The Annals of Probability
2016, Vol. 44, No. 4, 3111–3186
DOI: 10.1214/15-AOP1043

ON LARGE DEVIATIONS OF COUPLED DIFFUSIONS WITH TIME SCALE SEPARATION

BY ANATOLII A. PUHALSKII

Institute for Problems in Information Transmission (IITP)

We consider two Itô equations that evolve on different time scales. The equations are fully coupled in the sense that all of the coefficients may depend on both the "slow" and the "fast" variables and the diffusion terms may be correlated. The diffusion term in the slow process is small. A large deviation principle is obtained for the joint distribution of the slow process and of the empirical process of the fast variable. By projecting on the slow and fast variables, we arrive at new results on large deviations in the averaging framework and on large deviations of the empirical measures of ergodic diffusions, respectively. The proof relies on the property that an exponentially tight sequence of probability measures on a metric space is large deviation relatively compact. The identification of the large deviation rate function is accomplished by analyzing the large deviation limit of an exponential martingale.

1. Introduction. Consider the coupled diffusions specified by the stochastic differential equations

$$
\begin{aligned}
dX_t^\varepsilon &= A\big(X_t^\varepsilon, x_t^\varepsilon\big)\,dt + \sqrt{\varepsilon} B\big(X_t^\varepsilon, x_t^\varepsilon\big)\,dW_t^\varepsilon,\\
dx_t^\varepsilon &= \frac{1}{\varepsilon} a\big(X_t^\varepsilon, x_t^\varepsilon\big)\,dt + \frac{1}{\sqrt{\varepsilon}} b\big(X_t^\varepsilon, x_t^\varepsilon\big)\,dW_t^\varepsilon,
\end{aligned}
\tag{1.1}
$$

where $\varepsilon > 0$ is a small parameter. Here, $A(u, x)$, where $u \in \mathbb{R}^n$ and $x \in \mathbb{R}^l$, is an n-vector, $B(u, x)$ is an $n \times k$-matrix, $a(u, x)$ is an l-vector, $b(u, x)$ is an $l \times k$-matrix, and $W^\varepsilon = (W_t^\varepsilon, t \in \mathbb{R}_+)$ is an $\mathbb{R}^k$-valued standard Wiener process. Accordingly, the stochastic process $X^\varepsilon = (X_t^\varepsilon, t \in \mathbb{R}_+)$ takes values in $\mathbb{R}^n$ and the stochastic process $x^\varepsilon = (x_t^\varepsilon, t \in \mathbb{R}_+)$ takes values in $\mathbb{R}^l$. The processes X^ε and x^ε are seen to evolve on different time scales in that time for x^ε is accelerated by a factor of $1/\varepsilon$. In a number of application areas, one is concerned with finding the logarithmic asymptotics of large deviations for the "slow" process X^ε as $\varepsilon \to 0$, which is usually expressed in the form of the large deviation principle (LDP). (As a matter of fact, our interest in this setup has been aroused by an application to optimal portfolio selection.) When no diffusion term is present in the equation for the

Received February 2015; revised May 2015.

MSC2010 subject classifications. Primary 60F10, 60J60; secondary 60F17, 60G57.

Key words and phrases. The large deviation principle, diffusion processes, time scale separation, averaging, empirical processes.

slow process, this sort of result is usually referred to as “the averaging principle;” for contributions, see Freĭdlin [17], Veretennikov [51, 53, 58, 59], Feng and Kurtz [16], Section 11, and references therein. The results in the literature that obtain an LDP and identify the large deviation rate function for X^ε, with a nondegenerate diffusion term being present in the first of equations (1.1), concern the time homogeneous case where the diffusion coefficient in the equation for the fast process does not depend on the slow process, Veretennikov [55, 56, 59], Liptser [28], Feng and Kurtz [16], Section 11. The latter restriction can be removed in the setting of the averaging principle provided the state space of the fast process is compact, Veretennikov [51, 57, 58].

A different perspective has been offered by Liptser [28] whose insight was to consider the joint distribution of the slow process and of the empirical process associated with the fast variable. For the case where the processes X^ε and x^ε are one-dimensional, the coefficients $a(u,x)$ and $b(u,x)$ do not depend on the first variable and the Wiener processes driving the diffusions can be taken independent, they derived an LDP for the pair $(X^\varepsilon, \mu^\varepsilon)$ and identified the associated large deviation rate function, where μ^ε represents the empirical process associated with x_t^ε. The large deviation principle for the slow process then follows by projection.

In this paper, we extend the joint LDP in Liptser [28] to the multidimensional case. It is assumed that the process dimensions are arbitrary and that all coefficients may depend on both variables in a continuous fashion, on the time variable in a measurable fashion, and on ε. The diffusions driving the slow and the fast processes do not have to be uncorrelated. We obtain an LDP for the distribution of $(X^\varepsilon, \mu^\varepsilon)$ and produce the large deviation rate function. Projections on the first and second coordinates yield LDPs for X^ε and μ^ε, respectively.

For the time-homogeneous case, the continuity and nondegeneracy conditions on the coefficients are similar to those in the literature, except that additional smoothness properties are assumed of $b(u,x)$ as a function of x, as it is done in Liptser [28]. In return, we obtain that the probability measures for which the large deviation rate function is finite must have weakly differentiable densities whose square roots belong to the Sobolev space $\mathbb{W}^{1,2}(\mathbb{R}^l)$. In particular, additional insight is gained into the LDP for the empirical measures of ergodic diffusion processes. On the other hand, the ergodicity requirements on the fast process in the nongradiental case are more restrictive than those in some of the literature.

Also, this contribution fills in the gaps in the study of the LDP for X^ε by tackling a case of fully coupled diffusions in a noncompact state space. In addition, the coefficients may depend on the time variable explicitly. The results cover both the setup with a nondegenerate diffusion term and the setup with no diffusion term in the equation for the slow process. The form of the large deviation rate function for the slow process is new.

As in Liptser [28], an important part in our approach is played by the property that exponential tightness implies large deviation relative compactness so that once exponential tightness has been shown, establishing that a large deviation limit point is unique concludes an LDP proof. Liptser [28] identifies the large deviation rate function by evaluating limits of the probabilities that the process in question resides in small balls. We use a different device. The general idea is to consider a characterisation of stochastic processes that admits taking the large deviation limit. Such a characterisation may be the property that a certain process be a martingale, it may also arise out of the description of the process dynamic. The large deviation rate function is identified by the limiting relation; cf. Puhalskii [39, 40, 42], Puhalskii and Vladimirov [43]. In this paper, similar to Puhalskii [39, 40], the large deviation limit is taken in an exponential martingale problem that has the distribution of $(X^\varepsilon, \mu^\varepsilon)$ as a solution. We then undertake a study of the limit equation. On the one hand, regularity properties of solutions are investigated. That analysis has much in common with and uses the results and methods of the regularity theory of elliptic partial differential equations. On the other hand, the domain of the validity of the equation is expanded. Put together, those tools enable us to show that the equation has a unique solution and to identify that solution.

The rest of the paper is organized as follows. In Section 2, the main results are stated, their implications are discussed, and earlier contributions are given a more detailed consideration. Section 3 outlines the proof strategy. It is implemented in Sections 4–8. The proof is completed in Section 9. Thanks to constraints on the size of the publication, some pieces of reasoning are either omitted or merely outlined. More detail can be found in Puhalskii [41].

We conclude the Introduction by giving a list of notation and conventions adopted in the paper. The blackboard bold font is reserved for topological spaces, the boldface font is used for entities associated with probability. Vectors are treated as column vectors. The Euclidean length of vector $x = (x_1, \dots, x_d)$ from $\mathbb{R}^d$, where $d \in \mathbb{N}$, is denoted by $|x|$, ^{T} stands for the transpose of a matrix or a vector. For matrix A, $\|A\|$ denotes the operator norm and $A^\oplus$ denotes the Moore–Penrose pseudoinverse, if A is square then $\mathrm{tr}(A)$ represents the trace of A. Given positive definite symmetric matrix Δ and matrix z of a suitable dimension, which may be a vector, we define $\|z\|^2_\Delta = z^T \Delta z$. Derivatives are understood as weak, or Sobolev, derivatives. For the definitions and basic properties, the reader is referred either to Adams and Fournier [1] or to Gilbarg and Trudinger [21]. For an $\mathbb{R}$-valued function f on $\mathbb{R}^d$, Df denotes the gradient and $D^2 f$ denotes the Hessian matrix of f. If f assumes its values in $\mathbb{R}^{d_1}$, then Df is the $d \times d_1$-matrix with entries $\partial f_i/\partial x_j$ and $\mathrm{div}\, f$ represents the divergence of f, where $d_1 \in \mathbb{N}$. The divergence of a matrix is computed rowwise. Subscripts may be added to indicate that differentiation is carried out with respect to a specific variable. For instance, for an $\mathbb{R}$-valued function $f(t, u, x)$, where $u = (u_1, \dots, u_d)$ and $x = (x_1, \dots, x_{d_1})$, $D_x f$ and $D_u f$ refer to gradients in the third and the second variables, respectively, $D^2_{uu} f$ is the matrix with entries $\partial^2 f/\partial u_i\, \partial u_j$, $D^2_{xx} f$ is the matrix with entries $\partial^2 f/\partial x_i\, \partial x_j$, and $D^2_{ux} f$ is the matrix with entries $\partial^2 f/\partial u_i\, \partial x_j$. If $q > 1$, we will denote by q' the conjugate: $q' = q/(q-1)$. We use standard notation for spaces of differentiable functions, for example, $\mathbb{C}^{1,2}(\Upsilon)$ denotes the space of $\mathbb{R}$-valued functions that are continuously differentiable once in the first variable and twice in the second variable over a domain Υ in $\mathbb{R}^d$, $\mathbb{C}_0^{1,2}(\Upsilon)$ is the subspace of $\mathbb{C}^{1,2}(\Upsilon)$ of functions of compact support, $\mathbb{C}_0^1(\Upsilon)$ is the space of continuously differentiable functions of compact support, and $\mathbb{C}_0^\infty(\Upsilon)$ is the space of infinitely differentiable functions of compact support. Given a measurable function $c(x)$ on Υ with values in the set of positive definite symmetric $d \times d$-matrices and an $\mathbb{R}_+$-valued measurable function $m(x)$ on Υ, we will denote by $\mathbb{L}^2(\Upsilon, \mathbb{R}^d, c(x), m(x)\,dx)$ the Hilbert space of $\mathbb{R}^d$-valued measurable functions on Υ with the norm $\|f\|_{c(\cdot),m(\cdot)} = (\int_\Upsilon \|f(x)\|_{c(x)}^2 m(x)\,dx)^{1/2}$. If $c(x)$ is the identity matrix, the notation will be shortened to $\mathbb{L}^2(\Upsilon, \mathbb{R}^d, m(x)\,dx)$ and to $\mathbb{L}^2(\Upsilon, \mathbb{R}^d)$ if, in addition, $m(x) = 1$. Spaces $\mathbb{L}^2(\Upsilon, m(x)\,dx)$ and $\mathbb{L}^2(\Upsilon)$ are defined similarly and consist of $\mathbb{R}$-valued functions. Space $\mathbb{L}^2(\Upsilon, \mathbb{R}^d, \mu(dx))$ is defined via integration with respect to measure μ. Also, standard notation for Sobolev spaces is adhered to, for example, $\mathbb{W}^{1,2}(\Upsilon)$ is the Hilbert space of $\mathbb{R}$-valued functions f that possess the first Sobolev derivatives with the norm $\|f\|_{\mathbb{W}^{1,2}(\Upsilon)} = \|f\|_{\mathbb{L}^2(\Upsilon)} + \|Df\|_{\mathbb{L}^2(\Upsilon,\mathbb{R}^d)}$. The local version of a function space, for example, $\mathbb{W}_{\mathrm{loc}}^{1,2}(\Upsilon)$, consists of functions whose products with arbitrary $\mathbb{C}_0^\infty$-functions belong to that space, that is, $\mathbb{W}^{1,2}(\Upsilon)$ in this case, and is endowed with the weakest topology under which the mappings that associate with functions such products are continuous. We let $\mathbb{W}^{1,2}(\Upsilon, m(x)\,dx)$ denote the set of functions $f \in \mathbb{W}_{\mathrm{loc}}^{1,1}(\Upsilon)$ such that $f \in \mathbb{L}^2(\Upsilon, m(x)\,dx)$ and $Df \in \mathbb{L}^2(\Upsilon, \mathbb{R}^d, m(x)\,dx)$ equipped with the norm $\|f\|_{\mathbb{W}^{1,2}(\Upsilon,m(x)\,dx)} = \|f\|_{\mathbb{L}^2(\Upsilon,m(x)\,dx)} + \|Df\|_{\mathbb{L}^2(\Upsilon,\mathbb{R}^d,m(x)\,dx)}$ and let $\mathbb{H}^{1,2}(\Upsilon, m(x)\,dx)$ denote the completion of the set of functions from $\mathbb{C}^\infty(\Upsilon)$ having finite $\mathbb{W}^{1,2}(\Upsilon, m(x)\,dx)$-norms with respect to $\|\cdot\|_{\mathbb{W}^{1,2}(\Upsilon,m(x)\,dx)}$. Obviously, $\mathbb{H}^{1,2}(\Upsilon, m(x)\,dx) \subset \mathbb{W}^{1,2}(\Upsilon, m(x)\,dx)$. Spaces $\mathbb{W}^{1,2}(\Upsilon, c(x), m(x)\,dx)$ and $\mathbb{H}^{1,2}(\Upsilon, c(x), m(x)\,dx)$ are defined similarly. We let $\mathbb{L}_0^{1,2}(\Upsilon, \mathbb{R}^d, c(x), m(x)\,dx)$ represent the closure of the set of the gradients of functions from $\mathbb{C}_0^\infty(\Upsilon)$ in $\mathbb{L}^2(\Upsilon, \mathbb{R}^d, c(x), m(x)\,dx)$. The space of continuous functions on $\mathbb{R}_+$ with values in a metric space $\mathbb{S}$ is denoted by $\mathbb{C}(\mathbb{R}_+, \mathbb{S})$. It is endowed with the compact-open topology. If function $X = (X_s, s \in \mathbb{R}_+)$ from $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^d)$ is absolutely continuous with respect to Lebesgue measure, $\dot{X}_s$ denotes its derivative at s. We let $\mathbb{M}(\mathbb{R}^d)$ [resp., $\mathbb{M}_1(\mathbb{R}^d)$] represent the set of finite (resp., probability) measures on $\mathbb{R}^d$ endowed with the weak topology (see, e.g., Topsøe [50]); $\mathbb{P}(\mathbb{R}^d)$ denotes the set of probability densities $m(x)$ on $\mathbb{R}^d$ such that $m \in \mathbb{W}_{\mathrm{loc}}^{1,1}(\mathbb{R}^d)$ and $\sqrt{m} \in \mathbb{W}^{1,2}(\mathbb{R}^d)$. Topological spaces are equipped with Borel σ-algebras, except for $\mathbb{R}_+$ which is equipped with the Lebesgue σ-algebra, products of topological spaces are equipped with product topologies, and products of measurable spaces are equipped with product σ-algebras. The "overbar" notation is reserved for the closures of sets, $\mathbf{1}_\Gamma$ denotes the indicator function of set Γ, $\lfloor a \rfloor$ stands for the integer part of real number a, $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, and $a^+ = a \vee 0$.

Notation $U \subset\subset V$, where U and V are open subsets of $\mathbb{R}^d$, is to signify that the closure of U is a compact subset of V. Throughout, the conventions that $\inf_\varnothing = \infty$ and $0/0 = 0$ are adopted. The terms "absolutely continuous," "a.e.," "almost all" refer to Lebesgue measure unless specified otherwise. All suprema in the time variable are understood as essential suprema with respect to Lebesgue measure.

We say that a net of probability measures $\mathbf{P}^\varepsilon$, where $\varepsilon > 0$, defined on metric space $\mathbb{S}$ obeys the large deviation principle (LDP) with (tight) large deviation (rate) function $\mathbf{I}$ for rate $1/\varepsilon$ as $\varepsilon \to 0$ if $\mathbf{I}$ is a function from $\mathbb{S}$ to $[0, \infty]$ such that the sets $\{z \in \mathbb{S} : \mathbf{I}(z) \le \delta\}$ are compact for all $\delta \in \mathbb{R}_+$, $\liminf_{\varepsilon\to 0} \varepsilon \ln \mathbf{P}^\varepsilon(G) \ge -\inf_{z\in G} \mathbf{I}(z)$ for all open sets $G \subset \mathbb{S}$, and $\limsup_{\varepsilon\to 0} \varepsilon \ln \mathbf{P}^\varepsilon(F) \le -\inf_{z\in F} \mathbf{I}(z)$ for all closed sets $F \subset \mathbb{S}$. We say that the net $\mathbf{P}^\varepsilon$ is exponentially tight for rate $1/\varepsilon$ if $\inf_K \limsup_{\varepsilon\to 0} (\mathbf{P}^\varepsilon(\mathbb{S} \setminus K))^\varepsilon = 0$ where K ranges over the collection of compact subsets of $\mathbb{S}$.

2. Main results.

We will consider a time nonhomogeneous version of (1.1) in which the coefficients may depend on ε as well:

$$dX_t^\varepsilon = A_t^\varepsilon\big(X_t^\varepsilon, x_t^\varepsilon\big)\, dt + \sqrt{\varepsilon} B_t^\varepsilon\big(X_t^\varepsilon, x_t^\varepsilon\big)\, dW_t^\varepsilon, \tag{2.1a}$$

$$dx_t^\varepsilon = \frac{1}{\varepsilon} a_t^\varepsilon\big(X_t^\varepsilon, x_t^\varepsilon\big)\, dt + \frac{1}{\sqrt{\varepsilon}} b_t^\varepsilon\big(X_t^\varepsilon, x_t^\varepsilon\big)\, dW_t^\varepsilon. \tag{2.1b}$$

As above, $A_t^\varepsilon(u, x)$ is an n-vector, $B_t^\varepsilon(u, x)$ is an $n \times k$-matrix, $a_t^\varepsilon(u, x)$ is an l-vector, $b_t^\varepsilon(u, x)$ is an $l \times k$-matrix, and $W^\varepsilon = (W_t^\varepsilon, t \in \mathbb{R}_+)$ is an $\mathbb{R}^k$-valued standard Wiener process. The stochastic process $X^\varepsilon = (X_t^\varepsilon, t \in \mathbb{R}_+)$ takes values in $\mathbb{R}^n$ and the stochastic process $x^\varepsilon = (x_t^\varepsilon, t \in \mathbb{R}_+)$ takes values in $\mathbb{R}^l$. We assume that the functions $A_t^\varepsilon(u, x)$, $a_t^\varepsilon(u, x)$, $B_t^\varepsilon(u, x)$, and $b_t^\varepsilon(u, x)$ are measurable and locally bounded in (t, u, x) and are such that the equations (2.1a) and (2.1b) admit weak solution $(X^\varepsilon, x^\varepsilon)$ with trajectories in $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n \times \mathbb{R}^l)$ for every initial condition $(X_0^\varepsilon, x_0^\varepsilon)$. More specifically, we assume that there exists complete probability space $(\Omega^\varepsilon, \mathcal{F}^\varepsilon, \mathbf{P}^\varepsilon)$ with filtration $\mathbf{F}^\varepsilon = (\mathcal{F}_t^\varepsilon, t \in \mathbb{R}_+)$ such that $(W_t^\varepsilon, t \in \mathbb{R}_+)$ is a Wiener process relative to $\mathbf{F}^\varepsilon$, the processes $X^\varepsilon = (X_t^\varepsilon, t \in \mathbb{R}_+)$ and $x^\varepsilon = (x_t^\varepsilon, t \in \mathbb{R}_+)$ are $\mathbf{F}^\varepsilon$-adapted, have continuous trajectories, and the relations (2.1a) and (2.1b) hold for all $t \in \mathbb{R}_+$ $\mathbf{P}^\varepsilon$-a.s. (To ensure uniqueness which we do not assume apriori, one may require, in addition to the above hypotheses, that the coefficients be Lipschitz continuous.) For background information, see Ethier and Kurtz [15], Ikeda and Watanabe [24], Stroock and Varadhan [49]. We note that since the dimensions n, k, and l are arbitrary, the assumption that both X^ε and x^ε are driven by the same Wiener process does not constitute a loss of generality.

Let us denote $C_t^\varepsilon(u, x) = B_t^\varepsilon(u, x) B_t^\varepsilon(u, x)^T$ and $c_t^\varepsilon(u, x) = b_t^\varepsilon(u, x) b_t^\varepsilon(u, x)^T$. We introduce the boundedness and growth conditions that for all $N > 0$ and $t > 0$

$$\limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n : |u|\le N} \big\| c_s^\varepsilon(u, x) \big\| < \infty, \tag{2.2a}$$

$$\limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n:|u|\le N} \left|A_s^\varepsilon(u,x)\right| < \infty, \tag{2.2b}$$

$$\limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n} \frac{u^T A_s^\varepsilon(u,x)}{1+|u|^2} < \infty \tag{2.2c}$$

and

$$\limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n} \frac{\|C_s^\varepsilon(u,x)\|}{1+|u|^2} < \infty. \tag{2.2d}$$

We also assume as given "limit coefficients": $A_t(u,x)$ is an n-vector, $B_t(u,x)$ is an $n \times k$-matrix, $a_t(u,x)$ is an l-vector, and $b_t(u,x)$ is an $l \times k$-matrix. Let $C_t(u,x) = B_t(u,x)B_t(u,x)^T$ and $c_t(u,x) = b_t(u,x)b_t(u,x)^T$. The following regularity properties will be needed.

CONDITION 2.1. The functions $A_t(u,x)$, $B_t(u,x)$, and $b_t(u,x)$ are measurable and are bounded locally in (t,u) and globally in x and are continuous in (u,x), the function $a_t(u,x)$ is measurable and locally bounded in (t,u,x) and is Lipschitz continuous in x locally uniformly in (t,u), the functions $a_t(u,x)$ and $c_t(u,x)$ are continuous in u locally uniformly in t and uniformly in x, $c_t(u,x)$ is of class $\mathbb{C}^1$ in x, with the first partial derivatives being bounded and Lipschitz continuous in x locally uniformly in (t,u), and $\operatorname{div}_x c_t(u,x)$ is continuous in (u,x).

Another set of regularity requirements is furnished by the next condition. We introduce

$$G_t(u,x) = B_t(u,x)b_t(u,x)^T. \tag{2.3}$$

CONDITION 2.2. The matrix $c_t(u,x)$ is positive definite uniformly in x and locally uniformly in (t,u). Either $C_t(u,x) = 0$ for all (t,u,x) and $A_t(u,x)$ is locally Lipschitz continuous in u locally uniformly in t and uniformly in x, or the matrix $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u).

Finally, certain stability properties will be required: for all $N > 0$ and $t > 0$,

$$\lim_{M\to\infty} \limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l:|x|\ge M} \sup_{u\in\mathbb{R}^n:|u|\le N} a_s^\varepsilon(u,x)^T \frac{x}{|x|} = -\infty \tag{2.4a}$$

and

$$\lim_{|x|\to\infty} \sup_{s\in[0,t]} \sup_{u\in\mathbb{R}^n:|u|\le N} a_s(u,x)^T \frac{x}{|x|} = -\infty. \tag{2.4b}$$

Let $\mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ represent the subset of $\mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ of functions $\mu = (\mu_t, t \in \mathbb{R}_+)$ such that $\mu_t - \mu_s$ is an element of $\mathbb{M}(\mathbb{R}^l)$ for $t \ge s$ and $\mu_t(\mathbb{R}^l) = t$. It is endowed with the subspace topology and is a complete separable metric space, being closed in $\mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$. The stochastic process $\mu^\varepsilon = (\mu_t^\varepsilon, t \in \mathbb{R}_+)$, where

$$\mu_t^\varepsilon(\Theta) = \int_0^t \mathbf{1}_\Theta(x_s^\varepsilon)\, ds,$$

for $\Theta \in \mathcal{B}(\mathbb{R}^l)$, is a random element of $\mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$. We will regard $(X^\varepsilon, \mu^\varepsilon)$ as a random element of $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$. It is worth noting that the elements of $\mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ can be also regarded as σ-finite measures on $\mathbb{R}_+ \times \mathbb{R}^l$. We will then use notation $\mu(dt, dx)$ for μ.

Let Γ represent the set of (X, μ) such that the function $X = (X_s, s \in \mathbb{R}_+)$ from $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}_+$ and the function $\mu = (\mu_s, s \in \mathbb{R}_+)$ from $\mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$, when considered as a measure on $\mathbb{R}_+ \times \mathbb{R}^l$, is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}_+ \times \mathbb{R}^l$, that is, $\mu(ds, dx) = m_s(x)\, dx\, ds$, where $m_s(x)$, as a function of x, belongs to $\mathbb{P}(\mathbb{R}^l)$ for almost all s. Given $(X, \mu) \in \Gamma$, we define

$$\begin{aligned}
\mathbf{I}'(X, \mu) = \int_0^\infty & \sup_{\lambda \in \mathbb{R}^n} \Big(\lambda^T \Big(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s, x) m_s(x)\, dx \Big) \\
& - \frac{1}{2} \|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s, x) m_s(x)\, dx} \\
& + \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \Big(Dh(x)^T \Big(\frac{1}{2} \operatorname{div}_x \big(c_s(X_s, x) m_s(x) \big) \\
& - \big(a_s(X_s, x) + G_s(X_s, x)^T \lambda \big) m_s(x) \Big) \\
& - \frac{1}{2} \| Dh(x) \|^2_{c_s(X_s, x)} m_s(x) \Big)\, dx \Big)\, ds.
\end{aligned} \tag{2.5}$$

We let $\mathbf{I}'(X, \mu) = \infty$ if $(X, \mu) \notin \Gamma$. It follows, on letting $\lambda = 0$, that if $\mathbf{I}'(X, \mu) < \infty$, then

$$\begin{aligned}
\int_0^t \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} & \Big(Dh(x)^T \Big(\frac{1}{2} \operatorname{div}_x \big(c_s(X_s, x) m_s(x) \big) \\
& - a_s(X_s, x) m_s(x) \Big) - \frac{1}{2} \| Dh(x) \|^2_{c_s(X_s, x)} m_s(x) \Big)\, dx\, ds < \infty,
\end{aligned}$$

which is seen to imply [cf. (8.22) below], that for all $\lambda \in \mathbb{R}^n$,

$$\begin{aligned}
\int_0^t \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} & \Big(Dh(x)^T \Big(\frac{1}{2} \operatorname{div}_x \big(c_s(X_s, x) m_s(x) \big) \\
& - \big(a_s(X_s, x) + G_s(X_s, x)^T \lambda \big) m_s(x) \Big) - \frac{1}{2} \| Dh(x) \|^2_{c_s(X_s, x)} m_s(x) \Big)\, dx\, ds \\
& < \infty.
\end{aligned}$$

We introduce the following convergence condition.

CONDITION 2.3. If $\mathbf{I}'(X,\mu)<\infty$, then there exists a nonincreasing $[0,1]$-valued $\mathbb{C}_0^1(\mathbb{R}_+)$-function $\eta(y)$ such that $\eta(y)=1$ for $y\in[0,1]$, $\eta(y)=0$ for $y\ge 2$, and

$$\int_1^2 \frac{|D\eta(y)|^2}{1-\eta(y)}\,dy<\infty, \tag{2.6}$$

and, for arbitrary $t\in\mathbb{R}_+$ and $\lambda\in\mathbb{R}^n$,

$$\begin{aligned}
\lim_{r\to\infty}\int_0^t \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\Big(Dh(x)^T\Big(\frac{1}{2}\operatorname{div}_x\big(c_s(X_s,x)m_s(x)\big)-\big(a_s(X_s,x)\\
+G_s(X_s,x)^T\lambda\big)m_s(x)\Big)-\frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)}m_s(x)\Big)\eta^2\Big(\frac{|x|}{r}\Big)\,dx\,ds\\
=\int_0^t \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\Big(Dh(x)^T\Big(\frac{1}{2}\operatorname{div}_x\big(c_s(X_s,x)m_s(x)\big)\\
-\big(a_s(X_s,x)+G_s(X_s,x)^T\lambda\big)m_s(x)\Big)-\frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)}m_s(x)\Big)\,dx\,ds,
\end{aligned}$$

where $m_s(x)=\mu(ds,dx)/(ds\,dx)$.

We note that (2.6) is satisfied if $\eta(y)=1-e^{-1/(y-1)}$ in a right neighborhood of 1. By Theorem 6.1 below, if $\mathbf{I}'(X,\mu)<\infty$ then $\int_0^t\int_{\mathbb{R}^l}|D_xm_s(x)|^2/m_s(x)\,dx\,ds<\infty$. Therefore, if

$$\int_0^t\int_{\mathbb{R}^l}\big|a_s(X_s,x)\big|^2 m_s(x)\,dx\,ds<\infty, \tag{2.7}$$

then, assuming Condition 2.1 holds, Condition 2.3 is fulfilled. The next lemma, whose proof is relegated to the Appendix, shows that the square integrability in (2.7) holds if one requires that either a stronger version of the stability condition (2.4b) hold or that the drift of the fast process be gradiental.

LEMMA 2.1. *Let Conditions 2.1 and 2.2 hold. Suppose that $\mathbf{I}'(X,\mu)<\infty$. Let either*

$$\limsup_{|x|\to\infty}\sup_{s\in[0,t]} a_s(X_s,x)^T\frac{x}{|x|^2}<0 \tag{2.8}$$

or there exist real-valued function $\hat a_s(x)$ which belongs to $\mathbb{W}^{1,q}_{\text{loc}}(\mathbb{R}^l)$ in x, where $q>2$ and $q\ge l$, such that

$$c_s(X_s,x)^{-1}\big(a_s(X_s,x)-\tfrac{1}{2}\operatorname{div}_x c_s(X_s,x)\big)=D_x\hat a_s(x). \tag{2.9}$$

Then (2.7) holds, where $m_s(x)=\mu(ds,dx)/(dx\,ds)$.

We state the main result.

