

Chapter 3

Examples of Representations and Their Application



Our approach to the study of large deviations is based on convenient variational representations for expected values of nonnegative functionals. In this chapter we give three examples of such representations and show how they allow easy proofs of some classical results.

In Sect. 3.1 we present a representation for stochastic processes in discrete time. To illustrate the main idea we consider the simple setting in which the stochastic process is an iid sequence of random variables [Proposition 3.1]. We then show how this representation can be used to prove Sanov's theorem and Cramér's theorem. Analogous representations for more general noise models will be used many times in later chapters. In Sect. 3.2 we state a variational representation for functionals of a k -dimensional Brownian motion [Theorem 3.14]. This result will be generalized and proved in the setting of an infinite dimensional Brownian motion in Chap. 8, and we apply it here to give an elementary proof of the large deviation principle for small noise diffusions. Section 3.3 states a variational representation for functionals of a standard Poisson process [Theorem 3.23]. This result will also be extended in Chap. 8 to the setting of Poisson random measures with points in an arbitrary locally compact Polish space. As an application of Theorem 3.23 we prove the large deviation principle for stochastic differential equations driven by Poisson processes.

3.1 Representation for an IID Sequence

Owing to the role it plays in the representations, we sometimes refer to the measure appearing in the second position in relative entropy, i.e., θ in $R(\mu \parallel \theta)$, as the “base” measure. The starting point of all large deviation results in the book is the relative entropy representation in part (a) of Proposition 2.2. When the base measure is structured, for example when θ is a product measure or a Markov measure, a more useful, control-theoretic, representation can be found in terms of the component

measures that make up θ . Here is an example. Suppose that (X_1, X_2) is an $(S_1 \times S_2)$ -valued random variable with joint distribution $\theta(dx_1 \times dx_2) = \theta_1(dx_1)\theta_2(dx_2)$. Then the variational formula (2.1) says that if $G \in \mathcal{M}_b(S_1 \times S_2)$, then

$$-\log Ee^{-G(X_1, X_2)} = \inf_{\mu \in \mathcal{P}(S_1 \times S_2)} \left[\int_{S_1 \times S_2} G d\mu + R(\mu \parallel \theta) \right].$$

One can always disintegrate μ in the form

$$\mu(dx_1 \times dx_2) = [\mu]_1(dx_1)[\mu]_{2|1}(dx_2|x_1),$$

where $[\mu]_1$ is the marginal on S_1 and $[\mu]_{2|1}$ is the conditional distribution on S_2 given x_1 . Suppose that (\bar{X}_1, \bar{X}_2) is distributed according to μ , $\bar{\mu}_1(\cdot) = [\mu]_1(\cdot)$ and $\bar{\mu}_2(\cdot) = [\mu]_{2|1}(\cdot | \bar{X}_1)$ (and note that $\bar{\mu}_2$ is a random measure). It follows from the chain rule [Theorem 2.6] that

$$\begin{aligned} R(\mu \parallel \theta) &= R([\mu]_1 \parallel \theta_1) + \int_{S_1} R([\mu]_{2|1}(\cdot | x_1) \parallel \theta_2(\cdot)) [\mu]_1(dx_1) \\ &= E [R(\bar{\mu}_1 \parallel \theta_1) + R(\bar{\mu}_2 \parallel \theta_2)]. \end{aligned}$$

Here we have used that \bar{X}_1 has distribution $[\mu]_1$ to account for integration with respect to this measure. Then we can rewrite the representation as

$$-\log Ee^{-G(X_1, X_2)} = \inf_{\mu \in \mathcal{P}(S_1 \times S_2)} E \left[G(\bar{X}_1, \bar{X}_2) + \sum_{i=1}^2 R(\bar{\mu}_i \parallel \theta_i) \right]. \quad (3.1)$$

There is an obvious extension of (3.1) to any finite collection of independent random variables. The extension for the special case in which the random variables are iid is as follows. Let S^n denote the product space of n copies of S .

Proposition 3.1 *Let $\{X_i\}_{i \in \mathbb{N}}$ be iid S -valued random variables with distribution θ and let $n \in \mathbb{N}$. If $G \in \mathcal{M}_b(S^n)$, then*

$$-\frac{1}{n} \log Ee^{-nG(X_1, \dots, X_n)} = \inf E \left[G(\bar{X}_1^n, \dots, \bar{X}_n^n) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \theta) \right], \quad (3.2)$$

with the infimum over all collections of random probability measures $\{\bar{\mu}_i^n\}_{i \in \{1, \dots, n\}}$ that satisfy the following two conditions:

1. $\bar{\mu}_i^n$ is measurable with respect to the σ -algebra \mathcal{F}_{i-1}^n , where $\mathcal{F}_0^n = \{\emptyset, \Omega\}$ and for $i \in \{1, \dots, n\}$, $\mathcal{F}_i^n = \sigma\{\bar{X}_1^n, \dots, \bar{X}_i^n\}$;
2. the conditional distribution of \bar{X}_i^n , given \mathcal{F}_{i-1}^n , is $\bar{\mu}_i^n$.

Given any measure $\mu \in \mathcal{P}(S^n)$, if $\{\bar{X}_i^n\}_{i=1, \dots, n}$ has distribution μ , then $\bar{\mu}_i^n$ in the statement of the proposition would equal $[\mu]_{i|1, \dots, i-1}(\cdot | \bar{X}_1^n, \dots, \bar{X}_{i-1}^n)$. On the other

hand, given $\{\bar{X}_i^n\}$ and $\{\bar{\mu}_i^n\}$ as in the statement of the proposition, one can identify a $\mu \in \mathcal{P}(S^n)$ that corresponds to these conditional distributions. We consider $\{X_i^n\}_{i=1,\dots,n}$ to be a *controlled* version of the original sequence $\{X_i\}_{i=1,\dots,n}$, with control $\bar{\mu}_i^n$ selecting the (conditional) distribution of \bar{X}_i^n .

Notational convention. Throughout, we will use overbars to indicate the controlled analogue of any uncontrolled process.

3.1.1 Sanov's and Cramér's Theorems

First we recall the statement of the Glivenko–Cantelli lemma. The space of probability measures on S is denoted by $\mathcal{P}(S)$ and is equipped with the weak topology (see Appendix A).

Lemma 3.2 (GLIVENKO–CANTELLI LEMMA) *Let $\{X_i\}_{i \in \mathbb{N}}$ be iid S -valued random variables with distribution γ , and let L^n be the empirical measure of the first n variables:*

$$L^n(dx) \doteq \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(dx).$$

Then with probability one (w.p.1), L^n converges to γ .

The proof is a special case of the arguments we will use for Sanov's theorem, and in particular, the result follows from Lemmas 3.4 and 3.5. Sanov's theorem itself is the large deviation refinement of this law of large numbers (LLN) result.

Theorem 3.3 (SANOV'S THEOREM) *Let $\{X_i\}_{i \in \mathbb{N}}$ be iid S -valued random variables with distribution γ . Then $\{L^n\}_{n \in \mathbb{N}}$ satisfies the LDP on $\mathcal{P}(S)$ with rate function $I(\mu) = R(\mu \parallel \gamma)$.*

By Theorem 1.8, to prove Theorem 3.3 it is enough to show that

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log E \exp\{-nF(L^n)\} = \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)]$$

for every $F \in C_b(\mathcal{P}(S))$. The proof will use the control representation in Proposition 3.1 and will be completed in two steps. First, we will show that the left side in the last display is bounded below by the right side (which gives the Laplace upper bound), and then we will prove the reverse inequality (Laplace lower bound). The first inequality is proved in Sect. 3.1.3, while the second is proved in Sect. 3.1.4.

Taking $G(x_1, \dots, x_n) = F(\sum_{i=1}^n \delta_{x_i}(dx)/n)$ in the representation (3.2) gives

$$-\frac{1}{n} \log E \exp\{-nF(L^n)\} = \inf_{\{\bar{\mu}_i^n\}} E \left[F(\bar{L}^n) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right], \quad (3.3)$$

where $\bar{L}^n = \frac{1}{n} \sum_{i=1}^n \delta_{\bar{X}_i}$. Thus in order to prove Theorem 3.3, we need to show that

$$\inf_{\{\bar{\mu}_i^n\}} E \left[F(\bar{L}^n) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right] \rightarrow \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)].$$

Since F is bounded, the infimum in the representation is always bounded above by $\|F\|_\infty \doteq \sup_{x \in S} |F(x)| < \infty$. It follows that in the infimum in (3.3) we can always restrict to control sequences $\{\bar{\mu}_i^n\}_{i=1, \dots, n}$ for which

$$\sup_{n \in \mathbb{N}} E \left[\frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right] \leq 2 \|F\|_\infty + 1. \quad (3.4)$$

3.1.2 Tightness and Weak Convergence

The bound (3.4) on relative entropy costs is all that is *available*, but also all that is *needed*, to prove tightness.

Lemma 3.4 *Consider any collection of controls $\{\bar{\mu}_i^n, i = 1, \dots, n\}_{n \in \mathbb{N}}$ for which (3.4) is satisfied, and let $\hat{\mu}^n = \frac{1}{n} \sum_{i=1}^n \bar{\mu}_i^n$. Then $\{(\bar{L}^n, \hat{\mu}^n)\}_{n \in \mathbb{N}}$ is tight.*

Proof By the convexity of relative entropy and Jensen's inequality,

$$2 \|F\|_\infty + 1 \geq E \left[\frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right] \geq E [R(\hat{\mu}^n \parallel \gamma)].$$

Since $\mu \mapsto R(\mu \parallel \gamma)$ has compact level sets, it is a tightness function, and so the bound in the last display along with Lemmas 2.9 and 2.11 shows that both $\{\hat{\mu}^n\}_{n \in \mathbb{N}}$ and $\{E \hat{\mu}^n\}_{n \in \mathbb{N}}$ are tight. Since $\bar{\mu}_i^n$ is the conditional distribution used to select \bar{X}_i^n , it follows that for every bounded measurable function f ,

$$\begin{aligned} E \int_S f(x) \bar{L}^n(dx) &= E \left[\frac{1}{n} \sum_{i=1}^n f(\bar{X}_i^n) \right] \\ &= E \left[\frac{1}{n} \sum_{i=1}^n \int_S f(x) \bar{\mu}_i^n(dx) \right] \\ &= E \int_S f(x) \hat{\mu}^n(dx). \end{aligned}$$

Thus $E \bar{L}^n = E \hat{\mu}^n$, and so $\{\bar{L}^n\}$ and hence $\{(\bar{L}^n, \hat{\mu}^n)\}$ are tight. Here once more we have used Lemma 2.11. \square

Thus $(\bar{L}^n, \hat{\mu}^n)$ will converge, at least along subsequences. To prove the LDP we need to relate the limits of the controls $\hat{\mu}^n$ and the controlled process \bar{L}^n .

