The Annals of Probability
1998, Vol. 26, No. 4, 1641–1659

A VARIATIONAL REPRESENTATION FOR CERTAIN FUNCTIONALS OF BROWNIAN MOTION[1]

BY MICHELLE BOUÉ AND PAUL DUPUIS

University of Massachusetts and Brown University

In this paper we show that the variational representation

$$-\log Ee^{-f(W)} = \inf_v E\left\{\frac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(W + \int_0^{\cdot} v_s\, ds\right)\right\}$$

holds, where W is a standard d-dimensional Brownian motion, f is any bounded measurable function that maps $C([0,1]: \mathbb{R}^d)$ into $\mathbb{R}$ and the infimum is over all processes v that are progressively measurable with respect to the augmentation of the filtration generated by W. An application is made to a problem concerned with large deviations, and an extension to unbounded functions is given.

1. Introduction. In this paper we prove the following variational representation formula. Let W be a standard d-dimensional Brownian motion. Then for functions f: $C([0,1]: \mathbb{R}^d) \to \mathbb{R}$ that are bounded and measurable,

$$-\log Ee^{-f(W)} = \inf_v E\left\{\frac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(W + \int_0^{\cdot} v_s\, ds\right)\right\}.$$

In this equation E denotes expectation with respect to the probability space on which the Brownian motion is defined, and the infimum is over all processes v that are progressively measurable with respect to the augmentation of the filtration generated by the Brownian motion. (The definitions of progressive measurability and of the augmented filtration are recalled in the next section.)

Our main interest in this representation is due to its usefulness in deriving various asymptotic results of a large deviations nature. As is well known, large deviation theory allows one to characterize the behavior of certain exponential functionals (which include probabilities as a special case) as various parameters tend to their limits. The asymptotic behavior is shown to scale exponentially in the parameter, and usually with the coefficient C that multiplies the parameter represented as the solution to a variational problem.

The representation stated above can be used to prove such results whenever the functional of interest can be expressed as a measurable functional of a Brownian motion. Examples are integrals of small noise diffusions, inte-

---

Received February 1997; revised March 1998.
[1]Supported in part by NSF Grant DMS-94-03820 and by Consejo Nacional de Ciencia y Tecnología, Mexico.
AMS 1991 *subject classifications*. Primary 60H99; secondary 60J60, 60J65, 60F10.
Key words and phrases. Variational representation, Brownian motion, large deviations, relative entropy.

grals against the occupation measure for reflected diffusions and so on. All that is required to prove such a result, once one has such a representation in hand, is the convergence of the variational problems appearing in the representation to the limit variational problem that defines C. This can often be carried out efficiently using weak convergence methods [2], and an example that shows how easily one can obtain standard large deviation results is provided in Section 4. A large number of examples involving various classes of processes in the discrete time setting are given in [9].

While the representation formula provides an elegant method for deriving known results under weak assumptions, the primary motivation for its development comes from a desire to analyze problems for which the standard discretization methods of large deviation are awkward. An illustrative example is the derivation of large deviation properties of a small noise diffusion with discontinuous drift [5]. Standard methods for treating such problems generally rely on discretizing time. This can cause a number of difficulties, since it is difficult to approximate the continuous time process accurately by a discrete time analogue in a neighborhood of the discontinuity. The approach based on the representation given above bypasses the time discretization step and thus avoids this difficulty.

A second class of problems where standard large deviation methods are not convenient involves the asymptotic analysis of risk-sensitive stochastic control problems. Such problems are currently of significant interest (see [1], [11], [14], [17], and the references therein). The presence of a progressively measurable control process makes the burden of detail that results from discretization rather onerous. Such problems were in fact the main motivation for the development of the representation, and applications to such problems will appear elsewhere [4].

The connection between exponential functionals and variational representations has been exploited in a number of contexts. The first results in this area appear to be due to Fleming [10], who considered functionals of nondegenerate diffusions that satisfied certain parabolic partial differential equations and applied a representation to study certain large deviation problems. Extensions to more general classes of processes were given by Sheu [15]. More recently, Dupuis and Ellis [9] have obtained representations under very weak assumptions in discrete time and have applied these representations to study a variety of large deviation problems. Representations for continuous time Markov chains were developed in [8] and applied to problems of large deviations for queueing networks. Also, see [6] for representations for certain jump processes.

A summary of the paper is as follows. The second section is devoted to preliminary results and notation. In the third section we state and prove the representation for bounded functionals of Brownian motion on bounded time intervals. In Section 4 we specialize the result to the case where f is of the form $g(X)$, with g a bounded measurable functional of X, and where X is the sample path of a strong solution to a stochastic differential equation driven by W. We also show how one can derive the standard large deviation asymptotics for small noise diffusions using the representation. In the last section we show that the representation continues to hold if f is bounded from above. This extension is especially useful for problems from risk-sensitive control [3].

2. **Preliminaries.** Throughout the sequel we shall consider the canonical probability space $(\Omega, \mathcal{F}, \theta)$, where $\Omega = \mathcal{C}([0,1]: \mathbb{R}^d)$, $\mathcal{F} = \mathcal{B}(\mathcal{C}([0,1]: \mathbb{R}^d))$ is the Borel σ-algebra and θ is d-dimensional Wiener measure. Under θ the coordinate mapping process $W = \{W_t(\omega) \doteq \omega(t), 0 \le t \le 1\}$ together with the filtration $\{\mathcal{F}_t^W\} \doteq \{\sigma(W_s; 0 \le s \le t)\}$ is a d-dimensional Brownian motion starting at the origin. We construct the *augmented filtration* $\{\mathcal{F}_t\}$ by considering the collection of null sets $\mathcal{N} \doteq \{N \subseteq \mathcal{C}([0,1]: \mathbb{R}^d): \exists\ B \in \mathcal{B}(\mathcal{C}([0,1]: \mathbb{R}^d))$ with $N \subseteq B$ and $\theta(B) = 0\}$ and defining

$$\mathcal{F}_t \doteq \sigma\left(\mathcal{F}_t^W \cup \mathcal{N}\right), \qquad 0 \le t \le 1.$$

Relative to the continuous filtration $\{\mathcal{F}_t\}$, W is still a d-dimensional Brownian motion.

REMARK. For notational simplicity, we consider only the time interval $[0,1]$. All the results in this paper carry over with only notational changes to an arbitrary interval $[0,T]$ with $T < \infty$.

We begin by introducing some basic definitions and notation.

DEFINITION 2.1. A stochastic process X on $(\Omega, \mathcal{F})$ is *progressively measurable* with respect to the filtration $\{\mathcal{F}_t\}$ if for every t the map $(s, \omega) \mapsto X_s(\omega)$: $([0,t] \times \Omega, \mathcal{B}([0,t]) \otimes \mathcal{F}_t) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is measurable.

DEFINITION 2.2. We denote by $\mathcal{A}$ the set of all vectors of $\mathcal{F}_t$-progressively measurable processes $v = \{v_t = (v_t^{(1)}, \dots, v_t^{(d)}); 0 \le t \le 1\}$ which satisfy

$$\int_{\mathcal{C}([0,1]: \mathbb{R}^d)} \left[\int_0^1 \left(v_t^{(i)}\right)^2 dt\right] d\theta < \infty, \qquad 1 \le i \le d.$$

Further, we denote by $\mathcal{A}_b$ the set of all vectors of bounded $\mathcal{F}_t$-progressively measurable processes, in the sense that there exists $M < \infty$ such that $\|v_t\| \le M$ for all $t \in [0,1]$ with probability 1.

DEFINITION 2.3. A stochastic process X on $(\Omega, \mathcal{F})$ is *simple* if there exist $C < \infty$, a strictly increasing sequence of real numbers $\{t_i, i = 0, \dots, j\}$ with $t_0 = 0$ and $t_j = 1$, and a sequence of random variables $\{\xi_i, i = 0, \dots, j-1\}$ such that ξ_i is $\mathcal{F}_{t_i}$-measurable for every i, $\sup_{\omega \in \Omega} \max_{i \in \{0, \dots, j-1\}} \|\xi_i(\omega)\| \le C$ and

$$X_t(\omega) = \xi_0(\omega) 1_{\{0\}}(t) + \sum_{i=0}^{j-1} \xi_i(\omega) 1_{(t_i, t_{i+1}]}(t), \qquad 0 \le t \le 1,\ \omega \in \Omega.$$

The class of simple processes will be denoted by $\mathcal{A}_s$.

