Chapter 4
Recursive Markov Systems with Small Noise

In Chap. 3 we presented several examples of representations and how they could be used for large deviation analysis. A simplifying feature of all the examples of Chap. 3 is that the process models (e.g., empirical measure, solution to an SDE) could be thought of as a "nice" functional of a process that is "white" in the time variable, by which we mean independent in the setting of discrete time, and with independent increments in the setting of continuous time (see Sect. 3.5 for what is meant by a nice functional in the case of small noise SDEs).

In this chapter we study a model for which there is, in general, no convenient representation as a functional of white noise. Note that we do not claim that such a representation is impossible, but rather that it will not (in general) be useful, e.g., in proving law of large numbers limits. Because of this feature, a more complex representation and weak convergence analysis cannot be avoided. In particular, the "base" measure in the representation will be a Markov measure rather than a product measure, and the process model will be a general "small noise" Markov process. The model provides a substantial generalization of the random walk considered in Cramér's theorem. It occurs frequently in stochastic systems theory, e.g., stochastic approximation and related recursive algorithms [18, 182, 193], where the rate function can be used to define a rate of convergence [102]. The model also arises as a discrete time approximation to various continuous time models, such as the small noise SDE in Sect. 3.2.1 of Chap. 3, and indeed provides an alternative approach to proving large deviation estimates for such models (though we much prefer the direct approach of Chap. 3).

4.1 Process Model

We begin with a description of the process model. Suppose that $\theta(dy|x)$ is a stochastic kernel on $\mathbb{R}^d$ given $\mathbb{R}^d$. One can construct a probability space that supports **iid random vector fields** $\{v_i(x), i \in \mathbb{N}_0, x \in \mathbb{R}^d\}$, with the property that for all

$x \in \mathbb{R}^d$, $v_i(x)$ has distribution $\theta(\cdot|x)$. To be precise, there exists a probability space $(\Omega, \mathscr{F}, P)$ such that for each $i \in \mathbb{N}_0$, v_i is a measurable map from $\mathbb{R}^d \times \Omega$ to $\mathbb{R}^d$; for $k \in \mathbb{N}$ and distinct $i_1, \ldots, i_k \in \mathbb{N}_0$ and $x_{i_1}, \ldots, x_{i_k} \in \mathbb{R}^d$, the random vectors $v_{i_1}(x_{i_1}), \ldots, v_{i_k}(x_{i_k})$ are mutually independent; and for each $i \in \mathbb{N}_0$, $v_i(x)$ has distribution $\theta(\cdot|x)$. We then define for each $n \in \mathbb{N}$ a Markov process $\left\{X_i^n\right\}_{i=0,\ldots,n}$ by setting

$$X_{i+1}^n = X_i^n + \frac{1}{n}v_i(X_i^n), \quad X_0^n = x_0. \tag{4.1}$$

This discrete time process is interpolated into continuous time according to

$$X^n(t) = X_i^n + \left[X_{i+1}^n - X_i^n\right](nt - i), \quad t \in [i/n, i/n + 1/n]. \tag{4.2}$$

The goal of this chapter is to study a large deviation principle for the sequence $\{X^n\}_{n \in \mathbb{N}}$ of $\mathscr{C}([0, T] : \mathbb{R}^d)$-valued random variables.

Example 4.1 Suppose that for each $x \in \mathbb{R}^d$, $v_i(x)$ has a normal distribution with continuous mean $b(x)$ and covariance $\sigma(x)\sigma^T(x)$. Then $X^n(t)$ is the Euler approximation with step size $1/n$ to the SDE (3.15) with drift coefficient b, diffusion coefficient σ, and $\varepsilon = 1/n$.

Example 4.2 For an example in the form of a stochastic approximation algorithm, take $v_i(x) = -\nabla V(x) + w_i$, where the w_i are iid with $Ew_i = 0$ and V is a smooth function. In this case, $1/n$ is the "gain" of the algorithm [18, 182].

Of course, to prove an LDP for $\{X^n\}_{n \in \mathbb{N}}$, additional assumptions must be made. For $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}^d$, define

$$H(x, \alpha) \doteq \log Ee^{\langle \alpha, v_i(x) \rangle}.$$

Condition 4.3 (a) For each $\alpha \in \mathbb{R}^d$, $\sup_{x \in \mathbb{R}^d} H(x, \alpha) < \infty$.

(b) The mapping $x \mapsto \theta(\cdot|x)$ from $\mathbb{R}^d$ to $\mathscr{P}(\mathbb{R}^d)$ is continuous in the topology of weak convergence.

The first condition is not needed for an LDP to hold. However, if $H(x, \alpha) = \infty$ for some values of x and α, then $v_i(x)$ has relatively heavy tails in certain directions. Paths with jumps may be important from the perspective of large deviations, and the setting used here is no longer appropriate. The second condition can also be weakened. However, this often leads to a qualitatively different form of the rate function, and the process models that violate this condition are said to have "discontinuous statistics" [95, 98]. For an example of such a process but in continuous time, see Chap. 13.

4.2 The Representation

The first issue to resolve is the formulation of a representation that reflects the natural structure of the process model. As noted at the beginning of the chapter, it is possible to represent $\{X^n\}$ in terms of iid random variables, e.g., in the form $X_{i+1}^n = X_i^n + \frac{1}{n}g(X_i^n, U_i)$, where g is measurable and the $\{U_i, i \in \mathbb{N}_0\}$ are iid random variables with uniform distribution on $[0, 1]$. Although this form would allow a representation in terms of an iid base measure, it would not be useful. This is because the map g is not in general continuous in x, and hence this formulation is poorly suited for even a law of large numbers analysis.

An alternative and more useful representation follows from the form (4.1) and the continuity of $x \mapsto \theta(\cdot|x)$. Following our convention, we present only the representation needed to prove an LDP on $\mathscr{C}([0, 1] : \mathbb{R}^d)$, but the analogous representation holds with $[0, 1]$ replaced by any interval $[0, T]$, $T < \infty$. The line of argument used to prove the LDP will adapt the arguments used for Sanov's theorem and Cramér's theorem to this functional setting. However, obtaining "process-level" information requires a more complicated empirical measure than the one used for Sanov's theorem. Define L^n by

$$L^n(A \times B) \doteq \int_B L^n(A|t)dt, \quad L^n(A|t) \doteq \delta_{v_i(X_i^n)}(A) \text{ if } t \in [i/n, i/n + 1/n)$$

(4.3)

for Borel sets $A \subset \mathbb{R}^d$ and $B \subset [0, 1]$. This measure and controlled analogues to be introduced below record the joint empirical distribution of velocity and time. Owing to conflicting but standard usage, in this chapter L is used for both an empirical measure (as defined above) and the local rate function. The intended use should always be clear, since the former appears only as L^n, and the latter as L.

The following construction identifies quantities that will appear in the representation as well as others to be used in the convergence analysis. As first discussed in Sect. 3.1, we can consider $[\mu^n]_{i|0,\ldots,i-1}(dv_i|\bar{v}_0^n, \ldots, \bar{v}_{i-1}^n)$ to be simply a random measure on $\mathbb{R}^d$ that is measurable with respect to $\sigma(\bar{v}_j^n, j = 0, \ldots, i - 1) = \sigma(\bar{X}_j^n, j = 1, \ldots, i)$, and this ω-dependent measure is denoted by $\bar{\mu}_i^n(dv_i)$. Also as in Sect. 3.1, for notational convenience we assume that the original processes as well as controlled analogues are all defined on the same probability space. Note that the role of the "driving noises" played by X_i in Sect. 3.1 is here played by $v_i(X_i^n)$. The measure μ^n picks new distributions for these driving noises, as reflected by the notation. Another minor notational difference is that the noise index is from 0 to $n - 1$ rather than 1 to n.

Construction 4.4 *Suppose we are given a probability measure $\mu^n \in \mathscr{P}((\mathbb{R}^d)^n)$ and decompose it in terms of conditional distributions $[\mu^n]_{i|1,\ldots,i-1}$ on the ith variable given variables 0 through $i - 1$:*

$$\mu^n(dv_0 \times \cdots \times dv_{n-1}) = [\mu^n]_0(dv_0)[\mu^n]_{1|0}(dv_1|v_0)$$
$$\times \cdots \times [\mu^n]_{n-1|1,\dots,n-2}(dv_{n-1}|v_0,\dots,v_{n-2}).$$

Let $\{\bar{v}_i^n\}_{i=0,\dots,n-1}$ be random variables defined on a probability space $(\Omega, \mathscr{F}, P)$ and with joint distribution μ^n. Thus conditioned on $\bar{\mathscr{F}}_i^n \doteq \sigma(\bar{v}_j^n, j = 0, \dots, i - 1)$, $\bar{v}_i^n$ has distribution $\bar{\mu}_i^n(dv_i) \doteq [\mu^n]_{i|0,\dots,i-1}(dv_i|\bar{v}_0^n, \dots, \bar{v}_{i-1}^n)$. The collection $\{\bar{\mu}_i^n\}_{i=0,\dots,n-1}$ will be called a control. Then controlled processes $\bar{X}^n$ and measures $\bar{L}^n$ are recursively constructed as follows. Let $\bar{X}_0^n = x_0$, and for $i = 1, \dots, n$ define $\bar{X}_i^n$ recursively by

$$\bar{X}_{i+1}^n = \bar{X}_i^n + \frac{1}{n}\bar{v}_i^n.$$

When $\{\bar{X}_i^n\}_{i=1,\dots,n}$ has been constructed, $\bar{X}^n(t)$ is defined as in (4.2) as piecewise linear interpolation, and

$$\bar{L}^n(A \times B) \doteq \int_B \bar{L}^n(A|t)dt, \quad \bar{L}^n(A|t) \doteq \delta_{\bar{v}_i^n}(A) \text{ if } t \in [i/n, i/n + 1/n).$$

We also define

$$\bar{\mu}^n(A \times B) \doteq \int_B \bar{\mu}^n(A|t)dt, \quad \bar{\mu}^n(A|t) \doteq \bar{\mu}_i^n(A) \text{ if } t \in [i/n, i/n + 1/n)$$

and

$$\lambda^n(A \times B) \doteq \int_B \lambda^n(A|t)dt, \quad \lambda^n(A|t) \doteq \theta(A|\bar{X}_i^n) \text{ if } t \in [i/n, i/n + 1/n).$$

The measures $\bar{\mu}^n(dx \times dt)$ record the time dependence of the $\bar{\mu}_i^n$. When taking limits, we will also want to keep track of the corresponding $\theta(\cdot|\bar{X}_i^n)$, since the two appear together in the relative entropy representation. This information is recorded in $\lambda^n \in \mathscr{P}(\mathbb{R}^d \times [0, 1])$. Note also that, as remarked previously, $\bar{\mathscr{F}}_i^n = \sigma(\bar{X}_j^n, j = 1, \dots, i)$.

Theorem 4.5 *Let $G : \mathscr{P}(\mathbb{R}^d \times [0, 1]) \to \mathbb{R}$ be bounded from below and measurable. Let L^n be defined as in (4.3), and given a control $\{\bar{\mu}_i^n\}$, let $\{\bar{X}_i^n\}$ and $\{\bar{L}^n\}$ be defined as in Construction 4.4. Then*

$$-\frac{1}{n}\log Ee^{-nG(L^n)} = \inf_{\{\bar{\mu}_i^n\}} E\left[G(\bar{L}^n) + \frac{1}{n}\sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot|\bar{X}_i^n)\right) \right].$$

Proof The representation follows directly from the high-level variational representation for exponential integrals [part (a) of Proposition 2.3] and the chain rule [Theorem 2.6], and the argument is almost the same as that used to derive the representation

used to prove Sanov's theorem [Proposition 3.1]. The only difference is that the base measure in that case was product measure, reflecting the iid noise structure. Here the base measure is

$$\theta(dv_0|x_0^n)\theta(dv_1|x_1^n) \times \cdots \times \theta(dv_{n-1}|x_{n-1}^n),$$

where

$$x_i^n = x_0 + \frac{1}{n}\sum_{j=0}^{i-1} v_j.$$

One applies the chain rule exactly as was done in Proposition 3.1. The change in the base measure is reflected by a change in the measures appearing in the relative entropy cost, i.e., $\theta(\cdot|\bar{X}_i^n)$ rather than $\theta(\cdot)$ as in the iid case. $\qquad\square$

Note that the definition of $\bar{L}^n$ allows us to write

$$\bar{X}^n(t) = \int_{\mathbb{R}^d \times [0,t]} y\bar{L}^n(dy \times ds) + x_0.$$

Thus a special case of the representation in Theorem 4.5 occurs for F that is a bounded and measurable map from $\mathscr{C}([0,1] : \mathbb{R}^d)$ to $\mathbb{R}$:

$$-\frac{1}{n}\log Ee^{-nF(X^n)} = \inf_{\{\bar{\mu}_i^n\}} E\left[F(\bar{X}^n) + \frac{1}{n}\sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot|\bar{X}_i^n)\right)\right]. \qquad (4.4)$$

This representation will be used in the proof of the LDP for $\{X^n\}$. As in passing from Sanov's theorem to Cramér's theorem, convergence of $\bar{L}^n$ plus some uniform integrability will imply convergence of $\bar{X}^n$.

Remark 4.6 Although the proof of the LDP requires only bounded F (and hence bounded G), we state Theorem 4.5 so as to allow its use in the analysis of importance sampling in Chap. 15, where unbounded functionals cannot be avoided.

4.3 Form of the Rate Function

Before going further, we pause to comment on the expected form of the rate function. We give a completely heuristic calculation, based on a time scale separation due to the $1/n$ scaling of the noise and the weak continuity of $x \to \theta(\cdot|x)$, which suggests the correct form of the rate function. Over an interval $[s, s + \delta]$, with $\delta > 0$ small and $1/\delta \in \mathbb{N}$, the noise terms in the definition of $X^n(s + \delta) - X^n(s)$ are approximately iid with distribution $\theta(\cdot|X^n(s))$. Therefore,

$$\frac{X^n(s+\delta) - X^n(s)}{\delta} \approx \frac{1}{n\delta} \sum_{i=\lfloor ns \rfloor}^{\lfloor ns+n\delta \rfloor} v_i(X^n(s)),$$

and by Cramér's theorem, the right-hand side satisfies an LDP with the rate function $\delta L(X^n(s), \beta)$, where

$$L(x, \beta) = \inf \left[R\left(\mu(\cdot) \, \| \, \theta(\cdot | x)\right) : \int_{\mathbb{R}^d} y\mu(dy) = \beta \right]. \tag{4.5}$$

Suppose that $\sigma > 0$ is small, and that in the following display, $B(y, \sigma)$ denotes a (context-dependent) open ball of radius σ. Using the Markov property to combine estimates over small intervals, for a smooth trajectory $\phi \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ that starts at x_0, we have

$$
\begin{aligned}
&P\left\{X^n \in B(\phi, \sigma)\right\} \\
&\quad \approx P\left\{X^n(j\delta) \in B(\phi(j\delta), \sigma) \text{ all } 1 \le j \le 1/\delta\right\} \\
&\quad \approx P\left\{\frac{X^n(j\delta + \delta) - X^n(j\delta)}{\delta} \in B\left(\frac{\phi(j\delta + \delta) - \phi(j\delta)}{\delta}, \frac{2\sigma}{\delta}\right), \ 0 \le j < \frac{1}{\delta}\right\} \\
&\quad \approx \prod_{j=0}^{\frac{1}{\delta}-1} \exp\left\{-n\delta L\left(\phi(j\delta), \frac{\phi(j\delta + \delta) - \phi(j\delta)}{\delta}\right)\right\} \\
&\quad \approx \exp\left\{-n \int_0^1 L(\phi(s), \dot{\phi}(s))ds\right\}.
\end{aligned}
$$

Therefore, one may expect the rate function $I(\phi) = \int_0^1 L(\phi(s), \dot{\phi}(s))ds$ for such ϕ. Owing to this interpretation, $\beta \mapsto L(x, \beta)$ is often called a **local rate function** in this context.

4.4   Statement of the LDP

We now turn to the rigorous analysis. As was the case in Chap. 3 with Sanov's theorem and small noise diffusions, we first establish tightness, and then prove a result that links the limits of weakly converging controls and controlled processes. With these results in hand, the Laplace principle is proved by establishing upper and lower bounds. The conditions we assume and some of the arguments are close to those used in [97]. However, the perspective is somewhat different, with the main argument being a functional version of the one used to obtain Cramér's theorem from Sanov's theorem, and we also set the arguments up so they can easily be adapted to the problems of importance sampling considered later in the book.