Lemma 3.5 *Suppose $\{(\bar{L}^n, \hat{\mu}^n)\}_{n \in \mathbb{N}}$ converges along a subsequence to $(\bar{L}, \hat{\mu})$. Then $\bar{L} = \hat{\mu}$ w.p.1.*

The proof of this result, which is a martingale version of the proof of the Glivenko–Cantelli lemma, will be given in Sect. 3.1.5 after we complete the proof of Sanov’s theorem.

3.1.3 Laplace Upper Bound

The proof of Theorem 3.3 is partitioned into upper and lower bounds. In this section we will prove the Laplace upper bound, which is the same as the variational lower bound

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log E \exp\{-nF(L^n)\} \geq \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)]. \quad (3.5)$$

For $\varepsilon > 0$, let $\{\bar{\mu}_i^n\}_{i=1, \dots, n}$ and $\{\bar{X}_i^n\}_{i=1, \dots, n}$ come within ε of the infimum in (3.3):

$$-\frac{1}{n} \log E \exp\{-nF(L^n)\} + \varepsilon \geq E \left[F(\bar{L}^n) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right].$$

Recall that we assume (without loss of generality) that the uniform bound in (3.4) holds, and thus by Lemma 3.4, $\{(\bar{L}^n, \hat{\mu}^n)\}$ is tight.

Owing to tightness, for every subsequence of $\{(\bar{L}^n, \hat{\mu}^n)\}$ we can extract a further subsequence that converges weakly. It suffices to prove (3.5) for such a subsubsequence. To simplify notation, we denote this subsubsequence by n , and its limit by $(\bar{L}, \hat{\mu})$. According to Lemma 3.5, $\bar{L} = \hat{\mu}$ a.s. Using Jensen’s inequality for the second inequality, the convergence in distribution, Fatou’s lemma and lower semicontinuity of relative entropy for the third inequality, and the w.p.1 relation $\bar{L} = \hat{\mu}$ for the last inequality, we obtain

$$\begin{aligned} \liminf_{n \rightarrow \infty} -\frac{1}{n} \log E e^{-nF(L^n)} + \varepsilon &\geq \liminf_{n \rightarrow \infty} E \left[F(\bar{L}^n) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right] \\ &\geq \liminf_{n \rightarrow \infty} E [F(\bar{L}^n) + R(\hat{\mu}^n \parallel \gamma)] \\ &\geq E [F(\bar{L}) + R(\hat{\mu} \parallel \gamma)] \\ &\geq \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)]. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, (3.5) follows. \square

3.1.4 Laplace Lower Bound

Next we prove the variational upper bound

$$\limsup_{n \rightarrow \infty} -\frac{1}{n} \log E \exp\{-nF(L^n)\} \leq \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)], \quad (3.6)$$

which establishes the Laplace lower bound. For $\varepsilon > 0$ let μ^* satisfy

$$F(\mu^*) + R(\mu^* \parallel \gamma) \leq \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)] + \varepsilon.$$

Then let $\bar{\mu}_i^n = \mu^*$ for all $n \in \mathbb{N}$ and $i \in \{1, \dots, n\}$. By either Lemma 3.5 or the ordinary Glivenko–Cantelli lemma, the weak limit of \bar{L}^n equals μ^* w.p.1. The representation in Proposition 3.1 gives the first inequality below, and the dominated convergence theorem gives the equality

$$\begin{aligned} \limsup_{n \rightarrow \infty} -\frac{1}{n} \log E e^{-nF(L^n)} &\leq \limsup_{n \rightarrow \infty} E \left[F(\bar{L}^n) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right] \\ &= [F(\mu^*) + R(\mu^* \parallel \gamma)] \\ &\leq \inf_{\mu \in \mathcal{P}(S)} [F(\mu) + R(\mu \parallel \gamma)] + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the bound (3.6) follows. \square

Remark 3.6 When combined with the previous subsection, the argument just given shows that for asymptotic optimality one can restrict to controls of the form $\bar{\mu}_i^n = \mu^*$, i.e., product measure.

3.1.5 Proof of Lemma 3.5 and Remarks on the Proof of Sanov's Theorem

Since S is Polish, there exists a countable separating class $\{f_m\}_{m \in \mathbb{N}}$ of bounded continuous functions (see Appendix A). Define $K_m \doteq \|f_m\|_\infty$ and $\Delta_{m,i}^n \doteq f_m(\bar{X}_i^n) - \int_S f_m(x) \bar{\mu}_i^n(dx)$. For every $\varepsilon > 0$,

$$\begin{aligned} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n f_m(\bar{X}_i^n) - \frac{1}{n} \sum_{i=1}^n \int_S f_m(x) \bar{\mu}_i^n(dx) \right| > \varepsilon \right\} \\ \leq \frac{1}{\varepsilon^2} E \left[\frac{1}{n^2} \sum_{i,j=1}^n \Delta_{m,i}^n \Delta_{m,j}^n \right]. \end{aligned}$$

Recall that $\mathcal{F}_j^n = \sigma(\bar{X}_i^n, i = 1, \dots, j)$. By a standard conditioning argument, the off-diagonal terms vanish: for $i > j$,

$$E[\Delta_{m,i}^n \Delta_{m,j}^n] = E[E[\Delta_{m,i}^n \Delta_{m,j}^n | \mathcal{F}_{i-1}^n]] = E[E[\Delta_{m,i}^n | \mathcal{F}_{i-1}^n] \Delta_{m,j}^n] = 0.$$

Since $|\Delta_{m,i}^n| \leq 2K_m$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n f_m(\bar{X}_i^n) - \frac{1}{n}\sum_{i=1}^n \int_S f_m(x) \bar{\mu}_i^n(dx)\right| > \varepsilon\right\} \leq \frac{4K_m^2}{n\varepsilon^2}.$$

Since $(\bar{L}^n, \hat{\mu}^n) \Rightarrow (\bar{L}, \hat{\mu})$ and $\varepsilon > 0$ is arbitrary, by Fatou's lemma, we have

$$P\left\{\int_S f_m(x) \bar{L}(dx) = \int_S f_m(x) \hat{\mu}(dx)\right\} = 1.$$

Now use that $\{f_m\}$ is countable and separating to conclude that $\bar{L} = \hat{\mu}$ w.p.1. \square

Remark 3.7 There is a close relationship between the legitimate use of Jensen's inequality in the proof of any particular Laplace upper bound and the asymptotic independence of optimal controls with respect to one or more parameters. In the context of Sanov's theorem, the parameter is the time index i . In the proof of the upper bound, the inequality

$$E\left[\frac{1}{n}\sum_{i=1}^n R(\bar{\mu}_i^n \|\gamma)\right] \geq E[R(\hat{\mu}^n \|\gamma)]$$

was used, where $\hat{\mu}^n$ is the average (over i) of $\bar{\mu}_i^n$. In general, Jensen's inequality holds with a strict inequality. There is an exception when the quantity being averaged is independent of the parameter over which the averaging occurs. Since we consider the limit $n \rightarrow \infty$, this means that there should be no loss due to the use of Jensen's inequality if one restricts to controls that are independent of i in this limit. In any particular instance, a use of Jensen's inequality is appropriate only when one proves the corresponding lower bound with the same rate function, i.e., in the proof of the lower bound one should be able to restrict to controls that do not depend on the parameter being averaged. This of course occurs in the proof of Sanov's theorem, since for the lower bound we consider controls of the form $\bar{\mu}_i^n = \mu^*$ for a fixed measure μ^* .

Information on what control dependencies are asymptotically unimportant can be useful in various ways, including the construction of importance sampling schemes, which is considered later in the book. It typically simplifies the large deviation proofs considerably, since one needs to keep track in the weak convergence analysis of only the nontrivial dependencies, and often one has some a priori insight into which parameters should be unimportant. However, as noted previously, it is only after the proof of upper and lower bounds with the same rate function that one can claim that the use of Jensen's inequality was without loss.

3.1.6 Cramér's Theorem

Cramér's theorem states the LDP for the empirical mean of \mathbb{R}^d -valued iid random variables: $S_n \doteq \frac{1}{n}(X_1 + \dots + X_n)$. Of course, one can recover the empirical mean from the empirical measure via $S_n = \int_{\mathbb{R}^d} y L^n(dy)$. If the underlying distribution γ has compact support, then the mapping $\mu \rightarrow \int_{\mathbb{R}^d} y \mu(dy)$ is continuous on a subset of $\mathcal{P}(\mathbb{R}^d)$ that contains L^n w.p.1. In this case, the LDP for $\{S_n\}_{n \in \mathbb{N}}$ follows directly from the contraction principle [Theorem 1.16], with the rate function I given by

$$I(\beta) \doteq \inf \left[R(\mu \parallel \gamma) : \int_{\mathbb{R}^d} y \mu(dy) = \beta \right] \quad (3.7)$$

for $\beta \in \mathbb{R}^d$. However, in general the mapping $\mu \mapsto \int_{\mathbb{R}^d} y \mu(dy)$ is not continuous, and the contraction principle does not suffice. As we will see, the issue is that the conditions of Sanov's theorem are too weak to force continuity with high probability. They are sufficient to imply tightness of controls, but no more. Once the conditions are appropriately strengthened, the weak convergence arguments can be carried out just as before, with the only difference being in the qualitative properties of the convergence. For $\alpha \in \mathbb{R}^d$ let

$$H(\alpha) \doteq \log \int_{\mathbb{R}^d} e^{(\alpha, y)} \gamma(dy).$$

Theorem 3.8 (CRAMÉR'S THEOREM) *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of iid \mathbb{R}^d -valued random variables with common distribution γ , and let $S_n \doteq \frac{1}{n} \sum_{i=1}^n X_i$. Assume that $H(\alpha) < \infty$ for all $\alpha \in \mathbb{R}^d$. Then $\{S_n\}_{n \in \mathbb{N}}$ satisfies the LDP with rate function I defined in (3.7).*

To prove the LDP we need to calculate the limits of

$$-\frac{1}{n} \log E \exp \left\{ -nF \left(\int_{\mathbb{R}^d} y L^n(dy) \right) \right\}, \quad (3.8)$$

where $F \in \mathcal{C}_b(\mathbb{R}^d)$. From the representation in Proposition 3.1 we see that (3.8) equals

$$\inf_{\{\bar{\mu}_i^n\}} E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}^n(dy) \right) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right].$$

Once more, without loss of generality we can assume that the relative entropy cost is uniformly bounded, and in particular that (3.4) holds. The next lemma shows that as a consequence of this uniform bound and our assumption on H , the collection $\{\bar{L}^n\}_{n \in \mathbb{N}}$ is uniformly integrable.

Lemma 3.9 Assume (3.4) and that $H(\alpha) < \infty$ for all $\alpha \in \mathbb{R}^d$. Then

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

Before proving the lemma we complete the proof of Theorem 3.8.

Proof (of Theorem 3.8) The uniform integrability of Lemma 3.9 implies that if \bar{L}^n converges in distribution to \bar{L} and (3.4) holds, then

$$E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}^n(dy) \right) \right] \rightarrow E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}(dy) \right) \right]. \quad (3.9)$$

The limit of (3.8) will now be calculated using essentially the same argument as that used to prove Sanov's theorem.

