Chapter 5
Moderate Deviations for Recursive Markov Systems

In this chapter we consider $\mathbb{R}^d$-valued discrete time processes of the same form as in Chap. 4, but instead of analyzing the large deviation behavior, we consider deviations closer to the LLN limit. Since this will require centering on the limit, we assume that the process model has the form

$$X_{i+1}^n \doteq X_i^n + \frac{1}{n}b(X_i^n) + \frac{1}{n}v_i(X_i^n), \quad X_0^n = x_0, \tag{5.1}$$

where $\{v_i(\cdot)\}_{i \in \mathbb{N}_0}$ are iid random vector fields as in Chap. 4 but with zero mean, and $b : \mathbb{R}^d \to \mathbb{R}^d$ is continuously differentiable.

As in Chap. 4, we consider the continuous time piecewise linear interpolations $\{X^n(t)\}_{0 \le t \le T}$ with $X^n(i/n) = X_i^n$. Under moment conditions that are weaker than those of Chap. 4, there is a law of large numbers limit $X^0 \in \mathscr{C}([0, T] : \mathbb{R}^d)$. Closely related to X^0 is the noiseless version of (5.1) obtained by setting $v_i(\cdot) = 0$, which is denoted by $\{X_i^{n,0}\}_{n \in \mathbb{N}_0}$ with piecewise linear interpolation $\{X^{n,0}(t)\}_{0 \le t \le T}$. We introduce a scaling sequence $\varkappa(n)$ that satisfies

$$\varkappa(n) \to 0 \text{ and } \varkappa(n)n \to \infty, \tag{5.2}$$

and study the rescaled difference

$$Y^n \doteq \sqrt{\varkappa(n)n}(X^n - X^{n,0}).$$

Since under Condition 5.1, b is Lipschitz continuous, we have

$$\left\|X^0 - X^{n,0}\right\|_\infty \doteq \sup_{t \in [0,T]} \left\|X^0(t) - X^{n,0}(t)\right\| = O(1/n).$$

Thus

$$\sqrt{\varkappa(n)n} \left\|X^0 - X^{n,0}\right\|_\infty = O(\sqrt{\varkappa(n)/n}),$$

and hence Y^n behaves the same asymptotically as $\sqrt{\varkappa(n)n}(X^n - X^0)$. It will be shown that under weaker conditions on the noise $v_i(\cdot)$ than were used in Chap. 4, Y^n satisfies the large deviation principle on $\mathscr{C}([0, T] : \mathbb{R}^d)$ with a "Gaussian"-type rate function. As is customary for this type of scaling, we refer to this as moderate deviations.

While one might expect the proof of the moderate deviations result to be similar to that of the corresponding large deviations result, there are important differences. For example, the tightness proof is significantly more complicated in the case of moderate deviations than for the case of large deviations. In Chap. 4 we were able to establish an a priori bound on certain relative entropy costs associated with any sequence of nearly minimizing controls. Under this boundedness of the relative entropy costs, empirical measures of the controlled driving noises as well as the controlled processes themselves were tight. With the scaling used for moderate deviations, and even with the information that the analogous relative entropy costs decay like $O(1/\varkappa(n)n)$, tightness of the empirical measures of the driving noise does not hold. Instead, one must consider the empirical measures of conditional means of the noises, and additional effort is required to show that the limits of these measures determine the limit of the controlled processes. This extra difficulty arises for moderate deviations (even with the vanishing relative entropy costs), because the noise is amplified by the factor $\sqrt{\varkappa(n)n}$ in the definition of Y^n.

A second way in which the proofs for large and moderate deviations differ is in their treatment of degenerate noise, i.e., problems in which the support of $v_i(\cdot)$ is not all of $\mathbb{R}^d$. As we saw in Chap. 4, this leads to significant difficulties in the proof of the large deviation lower bound, requiring a somewhat involved mollification argument. In contrast, the proof in the setting of moderate deviations, though more involved than the nondegenerate case, is much more straightforward.

As a potential application of these results we mention their usefulness in the design and analysis of Monte Carlo schemes for events whose probability is small but not very small. For such problems, the performance of standard Monte Carlo may not be adequate, especially if the quantity must be computed for many different parameter settings, as in, say, an optimization problem. Another instance is the situation in which the cost for even a single sample is very high, as for example in the case of stochastic partial differential equations. Then accelerated Monte Carlo may be of interest, and as is well known, such schemes (e.g., importance sampling and splitting) benefit from the use of information contained in the large deviation rate function as part of the algorithm design (e.g., [28, 76, 114, 116]). In a situation in which one considers events of small but not too small probability, one may find the moderate deviation approximation both adequate and relatively easy to apply, since moderate deviations lead to situations in which the objects needed to design an efficient scheme can be explicitly constructed in terms of solutions to the linear–quadratic regulator. These issues were first explored in [101]. Other moderate deviation analyses are presented in Chaps. 10 and 13, and an example of how moderate deviation approximations can be used to construct accelerated Monte Carlo schemes is given in Sect. 17.5.

5.1 Assumptions, Notation, and Theorem Statement

Let

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

where the $\{v_i(\cdot)\}_{i \in \mathbb{N}_0}$ are zero-mean iid vector fields whose distribution is given by a stochastic kernel $\theta(dy|x)$ on $\mathbb{R}^d$ given $\mathbb{R}^d$. For $\alpha \in \mathbb{R}^d$, define

$$H_c(x, \alpha) \doteq \log E e^{\langle \alpha, v_i(x) \rangle}.$$

The subscript c reflects the fact that this log moment generating function uses the centered distribution $\theta(\cdot|x)$, rather than $H(x, \alpha) = H_c(x, \alpha) + \langle \alpha, b(x) \rangle$ as in Chap. 4. We use the following.

Condition 5.1 (a) *There exist $\lambda > 0$ and $K_{mgf} < \infty$ such that*

$$\sup_{x \in \mathbb{R}^d} \sup_{\|\alpha\| \leq \lambda} H_c(x, \alpha) \leq K_{mgf}. \tag{5.3}$$

(b) *The mapping $x \mapsto \theta(dy|x)$ from $\mathbb{R}^d$ to $\mathscr{P}(\mathbb{R}^d)$ is continuous with respect to the topology of weak convergence.*
(c) *$b(x)$ is continuously differentiable, and the norms of both $b(x)$ and its derivative are uniformly bounded by a constant $K_b < \infty$.*

Throughout this chapter we let $\|\alpha\|_A^2 \doteq \langle \alpha, A\alpha \rangle$ for $\alpha \in \mathbb{R}^d$ and a symmetric nonnegative definite matrix A. Define

$$A_{ij}(x) \doteq \int_{\mathbb{R}^d} y_i y_j \theta(dy|x),$$

and note that the weak continuity of $\theta(dy|x)$ with respect to x and (5.3) ensures that $A(x)$ is continuous in x and that its norm, $\|A(x)\| \doteq \sup_{v:\|v\|=1} \|A(x)v\|$, is uniformly bounded by some constant K_A. Note that

$$\frac{\partial H_c(x, 0)}{\partial \alpha_i} = \int_{\mathbb{R}^d} y_i \theta(dy|x) = 0 \tag{5.4}$$

and

$$\frac{\partial^2 H_c(x, 0)}{\partial \alpha_i \partial \alpha_j} = \int_{\mathbb{R}^d} y_i y_j \theta(dy|x) = A_{ij}(x) \tag{5.5}$$

for all $i, j \in \{1, \ldots, d\}$ and $x \in \mathbb{R}^d$, and that $A(x)$ is nonnegative definite and symmetric. It follows that for $x \in \mathbb{R}^d$,

$$A(x) = Q(x) \Lambda(x) Q^T(x),$$

where $Q(x)$ is an orthogonal matrix whose columns are the eigenvectors of $A(x)$, and $\Lambda(x)$ is the diagonal matrix consisting of the eigenvalues of $A(x)$ in descending order. The map $x \mapsto \Lambda(x)$ is continuous. In what follows, we define $\Lambda^{-1}(x)$ to be the diagonal matrix with diagonal entries each equal to the inverse of the corresponding eigenvalue for the positive eigenvalues, and equal to ∞ for the zero eigenvalues. Then when we write

$$\|\alpha\|^2_{A^{-1}(x)} = \|\alpha\|^2_{Q(x)\Lambda^{-1}(x)Q^T(x)}, \tag{5.6}$$

we mean a value of ∞ for $\alpha \in \mathbb{R}^d$ not in the linear span of the eigenvectors corresponding to the positive eigenvalues, and the standard value for vectors $\alpha \in \mathbb{R}^d$ in that linear span. (Note that even if the definition of $A^{-1}(x)$ is ambiguous, in that for α in the range of $A(x)$ there may be more than one v such that $A(x)v = \alpha$, the value of $\|\alpha\|^2_{A^{-1}(x)}$ is not ambiguous. Indeed, since the eigenvectors can be assumed orthogonal, for all such v, $\langle v, \alpha \rangle$ coincides with $\langle \bar{v}, \alpha \rangle$, where $\bar{v}$ is the solution in the span of eigenvectors corresponding to positive eigenvalues.) Condition 5.1(a) implies there exist $K_{DA} < \infty$ and $\lambda_{DA} \in (0, \lambda]$ such that

$$\sup_{x \in \mathbb{R}^d} \sup_{\|\alpha\| \le \lambda_{DA}} \max_{i,j,k} \left| \frac{\partial^3 H_c(x, \alpha)}{\partial \alpha_i \partial \alpha_j \partial \alpha_k} \right| \le \frac{K_{DA}}{d^3}, \tag{5.7}$$

and consequently for all $\|\alpha\| \le \lambda_{DA}$ and all $x \in \mathbb{R}^d$,

$$\frac{1}{2} \|\alpha\|^2_{A(x)} - \|\alpha\|^3 K_{DA} \le H_c(x, \alpha) \le \frac{1}{2} \|\alpha\|^2_{A(x)} + \|\alpha\|^3 K_{DA}. \tag{5.8}$$

Define the continuous time piecewise linear interpolation of X^n_i by

$$X^n(t) \doteq X^n_i + [X^n_{i+1} - X^n_i](nt - i), \quad t \in [i/n, i/n + 1/n].$$

In addition, define

$$X^{n,0}_{i+1} = X^{n,0}_i + \frac{1}{n} b(X^{n,0}_i), \quad X^{n,0}_0 = x_0,$$

and let $X^{n,0}(t)$ be the analogously defined continuous time interpolation. Then $X^{n,0} \to X^0$ in $\mathscr{C}([0, T] : \mathbb{R}^d)$, where for $t \in [0, T]$,

$$X^0(t) = \int_0^t b(X^0(s))ds + x_0.$$

Since $Ev_i(x) = 0$ for all $x \in \mathbb{R}^d$, we know that $X^n \to X^0$ in $\mathscr{C}([0, T] : \mathbb{R}^d)$ in probability.