We show that Condition 4.3 by itself suffices for the Laplace principle and large deviation upper bound. For the lower bound we need additional conditions. Two types of conditions will be used, and are formulated as Conditions 4.7 and 4.8 below. The Laplace principle lower bound under Conditions 4.3 and 4.7 will be proved in Sect. 4.7, and under Conditions 4.3 and 4.8 it will be proved in Sect. 4.8. The convex hull of the support of $\mu \in \mathscr{P}(\mathbb{R}^d)$ is the smallest closed and convex set $A \subset \mathbb{R}^d$ such that $\mu(A) = 1$.

Condition 4.7 For each $x \in \mathbb{R}^d$, the convex hull of the support of $\theta(\cdot|x)$ is $\mathbb{R}^d$.

Condition 4.8 For every compact $K \subset \mathbb{R}^d$ and $\varepsilon \in (0, 1)$, there exist $\eta = \eta(K, \varepsilon) \in (0, 1)$ and $m = m(K, \varepsilon) \in (0, \infty)$, such that whenever $\xi, \chi \in K$ satisfy $\|\xi - \chi\| \leq \eta$, we can find for each $\gamma \in \mathbb{R}^d$ a $\beta \in \mathbb{R}^d$ such that

$$L(\xi, \beta) - L(\chi, \gamma) \leq \varepsilon(1 + L(\chi, \gamma)), \quad \|\beta - \gamma\| \leq m(1 + L(\chi, \gamma))\|\xi - \chi\|.$$

Condition 4.7 can be weakened to the requirement that the relative interior of the convex hull of the support of $\theta(\cdot|x)$ be independent of x and contain 0 (see Sect. 6.3 of [97]). Condition 4.8 is very important in that it allows the noise to push the process in only a subset of all possible directions. For example, if the model of Example 4.1 corresponds to a degenerate diffusion, which means that $\sigma(x)\sigma^T(x)$ is only positive semidefinite, then Condition 4.7 is not valid, but under the assumption that b and σ are Lipschitz continuous, Condition 4.8 holds. Under similar Lipschitz-type assumptions, Condition 4.8 is satisfied for a broad range of models, and we refer the reader to Sect. 6.3 in [97] for additional illustrative examples.

Recall that $\mathscr{AC}_{x_0}([0, T] : \mathbb{R}^d)$ denotes the subset of $\mathscr{C}([0, T] : \mathbb{R}^d)$ consisting of all absolutely continuous functions satisfying $\phi(0) = x_0$.

Theorem 4.9 *Assume Condition 4.3 and define X^n by (4.2) and $L : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ by (4.5). Let*

$$I(\phi) = \int_0^T L(\phi(s), \dot{\phi}(s))ds \quad \text{if } \phi \in \mathscr{AC}_{x_0}([0, T] : \mathbb{R}^d),$$

and in all other cases set $I(\phi) = \infty$. Then the following conclusions hold.

(a) *I is a rate function and $\{X^n\}_{n\in\mathbb{N}}$ satisfies the Laplace principle upper bound with rate function I.*

(b) *Suppose that in addition, either Condition 4.7 or Condition 4.8 holds. Then $\{X^n\}_{n\in\mathbb{N}}$ satisfies the Laplace principle with rate function I.*

Remark 4.10 In the proofs to follow, the initial condition is fixed at x_0. However, the arguments apply with only notational changes if instead we consider a sequence of initial conditions $\{x_0^n\}_{n\in\mathbb{N}}$ with $x_0^n \to x_0$, and establish

$$\frac{1}{n} \log E_{x_0^n} e^{-nF(X^n)} + \inf_{\phi:\phi(0)=x_0^n} [F(\phi) + I(\phi)] \to 0. \tag{4.6}$$

Using an elementary argument by contradiction, this implies that the Laplace and large deviation principles hold uniformly for initial conditions in compact sets, as defined in Chap. 1. To be specific, if the uniform Laplace principle is not valid, then there exists a compact set $K \subset \mathbb{R}^d$, $\delta > 0$, and for each $n \in \mathbb{N}$, an initial condition $x_0^n \in K$ such that

$$\left| \frac{1}{n} \log E_{x_0^n} e^{-nF(X^n)} + \inf_{\phi:\phi(0)=x_0^n} [F(\phi) + I(\phi)] \right| \geq \delta. \tag{4.7}$$

However, since K is compact, there exist a subsequence $x_0^{n_k}$ and $x_0 \in K$ such that $x_0^{n_k} \to x_0$. Then (4.6) contradicts (4.7), and thus the uniform Laplace principle holds.

The rest of the chapter is organized as follows. In Sect. 4.5 we prove part (a) of Theorem 4.9. In preparation for the (two) proofs of the Laplace lower bound, Sect. 4.6 studies some basic properties of the function $L(x, \beta)$. The last two sections of the chapter, Sects. 4.7 and 4.8, contain the proof of the lower bound under Condition 4.7 and Condition 4.8, respectively. Throughout the chapter we assume Condition 4.3, and to simplify notation, proofs are given for $T = 1$.

4.5   Laplace Upper Bound

We begin with preliminary results on tightness and uniform integrability of the controlled processes from Sect. 4.2.

4.5.1   Tightness and Uniform Integrability

Lemma 4.11 *Assume Condition 4.3 and consider any sequence of controls $\{\bar{\mu}_i^n\}$ for which the relative entropy costs satisfy*

$$\sup_{n \in \mathbb{N}} E \left[\frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot | \bar{X}_i^n) \right) \right] \leq K < \infty.$$

Let $\{\bar{L}^n\}_{n \in \mathbb{N}}$, $\{\bar{X}^n\}_{n \in \mathbb{N}}$, $\{\bar{\mu}^n\}_{n \in \mathbb{N}}$, and $\{\lambda^n\}_{n \in \mathbb{N}}$ be defined as in Construction 4.4. Then the empirical measures $\{\bar{L}^n\}$ are tight and in fact uniformly integrable in the sense that

$$\lim_{M \to \infty} \limsup_{n \to \infty} E \left[\int_{\mathbb{R}^d \times [0,1]} \|y\| \, 1_{\{\|y\| \geq M\}} \bar{L}^n(dy \times dt) \right] = 0. \tag{4.8}$$

The measures $\{\bar{\mu}^n\}_{n \in \mathbb{N}}$ are also uniformly integrable in the sense of (4.8), and $\{\bar{X}^n\}$, $\{\bar{\mu}^n\}$, and $\{\lambda^n\}$ are all tight.

Proof Except for more complicated notation, the proof is almost the same as the analogous result needed for Cramér's theorem. From the inequality (2.9), it follows that if $\mu \in \mathscr{P}(\mathbb{R}^d \times [0, 1])$ satisfies $\mu \ll \lambda^n$, then for all $\sigma \in [1, \infty)$,

$$\int_{\mathbb{R}^d \times [0,1]} \|y\| \, 1_{\{\|y\| \geq M\}} \mu(dy \times dt)$$
$$\leq \int_{\mathbb{R}^d \times [0,1]} e^{\sigma\|y\|} 1_{\{\|y\| \geq M\}} \lambda^n(dy \times dt) + \frac{1}{\sigma} R\left(\mu \, \| \lambda^n\right).$$

By a conditioning argument it follows that $E \int f d\bar{L}^n = E \int f d\bar{\mu}^n$ for every bounded and measurable function f. Using the definitions of $\bar{\mu}^n$ and λ^n and the chain rule to get the first equality, we have

$$E\left[R\left(\bar{\mu}^n \, \| \lambda^n\right)\right] = E\left[\int_0^1 R\left(\bar{\mu}^n(\cdot|t) \, \| \lambda^n(\cdot|t)\right) dt\right] \tag{4.9}$$
$$= E\left[\frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \, \| \theta(\cdot|\bar{X}_i^n)\right)\right]$$
$$\leq K.$$

Therefore,

$$E\left[\int_{\mathbb{R}^d \times [0,1]} \|y\| \, 1_{\{\|y\| \geq M\}} \bar{L}^n(dy \times dt)\right]$$
$$= E\left[\int_{\mathbb{R}^d \times [0,1]} \|y\| \, 1_{\{\|y\| \geq M\}} \bar{\mu}^n(dy \times dt)\right]$$
$$\leq \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} e^{\sigma\|y\|} 1_{\{\|y\| \geq M\}} \theta(dy|x) + \frac{1}{\sigma} K. \tag{4.10}$$

From part (a) of Condition 4.3 it follows that for $\sigma \in \mathbb{R}$,

$$\sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} e^{2\sigma\|y\|} \theta(dy|x) < \infty \tag{4.11}$$

(for details see the analogous claim in the proof of Lemma 3.9). Since

$$\int_{\mathbb{R}^d} e^{\sigma\|y\|} 1_{\{\|y\| \geq M\}} \theta(dy|x) \leq e^{-\sigma M} \int_{\mathbb{R}^d} e^{2\sigma\|y\|} \theta(dy|x), \tag{4.12}$$

sending first $n \to \infty$, then $M \to \infty$, and finally $\sigma \to \infty$ in (4.10), the limit (4.8) holds for both $\{\bar{L}^n\}$ and $\{\bar{\mu}^n\}$. Tightness of $\{\bar{L}^n\}$ and $\{\bar{\mu}^n\}$ follows directly, and the tightness of $\{\lambda^n\}$ follows from part (a) of Condition 4.3.

To establish tightness of $\{\bar{X}^n\}$ we use the fact that

$$\bar{X}^n(t) = \int_{\mathbb{R}^d \times [0,t]} y\bar{L}^n(dy \times ds) + x_0. \tag{4.13}$$

Tightness will follow if given $\varepsilon > 0$ and $\eta > 0$, there is $\delta > 0$ such that

$$\limsup_{n \to \infty} P\left\{w^n(\delta) \geq \varepsilon\right\} \leq \eta, \tag{4.14}$$

where $w^n(\delta) \doteq \sup_{0 \leq s < t \leq 1 : t - s \leq \delta} \left\|\bar{X}^n(t) - \bar{X}^n(s)\right\|$. Using (4.8), choose $M < \infty$ such that

$$\limsup_{n \to \infty} E\left[\int_{\mathbb{R}^d \times [0,1]} \|y\|\, 1_{\{\|y\| \geq M\}} \bar{L}^n(dy \times dt)\right] \leq \frac{\varepsilon\eta}{2}.$$

Let $\delta \doteq (\varepsilon/2M) \wedge 1$. Then since $M\delta \leq \varepsilon/2$, we have

$$\sup_{0 \leq s < u \leq 1 : u - s \leq \delta} \int_{\mathbb{R}^d \times [s,u]} \|y\|\, 1_{\{\|y\| \leq M\}} \bar{L}^n(dy \times dt) \leq M\delta \leq \frac{\varepsilon}{2}.$$

Hence

$$P\left\{w^n(\delta) \geq \varepsilon\right\} \leq P\left\{\int_{\mathbb{R}^d \times [0,1]} \|y\|\, 1_{\{\|y\| \geq M\}} \bar{L}^n(dy \times dt) \geq \frac{\varepsilon}{2}\right\}$$
$$\leq \frac{2}{\varepsilon} E\left[\int_{\mathbb{R}^d \times [0,1]} \|y\|\, 1_{\{\|y\| \geq M\}} \bar{L}^n(dy \times dt)\right]$$
$$\leq \eta.$$

This proves (4.14), and tightness of $\{\bar{X}^n\}$ follows. $\square$

4.5.2 Weak Convergence

Lemma 4.11 proved tightness of $\{(\bar{L}^n, \bar{\mu}^n, \lambda^n, \bar{X}^n)\}_{n \in \mathbb{N}}$. The following lemma characterizes the weak limits of this collection.

Lemma 4.12 *Consider any sequence of controls $\{\bar{\mu}_i^n\}$ as in Construction 4.4 for which the relative entropy costs satisfy*

$$E\left[\frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot|\bar{X}_i^n)\right)\right] \leq K < \infty.$$

Let $\{(\bar{X}^n, \bar{L}^n, \bar{\mu}^n)\}$ denote a weakly converging subsequence, which for notational convenience we again label by n, with limit $(\bar{X}, \bar{L}, \bar{\mu})$. Then w.p.1, $\bar{L} = \bar{\mu}$, and

$\bar{\mu}(dy \times dt)$ can be decomposed as $\bar{\mu}(dy|t)dt$, where $\bar{\mu}(dy|t)$ is a stochastic kernel on $\mathbb{R}^d$ given $[0, 1]$, and w.p.1 for all $t \in [0, 1]$,

$$\bar{X}(t) = \int_{\mathbb{R}^d \times [0,t]} y\bar{\mu}(dy \times ds) + x_0 = \int_{\mathbb{R}^d \times [0,t]} y\bar{\mu}(dy|s)ds + x_0. \qquad (4.15)$$

In addition, λ^n converges weakly to a limit λ of the form

$$\lambda(A \times B) = \int_B \theta(A|\bar{X}(t))dt. \qquad (4.16)$$

Proof Recall that $\bar{\mu}_i^n$ picks the conditional distribution of $\bar{v}_i^n$. Hence a minor modification of the martingale argument used to prove the analogous result needed for Sanov's theorem (Lemma 3.5) can be used to show that $\bar{L} = \bar{\mu}$ w.p.1. The changes are mainly notational, and are needed, since in the present setting the measures must record time information. For completeness we give the details.

Now, $\mathbb{R}^d \times [0, 1]$ is a Polish space, and on such a space there exists a countable separating class of bounded uniformly continuous functions (see Appendix A). Thus to verify $\bar{L} = \bar{\mu}$ w.p.1, it suffices to show that for every bounded uniformly continuous f,

$$P\left\{\int_{\mathbb{R}^d \times [0,1]} f(v, t)\bar{L}(dv \times dt) = \int_{\mathbb{R}^d \times [0,1]} f(v, t)\bar{\mu}(dv \times dt)\right\} = 1. \qquad (4.17)$$

Define $K \doteq \|f\|_\infty$ and $\Delta_i^n \doteq f\left(\bar{v}_i^n, i/n\right) - \int_{\mathbb{R}^d} f(v, i/n)\bar{\mu}_i^n(dv)$. For all $\varepsilon > 0$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n f\left(\bar{v}_i^n, i/n\right) - \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} f(v, i/n)\bar{\mu}_i^n(dv)\right| > \varepsilon\right\}$$
$$\leq \frac{1}{\varepsilon^2}E\left[\frac{1}{n^2}\sum_{i,j=1}^n \Delta_i^n \Delta_j^n\right].$$

Recall that $\bar{\mathscr{F}}_i^n \doteq \sigma(\bar{v}_j^n, j = 0, \ldots, i-1)$. By a standard argument, for $i \neq j$, conditioning on $\bar{\mathscr{F}}_{j \wedge i - 1}^n$ gives $E[\Delta_i^n \Delta_j^n] = 0$. Since $|\Delta_i^n| \leq 2K$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n f(\bar{v}_i^n, i/n) - \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} f(v, i/n)\bar{\mu}_i^n(dv)\right| > \varepsilon\right\} \leq \frac{4K^2}{n\varepsilon^2}.$$

Let $\gamma(\delta)$ denote the modulus $\sup_{v \in \mathbb{R}^d, 0 \leq s \leq t \leq 1 : t - s \leq \delta}\{|f(v, t) - f(v, s)|\}$. Since f is uniformly continuous, $\gamma(\delta) \downarrow 0$ as $\delta \downarrow 0$, and the definition of $\gamma(\delta)$ implies

$$\left| \frac{1}{n} \sum_{i=1}^{n} f(\bar{v}_i^n, i/n) - \int_{\mathbb{R}^d \times [0,1]} f(v, t) \bar{L}^n(dv \times dt) \right| \leq \gamma(1/n)$$

$$\left| \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^d} f(v, i/n) \bar{\mu}_i^n(dv) - \int_{\mathbb{R}^d \times [0,1]} f(v, t) \bar{\mu}^n(dv \times dt) \right| \leq \gamma(1/n).$$

Letting first $n \to \infty$ and then $\varepsilon \to 0$, we obtain (4.17), which proves $\bar{L} = \bar{\mu}$ w.p.1.

