Large Deviations for a Random Walk Model with State-dependent Noise

Michelle Boué — ITAM
Daniel Hernández-Hernández — CIMAT
Richard S. Ellis — UMass, Amherst

November 27, 2000

Abstract

In this paper we prove the Laplace principle for a class of random walks with state-dependent noise. The Laplace principle implies a large deviations principle with the same rate function. This type of model has important applications in queueing and communication theory, and in the area of stochastic approximation.

1 Introduction

This paper is concerned with proving a large deviations 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 has 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 [9].

Let $\mathcal{S}$ be a compact Polish space and let $p(d\zeta|x,\xi)$ be a stochastic kernel on $\mathcal{S}$ given $I\!R^d \times \mathcal{S}$. For each $n \in I\!N$, 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 $I\!R^d \times \mathcal{S}$. For $x \in I\!R^d$, $\xi \in \mathcal{S}$ and b a function mapping $I\!R^d \times \mathcal{S}$ into $I\!R^d$, this sequence is defined through

$$\begin{aligned} X_0^n &\doteq x \\ Z_0^n &\doteq \xi \\ X_{j+1}^n &\doteq X_j^n + \frac{1}{n} b(X_j^n, Z_{j+1}^n), \end{aligned}$$

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

$$P_{x,\xi}\left\{Z_{j+1}^n \in d\zeta | (X_i^n, Z_i^n), i = 0, \ldots, j\right\} = p(d\zeta | X_j^n, Z_j^n). \tag{1.1}$$

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).

c) Given any compact set $\Delta \subset I\!R^d$, there exist a probability measure ϑ on $\mathcal{S}$, a function $\tilde{p}^x(\xi, \zeta)$ on $\mathcal{S} \times \mathcal{S}$, and constants $0 < a \le A < \infty$ such that

$$p(d\zeta | x, \xi) = \tilde{p}^x(\xi, \zeta)\vartheta(d\zeta) \quad \text{and} \quad a \le \tilde{p}^x(\xi, \zeta) \le A$$

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

Under this hypothesis, the following statements are proved in [10]. Consider the eigenvalue problem:

$$e^{\Lambda(\alpha) + \Psi(\xi)} = \int_{\mathcal{S}} e^{\langle \alpha, b(x,\zeta)\rangle + \Psi(\zeta)} p(d\zeta | x, \xi). \tag{1.2}$$

Under Hypothesis H.1 the solution to this problem exists. Moreover, the solution is bounded with Λ smooth. In fact, we can identify $\Lambda(\alpha)$ 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

$$\Lambda(x,\alpha) = \lim_{N\to\infty} \frac{1}{N} \log E_\xi \left\{ \exp\langle \alpha, \sum_{j=1}^{N} b(x, \xi_j^x)\rangle \right\}, \tag{1.3}$$

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 (1.1) is held constant at value x. This process is intimately connected with the process $\{X_j^n\}$: 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 taking X_j^n as constant (see [9]).

Let $X^n = \{X^n(t), 0 \leq t \leq 1\}$ 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 2.2, states the Laplace principle for the process $\{X^n, n \in \mathbb{N}\}$. The term Laplace principle refers to the asymptotic analysis of normalized logarithms of expectations involving continuous functions (a precise definition is given in the next section). Because a Laplace principle is equivalent to a large deviations principle with the same rate function, the large deviation principle that we are interested in is an immediate consequence of Theorem 2.2.

We note that the large deviations principle implied by Theorem 2.2 is covered by the results in [3]. The proof there relies on technical assumptions for the function Λ, which is assumed to exist. Under Hypothesis H.1, the function Λ exists indeed, and satisfies the technical assumptions required there, thereby implying the large deviations principle. Instead, an advantage of the proof presented here is that it depends on assumptions made on the evolution of the process itself (the transition kernels and the function b). The advantage is in several senses. 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 [9, 8]), 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

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

Definition 2.1 *Let $\{X^n, n \in I\!\!N\}$ be a sequence of random variables taking values in a Polish space $\mathcal{X}$ and let I be a rate funcion on $\mathcal{X}$. We say that $\{X^n, n \in I\!\!N\}$ satisfies a* **Laplace principle** *with rate function I if for every bounded continuous function h mapping $\mathcal{X}$ into $I\!\!R$*

$$\lim_{n\to\infty} \frac{1}{n} \log E \left\{ \exp[-nh(X^n)] \right\} = - \inf_{x\in\mathcal{X}} \left\{ h(x) + I(x) \right\}.$$

A Laplace principle is equivalent to a large deviations principle with the same rate function (see Theorems 2.2.1 and 2.2.3 in [4] for a proof). For our model of interest, we will prove the former in the theorem that follows, obtaining the latter as a direct consequence of it.

Theorem 2.2 *For $n \in I\!\!N$, let X^n be the piecewise linear interpolation of $\{X^n_j, j = 0, 1, \ldots, n\}$ as defined in* (1.4)*. Under Hypothesis* H.1 *the sequence $\{X^n, n \in I\!\!N\}$ satisfies* *the Laplace principle with rate function $I_{x,\xi}(\cdot)$, where*

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

and $L(x,\cdot)$ is the Legendre-Fenchel transform of the function $\Lambda(x,\cdot)$ given in (1.3) with respect to the second variable. That is, for x and β in $I\!R^d$

$$L(x,\beta) \doteq \sup_{\alpha\in I\!R^d} \{\langle \alpha,\beta\rangle - \Lambda(x,\alpha)\}. \tag{2.5}$$

Let $W^n(x,\xi) \doteq -1/n \log E_{x,\xi}\{\exp[-nh(X^n)]\}$. The proof of Theorem 2.2 is done in two parts. We start by proving the Laplace principle upper bound, equivalent to

$$\liminf_{n\to\infty} W^n(x,\xi) \geq \inf_{\phi\in\mathcal{C}([0,1]:I\!R^d)} \{I_{x,\xi}(\phi)+h(\phi)\}. \tag{2.6}$$

This is the content of Section 3. The lower bound

$$\limsup_{n\to\infty} W^n(x,\xi) \leq \inf_{\phi\in\mathcal{C}([0,1]:I\!R^d)} \{I_{x,\xi}(\phi)+h(\phi)\}. \tag{2.7}$$

is then proved in Section 4. In both cases, a key step in the proof involves the study (via weak convergence methods) of the limit properties of a sequence of associated stochastic control problems. For ease of presentation, the representation formulas required to associate $W^n(x,\xi)$ with an appropriate stochastic control problem are presented in the Appendix (Theorem .1 is used for the upper bound and Theorem .2 for the lower bound).

3 Proof of the Laplace principle upper bound

This section is devoted to the proof of (2.6). As was mentioned earlier, we shall use the representation formula found in Theorem .1 in the Appendix, so that for every $n \in I\!N$

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

The infimum is taken over all admissible control sequences $\{\nu_j^n, j = 0, \ldots, n-1\}$, with $\nu_j^n(\cdot) = \nu_j^n(\cdot|\bar{X}_0^n, \ldots, \bar{X}_j^n, \bar{Z}_j^n)$. The reader is referred to the Appendix for the definition of all the quantities involved in the representation.

Given $\varepsilon > 0$, for each $n \in I\!N$ let $\{\nu_j^n, j = 0, \ldots, n-1\}$ be a sequence of early optimal admissible controls, so that

$$W^n(x,\xi) \geq \bar{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot) \| p(\cdot|\bar{X}_j^n, \bar{Z}_j^n)) + h(\bar{X}^n) \right\} - \varepsilon. \tag{3.8}$$

Here $\{\bar{X}_j^n, j = 0, \ldots, n\}$ is the controlled process associated with this sequence of controls, and $\{\bar{X}^n, n \in I\!N\}$ is the sequence of piecewise linear interpolations of $\{\bar{X}_j^n, j = 0, \ldots, n\}$. The proof of (2.6) requires that we analyze the asymptotic behavior of $\{\bar{X}^n, n \in I\!N\}$ and of a sequence of control measures $\{\nu^n, n \in I\!N\}$ which we now construct using the sequences $\{\nu_j^n, j = 0, \ldots, n, n \in I\!N\}$.

Let $\{m_n, n \in I\!N\}$ 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 integral multiple of k_n. The basic idea will be to collect terms of the sum appearing in (3.8) in groups of size m_n for the purposes of averaging. The control measures that we now define will be the result of this grouping.

For $\xi \in \mathcal{S}$, let δ_ξ denote the unit point measure at ξ. For $l = 0, \ldots, 1/k_n - 1$, and Borel subsets B_1 and B_2 of $\mathcal{S}$, let

$$\nu_l^n(B_1 \times B_2) \doteq \frac{1}{m_n} \sum_{j=lm_n}^{(l+1)m_n - 1} \delta_{\bar{Z}_j^n}(B_1) \times \nu_j^n(B_2|\bar{X}_j^n, \bar{Z}_j^n).$$

The quantity ν_l^n is a stochastic kernel[1] on $\mathcal{S} \times \mathcal{S}$ with marginals

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

[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

For each $n \in I\!\!N$ and $t \in [0,1]$ define

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

Finally we 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

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

Theorem A.5.6 in [4] provides us with a convenient decomposition for the measures $\nu^n(d\zeta \times dy \times dt)$ in terms of the first marginal of $\nu^n(d\zeta \times dy|t)$. More precisely, 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 $[0,1]$ we can write

$$\begin{aligned} \nu^n(B_1 \times B_2 \times C) &= \int_C \nu^n(B_1 \times B_2|t)dt \\ &= \int_C \int_{B_1 \times B_2} \hat{\nu}_1(d\zeta|t)\hat{\nu}_2^n(dy|\zeta,t)dt \\ &= \int_C \int_{B_1} \hat{\nu}_2^n(B_2|\zeta,t)\hat{\nu}_1^n(d\zeta|t)dt, \end{aligned} \tag{3.9}$$

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 $\nu^n(d\zeta \times dy \times dt) = \hat{\nu}_1^n(d\zeta|t) \otimes \hat{\nu}_2^n(dy|\zeta,t) \otimes dt$. The marginals of $\nu^n(d\zeta \times dy \times dt)$ can also be represented in terms of the kernels $\hat{\nu}_1^n(d\zeta|t)$ and $\hat{\nu}_2^n(dy|\zeta,t)$ and of Lebesgue measure λ on $I\!\!R$. Namely, for Borel sets B_1 and B_2 of $\mathcal{S}$ and C of $[0,1]$ the marginal over (ζ,t) can be written as

$$\begin{aligned} \nu^n(B_1 \times \mathcal{S} \times C) &= \int_C \hat{\nu}_1^n(B_1|t)dt \\ &\doteq (\hat{\nu}_1^n \otimes \lambda)(B_1 \times C) \end{aligned}$$

and the marginal over (y,t) as

$$\begin{aligned} \nu^n(\mathcal{S} \times B_2 \times C) &= \int_C \int_{\mathcal{S}} \hat{\nu}_2^n(B_2|\zeta,t)\hat{\nu}_1^n(d\zeta|t)dt \\ &\doteq (\hat{\nu}_2^n \otimes \lambda)(B_2 \times C). \end{aligned}$$

We note that the marginal over t is simply Lebesgue measure λ.