In Chap. 4 we showed, under significantly stronger assumptions, that $X^n(t)$ satisfies a large deviation principle on $\mathscr{C}([0, T] : \mathbb{R}^d)$ with scaling sequence $r(n) = 1/n$. Letting I_L denote this rate function (with L for large deviation), it takes the form

$$I_L(\phi) \doteq \inf \left[\int_0^T L_c(\phi(s), u(s)) ds : \phi(t) = x_0 + \int_0^t b(\phi(s)) ds \right.$$
$$\left. + \int_0^t u(s) ds, \ t \in [0, T] \right],$$

where

$$L_c(x, \beta) \doteq \sup_{\alpha \in \mathbb{R}^d} [\langle \alpha, \beta \rangle - H_c(x, \alpha)] \tag{5.9}$$

is the Legendre transform of $H_c(x, \alpha)$. We see that I_L coincides with the rate function of Chap. 4, because

$$L(x, \beta) \doteq \sup_{\alpha \in \mathbb{R}^d} [\langle \alpha, \beta \rangle - H_c(x, \alpha) - \langle \alpha, b(x) \rangle] = L_c(x, \beta - b(x)),$$

so that $L_c(\phi(s), u(s)) = L_c(\phi(s), \dot{\phi}(s) - b(\phi(s))) = L(\phi(s), \dot{\phi}(s))$.

Assume that $\varkappa(n)$ satisfies (5.2):

$$\varkappa(n) \to 0 \text{ and } \varkappa(n)n \to \infty.$$

We define the rescaled difference

$$Y^n(t) \doteq \sqrt{\varkappa(n)n}(X^n(t) - X^{n,0}(t)).$$

Let $Db(x)$ denote the matrix of partial derivatives $(Db(x))_{ij} = \partial b_i(x)/\partial x_j$, and let $A^{1/2}(x)$ be the unique nonnegative definite square root of $A(x)$.

Theorem 5.2 *Assume Condition 5.1. Then* $\{Y^n\}_{n \in \mathbb{N}}$ *satisfies the Laplace principle on* $\mathscr{C}([0, T] : \mathbb{R}^d)$ *with scaling sequence* $\varkappa(n)$ *and rate function*

$$I_M(\phi) = \inf \left[\frac{1}{2} \int_0^T \|u(t)\|^2 dt : \phi(t) = \int_0^t Db(X^0(s))\phi(s) ds \right.$$
$$\left. + \int_0^t A^{1/2}(X^0(s))u(s) ds, \ t \in [0, T] \right].$$

The function I_M is essentially the same as what one would obtain using a linear approximation around the law of large numbers limit X^0 of the dynamics and a quadratic approximation of the costs in I_L. By our convention, proofs will be given for the case $T = 1$, with only notational modifications needed for the general case. An alternative form of the rate function that is consistent with expressions we use for continuous time models is

$$I_M(\phi) = \inf_{u \in U_\phi} \left[\frac{1}{2} \int_0^1 \|u(t)\|^2 dt \right], \tag{5.10}$$

where for $\phi \in \mathscr{A}\mathscr{C}_0([0, 1] : \mathbb{R}^d)$ U_ϕ is the subset of $\mathscr{L}^2([0, 1] : \mathbb{R}^d)$ given by

$$U_\phi \doteq \left\{ u : \phi(\cdot) = \int_0^\cdot Db(X^0(s))\phi(s)ds + \int_0^\cdot A^{1/2}(X^0(s))u(s)ds \right\}, \tag{5.11}$$

and U_ϕ is the empty set otherwise. Since $\{\phi : I_M(\phi) \le K\}$ is the image of the compact set $\{u : \int_0^1 \|u(t)\|^2\, dt \le K\}$ (with the weak topology on $\mathscr{L}^2([0, 1] : \mathbb{R}^d)$) under a continuous mapping, I_M has compact level sets. To complete the proof of Theorem 5.2, we must show that for every bounded and continuous F,

$$\lim_{n \to \infty} -\varkappa(n) \log E\left[e^{-\frac{1}{\varkappa(n)} F(Y^n)} \right] = \inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [I_M(\phi) + F(\phi)]. \tag{5.12}$$

The argument will follow the same layout as that used in Chaps. 3 and 4. In Sect. 5.2, a representation is derived for the exponential integral in (5.12), and in Sect. 5.3, tightness of empirical measures and identification of limits for these measures and controlled processes are carried out. It is here that the moderate deviation problem differs most from the corresponding large deviation problem, in that the definition of the empirical measures is not analogous to the definition used in Chap. 4. The Laplace principle upper and lower bounds that together imply (5.12) are proved in Sects. 5.4 and 5.5, respectively.

5.2   The Representation

As usual, the first step is to identify a useful representation for the Laplace functionals. Owing to the moderate deviation scaling, the construction of the controlled processes differs slightly from that of Chap. 4.

Construction 5.3 Suppose we are given a probability measure $\mu^n \in \mathscr{P}((\mathbb{R}^d)^n)$ and decompose it in terms of conditional distributions $[\mu^n]_{i|1,\dots,i-1}$ on the ith variable given variables 0 through $i - 1$:

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

Let $\{\bar{v}_i^n\}_{i=0,\dots,n-1}$ be random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and with joint distribution μ^n. Thus conditioned on $\tilde{\mathscr{F}}_i^n \doteq \sigma(\bar{v}_j^n, j = 0, \dots, i - 1)$, $\bar{v}_i^n$ 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. Controlled processes $\bar{X}^n$, $\bar{Y}^n$ and measures $\bar{M}^n$ 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

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

When $\{\bar{X}_i^n\}_{i=1,\dots,n}$ has been constructed, define

$$\bar{Y}_{i+1}^n = \bar{Y}_i^n + \sqrt{\frac{\varkappa(n)}{n}}\left(b(\bar{X}_i^n) - b(X_i^{n,0})\right) + \sqrt{\frac{\varkappa(n)}{n}}\bar{v}_i^n, \quad \bar{Y}_0^n = 0. \tag{5.13}$$

Note that

$$\bar{X}_i^n - X_i^{n,0} = \frac{1}{\sqrt{\varkappa(n)n}}\bar{Y}_i^n, \ i = 0, 1, \dots, n. \tag{5.14}$$

Let $\bar{X}^n$ and $\bar{Y}^n$ be as in (4.2) the piecewise linear interpolations with $\bar{X}^n(i/n) = \bar{X}_i^n$ and $\bar{Y}^n(i/n) = \bar{Y}_i^n$. Define also the interpolated conditional mean (provided it exists)

$$\bar{w}^n(t) \doteq \int_{\mathbb{R}^d} y\bar{\mu}_i^n(dy), \quad t \in [i/n, i/n + 1/n),$$

the scaled conditional mean

$$w^n(t) \doteq \sqrt{\varkappa(n)n}\,\bar{w}^n(t),$$

and random measures on $\mathbb{R}^d \times [0, 1]$ by

$$\bar{M}^n(dw \times dt) \doteq \delta_{w^n(t)}(dw)dt = \delta_{\sqrt{\varkappa(n)n}\bar{w}^n(t)}(dw)dt.$$

Note that $\bar{M}^n$ is the empirical measure of the scaled conditional means and not, in contrast to Chap. 4, of the $\bar{v}_i^n$, scaled or otherwise. This additional "averaging" will be needed for tightness. We will refer to this construction when we are given $\{\bar{\mu}_i^n\}_{i=1,\dots,n}$ to identify associated $\bar{X}^n$, $\bar{Y}^n$, w^n and $\bar{M}^n$. By Theorem 4.5, for every bounded, continuous $F : \mathscr{C}([0, 1] : \mathbb{R}^d) \to \mathbb{R}$,

$$-\varkappa(n) \log E\left[e^{-\frac{1}{\varkappa(n)}F(Y^n)}\right] \tag{5.15}$$

$$= \inf_{\{\bar{\mu}_i^n\}} E\left[\sum_{i=0}^{n-1}\varkappa(n)R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n)) + F(\bar{Y}^n)\right].$$

5.3 Tightness and Limits for Controlled Processes

5.3.1 Tightness and Uniform Integrability

When the moment-generating function is finite for all α, a variational characterization of its Legendre transform in terms of relative entropy is proved in Lemma 4.16. In this

chapter we will need only the following inequality, which holds when the moment-generating function is finite in some neighborhood of the origin. Recall that $L_c(x, \cdot)$ is the Legendre–Fenchel transform of $H_c(x, \cdot)$.

Lemma 5.4 *Assume Condition 5.1. Then for all $x, \beta \in \mathbb{R}^d$,*

$$L_c(x, \beta) \leq R(\eta(\cdot) \| \theta(\cdot|x))$$

for all $\eta \in \mathscr{P}(\mathbb{R}^d)$ satisfying $\int_{\mathbb{R}^d} y\eta(dy) = \beta$.

Proof Fix $x, \beta \in \mathbb{R}^d$ and consider any $\eta \in \mathscr{P}(\mathbb{R}^d)$ that satisfies $\int_{\mathbb{R}^d} y\eta(dy) = \beta$. If $R(\eta(\cdot) \| \theta(\cdot|x)) = \infty$, the lemma is automatically true, so we assume without loss that $R(\eta(\cdot) \| \theta(\cdot|x)) < \infty$. From (5.3) we have

$$\int_{\mathbb{R}^d} e^{\lambda \|y\|/d^{1/2}} \theta(dy|x) \leq 2de^{K_{\mathrm{mgf}}} < \infty.$$

Therefore, recalling the inequality $ab \leq e^a + \ell(b)$ for $a, b \geq 0$, we have

$$\int_{\mathbb{R}^d} \frac{\lambda}{d^{1/2}} \|y\| \frac{d\eta(\cdot)}{d\theta(\cdot|x)}(y)\theta(dy|x)$$
$$\leq \int_{\mathbb{R}^d} e^{\lambda \|y\|/d^{1/2}} \theta(dy|x) + \int_{\mathbb{R}^d} \ell\left(\frac{d\eta(\cdot)}{d\theta(\cdot|x)}(y)\right) \theta(dy|x)$$
$$\leq 2de^{K_{\mathrm{mgf}}} + R(\eta(\cdot) \| \theta(\cdot|x)).$$

Consequently, for all $\alpha \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|\alpha\| \|y\| \frac{d\eta(\cdot)}{d\theta(\cdot|x)}(y)\theta(dy|x) \leq \frac{d^{1/2} \|\alpha\|}{\lambda} \left(2de^{K_{\mathrm{mgf}}} + R(\eta(\cdot) \| \theta(\cdot|x))\right) < \infty. \tag{5.16}$$

Define the bounded continuous function $F_K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$F_K(y, \alpha) = \begin{cases} \langle \alpha, y \rangle & \text{if } |\langle \alpha, y \rangle| \leq K, \\ \frac{K \langle \alpha, y \rangle}{|\langle \alpha, y \rangle|} & \text{otherwise.} \end{cases}$$

From (5.16) and the dominated convergence theorem, we have

$$\lim_{K \to \infty} \int_{\mathbb{R}^d} F_K(y, \alpha)\eta(dy) = \left\langle \alpha, \int_{\mathbb{R}^d} y\eta(dy) \right\rangle = \langle \alpha, \beta \rangle.$$

Another application of the monotone convergence theorem gives

$$\lim_{K \to \infty} \int_{\{y:\langle \alpha, y \rangle < 0\}} e^{F_K(y,\alpha)} \theta(dy|x) = \int_{\{y:\langle \alpha, y \rangle < 0\}} e^{\langle \alpha, y \rangle} \theta(dy|x),$$

and the monotone convergence theorem gives

$$\lim_{K \to \infty} \int_{\{y: \langle \alpha, y \rangle \geq 0\}} e^{F_K(y, \alpha)} \theta(dy|x) = \int_{\{y: \langle \alpha, y \rangle \geq 0\}} e^{\langle \alpha, y \rangle} \theta(dy|x).$$