Note that simple processes are progressively measurable and bounded, and thus $A_s \subset A_b$.

DEFINITION 2.4. Let (X, F) be a measurable space, and let $P(X)$ denote the set of probability measures defined on it. For $\theta \in P(X)$, the *relative entropy function* $R(\cdot \| \theta)$ is the mapping from $P(X)$ into the extended real numbers given by

$$R(\gamma \| \theta) \doteq \int_X \left(\log \frac{d\gamma}{d\theta}(x) \right) \gamma(dx)$$

whenever $\gamma \in P(X)$ is absolutely continuous with respect to θ and $\log((d\gamma/d\theta)(x))$ is γ-integrable. In all other cases, we set $R(\gamma \| \theta) \doteq \infty$.

The following proposition is the starting point for the representation derived in Theorem 3.1. It states a variational formula involving the relative entropy function (see, e.g., [9], Proposition 2.4.2 for a proof).

PROPOSITION 2.5. *Let (X, F) be a measurable space, let f be a bounded measurable function mapping X into $\mathbb{R}$ and let θ be a probability measure on X. The following conclusions hold*:

(a) *We have the variational formula*

$$-\log \int_X e^{-f(x)} \theta(dx) = \inf_{\gamma \in P(X)} \left\{ R(\gamma \| \theta) + \int_X f(x) \gamma(dx) \right\}. \tag{1}$$

(b) *The infimum in* (1) *is uniquely attained at the probability measure γ_0 which is absolutely continuous with respect to θ and has Radon–Nikodym derivative*

$$\frac{d\gamma_0}{d\theta}(x) \doteq e^{-f(x)} \cdot \frac{1}{\int_X e^{-f(x)} \theta(dx)}.$$

We next state two approximation results which will be used in the sequel. The first result concerns measurable functions as approximated by continuous functions ([7], Theorem V.16a), and the second concerns progressively measurable processes as approximated by simple processes ([12], Lemma 3.2.4).

THEOREM 2.6. *Let (X, F, λ) be a probability space, with X a Polish space and F the associated Borel σ-algebra. Let f be a Borel-measurable function from this space into $\mathbb{R}$. There is a sequence of continuous functions $\{f_j, j \in \mathbb{N}\}$ from X into $\mathbb{R}$ with*

$$\lim_{j \to \infty} f_j = f, \qquad \lambda\text{-a.e.}$$

If the function f is bounded in absolute value by B, then all the approximating functions can be taken to be bounded by B as well.

PROPOSITION 2.7. *Let X be a bounded progressively measurable process, in the sense that $\|X_t(\omega)\| \leq B$ for all $\omega \in \Omega$ and $t \in [0,1]$. Then there exists a sequence $\{X^n,\ n \in \mathbb{N}\}$ of simple processes such that $\sup_{n\in\mathbb{N}} \|X_t^n(\omega)\| \leq B$ for every $\omega \in \Omega$ and $t \in [0,1]$, and also*

$$\lim_{n\to\infty} E\int_0^1 \|X_s^n - X_s\|^2\, ds = 0.$$

The final result in this section will be needed to justify interchanges of limits and expectations in several places. It is related to the fact that the level sets of $R(\cdot \| \theta)$ are compact in the τ-topology.

LEMMA 2.8. *Let (X, F) be a measurable space, with X a Polish space and F the associated Borel σ-algebra. Let θ be a probability measure defined on it and let $f\colon X \to \mathbb{R}$ be a bounded Borel-measurable function. Consider a sequence $\{\mu_n,\ n \in \mathbb{N}\}$ of measures in $\mathcal{P}(X)$ satisfying $\sup_{n\in\mathbb{N}} R(\mu_n \| \theta) \leq \alpha < \infty$. Assume μ_n converges weakly to a probability measure μ. Then the following hold*:

(a) $\lim_{n\to\infty} \int_X f\, d\mu_n = \int_X f\, d\mu$;

(b) *if $\{f_n,\ n \in \mathbb{N}\}$ is a sequence of uniformly bounded functions converging θ-a.s. to f, then*

$$\lim_{n\to\infty} \int_X f_n\, d\mu_n = \int_X f\, d\mu.$$

PROOF. As a first step we verify that the limit measure μ is absolutely continuous with respect to θ. Indeed, by the weak convergence of μ_n to μ and the lower semicontinuity of $R(\cdot \| \theta)$ (see, e.g., [9], Lemma 2.4.3),

$$R(\mu \| \theta) \leq \liminf_{n\to\infty} R(\mu_n \| \theta) \leq \alpha < \infty.$$

From the definition of relative entropy, this implies that μ is absolutely continuous with respect to θ. Theorem 2.6 enables us to find a sequence $\{\tilde{f}_j,\ j \in \mathbb{N}\}$ of bounded and continuous functions such that $\lim_{j\to\infty} \tilde{f}_j = f$, θ-a.e. Since $\mu \ll \theta$ the limit also holds μ-a.e. By the dominated convergence theorem $\int_X \tilde{f}_j\, d\mu$ converges to $\int_X f\, d\mu$. For each fixed $j \in \mathbb{N}$, $\int_X \tilde{f}_j\, d\mu_n$ converges to $\int_X \tilde{f}_j\, d\mu$ because of the weak convergence of μ_n to μ. Hence to prove part (a) of the lemma it only remains to verify

$$\lim_{j\to\infty} \sup_{n\in\mathbb{N}} \int_X |\tilde{f}_j - f|\, d\mu_n = 0. \tag{2}$$

Fix $\varepsilon > 0$. Define $\|f\|_\infty \doteq \sup_{x\in X} |f(x)|$ and let M be such that $\|f\|_\infty \leq M$ and $\sup_{j\in\mathbb{N}} \|\tilde{f}_j\|_\infty \leq M$. Then

$$\begin{aligned}
\int_X |\tilde{f}_j - f|\, d\mu_n &= \int_{\{|\tilde{f}_j - f| > \varepsilon\}} |\tilde{f}_j - f|\, d\mu_n + \int_{\{|\tilde{f}_j - f| \leq \varepsilon\}} |\tilde{f}_j - f|\, d\mu_n \\
&\leq \int_{\{|\tilde{f}_j - f| > \varepsilon\}} |\tilde{f}_j - f|\, d\mu_n + \varepsilon \\
&\leq 2M\mu_n\left\{|\tilde{f}_j - f| > \varepsilon\right\} + \varepsilon.
\end{aligned}$$

Since ε is arbitrary, (2) will follow if we can show that

$$\lim_{j\to\infty} \sup_{n\in\mathbb{N}} \mu_n\Big\{|\tilde{f}_j - f| > \varepsilon\Big\} = 0.$$

For any $c \in (1,\infty)$ we have

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \mu_n\Big\{|\tilde{f}_j - f| > \varepsilon\Big\} &= \sup_{n\in\mathbb{N}} \int_{\{|\tilde{f}_j - f| > \varepsilon\}} \frac{d\mu_n}{d\theta}\, d\theta \\
&\le \sup_{n\in\mathbb{N}} \int_{\{|\tilde{f}_j - f| > \varepsilon\}\cap\{d\mu_n/d\theta \le c\}} \frac{d\mu_n}{d\theta}\, d\theta \\
&\quad + \sup_{n\in\mathbb{N}} \int_{\{|\tilde{f}_j - f| > \varepsilon\}\cap\{d\mu_n/d\theta > c\}} \frac{d\mu_n}{d\theta}\, d\theta.
\end{aligned}$$

The first term in the last expression converges to 0 as $j \to \infty$ since it is bounded above by $c\theta\{|\tilde{f}_j - f| > \varepsilon\}$. Convergence of the second term to 0 follows from the upper bound

$$\begin{aligned}
&\sup_{n\in\mathbb{N}} \int_{\{|\tilde{f}_j - f| > \varepsilon\}\cap\{d\mu_n/d\theta > c\}} \frac{d\mu_n}{d\theta}\, d\theta \\
&\quad\le \frac{1}{\log c} \sup_{n\in\mathbb{N}} \int_{\{d\mu_n/d\theta > c\}} \frac{d\mu_n}{d\theta} \log \frac{d\mu_n}{d\theta}\, d\theta \\
&\quad\le \frac{1}{\log c} \sup_{n\in\mathbb{N}} \int \left(e^{-1} + \frac{d\mu_n}{d\theta} \log \frac{d\mu_n}{d\theta}\right) d\theta \\
&\quad= \frac{1}{\log c}\left(e^{-1} + \sup_{n\in\mathbb{N}} R(\mu_n \,\|\, \theta)\right) \\
&\quad\le \frac{e^{-1} + \alpha}{\log c}.
\end{aligned}$$

Since c can be taken arbitrarily large, (2) is established and the proof of (a) is complete.

