeta) - \Psi(x;\alpha,\xi) - \Lambda(x,\alpha)\} p(d\zeta|x,\xi).$$

The rest of the proof follows the same arguments given in the proof of Lemma B.2(b) after (B.2).

(d) Let $\Delta \subset \mathbb{R}^d$ compact, $x, y \in \Delta$, $\beta \in \mathbb{R}^d$, and $\varepsilon \in (0,1)$. We know from part (c) that there exist γ^* and υ^* such that the infimum in part (c) is attained for (x,β). Define $\bar{\beta} \doteq \int_{\mathcal{S}} b(y,\xi)\mu^{\gamma^*}(d\xi) + \int_{\mathbb{R}^d} y\nu^*(dy)$. Then, from the representation formula given in part (c),

$$||\beta - \bar{\beta}|| \leq \left|\int_{\mathcal{S}} (b(x,\xi) - b(y,\xi))\mu^{\gamma^*}(d\xi)\right| \leq K\|x-y\|.$$

Now, for any Borel set B of $\mathcal{S}$, we can use part (c) of Hypothesis H.1 to write

$$\gamma^*(B|x,\xi) = \int_B \exp\Big\{\langle\alpha, b(x,\zeta)\rangle + \Psi_\sigma(x;\alpha,\zeta) - \Psi_\sigma(x;\alpha,\xi) - \Lambda(x,\alpha)\Big\}\tilde{p}^x(\xi,\zeta)\vartheta(d\zeta).$$

From the bound that we have on $\tilde{p}^x(\cdot,\cdot)$, it follows that $\gamma^*(\cdot|x,\xi)$ is absolutely continuous with respect to $p(\cdot|y,\xi)$; from the uniform continuity of $\tilde{p}^x(\xi,\zeta)$, there exists $\eta > 0$ such that $\|x-y\| < \eta$ implies that $\tilde{p}^x(\xi,\zeta) \le \tilde{p}^y(\xi,\zeta)e^{\varepsilon}$ (this is as in (4.15)). Then, from the variational equivalence given in part (c), $L_\sigma(y,\bar{\beta}) < \infty$ and

$$\begin{aligned} L_\sigma(y,\bar{\beta}) &\le \int_{\mathcal{S}} R(\gamma^*(\cdot|x,\xi)\times\nu^*(\cdot)||p(\cdot|y,\xi)\times\rho_\sigma(\cdot))\mu^{\gamma^*}(d\xi) \\ &\le \left\langle \alpha, \int_{\mathcal{S}} b(x,\xi)\mu^{\gamma^*}(d\xi)\right\rangle - \Lambda(x,\alpha) + \left\langle \alpha, \int_{\mathbb{R}^d} y\nu^*(dy)\right\rangle - \frac{\sigma^2}{2}||\alpha||^2 + \varepsilon \\ &= \langle\alpha,\beta\rangle - \Lambda_\sigma(x,\alpha) + \varepsilon = L_\sigma(x,\beta)+\varepsilon. \end{aligned}$$

(e) The proof of this part is based on Lemmas 6.5.3 and 6.5.5 of [4], which in our case also hold due to the structural properties given in parts (a) and (b).

(f) Given x, $\beta \in \mathbb{R}^d$, part (b) and the differentiability of $\alpha \to \Lambda_\sigma(x,\alpha)$ imply that there exists a unique $\alpha(x,\beta)$ such that $L_\sigma(x,\beta) = \langle\alpha(x,\beta),\beta\rangle - \Lambda_\sigma(x,\alpha(x,\beta))$, $\beta = \nabla_\alpha\Lambda_\sigma(x,\alpha(x,\beta))$, and $\alpha(x,\beta) = \nabla_\beta L_\sigma(x,\beta)$. We observe that $\beta \to L_\sigma(x,\beta)$ is continuously differentiable thanks to [14, Corollary 25.5.1]. Moreover, $x \to \nabla_\beta L_\sigma(x,\beta)$ is continuous [14, Theorem 25.7] and, in fact, $(x,\beta) \to \nabla_\beta L_\sigma(x,\beta)$ is continuous by the same theorem. Therefore, $(x,\beta) \to \alpha(x,\beta)$ is continuous in both variables. This completes the proof of the lemma. ∎

Appendix C. Proofs of some limit results. This appendix is dedicated to the proofs of some limit results needed in the proof of Theorem 2.1.

THEOREM C.1. *Let $\mathcal{S}$ be compact. For any $x \in \mathbb{R}^d$, $\xi \in \mathcal{S}$ and each $n \in \mathbb{N}$, consider any admissible control sequence such that*

$$\sup_{n\in\mathbb{N}} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu_j^n(\cdot)\|p(\cdot|\bar{X}_j^n,\bar{Z}_j^n))\right\} < \infty,$$

where $\nu_j^n(\cdot) = \nu_j^n(\cdot|\bar{X}_0^n,\ldots,\bar{X}_j^n,\bar{Z}_j^n)$. In terms of these sequences we define the piecewise linear interpolation $\{\bar{X}^n\}$, the piecewise constant interpolation $\{\tilde{X}^n\}$, the sequence of admissible control measures $\{\nu^n\}$ and its marginals $\{\hat{\nu}_2^n\otimes\lambda, \hat{\nu}_1^n\otimes\lambda\}$, and the measures $\{\gamma^n\}$ as in section 3. Also, for each $n \in \mathbb{N}$ we define the process $S^n = \{S^n(t), t\in[0,1]\}$ by

$$S^n(t) \doteq x + \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(S^n(s),y)\nu^n(d\zeta\times dy\times ds).$$

The following conclusions hold.