THEOREM 2.1. *Let* (2.2a)–(2.2d), (2.4a), (2.4b) *and Conditions* 2.1, 2.2 *and* 2.3 *hold. If the net* X_0^ε *obeys the LDP in* $\mathbb{R}^n$ *with large deviation function* $\mathbf{I}_0$ *for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$, *the net* x_0^ε *is exponentially tight in* $\mathbb{R}^l$ *for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$, *and, for all* $t > 0$ *and* $N > 0$, *the convergences*

$$\lim_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l:|x|\le N} \sup_{u\in\mathbb{R}^n:|u|\le N} \Big(\big|A_s^\varepsilon(u,x) - A_s(u,x)\big| + \big|a_s^\varepsilon(u,x) - a_s(u,x)\big| + \big\|B_s^\varepsilon(u,x) - B_s(u,x)\big\| + \big\|b_s^\varepsilon(u,x) - b_s(u,x)\big\|\Big) = 0 \tag{2.10}$$

hold, then the net $(X^\varepsilon, \mu^\varepsilon)$ *obeys the LDP in* $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ *for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$ *with large deviation function* $\mathbf{I}$ *defined as follows*:

$$\mathbf{I}(X,\mu) = \begin{cases} \mathbf{I}_0(X_0) + \mathbf{I}'(X,\mu), & \text{if } (X,\mu) \in \Gamma, \\ \infty, & \text{otherwise}. \end{cases}$$

REMARK 2.1. Condition 2.3 may be superfluous as far as the validity of Theorem 2.1 is concerned. It is used at the final stage of the proof only; see Theorem 8.1.

REMARK 2.2. By Lemma 6.7 below, $\mathbf{I}(X,\mu) = 0$ provided that a.e.

$$\dot{X}_s = \int_{\mathbb{R}^l} A_s(X_s, x) m_s(x)\, dx$$

and

$$\int_{\mathbb{R}^l} \left(\frac{1}{2} \operatorname{tr}\big(c_s(X_s,x) D^2 p(x)\big) + Dp(x)^T a_s(X_s,x)\right) m_s(x)\, dx = 0,$$

where the latter equation holds for all $p \in \mathbb{C}_0^\infty(\mathbb{R}^l)$ and X_0 satisfies the equality $\mathbf{I}_0(X_0) = 0$. Consequently, $m_s(\cdot)$ is the invariant density of the diffusion process with the infinitesimal drift $a_s(X_s, \cdot)$ and diffusion matrix $c_s(X_s, \cdot)$.

REMARK 2.3. Conditions 2.1 and (2.10) imply that

$$\limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l:|x|\le N} \sup_{u\in\mathbb{R}^n:|u|\le N} \big|a_s^\varepsilon(u,x)\big| < \infty. \tag{2.11}$$

Conditions (2.2a)–(2.2d) and (2.10) imply that

$$\sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n:|u|\le N} \big\|c_s(u,x)\big\| < \infty, \tag{2.12a}$$

$$\sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n:|u|\le N} \big|A_s(u,x)\big| < \infty, \tag{2.12b}$$

$$\sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n} \frac{u^T A_s(u,x)}{1+|u|^2} < \infty \tag{2.12c}$$

and

$$\sup_{s\in[0,t]}\sup_{x\in\mathbb{R}^l}\sup_{u\in\mathbb{R}^n}\frac{\|C_s(u,x)\|}{1+|u|^2}<\infty. \tag{2.12d}$$

In particular, some of the boundedness requirements in Condition 2.1 are consequences of the other hypotheses of Theorem 2.1.

REMARK 2.4. If the matrices $c_t(u,x)$ and $C_t(u,x)$ are positive definite uniformly in x and locally uniformly in (t,u), then since $b_t(u,x)^T c_t(u,x)^{-1} b_t(u,x)$ is the orthogonal projection operator onto the range of $b_t(u,x)^T$, the condition that $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ be positive definite uniformly in x and locally uniformly in (t,u) is implied by the following "angle condition": for any bounded region of (t,u), there exists $\ell \in (0,1)$ such that $|y_1^T y_2| \le \ell|y_1||y_2|$ for all y_1 and y_2 from the ranges of $B_t(u,x)^T$ and $b_t(u,x)^T$, respectively, where x is arbitrary and (t,u) belongs to the region. To put it another way, the condition requires that the angles between the elements of the range of $B_t(u,x)^T$, on the one hand, and the elements of the range of $b_t(u,x)^T$, on the other hand, be bounded away from zero uniformly in x and locally uniformly in (t,u). It ensures that the processes X^ε and x^ε are "sufficiently random" in relation to each other. Under that condition, the ranges of $B_t(u,x)^T$ and $b_t(u,x)^T$ do not have common nontrivial subspaces and $k \ge n+l$. On the other hand, if $\|C_t(u,x)\|$ is bounded uniformly in x and locally uniformly in (t,u), as is the case under the hypotheses of Theorem 2.1 according to (2.12d), the converse is also true: if $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u), then the angle condition holds.

The solution of the variational problem in (2.5) plays an important part in the proof below, so we proceed with describing it. Let $c(x)$ represent a measurable function defined for $x \in \mathbb{R}^d$ and taking values in the space of positive definite symmetric $d \times d$-matrices, let $m(x)$ represent a probability density on $\mathbb{R}^d$, and let S_i represent an open ball of radius i centered at the origin in $\mathbb{R}^d$, where $d \in \mathbb{N}$ and $i \in \mathbb{N}$. For function $\psi_j \in \mathbb{L}_0^{1,2}(S_j, \mathbb{R}^d, c(x), m(x)\,dx)$ and $j \ge i$, where $j \in \mathbb{N}$, we let $\pi_{ji}\psi_j$ denote the orthogonal projection of the restriction of ψ_j to S_i onto $\mathbb{L}_0^{1,2}(S_i, \mathbb{R}^d, c(x), m(x)\,dx)$ in $\mathbb{L}^2(S_i, \mathbb{R}^d, c(x), m(x)\,dx)$. Thus, the function $\pi_{ji}\psi_j$ is the element of $\mathbb{L}_0^{1,2}(S_i, \mathbb{R}^d, c(x), m(x)\,dx)$ such that $\int_{S_i} Dp(x)^T c(x)\pi_{ji}\psi_j(x)m(x)\,dx = \int_{S_i} Dp(x)^T c(x)\psi_j(x)m(x)\,dx$ for all $p \in \mathbb{C}_0^\infty(S_i)$. We note that if the density $m(x)$ is locally bounded away from zero, then $\pi_{ji}\psi_j$ is a certain gradient: $\pi_{ji}\psi_j = D\chi_{ji}$, where χ_{ji} is the weak solution of the Dirichlet problem $\operatorname{div}(c(x)m(x)D\chi_{ji}(x)) = \operatorname{div}(c(x)m(x)\psi_j(x))$ for $x \in S_i$ with a zero boundary condition (cf. the proof of Lemma 7.3). Since, for $i \le j \le k$, $\pi_{ji} \circ \pi_{kj} = \pi_{ki}$, the family $(\mathbb{L}_0^{1,2}(S_j, \mathbb{R}^d, c(x), m(x)\,dx), \pi_{ji})$ is a projective (or inverse) system in the category of sets. Given a function

$\phi \in \mathbb{L}^2_{\text{loc}}(\mathbb{R}^d, \mathbb{R}^d, c(x), m(x)\,dx)$, the orthogonal projections ϕ_i of the restrictions of ϕ to S_i onto $\mathbb{L}_0^{1,2}(S_i, \mathbb{R}^d, c(x), m(x)\,dx)$ in $\mathbb{L}^2(S_i, \mathbb{R}^d, c(x), m(x)\,dx)$ are such that $\pi_{ji}\phi_j = \phi_i$, provided $i \le j$, so they specify an element of the projective (or inverse) limit of $(\mathbb{L}_0^{1,2}(S_j, \mathbb{R}^d, c(x), m(x)\,dx), \pi_{ji})$, which we denote by $\Pi_{c(\cdot),m(\cdot)}\phi$. On extending the ϕ_i by zero outside of S_i, one has that, for $i \le j$, $\|\phi_j\|^2_{c(\cdot),m(\cdot)} - \|\phi_i\|^2_{c(\cdot),m(\cdot)} = \|\phi_j - \phi_i\|^2_{c(\cdot),m(\cdot)}$, where the norms are taken in $\mathbb{R}^d$. Hence, if $\lim_{i\to\infty} \|\phi_i\|^2_{c(\cdot),m(\cdot)} < \infty$, then the sequence ϕ_i converges in $\mathbb{L}^2(\mathbb{R}^d, \mathbb{R}^d, c(x), m(x)\,dx)$ as $i \to \infty$ and one can identify $\Pi_{c(\cdot),m(\cdot)}\phi$ with the limit, so $\Pi_{c(\cdot),m(\cdot)}\phi \in \mathbb{L}^2(\mathbb{R}^d, \mathbb{R}^d, c(x), m(x)\,dx)$. It is uniquely specified by the requirements that $\Pi_{c(\cdot),m(\cdot)}\phi \in \mathbb{L}_0^{1,2}(\mathbb{R}^d, \mathbb{R}^d, c(x), m(x)\,dx)$ and that, for all $p \in \mathbb{C}_0^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} Dp(x)^T c(x) \Pi_{c(\cdot),m(\cdot)}\phi(x) m(x)\,dx = \int_{\mathbb{R}^d} Dp(x)^T c(x)\phi(x) m(x)\,dx. \tag{2.13}$$

In particular, if ϕ is an element of $\mathbb{L}^2(\mathbb{R}^d, \mathbb{R}^d, c(x), m(x)\,dx)$, then $\Pi_{c(\cdot),m(\cdot)}\phi$ is the orthogonal projection of ϕ onto $\mathbb{L}_0^{1,2}(\mathbb{R}^d, \mathbb{R}^d, c(x), m(x)\,dx)$. For results on the existence and uniqueness for equation (2.13) when $\Pi_{c(\cdot),m(\cdot)}\phi$ is a gradient; see Pardoux and Veretennikov [35].

In the setting of Theorem 2.1, $d = l$. Since, under the hypotheses of Theorem 2.1, the matrix functions $c_t(u,\cdot)^{-1} G_t(u,\cdot)^T$ are bounded, the matrix function $\Pi_{c_t(u,\cdot),m(\cdot)}(c_t(u,\cdot)^{-1} G_t(u,\cdot)^T)$, whose columns are the projections of the n columns of $c_t(u,\cdot)^{-1} G_t(u,\cdot)^T$ onto the space $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_t(x), m(x)\,dx)$, is a well defined element of the space $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^{l\times n}, c_t(u,x), m(x)\,dx)$ and we denote it by $\Psi_{t,m(\cdot),u}$. We also define

$$Q_{t,m(\cdot)}(u,x) = C_t(u,x) - \|\Psi_{t,m(\cdot),u}(x)\|^2_{c_t(u,x)}. \tag{2.14}$$

The function $Q_{t,m(\cdot)}(u,x)$ assumes values in the space of positive semi-definite $n \times n$-matrices. If the matrix $C_t(u,x) - G_t(u,x) c_t(u,x)^{-1} G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u), then the matrix $\int_{\mathbb{R}^l} Q_{t,m(\cdot)}(u,x) m(x)\,dx$ is positive definite locally uniformly in (t,u). We also introduce $\Phi_{t,m(\cdot),u} = \Pi_{c_t(u,\cdot),m(\cdot)}(c_t(u,\cdot)^{-1}(a_t(u,\cdot) - \operatorname{div}_x c_t(u,\cdot)/2))$. Since $a_t(u,\cdot)$ is not necessarily square integrable with respect to $m(x)\,dx$, the function $\Pi_{c_t(u,\cdot),m(\cdot)}(c_t(u,\cdot)^{-1}(a_t(u,\cdot) - \operatorname{div}_x c_t(u,\cdot)/2))$, as a function of $x \in \mathbb{R}^l$, might not be an element of $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, c_t(u,x), m(x)\,dx)$.

For future reference, we note that, according to (2.13), a.e.,

$$\begin{aligned} &\int_{\mathbb{R}^l} Dp(x)^T c_s(u,x) \Psi_{s,m(\cdot),u}(x) m(x)\,dx \\ &\quad = \int_{\mathbb{R}^l} Dp(x)^T G_s(u,x)^T m(x)\,dx \end{aligned} \tag{2.15a}$$

and, provided $\Phi_{s,m(\cdot),u} \in \mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, c_s(u,x), m(x)\,dx)$,

$$
\begin{aligned}
&\int_{\mathbb{R}^l} Dp(x)^T c_s(u,x)\Phi_{s,m(\cdot),u}(x)m(x)\,dx \\
&\qquad = \int_{\mathbb{R}^l} Dp(x)^T \Bigl(a_s(u,x) - \frac{1}{2}\operatorname{div}_x c_s(u,x)\Bigr) m(x)\,dx,
\end{aligned} \tag{2.15b}
$$

for all $p \in \mathbb{C}_0^\infty(\mathbb{R}^l)$. In addition, (2.15a) extends to Dp representing an arbitrary element of $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(u,x), m(x)\,dx)$. A similar extension property holds for (2.15b), provided $a_s(u,\cdot) \in \mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, c_s(u,x), m(x)\,dx)$.

PROPOSITION 2.1. *If, under the hypotheses of Theorem 2.1, $\mathbf{I}'(X,\mu) < \infty$, then $\Phi_{s,m_s(\cdot),X_s}$ belongs to the space $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$ for almost all s and*

$$
\begin{aligned}
&\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx \\
&\qquad - \int_{\mathbb{R}^l} G_s(X_s,x)\Bigl(\frac{D_x m_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\Bigr)m_s(x)\,dx
\end{aligned}
$$

belongs to the range of $\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx$ for almost all s. Furthermore, $\Phi_{s,m_s(\cdot),X_s}(x)$ and $\Psi_{s,m_s(\cdot),X_s}(x)$ are measurable in (s,x) so that in the statement of Theorem 2.1,

$$
\begin{aligned}
\mathbf{I}(X,\mu) &= \mathbf{I}_0(X_0) + \frac{1}{2}\int_0^\infty \biggl(\int_{\mathbb{R}^l} \Bigl\| \frac{D_x m_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\Bigr\|^2_{c_s(X_s,x)} m_s(x)\,dx \\
&\quad + \Bigl\| \dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx - \int_{\mathbb{R}^l} G_s(X_s,x)\Bigl(\frac{D_x m_s(x)}{2m_s(x)} \\
&\quad - \Phi_{s,m_s(\cdot),X_s}(x)\Bigr)m_s(x)\,dx\Bigr\|^2_{(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx)^\oplus}\biggr)\,ds.
\end{aligned} \tag{2.16}
$$

REMARK 2.5. If $m(x)$ is an element of $\mathbb{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^l)$, then $Dm(x) = 0$ for almost all x on the set where $m(x) = 0$, so we will assume throughout that $Dm(x)/m(x) = 0$ a.e. on that set.

REMARK 2.6. The expression on the right-hand side of (2.16) serves both the case where $C_t(u,x) = 0$ for all (t,u,x) and $A_t(u,x)$ is locally Lipschitz continuous in u locally uniformly in t and uniformly in x, and the case where $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u). In each of the two cases, however, it simplifies as follows. If $C_t(u,x) = 0$ for all (t,u,x) and $A_t(u,x)$ is locally Lipschitz continuous in u

locally uniformly in t and uniformly in x, then $Q_{s,m_s(\cdot)}(u,x)=0$ for all (s,u,x), so, in order for $\mathbf{I}(X,\mu)$ to be finite, it is necessary that, a.e.,

$$\dot{X}_s = \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx$$

so that

$$\begin{aligned}\mathbf{I}(X,\mu) &= \mathbf{I}_0(X_0)\\ &\quad + \frac{1}{2}\int_0^\infty \int_{\mathbb{R}^l} \left\| \frac{D_x m_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\right\|^2_{c_s(X_s,x)} m_s(x)\,dx\,ds.\end{aligned} \tag{2.17}$$

If $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u), then the matrix $\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx$ is invertible, so its pseudo-inverse is the same as the inverse and the range condition in the statement of Proposition 2.1 is superfluous.

REMARK 2.7. By Theorem 6.1, in order for $\mathbf{I}(X,\mu)$ to be finite it is necessary that $\int_0^t\int_{\mathbb{R}^l}(|D_x m_s(x)|^2/m_s(x) + |\Phi_{s,m_s(\cdot),X_s}(x)|^2)\,dx\,ds<\infty$ for all $t\in\mathbb{R}_+$.

REMARK 2.8. The large deviation function in (2.16) can also be written as

$$\mathbf{I}(X,\mu) = \mathbf{I}_0(X_0) + \frac{1}{2}\int_0^\infty\int_{\mathbb{R}^l}\left|B_s(X_s,x)^T\hat{\lambda}_s + b_s(X_s,x)^T\hat{g}_s(x)\right|^2 m_s(x)\,dx\,ds,$$

where the pair $(\hat{\lambda}_s,\hat{g}_s(x))$ attains the supremum in (2.5), with $\hat{g}_s$ assuming the role of Dh:

$$\begin{aligned}\hat{\lambda}_s = \left(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx\right)^{\oplus}\biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx\\ - \int_{\mathbb{R}^l} G_s(X_s,x)\left(\frac{D_x m_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\right)m_s(x)\,dx\biggr)\end{aligned}$$

and

$$\hat{g}_s(x) = \frac{D_x m_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x) - \Psi_{s,m_s(\cdot),X_s}(x)\hat{\lambda}_s.$$

In the symmetric case where $c_t(u,x)^{-1}(2a_t(u,x) - \operatorname{div}_x c_t(u,x)) = D_x\hat{m}_t(u,x)/\hat{m}_t(u,x)$, for some positive probability density $\hat{m}_t(u,\cdot)$ from $\mathbb{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^l)$, one can identify $\Phi_{t,m_t(\cdot),u}$ with $D_x\hat{m}_t(u,\cdot)/(2\hat{m}_t(u,\cdot))$. [We note that the diffusion process with the infinitesimal drift coefficient $a_t(u,\cdot)$ and diffusion matrix $c_t(u,\cdot)$ has $\hat{m}_t(u,\cdot)$ as an invariant density.] One can then write the large deviation function in (2.17) by using a Dirichlet form:

$$\mathbf{I}(X,\mu) = \mathbf{I}_0(X_0) + \frac{1}{2}\int_0^\infty\int_{\mathbb{R}^l}\left\|D_x\sqrt{\frac{m_s(x)}{\hat{m}_s(X_s,x)}}\right\|^2_{c_s(X_s,x)}\hat{m}_s(X_s,x)\,dx\,ds,$$

provided $D_x\hat{m}_t(u,\cdot)/\hat{m}_t(u,\cdot)\in\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,c_t(x),m_t(x)\,dx)$.

Let us look at a one-dimensional example:

$$dX_t^\varepsilon = A_t(X_t^\varepsilon, x_t^\varepsilon)\,dt + \sqrt{\varepsilon} B_t(X_t^\varepsilon, x_t^\varepsilon)\,dW_{1,t}^\varepsilon,$$
$$dx_t^\varepsilon = \frac{1}{\varepsilon} a_t(X_t^\varepsilon, x_t^\varepsilon)\,dt + \frac{1}{\sqrt{\varepsilon}} b_t(X_t^\varepsilon, x_t^\varepsilon)\,dW_{2,t}^\varepsilon,$$

where all coefficients are scalars and $W_{1,t}^\varepsilon$ and $W_{2,t}^\varepsilon$ are one-dimensional standard Wiener processes. Assuming that $\mathbf{E}^\varepsilon(W_{1,t}^\varepsilon W_{2,t}^\varepsilon) = \rho t$, where $|\rho| < 1$, this setup can be cast as (2.1a) and (2.1b) with $W_t^\varepsilon = (W_{1,t}^\varepsilon, W_{3,t}^\varepsilon)^T$, $B_t^\varepsilon(u,x) = (B_t(u,x), 0)$, and $b_t^\varepsilon(u,x) = (\rho b_t(u,x), \sqrt{1-\rho^2} b_t(u,x))$, where $W_{3,t}^\varepsilon$ represents a standard one-dimensional Wiener process that is independent of $W_{1,t}^\varepsilon$. If $B_t(u,x)$ is bounded away from zero, the large deviation function in (2.16) takes the form

$$\begin{aligned}
&\mathbf{I}(X,\mu)\\
&\quad = \mathbf{I}_0(X_0) + \int_0^\infty \biggl(\frac{1}{8}\int_{\mathbb{R}} \biggl|\frac{D_x m_s(x)}{m_s(x)} - \frac{D_x \hat{m}_s(X_s,x)}{\hat{m}_s(X_s,x)}\biggr|^2 b_s(X_s,x)^2 m_s(x)\,dx\\
&\qquad + \frac{1}{2(1-\rho^2)} \frac{1}{\int_{\mathbb{R}} B_s(X_s,x)^2 m_s(x)\,dx} \biggl|\dot{X}_s - \int_{\mathbb{R}} A_s(X_s,x) m_s(x)\,dx\\
&\qquad - \frac{\rho}{2}\int_{\mathbb{R}} B_s(X_s,x) b_s(X_s,x)\biggl(\frac{D_x m_s(x)}{m_s(x)} - \frac{D_x \hat{m}_s(X_s,x)}{\hat{m}_s(X_s,x)}\biggr) m_s(x)\,dx\biggr|^2\biggr)\,ds.
\end{aligned}$$

If $B_t(u,x) = 0$, then according to (2.17),

$$\begin{aligned}
\mathbf{I}(X,\mu) &= \mathbf{I}_0(X_0)\\
&\quad + \frac{1}{8}\int_0^\infty \int_{\mathbb{R}} \biggl|\frac{D_x m_s(x)}{m_s(x)} - \frac{D_x \hat{m}_s(X_s,x)}{\hat{m}_s(X_s,x)}\biggr|^2 b_s(X_s,x)^2 m_s(x)\,dx\,ds,
\end{aligned}$$

provided $\dot{X}_s = \int_{\mathbb{R}^l} A_s(X_s,x) m_s(x)\,dx$ a.e. For the special case that $A_s(u,x)$ and $B_s(u,x)$ do not depend on s, $a_s(u,x)$ and $b_s(u,x)$ do not depend on either s or u, and $\rho = 0$, this large deviation function appears in Liptser [28].

We now project to obtain an LDP for X^ε. The device of Lemma 6.5 and the minimax theorem (see, e.g., Theorem 7 on page 319 in Aubin and Ekeland [3]) yield the following expression for $\inf_\mu \mathbf{I}(X,\mu)$.

COROLLARY 2.1. *Under the hypotheses of Theorem 2.1, the net X^ε obeys the LDP in $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n)$ for rate $1/\varepsilon$ as $\varepsilon \to 0$ with large deviation function $\mathbf{I}^X$ defined as follows. If function $X = (X_s, s \in \mathbb{R}_+)$ from $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}_+$, then*

$$\begin{aligned}
\mathbf{I}^X(X) = \mathbf{I}_0(X_0) + \int_0^\infty \sup_{\lambda\in\mathbb{R}^n}\biggl(\lambda^T \dot{X}_s - \sup_{m\in\mathbb{P}(\mathbb{R}^l)}\biggl(\lambda^T\int_{\mathbb{R}^l} A_s(X_s,x) m(x)\,dx\\
+ \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x) m(x)\,dx}
\end{aligned}$$

$$- \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \biggl(Dh(x)^T \biggl(\frac{1}{2} \operatorname{div}_x \bigl(c_s(X_s, x) m(x)\bigr)$$

$$- \bigl(a_s(X_s, x) + G_s(X_s, x)^T \lambda\bigr) m(x) \biggr)$$

$$- \frac{1}{2} \| Dh(x)\|^2_{c_s(X_s,x)} m(x) \biggr) dx \biggr) \biggr) ds.$$

Otherwise, $\mathbf{I}^X(X) = \infty$.

If X^ε is decoupled from x^ε, that is, $A_t(u, x)$ and $B_t(u, x)$ do not depend on x, then Corollary 2.1 yields the LDP for Itô processes with small diffusions (cf. Freĭdlin and Wentzell [18]): with $A_s(u, x) = A_s(u)$, $B_s(u, x) = B_s(u)$, and $C_s(u) = B_s(u) B_s(u)^T$,

$$\mathbf{I}^X(X) = \mathbf{I}_0(X_0) + \int_0^\infty \frac{1}{2} \| \dot{X}_s - A_s(X_s) \|^2_{C_s(X_s)^\oplus} ds,$$

provided $\dot{X}_s - A_s(X_s)$ belongs to the range of $C_s(X_s)$ a.e. and $\mathbf{I}^X(X) = \infty$, otherwise.

If one projects the LDP of Theorem 2.1 on the second variable, then an LDP for μ^ε is obtained. In particular, if x^ε is decoupled from X^ε so that $a_t(u, x)$ and $b_t(u, x)$ do not depend on u, we have the following results on the large deviations of the empirical processes and empirical measures of diffusion processes.

COROLLARY 2.2. *Suppose that*

$$d\tilde{x}_t^\varepsilon = \frac{1}{\varepsilon} \tilde{a}_t^\varepsilon(\tilde{x}_t^\varepsilon)\, dt + \frac{1}{\sqrt{\varepsilon}} \tilde{b}_t^\varepsilon(\tilde{x}_t^\varepsilon)\, d\tilde{W}_t^\varepsilon,$$

where $\tilde{x}_t^\varepsilon \in \mathbb{R}^l$, $\tilde{a}_t^\varepsilon(x) \in \mathbb{R}^l$, $\tilde{b}_t^\varepsilon(x) \in \mathbb{R}^{l\times k}$, *and* $\tilde{W}_t^\varepsilon \in \mathbb{R}^k$, *with the coefficients being locally bounded. Assume that, for all* $t \in \mathbb{R}_+$,

$$\limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \| \tilde{b}_s^\varepsilon(x) \tilde{b}_s^\varepsilon(x)^T \| < \infty,$$

$$\lim_{M\to\infty} \limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l:|x|\ge M} \tilde{a}_s^\varepsilon(x)^T \frac{x}{|x|} = -\infty.$$

If, for all $t \in \mathbb{R}_+$ *and all* $N \in \mathbb{R}_+$,

$$\lim_{\varepsilon\to 0} \sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l:|x|\le N} \bigl(|\tilde{a}_s^\varepsilon(x) - \tilde{a}_s(x)| + \| \tilde{b}_s^\varepsilon(x) - \tilde{b}_s(x) \| \bigr) = 0,$$

the matrix $\tilde{c}_t(x) = \tilde{b}_t(x) \tilde{b}_t(x)^T$ *is positive definite uniformly in* x *and locally uniformly in* t, *is of class* $\mathbb{C}^1$ *in* x, *with the first partial derivatives being Lipschitz continuous and bounded in* x *locally uniformly in* t, $\tilde{a}_t(x)$ *is Lipschitz continuous in* x *locally uniformly in* t, $\sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \|\tilde{c}_s(x)\| < \infty$,

$\limsup_{|x|\to\infty}\sup_{s\in[0,t]}\tilde a_s(x)^T x/|x|^2<0$ *for all* $t\in\mathbb{R}_+$, *and the net* x_0^ε *is exponentially tight in* $\mathbb{R}^l$ *for rate* $1/\varepsilon$ *as* $\varepsilon\to 0$, *then the net* $\tilde\mu^\varepsilon$, *where* $\tilde\mu_t^\varepsilon(dx)=\int_0^t \mathbf{1}_{dx}(\tilde x_s^\varepsilon)\,ds$, *obeys the LDP in* $\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$ *for rate* $1/\varepsilon$ *as* $\varepsilon\to 0$ *with large deviation function* $\mathbf{J}$ *defined as follows.*

If function $\mu=(\mu_s,\,s\in\mathbb{R}_+)$ *from* $\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$, *when considered as a measure on* $\mathbb{R}_+\times\mathbb{R}^l$, *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{R}_+\times\mathbb{R}^l$, *that is,* $\mu(ds,dx)=m_s(x)\,dx\,ds$, $m_s(x)$, *as a function of* x, *belongs to* $\mathbb{P}(\mathbb{R}^l)$ *for almost all* s, *and* $\tilde\Phi_{s,m_s(\cdot)}$, *which represents* $\Pi_{\tilde c_s(\cdot),m_s(\cdot)}(\tilde c_s(\cdot)^{-1}(\tilde a_s(\cdot)-\operatorname{div}_x\tilde c_s(\cdot)/2))$, *is an element of* $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,\tilde c_s(x),m_s(x)\,dx)$ *for almost all* s, *then*

$$\begin{aligned}\mathbf{J}(\mu)&=\int_0^\infty \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\biggl(Dh(x)^T\biggl(\frac12\operatorname{div}_x\bigl(\tilde c_s(x)m_s(x)\bigr)-\tilde a_s(x)m_s(x)\biggr)\\&\qquad-\frac12\|Dh(x)\|^2_{\tilde c_s(x)}m_s(x)\biggr)\,dx\,ds\\&=\frac12\int_0^\infty\int_{\mathbb{R}^l}\biggl\|\frac{D_x m_s(x)}{2m_s(x)}-\tilde\Phi_{s,m_s(\cdot)}(x)\biggr\|^2_{\tilde c_s(x)}m_s(x)\,dx\,ds.\end{aligned}$$

Otherwise, $\mathbf{J}(\mu)=\infty$.

COROLLARY 2.3. *Suppose that*

$$dY_t=\breve a(Y_t)\,dt+\breve b(Y_t)\,d\breve W_t,\qquad Y_0=0,$$

where $Y_t\in\mathbb{R}^l$, $\breve a(x)\in\mathbb{R}^l$, $\breve b(x)\in\mathbb{R}^{l\times k}$, *and* $\breve W_t\in\mathbb{R}^k$, *with the coefficients being locally bounded.*

If the matrix $\breve c(x)=\breve b(x)\breve b(x)^T$ *is uniformly positive definite,* $\|\breve c(x)\|$ *is bounded,* $\breve c(\cdot)\in\mathbb{C}^1(\mathbb{R}^l,\mathbb{R}^{l\times l})$, *with Lipschitz continuous bounded first partial derivatives,* $\breve a(\cdot)$ *is Lipschitz continuous, and* $\limsup_{|x|\to\infty}\breve a(x)^Tx/|x|^2<0$, *then the empirical measures* $(1/t)\int_0^t\mathbf{1}_{dx}(Y_s)\,ds$ *obey the LDP in* $\mathbb{M}_1(\mathbb{R}^l)$ *for rate* t *as* $t\to\infty$ *with the large deviation function*

$$\begin{aligned}\breve{\mathbf{J}}(\mu)&=\sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\biggl(Dh(x)^T\biggl(\frac12\operatorname{div}\bigl(\breve c(x)m(x)\bigr)-\breve a(x)m(x)\biggr)\\&\qquad-\frac12\|Dh(x)\|^2_{\breve c(x)}m(x)\biggr)\,dx\\&=\frac12\int_{\mathbb{R}^l}\biggl\|\frac{Dm(x)}{2m(x)}-\breve\Phi_{m(\cdot)}(x)\biggr\|^2_{\breve c(x)}m(x)\,dx\end{aligned}$$

provided probability measure μ *on* $\mathbb{R}^l$ *has density* m, *which is an element of* $\mathbb{P}(\mathbb{R}^l)$, *and* $\breve\Phi_{m(\cdot)}=\Pi_{\breve c(\cdot),m(\cdot)}(\breve c(\cdot)^{-1}(\breve a(\cdot)-\operatorname{div}\breve c(\cdot)/2))$ *is an element of* $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,\breve c(x),m(x)\,dx)$. *Otherwise,* $\breve{\mathbf{J}}(\mu)=\infty$.