Note that both $\bar{L}^n$ and $\bar{\mu}^n$ have second marginals equal to Lebesgue measure. Since this property is inherited by the weak limits, $\bar{L}(\mathbb{R}^d \times \{t\}) = 0$ w.p.1. This property and the uniform integrability allow us to pass to the limit in (4.13) and obtain

$$\bar{X}(t) = \int_{\mathbb{R}^d \times [0,t]} y \bar{L}(dy \times ds) + x_0.$$

Now use that $\bar{L} = \bar{\mu}$ w.p.1 to get the first part of (4.15). Since each $\bar{\mu}$ has Lebesgue measure as its second marginal, both the decomposition and the second part of (4.15) follow. Finally, the weak convergence of λ^n and the form of the limit follow from the weak convergence of $\bar{X}^n$ to $\bar{X}$ and the assumption that $x \to \theta(\cdot | x)$ is continuous in the weak topology. $\square$

4.5.3   Completion of the Laplace Upper Bound

The large deviation and Laplace principle upper bounds correspond to the variational lower bound. To prove such a lower bound, we again follow the line of argument used for Cramér's theorem in Sect. 3.1.6. Fix a continuous and bounded $F : \mathscr{C}([0, 1] : \mathbb{R}^d) \to \mathbb{R}$ and $\varepsilon > 0$. Using (4.4), let $\{\bar{\mu}_i^n\}_{i=1,\dots,n}$ satisfy

$$-\frac{1}{n} \log E e^{-nF(X^n)} + \varepsilon \geq E \left[F(\bar{X}^n) + \frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot | \bar{X}_i^n) \right) \right].$$

Then since F is bounded, we have $\sup_n \frac{1}{n} E \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot | \bar{X}_i^n) \right) < \infty$, and therefore by Lemma 4.11, it follows that $\{ (\bar{L}^n, \bar{X}^n, \bar{\mu}^n, \lambda^n) \}$ is tight. Consider any subsequence that converges to a weak limit $(\bar{L}, \bar{X}, \bar{\mu}, \lambda)$, and denote the convergent subsequence by n. If the lower bound is demonstrated for this subsequence, then the standard argument by contradiction establishes the lower bound for the original sequence. Details of the following calculation are given after the display:

$$\liminf_{n \to \infty} -\frac{1}{n} \log E \exp\{-nF(X^n)\} + \varepsilon$$

$$\geq \liminf_{n\to\infty} E\left[F(\bar{X}^n) + \frac{1}{n}\sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot)\,\big\|\,\theta(\cdot|\bar{X}_i^n)\right) \right]$$

$$= \liminf_{n\to\infty} E\left[F(\bar{X}^n) + R\left(\bar{\mu}^n(dy\times dt)\,\big\|\,\lambda^n(dy\times dt)\right) \right]$$

$$\geq E\left[F\left(\bar{X}\right) + R\left(\bar{\mu}(dy\times dt)\,\big\|\,\lambda(dy\times dt)\right) \right]$$

$$= E\left[F\left(\bar{X}\right) + \int_{[0,1]} R\left(\bar{\mu}(\cdot|t)\,\big\|\,\theta(\cdot|\bar{X}(t))\right) dt \right]$$

$$\geq E\left[F(\bar{X}) + \int_{[0,1]} L(\bar{X}(t),\dot{\bar{X}}(t)) dt \right]$$

$$\geq \inf_{\phi}\left[F(\phi) + \int_{[0,1]} L(\phi(t),\dot{\phi}(t)) dt \right]. \tag{4.18}$$

The first equality uses the rewriting of the relative entropy in (4.9); the next inequality is due to the weak convergence, the lower semicontinuity of $R\left(\cdot\|\cdot\right)$, continuity of F, and Fatou's lemma; the next equality uses the decompositions $\bar{\mu}(dy\times dt) = \bar{\mu}(dy|t)dt$ and $\lambda(dy\times dt) = \theta(dy|\bar{X}(t))dt$ and the chain rule; the third inequality follows from (4.5) and (4.15); and the infimum in the last line is over all $\phi\in\mathscr{AC}_{x_0}([0,1]:\mathbb{R}^d)$. Since $\varepsilon > 0$ is arbitrary, we have proved the Laplace upper bound for $\{X^n\}$:

$$\limsup_{n\to\infty} \frac{1}{n}\log E \exp\{-nF(X^n)\} \leq -\inf_{\phi\in\mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I(\phi)].$$

$\square$

4.5.4 I is a Rate Function

As first noted in Chap. 3, in the weak convergence approach, a deterministic version of the argument used to prove the Laplace upper bound will usually show that the proposed rate function is indeed a rate function, i.e., that it has compact level sets.

Theorem 4.13 *Assume Condition 4.3, define $L(x,\beta)$ by (4.5), and let I be the function defined in Theorem 4.9. Then I has compact level sets in $\mathscr{C}([0,T]:\mathbb{R}^d)$.*

Proof As usual, the proof is given for $T = 1$. Suppose $\{\phi_j\}_{j\in\mathbb{N}}$ is given such that $I(\phi_j) \leq K < \infty$ for all $j\in\mathbb{N}$. Then we need to show that $\{\phi_j\}$ is precompact, and that if $\phi_j \to \phi$, then

$$\liminf_{j\to\infty} I(\phi_j) \geq I(\phi).$$

Since $I(\phi_j) < \infty$, we know ϕ_j is absolutely continuous. Define probability measures μ^j on $\mathbb{R}^d\times[0,1]$ by

$$\mu^j(A \times B) = \int_B \delta_{\dot{\phi}_j(t)}(A)dt, \ A \in \mathscr{B}(\mathbb{R}^d), B \in \mathscr{B}([0, 1]).$$

Note that

$$\phi_j(t) = x_0 + \int_{\mathbb{R}^d \times [0,t]} y\mu^j(dy \times ds).$$

Using $I(\phi_j) \leq K < \infty$, exactly the same argument as in Lemma 4.11 shows that $\{\mu^j\}_{j \in \mathbb{N}}$ is tight and uniformly integrable. By the usual subsequential argument, we can assume that μ^j converges along the full sequence, and a deterministic version of Lemma 4.12 shows that the limit μ can be factored in the form $\mu(dy \times dt) = \mu(dy|t)dt$, and that $\phi_j \to \phi$, with $\int_{\mathbb{R}^d} y\mu(dy|t) = \dot{\phi}(t)$. Thus $\{\phi_j\}$ is precompact. We now argue that $I(\phi) \leq K$. In Lemma 4.14 it will be shown that L is a lower semicontinuous function that is convex in the second variable. Using these properties, we obtain

$$K \geq \liminf_{j \to \infty} I(\phi_j)$$

$$= \liminf_{j \to \infty} \int_{\mathbb{R}^d \times [0,1]} L(\phi_j(t), y)\mu^j(dy \times dt)$$

$$\geq \int_{\mathbb{R}^d \times [0,1]} L(\phi(t), y)\mu(dy \times dt)$$

$$= \int_0^1 \int_{\mathbb{R}^d} L(\phi(t), y)\mu(dy|t)dt$$

$$\geq \int_0^1 L(\phi(t), \dot{\phi}(t))dt = I(\phi),$$

where the second inequality is a consequence of Fatou's lemma, the lower semicontinuity of L, and the convergence of (ϕ_j, μ^j) to (ϕ, μ), while the third inequality uses the convexity of L and Jensen's inequality. Thus I has compact level sets, and hence is a rate function. □

4.6 Properties of $L(x, \beta)$

To prove a Laplace lower bound, we must take a trajectory ϕ that nearly minimizes in $\inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I(\phi)]$ and show how to construct a control that can be applied in the representation that will give asymptotically the same cost. For the continuous-time models in Chap. 3 this was not very difficult, in part because the implementation of the control was straightforward. For example, in the case of the diffusion model, the construction of a solution to (3.17) is possible when v is measurable in t and has appropriate integrability properties; in particular, piecewise continuity or some similar form of regularity is not required. The situation is different in discrete time.

In general, $I(\phi) < \infty$ implies only that ϕ is absolutely continuous. As we will see, it is natural to define controls for the prelimit in terms of $\dot{\phi}(t)$, where ϕ is nearly minimizing. Since the derivative is well defined only up to a set of Lebesgue measure zero, this causes a number of problems. The solution is to show that one can always construct a "nice" nearly minimizing trajectory, e.g., one whose derivative is continuous from the left with right-hand limits. Such a construction requires some regularity properties of $L(x, \beta)$, which we now present.

Recall that for $x \in \mathbb{R}^d$, the Legendre–Fenchel transform of $H(x, \cdot)$ is defined by

$$H^*(x, \beta) = \sup_{\alpha \in \mathbb{R}^d} \left[\langle \alpha, \beta \rangle - H(x, \alpha) \right], \ \beta \in \mathbb{R}^d,$$

and $L(x, \beta)$ is defined as in (4.5). As noted in Remark 3.10 and shown in Lemma 4.16, for each fixed x these are dual representations of the same function.

Lemma 4.14 *Assume Condition 4.3. Then the following are valid.*

(a) *For each $x \in \mathbb{R}^d$, $\alpha \mapsto H(x, \alpha)$ is a finite convex function on $\mathbb{R}^d$ that is differentiable for all $\alpha \in \mathbb{R}^d$. Also, $(x, \alpha) \mapsto H(x, \alpha)$ is continuous on $\mathbb{R}^d \times \mathbb{R}^d$.*

(b) *For each $x \in \mathbb{R}^d$, $\beta \mapsto H^*(x, \beta)$ is a convex function on $\mathbb{R}^d$. Furthermore, $(x, \beta) \mapsto H^*(x, \beta)$ is a nonnegative lower semicontinuous function on $\mathbb{R}^d \times \mathbb{R}^d$.*

(c) *The map $\beta \mapsto H^*(x, \beta)$ is superlinear, uniformly in x, which means that*

$$\lim_{N \to \infty} \inf_{x \in \mathbb{R}^d} \inf_{\beta \in \mathbb{R}^d : \|\beta\| = N} \frac{H^*(x, \beta)}{\|\beta\|} = \infty.$$

(d) *The map $(x, \beta) \mapsto L(x, \beta)$ is a lower semicontinuous function on $\mathbb{R}^d \times \mathbb{R}^d$, and for each $x \in \mathbb{R}^d$, the map $\beta \mapsto L(x, \beta)$ is convex.*

Proof (a) Part (a) of Condition 4.3 ensures that $H(x, \alpha) \in (-\infty, \infty)$ for all $(x, \alpha) \in \mathbb{R}^d \times \mathbb{R}^d$. The convexity of $\alpha \mapsto H(x, \alpha)$ then follows from Hölder's inequality: if $\alpha_1, \alpha_2 \in \mathbb{R}^d$ and $\rho \in [0, 1]$, then

$$\int_{\mathbb{R}^d} e^{\langle \rho \alpha_1 + (1-\rho)\alpha_2, y \rangle} \theta(dy|x) \leq \left(\int_{\mathbb{R}^d} e^{\langle \alpha_1, y \rangle} \theta(dy|x) \right)^{\rho} \left(\int_{\mathbb{R}^d} e^{\langle \alpha_2, y \rangle} \theta(dy|x) \right)^{1-\rho}.$$

Taking the logarithm of both sides demonstrates convexity. Under part (a) of Condition 4.3 one can easily construct an appropriate dominating function, and hence show that $\alpha \mapsto H(x, \alpha)$ is differentiable, where for all $(x, \alpha) \in \mathbb{R}^d \times \mathbb{R}^d$, the gradient $\nabla_\alpha H(x, \alpha)$ is given by

$$\nabla_\alpha H(x, \alpha) = \frac{\int_{\mathbb{R}^d} y e^{\langle \alpha, y \rangle} \theta(dy|x)}{\int_{\mathbb{R}^d} e^{\langle \alpha, y \rangle} \theta(dy|x)}. \tag{4.19}$$

To see the continuity of $(x, \alpha) \mapsto H(x, \alpha)$, let $(x_n, \alpha_n) \to (x, \alpha)$ in $\mathbb{R}^d \times \mathbb{R}^d$. We write

$$e^{H(x_n,\alpha_n)} - e^{H(x,\alpha)} = \int_{\mathbb{R}^d} \left[e^{\langle \alpha_n, y \rangle} - e^{\langle \alpha, y \rangle} \right] \theta(dy|x_n)$$
$$+ \left[\int_{\mathbb{R}^d} e^{\langle \alpha, y \rangle} \theta(dy|x_n) - \int_{\mathbb{R}^d} e^{\langle \alpha, y \rangle} \theta(dy|x) \right],$$

and recall the bounds (4.11) and (4.12). The second term on the right converges to zero by the Feller property assumed in Condition 4.3(b) and the uniform integrability implied by Condition 4.3(a). For every $M < \infty$, we have $\sup_{\|y\| \le M} |\exp\langle \alpha_n, y \rangle - \exp\langle \alpha, y \rangle| \to 0$ as $n \to \infty$, and by part (a) of Condition 4.3,

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}^d} \int_{\{\|y\| \ge M\}} e^{\langle \alpha_n, y \rangle} \theta(dy|x) \to 0$$

as $M \to \infty$. Hence the first term on the right converges to zero, which completes the proof of the continuity of $(x, \alpha) \mapsto H(x, \alpha)$.

(b) By duality for Legendre–Fenchel transforms [217, Theorem 23.5],

$$H(x, \alpha) = \sup_{\beta \in \mathbb{R}^d} \left[\langle \alpha, \beta \rangle - H^*(x, \beta) \right], \ \alpha \in \mathbb{R}^d.$$

Taking $\alpha = 0$ in the last display shows that $\inf_{\beta \in \mathbb{R}^d} H^*(x, \beta) = -H(x, 0) = 0$, and thus H^* is nonnegative. For each $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}^d$, the mapping $\beta \mapsto \langle \alpha, \beta \rangle - H(x, \alpha)$ is convex (in fact affine), and for each $\alpha \in \mathbb{R}^d$, the mapping $(x, \beta) \mapsto \langle \alpha, \beta \rangle - H(x, \alpha)$ is continuous. Recalling that the pointwise supremum of convex functions is convex and that the pointwise supremum of continuous functions is lower semicontinuous, we see that $H^*(x, \cdot)$ is convex on $\mathbb{R}^d$ and H^* is lower semicontinuous on $\mathbb{R}^d \times \mathbb{R}^d$.

(c) From part (a) of Condition 4.3, for every $M < \infty$, we have

$$C_M \doteq \sup_{x \in \mathbb{R}^d} \sup_{\alpha \in \mathbb{R}^d : \|\alpha\| = M} H(x, \alpha) < \infty.$$

Also, for every $\beta \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$, the definition of H^* implies

$$H^*(x, \beta) \ge \langle M\beta/\|\beta\|, \beta \rangle - H(x, M\beta/\|\beta\|) \ge M\|\beta\| - C_M.$$

Thus

$$\inf_{x \in \mathbb{R}^d} \inf_{\beta \in \mathbb{R}^d : \|\beta\| = N} \frac{H^*(x, \beta)}{\|\beta\|} \ge M - \frac{C_M}{N}.$$

The asserted superlinearity follows by sending first $N \to \infty$ and then $M \to \infty$ in the last display.