Variational lower bound. For $\varepsilon > 0$ let $\{\bar{\mu}_i^n\}_{i=1, \dots, n}$ and $\{\bar{X}_i^n\}_{i=1, \dots, n}$ satisfy

$$-\frac{1}{n} \log E e^{-nF(\int_{\mathbb{R}^d} y L^n(dy))} + \varepsilon \geq E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}^n(dy) \right) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right].$$

Consider a subsubsequence as in Sect. 3.1.3 (denoted again by n) along which $(\bar{L}^n, \hat{\mu}^n)$ converges weakly to $(\bar{L}, \hat{\mu})$. Then as in Sect. 3.1.3, we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} -\frac{1}{n} \log E \exp \left\{ -nF \left(\int_{\mathbb{R}^d} y L^n(dy) \right) \right\} + \varepsilon \\ & \geq \liminf_{n \rightarrow \infty} E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}^n(dy) \right) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \parallel \gamma) \right] \\ & \geq E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}(dy) \right) + R(\hat{\mu} \parallel \gamma) \right] \\ & \geq E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}(dy) \right) + I \left(\int_{\mathbb{R}^d} y \bar{L}(dy) \right) \right] \\ & \geq \inf_{\beta \in \mathbb{R}^d} [F(\beta) + I(\beta)]. \end{aligned}$$

Here the second inequality follows from (3.9), and the third follows from the definition of I and $\bar{L} = \hat{\mu}$ a.s. Since $\varepsilon > 0$ is arbitrary, the lower bound follows.

Variational upper bound. For $\varepsilon \in (0, 1)$ let $\beta^* \in \mathbb{R}^d$ satisfy

$$F(\beta^*) + I(\beta^*) \leq \inf_{\beta \in \mathbb{R}^d} [F(\beta) + I(\beta)] + \varepsilon.$$

Next let $\mu^* \in \mathcal{P}(\mathbb{R}^d)$ be such that $\int_{\mathbb{R}^d} x \mu^*(dx) = \beta^*$ and

$$F(\beta^*) + R(\mu^* \|\gamma) \leq F(\beta^*) + I(\beta^*) + \varepsilon.$$

As in Sect. 3.1.4, let $\bar{\mu}_i^n = \mu^*$ for all $n \in \mathbb{N}$ and $i \in \{1, \dots, n\}$. Then the weak limit of \bar{L}^n equals μ^* a.s., and (3.4) is satisfied. Thus

$$\begin{aligned} & \limsup_{n \rightarrow \infty} -\frac{1}{n} \log E \exp \left\{ -nF \left(\int_{\mathbb{R}^d} y L^n(dy) \right) \right\} \\ & \leq \limsup_{n \rightarrow \infty} E \left[F \left(\int_{\mathbb{R}^d} y \bar{L}^n(dy) \right) + \frac{1}{n} \sum_{i=1}^n R(\bar{\mu}_i^n \|\gamma) \right] \\ & = F(\beta^*) + R(\mu^* \|\gamma) \\ & \leq F(\beta^*) + I(\beta^*) + \varepsilon \\ & \leq \inf_{\beta \in \mathbb{R}^d} [F(\beta) + I(\beta)] + 2\varepsilon. \end{aligned}$$

Here the equality follows from (3.9) and the a.s. convergence of \bar{L}^n to μ^* . Since $\varepsilon \in (0, 1)$ is arbitrary, the upper bound follows. \square

Finally, we give the proof of Lemma 3.9.

Proof (of Lemma 3.9) The uniform integrability stated in this lemma is essentially a consequence of the bound on relative entropy costs and the assumption $H(\alpha) < \infty$. For $b \geq 0$ let

$$\ell(b) \doteq b \log b - b + 1. \quad (3.10)$$

We recall a bound already used frequently in Chap. 2 [see (2.9)]: for $a \geq 0$, $b \geq 0$, and $\sigma \geq 1$,

$$ab \leq e^{\sigma a} + \frac{1}{\sigma} (b \log b - b + 1) = e^{\sigma a} + \frac{1}{\sigma} \ell(b).$$

Thus if $\theta \in \mathcal{P}(\mathbb{R}^d)$ satisfies $\theta \ll \gamma$, then for every $\sigma \geq 1$,

$$\begin{aligned} \int_{\mathbb{R}^d} \|y\| \mathbf{1}_{\{\|y\| \geq M\}} \theta(dy) &= \int_{\mathbb{R}^d} \|y\| \mathbf{1}_{\{\|y\| \geq M\}} \frac{d\theta}{d\gamma}(y) \gamma(dy) \\ &\leq \int_{\mathbb{R}^d} e^{\sigma \|y\|} \mathbf{1}_{\{\|y\| \geq M\}} \gamma(dy) + \frac{1}{\sigma} \int_{\mathbb{R}^d} \ell \left(\frac{d\theta}{d\gamma}(y) \right) \gamma(dy) \\ &= \int_{\mathbb{R}^d} e^{\sigma \|y\|} \mathbf{1}_{\{\|y\| \geq M\}} \gamma(dy) + \frac{1}{\sigma} R(\theta \|\gamma). \end{aligned}$$

Note that the inequality holds trivially if $\theta \ll \gamma$. Therefore,

$$\begin{aligned} E \int_{\mathbb{R}^d} \|y\| \mathbf{1}_{\{\|y\| \geq M\}} \bar{L}^n(dy) &= E \int_{\mathbb{R}^d} \|y\| \mathbf{1}_{\{\|y\| \geq M\}} \hat{\mu}^n(dy) \\ &\leq \int_{\mathbb{R}^d} e^{\sigma \|y\|} \mathbf{1}_{\{\|y\| \geq M\}} \gamma(dy) + \frac{1}{\sigma} ER(\hat{\mu}^n \|\gamma). \quad (3.11) \end{aligned}$$

Since $H(\alpha) < \infty$ for all $\alpha \in \mathbb{R}^d$, for each fixed σ the mapping $y \mapsto \exp\{\sigma \|y\|\}$ is integrable with respect to γ . To see this, for $\lambda > 0$ let

$$m(\lambda) \doteq \sup_{\alpha \in \mathbb{R}^d: \|\alpha\| \leq \lambda} e^{H(\alpha)} = \sup_{\alpha \in \mathbb{R}^d: \|\alpha\| \leq \lambda} \int_{\mathbb{R}^d} e^{(\alpha, y)} \gamma(dy).$$

From the continuity of $\alpha \mapsto H(\alpha)$ it follows that $m(\lambda) < \infty$. For $J \subset \{1, \dots, d\}$ let $\mathbb{R}_J^d \doteq \{x \in \mathbb{R}^d : x_i \geq 0 \text{ if and only if } i \in J\}$, and define $\alpha^J \in \mathbb{R}^d$ by

$$\alpha_i^J \doteq \frac{\lambda}{\sqrt{d}} \text{ if } i \in J \text{ and } \alpha_i^J \doteq -\frac{\lambda}{\sqrt{d}} \text{ if } i \in J^c.$$

Then $\|\alpha^J\| = \lambda$ for all J , and for all $y \in \mathbb{R}_J^d$,

$$\langle \alpha^J, y \rangle = \frac{\lambda}{\sqrt{d}} \sum_{i=1}^d |y_i| \geq \frac{\lambda}{\sqrt{d}} \|y\|.$$

Thus

$$m(\lambda) \geq \int_{\mathbb{R}^d} e^{(\alpha^J, y)} \gamma(dy) \geq \int_{\mathbb{R}_J^d} e^{(\alpha^J, y)} \gamma(dy) \geq \int_{\mathbb{R}_J^d} e^{\frac{\lambda}{\sqrt{d}} \|y\|} \gamma(dy),$$

and therefore

$$\int_{\mathbb{R}^d} e^{\frac{\lambda}{\sqrt{d}} \|y\|} \gamma(dy) = \sum_J \int_{\mathbb{R}_J^d} e^{\frac{\lambda}{\sqrt{d}} \|y\|} \gamma(dy) \leq 2^d m(\lambda). \quad (3.12)$$

Since $\lambda > 0$ is arbitrary, we get $\int_{\mathbb{R}^d} \exp\{\sigma \|y\|\} \gamma(dy) < \infty$ for every $\sigma \in \mathbb{R}$, as asserted.

The bound (3.4) on the relative entropy and Jensen's inequality imply that the last term in (3.11) is bounded by $(2 \|F\|_\infty + 1)/\sigma$. The conclusion of Lemma 3.9 follows by taking limits in (3.11), in the order $n \rightarrow \infty$, $M \rightarrow \infty$, and then $\sigma \rightarrow \infty$. \square

Remark 3.10 The proof most often given of Cramér's theorem (e.g., as in [239]) uses a change of measure argument for the large deviation lower bound and Chebyshev's inequality for the upper bound. This line of argument naturally produces the following alternative form of the rate function as the **Legendre-Fenchel transform** of H :

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

By the uniqueness of rate functions [Theorem 1.15] it must be that $I = L$, though one can also directly verify that the two coincide [Lemma 4.16]. Both characterizations of the rate are useful. For example, the description as a Legendre transform easily shows that I is convex, while the characterization in terms of relative entropy allows

an easy calculation of the domain of finiteness of I . Note also that in principle, the two different expressions can be used to obtain upper and lower bounds on $I(\beta)$ for any given β . The two descriptions are in fact dual to each other.

Remark 3.11 It is possible to prove Cramér's theorem under just the condition that there is $\delta > 0$ such that $H(\alpha) < \infty$ for all α with $\|\alpha\| \leq \delta$. The main difficulty imposed by this weaker condition is that boundedness of costs does not imply the uniform integrability of controls that is used in the proof of the variational lower bound. This can be bypassed by the use of unbounded test functions of the form $F(x) = \infty 1_{C^c}(x)$, where C is convex. In the proof of the variational upper bound (large deviation lower bound) we can take C to be an open ball of radius $\delta > 0$ about a point x . Tightness follows, since here one picks controls that correspond to product measure. For the lower bound one must first establish that lower bounds for convex sets, which correspond to large deviation upper bounds, suffice to establish the full large deviation upper bound. This can be shown by approximating the complement of a level set of the rate function by a finite union of half-spaces (see the proof of Cramér's theorem in [239]), which uses the compactness of the level sets and an open covering argument. Given that it is sufficient to prove the variational lower bound for just convex sets, Jensen's inequality can be used to move the expected value inside F in the representation, and all that is required to complete the proof is boundedness of $E \int_{\mathbb{R}^d} x \hat{\mu}^n(dx)$ when costs are bounded. Since boundedness of costs implies boundedness of $L(E \int_{\mathbb{R}^d} x \hat{\mu}^n(dx))$, this follows, since L has compact level sets.