In order to derive Corollary 2.3 from Corollary 2.2, one takes $\varepsilon = 1/t$ and defines $\tilde{x}_s^\varepsilon = Y_{st}$.

One can thus write the large deviation function of Theorem 2.1 as

$$\begin{aligned}\mathbf{I}(X, \mu) = \mathbf{I}_0(X_0) + \int_0^\infty \sup_{\lambda \in \mathbb{R}^n} \biggl(\lambda^T \biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s, x)\nu_s(dx)\biggr) \\ - \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x)\nu_s(dx)} + \mathbf{J}^{s,X_s,\lambda}(\nu_s)\biggr)\, ds,\end{aligned} \tag{2.18}$$

where $\nu_s(dx) = m_s(x)\, dx$, and the large deviation function of Corollary 2.1 as

$$\begin{aligned}\mathbf{I}^X(X) = \mathbf{I}_0(X_0) + \int_0^\infty \sup_{\lambda \in \mathbb{R}^n} \biggl(\lambda^T \dot{X}_s - \sup_{\nu \in \mathbb{M}_1(\mathbb{R}^l)} \biggl(\lambda^T \int_{\mathbb{R}^l} A_s(X_s, x)\nu(dx) \\ + \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x)\nu(dx)} - \mathbf{J}^{s,X_s,\lambda}(\nu)\biggr)\biggr)\, ds,\end{aligned} \tag{2.19}$$

where $\mathbf{J}^{s,u,\lambda}$ represents the large deviation function for the empirical measures $\nu_t^{s,u,\lambda}(dx) = (1/t)\int_0^t \mathbf{1}_{dx}(y_r^{s,u,\lambda})\, dr$ for rate t as $t \to \infty$ and

$$dy_t^{s,u,\lambda} = \bigl(a_s(u, y_t^{s,u,\lambda}) + G_s(u, y_t^{s,u,\lambda})^T \lambda\bigr)\, dt + b_s(u, y_t^{s,u,\lambda})\, dw_t, \qquad y_0^{s,u,\lambda} = 0,$$

(w_t) being a k-dimensional standard Wiener process. In particular, if $G_t(u, x) = 0$ so that the diffusions driving the slow and the fast processes are virtually uncorrelated, then $\mathbf{J}^{s,u,\lambda}$ does not depend on λ and by Corollaries 2.2, 2.3 and (2.18) the large deviation function $\mathbf{I}(X, \mu)$ is the sum of the large deviation function of the slow process, with the coefficients being averaged over the "current" empirical measure of the fast variable, and of the large deviation function of the empirical process of the fast variable, with the coefficients "frozen" at the current value of the slow variable.

The first results on large deviation asymptotics for the system (1.1) in the setup of the averaging principle available in the literature appear in Freĭdlin [17]; see also the exposition in Freĭdlin and Wentzell [18], Section 9 of Chapter 7. Freĭdlin [17] considers the equations

$$\begin{aligned}\dot{x}_t^\varepsilon &= b(x_t^\varepsilon, y_t^\varepsilon), \\ \dot{y}_t^\varepsilon &= \frac{1}{\varepsilon}[B(x_t^\varepsilon, y_t^\varepsilon) + g(y_t^\varepsilon)] + \frac{1}{\sqrt{\varepsilon}} c(y_t^\varepsilon)\dot{w}_t.\end{aligned}$$

It is assumed that the state space is a compact manifold. A noncompact setting is considered by Veretennikov [53]. Veretennikov [51, 57, 58] allows the diffusion coefficient in the fast process to depend on both variables:

$$\begin{aligned}dX_t^\varepsilon &= f(X_t^\varepsilon, Y_t^\varepsilon)\, dt, \\ dY_t^\varepsilon &= \varepsilon^{-2} B(X_t^\varepsilon, Y_t^\varepsilon)\, dt + \varepsilon^{-1} C(X_t^\varepsilon, Y_t^\varepsilon)\, dW_t.\end{aligned}$$

The state space of the fast process is a compact manifold.

Veretennikov [55, 56, 59] tackles the case where the slow process has a small diffusion term and the state space of the fast process may be noncompact but the diffusion coefficient in the equation for the fast process does not depend on the slow process so that

$$
\begin{aligned}
dX_t^\varepsilon &= f\big(X_t^\varepsilon, Y_t^\varepsilon\big)\,dt + \varepsilon\big(\sigma_1\big(X_t^\varepsilon, Y_t^\varepsilon\big)\,dW_t^1 + \sigma_3\big(X_t^\varepsilon, Y_t^\varepsilon\big)\,dW_t^3\big),\\
dY_t^\varepsilon &= \varepsilon^{-2}B\big(X_t^\varepsilon, Y_t^\varepsilon\big)\,dt + \varepsilon^{-1}\big(C_1\big(Y_t^\varepsilon\big)\,dW_t^1 + C_2\big(Y_t^\varepsilon\big)\,dW_t^2\big),
\end{aligned}
\tag{2.20}
$$

where the Wiener processes are independent. The stability condition on the slow process is similar to (2.4a) and (2.4b).

In those papers, results on the LDP for the slow processes are obtained in the space of continuous functions on the $[0, L]$ interval endowed with uniform norm, where $L > 0$. The large deviation rate functions are of the form

$$
\mathbf{I}(X) = \int_0^L \sup_\lambda\big(\lambda^T \dot{X}_t - H(X_t, \lambda)\big)\,dt,
$$

provided $X_t, t \in [0, L]$, is an absolutely continuous function with a suitable initial condition. Otherwise, $\mathbf{I}(X) = \infty$. Here, with the notation of (2.20),

$$
\begin{aligned}
H(u, \lambda) = \lim_{t\to\infty} \frac{1}{t} \ln \mathbf{E} \exp\biggl(\int_0^t \biggl(\lambda^T f\big(u, y_s^{u,\lambda}\big) + \frac{1}{2}\lambda^T \big(\sigma_1\sigma_1^T\big(u, y_s^{u,\lambda}\big)\\
+ \sigma_3\sigma_3^T\big(u, y_s^{u,\lambda}\big)\big)\lambda\biggr)\,ds\biggr),
\end{aligned}
\tag{2.21}
$$

where

$$
\begin{aligned}
dy_t^{u,\lambda} = \big(B\big(u, y_t^{u,\lambda}\big) + C_1\big(y_t^{u,\lambda}\big)\sigma_1\big(u, y_t^{u,\lambda}\big)^T\lambda\big)\,dt + \big(C_1\big(Y_t^{u,\lambda}\big)\,dW_t^1\\
+ C_2\big(Y_t^{u,\lambda}\big)\,dW_t^2\big), \qquad y_0^{u,\lambda} = 0.
\end{aligned}
$$

Let us note that if one assumes the LDP at rate t as $t \to \infty$ of the empirical measures $\nu_t^{u,\lambda}(dx) = (1/t)\int_0^t \mathbf{1}_{dx}(y_s^{u,\lambda})\,ds$ with large deviation rate function $\mathbf{J}^{u,\lambda}$, then in view of Varadhan's lemma and (2.21), under suitable assumptions,

$$
\begin{aligned}
H(u, \lambda) = \sup_{\nu\in\mathbb{M}_1(\mathbb{R}^l)} \biggl(\int_{\mathbb{R}^l} \biggl(\lambda^T f(u, x)\\
+ \frac{1}{2}\lambda^T \big(\sigma_1\sigma_1^T(u, x) + \sigma_3\sigma_3^T(u, x)\big)\lambda\biggr)\nu(dx) - \mathbf{J}^{u,\lambda}(\nu)\biggr),
\end{aligned}
$$

which is consistent with (2.19).

Section 11.6 of Feng and Kurtz [16] is concerned with the process X^ε satisfying equations (1.1). Conditions for the LDP to hold are obtained. They require the existence of functions with certain properties and are not easily translated into conditions on the coefficients. When the authors give explicit conditions on the coefficients, they need, in particular, $b(u, x)$ not to depend on u (see Lemma 11.60 on page 278). The large deviation rate function is identified as having the form (2.19) corresponding to the time-homogeneous setting, provided $B(u,x)b(u,x)^T = 0$ and certain additional hypotheses hold (see Theorem 11.6.5 on page 282). The authors choose not to pursue the setup of the averaging principle.

The LDP for the empirical measures of continuous-time Markov processes, such as in Corollary 2.3, is a well-explored subject; see Donsker and Varadhan [11, 12], Deuschel and Stroock [10]. The canonical form of the large deviation rate function is $\sup_f \int_{\mathbb{R}^l} -\mathcal{L}f/f\,d\mu$, where $\mathcal{L}$ represents the infinitesimal generator of the Markov process; see, for example, Theorem 4.2.43 in Deuschel and Stroock [10]. The form in Corollary 2.3 follows by taking $f(x) = e^{-h(x)}$. Gärtner [19] and Veretennikov [52] characterize the large deviation functions via limits similar to that in (2.21), the latter author allowing discontinuous coefficients. Theorem 12.7 on page 291 of Feng and Kurtz [16] tackles associated empirical processes; cf. Corollary 2.2.

3. Some generalities. This section contains general results on the LDP that underlie the proof of Theorem 2.1; cf. Puhalskii [40]. Let Σ represent a directed set, let $\mathbf{P}_\sigma$, where $\sigma \in \Sigma$, represent a net of probability measures on a metric space $\mathbb{S}$ indexed with the elements of Σ and let r_σ represent an $\mathbb{R}_+$-valued function which tends to infinity as $\sigma \in \Sigma$. A $[0,\infty]$-valued function $\mathbf{I}$ on $\mathbb{S}$ is referred to as a large deviation function if the sets $K_\delta = \{z \in \mathbb{S} : \mathbf{I}(z) \le \delta\}$ are compact for all $\delta \in \mathbb{R}_+$. We say that the net $\mathbf{P}_\sigma$ obeys the LDP with a large deviation function $\mathbf{I}$ for rate r_σ as $\sigma \in \Sigma$ if $\liminf_{\sigma\in\Sigma} r_\sigma^{-1} \ln \mathbf{P}_\sigma(G) \ge -\inf_{z\in G} \mathbf{I}(z)$ for all open sets $G \subset \mathbb{S}$ and $\limsup_{\sigma\in\Sigma} r_\sigma^{-1} \ln \mathbf{P}_\sigma(F) \le -\inf_{z\in F} \mathbf{I}(z)$ for all closed sets $F \subset \mathbb{S}$. We say that $\mathbf{I}$ is a large deviation (LD) limit point of $\mathbf{P}_\sigma$ for rate r_σ if there exists a subsequence σ_i, where $i \in \mathbb{N}$, such that $\mathbf{P}_{\sigma_i}$ satisfies the LDP with $\mathbf{I}$ for rate r_{σ_i} as $i \to \infty$. We say that the net $\mathbf{P}_\sigma$ is sequentially large deviation (LD) relatively compact for rate r_σ as $\sigma \in \Sigma$ if any subsequence $\mathbf{P}_{\sigma_i}$ of $\mathbf{P}_\sigma$ contains a further subsequence $\mathbf{P}_{\sigma_{i_j}}$ which satisfies the LDP for rate $r_{\sigma_{i_j}}$ with some large deviation function as $j \to \infty$. We say that the net $\mathbf{P}_\sigma$ is exponentially (or large deviation) tight for rate r_σ as $\sigma \in \Sigma$ if for arbitrary $\kappa > 0$ there exists compact $K \subset \mathbb{S}$ such that $\limsup_{\sigma\in\Sigma} \mathbf{P}_\sigma(\mathbb{S} \setminus K)^{1/r_\sigma} < \kappa$. We say that the net $\mathbf{P}_\sigma$ is sequentially exponentially tight for rate r_σ as $\sigma \in \Sigma$ if any subsequence $\mathbf{P}_{\sigma_i}$ is exponentially tight for rate r_{σ_i} as $i \to \infty$. We say that a net Y_σ of random elements of $\mathbb{S}$ obeys the LDP, respectively, is sequentially LD relatively compact, respectively, is exponentially tight, respectively, is sequentially exponentially tight if the net of their laws has the indicated property.

The cornerstone of our approach is the next result (Puhalskii [36, 37, 40, 44], see also Feng and Kurtz [16] and references therein).

THEOREM 3.1. *If the net $\mathbf{P}_\sigma$ is sequentially exponentially tight for rate r_σ as $\sigma \in \Sigma$, then the net $\mathbf{P}_\sigma$ is sequentially LD relatively compact for rate r_σ as $\sigma \in \Sigma$.*

The proof of the following theorem is standard.

THEOREM 3.2. *If the net $\mathbf{P}_\sigma$ is sequentially LD relatively compact for rate r_σ as $\sigma \in \Sigma$ and $\mathbf{I}$ is a unique LD limit point of the $\mathbf{P}_\sigma$, then the net $\mathbf{P}_\sigma$ satisfies the LDP with $\mathbf{I}$ for rate r_σ as $\sigma \in \Sigma$.*

The next theorem is essentially Varadhan's lemma; see, for example, Deuschel and Stroock [10]. It will be used to obtain equations for LD limit points.

THEOREM 3.3. *Suppose the net $\mathbf{P}_\sigma$ is sequentially exponentially tight for rate r_σ as $\sigma \in \Sigma$ and let $\mathbf{I}$ represent an LD limit point of $\mathbf{P}_\sigma$. Let U_σ be a net of uniformly bounded real valued functions on $\mathbb{S}$ such that $\int_{\mathbb{S}} \exp(r_\sigma U_\sigma(z))\mathbf{P}_\sigma(dz) = 1$. If $U_\sigma \to U$ uniformly on compact sets as $\sigma \in \Sigma$, where the function U is continuous, then $\sup_{z\in\mathbb{S}}(U(z) - \mathbf{I}(z)) = 0$.*

Identification of LD limit points will be carried out with the aid of the next result.

THEOREM 3.4. *Suppose $\mathbf{I}$ is a large deviation function on $\mathbb{S}$ and $\mathcal{U}$ is a collection of functions on $\mathbb{S}$ such that $\sup_{z\in\mathbb{S}}(U(z) - \mathbf{I}(z)) = 0$ for all $U \in \mathcal{U}$. Let $\mathbf{I}^{**}(z) = \sup_{U\in\mathcal{U}} U(z)$ and $K_\delta = \{z \in \mathbb{S} : \mathbf{I}(z) \le \delta\}$, where $\delta \in \mathbb{R}_+$.*

1. *Let $\tilde{\mathcal{U}}$ represent a set of functions U such that $\sup_{z\in K_\delta}(U(z) - \mathbf{I}(z)) = 0$ for suitable $\delta \in \mathbb{R}_+$. Suppose $\hat{z} \in \mathbb{S}$ is such that $\mathbf{I}^{**}(\hat{z}) = \hat{U}(\hat{z})$ for some function $\hat{U} \in \tilde{\mathcal{U}}$. Suppose there exists sequence $U_i \in \tilde{\mathcal{U}}$ with the following properties: $\sup_{z\in K_\delta}(U_i(z) - \mathbf{I}(z)) = 0$ for some common δ, the functions U_i are continuous when restricted to K_δ and if z_i is a convergent sequence of elements of K_δ such that $U_i(z_i) = \mathbf{I}(z_i)$, then $U_i(z_i) \to \hat{U}(\hat{z})$ and $z_i \to \hat{z}$ as $i \to \infty$. Then $\mathbf{I}(\hat{z}) = \mathbf{I}^{**}(\hat{z})$.*
2. *If for every $z \in \mathbb{S}$ such that $\mathbf{I}^{**}(z) < \infty$ there exists a sequence of points z_i such that $\mathbf{I}(z_i) = \mathbf{I}^{**}(z_i)$, $z_i \to z$, and $\mathbf{I}^{**}(z_i) \to \mathbf{I}^{**}(z)$ as $i \to \infty$, then $\mathbf{I}(z) = \mathbf{I}^{**}(z)$ for all $z \in \mathbb{S}$.*

PROOF. Let us first note that $\mathbf{I}(z) \ge \mathbf{I}^{**}(z)$ for all z, so, one needs to prove that $\mathbf{I}(z) \le \mathbf{I}^{**}(z)$ if $\mathbf{I}^{**}(z) < \infty$. We prove part 1. Since $\sup_{z\in K_\delta}(U_i(z) - \mathbf{I}(z)) = 0$, K_δ is compact, and $U_i(z) - \mathbf{I}(z)$ is upper semicontinuous when restricted to K_δ, there exist $z_i \in K_\delta$ such that $U_i(z_i) = \mathbf{I}(z_i)$. One may assume that the sequence converges. Since $U_i(z_i) \to \hat{U}(\hat{z})$, $z_i \to \hat{z}$ and $\mathbf{I}$ is lower semicontinuous, $\hat{U}(\hat{z}) \ge \mathbf{I}(\hat{z})$, so $\mathbf{I}^{**}(\hat{z}) \ge \mathbf{I}(\hat{z})$. The proof of part 2 is similar. □

In the rest of the paper, the above framework is used to prove Theorem 2.1. In Section 4, LD relative compactness is established; see Theorem 4.1. In Section 5, equations along the lines of Theorem 3.3 are derived; see Theorem 5.1. Section 6 is concerned with regularity properties of (X, μ) for which the function $\mathbf{I}^{**}$ as defined in Theorem 3.4 assumes finite values. It is also shown to be of the form given in Proposition 2.1; see Theorem 6.1. In Theorem 7.1 of Section 7, the large deviation function is identified for a large class of (X, μ), which implements the recipe of part 1 of Theorem 3.4. In Theorem 8.1 of Section 8, it is proved that that class is dense in the sense of part 2 of Theorem 3.4. In Section 9, the proof of Theorem 2.1 is completed.

4. LD relative compactness. The main result of this section is the following theorem.

THEOREM 4.1. *Suppose that conditions* (2.2a)–(2.2d) *and* (2.4a) *hold and that the net* $(X_0^\varepsilon, x_0^\varepsilon)$ *is exponentially tight for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$. *Then the net* $(X^\varepsilon, \mu^\varepsilon)$ *is sequentially LD relatively compact in* $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ *for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$.

We precede the proof with a criterion of sequential LD relative compactness in $\mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$. Let $d(\cdot, \cdot)$ represent the Lipschitz metric on $\mathbb{M}(\mathbb{R}^l)$: $d(\tilde{\mu}, \hat{\mu}) = \sup\{|\int_{\mathbb{R}^l} f(x)\tilde{\mu}(dx) - \int_{\mathbb{R}^l} f(x)\hat{\mu}(dx)|\}$, with the supremum being taken over functions $f : \mathbb{R}^l \to \mathbb{R}$ such that $\sup_{x\in\mathbb{R}^l} |f(x)| \le 1$ and $\sup_{x,y\in\mathbb{R}^l, x\neq y} |f(x) - f(y)|/|x - y| \le 1$; see, for example, page 395 in Dudley [13]. The proof of the next lemma is done in a standard fashion (cf., Billingsley [5], Chapter 2) and is omitted.

LEMMA 4.1. 1. *A net* $\{\nu_\varepsilon, \varepsilon > 0\}$, *where* $\nu_\varepsilon = (\nu_{\varepsilon,t}, t \in \mathbb{R}_+)$, *of random elements of* $\mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ *defined on respective probability spaces* $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, \mathbf{P}_\varepsilon)$ *is sequentially exponentially tight for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$ *if and only if for all* $t \in \mathbb{R}_+$ *and all* $\eta > 0$,

$$\lim_{N\to\infty} \limsup_{\varepsilon\to 0} \mathbf{P}_\varepsilon\big(\nu_{\varepsilon,t}\big(x \in \mathbb{R}^l : |x| > N\big) > \eta\big)^\varepsilon = 0$$

and

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \sup_{s_1\in[0,t]} \mathbf{P}_\varepsilon\Big(\sup_{s_2\in[s_1,s_1+\delta]} d(\nu_{\varepsilon,s_1}, \nu_{\varepsilon,s_2}) > \eta\Big)^\varepsilon = 0.$$

2. *A net* $\{Y_\varepsilon, \varepsilon > 0\}$, *where* $Y_\varepsilon = (Y_{\varepsilon,t}, t \in \mathbb{R}_+)$, *of random elements of* $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n)$ *defined on respective probability spaces* $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, \mathbf{P}_\varepsilon)$ *is sequentially exponentially tight for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$ *if and only if*

$$\lim_{N\to\infty} \limsup_{\varepsilon\to 0} \mathbf{P}_\varepsilon\big(|Y_{\varepsilon,0}| > N\big)^\varepsilon = 0$$

and, for all $t \in \mathbb{R}_+$ *and all* $\eta > 0$,

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \sup_{s_1\in[0,t]} \mathbf{P}_\varepsilon\Big(\sup_{s_2\in[s_1,s_1+\delta]} |Y_{\varepsilon,s_2} - Y_{\varepsilon,s_1}| > \eta\Big)^\varepsilon = 0.$$

REMARK 4.1. The form of the conditions is due to Feng and Kurtz [16].

PROOF OF THEOREM 4.1. Since $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ is a closed subset of $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ and $\mathbf{P}^\varepsilon((X^\varepsilon, \mu^\varepsilon) \in \mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))) = 1$, it is sufficient to prove that the net $((X^\varepsilon, \mu^\varepsilon), \varepsilon > 0)$ is sequentially LD relatively compact in $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$. By Theorem 3.1, the latter property holds if $(X^\varepsilon, \mu^\varepsilon)$ is sequentially exponentially tight, which is the case if the nets X^ε and μ^ε are each sequentially exponentially tight.

We show that the net X^ε is sequentially exponentially tight first. By (2.1a) and Itô's lemma, on denoting $g_1(x) = D^2 \ln(1 + |x|^2)$,

$$\begin{aligned}\ln(1 + |X_t^\varepsilon|^2) &= \ln(1 + |X_0^\varepsilon|^2) + \int_0^t \frac{2(X_s^\varepsilon)^T A_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)}{1 + |X_s^\varepsilon|^2}\, ds \\ &\quad + \frac{\varepsilon}{2}\int_0^t \operatorname{tr}(C_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon) g_1(X_s^\varepsilon))\, ds + \sqrt{\varepsilon}\int_0^t \frac{2(X_s^\varepsilon)^T}{1 + |X_s^\varepsilon|^2} B_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, dW_s^\varepsilon.\end{aligned}$$

Given $N > 0$, let $\tau_N^\varepsilon = \inf\{s \in \mathbb{R}_+ : |X_s^\varepsilon| \geq N\}$. Since τ_N^ε is an $\mathbf{F}^\varepsilon$-stopping time and

$$\exp\left(\frac{1}{\sqrt{\varepsilon}}\int_0^t \frac{2(X_s^\varepsilon)^T}{1 + |X_s^\varepsilon|^2} B_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, dW_s^\varepsilon - \frac{1}{2\varepsilon}\int_0^t \left\| \frac{2X_s^\varepsilon}{1 + |X_s^\varepsilon|^2}\right\|^2_{C_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)} ds\right), \quad t \in \mathbb{R}_+,$$

is an $\mathbf{F}^\varepsilon$-local martingale,

$$\begin{aligned}&\mathbf{E}^\varepsilon \exp\left(\frac{1}{\varepsilon}\ln(1 + |X^\varepsilon_{t\wedge\tau_N^\varepsilon}|^2) - \frac{1}{\varepsilon}\ln(1 + |X_0^\varepsilon|^2) - \frac{1}{\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon} \frac{2(X_s^\varepsilon)^T A_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)}{1 + |X_s^\varepsilon|^2}\, ds\right. \\ &\quad - \frac{1}{2}\int_0^{t\wedge\tau_N^\varepsilon} \operatorname{tr}(C_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon) g_1(X_s^\varepsilon))\, ds \\ &\quad \left. - \frac{1}{2\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon} \left\|\frac{2X_s^\varepsilon}{1 + |X_s^\varepsilon|^2}\right\|^2_{C_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)} ds\right) \leq 1. \end{aligned} \tag{4.1}$$

Since

$$\begin{aligned}\operatorname{tr}(C_s^\varepsilon(X_s^\varepsilon, x) g_1(X_s^\varepsilon)) &\leq \sqrt{\operatorname{tr} g_1(X_s^\varepsilon)^2}\sqrt{\operatorname{tr} C_s^\varepsilon(X_s^\varepsilon, x)^2} \\ &\leq 2n\sqrt{n}\frac{\|C_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\|}{1 + |X_s^\varepsilon|^2},\end{aligned} \tag{4.2}$$

on recalling (2.2c) and (2.2d), we have that there exists $L > 0$, which does not depend either on t or on N, such that for all $\varepsilon > 0$ small enough,

$$\mathbf{E}^\varepsilon \exp\left(\frac{1}{\varepsilon}\ln(1 + |X^\varepsilon_{t\wedge\tau_N^\varepsilon}|^2) - \frac{1}{\varepsilon}\ln(1 + |X_0^\varepsilon|^2) - \frac{Lt}{\varepsilon}\right) \leq 1.$$

For $\tilde{N} > 0$,

$$
\begin{aligned}
&\mathbf{P}^\varepsilon\Big(\sup_{s\in[0,t]}|X_s^\varepsilon| \geq N\Big)\\
&\quad = \mathbf{P}^\varepsilon\big(|X^\varepsilon_{t\wedge\tau_N^\varepsilon}| \geq N\big) \leq \mathbf{P}^\varepsilon\big(|X_0^\varepsilon| > \tilde{N}\big)\\
&\qquad + \mathbf{E}^\varepsilon \exp\Big(\frac{1}{\varepsilon}\ln\big(1+|X^\varepsilon_{t\wedge\tau_N^\varepsilon}|^2\big) - \frac{1}{\varepsilon}\ln\big(1+N^2\big)\Big)\mathbf{1}_{\{|X_0^\varepsilon|\leq\tilde{N}\}}\\
&\quad \leq \mathbf{P}^\varepsilon\big(|X_0^\varepsilon| > \tilde{N}\big) + \exp\Big(\frac{1}{\varepsilon}\ln\big(1+\tilde{N}^2\big) + \frac{Lt}{\varepsilon} - \frac{1}{\varepsilon}\ln\big(1+N^2\big)\Big),
\end{aligned}
$$

so

$$
\limsup_{N\to\infty}\limsup_{\varepsilon\to 0}\mathbf{P}^\varepsilon\Big(\sup_{s\in[0,t]}|X_s^\varepsilon| > N\Big)^\varepsilon \leq \limsup_{\varepsilon\to 0}\mathbf{P}^\varepsilon\big(|X_0^\varepsilon| > \tilde{N}\big)^\varepsilon.
$$

Since X_0^ε is exponentially tight and $\tilde{N}$ is arbitrary, we conclude that

$$
\lim_{N\to\infty}\limsup_{\varepsilon\to 0}\mathbf{P}^\varepsilon\Big(\sup_{s\in[0,t]}|X_s^\varepsilon| > N\Big)^\varepsilon = 0. \tag{4.3}
$$

By (2.1a), for $s \in [0, t]$, $\delta > 0$, and $\eta > 0$,

$$
\begin{aligned}
\mathbf{P}^\varepsilon\Big(\sup_{\tilde{s}\in[s,s+\delta]}|X_{\tilde{s}}^\varepsilon - X_s^\varepsilon| > \eta\Big) &\leq \mathbf{P}^\varepsilon\big(\tau_N^\varepsilon \leq t\big) + \mathbf{P}^\varepsilon\Big(\sup_{|u|\leq N}\sup_{x\in\mathbb{R}^l}|A_s^\varepsilon(u,x)|\delta\\
&\qquad + \sqrt{\varepsilon}\sup_{\tilde{s}\in[s,s+\delta]}\Big|\int_{s\wedge\tau_N^\varepsilon}^{\tilde{s}\wedge\tau_N^\varepsilon} B_r^\varepsilon(X_r^\varepsilon, x_r^\varepsilon)\,dW_r^\varepsilon\Big| > \eta\Big).
\end{aligned}
$$

Let e_i, for $i = 1, 2, \dots, n$, denote the ith unit vector of $\mathbb{R}^n$. Thanks to (2.2b) and (2.2d), for small enough δ and arbitrary $\alpha > 0$, provided $\varepsilon > 0$ is small enough, on using Doob's inequality,

$$
\begin{aligned}
&\mathbf{P}^\varepsilon\Big(\sup_{\tilde{s}\in[s,s+\delta]}|X_{\tilde{s}}^\varepsilon - X_s^\varepsilon| > \eta\Big)\\
&\quad \leq \mathbf{P}^\varepsilon\big(\tau_N^\varepsilon \leq t\big) + \mathbf{P}^\varepsilon\Big(\sqrt{\varepsilon}\sup_{\tilde{s}\in[s,s+\delta]}\Big|\int_{s\wedge\tau_N^\varepsilon}^{\tilde{s}\wedge\tau_N^\varepsilon} B_r^\varepsilon(X_r^\varepsilon, x_r^\varepsilon)\,dW_r^\varepsilon\Big| > \frac{\eta}{2}\Big)\\
&\quad \leq \mathbf{P}^\varepsilon\big(\tau_N^\varepsilon \leq t\big) + \sum_{i=1}^n \mathbf{P}^\varepsilon\Big(\sqrt{\varepsilon}\sup_{\tilde{s}\in[s,s+\delta]}\Big(e_i^T\int_{s\wedge\tau_N^\varepsilon}^{\tilde{s}\wedge\tau_N^\varepsilon} B_r^\varepsilon(X_r^\varepsilon, x_r^\varepsilon)\,dW_r^\varepsilon\Big) > \frac{\eta}{2n}\Big)\\
&\quad \leq \mathbf{P}^\varepsilon\big(\tau_N^\varepsilon \leq t\big) + \sum_{i=1}^n \mathbf{P}^\varepsilon\Big(\sup_{\tilde{s}\in[s,s+\delta]}\exp\Big(\frac{\alpha}{\sqrt{\varepsilon}}e_i^T\int_{s\wedge\tau_N^\varepsilon}^{\tilde{s}\wedge\tau_N^\varepsilon} B_r^\varepsilon(X_r^\varepsilon, x_r^\varepsilon)\,dW_r^\varepsilon\\
&\qquad - \frac{\alpha^2}{2\varepsilon}\int_{s\wedge\tau_N^\varepsilon}^{\tilde{s}\wedge\tau_N^\varepsilon} e_i^T C_r^\varepsilon(X_r^\varepsilon, x_r^\varepsilon)e_i\,dr\Big)
\end{aligned}
$$

$$> e^{\alpha\eta/(2n\varepsilon)} \exp\biggl(-\frac{\alpha^2\delta}{2\varepsilon} \sup_{r\in[0,t]} \sup_{|u|\le N} \sup_{x\in\mathbb{R}^l} \|C_r^\varepsilon(u,x)\|\biggr)\biggr)$$

$$\le \mathbf{P}^\varepsilon\Bigl(\sup_{\tilde{s}\in[0,t]} |X_{\tilde{s}}^\varepsilon| \ge N\Bigr) + ne^{-\alpha\eta/(2n\varepsilon)} \exp\biggl(\frac{\alpha^2\delta}{2\varepsilon} \sup_{r\in[0,t]} \sup_{|u|\le N} \sup_{x\in\mathbb{R}^l} \|C_r^\varepsilon(u,x)\|\biggr).$$

By (2.2d), (4.3) and the fact that α can be chosen arbitrarily great,

$$\limsup_{\delta\to 0} \limsup_{\varepsilon\to 0} \sup_{s\in[0,t]} \mathbf{P}^\varepsilon\Bigl(\sup_{\tilde{s}\in[s,s+\delta]} |X_{\tilde{s}}^\varepsilon - X_s^\varepsilon| > \eta\Bigr)^\varepsilon = 0.$$

The sequential exponential tightness of X^ε follows from part 2 of Lemma 4.1.