The next step is to manipulate the sum appearing in (3.8) in order to rewrite the right hand side of this inequality in terms of the measures ν^n. We first use the fact 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

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

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 above is no less than

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

which is clearly equivalent to

$$
\bar{E}_{x,\xi}\left\{\int_0^1 R(\nu^n(\cdot|t)\|\gamma^n(\cdot|t))dt\right\}, \tag{3.10}
$$

where, for each $n\in I\!N$ and $t\in[0,1]$, we define

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

for $l = 0,\ldots,1/k_n-1$. Define analogously the measure γ^n on $\mathcal{S}\times\mathcal{S}\times[0,1]$ as:

$$
\gamma^n(B_1\times B_2\times C) \doteq \int_C \gamma^n(B_1\times B_2|t)dt.
$$

Finally we apply the equality

$$
\int_{\mathcal{X}} R(\alpha(\cdot|x)\|\beta(\cdot|x))\gamma(dx) = R(\alpha\otimes\gamma\|\beta\otimes\gamma), \tag{3.11}
$$

valid for all stochastic kernels α and β on $\mathcal{S}$ given $\mathcal{X}$ and probability measures γ on $\mathcal{X}$ (see [4, Lemma 1.4.3 (f)]). In the present context, this equality implies that (3.10) can be rewritten as

$$\bar{E}_{x,\xi}\left\{R(\nu^n\|\gamma^n)\right\}.$$

Combining this series of inequalities with (3.8) we get

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

We now wish take the limit inferior as $n\to\infty$ of both terms in the last inequality. The asymptotic properties of the sequence $\{(\nu^n,\gamma^n,\bar{X}^n), n\in I\!N\}$ required to do this are proved in Theorem .7 in the Appendix. Accordingly, there exists a probability space where a subsequence of $\{(\nu^n,\gamma^n,\bar{X}^n), n\in I\!N\}$ converges in distribution to some limit $(\nu,\gamma,\bar{X})$. The stochastic kernels ν and γ and the random variable $\bar{X}$ (taking values in the continuous functions on [0,1]) satisfy all the conclusions stated in Lemma .7. Thanks to the Skorohod Representation Theorem [4, Theorem A.3.9], we can assume without loss of generality that convergence takes place with probability one. Along the convergent subsequence, lower semicontinuity of $R(\cdot\|\cdot)$ and Fatou's Lemma, plus continuity of h yield

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

The limit characterizations of ν,γ and $\bar{X}$ enable us to continue

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

$$= \bar{E}_{x,\xi}\left\{\int_0^1 L(\bar{X}(t),\dot{\bar{X}}(t))dt + h(\bar{X})\right\}$$
$$= \bar{E}_{x,\xi}\left\{I_{x,\xi}(\bar{X}) + h(\bar{X})\right\}$$
$$\geq \inf_{\phi\in\mathcal{C}([0,1]:I\!R^d)} \{I_{x,\xi}(\phi) + h(\phi)\}.$$

The third line uses part b) of Lemma .7; the fourth and fifth lines both use (3.11); the sixth line uses part h) of Theorem .4 and line seven follows from part e) of Lemma .7. Since the above inequality is valid for all $\varepsilon > 0$, (2.6) follows, concluding the proof of the upper bound.

4 Proof of the Laplace principle lower bound

This section is devoted to showing that (2.7) holds for all $h : \mathcal{C}([0,1] : I\!R^d) \mapsto I\!R^d$. ¿From part (b) of Corollary 1.2.5 in [4], it suffices to show that (2.7) holds for functions h in $C([0,1]; I\!R^d)$ that are Lipschitz continuous. Therefore, throughout this section we work under the assumption that h is Lipschitz continuous.

Following the proof of Proposition 6.6.1 in [4], the proof of (2.7) 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 the function L_σ, which is the analogue of the function L defined in (2.5) but for the perturbed process. Weak convergence arguments make use of these continuity properties, entailing the Laplace principle lower bound when taking the perturbation to be sufficiently small.

Given $\sigma > 0$, let $\{G_{j,\sigma}, j \in I\!N_0\}$ be a sequence of i.i.d. random variables on $I\!R^d$ with common Gaussian distribution ρ_σ, with mean zero and variance σI. We assume them to be independent of $\{\xi_j^x, x \in I\!R^d, j \in I\!N_0\}$, where ξ_j^x is the "fixed x" Markov process with transition kernel $p(\cdot|x,\xi_j^x)$. Given $n \in I\!N$ and $j \in \{0, 1, \ldots, n-1\}$, let X_j^n and Z_j^n be as before, and define

$$U_{0,\sigma}^n \doteq 0$$

$$U^n_{j+1,\sigma} \doteq U^n_{j,\sigma} + \frac{1}{n} G_{j,\sigma}.$$

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

$$Y^n_\sigma(t) \doteq X^n(t) + U^n_\sigma(t), \tag{4.12}$$

which is the piecewise linear interpolation of $\{X^n_j + U^n_{j,\sigma}\}$.

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 [10, p. 376]), namely, $b(x, \xi^x_j) + G_{j,\sigma}$. Here ξ^x_j is the "fixed x" Markov process described earlier. Let Q^x_σ be the stochastic kernel on $\mathcal{S} \times \mathbb{R}^d$ given $\xi \in \mathcal{S}$ defined by

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

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

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

These bounds on $Q^x_\sigma(\cdot, \cdot|\xi)$ and the fact that the convex hull of the support of $\upsilon_\sigma(\mathcal{S} \times \cdot|x) = \mathbb{R}^d$ guarantee [10, Lemma 3.1] the existence of a solution to the eigenvalue problem for each $x, \alpha \in \mathbb{R}^d$. That is, for ach $x, \alpha \in \mathbb{R}^d$ 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 Λ (see (1.2)). Finally, we define

$$L_\sigma(x, \beta) = \Lambda^*_\sigma(x, \alpha) \doteq \sup_{\alpha \in \mathbb{R}^d} \{\langle \alpha, \beta\rangle - \Lambda_\sigma(x, \alpha)\}. \tag{4.13}$$

Having introduced the necessary definitions, we proceed with the proof of (2.7). The properties of L_σ that will be used are proved in Lemma .5 in the Appendix.

Let K_1 be the Lipschitz constant of h and define $B \doteq 2||h||_\infty$. Then

$$h(Y^n_\sigma) = h(X^n + U^n_\sigma) \geq h(X^n) - (K_1||U^n_\sigma||_\infty \wedge B)$$

and, because of independence,

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

Hence

$$\begin{aligned} \limsup_{n\to\infty} W^n(x,\xi) &\leq \limsup_{n\to\infty}\Big(-\frac{1}{n}\log E_{x,\xi}\{\exp[-nh(Y^n_\sigma)]\}\Big) + \\ &\quad \limsup_{n\to\infty}\Big(\frac{1}{n}\log E_{x,\xi}\{\exp[n(K_1||U^n_\sigma|_\infty \wedge B)]\}\Big) \\ &\leq \limsup_{n\to\infty}\Big(-\frac{1}{n}\log E_{x,\xi}\{\exp[-nh(Y^n_\sigma)]\}\Big) + \frac{K_1^2\sigma^2}{2}, \end{aligned} \tag{4.14}$$

where the second inequality follows from [4, p.189]. This implies that (2.7) holds as long as we can show that

$$\limsup W^n_\sigma(x,\xi) \leq \inf_{\varphi\in C([0,1]:I\!\!R^d)}\{I_{x,\xi}(\varphi) + h(\varphi)\} + \theta(\sigma), \tag{4.15}$$

with $W^n_\sigma \doteq -\frac{1}{n}\log E_{x,\xi}\exp[-nh(Y^n_\sigma)]$, and $\theta(\sigma) \to 0$ when $\sigma \to 0$.

Now, given $\varepsilon > 0$, there exists $\psi \in C([0,1] : I\!\!R^d)$ such that

$$I_{x,\xi}(\psi) + h(\psi) \leq \inf_{\varphi\in C([0,1]:I\!\!R^d)}\{I_{x,\xi}(\varphi) + h(\varphi)\} + \varepsilon < \infty. \tag{4.16}$$

Then (4.15) will follow once we prove that

$$\limsup_{n\to\infty} W^n_\sigma(x,\xi) \leq I_{x,\xi}(\psi) + h(\psi) + \theta(\sigma). \tag{4.17}$$

The rest of this section is devoted to proving (4.17).

As was the case in the proof of the upper bound, a key step in the proof of the lower bound is the use of a variational representation. In this case we use a representation for the quantity $W^n_\sigma(x,\xi)$ for every $n \in I\!\!N$, so that

$$W^n_\sigma(x,\xi) = \inf_{\{\nu^n_j\}} \overline{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu^n_j(\cdot)||(p \times \rho_\sigma)(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + h(\overline{Y}^n) \right\}.$$

The admissible controls $\{\nu^n_j\}$, and the controlled processes $\bar{X}^n_j$, $\bar{Z}^n_j$ and $\bar{U}^n_j$ involved in the representation are described in detail in Theorem .2 in the Appendix. In particular, control sequences of the form

$$\begin{aligned}&\nu^n_{j,prod}(d\zeta \times dy)|\overline{X}^n_0, \ldots, \overline{X}^n_j, \overline{U}^n_0, \ldots, \overline{U}^n_j, \overline{Z}^n_j) \\ &\quad = \nu^{1,n}_j(d\zeta|\overline{X}^n_0, \ldots, \overline{X}^n_j, \overline{U}^n_0, \ldots, \overline{U}^n_j, \overline{Z}^n_j) \times \nu^{2,n}_j(dy|\overline{X}^n_0, \ldots, \overline{X}^n_j, \overline{U}^n_0, \ldots, \overline{U}^n_j, \overline{Z}^n_j)\end{aligned}$$

are admissible. Since for any probability measures γ and θ on $\mathcal{S}$, and λ and μ on $I\!\!R^d$ we have [4, Corollary C.3.3]

$$R(\gamma \times \lambda \| \theta \times \mu) = R(\gamma\|\theta) + R(\lambda\|\mu), \tag{4.18}$$

we can write

$$\begin{aligned}&\overline{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu^n_{j,prod}(\cdot)||(p \times \rho_\sigma)(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) \right\} \\ &\quad = \overline{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} [R(\nu^{1,n}_j(\cdot)||p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + R(\nu^{2,n}_j(\cdot)||\rho_\sigma(\cdot))] \right\}.\end{aligned}$$

Then the representation formula implies that

$$W^n_\sigma(x,\xi) \le \inf_{\{\nu^n_{j,prod}\}} \overline{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} \Big[R(\nu^{1,n}_j(\cdot)||p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + R(\nu^{2,n}_j(\cdot)||\rho_\sigma(\cdot)) \Big] + h(\overline{Y}^n) \right\}. \tag{4.19}$$

We will show that this inequality implies that

$$\limsup_{n\to\infty} W^n_\sigma(x,\xi) \le \int_0^1 L_\sigma(\psi(t), \dot{\psi}(t))dt + h(\psi) + \theta(\sigma), \tag{4.20}$$

for ψ as in (4.16). We observe that, since $L_\sigma(x,\beta) \le L(x,\beta)$ for all x and $\beta \in I\!\!R^d$ (Lemma .5 part a) in the Appendix), (4.17) will follow after that.