Thus

$$\lim_{K \to \infty} \log \left(\int_{\mathbb{R}^d} e^{F_K(y, \alpha)} \theta(dy|x) \right) = H_c(x, \alpha).$$

By the Donsker–Varadhan variational formula (Proposition 2.2), for every $K \in (0, \infty)$ and $\alpha \in \mathbb{R}^d$,

$$R(\eta(\cdot) \| \theta(\cdot|x)) \geq \int_{\mathbb{R}^d} F_K(y, \alpha) \eta(dy) - \log \left(\int_{\mathbb{R}^d} e^{F_K(y, \alpha)} \theta(dy|x) \right).$$

Sending $K \to \infty$ and taking the supremum over $\alpha \in \mathbb{R}^d$ yields

$$R(\eta(\cdot) \| \theta(\cdot|x)) \geq \sup_{\alpha \in \mathbb{R}^d} \left[\langle \alpha, \beta \rangle - H_c(x, \alpha) \right] = L_c(x, \beta),$$

which completes the proof of the lemma.   □

Theorem 5.5 *Assume Condition 5.1 and*

$$\sup_{n \in \mathbb{N}} \left[\varkappa(n) n E \left[\frac{1}{n} \sum_{i=0}^{n-1} R(\bar{\mu}_i^n(\cdot) \| \theta(\cdot | \bar{X}_i^n)) \right] \right] \leq K_E < \infty. \qquad (5.17)$$

Let $\{\bar{M}^n\}_{n \in \mathbb{N}}$, $\{\bar{w}^n\}_{n \in \mathbb{N}}$, and $\{w^n\}_{n \in \mathbb{N}}$ be defined as in Construction 5.3. Then

$$\sup_{n \in \mathbb{N}} E \left[\int_0^1 \sqrt{\varkappa(n) n} \, \|\bar{w}^n(t)\| \, dt \right] = \sup_{n \in \mathbb{N}} E \left[\int_0^1 \|w^n(t)\| \, dt \right] < \infty.$$

In addition, $\{\bar{M}^n\}$ is tight (as a sequence of random probability measures) and uniformly integrable in the sense that

$$\lim_{C \to \infty} \limsup_{n \to \infty} E \left[\int_{\mathbb{R}^d \times [0,1]} \|w\| \, 1_{\{\|w\| \geq C\}} \bar{M}^n(dw \times dt) \right] = 0. \qquad (5.18)$$

Proof We assume without loss of generality that $\inf_{n \in \mathbb{N}} \{\sqrt{\varkappa(n) n}\} > 0$. Let $B \in (0, \infty)$ be such that $B \leq \lambda_{DA} \inf_{n \in \mathbb{N}} \{\sqrt{\varkappa(n) n}\}$, so that $\lambda_{DA} \geq B / \sqrt{\varkappa(n) n}$ for all n. Recall L_c from (5.9), and let $\bar{K} \doteq \lambda_{DA} K_{DA} + K_A/2$, where we recall that K_A is the bound on $A(x)$ and K_{DA} was introduced in (5.8). Let e_i denote the standard unit vectors in $\mathbb{R}^d$. Then for all $i \in \{1, \ldots, d\}$ and each choice of $\pm$,

$$\varkappa(n)nL_c(x, \beta)$$

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

$$\geq \pm\sqrt{\varkappa(n)n} \left\langle \frac{B}{\sqrt{\varkappa(n)n}}e_i, \sqrt{\varkappa(n)n}\beta \right\rangle - \varkappa(n)nH_c \left(x, \pm\frac{B}{\sqrt{\varkappa(n)n}}e_i \right)$$

$$\geq \pm B\sqrt{\varkappa(n)n}\beta_i - \frac{1}{2}B^2 \|A(x)\| - B^2\lambda_{DA}K_{DA}$$

$$\geq \pm B\sqrt{\varkappa(n)n}\beta_i - B^2\bar{K},$$

where the first inequality follows from making a specific choice of α and the second uses (5.8). If we multiply both sides by $|\beta_i|$, sum on i, and then divide by $\sum_{i=1}^d |\beta_i|$, we obtain

$$d\varkappa(n)nL_c(x, \beta) + dB^2\bar{K} \geq Bd\sqrt{\varkappa(n)n}\frac{\|\beta\|^2}{\sum_{i=1}^d |\beta_i|} \geq B\sqrt{\varkappa(n)n}\|\beta\|. \quad (5.19)$$

To slightly simplify the notation, we let $s^n(t) \doteq \lfloor nt \rfloor / n$, where $\lfloor a \rfloor$ is the integer part of a. Using the bound relating L_c and relative entropy from Lemma 5.4 together with (5.17), we obtain

$$d\left(\frac{K_E}{B} + B\bar{K} \right) \geq \frac{d\varkappa(n)n}{B} E \left[\int_0^1 L_c \left(\bar{X}^n \left(s^n(t)\right), \bar{w}^n(t) \right) dt \right] + dB\bar{K}$$

$$\geq E \left[\int_0^1 \sqrt{\varkappa(n)n}\|\bar{w}^n(t)\| dt \right], \quad (5.20)$$

which proves the first statement in the theorem. Since by Theorem 2.10 the mapping

$$m \mapsto \int_{\mathbb{R}^d \times [0,1]} \|w\| m(dw \times dt)$$

defines a tightness function on $\mathscr{P}\left(\mathbb{R}^d \times [0, 1]\right)$, it follows from Lemma 2.9 and the first claim that $\{\bar{M}^n\}_{n \in \mathbb{N}}$ is tight.

For the uniform integrability, let $C \in (1, \infty)$ be arbitrary. We note that the estimates in (5.19) and (5.20) hold for any B and n such that $B \leq \lambda_{DA}\{\sqrt{\varkappa(n)n}\}$. Consider n large enough that

$$\min\{\lambda_{DA}, 1\} \geq \frac{C}{\sqrt{\varkappa(n)n}}.$$

Then for such n, the estimates (5.19) and (5.20) hold with $B = 1$ and $B = C$. Recalling $\bar{K} \doteq \lambda_{DA}K_{DA} + K_A/2$ and applying (5.20) with $B = 1$, we have for such n,

$$E\left[\int_0^1 \sqrt{\varkappa(n)n}\,\|\bar{w}^n(s)\|\,ds\right] \le K^* \doteq d\left(K_E + \frac{1}{2}K_A + \lambda_{DA}K_{DA}\right),$$

and therefore

$$E\left[\int_0^1 1_{\{\sqrt{\varkappa(n)n}\|\bar{w}^n(s)\|>C^2\}}ds\right] \le \frac{K^*}{C^2}.$$

In the following bound, (5.19) with $C = B$ is used to get the first inequality and the last display, and (5.20) with $B = 1$ for the third inequality:

$$CE\left[\int_{\mathbb{R}^d \times [0,1]} \|w\|\,1_{\{\|w\|\ge C^2\}}\bar{M}^n(dw \times dt)\right]$$

$$\le E\left[d\int_0^1 1_{\{\|w^n(s)\|>C^2\}}\left(\varkappa(n)nL_c\left(\bar{X}^n\left(s^n(t)\right),\bar{w}^n(t)\right)+C^2\bar{K}\right)dt\right]$$

$$\le d\varkappa(n)nE\left[\int_0^1 L_c\left(\bar{X}^n\left(s^n(t)\right),\bar{w}^n(t)\right)dt\right]+C^2d\bar{K}E\left[\int_0^1 1_{\{\|w^n(t)\|>C^2\}}dt\right]$$

$$\le K^*d\left(1+\bar{K}\right).$$

This proves the claimed uniform integrability. $\qquad\square$

5.3.2 Identification of Limits

The following theorem is a law of large numbers type result for the difference between the noises and their conditional means, and is the most complicated part of the analysis.

Theorem 5.6 *Assume Condition 5.1 and (5.17). Consider the sequence $\{\bar{v}_i^n\}$ of controlled noises and $\{\bar{w}^n(i/n)\}$ of means of the controlled noises as in Construction 5.3. For $i \in \{1, \ldots, n\}$ let*

$$W_i^n \doteq \frac{1}{n}\sum_{j=0}^{i-1}\sqrt{\varkappa(n)n}\left(\bar{v}_j^n - \bar{w}^n\left(j/n\right)\right).$$

Then for all $\delta > 0$,

$$\lim_{n\to\infty} P\left\{\max_{i\in\{1,\ldots,n\}}\left\|W_i^n\right\| \ge \delta\right\} = 0.$$

Proof According to (5.17),

$$\frac{1}{n}\sum_{i=0}^{n-1}E[R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))] \le \frac{K_E}{\varkappa(n)n}.$$

Because of this the (random) Radon–Nikodym derivatives,

$$f_i^n(y) = \frac{d\bar{\mu}_i^n(\cdot)}{d\theta(\cdot|\bar{X}_i^n)}(y)$$

are well defined and can be selected in a measurable way [79, Theorem V.58]. We will control the magnitude of the noise when the Radon–Nikodym derivative is large by bounding

$$\frac{1}{n}\sum_{i=0}^{n-1} E[1_{\{f_i^n(\bar{v}_i^n)\geq r\}}\|\bar{v}_i^n\|]$$

for large $r \in (0, \infty)$.

From the bound on the moment-generating function (5.3) [see (3.12)], we obtain

$$\sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d} e^{\frac{\lambda}{\sqrt{d}}\|y\|}\theta(dy|x) \leq 2de^{K_{\mathrm{mgf}}}.$$

Let $\sigma = \min\{\lambda/2\sqrt{d}, 1\}$ and recall $\ell(b) \doteq b\log b - b + 1$. Then

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}}\|\bar{v}_i^n\|\right] = \frac{1}{n}\sum_{i=0}^{n-1} E\left[\int_{\{y:f_i^n(y)\geq r\}}\|y\|\,f_i^n(y)\theta(dy|\bar{X}_i^n)\right],$$

and the bound $ab \leq e^a + \ell(b)$ for $a, b \geq 0$ with $a = \sigma\|y\|$ and $b = f_i^n(y)$ gives that for all i,

$$E\left[\int_{\{y:f_i^n(y)\geq r\}}\|y\|\,f_i^n(y)\theta(dy|\bar{X}_i^n)\right]$$
$$\leq \frac{1}{\sigma}E\left[\int_{\{y:f_i^n(y)\geq r\}} e^{\sigma\|y\|}\theta(dy|\bar{X}_i^n)\right] + \frac{1}{\sigma}E\left[\int_{\{y:f_i^n(y)\geq r\}}\ell(f_i^n(y))\bar{\mu}_i^n(dy)\right].$$