We prove now that μ^ε is sequentially exponentially tight. Let f represent an $\mathbb{R}$-valued twice continuously differentiable function on $\mathbb{R}^l$. By (2.1b) and Itô's lemma,

$$f(x_t^\varepsilon) = f(x_0^\varepsilon) + \frac{1}{\varepsilon}\int_0^t Df(x_s^\varepsilon)^T a_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\,ds$$
$$+ \frac{1}{2\varepsilon}\int_0^t \operatorname{tr}\bigl(c_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon) D^2 f(x_s^\varepsilon)\bigr)\,ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t Df(x_s^\varepsilon)^T b_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\,dW_s^\varepsilon.$$

Therefore, on identifying μ^ε with measure $\mu^\varepsilon(dt,dx)$, we have that, in analogy with (4.1),

$$\mathbf{E}^\varepsilon \exp\biggl(f(x_{t\wedge\tau_N^\varepsilon}^\varepsilon) - f(x_0^\varepsilon) - \frac{1}{\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon}\int_{\mathbb{R}^l} Df(x)^T a_s^\varepsilon(X_s^\varepsilon, x)\mu^\varepsilon(ds,dx)$$
$$- \frac{1}{2\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon}\int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s^\varepsilon(X_s^\varepsilon, x) D^2 f(x)\bigr)\mu^\varepsilon(ds,dx)$$
$$- \frac{1}{2\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon}\int_{\mathbb{R}^l} \|Df(x)\|^2_{c_s^\varepsilon(X_s^\varepsilon,x)}\mu^\varepsilon(ds,dx)\biggr) \le 1.$$

Let $g_2(u)$, where $u\in\mathbb{R}_+$, be an $\mathbb{R}_+$-valued nondecreasing $\mathbb{C}^2$-function with a bounded second derivative such that $Dg_2(0) = D^2 g_2(0) = 0$ and $g_2(u) = u$ for $u \ge 1$. For given $\breve{N} > 0$, we let $f(x) = g_2((|x| - \breve{N})^+)$, where $x\in\mathbb{R}^l$. By (2.4a), if $\breve{N}$ is great enough, then for all ε small enough, $(x/|x|)^T a_{s\wedge\tau_N^\varepsilon}^\varepsilon(X_{s\wedge\tau_N^\varepsilon}^\varepsilon, x) \le 0$ provided $|x| \ge \breve{N}$. Since g_2 is a nondecreasing function,

$$Df(x)^T a_{s\wedge\tau_N^\varepsilon}^\varepsilon(X_{s\wedge\tau_N^\varepsilon}^\varepsilon, x) = Dg_2\bigl((|x| - \breve{N})^+\bigr)(x/|x|)^T a_{s\wedge\tau_N^\varepsilon}^\varepsilon(X_{s\wedge\tau_N^\varepsilon}^\varepsilon, x) \le 0.$$

In addition, as in (4.2), $\operatorname{tr}(c_s^\varepsilon(X_s^\varepsilon, x) D^2|x|) \le l\sqrt{l-1}\|c_s^\varepsilon(X_s^\varepsilon, x)\|/|x|$. We obtain that

$$\mathbf{E}^\varepsilon \exp\biggl(-f(x_0^\varepsilon) - \frac{1}{\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon}\int_{|x|>\breve{N}+1} \frac{x^T}{|x|} a_s^\varepsilon(X_s^\varepsilon, x)\mu^\varepsilon(ds,dx)$$

$$\text{(4.4)} \quad - \frac{1}{2\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon}\int_{\breve{N}\le|x|\le\breve{N}+1} \bigl(\operatorname{tr}\bigl(c_s^\varepsilon(X_s^\varepsilon, x) D^2 f(x)\bigr)$$

$$+ \|Df(x)\|^2_{c_s^\varepsilon(X_s^\varepsilon,x)}\big)\mu^\varepsilon(ds,dx)$$
$$- \frac{1}{2\varepsilon}\int_0^{t\wedge\tau_N^\varepsilon}\int_{|x|>\breve{N}+1}\biggl(\frac{\sqrt{l-1}}{|x|}l\|c_s^\varepsilon(X_s^\varepsilon,x)\| + \biggl\|\frac{x}{|x|}\biggr\|^2_{c_s^\varepsilon(X_s^\varepsilon,x)}\biggr)\mu^\varepsilon(ds,dx)\biggr)$$
$$\le 1.$$

Since $\|c_s^\varepsilon(u,x)\|$ is asymptotically bounded locally in (s,u) and globally in x, see (2.2a), there exists $\tilde{L} > 0$ such that $|\operatorname{tr}(c^\varepsilon_{s\wedge\tau_N^\varepsilon}(X^\varepsilon_{s\wedge\tau_N^\varepsilon},x)D^2f(x)) + \|Df(x)\|^2_{c^\varepsilon_{s\wedge\tau_N^\varepsilon}(X^\varepsilon_{s\wedge\tau_N^\varepsilon},x)}| \le \tilde{L}$ for all $s \le t$, all $\breve{N}$, and all x such that $|x| \in [\breve{N}, \breve{N}+1]$, provided $\varepsilon > 0$ is small enough. We can also assume that $\tilde{L}$ is an upper bound for $\|c^\varepsilon_{s\wedge\tau_N^\varepsilon}(X^\varepsilon_{s\wedge\tau_N^\varepsilon},x)\|$. We thus obtain from (4.4), on recalling that $\mu^\varepsilon([0,t],\mathbb{R}^l) = t$, that provided ε is small enough and $\breve{N}$ is great enough,

$$\mathbf{E}^\varepsilon \exp\biggl(-f(x_0^\varepsilon) + \frac{1}{\varepsilon}M^\varepsilon\mu^\varepsilon\biggl([0,t\wedge\tau_N^\varepsilon],\{x:|x|>\breve{N}+1\} - \frac{3\tilde{L}t}{2\varepsilon}\biggr)\biggr) \le 1,$$

where

$$M^\varepsilon = -\sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le N}\sup_{x\in\mathbb{R}^l:|x|>\breve{N}+1}\frac{x^T}{|x|}a_s^\varepsilon(u,x) > 0.$$

It follows that for arbitrary $\delta > 0$, all ε small enough, and all $\breve{N}$ great enough:

$$\begin{aligned}
&\mathbf{P}^\varepsilon\bigl(\mu^\varepsilon\bigl([0,t\wedge\tau_N^\varepsilon],\{x\in\mathbb{R}^l:|x|>\breve{N}+1\}\bigr)>\delta\bigr)\\
&\quad\le \mathbf{P}^\varepsilon\bigl(|x_0^\varepsilon|>\breve{N}\bigr)\\
&\qquad + \mathbf{E}^\varepsilon\exp\biggl(\frac{M^\varepsilon}{\varepsilon}\mu^\varepsilon\bigl([0,t\wedge\tau_N^\varepsilon],\{x:|x|>\breve{N}+1\}\bigr)\biggr)\mathbf{1}_{\{|x_0^\varepsilon|\le\breve{N}\}}\exp\biggl(-\frac{M^\varepsilon}{\varepsilon}\delta\biggr)\\
&\quad\le \mathbf{P}^\varepsilon\bigl(|x_0^\varepsilon|>\breve{N}\bigr) + \exp\biggl(\frac{3\tilde{L}t}{2\varepsilon} - \frac{M^\varepsilon\delta}{\varepsilon} + g_2(0)\biggr),
\end{aligned}$$

so by the facts that $\liminf_{\varepsilon\to0}M^\varepsilon\to\infty$ and $\limsup_{\varepsilon\to0}\mathbf{P}^\varepsilon(|x_0^\varepsilon|\ge\breve{N})^\varepsilon\to0$ as $\breve{N}\to\infty$, and that (4.3) holds, we obtain that

$$\lim_{\breve{N}\to\infty}\limsup_{\varepsilon\to0}\mathbf{P}^\varepsilon\bigl(\mu^\varepsilon\bigl([0,t],\{x\in\mathbb{R}^l:|x|>\breve{N}+1\}\bigr)>\delta\bigr)^\varepsilon = 0.$$

Since $|\mu_t^\varepsilon(\Theta) - \mu_s^\varepsilon(\Theta)| \le |t-s|$, for $\Theta\in\mathcal{B}(\mathbb{R}^l)$, the sequential exponential tightness of μ^ε follows from part 1 of Lemma 4.1. $\square$

REMARK 4.2. Since $(X^\varepsilon,\mu^\varepsilon)$ is continuous in ε in distribution, one can prove that $(X^\varepsilon,\mu^\varepsilon)$ is exponentially tight.

5. The equation for the large deviation function. In this section, we derive the equation for large deviation limit points of $(X^\varepsilon, \mu^\varepsilon)$ that is to be used for identifying the large deviation function. For $0 = t_0 < t_1 < \cdots < t_i$, let

$$\lambda(t, X) = \sum_{j=1}^{i} \lambda_j(X_{t_{j-1}}) \mathbf{1}_{[t_{j-1}, t_j)}(t), \tag{5.1}$$

where $X = (X_s, s \in \mathbb{R}_+) \in \mathbb{C}(\mathbb{R}_+, \mathbb{R}^n)$ and the functions $\lambda_j(u)$, for $u \in \mathbb{R}^n$, are $\mathbb{R}^n$-valued and continuous. We define

$$\int_0^t \lambda(s, X)\, dX_s = \sum_{j=1}^{i} \lambda_i(X_{t_{j-1} \wedge t})^T (X_{t \wedge t_j} - X_{t \wedge t_{j-1}}). \tag{5.2}$$

Let $f(t, u, x)$ represent a $\mathbb{C}^{1,2,2}(\mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^l)$-function with compact support in x locally uniformly in (t, u) and let, with $(X, \mu) \in \mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$,

$$\begin{aligned} U_t^{\lambda(\cdot), f}(X, \mu) &= \int_0^t \lambda(s, X)\, dX_s - \int_0^t \int_{\mathbb{R}^l} \lambda(s, X)^T A_s(X_s, x) \mu(ds, dx) \\ &\quad - \int_0^t \int_{\mathbb{R}^l} D_x f(s, X_s, x)^T a_s(X_s, x) \mu(ds, dx) \\ &\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s(X_s, x) D_{xx}^2 f(s, X_s, x)\bigr) \mu(ds, dx) \\ &\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \|\lambda(s, X)\|^2_{C_s(X_s, x)} \mu(ds, dx) \\ &\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \|D_x f(s, X_s, x)\|^2_{c_s(X_s, x)} \mu(ds, dx) \\ &\quad - \int_0^t \int_{\mathbb{R}^l} \lambda(s, X)^T G_s(X_s, x) D_x f(s, X_s, x) \mu(ds, dx). \end{aligned} \tag{5.3}$$

Under Condition 2.1, $U_t^{\lambda(\cdot), f}(X, \mu)$ is a continuous function of (X, μ).

Let $\tau(X, \mu)$ represent a continuous function of $(X, \mu) \in \mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ that is also a stopping time relative to the flow $\mathbf{G} = (\mathcal{G}_t, t \in \mathbb{R}_+)$ on $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$, where the σ-algebra $\mathcal{G}_t$ is generated by the mappings $X \to X_s$ and $\mu \to \mu_s$ for $s \le t$. (We note that the flow $\mathbf{G}$ is not right continuous, so τ is a strict stopping time; see Jacod and Shiryaev [25].) Let us also assume that $X_{t \wedge \tau(X, \mu)}$ is a bounded function of (X, μ).

THEOREM 5.1. *Suppose that Conditions* 2.1, (2.2a), (2.2b), (2.2d) *and* (2.10) *hold. If* $\tilde{\mathbf{I}}$ *is a large deviation limit point of* $(X^\varepsilon, \mu^\varepsilon)$ *for rate* $1/\varepsilon$ *as* $\varepsilon \to 0$, *then*

$$\sup_{(X,\mu) \in \mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))} \bigl(U_{t \wedge \tau(X,\mu)}^{\lambda(\cdot), f}(X, \mu) - \tilde{\mathbf{I}}(X, \mu)\bigr) = 0. \tag{5.4}$$

PROOF. The process $(\lambda(t, X^\varepsilon), t \in \mathbb{R}_+)$ is $\mathbf{F}^\varepsilon$-adapted so that by (2.1a) and (5.2),

$$\int_0^t \lambda(s, X^\varepsilon)\, dX_s^\varepsilon = \int_0^t \lambda(s, X^\varepsilon)^T A_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, ds + \sqrt{\varepsilon} \int_0^t \lambda(s, X^\varepsilon)^T B_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, dW_s^\varepsilon. \tag{5.5}$$

By (2.1a), (2.1b) and Itô's lemma,

$$\begin{aligned} f(t, X_t^\varepsilon, x_t^\varepsilon) &= f(0, X_0^\varepsilon, x_0^\varepsilon) + \int_0^t \frac{\partial f(s, X_s^\varepsilon, x_s^\varepsilon)}{\partial s}\, ds \\ &\quad + \int_0^t D_u f(s, X_s^\varepsilon, x_s^\varepsilon)^T A_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, ds \\ &\quad + \sqrt{\varepsilon} \int_0^t D_u f(s, X_s^\varepsilon, x_s^\varepsilon)^T B_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, dW_s^\varepsilon \\ &\quad + \frac{1}{\varepsilon} \int_0^t D_x f(s, X_s^\varepsilon, x_s^\varepsilon)^T a_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, ds \\ &\quad + \frac{1}{\sqrt{\varepsilon}} \int_0^t D_x f(s, X_s^\varepsilon, x_s^\varepsilon)^T b_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon)\, dW_s^\varepsilon \\ &\quad + \frac{\varepsilon}{2} \int_0^t \operatorname{tr}\big(C_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon) D_{uu}^2 f(s, X_s^\varepsilon, x_s^\varepsilon)\big)\, ds \\ &\quad + \frac{1}{2\varepsilon} \int_0^t \operatorname{tr}\big(c_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon) D_{xx}^2 f(s, X_s^\varepsilon, x_s^\varepsilon)\big)\, ds \\ &\quad + \int_0^t \operatorname{tr}\big(G_s^\varepsilon(X_s^\varepsilon, x_s^\varepsilon) D_{ux}^2 f(s, X_s^\varepsilon, x_s^\varepsilon)\big)\, ds, \end{aligned} \tag{5.6}$$

where $G_s^\varepsilon(u, x) = B_s^\varepsilon(u, x) b_s^\varepsilon(u, x)^T$. We denote

$$\begin{aligned} U_t^\varepsilon(X, \mu) &= \int_0^t \lambda(s, X)\, dX_s - \int_0^t \int_{\mathbb{R}^l} \lambda(s, X)^T A_s^\varepsilon(X_s, x) \mu(ds, dx) \\ &\quad - \int_0^t \int_{\mathbb{R}^l} D_x f(s, X_s, x)^T a_s^\varepsilon(X_s, x) \mu(ds, dx) \\ &\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\big(c_s^\varepsilon(X_s, x) D_{xx}^2 f(s, X_s, x)\big) \mu(ds, dx) \\ &\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \|\lambda(s, X)\|_{C_s^\varepsilon(X_s, x)}^2 \mu(ds, dx) \\ &\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \|D_x f(s, X_s, x)\|_{c_s^\varepsilon(X_s, x)}^2 \mu(ds, dx) \\ &\quad - \int_0^t \int_{\mathbb{R}^l} \lambda(s, X)^T G_s^\varepsilon(X_s, x) D_x f(s, X_s, x) \mu(ds, dx) \end{aligned}$$

and

$$V_t^\varepsilon(X,\mu) = f(t, X_t, x_t^\varepsilon) - f(0, X_0, x_0^\varepsilon) - \int_0^t \int_{\mathbb{R}^l} \frac{\partial f(s, X_s, x)}{\partial s} \mu(ds, dx)$$
$$- \int_0^t \int_{\mathbb{R}^l} D_u f(s, X_s, x)^T A_s^\varepsilon(X_s, x) \mu(ds, dx)$$
$$- \frac{\varepsilon}{2} \int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\bigl(C_s^\varepsilon(X_s, x) D_{uu}^2 f(s, X_s, x)\bigr) \mu(ds, dx)$$
$$- \int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\bigl(G_s^\varepsilon(X_s, x) D_{ux}^2 f(s, X_s, x)\bigr) \mu(ds, dx)$$
$$- \frac{\varepsilon}{2} \int_0^t \int_{\mathbb{R}^l} \| D_u f(s, X_s, x) \|^2_{C_s^\varepsilon(X_s,x)} \mu(ds, dx)$$
$$- \int_0^t \int_{\mathbb{R}^l} \lambda(s, X)^T C_s^\varepsilon(X_s, x) D_u f(s, X_s, x) \mu(ds, dx)$$
$$- \int_0^t \int_{\mathbb{R}^l} D_u f(s, X_s, x)^T G_s^\varepsilon(X_s, x) D_x f(s, X_s, x) \mu(ds, dx).$$

Since the function $\lambda(s,u)$ is locally bounded, the function $f(s,u,x)$ and its derivatives are locally bounded and are of compact support in x, conditions (2.2a), (2.2b), (2.2d) and (2.11) hold, and $X_{t\wedge\tau(X,\mu)}$ is bounded; we have that there exists number $R(t) > 0$ such that for all ε small enough uniformly over (X,μ),

$$\bigl|U^\varepsilon_{t\wedge\tau(X,\mu)}(X,\mu)\bigr| + \bigl|V^\varepsilon_{t\wedge\tau(X,\mu)}(X,\mu)\bigr| \le R(t). \tag{5.7}$$

Since X_s^ε and μ_s^ε are $\mathcal{F}_s^\varepsilon$-measurable, $\tau(X^\varepsilon,\mu^\varepsilon)$ is a stopping time relative to $\mathbf{F}^\varepsilon$. By (5.5), (5.6) and (5.7), the process $(\exp((1/\varepsilon)U^\varepsilon_{t\wedge\tau(X^\varepsilon,\mu^\varepsilon)}(X^\varepsilon,\mu^\varepsilon) + V^\varepsilon_{t\wedge\tau(X^\varepsilon,\mu^\varepsilon)}(X^\varepsilon,\mu^\varepsilon)), t\in\mathbb{R}_+)$ is a bounded $\mathbf{F}^\varepsilon$-martingale, so

$$\mathbf{E}^\varepsilon \exp\biggl(\frac{1}{\varepsilon} U^\varepsilon_{t\wedge\tau(X^\varepsilon,\mu^\varepsilon)}(X^\varepsilon,\mu^\varepsilon) + V^\varepsilon_{t\wedge\tau(X^\varepsilon,\mu^\varepsilon)}(X^\varepsilon,\mu^\varepsilon)\biggr) = 1.$$

Since the function $f(s,u,x)$ is of compact support in x, the convergence hypotheses in (2.10) and the bound in (5.7) imply that $U^\varepsilon_{t\wedge\tau(X,\mu)}(X,\mu) \to U^{\lambda(\cdot),f}_{t\wedge\tau(X,\mu)}(X,\mu)$ as $\varepsilon\to 0$ uniformly over compact sets. By Theorem 3.3, $\sup_{(X,\mu)\in\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)\times\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))}(U^{\lambda(\cdot),f}_{t\wedge\tau(X,\mu)}(X,\mu) - \tilde{\mathbf{I}}(X,\mu)) = 0$. □

REMARK 5.1. One can see that there exists compact $K \subset \mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)\times\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$ such that $\sup_{(X,\mu)\in K}(U^{\lambda(\cdot),f}_{t\wedge\tau(X,\mu)}(X,\mu) - \tilde{\mathbf{I}}(X,\mu)) = 0$.

6. Regularity properties. Let $\tilde{\mathbf{I}}$ represent a large deviation limit point of $(X^\varepsilon,\mu^\varepsilon)$ for rate $1/\varepsilon$ as $\varepsilon\to 0$ under the hypotheses of Theorem 2.1 such that

$\tilde{\mathbf{I}}(X,\mu)=\infty$ unless $X_0=\hat{u}$, where $\hat{u}$ is a preselected element of $\mathbb{R}^n$. Let, as in Theorem 3.4, for $(X,\mu)\in\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)\times\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$,

$$\mathbf{I}^{**}(X,\mu)=\sup_{\lambda(\cdot),f,t,\tau} U^{\lambda(\cdot),f}_{t\wedge\tau(X,\mu)}(X,\mu), \tag{6.1}$$

with the supremum being taken over $\lambda(t,X)$, $f(t,u,x)$, and $\tau(X,\mu)$ satisfying the requirements of Theorem 5.1 and over $t\geq 0$. We note that, under Condition 2.1, $\mathbf{I}^{**}(X,\mu)$ is a lower semicontinuous function of (X,μ) and that by Theorem 5.1,

$$\mathbf{I}^{**}(X,\mu)\leq\tilde{\mathbf{I}}(X,\mu). \tag{6.2}$$

The rest of the paper is concerned mostly with proving that equality prevails in (6.2), provided $X_0=\hat{u}$. Since the case where $\mathbf{I}^{**}(X,\mu)<\infty$ needs to be considered only, in this section we undertake a study of the properties of (X,μ) such that $\mathbf{I}^{**}(X,\mu)<\infty$. We prove that if $\mathbf{I}^{**}(X,\mu)<\infty$ and $X_0=\hat{u}$, then $\mathbf{I}^{**}(X,\mu)=\mathbf{I}(X,\mu)$, where $\mathbf{I}(X,\mu)$ is given in the statements of Theorem 2.1 and Proposition 2.1 with $\mathbf{I}_0(\hat{u})=0$; see Theorem 6.1. We assume throughout Conditions 2.1, 2.2, (2.4b), (2.12c) and (2.12d) to hold.

LEMMA 6.1. *If $\mu\in\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$, then μ is of the form $\mu(ds,dx)=\nu_s(dx)\,ds$, where $\nu_s(dx)$ is a transition probability kernel from $\mathbb{R}_+$ to $\mathbb{R}^l$. If $(X,\mu)\in\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)\times\mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$ is such that $\mathbf{I}^{**}(X,\mu)<\infty$, then X is absolutely continuous with respect to Lebesgue measure.*

PROOF. We have that $\mu(ds,dx)=\nu_s(dx)\mu(ds,\mathbb{R}^l)$, where $\nu_s(dx)$ is a transition kernel from $\mathbb{R}_+$ to $\mathbb{R}^l$; see, for example, Theorem 8.1 on page 502 of Ethier and Kurtz [15]. Since $\mu(ds,\mathbb{R}^l)$ is Lebesgue measure on $\mathbb{R}_+$, $\mu(ds,dx)=\nu_s(dx)\,ds$.

On taking $f=0$ in (5.3) and assuming $\lambda(s,X)$ not to depend on X, so the piece of notation $\lambda(s)$ can be used instead, we have by (5.3), (6.1) and the part of the lemma just proved that if $\mathbf{I}^{**}(X,\mu)<\infty$, then

$$\int_0^t\lambda(s)\,dX_s\leq\int_0^t\int_{\mathbb{R}^l}\lambda(s)^T A_s(X_s,x)\nu_s(dx)\,ds + \frac{1}{2}\int_0^t\int_{\mathbb{R}^l}\|\lambda(s)\|^2_{C_s(X_s,x)}\nu_s(dx)\,ds+\mathbf{I}^{**}(X,\mu).$$

Replacing $\lambda(s)$ with $\delta\lambda(s)$, where $\delta>0$, dividing through by δ, and minimising the right-hand side over δ obtains that

$$\int_0^t\lambda(s)\,dX_s\leq\int_0^t\int_{\mathbb{R}^l}\lambda(s)^T A_s(X_s,x)\nu_s(dx)\,ds + \sqrt{2}\sqrt{\mathbf{I}^{**}(X,\mu)}\sqrt{\int_0^t\int_{\mathbb{R}^l}\|\lambda(s)\|^2_{C_s(X_s,x)}\nu_s(dx)\,ds}.$$

It follows that X is absolutely continuous with respect to ds. $\square$

By Lemma 6.1, if $\mathbf{I}^{**}(X,\mu)<\infty$, then (5.3) takes the form

$$
\begin{aligned}
U_t^{\lambda(\cdot),f}(X,\mu) = {} & \int_0^t \lambda(s,X)^T \dot{X}_s\,ds - \int_0^t\int_{\mathbb{R}^l} \lambda(s,X)^T A_s(X_s,x)\nu_s(dx)\,ds \\
& - \int_0^t\int_{\mathbb{R}^l} D_x f(s,X_s,x)^T a_s(X_s,x)\nu_s(dx)\,ds \\
& - \frac{1}{2}\int_0^t\int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s(X_s,x)D^2_{xx} f(s,X_s,x)\bigr)\nu_s(dx)\,ds \\
& - \frac{1}{2}\int_0^t\int_{\mathbb{R}^l} \|\lambda(s,X)\|^2_{C_s(X_s,x)}\nu_s(dx)\,ds \\
& - \frac{1}{2}\int_0^t\int_{\mathbb{R}^l} \|D_x f(s,X_s,x)\|^2_{c_s(X_s,x)}\nu_s(dx)\,ds \\
& - \int_0^t\int_{\mathbb{R}^l} \lambda(s,X)^T G_s(X_s,x) D_x f(s,X_s,x)\nu_s(dx)\,ds.
\end{aligned}
\tag{6.3}
$$

The next step is to show that $\nu_s(dx)$ has to be absolutely continuous with respect to dx and establish its integrability properties. We need, however, to lay the groundwork. The proofs of the following two key lemmas are omitted. The first one is essentially due to Röckner and Zhang [45], pages 204–205, [46]; see also Bogachev, Krylov and Röckner [6]. The second one is a local version of the result by Bogachev, Krylov and Röckner [6] that if $b\in\mathbb{L}^2(\mathbb{R}^d,m(x)\,dx)$ then $\sqrt{m}\in\mathbb{W}^{1,2}(\mathbb{R}^d)$, and is proved along similar lines; see also Metafune, Pallara and Rhandi [31].

LEMMA 6.2. *Let $d\in\mathbb{N}$ and let O represent either $\mathbb{R}^d$ or an open ball in $\mathbb{R}^d$. If $m(x)$ is an $\mathbb{R}_+$-valued measurable function on $\mathbb{R}^d$ such that $m\in\mathbb{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ and $\sqrt{m}\in\mathbb{W}^{1,2}(O)$, then $\mathbb{H}^{1,2}(O,m(x)\,dx)=\mathbb{W}^{1,2}(O,m(x)\,dx)$.*

LEMMA 6.3. *For $d\in\mathbb{N}$ and $x\in\mathbb{R}^d$, let $c(x)$ represent a locally Lipschitz continuous function with values in the set of symmetric positive definite $d\times d$-matrices and let $b(x)$ represent an $\mathbb{R}^d$-valued measurable function. Suppose $m(x)$ is a probability density on $\mathbb{R}^d$ such that $m(\ln m)^2\in\mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}^d)$, $b\in\mathbb{L}^2_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d,m(x)\,dx)$, and*

$$
\int_{\mathbb{R}^d}\operatorname{tr}\bigl(c(x)D^2p(x)\bigr)m(x)\,dx+\int_{\mathbb{R}^d}Dp(x)^Tb(x)m(x)\,dx=0
$$

for all $p\in\mathbb{C}_0^\infty(\mathbb{R}^d)$, where we assume that $0(\ln 0)^2=0$.

Then $m\in\mathbb{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ and $\sqrt{m}\in\mathbb{W}^{1,2}_{\mathrm{loc}}(\mathbb{R}^d)$. Furthermore, given open ball S from $\mathbb{R}^d$, there exists constant M which depends on S, on the Lipschitz constant of $c(x)$ on S, and on $\inf_{x\in S}x^Tc(x)x/|x|^2$ only, such that

$$
\int_S\frac{|Dm(x)|^2}{m(x)}\,dx\le M\biggl(1+\int_S\bigl(\ln m(x)\bigr)^2m(x)\,dx+\int_S\bigl|b(x)\bigr|^2m(x)\,dx\biggr).
\tag{6.4}
$$

For the next lemma, we recall that, according to our conventions, $q' = q/(q-1)$, provided $q > 1$.