To prove (b), observe that we may write

$$\int_X f_n\, d\mu_n = \int_X f\, d\mu_n + \left\{ \int_X f_n\, d\mu_n - \int_X f\, d\mu_n \right\}.$$

Since convergence of the first term to $\int_X f\, d\mu$ is a direct consequence of (a), we only need to show convergence of the second term to 0. This is done just as in the proof of (2) but replacing $\tilde{f}_j$ by f_n [note that the proof of (2) does not use continuity of $\tilde{f}_j$]. □

3. The main theorem. We are interested in a representation formula for the quantity

$$-\log E e^{-f(W)},$$

where f: $C([0,1]: \mathbb{R}^d) \to \mathbb{R}$ is a bounded Borel-measurable function. This formula will be stated in the next theorem. The boundedness assumption is stronger than needed, but it is included here to simplify the proof. An extension will be given in Section 5.

THEOREM 3.1. *Let f be a bounded Borel-measurable function mapping $C([0,1]: \mathbb{R}^d)$ into $\mathbb{R}$. Then*

$$-\log Ee^{-f(W)} = \inf_{v \in A} E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\,ds + f\left(W + \int_0^{\cdot} v_s\,ds\right)\right\}.$$

PROOF OF THE UPPER BOUND. Consider any $v \in A_b$. Since v is bounded, for each $1 \le i \le d$ the stochastic integral $\int_0^t v_s^{(i)}\,dW_s^{(i)}$ is well defined and is a square integrable martingale. If we define R_t by

$$R_t = \exp\left[\sum_{i=1}^d \int_0^t v_s^{(i)}\,dW_s^{(i)} - \tfrac{1}{2}\int_0^t \|v_s\|^2\,ds\right], \tag{3}$$

then R_t is a martingale. We define a probability measure γ_v on F_1 by

$$\gamma_v(A) \doteq \int_A R_1\,d\theta \quad \text{for } A \in F_1. \tag{4}$$

By Girsanov's theorem the process $\tilde{W} = \{\tilde{W}_t = (\tilde{W}_t^{(1)}, \ldots, \tilde{W}_t^{(d)});\ 0 \le t \le 1\}$ given by

$$\tilde{W}_t = W_t - \int_0^t v_s\,ds \tag{5}$$

is a d-dimensional Brownian motion under γ_v. Let T_v be the map defined by

$$T_v(\phi)_t \doteq \phi_t - \int_0^t v_s(\phi)\,ds.$$

Then, for any Borel set $A \subset C([0,1]: \mathbb{R}^d)$, $\theta(A) = \gamma_v(T_v^{-1}(A))$.

Using the definition of $R(\gamma_v \,\|\, \theta)$ and substituting (3) and (4), we obtain

$$\begin{aligned} R(\gamma_v \,\|\, \theta) &= \int \left(\log \frac{d\gamma_v}{d\theta}(\phi)\right)\gamma_v(d\phi) \\ &= \int \left\{\sum_{i=1}^d \int_0^1 v_s^{(i)}\,dW_s^{(i)} - \frac{1}{2}\int_0^1 \|v_s\|^2\,ds\right\}\gamma_v(d\phi). \end{aligned}$$

Thanks to (5),

$$\begin{aligned} R(\gamma_v \,\|\, \theta) &= E^v\left\{\sum_{i=1}^d \int_0^1 v_s^{(i)}\,d\tilde{W}_s^{(i)} + \sum_{i=1}^d \int_0^1 \left(v_s^{(i)}\right)^2 ds - \tfrac{1}{2}\int_0^1 \|v_s\|^2\,ds\right\} \\ &= E^v\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\,ds\right\}, \end{aligned} \tag{6}$$

where the last equality uses the martingale property of the stochastic integral and E^v denotes expectation with respect to the probability measure γ_v.

Consequently,

$$R(\gamma_v \parallel \theta) + \int f(\phi)\gamma_v(d\phi) = E^v\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(\widetilde{W} + \int_0^{\cdot} v_s\, ds\right)\right\}$$

and from (1) we obtain

$$-\log Ee^{-f(W)} \le \inf_{v\in A_b} E^v\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(\widetilde{W} + \int_0^{\cdot} v_s\, ds\right)\right\}. \tag{7}$$

We now use (7) to show that, for any $v \in A$,

$$-\log Ee^{-f(W)} \le E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(W + \int_0^{\cdot} v_s\, ds\right)\right\}, \tag{8}$$

where expectation is with respect to Wiener measure θ. The proof proceeds in three steps.

Step 1. *Simple v.* Suppose v can be written as

$$v_t(\phi) = \xi_0(\phi)1_{\{0\}}(t) + \sum_{i=0}^{j-1} \xi_i(\phi)1_{(t_i, t_{i+1}]}(t), \quad 0 \le t \le 1,\ \phi \in C([0,1]: \mathbb{R}^d),$$

where ξ_i is F_{t_i}-measurable for $i = 0, \dots, j-1$. We define a new family of random variables $\{\tilde{\xi}_i,\ i = 0, \dots, j-1\}$ as follows:

$$\tilde{\xi}_0(\phi) \doteq \xi_0(\phi),$$

and for $i = 1, \dots, j-1$,

$$\tilde{\xi}_i(\phi) \doteq \xi_i(\psi_i),$$

where ψ_i is any function which agrees with

$$\phi - \sum_{k=0}^{i-1} \tilde{\xi}_k(\phi)(t_{k+1} - t_k)$$

up to time t_i. Each $\tilde{\xi}_i$ is F_{t_i}-measurable, so the process $\tilde{v}$ given by

$$\tilde{v}_t(\phi) \doteq \tilde{\xi}_0(\phi)1_{\{0\}}(t) + \sum_{i=0}^{j-1} \tilde{\xi}_i(\phi)1_{(t_i, t_{i+1}]}(t), \quad 0 \le t \le 1,\ \phi \in C([0,1]: \mathbb{R}^d),$$

is simple, and thus an element of A_b. Moreover, $\tilde{v}$ satisfies $\tilde{v}(\phi) = v(T_{\tilde{v}}(\phi))$ with probability 1. This relation implies that, for $\widetilde{W}(\phi) \doteq W(\phi) - \int_0^{\cdot} \tilde{v}_s(\phi)\, ds$ and $A \in B(C([0,1]: \mathbb{R}^d))$, $B \in B(L^2([0,1]: \mathbb{R}^d))$,

$$\begin{aligned}
\gamma_{\tilde{v}}(\widetilde{W} \in A,\ \tilde{v} \in B) &= \gamma_{\tilde{v}}\left(\left\{\phi\colon \phi - \int_0^{\cdot} \tilde{v}_s(\phi)\, ds \in A,\ \tilde{v}(\phi) \in B\right\}\right) \\
&= \gamma_{\tilde{v}}(\{\phi\colon T_{\tilde{v}}(\phi) \in A,\ v(T_{\tilde{v}}(\phi)) \in B\}) \\
&= \theta(\{\psi\colon \psi \in A,\ v(\psi) \in B\}) \\
&= \theta(W \in A,\ v \in B),
\end{aligned}$$

which establishes that the distribution of $(\tilde{W}, \tilde{v})$ under the measure $\gamma_{\tilde{v}}$ is the same as the distribution of (W, v) under θ. We use this equivalence and (7) to obtain

$$\begin{aligned} -\log Ee^{-f(W)} &\le E^{\tilde{v}}\left\{\tfrac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds + f\left(\tilde{W} + \int_0^{\cdot} \tilde{v}_s\, ds\right)\right\} \\ &= E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(W + \int_0^{\cdot} v_s\, ds\right)\right\}, \end{aligned}$$

which implies (8) for all $v \in \mathcal{A}_s$. Let $L_\theta(W + \int_0^{\cdot} v_s\, ds)$ denote the measure on $\mathcal{C}([0,1]: \mathbb{R}^d)$ that is induced by $W + \int_0^{\cdot} v_s\, ds$ under θ. Using (6), the equality in the last display implies that

$$R\left(L_\theta\left(W + \int_0^{\cdot} v_s\, ds\right) \middle\| \theta\right) = E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds\right\} \tag{9}$$

for all $v \in \mathcal{A}_s$.

