Chapter 8
Representations for Continuous Time Processes

In previous chapters we developed and applied representations for the large deviation analysis of discrete time processes. The derivation of useful representations in this setting follows from a straightforward application of the chain rule. The only significant issue is to decide on the ordering used for the underlying "driving noises" when the chain rule is applied, since controls are allowed to depend on the "past," which is determined by this ordering.

In continuous time, the situation is both simpler and more complex. It is simpler in that most models in continuous time can be conveniently represented as systems driven by an exogenous noise process of either Gaussian or Poisson type. As we will see, useful representations hold in great generality for both types of noise. It is also more complex, in that the chain rule cannot be directly applied, and one must approximate and justify suitable limits to establish the representations. In the end, the representations take a form that is analogous to their discrete time counterparts, and we consider controls that are allowed to depend on the past, i.e., controls that are predictable with respect to a suitable filtration.[1]

This chapter consists of three sections, which present the representations for functionals of infinite dimensional Brownian motion, functionals of a Poisson random measure, and the combined case. The proofs given here differ from the first versions that appeared in [39, 45]. In particular, while the details are different, the approach to both models is very much the same.

[1] For special cases, one can consider the infimum of a smaller class (e.g., feedback controls), a result that is sometimes of interest.

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0_8

8.1 Representation for Infinite Dimensional Brownian Motion

The starting point of the proof of the representation is of course (2.1). Hence we will need to understand the form of $d\gamma/d\theta$ when θ is the measure induced by an infinite dimensional Brownian motion. Several formulations of infinite dimensional Brownian motion are commonly used. We focus for now on the formulation as a Hilbert space valued Wiener process,[2] and comment in Chap. 11 on how representations for other formulations follow easily from this one.

8.1.1 The Representation

Let $(\Omega, \mathscr{F}, P)$ be a probability space with a filtration $\{\mathscr{F}_t\}_{0\le t\le T}$ satisfying the usual conditions. We begin with the definition of a Hilbert space valued Wiener process. Let $(\mathscr{H}, \langle\cdot,\cdot\rangle)$ be a real separable Hilbert space. Let Λ be a symmetric strictly positive trace class operator on $\mathscr{H}$ (see Appendix E for definitions and terminology related to Hilbert spaces). This means that Λ is a bounded linear operator such that if $\{e_i\}_{i\in\mathbb{N}}$ is any complete orthonormal sequence (CONS) in $\mathscr{H}$, then for all $i, j \in \mathbb{N}$, we have $\langle e_i, \Lambda e_j\rangle = \langle e_j, \Lambda e_i\rangle$, $\langle e_i, \Lambda e_i\rangle > 0$, and $\sum_{i=1}^{\infty}\langle e_i, \Lambda e_i\rangle < \infty$.

Definition 8.1 An $\mathscr{H}$-valued continuous stochastic process $\{W(t)\}_{0\le t\le T}$ is called a Λ-Wiener process with respect to $\{\mathscr{F}_t\}_{0\le t\le T}$ if for every nonzero $h \in \mathscr{H}$, $\langle \Lambda h, h\rangle^{-1/2}$ $\langle W(t), h\rangle$ is a standard one-dimensional $\mathscr{F}_t$-Wiener process (see Sect. 3.2).

Define $\mathscr{H}_0 \doteq \Lambda^{1/2}\mathscr{H}$. Then $\mathscr{H}_0$ is a Hilbert space with the inner product

$$\langle h, k\rangle_0 \doteq \left\langle \Lambda^{-1/2}h, \Lambda^{-1/2}k\right\rangle$$

for $h, k \in \mathscr{H}_0$. Denote the norms in $\mathscr{H}$ and $\mathscr{H}_0$ by $\|\cdot\|$ and $\|\cdot\|_0$ respectively. Since Λ is trace class, the identity mapping from $\mathscr{H}_0$ to $\mathscr{H}$ is Hilbert–Schmidt. This Hilbert–Schmidt embedding of $\mathscr{H}_0$ in $\mathscr{H}$ will play a central role in many of the arguments to follow. An important consequence of the embedding is that if v^n is a sequence in $\mathscr{H}_0$ such that $v^n \to 0$ weakly in $\mathscr{H}_0$, then $\|v^n\| \to 0$. For an exposition of stochastic calculus with respect to an $\mathscr{H}$ valued Wiener process, we refer to [69]. Other useful references are [197, 198, 252].

We first present and prove a representation that uses controls that are predictable with respect to the filtration generated by the Wiener process, and later, in Sect. 8.1.5, we extend this representation to controls that are predictable with respect to $\{\mathscr{F}_t\}$. Let $\{\mathscr{G}_t\}_{0\le t\le T}$ be the filtration generated by $\{W(t)\}_{0\le t\le T}$ augmented with all P-null sets in $\mathscr{F}$.

[2] We will use the terms "Brownian motion" and "Wiener process" interchangeably.

Definition 8.2 Given $0 \le a < b \le T$ and a bounded $\mathscr{F}_a$-measurable real random variable ξ, let $g : [0, T] \times \Omega \to \mathbb{R}$ be defined by $g(s, \omega) \doteq \xi(\omega) 1_{(a,b]}(s), s \in [0, T]$, $\omega \in \Omega$. Denote by $\mathscr{P}\mathscr{F}$ the σ-field on $[0, T] \times \Omega$ generated by the collection of all such g. This σ-field is called the $\mathscr{F}_t$-predictable σ-field. For a Polish space $\mathscr{E}$, a $\mathscr{P}\mathscr{F}/\mathscr{B}(\mathscr{E})$-measurable map $v : [0, T] \times \Omega \to \mathscr{E}$ is referred to as an $\mathscr{E}$-valued $\mathscr{F}_t$-predictable process.

Define $\bar{\mathscr{A}}$ to be the class of $\mathscr{H}_0$-valued $\mathscr{F}_t$-predictable processes v that satisfy

$$P\left\{\int_0^T \|v(s)\|_0^2 ds < \infty\right\} = 1,$$

and let $\mathscr{A}$ denote the subset of those that are predictable with respect to $\{\mathscr{G}_t\}_{0 \le t \le T}$. We refer to [69, Chap. 4] for the definition of stochastic integrals of elements of $\bar{\mathscr{A}}$ with respect to W. Let $\mathscr{L}^2([0, T] : \mathscr{H}_0)$ denote the Hilbert space of all measurable maps $u : [0, T] \to \mathscr{H}_0$ for which $\int_{[0,T]} \|u(s)\|_0^2 ds$ is finite together with the usual inner product, and for $M \in \mathbb{N}$, define

$$S_M \doteq \left\{u \in \mathscr{L}^2([0, T] : \mathscr{H}_0) : \int_0^T \|u(s)\|_0^2 ds \le M\right\}. \tag{8.1}$$

We endow S_M with the weak topology, which makes it a compact Polish space (cf. [93]). In particular, a sequence $\{v_n\} \subset S_M$ converges to $v \in S_M$ if $\int_0^T \langle v_n(s), h(s)\rangle_0 ds$ converges to $\int_0^T \langle v(s), h(s)\rangle_0 ds$ for all $h \in \mathscr{L}^2([0, T] : \mathscr{H}_0)$. Finally, let

$$\mathscr{A}_{b,M} \doteq \{v \in \mathscr{A} : v(\omega) \in S_M \ \theta\text{-}a.s.\}, \qquad \mathscr{A}_b \doteq \cup_{M \in \mathbb{N}} \mathscr{A}_{b,M}. \tag{8.2}$$

Let $\bar{\mathscr{A}}_{b,M}$ [resp. $\bar{\mathscr{A}}_b$] be defined exactly as $\mathscr{A}_{b,M}$ [resp. $\mathscr{A}_b$], except that $\{\mathscr{G}_t\}$ is replaced by $\{\mathscr{F}_t\}$.

We next state the main result of this section. Let E denote expectation with respect to P. Though in the theorem we take G to be a bounded function, it can be shown that the representation holds if G is bounded from above. The fact that the representation also holds with respect to the smaller class $\mathscr{A}_b \subset \mathscr{A}$ is quite convenient in applications, since these are in some sense very well behaved processes.

Theorem 8.3 *Let W be a Λ-Wiener process and let G be a bounded Borel measurable function mapping $\mathscr{C}([0, T] : \mathscr{H})$ into $\mathbb{R}$. Then*

$$-\log E \exp\{-G(W)\} = \inf_{v \in \mathscr{R}} E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s) ds\right)\right], \tag{8.3}$$

where $\mathscr{R}$ can be either $\mathscr{A}_b$, $\mathscr{A}$, $\bar{\mathscr{A}}_b$ or $\bar{\mathscr{A}}$.

Using (8.3), one can prove the following in an identical manner as Theorem 3.17, and we therefore omit the proof.

Theorem 8.4 *Let W and G be as in Theorem 8.3 and let $\delta > 0$. Then there exists $M < \infty$ depending on $\|G\|_\infty$ and δ such that for all $\varepsilon \in (0,1)$,*

$$
\begin{aligned}
&-\varepsilon \log E \exp\left\{-\frac{1}{\varepsilon} G(\sqrt{\varepsilon} W)\right\} \\
&\geq \inf_{v \in \mathscr{A}_{b,M}} E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(\sqrt{\varepsilon} W + \int_0^{\cdot} v(s) ds\right)\right] - \delta.
\end{aligned}
$$

The rest of Sect. 8.1 is devoted to the proof of Theorem 8.3. After developing the needed preliminary results, the proof for $\mathscr{A}_b$ and $\mathscr{A}$ is given in Sects. 8.1.3 and 8.1.4, and the extension to $\bar{\mathscr{A}}_b$ and $\bar{\mathscr{A}}$ is completed in Sect. 8.1.5.

8.1.2 Preparatory Results

In this section we present several theorems and approximations that will be used in the proof of the representation. For use later on, some results are stated for the more general class of processes $\bar{\mathscr{A}}$. The following result follows from Theorem 10.14 of [69].

Theorem 8.5 (Girsanov) *Let $\psi \in \bar{\mathscr{A}}$ be such that*

$$
E\left[\exp\left\{\int_0^T \langle \psi(s), dW(s)\rangle_0 - \frac{1}{2}\int_0^T \|\psi(s)\|_0^2 ds\right\}\right] = 1.
$$

Then the process

$$
\tilde{W}(t) \doteq W(t) - \int_0^t \psi(s) ds,
$$

$t \in [0,T]$, is a Λ-Wiener process with respect to $\{\mathscr{F}_t\}$ on $(\Omega, \mathscr{F}, Q)$, where Q is the probability measure defined by

$$
\frac{dQ}{dP} = \exp\left\{\int_0^T \langle \psi(s), dW(s)\rangle_0 - \frac{1}{2}\int_0^T \|\psi(s)\|_0^2 ds\right\}.
$$

We record a result that will be used in proving tightness for a sequence of Hilbert space valued processes. Recall the topology on S_N introduced below (8.1).

Lemma 8.6 *Let $\{v^n\}_{n\in\mathbb{N}}$ be a sequence of elements of $\bar{\mathscr{A}}$. Assume that there is $M < \infty$ such that*

$$
\sup_{n\in\mathbb{N}} \int_0^T \|v^n(s)\|_0^2 ds \leq M
$$

a.s. Suppose further that $\{v^n\}$ converges in distribution to v as S_M-valued random variables. Then $\int_0^{\cdot} v^n(s)ds$ converges in distribution to $\int_0^{\cdot} v(s)ds$ in $\mathscr{C}([0,T] : \mathscr{H})$.

Proof It suffices to show that the map from S_M to $\mathscr{C}([0,T]:\mathscr{H})$ defined by $u \mapsto \int_0^{\cdot} u(s)ds$ is continuous. Let $\{\phi^n\}$ be a sequence in S_M that converges to ϕ. Let $\{e_j\}_{j\in\mathbb{N}}$ be a CONS of eigenvectors of Λ with corresponding eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$. Then for every $t \in [0,T]$ and $i \in \mathbb{N}$,

$$\int_0^t \langle \phi^n(s) - \phi(s), e_i\rangle ds = \lambda_i \int_0^t \langle \phi^n(s) - \phi(s), e_i\rangle_0 ds,$$

and by assumption, the right side converges to 0 as $n \to \infty$. Also,

$$\left\| \int_0^t [\phi^n(s) - \phi(s)]ds \right\|^2 = \sum_{i=1}^{\infty} \left(\int_0^t \langle \phi^n(s) - \phi(s), e_i\rangle ds \right)^2.$$

By Hölder's inequality and $\Lambda^{1/2} e_i = \lambda_i e_i$,

$$\begin{aligned} \left(\int_0^t \langle \phi^n(s) - \phi(s), e_i\rangle ds \right)^2 &\le \lambda_i T \int_0^T \|\Lambda^{-1/2}[\phi^n(s) - \phi(s)]\|^2 ds \\ &= \lambda_i T \int_0^T \|\phi^n(s) - \phi(s)\|_0^2 ds \\ &\le 4MT\lambda_i. \end{aligned}$$

Since $\sum_{i=1}^{\infty} \lambda_i < \infty$, it follows from the dominated convergence theorem that for each $t \in [0,T]$, $\int_0^t \phi^n(s)ds$ converges to $\int_0^t \phi(s)ds$ in $\mathscr{H}$. To prove that this convergence is uniform on $[0,T]$, we need an equicontinuity estimate. This follows by noting that for $0 \le s \le t \le T$,

$$\begin{aligned} \left\| \int_0^t \phi^n(r)dr - \int_0^s \phi^n(r)dr \right\| &\le \sqrt{t-s} \left(\int_0^T \|\phi^n(s)\|^2 ds \right)^{1/2} \\ &\le \sqrt{t-s}\, \|\Lambda\|^{1/2} \left(\int_0^T \|\Lambda^{-1/2}\phi^n(s)\|^2 ds \right)^{1/2} \\ &= \sqrt{t-s}\, \|\Lambda\|^{1/2} \left(\int_0^T \|\phi^n(s)\|_0^2 ds \right)^{1/2} \\ &\le \sqrt{t-s}\, \|\Lambda\|^{1/2} M^{1/2}, \end{aligned}$$

where $\|\Lambda\| \doteq \sup_{h\in\mathscr{H}:\|h\|=1} \|\Lambda h\|$ is the operator norm. □

Before turning to the proof of Theorem 8.3, we state one last result. A process $v \in \mathscr{A}$ is called **simple** if there exist $k \in \mathbb{N}$, $0 = t_1 \le \cdots \le t_{k+1} = T$ and $N \in \mathbb{N}$ such that

$$v(s,\omega) \doteq \sum_{j=1}^{k} X_j(\omega) 1_{(t_j,t_{j+1}]}(s),$$

where the X_j are $\mathcal{H}_0$-valued $\mathcal{G}_{t_j}$-measurable random variables satisfying $\|X_j(\omega)\|_0 \le N$ for all $j \in \{1, \ldots, k\}$. Let $\mathcal{A}_s$ denote the collection of simple processes, and note that $\mathcal{A}_s \subset \mathcal{A}_b$. Given any $v \in \mathcal{A}_s$, it is straightforward that

$$E\left[\exp\left\{\int_0^T \langle v(s), dW(s)\rangle_0 - \frac{1}{2}\int_0^T \|v(s)\|_0^2 ds\right\}\right] = 1,$$

and thus by Theorem 8.5, the process

$$W^v(t) \doteq W(t) - \int_0^t v(s)ds,$$

$t \in [0, T]$, is a Λ-Wiener process with respect to $\{\mathcal{G}_t\}$ on $(\Omega, \mathcal{F}, Q^v)$, where Q^v is the probability measure defined by

$$\frac{dQ^v}{dP} = \exp\left\{\int_0^T \langle v(s), dW(s)\rangle_0 - \frac{1}{2}\int_0^T \|v(s)\|_0^2 ds\right\}.$$

Let E^v denote integration with respect to Q^v.