(a) *Given any subsequence of $\{(\nu^n,\hat{\nu}_2^n\otimes\lambda,\hat{\nu}_1^n\otimes\lambda,\gamma^n,\bar{X}^n,\tilde{X}^n,S^n), n\in\mathbb{N}\}$, there exist a subsubsequence, a stochastic kernel ν on $\mathcal{S}\times\mathcal{S}\times[0,1]$ (given $\bar{\Omega}$) with marginals μ_1, μ_2, a stochastic kernel γ on $\mathcal{S}\times\mathcal{S}\times[0,1]$, and random variables $\bar{X}$ and S mapping $\bar{\Omega}$ into $\mathcal{C}([0,1]:\mathbb{R}^d)$ such that the subsubsequence converges in distribution to $(\nu,\mu_1,\mu_2,\gamma,\bar{X},\bar{X},S)$.*

(b) *The stochastic kernel ν has the decomposition*

$$\nu(B_1\times B_2\times C) = \int_C \nu(B_1\times B_2|t)dt = \int_C\int_{B_1}\hat{\nu}_1(d\zeta|t)\hat{\nu}_2(B_2|\zeta,t)dt$$

for some stochastic kernels $\nu(\cdot|t)$ on $\mathcal{S}\times\mathcal{S}$ given $[0,1]$, $\hat{\nu}_1(\cdot|t)$ on $\mathcal{S}$ given $[0,1]$, and $\hat{\nu}_2(\cdot|\zeta,t)$ on $\mathcal{S}$ given $\mathcal{S}\times[0,1]$.

(c) *We have the equality $\hat{\nu}_1\otimes\lambda = \hat{\nu}_2\otimes\lambda$.*

(d) *$\hat{\nu}_1(d\zeta|t)$ is an invariant measure of $\hat{\nu}_2(dy|\zeta,t)$ for each $t\in[0,1]$.*

(e) *W.p.1 for every* $t \in [0,1]$

$$\bar{X}(t) = x + \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu(d\zeta \times dy \times ds) \tag{C.1}$$
$$= x + \int_0^t \int_{\mathcal{S}} b(\bar{X}(s), \zeta)\hat{\nu}_1(d\zeta|s)ds.$$

(f) *The stochastic kernel* γ *has the decomposition*

$$\gamma(B_1 \times B_2 \times C) = \int_C \int_{B_1\times B_2} \hat{\nu}_1(d\zeta|t) \otimes p(dy|\bar{X}(t), \zeta)dt.$$

Proof. (a) Given the compactness of $\mathcal{S}$, we immediately get tightness of ν^n, of γ^n, and of all the marginals. Tightness of $\{\bar{X}^n\}$ and of $\{\tilde{X}^n\}$ on $\mathcal{C}([0,1] : \mathbb{R}^d)$ follows from the bound

$$w_{\bar{X}^n}(\delta) \doteq \sup_{\{s,t\in[0,1]:|s-t|\le\delta\}} \|\bar{X}^n(t) - \bar{X}^n(s)\| \le 2\|b\|_\infty \delta, \tag{C.2}$$

and tightness of $\{S^n\}$ can be verified similarly. Since the function mapping ν^n into $(\nu^n, \hat{\nu}_1^n \otimes \lambda, \hat{\nu}_2^n \otimes \lambda)$ is continuous, there exist measures ν and γ over $\mathcal{S}\times\mathcal{S}\times[0,1]$, measures μ_1 and μ_2 over $\mathcal{S}\times[0,1]$, and random variables $\bar{X}$ and S on $\mathcal{C}([0,1] : \mathbb{R}^d)$ such that $(\nu^n, \hat{\nu}_1^n \otimes \lambda, \hat{\nu}_2^n \otimes \lambda, \gamma^n, \bar{X}^n, \tilde{X}^n, S^n) \xrightarrow{\mathcal{D}} (\nu, \mu_1, \mu_2, \gamma, \bar{X}, \bar{X}, S)$ [4, Theorem A.3.6]. Moreover, w.p.1 μ_1 and μ_2 equal the marginals of ν over (ζ, t) and over (y, t), respectively. For the developments below, we note that by the Skorohod representation theorem we can assume that convergence takes place w.p.1 on some probability space, which we also denote by $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$.