The proof of (4.20) proceeds as follows. Let ψ^* be as in part e) of Lemma .5 in the Appendix. We will design an admissible control sequence based on ψ^* with the following properties: the running costs are nearly optimal and, with probability converging to 1, the process $\overline{Y}^n \doteq \overline{X}^n + \overline{U}^n$ enters a small neighborhood of ψ^* as $n \to \infty$. The estimates that we obtain for the running costs will lead directly to (4.20).

Define the compact set

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

Let $\eta = \eta(\Delta,\sigma) \in (0,1)$ satisfy the conclusions of part (d) of Lemma .5 when taking $\varepsilon = \sigma$. Also, let $\{x_j, j = 1\ldots,n\}$ be a sequence in Δ satisfying $\|\psi^*(j/n) - x_j\| < \eta$. For every $n \in I\!\!N$, $j = 1,\ldots,n$ and with $x = \psi^*(j/n)$, $y = x_j$ and $\beta = \dot{\psi}^*(j/n)$, part (d) of that lemma implies that there exists $\overline{\beta}_j^n \in I\!\!R^d$ such that

$$L_\sigma(x_j,\overline{\beta}_j^n) - L_\sigma(\psi^*(j/n),\dot{\psi}^*(j/n)) \le \sigma \tag{4.21}$$

and

$$||\overline{\beta}_j^n - \dot{\psi}^*(j/n)|| \le K\|\psi^*(j/n) - x_j\|.$$

Further, $\overline{\beta}_j^n = \overline{\beta}_j^{1,n} + \overline{\beta}_j^{2,n}$, with

$$\overline{\beta}_j^{1,n} = \int_{\mathcal{S}} b(x_j,\xi)\mu^*_{j,n}(d\xi), \qquad \text{and} \qquad \overline{\beta}_j^{2,n} = \int_{I\!\!R^d} y\nu^*_{j,n}(dy).$$

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

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

and for $B_2 \in \mathcal{B}(I\!\!R^d)$

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

Here $\alpha = \alpha(\psi^*(j/n), \dot{\psi}^*(j/n))$ and $\dot{\psi}^*(j/n) = \int_{\mathcal{S}} b(\psi^*(j/n), \xi)\mu^*_{j,n}(d\xi) + \int_{I\!R^d} y\nu^*_{j,n}(dy)$.

We observe that, from part (h) of Lemma .4 in the Appendix,

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

Now, from part (c) of Lemma .5 and for $\overline{\beta}^n_j$ as in (4.21), there exist a stochastic kernel $\gamma^{1,n}_j(\cdot|x_j, \xi)$ on $\mathcal{S}$ given $\mathcal{S} \times I\!R^d$, with invariant measure μ^n_j, and a measure $\gamma^{2,n}_j$ on $I\!R^d$ which achieve the infimum in the representation for L_σ. That is,

$$L_\sigma(x_j, \overline{\beta}^n_j) = \int_{\mathcal{S}} R(\gamma^{1,n}_j(\cdot|x_j, \xi)||p(\cdot|x_j, \xi))\mu^n_j(d\xi) + R(\gamma^{2,n}_j||\rho_\sigma), \tag{4.22}$$

and

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

where we have used formula (4.18). The kernel $\gamma^{1,n}_j$ and the measure $\gamma^{2,n}_j$ are defined, respectively, as

$$\gamma^{1,n}_j(B_1|x_j, \xi) = \int_{B_1} e^{\langle \alpha, b(x_j,\zeta)\rangle + \Psi_\sigma(x_j;\alpha,\zeta) - \Psi_\sigma(x_j;\alpha,\xi) + \Lambda(x_j,\alpha)} p(d\zeta|x_j, \xi) \tag{4.23}$$

and

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

with $\alpha = \alpha(x_j, \overline{\beta}^n_j) = \alpha(x_j, \psi^*(j/n), \dot{\psi}^*(j/n))$ in both $\gamma^{1,n}_j$ and $\gamma^{2,n}_j$. Note that $\gamma^{2,n}_j$ depends implicitly of 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 build a sequence of admissible controls that will enable us to prove (4.17). The construction is similar to the one used in Section 3. Let $\{m_n, n \in I\!N\}$ be a sequence as the one used there, so that $m_n \to \infty$ as $n \to \infty$. Let $k_n = m_n/n$ and $l \in \{0, \ldots, \frac{1}{k_n} - 1\}$. Then for $lm_n \leq j < (l+1)m_n - 1$ we define

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

and

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

To simplify notation we have not made explicit the dependence on σ of $\nu_j^{1,n}$ and $\nu_j^{2,n}$.

Let $\overline{X}_0^n = x \in I\!R^d$, $\overline{U}_0^n = 0$, $\overline{Z}_0^n = \xi \in \mathcal{S}$, and define

$$\overline{X}_{j+1}^n = \overline{X}_j^n + \frac{1}{n}b(\overline{X}_j^n, \overline{Z}_{j+1}^n), \quad j=0,\ldots,n-1$$

and

$$\overline{U}_{j+1}^n = \overline{U}_j^n + \frac{1}{n}\overline{G}_j^n, \quad j=0,\ldots,n-1,$$

where $\overline{G}_j^n$, $\overline{Z}_j^n$, $lm_n \le j < (l+1)m_n - 1$ are random variables with joint distribution

$$\overline{P}_{x,\xi}\{(\overline{Z}_{j+1}^n, \overline{G}_j^n) \in (d\xi\times dy)|\overline{X}_0^n,\ldots,\overline{X}_j^n,\overline{U}_0,\ldots\overline{U}_j^n,\overline{Z}_j^n\} = \nu_j^{1,n}(d\zeta)\times\nu_j^{2,n}(dy).$$

Now, for $B_1, B_2 \in \mathcal{B}(S)$, $B \in \mathcal{B}(I\!R^d)$, define

$$\begin{aligned} \nu_l^{1,n}(B_1\times B_2) &\doteq \frac{1}{m_n}\sum_{j=l_n}^{(l+1)m_n-1}\delta_{\overline{Z}_j^n}(B_1)\times\nu_j^{1,n}(B_2|\overline{X}_0^n,\ldots\overline{X}_j^n,\overline{Z}_j^n), \\ \nu_l^{2,n}(B) &\doteq \frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1}\nu_j^{2,n}B|\overline{X}_0^n,\ldots,\overline{X}_j^n), \\ \nu^{1,n}(\cdot|t) &\doteq \begin{cases} \nu_l^{1,n}(\cdot) & \text{if} \quad t\in[lk_n,(l+1)k_n) \quad \text{for } l=0,\ldots,1/k_n-2 \\ \nu_{(\frac{1}{k_n}-1)}^{1,n}(\cdot) & \text{if} \quad t\in[1-k_n,1] \end{cases} \end{aligned}$$

and

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

Define the random measures $\nu^{1,n}$ and $\nu^{2,n}$ on $\mathcal{S}\times\mathcal{S}\times[0,1]$ and $I\!R^d\times[0,1]$, respectively, by:

$$\nu^{1,n}(B_1\times B_2\times C) \doteq \int_C \nu^{1,n}(B_1\times B_2|t)dt$$

and

$$\nu^{2,n}(B\times C) \doteq \int_C \nu^{2,n}(B|t)dt.$$

Further, let ν^n_{prod} be the random measure on $\mathcal{S} \times \mathcal{S} \times I\!\!R^d \times [0,1]$ defined as

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

Finally, let $\tau^n = \frac{1}{n}(\min\{i \in \{0,1,\ldots,n\} : \|\overline{X}^n_i - \psi^*(i/n)\| > \eta\} \wedge n)$.

With these definitions we have that for $j \le n\tau^n - 1$, and $lm_n \le j \le (l+1)m_n - 1$ for some $l \in \{0,\ldots,\frac{1}{k_n} - 1\}$,

$$R(\nu^{1,n}_j(\cdot)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) = R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j))$$

and

$$R(\nu^{2,n}_j(\cdot)\|\rho_\sigma(\cdot)) = R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot)).$$

Moreover, the running cost has the form

$$\begin{aligned}
&\overline{E}_{x,\xi}\Big\{\frac{1}{n}\sum_{j=0}^{n-1} R\left(\nu^{1,n}_j(\cdot) \times \nu^{2,n}_j(\cdot)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j) \times \rho_\sigma(\cdot)\right)\Big\} \\
&= \overline{E}_{x,\xi}\Big\{\frac{1}{n}\sum_{j=0}^{n\tau^n-1}[R(\nu^{1,n}_j(\cdot)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + R(\nu^{2,n}_j(\cdot)\|\rho_\sigma(\cdot))]\Big\} \\
&= \overline{E}_{x,\xi}\Big\{\sum_{l=0}^{q_n-1} k_n\left[\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + R(\gamma^{2,n}_{lm_n}(\cdot)\|\rho_\sigma(\cdot))\right] \\
&\quad + \frac{1}{n}\sum_{j=q_nm_n}^{n\tau^n-1}[R(\gamma^{1,n}_{q_nm_n}(\cdot|\overline{X}^n_{q_nm_n}, \overline{Z}^n_j)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j) + R(\gamma^{2,n}_{q_nm_n}(\cdot)||\rho_\sigma(\cdot))]\Big\}, \qquad (4.24)
\end{aligned}$$

where q_n is such that $n\tau^n = q_nm_n + r_n$, with $0 \le r_n < m_n$, and $q_n, r_n \in I\!\!N_0$.

We now 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,

$$R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)\|p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) \le R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)\|p(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)) + \sigma. \qquad (4.25)$$

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 $\|x - y\| < \delta$

$$\frac{\tilde{p}^y(\xi,\zeta)}{\tilde{p}^x(\xi,\zeta)} = 1 + \frac{\tilde{p}^y(\xi,\zeta) - \tilde{p}^x(\xi,\zeta)}{\tilde{p}^x(\xi,\zeta)} \le 1 + \frac{\tilde{p}^y(\xi,\zeta) - \tilde{p}^x(\xi,\zeta)}{a} \le e^\sigma. \qquad (4.26)$$

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

$$\begin{aligned}
&\gamma^{1,n}_{lm_n}(B_1|\overline{X}^n_{lm_n}, \overline{Z}^n_j) \\
&= \int_{B_1} e^{\langle \alpha, b(\overline{X}^n_{lm_n}, \zeta)\rangle + \Psi_\sigma(\overline{X}^n_{lm_n};\alpha,\zeta) - \Psi_\sigma(\overline{X}^n_{lm_n};\alpha,\xi) - \Lambda(\overline{X}^n_{lm_n},\alpha)} p(d\zeta|\overline{X}^n_{lm_n}, \overline{Z}^n_j) \\
&= \int_{B_1} e^{\langle \alpha, b(\overline{X}^n_{lm_n}, \zeta)\rangle + \Psi_\sigma(\overline{X}^n_{lm_n};\alpha,\zeta) - \Psi_\sigma(\overline{X}^n_{lm_n};\alpha,\xi) - \Lambda(\overline{X}^n_{lm_n},\alpha)} \tilde{p}^{\overline{X}^n_{lm_n}}(\overline{Z}^n_j, \zeta) \frac{\tilde{p}^{\overline{X}^n_j}(\overline{Z}^n_j, \zeta)}{\tilde{p}^{\overline{X}^n_j}(\overline{Z}^n_j, \zeta)} \vartheta(d\zeta) \\
&\leq e^\sigma \int_{B_1} e^{\langle \alpha, b(\overline{X}^n_{lm_n}, \zeta)\rangle + \Psi_\sigma(\overline{X}^n_{lm_n};\alpha,\zeta) - \Psi_\sigma(\overline{X}^n_{lm_n};\alpha,\xi) - \Lambda(\overline{X}^n_{lm_n},\alpha)} \tilde{p}^{\overline{X}^n_j}(\overline{Z}^n_j, \zeta) \vartheta(d\zeta) \qquad (4.27)
\end{aligned}$$

for any $B_1 \in \mathcal{B}(\mathcal{S})$. ¿From the above we get that $\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j) << p(\cdot|\overline{X}^n_j, \overline{Z}^n_j)$ and that

$$\begin{aligned}
\frac{d\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)}{dp(\cdot|\overline{X}^n_j, \overline{Z}^n_j)} &= \frac{d\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)}{d\tilde{p}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)} \cdot \frac{d\tilde{p}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)}{d\tilde{p}(\cdot|\overline{X}^n_j, \overline{Z}^n_j)} \\
&\leq e^\sigma \cdot \frac{d\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)}{d\tilde{p}(\cdot|\overline{X}^n_{lm_n}, \overline{Z}^n_j)},
\end{aligned}$$

which implies (4.25).