Lemma 8.7 *For every $v \in \mathcal{A}_s$, there is $\tilde{v} \in \mathcal{A}_s$ such that $(W^{\tilde{v}}, \tilde{v})$ has the same distribution under $Q^{\tilde{v}}$ as (W, v) does under P.*

Proof Let v be simple and of the form

$$v(s, \omega) \doteq \sum_{j=1}^{k} X_j(\omega) 1_{(t_j, t_{j+1}]}(s),$$

where $k \in \mathbb{N}$, $0 = t_1 \le \cdots \le t_{k+1} = T$ and X_j are $\mathcal{H}_0$-valued $\mathcal{G}_{t_j}$-measurable random variables satisfying $\|X_j(\omega)\|_0 \le N$ for all $j \in \{0, \ldots, k\}$ and some $N \in \mathbb{N}$. New random variables $\bar{X}_j$, $j \in \{0, \ldots, k\}$, are defined as follows. Since $X_1(\omega)$ is $\mathcal{G}_0$-measurable, there exists measurable $G_1 : \mathcal{H}_0 \to \mathcal{H}_0$ such that $X_1(\omega) = G_1(W(0, \omega))$ a.s. Let $\bar{X}_1 \doteq G_1(W(0)) = X_1$. For $j \in \{2, \ldots, k\}$, there are measurable $G_j : \mathcal{C}([0, t_j] : \mathcal{H}_0) \to \mathcal{H}_0$ such that $X_j(\omega) = G_j(W(t, \omega), 0 \le t \le t_j)$ a.s. We can also consider G_j as a mapping $\mathcal{C}([0, T] : \mathcal{H}_0) \to \mathcal{H}_0$, which depends on $w \in \mathcal{C}([0, T] : \mathcal{H}_0)$ only though the restriction to $[0, t_j]$, and we do so with the notation $G_j(w)$. We then recursively define

$$\bar{X}_j \doteq G_j\left(W(\cdot) - \int_0^{\cdot} \sum_{i=1}^{j-1} \bar{X}_i 1_{(t_i, t_{i+1}]}(s)ds\right).$$

By construction, each $\bar{X}_j$ is $\mathcal{G}_{t_j}$-measurable and satisfies $\|\bar{X}_j(\omega)\|_0 \le N$ for a.e. ω. Now let

$$\tilde{v}(s,\omega) \doteq \sum_{j=1}^{k} \bar{X}_j(\omega) 1_{(t_j,t_{j+1}]}(s),$$

and note that

$$\tilde{v}(s) \doteq \sum_{j=1}^{k} G_j \left(W(\cdot) - \int_0^{\cdot} \tilde{v}(s)ds \right) 1_{(t_j,t_{j+1}]}(s). \tag{8.4}$$

By Theorem 8.5, $W(t) - \int_0^t \tilde{v}(s)ds$ is a Λ-Wiener process under $Q^{\tilde{v}}$. Since $\tilde{v}$ has the form given in (8.4), it follows that $(W^{\tilde{v}}, \tilde{v})$ has the same distribution under $Q^{\tilde{v}}$ as (W, v) does under P. $\square$

8.1.3 Proof of the Upper Bound in the Representation

In this subsection we prove

$$-\log E \exp\{-G(W)\} \le \inf_{v \in \mathscr{A}} E \left[\frac{1}{2} \int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s)ds \right) \right].$$

Note that this automatically gives the corresponding bound for the smaller class $\mathscr{A}_b$ in (8.3). The proof is in two steps.

Step 1. Simple v. According to (2.1), for every probability measure Q on $(\Omega, \mathscr{F})$,

$$-\log E \exp\{-G(W)\} \le R(Q \,\|\, P) + \int_{\Omega} G(W) dQ. \tag{8.5}$$

If $v \in \mathscr{A}_s$, then by Lemma 8.7 there is $\tilde{v} \in \mathscr{A}_s$ such that the distribution of (W, v) under P is the same as that of $(W^{\tilde{v}}, \tilde{v})$ under $Q^{\tilde{v}}$. Since $\tilde{v}$ is bounded, it follows from Theorem 8.5 that

$$\begin{aligned} R(Q^{\tilde{v}} \,\|\, P) &= E^{\tilde{v}} \left[\int_0^T \langle \tilde{v}(s), dW(s) \rangle_0 - \frac{1}{2} \int_0^T \|\tilde{v}(s)\|_0^2 ds \right] \\ &= E^{\tilde{v}} \left[\int_0^T \left\langle \tilde{v}(s), dW^{\tilde{v}}(s) \right\rangle_0 + \frac{1}{2} \int_0^T \|\tilde{v}(s)\|_0^2 ds \right] \\ &= E^{\tilde{v}} \left[\frac{1}{2} \int_0^T \|\tilde{v}(s)\|_0^2 ds \right] \\ &= E \left[\frac{1}{2} \int_0^T \|v(s)\|_0^2 ds \right]. \end{aligned}$$

Taking $Q = Q^{\tilde{v}}$ in (8.5) together with $E^{\tilde{v}} G(W) = E^{\tilde{v}} G(W^{\tilde{v}} + \int_0^{\cdot} \tilde{v}(s)ds) = EG(W + \int_0^{\cdot} v(s)ds)$ gives

$$-\log E\exp\{-G(W)\} \le E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s)ds\right)\right]. \quad (8.6)$$

Step 2. General v. Next consider $v \in \mathscr{A}$. We can assume without loss of generality that $E[\int_0^T \|v(s)\|_0^2 ds] < \infty$. Then (see, for example, [159, Lemma II.1.1]) there is a sequence $\{v_n\} \subset \mathscr{A}_s$ such that

$$E\int_0^T \|v_n(s) - v(s)\|_0^2 ds \to 0. \quad (8.7)$$

In particular, $N \doteq \sup_{n\in\mathbb{N}} E\int_0^T \|v_n(s)\|_0^2 ds < \infty$. From Step 1, for all n,

$$-\log E\exp\{-G(W)\} \le E\left[\frac{1}{2}\int_0^T \|v_n(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v_n(s)ds\right)\right]. \quad (8.8)$$

We would like to apply Lemma 2.5, where μ_n and θ are the distributions induced by $W + \int_0^{\cdot} v_n(s)ds$ and W under P, respectively. Since μ_n is also the distribution induced by W under $Q^{\tilde{v}_n}$, part (f) of Lemma 2.4 implies

$$R(\mu_n \| \theta) \le R(Q^{\tilde{v}_n} \| P) = E^{\tilde{v}_n}\left[\frac{1}{2}\int_0^T \|\tilde{v}_n\|_0^2 ds\right] = E\left[\frac{1}{2}\int_0^T \|v_n\|_0^2 ds\right].$$

Thus $\sup_n R(\mu_n \| \theta) \le N/2$. From (8.7), it follows that

$$E\sup_{0\le t\le T}\left\|\int_0^t v_n(s)ds - \int_0^t v(s)ds\right\|^2 \le T\|\Lambda\|_{\text{op}}^2 E\int_0^T \|v_n(s) - v(s)\|_0^2 ds \to 0,$$

and therefore μ_n converges weakly to μ, where μ is the distribution of $W + \int_0^{\cdot} v(s)ds$. Since G is bounded and measurable, we now obtain (8.6) using Lemma 2.5 and sending $n \to \infty$ in (8.8). ☐

8.1.4 Proof of the Lower Bound in the Representation

In this subsection we prove

$$-\log E\exp\{-G(W)\} \ge \inf_{v\in\mathscr{A}_b} E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s)ds\right)\right]. \quad (8.9)$$

This automatically gives the corresponding bound for the larger class $\mathscr{A}$ in (8.3). The proof is in two steps.

Step 1. G of a particular form. We first consider G of a special form. Recall that $\{e_n\}_{n\in\mathbb{N}}$ denotes a CONS in $\mathscr{H}$. Let $K, N \in \mathbb{N}$ be arbitrary, and consider any collection $0 = t_1 < t_2 < \cdots < T_K = T$. Let $h : \mathbb{R}^{KN} \to \mathbb{R}$ have compact support and continuous derivatives of all orders. Then G is of the form

$$G(W) = h(w(t_1), w(t_2) - w(t_1), \ldots, w(t_K) - w(t_{K-1})), \tag{8.10}$$

where for $0 \le t \le T$,

$$w(t) = (\lambda_1^{-1/2} \langle e_1, W(t) \rangle, \ldots, \lambda_N^{-1/2} \langle e_N, W(t) \rangle). \tag{8.11}$$

Note that $\{w(t)\}_{0 \le t \le T}$ is an N-dimensional standard $\mathscr{G}_t$-Wiener process. Using methods from stochastic control theory, we will construct a process $v \in \mathscr{A}_b$ that gives equality in (8.9). The following lemma, whose proof is omitted, follows by classical and elementary stochastic control arguments that apply when there is a smooth value function (see Sect. VI.2 of [134]).

Lemma 8.8 *Let $g : \mathbb{R}^m \times \mathbb{R}^N \to \mathbb{R}$ have compact support and continuous derivatives of all orders. Let $\{w(t)\}_{0 \le t \le T}$ be an N-dimensional standard Brownian motion, and let $V : [0, T] \times \mathbb{R}^m \times \mathbb{R}^N \to \mathbb{R}$ be defined by*

$$V(t, z, x) \doteq - \log E e^{-g(z, x + w(T-t))}.$$

Then the following hold.

(a) *For all $(t, x) \in [0, T] \times \mathbb{R}^N$, $z \mapsto V(t, z, x)$ has compact support and derivatives of all orders that are continuous functions of (t, z, x).*

(b) *For all $(t, z) \in [0, T] \times \mathbb{R}^m$, $x \mapsto V(t, z, x)$ has derivatives of all orders that are continuous and bounded functions of (t, z, x).*

(c) *For $z \in \mathbb{R}^m$, let $\{X(z, t)\}_{0 \le t \le T}$ be the unique solution of*

$$X(z, t) = - \int_0^t D_x V(s, z, X(z, s)) ds + w(t), \ \ t \in [0, T].$$

Then with $u(t) = -D_x V(t, z, X(z, t))$ for $t \in [0, T]$,

$$- \log E \exp\{-g(z, w(T))\} = E \left[\frac{1}{2} \int_0^T \|u\|^2 ds + g\left(z, w + \int_0^{\cdot} u ds\right) \right]. \tag{8.12}$$

Remark 8.9 For the proof of Lemma 8.8, one starts with the linear partial differential equation (PDE) for which $(t, z, x) \mapsto E e^{-g(z, x + w(T-t))}$ is a classical-sense solution. From this, one obtains the nonlinear PDE (Hamilton–Jacobi–Bellman equation) for which V is a classical-sense solution. As such, V also has an interpretation as the minimal cost in a stochastic optimal control problem. Using a classical verification argument [134], it is straightforward to show that u as defined in the lemma is the optimal control, and the right-hand side of (8.12) is the minimal cost starting from $x = 0$, which establishes (8.12).

Now for $j = 1, \ldots, K$, define $V_j : \mathbb{R}^{jN} \to \mathbb{R}$ as follows: $V_K = h$ and

$$V_j(z_j) = -\log E e^{-V_{j+1}(z_j, w(t_{j+1})-w(t_j))}, \; z_j \in \mathbb{R}^{jN}, \; j = 1, \ldots, K-1.$$

By successive conditioning, it is easily checked that

$$V_0 \doteq -\log E e^{-V_1(w(t_1)-w(t_0))} = -\log E e^{-G(W)},$$

where G is as in (8.10). From part (a) of Lemma 8.8 it follows that for all $j = 1, \ldots, K$, V_j has continuous and bounded derivatives of all orders. For $j = 1, \ldots, K$, let $Z_j = (w(t_1), w(t_2) - w(t_1), \ldots, w(t_j) - w(t_{j-1}))$, and note that Z_j is an $\mathbb{R}^{jN}$-valued random variable. For $z_j \in \mathbb{R}^{jN}$, let $\{Y(z_j, t)\}_{t \in [t_j, t_{j+1}]}$, $j = 1, \ldots, K-1$, be the unique solution of

$$Y(z_j, t) = -\int_{t_j}^{t} D_x V_{j+1}(s, z_j, Y(z_j, s))ds + w(t) - w(t_j), \; t \in [t_j, t_{j+1}].$$

The existence and uniqueness of the solution is a consequence of the smoothness property of V_{j+1} noted earlier. Now define

$$u(t) = -D_x V_{j+1}(t, Z_j, Y(Z_j, t)), \; t \in [t_j, t_{j+1}), \; j = 0, \ldots, K-1.$$

Then by a straightforward recursive argument using Lemma 8.8, we see that

$$-\log E e^{-G(W)} = E\left[\frac{1}{2}\int_0^T \|u(s)\|^2 ds + h\left(w(t_1) + \int_0^{t_1} u(s)ds, \ldots \right.\right.$$
$$\left.\left., w(t_K) - w(t_{K-1}) + \int_{t_{K-1}}^{t_K} u(s)ds\right)\right].$$

Let $v(s) \doteq \sum_{i=1}^{N} \lambda_i^{1/2} u_i(s) e_i$, $s \in [0, T]$. Then $v \in \mathscr{A}_b$, and by (8.10) and (8.11),

$$-\log E \exp\{-G(W)\} = E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s)ds\right)\right].$$

Thus we have proved (8.9) for all G of the form (8.10)–(8.11).