(d) The claimed properties of $L(x, \beta)$ follow from its definition in (4.5) and the corresponding properties of $R(\cdot \| \cdot)$ [part (b) of Lemma 2.4]. We first consider convexity. Fix x, let $\beta_1, \beta_2 \in \mathbb{R}^d$, $\delta > 0$, $\rho \in [0, 1]$ be given, and suppose that μ_i

are within δ of the infimum in (4.5) for β_i, $i = 1, 2$. Then since the mean under $\rho\mu_1 + (1 - \rho)\mu_2$ is $\rho\beta_1 + (1 - \rho)\beta_2$, we have

$$
\begin{aligned}
L(x, \rho\beta_1 + (1 - \rho)\beta_2) &\leq R(\rho\mu_1(\cdot) + (1 - \rho)\mu_2(\cdot) \, \| \theta(\cdot|x)) \\
&\leq \rho R(\mu_1(\cdot) \, \| \theta(\cdot|x)) + (1 - \rho)R(\mu_2(\cdot) \, \| \theta(\cdot|x)) \\
&\leq \rho L(x, \beta_1) + (1 - \rho)L(x, \beta_2) + \delta.
\end{aligned}
$$

Convexity follows, since $\delta > 0$ is arbitrary. Next suppose that $x_j \to x$ and $\beta_j \to \beta$ as $j \to \infty$. By an argument by contradiction based on subsequences, we can assume without loss that $L(x_j, \beta_j)$ converges in $[0, \infty]$, and we need to prove that $L(x, \beta) \leq \lim_{j \to \infty} L(x_j, \beta_j)$. If the limit is ∞, there is nothing to prove, and hence we assume $L(x_j, \beta_j) \leq M < \infty$ for all j. Choose μ_j that is within $\delta > 0$ of the infimum in the definition of $L(x_j, \beta_j)$. Then by part (d) of Lemma 2.4, $\{\mu_j\}$ is tight and uniformly integrable. Thus if μ^* is the limit of any convergence subsequence, then the mean of μ^* is β, and so (along this subsequence)

$$
L(x, \beta) \leq R(\mu^*(\cdot) \, \| \theta(\cdot|x)) \leq \liminf_{j \to \infty} R(\mu_j(\cdot) \, \| \theta(\cdot|x_j)) \leq \liminf_{j \to \infty} L(x_j, \beta_j) + \delta.
$$

This establishes the lower semicontinuity. $\qquad\square$

Remark 4.15 For the next result we will assume, in addition to Condition 4.3, that the support of $\theta(\cdot|x)$ is all of $\mathbb{R}^d$. Part (d) of the lemma was mentioned in Remark 3.10, which noted that the rate function for Cramér's theorem (which plays the role of the local rate function here) has two variational representations. These correspond here to H^* (as a supremum involving a moment-generating function) and L (as an infimum involving relative entropy). Although for our needs it suffices to prove this assuming Conditions 4.3 and 4.7, the functions H^* and L coincide when $\sup_{x \in \mathbb{R}^d} H(x, \alpha) < \infty$ for α in some open neighborhood of the origin. This will be proved in Lemma 5.4. Several statements in the lemma below hold assuming only Condition 4.3 (cf. [97, Lemma 6.2.3]). However, for simplicity we assume here that both Conditions 4.3 and 4.7 are satisfied.

Lemma 4.16 *Assume Conditions 4.3 and 4.7. Then the following conclusions hold.*
  (a) H^* *is finite on* $\mathbb{R}^d \times \mathbb{R}^d$.
  (b) *For every* $x \in \mathbb{R}^d$, $\alpha \mapsto H(x, \alpha)$ *is strictly convex on* $\mathbb{R}^d$.
  (c) *For every* $(x, \beta) \in \mathbb{R}^d \times \mathbb{R}^d$, *there is a unique* $\alpha = \alpha(x, \beta) \in \mathbb{R}^d$ *such that* $\nabla_\alpha H(x, \alpha(x, \beta)) = \beta$.
  (d) $H^* = L$.
  (e) *For every* $(x, \beta) \in \mathbb{R}^d \times \mathbb{R}^d$, *with* $\alpha(x, \beta)$ *as in part (c),*

$$
L(x, \beta) = \langle \alpha(x, \beta), \beta \rangle - H(x, \alpha(x, \beta)).
$$

  (f) $(x, \beta) \mapsto L(x, \beta)$ *is continuous on* $\mathbb{R}^d \times \mathbb{R}^d$.

(g) *There exists a measurable* $\alpha : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ *such that the stochastic kernel* $\gamma(dy|x, \beta)$ *on* $\mathbb{R}^d$ *given* $\mathbb{R}^d \times \mathbb{R}^d$ *defined by*

$$\gamma(dy|x, \beta) \doteq e^{\langle \alpha(x,\beta), y \rangle - H(x, \alpha(x,\beta))} \theta(dy|x),$$

satisfies

$$R(\gamma(\cdot|x, \beta)\|\theta(\cdot|x)) = L(x, \beta) \text{ and } \int_{\mathbb{R}^d} y \gamma(dy|x, \beta) = \beta \text{ for all } x \in \mathbb{R}^d, \beta \in \mathbb{R}^d.$$
(4.20)

Proof Since x plays no role other than as a parameter in parts (a)–(e), it is dropped from the notation in the proofs of these parts.

(a) Let Y be a random variable with distribution θ, and let Y_i denote the ith component. We first claim that the map $\alpha \to H(\alpha)$ is superlinear. By Condition 4.7, the support of θ is all of $\mathbb{R}^d$. Thus for each $M < \infty$,

$$\Theta_M \doteq \min_{\mathscr{I} \subset \{1,\dots,d\}} P \left\{ [\cap_{i \in \mathscr{I}} \{Y_i \geq M\}] \cap [\cap_{i \in \mathscr{I}^c} \{Y_i \leq -M\}] \right\} > 0.$$

Therefore, for every $\alpha \in \mathbb{R}^d$,

$$\frac{1}{\|\alpha\| + 1} \log E e^{\langle \alpha, Y \rangle} \geq \frac{1}{\|\alpha\| + 1} \log \left[\Theta_M e^{M \sum_{i=1}^d |\alpha_i|} \right] = M \frac{\sum_{i=1}^d |\alpha_i|}{\|\alpha\| + 1} + \frac{\log \Theta_M}{\|\alpha\| + 1}.$$

The superlinearity of H now follows by sending $\|\alpha\| \to \infty$ and then $M \to \infty$. The superlinearity implies that the Legendre–Fenchel transform H^* of H is finite everywhere, since for each $\beta \in \mathbb{R}^d$, one can find a compact set $K \subset \mathbb{R}^d$ such that

$$H^*(\beta) = \sup_{\alpha \in \mathbb{R}^d} [\langle \alpha, \beta \rangle - H(\alpha)] = \sup_{\alpha \in K} [\langle \alpha, \beta \rangle - H(\alpha)].$$

Since H is continuous, the last expression is finite, and thus (a) follows.

(b) As shown in Lemma 4.14, H is convex. Suppose that for some $\alpha_1, \alpha_2 \in \mathbb{R}^d$, $\alpha_1 \neq \alpha_2$, and $\rho \in (0, 1)$, we have

$$H(\rho\alpha_1 + (1 - \rho)\alpha_2) = \rho H(\alpha_1) + (1 - \rho)H(\alpha_2).$$

Then the condition for equality in Hölder's inequality requires that for $\theta(dy)$ a.e. y,

$$\frac{\exp\langle \alpha_1, y \rangle}{\int_{\mathbb{R}^d} \exp\langle \alpha_1, z \rangle \theta(dz)} = \frac{\exp\langle \alpha_2, y \rangle}{\int_{\mathbb{R}^d} \exp\langle \alpha_2, z \rangle \theta(dz)},$$

which implies that

$$\langle \alpha_1 - \alpha_2, y \rangle = H(\alpha_1) - H(\alpha_2) \text{ a.s. } \theta(dy).$$

In other words, θ is supported on a hyperplane of dimension $d - 1$. But this contradicts the fact that the support of $\theta(dy)$ is all of $\mathbb{R}^d$, which proves the strict convexity of $\alpha \mapsto H(\alpha)$.

(c) From Corollary 26.4.1 in [217] it now follows that the gradient $\nabla_\alpha H(\alpha)$ is onto $\mathbb{R}^d$. Thus given β, there exists a vector $\alpha(\beta)$ such that $\nabla_\alpha H(\alpha(\beta)) = \beta$. We claim that $\alpha(\beta)$ is unique. Suppose $\alpha_1 \neq \alpha_2$ are such that

$$\nabla_\alpha H(\alpha_1) = \nabla_\alpha H(\alpha_2) = \beta. \tag{4.21}$$

Define $\zeta : \mathbb{R} \to \mathbb{R}$ by $\zeta(\lambda) \doteq H(\alpha_1 + \lambda(\alpha_2 - \alpha_1))$, $\lambda \in \mathbb{R}$. From part (b), ζ is strictly convex on $\mathbb{R}$, and so

$$\zeta'(0) = \langle \nabla_\alpha H(\alpha_1), \alpha_2 - \alpha_1 \rangle < \zeta'(1) = \langle \nabla_\alpha H(\alpha_2), \alpha_2 - \alpha_1 \rangle.$$

This contradicts (4.21), and thus there is only one α that satisfies $\nabla_\alpha H(\alpha) = \beta$.

(d,e) Setting $\gamma(dy) = e^{\langle \alpha(\beta), y \rangle} \theta(dy) / e^{H(\alpha(\beta))}$, we have from (4.19),

$$\int_{\mathbb{R}^d} y\gamma(dy) = \frac{1}{e^{H(\alpha(\beta))}} \int_{\mathbb{R}^d} y e^{\langle \alpha(\beta), y \rangle} \theta(dy) = \nabla_\alpha H(\alpha(\beta)) = \beta. \tag{4.22}$$

A direct calculation using the form of $\gamma(dy)$, the definition of relative entropy, and the definition of L in (4.5) then gives

$$L(\beta) \leq R(\gamma \| \theta) = \langle \alpha(\beta), \beta \rangle - H(\alpha(\beta)) \leq H^*(\beta). \tag{4.23}$$

Using the definition of H and part (c) of Proposition 2.3, we have

$$H(\alpha) = \sup_{\mu \in \mathscr{P}(\mathbb{R}^d) : R(\mu \| \theta) < \infty} \left[\int_{\mathbb{R}^d} \langle \alpha, y \rangle \mu(dy) - R(\mu \| \theta) \right].$$

Therefore, for all $\alpha \in \mathbb{R}^d$ and $\mu \in \mathscr{P}(\mathbb{R}^d)$,

$$R(\mu \| \theta) \geq \left\langle \alpha, \int_{\mathbb{R}^d} y\mu(dy) \right\rangle - H(\alpha),$$

and consequently

$$L(\beta) \geq \langle \alpha, \beta \rangle - H(\alpha).$$

Since $\alpha \in \mathbb{R}^d$ is arbitrary, we have $L(\beta) \geq H^*(\beta)$. By (4.23), the reverse inequality holds, which shows that $L(\beta) = H^*(\beta)$ and also proves part (e).

(f) For the last two parts of the lemma we include the x dependence. From Lemma 4.14, $(x, \alpha) \mapsto H(x, \alpha)$ is continuous. We now show that joint continuity of $L(x, \beta)$ follows from this. If a sequence of differentiable convex functions g_i with Legendre transforms g_i^* converges pointwise to another differentiable convex function g with transform g^*, and if β is any point such that $g^*(\beta) < \infty$, then when-

ever $\beta_i \to \beta$, we have $g_i^*(\beta_i) \to g^*(\beta)$ [97, Lemma C.8.1]. We apply this result with $g_i(\alpha) = H(x_i, \alpha)$ and $g(\alpha) = H(x, \alpha)$ to conclude that if $x_i \to x$ and $\beta_i \to \beta$, then $L(x_i, \beta_i) \to L(x, \beta)$.

(g) To see the measurability of $(x, \beta) \mapsto \alpha(x, \beta)$, note that for each x, the strict convexity of $\alpha \mapsto H(x, \alpha)$ and the fact $L(x, \beta) = H^*(x, \beta) < \infty$ imply that $\beta \mapsto L(x, \beta)$ is differentiable for all $\beta \in \mathbb{R}^d$ [217, Theorem 26.3]. The characterization of $H(x, \alpha)$ as the Legendre–Fenchel transform of $L(x, \beta)$ then gives $\alpha(x, \beta) = \nabla_\beta L(x, \beta)$, from which measurability follows.

The second equality in (4.20) follows from (4.22), while the first equality follows on noting that the first inequality in (4.23) was shown to be an equality.   □

4.7 Laplace Lower Bound Under Condition 4.7

In this section we prove the Laplace principle lower bound under Conditions 4.3 and 4.7. For the proof of this lower bound we construct a nearly optimal trajectory ϕ^* for $\inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I(\phi)]$ that has a simple form. Based on ϕ^*, a control is constructed for use in the representation, so that the running cost is close to $I(\phi^*)$ and the associated controlled process converges to the nearly optimal trajectory ϕ^* as $n \to \infty$.

Fix $\varepsilon > 0$. Then there is $\zeta \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ such that

$$[F(\zeta) + I(\zeta)] \leq \inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I(\phi)] + \varepsilon. \tag{4.24}$$

While $\{\zeta(t) : 0 \leq t \leq 1\}$ is bounded by continuity, we also claim that without loss of generality, we can assume that

$$\{\dot{\zeta}(t) : 0 \leq t \leq 1\} \text{ is bounded.}$$

This claim will be established in Sect. 4.7.3.

Recall from part (f) of Lemma 4.16 that L is continuous. Let $M < \infty$ and $K < \infty$ be such that

$$\sup_{t \in [0,1]} \|\zeta(t)\| \vee \sup_{t \in [0,1]} \|\dot{\zeta}(t)\| \leq M, \qquad \sup_{\{(x,\beta):\|x\|\leq M+1, \|\beta\|\leq M+1\}} L(x, \beta) \leq K.$$

For $\delta > 0$, let ζ^δ be the piecewise linear interpolation of ζ, with interpolation points $t = k\delta$. Since ζ is absolutely continuous, $\dot{\zeta}^\delta(t)$ converges to $\dot{\zeta}(t)$ for a.e. $t \in [0, 1]$. Also, since $\sup_{t \in [0,1]} \|\zeta^\delta(t)\| \leq M$ and $\sup_{t \in [0,1]} \|\dot{\zeta}^\delta(t)\| \leq M$, the continuity of L and the dominated convergence theorem imply that there is $\delta > 0$ such that $\left[F(\zeta^\delta) + I(\zeta^\delta)\right] \leq [F(\zeta) + I(\zeta)] + \varepsilon$. We set $\phi^* = \zeta^\delta$ for such a δ.

4.7.1 Construction of a Nearly Optimal Control

The construction of a control to apply in the representation is now straightforward. Let $\gamma(dy|x, \beta)$ be as in part (g) of Lemma 4.16. Recall that $\bar{\mu}_i^n$ is allowed to be any measurable function of $\bar{X}_j^n$, $j = 0, \ldots, i$. Define $N^n \doteq \inf\{j : \|\bar{X}_j^n - \phi^*(j/n)\| > 1\} \wedge n$. Then we set

$$\bar{\mu}_i^n(\cdot) = \begin{cases} \gamma(\cdot|\bar{X}_i^n, \dot{\phi}^*(i/n)) & \text{if } i < N^n, \\ \theta(\cdot|\bar{X}_i^n) & \text{if } i \geq N^n. \end{cases} \tag{4.25}$$

The cost under this control satisfies

$$E\left[\frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \, \big\| \, \theta(\cdot|\bar{X}_i^n)\right)\right] = E\left[\frac{1}{n} \sum_{i=0}^{N^n-1} L(\bar{X}_i^n, \dot{\phi}^*(i/n))\right] \leq K,$$

and therefore Lemma 4.11 applies. Since $\tau^n \doteq N^n/n$ takes values in a compact set, given any subsequence of $\mathbb{N}$ we can find a further subsequence (again denoted by n) such that $(\bar{X}^n, \bar{\mu}^n, \tau^n)$ converges in distribution to a limit $(\bar{X}, \bar{\mu}, \tau)$. Also, it follows from the fact that the mean of $\bar{\mu}_i^n$ is $\dot{\phi}^*(i/n)$ for $i < N^n$ that

$$x_0 + \int_{\mathbb{R}^d \times [0, t \wedge \tau_n]} y\bar{\mu}^n(dy \times ds) = x_0 + \frac{1}{n} \sum_{i=0}^{\lfloor nt \rfloor \wedge N^n - 1} \dot{\phi}^*(i/n) + O(1/n)$$

$$= \phi^*(t \wedge \tau^n) + O(1/n).$$

Using the uniform integrability (4.8), we can send $n \to \infty$ and obtain

$$x_0 + \int_{\mathbb{R}^d \times [0, t]} y\bar{\mu}(dy|s)ds = \phi^*(t)$$

for all $t \in [0, \tau]$. It then follows from Lemma 4.12 [see (4.15)] that $\bar{X}(t) = \phi^*(t)$ for all $t \in [0, \tau]$, w.p.1. However, since $\|\bar{X}^n(\tau^n) - \phi^*(\tau^n)\|$ converges in distribution to $\|\bar{X}(\tau) - \phi^*(\tau)\|$, the definition of N^n implies that on the set $\tau < 1$, we have $\|\bar{X}(\tau) - \phi^*(\tau)\| \geq 1$. Thus $P(\tau < 1) = 0$, and so $\bar{X}(t) = \phi^*(t)$ for $t \in [0, 1]$. We conclude that along the full sequence $\mathbb{N}$, $\bar{X}^n$ converges in distribution to ϕ^*.