3.2 Representation for Functionals of Brownian Motion

Let (Ω, \mathcal{F}, P) be a probability space and $T \in (0, \infty)$. A filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ is a collection of sub-sigma fields of \mathcal{F} with the property $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$. A filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ is called right continuous if $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$ for every $t \in [0, T)$. A filtration $\{\mathcal{F}_t\}_{t \in [0, T]}$ is said to satisfy **the usual conditions** if it is right continuous and for every $t \in [0, T]$, \mathcal{F}_t contains all P -null sets in \mathcal{F} . All filtrations in this book will satisfy the usual conditions. Suppose we are given such a filtration $\{\mathcal{F}_t\}$ on (Ω, \mathcal{F}, P) and that $\{W(t)\}_{0 \leq t \leq T}$ is a k -dimensional \mathcal{F}_t -**Brownian motion**, i.e., $W(0) = 0$; W has continuous trajectories; $W(t)$ is \mathcal{F}_t -measurable for every $t \in [0, T]$; and $W(t) - W(s)$ is independent of \mathcal{F}_s for all $0 \leq s \leq t \leq T$ and is normally distributed with mean zero and variance $(t - s)$. A standard choice of \mathcal{F}_t is the sigma-field $\sigma\{W(s) : 0 \leq s \leq t\}$, augmented with all P -null sets, i.e.,

$$\mathcal{G}_t \doteq \sigma\{\sigma\{W(s) : 0 \leq s \leq t\} \vee \mathcal{N}\},$$

where $\mathcal{N} = \{A \subset \Omega : \text{there is } B \in \mathcal{F} \text{ with } A \subset B \text{ and } P(B) = 0\}$.

Definition 3.12 An \mathbb{R}^k -valued stochastic process $\{v(t)\}_{0 \leq t \leq T}$ on (Ω, \mathcal{F}, P) is said to be **\mathcal{F}_t -progressively measurable** if for every $t \in [0, T]$, the map $(s, \omega) \mapsto v(s, \omega)$ from $([0, t] \times \Omega, \mathcal{B}([0, t]) \otimes \mathcal{F}_t)$ to $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is measurable.

Definition 3.13 Let \mathcal{A} [resp., $\tilde{\mathcal{A}}$] denotes the collection of all \mathcal{G}_t -progressively [resp., \mathcal{F}_t -progressively] measurable processes $\{v(t)\}_{0 \leq t \leq T}$ that satisfy the integrability condition $E[\int_0^T \|v(t)\|^2 dt] < \infty$.

The following representation theorem for bounded measurable functionals of a Brownian motion is analogous to the one stated in Proposition 3.1 for functionals of an iid sequence. It is a special case of a representation that will be proved in Chap. 8 [Theorem 8.3]. In the representation, the controlled measures have been replaced by just a control process, and the relative entropy cost is the expected L^2 -norm of this process. Recall that $\mathcal{C}([0, T] : \mathbb{R}^k)$ denotes the space of \mathbb{R}^k -valued continuous functions on $[0, T]$. This space is equipped with the uniform metric, which makes it a Polish space.

Theorem 3.14 Let G be a bounded Borel measurable function mapping $\mathcal{C}([0, T] : \mathbb{R}^k)$ into \mathbb{R} . Then

$$-\log Ee^{-G(W)} = \inf_{v \in \mathcal{A}} E \left[G \left(W + \int_0^T v(s) ds \right) + \frac{1}{2} \int_0^T \|v(s)\|^2 ds \right]. \quad (3.13)$$

Remark 3.15 The proof of this representation first appeared in [32]. The form of the representation closely parallels the corresponding discrete time result for product measure, reflecting the fact that Brownian motion is the integral of “white” noise, and progressive measurability is analogous to the fact that in the representation for iid noises, $\tilde{\mu}_i^n$ is allowed to depend on all controlled noises up to time $i - 1$. In fact, if one replaces W by the corresponding piecewise linear interpolation with interpolation interval $\delta > 0$ (which is equivalent to a collection of $1/\delta$ iid $N(0, \delta)$ random variables) and assumes that the minimizing measures are Gaussian with means $\delta \tilde{v}_i^n$, then the L^2 cost in (3.13) corresponds to $R(N(\delta \tilde{v}_i^n, \delta) \| N(0, \delta)) = \delta \|\tilde{v}_i^n\|^2/2$. The assumption that one can restrict the discrete time measures to those of the form $N(\delta \tilde{v}_i^n, \delta)$ is valid in the limit $\delta \rightarrow 0$, which is why the continuous time representation is in some ways simpler than the corresponding discrete time representation.

Remark 3.16 One can replace the class \mathcal{A} with $\tilde{\mathcal{A}}$ in (3.13) (see Chap. 8). Although in this chapter we use progressively measurable controls (as in [32]), in Chap. 8 these are replaced by predictable controls. For the case of Brownian motion, the two are interchangeable, since any \mathcal{G}_t [resp., \mathcal{F}_t] predictable process satisfying the square integrability condition in Definition 3.13 is in \mathcal{A} [resp., $\tilde{\mathcal{A}}$], and conversely, to any v in \mathcal{A} [resp., $\tilde{\mathcal{A}}$] there is a predictable \tilde{v} in \mathcal{A} [resp., $\tilde{\mathcal{A}}$] such that $v(t, \omega) = \tilde{v}(t, \omega)$ a.s. $dt \times P$; see [168, Remark 3.3.1]. However, predictability is needed for the case of processes with jumps, e.g., systems driven by a Poisson random measure.

We next state a version of the representation that restricts the class of controls to a compact set. For $M \in [0, \infty)$ let

$$S_M \doteq \left\{ \phi \in \mathcal{L}^2([0, T] : \mathbb{R}^k) : \int_0^T \|\phi(s)\|^2 ds \leq M \right\},$$

where $\mathcal{L}^2([0, T] : \mathbb{R}^k)$ is the Hilbert space of square integrable functions from $[0, T]$ to \mathbb{R}^k , and define $\mathcal{A}_{b,M}$ to be the subset of \mathcal{A} such that $v \in \mathcal{A}_{b,M}$ if $v(\omega) \in S_M$ for all $\omega \in \Omega$. Let $\mathcal{A}_b = \bigcup_{M=1}^{\infty} \mathcal{A}_{b,M}$. In the statement of the theorem, we introduce a scaling that will be appropriate for large deviation analysis of small noise diffusions.

Theorem 3.17 *Let G be a bounded Borel measurable function mapping $\mathcal{C}([0, T] : \mathbb{R}^k)$ into \mathbb{R} and let $\delta > 0$. Then there exist $M < \infty$ depending on $\|G\|_{\infty}$ and δ such that for all $\varepsilon \in (0, 1)$,*

$$\begin{aligned} & -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} G(\sqrt{\varepsilon} W) \right\} \\ & \geq \inf_{v \in \mathcal{A}_{b,M}} E \left[G \left(\sqrt{\varepsilon} W + \int_0^T v(s) ds \right) + \frac{1}{2} \int_0^T \|v(s)\|^2 ds \right] - \delta. \end{aligned} \quad (3.14)$$

Proof To consolidate notation, for $v \in \mathcal{A}$ let $W^v \doteq W + \int_0^T v(s) ds$. For the given $\varepsilon \in (0, 1)$ and $\eta \in (0, 1)$, choose $\tilde{v}^\varepsilon \in \mathcal{A}$ such that

$$\begin{aligned} & \inf_{v \in \mathcal{A}} E \left[G \left(\sqrt{\varepsilon} W^{v/\sqrt{\varepsilon}} \right) + \frac{1}{2} \int_0^T \|v\|^2 ds \right] \\ & \geq E \left[G \left(\sqrt{\varepsilon} W^{\tilde{v}^\varepsilon/\sqrt{\varepsilon}} \right) + \frac{1}{2} \int_0^T \|\tilde{v}^\varepsilon\|^2 ds \right] - \eta. \end{aligned}$$

From the boundedness of G it follows that

$$\infty > C_G \doteq 2(2\|G\|_{\infty} + 1) \geq \sup_{\varepsilon \in (0,1)} E \left[\int_0^T \|\tilde{v}^\varepsilon(s)\|^2 ds \right].$$

We next show using an approximation argument that one can in fact assume an almost sure bound. For $M \in (0, \infty)$ let

$$\tau_M^\varepsilon \doteq \inf \left[t \in [0, T] : \int_0^t \|\tilde{v}^\varepsilon(s)\|^2 ds \geq M \right] \wedge T.$$

Note that v^ε defined by $v^\varepsilon(s) \doteq \tilde{v}^\varepsilon(s) 1_{[0, \tau_M^\varepsilon]}(s)$, $s \in [0, T]$ is an element of \mathcal{A} , and that $v^\varepsilon \in S_M$ a.s. Note also that

$$\begin{aligned} & E \left[G \left(\sqrt{\varepsilon} W^{v^\varepsilon/\sqrt{\varepsilon}} \right) + \frac{1}{2} \int_0^T \|\tilde{v}^\varepsilon(s)\|^2 ds \right] \\ & \geq E \left[G \left(\sqrt{\varepsilon} W^{v^\varepsilon/\sqrt{\varepsilon}} \right) + \frac{1}{2} \int_0^T \|v^\varepsilon(s)\|^2 ds \right] \end{aligned}$$

$$+ E \left[G \left(\sqrt{\varepsilon} W^{\tilde{v}^\varepsilon / \sqrt{\varepsilon}} \right) - G \left(\sqrt{\varepsilon} W^{v^\varepsilon / \sqrt{\varepsilon}} \right) \right].$$

By Chebyshev's inequality,

$$E \left[\left| G \left(\sqrt{\varepsilon} W^{\tilde{v}^\varepsilon / \sqrt{\varepsilon}} \right) - G \left(\sqrt{\varepsilon} W^{v^\varepsilon / \sqrt{\varepsilon}} \right) \right| \right] \leq 2 \|G\|_\infty P\{\tau_M^\varepsilon < T\} \leq 2 \|G\|_\infty \frac{C_G}{M}.$$

For $\delta > 0$, let $M = (2 \|G\|_\infty C_G + 1) / \delta$. Then for all $\varepsilon \in (0, 1)$,

$$\begin{aligned} & E \left[G \left(\sqrt{\varepsilon} W^{\tilde{v}^\varepsilon / \sqrt{\varepsilon}} \right) + \frac{1}{2} \int_0^T \|\tilde{v}^\varepsilon(s)\|^2 ds \right] \\ & \geq E \left[G \left(\sqrt{\varepsilon} W^{v^\varepsilon / \sqrt{\varepsilon}} \right) + \frac{1}{2} \int_0^T \|v^\varepsilon(s)\|^2 ds \right] - \delta. \end{aligned}$$

Since $\eta > 0$ is arbitrary, the conclusion of the theorem follows from the last display and Theorem 3.14. \square

3.2.1 Large Deviation Theory of Small Noise Diffusions

The representation (3.13) and its variant (3.14) are very convenient for weak convergence large deviation analysis, and in many ways they make the continuous time setting simpler than the corresponding discrete time setting. As an illustration of their use we prove the large deviation principle for a class of small noise diffusions. While fairly general, the assumptions on the coefficients are chosen to make the presentation simple, and they can be significantly relaxed.