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

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

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

$$\frac{a}{A} \leq e^{\Psi_\sigma(x;,\alpha,\xi)} \leq \frac{A}{a}, \quad \forall \xi \in \mathcal{S},\ x,\ \alpha \in I\!R^d. \qquad (4.28)$$

Then, from (4.23),

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

with $\alpha = \alpha(x, \overline{\beta}^n_{lm_n})$, $x \in \Delta$, and $\overline{\beta}^n_{lm_n}$ to satisfy (4.21), and hence

$$\begin{aligned}
\gamma^{1,n}_{lm_n}(d\zeta|x,\xi) &\geq \frac{a^3}{A^2} e^{\langle \alpha, b(x,\zeta)\rangle - \Lambda(x,\alpha)} \vartheta(d\zeta) \\
&\geq \frac{a^3}{A^2} e^{-2||b||_\infty \max_x ||\alpha||} \vartheta(d\zeta).
\end{aligned}$$

Using the fact that $(x,\beta)\mapsto\alpha(x,\beta)$ is continuous, we get that for $x\in\Delta,\overline{\beta}^n_{\ell m_n}$ belongs to the set $\Theta\doteq\bigcup_{t\in[0,1]}\left\{\beta\in I\!R^d:||\beta-\dot{\psi}^*(t)||\le K\right\}$ and, moreover, that $\max_{\left\{\substack{x\in\Delta\\ \beta\in\Theta}\right\}}||\alpha||$ is bounded.

We complete the estimate on the running cost for our admissible control sequence in the inequalities that follow. Here $\gamma^{1,n,j}_{lm_n}$ shall denote the jth iteration of the kernel $\gamma^{1,n}_{lm_n}$. We observe that, from 16.0.2 in [11]:

$$\|\gamma^{1,n,j}_{lm_n}(\cdot|x,\zeta)-\nu_{lm_n}(\cdot)\|\le\rho^n,\quad\text{with}\quad\rho=1-\frac{a^3}{A^2}e^{-2\max_{x,\zeta}|\langle\alpha,b(x,\zeta)\rangle|}$$

We have that (4.24) is less than or equal to

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

$$
\begin{aligned}
& -\Lambda(\overline{X}^n_{lm_n}, \alpha_{lm_n})] + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))|\overline{Z}^n_0, \dots, \overline{Z}^n_{lm_n}, \overline{X}^n_0, \dots, \overline{X}^n_{lm_n}\}\Big\} + \sigma \\
= \; & \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \overline{E}_{x,\xi}\Big\{1_{[(r-1)m_n < n\tau^n \le rm_n]}\Big[\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1}(\langle\alpha_{lm_n}, \int_{\mathcal{S}} b(\overline{X}^n_{lm_n}, \zeta)\gamma^{1,n,j}_{lm_n}(d\zeta|\overline{X}^n_{lm_n}, \overline{Z}^n_{lm_n})\rangle \\
& + \int_{\mathcal{S}} \Psi_\sigma(\overline{X}^n_{lm_n}; \alpha_{lm_n}, \zeta)\gamma^{1,n,j+1}_{lm_n}(d\zeta|\overline{X}^n_{lm_n}, \overline{Z}^n_{lm_n}) - \int_{\mathcal{S}} \Psi_\sigma(\overline{X}^n_{lm_n}; \alpha_{lm_n}, \zeta)\gamma^{1,n,j}_{lm_n}(d\zeta|\overline{X}^n_{lm_n}, \overline{Z}^n_{lm_n}) \\
& -\Lambda(\overline{X}^n_{lm_n}, \alpha_{lm_n})) \pm \int_{\mathcal{S}} R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \xi)||p(\cdot|\overline{X}^n_{lm_n}, \xi))\mu_{lm_n}(d\xi) + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))\Big]\Big\} + \sigma \\
\le \; & \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \overline{E}_{x,\xi}\Big\{1_{[(r-1)m_n < n\tau^n \le rm_n]}\Big[\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1}\langle\alpha_{lm_n}, \int_{\mathcal{S}} b(\overline{X}^n_{lm_n}, \zeta)\gamma^{1,n,j+1}_{lm_n}(d\zeta|\overline{X}^n_{lm_n}, \overline{Z}^n_{lm_n}) \\
& - \int_{\mathcal{S}} b(\overline{X}^n_{lm_n}, \zeta)\mu_{lm_n}(d\zeta)\rangle + \int_{\mathcal{S}} R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \xi)||p(\cdot|\overline{X}^n_{lm_n}, \xi))\mu_{lm_n}(d\xi) \\
& + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))\Big]\Big\} + \frac{4}{n}\ln\frac{A}{a} + \sigma \\
\le \; & \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \overline{E}_{x,\xi}\Big\{1_{[(r-1)m_n < n\tau^n \le rm_n]}\Big[\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)\|\}} \\
& + \int_{\mathcal{S}} R(\gamma^{1,n}_{lm_n}(\cdot|\overline{X}^n_{lm_n}, \xi)||p(\cdot|\overline{X}^n_{lm_n}, \xi))\mu_{lm_n}(d\xi) + R(\gamma^{2,n}_{lm_n}(\cdot)||\rho_\sigma(\cdot))\Big]\Big\} + \frac{4}{n}\ln\frac{A}{a} + \sigma \\
\le \; & \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)\|\}} \\
& + \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \overline{E}_{x,\xi}\Big\{1_{[(r-1)m_n < n\tau^n \le rm_n]}\Big[L_\sigma(\overline{X}^n_{lm_n}, \overline{\beta}^n_{lm_n})\Big]\Big\} + \frac{4}{n}\ln\frac{A}{a} + \sigma \\
\le \; & \sum_{r=1}^{1/k_n} k_n \sum_{l=0}^{r-1} \overline{E}_{x,\xi}\Big\{1_{[(r-1)m_n < n\tau^n \le rm_n]} L_\sigma(\psi^*(\frac{lm_n}{n}), \dot{\psi}^*(\frac{lm_n}{n}))\Big\} \\
& + \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 \; & \overline{E}_{x,\xi}\Big\{\sum_{l=0}^{[\frac{\tau^n}{k_n}]} k_n L_\sigma(\psi^*(\frac{lm_n}{n}), \dot{\psi}^*(\frac{lm_n}{n}))\Big\} + 3\sigma, \quad \text{for } n \text{ large enough}, \\
\le \; & k_n \sum_{l=0}^{[1/k_n]} L_\sigma(\psi^*(\frac{lm_n}{n}), \dot{\psi}^*(\frac{lm_n}{n})) + 3\sigma. \qquad (4.29)
\end{aligned}
$$

We conclude that the admissible control sequence that we constructed has running cost which is nearly optimal, as we had claimed.

We can now return to the proof of (4.20). All of the claims that follow refer to the nearly optimal admissible control sequence constructed above. Since L_σ is continuous, ψ^*

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

$$\sup_n \overline{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^{1,n}(\cdot) \times \nu_j^{2,n}(\cdot) || p(\cdot|\overline{X}_j^n, \overline{Z}_j^n) \times \rho_\sigma(\cdot)) \right\} < \infty. \tag{4.30}$$

Then, Theorem 5.3.5 in [4] and the fact that $\mathcal{S}$ is compact imply that, given any subsequence of $\{(\nu^{1,n}, \nu^{2,n}, \overline{X}^n, \overline{U}^n, \tau^n), n \in I\!N\}$, there exists a subsubsequence such that $(\nu^{1,n}, \nu^{2,n}, \overline{X}^n, \overline{U}^n, \tau^n)$ converges in distribution to $(\nu^1, \nu^2, \overline{X}, \overline{U}, \tau)$, when $n \to \infty$. ¿From part e) of Theorem .7 in the Appendix we know that

$$\overline{X}(t) = x + \int_0^t \int_{\mathcal{S}} b(\overline{X}(s), \zeta) \hat{\nu}_1(d\zeta|s) ds,$$

where $\nu^1(d\zeta \times dy \times ds) = \hat{\nu}_1(d\zeta|s) \otimes \hat{\nu}_2(dy|\zeta, x) \otimes ds$ and $\hat{\nu}_1(d\zeta|s)$ is an invariant measure of $\hat{\nu}_2(dy|\zeta, s)$. Furthermore,

$$\overline{U}(t) = \int_{I\!R^d \times [0,1]} y \nu^2(dy \times ds),$$

and

$$\overline{Y}(t) \doteq \lim_{n\to\infty} \overline{Y}^n(t) = \lim_{n\to\infty} \overline{X}(t) + \lim_{n\to\infty} \overline{U}(t) = \overline{X}(t) + \overline{U}(t).$$

Let $\{(\nu^{1,n}, \nu^{2,n}, \overline{X}^n, \overline{U}^n, \tau^n)\}$ be a convergent subsubsequence. Then, (4.29) and Lemma .5(e) (with $\varepsilon = \sigma$) imply that

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

Thus, the proof of (4.20) will be complete once we prove that

$$\limsup \overline{E}_{x,\xi} h(\overline{Y}^n) \leq h(\psi) + \tilde{\theta}(\sigma),$$

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

$$\lim_{\sigma\to 0} \limsup_{n\to\infty} \overline{P}_{x,\xi}\{ \sup_{t\in[0,1]} ||\overline{Y}^n(t) - \psi^*(t)|| \geq \sigma\} = 0, \tag{4.31}$$

because of the Lipschitz property of h and Lemma .5(e).