LEMMA 6.4. *Suppose that* $\mathbf{I}^{**}(X,\mu) < \infty$, *where* $\mu(ds,dx) = \nu_s(dx)\,ds$. *Then, for almost all* s, *the transition kernel* $\nu_s(dx)$ *is absolutely continuous with respect to Lebesgue measure, the density* $m_s(x) = \nu_s(dx)/dx$ *is an element of* $\mathbb{L}^{\beta}_{\mathrm{loc}}(\mathbb{R}^l)$ *for all* $\beta \in [1, l/(l-1))$ *and is an element of* $\mathbb{W}^{1,\alpha}_{\mathrm{loc}}(\mathbb{R}^l)$ *for all* $\alpha \in [1, 2l/(2l-1))$, *and* $\sqrt{m_s(\cdot)} \in \mathbb{W}^{1,2}_{\mathrm{loc}}(\mathbb{R}^l)$. *Furthermore, for arbitrary* $t > 0$ *and open ball* $S \subset \mathbb{R}^l$,

$$\int_0^t \int_S \frac{|Dm_s(x)|^2}{m_s(x)}\,dx\,ds < \infty. \tag{6.5}$$

If, in addition, $\sqrt{m_s(\cdot)} \in \mathbb{W}^{1,2}(\mathbb{R}^l)$, *then* $Dm_s(\cdot)/m_s(\cdot) \in \mathbb{L}^{1,2}_0(\mathbb{R}^l, \mathbb{R}^l, m_s(x)\,dx)$.

If $\kappa \geq 0$, $q \geq 2$, *and* $q > l$, *then*

$$\sup_{(X,\mu):\mathbf{I}^{**}(X,\mu)\leq\kappa} \int_0^t \int_S m_s(x)^{q'}\,dx\,ds < \infty. \tag{6.6}$$

PROOF. By taking $\lambda(s,X) = 0$ and $f(s,u,x) = \phi(s,x)$ in (6.1) and (6.3), where $\phi \in \mathbb{C}^{1,2}(\mathbb{R}_+ \times \mathbb{R}^l)$ and the support of ϕ in x is bounded locally uniformly in s, we have that

$$\begin{aligned}
&-\frac{1}{2}\int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s(X_s,x)D^2_{xx}\phi(s,x)\bigr)\nu_s(dx)\,ds \\
&\quad - \int_0^t \int_{\mathbb{R}^l} D_x\phi(s,x)^T a_s(X_s,x)\nu_s(dx)\,ds \\
&\quad \leq \mathbf{I}^{**}(X,\mu) + \frac{1}{2}\int_0^t \int_{\mathbb{R}^l} \|D_x\phi(s,x)\|^2_{c_s(X_s,x)}\nu_s(dx)\,ds.
\end{aligned}$$

Replacing $\phi(s,x)$ with $\delta\phi(s,x)$, where $\delta > 0$, dividing through by δ, and minimizing the right-hand side over δ yields

$$\begin{aligned}
&-\frac{1}{2}\int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s(X_s,x)D^2_{xx}\phi(s,x)\bigr)\nu_s(dx)\,ds \\
&\quad - \int_0^t \int_{\mathbb{R}^l} D_x\phi(s,x)^T a_s(X_s,x)\nu_s(dx)\,ds \\
&\quad \leq \sqrt{2}\mathbf{I}^{**}(X,\mu)^{1/2}\biggl(\int_0^t \int_{\mathbb{R}^l} \|D_x\phi(s,x)\|^2_{c_s(X_s,x)}\nu_s(dx)\,ds\biggr)^{1/2}.
\end{aligned} \tag{6.7}$$

Let $\mathbb{L}^{1,2}_0([0,t]\times\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), \nu_s(dx)\,ds)$ denote the closure in $\mathbb{L}^2([0,t]\times\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), \nu_s(dx)\,ds)$ of the space of functions $D_x\phi$. By (6.7), the left-hand side extends to a continuous functional $T_t(g)$ on $\mathbb{L}^{1,2}_0([0,t]\times\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x),$

$\nu_s(dx)\,ds)$. By the Riesz representation theorem, there exists a unique $\psi \in \mathbb{L}_0^{1,2}([0,t]\times\mathbb{R}^l,\mathbb{R}^l,c_s(X_s,x),\nu_s(dx)\,ds)$ such that

$$T_t(g) = \int_0^t \int_{\mathbb{R}^l} g(s,x)^T c_s(X_s,x)\psi(s,x)\nu_s(dx)\,ds,$$

for all $g \in \mathbb{L}_0^{1,2}([0,t]\times\mathbb{R}^l,\mathbb{R}^l,c_s(X_s,x),\nu_s(dx)\,ds)$, and

$$\left(\int_0^t \int_{\mathbb{R}^l} \|\psi(s,x)\|^2_{c_s(X_s,x)} \nu_s(dx)\,ds\right)^{1/2} \le \sqrt{2}\mathbf{I}^{**}(X,\mu)^{1/2}. \tag{6.8}$$

By uniqueness, ψ can be extended to a function on $\mathbb{R}_+\times\mathbb{R}^l$ so that for all $t>0$,

$$\begin{aligned} &-\frac{1}{2}\int_0^t \int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s(X_s,x)D^2_{xx}\phi(s,x)\bigr)\nu_s(dx)\,ds \\ &\quad - \int_0^t \int_{\mathbb{R}^l} D_x\phi(s,x)^T a_s(X_s,x)\nu_s(dx)\,ds \\ &\quad = \int_0^t \int_{\mathbb{R}^l} D_x\phi(s,x)^T c_s(X_s,x)\psi(s,x)\nu_s(dx)\,ds. \end{aligned} \tag{6.9}$$

It follows that for almost all s and for all $h \in \mathbb{C}_0^2(\mathbb{R}^l)$,

$$\begin{aligned} &-\frac{1}{2}\int_{\mathbb{R}^l} \operatorname{tr}\bigl(c_s(X_s,x)D^2h(x)\bigr)\nu_s(dx) \\ &\quad = \int_{\mathbb{R}^l} Dh(x)^T a_s(X_s,x)\nu_s(dx) + \int_{\mathbb{R}^l} Dh(x)^T c_s(X_s,x)\psi(s,x)\nu_s(dx). \end{aligned} \tag{6.10}$$

Since $\psi \in \mathbb{L}_0^{1,2}([0,t]\times\mathbb{R}^l,\mathbb{R}^l,c_s(X_s,x),\nu_s(dx)\,ds)$, we have that, for almost all s, $\psi(s,\cdot)$ belongs to the closure of the set of the $D_x h$ in $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,c_s(X_s,x),\nu_s(dx))$. In particular, $\int_{\mathbb{R}^l} |\psi(s,x)|^2\nu_s(dx) < \infty$. Since $a_s(X_s,\cdot)$ and $\psi(s,\cdot)$ are locally integrable with respect to $\nu_s(dx)$ and $c_s(X_s,\cdot)$ is uniformly positive definite and is of class $\mathbb{C}^1$, (6.10) and Theorem 2.1 in Bogachev, Krylov and Röckner [8] imply that the measure $\nu_s(dx)$ has density $m_s(x)$ with respect to Lebesgue measure which belongs to $L^\beta_{\mathrm{loc}}(\mathbb{R}^l)$ for all $\beta < l'$. It follows since $a_s(X_s,\cdot)$ and $c_s(X_s,\cdot)$ are locally bounded and $\int_{\mathbb{R}^l} |\psi(s,x)|^2\nu_s(dx) < \infty$, that for arbitrary open ball S in $\mathbb{R}^l$, there exists $M>0$ such that for all $h \in \mathbb{C}_0^2(S)$:

$$\left|\int_S \operatorname{tr}\bigl(c_s(X_s,x)D^2h(x)\bigr)m_s(x)\,dx\right| \le M\|Dh\|_{\mathbb{L}^{2\beta'}(S,\mathbb{R}^l)}.$$

Since $c_s(u,\cdot)$ is uniformly positive definite and is of class $\mathbb{C}^1$, by Theorem 6.1 in Agmon [2], the density $m_s(\cdot)$ belongs to $\mathbb{W}^{1,\alpha}_{\mathrm{loc}}(S)$ for all $\alpha < 2l/(2l-1)$. The inclusion $\sqrt{m_s(\cdot)} \in \mathbb{W}^{1,2}_{\mathrm{loc}}(\mathbb{R}^l)$ follows from Lemma 6.3 and (6.10). For inequality (6.5), we also recall (6.4) and (6.8). The property that $Dm_s(\cdot)/m_s(\cdot) \in \mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^l,m_s(x)\,dx)$ when $\sqrt{m_s(\cdot)} \in \mathbb{W}^{1,2}(\mathbb{R}^l)$ follows from Lemma 6.2 and the fact that $\mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^l,m_s(x)\,dx)$ is the closure of the space of the gradients of $\mathbb{C}^\infty$-functions whose gradients belong to $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, m_s(x)\,dx)$ in $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, m_s(x)\,dx)$; cf., Theorem 1.27 on page 23 of Heinonen, Kilpeläinen and Martio [23].

We now adapt the proof of Theorem 2.1 in Bogachev, Krylov and Röckner [8] in order to obtain the bound in (6.6). Let S_1 represent an open ball which contains S. By (6.8), (6.9) and local boundedness of $a_s(u,x)$ and $c_s(u,x)$, assuming that $\phi(s,x)$ in (6.9) is supported by S_1 in x for all $s \in [0,t]$, we have that there exists $L_1 > 0$ such that for all (X,μ) that satisfy the inequality $\mathbf{I}^{**}(X,\mu) \le \delta$,

$$\begin{aligned}\left|\int_0^t \int_{S_1} \operatorname{tr}\bigl(c_s(X_s,x) D^2_{xx}\phi(s,x)\bigr) m_s(x)\,dx\,ds\right| \\ \le L_1\left(\int_0^t \sup_{x\in S_1}\bigl|D_x\phi(s,x)\bigr|^2\,ds\right)^{1/2}.\end{aligned} \tag{6.11}$$

An approximation argument shows that one may assume that $\phi(s,x)$ is measurable in (s,x) and is of class $\mathbb{C}^2$ in x. Let $\zeta(x)$ represent a $\mathbb{C}_0^\infty$-function on $\mathbb{R}^l$ with support in S_1 that equals 1 on S and let $\varphi(s,x)$ be a measurable function that is of class $\mathbb{C}^\infty$ in x. On letting $\phi(s,x) = \zeta(x)\varphi(s,x)$ in (6.11), we have that there exists $L_2 > 0$ such that for all $\varphi(s,x)$,

$$\begin{aligned}\left|\int_0^t \int_{S_1} \operatorname{tr}\bigl(c_s(X_s,x) D^2_{xx}\varphi(s,x)\bigr)\zeta(x) m_s(x)\,dx\,ds\right| \\ \le L_2\left(\int_0^t \Bigl(\sup_{x\in S_1}\bigl|\varphi(s,x)\bigr|^2 + \sup_{x\in S_1}\bigl|D_x\varphi(s,x)\bigr|^2\Bigr)\,ds\right)^{1/2}.\end{aligned}$$

By Sobolev's imbedding, $\mathbb{W}^{2,q}(S_1)$ is continuously imbedded in $\mathbb{W}^{1,\infty}(S_1)$ provided $q > l$ (see, e.g., Theorem 4.12 on page 85 in Adams and Fournier [1]), hence,

$$\begin{aligned}\left|\int_0^t \int_{S_1} \operatorname{tr}\bigl(c_s(X_s,x) D^2_{xx}\varphi(s,x)\bigr)\zeta(x) m_s(x)\,dx\,ds\right| \\ \le L_3\left(\int_0^t \|\varphi(s,\cdot)\|^2_{\mathbb{W}^{2,q}(S_1)}\,ds\right)^{1/2},\end{aligned} \tag{6.12}$$

where $L_3 > 0$. The latter inequality extends to $\varphi(s,\cdot) \in \mathbb{C}^2(\overline{S_1})$. Given a bounded continuous function $f(s,x)$ such that $f(s,\cdot) \in \mathbb{C}_0^\infty(S_1)$, let $\varphi(s,\cdot) \in \mathbb{C}^2(\overline{S_1})$ be such that $\operatorname{tr}(c_s(X_s,x)D^2_{xx}\varphi(s,x)) = f(s,x)$ and $\varphi(s,x) = 0$ on the boundary of S_1; see Theorem 6.14 on page 107 of Gilbarg and Trudinger [21]. By Theorem 9.13 on page 239 in Gilbarg and Trudinger [21], where we take $\Omega' = \Omega = S_1$, and on recalling that the norms $\|c_s(u,\cdot)\|_{\mathbb{W}^{2,q}(S_1)}$ are bounded locally in (s,u), we have that

$$\|\varphi(s,\cdot)\|_{\mathbb{W}^{2,q}(S_1)} \le L_4\bigl(\|\varphi(s,\cdot)\|_{\mathbb{L}^q(S_1)} + \|f(s,\cdot)\|_{\mathbb{L}^q(S_1)}\bigr)$$

locally uniformly in s. By Theorem 9.1 on page 220 in Gilbarg and Trudinger [21], $\sup_{x\in S_1}|\varphi(s,x)| \le L_5\|f(s,\cdot)\|_{\mathbb{L}^q(S_1)}$ locally uniformly in s. We obtain that there exists $L_6 > 0$ such that $\|\varphi(s,\cdot)\|_{\mathbb{W}^{2,q}(S_1)} \le L_6 \|f(s,\cdot)\|_{\mathbb{L}^q(S_1)}$. By (6.12), if $q \ge 2$, then, for some $L_7 > 0$,

$$\left|\int_0^t \int_{S_1} f(s,x)\zeta(x)m_s(x)\,dx\,ds\right| \le L_7 \left(\int_0^t \int_{S_1} |f(s,x)|^q\,dx\,ds\right)^{1/q}.$$

Since the functions $f(s,x)$ are dense in $\mathbb{L}^q([0,t]\times S_1)$,

$$\left(\int_0^t \int_{S_1} |\zeta(x)m_s(x)|^{q'}\,dx\,ds\right)^{1/q'} \le L_7,$$

which yields the required bound (6.6) if one recalls that $\zeta(x) = 1$ on S. □

REMARK 6.1. As a byproduct of the proof, the function $\psi(s,\cdot)$ is an element of $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$ for almost all s.

We now work toward proving that $\mathbf{I}^{**}$ is the same as $\mathbf{I}$ in Theorem 2.1 and Proposition 2.1. The following lemma will be useful for calculating $\mathbf{I}^{**}$; cf. Lemma A.2 on page 460 in Puhalskii [40].

LEMMA 6.5. *Let V represent a complete separable metric space, let U represent a dense subspace, and let $\mathbb{R}$-valued function $f(s,y)$ be defined on $\mathbb{R}_+ \times V$, be measurable in s and continuous in y. Suppose also that $f(s,\lambda(s))$ is locally integrable with respect to Lebesgue measure for all measurable functions $\lambda(s)$ that assume values in U. Then, for all $t \in \mathbb{R}_+$,*

$$\sup_{\lambda(\cdot)\in\Lambda} \int_0^t f\big(s,\lambda(s)\big)\,ds = \int_0^t \sup_{y\in U} f(s,y)\,ds,$$

where Λ represents the set of measurable functions assuming values in U.

In the rest of the paper, we denote D_x by D, divergencies are understood with respect to x. The next lemma is a key to proving that $\sqrt{m_s(\cdot)} \in \mathbb{W}^{1,2}(\mathbb{R}^l)$ in the statement of Theorem 2.1.

LEMMA 6.6. *Let $m_s(x)$, where $x \in \mathbb{R}^l$ and $s \in \mathbb{R}_+$, represent an $\mathbb{R}_+$-valued measurable function which is a probability density on $\mathbb{R}^l$ and an element of $\mathbb{W}^{1,1}_{\text{loc}}(\mathbb{R}^l)$ for almost all s. If, for some $t > 0$ and $L_1 > 0$, we have that $\int_0^t \int_S |Dm_s(x)|^2/m_s(x)\,dx\,ds < \infty$, for all open balls S, and*

$$\begin{aligned}\int_0^t \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \bigg(& Dh(x)^T\Big(\frac{1}{2}\operatorname{div}\big(c_s(X_s,x)m_s(x)\big) - a_s(X_s,x)m_s(x)\Big) \\ & - \frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)} m_s(x)\bigg)\,dx\,ds \le L_1,\end{aligned} \tag{6.13}$$

then there exists $L_2 > 0$, which depends on L_1 and t only, such that

$$\int_0^t \int_{\mathbb{R}^l} \frac{|Dm_s(x)|^2}{m_s(x)}\, dx\, ds \le L_2.$$

PROOF. Due to space constraints, we resort to a proof outline. Let $\eta(x)$ represent a $[0,1]$-valued twice continuously differentiable nonincreasing function defined for $x \ge 0$ such that $\eta(x) = 1$ for $x \in [0,1]$ and $\eta(x) = 0$ for $x \ge 2$. Let $\eta_r(x) = \eta(|x|/r)$ where $x \in \mathbb{R}^l$ and $r > 0$. We note that the bound in (6.13) extends to functions $h(x)$ from the closure $\mathbb{H}_0^{1,2}(S_{2r+1}, m_s(x)\,dx)$ of $\mathbb{C}_0^\infty(S_{2r+1})$ in $\mathbb{W}^{1,2}(S_{2r+1}, m_s(x)\,dx)$, where S_{2r+1} represents the open ball of radius $2r+1$ centered at the origin in $\mathbb{R}^l$. Let $\delta > 1$. Since (the restriction of) $\ln(m_s(\cdot) \wedge \delta \vee \delta^{-1})$ to S_{2r+1} is an element of $\mathbb{W}^{1,2}(S_{2r+1}, m_s(x)\,dx)$ a.e. and since by Lemma 6.2, $\mathbb{W}^{1,2}(S_{2r+1}, m_s(x)\,dx) = \mathbb{H}^{1,2}(S_{2r+1}, m_s(x)\,dx)$, we have that $\ln(m_s(\cdot) \wedge \delta \vee \delta^{-1}) \in \mathbb{H}^{1,2}(S_{2r+1}, m_s(x)\,dx)$. So, $\ln(m_s(\cdot) \wedge \delta \vee \delta^{-1})\eta_r(\cdot)^2$ is an element of $\mathbb{H}_0^{1,2}(S_{2r+1}, m_s(x)\,dx)$. Hence, one can take $h(x) = (1/4)\ln(m_s(x) \wedge \delta \vee \delta^{-1})\eta_r(x)^2$ in (6.13). The bound in (6.13) implies that there exist $L_1 > 0$ and $M_1 > 0$ such that, given arbitrary $\delta > 1$ and $\kappa \in (0, 1/2)$, for all r great enough (depending on δ),

$$\int_0^t \int_{\mathbb{R}^l} \frac{\|Dm_s(x)\|^2_{c_s(X_s,x)}}{m_s(x)} \eta_r(x)^2 \mathbf{1}_{\{\delta^{-1} \le m_s(x) \le \delta\}}(x)\, dx\, ds \le \frac{16}{1-2\kappa}(L_1 + M_1 t),$$

which implies the assertion of the lemma by letting $r \to \infty$ and $\delta \to \infty$. □

The next theorem establishes the equality $\mathbf{I}^{**}(X,\mu) = \mathbf{I}(X,\mu)$ provided $\mathbf{I}^{**}(X,\mu) < \infty$, $X_0 = \hat{u}$, and $\mathbf{I}_0(\hat{u}) = 0$.

THEOREM 6.1. *Suppose that Conditions 2.1, 2.2, (2.4b), (2.12c) and (2.12d) hold and that $\mathbf{I}^{**}(X,\mu) < \infty$. Then $\mu(ds,dx) = m_s(x)\,dx\,ds$, where $m_s(\cdot) \in \mathbb{P}(\mathbb{R}^l)$ a.e. We have that $\int_0^t \int_{\mathbb{R}^l} |x^T a_s(X_s,x)|/|x| m_s(x)\,dx\,ds < \infty$ and $\int_0^t \int_{\mathbb{R}^l} |Dm_s(x)|^2/m_s(x)\,dx\,ds < \infty$ for all $t \in \mathbb{R}_+$. The projection $\Phi_{s,m_s(\cdot),X_s}(x)$ belongs to $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$ as a function of x for almost every s, $\Phi_{s,m_s(\cdot),X_s}(x)$ and $\Psi_{s,m_s(\cdot),X_s}(x)$ are measurable in (s,x), and $\int_0^t \int_{\mathbb{R}^l} \|\Phi_{s,m_s(\cdot),X_s}(x)\|^2 m_s(x)\,dx\,ds < \infty$ for all $t \in \mathbb{R}_+$. We also have that*

$$\mathbf{I}^{**}(X,\mu) = \int_0^\infty \sup_{\lambda \in \mathbb{R}^n} \biggl(\lambda^T\biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x) m_s(x)\,dx\biggr) - \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x) m_s(x)\,dx} \tag{6.14}$$

$$
\begin{aligned}
&+ \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l}\Big(Dh(x)^T\Big(\frac{1}{2}\operatorname{div}\big(c_s(X_s,x)m_s(x)\big)\\
&- \big(a_s(X_s,x)+G_s(X_s,x)^T\lambda\big)m_s(x)\Big)\\
&- \frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)}m_s(x)\Big)\,dx\Big)\,ds\\
= \int_0^\infty \sup_{\lambda\in\mathbb{R}^n}&\Big(\lambda^T\Big(\dot X_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx\Big)\\
&- \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x)m_s(x)\,dx}\\
&+ \sup_{g\in\mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^l,c_s(X_s,x),m_s(x)\,dx)} \int_{\mathbb{R}^l}\Big(g(x)^Tc_s(X_s,x)\Big(\frac{Dm_s(x)}{2m_s(x)}\\
&- \Phi_{s,m_s(\cdot),X_s}(x) - \Psi_{s,m_s(\cdot),X_s}(x)\lambda\Big) - \frac{1}{2}\|g(x)\|^2_{c_s(X_s,x)}\Big)m_s(x)\,dx\Big)\,ds.
\end{aligned}
\tag{6.15}
$$

The vector $\dot X_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx - \int_{\mathbb{R}^l} G_s(X_s,x)(Dm_s(x)/(2m_s(x)) - \Phi_{s,m_s(\cdot),X_s}(x))m_s(x)\,dx$ *is in the range of* $\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx$ *a.e. and the supremum in* (6.15) *is attained at*

$$
\begin{aligned}
\hat\lambda_s = \Big(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx\Big)^{\oplus}&\Big(\dot X_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx\\
&- \int_{\mathbb{R}^l} G_s(X_s,x)\Big(\frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\Big)m_s(x)\,dx\Big)
\end{aligned}
\tag{6.16}
$$

and

$$
\hat g_s(x) = \frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x) - \Psi_{s,m_s(\cdot),X_s}(x)\hat\lambda_s
\tag{6.17}
$$

so that

$$
\begin{aligned}
\mathbf{I}^{**}(X,\mu) = \int_0^\infty\Big(\frac{1}{2}\int_{\mathbb{R}^l}&\Big\|\frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\Big\|^2_{c_s(X_s,x)}m_s(x)\,dx\\
&+ \frac{1}{2}\Big\|\dot X_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx - \int_{\mathbb{R}^l} G_s(X_s,x)\Big(\frac{Dm_s(x)}{2m_s(x)}\\
&- \Phi_{s,m_s(\cdot),X_s}(x)\Big)m_s(x)\,dx\Big\|^2_{(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx)^{\oplus}}\Big)\,ds.
\end{aligned}
\tag{6.18}
$$

PROOF. We recall the expression (6.1) for $\mathbf{I}^{**}(X,\mu)$, where the supremum is taken over $t\in\mathbb{R}_+$, functions $\lambda(s,X)$ given by (5.1), and $\mathbb{C}^{1,2,2}(\mathbb{R}_+\times\mathbb{R}^n\times\mathbb{R}^l)$-functions $f(s,u,x)$ that are compactly supported in x locally uniformly in (t,u).

According to Lemma 6.4, if $\mathbf{I}^{**}(X,\mu) < \infty$, then $\nu_s(dx) = m_s(x)\,dx$, where $m_s(\cdot) \in \mathbb{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^l)$, so one can integrate by parts in (6.3) to obtain

$$
\begin{aligned}
&U_t^{\lambda(\cdot),f}(X,\mu) \\
&\quad = \int_0^t \biggl(\lambda(s,X)^T \biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x) m_s(x)\,dx\biggr) \\
&\qquad - \frac{1}{2}\int_{\mathbb{R}^l} \|\lambda(s,X)\|^2_{C_s(X_s,x)} m_s(x)\,dx \\
&\qquad + \int_{\mathbb{R}^l} Df(s,X_s,x)^T \biggl(\frac{1}{2}\operatorname{div}\bigl(c_s(X_s,x)m_s(x)\bigr) - a_s(X_s,x)m_s(x)\biggr)\,dx \\
&\qquad - \frac{1}{2}\int_{\mathbb{R}^l} \|Df(s,X_s,x)\|^2_{c_s(X_s,x)} m_s(x)\,dx \\
&\qquad - \int_{\mathbb{R}^l} \lambda(s,X)^T G_s(X_s,x) Df(s,X_s,x) m_s(x)\,dx\biggr)\,ds.
\end{aligned}
\tag{6.19}
$$

An approximation argument using mollifiers implies that the supremum will not change if $\lambda(s,X)$ is assumed bounded and measurable in s and if $f(s,u,x)$ is assumed measurable, continuously differentiable in x with bounded first partial derivatives and compactly supported in x locally uniformly in (s,u). Therefore, on noting that X is kept fixed,

$$
\begin{aligned}
\mathbf{I}^{**}(X,\mu) = \sup \int_0^t \biggl(&\lambda(s)^T \biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x) m_s(x)\,dx\biggr) \\
&- \frac{1}{2}\|\lambda(s)\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x) m_s(x)\,dx} \\
&+ \int_{\mathbb{R}^l} D\phi(s,x)^T \biggl(\frac{1}{2}\operatorname{div}_x\bigl(c_s(X_s,x)m_s(x)\bigr) - a_s(X_s,x)m_s(x)\biggr)\,dx \\
&- \frac{1}{2}\int_{\mathbb{R}^l} \|D\phi(s,x)\|^2_{c_s(X_s,x)} m_s(x)\,dx \\
&- \int_{\mathbb{R}^l} \lambda(s)^T G_s(X_s,x) D\phi(s,x) m_s(x)\,dx\biggr)\,ds,
\end{aligned}
$$

where the supremum is taken over $t \in \mathbb{R}_+$, bounded measurable functions $\lambda(s)$, and measurable functions $\phi(s,x)$ that are continuously differentiable in x with bounded first partial derivatives and are compactly supported in x locally uniformly in s. By Lemma 6.5, one can optimize with respect to $\lambda(s)$ and $D\phi(s,x)$ inside the ds-integral which yields (6.14). In some more detail, we apply Lemma 6.5 with U being the Cartesian product of the closed ball of radius i in $\mathbb{R}^n$ and of the set $U_i = \{Dh : h \in \mathbb{C}_0^1(\mathbb{R}^l), \sup_{x\in\mathbb{R}^l} |Dh(x)| \le i \text{ and } h(x) = 0 \text{ if } |x| \ge i\}$ and with V being the Cartesian product of the closed ball of radius i and of the closure of U_i in the space of continuous functions with support in the open ball of radius i

centered at the origin in $\mathbb{R}^l$ that are bounded above by i in absolute value; the latter space being endowed with the sup-norm topology, where $i \in \mathbb{N}$, and let $i \to \infty$.

Integration by parts in (6.10), with $\nu_s(dx) = m_s(x)\,dx$, yields

$$\int_{\mathbb{R}^l} Dh(x)^T \left(\frac{1}{2}\operatorname{div}\big(c_s(X_s,x)m_s(x)\big) - a_s(X_s,x)m_s(x)\right)dx$$
$$= \int_{\mathbb{R}^l} Dh(x)^T c_s(X_s,x)\psi(s,x)m_s(x)\,dx.$$

On recalling that $\psi(s,\cdot) \in \mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$ for almost all s by Remark 6.1, we have that the function $-\psi(s,x)$ represents the orthogonal projection of $c_s(X_s,x)^{-1}(a_s(X_s,x) - (1/2)\operatorname{div}(c_s(X_s,x)m_s(x))/m_s(x))$ onto $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$. Since by (6.14), Lemmas 6.4 and 6.6, $Dm_s(x)/m_s(x)$ is a member of $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$ for almost all s, we have that the function $-\psi(s,x) + (1/2)Dm_s(x)/m_s(x)$ belongs to $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$ for almost all s, so, by (2.15b), it equals $\Phi_{s,m_s(\cdot),X_s}(x)$.

We show that $\Phi_{s,m_s(\cdot),X_s}(x)$ and $\Psi_{s,m_s(\cdot),X_s}(x)$ are properly measurable. Let $\mathcal{U}_s$ represent the closure of the set $\{c_s(X_s,\cdot)^{1/2}\sqrt{m_s(\cdot)}Dp(\cdot) : p \in \mathbb{C}_0^\infty(\mathbb{R}^l, \mathbb{R}^n)\}$ in $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^{l\times n})$. Introducing $\varphi_s(x) = c_s(X_s,x)^{-1/2}G_s(X_s,x)^T\sqrt{m_s(x)}$ and $\hat\varphi_s(x) = c_s(X_s,x)^{1/2}\Psi_{s,m_s(\cdot),X_s}(x)\sqrt{m_s(x)}$, we have that $\hat\varphi_s$ is the orthogonal projection of φ_s onto $\mathcal{U}_s$ [see (2.15a) and (2.3)]. By Corollary 8.2.13 on page 317 in Aubin and Frankowska [4], $\hat\varphi_s$ is a measurable function from $\mathbb{R}_+$ to $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^{l\times n})$. (We note that $s \to \mathcal{U}_s$ is a measurable set-valued map by part (vi) of Theorem 8.1.4 on page 310 in Aubin and Frankowska [4].) This implies that the mapping $(s,x) \to \Psi_{s,m_s(\cdot),X_s}(x)$ is measurable. The reasoning for $\Phi_{s,m_s(\cdot),X_s}$ is similar.