Step 2. G that is bounded and measurable. Now suppose that G is simply bounded and measurable. We claim that there exist functions $\{G_n\}_{n \in \mathbb{N}}$ such that for each n, G_n is of the form assumed in Step 1, $\|G_n\|_\infty \leq \|G\|_\infty$, and $G_n \to G$ a.s. with respect to θ. This can be seen most easily by considering the approximation in stages. We note that each of the following classes admits an approximation of this form relative to elements of the preceding class, save of course the first:

- G bounded and measurable;
- G bounded and continuous;

- $G(W) = H(W(t_1), W(t_2), \ldots, W(t_K))$, where $H : \mathscr{H}^K \to \mathbb{R}$ is continuous and bounded and $K \in \mathbb{N}$ and $0 = t_1 < t_2 < \cdots < T_K = T$ are arbitrary;
- G of the form (8.10)–(8.11) where h is a bounded and continuous function from $\mathbb{R}^{NK} \to \mathbb{R}$, $K, N \in \mathbb{N}$ and $0 = t_1 < t_2 < \cdots < t_K = T$ are arbitrary;
- G as above and in addition, h has compact support;
- G as above and in addition, h has continuous and bounded derivatives of all orders.

All of these approximations follow by standard arguments. The first approximation statement (i.e., a bounded measurable G can be approximated by a bounded continuous G) is the conclusion of a result due to Doob and presented in the appendix as Theorem E.4. For the second statement we use the martingale convergence theorem, which states that if $\{\mathscr{F}_n\}$ is a filtration increasing to the σ-field $\mathscr{F}_\infty$ and X is an integrable $\mathscr{F}_\infty$-measurable random variable, then $E[X \mid \mathscr{F}_n]$ converges to X a.s. Consider a sequence of partitions $\pi_n = \{0 = t_1^n < t_2^n < \cdots < t_{K_n}^n = T\}$ such that $\pi_n \subset \pi_{n+1}$ and $|\pi_n| \doteq \max_{1 \le j \le K_n - 1}(t_{j+1}^n - t_j^n) \to 0$ as $n \to \infty$. Let $\mathscr{F}_n \doteq \sigma\{W(t_i^n), 1 \le i \le K_n\}$. Clearly, $\mathscr{F}_n$ is a filtration and G is $\mathscr{F}_\infty \doteq \sigma(\cup_{n \ge 1} \mathscr{F}_n)$-measurable. Thus by the martingale convergence theorem, $G_n = E[G \mid \mathscr{F}_n]$ converges a.s. to G. Clearly, $\|G_n\|_\infty \le \|G\|_\infty$ a.s. The second approximation statement now follows from another application of Theorem E.4 if G_n is not continuous. The proof of the third approximation statement is similar but uses the filtration

$$\mathscr{F}_n \doteq \sigma\{\langle W(t_i), e_j \rangle, j = 1, \ldots, n, \ i = 1, \ldots, K\},$$

where $\{e_j\}_{j \in \mathbb{N}}$ is a CONS in $\mathscr{H}$. The fourth approximation statement involves replacing h in (8.10)–(8.11) by $h\psi_n$ in defining G_n, where ψ_n is a continuous function with values in $[0, 1]$ such that $\psi_n(x) = 1$ when x is in a ball of radius n and $\psi_n(x) = 0$ outside a ball of radius $n + 1$. Finally, the last statement follows by replacing h with $h * \eta_n$ in (8.10)–(8.11), where $\eta_n(x) = n^{-NK}\eta(nx)$, $x \in \mathbb{R}^{NK}$,

$$\eta(x) \doteq c \exp\left\{-\frac{1}{1 - |x|^2}\right\} 1_{\{|x|<1\}},$$

and c is the normalizing constant such that $\int \eta(x)dx = 1$.

With the claim verified, we now complete the lower bound. With each $n \in \mathbb{N}$ we can associate $v_n \in \mathscr{A}_b$ such that

$$-\log E \exp\{-G_n(W)\} = E\left[\frac{1}{2}\int_0^T \|v_n(s)\|_0^2 ds + G_n\left(W + \int_0^{\cdot} v_n(s)ds\right)\right].$$

As in the proof of the upper bound, if μ_n is the distribution induced by $W + \int_0^{\cdot} v_n(s)ds$, then

$$R(\mu_n \| \theta) \le E\left[\frac{1}{2}\int_0^T \|v_n(s)\|_0^2 ds\right] \le 2\|G\|_\infty,$$

where the last inequality is valid because $\|G_n\|_\infty \le \|G\|_\infty$. Thus $\{\mu_n\}$ is tight, and from part (b) of Lemma 2.5, we have

$$\lim_{n\to\infty} E\left|G_n\left(W + \int_0^{\cdot} v_n(s)ds\right) - G\left(W + \int_0^{\cdot} v_n(s)ds\right)\right| = 0.$$

By the dominated convergence theorem,

$$\lim_{n\to\infty} |\log E\exp\{-G_n(W)\} - \log E\exp\{-G(W)\}| = 0.$$

Therefore, given $\varepsilon > 0$, we can find $n \in \mathbb{N}$ such that

$$\begin{aligned}
-\log Ee^{-G(W)} &\ge -\log Ee^{-G_n(W)} - \varepsilon \\
&= E\left[\frac{1}{2}\int_0^T \|v_n(s)\|_0^2 ds + G_n\left(W + \int_0^{\cdot} v_n(s)ds\right)\right] - \varepsilon \\
&\ge E\left[\frac{1}{2}\int_0^T \|v_n(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v_n(s)ds\right)\right] - 2\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary and $v_n \in \mathscr{A}_b$, we have (8.9), completing the proof. □

8.1.5 Representation with Respect to a General Filtration

We now return to the issue of whether the representation holds when $\{\mathscr{G}_t\}_{0\le t\le T}$, the filtration generated by the Wiener process, is replaced by any filtration $\{\mathscr{F}_t\}_{0\le t\le T}$ that satisfies the usual conditions and such that W is a Λ-Wiener process with respect to this larger filtration.

We will make use of the following lemma on measurable selections.

Lemma 8.10 *Let $\mathscr{E}_1, \mathscr{E}_2$ be Polish spaces and let $g : \mathscr{E}_1 \times \mathscr{E}_2 \to \mathbb{R}$ be a bounded continuous function. Let K be a compact set in $\mathscr{E}_2$. For each $x \in \mathscr{E}_1$, define*

$$\Gamma_x \doteq \left\{y \in K : \inf_{y_0\in K} g(x, y_0) = g(x, y)\right\}.$$

Then there exists a Borel measurable function $g_1 : \mathscr{E}_1 \to \mathscr{E}_2$ such that $g_1(x) \in \Gamma_x$ for all $x \in \mathscr{E}_1$.

Proof Let x_n be a sequence in $\mathscr{E}_1$ converging to $\bar{x}$. For each $n \in \mathbb{N}$, let $y_n \in \Gamma_{x_n}$. In view of Corollary E.3, it suffices to show that $\{y_n\}$ has a limit point in $\Gamma_{\bar{x}}$. Let $\bar{y}$ be a limit point of $\{y_n\}$. For each n, $g(x_n, y_n) - \inf_{y_0\in K} g(x_n, y_0)$ equals zero. Since the map $(x, y) \mapsto g(x, y) - \inf_{y_0\in K} g(x, y_0)$ is continuous, letting $n \to \infty$ shows that $\bar{y} \in \Gamma_x$. □

Recall that $\bar{\mathscr{A}}$ was defined exactly as $\mathscr{A}$, except with $\{\mathscr{G}_t\}$ replaced by $\{\mathscr{F}_t\}$. The only issue to check is whether the upper bound

$$-\log E\exp\{-G(W)\} \leq \inf_{v\in\bar{\mathscr{A}}} E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s)ds\right)\right] \tag{8.13}$$

continues to hold.

The only place where the structure of $\{\mathscr{G}_t\}_{0\leq t\leq T}$ is used in Sect. 8.1.3 is in the proof of (8.6), where we appeal to Lemma 8.7 to argue that if $v \in \mathscr{A}_s$, then there is $\tilde{v} \in \mathscr{A}_s$ such that the distribution of (W, v) under P is the same as that of $(W^{\tilde{v}}, \tilde{v})$ under $Q^{\tilde{v}}$. We can reduce to that case if we show that given $\varepsilon > 0$ and any control v that is simple with respect to $\{\mathscr{F}_t\}_{0\leq t\leq T}$, there is a $\bar{v} \in \mathscr{A}_s$ such that

$$\begin{aligned} &E\left[\frac{1}{2}\int_0^T \|\bar{v}(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} \bar{v}(s)ds\right)\right] \\ &\quad\leq E\left[\frac{1}{2}\int_0^T \|v(s)\|_0^2 ds + G\left(W + \int_0^{\cdot} v(s)ds\right)\right] + \varepsilon. \end{aligned} \tag{8.14}$$

For simplicity, we consider the case $v(s) = 0$ for $s \in [0, t]$ and $v(s) = X$ for $s \in (t, T]$, where X is $\mathscr{F}_t$-measurable, and also assume that $\|X\|_0 \leq M < \infty$ a.s. The generalization to the finite collection of random variables that appear in a simple control is straightforward (see [39]). For the moment, we also assume that G is continuous as well as bounded. Consider the mapping

$$g(\phi, x) = E\left[\frac{(T-t)}{2}\|x\|_0^2 + G\left(\phi^B + \int_0^{\cdot} 1_{[t,T]}(s)x\,ds\right)\right],$$

where $\phi \in \mathscr{C}([0,t] : \mathscr{H}_0)$, $x \in \{x \in \mathscr{H}_0 : \|x\|_0 \leq M\}$, and

$$\phi^B(s) = \begin{cases} \phi(s), & s \in [0,t], \\ \phi(t) + B(s-t) - B(0), & s \in [t,T], \end{cases}$$

with B a Λ-Wiener process.

Note that g is bounded, and that by the dominated convergence theorem, it is also continuous in (ϕ, x). Consider the $\mathscr{C}([0,t] : \mathscr{H}_0)$-valued random variable $Z \doteq \{W(s)\}_{0\leq s\leq t}$. Then

$$E\left[\int_0^T \frac{1}{2}\|v(s)\|_0^2 + G\left(W + \int_0^{\cdot} v(s)ds\right)\right] = E[g(Z, X)].$$

Since a single probability measure on a Polish space is tight, there is a compact subset K_0 of $\mathscr{H}_0$ such that $P\{X \in K_0^c\} \leq \varepsilon/(2\|g\|_\infty + 1)$. Then

$$E[g(Z, X)] \geq E[g(Z, X)1_{K_0}(X)] - \frac{\varepsilon}{2}.$$

Now we apply Lemma 8.10 with $\mathscr{E}_1 = \mathscr{C}([0,t] : \mathscr{H}_0)$, $\mathscr{E}_2 = \mathscr{H}_0$, and $K = K_0 \cap \{x \in \mathscr{H}_0 : \|x\|_0 \le M\}$. Then there is a measurable map $g_1 : \mathscr{C}([0,t] : \mathscr{H}_0) \to K$ such that with $\bar{X} \doteq g_1(Z)$,

$$
\begin{aligned}
E[g(Z,X)] &\ge E[g(Z,X)1_{K_0}(X)] - \frac{\varepsilon}{2} \\
&\ge E[g(Z,g_1(Z))1_{K_0}(X)] - \frac{\varepsilon}{2} \\
&\ge E[g(Z,g_1(Z))] - \varepsilon \\
&= E\left[\frac{[T-t]}{2}\left\|\bar{X}\right\|_0^2 + G\left(W + \int_0^{\cdot} 1_{[t,T]}(s)\bar{X}ds\right)\right] - \varepsilon.
\end{aligned}
$$

Letting $\bar{v}(s) = 0$ for $s \in [0,t)$ and $\bar{v}(s) = \bar{X}$ for $s \in [t,T]$, we now have that

$$
\begin{aligned}
&E\left[\int_0^T \frac{1}{2}\|v(s)\|_0^2 + G\left(W + \int_0^{\cdot} v(s)ds\right)\right] \\
&\quad\ge E\left[\int_0^T \frac{1}{2}\|\bar{v}(s)\|_0^2 + G\left(W + \int_0^{\cdot} \bar{v}(s)ds\right)\right] - \varepsilon.
\end{aligned}
\tag{8.15}
$$

This completes the argument for the case that G is continuous.

Finally, we remove the assumption that G is continuous. Let v take the same form as previously, and suppose that G is bounded and measurable. It then follows from Theorem E.4 that there are bounded and continuous G_j that converge to G as $j \to \infty$ almost surely with respect to the distribution of W and that have the same uniform bound as G. Thus by the dominated convergence theorem, given $\varepsilon > 0$, we have for all sufficiently large $j \in \mathbb{N}$ that

$$
E\left[G_j\left(W + \int_0^{\cdot} v(s)ds\right)\right] \le E\left[G\left(W + \int_0^{\cdot} v(s)ds\right)\right] + \frac{\varepsilon}{2}.
$$

We have shown that there is $\bar{v}_j \in \mathscr{A}_s$ such that (8.15) holds with G replaced by G_j. Since $\sup_j R(\mu_j \| \theta) \le TM^2/2$, where μ_j is the probability distribution of $W + \int_0^{\cdot} \bar{v}_j(s)ds$, an application of Lemma 2.5 shows that for sufficiently large $j \in \mathbb{N}$,

$$
E\left[G\left(W + \int_0^{\cdot} \bar{v}_j(s)ds\right)\right] \le E\left[G_j\left(W + \int_0^{\cdot} \bar{v}_j(s)ds\right)\right] + \frac{\varepsilon}{2}.
$$

Thus for j that satisfy the last two displays, we have

$$
\begin{aligned}
&E\left[\int_0^T \frac{1}{2}\left\|\bar{v}_j(s)\right\|_0^2 + G\left(W + \int_0^{\cdot} \bar{v}_j(s)ds\right)\right] \\
&\quad\le E\left[\int_0^T \frac{1}{2}\left\|\bar{v}_j(s)\right\|_0^2 + G_j\left(W + \int_0^{\cdot} \bar{v}_j(s)ds\right)\right] + \frac{\varepsilon}{2}
\end{aligned}
$$

$$\leq E\left[\int_0^T \frac{1}{2}\,\|v(s)\|_0^2 + G_j\left(W + \int_0^\cdot v(s)ds\right)\right] + \frac{\varepsilon}{2}$$
$$\leq E\left[\int_0^T \frac{1}{2}\,\|v(s)\|_0^2 + G\left(W + \int_0^\cdot v(s)ds\right)\right] + \varepsilon.$$

Since $\bar{v}_j \in \mathscr{A}_s$, we have (8.14), and the desired upper bound (8.13) follows. □

8.2 Representation for Poisson Random Measure

In this section we present the analogous representations for functionals of a Poisson random measure (PRM), the other important driving noise in continuous time. In contrast to the case of a Wiener process, for PRM it is convenient and in some sense necessary to enlarge the underlying probability space. The enlargement is needed to define a very general class of controlled Poisson random measures. An alternative approach that has been considered is to dilate time, thereby increasing or decreasing rates, but this method of producing a controlled PRM does not allow for a representation general enough. For further discussion, see [45, p. 726].