¿From the definition of $\overline{\beta}_j^{1,n}$, and Lemma .5 (d), for each $l \in \{0, \ldots, [\frac{\tau^n}{k_n}]\}$,

$$\begin{aligned} ||\overline{\beta}^n_{lm_n} - \dot{\psi}^*(\frac{lm_n}{n})|| &= ||\overline{\beta}^{1,n}_{lm_n} + \overline{\beta}^{2,n}_{lm_n} - \dot{\psi}^*(\frac{lm_n}{n})|| \\ &= ||\int_{\mathcal{S}} b(\overline{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi) + \int_{I\!\!R^d} y\nu^{2,n}_{lm_n}(dy) - \dot{\psi}^*(\frac{lm_n}{n})|| \\ &\leq K||\overline{X}^n_{lm_n} - \psi^*(\frac{lm_n}{n})||. \end{aligned} \tag{4.32}$$

Also, from the proof Lemma .7(e), when $n \to \infty$

$$S^{1,n}(t) \doteq x + \int_0^t \int_{\mathcal{S}\times} b(S^{1,n}(s), \xi_2)\nu^{1,n}(d\xi_1 \times d\xi_2|s)ds \to \overline{X}(t)$$

and

$$S^{2,n}(t) \doteq \int_{I\!\!R^d \times [0,t]} y\nu^{2,n}(dy \times ds) \to \overline{U}(t)$$

in distribution for almost all $t \in [0, 1]$. Let $t \in [0, \tau]$. Then,

$$\begin{aligned} ||\overline{Y}(\,t\,) - \psi^*(t)|| &= \lim_{n\to\infty} ||\int_0^t \int_{\mathcal{S}\times\mathcal{S}} b(S^n(s), \xi_2)\nu^{1,n}(d\xi_1 \times d\xi_2|s)ds \\ &- \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\tilde{X}^n(s), \xi_2)\nu^{1,n}(d\xi_1 \times d\xi_2|s)ds + \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\tilde{X}^n(s), \xi_2)\nu^{1,n}(d\xi_1 \times d\xi_2|s)ds \\ &+ \sum_{l=0}^{[\frac{t}{k_n}]} k_n \int_{I\!\!R^d} y\nu^{2,n}_{lm_n}(dy) - \sum_{l=0}^{[\frac{t}{k_n}]} k_n \int_{\mathcal{S}} b(\tilde{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi) \\ &+ \sum_{l=0}^{[\frac{t}{k_n}]} k_n \int_{\mathcal{S}} b(\tilde{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi) - \int_0^t \dot{\psi}^*(s)ds|| \\ &\leq \limsup_{n\to\infty} \int_0^1 (K||S^{1,n}(s) - \tilde{X}^n(s)|| \wedge 2\|b\|_\infty)ds \\ &+ \limsup_{n\to\infty} \sum_{l=0}^{[\frac{t}{k_n}]} k_n || \frac{1}{m_n} \sum_{j=l}^{(l+1)m_n} \int_{\mathcal{S}} b(\overline{X}_{lm_n}, \zeta)\gamma^{1,n,j}_{lm_n}(d\zeta|\overline{X}^n_{lm_n}, \overline{Z}_{lm_n}) \\ &- \int_{\mathcal{S}} b(\overline{X}^n_{lm_n}, \xi)\mu^n_{lm_n}(d\xi)|| + \limsup_{n\to\infty} K \int_0^t ||\overline{X}^n(s) - \psi^*(s)||ds \\ &\leq \limsup_{n\to\infty} \sum_{l=0}^{[\frac{t}{k_n}]} k_n||b||_\infty \frac{A^2}{a_n^3} e^{2||b||_\infty \max_{x\in\Gamma,\beta\in\Theta}\{\|\alpha(x,\beta)\|\}} \\ &+K \int_0^t ||\overline{X}(s) - \psi^*(s)||ds \end{aligned}$$

$$
\begin{aligned}
&= K\int_0^t ||\overline{X}(s)-\psi^*(s)||ds \\
&\leq \int_0^t ||\overline{X}(s)-\overline{Y}(s)||ds + K\int_0^t ||\overline{Y}(s)-\psi^*(s)||ds \\
&\leq K\sup_{s\in[0,\tau]}||\overline{X}(s)-\overline{Y}(s)|| + K\int_0^t ||\overline{Y}(s)-\psi^*(s)||ds.
\end{aligned}
$$

Let $\overline{K} \doteq Ke^K$. By Gronwall's inequality,

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

which together with Lemma .6 implies that, for any $\sigma > 0$,

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

Since $\overline{X} = \overline{Y} - \overline{U}$, for $0 \leq \sigma \leq \eta/2$,

$$
\begin{aligned}
\lim_{\sigma\to 0}\overline{P}_{x,\xi}\{\sup_{t\in[0,\tau]}||\overline{X}(t)-\psi^*(t)|| \geq \sigma\} &\leq \lim_{\sigma\to 0}\overline{P}_{x,\xi}\{\sup_{t\in[0,\tau]}||\overline{U}(t)|| \geq \sigma/2\} \\
&\quad + \lim_{\sigma\to 0}\overline{P}_{x,\xi}\{\sup_{t\in[0,\tau]}||\overline{Y}(t)-\psi^*(t)|| \geq \sigma/2\} \\
&= 0.
\end{aligned}
$$

Finally, writing $\tau = \tau_\sigma$, it can be proved, following the same arguments given in [4, p. 205-206], that

$$\lim_{\sigma\to 0}\overline{P}_{x,\xi}\{\tau_\sigma < 1\} = 0,$$

which implies (4.31). This completes the proof of the lower bound.

Appendix

Representation Formulas

In this appendix we discuss the representation formulas used in the proof of Theorem 2.2. In particular, we shall establish variational representations for the quantity

$$W^n(x,\xi) \doteq -\frac{1}{n}\log E_{x,\xi}\left\{\exp[-nh(X^n)]\right\}, \tag{.33}$$

and for the quantity

$$W^n_\sigma(x,\xi) \doteq -\frac{1}{n}\log E_{x,\xi}\left\{\exp[-nh(Y^n_\sigma)]\right\}, \tag{.34}$$

where X^n and Y^n_σ are defined by (1.4) and (4.12), respectively, $E_{x,\xi}$ denotes expectation conditioned on $X^n_0 = x$ and $Z^n_0 = \xi$ and h is bounded and continuous. We will only discuss the representation for the first quantity in detail, since the representation for the second quantity is derived in a completely analogous way.

Theorem .1 states the representation for the quantity $W^n(x,\xi)$ in terms of the minimal cost function of a certain stochastic control problem. Before stating the theorem, we introduce the appropriate control problem.

We define a discrete-time controlled process taking values in $I\!R^d \times \mathcal{S}$ denoted by $\{(\bar{X}^n_j, \bar{Z}^n_j), j = 0, \ldots, n\}$. The control at time j is given by the distribution of the controlled random variable $\bar{Z}^n_j$, namely, a stochastic kernel $\nu(d\zeta|\bar{X}^n_0, \ldots, \bar{X}^n_j, \bar{Z}^n_j)$ on $\mathcal{S}$ given $(I\!R^d)^{j+1} \times \mathcal{S}$. A sequence of controls $\{\nu^n_j, j = 0, \ldots, n-1\}$ is what we refer to as an admissible control sequence.

Now, setting $\bar{Z}^n_0 = \xi$ and $\bar{X}^n_0 = x$, the evolution of the controlled process is through the relation

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

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

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

Finally, we define $\bar{X}^n = \{\bar{X}^n(t), t \in [0,1]\}$ to be the piecewise linear interpolation of $\{\bar{X}^n_j, j = 0, \ldots, n\}$ (see (1.4)). We consider the minimal cost function

$$V^n(x,\xi) \doteq \inf_{\{\nu^n_j\}} \bar{E}_{x,\xi} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu^n_j(\cdot) \| p(\cdot|\bar{X}^n_j, \bar{Z}^n_j)) + h(\bar{X}^n) \right\}. \qquad (.35)$$

Here R is the relative entropy function; $\nu^n_j(\cdot) = \nu^n_j(\cdot|\bar{X}^n_0, \ldots, \bar{X}^n_j, \bar{Z}^n_j)$; the infimum is taken over all admissible control sequences $\{\nu^n_j, j = 0, \ldots, n-1\}$; $\bar{E}_{x,\xi}$ denotes expectation conditioned on $\bar{X}^n_0 = x$ and $\bar{Z}^n_0 = \xi$; $\{(\bar{X}^n_j, \bar{Z}^n_j)\}$ is the controlled process associated with a particular control sequence $\{\nu^n_j\}$; and h is the same function appearing in the definition (.33) of $W^n(x,\xi)$. We can now state the representation formula.

Theorem .1 *Let h be a bounded measurable function mapping $\mathcal{C}([0,1]:I\!\!R^d) \mapsto I\!\!R$. Then for all $n \in I\!\!N, x \in I\!\!R^d$ and $\xi \in \mathcal{S}$ the quantity $W^n(x,\xi)$ defined in (.33) equals the minimal cost function $V^n(x,\xi)$ defined in* (.35), so that

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

Proof. For any starting time $i \in \{0,\ldots,n\}$ and any points $x_0, x_1, \ldots, x_i \in I\!\!R^d$ and $\xi_i \in \mathcal{S}$, we start by defining

$$W^n(i,\{x_0,\ldots,x_i\},\xi_i) \doteq -\frac{1}{n}\log E_\xi\left\{\exp\left[-nh(X^n)|X_0^n = x_0,\ldots,X_i^n = x_i, Z_i^n = \xi_i\right]\right\}.$$

We identify x_0 with x. Then $W^n(0,\{x\},\xi)$ equals $W^n(x,\xi)$, while $W^n(n,\{x,\ldots,x_n\},\xi_n)$ equals $h(X^n)$. Taking $i \in \{0,\ldots,n-1\}$, we use the Markov property to write:

$$\begin{aligned}
&\exp\left[-nW^n(i,\{x,\ldots,x_i\},\xi_i)\right] \\
&\quad = E_\xi\left\{\exp\left[-nh(X^n)\right]|X_0^n = x,\ldots,X_i^n = x_i, Z_0^n = \xi_0,\ldots,Z_i^n = \xi_i\right\} \\
&\quad = E_\xi\left\{E\left\{\exp\left[-nh(X^n)\right]|X_0^n = x,\ldots,X_{i+1}^n, Z_{i+1}^n, Z_i^n\right\}|X_0^n = x,\ldots,Z_i^n = \xi_i\right\} \\
&\quad = E_\xi\left\{\exp\left[-nW^n(i+1,\{x,\ldots,x_i+\frac{1}{n}b(x_i,\xi_{i+1})\},\xi_{i+1})\right]|X_0^n = x,\ldots,Z_i^n = \xi_i\right\} \\
&\quad = \int_{\mathcal{S}}\exp\left[-nW^n(i+1,\{x,\ldots,x_i+\frac{1}{n}b(x_i,\zeta)\},\zeta)\right]p(d\zeta|x_i,\xi_i).
\end{aligned}$$

Therefore

$$W^n(i,\{x,\ldots,x_i\},\xi_i) = -\frac{1}{n}\log\int_{\mathcal{S}}\exp\left[-nW^n(i+1,\{x,\ldots,x_i+\frac{1}{n}b(x_i,\zeta)\},\zeta)\right]p(d\zeta|x_i,\xi_i).$$

Applying the variational formula in Proposition 1.4.2(a) of [4] we obtain a dynamic programming equation (DPE) of the form

$$\left\{\begin{array}{l}
W^n(i,\{x,\ldots,x_i\},\xi_i) \\
\qquad = \inf_{\nu\in\mathcal{P}(\mathcal{S})}\left\{\frac{1}{n}R(\nu(\cdot)\|p(\cdot|x_i,\xi_i)) + \int_{\mathcal{S}}W^n(i,\{x,\ldots,x_i+\frac{1}{n}b(x_i,\zeta)\},\zeta)\nu(d\zeta)\right\} \\
W^n(n,\{x,\ldots,x_n\},\xi_n) = h(X^n).
\end{array}\right.$$

What we will show is that the quantities $W^n(i,\{x,\dots,x_i\},\xi)$ are not the only solution to the DPE, but that it is also satisfied by the minimal cost functions

$$V^n(i,\{x,\dots,x_i\},\xi_i) \doteq \inf_{\{\nu_j^n\}} \bar{E}_{i,\{x,\dots,x_i\}}\left\{\sum_{j=i}^{n-1}\frac{1}{n}R(\nu_j^n(\cdot)\|p(\cdot|\bar{X}_j^n,\bar{Z}_j^n)+h(X^n)\right\}.$$

Since the DPE and the terminal condition have a unique solution, then the conclusion of the theorem will follow.