The representation in (6.15) follows from (2.15a), (2.15b), (2.3) and (6.14). Since the function

$$\tilde g_s(x) = \frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x) - \Psi_{s,m_s(\cdot),X_s}(x)\lambda$$

is a member of $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, c_s(X_s,x), m_s(x)\,dx)$, it attains the supremum in (6.15), which yields

$$\begin{aligned}
&\mathbf{I}^{**}(X,\mu)\\
&\quad= \int_0^\infty \sup_{\lambda\in\mathbb{R}^n}\Bigg(\lambda^T\left(\dot X_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx\right) - \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x)m_s(x)\,dx}\\
&\qquad + \frac{1}{2}\int_{\mathbb{R}^l}\bigg\|\Phi_{s,m_s(\cdot),X_s}(x) - \frac{Dm_s(x)}{2m_s(x)}\\
&\qquad - \Psi_{s,m_s(\cdot),X_s}(x)\lambda\bigg\|^2_{c_s(X_s,x)} m_s(x)\,dx\Bigg)\,ds.
\end{aligned} \tag{6.20}$$

Since the matrix $Q_{s,m_s(\cdot)}(u,x) = C_s(u,x) - \|\Psi_{s,m_s(\cdot),u}(x)\|^2_{c_s(u,x)}$ [see (2.14)] is positive semidefinite, the supremum over λ in (6.20) is attained at

$$\tilde{\lambda} = \left(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx\right)^{\oplus}\biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx \\ - \int_{\mathbb{R}^l} \Psi_{s,m_s(\cdot),X_s}(x)^T c_s(X_s,x)\left(\Phi_{s,m_s(\cdot),X_s}(x) - \frac{Dm_s(x)}{2m_s(x)}\right)m_s(x)\,dx\biggr)$$

and equals

$$\frac{1}{2}\int_{\mathbb{R}^l}\left\|\frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}\right\|^2_{c_s(X_s,x)} m_s(x)\,dx \\ + \frac{1}{2}\|\tilde{\lambda}\|^2_{(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx)^{\oplus}},$$

provided

$$\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx \\ - \int_{\mathbb{R}^l} \Psi_{s,m_s(\cdot),X_s}(x)^T c_s(X_s,x)\left(\frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\right)m_s(x)\,dx$$

is in the range of $\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx$ a.e. Otherwise, the supremum equals infinity. The fact that $\tilde{\lambda} = \hat{\lambda}_s$ and the expression in (6.18) follow from (2.15a) and (6.20). The properties that $\int_0^t\int_{\mathbb{R}^l} |Dm_s(x)|^2/m_s(x)\,dx\,ds$ and $\int_0^t\int_{\mathbb{R}^l} \|\Phi_{s,m_s(\cdot),X_s}(x)\|^2\,dx\,ds$ are finite follow from Lemma 6.6, (6.14), and (6.18). The integral $\int_0^t\int_{\mathbb{R}^l} |x^T a_s(X_s,x)|/|x| m_s(x)\,dx\,ds$ being finite follows from (6.14) if one lets $\lambda = 0$, takes as $h(x)$ a smoothing of the function $-(|x|\wedge\delta)\eta(|x|/r)$, where $\eta(y)$ satisfies the hypotheses of Condition 2.3, and lets $r\to\infty$, first, and $\delta\to\infty$, next. □

Motivated by (6.18) in Theorem 6.1, let us introduce, provided $\mathbf{I}^{**}(X,\mu) < \infty$ so that $\mu(ds,dx) = m_s(x)\,dx\,ds$, where $m_s(\cdot)\in\mathbb{P}(\mathbb{R}^l)$, and $\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx - \int_{\mathbb{R}^l} G_s(X_s,x)(Dm_s(x)/(2m_s(x)) - \Phi_{s,m_s(\cdot),X_s})m_s(x)\,dx$ is in the range of $\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx$ a.e.,

$$\mathbf{I}_t^{**}(X,\mu) = \int_0^t\biggl(\frac{1}{2}\int_{\mathbb{R}^l}\left\|\frac{Dm_s(x)}{2m_s(x)} - \Phi_{s,m_s(\cdot),X_s}(x)\right\|^2_{c_s(X_s,x)} m_s(x)\,dx \\ + \frac{1}{2}\biggl\|\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx - \int_{\mathbb{R}^l} G_s(X_s,x)\biggl(\frac{Dm_s(x)}{2m_s(x)} \\ - \Phi_{s,m_s(\cdot),X_s}(x)\biggr)m_s(x)\,dx\biggr\|^2_{(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx)^{\oplus}}\biggr)\,ds. \tag{6.21}$$

As in the proof of Theorem 6.1, we also have that

$$
\begin{aligned}
\mathbf{I}_t^{**}(X,\mu) = \int_0^t \sup_{\lambda\in\mathbb{R}^n}\biggl(\lambda^T\Bigl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx\Bigr) \\
- \frac{1}{2}\|\lambda\|^2_{\int_{\mathbb{R}^l} C_s(X_s,x)m_s(x)\,dx} \\
+ \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\biggl(Dh(x)^T\Bigl(\frac{1}{2}\operatorname{div}\bigl(c_s(X_s,x)m_s(x)\bigr) \\
- \bigl(a_s(X_s,x)+G_s(X_s,x)^T\lambda\bigr)m_s(x)\Bigr) \\
- \frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)}m_s(x)\biggr)\,dx\biggr)\,ds.
\end{aligned}
\tag{6.22}
$$

For the proof of Theorem 8.1, it will be needed to extend (X,μ) defined on $[0,t]$ past t in such a way that $\mathbf{I}_t^{**}(X,\mu)=\mathbf{I}^{**}(X,\mu)$. That is done in the following lemma which also concerns the zeros of $\mathbf{I}^{**}(X,\mu)$.

LEMMA 6.7. *For $t\in\mathbb{R}_+$ and $z\in\mathbb{R}^n$, the system of equations*

$$
\dot{X}_s = \int_{\mathbb{R}^l} A_{s+t}(X_s,x)m_s(x)\,dx, \qquad X_0=z, \tag{6.23}
$$

$$
\int_{\mathbb{R}^l}\Bigl(\frac{1}{2}\operatorname{tr}\bigl(c_{s+t}(X_s,x)D^2p(x)\bigr)+a_{s+t}(X_s,x)^T Dp(x)\Bigr)m_s(x)\,dx=0, \tag{6.24}
$$

where $p\in\mathbb{C}_0^\infty(\mathbb{R}^l)$ is otherwise arbitrary, has a solution $(X^\dagger,(m_s^\dagger(x)))$ such that $X^\dagger$ is locally Lipschitz continuous, $m_s^\dagger(x)$ is measurable, and $m_s^\dagger(\cdot)\in\mathbb{P}(\mathbb{R}^l)$. If, given (X,μ) such that $\mathbf{I}^{**}(X,\mu)<\infty$, one defines $(\hat{X},\hat{\mu})$ by the relations $\hat{X}_s=X_s$ and $\hat{\mu}_s=\mu_s$ for $s\le t$, and $\hat{X}_s=X^\dagger_{s-t}$ and $\hat{\mu}_s(dx)=\mu_t(dx)+\int_0^{s-t}m_r^\dagger(dx)\,dr$ for $s>t$, where $z=X_t$, then $\mathbf{I}^{**}(\hat{X},\hat{\mu})=\mathbf{I}_t^{**}(X,\mu)$. In particular, if $t=0$, then $\mathbf{I}^{**}(X^\dagger,\mu^\dagger)=0$.

PROOF. Since $a_s(u,x)$ is locally bounded, since $c_s(u,x)$ is bounded, is positive definite and is of class $\mathbb{C}^1$ in x, and since $a_s(u,x)^T x/|x|\to-\infty$ as $|x|\to\infty$ by (2.4b), applications of Theorem 1.4.1 in Bogachev, Krylov and Rëkner [7] [with $V(x)=\sqrt{1+|x|^2}$] and of Theorem 2.2 and Proposition 2.4 in Metafune, Pallara and Rhandi [31], show that for every $s,t\in\mathbb{R}_+$ and $u\in\mathbb{R}^n$ there exists a unique probability density $m_s(x)$ satisfying the equation

$$
\int_{\mathbb{R}^l}\Bigl(\frac{1}{2}\operatorname{tr}\bigl(c_{s+t}(u,x)D^2p(x)\bigr)+Dp(x)^T a_{s+t}(u,x)\Bigr)m_s(x)\,dx=0. \tag{6.25}
$$

We apply the method of successive approximations: let $X_s^0 = z$ and, for $i \in \mathbb{N}$,

$$\int_{\mathbb{R}^l} \left(\frac{1}{2} \operatorname{tr}\big(c_{s+t}(X_s^i, x) D^2 p(x)\big) + Dp(x)^T a_{s+t}(X_s^i, x) \right) m_s^i(x)\, dx = 0, \tag{6.26}$$

$$\dot{X}_s^{i+1} = \int_{\mathbb{R}^l} A_{s+t}(X_s^{i+1}, x) m_s^i(x)\, dx, \qquad X_0^{i+1} = z. \tag{6.27}$$

We note that $m_s^i(x)$ is a measurable function of (s, x) (one can use, e.g., Theorem 8.2.9 on page 315 in Aubin and Frankowska [4]). By (2.12c), we have that given $L > 0$, there exists $M > 0$ such that a.e. in $s \in [0, L]$, $d|X_s^{i+1}|^2/ds \le M(1 + |X_s^{i+1}|^2)$. Gronwall's inequality implies that $\sup_{i\in\mathbb{N}} \sup_{s\in[0,L]} |X_s^i| < \infty$. By (6.27) and (2.12b), the derivatives $\dot{X}_s^{i+1}$ are bounded uniformly in $i \in \mathbb{N}$ and $s \in [0, L]$, so the sequence $(X_s^i, s \in [0, L])$ is relatively compact for the uniform norm on $[0, L]$. Let $X_s^\dagger$ represent a limit point. It is a locally Lipschitz continuous function.

As in Metafune, Pallara and Rhandi [31], Proposition 2.4, we have that, for arbitrary $\delta > 0$ and $L > 0$,

$$\sup_{s\in[0,L]} \sup_{i\in\mathbb{N}} \int_{\mathbb{R}^l} e^{\delta|x|} m_s^i(x)\, dx < \infty. \tag{6.28}$$

In some more detail, let for a function p which is twice differentiable at x,

$$\mathcal{L}_s^i p(x) = \tfrac{1}{2} \operatorname{tr}\big(c_{s+t}(X_s^i, x) D^2 p(x)\big) + Dp(x)^T a_{s+t}(X_s^i, x).$$

Since, for $|x| > 0$,

$$\begin{aligned} \mathcal{L}_s^i e^{\delta|x|} = \Bigg(\frac{1}{2} \operatorname{tr}\bigg(c_{s+t}(X_s^i, x) \bigg(\frac{\delta}{|x|} \bigg(I - \frac{x x^T}{|x|^2} \bigg) + \delta^2 \frac{x x^T}{|x|^2} \bigg) \bigg) \\ + \delta a_{s+t}^i (X_s^i, x)^T \frac{x}{|x|} \Bigg) e^{\delta|x|}, \end{aligned}$$

where I represents the $l \times l$ identity matrix, and $\sup_{i\in\mathbb{N}} a_{s+t}(X_s^i, x)^T x/|x| \to -\infty$ as $|x| \to \infty$, there exists $R > 1$ such that $\mathcal{L}_s^i e^{\delta|x|} \le 0$ and $e^{\delta|x|} \le |\mathcal{L}_s^i e^{\delta|x|}|$ for all $s \in [0, t]$ and all $i \in \mathbb{N}$ provided $|x| > R$. Let F be a $\mathbb{C}^\infty(\mathbb{R}^l)$-function such that $F(x) = e^{\delta|x|}$ if $|x| \ge 1$. Arguing as in the proof of Proposition 2.3 in Metafune, Pallara and Rhandi [31], one can see that

$$\int_{x\in\mathbb{R}^l:|x|>R} \big|\mathcal{L}_s^i e^{\delta|x|}\big| m_s^i(x)\, dx \le \int_{x\in\mathbb{R}^l:|x|\le R} \mathcal{L}_s^i F(x) m_s^i(x)\, dx$$

so that

$$\int_{x\in\mathbb{R}^l:|x|>R} e^{\delta|x|} m_s^i(x)\, dx \le \int_{x\in\mathbb{R}^l:|x|\le R} \mathcal{L}_s^i F(x) m_s^i(x)\, dx, \tag{6.29}$$

which implies (6.28).

Hence, given $s \in [0, L]$, the sequence of probability measures $m_s^i(x)\,dx$ is tight. Proposition 2.16 in Bogachev, Krylov and Röckner [8] implies that the $m_s^i(x)$ converge in the variation norm along a subsequence to a density $m_s^\dagger(x)$. Since the local $\mathbb{L}^q$-norms of the $m_s^i(x)$ are uniformly bounded for all $q > 1$ (see (2.26) in Bogachev, Krylov and Röckner [8]), $\sup_{i\in\mathbb{N}} |a_{s+t}(X_s^i, x)|$ grows no faster than linearly with x by Lipschitz continuity and $\sup_{x\in\mathbb{R}^l}\sup_{i\in\mathbb{N}} \|c_{s+t}^i(X_s^i, x)\| < \infty$ (see Condition 2.1), and $\sup_{i\in\mathbb{N}} \int_{\mathbb{R}^l} e^{\delta|x|} m_s^i(x)\,dx < \infty$, on taking a limit in (6.26), we have by dominated convergence that (6.24) holds. Since density $m_s^\dagger(x)$ is specified uniquely by (6.24) $m_s^i(x) \to m_s^\dagger(x)$ as $i \to \infty$ along a subsequence such that the X^i converge to $X^\dagger$. Since $\sup_{i\in\mathbb{N}}\sup_{x\in\mathbb{R}^l} |A_{s+t}(X_s^i, x)| < \infty$ by (2.12b), a similar reasoning shows that taking the above subsequential limit in (6.27) obtains (6.23). Since (6.28) implies that $\int_{\mathbb{R}^l} |a_s(X_s, x)|^2 m_s^\dagger(x)\,dx < \infty$, by Theorem 1.1 in Bogachev, Krylov and Röckner [6], $\sqrt{m_s^\dagger(\cdot)} \in \mathbb{W}^{1,2}(\mathbb{R}^l)$.

On noting that (6.25) can be written as

$$\int_{\mathbb{R}^l} Dp(x)^T \Big(a_{s+t}(u, x) - \frac{1}{2}\operatorname{div} c_{s+t}(u, x)\Big) m_s(x)\,dx \\ = \frac{1}{2}\int_{\mathbb{R}^l} Dp(x)^T c_{s+t}(u, x) Dm_s(x)\,dx,$$

we have that $\Phi_{s+t, m_s^\dagger(\cdot), X_s^\dagger}(x) = Dm_s^\dagger(x)/(2m_s^\dagger(x))$ which implies, by (6.18) and (6.21), that $\mathbf{I}^{**}(\hat{X}, \hat{\mu}) = \mathbf{I}_t^{**}(X, \mu)$. □

REMARK 6.2. By Proposition 2.4 and Theorem 6.1 (with $\beta = 1$) in Metafune, Pallari and Randi [31], $m_s^\dagger(\cdot)$ decays exponentially at infinity. It is also positive and Hölder continuous; see Bogachev, Krylov and Röckner [8], Theorem 2.8, Corollaries 2.10, 2.11 and Bogachev, Krylov and Rëkner [7].

7. Identifying the large deviation function. The purpose of this section is to show that $\tilde{\mathbf{I}} = \mathbf{I}^{**}$ for sufficiently regular functions (X, μ). More specifically, we will prove the following theorem.

THEOREM 7.1. *Suppose that Conditions 2.1, 2.2, (2.4b) and (2.12d) hold. Suppose that $\tilde{\mathbf{I}}$ is a large deviation function that satisfies the assertion of Theorem 5.1 and is such that $\tilde{\mathbf{I}}(X, \mu) = \infty$ unless $X_0 = \hat{u}$. Suppose that $(\hat{X}, \hat{\mu})$ is such that $\hat{X}_0 = \hat{u}$, $\mathbf{I}^{**}(\hat{X}, \hat{\mu}) < \infty$, $\hat{X}$ is locally Lipschitz continuous and that $\hat{m}_s(x) = \hat{\mu}(ds, dx)/(ds\,dx)$ is of the form*

$$\hat{m}_s(x) = M_s\Big(\tilde{m}_s(x)\hat{\eta}^2\Big(\frac{|x|}{r}\Big) + e^{-\alpha|x|}\Big(1 - \hat{\eta}^2\Big(\frac{|x|}{r}\Big)\Big)\Big),$$

where $\tilde{m}_s(x)$ is a probability density in x which is locally bounded away from zero and belongs to $\mathbb{C}^1(\mathbb{R}^l)$ as a function of x, with $|Dm_s(x)|$ being locally bounded *in* (s,x), $\hat{\eta}(y)$ *is a nonincreasing* $[0,1]$-*valued* $\mathbb{C}_0^1(\mathbb{R}_+)$-*function*, *where* $y \in \mathbb{R}_+$, *that equals* 1 *for* $y \in [0,1]$ *and equals* 0 *for* $y \geq 2$, $r > 0$, $\alpha > 0$, *and* M_s *is the normalizing constant*. *Then*, *for given* $\tilde{m}_s(x)$, $\hat{\eta}(y)$, *and* r, *there exists* $\alpha_0 > 0$ *such that* $\tilde{\mathbf{I}}(\hat{X},\hat{\mu}) = \mathbf{I}^{**}(\hat{X},\hat{\mu})$ *for all* $\alpha > \alpha_0$.

We assume throughout the section the hypotheses of Theorem 7.1 to hold. We start by extending the assertion of Theorem 5.1 to a larger set of functions $(\lambda(\cdot), f)$. For economy of notation, we denote $\gamma = (X,\mu)$ and recall that Γ represents the set of γ such that X is absolutely continuous and μ admits density $m_s(x)$ that is an element of $\mathbb{P}(\mathbb{R}^l)$ in x, for almost all s. Let $\lambda(s,X)$, where $s \in \mathbb{R}_+$ and $X \in \mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)$, represent an $\mathbb{R}^n$-valued measurable function and let $h_s(u,x)$, where $s \in \mathbb{R}_+, u \in \mathbb{R}^n$ and $x \in \mathbb{R}^l$, represent an $\mathbb{R}$-valued measurable function, which is an element of $\mathbb{W}^{1,1}_{\text{loc}}(\mathbb{R}^l)$ in x and is of bounded support in x locally uniformly over (s,u). If, for all $t \in \mathbb{R}_+$ and all $\gamma \in \mathbb{C}(\mathbb{R}_+,\mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$, $\int_0^t\int_{\mathbb{R}^l}(|\lambda(s,X)|^2 + |Dh_s(X_s,x)|^2)\mu(dx,ds) < \infty$, we define, given $N \in \mathbb{N}$,

$$\tau^N(\gamma) = \inf\biggl\{t \in \mathbb{R}_+ : \int_0^t\int_{\mathbb{R}^l}\bigl(\|\lambda(s,X)\|^2_{C_s(X_s,x)} + \|Dh_s(X_s,x)\|^2_{c_s(X_s,x)}\bigr)\mu(dx,ds) + X_t^* + t \geq N\biggr\} \tag{7.1a}$$

and, provided $\gamma \in \Gamma$,

$$\begin{aligned}\theta^N(\gamma) = \int_0^{\tau^N(\gamma)}\biggl(&\lambda(s,X)^T\biggl(\dot{X}_s - \int_{\mathbb{R}^l}A_s(X_s,x)m_s(x)\,dx\biggr)\\ &- \frac{1}{2}\|\lambda(s,X)\|^2_{\int_{\mathbb{R}^l}C_s(X_s,x)m_s(x)\,dx}\\ &+ \int_{\mathbb{R}^l}\biggl(Dh_s(X_s,x)^T\biggl(\frac{1}{2}\operatorname{div}\bigl(c_s(X_s,x)m_s(x)\bigr) - a_s(X_s,x)m_s(x)\biggr)\\ &- \frac{1}{2}\|Dh_s(X_s,x)\|^2_{c_s(X_s,x)}m_s(x)\biggr)\,dx\\ &- \int_{\mathbb{R}^l}\lambda(s,X)^TG_s(X_s,x)Dh_s(X_s,x)m_s(x)\,dx\biggr)\,ds.\end{aligned} \tag{7.1b}$$

For the latter definition, we assume that, in addition,

$$\int_0^t\biggl(|\dot{X}_s|^2 + \int_{\mathbb{R}^l}\frac{|Dm_s(x)|^2}{m_s(x)}\,dx\biggr)\,ds < \infty, \tag{7.2}$$

for all $t \in \mathbb{R}_+$, and use the piece of notation $X_t^* = \sup_{s\in[0,t]}|X_s|$. [The definition of $\theta^N(\gamma)$ is modeled on the expression for $U_t^{\lambda(\cdot),f}(X,\mu)$ in (6.19).] We note that $\tau^N(\gamma) \leq N$. Furthermore, we have the following lemma, for which we reuse the piece of notation of Theorem 3.4 that, for $\delta \in \mathbb{R}_+$,

$$K_\delta = \{\gamma : \tilde{\mathbf{I}}(\gamma) \leq \delta\}$$

and recall that K_δ is a compact in $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ and that $K_\delta \subset \Gamma$. Theorem 6.1 implies that (7.2) holds on K_δ. For the definition of the essential supremum of a family of measurable functions used in the next lemma, see, for example, Proposition II.4.1 on page 44 of Neveu [34].

LEMMA 7.1. *Let $\lambda^i(s, X)$ and $h_s^i(u, x)$ be sequences of functions satisfying the same hypotheses as $\lambda(s, X)$ and $h_s(u, x)$, respectively, and let $\tau^{N,i}(\gamma)$ and $\theta^{N,i}(\gamma)$ be defined by the respective equations* (7.1a) *and* (7.1b), *with $\lambda^i(s, X)$ and $h_s^i(u, x)$ being substituted for $\lambda(s, X)$ and $h_s(u, x)$, respectively. If, in addition, the functions $h_s^i(u, x)$ are of bounded support in x uniformly over i and locally uniformly over (s, u),*

$$\int_0^N \operatorname*{ess\,sup}_{\gamma \in K_\delta} |\lambda(s, X)|^2 \, ds + \int_0^N \operatorname*{ess\,sup}_{\gamma \in K_\delta} \int_{\mathbb{R}^l} |Dh_s(X_s, x)|^2 m_s(x)\, dx\, ds < \infty, \tag{7.3}$$

$$\lim_{i\to\infty} \sup_{\gamma \in K_\delta} \int_0^N |\lambda(s, X) - \lambda^i(s, X)|^2 \, ds = 0 \tag{7.4a}$$

and

$$\lim_{i\to\infty} \sup_{\gamma \in K_\delta} \int_0^N \int_{\mathbb{R}^l} |Dh_s(X_s, x) - Dh_s^i(X_s, x)|^2 m_s(x)\, dx\, ds = 0, \tag{7.4b}$$

then

$$\lim_{i\to\infty} \sup_{\gamma \in K_\delta} |\tau^N(\gamma) - \tau^{N,i}(\gamma)| = 0 \tag{7.5a}$$

and

$$\lim_{i\to\infty} \sup_{\gamma \in K_\delta} |\theta^N(\gamma) - \theta^{N,i}(\gamma)| = 0. \tag{7.5b}$$

PROOF. Let us note that under the hypotheses,

$$\lim_{i\to\infty} \sup_{\gamma \in K_\delta} \int_0^N \int_{\mathbb{R}^l} \big| \|\lambda^i(s, X)\|^2_{C_s(X_s, x)} - \|\lambda(s, X)\|^2_{C_s(X_s, x)} \big| m_s(x)\, dx\, ds = 0, \tag{7.6a}$$

$$\lim_{i\to\infty} \sup_{\gamma \in K_\delta} \int_0^N \int_{\mathbb{R}^l} \big| \|Dh_s^i(X_s, x)\|^2_{c_s(X_s, x)} - \|Dh_s(X_s, x)\|^2_{c_s(X_s, x)} \big| m_s(x)\, dx\, ds = 0, \tag{7.6b}$$

$$\lim_{i\to\infty}\sup_{\gamma\in K_\delta}\int_0^N \Big|\big(\lambda^i(s,X)-\lambda(s,X)\big)^T \times\Big(\dot X_s-\int_{\mathbb{R}^l}A_s(X_s,x)m_s(x)\Big)\,dx\Big|\,ds=0 \tag{7.6c}$$

and

$$\lim_{i\to\infty}\sup_{\gamma\in K_\delta}\int_0^N \Big|\int_{\mathbb{R}^l}\big(Dh_s^i(X_s,x)-Dh_s(X_s,x)\big)^T \times\Big(\frac{1}{2}\operatorname{div}\big(c_s(X_s,x)m_s(x)\big)-a_s(X_s,x)m_s(x)\Big)\,dx\Big|\,ds=0. \tag{7.6d}$$

The first two convergences are implied by (7.4a), (2.12d) and (7.4b), (2.12a), respectively, and (7.3). The convergence in (7.6c) follows via Cauchy's inequality from (7.4a) and the fact that, according to (6.18) in Theorem 6.1,

$$\sup_{(X,\mu):\mathbf{I}^{**}(X,\mu)\le\delta}\int_0^N\Big|\dot X_s-\int_{\mathbb{R}^l}A_s(X_s,x)m_s(x)\,dx\Big|^2\,ds<\infty. \tag{7.7}$$

Similarly, (7.6a) is a consequence of (7.4b), if one recalls that the functions involved are of uniformly bounded support in x and takes into account part (6.5) of Lemma 6.4.

The convergence in (7.5a) follows from (7.6a), (7.6b) and the observation that by (7.1a)

$$\begin{aligned}
&\big|\tau^N(\gamma)-\tau^{N,i}(\gamma)\big|\\
&\quad\le\int_0^N\Big|\int_{\mathbb{R}^l}\big(\|\lambda(s,X)\|^2_{C_s(X_s,x)}-\|\lambda^i(s,X)\|^2_{C_s(X_s,x)}\\
&\qquad+\|Dh_s(X_s,x)\|^2_{c_s(X_s,x)}-\|Dh_s^i(X_s,x)\|^2_{c_s(X_s,x)}\big)m_s(x)\,dx\Big|\,ds.
\end{aligned}$$

The convergence in (7.5b) follows by (7.1b), (7.5a), (7.6a)–(7.6d) and (7.7), if one notes that, thanks to (7.3),

$$\sup_{\gamma\in K_\delta}\int_0^t|\lambda(s,X)|^2\,ds,\qquad \sup_{\gamma\in K_\delta}\int_0^t\int_{\mathbb{R}^l}|Dh_s(X_s,x)|^2m_s(x)\,dx\,ds$$

and

$$\sup_{\gamma\in K_\delta}\int_0^t\int_{\mathbb{R}^l}Dh_s(X_s,x)^T\Big(\frac{1}{2}\operatorname{div}\big(c_s(X_s,x)m_s(x)\big)-a_s(X_s,x)m_s(x)\Big)\,dx\,ds$$

are continuous functions of $t\in[0,N]$. □

LEMMA 7.2. *Let $\lambda_s(u)$ represent an $\mathbb{R}^n$-valued function of $(s,u)\in\mathbb{R}_+\times\mathbb{R}^n$, which is measurable in s, is continuous in u for almost all s and is* *such that* $\int_0^N \sup_{|u|\le L} |\lambda_s(u)|^2\, ds < \infty$ *for all* $L > 0$. *Suppose that the function* $h_s(u,x)$, *in addition to being measurable and being of class* $\mathbb{W}^{1,1}_{\mathrm{loc}}$ *in* x, *vanishes when* x *is outside of some open ball in* $\mathbb{R}^l$ *locally uniformly in* (s,u), *that the function* $Dh_s(u,x)$ *is continuous in* (u,x) *for almost all* $s \in \mathbb{R}_+$, *and that* $\int_0^N \sup_{u\in\mathbb{R}^n:|u|\le L} \int_{\mathbb{R}^l} |Dh_s(u,x)|^q\, dx\, ds < \infty$ *for all* $q > 1$ *and* $L > 0$. *Then, under the hypotheses of Theorem* 7.1, *the function* $\theta^N(\gamma)$, *where* $\lambda(s,X) = \lambda_s(X_s)$, *is continuous in* γ *when restricted to* K_δ,

$$\sup_{\gamma\in\Gamma}\big(\theta^N(\gamma) - \tilde{\mathbf{I}}(\gamma)\big) = 0$$

and the latter supremum is attained. Furthermore,

$$\sup_{\gamma\in K_{2N+2}}\big(\theta^N(\gamma) - \tilde{\mathbf{I}}(\gamma)\big) = 0.$$

PROOF. The functions $|\lambda_s(u)|\mathbf{1}_{\{|\lambda_s(u)|\ge r\}}(s,u)$ are upper semicontinuous in u and monotonically decreasing in r, so by Dini's theorem $|\lambda_s(u)|^2 \times \mathbf{1}_{\{|\lambda_s(u)|\ge r\}}(s,u) \to 0$ as $r \to \infty$ uniformly on $\{u \in \mathbb{R}^n : |u| \le L\}$. Let r_i be such that $\int_0^N \sup_{u\in\mathbb{R}^n:|u|\le L} |\lambda_s(u)|^2 \mathbf{1}_{\{|\lambda_s(u)|\ge r_i\}}(s,u)\, ds < 1/i$, where $L = \sup_{\gamma\in K_\delta}\sup_{s\in[0,t]} |X_s|$ and $i \in \mathbb{N}$. Since $\lambda_s(u)$ is a Carathéodory function, as a consequence of the Scorza–Dragoni theorem (see, e.g., page 235 in Ekeland and Temam [14]), there exists a measurable function $\breve{\lambda}^i_s(u)$ that is continuous in (s,u), is bounded above in absolute value by r_i, and is such that $\int_0^N \mathbf{1}_{\{\lambda_s(\cdot)\ne\breve{\lambda}^i_s(\cdot)\}}(s)\, ds < 2/(ir_i^2)$. Letting $\lambda^i(s,X) = \breve{\lambda}^i_{\lfloor j(i)s\rfloor/j(i)}(X_{\lfloor j(i)s\rfloor/j(i)})$, where $j(i)$ is great enough and $j(i) \to \infty$ as $i \to \infty$, we have that (7.4a) holds.