8.2.1 The Representation

For a locally compact Polish space $\mathscr{S}$, we denote by $\Sigma(\mathscr{S})$ the space of all measures ν on $(\mathscr{S}, \mathscr{B}(\mathscr{S}))$ satisfying $\nu(K) < \infty$ for every compact $K \subset \mathscr{S}$. We endow $\Sigma(\mathscr{S})$ with the vague topology, namely the weakest topology such that for every $f \in \mathscr{C}_c(\mathscr{S})$ (the space of real continuous functions on $\mathscr{S}$ with compact support), the function $\nu \mapsto \langle f, \nu\rangle = \int_{\mathscr{S}} f(u)\,\nu(du)$, $\nu \in \Sigma(\mathscr{S})$ is continuous. This topology can be metrized such that $\Sigma(\mathscr{S})$ is a Polish space. For details, see Sect. A.4.1.

Definition 8.11 Fix $T \in (0, \infty)$, let $\mathscr{X}$ be a locally compact Polish space, and let $\mathscr{X}_T = [0, T] \times \mathscr{X}$. Let $(\Omega, \mathscr{F}, P)$ be a probability space with a filtration $\{\mathscr{F}_t\}_{0\leq t\leq T}$. Consider any measure $\nu \in \Sigma(\mathscr{X})$ and let $\nu_T = \lambda_T \times \nu$, where λ_T is Lebesgue measure on $[0, T]$. Then an $\mathscr{F}_t$-Poisson random measure with intensity measure ν_T is a measurable mapping N from Ω into $\Sigma(\mathscr{X}_T)$ such that the following properties hold.

- For every $t \in [0, T]$ and every Borel subset $A \subset [0, t] \times \mathscr{X}$, $N(A)$ is $\mathscr{F}_t$-measurable.
- For every $t \in [0, T]$ and every Borel subset $A \subset (t, T] \times \mathscr{X}$, $N(A)$ is independent of $\mathscr{F}_t$.
- If $k \in \mathbb{N}$ and $A_i \in \mathscr{B}(\mathscr{X}_T)$, $i = 1, \ldots, k$, are such that $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\nu_T(A_i) < \infty$, then $N(A_1), \ldots, N(A_k)$ are mutually independent Poisson random variables with parameters $\nu_T(A_1), \ldots, \nu_T(A_k)$.

As with the case of functionals of a Wiener process, it is convenient to first discuss representations with respect to the canonical filtration. Thus we let $\mathbb{M} \doteq \Sigma(\mathscr{X}_T)$ and let P denote the unique probability measure on $(\mathbb{M}, \mathscr{B}(\mathbb{M}))$ under which the canonical map, $N : \mathbb{M} \to \mathbb{M}$, $N(m) \doteq m$, is a Poisson random measure with intensity measure ν_T. With applications to large deviations in mind, we also consider, for $\theta > 0$, the analogous probability measures P_θ on $(\mathbb{M}, \mathscr{B}(\mathbb{M}))$ under which N is a Poisson random measure with intensity $\theta\nu_T$. (In contrast to the case of a Wiener process, there is no simple transformation of a PRM with intensity ν_T that produces a PRM with intensity $\theta\nu_T$, $\theta \neq 1$.) The corresponding expectation operators will be denoted by E and E_θ, respectively. At the end of this section we state the representation for a general filtration.

We will obtain representations for $-\log E_\theta \exp\{-G(N)\}$, where $G \in \mathscr{M}_b(\mathbb{M})$, in terms of a "controlled" Poisson random measure constructed on a larger space. We now describe this construction. Let $\mathscr{Y} \doteq \mathscr{X} \times [0, \infty)$ and $\mathscr{Y}_T \doteq [0, T] \times \mathscr{Y}$. Let $\bar{\mathbb{M}} \doteq \Sigma(\mathscr{Y}_T)$ and let $\bar{P}$ be the unique probability measure on $(\bar{\mathbb{M}}, \mathscr{B}(\bar{\mathbb{M}}))$ such that the canonical map, $\bar{N} : \bar{\mathbb{M}} \to \bar{\mathbb{M}}$, $\bar{N}(m) \doteq m$, is a Poisson random measure with intensity measure $\bar{\nu}_T \doteq \lambda_T \times \nu \times \lambda_\infty$, where λ_∞ is Lebesgue measure on $[0, \infty)$. The corresponding expectation operator will be denoted by $\bar{E}$. The control will act through this additional component of the underlying point space.

Let $\mathscr{G}_t$ denote the augmentation of $\sigma\left\{\bar{N}((0, s] \times A) : 0 \le s \le t, A \in \mathscr{B}(\mathscr{Y})\right\}$ with all $\bar{P}$ null sets in $\mathscr{B}(\bar{\mathbb{M}})$, and denote by $\mathscr{PF}$ the predictable σ-field on $[0, T] \times \bar{\mathbb{M}}$ with the filtration $\{\mathscr{G}_t\}_{0 \le t \le T}$ on $(\bar{\mathbb{M}}, \mathscr{B}(\bar{\mathbb{M}}))$. Let $\mathscr{A}$ be the class of all maps $\varphi : \mathscr{X}_T \times \bar{\mathbb{M}} \to [0, \infty)$ that are $(\mathscr{PF} \otimes \mathscr{B}(\mathscr{X}))\backslash\mathscr{B}[0, \infty)$ measurable. [Note that there is a slight inconsistency in the notation, since $\mathscr{PF}$ concerns t, ω, while $\mathscr{B}(\mathscr{X})$ concerns x, but we write them in the order (t, x, ω).] Since $\bar{\mathbb{M}}$ is the underlying probability space, following standard convention, we will at times suppress the dependence of $\varphi(t, x, \omega)$ on ω, $(t, x, \omega) \in \mathscr{X}_T \times \bar{\mathbb{M}}$, and merely write $\varphi(t, x)$. For $\varphi \in \mathscr{A}$, define a counting process N^φ on $\mathscr{X}_T$ by setting

$$N^\varphi((0, t] \times U) \doteq \int_{(0,t]\times U} \int_{(0,\infty)} 1_{[0,\varphi(s,x)]}(r)\bar{N}(ds \times dx \times dr) \tag{8.16}$$

for all $t \in [0, T]$, $U \in \mathscr{B}(\mathscr{X})$. Here N^φ is to be thought of as a controlled random measure, with $\varphi(s, x)$ selecting the intensity for the points at location x and time s, in a possibly random but nonanticipating way. Figure 8.1 illustrates how, for some particular value x, the control modulates the jump rate by "thinning", i.e., keeping only the jumps corresponding to atoms of $\bar{N}$ that lie below $\varphi(t, x)$ at time t.

When $\varphi(s, x, \omega) = \theta$ for all $(s, x, \omega) \in \mathscr{X}_T \times \bar{\mathbb{M}}$ and some $\theta > 0$, we write N^φ as N^θ. Note that N^θ has the same distribution on $\bar{\mathbb{M}}$ with respect to $\bar{P}$ as N has on $\mathbb{M}$ with respect to P_θ. Therefore, N^θ plays the role of N on $\bar{\mathbb{M}}$.

Define $\ell : [0, \infty) \to [0, \infty)$ by

$$\ell(r) \doteq r \log r - r + 1, \;\; r \in [0, \infty),$$

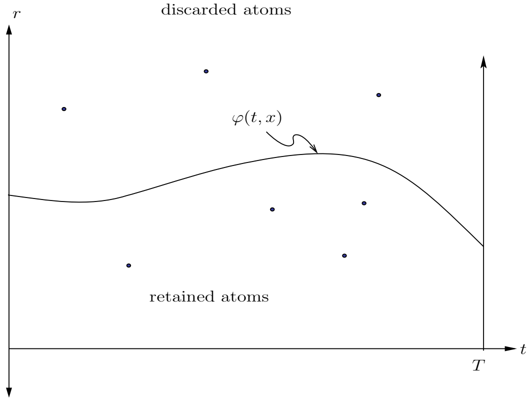

Fig. 8.1 Thinning by the control for a particular x

with the convention that $0 \log 0 = 0$. As is well known, ℓ is the local rate function (see Sect. 4.3 for this terminology) for a scaled standard Poisson process, and so it is not surprising that it plays a key role in our analysis. For $\varphi \in \mathscr{A}$, define a $[0, \infty]$-valued random variable $L_T(\varphi)$ by

$$L_T(\varphi)(\omega) \doteq \int_{\mathscr{X}_T} \ell(\varphi(t, x, \omega))\, \nu_T(dt \times dx), \;\; \omega \in \bar{\mathbb{M}}. \tag{8.17}$$

As in the setting of a Wiener process, it is convenient that the representation hold with a more restrictive class of controls. Let $\{K_n\}_{n\in\mathbb{N}}$ be an increasing sequence of compact subsets of $\mathscr{X}$ such that $\cup_{n=1}^{\infty} K_n = \mathscr{X}$. For each $M \in (0, \infty)$, let

$$\begin{aligned} \mathscr{A}_{b,M} \doteq \{\varphi \in \mathscr{A} : L_T(\varphi) \leq M \text{ a.e. and for some } n \in \mathbb{N},\; n \geq \varphi(t, x, \omega) \geq 1/n \\ \text{and } \varphi(t, x, \omega) = 1 \text{ if } x \in K_n^c, \text{ for all } (t, \omega) \in [0, T] \times \bar{\mathbb{M}}\}, \end{aligned} \tag{8.18}$$

and let

$$\mathscr{A}_b \doteq \cup_{M=1}^{\infty} \mathscr{A}_{b,M}. \tag{8.19}$$

As before, we let $\bar{\mathscr{A}}_{b,M}$, $\bar{\mathscr{A}}$, and $\bar{\mathscr{A}}_b$ denote the analogous spaces of controls when the canonical filtration $\{\mathscr{G}_t\}_{0\leq t\leq T}$ is replaced by a filtration $\{\mathscr{F}_t\}_{0\leq t\leq T}$ with the property that $\bar{N}$ is an $\mathscr{F}_t$-PRM with the same intensity.

The following is the representation theorem for PRM. The first equality holds because N under E_θ has the same distribution as N^θ under $\bar{E}$, as was discussed below (8.16).

Theorem 8.12 *Let $G \in \mathscr{M}_b(\mathbb{M})$. Then for $\theta > 0$,*

$$\begin{aligned} -\log E_\theta \exp\{-G(N)\} &= -\log \bar{E} \exp\{-G(N^\theta)\} \\ &= \inf_{\varphi \in \mathscr{R}} \bar{E}\left[\theta L_T(\varphi) + G(N^{\theta\varphi})\right], \end{aligned} \tag{8.20}$$

where $\mathscr{R}$ can be either $\mathscr{A}_b$, $\mathscr{A}$, $\bar{\mathscr{A}}_b$ or $\bar{\mathscr{A}}$.

The following is the analogue of Theorem 8.4 for the Poisson noise case.

Theorem 8.13 *Let $G \in \mathscr{M}_b(\mathbb{M})$ and let $\delta > 0$. Then there exist $M < \infty$ depending on $\|G\|_\infty$ and δ such that for all $\varepsilon \in (0, 1)$,*

$$-\varepsilon \log \bar{E} \exp\left\{-\frac{1}{\varepsilon} G(\varepsilon N^{1/\varepsilon})\right\} \geq \inf_{\varphi \in \mathscr{A}_{b,M}} \bar{E}\left[L_T(\varphi) + G(\varepsilon N^{\varphi/\varepsilon})\right] - \delta.$$

Proof Let $\delta > 0$. Using Theorem 8.12 with $\mathscr{R} = \mathscr{A}_b$, we can find, for each $\varepsilon \in (0, 1)$, $\tilde{\varphi}^\varepsilon \in \mathscr{A}_b$ such that

$$-\varepsilon \log \bar{E} \exp\left\{-\frac{1}{\varepsilon} G(\varepsilon N^{1/\varepsilon})\right\} \geq \bar{E}\left[L_T(\tilde{\varphi}^\varepsilon) + G(\varepsilon N^{\tilde{\varphi}^\varepsilon/\varepsilon})\right] - \delta/2.$$

From the boundedness of G, we obtain

$$\sup_{\varepsilon \in (0,1)} \bar{E}\left[L_T(\tilde{\varphi}^\varepsilon)\right] \leq C_G \doteq (2\|G\|_\infty + 1).$$

For $M \in \mathbb{N}$, let

$$\tau_M^\varepsilon(\omega) = \inf\left[t \in [0, T] : \int_{[0,t]\times\mathscr{X}} \ell(\tilde{\varphi}^\varepsilon(s, x, \omega))\, \nu_T(ds \times dx) \geq M\right] \wedge T.$$

Note that

$$\varphi^\varepsilon(s, x) \doteq 1 + (\tilde{\varphi}^\varepsilon(s, x) - 1)\, 1_{[0,\tau_M^\varepsilon]}(s), \;\; (s, x) \in \mathscr{X}_T$$

is an element of $\mathscr{A}_{b,M}$. Also,

$$\begin{aligned}\bar{E}&\left[L_T(\tilde{\varphi}^\varepsilon) + G(\varepsilon N^{\tilde{\varphi}^\varepsilon/\varepsilon})\right] \\ &\geq \bar{E}\left[L_T(\varphi^\varepsilon) + G(\varepsilon N^{\varphi^\varepsilon/\varepsilon})\right] + \bar{E}\left[G(\varepsilon N^{\tilde{\varphi}^\varepsilon/\varepsilon}) - G(\varepsilon N^{\varphi^\varepsilon/\varepsilon})\right].\end{aligned}$$

By Chebyshev's inequality,

$$\bar{E}\left|G(\varepsilon N^{\tilde{\varphi}^\varepsilon/\varepsilon}) - G(\varepsilon N^{\varphi^\varepsilon/\varepsilon})\right| \leq 2\|G\|_\infty \bar{P}\left\{\tau_M^\varepsilon < T\right\} \leq 2\|G\|_\infty \frac{C_G}{M}.$$

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

$$-\varepsilon \log \bar{E} \exp\left\{-\frac{1}{\varepsilon} G(\varepsilon N^{1/\varepsilon})\right\} \geq \bar{E}\left[L_T(\varphi^\varepsilon) + G(\varepsilon N^{\varphi^\varepsilon/\varepsilon})\right] - \delta,$$

as desired. □

Remark 8.14 We note that Theorem 3.23 is a special case of Theorems 8.12 and 8.13. To see this, consider the case $\mathscr{X} \doteq \{0\}$ and $\nu \doteq \delta_0$. Define $\gamma : \mathbb{M} \to \mathscr{D}([0,T] : \mathbb{R})$ by $\gamma(m)(t) \doteq m((0,t] \times \{0\})$, $t \in [0,T]$, $m \in \mathbb{M}$. Then γ is a Borel measurable map, and thus by Theorem 8.12, for every bounded Borel measurable function G mapping $\mathscr{D}([0,T] : \mathbb{R})$ to $\mathbb{R}$ and $\theta \in (0,\infty)$, we have

$$-\log \bar{E} \exp\{-G \circ \gamma(N^\theta)\} = \inf_{\varphi \in \mathscr{A}} \bar{E}\left[\theta L_T(\varphi) + G \circ \gamma(N^{\theta\varphi})\right]. \tag{8.21}$$

Recalling the definition of $\mathscr{X}$ and ν, any $\varphi \in \mathscr{A}$ can be identified with a nonnegative predictable process, and $\gamma(N^{\theta\varphi})$ is a controlled Poisson process in the sense of Sect. 3.3; in particular, $\gamma(N^\theta)$ is a Poisson process with rate θ (denoted in Sect. 3.3 by N^θ). Thus the first representation in Theorem 3.23 follows readily from (8.21) with $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}, \{\bar{\mathscr{F}}_t\}) = (\bar{M}, \mathscr{B}(\bar{M}), \bar{P}, \{\mathscr{G}_t\})$. The second representation in Theorem 3.23 follows similarly from Theorem 8.13.