We know that if a certain attainment condition is satisfied, then Theorem 1.5.2 in [4] guarantees that indeed $V^n(i,\{x,\dots,x_i\},\xi)$ are bounded, measurable and are the unique solution to the DPE. This attainment condition requires that there exist an admissible control sequence $\{\tilde{\nu}_j^n, j=0,\dots,n\}$ such that the infimum in the DPE is attained. Since such a sequence exists - it can be checked that the sequence of kernels $\{\nu_j^n, j=0,\dots n\}$ on $\mathcal{S}$ given $(I\!R^d)^i\times\mathcal{S}$ defined for every Borel set B of $\mathcal{S}$ through

$$\tilde{\nu}_j^n(B|x,\dots,x_j,\xi) \doteq \frac{\int_B \exp\left[-nW^n(j+1,\{x,\dots,x_j,x_j+\frac{1}{n}b(x_j,\zeta)\},\zeta)\right]p(d\zeta|x_j,\xi)}{\int_{\mathcal{S}} \exp\left[-nW^n(j+1,\{x,\dots,x_j,x_j+\frac{1}{n}b(x_j,\zeta)\},\zeta)\right]p(d\zeta|x_j,\xi)}$$

is indeed infimizing - the proof of the theorem is complete.

Using the same approach, one can prove the representation in Theorem .2 for the quantity $W_\sigma^n(x,\xi)$. We first state the appropriate stochastic control problem.

Let $p\times\rho_\sigma$ be the stochastic kernel on $S\times I\!R^d$, given $\xi\in\mathcal{S}$, $x\in I\!R^d$, 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

$\nu_j^n(d\zeta\times dy|x_0,\dots,x_j,u_0,\dots,u_j,z_j)$ on $\mathcal{S}\times I\!R^d$ given $(x_0,\dots,x_j),(u_0,\dots,u_j)\in(I\!R^d)^{j+1}$ and $z_j\in\mathcal{S}$. Once again we identify x_0 with x. For each admissible control sequence $\{\nu_j^n\}$, the controlled system is defined through

$$\begin{aligned}
\overline{X}_0^n &\doteq x & \overline{U}_0^n &\doteq 0\\
\overline{X}_{j+1}^n &\doteq \overline{X}_j^n+\tfrac{1}{n}b(\overline{X}_j^n,\overline{Z}_{j+1}^n), & \overline{U}_{j+1}^n &\doteq \overline{U}_j^n+\tfrac{1}{n}\overline{G}_j^n\\
\overline{Y}_j^n &= \overline{X}_j^n+\overline{U}_j^n,
\end{aligned}$$

where the conditional distribution of $(\overline{Z}^n_{j+1}, \overline{G}^n_j)$ is given by

$$\overline{P}_{x,\xi}\Big((\overline{Z}^n_{j+1}, \overline{G}^n_j) \in d\zeta \times dy | \overline{X}^n_0, \ldots, \overline{X}^n_j, \overline{U}^n_0, \ldots, \overline{U}^n_j, \overline{Z}^n_j\Big) = \nu^n_j(d\zeta \times dy | \overline{X}^n_0, \ldots, \overline{X}^n_j, \overline{U}^n_0, \ldots, \overline{U}^n_j, \overline{Z}^n_j).$$

Define now the processes $\overline{X}^n \doteq \{\overline{X}^n(t), t \in [0,1]\}$, $\overline{U}^n = \{\overline{U}^n(t), t \in [0,1]\}$ and $\overline{Y}^n \doteq \{\overline{Y}^n(t), t \in [0,1]\}$ as the linear interpolations of $\{\overline{X}^n_j\}$, $\{\overline{U}^n_j\}$ and $\{\overline{Y}^n_j\}$, respectively (see (1.4)). In this case the minimal cost function takes the form

$$V^n_\sigma(x,\xi) \doteq \inf_{\{\nu^n_j\}} \overline{E}_{x,\xi}\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu^n_j(\cdot)||(p \times \rho_\sigma)(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + h(\overline{Y}^n)\}.$$

Theorem .2 *Let h be a bounded measurable function mapping $\mathcal{C}([0,1] : I\!R^d) \mapsto I\!R$. Then for all $n \in I\!N, x \in I\!R^d, \xi \in \mathcal{S}$ and $\sigma > 0$, the quantity $W^n_\sigma(x,\xi)$ defined in (.34) equals the minimal cost function $V^n_\sigma(x,\xi)$, so that*

$$W^n_\sigma(x,\xi) = \inf_{\{\nu^n_j\}} \overline{E}_{x,\xi}\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu^n_j(\cdot)||(p \times \rho_\sigma)(\cdot|\overline{X}^n_j, \overline{Z}^n_j)) + h(\overline{Y}^n)\}.$$

Properties of the functions L and L_σ.

In this appendix we establish properties of the functions $\Lambda(x,\alpha)$ and $L(x,\beta)$ defined in (1.3) and (2.5) respectively, and of the function L_σ defined in (4.13).

Lemma .3 *Under Hypothesis H.1, the function $\Lambda(x,\alpha)$ defined in* (1.3) satisfies the following properties. For each $x \in I\!R^d$, $\Lambda(x,\alpha)$ is a finite strictly convex function of $\alpha \in I\!R^d$ which is differentiable for all α. In addition, $\Lambda(x,\alpha)$ is a continuous function of $(x,\alpha) \in I\!R^d \times I\!R^d$.

These properties follow from Lemmas 3.1 and 3.4 in [10] given the relation between the function Λ and the solution to the eigenvalue problem given in (1.2).

The next lemma establishes important properties of the function L which is the Frechel-Legendre transform of the function Λ.

Lemma .4 *Under Hypothesis H.1, the function $L(x,\beta)$ defined in* (2.5) satisfies the following properties.

a) For each $x \in I\!R^d$, $L(x,\beta)$ is a convex and lowersemicontinuous function of $\beta \in I\!R^d$.

b) $L(x,\beta) \geq -\Lambda(x,0)$.

c) $L(x,\beta)$ achieves its minimum of $-\Lambda(x,0)$ (equal to 0) at $\bar{\beta}$ if and only if $\bar{\beta} \in \partial\Lambda(0)$ and, furthermore, such a $\bar{\beta}$ exists.

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

e) There is uniqueness in the sense that there is only one convex and lower semicontinuous function Λ such that $L = \Lambda^*$.

f) If L achieves its minimum at $\beta = 0$ then $\Lambda(x,\alpha) \geq \Lambda(x,0)$.

g) $L(x,\beta) \to \infty$ as $\|\beta\| \to \infty$.

h) For each x and β in $I\!R^d$

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

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

Proof. Since Λ is finite and convex, properties a) - g) above are proved in [3, Lemma 2.1]. We now show h).

We first note that for any $\beta \in I\!R^d$ such that $L(x,\beta) = +\infty$ then the infimum is attained at any kernel γ whose invariant measure has mean β. Hence we can restrict our attention to $\beta \in \mathrm{ri}(\mathrm{dom} L(x,\cdot))$.

For $\alpha \in I\!R^d$ let γ_α be the stochastic kernel defined by

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

In terms of the function Λ defined in (1.2) we can write

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

Let μ_α be the unique invariant measure of γ_α. (Proposition 4.1 in [10] guarantees that such a measure exists). Part d) of the present lemma implies that there exists a unique $\alpha = \alpha(x,\beta)$ such that

$$L(x,\beta) = \langle \alpha(x,\beta), \beta \rangle - \Lambda(x, \alpha(x,\beta)) \tag{.36}$$

with $\alpha(x,\beta) \in \partial L(\beta)$ if and only if $\beta \in \partial H(\alpha(x,\beta))$, so that $\beta = \int_{\mathcal{S}} b(x,\xi)\mu^\alpha(d\xi)$. Since $\beta = \nabla \Lambda(\alpha(x,\beta))$, Proposition 4.1 in [10] gives

$$E^{\gamma_\alpha}_{\mu^\alpha} b(x,\xi) = \int_{\mathcal{S}} b(x,\xi)\mu^\alpha(d\xi) = \beta.$$

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

$$\int_{\mathcal{S}} b(x,\xi)\mu^\gamma(d\xi) = \beta$$

and

$$\int_{\mathcal{S}} R(\gamma(\xi|\cdot)\|p(\xi|\cdot))\mu^\gamma(\xi) < \infty.$$

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

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

$$= \int R(\gamma(\xi|\cdot)\|\gamma_\alpha(\xi|\cdot)\mu^\gamma(d\xi) - \Lambda(\alpha) + \langle \alpha, \beta\rangle$$
$$= \int R(\gamma(\xi|\cdot)\|\gamma_\alpha(\xi|\cdot)\mu^\gamma(d\xi) + L(x,\beta)$$
$$\geq L(x,\beta).$$

Equality is obtained if and only if $\gamma \equiv \gamma_\alpha$.

Remark.

a) Since $\Lambda(x,\cdot)$ is strictly convex on $I\!R^d$, $L(x,\cdot)$ is differentiable on $\text{int}(\text{dom}L(x,\cdot))$. See Theorem D.2.8 in [4].

b) For each $\beta \in \text{int}(\text{dom}L(x,\cdot))$ there exists $\alpha = \alpha(x,\beta) \in I\!R^d$ with $\alpha(x,\beta) = \nabla L(x,\beta)$ and $L(x,\beta) = \langle \alpha(x,\beta), \beta\rangle - \Lambda(x, \alpha(x,\beta))$. This follows from part (d) of Lemma .4

The next result establishes properties of the function L_σ defined in (4.13) that are needed in the proof of the lower bound.

Lemma .5 *Given $\sigma > 0$, the function $L_\sigma(x,\beta)$ satisfies the following properties.*

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

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

(c)

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

Further, for each $x, \beta \in I\!R^d$, 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 $I\!R^d$, 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 $\alpha = \alpha(x,\beta) \in \operatorname{argmax}\{\langle\alpha,\beta\rangle - \Lambda_\sigma(x,\alpha) : \alpha \in I\!R^d\}$.

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

a) $L_\sigma(y,\overline{\beta}) - L_\sigma(x,\beta) \le \varepsilon$

b) $||\overline{\beta} - \beta|| \le K||x-y||$,

where K *is the Lipschitz constant of* b.

(e) Given $\psi \in C([0,1], I\!R^d)$ *satisfying* $I_{x,\xi}(\psi) < \infty$ *and* $\varepsilon > 0$, *there exists* $\psi^* \in \mathcal{N}$ *such that* $||\psi - \psi^*||_\infty < \varepsilon$ *and*

$$\begin{aligned}\int_0^1 L_\sigma(\psi^*(t), \dot{\psi}^*(t))dt &\le \int_0^1 L_\sigma(\psi(t), \dot{\psi}(t))dt + \varepsilon \\ &\le I_{x,\xi}(\psi) + \varepsilon.\end{aligned}$$

(f) The function $(x,\beta) \to \alpha(x,\beta) \in argmax_{\alpha\in I\!R^d}\{\langle\alpha,\beta\rangle - \Lambda_\sigma(x,\alpha)\}$ *is continuous.*

Proof.