4.7.2 Completion of the Proof of the Laplace Lower Bound

We now put the pieces together to prove the Laplace lower bound. For the particular control $\{\bar{\mu}_i^n\}$ just constructed, we have

$$\limsup_{n \to \infty} -\frac{1}{n} \log E \exp\{-nF(X^n)\}$$

$$\leq \limsup_{n \to \infty} E\left[F(\bar{X}^n) + \frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(\cdot) \,\big\|\, \theta(\cdot | \bar{X}_i^n)\right)\right]$$

$$= \limsup_{n \to \infty} E\left[F(\bar{X}^n) + \frac{1}{n} \sum_{i=0}^{N^n-1} L(\bar{X}_i^n, \dot{\phi}^*(i/n))\right]$$

$$= \left[F(\phi^*) + \int_0^1 L(\phi^*(t), \dot{\phi}^*(t))dt\right]$$

$$\leq \inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I(\phi)] + 2\varepsilon.$$

The first inequality follows since the representation considers the infimum over all controls, and the first equality is due to the definition of $\bar{\mu}_i^n$ in (4.25). The second equality follows from the weak convergence $\bar{X}^n \Rightarrow \phi^*$ [Lemma 4.12], the uniform bound on $L(\bar{X}_i^n, \dot{\phi}^*(i/n))$ for $i \leq N^n - 1$, and the dominated convergence theorem. The last inequality uses the fact that ϕ^* as constructed is within 2ε of the infimum. Since $\varepsilon > 0$ is arbitrary, the Laplace lower bound

$$\liminf_{n \to \infty} \frac{1}{n} \log E \exp\{-nF(X^n)\} \geq - \inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I(\phi)]$$

follows.   □

4.7.3 Approximation by Bounded Velocity Paths

We now prove the claim that for ζ satisfying (4.24), $\dot{\zeta}$ can also be assumed bounded.

Lemma 4.17 *Consider* $\zeta \in \mathscr{C}([0,1] : \mathbb{R}^d)$ *such that* $[F(\zeta) + I(\zeta)] < \infty$. *Then given* $\varepsilon > 0$, *there is* ζ^* *such that* $\{\dot{\zeta}^*(t) : 0 \leq t \leq 1\}$ *is bounded and*

$$\sup_{0 \leq t \leq 1} \|\zeta(t) - \zeta^*(t)\| \leq \varepsilon, \quad I(\zeta^*) \leq I(\zeta) + \varepsilon.$$

Proof Since F is bounded, $I(\zeta) < \infty$. For $\lambda \in (0, 1)$ let $D_\lambda \doteq \{t : \|\dot{\zeta}(t)\| \geq 1/\lambda\}$, and define a time rescaling $S_\lambda : [0, 1] \to [0, \infty)$ by $S_\lambda(0) = 0$ and

$$\dot{S}_\lambda(t) = \begin{cases} \|\dot{\zeta}(t)\|/(1-\lambda), & t \in D_\lambda, \\ 1 & \text{otherwise.} \end{cases}$$

Then $S_\lambda(t)$ is continuous and strictly increasing. Let T_λ be the inverse of S_λ, which means that T_λ satisfies $T_\lambda(S_\lambda(t)) = t$ for all $t \in [0, 1]$ and $S_\lambda(T_\lambda(t)) = t$ for all $t \in [0, S_\lambda(1)] \supset [0, 1]$. Also define ζ_λ on $[0, S_\lambda(1)]$ by

$$\zeta_\lambda(t) \doteq \zeta(T_\lambda(t)),$$

which is a "slowed" version of ζ. By the ordinary chain rule, $\dot{\zeta}_\lambda(S_\lambda(t)) = \dot{\zeta}(t)/\dot{S}_\lambda(t)$, and therefore $\dot{\zeta}_\lambda(t)$ has uniformly bounded derivative for $t \in [0, 1]$. The ζ^* in the lemma will be ζ_λ for a positive λ.

From part (c) of Lemma 4.14 and part (d) of Lemma 4.16, $L(x, \beta)$ is uniformly superlinear in β: $L(x, \beta)/\|\beta\| \to \infty$ uniformly in x as $\|\beta\| \to \infty$. This property and $I(\zeta) < \infty$ imply $\int_0^1 \|\dot{\zeta}(t)\| dt < \infty$, and consequently

$$\lim_{\lambda \to 0} \int_0^1 1_{D_\lambda}(t) \|\dot{\zeta}(t)\| dt = 0. \tag{4.26}$$

It follows that $\lim_{\lambda \to 0} S_\lambda(s) = s$ uniformly for $s \in [0, 1]$. Since

$$\sup_{t \in [0,1]} \|\zeta_\lambda(t) - \zeta(t)\| = \sup_{t \in [0,1]} \|\zeta(T_\lambda(t)) - \zeta(t)\| = \sup_{s \in [0, T_\lambda(1)]} \|\zeta(s) - \zeta(S_\lambda(s))\|,$$

it follows that $\sup_{t \in [0,1]} \|\zeta_\lambda(t) - \zeta(t)\| \to 0$ as $\lambda \to 0$.

Thus we need only show that $I(\zeta_\lambda)$ is close to $I(\zeta)$. Let

$$\Gamma \doteq \sup_{t \in [0,1]} \sup_{\beta:\|\beta\| \leq 1} L(\zeta(t), \beta) < \infty.$$

For $t \in D_\lambda$, the nonnegativity of L implies

$$L\left(\zeta(t), \frac{\dot{\zeta}(t)}{\dot{S}_\lambda(t)}\right) \dot{S}_\lambda(t) - L(\zeta(t), \dot{\zeta}(t)) \leq L\left(\zeta(t), \frac{(1-\lambda)\dot{\zeta}(t)}{\|\dot{\zeta}(t)\|}\right) \frac{\|\dot{\zeta}(t)\|}{1-\lambda}$$

$$\leq \frac{\Gamma}{1-\lambda}\|\dot{\zeta}(t)\|,$$

and therefore

$$I(\zeta_\lambda) - I(\zeta) \leq \int_0^{S_\lambda(1)} L(\zeta_\lambda(t), \dot{\zeta}_\lambda(t)) dt - \int_0^1 L(\zeta(t), \dot{\zeta}(t)) dt$$

$$= \int_0^1 L(\zeta_\lambda(S_\lambda(t)), \dot{\zeta}_\lambda(S_\lambda(t))) \dot{S}_\lambda(t) dt - \int_0^1 L(\zeta(t), \dot{\zeta}(t)) dt$$

$$= \int_0^1 L\left(\zeta(t), \frac{\dot{\zeta}(t)}{\dot{S}_\lambda(t)}\right) \dot{S}_\lambda(t) dt - \int_0^1 L(\zeta(t), \dot{\zeta}(t)) dt$$

$$\leq \frac{\Gamma}{1-\lambda} \int_0^1 1_{D_\lambda}(t) \|\dot{\zeta}(t)\| dt.$$

From (4.26), the last expression converges to 0 as $\lambda \to 0$. Thus $\limsup_{\lambda \to 0} I(\zeta_\lambda) \leq I(\zeta)$ and $\zeta_\lambda \to \zeta$ as $\lambda \to 0$, and the claim is proved. $\qquad\square$

With the proof that we can assume that $\dot{\zeta}$ is bounded for ζ appearing in (4.24), the proof of the Laplace lower bound under Condition 4.7 is complete.

4.8 Laplace Lower Bound Under Condition 4.8

In this section we prove the Laplace principle lower bound without assuming the support condition on $\theta(\cdot|x)$. Instead, besides Condition 4.3, we use the Lipschitz-type assumption of Condition 4.8. The main difficulty in the proof for this case is that the construction of a piecewise linear nearly minimizing trajectory, which simplified the proof in Sect. 4.7, is not directly available here. The arguments of Sect. 4.7 relied on the finiteness and continuity of the function $(x, \beta) \mapsto L(x, \beta)$, properties that in general will not hold for the setting considered in this section.

In order to overcome this difficulty, we introduce a mollification that in a sense reduces the problem to the form studied in Sect. 4.7. The mollification introduces an error that needs to be carefully controlled. This is the main technical challenge in the proof. We will also make use of Lemma 1.10, which shows that the Laplace principle lower bound holds if and only if it holds for F that are Lipschitz continuous. Mollification techniques are often used in large deviation analysis, and are especially useful in proving lower bounds.

The section is organized as follows. We begin in Sect. 4.8.1 by introducing the mollification of the state dynamics. This takes the form of a small additive Gaussian perturbation, parametrized by $\sigma > 0$, to the noise sequence $\{v_i(X_i^n)\}$ in (4.1). We then estimate the asymptotics of $-\frac{1}{n} \log E \exp\{-nF(X^n)\}$ through an analogous expression when X^n is replaced by the perturbed state process Z_σ^n. Next in Sect. 4.8.2 we give a variational upper bound for functionals of the perturbed process in terms of a convenient family of controls. The limits of cost functions in this representation are given in terms of a perturbation L_σ of the function L introduced in (4.5). Section 4.8.3 studies properties of L_σ. In particular, we show that L_σ is a finite continuous function, is bounded above by L, and satisfies properties analogous to those assumed of L in Condition 4.8. Using these results, in Sect. 4.8.4 we construct a piecewise linear nearly optimal trajectory for $\inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\phi) + I_\sigma(\phi)]$, where I_σ is the rate function associated with the local rate function L_σ, which is then used to construct an asymptotically nearly optimal control sequence for the representation. Section 4.8.5 studies tightness and convergence properties of the associated controlled processes. Finally, Sect. 4.8.6 uses these convergence results and estimates from Sect. 4.8.1 to complete the proof of the variational upper bound. Throughout this section, F will be a real-valued bounded Lipschitz continuous function on $\mathscr{C}([0, 1] : \mathbb{R}^d)$.

4.8.1 Mollification

For $\sigma > 0$, let $\{w_{i,\sigma}\}_{i \in \mathbb{N}_0}$ be an iid sequence of Gaussian random variables with mean 0 and covariance σI that is independent of the random vector fields $\{v_i(\cdot)\}_{i \in \mathbb{N}_0}$.

For $n \in \mathbb{N}$ and $\sigma > 0$, consider along with the sequence $\{X_i^n\}_{i=0,\ldots,n}$, the sequence $\{U_{i,\sigma}^n\}_{i=0,\ldots,n}$ defined by

$$U_{i+1,\sigma}^n \doteq U_{i,\sigma}^n + \frac{1}{n}w_{i,\sigma}, \quad U_{0,\sigma}^n = 0.$$

Define the piecewise linear process $\{U_\sigma^n(t)\}_{t\in[0,1]}$ by

$$U_\sigma^n(t) \doteq U_{i,\sigma}^n + \left[U_{i+1,\sigma}^n - U_{i,\sigma}^n\right](nt - i), \quad t \in [i/n, i/n + 1/n].$$

Let $Z_\sigma^n = X^n + U_\sigma^n$, where X^n is as in (4.2). Note that Z_σ^n is the piecewise linear interpolation of the sequence $\{Z_{i,\sigma}^n\}$, where $Z_{i,\sigma}^n = X_i^n + U_{i,\sigma}^n$.

The following result shows that the Laplace lower bound properties of $\{X^n\}$ can be bounded in terms of the lower bound properties of $\{Z_\sigma^n\}$. For $\phi \in \mathscr{C}([0, 1] : \mathbb{R}^d)$, recall that $\|\phi\|_\infty \doteq \sup_{0 \leq t \leq 1} \|\phi(t)\|$.

Lemma 4.18 *For every $\sigma > 0$,*

$$\limsup_{n \to \infty} -\frac{1}{n} \log E e^{-nF(X^n)} \leq \limsup_{n \to \infty} -\frac{1}{n} \log E e^{-nF(Z_\sigma^n)} + \frac{M^2 \sigma^2}{2},$$

where

$$M \doteq \sup_{\phi, \eta \in \mathscr{C}([0,1]:\mathbb{R}^d), \phi \neq \eta} \frac{|F(\phi) - F(\eta)|}{\|\phi - \eta\|_\infty}.$$

Proof Let $B = 2\|F\|_\infty$. Then since

$$F(Z_\sigma^n) \geq F(X^n) - (M\|U_\sigma^n\|_\infty) \wedge B,$$

we see that

$$-\frac{1}{n} \log E e^{-nF(X^n)} \leq -\frac{1}{n} \log E e^{-nF(Z_\sigma^n)} + \frac{1}{n} \log E e^{n[(M\|U_\sigma^n\|_\infty) \wedge B]}. \qquad (4.27)$$

We now estimate the second term on the right side of (4.27) using the Laplace principle upper bound (which was proved in Sect. 4.5) with $\theta(\cdot|x) = \rho_\sigma(\cdot)$, where ρ_σ is the law of a d-dimensional normal random variable with mean 0 and covariance σI. Let

$$H_\sigma(\alpha) \doteq \log \int_{\mathbb{R}^d} \exp\langle \alpha, y\rangle \rho_\sigma(dy) = \frac{\sigma^2}{2}\|\alpha\|^2, \quad \alpha \in \mathbb{R}^d.$$

The Legendre–Fenchel transform of H_σ is given by

$$L_\sigma(\beta) \doteq \sup_{\alpha \in \mathbb{R}^d} [\langle \alpha, \beta\rangle - H_\sigma(\alpha)] = \frac{1}{2\sigma^2}\|\beta\|^2.$$

Then $\{U_\sigma^n\}_{n\in\mathbb{N}}$ satisfies the Laplace upper bound with rate function

$$I_{0,\sigma}(\varphi) \doteq \frac{1}{2\sigma^2}\int_0^1 \|\dot\varphi(s)\|^2 ds$$

if $\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)$ is absolutely continuous and $\varphi(0)=0$, and $I_{0,\sigma}(\varphi) \doteq \infty$ otherwise. This upper bound yields

$$\limsup_{n\to\infty}\frac{1}{n}\log E e^{n[(M\|U_\sigma^n\|_\infty)\wedge B]}$$

$$\leq -\inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)}\left[I_{0,\sigma}(\varphi)-(M\|\varphi\|_\infty)\wedge B\right]$$

$$\leq -\inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)}\left[I_{0,\sigma}(\varphi)-M\|\varphi\|_\infty\right]. \tag{4.28}$$

For all $\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)$ with $I_{0,\sigma}(\varphi) < \infty$, we have

$$\|\varphi\|_\infty^2 = \sup_{t\in[0,1]}\left\|\int_0^t \dot\varphi(s)ds\right\|^2 \leq \int_0^1 \|\dot\varphi(s)\|^2 ds,$$

and thus

$$\inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)}\left[I_{0,\sigma}(\varphi)-M\|\varphi\|_\infty\right] \geq \inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)}\left[\frac{\|\varphi\|_\infty^2}{2\sigma^2}-M\|\varphi\|_\infty\right]$$

$$= \inf_{r\geq 0}\left[\frac{r^2}{2\sigma^2}-Mr\right]$$

$$= -\frac{M^2\sigma^2}{2}.$$

The claim of the lemma now follows from the last display, (4.27), and (4.28). □

4.8.2 Variational Bound for the Mollified Process

In this section we present a variational bound for $Ee^{-nF(Z_\sigma^n)}$. The basic idea is to apply Theorem 4.5 with the Markov chain $\{X_i^n\}_{i\in\mathbb{N}_0}$ replaced by the $\mathbb{R}^{2d}$-valued Markov chain $\{(X_i^n, U_{i,\sigma}^n)\}_{i\in\mathbb{N}_0}$. Let $Y_{i,\sigma}^n \doteq (X_i^n, U_{i,\sigma}^n)$. The following construction is analogous to Construction 4.4, but for the doubled set of noises appearing in the mollification. In addition, we will build in the restriction on controls just mentioned.

Construction 4.19 *Suppose we are given a probability measure $\mu^n \in P((\mathbb{R}^d \times \mathbb{R}^d)^n)$, and decompose it into a collection of stochastic kernels. With a point in $(\mathbb{R}^d \times \mathbb{R}^d)^n$ denoted by $(v_1, w_1, v_2, w_2, \ldots, v_n, w_n)$, $[\mu^n]_{i|0,\ldots,i-1}^1$ will denote the*

marginal distribution of μ^n on v_i given (v_j, w_j), $j < i$, and $[\mu^n]_{i|0,\ldots,i-1}^2$ will denote the marginal distribution of μ^n on w_i given (v_j, w_j), $j < i$ and v_i.