Condition 3.18 *There is $C \in (0, \infty)$ such that $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times k}$ satisfy*

$$\|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| \leq C \|x - y\|, \quad \|b(x)\| + \|\sigma(x)\| \leq C(1 + \|x\|)$$

for all $x, y \in \mathbb{R}^d$.

Fix $x \in \mathbb{R}^d$, and for $\varepsilon > 0$ let $X^\varepsilon = \{X^\varepsilon(t)\}_{0 \leq t \leq T}$ be the strong solution of the stochastic differential equation (SDE) (cf. [172, Sect. 5.2])

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \sqrt{\varepsilon}\sigma(X^\varepsilon(t))dW(t), \quad X^\varepsilon(0) = x. \quad (3.15)$$

Let $\mathcal{A}\mathcal{C}_x([0, T] : \mathbb{R}^d)$ denote the space of \mathbb{R}^d -valued absolutely continuous functions φ on $[0, T]$ with $\varphi(0) = x$. Also, for $\varphi \in \mathcal{A}\mathcal{C}_x([0, T] : \mathbb{R}^d)$, let

$$U_\varphi = \left\{ u \in \mathcal{L}^2([0, T] : \mathbb{R}^k) : \varphi(\cdot) = x + \int_0^\cdot b(\varphi(s))ds + \int_0^\cdot \sigma(\varphi(s))u(s)ds \right\}. \quad (3.16)$$

For all other $\varphi \in \mathcal{C}([0, T] : \mathbb{R}^d)$ let U_φ be the empty set. The following large deviation principle for such small noise diffusions is one of the classical results in the theory [140]. Following our standard convention, the infimum over the empty set is taken to be ∞ .

Theorem 3.19 *Assume Condition 3.18. Then the collection $\{X^\varepsilon\}_{\varepsilon \in (0,1)}$ satisfies the LDP on $\mathcal{C}([0, T] : \mathbb{R}^d)$ with rate function*

$$I(\varphi) \doteq \inf_{u \in U_\varphi} \left[\frac{1}{2} \int_0^T \|u(t)\|^2 dt \right].$$

To prove the theorem, we must show that I is a rate function and for bounded and continuous $F : \mathcal{C}([0, T] : \mathbb{R}^d) \rightarrow \mathbb{R}$,

$$\lim_{\varepsilon \rightarrow 0} -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} F(X^\varepsilon) \right\} = \inf_{\varphi \in \mathcal{C}([0, T] : \mathbb{R}^d)} [F(\varphi) + I(\varphi)].$$

Following a convention that is used here for the first time, we present the proof just for the case $T = 1$, noting that the general case involves only notational differences. The first step is to interpret $F(X^\varepsilon)$ as a bounded measurable function of W . From unique pathwise solvability of the SDE in (3.15) [172, Definition 5.3.2 and Corollary 5.3.23] it follows that for each $\varepsilon > 0$, there is a measurable map $\mathcal{G}^\varepsilon : \mathcal{C}([0, 1] : \mathbb{R}^k) \rightarrow \mathcal{C}([0, 1] : \mathbb{R}^d)$ such that whenever \tilde{W} is a k -dimensional standard Brownian motion given on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, then $\tilde{X}^\varepsilon = \mathcal{G}^\varepsilon(\sqrt{\varepsilon}\tilde{W})$ is the unique solution of the SDE (3.15) with W replaced by \tilde{W} . Recalling the notation $W^v \doteq W + \int_0^\cdot v(s)ds$, this says that

$$\begin{aligned} -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} F(X^\varepsilon) \right\} &= -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} F \circ \mathcal{G}^\varepsilon(\sqrt{\varepsilon}W) \right\} \\ &= \inf_{v \in \mathcal{A}} E \left[F \circ \mathcal{G}^\varepsilon(\sqrt{\varepsilon}W^{v/\sqrt{\varepsilon}}) + \frac{1}{2} \int_0^1 \|v(s)\|^2 ds \right]. \end{aligned}$$

Assume that $v \in \mathcal{A}_{b,M}$ for some $M < \infty$, and consider the probability measure Q^ε on (Ω, \mathcal{F}) defined by

$$\frac{dQ^\varepsilon}{dP} = \exp \left[-\frac{1}{\sqrt{\varepsilon}} \int_0^1 v(s)dW(s) - \frac{1}{2\varepsilon} \int_0^1 \|v(s)\|^2 ds \right].$$

From Girsanov's theorem (see Theorem D.1) it follows that $Q^\varepsilon\{\sqrt{\varepsilon}W^{v/\sqrt{\varepsilon}} \in \cdot\} = P\{\sqrt{\varepsilon}W \in \cdot\}$. Consequently $\tilde{X}^\varepsilon = \mathcal{G}^\varepsilon(\sqrt{\varepsilon}W^{v/\sqrt{\varepsilon}})$ solves the SDE

$$d\tilde{X}^\varepsilon(t) = b(\tilde{X}^\varepsilon(t))dt + \sqrt{\varepsilon}\sigma(\tilde{X}^\varepsilon(t))dW^{v/\sqrt{\varepsilon}}(t), \quad \tilde{X}^\varepsilon(0) = x$$

on the filtered probability space $(\Omega, \mathcal{F}, Q^\varepsilon, \{\mathcal{F}_t\})$. Since Q^ε is mutually absolutely continuous with respect to P , it follows that \bar{X}^ε is the unique solution of the following SDE on $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\})$:

$$d\bar{X}^\varepsilon(t) = b(\bar{X}^\varepsilon(t))dt + \sqrt{\varepsilon}\sigma(\bar{X}^\varepsilon(t))dW(t) + \sigma(\bar{X}^\varepsilon(t))v(t)dt, \quad \bar{X}^\varepsilon(0) = x. \quad (3.17)$$

Thus whenever $v \in \mathcal{A}_{b,M}$, we have that $\mathcal{G}^\varepsilon(\sqrt{\varepsilon}W^{v/\sqrt{\varepsilon}})$ and the solution to (3.17) coincide. A collection of controls $\{v^\varepsilon\} \subset \mathcal{A}_{b,M}$ for fixed $M < \infty$ will be regarded as a collection of S_M -valued random variables, where S_M is equipped with the weak topology on the Hilbert space $\mathcal{L}^2([0, 1] : \mathbb{R}^d)$. Recall that in a Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, $f_n \rightarrow f$ under **the weak topology** if for all $g \in \mathcal{H}$, $\langle f_n - f, g \rangle \rightarrow 0$. Since S_M is weakly compact in $\mathcal{L}^2([0, 1] : \mathbb{R}^d)$, such a collection is automatically tight.

We now turn to the proof of the LDP, which will follow the same scheme of proof as in Sanov's theorem. Thus we first prove a tightness result and show how to relate the weak limits of controls and controlled processes. The proof of the variational lower bound (which corresponds to the Laplace upper bound) as well as the proof that I is a rate function follows, and we conclude with the proof of the variational upper bound (Laplace lower bound).

3.2.2 Tightness and Weak Convergence

As noted above, a collection of controls $\{v^\varepsilon\} \subset \mathcal{A}_{b,M}$ is trivially tight, since S_M is compact. The following lemma shows that the corresponding collection of solutions of controlled SDEs is also tight.

Lemma 3.20 *Assume Condition 3.18. Consider any collection of controls $\{v^\varepsilon\} \subset \mathcal{A}_{b,M}$ for fixed $M < \infty$, and define \bar{X}^ε by (3.17) with $v = v^\varepsilon$. Then $\{(\bar{X}^\varepsilon, v^\varepsilon)\}_{\varepsilon \in (0,1)}$ is a tight collection of $\mathcal{C}([0, 1] : \mathbb{R}^d) \times S_M$ -valued random variables.*

Proof Tightness of $\{v^\varepsilon\}$ is immediate. Since for $\varepsilon \in (0, 1)$, $\int_0^1 \|v^\varepsilon(s)\|^2 ds \leq M$ a.s., it follows on using the linear growth properties of the coefficients and an application of Gronwall's lemma [Lemma E.2] that

$$\sup_{\varepsilon \in (0,1)} E \sup_{0 \leq t \leq 1} \|\bar{X}^\varepsilon(t)\|^2 < \infty. \quad (3.18)$$

Also note that

$$\bar{X}^\varepsilon(t) - x = \int_0^t b(\bar{X}^\varepsilon(s))ds + \sqrt{\varepsilon} \int_0^t \sigma(\bar{X}^\varepsilon(s))dW(s) + \int_0^t \sigma(\bar{X}^\varepsilon(s))v^\varepsilon(s)ds. \quad (3.19)$$

The first and second terms on the right side are easily seen to be tight in $\mathcal{C}([0, 1] : \mathbb{R}^d)$ using the moment bound (3.18). Tightness of the third follows on using the inequality

$$\begin{aligned} \left\| \int_s^t \sigma(\bar{X}^\varepsilon(r)) v^\varepsilon(r) dr \right\| &\leq C(t-s)^{1/2} \left(1 + \sup_{0 \leq t \leq 1} \|\bar{X}^\varepsilon(t)\| \right) \left(\int_0^1 \|v^\varepsilon(r)\|^2 dr \right)^{1/2} \\ &\leq C(t-s)^{1/2} M^{1/2} \left(1 + \sup_{0 \leq t \leq 1} \|\bar{X}^\varepsilon(t)\| \right) \end{aligned}$$

for $0 \leq s \leq t \leq 1$ and once more using the moment bound. \square

The following lemma will be used to characterize the limit points of $\{(\bar{X}^\varepsilon, v^\varepsilon)\}$.