The rest of this section is devoted to the proof of Theorem 8.12. For notational convenience we provide details only for the case $\theta = 1$. The general case is treated similarly. As in the case of Brownian motion, we first consider $\mathscr{A}_b$ and $\mathscr{A}$, and then extend to $\bar{\mathscr{A}}_b$ and $\bar{\mathscr{A}}$.

8.2.2 Preparatory Results

Recall that $\mathscr{PF}$ denotes the predictable σ-field associated with the augmented PRM $\bar{N}$, and that $\mathscr{Y} = \mathscr{X} \times [0,\infty)$. A class of processes that will be used as test functions is defined as follows. Let $\hat{A}_b$ be the set of all $(\mathscr{PF} \otimes \mathscr{B}(\mathscr{Y})) \backslash \mathscr{B}(\mathbb{R})$-measurable maps $\vartheta : \mathscr{Y}_T \times \bar{\mathbb{M}} \to \mathbb{R}$ that are bounded and such that for some compact $K \subset \mathscr{Y}$, $\vartheta(s,x,r,\omega) = 0$ whenever $(x,r) \in K^c$. Once again ω will usually be suppressed in the notation. The following result is standard (see, e.g., Theorem III.3.24 of [161]), and the analogue with $\mathscr{A}_b$ replaced by $\bar{\mathscr{A}}_b$ and $\mathscr{G}_t$ by $\mathscr{F}_t$ also holds. Let N_c^1 be the compensated version of N^1, which is defined by $N_c^1(A) \doteq N^1(A) - \nu_T(A)$ for all $A \in \mathscr{B}(\mathscr{X}_T)$ such that $\nu_T(A) < \infty$.

Theorem 8.15 (Girsanov) *Let $\varphi \in \mathscr{A}_b$. Then*

$$\begin{aligned}\mathscr{E}^\varphi(t) &\doteq \exp\left\{\int_{(0,t]\times\mathscr{X}} \log(\varphi(s,x)) N_c^1(ds \times dx)\right. \\ &\qquad \left. + \int_{(0,t]\times\mathscr{X}} (\log(\varphi(s,x)) - \varphi(s,x) + 1)\, \nu_T(ds \times dx)\right\} \\ &= \exp\left\{\int_{(0,t]\times\mathscr{X}\times[0,1]} \log(\varphi(s,x)) N(ds \times dx \times dr)\right. \\ &\qquad \left. + \int_{(0,t]\times\mathscr{X}\times[0,1]} (-\varphi(s,x) + 1)\, \bar{\nu}_T(ds \times dx \times dr)\right\}\end{aligned} \tag{8.22}$$

is a $\mathcal{G}_t$-martingale. Define a probability measure $\bar{Q}^\varphi$ on $\bar{\mathbb{M}}$ by

$$\bar{Q}^\varphi(H) = \int_H \mathcal{E}^\varphi(T)d\bar{P} \text{ for } H \in \mathcal{B}(\bar{\mathbb{M}}),$$

and let $\bar{E}^\varphi$ denote integration with respect to $\bar{Q}^\varphi$. Then for every $\vartheta \in \hat{A}_b$,

$$\begin{aligned}\bar{E}^\varphi & \int_{\mathcal{Y}_T} \vartheta(s,x,r)\bar{N}(ds \times dx \times dr) \\ &= \bar{E}^\varphi \int_{\mathcal{Y}_T} \vartheta(s,x,r)\left[\varphi(s,x)1_{(0,1]}(r) + 1_{(1,\infty)}(r)\right]\bar{\nu}_T(ds \times dx \times dr).\end{aligned}$$

The last statement in the lemma says that under $\bar{Q}^\varphi$, $\bar{N}$ is a random counting measure with compensator $\left[\varphi(s,x)1_{(0,1]}(r) + 1_{(1,\infty)}(r)\right]\bar{\nu}_T(ds \times dx \times dr)$.

Recall that $\mathcal{X} = \cup_{n=1}^\infty K_n$ for increasing compact sets K_n. A process $\varphi \in \mathcal{A}_{b,M}$ is in the set $\mathcal{A}_{s,M}$ if the following holds. There exist $n, \ell, n_1, \ldots, n_\ell \in \mathbb{N}$; a partition $0 = t_0 < t_1 < \cdots < t_\ell = T$; for each $i = 1, \ldots, \ell$ a disjoint measurable partition E_{ij} of K_n, $j = 1, \ldots, n_i$; $\mathcal{G}_{t_{i-1}}$-measurable random variables X_{ij}, $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$, such that $1/n \le X_{ij} \le n$; and

$$\varphi(t,x,\bar{m}) = 1_{\{0\}}(t) + \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} 1_{(t_{i-1},t_i]}(t)X_{ij}(\bar{m})1_{E_{ij}}(x) + 1_{K_n^c}(x)1_{(0,T]}(t). \quad (8.23)$$

We let $\mathcal{A}_s \doteq \cup_{M=1}^\infty \mathcal{A}_{s,M}$ and refer to elements in $\mathcal{A}_s$ as *simple* processes.

Lemma 8.16 *Let $\varphi \in \mathcal{A}_b$. Then there exists a sequence of processes $\varphi_k \in \mathcal{A}_s$ with the following properties.*

(a) *N^{φ_k} converges in distribution to N^φ as $k \to \infty$.*

(b) *$\bar{E}|L_T(\varphi_k) - L_T(\varphi)| \to 0$ and $\bar{E}|\mathcal{E}^{\varphi_k}(T) - \mathcal{E}^\varphi(T)| \to 0$, as $k \to \infty$.*

Proof We first construct processes φ_k that satisfy parts (a) and (b) of the lemma but that instead of being simple are continuous in t. Since $\varphi \in \mathcal{A}_b$, we have for some $n \in \mathbb{N}$ that $n \ge \varphi(t,x,\omega) \ge 1/n$ and $\varphi(t,x,\omega) = 1$ if $x \in K_n^c$ for all $(t,\omega) \in [0,T] \times \bar{\mathbb{M}}$. For $k \in \mathbb{N}$, define

$$\varphi_k(t,x,\omega) = k\left(\frac{1}{k} - t\right)^+ + k\int_{(t-\frac{1}{k})^+}^t \varphi(s,x,\omega)ds, \;\; (t,x,\omega) \in \mathcal{X}_T \times \bar{\mathbb{M}}.$$

An application of Lusin's theorem gives that for $\nu \times \bar{P}$-a.e. (x,ω), as $k \to \infty$,

$$\begin{aligned}\int_{[0,T]} |\varphi_k(t,x,\omega) - \varphi(t,x,\omega)|dt \to 0 \\ \int_{[0,T]} |\ell(\varphi_k(t,x,\omega)) - \ell(\varphi(t,x,\omega))|dt \to 0.\end{aligned} \quad (8.24)$$

In particular, $\varphi_k \in \mathscr{A}_b$ for every k and $\bar{E}|L_T(\varphi_k) - L_T(\varphi)| \to 0$, as $k \to \infty$. It follows from (D.6) and the definition of the controlled PRM that for $g \in \mathscr{C}_c(\mathscr{X}_T)$,

$$\begin{aligned}
&\bar{E}\,|\langle g, N^{\varphi_k}\rangle - \langle g, N^{\varphi}\rangle| \\
&\quad \le \bar{E}\int_{\mathscr{Y}_T} |g(s,x)||1_{[0,\varphi_k(s,x,\omega)]}(r) - 1_{[0,\varphi(s,x,\omega)]}(r)|\,\bar{\nu}_T(ds\times dx\times dr) \\
&\quad \le \|g\|_\infty\,\bar{E}\int_{[0,T]\times K_n} |\varphi_k(s,x,\omega) - \varphi(s,x,\omega)|\,\nu_T(ds\times dx).
\end{aligned}$$

Using (8.24), $\nu(K_n) < \infty$, and the uniform bounds on φ_k and φ shows that the last quantity approaches 0 as $k \to \infty$, and hence by Lemma A.10, $N^{\varphi_k} \Rightarrow N^{\varphi}$.

Next we consider the convergence of $\mathscr{E}^{\varphi_k}(T)$ in $\mathscr{L}^1(\bar{P})$. By Scheffe's lemma [249, Sect. 5.11], if $f_k(\omega)$ and $f(\omega)$ are densities with respect to $\bar{P}$ such that $f_k \to f$ in probability, then the convergence is also in $\mathscr{L}^1(\bar{P})$. Thus it suffices to show that

$$\mathscr{E}^{\varphi_k}(T) \to \mathscr{E}^{\varphi}(T) \text{ in } \bar{P}\text{-probability.} \tag{8.25}$$

For this, it is enough to show [see (8.22)] that

$$\int_{\mathscr{X}_T} (1-\varphi_k(s,x))\,\nu_T(ds\times dx) \to \int_{\mathscr{X}_T} (1-\varphi(s,x))\,\nu_T(ds\times dx)$$

and since $N^1 = N^1_c + \nu_T$, that

$$\int_{\mathscr{X}_T} \log(\varphi_k(s,x))N^1(ds\times dx) \to \int_{\mathscr{X}_T} \log(\varphi(s,x))N^1(ds\times dx)$$

in probability as $k \to \infty$. The first convergence is immediate from (8.24), the uniform bounds on φ_k, φ, $\nu(K_n) < \infty$, and the fact that $1-\varphi_k(s,x) = 1-\varphi(s,x) = 0$ for $x \notin K_n$. The second convergence follows similarly on noting that $\varphi_k(s,x) \wedge \varphi(s,x) \ge 1/n$ implies

$$|\log(\varphi_k(s,x)) - \log(\varphi(s,x))| \le n|\varphi_k(s,x) - \varphi(s,x)|.$$

This proves (8.25) and so $\bar{E}|\mathscr{E}^{\varphi_k}(T) - \mathscr{E}^{\varphi}(T)| \to 0$ as $k \to \infty$. This completes the construction of φ_k that satisfy parts (a) and (b) of the lemma.

Next we show that the processes can be assumed to be simple. Note that by construction, $t \mapsto \varphi_k(t,x,\omega)$ is continuous for $\nu \times \bar{P}$-a.e. (x,ω). Consider any φ_k as constructed previously, and to simplify the notation, drop the k subscript. Two more levels of approximation will be used, and indexed by q and r. Thus for the fixed φ and $q \in \mathbb{N}$, define

$$\varphi_q(t,x,\omega) = \sum_{m=0}^{\lfloor qT\rfloor} \varphi\left(\frac{m}{q}, x, \omega\right) 1_{(\frac{m}{q},\frac{m+1}{q}]}(t),\;\; (t,x,\omega) \in \mathscr{X}_T \times \bar{\mathbb{M}}.$$

It is easily checked that (8.24) is satisfied as $q \to \infty$, and so arguing as above, the sequence $\{\varphi_q\}$ satisfies parts (a) and (b) of the lemma. Note that for fixed q and m, $g(x, \omega) = \varphi(m/q, x, \omega)$ is a $\mathscr{B}(\mathscr{X})\otimes\bar{\mathscr{F}}_{m/q}$-measurable map with values in $[1/n, n]$ and $g(x, \omega) = 1$ for $x \in K_n^c$. By a standard approximation procedure one can find $\mathscr{B}(\mathscr{X})\otimes\bar{\mathscr{F}}_{m/q}$-measurable maps g_r, $r \in \mathbb{N}$ with the following properties:

$$g_r(x, \omega) = \sum_{j=1}^{a(r)} c_j^r(\omega) 1_{E_j^r}(x) \text{ for } x \in K_n,$$

where for each r, $a(n) \in \mathbb{N}$, $\{E_j^r\}_{j=1,\ldots,a(r)}$ is some measurable partition of K_n, and for all j, r, $c_j^r(\omega) \in [1/n, n]$ a.s.; $g_r(x, \omega) = 1$ for $x \in K_n^c$; $g_r \to g$ a.s. $\nu \times \bar{P}$. If we make such an approximation for each m and label the process obtained when the g's are replaced by g_r's as φ_q^r, then $\varphi_q^r \in \mathscr{A}_s$. Hence by the triangle inequality we can find a sequence $\{\varphi_k\} \subset \mathscr{A}_s$ such that (a) and (b) hold. $\square$

A last result is needed before we can prove the main theorem. As in the case of the Wiener process, we need to know that controls under the original probability space can be replicated on a new space. The proof of the lemma, which uses an elementary but detailed argument, is given at the end of the chapter.

Lemma 8.17 *For every $\varphi \in \mathscr{A}_s$ there is $\hat{\varphi} \in \mathscr{A}_s$ such that $(L_T(\hat{\varphi}), N^1)$ has the same distribution under $\bar{Q}^{\hat{\varphi}}$ as $(L_T(\varphi), N^{\varphi})$ does under $\bar{P}$.*

We now proceed to the proof of Theorem 8.12. We will provide the proof only for the case that $\mathscr{R}$ is $\mathscr{A}$ or $\mathscr{A}_b$. The general filtration setting, i.e., when $\mathscr{R} = \bar{\mathscr{A}}, \bar{\mathscr{A}}_b$, can be treated as in Sect. 8.1.5.