(b) We let μ_3 denote the marginal of ν over t. Using the fact that the marginal of ν^n over t is Lebesgue measure λ, we have that w.p.1 for any bounded continuous function g mapping $[0,1]$ into $\mathbb{R}$,

$$\begin{aligned}\int_0^1 g(t)\mu_3(dt) &= \int_{\mathcal{S}\times\mathcal{S}\times[0,1]} g(t)\nu(d\zeta \times dy \times dt)\\ &= \lim_{n\to\infty} \int_{\mathcal{S}\times\mathcal{S}\times[0,1]} g(t)\nu^n(d\zeta \times dy \times dt)\\ &= \lim_{n\to\infty} \int_0^1 g(t)dt\\ &= \int_0^1 g(t)dt.\end{aligned}$$

Since the class of bounded and continuous functions is a measure determining class [1, Theorem 1.3], this implies that w.p.1 $\mu_3(\cdot)$ equals $\lambda(\cdot)$. By Theorem A.5.6 in [4], there exists a stochastic kernel $\nu(d\zeta \times dy|t)$ on $\mathcal{S}\times\mathcal{S}$ given $[0,1]$ such that w.p.1

$$\nu(B_1 \times B_2 \times C) = \int_C \nu(B_1 \times B_2|t)\, dt.$$

Once more Theorem A.5.6 in [4] gives the existence of stochastic kernels $\hat{\nu}_2(dy|\zeta, t)$ on $\mathcal{S}$ given $\mathcal{S}\times[0,1]$ and $\hat{\nu}_1(d\zeta|t)$ on $\mathcal{S}$ given $[0,1]$ (the second and first marginals of $\nu(d\zeta \times dy|t)$, respectively) such that

$$\nu(B_1 \times B_2 \times C) = \int_C \int_{B_1} \hat{\nu}_1(d\zeta|t)\hat{\nu}_2(B_2|\zeta, t)dt.$$

This gives the decomposition of $\nu(d\zeta \times dy \times dt)$ given in part (b).

(c) Consider a function f of the form $f(y,t) = g(y)h(t)$, with $g \in \mathcal{C}(\mathcal{S} : \mathbb{R}^d)$ and $h \in \mathcal{C}([0,1] : \mathbb{R}^d)$. Since

$$\bar{E}_{x,\xi}\left\{ g(\bar{Z}^n_{j+1}) - \int_{\mathcal{S}} g(y)\nu^n_j(dy|\bar{X}_0, \ldots, \bar{X}^n_j, \bar{Z}^n_j) \right\} = 0,$$

we have that

$$\left\{ g(\bar{Z}^n_{j+1}) - \int_{\mathcal{S}} g(y)\nu^n_j(dy|\bar{X}^n_0, \ldots, \bar{X}^n_j, \bar{Z}^n_j) \right\}$$

is a martingale difference sequence. Moreover,

$$\begin{aligned}\int_{\mathcal{S}\times[0,1]} f(\zeta,t)(\hat{\nu}^n_1 \otimes \lambda)(d\zeta \times dt) &= \sum_{l=0}^{\frac{1}{k_n}-1} \int_{lk_n}^{(l+1)k_n} h(t)dt \cdot \int_{\mathcal{S}} g(\zeta)(\tilde{\nu}^n_l)_1(d\zeta) \\ &= \sum_{l=0}^{\frac{1}{k_n}-1} \int_{lk_n}^{(l+1)k_n} h(t)dt \cdot \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n-1} g(\bar{Z}^n_j)\end{aligned}$$

and similarly

$$\begin{aligned}&\int_{\mathcal{S}\times[0,1]} f(y,t)(\hat{\nu}^n_2 \otimes \lambda)(dy \times dt) \\ &= \sum_{l=0}^{\frac{1}{k_n}-1} \int_{lk_n}^{(l+1)k_n} h(t)dt \cdot \left[\frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n-1} \int_{\mathcal{S}} g(y)\nu^n_j(dy) \right].\end{aligned}$$

Noting that

$$\left\| \sum_{l=0}^{\frac{1}{k_n}-1} \frac{1}{m_n} \left[\int_{lk_n}^{(l+1)k_n} h(t)dt \right] \left[g(\bar{Z}^n_{lm_n}) - g(\bar{Z}^n_{(l+1)m_n}) \right] \right\| \leq \frac{2\|g\|\|h\|}{m_n},$$

we have that if $m_n \geq \frac{4\|g\|\|h\|}{\varepsilon}$, then

$$\begin{aligned}&\bar{P}_{x,\xi}\left\{ \left| \int_{\mathcal{S}\times[0,1]} fd(\hat{\nu}^n_1 \otimes \lambda) - \int_{\mathcal{S}\times[0,1]} fd(\hat{\nu}^n_2 \otimes \lambda) \right| \geq \varepsilon \right\} \\ &\quad\leq \bar{P}_{x,\xi}\left\{ \left| \sum_{l=0}^{\frac{1}{k_n}-1} \frac{1}{m_n} \left[\int_{lk_n}^{(l+1)k_n} h(t)dt \right] \left[\sum_{j=lm_n}^{(l+1)m_n-1} \left(g(\bar{Z}^n_{j+1}) - \int g(y)\nu^n_j(dy) \right) \right] \right| \geq \frac{\varepsilon}{2} \right\} \\ &\quad\leq \bar{P}_{x,\xi}\left\{ \left| \frac{1}{n} \sum_{j=0}^{n-1} \left[g(\bar{Z}^n_{j+1}) - \int_{\mathcal{S}} g(y)\nu^n_j(dy) \right] \right| \geq \frac{\varepsilon}{2\|h\|} \right\} \\ &\quad\leq \frac{4\|h\|^2}{\varepsilon^2} \bar{E}_{x,\xi}\left\{ \frac{1}{n^2} \left(\sum_{j=0}^{n-1} \left[g(\bar{Z}^n_{j+1}) - \int_{\mathcal{S}} g(y)\nu^n_j(dy) \right] \right)^2 \right\}\end{aligned}$$