(a) Follows from Corollay D.4.2 in [D–E], while for the second part we take $z = 0$.

(b) From [12, Corollary 2.6.4] and Lemma 3.4 (iv) in [10], int(Dom $L_\sigma(x,\cdot)) =$ Range$(\nabla\Lambda_\sigma(x,\cdot)) = I\!R^d$. So $L_\sigma(x,\beta) < \infty$ for all $(x,\beta) \in I\!R^d \times I\!R^d$. 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 $x, \beta \in I\!R^d$. From Part (b) there exists $\alpha \in I\!R^d$, 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

$$\begin{aligned}\gamma(x,\alpha;\xi|d\zeta \times dy) &= \exp\{\langle\alpha, b(x,\zeta) + y\rangle + \Psi(x;\alpha,\zeta) - \Psi(x;\alpha,\xi) - \Lambda(x,\alpha) \\ &\quad -\frac{\sigma^2}{2}||\alpha||^2\} p(d\zeta|x,\xi)\rho_\sigma(dy).\end{aligned}$$

¿From Proposition 4.1 in [10],

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

where μ is the unique invariant measure of the kernel

$$\gamma_1(x,\alpha,\xi|d\zeta) = \exp\{\langle \alpha, b(x,\zeta) + y\rangle + \Psi(x;\alpha,\zeta) - \Psi(x;\alpha,\xi) - \Lambda(x,\alpha)\} p(d\zeta|x,\xi)$$

The rest of the proof follows the same arguments given in Lemma .4 (h).
(d) Let $\Delta \subset I\!\!R^d$ compact, $x, y \in \Delta$, $\beta \in I\!\!R^d$, and $\varepsilon > 0$. We know, from Part (c), that there exist γ^* and v^* such that the infimum in part (c) is attained for (x,β). Define $\overline{\beta} \doteq \int_{\mathcal{S}} b(y,\xi)\mu^{\gamma^*}(d\xi) + \int_{I\!\!R^d} y\nu^*(dy)$. Then, from the representation formula given in Part (c),

$$\begin{aligned} ||\beta - \overline{\beta}|| &\leq \| \int (b(x,\xi) - b(y,\xi))\mu^{\gamma^*}(d\xi)\| \\ &\leq K\|x-y\|. \end{aligned}$$

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 $p(\cdot|y,\xi)$; and from the uniform continuity of $\tilde{p}^x(\xi,\zeta)$, there exists $\eta > 0$ such that $\|x-y\| < \eta$ implies that

$$\tilde{p}^x(\xi,\zeta) \leq \tilde{p}^y(\xi,\zeta)e^{\varepsilon}. \qquad (.37)$$

(this is as in lines (4.26)). Then, from the variational equivalence given in Part (c), $L_\sigma(y,\overline{\beta}) < \infty$, and

$$\begin{aligned} L_\sigma(y,\overline{\beta}) &\leq \int R(\gamma^*(\cdot|x,\xi)\times \nu^*(\cdot)||p(\cdot|y,\xi)\times\rho_\sigma(\cdot))\mu^{\gamma^*}(d\xi) \\ &\leq \langle \alpha, \int b(x,\xi)\mu^{\gamma^*}(d\xi)\rangle - \Lambda(x,\alpha) + \langle \alpha, \int_{I\!\!R^d} y\nu^*(dy)\rangle - \frac{\sigma^2}{2}||\alpha|| + \varepsilon \\ &= \langle \alpha, \beta\rangle - \Lambda_\sigma(x,\alpha) + \varepsilon \\ &= L_\sigma(x,\beta) + \varepsilon \end{aligned}$$

(e) The proof of this part is based in 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 $x, \beta \in I\!R^d$, Part (b) and the differenciability 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 [12, Corollary 25.5.1]. Moreover, $x \to \nabla_\beta L_\sigma(x,\beta)$ is continuous [12, 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.

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

Lemma .6 *For any $\delta > 0$,*

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

Proof. Let $(\nu^{1,n}, \nu^{2,n}, \overline{X}^n, \overline{U}^n, \tau^n) \to (\nu^1, \nu^2, \overline{X}, \overline{U}, \tau)$ and define

$$S^n(t) \doteq \int_{I\!R^d\times[0,1]} y\nu^{2,n}(dy\times ds).$$

¿From (4.30), $\sup_n \overline{E}_{x,\xi}\{\frac{1}{n}\sum_{j=0}^{n-1} R(\nu_j^{2,n}||\rho)\} < \infty$, and Proposition 5.3.8 in [4] implies that $S^n \to \overline{U}$ in distribution in $C([0,1]; I\!R^d)$. The next inequality follows from the Skorohod Representation Theorem and Fatou's Lemma:

$$\overline{E}_{x,\xi}\{(\sup_{t\in[0,1]} ||\overline{U}(t)||)^2\} \leq \liminf_{n\to\infty} \overline{E}_{x,\xi}\{(\sup_{t\in[0,1]} ||S^n(t)||)^2\}. \quad (.38)$$

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

$$\begin{aligned}
\overline{P}_{x,\xi}[\sup_{t\in[0,1]} ||\overline{U}(t)|| \geq \delta] &\leq \frac{1}{\delta^2}\overline{E}_{x,\xi}\{(\sup_{t\in[0,1]} ||\overline{U}(t)||)^2\} \\
&\leq \frac{1}{\delta^2}\liminf_{n\to\infty} \overline{E}_{x,\xi}\{(\sup_{t\in[0,1]} ||\int_{I\!R^d\times[0,1]} y\nu^{2,n}(dy\times ds)||)^2\} \\
&\leq \frac{1}{\delta^2}\liminf_{n\to\infty} \overline{E}_{x,\xi}\{\frac{1}{n}\sum_{j=0}^{n-1} ||\int_{I\!R^d} y\nu_j^{2,n}(dy)||^2\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\delta^2} \liminf_{n\to\infty} \overline{E}_{x,\xi}\{\frac{1}{n}\sum_{j=0}^{n\tau^n-1} ||\beta_j^{2,n}||^2\} \\
&\le \frac{2\sigma^2}{\delta^2} \liminf_{n\to\infty} \overline{E}_{x,\xi}\{\frac{1}{n}\sum_{j=0}^{n\tau^n-1} R(\gamma_j^{2,n}||\rho_\sigma)\} \\
&\le \frac{2\sigma^2}{\delta^2} \liminf_{n\to\infty}\{k_n \sum_{l=0}^{[1/k_n]} L_\sigma(\psi^*(\frac{lm_n}{n}), \dot{\psi}^*(\frac{lm_n}{n})) + 3\sigma\} \\
&= \frac{2\sigma^2}{\delta^2}[\int_0^1 L_\sigma(\psi^*(t), \dot{\psi}^*(t))dt + 3\sigma] \\
&\le \frac{2\sigma^2}{\delta^2}[I_{x,\xi}(\psi) + 4\sigma].
\end{aligned}
$$

The fifth inequality follows noting that $\frac{1}{2\sigma^2}||\beta_j^{2,n}||^2 = L_\sigma(\beta_j^{2,n}) \le R(\gamma_j^{2,n}||\rho)$, where L_σ is the Frechel-Legendre of the moment generating function of ρ_σ. The sixth line follows from (4.24) and (4.29), while in line eight we used Part (e) of Lemma .5 with $\varepsilon = \sigma$. Finally we et

$$
\overline{P}_{x,\xi}(\sup_{t\in[0,1]} ||\overline{U}(t)|| \ge \delta) \le \lim_{\sigma\to 0} \frac{2\sigma^2}{\delta^2}[I_{x,\xi}(\psi) + 4\sigma] = 0,
$$

which proves the lemma.

Proofs of some limit results

This appendix is dedicated to the proofs of some limit results needed in the proof of Theorem 2.2.

Theorem .7 *For any $x \in I\!R^d, \xi \in \mathcal{S}$ and each $n \in I\!N$, consider any admissible control sequence such that*

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

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 $\{\tilde{\nu}_2^n \otimes \lambda, \tilde{\nu}_1^n \otimes \lambda\}$, and the measures $\{\gamma^n\}$ as in Section 3. The following conclusions hold.

a) *Given any subsequence of* $\{(\nu^n, \tilde{\nu}_2^n \otimes \lambda, \tilde{\nu}_1^n \otimes \lambda, \gamma^n, n \in I\!N\}$, *there exist a subsubsequence, a probability space* $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}_{x,\xi})$, *a a measure* ν *on* $\mathcal{S}\times\mathcal{S}\times[0,1]$ *with marginals* μ_1, μ_2, *and a measure* γ *on* $\mathcal{S}\times\mathcal{S}\times[0,1]$ *such that the subsubsequence converges in distribution to* $(\nu, \mu_1, \mu_2, \gamma)$.

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* $[0,1]\times\bar{\Omega}$, $\hat{\nu}_1(\cdot|t)$ *on* $\mathcal{S}$ *given* $[0,1]\times\bar{\Omega}$ *and* $\hat{\nu}_2(\cdot|\zeta,t)$ *on* $\mathcal{S}$ *given* $\mathcal{S}\times[0,1]\times\bar{\Omega}$.

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

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

e) *Let* $\{n\}$ *be the subsubsequence obtained in Part (a). Then* $\{\bar{X}^n, \tilde{X}^n\}$ *converges to* $\bar{X}$ *almost surely, where for every* $t \in 0,1]$

$$\begin{aligned}\bar{X}(t) &\doteq x + \int_{\mathcal{S}\times\mathcal{S}\times[0,t]} b(\bar{X}(s), y)\nu(d\zeta \times dy \times ds) \\ &= x + \int_0^t \int_{\mathcal{S}} b(\bar{X}(s), \xi)\hat{\nu}_1(d\zeta|s)ds,\end{aligned}$$

with $\hat{\nu}_1$ *an invariant measure for* $\bar{q}$.

f) *The kernel* γ *is has the decomposition*

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

Proof.

a) Given the compactness of $\mathcal{S}$, we immediately get tightness of ν^n, of γ^n and of all the marginals. Since the function mapping ν^n into $(\nu^n, \tilde{\nu}_1^n \otimes \lambda, \tilde{\nu}_2^n \otimes \lambda)$ is continuous, there exist measures ν and γ over $\mathcal{S}\times\mathcal{S}\times[0,1]$ and measures μ_1 and μ_2 over $\mathcal{S}\times[0,1]$ such that $(\nu^n, \tilde{\nu}_1^n\otimes\lambda, \tilde{\nu}_2^n\otimes\lambda, \gamma^n) \xrightarrow{\mathcal{D}} (\nu, \mu_1, \mu_2)$ [4, Theorem A.3.6]. Moreover, w.p.1 μ_1 and μ_2 equal the marginals of ν over (z,t) and over (y,t) respectively.

b) We let μ_3 denote the marginal of ν over t. By the Skorohod Representation Theorem we can assume that $\nu^n \Longrightarrow \nu$ and for $i = 1,2$ $\tilde{\nu}_i^n \otimes \lambda \Longrightarrow \mu_i$ w.p.1 on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$. Using the fact that the marginal of ν^n is Lebesgue measure λ, we have that w.p.1 for any bounded continuous function g mapping $[0,1]$ into $I\!\!R$