8.2.3 Proof of the Upper Bound in the Representation

In this subsection we prove (recall that we present the argument only for $\theta = 1$)

$$\begin{aligned} -\log E_1 \exp\{-G(N)\} &= -\log \bar{E} \exp\{-G(N^1)\} \\ &\leq \inf_{\varphi \in \mathscr{A}} \bar{E}\left[L_T(\varphi) + G(N^{\varphi})\right]. \end{aligned} \tag{8.26}$$

Note that this automatically gives the corresponding bound for the smaller class $\mathscr{A}_b$ in (8.20).

The proof parallels that of the case of a Wiener process. Let $h : \bar{\mathbb{M}} \to \mathbb{M}$ be defined by

$$h(\bar{m})((0, t] \times U) = \int_{(0,t]\times U\times(0,\infty)} 1_{[0,1]}(r)\bar{m}(ds \times dx \times dr)$$

for $t \in [0, T]$, $U \in \mathscr{B}(\mathscr{X})$. Thus $N^1 = h(\bar{N})$. Recalling (2.1), we have

$$
\begin{aligned}
-\log \bar{E}\exp\{-G(N^1)\} &= -\log \int_{\bar{\mathbb{M}}} \exp\{-G(h(\bar{m}))\}\bar{P}(d\bar{m}) \\
&= \inf_{\bar{Q}\in\mathscr{P}(\bar{\mathbb{M}})} \left[R\left(\bar{Q}\,\|\,\bar{P}\right) + \int_{\bar{\mathbb{M}}} G(h(\bar{m}))\bar{Q}(d\bar{m})\right]. \qquad (8.27)
\end{aligned}
$$

We begin by evaluating $R\left(\bar{Q}^{\varphi}\,\|\,\bar{P}\right)$ for $\varphi \in \mathscr{A}_b$. By Theorem 8.15, $\{\mathscr{E}^{\varphi}(t)\}$ [defined in (8.22)] is an $\mathscr{G}_t$-martingale, and under $\bar{Q}^{\varphi}$, $\bar{N}$, it is a random counting measure with compensator $\left[\varphi(s,x)1_{(0,1]}(r) + 1_{(1,\infty)}(r)\right]\bar{\nu}_T(ds\times dx\times dr)$. It follows from the definition of relative entropy and L_T in (8.17) that

$$
\begin{aligned}
&R\left(\bar{Q}^{\varphi}\,\|\,\bar{P}\right) \\
&\quad= \bar{E}^{\varphi}\left[\int_{\mathscr{X}_T} \log(\varphi(s,x))N_c^1(ds\times dx) \right.\\
&\qquad\qquad \left. + \int_{\mathscr{X}_T} \left(\log(\varphi(s,x)) - \varphi(s,x) + 1\right)\nu_T(ds\times dx)\right] \\
&\quad= \bar{E}^{\varphi}\left[\int_{\mathscr{X}_T} \log(\varphi(s,x))N^1(ds\times dx) + \int_{\mathscr{X}_T} \left(-\varphi(s,x) + 1\right)\nu_T(ds\times dx)\right] \\
&\quad= \bar{E}^{\varphi}\left[\int_{\mathscr{X}_T} \left(\varphi(s,x)\log(\varphi(s,x)) - \varphi(s,x) + 1\right)\nu_T(ds\times dx)\right] \\
&\quad= \bar{E}^{\varphi}L_T(\varphi). \qquad (8.28)
\end{aligned}
$$

Thus by (8.27), for $\varphi \in \mathscr{A}_b$, we have

$$
\begin{aligned}
-\log \bar{E}\exp\{-G(N^1)\} &\le \left[R\left(\bar{Q}^{\varphi}\,\|\,\bar{P}\right) + \int_{\bar{\mathbb{M}}} G(h(\bar{m}))\bar{Q}^{\varphi}(d\bar{m})\right] \qquad (8.29)\\
&= \bar{E}^{\varphi}\left[L_T(\varphi) + G(N^1)\right].
\end{aligned}
$$

The rest of the proof is in three steps.

Step 1. Simple φ. Suppose one is given $\varphi \in \mathscr{A}_s$. According to Lemma 8.17, one can find $\tilde{\varphi}$ that is $\mathscr{G}_t$-predictable and simple and such that $(\tilde{\varphi}, N^1)$ under $\bar{Q}^{\tilde{\varphi}}$ has the same distribution as (φ, N^{φ}) under $\bar{P}$. This implies

$$
\bar{E}^{\tilde{\varphi}}\left[L_T(\tilde{\varphi}) + G(N^1)\right] = \bar{E}\left[L_T(\varphi) + G(N^{\varphi})\right],
$$

and thus the desired inequality follows directly from (8.29).

Step 2. Bounded φ. Given $\varphi \in \mathscr{A}_b$, let $\varphi_k \in \mathscr{A}_s$ be the sequence constructed in Lemma 8.16. By Step 1, for every $k \in \mathbb{N}$,

$$
-\log \bar{E}\exp\{-G(N^1)\} \le \bar{E}\left[L_T(\varphi_k) + G(N^{\varphi_k})\right]. \qquad (8.30)
$$

From Lemma 8.16, under $\bar{P}$, we have $N^{\varphi_k} \Rightarrow N^{\varphi}$, and $\bar{E}\left[L_T(\varphi_k)\right] \to \bar{E}\left[L_T(\varphi)\right]$. However, G is not assumed continuous, and so we cannot simply pass to the limit in the last display. Instead, we will apply Lemma 2.5, which requires bounds on relative entropies. The first and the last equalities in the following display follow from Lemma 8.17, the second equality is a consequence of (8.28), and the inequality follows from the fact that relative entropy can only decrease when one is considering measures induced by the same mapping (in this case the random variable N^1) [see part (f) of Lemma 2.4]:

$$\begin{aligned} R(\bar{P} \circ (N^{\varphi_k})^{-1} \,\|\, \bar{P} \circ (N^1)^{-1}) &= R(\bar{Q}^{\tilde{\varphi}_k} \circ (N^1)^{-1} \,\|\, \bar{P} \circ (N^1)^{-1}) \\ &\leq R(\bar{Q}^{\tilde{\varphi}_k} \,\|\, \bar{P}) \\ &= \bar{E}^{\tilde{\varphi}_k} \left[L_T(\tilde{\varphi}_k) \right] \\ &= \bar{E} \left[L_T(\varphi_k) \right]. \end{aligned} \tag{8.31}$$

Since $\bar{E}[L_T(\varphi_k)] \to \bar{E}[L_T(\varphi)] < \infty$, the relative entropies in (8.31) are uniformly bounded in k. By Lemma 8.16 we can pass to the limit in (8.30) and obtain (8.26) when $\mathscr{A}$ is replaced by $\mathscr{A}_b$. For future use, note that the lower semicontinuity of relative entropy and (8.31) imply

$$R(\bar{P} \circ (N^{\varphi})^{-1} \,\|\, \bar{P} \circ (N^1)^{-1}) \leq \bar{E}[L_T(\varphi)] \text{ for } \varphi \in \mathscr{A}_b.$$

Step 3. General φ. For $\varphi \in \mathscr{A}$, define

$$\varphi_n(t,x,\omega) = \begin{cases} [\varphi(t,x,\omega) \vee (1/n)] \wedge n, & x \in K_n, \\ 1 & \text{otherwise.} \end{cases}$$

Note that $\varphi_n \in \mathscr{A}_b$, and so (8.30) holds with φ_k replaced by φ_n. Since the definition of φ_n implies that $\ell(\varphi_n(x,t,\omega))$ is nondecreasing in n, by the monotone convergence theorem, we have $\bar{E} L_T(\varphi_n) \uparrow \bar{E} L_T(\varphi)$. If $\bar{E} L_T(\varphi) = \infty$, there is nothing to prove. Assume therefore that

$$\bar{E} L_T(\varphi) < \infty. \tag{8.32}$$

Then $R(\bar{P} \circ (N^{\varphi_n})^{-1} \,\|\, \bar{P} \circ (N^1)^{-1}) \leq \bar{E} L_T(\varphi_n) \leq \bar{E} L_T(\varphi)$. We claim that N^{φ_n} converges in distribution to N^{φ}. If true, then using the uniform bound on relative entropies just noted, we can once again apply Lemma 2.5, pass to the limit on n, and thereby obtain (8.26).

Let $g \in \mathscr{C}_c(\mathscr{X}_T)$ and let n_0 be large enough that the support of g is contained in $[0,T] \times K_{n_0}$. Then for all $n \geq n_0$,

$$\bar{E}\, |\langle g, N^{\varphi_n} \rangle - \langle g, N^{\varphi} \rangle| \leq \|g\|_\infty \; \bar{E} \int_{[0,T]\times K_{n_0}} \left(\frac{1}{n} + (\varphi(s,x) - n)^+ \right) \nu_T(ds \times dx).$$

Note that $\nu_T([0,T] \times K_{n_0}) < \infty$, $(\varphi(t,x) - n)^+ \to 0$ as $n \to \infty$, and $(\varphi(t,x) - n)^+ \leq \ell(\varphi(t,x))$. These observations together with (8.32) show that the right-hand side in the last display approaches 0 as $n \to \infty$. We can therefore apply Lemma A.10, and $N^{\varphi_n} \Rightarrow N^{\varphi}$ follows. □

8.2.4 Proof of the Lower Bound in the Representation

In this subsection we prove (again only for $\theta = 1$)

$$-\log E_1 \exp\{-G(N)\} = -\log \bar{E} \exp\{-G(N^1)\} \\ \geq \inf_{\varphi \in \mathcal{A}_b} \bar{E}\left[L_T(\varphi) + G(N^{\varphi})\right], \tag{8.33}$$

which automatically gives the lower bound for the larger class $\mathcal{A}$ in (8.20). As in the Brownian motion case, the proof is in two steps.

Step 1. G of a particular form. Let $K, M \in \mathbb{N}$ be arbitrary, and consider any collection $0 = t_1 < t_2 < \cdots < T_K = T$. Let $h : \mathbb{N}_0^{KM} \to \mathbb{R}$ be a bounded map. Let $C_1, \ldots, C_M$ be precompact sets in $\mathscr{X}$ such that $C_i \cap C_j = \emptyset$ if $i \neq j$. Then G is of the form

$$G(N^1) = h(n(t_1), n(t_2) - n(t_1), \ldots, n(t_K) - n(t_{K-1})), \tag{8.34}$$

where for $0 \leq t \leq T$,

$$n(t) = (n_1(t), \ldots, n_M(t)) = (N^1((0,t] \times C_1), \ldots, N^1((0,t] \times C_M)). \tag{8.35}$$

For G of this particular form, we will construct $\varphi \in \mathcal{A}_b$ such that

$$-\log \bar{E} \exp\{-G(N^1)\} = \bar{E}\left[L_T(\varphi) + G(N^{\varphi})\right],$$

from which (8.33) is immediate. The underlying idea is the same as in the Brownian case. Using a conditioning argument, over each interval of the form $[t_i, t_{i+1}]$ we can interpret the logarithm of an exponential integral as the value function of a stochastic control problem. From the boundedness of h, the integral is smooth in the time variable, which means that the control problem has a classical-sense solution, and then an optimal control can be found from this solution and the corresponding dynamic programming equation. The controls over the various intervals are concatenated to produce $\varphi \in \mathcal{A}_b$, which actually achieves the infimum for the given G. The following lemma is analogous to Lemma 8.8 for the Brownian motion case and can be proved in a similar manner. When applied, the k in the statement of the lemma will be of the form jM, $j = 0, \ldots, K-1$.

Lemma 8.18 *Let $g : \mathbb{N}_0^k \times \mathbb{N}_0^M \to \mathbb{R}$ be uniformly bounded, and let $\{n(t)\}_{0 \leq t \leq T}$ be as in (8.35). Define $V : [0,T] \times \mathbb{N}_0^k \times \mathbb{N}_0^M \to \mathbb{R}$ by*

$$V(t,z,x) \doteq -\log \bar{E} e^{-g(z, x+n(T-t))}, \ \ (t,z,x) \in [0,T] \times \mathbb{N}_0^k \times \mathbb{N}_0^M.$$

For $(t, z, x) \in [0, T] \times \mathbb{N}_0^k \times \mathbb{N}_0^M$, *let*

$$\partial_{x_i} V(t, z, x) \doteq V(t, z, x + e_i) - V(t, z, x),$$

where $\{e_i\}_{i=1}^M$ *is the coordinate basis in* $\mathbb{R}^M$. *Then for each* z, x *and* i, $\partial_{x_i} V(t, z, x)$ *is continuous in* $t \in [0, T]$. *Let* $\{X(z, t)\}_{0 \le t \le T}$ *with* $X(z, t) = (X_1(z, t), \ldots, X_M(z, t))$ *be the unique solution of*

$$X_i(z, t) = \int_{(0,t]\times C_i \times [0,\infty)} 1_{[0,\exp\{-\partial_{x_i} V(s,z,X(z,s-))\}]}(r) \bar{N}(ds \times dx \times dr), \quad (8.36)$$

$i = 1, \ldots, M$. *Define*

$$\varphi(t, x) = \sum_{i=1}^{M} \varphi_i(t) 1_{C_i}(x) + 1_{C^c}(x), \; (t, x) \in [0, T] \times \mathscr{X},$$

where $C = \cup_{i=1}^M C_i$ *and* $\varphi_i(t) = \exp\{-\partial_{x_i} V(t, z, X(z, t-))\}$. *Then*

$$-\log \bar{E} \exp\{-g(z, n(T))\} = \bar{E}\left[L_T(\varphi) + g\left(z, n^{\varphi}(T)\right)\right],$$

where for $t \in [0, T]$,

$$n^{\varphi}(t) = (n_1^{\varphi}(t), \ldots, n_M^{\varphi}(t)) = (N^{\varphi}((0, t] \times C_1), \ldots, N^{\varphi}((0, t] \times C_M)). \quad (8.37)$$

Since $t \mapsto \partial_{x_i} V(t, z, x)$ is continuous and $\nu(C_i) < \infty$ for all (t, z, x) and $i = 1, \ldots, M$, the solution to (8.36) will jump a finite number of times a.s. over $[0, T]$. The equations can be solved recursively by updating the relevant component $X_i(z, \bar{t})$, $i = 1, \ldots, M$ if a jump occurs at time $\bar{t}$, using (8.36) to identify the next time that one of the components will jump, and repeating.