Step 2. *Bounded* v. Let $v \in \mathcal{A}_b$, so that $\|v_s(\omega)\| \le M < \infty$ for $0 \le s \le 1$, $\omega \in \Omega$. According to Proposition 2.7, there exists a sequence of simple processes $\{v^n, n \in \mathbb{N}\}$ such that $\|v_s^n(\omega)\| \le M < \infty$ for all $0 \le s \le 1$ and $\omega \in \Omega$, and

$$\lim_{n\to\infty} E\int_0^1 \|v_s^n - v_s\|^2\, ds = 0.$$

It follows that $(W, \int_0^{\cdot} v_s^n\, ds)$ converges in distribution to $(W, \int_0^{\cdot} v_s\, ds)$ in $(\mathcal{C}([0,1]: \mathbb{R}^d))^2$.

By virtue of Step 1, for every $n \in \mathbb{N}$,

$$-\log Ee^{-f(W)} \le E\left\{\tfrac{1}{2}\int_0^1 \|v_s^n\|^2\, ds + f\left(W + \int_0^{\cdot} v_s^n\, ds\right)\right\}. \tag{10}$$

It remains to show that the inequality above continues to hold in the limit as $n \to \infty$. To this end, let $\mu_n \doteq L_\theta(W + \int_0^{\cdot} v_s^n\, ds)$. Then (9) implies that

$$\sup_{n\in\mathbb{N}} R(\mu_n \| \theta) = \sup_{n\in\mathbb{N}} E\left\{\frac{1}{2}\int_0^1 \|v_s^n\|^2\, ds\right\} \le \frac{M^2}{2} < \infty.$$

Hence we can apply Lemma 2.8(a) to obtain

$$\lim_{n\to\infty} Ef\left(W + \int_0^{\cdot} v_s^n\, ds\right) = Ef\left(W + \int_0^{\cdot} v_s\, ds\right).$$

Letting $n \to \infty$ in (10) we conclude that (8) is valid for the limit process v, and thus for any $v \in \mathcal{A}_b$. Using the lower semicontinuity of $R(\cdot \| \theta)$, we also obtain

$$R\left(L_\theta\left(W + \int_0^{\cdot} v_s\, ds\right) \middle\| \theta\right) \le E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds\right\} \tag{11}$$

for all $v \in \mathcal{A}_b$.

Step 3. *General* $v \in A$. We define

$$v_s^n(\phi) \doteq v_s(\phi) 1_{\{\|v_s(\phi)\| \le n\}}, \qquad 0 \le s \le 1, \ \phi \in C([0,1]: \mathbb{R}^d).$$

Then v^n is bounded for every $n \in \mathbb{N}$ and thus Step 2 guarantees that (10) holds for each v^n. Let $\mu_n \doteq L_\theta(W + \int_0^\cdot v_s^n \, ds)$. Then (11) implies that

$$\sup_{n \in \mathbb{N}} R(\mu_n \,\|\, \theta) \le \sup_{n \in \mathbb{N}} E\left\{ \tfrac{1}{2} \int_0^1 \|v_s^n\|^2 \, ds \right\} \le E\left\{ \tfrac{1}{2} \int_0^1 \|v_s\|^2 \, ds \right\} < \infty.$$

As in Step 2, Lemma 2.8 and the dominated convergence theorem yield (8) for any $v \in A$, which in turn implies the desired upper bound. For use in Section 5, we note that the lower semicontinuity of $R(\cdot \,\|\, \theta)$ implies that, for all $v \in A$,

$$R\left(L_\theta\left(W + \int_0^\cdot v_s \, ds \right) \middle\| \theta \right) \le E\left\{ \tfrac{1}{2} \int_0^1 \|v_s\|^2 \, ds \right\}. \tag{12}$$

PROOF OF THE LOWER BOUND. Consider the measure γ_0 which infimizes in the variational formula (1). Then γ_0 is not only absolutely continuous with respect to θ, but it is in fact equivalent to θ on F_1. It follows that, for each $t \in [0,1]$, the restriction of γ_0 to F_t is equivalent to the restriction of θ to F_t. Let R_t be the corresponding Radon–Nikodym derivative. Then $R_t = E[(d\gamma_0/d\theta) \mid F_t]$, and the process $\{R_t; 0 \le t \le 1\}$ forms a θ-martingale that is bounded from below and above θ-a.s. by constants $\exp(-2\|f\|_\infty)$ and $\exp(2\|f\|_\infty)$, respectively. Moreover, since R_t is a martingale with respect to the augmentation under θ of the filtration generated by a Brownian motion, it can be represented as the stochastic integral $R_t = 1 + \int_0^t u_s \, dW_s$, where u_s is progressively measurable ([12], Theorems 3.4.15–16). Thanks to the boundedness from below of R_t we can define $\tilde{v}_t = u_t/R_t$ and write

$$R_t = 1 + \int_0^t \tilde{v}_s R_s \, dW_s. \tag{13}$$

The process R_t is uniformly bounded, and thus $ER_1^2 < \infty$. This and (13) yield $E \int_0^1 \|\tilde{v}_s\|^2 R_s^2 \, ds < \infty$. Because $R_t \ge \exp(-2\|f\|_\infty)$, θ-a.s., this implies $E \int_0^1 \|\tilde{v}_s\|^2 \, ds < \infty$, and since $d\gamma_0/d\theta$ is also bounded,

$$\int_{C([0,1]: \mathbb{R}^d)} \int_0^1 \|\tilde{v}_s\|^2 \, ds \, d\gamma_0 < \infty. \tag{14}$$

These bounds and (13) give ([12], page 191)

$$R_t = \exp\left[\int_0^t \tilde{v}_s \, dW_s - \tfrac{1}{2} \int_0^t \|\tilde{v}_s\|^2 \, ds \right]. \tag{15}$$

Since R_t is a martingale, Girsanov's theorem identifies γ_0 as the measure under which the process $\widetilde{W} \doteq W - \int_0^\cdot \tilde{v}_s \, ds$ is a Brownian motion. Using (14) and (15) as in (5)–(6) to evaluate $R(\gamma_0 \,\|\, \theta)$, we obtain

$$-\log Ee^{-f(W)} = E^{\gamma_0}\left\{ \tfrac{1}{2} \int_0^1 \|\tilde{v}_s\|^2 \, ds + f\left(\widetilde{W} + \int_0^\cdot \tilde{v}_s \, ds \right) \right\}.$$

Assume for a moment that f is continuous. Let $\{\tilde{v}^n,\ n \in N\}$ be a sequence of bounded simple processes such that

$$\lim_{n\to\infty} E^{\gamma_0}\left\{\int_0^1 \|\tilde{v}_s^n - \tilde{v}_s\|^2\, ds\right\} = 0.$$