Since $\ell(b) \geq 0$ for all $b \geq 0$, we have

$$E\left[\int_{\{y:f_i^n(y)\geq r\}}\ell\left(f_i^n(y)\right)\theta(dy|\bar{X}_i^n)\right] \leq E\left[\int_{\mathbb{R}^d}\ell(f_i^n(y))\theta(dy|\bar{X}_i^n)\right]$$
$$= E[R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))],$$

and by Hölder's inequality,

$$E\left[\int_{\{y:f_i^n(y)\geq r\}} e^{\sigma\|y\|}\theta(dy|\bar{X}_i^n)\right]$$

$$\leq E\left[\left(\int_{\mathbb{R}^d} 1_{\{f_i^n(y)\geq r\}}\theta(dy|\bar{X}_i^n)\right)^{1/2}\left(\int_{\mathbb{R}^d} e^{2\sigma\|y\|}\theta(dy|\bar{X}_i^n)\right)^{1/2}\right]$$

$$= E\left[\theta(\{y:f_i^n(y)\geq r\}|\bar{X}_i^n)^{1/2}\right]\left(2de^{K_{\mathrm{mgf}}}\right)^{1/2}.$$

In addition, for all $r > 1$, Markov's inequality gives

$$\theta(\{y:f_i^n(y)\geq r\}|\bar{X}_i^n) \leq \frac{1}{r\log r}\int \log(f_i^n(y))f_i^n(y)\theta(dy|\bar{X}_i^n)$$

$$= \frac{R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))}{r\log r}.$$

The last four displays give the bound

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[\int_{\{f_i^n(y)\geq r\}} \|y\|\, f_i^n(y)\theta(dy|\bar{X}_i^n)\right]$$

$$\leq \frac{1}{\sigma}\left(2de^{K_{\mathrm{mgf}}}\right)^{1/2}\frac{1}{n}\sum_{i=0}^{n-1} E\left[\left(\frac{R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))}{r\log r}\right)^{1/2}\right]$$

$$+ \frac{1}{\sigma}\frac{1}{n}\sum_{i=0}^{n-1} E[R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))].$$

Since by Jensen's inequality,

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[\left(\frac{R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))}{r\log r}\right)^{1/2}\right]$$

$$\leq \left(\frac{1}{r\log r}\right)^{1/2}\left(\frac{1}{n}\sum_{i=0}^{n-1} E[R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))]\right)^{1/2},$$

we obtain the overall bound

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}}\|\bar{v}_i^n\|\right]$$

$$\leq \frac{1}{\sigma}\left(2de^{K_{\mathrm{mgf}}}\right)^{1/2}\left(\frac{1}{r\log r}\right)^{1/2}\left(\frac{1}{n}\sum_{i=0}^{n-1} E[R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))]\right)^{1/2}$$

$$+ \frac{1}{\sigma}\frac{1}{n}\sum_{i=0}^{n-1} E[R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))]$$

$$\leq \frac{1}{\sigma} \frac{K_E^{1/2}}{\sqrt{\varkappa(n)n}} \left(2de^{K_{\mathrm{mgf}}}\right)^{1/2} \left(\frac{1}{r \log r}\right)^{1/2} + \frac{1}{\sigma} \frac{K_E}{\varkappa(n)n}. \tag{5.21}$$

Using this result, we can complete the proof. Define

$$\xi_i^{n,r} \doteq \begin{cases} \bar{v}_i^n & \text{if } f_i^n(\bar{v}_i^n) < r, \\ 0 & \text{otherwise.} \end{cases}$$

For all $\delta > 0$,

$$P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n}\left(\bar{v}_i^n - \bar{w}^n\left(\frac{i}{n}\right)\right)\right\| \geq 3\delta\right\}$$

$$\leq P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n}(\bar{v}_i^n - \xi_i^{n,r})\right\| \geq \delta\right\}$$

$$+ P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n}\left(\xi_i^{n,r} - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\| \geq \delta\right\}$$

$$+ P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n}\left(\bar{w}^n\left(\frac{i}{n}\right) - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\| \geq \delta\right\}.$$

The first term satisfies

$$P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n}(\bar{v}_i^n - \xi_i^{n,r})\right\| \geq \delta\right\}$$

$$\leq \frac{1}{\delta} \sqrt{\varkappa(n)n} \frac{1}{n} \sum_{i=0}^{n-1} E\left[\left\|\bar{v}_i^n - \xi_i^{n,r}\right\|\right]$$

$$= \frac{1}{\delta} \sqrt{\varkappa(n)n} \frac{1}{n} \sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}} \left\|\bar{v}_i^n\right\|\right].$$

The norm in the second term is a submartingale in k, and so by Doob's submartingale inequality [see (D.1)],

$$P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n}\left(\xi_i^{n,r} - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\| \geq \delta\right\}$$

$$\leq \frac{1}{\delta^2} E\left[\left\|\frac{1}{n} \sum_{i=0}^{n-1} \sqrt{\varkappa(n)n}\left(\xi_i^{n,r} - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\|^2\right]$$

$$= \frac{1}{\delta^2} \frac{\varkappa(n)}{n} \sum_{i=0}^{n-1} E\left[\left\|\left(\xi_i^{n,r} - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\|^2\right]$$

$$\leq \frac{1}{\delta^2} \frac{\varkappa(n)}{n} \sum_{i=0}^{n-1} E\left[\|\xi_i^{n,r}\|^2\right]$$

$$= \frac{1}{\delta^2} \frac{\varkappa(n)}{n} \sum_{i=0}^{n-1} E\left[\int_{\{y:f_i^n(y)<r\}} \|y\|^2 f_i^n(y)\theta(dy|\bar{X}_i^n)\right]$$

$$\leq \frac{r}{\delta^2} \frac{\varkappa(n)}{n} \sum_{i=0}^{n-1} E\left[\int_{\mathbb{R}^d} \|y\|^2 \theta(dy|\bar{X}_i^n)\right]$$

$$\leq \frac{r}{\delta^2} \varkappa(n) K_{\mu,2},$$

where

$$K_{\mu,2} = \sup_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d} \|y\|^2 \theta(dy|x) < \infty,$$

and the finiteness is due to (5.3). We can use Jensen's inequality with the third term and get the same bound that was proved for the first:

$$P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n}\sum_{i=0}^{k} \sqrt{\varkappa(n)n}\left(\bar{w}^n\left(\frac{i}{n}\right) - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\| \geq \delta\right\}$$

$$\leq \frac{1}{\delta}\sqrt{\varkappa(n)n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[\left\|\left(\bar{w}^n\left(\frac{i}{n}\right) - \int_{\{y:f_i^n(y)<r\}} y\bar{\mu}_i^n(dy)\right)\right\|\right]$$

$$= \frac{1}{\delta}\sqrt{\varkappa(n)n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[\left\|\int_{\{y:f_i^n(y)\geq r\}} y\bar{\mu}_i^n(dy)\right\|\right]$$

$$\leq \frac{1}{\delta}\sqrt{\varkappa(n)n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[\int_{\{y:f_i^n(y)\geq r\}} \|y\| \bar{\mu}_i^n(dy)\right]$$

$$= \frac{1}{\delta}\sqrt{\varkappa(n)n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}} \|\bar{v}_i^n\|\right].$$

Combining the bounds for these three terms with (5.21) gives

$$P\left\{\max_{k=0,\ldots,n-1} \left\|\frac{1}{n}\sum_{i=0}^{k} \sqrt{\varkappa(n)n}\left(\bar{v}_i^n - \bar{w}^n\left(\frac{i}{n}\right)\right)\right\| \geq 3\delta\right\}$$

$$\leq \frac{2}{\delta}\sqrt{\varkappa(n)n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}} \|\bar{v}_i^n\|\right] + \frac{r}{\delta^2}\varkappa(n)K_{\mu,2}$$

$$\leq \frac{2}{\sigma \delta} K_E^{1/2} \left(2de^{K_{\mathrm{mgf}}}\right)^{1/2} \left(\frac{1}{r \log r}\right)^{1/2} + \frac{2}{\sigma \delta} \frac{K_E}{\sqrt{\varkappa(n)n}} + \varkappa(n) \frac{r}{\delta^2} K_{\mu,2}.$$

Sending $n \to \infty$ and then $r \to \infty$ (and using $\varkappa(n) \to 0$, $\varkappa(n)n \to \infty$) gives

$$P \left\{ \max_{k=0,\dots,n-1} \left\| \frac{1}{n} \sum_{i=0}^{k} \sqrt{\varkappa(n)n} \left(\bar{v}_i^n - \bar{w}^n \left(\frac{i}{n}\right) \right) \right\| \geq 3\delta \right\} \to 0$$

as $n \to \infty$, which completes the proof. $\square$

The next result identifies the weak limits of controlled processes. We recall that for a probability measure γ on $\mathbb{R}^d \times [0, 1]$, the marginal distribution on the second coordinate is denoted by $[\gamma]_2$, and the conditional distribution on the first coordinate given the second is given by $[\gamma]_{1|2}$. Thus for Borel sets A and B,

$$\gamma(\mathbb{R}^d \times B) = [\gamma]_2(B) \text{ and } \gamma(A \times B) = \int_B [\gamma]_{1|2}(A|s)[\gamma]_2(ds).$$

Theorem 5.7 *Let $\{\bar{\mu}_i^n\}_{i=1,\dots,n}$ be a sequence of controls, and define the corresponding random variables as in Construction 5.3. Assume Condition 5.1 and that (5.17) holds for some $K_E < \infty$. Then $\{(\bar{M}^n, \bar{Y}^n)\}_{n \in \mathbb{N}}$ is tight in $\mathscr{P}(\mathbb{R}^d \times [0, 1]) \times \mathscr{C}([0, 1] : \mathbb{R}^d)$. Consider a subsequence (keeping the index n for convenience) such that $\{(\bar{M}^n, \bar{Y}^n)\}$ converges weakly to $(\bar{M}, \bar{Y})$. Then with probability 1, $[\bar{M}]_2(dt)$ is Lebesgue measure and*

$$\bar{Y}(t) = \int_0^t Db(X^0(s))\bar{Y}(s)ds + \int_0^t w(s)ds, \tag{5.22}$$

where

$$w(t) = \int_{\mathbb{R}^d} w[\bar{M}]_{1|2}(dw\,|t). \tag{5.23}$$

In addition,

$$\liminf_{n \to \infty} \varkappa(n)nE \left[\frac{1}{n} \sum_{i=0}^{n-1} R(\bar{\mu}_i^n(\cdot) \,\|\, \theta(\cdot|\bar{X}_i^n)) \right] \geq E \left[\int_0^1 \frac{1}{2} \|w(s)\|_{A^{-1}(X^0(s))}^2 \, ds \right]. \tag{5.24}$$

The proof of this theorem is lengthy. After some preliminary discussion, several lemmas will be presented. After stating and proving the lemmas, we will return to complete the argument for Theorem 5.7.

It was shown in Theorem 5.5 that $\{\bar{M}^n\}_{n \in \mathbb{N}}$ is tight. If $\bar{M}$ is any weak limit of a subsequence of $\{\bar{M}^n\}_{n \in \mathbb{N}}$, then since for all n the second marginal of $\bar{M}^n(dw \times dt)$ is Lebesgue measure, it follows that $[\bar{M}]_2(dt)$ is Lebesgue measure with probability 1.