Lemma 3.21 *Assume Condition 3.18. Suppose for each $\varepsilon \in (0, 1)$ that $(\bar{X}^\varepsilon, v^\varepsilon)$ solves (3.19), and that $(\bar{X}^\varepsilon, v^\varepsilon)$ converges weakly to (\bar{X}, v) as $\varepsilon \rightarrow 0$. Then w.p.1,*

$$\bar{X}(t) - x = \int_0^t b(\bar{X}(s)) ds + \int_0^t \sigma(\bar{X}(s)) v(s) ds. \quad (3.20)$$

Proof By a standard martingale bound (see (D.3) and Sect. D.2.1),

$$E \sup_{0 \leq t \leq T} \left\| \int_0^t \sigma(\bar{X}^\varepsilon(r)) dW(r) \right\|^2 \leq C \int_0^T E (1 + \|\bar{X}^\varepsilon(r)\|^2) dr,$$

and thus using the moment bound in (3.18), the stochastic integral term in (3.19) converges to 0 as $\varepsilon \rightarrow 0$. By the continuous mapping theorem, it suffices to check that for each $t \in [0, 1]$, the maps $\phi \mapsto \int_0^t b(\phi(s)) ds$ and $(\phi, u) \mapsto \int_0^t \sigma(\phi(s)) u(s) ds$, from $\mathcal{C}([0, 1] : \mathbb{R}^d)$ to \mathbb{R}^d and from $\mathcal{C}([0, 1] : \mathbb{R}^d) \times S_M$ to \mathbb{R}^d , are continuous. The continuity of the first map is immediate from the Lipschitz property of b . Consider now the second map. Suppose $\phi_n \rightarrow \phi$ in $\mathcal{C}([0, 1] : \mathbb{R}^d)$ and $u_n \rightarrow u$ in S_M as $n \rightarrow \infty$. We can write

$$\begin{aligned} &\int_0^t \sigma(\phi_n(s)) u_n(s) ds - \int_0^t \sigma(\phi(s)) u(s) ds \\ &= \int_0^t [\sigma(\phi_n(s)) - \sigma(\phi(s))] u_n(s) ds + \int_0^t \sigma(\phi(s)) [u_n(s) - u(s)] ds. \end{aligned}$$

The first term tends to zero by Hölder's inequality and since $u_n \in S_M$, and the second converges to zero since $s \mapsto \sigma(\phi(s)) 1_{[0, t]}(s)$ is in $\mathcal{L}^2([0, 1] : \mathbb{R}^d)$ and $u_n \rightarrow u$ in S_M . \square

3.2.3 Laplace Upper Bound

We now prove the Laplace upper bound by establishing the lower bound

$$\liminf_{\varepsilon \rightarrow 0} -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} F(X^\varepsilon) \right\} \geq \inf_{\varphi \in \mathcal{C}([0, 1] : \mathbb{R}^d)} [F(\varphi) + I(\varphi)]. \quad (3.21)$$

We prove (3.21) using the variational representation. It suffices to show that for every sequence $\varepsilon_k \rightarrow 0$ there is a further subsequence for which (3.21) holds when the limit inferior on the left side is taken along the particular subsequence. Let $\delta > 0$, and with $G = F \circ \mathcal{G}^\varepsilon$ choose M according to Theorem 3.17 (note that M does not depend on ε), and choose a sequence $\{v^\varepsilon\} \subset \mathcal{A}_{b,M}$ that is within δ of the infimum in (3.14). We now fix a sequence $\{\varepsilon_k\}$. From Lemma 3.20 we can find a subsequence along which $(\bar{X}^{\varepsilon_k}, v^{\varepsilon_k})$ converges in distribution. For notational convenience, we index this subsequence once more by ε . Denoting the weak limit of $(\bar{X}^\varepsilon, v^\varepsilon)$ by (\bar{X}, v) , we have from Lemma 3.21 that \bar{X} is the unique solution of (3.20). Therefore

$$\begin{aligned} & \liminf_{\varepsilon \rightarrow 0} -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} F(X^\varepsilon) \right\} + 2\delta \\ & \geq \liminf_{\varepsilon \rightarrow 0} E \left[F(\bar{X}^\varepsilon) + \frac{1}{2} \int_0^1 \|v^\varepsilon(s)\|^2 ds \right] \\ & \geq E \left[F(\bar{X}) + \frac{1}{2} \int_0^1 \|v(s)\|^2 ds \right] \\ & \geq \inf_{\varphi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [F(\varphi) + I(\varphi)]. \end{aligned}$$

Here the second inequality is a consequence of Fatou's lemma and the lower semi-continuity of the map $\phi \mapsto \int_0^1 \|\phi(s)\|^2 ds$ from $\mathcal{L}^2([0,1]; \mathbb{R}^d)$ to \mathbb{R} with the weak topology on $\mathcal{L}^2([0,1]; \mathbb{R}^d)$. Recalling the definition of U_φ in (3.16), the last inequality follows from the a.s. inequality

$$\begin{aligned} F(\bar{X}) + \frac{1}{2} \int_0^1 \|v(s)\|^2 ds & \geq F(\bar{X}) + \inf_{u \in U_{\bar{X}}} \left[\frac{1}{2} \int_0^1 \|u(s)\|^2 ds \right] \\ & \geq \inf_{\varphi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [F(\varphi) + I(\varphi)]. \end{aligned}$$

Since $\delta > 0$ is arbitrary, (3.21) follows. \square

3.2.4 Compactness of Level Sets

We now argue that I introduced in Theorem 3.19 is a rate function, which requires that we show that it has compact level sets. As we will see, it is essentially just a deterministic version of the argument used for the Laplace upper bound (variational lower bound). This is in fact generic in the weak convergence approach to large deviations and not at all surprising, in that the main difference between these two arguments is that the variational lower bound has the additional complication of a law of large numbers limit as the large deviation parameter tends to its limit, an item missing in the corresponding and purely deterministic analysis of the rate function.

Let $M \in (0, \infty)$ and let $\{\varphi_n\} \subset \mathcal{C}([0, 1] : \mathbb{R}^d)$ be a sequence such that $I(\varphi_n) \leq M$ for all $n \in \mathbb{N}$. Choose $u_n \in U_{\varphi_n}$ such that $\frac{1}{2} \int_0^1 \|u_n(s)\|^2 ds \leq M + 1/n$. Then the sequence $\{u_n\}$ is contained in the (weakly) compact set $\mathcal{S}_{2(M+1)}$. Let u be a limit point of u_n along some subsequence. Then $\frac{1}{2} \int_0^1 \|u(s)\|^2 ds \leq M$. Also, a simpler version of an argument in the proof of Lemma 3.21 shows that along the same subsequence, $\varphi_n(\cdot)$ converges to $\varphi(\cdot)$, where φ is the unique solution of $\varphi(t) = x + \int_0^t (b(\varphi(s)) + \sigma(\varphi(s))u(s)) ds$. In particular, $u \in U_\varphi$ and thus $I(\varphi) \leq M$. This proves the compactness of level sets of I . \square

3.2.5 Laplace Lower Bound

To prove the Laplace lower bound we use the variational representation to show that

$$\limsup_{\varepsilon \rightarrow 0} -\varepsilon \log E \exp \left\{ -\frac{1}{\varepsilon} F(X^\varepsilon) \right\} \leq \inf_{\varphi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [F(\varphi) + I(\varphi)].$$

For $\delta > 0$ choose $\varphi^* \in \mathcal{C}([0, 1] : \mathbb{R}^d)$ such that

$$F(\varphi^*) + I(\varphi^*) \leq \inf_{\varphi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [F(\varphi) + I(\varphi)] + \delta.$$

Let $u \in U_{\varphi^*}$ be such that $\frac{1}{2} \int_0^1 \|u(s)\|^2 ds \leq I(\varphi^*) + \delta$, so that in particular, $u \in \mathcal{A}_{b, 2(I(\varphi^*) + \delta)}$. Let \bar{X}^ε be the unique solution of (3.17) when we replace v on the right side of the equation by u . By Lemmas 3.20 and 3.21 on tightness and weak convergence, \bar{X}^ε converges in probability to φ^* . Thus

$$\begin{aligned} & \limsup_{\varepsilon \rightarrow 0} -\varepsilon \log E e^{-\frac{1}{\varepsilon} F(X^\varepsilon)} \\ &= \limsup_{\varepsilon \rightarrow 0} \inf_{v \in \mathcal{A}} E \left[F \circ \mathcal{G}^\varepsilon(\sqrt{\varepsilon} W^{v/\sqrt{\varepsilon}}) + \frac{1}{2} \int_0^1 \|v(s)\|^2 ds \right] \\ &\leq \limsup_{\varepsilon \rightarrow 0} E \left[F(\bar{X}^\varepsilon) + \frac{1}{2} \int_0^1 \|u(s)\|^2 ds \right] \\ &= F(\varphi^*) + \frac{1}{2} \int_0^1 \|u(s)\|^2 ds \\ &\leq F(\varphi^*) + I(\varphi^*) + \delta \\ &\leq \inf_{\varphi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [F(\varphi) + I(\varphi)] + 2\delta. \end{aligned}$$

Since $\delta > 0$ is arbitrary, the upper bound follows. \square

Remark 3.22 One can consider the form

$$I(\varphi) \doteq \inf_{u \in U_\varphi} \left[\frac{1}{2} \int_0^T \|u(t)\|^2 dt \right]$$

of the rate function, where U_φ are those u satisfying $\varphi(t) = x + \int_0^t b(\varphi(s))ds + \int_0^t \sigma(\varphi(s))u(s)ds$, as a “control” formulation. If $\sigma(x)$ is $d \times d$ and invertible for all $x \in \mathbb{R}^d$, then one can solve for u and obtain the calculus of variations form

$$I(\varphi) \doteq \int_0^T \frac{1}{2} \langle (\dot{\varphi}(t) - b(\varphi(t))), [\sigma \sigma^T(\varphi(t))]^{-1} (\dot{\varphi}(t) - b(\varphi(t))) \rangle dt,$$

where σ^T is the transpose of σ .

3.3 Representation for Functionals of a Poisson Process

Our final example in this chapter is the representation for positive functionals of a Poisson process. This example will be substantially generalized in Chap. 8, where we prove the representation for a Poisson random measure (PRM) on an arbitrary locally compact Polish space. The representation for a PRM allows the treatment of a much broader class of process models, and in particular when used as a driving noise, a PRM can easily accommodate both state-dependent jump rates and jumps sizes, while a Poisson process (which is essentially a PRM with only one “type” of point) is limited to state dependence of jump sizes. However, the purpose of this chapter is to illustrate the use of representations, and we prefer to postpone the notation and terminology required for the general case of a PRM.

Fix $T \in (0, \infty)$ and let (Ω, \mathcal{F}, P) be a probability space with filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ satisfying the usual conditions. Recall that $\mathcal{D}([0, T] : \mathbb{R})$ is the space of functions from $[0, T]$ to \mathbb{R} that are right continuous and with limits from the left at each $t \in (0, T]$. As noted in Chap. 2, there is a metric that is consistent with the usual Skorohod topology that makes this a Polish space [24, Chap. 3, Sect. 12]. An \mathcal{F}_t -**Poisson process** is a measurable mapping N from Ω into $\mathcal{D}([0, T] : \mathbb{R})$ such that $N(t)$ is \mathcal{F}_t -measurable for every $t \in [0, T]$, and for all $0 \leq t < s \leq T$, $N(s) - N(t)$ is independent of \mathcal{F}_t and has a Poisson distribution with parameter $s - t$: $P(N(s) - N(t) = j) = (s - t)^j e^{-(s-t)} / j!$. We say that such a standard Poisson process has **jump intensity** or **jump rate** 1, since the probability that $N(s) - N(t) = 1$ is approximately $s - t$ when this difference is small [and the probability of more than one jump is $o(s - t)$].

In contrast to the case of Brownian motion, in which the natural controlled version shifts the mean, here the controlled version will shift the jump rate and pay the appropriate cost suggested by Girsanov’s theorem for Poisson processes (see, for example, Theorem 8.15). There are various ways to construct Poisson processes with general jump rates on a common probability space. The most convenient one requires the use of a PRM on the space $[0, T] \times [0, \infty)$ and with intensity measure equal to Lebesgue measure on this space (see Chap. 8 for definitions and associated terminology). In this framework the Poisson process on $[0, T]$ is considered a PRM

on $[0, T]$, and to accommodate general controls we suitably enlarge the space. We do not give the details here, but instead just state the outcome of this construction.