Then given $\varepsilon > 0$ there exist $n < \infty$ and $C < \infty$ such that $\|\tilde{v}_s^n(\omega)\| \le C$ for $0 \le s \le 1$, $\omega \in \Omega$, and

$$\begin{aligned} &E^{\gamma_0}\left\{\frac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds + f\left(\tilde{W} + \int_0^{\cdot} \tilde{v}_s\, ds\right)\right\} \\ &\quad \ge E^{\gamma_0}\left\{\frac{1}{2}\int_0^1 \|\tilde{v}_s^n\|^2\, ds + f\left(\tilde{W} + \int_0^{\cdot} \tilde{v}_s^n\, ds\right)\right\} - \frac{\varepsilon}{2}. \end{aligned} \tag{16}$$

Let us write $\tilde{v}^n$ in the form

$$\tilde{v}_t^n(\omega) = \xi_0(\omega) 1_{\{0\}}(t) + \sum_{i=0}^{l-1} \xi_i(\omega) 1_{(t_i, t_{i+1}]}(t), \qquad 0 \le t \le 1,\ \omega \in \Omega,$$

where $0 = t_0 < t_1 < \cdots < t_l = 1$ and ξ_i is F_{t_i}-measurable for every $i = 0, \dots, (l-1)$. For $i = 1, \dots, l$ let $M_i \doteq C([0, t_i - t_{i-1}] : \mathbb{R}^d)$ and define measurable maps $\tilde{Z}^{(i)}: \Omega \mapsto M_i$ by

$$\tilde{Z}_s^{(i)}(\omega) \doteq \tilde{W}_{t_{i-1}+s}(\omega) - \tilde{W}_{t_{i-1}}(\omega), \qquad 0 \le s \le t_i - t_{i-1}.$$

The continuity of f implies that there exist continuous functions $F_1: (\mathbb{R}^d)^l \mapsto \mathbb{R}$ and $F_2: ((\mathbb{R}^d)^l \times (\prod_{i=1}^l M_i)) \mapsto \mathbb{R}$ such that, θ-a.s.,

$$F_1(\xi_0, \dots, \xi_{l-1}) = \tfrac{1}{2}\int_0^1 \|\tilde{v}_s^n\|^2\, ds$$

and

$$F_2\left(\xi_0, \dots, \xi_{l-1}, \tilde{Z}^{(1)}, \dots, \tilde{Z}^{(l)}\right) = f\left(\tilde{W} + \int_0^{\cdot} \tilde{v}_s^n\, ds\right).$$

Define $\xi \doteq (\xi_0, \dots, \xi_{l-1})$ and, for $i = 1, \dots, l$, let $\tilde{\mathbf{Z}}^i \doteq (\tilde{Z}^{(1)}, \dots, \tilde{Z}^{(i)})$. With this notation, (16) can be rewritten as

$$\begin{aligned} &E^{\gamma_0}\left\{\frac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds + f\left(\tilde{W} + \int_0^{\cdot} \tilde{v}_s\, ds\right)\right\} \\ &\quad \ge E^{\gamma_0}\left\{F_1(\xi) + F_2\left(\xi, \tilde{\mathbf{Z}}^l\right)\right\} - \frac{\varepsilon}{2}. \end{aligned} \tag{17}$$

Let us observe that each ξ_i can be written as a bounded measurable function of $\tilde{Z}^{(j)}$, $j < i$. A construction that makes use of the independence of increments of Brownian motion, this F_{t_i}-measurability of ξ_i and the smoothing allowed by Theorem 2.6 gives the existence of a continuous mapping $\Gamma: \prod_{i=1}^l M_i \mapsto (\mathbb{R}^d)^l$ such that the following conclusions hold:

(i) $\Gamma(\mathrm{z}^l)$ can be written in the form $(\Gamma_0, \Gamma_1(\mathrm{z}^1), \dots, \Gamma_l(\mathrm{z}^l))$, where $\mathrm{z}^i \doteq (z_1, \dots, z_i) \in \prod_{j=1}^i M_j$;

(ii) $\Gamma_0 \in \mathbb{R}^d$ is deterministic with $\|\Gamma_0\| \le C$;

(iii) for $i = 1, \dots, l$, $\Gamma_i \colon \prod_{j=1}^{i} M_j \mapsto \mathbb{R}^d$ satisfies $\|\Gamma_i(\mathbf{u})\| \le C$ for $\mathbf{u} \in \prod_{j=1}^{i} M_j$;

(iv) we have the inequality

$$E^{\gamma_0}\Big\{F_1(\xi) + F_2(\xi, \widetilde{\mathbf{Z}}^l)\Big\} \ge E^{\gamma_0}\Big\{F_1\big(\Gamma(\widetilde{\mathbf{Z}}^l)\big) + F_2\big(\Gamma(\widetilde{\mathbf{Z}}^l), \widetilde{\mathbf{Z}}^l\big)\Big\} - \frac{\varepsilon}{2}.$$

Details of the construction are given in the Appendix. Now define measurable maps $Z^{(i)} \colon \Omega \mapsto M_i$ by setting

$$Z_s^{(i)}(\omega) \doteq W_{t_{i-1}+s}(\omega) - W_{t_{i-1}}(\omega), \qquad 0 \le s \le t_i - t_{i-1},\ i = 1, \dots, l,$$

and let $\mathbf{Z}^i \doteq (Z^{(1)}, \dots, Z^{(i)})$. Finally define

$$\bar{v}_t(\omega) \doteq \Gamma_0 1_{[0, t_1]}(t) + \sum_{i=1}^{l-1} \Gamma_i\big(\mathbf{Z}^i(\omega)\big) 1_{(t_i, t_{i+1}]}(t), \qquad 0 \le t \le 1.$$

By construction, $\bar{v}$ is a simple process in A which satisfies

$$E^{\gamma_0}\left\{\frac{1}{2}\int_0^1 \|\tilde{v}_s^n\|^2\, ds + f\left(\widetilde{W} + \int_0^{\cdot} \tilde{v}_s^n\, ds\right)\right\} \ge E\left\{\frac{1}{2}\int_0^1 \|\bar{v}_s\|^2\, ds + f\left(W + \int_0^{\cdot} \bar{v}_s\, ds\right)\right\} - \frac{\varepsilon}{2}.$$

Combining this inequality with (17) yields

$$\begin{aligned} -\log Ee^{-f(W)} &= E^{\gamma_0}\left\{\tfrac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds + f\left(\widetilde{W} + \int_0^{\cdot} \tilde{v}_s\, ds\right)\right\} \\ &\ge E\left\{\tfrac{1}{2}\int_0^1 \|\bar{v}_s\|^2\, ds + f\left(W + \int_0^{\cdot} \bar{v}_s\, ds\right)\right\} - \varepsilon, \end{aligned} \tag{18}$$

which proves the lower bound for continuous f.

If f is not continuous, let $\{f_j,\ j \in \mathbb{N}\}$ be a sequence of bounded and continuous functions such that $\|f_j\|_\infty \le \|f\|_\infty < \infty$ and $\lim_{j\to\infty} f_j = f$, θ-a.s. The preceding argument applied to each of the functions f_j implies that there exists a sequence $\{\bar{v}^j,\ j \in \mathbb{N}\}$ of processes in A such that $\bar{v}^j$ satisfies (18) for each j but with f replaced by f_j; that is,

$$-\log Ee^{-f_j(W)} \ge E\left\{\tfrac{1}{2}\int_0^1 \|\bar{v}_s^j\|^2\, ds + f_j\left(W + \int_0^{\cdot} \bar{v}_s^j\, ds\right)\right\} - \varepsilon. \tag{19}$$

Thanks to (9) we have

$$\sup_j R\left(L_\theta\left(W + \int_0^{\cdot} \bar{v}_s^j\, ds\right)\middle\|\theta\right) = \sup_j E\left\{\tfrac{1}{2}\int_0^1 \|\bar{v}_s^j\|^2\, ds\right\} \le \|f\|_\infty.$$

It follows from this bound that the pair $(\int_0^{\cdot} \bar{v}_s^j\, ds, W)$ is tight. Hence there exists a subsequence such that $(\int_0^{\cdot} \bar{v}_s^j\, ds, W)$ converges in distribution to $(\int_0^{\cdot} \bar{v}_s\, ds, W)$. It follows from (19), the dominated convergence theorem and Lemma 2.8 that, for all sufficiently large j,

$$-\log E^{-f(W)} \ge E\left\{\tfrac{1}{2}\int_0^1 \|\bar{v}_s^j\|^2\, ds + f\left(W + \int_0^{\cdot} \bar{v}_s^j\, ds\right)\right\} - 2\varepsilon.$$

Since each $\bar{v}^j$ is in A, this completes the proof of the lower bound. □

4. An application to diffusions. In this section we specialize the result of Theorem 3.1 to the case of bounded measurable functionals of strong solutions to stochastic differential equations (Theorem 4.1). We then apply this formula to prove a well-known result about large deviations for small noise diffusions.