The ultimate goal is to show that $\bar{Y}^n \to \bar{Y}$ weakly in $\mathscr{C}([0, 1] : \mathbb{R}^d)$, where $\bar{Y}(t)$ is given by (5.22) in terms of the weak limit $\bar{M}$. To achieve this, we introduce the following processes, which serve as intermediate steps. Let $\check{Y}_0^n = 0$ and

$$\check{Y}_{i+1}^n = \check{Y}_i^n + \sqrt{\frac{\varkappa(n)}{n}} \left(b\left(X_i^{n,0} + \frac{1}{\sqrt{\varkappa(n)n}}\check{Y}_i^n\right) - b\left(X_i^{n,0}\right)\right) + \sqrt{\frac{\varkappa(n)}{n}}\bar{w}^n\left(\frac{i}{n}\right),$$

together with its continuous time linear interpolation defined for $t \in [i/n, i/n + 1/n]$ by

$$\check{Y}^n(t) = (i + 1 - nt)\check{Y}_i^n + (nt - i)\check{Y}_{i+1}^n.$$

Also let

$$\hat{Y}^n(t) = \int_0^t Db\left(X^0(s)\right)\hat{Y}^n(s)ds + \int_0^t w^n(s)ds, \tag{5.25}$$

where

$$w^n(t) = \int_{\mathbb{R}^d} w[\bar{M}]_{1|2}^n(dw\,|t)$$

as in Construction 5.3. Then both $\check{Y}^n$ and $\hat{Y}^n$ are random variables taking values in $\mathscr{C}([0, 1] : \mathbb{R}^d)$. Note that $\bar{Y}^n$ differs from $\check{Y}^n$, because $\bar{Y}^n$ is driven by the actual noises and $\check{Y}^n$ is driven by their conditional means. While the driving terms of $\hat{Y}^n$ and $\check{Y}^n$ are the same [recall that $\sqrt{\varkappa(n)n}\bar{w}^n(t) = w^n(t)$], they differ in that $\check{Y}^n$ is still a linear interpolation of a discrete time process, whereas $\hat{Y}^n$ satisfies an ODE. We will show that along any subsequence where $\bar{M}^n \to \bar{M}$ weakly,

$$\bar{Y}^n - \check{Y}^n \to 0, \; \check{Y}^n - \hat{Y}^n \to 0, \; \text{and} \; \hat{Y}^n \to \bar{Y}$$

in $\mathscr{C}([0, 1] : \mathbb{R}^d)$, all in distribution, where $\bar{Y}$ is the unique solution of (5.22).

To show that $\hat{Y}^n \to \bar{Y}$, we show that $\{\hat{Y}^n\}$ is tight in $\mathscr{C}([0, 1] : \mathbb{R}^d)$ and use the mapping defined by (5.25) from $\int_0^\cdot w^n$ to $\hat{Y}^n$. Recall that $\sup_{x \in \mathbb{R}^d} \|Db(x)\| \le K_b$. The following lemma uses the uniform integrability of $\{\bar{M}^n\}$ given in Theorem 5.5 to prove tightness of $\{\hat{Y}^n\}$.

Lemma 5.8 *Assume Conditions 5.1 and (5.17). The sequence $\{\hat{Y}^n\}$ defined in (5.25) in terms of the measures $\{\bar{M}^n\}$ via Construction 5.3 is tight in $\mathscr{C}([0, 1] : \mathbb{R}^d)$, as is $\{\int_0^\cdot w^n ds\}$.*

Proof Tightness of $\{\int_0^\cdot w^n ds\}$ is a consequence of the fact that for $\delta, C \in (0, \infty)$,

$$\limsup_{n\to\infty} P\left(\sup_{|s-t|\le\delta}\int_s^t \|w^n(r)\|\,dr > \varepsilon\right) \le \delta\frac{C}{\varepsilon} + \frac{1}{\varepsilon}T(C),$$

where

$$T(C) \doteq \limsup_{n \to \infty} E \left[\int_0^1 1_{\{\|w^n(t)\| > C\}} \left\| w^n(t) \right\| dt \right]$$

$$= \limsup_{n \to \infty} E \left[\int_{\{\|w\| > C\}} \|w\| \, \bar{M}^n(dw \times dt) \right],$$

and the fact that by Theorem 5.5, $T(C) \to 0$ as $C \to \infty$. For tightness of $\{\hat{Y}^n\}$ it suffices to check that the map from $\mathscr{C}([0, 1] : \mathbb{R}^d)$ to itself that takes $z \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ to the unique solution of

$$\phi(t) = \int_0^t Db(X^0(s))\phi(s)ds + z(t), \ t \in [0, 1],$$

is continuous. However, this continuity follows directly from Gronwall's inequality. $\square$

We still need to show that $\hat{Y}^n$ converges to $\bar{Y}$. This also relies on the uniform integrability given by Theorem 5.5.

Lemma 5.9 *Assume Conditions 5.1 and (5.17). Let the sequence $\{\hat{Y}^n(t)\}$ be defined by (5.25) and consider a weakly convergent subsequence $\{(\hat{Y}^n, \bar{M}^n)\}$ with limit $(\hat{Y}, \bar{M})$. Then w.p.1, $\hat{Y} = \bar{Y}$, where $\bar{Y}$ is defined by (5.22)–(5.23).*

Proof We can write

$$\hat{Y}^n(t) = \int_0^t Db(X^0(s))\hat{Y}^n(s)ds + \int_0^t \int_{\mathbb{R}^d} w \bar{M}^n(dw \times ds).$$

The uniform integrability proved in Theorem 5.5 and that $[\bar{M}]_2$ is Lebesgue measure w.p.1 will be used. The latter implies $E\bar{M}(\mathbb{R}^d \times \{t\}) = 0$ for $t \in [0, 1]$. Sending $n \to \infty$ and using the definition of $w(s)$ in (5.23) gives

$$\hat{Y}(t) = \int_0^t Db(X^0(s))\hat{Y}(s)ds + \int_0^t \int_{\mathbb{R}^d} w \bar{M}(dw \times ds)$$

$$= \int_0^t Db(X^0(s))\hat{Y}(s)ds + \int_0^t w(s)ds.$$

By uniqueness of the solution, $\hat{Y} = \bar{Y}$ follows. $\square$

It remains to show that $\bar{Y}^n - \check{Y}^n \to 0$ and $\check{Y}^n - \hat{Y}^n \to 0$. We begin with $\bar{Y}^n - \check{Y}^n \to 0$. Recall that the difference between $\bar{Y}^n$ and $\check{Y}^n$ is that the first is driven by the actual noises, and the second is driven by their conditional means. The following discrete version of Gronwall's inequality will be used to prove $\bar{Y}^n - \check{Y}^n \to 0$. A proof can be found in [83, p. 283].

Lemma 5.10 *If $\{z_n\}$, $\{u_n\}$, and $\{v_n\}$ are nonnegative sequences defined for $n \in \mathbb{N}_0$ that satisfy*

$$z_k \le v_k + \sum_{k=0}^{n-1} u_i z_i,$$

then

$$z_k \le v_k + \sum_{k=0}^{n-1} u_i v_i \exp\left\{\sum_{i=k+1}^{n-1} u_j\right\}.$$

Lemma 5.11 *Assume Conditions 5.1 and (5.17). Then* $\check{Y}^n - \bar{Y}^n \to 0$ *in probability.*

Proof Recall from (5.13) and (5.14) that

$$\bar{Y}_k^n = \sum_{i=0}^{k-1}\sqrt{\frac{\varkappa(n)}{n}}\left(b\left(X_i^{n,0} + \frac{1}{\sqrt{\varkappa(n)n}}\bar{Y}_i^n\right) - b\left(X_i^{n,0}\right)\right) + \sum_{i=0}^{k-1}\sqrt{\frac{\varkappa(n)}{n}}\bar{v}_i^n$$

and

$$\check{Y}_k^n = \sum_{i=0}^{k-1}\sqrt{\frac{\varkappa(n)}{n}}\left(b\left(X_i^{n,0} + \frac{1}{\sqrt{\varkappa(n)n}}\check{Y}_i^n\right) - b\left(X_i^{n,0}\right)\right) + \sum_{i=0}^{k-1}\sqrt{\frac{\varkappa(n)}{n}}\bar{w}^n\left(\frac{i}{n}\right),$$

so with W_k^n defined as in Theorem 5.6,

$$\left\|\bar{Y}_k^n - \check{Y}_k^n\right\| \le \left\|W_k^n\right\| + \sum_{i=0}^{k-1}\frac{K_b}{n}\left\|\bar{Y}_i^n - \check{Y}_i^n\right\|.$$

Using Lemma 5.10 gives, for $k \le n$,

$$\left\|\bar{Y}_k^n - \check{Y}_k^n\right\| \le \left\|W_k^n\right\| + \sum_{i=0}^{k-1}\left\|W_i^n\right\|\frac{K_b}{n}\exp\left\{\frac{K_b}{n}(k-i-1)\right\}$$

$$\le (1 + K_b e^{K_b})\max_{i\in\{0,1,\ldots,k\}}\left\|W_i^n\right\|.$$

From Theorem 5.6, we have $\max_{i\in\{1,\ldots,n\}}\left\|W_i^n\right\| \to 0$ in probability, and therefore

$$\max_{i\in\{1,\ldots,n\}}\left\|\bar{Y}_i^n - \check{Y}_i^n\right\| \to 0,$$

and hence $\sup_{t\in[0,1]}\left\|\bar{Y}^n(t) - \check{Y}^n(t)\right\| \to 0$ in probability.   □

To complete the proof of the convergence we need to show that $\check{Y}^n - \hat{Y}^n \to 0$. Recall that these two processes have the same driving terms but different drifts, in that $\hat{Y}^n$ satisfies the ODE

$$\hat{Y}^n(t) = \int_0^t Db(X^0(s))\hat{Y}^n(s)ds + \int_0^t w^n(s)ds,$$

while $\check{Y}^n$ is the linear interpolation of the discrete time process defined by $\check{Y}_0^n = 0$ and

$$\check{Y}_{i+1}^n = \check{Y}_i^n + \sqrt{\frac{\varkappa(n)}{n}} \left(b\left(X_i^{n,0} + \frac{1}{\sqrt{\varkappa(n)n}}\check{Y}_i^n\right) - b\left(X_i^{n,0}\right) \right) + \frac{1}{n}w^n\left(\frac{i}{n}\right).$$

However, essentially the same arguments as those used in Lemma 5.8 to show tightness of $\{\hat{Y}^n\}$ can be used to prove tightness of $\{\check{Y}^n\}$, and then it easily follows as in Lemma 5.9 that any limit will satisfy the same ODE (5.22) as the limit of $\{\hat{Y}^n\}$, and therefore $\check{Y}^n - \hat{Y}^n \to 0$ follows.

Combining $\bar{Y}^n - \check{Y}^n \to 0$, $\check{Y}^n - \hat{Y}^n \to 0$, and $\hat{Y}^n \to \bar{Y}$ demonstrates that along the subsequence where $\bar{M}^n \to \bar{M}$ weakly, $\bar{Y}^n \to \bar{Y}$ in distribution, which implies that along this subsequence, $(\bar{M}^n, \bar{Y}^n) \to (\bar{M}, \bar{Y})$ weakly. We have already shown that with probability 1, $[\bar{M}]_2(dt)$ is Lebesgue measure and

$$\bar{Y}(t) = \int_0^t Db(X^0(s))\bar{Y}(s)ds + \int_0^t \int_{\mathbb{R}^d} w[\bar{M}]_{1|2}(dw\,|s)ds,$$

so the proof of convergence (i.e., the first part of Theorem 5.7) is complete.