One can construct a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, and on this space a filtration $\{\bar{\mathcal{F}}_t\}_{0 \leq t \leq T}$ satisfying the usual conditions, such that the following properties hold. Let $\theta \in (0, \infty)$ (later θ will play the role of a large deviation parameter). Denote by \mathcal{A} the collection of predictable processes $\varphi : [0, T] \times \bar{\Omega} \rightarrow [0, \infty)$ (see Definition 8.2 for the definition of predictability in a general setting) such that $\int_0^T \varphi(s) ds < \infty$ a.s. Predictable processes are in a suitable way not allowed to anticipate the jumps of a Poisson process with respect to the same filtration, and hence are the appropriate analogue of the class of controls used for representations in discrete time. Associated with each $\varphi \in \mathcal{A}$ one can construct a “controlled” Poisson process $N^{\theta\varphi}$ with jump intensity $\theta\varphi$ and jump size 1. To be precise, $N^{\theta\varphi}$ is an $\bar{\mathcal{F}}_t$ -adapted stochastic process with trajectories in $\mathcal{D}([0, T] : \mathbb{R})$ such that for every bounded function $f : [0, \infty) \rightarrow [0, \infty)$,

$$f(N^{\theta\varphi}(t)) - f(0) - \theta \int_0^t \varphi(s) [f(N^{\theta\varphi}(s) + 1) - f(N^{\theta\varphi}(s))] ds$$

is an $\bar{\mathcal{F}}_t$ -martingale, and $N^{\theta\varphi}(0) = 0$. Note that N^θ is an ordinary Poisson process with constant jump intensity θ and jump size 1.

In terms of these controls and controlled processes, we have the following representation. Recall the function ℓ introduced in (3.10): for $r \in [0, \infty)$, we have $\ell(r) \doteq r \log r - r + 1$, with the convention that $0 \log 0 = 0$. We consider all processes $N^{\theta\varphi}$ to be random variables with values in $\mathcal{D}([0, T] : \mathbb{R})$. We also introduce

$$S_M \doteq \left\{ \phi \in \mathcal{L}^0([0, T] : \mathbb{R}_+) : \int_0^T \ell(\phi(s)) ds \leq M \right\},$$

where $\mathcal{L}^0([0, T] : \mathbb{R}_+)$ denotes the space of Borel-measurable functions from $[0, T]$ to $[0, \infty)$, and given $M \in (0, \infty)$ define $\mathcal{A}_{b,M}$ to be the subset of \mathcal{A} such that $\varphi \in \mathcal{A}_{b,M}$ implies $\varphi(\omega) \in S_M$ for all $\omega \in \bar{\Omega}$ and for some $K \in (0, \infty)$ (possibly depending on φ), $K^{-1} \leq \varphi \leq K$, a.s. Also, let $\mathcal{A}_b = \cup_{M=1}^\infty \mathcal{A}_{b,M}$. The spaces S_M , $\mathcal{A}_{b,M}$, \mathcal{A}_b in this section play an analogous role for Poisson processes to that of the corresponding spaces introduced in Sect. 3.2 for the Brownian motion case.

Theorem 3.23 *Let G be a bounded Borel measurable function mapping $\mathcal{D}([0, T] : \mathbb{R})$ into \mathbb{R} and let $\theta \in (0, \infty)$. Then*

$$-\log E \exp\{-G(N^\theta)\} = \inf_{\varphi \in \mathcal{A}} E \left[G(N^{\theta\varphi}) + \theta \int_0^T \ell(\varphi(s)) ds \right].$$

If $\delta > 0$, then there exists $M < \infty$ depending on $\|G\|_\infty$ and δ such that for all $\theta \in (0, \infty)$,

$$-\frac{1}{\theta} \log E \exp \{-\theta G(N^\theta)\} \geq \inf_{\varphi \in \mathcal{A}_{b,M}} E \left[G(N^{\theta\varphi}) + \int_0^T \ell(\varphi(s)) ds \right] - \delta. \quad (3.22)$$

The proof of Theorem 3.23 follows as a special case of more general results [Theorems 8.12 and 8.13] that will be proved in Chap. 8. In particular, the general result will show that \mathcal{A} can be replaced by \mathcal{A}_b in the first representation. We now show how this representation can be used to obtain a large deviation principle for SDEs driven by a Poisson process. We begin with a condition on the coefficients that can be relaxed substantially (see, for example, Chap. 10).

Condition 3.24 *There is $C \in (0, \infty)$ such that $b : \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ satisfy*

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq C |x - y| \quad \text{and} \quad |b(x)| + |\sigma(x)| \leq C$$

for all $x, y \in \mathbb{R}$.

Fix $x \in \mathbb{R}$, and for $n \in \mathbb{N}$ let $X^n = \{X^n(t)\}_{0 \leq t \leq T}$ be the pathwise solution of the SDE

$$dX^n(t) = b(X^n(t))dt + \frac{1}{n}\sigma(X^n(t-))dN^n(t), \quad X^n(0) = x, \quad (3.23)$$

where $X^n(t-)$ denotes the limit from the left. One can explicitly construct the solution in terms of the jump times $\{t_i^n\}_{i \in \mathbb{N}}$ of $N^n(\cdot)$. With probability 1, these jump times satisfy $0 < t_1^n < t_2^n < \dots$ and $t_i^n \rightarrow \infty$. Letting $t_0^n = 0$ and $X^n(t_0^n) = x$, we then recursively define $X^n(t)$ as follows. Assuming that $X^n(t_i^n)$ is given, let

$$\dot{X}^n(t) = b(X^n(t)) \text{ for } t \in (t_i^n, t_{i+1}^n)$$

and then set $X^n(t_{i+1}^n) \doteq X^n(t_{i+1}^n-) + \sigma(X^n(t_{i+1}^n-))/n$. With $X^n(t_{i+1}^n)$ now given, we repeat the procedure, and since $t_i^n \rightarrow \infty$, the construction on $[0, T]$ is well defined.

For $\psi \in \mathcal{A}\mathcal{C}_x([0, T] : \mathbb{R})$, let

$$U_\psi = \left\{ \gamma \in \mathcal{L}^1([0, T] : \mathbb{R}_+) : \psi(\cdot) = x + \int_0^\cdot b(\psi(s))ds + \int_0^\cdot \sigma(\psi(s))\gamma(s)ds \right\}, \quad (3.24)$$

where $\mathcal{L}^1([0, T] : \mathbb{R}_+)$ is the space of \mathbb{R}_+ -valued integrable functions on $[0, T]$.

Theorem 3.25 *Assume Condition 3.24. Then the collection $\{X^n\}_{n \in \mathbb{N}}$ satisfies the LDP on $\mathcal{D}([0, T] : \mathbb{R})$ with rate function*

$$I(\psi) \doteq \inf_{\gamma \in U_\psi} \left[\int_0^T \ell(\gamma(t))dt \right].$$

The proof of this theorem is a close parallel to that of Brownian motion, and because of this we do not separate the proof into a series of statements (lemmas, propositions, etc.) and their proofs. We must show that I is a rate function and that for every bounded and continuous $F : \mathcal{D}([0, T] : \mathbb{R}) \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log E \exp \{-nF(X^n)\} = \inf_{\psi \in \mathcal{D}([0, T]: \mathbb{R})} [F(\psi) + I(\psi)].$$

Following our convention, we consider just the case $T = 1$. We have already explicitly identified the measurable map $\mathcal{G}^n : \mathcal{D}([0, 1] : \mathbb{R}) \rightarrow \mathcal{D}([0, 1] : \mathbb{R})$ such that whenever \tilde{N}^n is a Poisson process with rate n on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, then $\bar{X}^n = \mathcal{G}^n(\tilde{N}^n)$ is the unique solution of the SDE (3.23) with N^n replaced by \tilde{N}^n . Hence by Theorem 3.23 with $\theta = n$,

$$\begin{aligned} -\frac{1}{n} \log E \exp \{-nF(X^n)\} &= -\frac{1}{n} \log E \exp \{-nF \circ \mathcal{G}^n(N^n)\} \\ &= \inf_{\varphi \in \mathcal{A}} E \left[F \circ \mathcal{G}^n(N^{n\varphi}) + \int_0^1 \ell(\varphi(t)) dt \right]. \end{aligned}$$

Analogous to the case of Brownian motion, if $\varphi \in \mathcal{A}_{b, M}$ for some $M < \infty$, then $\bar{X}^n = \mathcal{G}^n(N^{n\varphi})$ is the solution of the SDE

$$d\bar{X}^n(t) = b(\bar{X}^n(t))dt + \sigma(\bar{X}^n(t-))dN^{n\varphi}(t), \quad \bar{X}^n(0) = x. \quad (3.25)$$

Here, the important property that follows from $\varphi \in \mathcal{A}_{b, M}$ is that it guarantees (as easily follows from Girsanov's formula) that the jump times of $N^{n\varphi}$ tend to ∞ w.p.1, and so the recursive construction of \bar{X}^n is well defined on $[0, 1]$.

A distinction with respect to the case of Brownian motion is that it is no longer appropriate to consider S_M as a subset of a Hilbert space. Instead, we will identify S_M with a compact space of measures. In particular, associated with each element γ of S_M is a measure ν^γ on $([0, 1], \mathcal{B}([0, 1]))$ defined by $\nu^\gamma(ds) \doteq \gamma(s)m(ds)$, where m denotes Lebesgue measure. As discussed in Lemma A.11, when considered with the natural generalization of the weak topology from probability measures to measures with finite total measure, S_M is a compact Polish space.