We next apply Lemma 8.18 recursively. For $j = 1, \ldots, K$, define $V_j : \mathbb{N}_0^{jM} \to \mathbb{R}$ as follows: $V_K = h$ and

$$V_j(z_j) = -\log \bar{E} e^{-V_{j+1}(z_j, n(t_{j+1}) - n(t_j))}, \; z_j \in \mathbb{N}_0^{jM}, \; j = 1, \ldots, K-1.$$

By successive conditioning it is easily checked that

$$V_0 \doteq -\log \bar{E} e^{-V_1(n(t_1) - n(t_0))} = -\log E e^{-G(N)}.$$

Note that V_j is a bounded map for each j. For $j = 1, \ldots, K$, let $Z_j = (n(t_1), n(t_2) - n(t_1), \ldots, n(t_j) - n(t_{j-1}))$ and note that Z_j is a $\mathbb{N}_0^{jM}$-valued random variable. For $z_j \in \mathbb{N}_0^{jM}$ and $j = 1, \ldots, K-1$, let $\{Y(z_j, t)\}_{t \in [t_j, t_{j+1}]}$ be the unique solution of

$$Y_i(z_j, t) = \int_{(t_j,t]\times C_i \times [0,\infty)} 1_{[0,\exp\{-\partial_{x_i} V_{j+1}(s,z_j,Y(z_j,s-))\}]}(r) \bar{N}(ds \times dx \times dr)$$

for $t \in [t_j, t_{j+1}]$, where $Y(z_j, t) = (Y_1(z_j, t), \ldots, Y_M(z_j, t))$. For $i = 1, \ldots, M$, define

$$\varphi_i(t) = \exp\{-\partial_{x_i} V_{j+1}(t, Z_j, Y(Z_j, t-))\},\ t \in [t_j, t_{j+1}),\ j = 0, \ldots, K-1$$

and

$$\varphi(t,x) = \sum_{i=1}^{M} \varphi_i(t,x) 1_{C_i}(x) + 1_{C^c}(x),\ (t,x) \in [0,T] \times \mathscr{X}.$$

Then by a recursive argument using Lemma 8.18, we see that

$$\begin{aligned} -\log E e^{-G(N)} &= \bar{E}\left[L_T(\varphi) + h(n^{\varphi}(t_1), \ldots, n^{\varphi}(t_K) - n^{\varphi}(t_{K-1}))\right] \\ &= \bar{E}\left[L_T(\varphi) + G(N^{\varphi})\right], \end{aligned}$$

where in the first line, n^{φ} is defined as in (8.37). Note that by construction, $\varphi \in \mathscr{A}_b$. Thus we have proved (8.33) for all G of the form (8.34).

Step 2. G that is bounded and measurable. Now suppose that G is simply bounded and measurable. We claim that there exist functions $\{G_n\}_{n\in\mathbb{N}}$ such that for each n, G_n is of the form assumed in Step 1, $\|G_n\|_\infty \le \|G\|_\infty$, and $G_n \to G$ a.s. with respect to P. This is shown, as in the Brownian case, by noting that each of the following classes admits an approximation of this form relative to elements of the preceding class, save of course the first:

- G bounded and measurable;
- G bounded and continuous;
- G bounded and continuous and depending on N^1 only through

$$\{N^1((0,t_k] \times \cdot)\}_{k=1,\ldots,K};$$

  where $K \in \mathbb{N}$ and $0 = t_1 < t_2 < \cdots < T_K = T$ are arbitrary;
- G bounded and depending on N^1 only through

$$\{N^1((0,t_k] \times C_i)\}_{k=1,\ldots,K, i=1,\ldots,M},$$

  where $K, M \in \mathbb{N}$, $0 = t_1 < t_2 < \cdots < T_K = T$ and $C_1, \ldots, C_M$ are precompact sets in $\mathscr{X}$ such that $C_i \cap C_j = \emptyset$ if $i \neq j$.

As before, these approximations follow by standard arguments based on Theorem E.4 and the martingale convergence theorem. We now complete the lower bound in exactly the same way as in the case of Brownian motion. With each $n \in \mathbb{N}$ we can associate $\varphi_n \in \mathscr{A}_b$ such that

$$-\log E \exp\{-G_n(N)\} = \bar{E}\left[L_T(\varphi_n) + G_n(N^{\varphi_n})\right].$$

If Q_n is the distribution induced by N^{φ_n}, then [see (8.31)]

$$R(Q_n \| P) \leq \bar{E}\,[L_T(\varphi_n)] \leq 2\,\|G\|_\infty\,.$$

Thus $\{Q_n\}$ is tight, and from part (b) of Lemma 2.5,

$$\lim_{n\to\infty} \bar{E}\,|G_n(N^{\varphi_n}) - G(N^{\varphi_n})| = 0.$$

By the dominated convergence theorem,

$$\lim_{n\to\infty} |\log E \exp\{-G_n(N)\} - \log E \exp\{-G(N)\}| = 0.$$

Therefore, given $\varepsilon > 0$, we can find $n \in \mathbb{N}$ such that

$$\begin{aligned}
-\log E e^{-G(N)} &\geq -\log E e^{-G_n(N)} - \varepsilon \\
&= \bar{E}\left[L_T(\varphi_n) + G_n\left(N^{\varphi_n}\right)\right] - \varepsilon \\
&\geq \bar{E}\left[L_T(\varphi_n) + G\left(N^{\varphi_n}\right)\right] - 2\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary and $v_n \in \mathscr{A}_b$, we have (8.33), completing the proof. □

8.2.5 Construction of Equivalent Controls

In this section we give the proof of Lemma 8.17. We need to show that given $\varphi \in \mathscr{A}_s$, there is $\hat{\varphi} \in \mathscr{A}_s$ such that the distribution of $(L_T(\hat{\varphi}), N^1)$ under $\bar{\mathbb{Q}}^{\hat{\varphi}}$ is the same as that of $(L_T(\varphi), N^{\varphi})$ under $\bar{\mathbb{P}}$. Let φ be as in (8.23):

$$\varphi(t, x, \bar{m}) = 1_{\{0\}}(t) + \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} 1_{(t_{i-1},t_i]}(t) X_{ij}(\bar{m}) 1_{E_{ij}}(x) + 1_{K_n^c}(x) 1_{(0,T]}(t).$$

We will need some notation to describe how measures on $[0, T] \times \mathscr{Y}$ are decomposed into parts on subintervals of the form $(t_{i-1}, t_i]$, and also how after some manipulation such quantities can be recombined. For $i = 1, \ldots, \ell$, let $\mathbb{I}_i \doteq (t_{i-1}, t_i]$ and let $\mathscr{Y}_i \doteq \mathbb{I}_i \times \mathscr{Y}$. Denote by $\bar{\mathbb{M}}_i$ the space of nonnegative σ-finite integer-valued measures $\bar{m}_i$ on $(\mathscr{Y}_i, \mathscr{B}(\mathscr{Y}_i))$ that satisfy $\bar{m}_i(\mathbb{I}_i \times K) < \infty$ for all compact $K \subset \mathscr{Y}$. Endow $\bar{\mathbb{M}}_i$ with the weakest topology making the functions $m \mapsto \langle f, m\rangle$, $m \in \bar{\mathbb{M}}_i$ continuous, for every f in $\mathscr{C}_c(\mathscr{Y}_i)$ vanishing outside some compact subset of $\mathscr{Y}_i$. Denote by $\bar{\mathscr{M}}_i$ the corresponding Borel σ-field. Let $\bar{N}_i$ be the $\bar{\mathbb{M}}_i$-valued random variable on $(\bar{\mathbb{M}}, \mathscr{B}(\bar{\mathbb{M}}))$ defined by $\bar{N}_i(A) \doteq \bar{N}(A)$, $A \in \mathscr{B}(\mathscr{Y}_i)$. Also, define $\mathbb{J}_i \doteq [1/n, n]^{n_i}$, and the $\mathbb{J}_i$-valued random variable X_i by $X_i \doteq (X_{i1}, \ldots, X_{in_i})$. Let $\hat{\mathbb{M}} \doteq \bar{\mathbb{M}}_1 \times \cdots \times \bar{\mathbb{M}}_\ell$, and define $\varpi : \hat{\mathbb{M}} \to \bar{\mathbb{M}}$ by $\varpi(\hat{m}) = m$, where for $B \in \mathscr{B}(\mathscr{Y})$, $A \in \mathscr{B}[0, T]$, and with $\hat{m} = (m_1, \ldots, m_\ell)$, $m_i \in \bar{\mathbb{M}}_i$, we have

$$m(A \times B) = \sum_{i=1}^{q} m_i((A \cap \mathbb{I}_i) \times B).$$

With these definitions, ϖ concatenates the measures back together, and in particular, $\varpi((\bar{N}_1, \ldots, \bar{N}_\ell)) = \bar{N}$.

From the predictability properties of φ it follows that for $i = 2, \ldots, \ell$, there are measurable maps $\xi_i : \bar{\mathbb{M}}_1 \times \cdots \times \bar{\mathbb{M}}_{i-1} \to \mathbb{J}_i$, which can be written in component form $\xi_i = (\xi_{i1}, \ldots, \xi_{in_i})$ such that

$$X_{ij}(\bar{m}) = \xi_{ij}(\bar{N}_1(\bar{m}), \ldots, \bar{N}_{i-1}(\bar{m})).$$

Also, for $i = 1$, we set $X_1 = \xi_1$ a.s.-$\bar{\mathbb{P}}$ for some fixed vector ξ_1 in $\mathbb{J}_1$. The construction of $\hat{\varphi}$, which takes the same form as φ, is recursive. For $s \in \mathbb{I}_1$ we set $\hat{\varphi}(s, x, \bar{m}) = \varphi(s, x, \bar{m})$. As we will see, if there were only one time interval, we would be done, in that N^{φ} under $\bar{\mathbb{P}}$ and N^1 under $\bar{\mathbb{Q}}^{\hat{\varphi}}$ would have the same distribution, and the costs $L_T(\varphi)$ and $L_T(\hat{\varphi})$ would obviously be the same. The definition on subsequent intervals will depend on maps $T_i : \bar{\mathbb{M}}_1 \times \cdots \times \bar{\mathbb{M}}_i \to \bar{\mathbb{M}}_i$ for $i = 1, \ldots, \ell$, which must also be defined recursively.

Observe that under $\bar{\mathbb{P}}$, $\bar{N}_1$ has intensity $ds \times \nu(dx) \times dr$. Under $\bar{\mathbb{Q}}^{\hat{\varphi}}$, regardless of the definition of $\hat{\varphi}$ on later intervals, $\bar{N}_1$ has intensity

$$ds \times \nu(dx) \times \left[\sum_{j=1}^{n_1} \xi_{1j} 1_{E_{1j}}(x) 1_{(0,1]}(r) + 1_{(1,\infty)}(r) \right] dr.$$

The task of T_1 is to "undo" the effect of the change of measure, so that under $\bar{\mathbb{Q}}^{\hat{\varphi}}$, $\hat{N}_1 = T_1\left[\bar{N}_1\right]$ has intensity $ds \times \nu(dx) \times dr$. For $\bar{m}_1 \in \bar{\mathbb{M}}_1$, $\hat{m}_1 = T_1[\bar{m}_1]$ is defined as follows: for all $j \in \{1, \ldots, n_1\}$ and Borel subsets $A \subset \mathbb{I}_1$, $B \subset E_{1j}$, $C_1 \subset [0, \xi_{1j}]$ and $C_2 \subset (\xi_{1j}, \infty)$,

$$\hat{m}_1\left(A \times B \times [C_1 \cup C_2]\right) = \bar{m}_1\left(A \times B \times \left[\frac{1}{\xi_{1j}} C_1 \cup \left(C_2 - \xi_{1j} + 1\right)\right]\right).$$

The mapping T_1 can thus be viewed as a transformation on the underlying space $\mathscr{Y}_1$ on which m_1 is defined. An equivalent characterization of $\hat{m}_1 = T_1(\bar{m}_1)$ that will be used below is that $\hat{m}_1$ is the unique measure that for all nonnegative $\psi \in M_b(\mathscr{Y}_1)$ satisfies

$$\int_{\mathscr{Y}_1} \psi(s, x, r) \hat{m}_1(ds \times dx \times dr) = \sum_{j=1}^{n_1} \int_{\mathscr{Y}_1} 1_{E_{1j}}(x) \left[\psi(s, x, \xi_{1j} r) 1_{(0,1]}(r) \right.$$
$$\left. + \ \psi(s, x, r + \xi_{1j} - 1) 1_{(1,\infty)}(r)\right] \bar{m}_1(ds \times dx \times dr).$$

With T_1 in hand, the definition of $\hat{\varphi}(s, x, \bar{m})$ for $s \in \mathbb{I}_2$ is straightforward. Indeed, since $\hat{N}_1$ has the same distribution under $\bar{\mathbb{Q}}^{\hat{\varphi}}$ that $\bar{N}_1$ has under $\bar{\mathbb{P}}$, and since each ξ_{2j} is a function only of $\bar{N}_1$, with the definition $\hat{X}_{2j} = \xi_{2j}(T_1[\bar{N}_1]) = \xi_{2j}(\hat{N}_1)$, $\hat{X}_{2j}$ under $\bar{\mathbb{Q}}^{\hat{\varphi}}$ has the same distribution as X_{2j} under $\bar{\mathbb{P}}$. We now define $\hat{\varphi}$ on $\mathbb{I}_1 \cup \mathbb{I}_2$ as in (8.23) but with X_{2j} replaced by $\hat{X}_{2j}$. Then $\left\{\hat{\varphi}(s, x, \bar{m}), s \in \mathbb{I}_1 \cup \mathbb{I}_2, x \in \mathscr{X}\right\}$ under $\bar{\mathbb{Q}}^{\hat{\varphi}}$ has the same distribution as $\{\varphi(s, x, \bar{m}), s \in \mathbb{I}_1 \cup \mathbb{I}_2, x \in \mathscr{X}\}$ under $\bar{\mathbb{P}}$.