To finish the proof of Theorem 5.7, we must prove the bound (5.24). Recall the notation $s^n(t) \doteq \lfloor nt \rfloor /n$, and note from (5.14) that the weak convergence of $\bar{Y}^n$ implies

$$\sup_{t\in[0,1]} \left\| \bar{X}^n(s^n(t)) - X^0(t) \right\| \to 0 \text{ in probability.} \qquad (5.26)$$

Now define random measures on $\mathbb{R}^d \times \mathbb{R}^d \times [0, 1]$ by

$$\gamma^n(dx \times dw \times dt) = \delta_{\bar{X}^n(s^n(t))}(dx)\,\bar{M}^n(dw \times dt).$$

Note that the tightness of $\{\gamma^n\}$ follows easily from (5.26) and the tightness of $\{\bar{M}^n\}$. Thus given any subsequence, we can choose a further subsequence (again we will retain n as the index for simplicity) along which $\{\gamma^n\}$ converges weakly to some $\mathscr{P}\left(\mathbb{R}^d \times \mathbb{R}^d \times [0, 1]\right)$-valued random variable γ with

$$[\gamma]_{2,3}(dw \times dt) = \bar{M}(dw \times dt),$$

where $[\gamma]_{2,3}$ is the second and third marginal of γ. If we establish (5.24) for this subsequence, it holds for the original one using a standard argument by contradiction. For $\sigma > 0$, let

$$G_\sigma^{X^0} \doteq \left\{(x, w, t) : \left\| x - X^0(t) \right\| \leq \sigma\right\}$$

be closed sets centered on $X^0(t)$ in the x variable, and note that by (5.26) and weak convergence, for all $\sigma > 0$,

$$1 = \limsup_{n \to \infty} E\left[\gamma^n\left(G_\sigma^{X^0}\right)\right] \le E\left[\gamma\left(G_\sigma^{X^0}\right)\right].$$

Thus

$$E\left[\gamma\left(\bigcap_{n \in \mathbb{N}} G_{1/n}^{X^0}\right)\right] = 1,$$

so with probability 1, γ puts all its mass on $\{(x, w, t) : x = X^0(t)\}$. Therefore, with probability 1, for a.e. (w, t) under $[\gamma]_{2,3}(dw \times dt)$,

$$[\gamma]_{1|2,3}(dx \,|\, w, t) = \delta_{X^0(t)}(dx).$$

Combined with the fact that the second marginal of $\bar{M}(dw \times dt)$ is Lebesgue measure, this gives

$$\gamma(dx \times dw \times dt) = \delta_{X^0(t)}(dx)\,\bar{M}(dw|t)\,dt. \qquad (5.27)$$

For $\kappa \in (0, \infty)$, define

$$\bar{L}_\kappa(x, \beta) \doteq \sup_{\alpha \in \mathbb{R}^d}\left[\langle \alpha, \beta \rangle - \frac{1}{2}\|\alpha\|_{A(x)}^2 - \frac{1}{2\kappa}\|\alpha\|^2\right].$$

Using (5.8),

$$
\begin{aligned}
\varkappa(n)n L_c&\left(x, \frac{1}{\sqrt{\varkappa(n)n}}\beta\right)\\
&= \sup_{\alpha \in \mathbb{R}^d}\left[\sqrt{\varkappa(n)n}\langle \alpha, \beta \rangle - \varkappa(n)n H_c(x, \alpha)\right]\\
&\ge \sup_{\alpha \in \mathbb{R}^d}\left[\sqrt{\varkappa(n)n}\langle \alpha, \beta \rangle - \frac{\varkappa(n)n}{2}\|\alpha\|_{A(x)}^2 - \varkappa(n)n K_{DA}\|\alpha\|^3\right]\\
&\ge \sup_{\alpha \in \mathbb{R}^d}\left[\langle \alpha, \beta \rangle - \|\alpha\|_{A(x)}^2 - \frac{1}{2\kappa}\|\alpha\|^2 - \frac{K_{DA}}{\sqrt{\varkappa(n)n}}\|\alpha\|^3\right]. \qquad (5.28)
\end{aligned}
$$

Let K_1 be an arbitrary compact subset of $\mathbb{R}^d$. Since $\|\alpha\|^2$ is superlinear, there exists another compact set K_2 of $\mathbb{R}^d$, depending only on κ and K_1, such that whenever $\beta \in K_1$ and $x \in \mathbb{R}^d$,

$$
\begin{aligned}
\sup_{\alpha \in K_2}\left[\langle \alpha, \beta \rangle - \frac{1}{2}\|\alpha\|_{A(x)}^2 - \frac{1}{2\kappa}\|\alpha\|^2\right] &= \sup_{\alpha \in \mathbb{R}^d}\left[\langle \alpha, \beta \rangle - \frac{1}{2}\|\alpha\|_{A(x)}^2 - \frac{1}{2\kappa}\|\alpha\|^2\right]\\
&= \bar{L}_\kappa(x, \beta).
\end{aligned}
$$

Also, from (5.28),

$$\Lambda^n_{K_2}(x, \beta) \doteq \sup_{\alpha \in K_2} \left[\langle \alpha, \beta \rangle - \|\alpha\|^2_{A(x)} - \frac{1}{2\kappa} \|\alpha\|^2 - \frac{K_{DA}}{\sqrt{\varkappa(n)n}} \|\alpha\|^3 \right]$$

$$\leq \varkappa(n)n L_c \left(x, \frac{1}{\sqrt{\varkappa(n)n}} \beta \right)$$

Note that as $n \to \infty$,

$$\sup_{(x, \beta) \in \mathbb{R}^d \times K_1} |\Lambda^n_{K_2}(x, \beta) - \bar{L}_\kappa(x, \beta)| \to 0. \tag{5.29}$$

By Lemma 5.4 and the definitions of γ^n and $\bar{M}^n$, we have

$$\liminf_{n \to \infty} \varkappa(n) n E \left[\sum_{i=0}^{n-1} \frac{1}{n} R(\bar{\mu}^n_i(\cdot) \| \theta(\cdot | \bar{X}^n_i)) \right]$$

$$\geq \liminf_{n \to \infty} E \left[\int_{\mathbb{R}^d \times \mathbb{R}^d \times [0,1]} \varkappa(n) n L_c \left(x, \frac{1}{\sqrt{\varkappa(n)n}} w \right) \gamma^n \, (dx \times dw \times dt) \right]$$

$$\geq \liminf_{n \to \infty} E \left[\int_{\mathbb{R}^d \times K_1 \times [0,1]} \Lambda^n_{K_2}(x, w) \gamma^n \, (dx \times dw \times dt) \right].$$

For fixed K_1, since K_2 is bounded, there is $c \in (0, \infty)$ such that $|\Lambda^n_{K_2}(x, w)| \leq c(1 + \|w\|)$ for all $n \in \mathbb{N}$ and also $|\bar{L}_\kappa(x, w)| \leq c(1 + \|w\|)$. Using these bounds and (5.18) to control contributions to the integrals from large values of $\|w\|$, it follows from (5.29) that the last quantity in the previous display is the same as

$$\liminf_{n \to \infty} E \left[\int_{\mathbb{R}^d \times K_1 \times [0,1]} \bar{L}_\kappa(x, w) \gamma^n \, (dx \times dw \times dt) \right].$$

Using the continuity of $(x, \beta) \mapsto \bar{L}_\kappa(x, \beta)$ and Fatou's lemma thus gives

$$\liminf_{n \to \infty} \varkappa(n) n E \left[\sum_{i=0}^{n-1} \frac{1}{n} R(\bar{\mu}^n_i(\cdot) \| \theta(\cdot | \bar{X}^n_i)) \right]$$

$$\geq E \left[\int_{\mathbb{R}^d \times K_1 \times [0,1]} \bar{L}_\kappa(x, w) \gamma \, (dx \times dw \times dt) \right].$$

Since $\bar{L}_\kappa \geq 0$, by the monotone convergence theorem we can replace K_1 by $\mathbb{R}^d$ in the last display. Next note that as $\kappa \to \infty$,

$$\bar{L}_\kappa(x, \beta) \uparrow \frac{1}{2} \|\beta\|^2_{A^{-1}(x)}$$

for all $(x, \beta) \in \mathbb{R}^{2d}$. Finally, using the monotone convergence theorem, the decomposition (5.27), and Jensen's inequality in that order shows that

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

$$\geq \lim_{\kappa\to\infty} E\left[\int_{\mathbb{R}^d\times\mathbb{R}^d\times[0,1]} \bar{L}_\kappa(x, w)\,\gamma\,(dx \times dw \times dt)\right]$$

$$= E\left[\int_{\mathbb{R}^d\times\mathbb{R}^d\times[0,1]} \frac{1}{2}\|w\|_{A^{-1}(x)}^2\,\gamma\,(dx \times dw \times dt)\right]$$

$$= E\left[\int_0^1\int_{\mathbb{R}^d}\frac{1}{2}\|w\|_{A^{-1}(X^0(t))}^2\,\bar{M}\,(dw|\,t)\,dt\right]$$

$$\geq E\left[\frac{1}{2}\int_0^1\|w(t)\|_{A^{-1}(X^0(t))}^2\,dt\right],$$

which is (5.24). This concludes the proof of Theorem 5.7.   □

5.4   Laplace Upper Bound

In this section we prove the variational lower bound

$$\liminf_{n\to\infty} -\varkappa(n)\log E\left[e^{-\frac{1}{\varkappa(n)}F(Y^n)}\right] \geq \inf_{\phi\in\mathscr{C}([0,1]:\mathbb{R}^d)}[I_M(\phi) + F(\phi)],\qquad (5.30)$$

which corresponds to the Laplace upper bound.

Suppose for each n that $\{\bar{\mu}_i^n\}$ comes within $1/n$ of achieving the infimum in (5.15), so that

$$\liminf_{n\to\infty} -\varkappa(n)\log E\left[e^{-\frac{1}{\varkappa(n)}F(Y^n)}\right]$$

$$\geq \liminf_{n\to\infty} E\left[\sum_{i=0}^{n-1}\varkappa(n)R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n)) + F(\bar{Y}^n)\right].\qquad (5.31)$$

We also have

$$\sup_{n\in\mathbb{N}} \varkappa(n)nE\left[\sum_{i=0}^{n-1}\frac{1}{n}R(\bar{\mu}_i^n(\cdot)\,\|\,\theta(\cdot|\bar{X}_i^n))\right] \leq 2\|F\|_\infty + 1.$$

Consequently, (5.17) is satisfied with $K_E = 2\|F\|_\infty + 1$, and from Theorem 5.7 we can choose for every subsequence of $\{(\bar{M}^n, \bar{Y}^n)\}$ a further subsequence (we retain n as the index for convenience) along which $(\bar{M}^n, \bar{Y}^n)$ converges to $(\bar{M}, \bar{Y})$ in distribution,

$\bar{M}$, $\bar{Y}$ are related by (5.22)–(5.23), and such that (5.24) is satisfied. Combining this with (5.31) gives

$$\liminf_{n\to\infty} -\varkappa(n) \log E\left[e^{-\frac{1}{\varkappa(n)}F(Y^n)}\right]$$

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

$$\geq E\left[\int_0^1 \frac{1}{2} \|w(s)\|_{A^{-1}(X^0(s))}^2 \, ds + F(\bar{Y})\right].$$