Next suppose that Condition 3.24 holds. Consider any collection of controls $\{\varphi^n\} \subset \mathcal{A}_{b, M}$ for fixed $M < \infty$, and define \bar{X}^n by (3.25) with $\varphi = \varphi^n$. We claim that $\{(\bar{X}^n, \varphi^n)\}_{n \in \mathbb{N}}$ is a tight collection of $\mathcal{D}([0, 1] : \mathbb{R}) \times S_M$ -valued random variables. Tightness of $\{\varphi^n\}$ follows from the compactness of S_M . For the tightness of $\{\bar{X}^n\}$, we consider the Doob decomposition

$$\begin{aligned} \bar{X}^n(t) - x &= \int_0^t b(\bar{X}^n(s))ds + \frac{1}{n} \int_0^t \sigma(\bar{X}^n(s-))dN^{n\varphi^n}(s) \\ &= \int_0^t b(\bar{X}^n(s))ds + \int_0^t \sigma(\bar{X}^n(s))\varphi^n(s)ds \\ &\quad + \int_0^t \sigma(\bar{X}^n(s-))[dN^{n\varphi^n}(s)/n - \varphi^n(s)]ds. \end{aligned} \quad (3.26)$$

Since the restriction of the Skorohod metric to $\mathcal{C}([0, 1] : \mathbb{R})$ is equivalent to the standard uniform metric, it suffices, for the first two terms, to show tightness in

$\mathcal{C}([0, 1] : \mathbb{R})$. Tightness of the first follows from $\|b\|_\infty \leq C$. For the second term we use the bound $ab \leq e^{ca} + \ell(b)/c$, valid for $a \geq 0, b \geq 0$ and $c \geq 1$ [see (2.9)]. For all $0 \leq s \leq t \leq 1$,

$$\int_s^t \sigma(\bar{X}^n(r))\varphi^n(r)dr \leq \int_s^t [e^{c\|\sigma\|_\infty} + \ell(\varphi^n(r))/c]dr \leq (t-s)e^{c\|\sigma\|_\infty} + \frac{1}{c}M.$$

This shows equicontinuity of the second term in (3.26) that is uniform in ω , and tightness of that term follows. Let $Q^n(t)$ denote the third term. This term is a martingale with quadratic variation (see Sects. D.1 and D.2.2) $[Q^n]_t$ bounded above by

$$\frac{1}{n^2} \|\sigma\|_\infty^2 EN^{n\varphi^n}(1) = \frac{1}{n} \|\sigma\|_\infty^2 E \int_0^1 \varphi^n(s)ds \leq \frac{1}{n} \|\sigma\|_\infty^2 (e + M),$$

where $b \leq e + \ell(b)$ is used for the last inequality. By the Burkholder–Gundy–Davis inequality (see (D.3) in Sect. D.1), $E \sup_{t \in [0,1]} |Q^n(t)| \leq C_1 E[Q^n]_1^{1/2} \rightarrow 0$ for some $C_1 \in (0, \infty)$. Thus by Chebyshev's inequality, Q^n converges weakly to zero uniformly in t , which both shows tightness and identifies the limit. Since all three terms on the right-hand side of (3.26) are tight (and limit points are continuous a.s.), so is $\{\bar{X}^n\}$.

To identify weak limits along any convergent subsequence, we need to know that if $\gamma_n \rightarrow \gamma$ in S_M and $\psi_n \rightarrow \psi$ uniformly, then

$$\int_0^t \sigma(\psi_n(s))\gamma_n(s)ds \rightarrow \int_0^t \sigma(\psi(s))\gamma(s)ds. \quad (3.27)$$

Again using $b \leq e + \ell(b)$, we have

$$\begin{aligned} \left| \int_0^t [\sigma(\psi_n(s)) - \sigma(\psi(s))]\gamma_n(s)ds \right| &\leq \sup_{s \in [0,1]} |\sigma(\psi_n(s)) - \sigma(\psi(s))| \int_0^t \gamma_n(s)ds \\ &\leq \sup_{s \in [0,1]} |\sigma(\psi_n(s)) - \sigma(\psi(s))| (e + M) \\ &\rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. To show that

$$\int_0^t \sigma(\psi(s))[\gamma_n(s) - \gamma(s)]ds \rightarrow 0,$$

we use that $v^{\gamma_n}(ds) \doteq \gamma_n(s)m(ds)$ converges in the weak topology to $v^\gamma(ds)$. Since $s \mapsto 1_{[0,t]}(s)\sigma(\psi(s))$ is bounded and discontinuous only at $s = t$ and $v^\gamma(\{t\}) = 0$, the last display is valid, and this completes the proof of (3.27).

Consider any subsequence of $\{(\bar{X}^n, \varphi^n)\}_{n \in \mathbb{N}}$ that converges in distribution with limit (\bar{X}, φ) . Sending $n \rightarrow \infty$ in (3.26) and using (3.27) establishes the w.p.1 relation

$$\bar{X}(t) - x = \int_0^t b(\bar{X}(s))ds + \int_0^t \sigma(\bar{X}(s))\varphi(s)ds. \quad (3.28)$$

The rest of the proof is now essentially identical to that for Brownian motion. For the Laplace upper bound, we need to show that

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log E \exp \{-nF(X^n)\} \geq \inf_{\psi \in \mathcal{D}([0,1];\mathbb{R})} [F(\psi) + I(\psi)].$$

Let $\delta > 0$, choose M according to Theorem 3.23, and choose a sequence $\{\varphi^n\} \subset \mathcal{A}_{b,M}$ that is within δ of the infimum in the representation (3.22) (with G replaced by $F \circ \mathcal{G}^n$ and θ replaced by n). Fix any subsequence of n and choose a further subsequence (again denoted by n) along which (\bar{X}^n, φ^n) converges in distribution to (\bar{X}, φ) . Then

$$\begin{aligned} & \liminf_{n \rightarrow \infty} -\frac{1}{n} \log E \exp \{-nF(X^n)\} + 2\delta \\ & \geq \liminf_{n \rightarrow \infty} E \left[F(\bar{X}^n) + \int_0^1 \ell(\varphi^n(s))ds \right] \\ & \geq E \left[F(\bar{X}) + \int_0^1 \ell(\varphi(s))ds \right] \\ & \geq \inf_{\psi \in \mathcal{D}([0,1];\mathbb{R})} [F(\psi) + I(\psi)], \end{aligned}$$

where the second inequality uses Fatou's lemma and the lower semicontinuity of the map $\varphi \mapsto \int_0^1 \ell(\varphi(s))ds$ from S_M to $[0, \infty)$. Recalling the definition of U_ψ in (3.24), the last inequality is a consequence of the a.s. inequality

$$\begin{aligned} F(\bar{X}) + \int_0^1 \ell(\varphi(s))ds & \geq F(\bar{X}) + \inf_{\varphi \in U_{\bar{X}}} \left[\int_0^1 \ell(\varphi(s))ds \right] \\ & \geq \inf_{\psi \in \mathcal{D}([0,1];\mathbb{R})} [F(\psi) + I(\psi)]. \end{aligned}$$

Since $\delta > 0$ is arbitrary, the Laplace upper bound follows.

As in Sect. 3.2.4, a deterministic version of the argument used for the Laplace upper bound gives the compactness of level sets for the rate function, and so this argument is omitted. To complete the proof, all that remains is the Laplace lower bound, which requires that for bounded and continuous F ,

$$\limsup_{n \rightarrow \infty} -\frac{1}{n} \log E \exp \{-nF(X^n)\} \leq \inf_{\psi \in \mathcal{D}([0,1];\mathbb{R})} [F(\psi) + I(\psi)].$$

For $\delta > 0$ choose $\psi^* \in \mathcal{D}([0,1];\mathbb{R})$ such that

$$F(\psi^*) + I(\psi^*) \leq \inf_{\psi \in \mathcal{D}([0,1];\mathbb{R})} [F(\psi) + I(\psi)] + \delta.$$

Let $\varphi \in U_{\psi^*}$ be such that $\int_0^1 \ell(\varphi(s))ds \leq I(\psi^*) + \delta$. We now approximate φ with an element in $\mathcal{A}_{b,M}$, where $M = I(\psi^*) + \delta$. For $q \in \mathbb{N}$ let

$$\varphi_q(t) = \left(\varphi(t) \vee \frac{1}{q} \right) \wedge q.$$

Then $\varphi_q \in \mathcal{A}_{b,M}$ and $\int_0^1 \ell(\varphi_q(s))ds \uparrow \int_0^1 \ell(\varphi(s))ds$ as $q \rightarrow \infty$. Let ψ_q^* be the solution of (3.28) with φ replaced by φ_q . It is easily seen that $\psi_q^* \rightarrow \psi^*$ in $\mathcal{C}([0, 1] : \mathbb{R})$ as $q \rightarrow \infty$. Let \bar{X}^n be the unique solution of (3.25) with φ replaced by φ_q . The tightness of (\bar{X}^n, φ_q) and identification of limits is exactly as in the proof of the Laplace upper bound, since $\varphi_q \in \mathcal{A}_{b,M}$. Using the uniqueness of solutions to the limit ordinary differential equation (ODE), \bar{X}^n converges in probability to ψ_q^* . Thus

$$\begin{aligned} \limsup_{n \rightarrow \infty} -\frac{1}{n} \log E e^{-nF(X^n)} &\leq \limsup_{n \rightarrow \infty} E \left[F(\bar{X}^n) + \int_0^1 \ell(\varphi_q(s))ds \right] \\ &= F(\psi_q^*) + \int_0^1 \ell(\varphi_q(s))ds. \end{aligned}$$

Sending $q \rightarrow \infty$, we now have

$$\begin{aligned} \limsup_{n \rightarrow \infty} -\frac{1}{n} \log E e^{-nF(X^n)} &\leq F(\psi^*) + \int_0^1 \ell(\varphi(s))ds \\ &\leq F(\psi^*) + I(\psi^*) + \delta \\ &\leq \inf_{\psi \in \mathcal{D}([0, 1]; \mathbb{R})} [F(\psi) + I(\psi)] + 2\delta. \end{aligned}$$

Since $\delta > 0$ is arbitrary, the upper bound follows, thus completing the proof of Theorem 3.25. \square

3.4 Notes

Our treatment of Sanov's theorem [228] follows very closely the one in [97], though as noted in the introduction we use the representation based on the chain rule rather than that based on dynamic programming. The proof of Cramér's theorem [68] differs from that of [97] and follows a line of argument that will be used elsewhere, which is that to analyze discrete time "small noise" problems, we first establish an empirical-measure-type large deviation result for the driving noises, and then (in combination with integrability properties of the noise) obtain the large deviation properties of a process driven by these noises through a continuous-mapping-type argument.

The proof of large deviation estimates for small noise diffusions is taken from [32], and it suggests why the more highly structured setting of continuous time Markov processes is, given the appropriate representations, easier than for discrete time. The

solution mapping to the SDE (3.15) is not continuous, since if it were, we would just use the contraction principle and the large deviation theory for scaled Brownian motion (Schilder's theorem [229]). However, it is in some sense almost continuous on the support of the measure induced by $\sqrt{\varepsilon}W$, in that the mapping $u \mapsto \varphi$ when $\varphi(t) = x + \int_0^t (b(\varphi(s)) + \sigma(\varphi(s))u(s)) ds$ is continuous on S_M for all $M < \infty$, a fact that was key in proving the convergence of the variational representation. An analogous continuity applies only to particular models in the setting of discrete time. In particular, the reader will note that the arguments of Chap. 4 are considerably more involved than those for SDEs in continuous time.

The idea of viewing a diffusion as a nearly continuous mapping on Brownian motion (in the small noise limit) originates with Azencott [7]. The first proofs of an LDP for diffusions appear in the papers of Wentzell [245–248], where they appear as just a special case of a more general treatment. Fleming [133] considers certain problems of large deviations involving diffusion process and computes the desired limits using ideas from stochastic control. His approach is closely related to the approach of this book and in many ways inspired it.

In the final example of an SDE driven by a Poisson process we have attempted to emphasize the similarity with the case of Brownian motion, and indeed, the arguments are very close, with the main differences due to the weaker control one obtains from bounded costs and the need to place the controls in a space more complicated than $\mathcal{L}^2([0, T] : \mathbb{R}^k)$ with the weak topology. This example is a simplified form of the problem considered in [45].