We now assume that μ^n also has the property that $[\mu^n]_{i|0,\ldots,i-1}^2$ does not depend on v_i, which implies that the distributions on v_i and w_i are conditionally independent given (v_j, w_j), $j < i$. Let $\{(\bar{v}_i^n, \bar{w}_i^n)\}_{i=0,\ldots,n-1}$ be random variables defined on a probability space $(\Omega, \mathscr{F}, P)$ and with joint distribution μ^n. Let $\bar{\mathscr{F}}_i^n \doteq \sigma((\bar{v}_j^n, \bar{w}_j^n), j = 0, \ldots, i-1)$, and define

$$\bar{\mu}_i^{1,n}(dv_i) \doteq [\mu^n]_{i|0,\ldots,i-1}^1(dv_i | \bar{v}_0^n, \bar{w}_0^n, \ldots, \bar{v}_{i-1}^n, \bar{w}_{i-1}^n)$$
$$\bar{\mu}_i^{2,n}(dw_i) \doteq [\mu^n]_{i|0,\ldots,i-1}^2(dw_i | \bar{v}_0^n, \bar{w}_0^n, \ldots, \bar{v}_{i-1}^n, \bar{w}_{i-1}^n),$$

so that these controls pick the distributions of $\bar{v}_i^n$ and $\bar{w}_i^n$ conditioned on $\bar{\mathscr{F}}_i^n$. Controlled processes $\bar{X}^n$, $\bar{U}_\sigma^n$ and measures $\bar{L}^n$ are then recursively constructed as follows. Let $(\bar{X}_0^n, \bar{U}_{0,\sigma}^n) = (x_0, 0)$ and define

$$\bar{X}_{i+1}^n = \bar{X}_i^n + \frac{1}{n}\bar{v}_i^n, \quad \bar{U}_{i+1,\sigma}^n = \bar{U}_{i,\sigma}^n + \frac{1}{n}\bar{w}_i^n. \tag{4.29}$$

Note that $\bar{\mathscr{F}}_i^n = \sigma((\bar{X}_j^n, \bar{U}_{j,\sigma}^n), j = 1, \ldots, i)$. When $\{(\bar{X}_i^n, \bar{U}_{i,\sigma}^n)\}_{i=1,\ldots,n}$ has been constructed, $\bar{X}^n(t)$ and $\bar{U}_\sigma^n(t)$ are defined as in (4.2) as the piecewise linear interpolations, and we set $\bar{Z}_\sigma^n(t) \doteq \bar{X}^n(t) + \bar{U}_\sigma^n(t)$ for $t \in [0, 1]$. In addition, define

$$\bar{L}^n(A \times B) \doteq \int_B \bar{L}^n(A|t)dt, \quad \bar{L}^n(A|t) = \delta_{(\bar{v}_i^n, \bar{w}_i^n)}(A) \text{ if } t \in [i/n, i/n + 1/n).$$

The following is the main result of this section. Owing to the restriction placed on the controls in Construction 4.19, we obtain only an inequality, but the inequality is in the right direction to establish a Laplace lower bound.

Proposition 4.20 *Let $F : \mathscr{C}([0, 1] : \mathbb{R}^d) \to \mathbb{R}$ be Lipschitz continuous. Given a control $\{(\bar{\mu}_i^{1,n}, \bar{\mu}_i^{2,n})\}_{i=0,\ldots,n-1}$, let $\{\bar{X}_i^n\}$ and $\{\bar{Z}_\sigma^n\}$ be defined as in Construction 4.19. Then for all $n \in \mathbb{N}$ and $\sigma > 0$,*

$$-\frac{1}{n} \log E e^{-nF(Z_\sigma^n)}$$

$$\leq \inf_{\{\bar{\mu}_i^{1,n}, \bar{\mu}_i^{2,n}\}} E \left[F(\bar{Z}_\sigma^n) + \frac{1}{n} \sum_{i=0}^{n-1} \left[R\left(\bar{\mu}_i^{1,n}(\cdot) \| \theta(\cdot | \bar{X}_i^n)\right) + R\left(\bar{\mu}_i^{2,n} \| \rho_\sigma\right) \right] \right].$$

Proof We apply Theorem 4.5 with d replaced by $2d$, $\{X_i^n\}$ replaced by $\{(X_i^n, U_{i,\sigma}^n)\}$, and $G : \mathscr{P}(\mathbb{R}^{2d} \times [0, 1]) \to \mathbb{R}$ defined by

$$G(\gamma) \doteq F(\varphi_\gamma), \quad \gamma \in \mathscr{P}(\mathbb{R}^{2d} \times [0, 1]),$$

where $\varphi_\gamma \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ is defined by

$$\varphi_\gamma(t) \doteq \int_{\mathbb{R}^{2d} \times [0,t]} (y + z)\gamma(dy \times dz \times ds) + x_0, \ t \in [0, 1]$$

if $\|y + z\|$ has finite integral under γ, and φ_γ identically zero otherwise. Let $Y_{i,\sigma}^n = (X_i^n, U_{i,\sigma}^n)$ and let $\bar{Y}_{i,\sigma}^n$ be its controlled analogue according to (4.29), with the appropriate replacements, and in particular $\{\bar{v}_i^n\}$ replaced by $\{(\bar{v}_i^n, \bar{w}_i^n)\}$. Let

$$\bar{\mu}_i^n(A \times B) \doteq \bar{\mu}_i^{1,n}(A)\bar{\mu}_i^{2,n}(B), \ A, B \in \mathscr{B}(\mathbb{R}^d).$$

Since we have placed restrictions on the measures μ^n (equivalently on the controls $\{(\bar{\mu}_i^{1,n}, \bar{\mu}_i^{2,n})\}$), Theorem 4.5 yields the inequality

$$-\frac{1}{n} \log Ee^{-nF(Z_\sigma^n)} \tag{4.30}$$

$$\leq \inf_{\{\bar{\mu}_i^{1,n}, \bar{\mu}_i^{2,n}\}} E\left[F(\varphi_{\bar{L}^n}) + \frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^n(dv_i \times dw_i) \| \theta(dv_i | \bar{X}_i^n)\rho_\sigma(dw_i)\right) \right].$$

Here we have used that the distribution of the original process on (v_i, w_i) depends on $Y_{i,\sigma}^n$ only through X_i^n. The chain rule implies

$$R\left(\bar{\mu}_i^n(dv_i \times dw_i) \| \theta(dv_i | \bar{X}_i^n)\rho_\sigma(dw_i)\right) = R\left(\bar{\mu}_i^{1,n}(\cdot) \| \theta(\cdot | \bar{X}_j^n)\right) + R\left(\bar{\mu}_i^{2,n} \| \rho_\sigma\right).$$

Finally, from the definition of φ_γ it follows that if the relative entropy cost is finite, then $F(\varphi_{\bar{L}^n}) = F(\bar{Z}_\sigma^n)$ w.p.1. Inserting these into (4.30) completes the proof of the lemma.   □

4.8.3 Perturbation of L and Its Properties

In order to characterize the limits of the relative entropy terms in (4.30), we use a perturbation of the function L introduced in (4.5). For $\sigma > 0$, let

$$L_\sigma(x, \beta) \doteq \sup_{\alpha \in \mathbb{R}^d}\left[\langle \alpha, \beta \rangle - H(x, \alpha) - \frac{\sigma^2}{2}\|\alpha\|^2 \right].$$

Note that for each $x \in \mathbb{R}^d$, $\beta \mapsto L_\sigma(x, \beta)$ is the Legendre–Fenchel transform of

$$H_\sigma(x, \alpha) \doteq \log \int_{\mathbb{R}^d} e^{\langle \alpha, y \rangle}\theta_\sigma(dy|x),$$

where $\theta_\sigma(\cdot|x)$ is the distribution of $v_0(x) + w_{0,\sigma}$. The following lemma records some important properties of L_σ.

Lemma 4.21 *Assume Conditions 4.3 and 4.8 and fix $\sigma > 0$. Then the following conclusions hold.*

(a) *For all $(x, \beta) \in \mathbb{R}^d \times \mathbb{R}^d$,*

$$L_\sigma(x, \beta) = \inf_{b \in \mathbb{R}^d} \left[L(x, \beta - b) + \frac{\|b\|^2}{2\sigma^2} \right].$$

(b) *For all $(x, \beta) \in \mathbb{R}^d \times \mathbb{R}^d$, $L_\sigma(x, \beta) \le L(x, \beta)$.*

(c) *$(x, \beta) \mapsto L_\sigma(x, \beta)$ is a finite nonnegative continuous function on $\mathbb{R}^d \times \mathbb{R}^d$.*

(d) *Condition 4.8 is satisfied with L replaced by L_σ uniformly in σ in the following sense: for every compact $K \subset \mathbb{R}^d$ and $\varepsilon \in (0, 1)$ there exist $\bar{\eta} = \bar{\eta}(K, \varepsilon) \in (0, 1)$ and $\bar{m} = \bar{m}(K, \varepsilon) \in (0, \infty)$ such that whenever $\xi, \chi \in K$ satisfy $\|\xi - \chi\| \le \bar{\eta}$, for every $\bar{\gamma} \in \mathbb{R}^d$ and $\sigma > 0$ we can find $\bar{\beta} \in \mathbb{R}^d$ such that*

$$L_\sigma(\xi, \bar{\beta}) - L_\sigma(\chi, \bar{\gamma}) \le \varepsilon(1 + L_\sigma(\chi, \bar{\gamma})), \quad \|\bar{\beta} - \bar{\gamma}\| \le \bar{m}(1 + L_\sigma(\chi, \bar{\gamma}))\|\xi - \chi\|.$$
(4.31)

(e) *Given $\zeta \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ satisfying $I(\zeta) < \infty$, there is a $\zeta^* \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ that is piecewise linear with finitely many pieces such that $\|\zeta^* - \zeta\|_\infty < \sigma$ and*

$$\int_0^1 L_\sigma(\zeta^*(t), \dot{\zeta}^*(t))dt \le \int_0^1 L_\sigma(\zeta(t), \dot{\zeta}(t))dt + \sigma \le I(\zeta) + \sigma. \qquad (4.32)$$

Proof (a) For each $x \in \mathbb{R}^d$, $\beta \mapsto L_\sigma(x, \beta)$ is the Legendre–Fenchel transform of the sum of the convex functions $H(x, \cdot)$ and $\bar{H}_\sigma(\cdot)$, where $\bar{H}_\sigma(\alpha) = \sigma^2\|\alpha\|^2/2$, $\alpha \in \mathbb{R}^d$. The Legendre transform of the first function is $L(x, \cdot)$, and that of the second function is $L_\sigma(\beta) = \|\beta\|^2/2\sigma^2$, $\beta \in \mathbb{R}^d$. From Theorem 16.4 of [217] it follows that

$$L_\sigma(x, \beta) = \inf \left[L(x, \beta_1) + \frac{\|\beta_2\|^2}{2\sigma^2} : \beta_i \in \mathbb{R}^d, i = 1, 2, \beta_1 + \beta_2 = \beta \right],$$

as claimed.

(b) This is an immediate consequence of part (a).

(c) Note that Conditions 4.3 and 4.7 are satisfied with $H(x, \alpha)$ replaced by $H_\sigma(x, \alpha) \doteq H(x, \alpha) + H_\sigma(\alpha)$ and $\theta(\cdot|x)$ replaced by $\theta_\sigma(\cdot|x)$. Part (c) now follows from parts (a), (d), and (f) of Lemma 4.16.

(d) Fix a compact $K \subset \mathbb{R}^d$ and $\varepsilon \in (0, 1)$. From Condition 4.8, we can find $\eta \in (0, 1)$ and $m \in (0, \infty)$ such that whenever $\xi, \chi \in K$ satisfy $\|\xi - \chi\| \le \eta$, we can find for every $\gamma \in \mathbb{R}^d$ a $\beta \in \mathbb{R}^d$ such that

$$L(\xi, \beta) - L(\chi, \gamma) \le \frac{\varepsilon}{2}(1 + L(\chi, \gamma)), \quad \|\beta - \gamma\| \le m(1 + L(\chi, \gamma))\|\xi - \chi\|.$$
(4.33)

We claim that the statement in (d) holds for $\bar{\eta} = \eta$ and $\bar{m} = 2m$. To see this, suppose $\xi, \chi \in K$ satisfy $\|\xi - \chi\| \le \eta$ and $\bar{\gamma} \in \mathbb{R}^d$. Using part (a), we can find $\bar{b} \in \mathbb{R}^d$ such that

$$L_\sigma(\chi, \bar\gamma) \geq L(\chi, \bar\gamma - \bar b) + \frac{\|\bar b\|^2}{2\sigma^2} - \frac{\varepsilon}{4}. \tag{4.34}$$

Using Condition 4.8 and taking $\gamma = \bar\gamma - \bar b$, we can find $\beta \in \mathbb{R}^d$ such that (4.33) holds. Letting $\bar\beta = \beta + \bar b$, we have

$$\begin{aligned}
\|\bar\beta - \bar\gamma\| &= \|\beta - \gamma\| \\
&\leq m(1 + L(\chi, \bar\gamma - \bar b))\|\xi - \chi\| \\
&\leq m(1 + L_\sigma(\chi, \bar\gamma) + \varepsilon/4)\|\xi - \chi\| \\
&\leq 2m(1 + L_\sigma(\chi, \bar\gamma))\|\xi - \chi\|,
\end{aligned}$$

where the second inequality follows from (4.34). This proves the second inequality in (4.31). Also, from part (a),

$$\begin{aligned}
L_\sigma(\xi, \bar\beta) - L_\sigma(\chi, \bar\gamma) &\leq L(\xi, \bar\beta - \bar b) + \frac{\|\bar b\|^2}{2\sigma^2} - L(\chi, \bar\gamma - \bar b) - \frac{\|\bar b\|^2}{2\sigma^2} + \frac{\varepsilon}{4} \\
&\leq \frac{\varepsilon}{2}\left(1 + L(\chi, \bar\gamma - \bar b) + \frac{\|\bar b\|^2}{2\sigma^2}\right) + \frac{\varepsilon}{4} \\
&\leq \varepsilon\left(1 + L_\sigma(\chi, \bar\gamma)\right),
\end{aligned}$$

where the second inequality is from (4.33) and the third is from (4.34). This proves the first inequality in (4.31) and completes the proof of part (d).

(e) Recall that for all $\sigma > 0$, $H_\sigma(\cdot, \cdot)$ and θ_σ satisfy Conditions 4.3 and 4.7. Fix $\zeta \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ satisfying $I(\zeta) < \infty$. Note that $I_\sigma(\zeta) \doteq \int_0^1 L_\sigma(\zeta(t), \dot\zeta(t))dt \leq I(\zeta)$. Then applying Lemma 4.17 to I_σ, we can find $\zeta_1^* \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ such that $\{\dot\zeta_1^*(t) : 0 \leq t \leq 1\}$ is bounded, $\|\zeta_1^* - \zeta\|_\infty < \frac{\sigma}{2}$, and (4.32) holds with ζ^* replaced by ζ_1^* and σ replaced by $\sigma/2$. The statement in part (e) now follows by taking ζ^* to be a piecewise linear approximation of ζ_1^* and using the continuity of $(x, \beta) \mapsto L_\sigma(x, \beta)$. $\qquad\square$

4.8.4 A Nearly Optimal Trajectory and Associated Control Sequence

For $\varepsilon > 0$ let $\zeta \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ be such that

$$F(\zeta) + I(\zeta) \leq \inf_{\varphi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [F(\varphi) + I(\varphi)] + \varepsilon. \tag{4.35}$$

Here F, as in Sect. 4.8.1, is a Lipschitz continuous function from $\mathscr{C}([0, 1] : \mathbb{R}^d)$ to $\mathbb{R}$, and I is the expected rate function for the system without mollification. From part (e) of Lemma 4.21, for each fixed $\sigma \in (0, 1)$, we can find a $\zeta^* \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ that

is piecewise linear with finitely many pieces such that $\|\zeta^* - \zeta\|_\infty \leq \sigma$ and (4.32) holds, i.e., $\int_0^1 L_\sigma(\zeta^*(t), \dot{\zeta}^*(t)) dt \leq I(\zeta) + \sigma$.

We now construct a control sequence to be applied to the mollified process for which the running cost is asymptotically close to the left side in (4.32) and the associated controlled process $\bar{Z}_\sigma^n$ tracks the nearly optimal trajectory ζ^* closely as $n \to \infty$. Let

$$K \doteq \bigcup_{t \in [0,1]} \left\{ y \in \mathbb{R}^d : \|y - \zeta(t)\| \leq 2 \right\}.$$

Since ζ is continuous, K is compact. We apply part (d) of Lemma 4.21 with K defined as in the last display and ε as in (4.35). Thus there exist $\bar{\eta} \in (0, 1)$ and $\bar{m} \in (0, \infty)$ such that whenever $\xi, \chi \in K$ satisfy $\|\xi - \chi\| \leq \bar{\eta}$, we can find for every $\bar{\gamma} \in \mathbb{R}^d$ and $\sigma \in (0, 1)$, a $\bar{\beta} \in \mathbb{R}^d$ such that (4.31) holds. The following lemma says that the selection of $\bar{\beta}$ can be done in a measurable way. Note that the choice of $\bar{\eta}$ and $\bar{m}$ depends on ε and K, but is independent of σ.