$$\text{(C.3)} \quad \leq \frac{4\|h\|^2}{\varepsilon^2} \bar{E}_{x,\xi}\left\{ \frac{1}{n^2} \sum_{j=0}^{n-1} \left(g(\bar{Z}^n_{j+1}) - \int_{\mathcal{S}} g(y)\nu^n_j(dy) \right)^2 \right\} \leq \frac{16\|h\|^2\|g\|^2}{n\varepsilon^2}.$$

This implies that $\int_{\mathcal{S}\times[0,1]} f d(\hat{\nu}_1^n \otimes \lambda) - \int_{\mathcal{S}\times[0,1]} f d(\hat{\nu}_2^n \otimes \lambda)$ converges to zero in probability and hence in distribution [1, Theorem 4.3]. Given the convergence w.p.1 of $\hat{\nu}_1^n \otimes \lambda$ and $\hat{\nu}_2^n \otimes \lambda$ to $\hat{\nu}_1 \otimes \lambda$ and $\hat{\nu}_2 \otimes \lambda$, respectively, we get w.p.1

$$\lim_{n\to\infty} \int_{\mathcal{S}\times[0,1]} f(y,t)(\hat{\nu}_r^n \otimes \lambda)(dy \times dt) = \int_{\mathcal{S}\times[0,1]} f(y,t)(\hat{\nu}_r \otimes \lambda)(dy \times dt)$$

for $r = 1, 2$. Therefore w.p.1

$$\int_{\mathcal{S}\times[0,1]} f(y,t)(\hat{\nu}_1 \otimes \lambda)(dy \times dt) = \int_{\mathcal{S}\times[0,1]} f(y,t)(\hat{\nu}_2 \otimes \lambda)(dy \times dt). \tag{C.4}$$

Theorem A.3.14 in [4] implies that we can extend the equality in (C.4) from f of the form $f(y,t) = g(y)h(t)$ to all $f : \mathcal{S} \times [0,1] \mapsto \mathbb{R}$ that are bounded and continuous. Since the class of bounded continuous functions is measure-determining, we have that $\hat{\nu}_1 \otimes \lambda = \hat{\nu}_2 \otimes \lambda$, as we wanted to show.

(d) We now show that for each $t \in [0,1]$ and Borel subset B_1 of $\mathcal{S}$, we have

$$\hat{\nu}_1(B_1|t) = \int_{\mathcal{S}} \hat{\nu}_2(B_1|\zeta, t)\hat{\nu}_1(d\zeta|t). \tag{C.5}$$

Let $\mathcal{U}_b(\mathcal{S})$ denote the space of bounded, uniformly continuous functions mapping $\mathcal{S}$ into $\mathbb{R}$. Since $\mathcal{S}$ is Polish, there exists an equivalent metric m under which $\mathcal{U}_b(\mathcal{S}, m)$ is separable with respect to the uniform metric. Let $\mathcal{E}$ be a countable dense subset of $\mathcal{U}_b(\mathcal{S}, m)$, and let g be any function in $\mathcal{E}$. For each $s \in [0,1]$ let $E_i \subset [0,1]$ be a sequence of sets which *shrinks nicely* to s (see [2, p. 353]), and define

$$f(t,y) \doteq g(y) \cdot \frac{1}{\lambda(E_i)} I_{E_i}(t).$$

Since $\hat{\nu}_1 \otimes \lambda = \hat{\nu}_2 \otimes \lambda$,

$$\frac{1}{\lambda(E_i)} \int_{E_i} \int_{\mathcal{S}} g(y)\hat{\nu}_1(dy|t)\lambda(dt) = \frac{1}{\lambda(E_i)} \int_{E_i} \int_{\mathcal{S}} \int_{\mathcal{S}} g(y)\hat{\nu}_2(dy|\zeta, t)\hat{\nu}_1(d\zeta|t)\lambda(dt).$$