Similarly, let

$$h^i_s(u,x) = \int_{\mathbb{R}\times\mathbb{R}^l} \rho_{1/i}(\tilde{s},y) h_{s-\tilde{s}}(u,x-y)\, d\tilde{s}\, dy,$$

where $\rho_\kappa(\tilde{s},y) = (\tilde{\rho}_1(\tilde{s}/\kappa)/\kappa)(\tilde{\rho}_2(y/\kappa)/\kappa^l)$, $\tilde{\rho}_1(\tilde{s})$ is a mollifier on $\mathbb{R}$ such that $\tilde{\rho}_1(\tilde{s}) = 0$ if $|\tilde{s}| > 1$, $\tilde{\rho}_2(y)$ is a mollifier on $\mathbb{R}^l$ such that $\tilde{\rho}_2(y) = 0$ if $|y| > 1$, and $h_s(u,x) = 0$ if $s \le 0$. The function $h^i_s(u,x)$ is an element of $\mathbb{C}^\infty(\mathbb{R}_+ \times \mathbb{R}^l)$ in (s,x) for all u and $Dh^i_s(u,x) = \int_{\mathbb{R}\times\mathbb{R}^l} \rho_{1/i}(\tilde{s},y) Dh_{s-\tilde{s}}(u,x-y)\, d\tilde{s}\, dy$; cf. Theorem 2.29 on page 36 in Adams and Fournier [1]. In addition, $Dh^i_s(u,x)$ is a continuous function for every i. We also have that, for all open balls S, all $L > 0$ and all $q > 1$,

$$\lim_{i\to\infty}\int_0^N \sup_{u\in\mathbb{R}^n:|u|\le L}\int_S \big|Dh_s(u,x) - Dh^i_s(u,x)\big|^q\, dx\, ds = 0, \tag{7.8}$$

which can be shown as follows. If, in addition, $Dh_s(u,x)$ is continuous in all variables, then $Dh^i_s(u,x)$ converges to $Dh_s(u,x)$ locally uniformly in (s,u,x) (cf. Theorem 2.29 on page 36 in Adams and Fournier [1]). So, (7.8) holds.

In the general case, in analogy with the above reasoning, there exist r_j such that $\int_0^{N+1} \sup_{u \in \mathbb{R}^n : |u| \le L} \int_{\tilde{S}} |Dh_s(u,x)|^q \mathbf{1}_{\{|Dh_s(u,x)| \ge r_j\}}(s,u,x)\, dx\, ds < 1/j$ where $\tilde{S}$ represents the open ball in $\mathbb{R}^l$ centered at the origin of radius one greater than that of S, and there exists a continuous function $\breve{h}_s^j(u,x)$, which is bounded above in absolute value by r_j, such that $\int_0^{N+1} \mathbf{1}_{\{Dh_s(\cdot,\cdot) \ne \breve{h}_s^j(\cdot,\cdot)\}}(s)\, ds < 2/(j r_j^q)$. Calculations show that

$$\int_0^N \sup_{u \in \mathbb{R}^n : |u| \le L} \int_S \left|Dh_s(u,x) - \breve{h}_s^j(u,x)\right|^q dx\, ds \le \frac{2^{q-1}}{j} + \frac{2^{q+1}}{j}$$

and

$$\int_0^N \sup_{u \in \mathbb{R}^n : |u| \le L} \int_S \left| Dh_s^i(u,x) - \int_{\mathbb{R} \times \mathbb{R}^l} \rho_{1/i}(\tilde{s},y) \breve{h}^j_{s-\tilde{s}}(u, x-y)\, d\tilde{s}\, dy \right|^q dx\, ds$$
$$\le \frac{2^{q-1}}{j} + \frac{2^{q+1} V(S)}{j},$$

where $V(S)$ represents the volume of the ball S. Hence, (7.8) holds.

By an application of Hölder's inequality, it follows from (7.8), (6.6) in Lemma 6.4 and $h_s(u,x)$ having compact support in x locally uniformly over (s,u) that (7.4b) holds. Also, (7.3) holds.

Let $\tau^{N,i}$ and $\theta^{N,i}$ be defined as in Lemma 7.1. The functions $h_s^i(u,x)$, $\lambda^i(s,X)$, and $\tau^{N,i}(X,\mu)$ satisfy the requirements imposed on the respective functions $f(s,u,x)$, $\lambda(s,X)$ and $\tau(X,\mu)$ when deriving (5.4). Furthermore, integration by parts on the right-hand side of (5.3) with $\mu(ds,dx) = m_s(x)\, dx\, ds$, implies that $\theta^{N,i}(\gamma) = U^{\lambda^i(\cdot),h^i}_{N \wedge \tau^{N,i}(\gamma)}(\gamma)$ provided $\gamma \in \Gamma$. In addition, by (7.1a), $|X_{\tau^{N,i}(\gamma)}| \le N$ and $\tau^{N,i}(\gamma)$ is a continuous function of $\gamma \in \mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$; cf. Theorem 2 on page 510 and Theorem 3 on page 511 in Liptser and Shiryayev [29]. We obtain by equation (5.4) of Theorem 5.1 and the fact that $\tilde{\mathbf{I}}(\gamma) = \infty$ unless $\gamma \in \Gamma$ (see Theorem 6.1) that

$$\sup_{\gamma \in \Gamma} \left(\theta^{N,i}(\gamma) - \tilde{\mathbf{I}}(\gamma)\right) = 0. \tag{7.9}$$

Let us show that, for all $\delta > 2N+1$,

$$\sup_{\gamma \in K_\delta} \left(\theta^{N,i}(\gamma) - \tilde{\mathbf{I}}(\gamma)\right) = 0. \tag{7.10}$$

Let, for $\gamma \in \Gamma$,

$$\tilde{\theta}^{N,i}(\gamma) = \int_0^{\tau^{N,i}(\gamma)} \biggl(2\lambda^{i,j}(s,X)^T \biggl(\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s,x) m_s(x)\, dx\biggr)$$
$$- \frac{1}{2} \int_{\mathbb{R}^l} \|2\lambda^{i,j}(s,X)\|^2_{C_s(X_s,x)} m_s(x)\, dx$$

$$+\int_{\mathbb{R}^l}\biggl(2Dh_s^i(X_s,x)^T\biggl(\frac{1}{2}\operatorname{div}\bigl(c_s(X_s,x)m_s(x)\bigr)-a_s(X_s,x)m_s(x)\biggr)$$
$$-\frac{1}{2}\|2Dh_s^i(X_s,x)\|^2_{c_s(X_s,x)}m_s(x)\biggr)\,dx$$
$$-\int_{\mathbb{R}^l}4\lambda^{i,j}(s,X)^TG_s(X_s,x)Dh_s^i(X_s,x)m_s(x)\,dx\biggr)\,ds.$$

By (5.3), $\tilde{\theta}^{N,i}(\gamma)=U^{2\lambda^i(\cdot),2h^i}_{N\wedge\tau^{N,i}(\gamma)}(\gamma)$, provided $\gamma\in\Gamma$, so in analogy with (7.9),

$$\sup_{\gamma\in\Gamma}\bigl(\tilde{\theta}^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)=0.$$

On noting that $\tilde{\theta}^{N,i}(\gamma)\ge 2\theta^{N,i}(\gamma)-2N$, we have that, for $M>0$,

$$\begin{aligned}\sup_{\gamma:\theta^{N,i}(\gamma)\ge M}\bigl(\theta^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)&\le\sup_{\gamma:\theta^{N,i}(\gamma)\ge M}\bigl(2\theta^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)-M\\&\le\sup_{\gamma:\theta^{N,i}(\gamma)\ge M}\bigl(\tilde{\theta}^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)+2N-M\\&\le 2N-M.\end{aligned}$$

Since, by (7.9),

$$\begin{aligned}0&=\sup_{\gamma\in\Gamma}\bigl(\theta^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)\\&\le\sup_{\gamma\in K_\delta}\bigl(\theta^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)\vee\sup_{\gamma:\theta^{N,i}(\gamma)\ge M}\bigl(\theta^{N,i}(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)\vee(M-\delta),\end{aligned}$$

we conclude, on choosing $M=2N+1$, that (7.10) holds for $\delta>2N+1$.

Since by Lemma 7.1, for arbitrary $\delta\in\mathbb{R}_+$,

$$\lim_{i\to\infty}\sup_{\gamma\in K_\delta}\bigl|\theta^N(\gamma)-\theta^{N,i}(\gamma)\bigr|=0,\tag{7.11}$$

we obtain by (7.10) that

$$\sup_{\gamma\in K_\delta}\bigl(\theta^N(\gamma)-\tilde{\mathbf{I}}(\gamma)\bigr)=0.\tag{7.12}$$

Since $\theta^{N,i}(\gamma)=U^{\lambda^i(\cdot),h^i}_{\tau^{N,i}(\gamma)}(\gamma)$, the latter function is continuous in γ, and K_δ is compact, (7.11) implies that $\theta^N(\gamma)$ is continuous on K_δ. Since $\tilde{\mathbf{I}}(\gamma)$ is a lower semicontinuous function of γ, the supremum in (7.12) is attained. On the other hand, if $\tilde{\mathbf{I}}(\gamma)<\infty$, then by (7.9) and (7.11), $\sup_{\gamma\in\Gamma}(\theta^N(\gamma)-\tilde{\mathbf{I}}(\gamma))\le 0$. □

In order to prove that $\tilde{\mathbf{I}}=\mathbf{I}^{**}$, we will use $\hat{\lambda}$ and $\hat{g}$ defined in (6.16) and (6.17), respectively, as $\lambda_s(u)$ and $Dh_s(u,x)$ in the preceding lemma. We therefore need $\Phi_{s,m_s(\cdot),u}$ and $\Psi_{s,m_s(\cdot),u}$ to be sufficiently regular. The next lemma addresses both regularity and growth-rate properties.

LEMMA 7.3. *Suppose that Conditions 2.1, 2.2, (2.4b) and (2.12d) hold. Let $m_s(x)$ represent an $\mathbb{R}_+$-valued measurable function that is a probability density in x for almost every s. Suppose $m_s(x)$ is bounded away from zero on bounded sets of (s,x), $m_s(\cdot)\in\mathbb{C}^1(\mathbb{R}^l)$, with $|Dm_s(x)|$ being locally bounded in (s,x), and $m_s(x)=M_se^{-\alpha|x|}$ for all $|x|$ great enough locally uniformly in s, where $\alpha>0$. Then there exist $\mathbb{R}$-valued measurable function $w_s(u,x)$ and $\mathbb{R}^n$-valued measurable function $v_s(u,x)$ such that $w_s(u,\cdot)\in\mathbb{W}^{2,q}_{\mathrm{loc}}(\mathbb{R}^l)$ and $v_s(u,\cdot)\in\mathbb{W}^{2,q}_{\mathrm{loc}}(\mathbb{R}^l,\mathbb{R}^n)$, where $q>1$ is otherwise arbitrary, $\Phi_{s,m_s(\cdot),u}(\cdot)=Dw_s(u,\cdot)$ and $\Psi_{s,m_s(\cdot),u}(\cdot)=Dv_s(u,\cdot)$ for almost all $s\in\mathbb{R}_+$ and all $u\in\mathbb{R}^n$, that is,*

$$\int_{\mathbb{R}^l} Dp(x)^T\Big(a_s(u,x)-\frac{1}{2}\operatorname{div} c_s(u,x)\Big)m_s(x)\,dx = \int_{\mathbb{R}^l} Dp(x)^T c_s(u,x)Dw_s(u,x)m_s(x)\,dx \tag{7.13a}$$

and

$$\int_{\mathbb{R}^l} Dp(x)^T G_s(u,x)^T m_s(x)\,dx = \int_{\mathbb{R}^l} Dp(x)^T c_s(u,x)Dv_s(u,x)m_s(x)\,dx \tag{7.13b}$$

for all $p\in\mathbb{C}_0^\infty(\mathbb{R}^l)$. Furthermore, $w_s(u,x)$, $Dw_s(u,x)$, $v_s(u,x)$, and $Dv_s(u,x)$ are continuous in (u,x) for almost all $s\in\mathbb{R}_+$, and, for all open balls $S\subset\mathbb{R}^l$, all $L>0$ and all $t>0$,

$$\sup_{s\in[0,t]}\sup_{u:|u|\le L}\big(\|w_s(u,\cdot)\|_{\mathbb{W}^{2,q}(S)}+\|v_s(u,\cdot)\|_{\mathbb{W}^{2,q}(S,\mathbb{R}^n)} + \|Dw_s(u,\cdot)\|_{\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,m_s(x)\,dx)}+\|Dv_s(u,\cdot)\|_{\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^{l\times n},m_s(x)\,dx)}\big)<\infty.$$

Also, there exists α_0 which depends on the functions $a_s(u,x)$ and $c_s(u,x)$ only such that, if $\alpha>\alpha_0$, then for all $L>0$ and all $t>0$,

$$\sup_{s\in[0,t]}\sup_{x\in\mathbb{R}^l}\sup_{u\in\mathbb{R}^n:|u|\le L} w_s(u,x)<\infty \tag{7.14a}$$

and

$$\sup_{s\in[0,t]}\sup_{x\in\mathbb{R}^l}\sup_{u\in\mathbb{R}^n:|u|\le L}\bigg(\frac{|w_s(u,x)|+|Dw_s(u,x)|}{1+|x|^2} + \frac{|v_s(u,x)|+\|Dv_s(u,x)\|}{1+|x|}\bigg)<\infty, \tag{7.14b}$$

and, for all $|x|$ great enough locally uniformly in s,

$$x^T\big(a_s(u,x)-\tfrac{1}{2}\operatorname{div} c_s(u,x)-c_s(u,x)Dw_s(u,x)\big)=0 \tag{7.15a}$$

and

$$\big(G_s(u,x)-Dv_s(u,x)^Tc_s(u,x)\big)x=0. \tag{7.15b}$$

PROOF. Since $a_s(u,\cdot)\in\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,m_s(x)\,dx)$ by the fact that $a_s(u,x)$ grows at most linearly in x and $m_s(x)$ decays exponentially, and $\operatorname{div} c_s(u,\cdot)\in\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l, m_s(x)\,dx)$ for a similar reason, $\Phi_{s,m_s(\cdot),u}$ as defined by (2.15b), is an element of $\mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^l,c_s(u,x),m_s(x)\,dx)$, being a projection in the Hilbert space $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,c_s(X_s,x),m_s(x)\,dx)$. In addition,

$$\begin{aligned}&\int_{\mathbb{R}^l}\|\Phi_{s,m_s(\cdot),u}(x)\|^2_{c_s(u,x)}m_s(x)\,dx\\ &\qquad\le\int_{\mathbb{R}^l}\left\|c_s(u,x)^{-1}\left(a_s(u,x)-\frac{1}{2}\operatorname{div} c_s(u,x)\right)\right\|^2_{c_s(u,x)}m_s(x)\,dx.\end{aligned}\tag{7.16}$$

By Conditions 2.1 and 2.2 and by $m_s(\cdot)$ decaying exponentially,

$$\sup_{s\in[0,t]}\ \sup_{u\in\mathbb{R}^n:|u|\le L}\int_{\mathbb{R}^l}\left|\Phi_{s,m_s(\cdot),u}(x)\right|^2m_s(x)\,dx<\infty.\tag{7.17}$$

We prove that $\Phi_{s,m_s(\cdot),u}(\cdot)$ is a gradient. Let $Dw_i\to\Phi_{s,m_s(\cdot),u}$ in $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l, c_s(u,x),m_s(x)\,dx)$ as $i\to\infty$, where $w_i\in\mathbb{C}_0^\infty(\mathbb{R}^l)$. Then for every $f\in\mathbb{C}_0^\infty(\mathbb{R}^l, \mathbb{R}^l)$ such that $\operatorname{div} f(x)=0$, we have that $\int_{\mathbb{R}^l}Dw_i(x)^Tf(x)\,dx=0$. Since $m_s(\cdot)$ is bounded away from zero locally and $c_s(u,x)$ is positive definite, convergence in $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,c_s(u,x),m_s(x)\,dx)$ implies convergence in $\mathbb{L}^2_{\mathrm{loc}}(\mathbb{R}^l,\mathbb{R}^l)$, so $Dw_i\to \Phi_{s,m_s(\cdot),u}$ in $\mathbb{L}^2_{\mathrm{loc}}(\mathbb{R}^l,\mathbb{R}^l)$. Therefore, $\int_{\mathbb{R}^l}\Phi_{s,m_s(\cdot),u}(x)^Tf(x)\,dx=0$. It follows that $\Phi_{s,m_s(\cdot),u}(x)=D\tilde w_s(u,x)$ in the sense of distributions, where $\tilde w_s(u,\cdot)\in \mathbb{L}^2_{\mathrm{loc}}(\mathbb{R}^l)$; see, for example, Lemma 2.2.1 on page 73 in Sohr [48]. Consequently, $\int_{\mathbb{R}^l}\chi(x)^T\Phi_{s,m_s(\cdot),u}(x)\,dx=-\int_{\mathbb{R}^l}\operatorname{div}\chi(x)\tilde w_s(u,x)\,dx$, for all $\chi\in\mathbb{C}_0^1(\mathbb{R}^l,\mathbb{R}^l)$.

By (2.15b) and Condition 2.1, for $p\in\mathbb{C}_0^\infty(\mathbb{R}^l)$,

$$\begin{aligned}&-\int_{\mathbb{R}^l}\operatorname{div}\bigl(c_s(u,x)m_s(x)Dp(x)^T\bigr)\tilde w_s(u,x)\,dx\\ &\qquad=\int_{\mathbb{R}^l}Dp(x)^T\left(a_s(u,x)-\frac{1}{2}\operatorname{div} c_s(u,x)\right)m_s(x)\,dx.\end{aligned}$$

By Theorem 6.1 in Agmon [2], $\tilde w_s(u,\cdot)\in\mathbb{W}^{1,2}_{\mathrm{loc}}(\mathbb{R}^l)$ so that $\tilde w_s(u,\cdot)$ is a weak solution to the equation

$$\operatorname{div}\bigl(c_s(u,x)D\tilde w_s(u,x)m_s(x)\bigr)=\operatorname{div}\bigl(\bigl(a_s(u,x)-\tfrac{1}{2}\operatorname{div} c_s(u,x)\bigr)m_s(x)\bigr)\tag{7.18}$$

in that

$$\begin{aligned}&\int_{\mathbb{R}^l}Dp(x)^Tc_s(u,x)D\tilde w_s(u,x)m_s(x)\,dx\\ &\qquad=\int_{\mathbb{R}^l}Dp(x)^T\left(a_s(u,x)-\frac{1}{2}\operatorname{div} c_s(u,x)\right)m_s(x)\,dx.\end{aligned}\tag{7.19}$$

We note that (7.19) uniquely specifies $D\tilde w_s(u,\cdot)$ as an element of $\mathbb{L}_0^{1,2}(\mathbb{R}^l, c_s(u,x),m_s(x)\,dx)$.

Let S and $\tilde{S}$ represent open balls in $\mathbb{R}^l$ such that $S \subset\subset \tilde{S}$, let $\zeta(x)$ represent a $\mathbb{C}_0^\infty$-function with support in $\tilde{S}$ such that $\zeta(x) = 1$ for $x \in S$, and let $\varphi(x)$ represent a $\mathbb{C}_0^\infty(\tilde{S})$ function. On letting $p(x) = \varphi(x)\zeta(x)$ in (7.19) and integrating by parts, we obtain that $\zeta(x)\tilde{w}_s(u, x)$ is a weak solution f to the Dirichlet problem:

$$\begin{aligned} &\operatorname{div}\bigl(c_s(u,x)m_s(x)Df(x)\bigr) \\ &\quad = \operatorname{div}\bigl(c_s(u,x)D\zeta(x)\tilde{w}_s(u,x)m_s(x)\bigr) \\ &\qquad + D\zeta(x)^T c_s(u,x)D\tilde{w}_s(u,x)m_s(x) \\ &\qquad + \operatorname{div}\bigl(\bigl(a_s(u,x) - \tfrac{1}{2}\operatorname{div} c_s(u,x)\bigr)\zeta(x)m_s(x)\bigr) \\ &\qquad - D\zeta(x)^T\bigl(a_s(u,x) - \tfrac{1}{2}\operatorname{div} c_s(u,x)\bigr)m_s(x) \end{aligned} \tag{7.20}$$

on $\tilde{S}$ with a zero boundary condition. By Theorem 8.3 on page 181 and Theorem 8.8 on page 183 in Gilbarg and Trudinger [21], $\zeta(x)\tilde{w}_s(u, x)$ is an element of $\mathbb{W}^{2,2}(S)$ and is a strong solution of (7.20). Therefore, $\tilde{w}_s(u, \cdot) \in \mathbb{W}^{2,2}_{\mathrm{loc}}(\mathbb{R}^l)$ and (7.18) holds a.e. in x.

Differentiation in (7.18) and division by $m_s(x)$ yield

$$\begin{aligned} &\operatorname{tr}\bigl(c_s(u,x)D^2\tilde{w}_s(u,x)\bigr) + \left(c_s(u,x)\frac{Dm_s(x)}{m_s(x)} + \operatorname{div} c_s(u,x)\right)^T D\tilde{w}_s(u,x) \\ &\quad = \operatorname{div}\left(a_s(u,x) - \frac{1}{2}\operatorname{div} c_s(u,x)\right) + \left(a_s(u,x) - \frac{1}{2}\operatorname{div} c_s(u,x)\right)\frac{Dm_s(x)}{m_s(x)}. \end{aligned} \tag{7.21}$$

On writing the left-hand side as $\mathcal{L}_{s,u}(x)\tilde{w}_s(u, x)$ and letting $f_s(u, x)$ represent the right-hand side, we have that $\mathcal{L}_{s,u}(x)\tilde{w}_s(u, x) = f_s(u, x)$. Let $Y^y_{s,u}(t)$ represent the diffusion process in t with the infinitesimal generator $\mathcal{L}_{s,u}(\cdot)$ and initial condition y, defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with expectation denoted by $\mathbf{E}$. It is a strong Markov process by Conditions 2.1 and 2.2 and the hypotheses of the lemma. One can also choose $Y^y_{s,u}(t)$ to be measurable in all variables. (A possible line of reasoning invokes continuous dependence of solutions of stochastic differential equations on parameters; see, e.g., Gikhman and Skorokhod [20], or Krylov [26], and the Scorza–Dragoni theorem.) If $|x|$ is great enough so that $m_s(x) = M_s e^{-\alpha|x|}$, then

$$\frac{Dm_s(x)}{m_s(x)} = -\alpha\frac{x}{|x|}. \tag{7.22}$$

Hence, on recalling Condition 2.1, in particular that $|\operatorname{div} c_s(u, x)|$ is bounded in x locally uniformly in (s, u), and (2.12a), we have that there exists α_0 which depends on $a_t(u, x)$ and $c_t(u, x)$ only such that if $\alpha > \alpha_0$, then $\limsup_{|x|\to\infty}(x/|x|)^T(c_s(u, x)Dm_s(x)/m_s(x) + \operatorname{div} c_s(u, x)) < 0$, so $Y^y_{s,u}(t)$ is an ergodic process; see, for example, Has'minskiĭ [22], Veretennikov [54] and Malyshkin [30]. Since, by the divergence theorem,

$$\int_{\mathbb{R}^l} \mathcal{L}_{s,u}(x)p(x)m_s(x)\,dx = \int_{\mathbb{R}^l} \operatorname{div}\bigl(c_s(u,x)Dp(x)m_s(x)\bigr)\,dx = 0$$

for all $p \in \mathbb{C}_0^\infty(\mathbb{R}^l)$, $m_s(x)\,dx$ is the unique invariant measure. Similarly,

$$\int_{\mathbb{R}^l} f_s(u,x) m_s(x)\,dx = \int_{\mathbb{R}^l} \operatorname{div}\Big(\Big(a_s(u,x) - \frac{1}{2}\operatorname{div} c_s(u,x)\Big) m_s(x)\Big)\,dx = 0,$$

the latter equality being a consequence of $m_s(x)$ decaying exponentially as $|x| \to \infty$. By (7.21), (7.22), Lipschitz continuity of $a_s(u,\cdot)$ and of $\operatorname{div} c_s(u,\cdot)$, the boundedness property of $\operatorname{div} c_s(u,\cdot)$, and by (2.4b), we may assume that α_0 is such that if $\alpha > \alpha_0$, then $f_s(u,x) > 0$ for all $|x|$ great enough locally uniformly in (s,u). Also,

$$\sup_{s\in[0,t]} \sup_{x\in\mathbb{R}^l} \sup_{u\in\mathbb{R}^n:|u|\le L} \frac{|f_s(u,x)|}{1+|x|} < \infty.$$

By Theorem 1 in Pardoux and Veretennikov [35], the function

$$\breve{w}_s(u,x) = -\int_0^\infty \mathbf{E} f_s\big(u, Y^x_{s,u}(t)\big)\,dt \tag{7.23}$$

is well defined, belongs to $\mathbb{W}^{2,q}_{\text{loc}}(\mathbb{R}^l)$, for all $q > 1$, as a function of x, $D\breve{w}_s(u,x)$ is of polynomial growth in x, in particular, $D\breve{w}_s(u,\cdot) \in \mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, m_s(x)\,dx)$, and $\mathcal{L}_{s,u}(x)\breve{w}_s(u,x) = f_s(u,x)$. Since $D\breve{w}_s(u,x)$ also satisfies (7.19), we have that $D\breve{w}_s(u,x) = D\tilde{w}_s(u,x)$. In addition, $\breve{w}_s(u,x)$ is measurable in (s,u,x).

As in Pardoux and Veretennikov [35], by (7.23) and the strong Markov property, for $R > 0$,

$$\breve{w}_s(u,x) = \mathbf{E}\breve{w}_s\big(u, Y^x_{s,u}(\tau^R)\big) - \mathbf{E}\int_0^{\tau^R} f_s\big(u, Y^x_{s,u}(t)\big)\,dt, \tag{7.24}$$

where $\tau^R = \inf\{t \in \mathbb{R}_+ : |Y^x_{s,u}(t)| \le R\} < \infty$. Since $|Y^x_{s,u}(\tau^R)| = R$ if $|x| > R$, by $f_s(u,x)$ being positive for all $|x|$ great enough, we have that if R is great enough then $\breve{w}_s(u,x) \le \breve{w}_s(u,R)$, provided $|x| > R$. One can see that the bounds in the calculation of part (a) of the proof of Theorem 1 in Pardoux and Veretennikov [35] hold uniformly over $u \in [0,L]$ and $s \in [0,t]$, which shows that $\sup_{x:|x|\le R} \sup_{s\in[0,t],\, u\in\mathbb{R}^n:|u|\le L} |\breve{w}_s(u,x)| < \infty$. The bound (7.14a) follows. Since the right-hand side of (7.21) grows at most linearly in $|x|$ locally uniformly in (s,u), the arguments of part (b) of the proof of Theorem 2 (with $\beta = 2$ and $\alpha = 0$) and of part (e) of the proof of Theorem 1 in Pardoux and Veretennikov [35], along with (7.24), show that the functions $|\breve{w}_s(u,x)|$ and $|D\breve{w}_s(u,x)|$ grow at most quadratically in $|x|$ locally uniformly in (s,u).

We define $w_s(u,x) = \breve{w}_s(u,x) - V_1^{-1}\int_{S_1} \breve{w}_s(u,y)\,dy$, where S_1 represents the unit open ball centered at the origin in $\mathbb{R}^l$ and V_1 represents the volume of that ball. Obviously, the bounds on $\breve{w}_s(u,x)$ we have found are also valid for $w_s(u,x)$. It also satisfies (7.13a). We prove that, for all $q > 1$,

$$\sup_{s\in[0,t]} \sup_{u\in\mathbb{R}^n:|u|\le L} \big\| w_s(u,\cdot)\big\|_{\mathbb{W}^{2,q}(S)} < \infty. \tag{7.25}$$

Since $Dw_s(u,x) = \Phi_{s,m_s(\cdot),u}$ and (7.17) holds, $\int_{S_1} w_s(u,x)\,dx = 0$, and $m_s(x)$ is locally bounded away from zero, an application of Poincaré's inequality yields $\sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le L}\|w_s(u,\cdot)\|_{\mathbb{L}^2(S_1)} < \infty$. If S_2 is a ball containing S_1, then, for some $L_{S_1,S_2} > 0$,

$$\|w_s(u,\cdot)\|^2_{\mathbb{W}^{1,2}(S_2)} \le L_{S_1,S_2}\big(\|Dw_s(u,\cdot)\|^2_{\mathbb{L}^2(S_2)} + \|w_s(u,\cdot)\|^2_{\mathbb{L}^2(S_1)}\big);$$

see page 299 in Kufner, John and Fučík [27], and also Theorem 7.4 on page 109 in Nečas [33]. Thus, on recalling that $S \subset\subset \tilde{S}$ and letting $\breve{S}$ represent an open ball in $\mathbb{R}^l$ such that $\tilde{S} \subset\subset \breve{S}$, we have that

$$\sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le L}\|w_s(u,\cdot)\|_{\mathbb{W}^{1,2}(\breve{S})} < \infty. \tag{7.26}$$

By (7.13a), Theorem 5.5.5$'$(a) on page 156 in Morrey [32], the discussion on page 12 of Bogachev, Krylov and Rëkner [7], Shaposhnikov [47] and the fact that $\sup_{x\in\breve{S}}|a_s(u,x)|$ and $\|c_s(u,\cdot)\|_{\mathbb{W}^{1,\infty}(\breve{S},\mathbb{R}^{l\times l})}$ are bounded locally uniformly in (s,u), we have that $\|w_s(u,\cdot)\|_{\mathbb{W}^{1,q}(\tilde{S})} \le M_{\tilde{S},\breve{S},q}(1 + \|w_s(u,\cdot)\|_{\mathbb{L}^1(\breve{S})})$ locally uniformly in (s,u). By (7.26), $\sup_{s\in[0,t]}\sup_{|u|\le L}\|w_s(u,\cdot)\|_{\mathbb{W}^{1,q}(\tilde{S})} < \infty$. By (7.13a), via a similar argument to the one used for $\zeta(\cdot)\tilde{w}_s(u,\cdot)$ above, $\zeta(\cdot)w_s(u,\cdot)$ is a strong solution to (7.20). By Theorem 9.15 on page 241 in Gilbarg and Trudinger [21], $\zeta(\cdot)w_s(u,\cdot) \in \mathbb{W}^{2,q}(\tilde{S})$. By Theorem 9.11 on page 235 in Gilbarg and Trudinger [21], locally uniformly in (s,u), for some $\tilde{M}_{S,\tilde{S},q} > 0$,

$$\|w_s(u,\cdot)\|_{\mathbb{W}^{2,q}(S)} \le \tilde{M}_{S,\tilde{S},q}\big(1 + \|w_s(u,\cdot)\|_{\mathbb{L}^q(\tilde{S})}\big)$$

which implies (7.25).