Lemma 4.22 (a) *Fix* $\chi, \bar{\gamma} \in \mathbb{R}^d$ *and* $\sigma > 0$. *Given* $\xi \in K(\chi) \doteq \{x \in K : \|x - \chi\| \leq \bar{\eta}\}$, *define* Γ_ξ *to be the set of all* $\bar{\beta} \in \mathbb{R}^d$ *such that* (4.31) *holds. Then there is a measurable map* $B : K(\chi) \to \mathbb{R}^d$ *such that* $B(\xi) \in \Gamma_\xi$ *for all* $\xi \in K(\chi)$ *and* $\chi \in K$.
(b) *Let* $K_{\bar{M}} = \{\beta \in \mathbb{R}^d : \|\beta\| \leq \bar{M}\}$, *where*

$$\bar{M} = \sup_{s \in [0,1], \xi \in K} \left[\bar{m} \|\xi - \zeta^*(s)\| \left(1 + L_\sigma(\zeta^*(s), \dot{\zeta}^*(s)) \right) + \|\dot{\zeta}^*(s)\| \right].$$

Given $(\xi, \beta) \in K \times K_{\bar{M}}$, *define* $\tilde{\Gamma}_{(\xi, \beta)}$ *to be the set of all* $(\beta^1, \beta^2) \in \mathbb{R}^d \times \mathbb{R}^d$ *such that*

$$L(\xi, \beta^1) + \frac{1}{2\sigma^2} \|\beta^2\|^2 \leq L_\sigma(\xi, \beta) + \sigma, \ \beta^1 + \beta^2 = \beta.$$

Then there are measurable maps $B^i : K \times K_{\bar{M}} \to \mathbb{R}^d$, $i = 1, 2$, *such that* $(B^1(\xi, \beta), B^2(\xi, \beta)) \in \tilde{\Gamma}_{(\xi, \beta)}$ *for all* $(\xi, \beta) \in K \times K_{\bar{M}}$.

Proof Corollary E.3 in the appendix is concerned with measurable selections. The proof of part (a) is immediate from this corollary and the continuity of $L_\sigma(\cdot, \cdot)$. The second part also follows from Corollary E.3 and the lower semicontinuity of L proved in part (b) of Lemma 4.14. Indeed, suppose $(\xi_n, \beta_n) \in K \times K_{\bar{M}}$ are such that $\xi_n \to \xi$ and $\beta_n \to \beta$, and let $(\beta_n^1, \beta_n^2) \in \tilde{\Gamma}_{(\xi_n, \beta_n)}$. Since $K \times K_{\bar{M}}$ is compact and L_σ is continuous, $\sup_{n \in \mathbb{N}} L_\sigma(\xi_n, \beta_n) < \infty$. Using the inequality

$$L(\xi_n, \beta_n^1) + \frac{1}{2\sigma^2} \|\beta_n^2\|^2 \leq L_\sigma(\xi_n, \beta_n) + \sigma, \tag{4.36}$$

we see that $\{\beta_n^2\}$ is bounded, and since $\beta_n^1 + \beta_n^2 = \beta_n$, $\{\beta_n^1\}$ is bounded as well. Suppose now that $\beta_n^1 \to \beta^1$ and $\beta_n^2 \to \beta^2$ along a subsequence. Clearly $\beta^1 + \beta^2 = \beta$, and from the lower semicontinuity of L and continuity of L_σ, (4.36) holds

with $(\xi_n, \beta_n, \beta_n^1, \beta_n^2)$ replaced by $(\xi, \beta, \beta^1, \beta^2)$. Thus $(\beta^1, \beta^2) \in \tilde{\Gamma}_{(\xi,\beta)}$. Hence the assumptions of Corollary E.3 are satisfied, and the result follows. $\qquad\square$

As shown in Appendix B, it follows from part (g) of Lemma 4.16 that there are stochastic kernels γ^i, $i = 1, 2$, from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathscr{P}(\mathbb{R}^d)$ and $\mathbb{R}^d$ to $\mathscr{P}(\mathbb{R}^d)$, respectively, such that for all $(\xi, \beta^1) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\beta^2 \in \mathbb{R}^d$,

$$R\left(\gamma^1(\cdot|\xi, \beta^1) \,\|\, \theta(\cdot|\xi)\right) = L(\xi, \beta^1) \quad \text{and} \quad \int_{\mathbb{R}^d} y\gamma^1(dy|\xi, \beta^1) = \beta^1$$

and

$$R\left(\gamma^2(\cdot|\beta^2) \,\|\, \rho_\sigma(\cdot)\right) = \frac{1}{2\sigma^2}\|\beta^2\|^2 \quad \text{and} \quad \int_{\mathbb{R}^d} y\gamma^2(dy|\beta^2) = \beta^2.$$

Using these kernels, we recursively define a control sequence $\{(\bar{\mu}_i^{1,n}, \bar{\mu}_i^{2,n})\}$, controlled processes $\{(\bar{X}_i^n, \bar{U}_{i,\sigma}^n)\}$, and a stopping time κ^n as follows. We initialize with $(\bar{X}_0^n, \bar{U}_{0,\sigma}^n) = (x_0, 0)$, and set

$$\kappa^n \doteq \inf\left\{i \in \mathbb{N}_0 : \|\bar{X}_i^n - \zeta^*(i/n)\| > \bar{\eta}\right\} \wedge n.$$

At each discrete time $j = 0, \ldots, \kappa^n - 1$ we apply part (a) of Lemma 4.22 with $(\chi, \bar{\gamma}) = (\zeta^*(j/n), \dot{\zeta}^*(j/n))$. Noting that $\bar{X}_j^n \in K(\zeta^*(j/n))$, we define $\bar{\beta}_j^n = B(\bar{X}_j^n)$. Note that $\bar{\beta}_j^n \in K_{\bar{M}}$. With B^i as in part (b) of Lemma 4.22, let $\beta_j^{i,n} = B^i(\bar{X}_j^n, \bar{\beta}_j^n)$, $i = 1, 2$. Define

$$\bar{\mu}_j^{1,n}(\cdot) = 1_{\{j < \kappa^n\}}\gamma^1(\cdot|\bar{X}_j^n, \beta_j^{1,n}) + 1_{\{j \geq \kappa^n\}}\theta(\cdot|\bar{X}_j^n),$$

$$\bar{\mu}_j^{2,n}(\cdot) = 1_{\{j < \kappa^n\}}\gamma^2(\cdot|\beta_j^{2,n}) + 1_{\{j \geq \kappa^n\}}\rho_\sigma(\cdot)$$

and define $\bar{v}_j^n$, $\bar{w}_{j,\sigma}^n$, $\bar{X}_{j+1}^n$, $\bar{U}_{j+1,\sigma}^n$, $\bar{X}^n$, $\bar{U}_\sigma^n$, and $\bar{Z}_\sigma^n$ according to Construction 4.19. As in the previous proof of a Laplace lower bound, we revert to the original distributions when $\bar{X}_i^n$ wanders farther than $\bar{\eta}$ from $\zeta^*(i/n)$ to keep the relative entropy costs uniformly bounded. This will be needed when we study the convergence of the controlled processes.

Note that the choice of the control sequence ensures that for $j \in \{0, 1, \ldots, \kappa^n - 1\}$,

$$R\left(\bar{\mu}_j^{1,n}(\cdot) \,\|\, \theta(\cdot|\bar{X}_j^n)\right) = L(\bar{X}_j^n, \beta_j^{1,n}), \quad \int_{\mathbb{R}^d} y\bar{\mu}_j^{1,n}(dy) = \beta_j^{1,n} \tag{4.37}$$

and

$$R\left(\bar{\mu}_j^{2,n} \,\|\, \rho_\sigma\right) = \frac{1}{2\sigma^2}\|\beta_j^{2,n}\|^2, \quad \int_{\mathbb{R}^d} y\bar{\mu}_j^{2,n}(dy) = \beta_j^{2,n}. \tag{4.38}$$

Also for $j \in \{\kappa^n, \ldots, n-1\}$, $R(\bar{\mu}_j^{1,n}\|\theta(\cdot|\bar{X}_j^n)) = R(\bar{\mu}_j^{2,n}\|\rho_\sigma) = 0$. From the choice of $(\chi, \bar{\gamma}) = (\zeta^*(j/n), \dot{\zeta}^*(j/n))$ in the definition of $\bar{\beta}_j^n$,

$$L_\sigma(\bar{X}_j^n, \bar{\beta}_j^n) - L_\sigma(\zeta^*(j/n), \dot{\zeta}^*(j/n)) \leq \varepsilon \left[1 + L_\sigma(\zeta^*(j/n), \dot{\zeta}^*(j/n))\right], \quad (4.39)$$

$$\|\bar{\beta}_j^n - \dot{\zeta}^*(j/n)\| \leq \bar{m}\left[1 + L_\sigma(\zeta^*(j/n), \dot{\zeta}^*(j/n))\right]\|\bar{X}_j^n - \zeta^*(j/n)\| \quad (4.40)$$

and

$$L(\bar{X}_j^n, \beta_j^{1,n}) + \frac{1}{2\sigma^2}\|\beta_j^{2,n}\|^2 \leq L_\sigma(\bar{X}_j^n, \bar{\beta}_j^n) + \sigma. \quad (4.41)$$

It follows that

$$E\left[\frac{1}{n}\sum_{j=0}^{n-1}\left(R\left(\bar{\mu}_j^{1,n}(\cdot)\|\theta(\cdot|\bar{X}_j^n)\right) + R\left(\bar{\mu}_j^{2,n}(\cdot)\|\rho_\sigma(\cdot)\right)\right)\right]$$

$$= E\left[\frac{1}{n}\sum_{j=0}^{\kappa^n-1}\left(R\left(\bar{\mu}_j^{1,n}(\cdot)\|\theta(\cdot|\bar{X}_j^n)\right) + R\left(\bar{\mu}_j^{2,n}(\cdot)\|\rho_\sigma(\cdot)\right)\right)\right]$$

$$= E\left[\frac{1}{n}\sum_{j=0}^{\kappa^n-1}\left(L(\bar{X}_j^n, \beta_j^{1,n}) + \frac{1}{2\sigma^2}\|\beta_j^{2,n}\|^2\right)\right]$$

$$\leq E\left[\frac{1}{n}\sum_{j=0}^{\kappa^n-1}L_\sigma(\bar{X}_j^n, \bar{\beta}_j^n)\right] + \sigma$$

$$\leq \frac{1}{n}\sum_{j=0}^{n-1}\left[L_\sigma(\zeta^*(j/n), \dot{\zeta}^*(j/n)) + \varepsilon\left[1 + L_\sigma(\zeta^*(j/n), \dot{\zeta}^*(j/n))\right]\right] + \sigma,$$

where the first equality follows from observing that for $j \geq \kappa^n$, the relative entropy terms in the first line are zero, the second from (4.37) and (4.38), the inequality on the third line from (4.41), and the last line from (4.39). Taking limits as $n \to \infty$ in the last display and using the continuity of L_σ and that ζ^* is piecewise linear, it follows that

$$\limsup_{n\to\infty} E\left[\frac{1}{n}\sum_{j=0}^{n-1}\left(R\left(\bar{\mu}_j^{1,n}(\cdot)\|\theta(\cdot|\bar{X}_j^n)\right) + R\left(\bar{\mu}_j^{2,n}(\cdot)\|\rho_\sigma(\cdot)\right)\right)\right]$$

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

$$\leq (1+\varepsilon)I(\zeta) + 2\sigma + \varepsilon(1+\sigma), \quad (4.42)$$

where the last inequality follows on recalling that ζ^* was chosen so that (4.32) is satisfied.

We recall that Lemma 4.18 gave a bound for $Ee^{-nF(X^n)}$ in terms of $Ee^{-nF(Z^n_\sigma)}$, and that Proposition 4.20 gave a variational bound for $-\frac{1}{n}\log Ee^{-nF(Z^n_\sigma)}$. If we combine these with the last display and (4.35), then

$$
\begin{aligned}
\limsup_{n\to\infty} -\frac{1}{n}\log Ee^{-nF(X^n)} \leq\ & F(\zeta) + I(\zeta) + 2\sigma + \varepsilon(1+\sigma+I(\zeta)) \\
& + \limsup_{n\to\infty} E\left[F(\bar{Z}^n_\sigma) - F(\zeta)\right] + M^2\sigma^2/2 \\
\leq\ & \inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)} [F(\varphi)+I(\varphi)] + 2\sigma + \varepsilon(2+\sigma+I(\zeta)) \\
& + \limsup_{n\to\infty} E\left[F(\bar{Z}^n_\sigma)-F(\zeta)\right] + M^2\sigma^2/2,
\end{aligned}
$$

where M, as in the statement of Lemma 4.18, is the Lipschitz constant of F. Taking the limit as $\sigma \to 0$ and then $\varepsilon \to 0$ gives

$$
\begin{aligned}
&\limsup_{n\to\infty} -\frac{1}{n}\log Ee^{-nF(X^n)} \\
&\leq \inf_{\varphi\in\mathscr{C}([0,1]:\mathbb{R}^d)} [F(\varphi)+I(\varphi)] + \limsup_{\varepsilon\to 0}\limsup_{\sigma\to 0}\limsup_{n\to\infty} E\left[F(\bar{Z}^n_\sigma) - F(\zeta)\right].
\end{aligned}
$$

Hence in order to complete the proof of the Laplace principle lower bound, it now suffices to argue that

$$
\limsup_{\varepsilon\to 0}\limsup_{\sigma\to 0}\limsup_{n\to\infty} E\left[F(\bar{Z}^n_\sigma) - F(\zeta)\right] = 0. \tag{4.43}
$$

To do this we must analyze the asymptotic properties of the controls and controlled processes.

4.8.5 Tightness and Convergence of Controlled Processes

To prove (4.43) we will need to establish tightness and characterize the limits of $\{\bar{Z}^n_\sigma\}$. The main results that are needed have already been established in Lemmas 4.11 and 4.12. To apply these results, we first must identify correspondences between objects here [on $\mathbb{R}^d \times \mathbb{R}^d$] and those of the lemmas [on $\mathbb{R}^d$]. We recall the definitions

$$
\bar{L}^n(A\times C) \doteq \int_C \bar{L}^n(A|t)dt, \quad \bar{L}^n(A|t) \doteq \delta_{(\bar{v}^n_i,\bar{w}^n_{i,\sigma})}(A) \text{ if } t\in[i/n, i/n+1/n)
$$

and $\bar{\mu}^n_i(A\times B) \doteq \bar{\mu}^{1,n}_i(A)\bar{\mu}^{2,n}_i(B)$. Define random probability measures by

$$\bar{\mu}^{1,n}(A \times C) \doteq \int_C \bar{\mu}^{1,n}(A|t)dt, \quad \bar{\mu}^{1,n}(A|t) \doteq \bar{\mu}_i^{1,n}(A) \text{ if } t \in [i/n, i/n + 1/n),$$

$$\bar{\mu}^{2,n}(B \times C) \doteq \int_C \bar{\mu}^{2,n}(B|t)dt, \quad \bar{\mu}^{2,n}(B|t) \doteq \bar{\mu}_i^{2,n}(B) \text{ if } t \in [i/n, i/n + 1/n),$$

and

$$\bar{\mu}^n(A \times B \times C) \doteq \int_C \bar{\mu}^{1,n}(A|t)\bar{\mu}^{2,n}(B|t)dt, \ \lambda^n(A \times B \times C) \doteq \int_C \lambda^n(A \times B|t)dt,$$

$$\lambda^n(A \times B|t) \doteq \theta(A|\bar{X}_i^n)\rho_\sigma(B) \text{ if } t \in [i/n, i/n + 1/n).$$

Here $A, B \in \mathscr{B}(\mathbb{R}^d)$ and $C \in \mathscr{B}([0, 1])$. Also, let $\bar{\kappa}^n = \kappa^n/n$.