We now proceed recursively, and having defined $T_1, \ldots, T_{p-1}$ for some $1 < p \leq \ell$, we define T_p by $T_p(\bar{m}_1, \ldots, \bar{m}_p) = \hat{m}_p$, where $\hat{m}_p$ is the unique measure satisfying, for all nonnegative $\psi \in M_b(\mathscr{Y}_p)$,

$$\int_{\mathscr{Y}_p} \psi(s, x, r)\hat{m}_p(ds \times dx \times dr) = \sum_{j=1}^{n_p} \int_{\mathscr{Y}_p} 1_{E_{pj}}(x) \Big[\psi(s, x, \hat{\xi}_{pj} r) 1_{(0,1]}(r) + \psi(s, x, r + \hat{\xi}_{pj} - 1) 1_{(1,\infty)}(r)\Big] \bar{m}_p(ds \times dx \times dr),$$

where $\hat{\xi}_p = \xi_p(\hat{m}_1, \ldots, \hat{m}_{p-1})$ and $\hat{m}_i = T_i(\bar{m}_1, \ldots, \bar{m}_i)$. We define the transformation $T : \bar{\mathbb{M}} \to \bar{\mathbb{M}}$ by

$$T(\bar{m}) = \varpi \left(T_1(\bar{N}_1(\bar{m})), \ldots, T_\ell(\bar{N}_1(\bar{m}), \ldots, \bar{N}_\ell(\bar{m}))\right),$$

and define $\hat{\varphi} \in \mathscr{A}_s$ for all times s by replacing X_{ij} with $\hat{X}_{ij}$ in the right side of (8.23), where

$$\hat{X}_i(\bar{m}) = X_i(T(\bar{m})) = \xi_i(T_1(\bar{N}_1(\bar{m})), \ldots, T_i(\bar{N}_1(\bar{m}), \ldots, \bar{N}_i(\bar{m}))). \tag{8.38}$$

Denoting $T(\bar{N})$ by $\hat{N}$, we see that for ϑ in the class $\hat{A}_b$ defined above Theorem 8.15,

$$\int \vartheta(s, x, r)\hat{N}(ds \times dx \times dr) = \int \big[\vartheta(s, x, \hat{\varphi}(s, x)r) 1_{(0,1]}(r) + \vartheta(s, x, r + \hat{\varphi}(s, x) - 1) 1_{(1,\infty)}(r)\big] \bar{N}(ds \times dx \times dr). \tag{8.39}$$

Also, let $h_\varphi : \bar{\mathbb{M}} \to \mathbb{M}$ be defined by

$$h_\varphi(\bar{m})(A \times B) \doteq \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \bar{m}((A \cap \mathbb{I}_i) \times (B \cap E_{ij}) \times [0, X_{ij}(\bar{m})])$$

for $A \times B \in \mathscr{B}(\mathscr{X}_T)$. Recall that L_T was defined in (8.17). In order to complete the proof of the lemma, we will prove the following:

(a) the distribution of $\hat{N} = T(\bar{N})$ under $\bar{\mathbb{Q}}^{\hat{\varphi}}$ is the same as that of $\bar{N}$ under $\bar{\mathbb{P}}$;

(b) $h_\varphi(\bar{N}) = N^\varphi$ and $h_\varphi(T(\bar{N})) = h_\varphi(\hat{N}) = N^1$;

(c) for some measurable map $\Theta : \bar{\mathbb{M}} \to [0, \infty)$, $L_T(\varphi) = \Theta(\bar{N})$ and $L_T(\hat{\varphi}) = \Theta(T(\bar{N}))$, a.s. $\bar{\mathbb{P}}$.

Item (c) is an immediate consequence of the definition of $\hat{\varphi}$ via (8.38). We next consider (b). Noting that $\bar{N}(\bar{m}) = \bar{m}$, suppressing $\bar{m}$ in notation, and recalling the form of φ in (8.23), we have for $A \times B \in \mathscr{B}(\mathscr{X}_T)$,

$$\begin{aligned}
h_\varphi(\bar{N})(A \times B) &= \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \bar{N}((A \cap \mathbb{I}_i) \times (B \cap E_{ij}) \times [0, X_{ij}]) \\
&= \sum_{i=1}^{\ell} \int_{(t_{i-1},t_i] \times \mathscr{X} \times [0,\infty)} 1_{A\times B}(s,x) 1_{[0,\varphi(s,x)]}(r) \bar{N}(ds \times dx \times dr) \\
&= \int_{\mathscr{Y}_T} 1_{A\times B}(s,x) 1_{[0,\varphi(s,x)]}(r) \bar{N}(ds \times dx \times dr) \\
&= N^\varphi(A \times B).
\end{aligned}$$

This proves the first statement in (b). Next, using (8.38), (8.39), and that $r > 1$ implies $r + \hat{\varphi}(s,x) - 1 > \hat{\varphi}(s,x)$, we obtain

$$\begin{aligned}
h_\varphi(T(\bar{N}))(A \times B) &= \int 1_{A\times B}(s,x) 1_{[0,\hat{\varphi}(s,x)]}(r) \hat{N}(ds \times dx \times dr) \\
&= \int 1_{A\times B}(s,x) [1_{[0,\hat{\varphi}(s,x)]}(\hat{\varphi}(s,x) r) 1_{[0,1]}(r) \\
&\quad + 1_{[0,\hat{\varphi}(s,x)]}(r + \hat{\varphi}(s,x) - 1) 1_{(1,\infty)}(r)] \bar{N}(ds \times dx \times dr) \\
&= \int 1_{A\times B}(s,x) 1_{[0,1]}(r) \bar{N}(ds \times dx \times dr) \\
&= N^1(A \times B).
\end{aligned}$$

This proves the second statement in (b). Lastly, we prove (a). It suffices to show that for every $\vartheta \in \hat{A}_b$,

$$\bar{E}^{\hat{\varphi}} \int \vartheta(s,x,r) \hat{N}(ds \times dx \times dr) = \bar{E}^{\hat{\varphi}} \int \vartheta(s,x,r) \bar{\nu}_T(ds \times dx \times dr).$$

Using (8.39) and the last part of Theorem 8.15 for the first equality and that the marginal of $\bar{\nu}_T(ds \times dx \times dr)$ in r is Lebesgue measure, we have

$$\begin{aligned}
&\bar{E}^{\hat{\varphi}} \int \vartheta(s,x,r) \hat{N}(ds \times dx \times dr) \\
&= \bar{E}^{\hat{\varphi}} \int \big[\vartheta(s,x,\hat{\varphi}(s,x) r) \hat{\varphi}(s,x) 1_{(0,1]}(r) \\
&\quad + \vartheta(s,x,r + \hat{\varphi}(s,x) - 1) 1_{(1,\infty)}(r)\big] \bar{\nu}_T(ds \times dx \times dr) \\
&= \bar{E}^{\hat{\varphi}} \int \big[\vartheta(s,x,r) 1_{(0,\hat{\varphi}(s,x)]}(r) \\
&\quad + \vartheta(s,x,r) 1_{(\hat{\varphi}(s,x),\infty)}(r)\big] \bar{\nu}_T(ds \times dx \times dr)
\end{aligned}$$

$$= \bar{E}^{\hat{\varphi}} \int \vartheta(s, x, r) \bar{\nu}_T(ds \times dx \times dr),$$

which proves (a), and completes the proof of the lemma. □

8.3 Representation for Functionals of PRM and Brownian Motion

In this section we state the representation for functionals of both a PRM and a Hilbert space valued Wiener process. In Chap. 11 we will show how representations for a Hilbert space valued Brownian motion can be translated into representations for related objects, such as a collection of infinitely many independent scalar Brownian motions and the Brownian sheet. The analogous conversions can also be done when one is considering a PRM along with a Hilbert space valued Brownian motion.

The independent processes we consider are thus a Λ-Wiener process as in Definition 8.1 and a PRM as in Definition 8.11. In the proof of such a result we would follow the same procedure as in the separate cases and consider first the canonical space and filtration and then generalize. However, in this instance we skip the proof because it is simply a combination of the arguments used for the two separate cases, and we instead present a result that holds for a general filtration on a general probability space.

Thus we let $(\Omega, \mathscr{F}, P)$ be a probability space with a filtration $\{\mathscr{F}_t\}_{0\le t\le T}$ satisfying the usual conditions, and assume that $(\Omega, \mathscr{F}, P)$ supports all the following processes. Let W be a Λ-Wiener process with respect to $\{\mathscr{F}_t\}$. Let ν be a σ-finite measure on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ and let $\bar{N}$ be a PRM, with respect to $\{\mathscr{F}_t\}$, on $\mathscr{Y}_T \doteq [0, T] \times \mathscr{Y}$, where $\mathscr{Y} \doteq \mathscr{X} \times [0, \infty)$, with intensity measure $\bar{\nu}_T \doteq \lambda_T \times \nu \times \lambda_\infty$. We assume that for all $0 \le s \le t < \infty$, $(\bar{N}((s, t] \times \cdot), W(t) - W(s))$ is independent of $\mathscr{F}_s$. Let $\mathscr{PF}$ be the predictable σ-field on $[0, T] \times \Omega$. Let $\bar{\mathscr{A}}^W$ and $\bar{\mathscr{A}}_b^W$ be the collections of controls for the Wiener process defined as $\bar{\mathscr{A}}$ was below Definition 8.2 and $\bar{\mathscr{A}}_b$ was below (8.2) respectively, and let $\bar{\mathscr{A}}^N$, $\bar{\mathscr{A}}_b^N$ be controls for the PRM defined as $\bar{\mathscr{A}}$ and $\bar{\mathscr{A}}_b$ were below (8.19). The classes $\bar{\mathscr{A}}_{b,M}^N$ and $\bar{\mathscr{A}}_{b,M}^W$, which give uniform (in ω) bounds, are defined as they were in (8.18) and (8.2). For each $\varphi \in \bar{\mathscr{A}}^N$, N^φ will be a counting process on $\mathscr{X}_T$ defined as in (8.16) with φ as its controlled intensity measure.

Let $\bar{\mathscr{A}}_{b,M} \doteq \bar{\mathscr{A}}_{b,M}^W \times \bar{\mathscr{A}}_{b,M}^N$, $\bar{\mathscr{A}}_b \doteq \bar{\mathscr{A}}_b^W \times \bar{\mathscr{A}}_b^N$, and $\bar{\mathscr{A}} \doteq \bar{\mathscr{A}}^W \times \bar{\mathscr{A}}^N$. For $\psi \in \bar{\mathscr{A}}^W$, define $L_T^W(\psi) \doteq \frac{1}{2}\int_0^T \|\psi(s)\|_0^2\, ds$, with the norm $\|\cdot\|_0$ as in Sect. 8.1.1. For $\varphi \in \bar{\mathscr{A}}^N$, define $L_T^N(\varphi) \doteq \int_{\mathscr{X}_T} \ell(\varphi(t, x))\, \nu_T(dt \times dx)$, and for $u = (\psi, \varphi) \in \bar{\mathscr{A}}$, set $\bar{L}_T(u) \doteq L_T^N(\varphi) + L_T^W(\psi)$. For $\psi \in \bar{\mathscr{A}}^W$, let W^ψ be defined by $W^\psi(t) = W(t) + \int_0^t \psi(s)ds$, $t \in [0, T]$. We recall the definition of the space of measures $\mathbb{M} = \Sigma(\mathscr{X}_T)$ from Sect. 8.2.1 and its associated topology. With these definitions, the following representation holds. The proof of the second part of the theorem is similar to the proofs of Theorems 8.4 and 8.13.

Theorem 8.19 *Let $G \in M_b(\mathscr{C}([0,T] : \mathscr{H}) \times \mathbb{M})$. Then for $\theta \in (0, \infty)$,*

$$-\log E \exp\{-G(W, N^\theta)\} = \inf_{u=(\psi,\varphi)\in\mathscr{R}} E\left[\theta \bar{L}_T(u) + G(W^{\sqrt{\theta}\psi}, N^{\theta\varphi})\right],$$

where $\mathscr{R}$ can be either $\bar{\mathscr{A}}_b$ or $\bar{\mathscr{A}}$. Furthermore, for every $\delta > 0$, there exists $M < \infty$ depending on $\|G\|_\infty$ and δ such that for all $\varepsilon \in (0, 1)$,

$$\begin{aligned} &-\varepsilon \log E \exp\left\{-\frac{1}{\varepsilon} G(\sqrt{\varepsilon} W, \varepsilon N^{1/\varepsilon})\right\} \\ &\quad \geq \inf_{u=(\psi,\varphi)\in\bar{\mathscr{A}}_{b,M}} E\left[\bar{L}_T(u) + G(\sqrt{\varepsilon} W^{\psi/\sqrt{\varepsilon}}, \varepsilon N^{\varphi/\varepsilon})\right] - \delta. \end{aligned}$$

8.4 Notes

For basic results on Hilbert space valued Brownian motions see [69], and for Poisson random measures see [159, 161].

The representation for a finite dimensional Brownian motion first appeared in [32]. In Sect. 3.2, we saw how this representation allowed a straightforward large deviation analysis of small noise diffusions using weak convergence arguments. Other applications of the finite dimensional case include large deviation analysis of small noise diffusions with discontinuous statistics [33] and homogenization [111], and also the analysis of importance sampling for accelerating Monte Carlo in estimating rare events [112].

With respect to its application to large deviation analysis, the representation is convenient because it eliminates the need for superexponentially close approximation and exponential tightness results used by other methods. A special case of the representation, rediscovered by Borell in [31], has found use in proving various functional inequalities, as in [184].

While convenient in the finite dimensional setting, the representation for functionals of Brownian motion and associated weak convergence methods are even more important for processes with an infinite dimensional state, where the proof of approximation and tightness results can be very demanding, and which often require assumptions beyond those needed for the large deviation result itself to be valid. Representations for infinite dimensional problems first appeared in [39] for the case of infinite dimensional Brownian motion, and in [45] for the case of Poisson random measures. The proof given here differs substantially from those of [39, 45], in particular in that they use classical-sense solutions to dynamic programming equations to establish the first step in the proof of the lower bound. As noted in the overview of Part III of the book, other authors have made numerous and varied applications of these representations and the associated abstract large deviation theorems that can be based on them. These abstract large deviation theorems are the topic of the next chapter.

A generalization of the representation that is sometimes useful (see, e.g., [11] for its use in a problem studying large deviations from a hydrodynamic limit) is that the infimum in the representation can be restricted to simple adapted processes [39]. Another extension is to the case in which the functional, in addition to depending on a BM and PRM, depends also on a $\mathscr{F}_0$-valued random variable, such as an initial condition (see [11, 46]).

A somewhat different variational representation for functionals of a PRM is presented in [269]. This representation is given in terms of some predictable transformations on the canonical Poisson space whose existence relies on solvability of certain nonlinear partial differential equations from the theory of mass transportation. This imposes restrictive conditions on the intensity measure (e.g., absolute continuity with respect to Lebesgue measure) of the PRM, and in particular, a standard Poisson process is not covered. The use of such a representation for proving large deviation results for general continuous time models with jumps appears to be unclear.