Define $h_1(t) \doteq \int_{\mathcal{S}} g(y)\hat{\nu}_1(dy|t)$ and $h_2(t) \doteq \int_{\mathcal{S}} \int_{\mathcal{S}} g(y)\hat{\nu}_2(dy|\zeta, t)\hat{\nu}_1(d\zeta|t)$. Then we have

$$|h_1(s) - h_2(s)| \le \frac{1}{\lambda(E_i)} \int_{E_i} |h_1(t) - h_1(s)|\, dt + \frac{1}{\lambda(E_i)} \int_{E_i} |h_2(t) - h_2(s)|\, dt,$$

which tends to 0 as $n \to \infty$ for almost all $s \in [0,1]$ (see Theorem C.13 in [2]). This implies that $h_1(s) = h_2(s)$ a.s., so that there exists a set $B_g \in \mathcal{B}([0,1])$ with $\lambda(B_g) = 0$ and such that $\int_{\mathcal{S}} g(y)\hat{\nu}_1(dy|t) = \int_{\mathcal{S}} \int_{\mathcal{S}} g(y)\hat{\nu}_2(dy|\zeta, t)\hat{\nu}_1(d\zeta|t)$ for all $t \notin B_g$. Now define $B \doteq \cup_{g\in\mathcal{E}} B_g$. Then $\lambda(B) = 0$ and for all $t \notin B$ and $g \in \mathcal{E}$ the same equality holds. The equality can then be extended to $g \in \mathcal{U}_b(\mathcal{S}, m)$, which implies that $\hat{\nu}_1(dy|t) = \hat{\nu}_2(dy|t)$ for all $t \notin B$. Finally, redefining $\hat{\nu}_1$ and $\hat{\nu}_2$ in an obvious way for $t \in B$, we get (C.5).

(e) Let $\{\bar{Y}^n, n \in \mathbb{N}\}$ and $\{\tilde{Y}^n, n \in \mathbb{N}\}$ be the sequences of piecewise linear and piecewise constant interpolations, respectively, of the process $\{X_j^n, j = 0, \dots, n\}$ but when observed only at the endpoints of the intervals of size k_n. That is, they are the interpolations of a process $\{\bar{Y}_j^n, j = 0, \dots, n\}$ defined through $\bar{Y}_l^n \doteq \bar{X}_{lm_n}^n$ for $l = 0, \dots, \frac{1}{k_n}$. The intuition behind this idea is described clearly below (2.3). We will relate $\bar{Y}^n$ to $\bar{X}^n$, $\tilde{Y}^n$ to $\tilde{X}^n$, and both $\bar{Y}^n$ and $\tilde{Y}^n$ to S^n in a way that forces all five sequences to have the same limit [1, Theorem 4.1]. By showing that S and $\bar{X}$ as defined in (C.1) are the same w.p.1, the characterization of the limit process $\bar{X}$ will follow.

By definition, the process $\{\bar{Y}_l^n\}$ follows the evolution

$$\bar{Y}_{l+1}^n = \bar{Y}_l^n + \frac{1}{n} \sum_{j=lm_n}^{(l+1)m_n-1} b(\bar{X}_j^n, \bar{Z}_{j+1}^n).$$

Moreover, for every $l = 0, \ldots, \frac{1}{k_n} - 1$ and with $i = 1, \ldots, m_n - 1$, we have

$$\|\bar{Y}_l^n - \bar{X}_{lm_n+i}^n\| = \left\| \frac{1}{n} \sum_{j=0}^{i-1} b(\bar{X}_{lm_n+j}^n, \bar{Z}_{lm_n+j+1}^n) \right\| \le \frac{i\|b\|_\infty}{n}. \tag{C.6}$$

Hence

$$\begin{aligned} \sup_{t\in[0,1]} \|\bar{Y}^n(t) - \bar{X}^n(t)\| &\le w_{\bar{Y}^n}(k_n) + w_{\bar{X}^n}(1/n) \\ &\quad + \max_{l\in\{0,\ldots,\frac{1}{k_n}-1\}} \max_{i\in\{1,\ldots,m_n-1\}} \|\bar{Y}_l^n - \bar{X}_{lm_n+i}^n\| \\ &\le \left(3k_n + \frac{2}{n}\right) \|b\|_\infty, \end{aligned}$$

where we have used (C.6), (C.2), and the bound

$$w_{\bar{Y}^n}(k_n) \doteq \sup_{\{s,t\in[0,1]:|s-t|\le k_n\}} \|\bar{Y}^n(t) - \bar{Y}^n(s)\| \le 2k_n\|b\|_\infty. \tag{C.7}$$

This implies that in $\mathcal{C}([0,1] : \mathbb{R}^d)$ under the uniform metric, $d(\bar{X}^n, \bar{Y}^n)$ converges to 0 in probability. Similarly, in $\mathcal{D}([0,1] : \mathbb{R}^d)$ under the Skorohod metric, $d(\tilde{X}^n, \tilde{Y}^n)$ converges to 0 in probability.