Lemma 4.23 (a) *The collection* $\{(\bar{\mu}^n, \lambda^n, \bar{\mu}^{1,n}, \bar{\mu}^{2,n}, \bar{X}^n, \bar{U}_\sigma^n, \bar{\kappa}^n)\}_{n\in\mathbb{N}}$ *is a tight family of random variables with values in*

$$\mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d \times [0, 1])^2 \times \mathscr{P}(\mathbb{R}^d \times [0, 1])^2 \times \mathscr{C}([0, 1] : \mathbb{R}^d)^2 \times [0, 1].$$

(b) *Suppose* $(\bar{\mu}^n, \lambda^n, \bar{\mu}^{1,n}, \bar{\mu}^{2,n}, \bar{X}^n, \bar{U}_\sigma^n, \bar{\kappa}^n)$ *converges along a subsequence in distribution to* $(\bar{\mu}, \lambda, \bar{\mu}^1, \bar{\mu}^2, \bar{X}, \bar{U}_\sigma, \bar{\kappa})$. *Then a.s., for every* $t \in [0, 1]$,

$$\bar{X}(t) = x_0 + \int_{\mathbb{R}^d \times [0,t]} y\bar{\mu}^1(dy \times ds) \text{ and } \bar{U}_\sigma(t) = \int_{\mathbb{R}^d \times [0,t]} z\bar{\mu}^2(dz \times ds).$$

Proof It follows from (4.42) that

$$\sup_{n\in\mathbb{N}} E\left[R\left(\bar{\mu}^n\|\lambda^n\right)\right] < \infty.$$

Also, from part (a) of Condition 4.3, for every $\alpha = (\alpha^1, \alpha^2) \in \mathbb{R}^{2d}$,

$$\sup_{x\in\mathbb{R}^d} \log \int_{\mathbb{R}^{2d}} \exp\left(\langle\alpha^1, y\rangle + \langle\alpha^1, z\rangle\right) \theta(dy|x)\rho_\sigma(dz) < \infty.$$

We can therefore apply Lemma 4.11 with $\theta(dy|x)$ replaced with $\theta(dy|x)\rho_\sigma(dz)$ and $\bar{\mu}_i^n$ now given by $\bar{\mu}_i^{1,n} \times \bar{\mu}_i^{2,n}$. The lemma implies that the families $\{\bar{\mu}^n\}$ and $\{\bar{L}^n\}$ are tight and satisfy the uniform integrability property

$$\lim_{M\to\infty} \limsup_{n\to\infty} E\left[\int_{\mathbb{R}^{2d} \times [0,1]} \|(y, z)\| \, 1_{\{\|(y,z)\|\geq M\}}\bar{\mu}^n(dy \times dz \times dt)\right]$$

$$= \lim_{M\to\infty} \limsup_{n\to\infty} E\left[\int_{\mathbb{R}^{2d} \times [0,1]} \|(y, z)\| \, 1_{\{\|(y,z)\|\geq M\}}\bar{L}^n(dy \times dz \times dt)\right]$$

$$= 0. \tag{4.44}$$

Tightness of $\bar{X}^n$ and $\bar{U}^n_\sigma$ follows from Lemma 4.11 with the identities

$$(\bar{X}^n(t), \bar{U}^n_\sigma(t)) = (x_0, 0) + \int_{\mathbb{R}^{2d} \times [0,t]} (y, z)\bar{L}^n(dy \times dz \times ds).$$

Finally, tightness of $\{(\bar{\mu}^{1,n}, \bar{\mu}^{2,n})\}$ is immediate from that of $\{\bar{\mu}^n\}$, and the tightness of $\{\bar{\kappa}^n\}$ holds trivially due to the compactness of $[0, 1]$.

It follows from Lemma 4.12 that

$$(\bar{X}(t), \bar{U}_\sigma(t)) = (x_0, 0) + \int_{\mathbb{R}^{2d} \times [0,t]} (y, z)\bar{\mu}(dy \times dz \times ds).$$

We then use that w.p.1 $\bar{\mu}^1(dy \times ds) = \bar{\mu}(dy \times \mathbb{R}^d \times ds)$ and $\bar{\mu}^2(dz \times ds) = \bar{\mu}(\mathbb{R}^d \times dz \times ds)$ to get part (b). $\qquad\square$

4.8.6 Completion of the Proof of the Laplace Lower Bound

We now return to the proof of (4.43). We will argue that

$$\limsup_{n \to \infty} E\left[F(\bar{Z}^n_\sigma) - F(\zeta)\right] \leq h(\sigma, \varepsilon), \qquad (4.45)$$

where $h : (0, \infty) \times (0, \infty) \to [0, \infty)$ satisfies $\lim_{\sigma \to 0} h(\sigma, \varepsilon) = 0$ for all $\varepsilon \in (0, 1)$. For this, by a usual subsequential argument it is enough to argue that (4.45) holds along any subsequence as in part (b) of Lemma 4.23 with a function h that is independent of the choice of the subsequence. Using the Skorohod representation theorem [Appendix A, Theorem A.8], we can assume that the convergence in part (b) of Lemma 4.23 is a.s., and without loss we can also assume that it holds along the full sequence. Then for all $t \in [0, 1]$,

$$\bar{Z}_\sigma(t) \doteq \lim_{n \to \infty} \bar{Z}^n_\sigma(t) = \lim_{n \to \infty} \bar{X}^n(t) + \lim_{n \to \infty} \bar{U}^n_\sigma(t) = \bar{X}(t) + \bar{U}_\sigma(t).$$

The following lemma estimates the difference between $\bar{Z}_\sigma$ and $\bar{X}$.

Lemma 4.24 *Let* $m(\sigma, \varepsilon) \doteq 2\sigma^2 \left((1 + \varepsilon)(2\|F\|_\infty + \sigma + \varepsilon) + \sigma + \varepsilon\right)$. *For every* $\sigma > 0$,
$$E\|\bar{Z}_\sigma - \bar{X}\|_\infty^2 = E\|\bar{U}_\sigma\|_\infty^2 \leq m(\sigma, \varepsilon).$$

Proof We use the convergence of $\bar{\mu}^{2,n}$ to $\bar{\mu}^2$ and the uniform integrability stated in (4.44). The identities in part (b) of Lemma 4.23 and an application of Fatou's lemma then imply

$$E\|\bar{U}_\sigma\|_\infty^2 \le \liminf_{n\to\infty} E\left\|\int_{\mathbb{R}^d\times[0,\cdot]} z\bar{\mu}^{2,n}(dz\times ds)\right\|_\infty^2$$

$$\le \liminf_{n\to\infty} E\frac{1}{n}\sum_{j=0}^{n-1}\left\|\int_{\mathbb{R}^d} z\bar{\mu}_j^{2,n}(dz)\right\|^2$$

$$= \liminf_{n\to\infty} E\frac{1}{n}\sum_{j=0}^{\kappa^n-1}\left\|\beta_j^{2,n}\right\|^2$$

$$\le 2\sigma^2 \liminf_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\left[L_\sigma(\zeta^*(j/n),\dot\zeta^*(j/n))\right.$$

$$\left. + \varepsilon\left[1 + L_\sigma(\zeta^*(j/n),\dot\zeta^*(j/n))\right]\right] + 2\sigma^3\,,$$

where the second inequality follows from Jensen's inequality and the third uses (4.41) and (4.39). Using the continuity of L_σ leads to

$$E\|\bar{U}_\sigma\|_\infty^2 \le 2\sigma^2\left((1+\varepsilon)\int_0^1 L_\sigma(\zeta^*(t),\dot\zeta^*(t))dt\right) + 2\sigma^2(\sigma+\varepsilon)$$

$$\le 2\sigma^2\left((1+\varepsilon)(I(\zeta)+\sigma)\right) + 2\sigma^2(\sigma+\varepsilon)$$

$$\le 2\sigma^2\left((1+\varepsilon)(2\|F\|_\infty+\sigma+\varepsilon)\right) + 2\sigma^2(\sigma+\varepsilon),$$

where the second inequality follows from (4.32) and the third from (4.35). This is the claim of the lemma. $\square$

We recall that $\beta_j^{1,n} + \beta_j^{2,n} = \bar{\beta}_j^n$ and that $\bar{\beta}_j^n$ is chosen equal to $B(\bar{X}_j^n)$, which implies (4.31) with $(\xi,\bar{\beta},\chi,\bar{\gamma}) = (\bar{X}_j^n, \bar{\beta}_j^n, \zeta^*(j/n), \dot\zeta^*(j/n))$ for $j \le \kappa^n - 1$. It follows from (4.37), (4.38), and (4.40) that for $j = 0, 1, \ldots, \kappa^n - 1$,

$$\left\|\int_{\mathbb{R}^d} y\left[\bar{\mu}_j^{1,n}(dy) + \bar{\mu}_j^{2,n}(dy)\right] - \dot\zeta^*(j/n)\right\|$$

$$= \left\|\beta_j^{1,n} + \beta_j^{2,n} - \dot\zeta^*(j/n)\right\|$$

$$= \left\|\bar{\beta}_j^n - \dot\zeta^*(j/n)\right\|$$

$$\le \bar{m}\left(1 + L_\sigma(\zeta^*(j/n),\dot\zeta^*(j/n))\right)\|\bar{X}_j^n - \zeta^*(j/n)\|.$$

From the uniform integrability property in (4.44) and the a.s. convergence of $(\bar{\mu}^{1,n}, \bar{\mu}^{2,n}, \kappa^n)$ to $(\bar{\mu}^1, \bar{\mu}^2, \kappa)$, we have for all $t \in [0,1]$ that

$$\int_{\mathbb{R}^d\times[0,t\wedge\kappa^n]} y\left[\bar{\mu}^{1,n}(dy\times ds) + \bar{\mu}^{2,n}(dy\times ds)\right]$$

converges in probability to

$$\int_{\mathbb{R}^d \times [0, t \wedge \bar{\kappa}]} y \left[\bar{\mu}^1 (dy \times ds) + \bar{\mu}^2 (dy \times ds) \right].$$

Thus for every $t \in [0, \bar{\kappa}]$,

$$\left\| \int_{\mathbb{R}^d \times [0,t]} y \left[\bar{\mu}^1 (dy \times ds) + \bar{\mu}^2 (dy \times ds) \right] - \int_0^t \dot{\zeta}^*(s) ds \right\|$$

$$= \lim_{n \to \infty} \left\| \int_{\mathbb{R}^d \times [0,t]} y \left[\bar{\mu}^{1,n} (dy \times ds) + \bar{\mu}^{2,n} (dy \times ds) \right] - \int_0^t \dot{\zeta}^*(s) ds \right\|$$

$$\leq \bar{m} \int_0^t \| \bar{X}(s) - \zeta^*(s) \| \left(1 + L_\sigma(\zeta^*(s), \dot{\zeta}^*(s)) \right) ds.$$

For $s \in [0, 1]$, let $a(s) = 1 + L_\sigma(\zeta^*(s), \dot{\zeta}^*(s))$ and $b(s) = \| \bar{Z}_\sigma(s) - \zeta^*(s) \|$. Then from part (b) of Lemma 4.23 and since $\bar{Z}_\sigma = \bar{X} + \bar{U}_\sigma$, the last display implies that for $t \in [0, \bar{\kappa}]$,

$$b(t) = \left\| \int_{\mathbb{R}^d \times [0,t]} y \left[\bar{\mu}^1 (dy \times ds) + \bar{\mu}^2 (dy \times ds) \right] - \int_0^t \dot{\zeta}^*(s) ds \right\|$$

$$\leq \bar{m} \int_0^t \| \bar{X}(s) - \zeta^*(s) \| a(s) ds$$

$$\leq \bar{m} \int_0^t b(s) a(s) ds + \bar{m} \| \bar{X} - \bar{Z}_\sigma \|_\infty \int_0^1 a(s) ds.$$

Using that [again by (4.32) and (4.35)] $\int_0^1 a(s) ds \leq 1 + 2\|F\|_\infty + \sigma + \varepsilon$, for all $\varepsilon, \sigma \in (0, 1)$, Gronwall's lemma [Lemma E.2] implies

$$\| \bar{Z}_\sigma(\cdot \wedge \bar{\kappa}) - \zeta^*(\cdot \wedge \bar{\kappa}) \|_\infty \leq \bar{m} \| \bar{X} - \bar{Z}_\sigma \|_\infty \int_0^1 a(s) ds \int_0^1 e^{\bar{m} \int_0^t a(s) ds} dt$$

$$\leq \bar{m}(2\|F\|_\infty + 3) \| \bar{X} - \bar{Z}_\sigma \|_\infty e^{\bar{m}(2\|F\|_\infty + 3)}$$

$$= \bar{m}_1 \| \bar{X} - \bar{Z}_\sigma \|_\infty, \tag{4.46}$$

where $\bar{m}_1 \doteq \bar{m}(2\|F\|_\infty + 3) e^{\bar{m}(2\|F\|_\infty + 3)}$. Finally, with M as in Lemma 4.18 equal to the Lipschitz constant of F, we obtain

$$\limsup_{n \to \infty} E\left[F(\bar{Z}_\sigma^n) - F(\zeta) \right] \leq \limsup_{n \to \infty} E\left[M \| \bar{Z}_\sigma^n - \zeta \|_\infty \wedge 2\|F\|_\infty \right]$$

$$\leq M\left[\bar{m}_1 E \| \bar{X} - \bar{Z}_\sigma \|_\infty + \sigma \right] + 2\|F\|_\infty P(\bar{\kappa} < 1)$$

$$\leq M\left(\bar{m}_1 (m(\sigma, \varepsilon))^{1/2} + \sigma \right) + 2\|F\|_\infty P(\bar{\kappa} < 1).$$

For the second inequality we use the fact that $\| \zeta - \zeta^* \|_\infty \leq \sigma$, and we partition according to whether $\bar{\kappa} < 1$, using (4.46) when this is not the case. The third inequality follows from Lemma 4.24.

The last quantity we need to control is $P(\bar\kappa < 1)$. Since the convergence in path space is with respect to the uniform topology, we have

$$P(\bar\kappa < 1) \le P\left(\left\|\bar{Z}_\sigma(\cdot \wedge \bar\kappa) - \zeta^*(\cdot \wedge \bar\kappa)\right\|_\infty \ge \bar\eta/2\right) \le \frac{4}{\bar\eta^2}\bar{m}_1^2\, m(\sigma, \varepsilon),$$

where the first inequality follows from the definition of κ^n and the second inequality uses Chebyshev's inequality, (4.46) and Lemma 4.24. Thus (4.45) holds with

$$h(\sigma, \varepsilon) = M\left(\bar{m}_1(m(\sigma, \varepsilon))^{1/2} + \sigma\right) + 2\|F\|_\infty \frac{4}{\bar\eta^2(\varepsilon)}\bar{m}_1^2\, m(\sigma, \varepsilon),$$

where we write $\bar\eta = \bar\eta(\varepsilon)$ to emphasize its dependence on ε (recall from Lemma 4.21 that $\bar\eta$ does not depend on σ). From the definition of $m(\sigma, \varepsilon)$ in Lemma 4.24 it follows that $\lim_{\sigma \to 0} m(\sigma, \varepsilon) = 0$ for all $\varepsilon \in (0, 1)$. This proves that $\lim_{\sigma \to 0} h(\sigma, \varepsilon) = 0$ for every $\varepsilon \in (0, 1)$, and hence (4.43) holds. With the limit (4.43) demonstrated, we have completed the proof of the Laplace lower bound under Condition 4.8. $\square$

4.9   Notes

Among the very first papers to treat models of this type are those of Wentzell [245–248], Freidlin and Wentzell [140], and Azencott and Ruget [8]. The conditions we use are weaker than those of the cited references. The statement of assumptions and conclusions for this chapter parallels that of Chapter 6 of [97], though the proof is different and is directly analogous to the way Cramér's theorem was obtained in Chap. 3. In particular, we first prove a process-level generalization of Sanov's theorem for the driving noises that define the recursive stochastic model. Then, assuming appropriate integrability conditions on the distribution of these noises, one obtains Laplace asymptotics for the process of interest by viewing it as a mapping on this process-level empirical measure. This leads to a somewhat simpler proof, though in all approaches the mollification aspect of the proof under Conditions 4.3 and 4.8 is technical. We also note that the proof given here corrects a gap in a proof in [97], in that the constant M defined on page 205 of [97] depends on σ, and therefore the claim in equation (6.65) of [97] is not really established.

As noted several times already, the analysis of the discrete time model of this chapter is in many ways more difficult than that of the corresponding continuous-time models, largely because for continuous time models the noise enters in an additive and affine manner.