4.1. *The representation formula.* Let $b_i(t, x)$ and $\sigma_{ij}(t, x)$, $1 \le i \le m$, $1 \le j \le d$, be Borel-measurable functions from $[0, 1] \times \mathbb{R}^m$ into $\mathbb{R}$, and define the $m \times 1$ vector $b(t, x) = \{b_i(t, x)\}$ and the $m \times d$ matrix $\sigma(t, x) = \{\sigma_{ij}(t, x)\}$. Let $v \in A$ be given, and consider the stochastic differential equations

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \qquad X_0 = x, \tag{20}$$

and

$$dX_t^v = b(t, X_t^v)\, dt + \sigma(t, X_t^v)\, v_t\, dt + \sigma(t, X_t^v)\, dW_t, \qquad X_0^v = x, \tag{21}$$

where $x \in \mathbb{R}^m$ is deterministic and W is d-dimensional Brownian motion. We assume that (20) and (21) have a unique strong solution. For instance, this is the case if $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ satisfy Lipschitz and linear growth conditions, that is, there exists a constant $K < \infty$ such that

$$\|b(t, x) - b(t, y)\| + \|\sigma(t, x) - \sigma(t, y)\| \le K\|x - y\| \tag{22}$$

and

$$\|b(t, x)\|^2 + \|\sigma(t, x)\|^2 \le K^2(1 + \|x\|^2). \tag{23}$$

However, these conditions can be significantly weakened (see [16]).

The next theorem corresponds to Theorem 3.1 for the special case of diffusions which are strong solutions to stochastic differential equations.

THEOREM 4.1. *Let X be the diffusion process that is the unique strong solution to the stochastic differential equation* (20). *Then for any bounded Borel-measurable function f: $C([0, 1]: \mathbb{R}^m) \to \mathbb{R}$ the following representation holds*:

$$-\log Ee^{-f(X)} = \inf_{v \in A} E\left\{ \frac{1}{2}\int_0^1 \|v_s\|^2\, ds + f(X^v) \right\}, \tag{24}$$

where X^v is the unique solution to (21) *and E_x denotes expectation conditioned on $X_0 = x$.*

PROOF. Since X is a strong solution to (20), there exists a $\mathcal{B}(C([0, 1]: \mathbb{R}^d))/\mathcal{B}(C([0, 1]: \mathbb{R}^m))$-measurable function h: $C([0, 1]: \mathbb{R}^d) \to C([0, 1]: \mathbb{R}^m)$ such that

$$X_{\cdot} = h(W)_{\cdot}, \qquad \theta\text{-a.s.}$$

More precisely, the relation

$$h[W]_t = x + \int_0^t b(s, h[W]_s)\, ds + \int_0^t \sigma(s, h[W]_s)\, dW_s$$

is satisfied almost surely with respect to Wiener measure for all $t \in [0,1]$. Hence the left-hand side of (24) can be rewritten as $-\log Ee^{-f \circ h[W]}$. Representation (24) will be established once we verify that $h[W + \int_0^\cdot v_s\, ds] = X^v$. This follows easily from the strong existence and the uniqueness of solutions: letting $\tilde{W} = W + \int_0^\cdot v_s\, ds$, we have

$$\begin{aligned} h[\tilde{W}]_t &= x + \int_0^t b(s, h[\tilde{W}]_s)\, ds + \int_0^t \sigma(s, h[\tilde{W}]_s)\, d\tilde{W}_s \\ &= x + \int_0^t b(s, h[\tilde{W}]_s)\, ds + \int_0^t \sigma(s, h[\tilde{W}]_s)\, dW_s \\ &\quad + \int_0^t \sigma(s, h[\tilde{W}]_s) v_s\, ds, \end{aligned}$$

and, by the uniqueness of solutions to (21), this implies $h[W + \int_0^\cdot v_s\, ds] = X^v$. □

4.2. *Small noise diffusions.* As an elementary but elegant application of Theorem 4.1, we prove the following well-known theorem on large deviations for small noise diffusions. A more demanding problem for which the representation seems to be very useful, concerns diffusion processes with discontinuous coefficients [5]. For a discrete time analogue, see [9], Chapter 7.

We will use the concept of a Laplace principle, which we now define.

DEFINITION 4.2. Let $\{Y^\varepsilon\}$ be a family of random variables taking values in a Polish space $\mathbf{Y}$ and let I: $\mathbf{Y} \to [0, \infty]$.

We say the sequence $\{Y^\varepsilon\}$ satisfies a *Laplace principle* with rate function I if, for every bounded continuous function g mapping $\mathbf{Y}$ into $\mathbb{R}$,

$$\lim_{\varepsilon \to 0} -\varepsilon \log E\left\{\exp\left[-\frac{g(Y^\varepsilon)}{\varepsilon}\right]\right\} = \inf_{y \in \mathbf{Y}} \{g(y) + I(y)\},$$

and if for every $M \in [0, \infty)$ the set $\{y\colon I(y) \le M\}$ is compact.

Let $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ be as in the previous subsection, and for $\varepsilon > 0$ consider the stochastic differential equation

$$dX_t^\varepsilon = b(t, X_t^\varepsilon)\, dt + \varepsilon^{1/2}\sigma(t, X_t^\varepsilon)\, dW_t, \qquad 0 \le t \le 1,\ X_0^\varepsilon = x. \tag{25}$$

Let $\gamma_{\varepsilon, x}$ be the measure induced by X^ε on $(\mathbf{C}([0,1]\colon \mathbb{R}^m), \mathbf{B}(\mathbf{C}([0,1]\colon \mathbb{R}^m)))$. As $\varepsilon \to 0$, the measure $\gamma_{\varepsilon, x}$ converges weakly to the Dirac measure concentrated at the single trajectory which solves the ordinary differential equation

$$\dot{\phi}_t = b(t, \phi_t), \qquad \phi_0 = x.$$

Theorem 4.3 states the Laplace principle for the family $\{X^\varepsilon\}$. A Laplace principle is equivalent to a large deviation principle [if the definition of a rate function includes the requirement of compact level sets ([9], Section 2.2)]. It thus measures the probability of deviations away from the trajectory ϕ.

THEOREM 4.3. *Let $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ satisfy* (22) *and* (23). *Then* $\{X^\varepsilon\}$, *the solution to* (25), *satisfies the Laplace principle in* $C([0,1]:\mathbb{R}^m)$ *with rate function*

$$I_x(f) = \inf_{\{v\in L^2([0,1]:\,\mathbb{R}^m):\ f_t = x+\int_0^t b(s,f_s)\,ds+\int_0^t \sigma(s,f_s)v_s\,ds\}} \frac{1}{2}\int_0^1 \|v_t\|^2\,dt$$

whenever $\{v\in L^2([0,1]:\mathbb{R}^m):\ f_t = x+\int_0^t b(s,f_s)\,ds+\int_0^t \sigma(s,f_s)v_s\,ds\}\neq\varnothing$, *and* $I_x(f)=\infty$ *otherwise.*

PROOF OF THE LOWER BOUND. Fix $x\in\mathbb{R}^m$ and let g be any bounded and continuous function mapping $C([0,1]:\mathbb{R}^m)$ into $\mathbb{R}$. We will prove the Laplace principle lower bound

$$\liminf_{\varepsilon\to 0} -\varepsilon\log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\} \ge \inf_{\phi\in C([0,1]:\,\mathbb{R}^m)}\{g(\phi)+I_x(\phi)\}. \tag{26}$$