Next we prove that

$$\lim_{n\to\infty} \bar{P}_{x,\xi} \left\{ \sup_{t\in[0,1]} \left\|S^n(t) - \bar{Y}^n(t)\right\| \ge \varepsilon \right\} = 0, \tag{C.8}$$

which immediately implies the limit $\lim_{n\to\infty} \bar{P}_{x,\xi}\{\sup_{t\in[0,1]} \|S^n(t) - \tilde{Y}^n(t)\| \ge \varepsilon\} = 0$. For any $t \in [0,1]$ with $lk_n \le t < (l+1)k_n$, we have

$$\begin{aligned} &\|\bar{Y}^n(t) - S^n(t)\| \\ &\le w_{\bar{Y}^n}(k_n) + w_{S^n}(k_n) + Kk_n\|b\|_\infty \\ &\quad + \left\| \sum_{j=0}^{l} \frac{1}{n} \sum_{i=jm_n}^{(j+1)m_n} b(\bar{Y}_j^n, \bar{Z}_{i+1}^n) - \int_{\mathcal{S}\times\mathcal{S}\times[0,lk_n]} b(\tilde{Y}^n(s), y)\nu^n(d\zeta\times dy\times ds) \right\| \\ &\quad + \left\| \int_{\mathcal{S}\times\mathcal{S}\times[0,lk_n]} \left[b(\tilde{Y}^n(s), y) - b(\bar{Y}^n(s), y)\right] \nu^n(d\zeta\times dy\times ds) \right\| \\ &\quad + \left\| \int_{\mathcal{S}\times\mathcal{S}\times[0,lk_n]} \left[b(\bar{Y}^n(s), y) - b(S^n(s), y)\right] \nu^n(d\zeta\times dy\times ds) \right\| \\ &\le 3(1+K)k_n\|b\|_\infty + A(t) + \int_0^t K\|\bar{Y}^n(s) - S^n(s)\|ds, \end{aligned}$$

where we have defined

$$A(t) \doteq \left\| \sum_{j=0}^{l} \frac{1}{n} \sum_{i=jm_n}^{(j+1)m_n} b(\bar{Y}_j^n, \bar{Z}_{i+1}^n) - \int_{\mathcal{S}\times\mathcal{S}\times[0,lk_n]} b(\tilde{Y}^n(s), y)\nu^n(d\zeta\times dy\times ds) \right\|.$$

(Note that dependence on t comes through l.) By Gronwall's inequality we can then write

$$\begin{aligned}
&\|\bar{Y}^n(t) - S^n(t)\| \\
&\leq 3(1+K)k_n\|b\|_\infty + A(t) + \int_0^t K\left[3(1+K)k_n\|b\|_\infty + A(s)\right]e^{K(t-s)}ds. \\
&\leq e^K\left[3(1+K)k_n\|b\|_\infty + A(lk_n)\right] + \sum_{j=0}^{l-1} A(jk_n)Ke^{Kt}\int_{jk_n}^{(j+1)k_n} e^{-Ks}ds \\
&\leq e^K\left[3(1+K)k_n\|b\|_\infty + (1+K)\max_{j\in\{0,\ldots,1/k_n\}} A(jk_n)\right],
\end{aligned} \tag{C.9}$$

where in the last step we have used the inequality $e^{jk_n} - e^{(j+1)k_n} \leq k_n K e^{-Kjk_n}$, valid because of the mean value theorem. Given (C.9), all that remains to show is that $\{\max_{l\in\{0,\ldots,1/k_n\}} A(lk_n), n \in \mathbb{N}\}$ converges to zero in probability as $n \to \infty$.

Now for all $l = 0, \ldots, 1/k_n$ we have that

$$\int_{\mathcal{S}\times\mathcal{S}\times[0,lk_n]} b(\tilde{Y}^n(s), y)\nu^n(d\zeta\times dy\times ds) = \sum_{j=0}^{l} k_n \frac{1}{m_n} \sum_{i=jm_n}^{(j+1)m_n} \int_{\mathcal{S}} b(\bar{Y}_j^n, y)\nu_i^n(dy|\bar{X}_i^n, \bar{Z}_i^n).$$

Moreover, the sequence

$$\left\{ b(\bar{Y}_j^n, \bar{Z}_{i+1}^n) - \int_{\mathcal{S}} b(\bar{Y}_j^n, y)\nu_i^n(dy|\bar{X}_i^n, \bar{Z}_i^n), j = 0, \ldots, \frac{1}{k_n}, i = jm_n, \ldots, (j+1)m_n - 1 \right\}$$

forms a martingale difference sequence with respect to the sequence of sigma fields generated by $\{(\bar{X}_r^n, \bar{Z}_r^n), r = 0, \ldots, i\}$ for $i = 0, \ldots, n-1$. Therefore, the submartingale inequality [5, Lemma 2.2.3] applied to the submartingale

$$\left\{ \left\| b(\bar{Y}_j^n, \bar{Z}_{i+1}^n) - \int_{\mathcal{S}} b(\bar{Y}_j^n, y)\nu_i^n(dy|\bar{X}_i^n, \bar{Z}_i^n) \right\|^2 \right\}$$

implies that for any $\varepsilon > 0$

$$\begin{aligned}
&\bar{P}_{x,\xi}\left\{ \max_{l\in\{0,\ldots,1/k_n\}} A(lk_n) \geq \varepsilon \right\} \\
&\leq \frac{1}{\varepsilon^2}\bar{E}_{x,\xi}\{A(1)^2\} \\
&\leq \frac{1}{\varepsilon^2}\sum_{j=0}^{l}\frac{1}{n^2}\sum_{i=jm_n}^{(j+1)m_n} \bar{E}_{x,\xi}\left\| b(\bar{Y}_j^n, \bar{Z}_{i+1}^n) - \int_{\mathcal{S}} b(\bar{Y}_j^n, y)\nu_i^n(dy|\bar{X}_i^n, \bar{Z}_i^n)\right\|^2 \\
&\leq \frac{4\|b\|_\infty^2}{\varepsilon^2 n},
\end{aligned}$$

which gives (C.8).