Define ϕ^u for $u \in \mathcal{L}^2([0,1] : \mathbb{R}^d)$ by

$$\phi^u(t) = \int_0^t Db(X^0(s))\phi^u(s) ds + \int_0^t A^{1/2}(X^0(s))u(s) ds. \qquad (5.32)$$

Recalling

$$\bar{Y}(t) = \int_0^t Db(X^0(s))\bar{Y}(s) ds + \int_0^t w(s) ds,$$

it follows using the expression for I_M in (5.10) and (5.11) that

$$E\left[\int_0^1 \frac{1}{2} \|w(s)\|_{A^{-1}(X^0(s))}^2 \, ds + F(\bar{Y})\right]$$

$$\geq \inf_{u \in \mathcal{L}^2([0,1]:\mathbb{R}^d)} \left[\int_0^1 \frac{1}{2} \|u(s)\|^2 \, ds + F(\phi^u)\right]$$

$$= \inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [I_M(\phi) + F(\phi)],$$

which is the lower bound (5.30).   □

5.5 Laplace Lower Bound

In this section we prove the variational upper bound

$$\limsup_{n\to\infty} -\varkappa(n) \log E\left[e^{-\frac{1}{\varkappa(n)}F(Y^n)}\right] \leq \inf_{\phi \in \mathscr{C}([0,1]:\mathbb{R}^d)} [I_M(\phi) + F(\phi)], \qquad (5.33)$$

which is the Laplace lower bound. Note that for $u, v \in \mathcal{L}^2([0,1] : \mathbb{R}^d)$,

$$\phi^u(t) - \phi^v(t) = \int_0^t Db(X^0(s)) \left(\phi^u(s) - \phi^v(s)\right) ds$$

$$+ \int_0^t A^{1/2}(X^0(s))(u(s) - v(s)) ds.$$

Thus by Gronwall's inequality,

$$\sup_{t\in[0,1]} \left\| \phi^u(t) - \phi^v(t) \right\|$$

$$\leq e^{K_b} \int_0^1 \left\| A^{1/2}(X^0(s))(u(s) - v(s)) \right\| ds$$

$$\leq e^{K_b} \left(\int_0^1 \left\| A^{1/2}(X^0(s))(u(s) - v(s)) \right\|^2 ds \right)^{1/2}$$

$$\leq e^{K_b} K_A^{1/2} \left(\int_0^1 \| u(s) - v(s) \|^2 ds \right)^{1/2}. \tag{5.34}$$

Since $\mathscr{C}([0, 1] : \mathbb{R}^d)$ is dense in $\mathscr{L}^2([0, 1] : \mathbb{R}^d)$, the proof of the Laplace lower bound is reduced to showing that for an arbitrary $u \in \mathscr{C}([0, 1] : \mathbb{R}^d)$,

$$\limsup_{n\to\infty} -\varkappa(n) \log E \left[e^{-\frac{1}{\varkappa(n)} F(Y^n)} \right] \leq \frac{1}{2} \int_0^1 \| u(s) \|^2 ds + F(\phi^u). \tag{5.35}$$

We fix $u \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ for the remainder of the proof.

Remark 5.12 The proof of the lower bound for the moderate deviation problem differs substantially from the corresponding proof of the large deviation problem, especially in regard to the treatment of degenerate noise. For the case of large deviations, this was handled in Chap. 4 using a mollification. In the moderate deviations setting a simpler argument is possible. This is largely due to the form of the rate function, which is the same as that of the small noise diffusion model of Sect. 3.2, but with time-dependent drift $Db(X^0(t))$ and diffusion matrix $A^{1/2}(X^0(t))$. As just discussed, with this form one can find a nearly optimal trajectory for the limit variational problem of the form ϕ^u, with u continuous rather than just measurable, which greatly facilitates the construction of nearly optimal controls for the prelimit in the proof of the lower bound. This is not possible for the general model of Chap. 4, since it is not useful to view X^n there as a continuous or nearly continuous mapping on an "exogenous" noise process. In this sense, the moderate deviation problem shares some of the simplifying features of the continuous time models discussed in Sect. 3.2 and at greater length in later chapters.

We now turn to the proof of (5.35) for the given $u \in \mathscr{C}([0, 1] : \mathbb{R}^d)$. The main difficulty related to the possible degeneracy of the noise is the following. Since at the prelimit, the controlled processes $\bar{X}^n$ may be close to but not precisely equal to X^0, the range of $A(\bar{X}^n(i/n))$ can differ from that of $A(X^0(i/n))$ (at least in the degenerate case). Because of this, the construction of a control that approximates $A^{1/2}(X^0(i/n))u(i/n)$ with nearly optimal cost is not as straightforward as in the nondegenerate case [it is simple in that case due to the invertibility of $A^{1/2}(\bar{X}^n(i/n))$].

Recall the orthogonal decomposition of $A^{-1}(x)$ discussed above (5.6). For $\kappa \in (0, \infty)$, define

$$A_\kappa^{-1}(x) = Q(x)\Lambda_\kappa^{-1}(x)Q^T(x),$$

where $\Lambda_\kappa^{-1}(x)$ is the diagonal matrix such that $\Lambda_{ii,\kappa}^{-1}(x) = \Lambda_{ii}^{-1}(x)$ when $\Lambda_{ii}^{-1}(x) \leq \kappa^2$ and $\Lambda_{ii,\kappa}^{-1}(x) = \kappa^2$ when $\Lambda_{ii}^{-1}(x) > \kappa^2$. Note that by [155, Theorem 6.2.37], $A^{1/2}(x)$, $A_\kappa^{-1}(x)$, and $A_\kappa^{1/2}(x)$ are continuous functions of $A(x)$, and consequently they are also continuous functions of $x \in \mathbb{R}^d$. In addition, define

$$u_\kappa(s) = \begin{cases} u(s) & \text{for } \|u(s)\| \leq \kappa, \\ \frac{\kappa u(s)}{\|u(s)\|} & \text{for } \|u(s)\| > \kappa. \end{cases}$$

Let $\phi^{u,\kappa}(t) = \phi^{A_\kappa^{-1/2}(X^0)u_\kappa}(t)$, and note that $\phi^{u,\kappa}$ solves

$$\phi^{u,\kappa}(t) = \int_0^t Db(X^0(s))\phi^{u,\kappa}(s)ds$$

$$+ \int_0^t A(X^0(s))A_\kappa^{-1/2}(X^0(s))u_\kappa(s)ds. \qquad (5.36)$$

For n sufficiently large,

$$\max_{0 \leq i \leq n-1} \frac{1}{\sqrt{\varkappa(n)n}} \left\| A_\kappa^{-1/2}\left(X^0(i/n)\right)u_\kappa(i/n)\right\| \leq \frac{1}{\sqrt{\varkappa(n)n}}\kappa^2 \leq \lambda_{DA},$$

and we can define the sequence $\{(\bar{X}^{n,\kappa}, \bar{Y}^{n,\kappa}, \bar{M}^{n,\kappa}, w^{n,\kappa})\}$ as in Construction 5.3 with

$$\bar{\mu}_i^{n,\kappa}(dy) = \exp\left\{\left\langle y, \frac{1}{\sqrt{\varkappa(n)n}}A_\kappa^{-1/2}\left(X^0(i/n)\right)u_\kappa(i/n)\right\rangle\right.$$

$$\left. - H_c\left(\bar{X}_i^{n,\kappa}, \frac{1}{\sqrt{\varkappa(n)n}}A_\kappa^{-1/2}\left(X^0(i/n)\right)u_\kappa(i/n)\right)\right\}\theta(dy|\bar{X}_i^{n,\kappa}).$$

We will use

$$\int_{\mathbb{R}^d} y \exp\{\langle y, \alpha\rangle - H_c(x,\alpha)\}\theta(dy|x) = D_\alpha H_c(x,\alpha)$$

and the formula

$$D_\alpha H_c(x,\alpha) = D_\alpha H_c(x,\alpha) - D_\alpha H_c(x,0) = \int_0^1 \left(\frac{d}{ds}D_\alpha H_c(x,s\alpha)\right)ds,$$

where $D_\alpha H_c(x,0) = 0$ follows from (5.4). Using (5.5) to approximate second derivatives that appear on the right side of the last display, the bounds (5.7) imply that for $\|\alpha\| \leq \lambda_D A$,

$$\left\| \int_{\mathbb{R}^d} y \exp\{\langle y, \alpha \rangle - H_c(x, \alpha)\} \theta(dy|x) - A(x)\alpha \right\| \le K_{DA} \|\alpha\|^2 . \tag{5.37}$$

The next result identifies the limit in probability of the controlled processes and an asymptotic bound for the relative entropies. Recall that $u \in \mathcal{C}([0, 1] : \mathbb{R}^d)$ has been fixed.

Theorem 5.13 *Let $\kappa \in (0, \infty)$ be given. Consider the controls $\{\bar{\mu}_i^{n,\kappa}\}$ and random variables $\{(\bar{X}^{n,\kappa}, \bar{Y}^{n,\kappa}, \bar{M}^{n,\kappa}, w^{n,\kappa})\}$ as in Construction 5.3 with $\{\bar{\mu}_i^n\}$ replaced by $\{\bar{\mu}_i^{n,\kappa}\}$, and define $\phi^{u,\kappa}$ by (5.36). Then*

$$\bar{Y}^{n,\kappa} \to \phi^{u,\kappa} \tag{5.38}$$

in $\mathcal{C}([0, 1] : \mathbb{R}^d)$ in probability, and

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

$$\le \frac{1}{2} \int_0^1 \left\| A_\kappa^{-1/2}(X^0(s)) u_\kappa(s) \right\|_{A(X^0(s))}^2 ds. \tag{5.39}$$

Proof From (5.8) and (5.37), for all n large enough that $\kappa^2 / \sqrt{\varkappa(n)n} \le \lambda_{DA}$ and with $s_i^n \doteq i/n$,

$$R\left(\bar{\mu}_i^{n,\kappa}(\cdot) \,\big\|\, \theta(\cdot|\bar{X}_i^{n,\kappa}) \right)$$

$$= \int_{\mathbb{R}^d} \left\langle y, \frac{1}{\sqrt{\varkappa(n)n}} A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right) \right\rangle \bar{\mu}_i^{n,\kappa}(dy)$$

$$\quad - H_c\left(\bar{X}_i^{n,\kappa}, \frac{1}{\sqrt{\varkappa(n)n}} A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right) \right)$$

$$\le \frac{1}{\varkappa(n)n} \left\langle A\left(\bar{X}_i^{n,\kappa}\right) A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right), A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right) \right\rangle$$

$$\quad - \frac{1}{2\varkappa(n)n} \left\langle A\left(\bar{X}_i^{n,\kappa}\right) A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right), \right.$$

$$\left. A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right) \right\rangle + \frac{2}{(\varkappa(n)n)^{3/2}} K_{DA}\kappa^6$$

$$= \frac{1}{2\varkappa(n)n} \left\| A_\kappa^{-1/2} \left(X^0\left(s_i^n\right) \right) u_\kappa\left(s_i^n\right) \right\|_{A(\bar{X}_i^{n,\kappa})}^2 + \frac{2}{(\varkappa(n)n)^{3/2}} K_{DA}\kappa^6.$$