It suffices to show that every subsequence of $\{X^\varepsilon\}$ has a subsubsequence satisfying (26). Since g is bounded, we can assume that along the given subsequence

$$-\varepsilon\log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\}$$

converges. Applying Theorem 4.1 to the function g we obtain

$$-\varepsilon\log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\} = \inf_{v\in A} E_x\left\{\frac{1}{2}\int_0^1\|v_s\|^2\,ds + g(X^{\varepsilon,v})\right\},$$

where $X^{\varepsilon,v}$ is the unique solution to

$$X_t^{\varepsilon,v} = x+\int_0^t b(s,X_s^{\varepsilon,v})\,ds+\int_0^t\sigma(s,X_s^{\varepsilon,v})\,v_s\,ds+\int_0^t\varepsilon^{1/2}\sigma(s,X_s^{\varepsilon,v})\,dW_s.$$

For every ε let $v^\varepsilon\in A$ come within ε of the infimum, so that

$$-\varepsilon\log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\} \ge E_x\left\{\frac{1}{2}\int_0^1\|v_s^\varepsilon\|^2\,ds + g(X^{\varepsilon,v^\varepsilon})\right\}-\varepsilon.$$

Tightness of $\{\frac{1}{2}\int_0^1\|v_s^\varepsilon\|^2\,ds\}$ follows immediately from the fact that, for all $\varepsilon>0$, $E_x\{\frac{1}{2}\int_0^1\|v_s^\varepsilon\|^2\,ds\}$ is bounded above by $2M$, where $M=\|g\|_\infty$. Given the tightness of $\{\frac{1}{2}\int_0^1\|v_s^\varepsilon\|^2\,ds\}$, the tightness of $X^{\varepsilon,v^\varepsilon}$ is standard. We would lastly like to show tightness of $\{v^\varepsilon\}$ in $L^2([0,1]:\mathbb{R}^m)$ under the weak topology. A slight nuisance here is the fact that $L^2([0,1]:\mathbb{R}^m)$ is not metrizable as a Polish space with this topology. However, for any $N<\infty$, the set

$$S_N=\left\{f\in L^2([0,1]:\mathbb{R}^m):\ \int_0^1\|f_s\|^2\,ds\le N\right\}$$

is metrizable as a compact Polish space with this topology ([13], Theorem III.1). For a given $\delta > 0$, consider the set $K \doteq S_{2M/\delta}$. Then by Chebyshev's inequality, for all $\varepsilon > 0$,

$$\theta\{v^\varepsilon \notin K\} = \theta\left\{\int_0^1 \|v_s^\varepsilon\|^2 \, ds > \frac{2M}{\delta}\right\} \le \frac{\delta}{2M} E\left\{\frac{1}{2}\int_0^1 \|v_s^\varepsilon\|^2 \, ds\right\} < \delta.$$

This implies (modulo an approximation argument which we will omit) that v^ε can be assumed to take values in a compact Polish space with probability 1. Hence there exists a subsubsequence for which $(X^{\varepsilon, v^\varepsilon}, v^\varepsilon)$ converges in distribution to (X^v, v), where X^v is the unique solution to

$$X_t^v = x + \int_0^t b(s, X_s^v) \, ds + \int_0^t \sigma(s, X_s^v) v_s \, ds.$$

Since g is bounded and continuous, applying Fatou's lemma to this subsubsequence yields

$$\begin{aligned}
&\liminf_{\varepsilon \to 0} -\varepsilon \log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\} \\
&\quad\ge \liminf_{\varepsilon \to 0}\left[E_x\left\{\frac{1}{2}\int_0^1 \|v_s^\varepsilon\|^2 \, ds + g(X^{\varepsilon, v^\varepsilon})\right\} - \varepsilon\right] \\
&\quad\ge E_x\left\{\frac{1}{2}\int_0^1 \|v_s\|^2 \, ds + g(X^v)\right\} \\
&\quad\ge \left\{\inf_{\{I_1(v,\phi):\ \phi_t = x + \int_0^t b(s,\phi_s)\,ds + \int_0^t \sigma(s,\phi_s) v_s\,ds\}} \frac{1}{2}\int_0^1 \|v_s\|^2 \, ds + g(\phi)\right\} \\
&\quad\ge \inf_{\phi \in C([0,1]:\, \mathbb{R}^m)} \{I_x(\phi) + g(\phi)\}.
\end{aligned}$$

Compactness of level sets. The fact that $I_x(\cdot)$ has compact level sets follows from the compactness of S_N for each N, and the continuity of the map $v \mapsto \phi$ defined by $\phi_t = x + \int_0^t b(s, \phi_s)\, ds + \int_0^t \sigma(s, \phi_s) v_s \, ds$.

PROOF OF THE UPPER BOUND. We now prove

$$\limsup_{\varepsilon \to 0} -\varepsilon \log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\} \le \inf_{\phi \in C([0,1]:\, \mathbb{R}^m)} \{I_x(\phi) + g(\phi)\}. \tag{27}$$

Fix $\delta > 0$. For any g bounded and continuous there exists $\phi \in C([0,1]: \mathbb{R}^m)$ such that

$$I_x(\phi) + g(\phi) \le \inf_{\phi \in C([0,1]:\, \mathbb{R}^m)} \{I_x(\phi) + g(\phi)\} + \frac{\delta}{2} < \infty.$$

For such ϕ, choose $\tilde{v}$ such that

$$\frac{1}{2}\int_0^1 \|\tilde{v}_s\|^2 \, ds \le I_x(\phi) + \frac{\delta}{2},$$

and $\phi_t = x + \int_0^t b(s, \phi_s)\, ds + \int_0^t \sigma(s, \phi_s)\tilde{v}_s\, ds$. For ϕ, $\tilde{v}$ and $X^{\varepsilon, \tilde{v}}$ as above, Theorem 4.1 implies

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} -\varepsilon \log E_x\left\{\exp\left[-\frac{g(X^\varepsilon)}{\varepsilon}\right]\right\} \\
&\quad = \limsup_{\varepsilon\to 0} \inf_{v\in A} E_x\left\{\frac{1}{2}\int_0^1 \|v_s\|^2\, ds + g(X^{\varepsilon, v})\right\} \\
&\quad \le \limsup_{\varepsilon\to 0} E_x\left\{\frac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds + g(X^{\varepsilon, \tilde{v}})\right\} \\
&\quad = \frac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds + \limsup_{\varepsilon\to 0} E_x\, g(X^{\varepsilon, \tilde{v}}) \\
&\quad \le I_x(\phi) + \frac{\delta}{2} + \limsup_{\varepsilon\to 0} E_x\, g(X^{\varepsilon, \tilde{v}}).
\end{aligned}$$

Since g is bounded and continuous and $X^{\varepsilon, \tilde{v}}$ converges in distribution to ϕ, the inequalities can be continued as

$$I_x(\phi) + \frac{\delta}{2} + g(\phi) \le \inf_{\phi\in C([0,1]:\, \mathbb{R}^m)} \{I_x(\phi) + g(\phi)\} + \delta.$$

Since δ is arbitrary, the proof of (27) is complete. □

5. Extensions of the representation. The following extension is needed for applications to risk-sensitive control [4].

THEOREM 5.1. *Let f be a Borel-measurable function that is bounded from above. Then*

$$-\log Ee^{-f(W)} = \inf_{v\in A} E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(W + \int_0^\cdot v_s\, ds\right)\right\}.$$

PROOF. Without loss of generality we may assume $f \le 0$. Let $f_N = f \vee (-N)$. Then by the monotone convergence theorem and Theorem 3.1 we have

$$\begin{aligned}
-\log Ee^{-f(W)} &= \lim_{N\to\infty} -\log Ee^{-f_N(W)} \\
&= \lim_{N\to\infty} \inf_{v\in A} E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f_N\left(W + \int_0^\cdot v_s\, ds\right)\right\}.
\end{aligned}$$

It remains to verify that the last expression coincides with

$$\inf_{v\in A} E\left\{\tfrac{1}{2}\int_0^1 \|v_s\|^2\, ds + f\left(W + \int_0^\cdot v_s\, ds\right)\right\}. \tag{28}$$