Having shown that all five sequences must converge to the same limit, it remains to show that S and $\bar{X}$ as defined in (C.1) are the same w.p.1. We will show that for each fixed t, $S(t) = \bar{X}(t)$ w.p.1. Equality for all $t \in [0,1]$ w.p.1 follows by considering the rationals and then extending by continuity.

Fix $t \in [0.1]$. We have

$$\|S^n(t) - \bar{X}(t)\| \leq \int_0^t K \left\|S^n(s) - \bar{X}(s)\right\| ds$$
$$+ \left\| \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu^n(d\zeta\times dy\times ds) - \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu(d\zeta\times dy\times ds) \right\|.$$

Using Gronwall's inequality if follows that

$$\text{(C.10)} \qquad \left\|S^n(t) - \bar{X}(t)\right\|$$
$$\leq \left\| \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu^n(d\zeta\times dy\times ds) - \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu(d\zeta\times dy\times ds) \right\|$$
$$+ K\int_0^t \left\| \int_{\mathcal{S}\times\mathcal{S}\times[0,r]} b(\bar{X}(s), y)\nu^n(d\zeta\times dy\times ds) - \int_{\mathcal{S}\times\mathcal{S}\times[0,r]} b(\bar{X}(s), y)\nu(d\zeta\times dy\times ds) \right\| e^{K(t-r)}dr.$$

Weak convergence of ν^n to ν implies that w.p.1

$$\lim_{n\to\infty} \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu^n(d\zeta \times dy \times ds) = \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu(d\zeta \times dy \times ds).$$

This statement is true since we can identify the points of discontinuity of the bounded function $b(\bar{X}(s), y)1_{[0,t]}(s)$ to be $\mathcal{S} \times \mathcal{S} \times \{t\}$, which form a set of measure zero under the limit ν. Hence it follows from (C.10) and the dominated convergence theorem that for any given $\varepsilon > 0$

$$\lim_{n\to\infty} \bar{P}_{x,\xi}\left\{\|S^n(t) - \bar{X}(t)\| \geq \varepsilon\right\} \leq \lim_{n\to\infty} \frac{1}{\varepsilon}\bar{E}_{x,\xi}\left\{\left\|S^n(t) - \bar{X}(t)\right\|\right\} = 0.$$

Uniqueness of limits implies $S(t) = \bar{X}(t)$ w.p.1, as we wanted to show.

(f) Let $f : \mathcal{S} \times \mathcal{S} \times [0,1]$ be bounded and continuous. Then for each $n \in \mathbb{N}$ the function g_n mapping (ζ, t) to $\int_{\mathcal{S}} f(\zeta, y, t)p(dy|\bar{Y}^n(t), \zeta)$ is also bounded and continuous. Define the bounded and continuous function g mapping (ζ, t) into $\int_{\mathcal{S}} f(\zeta, y, t)$ $p(dy|\bar{X}(t), \zeta)$. Weak continuity of p and a.s. convergence of $\bar{Y}^n$ to $\bar{X}$ imply that for all $t \in [0,1]$ and $\zeta \in \mathcal{S}$ we have $\lim_{n\to\infty} g_n(\zeta, t) = g(\zeta, t)$ a.s. We argue that convergence is, in fact, uniform in $\mathcal{S} \times [0,1]$.

Indeed, let $\varepsilon > 0$ be given. Fix $N \in \mathbb{N}$ satisfying $k_N < \varepsilon/(3\|b\|_\infty)$ so that (see (C.7)) for all $n > N$ and r, s with $|r - s| \leq k_N$, we have $\|\bar{Y}^n(r) - \bar{Y}^n(s)\| < \varepsilon/3$ and $\|\bar{X}(r) - \bar{X}(s)\| < \varepsilon/3$. Further, N can be chosen so that for all $n \geq N$ and $i = 0, \dots, \frac{1}{k_N} - 1$ we have $\|\bar{Y}^n(ik_N) - \bar{X}(ik_N)\| < \varepsilon/3$. Hence we have that for all $t \in [0,1]$ and any $n \geq N$, $\|\bar{Y}^n(t) - \bar{X}(t)\| < \varepsilon$. By Hypothesis H.1 we can then write

$$g_n(\zeta, t) = \int_{\mathcal{S}} f(\zeta, y, t)p(dy|\bar{Y}^n(t), \zeta) = \int_{\mathcal{S}} f(\zeta, y, t)\tilde{p}^{\bar{Y}^n(t)}(\zeta, y)\vartheta(dy).$$

By the dominated convergence theorem, this last quantity converges (a.s. uniformly in ζ and t) to $\int_{\mathcal{S}} f(\zeta, y, t)\tilde{p}^{\bar{X}(t)}(\zeta, y)\vartheta(dy) = g(\zeta, t)$.