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

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 on $\mathcal{S}$ given $\mathcal{S}\times[0,1]$ $\tilde{\nu}_2(dy|\zeta,t)$, $\tilde{\nu}_1(d\zeta|t)$ the second and first marginals of $\nu(d\zeta\times dy|t)$ such that

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

This gives the decomposition of $\nu(d\zeta\times dy\times dt)$ given in part (b). In terms of these kernels we can write $\mu_1(d\zeta\times dt) = \tilde{\nu}_1(d\zeta|t)\otimes\lambda(dt)$ and $\mu_2(dy\times dt) = \tilde{\nu}_2(dy|\zeta,t)\times\tilde{\nu}_1(\mathcal{S}|t)\times\lambda(dt)$.

c) It is enough to check that

$$\int f(y,t)(\tilde{\nu}_1\otimes\lambda)(dy\times dt) = \int f(y,t)(\tilde{\nu}_2\otimes\lambda)(dy\times dt)$$

with $f(y,t) = g(y)h(t)$, $g \in \mathcal{C}(\mathcal{S} : I\!\!R^d)$ and $h \in \mathcal{C}([0,1] : I\!\!R^d)$ (see Theorem A.3.14 in [4]). 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. Therefore

$$\begin{aligned}
\int_{[0,1]\times\mathcal{S}} f(\zeta,t)(\tilde{\nu}^n_1\otimes\lambda)(d\zeta\times dt) &= \int_0^1 h(t)\int_{\mathcal{S}} g(\zeta)\tilde{\nu}^n_1(d\zeta|t)dt \\
&= \sum_{l=0}^{\frac{1}{k_n}-1}\int_{lk_n}^{(l+1)k_n} h(t)\int_{\mathcal{S}} g(\zeta)\tilde{\nu}^n_1(d\zeta|t)dt \\
&= \sum_{l=0}^{\frac{1}{k_n}-1}\int_{lk_n}^{(l+1)k_n} h(t)dt\cdot\int_{\mathcal{S}} g(\zeta)(\nu^n_l)_1(d\zeta) \\
&= \sum_{l=0}^{\frac{1}{k_n}-1}\int_{lk_n}^{(l+1)k_n} h(t)dt\cdot\frac{1}{m_n}\sum_{j=lm_n}^{(l+1)m_n-1} g(\bar{Z}^n_j).
\end{aligned}$$

Similarly,

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

Therefore

$$\begin{aligned}
&\int_0^1\int_{\mathcal{S}} f(y,t)(\tilde{\nu}^n_1\otimes\lambda)(dy\times dt) - \int_0^1\int_{\mathcal{S}} f(y,t)(\tilde{\nu}^n_2\otimes\lambda)(dy\times dt) \\
&\quad = \sum_{l=0}^{\frac{1}{k_n}-1}\frac{1}{m_n}\left[\int_{lk_n}^{(l+1)k_n} h(t)dt\right]\left[\sum_{j=lm_n}^{(l+1)m_n-1}\left(g(\bar{Z}^n_{j+1}) - \int g(y)\nu^n_j(dy|\bar{X}^n_0,\ldots,\bar{X}^n_j,\bar{Z}^n_j)\right)\right. \\
&\qquad \left. + [g(\bar{Z}^n_{lm_n}) - g(\bar{Z}^n_{(l+1)m_n})]\right].
\end{aligned}$$

Note that

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

so that if $m_n \geq \frac{4\|g\|\|h\|}{\varepsilon}$ then

$$\begin{aligned}
&\bar{P}_{x,\xi}\left\{\left|\int_{\mathcal{S}} f d(\tilde{\nu}_1^n \otimes \lambda) - \int_{\mathcal{S}} f d(\tilde{\nu}_2^n \otimes \lambda)\right| \geq \varepsilon\right\} \\
&\quad \leq \bar{P}_{x,\xi}\left\{\left|\sum_{l=0}^{\frac{1}{k_n}-1} \frac{1}{m_n}\left[\int_{lk_n}^{(l+1)k_n} h(t)dt\right]\left[\sum_{j=lm_n}^{(l+1)m_n-1}\Big(g(\bar{Z}_{j+1}^n) - \int g(y)\nu_j^n(dy)\Big)\right]\right| \geq \frac{\varepsilon}{2}\right\} \\
&\quad \leq \bar{P}_{x,\xi}\left\{\left|\frac{1}{n}\sum_{j=0}^{n-1}\left[g(\bar{Z}_{j+1}^n) - \int g(y)\nu_j^n(dy)\right]\right| \geq \frac{\varepsilon}{2\|h\|}\right\} \\
&\quad \leq \frac{4\|h\|^2}{\varepsilon^2}\bar{E}_{x,\xi}\left\{\frac{1}{n^2}\left(\sum_{j=0}^{n-1}\left[g(\bar{Z}_{j+1}^n) - \int g(y)\nu_j^n(dy)\right]\right)^2\right\} \\
&\quad \leq \frac{4\|h\|^2}{\varepsilon^2}\bar{E}_{x,\xi}\left\{\frac{1}{n^2}\sum_{j=0}^{n-1}\left(g(\bar{Z}_{j+1}^n) - \int g(y)\nu_j^n(dy)\right)^2\right\} \\
&\quad \leq \frac{4\|h\|^2}{\varepsilon^2}\cdot\frac{1}{n^2}\cdot 4\|g\|^2\cdot n = \frac{16\|h\|^2\|g\|^2}{n\varepsilon^2}.
\end{aligned}$$

This implies that $\int_{\mathcal{S}} f d(\tilde{\nu}_1^n \otimes \lambda) - \int_{\mathcal{S}} f d(\tilde{\nu}_2^n \otimes \lambda)$ converges to zero in probability, so that ([4, Theorem A.3.7])

$$\int_{\mathcal{S}} f d(\tilde{\nu}_1^n \otimes \lambda) - \int_{\mathcal{S}} f d(\tilde{\nu}_2^n \otimes \lambda) \xrightarrow{\mathcal{D}} 0.$$

Given the convergence in distribution of $\tilde{\nu}_1^n \otimes \lambda$ and $\tilde{\nu}_2^n \otimes \lambda$ to $\nu_1 \otimes \lambda$ and $\nu_2 \otimes \lambda$ respectively, we can appply the Skorohod Representation Theorem to get w.p.1

$$\lim_{n\to\infty}\int_{\mathcal{S}} f d(\tilde{\nu}_1^n \otimes \lambda) = \int_{\mathcal{S}} f d(\tilde{\nu}_1 \otimes \lambda)$$

and

$$\lim_{n\to\infty}\int_{\mathcal{S}} f d(\tilde{\nu}_2^n \otimes \lambda) = \int_{\mathcal{S}} f d(\tilde{\nu}_2 \otimes \lambda).$$

Therefore w.p.1

$$\int_{\mathcal{S}} f d(\tilde{\nu}_1 \otimes \lambda) = \int_{\mathcal{S}} f d(\tilde{\nu}_2 \otimes \lambda). \tag{.39}$$

Theorem A.3.14 in [4] implies that we can extend the equality in (.39) from f of the form $f(x,t) = g(x)h(t)$ to all $f : \mathcal{S} \times [0,1] \mapsto I\!R$ that is bounded and continuous. Since the class of bounded continous functions is measure-determing, we have that $\tilde{\nu}_1 \otimes \lambda = \tilde{\nu}_2 \otimes \lambda$, as we wanted to show.

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

$$\tilde{\nu}_1(B_1|t) = \int_{\mathcal{S}} \tilde{\nu}_2(B_1|\zeta, t)\tilde{\nu}_1(d\zeta|t). \tag{.40}$$

Let $\mathcal{U}_b(\mathcal{S})$ denote the space of bounded, uniformly continuous functions mapping $\mathcal{S}$ into $I\!R$. 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, and define

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

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

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

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

$$\begin{aligned}
&|h_1(s) - h_2(s)| \\
&\quad \le \left| \frac{1}{\lambda(E_i)} \int_{E_i} h_1(t)dt - h_1(s) \right| + \left| \frac{1}{\lambda(E_i)} \int_{E_i} h_2(t)dt - h_2(s) \right| \\
&\quad \le \frac{1}{\lambda(E_i)} \int_{E_i} |h_1(t) - h_1(s)|\, dt + \frac{1}{\lambda(E_i)} \int_{E_i} |h_2(t) - h_2(s)|\, dt,
\end{aligned}$$

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., that is,

$$\int_{\mathcal{S}} g(y)\tilde{\nu}_1(dy|t) = \int_{\mathcal{S}} \int_{\mathcal{S}} g(y)\tilde{\nu}_2(dy|\zeta,t)\tilde{\nu}_1(d\zeta|t) \qquad \text{a.s.}$$

This in turn implies that there exists a set $B_g \in \mathcal{B}([0,1])$ with $\lambda(B_g) = 0$ and such that $\int_{\mathcal{S}} g(y)\tilde{\nu}_1(dy|t) = \int_{\mathcal{S}} \int_{\mathcal{S}} g(y)\tilde{\nu}_2(dy|\zeta,t)\tilde{\nu}_1(d\zeta|t)$ for all $t \notin B_g$. Now define the set B as

$$B \doteq \cup_{g \in \mathcal{E}} B_g.$$

Then $\lambda(B) = 0$ and for all $t \notin B$ and $g \in \mathcal{E}$ we have

$$\int g(y)\tilde{\nu}_1(dy|t) = \int_{\mathcal{S}}\int_{\mathcal{S}} g(y)\tilde{\nu}_2(dy|\zeta, t)\tilde{\nu}_1(d\zeta|t).$$

This equality, valid for $g \in \mathcal{E}$, is extended to $g \in \mathcal{U}_b(\mathcal{S}, m)$, which implies that $\tilde{\nu}_1(dy|t) = \tilde{\nu}_2(dy|t)$ for all $t \notin B$. Finally, redefining in an obvious way $\tilde{\nu}_1$ and $\tilde{\nu}_2$ for $t \in B$, we get .40.

References

[1] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.

[2] A. Dembo, O. Zeitouni. *large Deviations Techniques and Applications* Second Edition. Springer Verlag, 1998.

[3] P. Dupuis. Large deviations analysis of some recursive algorithms with state dependent noise. *The Annals of Probability*, **16**, 1509-1536, 1988.

[4] P. Dupuis, R. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, 1996.

[5] S.N. Ethier, T.G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, New York, 1986.

[6] M.I. Freidlin, A.D. Wentzell. *Random perturbations of dynamical systems*. Springer–Verlag, Berlin, 1984.

[7] H.J. Kushner. *Approximation and weak convergence methods for random processes with applications to stochastic system theory*. Cambridge, MIT Press, 1984.

[8] H.J. Kushner and F.J. Vázquez-Abad. Stochastic approximation methods for systems over an infinite horizon, SIAM Journal on Control and Optimization **34**, 712-756, 1996.

[9] H.Y. Kushner and G.G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, NY, 1997.

[10] I. Iscoe, P. Ney, E. Nummelin. Large deviations of uniformly recurrent Markov additive processes. Advances in Applied Mathematics **6**, 373-412, 1985.

[11] S.P. Meyn and R.L. Tweedie. Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993.

[12] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.