We first treat the case when (28) is finite. Fix $\varepsilon > 0$ and choose $\tilde{v} \in A$ to be ε-optimal in (28). The nonpositivity of f implies $E\frac{1}{2}\int_0^1 \|\tilde{v}_s\|^2\, ds < \infty$. By the monotone convergence theorem (applied to $E\{-f_N(W+\int_0^{\cdot}\tilde{v}_s\,ds)\}$)

$$\lim_{N\to\infty}\inf_{v\in A} E\left\{\tfrac{1}{2}\int_0^1\|v_s\|^2\,ds + f_N\left(W+\int_0^{\cdot} v_s\,ds\right)\right\}$$

$$\le \lim_{N\to\infty} E\left\{\tfrac{1}{2}\int_0^1\|\tilde{v}_s\|^2\,ds + f_N\left(W+\int_0^{\cdot} \tilde{v}_s\,ds\right)\right\}$$

$$= E\left\{\tfrac{1}{2}\int_0^1\|\tilde{v}_s\|^2\,ds + f\left(W+\int_0^{\cdot} \tilde{v}_s\,ds\right)\right\}$$

$$\le \inf_{v\in A} E\left\{\tfrac{1}{2}\int_0^1\|v_s\|^2\,ds + f\left(W+\int_0^{\cdot} v_s\,ds\right)\right\} + \varepsilon.$$

Since ε is arbitrary, the upper bound follows.

For the case when (28) is equal to $-\infty$, we choose $\tilde{v}\in A$ such that $E\{\frac{1}{2}\int_0^1\|\tilde{v}_s\|^2\,ds + f(W+\int_0^{\cdot}\tilde{v}_s\,ds)\}$ has an arbitrarily large negative value. The same argument as above gives us the desired upper bound. Finally, we note that the lower bound is elementary, since $f_N \ge f$. □

APPENDIX

In this section we provide details for the proof of the lower bound in Theorem 3.1. We first state a lemma concerning measurable selections.

LEMMA A.1. *Let E_1, E_2 be Polish spaces and let f: $E_1\times E_2\mapsto\mathbb{R}$ be bounded and continuous. Let K be a compact set in E_2. For each $x\in E_1$ define*

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

Then there exists a Borel-measurable function g: $E_1\mapsto E_2$ such that $g(x)\in\Gamma_x$ for all $x\in E_1$.

We now prove the existence of a function Γ satisfying the properties (i)–(iv) stated below (17).

Since every probability measure on a Polish space is tight, there exists a compact set $K_0\subset\mathbb{R}^d$ such that

$$E^{\gamma_0}\left\{F_1(\xi^l) + F_2(\xi^l,\tilde{\mathbf{Z}}^l)\right\} \ge E^{\gamma_0}\left\{1_{(K_0)^{\otimes l}}(\xi^l)\left(F_1(\xi^l)+F_2(\xi^l,\tilde{\mathbf{Z}}^l)\right)\right\} - \frac{\varepsilon}{4l}.$$

Since $(\tilde{W}_t, F_t)$ is a Wiener process under γ_0, $\tilde{W}_{u_2}-\tilde{W}_{u_1}$ is independent of F_{u_1} for $0\le u_1\le u_2\le 1$. Therefore, $\tilde{Z}^{(j)}$ is independent of $(\xi^j,\tilde{\mathbf{Z}}^{j-1})$ under γ_0. Let μ_j denote standard Wiener measure on M_j and let $F_2^{(1)}$ be the real-valued continuous map on $((\mathbb{R}^d)^l\times(\prod_{j=1}^{l-1}M_j))$ obtained by integrating out $\tilde{Z}^{(l)}$ from F_2, that is, $F_2^{(1)}(y)\doteq\int F_2(y,z)\mu_l(dz)$, where $y\in((\mathbb{R}^d)^l\times(\prod_{j=1}^{l-1}M_j))$. Using the fact that $\|\xi_{l-1}\|\le C$, θ-a.s., and Lemma 6.1 with $E_2\doteq\mathbb{R}^d$, $E_1\doteq((\mathbb{R}^d)^{l-1}\times(\prod_{j=1}^{l-1}M_j))$, $K\doteq K_0\cap\{x\in\mathbb{R}^d:\|x\|\le C\}$ and $f\doteq F_1+F_2^{(1)}$, we obtain the existence of a measurable function h: $((\mathbb{R}^d)^{l-1}\times(\prod_{j=1}^{l-1}M_j))\mapsto\mathbb{R}^d$ with $\|h\|_\infty\le C$ such that the right-hand side of (17) is bounded below by

$$E^{\gamma_0}\left\{F_1\left(\xi^{l-1},h(\xi^{l-1},\tilde{\mathbf{Z}}^{l-1})\right) + F_2^{(1)}\left(\xi^{l-1},h(\xi^{l-1},\tilde{\mathbf{Z}}^{l-1}),\tilde{\mathbf{Z}}^{l-1}\right)\right\} - \frac{\varepsilon}{2l}.$$

Thanks to Lemma 2.6 and the dominated convergence theorem, we can take h to be continuous if we subtract an additional $\varepsilon/2l$ from this lower bound.

If we iterate this procedure $l-1$ times, we obtain the existence of a continuous mapping $\Gamma\colon \prod_{i=1}^{l} M_i \mapsto (\mathbb{R}^d)^l$ such that the conclusions (i)–(iv) stated below equation (17) hold.

Acknowledgment. The authors wish to thank an anonymous referee for a careful reading of the paper and for suggesting several improvements.

REFERENCES

[1] BENSOUSSAN, A. and VAN SCHUPPEN, J. H. (1985). Optimal control of partially observable stochastic systems with an exponential-of-integral performance index. *SIAM J. Control Optim.* 23 599–613.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] BOEL, R. K., JAMES, M. R. and PETERSEN, I. R. (1996). Robust filtering. Technical report, Australian National Univ.

[4] BOUÉ, M. and DUPUIS, P. (1998). Risk-sensitive and robust escape control for degenerate processes. Unpublished manuscript.

[5] BOUÉ, M., DUPUIS, P. and ELLIS, R. S. (1997). Large deviations for diffusions with discontinuous statistics. Unpublished manuscript.

[6] DAI PRA, P., MENEGHINI, L. and RUNGGALDIER, W. J. (1997). Some connections between stochastic control and dynamic games. *Math. Control Signals Systems* 9 303–326.

[7] DOOB, J. L. (1994). *Measure Theory.* Springer, New York.

[8] DUPUIS, P. and ELLIS, R. S. (1995). The large deviation principle for a general class of queueing systems, I. *Trans. Amer. Math. Soc.* 347 2689–2751.

[9] DUPUIS, P. and ELLIS, R. (1997). *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley, New York.

[10] FLEMING, W. H. (1978). Exit probabilities and optimal stochastic control. *Appl. Math. Optim.* 4 329–346.

[11] FLEMING, W. H. and MCENEANEY, W. M. (1992). Risk sensitive optimal control and differential games. *Lecture Notes in Control and Inform. Sci.* 184 185–197. Springer, Berlin.

[12] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York.

[13] KOLMOGOROV, A. N. and FOMIN, S. F. (1957). *Elements of the Theory of Functions and Functional Analysis.* Graylock, Rochester.

[14] KUMAR, P. R. and VAN SCHUPPEN, J. H. (1981). On the optimal control of stochastic systems with an exponential-of-integral performance index. *SIAM J. Math. Anal. Appl.* 80 312–332.

[15] SHEU, S-J. (1985). Stochastic control and exit probabilities of jump processes. *SIAM J. Control Optim.* 23 306–328.

[16] VERETENNIKOV, A. J. (1981). On strong solutions and explicit formulas for solutions of stochastic integral equations. *Math. USSR-Sb.* 39 387–403.

[17] WHITTLE, P. (1990). *Risk-Sensitive Optimal Control.* Wiley, New York.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003
E-MAIL: boue@math.umass.edu

LEFSCHETZ CENTER FOR DYNAMICAL SYSTEMS
DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912
E-MAIL: dupuis@cfm.brown.edu