It now follows from the weak convergence of $\hat{\nu}_1^n \otimes \lambda$ to $\hat{\nu}_1 \otimes \lambda$ that

$$\lim_{n\to\infty} \int_0^1 \int_{\mathcal{S}} g_n(\zeta,t)\hat{\nu}_1^n(d\zeta|t)dt = \int_0^1 \int_{\mathcal{S}} g(\zeta,t)\hat{\nu}_1(d\zeta|t)dt.$$

This in turn implies that

$$\lim_{n\to\infty} \int_{\mathcal{S}\times\mathcal{S}\times[0,1]} f(\zeta,y,t)\gamma^n(d\zeta\times dy\times dt) = \int_{\mathcal{S}\times\mathcal{S}\times[0,1]} f(\zeta,y,t)\gamma(d\zeta\times dy\times dt),$$

which completes the proof. □

The last lemma in this section establishes an estimate needed in the proof of the lower bound.

LEMMA C.2. *For any* $\delta > 0$,

$$\lim_{\sigma\to 0} \bar{P}_{x,\xi}\left\{\sup_{t\in[0,1]} ||\bar{U}(t)|| \geq \delta\right\} = 0.$$

Proof. As was discussed in the proof of the lower bound, $S^{2,n} \to \bar{U}$ in distribution in $\mathcal{C}([0,1]:\mathbb{R}^d)$, so that by the Skorohod representation theorem and Fatou's lemma,

$$\bar{E}_{x,\xi}\left\{\left(\sup_{t\in[0,1]} ||\bar{U}(t)||\right)^2\right\} \leq \liminf_{n\to\infty} \bar{E}_{x,\xi}\left\{\left(\sup_{t\in[0,1]} ||S^n(t)||\right)^2\right\}.$$

Hence we have that for any $\delta > 0$,

$$\begin{aligned}
\bar{P}_{x,\xi}\left\{\sup_{t\in[0,1]} ||\bar{U}(t)|| \geq \delta\right\} &\leq \frac{1}{\delta^2}\bar{E}_{x,\xi}\left\{\left(\sup_{t\in[0,1]} ||\bar{U}(t)||\right)^2\right\} \\
&\leq \frac{1}{\delta^2}\liminf_{n\to\infty} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n-1}\left\|\int_{\mathbb{R}^d} y\nu_j^{2,n}(dy)\right\|^2\right\} \\
&= \frac{1}{\delta^2}\liminf_{n\to\infty} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n\tau^n-1} ||\beta_j^{2,n}||^2\right\} \\
&\leq \frac{2\sigma^2}{\delta^2}\liminf_{n\to\infty} \bar{E}_{x,\xi}\left\{\frac{1}{n}\sum_{j=0}^{n\tau^n-1} R(\gamma_j^{2,n}(\cdot)||\rho_\sigma(\cdot))\right\} \\
&\leq \frac{2\sigma^2}{\delta^2}\liminf_{n\to\infty}\left\{k_n\sum_{l=0}^{[1/k_n]} L_\sigma(\psi^*(lm_n/n),\dot{\psi}^*(lm_n/n)) + 3\sigma\right\} \\
&\leq \frac{2\sigma^2}{\delta^2}[I_x(\psi) + 4\sigma].
\end{aligned}$$

The fourth inequality follows from $\frac{1}{2\sigma^2}||\beta_j^{2,n}||^2 = \hat{L}_\sigma(\beta_j^{2,n}) \leq R(\gamma_j^{2,n}(\cdot)||\rho_\sigma(\cdot))$, where $\hat{L}_\sigma$ is the Legendre–Fenchel of the moment generating function of ρ_σ. The fifth line follows from (4.13) and (4.18), while in line six we have used part (e) of Lemma B.3 with $\varepsilon = \sigma$. Letting $\sigma \to 0$ completes the proof. □

REFERENCES

[1] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.
[2] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer-Verlag, New York, 1998.
[3] P. Dupuis, *Large deviations analysis of some recursive algorithms with state dependent noise*, Ann. Probab., 16 (1988), pp. 1509–1536.
[4] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley, New York, 1996.
[5] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley, New York, 1986.
[6] T. E. Harris, *Theory of Branching Processes*, Springer-Verlag, Berlin, 1963.
[7] I. Iscoe, P. Ney, and E. Nummelin, *Large deviations of uniformly recurrent Markov additive processes*, Adv. in Appl. Math., 6 (1985), pp. 373–412.
[8] H. J. Kushner and F. J. Vázquez-Abad, *Stochastic approximation methods for systems over an infinite horizon*, SIAM J. Control Optim., 34 (1996), pp. 712–756.
[9] H. J. Kushner and J. Yang, *Analysis of adaptive step-size SA algorithms for parameter tracking*, IEEE Trans. Automat. Control, 40 (1995), pp. 1403–1410.
[10] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, 1997.
[11] L. Ljung, *Analysis of recursive stochastic algorithms*, IEEE Trans. Automat. Control, 22 (1977), pp. 551–575.
[12] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.
[13] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.
[14] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.