Consequently,

$$\limsup_{n \to \infty} \varkappa(n) n E \left[\frac{1}{n} \sum_{i=0}^{n-1} R\left(\bar{\mu}_i^{n,\kappa}(\cdot) \,\big\|\, \theta(\cdot|\bar{X}_i^{n,\kappa}) \right) \right] \tag{5.40}$$

$$\leq \limsup_{n\to\infty} \frac{1}{2} E\left[\frac{1}{n}\sum_{i=0}^{n-1}\left\|A_\kappa^{-1/2}\left(X^0\left(i/n\right)\right)u_\kappa\left(i/n\right)\right\|^2_{A(\bar{X}_i^{n,\kappa})}\right],$$

where in fact,

$$\limsup_{n\to\infty} \frac{1}{2} E\left[\frac{1}{n}\sum_{i=0}^{n-1}\left\|A_\kappa^{-1/2}\left(X^0\left(i/n\right)\right)u_\kappa\left(i/n\right)\right\|^2_{A(\bar{X}_i^{n,\kappa})}\right] \leq \frac{1}{2}\kappa^4 K_A.$$

Therefore, (5.17) is satisfied by $\{\bar{\mu}_j^{n,\kappa}\}$. Thus the conclusions of Theorem 5.7 hold with $\bar{Y}^n$, $\bar{M}^n$ replaced by $\bar{Y}^{n,\kappa}$, $\bar{M}^{n,\kappa}$. Choose a subsequence (keeping n as the index for convenience) along which $\{(\bar{M}^{n,\kappa}, \bar{Y}^{n,\kappa})\}$ converges weakly to some limit $(\bar{M}^\kappa, \bar{Y}^\kappa)$, where $[\bar{M}^\kappa]_2$ is Lebesgue measure and

$$\bar{Y}^\kappa(t) = \int_0^t Db(X^0(s))\bar{Y}^\kappa(s)ds + \int_0^t \int_{\mathbb{R}^d} w[\bar{M}^\kappa]_{1|2}(dw\,|s)ds.$$

Then $\bar{Y}^{n,\kappa} \to \bar{Y}^\kappa$ implies

$$\sup_{t\in[0,1]}\left\|\bar{X}^{n,\kappa}(t) - X^0(t)\right\| \to 0$$

in probability. Because of this and the continuity of $A^{1/2}(x)$, we have (recall $s^n(t) \doteq \lfloor nt\rfloor/n$)

$$\sup_{t\in[0,1]}\left\|A^{1/2}(\bar{X}^{n,\kappa}(s^n(t))) - A^{1/2}(X^0(s^n(t)))\right\| \to 0$$

in probability. However, the continuity of $t \mapsto A^{1/2}(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)$ gives

$$\sup_{t\in[0,1]}\left\|A^{1/2}(X^0(s^n(t)))A_\kappa^{-1/2}(X^0(s^n(t)))u_\kappa(s^n(t))\right.$$
$$\left.-A^{1/2}(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)\right\| \to 0.$$

Combining these limits, and using the fact that $A_\kappa^{-1/2}(X^0(t))u_\kappa(t)$ is uniformly bounded, shows that

$$\sup_{t\in[0,1]}\left\|A^{1/2}(\bar{X}^{n,\kappa}(s^n(t)))A_\kappa^{-1/2}(X^0(s^n(t)))u_\kappa(s^n(t))\right. \tag{5.41}$$
$$\left.-A^{1/2}(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)\right\| \to 0$$

in probability. This combined with the uniform bounds allows the use of the dominated convergence theorem to show that

$$\limsup_{n\to\infty} E\left[\frac{1}{2}\int_0^1 \left\|A_\kappa^{-1/2}(X^0(s^n(t)))u_\kappa(s^n(t))\right\|^2_{A(\bar{X}^{n,\kappa}(s^n(t)))} dt\right]$$
$$= \frac{1}{2}\int_0^1 \left\|A_\kappa^{-1/2}(X^0(t))u_\kappa(t)\right\|^2_{A(X^0(t))} dt.$$

Combining this with (5.40) demonstrates (5.39).

To prove (5.38), we will show that in fact,

$$\bar{M}^\kappa(dw \times dt) = \delta_{A(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)}(dw)dt.$$

For all $\sigma > 0$, let

$$G_\sigma \doteq \left\{(z,t) \in \mathbb{R}^d \times [0,1] : \left\|z - A(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)\right\| \leq \sigma\right\},$$

and note that by weak convergence, $\limsup_{n\to\infty} E[\bar{M}^{n,\kappa}(G_\sigma)] \leq E[\bar{M}^\kappa(G_\sigma)]$. Note also that

$$E[\bar{M}^{n,\kappa}(G_\sigma)]$$
$$\geq P\left[\sup_{t\in[0,1]}\left\|\sqrt{\varkappa(n)n}\int_{\mathbb{R}^d} y\bar{\mu}_{\lfloor nt\rfloor}^{n,\kappa}(dy) - A(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)\right\| \leq \sigma\right].$$

However, by (5.37) we can choose n large enough to make

$$\sup_{t\in[0,1]}\left\|\sqrt{\varkappa(n)n}\int_{\mathbb{R}^d} y\bar{\mu}_{\lfloor nt\rfloor}^{n,\kappa}(dy) - A\left(\bar{X}^{n,\kappa}\left(s^n(t)\right)\right)A_\kappa^{-1/2}\left(X^0\left(s^n(t)\right)\right)u_\kappa\left(s^n(t)\right)\right\|$$

arbitrarily small, and the proof that

$$\sup_{t\in[0,1]}\left\|A(\bar{X}^{n,\kappa}(s^n(t)))A_\kappa^{-1/2}(X^0(s^n(t)))u_\kappa(s^n(t))\right.$$
$$\left. -A(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa(t)\right\| \to 0$$

in probability is identical to the proof of (5.41). Hence $\limsup_{n\to\infty} E[\bar{M}^{n,\kappa}(G_\sigma)] = 1$ for all $\sigma > 0$, and so $E[\bar{M}^\kappa(\cap_{n\in\mathbb{N}}G_{1/n})] = 1$. This implies that with probability 1,

$$[\bar{M}]_{1|2}^\kappa(dw\,|t) = \delta_{A(X^0(t))A_\kappa^{-1/2}(X^0(t))u_\kappa}(dw)$$

for a.e. t. It follows that

$$\bar{Y}^\kappa(t) = \int_0^t Db(X^0(s))\bar{Y}^\kappa(s)ds + \int_0^t A(X^0(s))A_\kappa^{-1/2}(X^0(s))u_\kappa(s)ds,$$

and therefore $\bar{Y}^{n,\kappa} \to \phi^{u,\kappa}$ weakly. This implies (5.38) and completes the proof. $\square$

The second theorem in this section allows us to approximate $F(\phi^u)$ by $F(\phi^{u,\kappa})$ and $\frac{1}{2} \int_0^1 \|u(s)\|^2 \, ds$ by

$$\frac{1}{2} \int_0^1 \left\| A_\kappa^{-1/2}(X^0(s))u_\kappa(s) \right\|_{A(X^0(s))}^2 \, ds.$$

Recall that $u \in \mathscr{C}([0, 1] : \mathbb{R}^d)$ has been given.

Theorem 5.14 *Define ϕ^u by (5.32) and $\phi^{u,\kappa}$ by (5.36). Then $\phi^{u,\kappa} \to \phi^u$ in $\mathscr{C}([0, 1] : \mathbb{R}^d)$ and*

$$\limsup_{\kappa \to \infty} \frac{1}{2} \int_0^1 \left\| A_\kappa^{-1/2}(X^0(s))u_\kappa(s) \right\|_{A(X^0(s))}^2 \, ds \le \frac{1}{2} \int_0^1 \|u(s)\|^2 \, ds.$$

Proof Note that

$$\left\| A^{1/2}(X^0(s)) A_\kappa^{-1/2}(X^0(s))u_\kappa(s) \right\| \le \|u(s)\|$$

for all $s \in [0, 1]$. Thus it is automatic that

$$\limsup_{\kappa \to \infty} \frac{1}{2} \int_0^1 \left\| A_\kappa^{-1/2}(X^0(s))u_\kappa(s) \right\|_{A(X^0(s))}^2 \, ds \le \frac{1}{2} \int_0^1 \|u(s)\|^2 \, ds.$$

Also, for $x \in \mathbb{R}^d$,

$$A(x)A_\kappa^{-1/2}(x) = Q(x)\Lambda(x)\Lambda_\kappa^{-1/2}(x)Q^T(x) \to Q(x)\Lambda^{1/2}(x)Q^T(x) = A^{1/2}(x).$$

Since $u_\kappa(s) \to u(s)$ for all $s \in [0, 1]$,

$$A(X^0(s))A_\kappa^{-1/2}(X^0(s))u_\kappa(s) \to A^{1/2}(X^0(s))u(s) \tag{5.42}$$

pointwise. Since $u \in \mathscr{L}^2([0, 1] : \mathbb{R}^d)$, by the dominated convergence theorem, (5.42) also holds in $\mathscr{L}^2([0, 1] : \mathbb{R}^d)$. Combining this with the second inequality in (5.34) shows that $\phi^{u,\kappa} \to \phi^u$ in $\mathscr{C}([0, 1] : \mathbb{R}^d)$. $\qquad\square$

Using (5.15) and the fact that any given control is suboptimal yields

$$-\varkappa(n) \log E \left[e^{-\frac{1}{\varkappa(n)} F(Y^n)} \right] \le E \left[\sum_{i=0}^{n-1} \varkappa(n) R \left(\bar{\mu}_i^{n,\kappa}(\cdot) \,\middle\|\, \theta(\cdot | \bar{X}_i^{n,\kappa}) \right) + F(\bar{Y}^{n,\kappa}) \right].$$

Using Theorem 5.13, this implies

$$\limsup_{n \to \infty} -\varkappa(n) \log E\left[e^{-\frac{1}{\varkappa(n)} F(Y^n)}\right]$$

$$\leq \frac{1}{2} \int_0^1 \left\| A_\kappa^{-1/2}(X^0(s)) u_\kappa(s) \right\|_{A(X^0(s))}^2 ds + F(\phi^{u,\kappa}).$$

Sending $\kappa \to \infty$ and using Theorem 5.14 gives (5.35), which completes the proof of (5.33). □

5.6 Notes

Among the earliest papers to study moderate deviations are those by Rubin and Sethuraman [222], Ghosh [146], and Michel [200]. See the introduction of [41] for a more complete discussion of work in this area. While a number of settings have been considered, to the authors' knowledge the first papers to consider moderate deviations for small noise processes around solutions to general nonlinear ODEs (rather than constant-velocity trajectories) are [100] for discrete time models, upon which this chapter is based, and [41] for continuous time processes. As noted in the introduction to the chapter, the proof of the moderate deviation principle presented here is neither uniformly harder nor easier than its large deviation counterpart, at least when one is using weak convergence methods. In particular, the large deviation upper bound is made more difficult due to difficulties in using tightness in the convergence analysis. (The case of solutions to SDEs driven by Brownian motion, which is given as an example in Chap. 3, is in fact much easier, owing to the fact that the driving noise is already Gaussian.) Also, the assumed conditions are not strictly weaker, mainly in that additional smoothness is needed for the proper centering and rescaling. Moderate deviation principles will also appear in Chaps. 10 and 13.