We now address the continuity of $w_s(u,x)$. Let $u_i \to u$. By (7.25) and Sobolev's imbedding, the sequences $w_s(u_i,\cdot)$ and $Dw_s(u_i,\cdot)$ are equicontinuous in $x \in S$, so they are relatively compact in $\mathbb{C}(S,\mathbb{R}^l)$. A similar property holds for $(a_s(u_i,\cdot) - (1/2)\operatorname{div} c_s(u_i,\cdot))m_s(\cdot)$. Taking a subsequential limit in (7.13a) implies that $Dw_s(u_i,\cdot) \to Dw_s(u,\cdot)$ in $\mathbb{C}(S,\mathbb{R}^l)$. By Poincaré's inequality for $S = S_1$ and the fact that $\int_{S_1} w_s(u,x)\,dx = 0$, $w_s(u_i,\cdot) \to w_s(u,\cdot)$ in $\mathbb{L}^2(S_1)$. The bound

$$\begin{aligned}&\|w_s(u_i,\cdot) - w_s(u,\cdot)\|^2_{\mathbb{W}^{1,2}(S_2)}\\ &\qquad \le L_{S_1,S_2}\big(\|Dw_s(u_i,\cdot) - Dw_s(u,\cdot)\|^2_{\mathbb{L}^2(S_2)} + \|w_s(u_i,\cdot) - w_s(u,\cdot)\|^2_{\mathbb{L}^2(S_1)}\big)\end{aligned}$$

shows that $w_s(u_i,\cdot) \to w_s(u,\cdot)$ in $\mathbb{L}^2(S_2)$. Since S_2 is an arbitrary ball that contains S_1, $w_s(u_i,\cdot) \to w_s(u,\cdot)$ in $\mathbb{C}(S,\mathbb{R}^l)$. Hence, $w_s(u,x)$ and $Dw_s(u,x)$ are continuous in (u,x) for almost all s.

We outline a proof of (7.15a). Since $a_s(u,\cdot) \in \mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,m_s(x)\,dx)$ and $Dw_s(u,\cdot) \in \mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,m_s(x)\,dx)$, the equality in (7.13a) extends to $\mathbb{C}^1(\mathbb{R}^l)$-functions $p(x)$ such that $\int_{\mathbb{R}^l}(p(x)^2 + |Dp(x)|^2)m_s(x)\,dx < \infty$. We choose $p(x) = |x|^2\exp(-\delta[(|x-x_0|^2/\kappa - 1)^+]^2)$ in (7.13a), where $\kappa > 0$, $\delta > 0$, and $x_0 \in \mathbb{R}^l$, let

$\delta \to \infty$, divide the limits by the volume of the ball of radius κ centered at x_0, let $\kappa \to 0$ and accounting for $Dw_s(u,x)$, $m_s(x)$, $c_s(u,x)$ and $a_s(u,x)$ being continuous in x, obtain (7.15a).

The part that concerns $v_s(u,x)$ is dealt with similarly, except that one uses Theorem 2 of Pardoux and Veretennikov [35] with $\beta = 1$ in order to bound the growth rate of the second term of the sum in (7.14b). □

We now take on the proof of Theorem 7.1. Let $\hat{w}_s(u,x)$ and $\hat{v}_s(u,x)$ represent $w_s(u,x)$ and $v_s(u,x)$, respectively, in the statement of Lemma 7.3 for $m_s(x) = \hat{m}_s(x)$. We define, guided by (6.16) and (6.17), on recalling (2.15a) and (2.3),

$$\begin{aligned}\hat{\lambda}_s(u) = \biggl(\int_{\mathbb{R}^l} Q_{s,\hat{m}_s(\cdot)}(u,x)\hat{m}_s(x)\,dx\biggr)^{-1}\biggl(\dot{\hat{X}}_s - \int_{\mathbb{R}^l} A_s(u,x)\hat{m}_s(x)\,dx \\ - \int_{\mathbb{R}^l} G_s(u,x)\biggl(\frac{D\hat{m}_s(x)}{2\hat{m}_s(x)} - D\hat{w}_s(u,x)\biggr)\hat{m}_s(x)\,dx\biggr)\end{aligned} \tag{7.27}$$

if $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u) and $\hat{\lambda}_s(u) = 0$ if $C_t(u,x) = 0$ for all (t,u,x), and

$$\hat{h}_s(u,x) = \tfrac{1}{2}\ln \hat{m}_s(x) - \hat{w}_s(u,x) - \hat{v}_s(u,x)^T\hat{\lambda}_s(u) \tag{7.28}$$

so that

$$D\hat{h}_s(u,x) = \frac{D\hat{m}_s(x)}{2\hat{m}_s(x)} - D\hat{w}_s(u,x) - D\hat{v}_s(u,x)^T\hat{\lambda}_s(u). \tag{7.29}$$

We note that by (2.14),

$$Q_{s,\hat{m}_s(\cdot)}(u,x) = C_s(u,x) - \|D\hat{v}_s(u,x)\|^2_{c_s(u,x)}. \tag{7.30}$$

The continuity properties of $w_s(u,x)$ and $v_s(u,x)$ established in Lemma 7.3 imply that $\hat{\lambda}_s(u)$ is continuous in u and that $\hat{h}_s(u,x)$ and $D\hat{h}_s(u,x)$ are continuous in (u,x), for almost all $s \in \mathbb{R}_+$.

If $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u), then the analogue of (7.16) for $D\hat{w}_s(u,x)$, (7.27) and Condition 2.1 imply that, for some $\vartheta_1 > 0$,

$$\begin{aligned}|\hat{\lambda}_s(u)| \le \vartheta_1\biggl(|\dot{\hat{X}}_s| + \sup_{x\in\mathbb{R}^l}|A_s(u,x)| \\ + \biggl(\int_{\mathbb{R}^l}|a_s(u,x)|^2\hat{m}_s(x)\,dx\biggr)^{1/2} + \sup_{x\in\mathbb{R}^l}\frac{|D\hat{m}_s(x)|}{\hat{m}_s(x)}\biggr).\end{aligned} \tag{7.31}$$

Since $\hat{m}_s(x) = M_s e^{-\alpha|x|}$ for $|x| > 2r$ and $|a_s(u,x)|$ grows at most linearly in x, we conclude that $|\hat{\lambda}_s(u)|$ is locally bounded in (s,u).

Therefore, by Lemma 7.3, for all $L > 0$, all open balls S in $\mathbb{R}^l$, and all $q > 1$,

$$\sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le L}|\hat{\lambda}_s(u)| + \sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le L}\int_S |D\hat{h}_s(u,x)|^q\,dx < \infty. \tag{7.32}$$

By Theorem 6.1, the supremum in (6.15) is attained at $\lambda = \hat{\lambda}_s(u)$ and $g = D\hat{h}_s(u,x)$, however, the function $\hat{h}_s(u,x)$ might not be of compact support in x so in order to use it in Lemma 7.2, we need to restrict it to a compact set. Let $\eta(y)$ represent an $\mathbb{R}_+$-valued nonincreasing $\mathbb{C}_0^\infty(\mathbb{R}_+)$-function such that $\eta(y) = 1$ for $0 \le y \le 1$ and $\eta(y) = 0$ for $y \ge 2$. Let $\hat{w}_s^i(u,x) = \hat{w}_s(u,x)\eta(|x|/i)$ and $\hat{v}_s^i(u,x) = \hat{v}_s(u,x)\eta(|x|/i)$. We note that

$$D\hat{w}_s^i(u,x) = \eta\left(\frac{|x|}{i}\right)D\hat{w}_s(u,x) + \frac{x}{i|x|}D\eta\left(\frac{|x|}{i}\right)\hat{w}_s(u,x) \tag{7.33a}$$

and

$$D\hat{v}_s^i(u,x) = D\hat{v}_s(u,x)\eta\left(\frac{|x|}{i}\right) + \frac{x}{i|x|}\hat{v}_s(u,x)D\eta\left(\frac{|x|}{i}\right)^T. \tag{7.33b}$$

We define, in analogy with (7.27),

$$\begin{aligned}\hat{\lambda}_s^i(u) = {}&\left(\int_{\mathbb{R}^l} Q_{s,\hat{m}_s(\cdot)}(u,x)\hat{m}_s(x)\,dx\right)^{-1}\left(\dot{\hat{X}}_s - \int_{\mathbb{R}^l} A_s(u,x)\hat{m}_s(x)\,dx\right.\\ &\left.- \int_{\mathbb{R}^l} G_s(u,x)\left(\frac{1}{2}D\left(\eta\left(\frac{|x|}{i}\right)\ln\hat{m}_s(x)\right) - D\hat{w}_s^i(u,x)\right)\hat{m}_s(x)\,dx\right)\end{aligned} \tag{7.34}$$

if $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u), and $\hat{\lambda}_s^i(u) = 0$ if $C_t(u,x) = 0$. We let, similar to (7.28),

$$\hat{h}_s^i(u,x) = \frac{1}{2}\eta\left(\frac{|x|}{i}\right)\ln\hat{m}_s(x) - \hat{w}_s^i(u,x) - \hat{v}_s^i(u,x)^T\hat{\lambda}_s^i(u). \tag{7.35}$$

In analogy with (7.31) and in view of (7.33a) and (7.14b) in Lemma 7.3, one can see that the $|\hat{\lambda}_s^i(u)|$ are bounded uniformly in i and locally uniformly in (s,u), where the bound may depend on α. Also, $\hat{\lambda}_s^i(u)$ is continuous in u, so it satisfies the hypotheses of Lemma 7.2.

If $C_t(u,x) - G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u), then by (7.33a), Lemma 7.3, (7.27) and (7.34),

$$\lim_{i\to\infty}\sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le L}|\hat{\lambda}_s^i(u) - \hat{\lambda}_s(u)| = 0. \tag{7.36}$$

The latter convergence also holds if $C_t(u,x) = 0$ in that $\hat{\lambda}_s^i(u) = \hat{\lambda}_s(u) = 0$.

Similarly, since by (7.33a), (7.33b) and (7.35),

$$\begin{aligned}D\hat{h}_s^i(u,x) = {}&\eta\left(\frac{|x|}{i}\right)D\hat{h}_s(u,x)\\ &+ \frac{1}{i}\frac{x}{|x|}D\eta\left(\frac{|x|}{i}\right)\left(\frac{1}{2}\ln\hat{m}_s(x) - \hat{w}_s(u,x) - \hat{v}_s(u,x)^T\hat{\lambda}_s^i(u)\right),\end{aligned} \tag{7.37}$$

we have that

$$\lim_{i\to\infty}\sup_{s\in[0,t]}\sup_{u\in\mathbb{R}^n:|u|\le L}\int_S \big|D\hat{h}^i_s(u,x)-D\hat{h}_s(u,x)\big|^q\,dx=0, \tag{7.38}$$

for all $L>0$, all open balls S in $\mathbb{R}^l$, and all $q>1$. The functions $\hat{h}^i_s(u,x)$ also satisfy the hypotheses of Lemma 7.2.

Another auxiliary lemma is in order.

LEMMA 7.4. *Suppose, for* $i\in\mathbb{N}$, $X^i\in\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)$, $X\in\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)$, *and* $m^i_s(x)$ *and* $m_s(x)$ *are measurable functions which are probability densities in* x *on* $\mathbb{R}^l$ *for almost all* s *such that*

$$\int_{\mathbb{R}^l}\biggl(\frac{1}{2}\operatorname{tr}\bigl(c_s(X^i_s,x)D^2p(x)\bigr)+Dp(x)^T\biggl(a_s(X^i_s,x)+G_s(X^i_s,x)^T\hat{\lambda}^i_s(X^i_s) - \frac{1}{2}\operatorname{div}c_s(X^i_s,x)+c_s(X^i_s,x)D\hat{h}^i_s(X^i_s,x)\biggr)\biggr)m^i_s(x)\,dx=0$$

and

$$\int_{\mathbb{R}^l}\biggl(\frac{1}{2}\operatorname{tr}\bigl(c_s(X_s,x)D^2p(x)\bigr)+Dp(x)^T\biggl(a_s(X_s,x)+G_s(X_s,x)^T\hat{\lambda}_s(X_s) - \frac{1}{2}\operatorname{div}c_s(X_s,x)+c_s(X_s,x)D\hat{h}_s(X_s,x)\biggr)\biggr)m_s(x)\,dx=0, \tag{7.39}$$

for all $p\in\mathbb{C}_0^\infty(\mathbb{R}^l)$. *If* $X^i\to X$ *as* $i\to\infty$, *then, for all* α *great enough and for all* $t>0$,

$$\lim_{i\to\infty}\int_0^t\int_{\mathbb{R}^l}\big|m^i_s(x)-m_s(x)\big|\,dx\,ds=0, \tag{7.40a}$$

$$\lim_{i\to\infty}\int_0^t\int_{\mathbb{R}^l}\|D\hat{h}^i_s(X^i_s,x)\|^2_{c_s(X^i_s,x)}m^i_s(x)\,dx\,ds = \int_0^t\int_{\mathbb{R}^l}\|D\hat{h}_s(X_s,x)\|^2_{c_s(X_s,x)}m_s(x)\,dx\,ds \tag{7.40b}$$

and

$$\lim_{i\to\infty}\int_0^t\|\hat{\lambda}^i_s(X^i_s)\|^2_{\int_{\mathbb{R}^l}C_s(X^i_s,x)m^i_s(x)\,dx}\,ds = \int_0^t\|\hat{\lambda}_s(X_s)\|^2_{\int_{\mathbb{R}^l}C_s(X_s,x)m_s(x)\,dx}\,ds. \tag{7.40c}$$

PROOF. We must again resort to a proof outline. We first address existence and uniqueness of $m^i_s(x)$ and $m_s(x)$. Since $\sup_{|u|\le L}(x/|x|)^Ta_s(u,x)\to-\infty$ as $|x|\to\infty$, the function $G_s(u,x)$ is bounded, the function $\hat{\lambda}^i_s(u)$ is bounded locally in (s,u) and the function $\hat{h}^i_s(u,x)$ is of compact support in x locally uniformly in (s,u), we have that

$$\lim_{|x|\to\infty}\frac{x^T}{|x|}\Big(a_s(X^i_s,x)+G_s(X^i_s,x)^T\hat{\lambda}^i_s(X^i_s)-\frac{1}{2}\operatorname{div}c_s(X^i_s,x) + c_s(X^i_s,x)D\hat{h}^i_s(X^i_s,x)\Big)=-\infty, \tag{7.41}$$

which implies that $m^i_s(x)$ is well defined and is specified uniquely; see, for example, Metafune, Pallara and Rhandi [31], Theorem 2.2, Proposition 2.4.

By (7.29), relations (7.15a) and (7.15b) of Lemma 7.3 imply that

$$x^T\Big(a_s(X_s,x)+G_s(X_s,x)^T\hat{\lambda}_s(X_s)-\frac{1}{2}\operatorname{div}c_s(X_s,x)+c_s(X_s,x)D\hat{h}_s(X_s,x)\Big) = \frac{x^T}{2}c_s(X_s,x)\frac{D\hat{m}_s(x)}{\hat{m}_s(x)}.$$

If $|x|>2r$, then $D\hat{m}_s(x)/\hat{m}_s(x)=-\alpha x/|x|$, so, locally uniformly in s,

$$\limsup_{|x|\to\infty}\frac{x^T}{|x|}\Big(a_s(X_s,x)+G_s(X_s,x)^T\hat{\lambda}_s(X_s)-\frac{1}{2}\operatorname{div}c_s(X_s,x) + c_s(X_s,x)D\hat{h}_s(X_s,x)\Big)<0,$$

which ensures the existence and uniqueness of $m_s(x)$.

As in the proof of Lemma 6.7, it is then shown that, for arbitrary $\delta>0$, there exists $\alpha>0$ such that for all $t>0$

$$\sup_{s\in[0,t]}\sup_{i\in\mathbb{N}}\int_{\mathbb{R}^l}e^{\delta|x|}m^i_s(x)\,dx<\infty. \tag{7.42}$$

First, a uniform version of (7.41) is established:

$$\lim_{\alpha\to\infty}\limsup_{|x|\to\infty}\sup_{s\in[0,t]}\sup_{i\in\mathbb{N}}\frac{x^T}{|x|}\Big(a_s(X^i_s,x)+G_s(X^i_s,x)^T\hat{\lambda}^i_s(X^i_s)-\frac{1}{2}\operatorname{div}c_s(X^i_s,x) + c_s(X^i_s,x)D\hat{h}^i_s(X^i_s,x)\Big)=-\infty. \tag{7.43}$$

The proof of the bound in (7.42) is similar to the argument in the proof of Lemma 6.7, with (7.43) assuming the role of the condition that

$$\sup_{i\in\mathbb{N}}a_{s+t}(X^i_s,x)^T\frac{x}{|x|}\to-\infty.$$

Since the $\lambda^i_s(u)$ are bounded uniformly in i and locally uniformly in (s,u) and (7.32) and (7.42) hold, by Proposition 2.16 in Bogachev, Krylov and Röckner [8], for almost all s the functions $m_s^i(\cdot)$ converge in the variation norm along a subsequence to probability density $\tilde{m}_s(\cdot)$. By (7.29), (7.37), the bounds (7.14b), and by (7.42), we have that

$$\sup_{s\in[0,t]}\sup_{i\in\mathbb{N}}\int_{\mathbb{R}^l}\|D\hat{h}_s^i(X_s^i,x)\|^3_{c_s(X_s^i,x)}m_s^i(x)\,dx<\infty. \tag{7.44}$$

Since $\sup_{i\in\mathbb{N}}|\hat{\lambda}_s^i|<\infty$, the convergences in (7.36) and (7.38) imply that $\tilde{m}_s(x)$ must satisfy (7.39), so $\tilde{m}_s(x)=m_s(x)$ and $m_s^i(\cdot)\to m_s(\cdot)$ in the variation norm. The limit in (7.40a) follows by dominated convergence. The convergence in (7.40c) follows from (7.36), (7.32) and (2.12d). For (7.40b), we also take into account (7.44). $\square$

We complete the proof of Theorem 7.1. Let, given $N\in\mathbb{N}$, $\hat{\tau}^{N,i}$ and $\hat{\theta}^{N,i}$ be defined by the respective equations (7.1a) and (7.1b) with $\hat{\lambda}_s^i(u)$ and $\hat{h}_s^i(u,x)$ as $\hat{\lambda}_s(u)$ and $\hat{h}_s(u,x)$, respectively. Since the functions $\hat{\lambda}_s^i(u)$ and $\hat{h}_s^i(u,x)$ satisfy the hypotheses of Lemma 7.2, there exist $\gamma^{N,i}=(X^{N,i},\mu^{N,i})\in\Gamma$ such that $\hat{\theta}^{N,i}(\gamma^{N,i})=\tilde{\mathbf{I}}(\gamma^{N,i})$ and $\gamma^{N,i}\in K_{2N+2}$ for all i. In particular, $X_0^{N,i}=\hat{u}$, $\mu^{N,i}(ds,dx)=m_s^{N,i}(x)\,dx\,ds$, where $m_s^{N,i}(\cdot)\in\mathbb{P}(\mathbb{R}^l)$ (see Theorem 6.1), and the set $\{\gamma^{N,i},i=1,2,\dots\}$ is relatively compact. Since $\tilde{\mathbf{I}}(\gamma^{N,i})\ge\mathbf{I}^{**}(\gamma^{N,i})$, on the one hand, and $\hat{\theta}^{N,i}(\gamma^{N,i})\le\mathbf{I}^{**}(\gamma^{N,i})$ by (6.14) and (7.1b), on the other hand, we have that

$$\hat{\theta}^{N,i}(\gamma^{N,i})=\mathbf{I}^{**}(\gamma^{N,i})=\tilde{\mathbf{I}}(\gamma^{N,i}). \tag{7.45}$$

Let $\mu^{N,i}\to\mu^N$ in $\mathbb{C}(\mathbb{R}_+,\mathbb{M}(\mathbb{R}^l))$ and $X^{N,i}\to X^N$ in $\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)$ along a subsequence of i, which we still denote by i.

By (7.1b) and (7.45), the suprema in (6.14) for $(X,\mu)=(X^{N,i},\mu^{N,i})$ are attained at $(\hat{\lambda}_s^i(X_s^{N,i}),\hat{h}_s^i(X_s^{N,i},x))$ when $s\le\hat{\tau}^{N,i}(\gamma^{N,i})$. In particular, since the supremum over h for $\lambda=\hat{\lambda}_s^i(X_s^{N,i})$ is attained at $h(x)=\hat{h}_s^i(X_s^{N,i},x)$, we have that

$$\begin{aligned}&D\hat{h}_s^i(X_s^{N,i},x)\\&\quad=\Pi_{c_s(X_s^{N,i},\cdot),m_s^{N,i}(\cdot)}\biggl(\frac{Dm_s^{N,i}(x)}{2m_s^{N,i}(x)}+c_s(X_s^{N,i},\cdot)^{-1}\\&\qquad\times\biggl(\frac{1}{2}\operatorname{div}c_s(X_s^{N,i},\cdot)-a_s(X_s^{N,i},\cdot)-G_s(X_s^{N,i},\cdot)^T\hat{\lambda}_s^i(X_s^{N,i})\biggr)\biggr)(x).\end{aligned} \tag{7.46}$$

Recalling the definition of Π and integrating by parts obtains, for $p\in\mathbb{C}_0^\infty(\mathbb{R}^l)$,

$$\begin{aligned}&\int_{\mathbb{R}^l}\biggl(\frac{1}{2}\operatorname{tr}(c_s(X_s^{N,i},x)D^2p(x))+Dp(x)^T\biggl(a_s(X_s^{N,i},x)-\frac{1}{2}\operatorname{div}c_s(X_s^{N,i},x)\\&\quad+G_s(X_s^{N,i},x)^T\hat{\lambda}_s^i(X_s^{N,i})+c_s(X_s^{N,i},x)D\hat{h}_s^i(X_s^{N,i},x)\biggr)\biggr)m_s^{N,i}(x)\,dx=0.\end{aligned} \tag{7.47}$$

Thus, $m_s^{N,i}(x)\,dx$ is an invariant probability for a diffusion. By (7.13a), (7.13b) [for $\hat{w}_s(u,x)$ and $\hat{v}_s(u,x)$], and (7.29), via a similar manipulation,

$$\int_{\mathbb{R}^l}\biggl(\frac{1}{2}\operatorname{tr}\bigl(c_s(X_s^N,x)D^2p(x)\bigr)+Dp(x)^T\biggl(a_s(X_s^N,x)-\frac{1}{2}\operatorname{div}c_s(X_s^N,x)$$

$$+G_s(X_s^N,x)^T\hat{\lambda}_s(X_s^N)+c_s(X_s^N,x)D\hat{h}_s(X_s^N,x)\biggr)\biggr)\hat{m}_s(x)\,dx=0.$$

Let $\tilde{m}_s^{N,i}(x)$ represent a probability density that solves (7.47) for all $s\in\mathbb{R}_+$ rather than for $s\le\hat{\tau}^{N,i}(\gamma^{N,i})$. The existence of $\tilde{m}_s^{N,i}(x)$ is established as in the proof of Lemma 7.4; more specifically, see (7.41). Lemma 7.4 implies that $\tilde{m}_s^{N,i}(x)\to\hat{m}_s(x)$ in $\mathbb{L}^1([0,t]\times\mathbb{R}^l)$ as $i\to\infty$, that

$$\begin{aligned}\lim_{i\to\infty}&\int_0^t\int_{\mathbb{R}^l}\|D\hat{h}_s^i(X_s^{N,i},x)\|^2_{c_s(X_s^{N,i},x)}\tilde{m}_s^{N,i}(x)\,dx\,ds\\&=\int_0^t\int_{\mathbb{R}^l}\|D\hat{h}_s(X_s^N,x)\|^2_{c_s(X_s^N,x)}\hat{m}_s(x)\,dx\,ds\end{aligned}\tag{7.48a}$$

and that

$$\begin{aligned}\lim_{i\to\infty}&\int_0^t\|\hat{\lambda}_s^i(X_s^{N,i})\|^2_{\int_{\mathbb{R}^l}C_s(X_s^{N,i},x)\tilde{m}_s^{N,i}(x)\,dx}\,ds\\&=\int_0^t\|\hat{\lambda}_s(X_s^N)\|^2_{\int_{\mathbb{R}^l}C_s(X_s^N,x)\hat{m}_s(x)\,dx}\,ds.\end{aligned}\tag{7.48b}$$

By Lemma 7.1, $\hat{\tau}^{N,i}(\tilde{\gamma}^{N,i})\to\hat{\tau}^N(\gamma^N)$ as $i\to\infty$, where $\tilde{\gamma}^{N,i}=(X^{N,i},\tilde{\mu}^{N,i})$ and $\tilde{\mu}^{N,i}(dx,ds)=\tilde{m}_s^{N,i}(x)\,dx\,ds$. Since $\hat{\tau}^{N,i}(\tilde{\gamma}^{N,i})=\hat{\tau}^{N,i}(\gamma^{N,i})$, we obtain that $\hat{\tau}^{N,i}(\gamma^{N,i})\to\tau^N(\gamma^N)$ and that $m_s^{N,i}(x)\to\hat{m}_s(x)$ in $\mathbb{L}^1([0,\tau^N(\gamma^N)]\times\mathbb{R}^l)$, so $\mu_s^N(dx)=\hat{m}_s(x)\,dx$ for almost all $s\le\tau^N(\gamma^N)$.

We now use the fact that the supremum in (6.14) over λ for $h(x)=\hat{h}_s^i(X_s^{N,i},x)$ is attained at $\lambda=\hat{\lambda}_s^i(X_s^{N,i})$. If $C_t(u,x)=0$ and $A_t(u,x)$ is locally Lipschitz continuous in u locally uniformly in t and uniformly in x, then $\hat{\lambda}_s^i(X_s^{N,i})=0$, so $\dot{X}_s^{N,i}=\int_{\mathbb{R}^l}A_s(X_s^{N,i},x)m_s^{N,i}(x)\,dx$, which, as in the proof of Lemma 6.7, implies since $X^{N,i}\to X^N$ in $\mathbb{C}(\mathbb{R}_+,\mathbb{R}^n)$ and $(m_s^{N,i}(x))\to(\hat{m}_s(x))$ in $\mathbb{L}^1([0,\tau^N(\gamma^N)]\times\mathbb{R}^l)$ as $i\to\infty$ that $\dot{X}_s^N=\int_{\mathbb{R}^l}A_s(X_s^N,x)\hat{m}_s(x)\,dx$ a.e. for $s\le\tau^N(\gamma^N)$. By uniqueness, $X_s^N=\hat{X}_s$ for $s\le\tau^N(\gamma^N)$. As a byproduct, $\dot{X}_s^{N,i}\to\dot{\hat{X}}_s$ as $i\to\infty$ a.e. on $[0,\tau^N(\gamma^N)]$.

Suppose that $C_t(u,x)-G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite locally uniformly in (t,u) and uniformly in x. Then the maximization condition is

$$\dot{X}_s^{N,i}=\int_{\mathbb{R}^l}A_s(X_s^{N,i},x)m_s^{N,i}(x)\,dx+\int_{\mathbb{R}^l}G_s(X_s^{N,i},x)D\hat{h}_s^i(X_s^{N,i},x)m_s^{N,i}(x)\,dx$$

$$+\int_{\mathbb{R}^l}C_s(X_s^{N,i},x)\hat{\lambda}_s^i(X_s^{N,i})m_s^{N,i}(x)\,dx.$$

On integrating both sides from 0 to t and letting $i\to\infty$, we have by the facts that $\gamma^{N,i}\to\gamma^N$, that $X^{N,i}\to X^N$, that $m_s^{N,i}(x)\to\hat m_s(x)$ in $\mathbb{L}^1([0,\tau^N(\gamma^N)]\times\mathbb{R}^l)$, and that $\hat\lambda_s^i(u)\to\hat\lambda_s(u)$ locally uniformly in (s,u) as $i\to\infty$ [see (7.36)], by (7.38), by (7.44) and by (7.29) that, for almost all $s\le\tau^N(\gamma^N)$,

$$\begin{aligned}\dot X_s^N&=\int_{\mathbb{R}^l}A_s(X_s^N,x)\hat m_s(x)\,dx\\&\quad+\int_{\mathbb{R}^l}G_s(X_s^N,x)\biggl(\frac{D\hat m_s(x)}{2\hat m_s(x)}-D\hat w_s(X_s^N,x)\biggr)\hat m_s(x)\,dx\\&\quad+\int_{\mathbb{R}^l}(C_s(X_s^N,x)-G_s(X_s^N,x)D\hat v_s(X_s^N,x))\hat m_s(x)\,dx\hat\lambda_s(X_s^N).\end{aligned}\tag{7.49}$$

Since $D\hat v_s(u,\cdot)\in\mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^{n\times l},c_s(x),\hat m_s(x)\,dx)$ and the function $G_s(u,\cdot)$ is bounded, (7.13b) extends to Dp representing an arbitrary element of $\mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^l,c_t(x),\hat m_s(x)\,dx)$, so by (7.30),

$$\int_{\mathbb{R}^l}Q_{s,\hat m_s(\cdot)}(u,x)\hat m_s(x)\,dx=\int_{\mathbb{R}^l}(C_s(u,x)-G_s(u,x)D\hat v_s(u,x))\hat m_s(x)\,dx.$$

Substitution of the latter expression in (7.27) and of (7.27) into (7.49) obtains that $\dot X_s^N=\dot{\hat X}_s$ a.e. on $[0,\tau^N(\gamma^N)]$, so on recalling that $X_0^N=\hat X_0=\hat u$ we conclude that $X_s^N=\hat X_s$ for $s\le\tau^N(\gamma^N)$. In addition, $\dot X_s^{N,i}\to\dot{\hat X}_s$ as $i\to\infty$ a.e. on $[0,\tau^N(\gamma^N)]$.

Hence, in either case, $\tau^N(\gamma^N)=\tau^N(\hat\gamma)$ and $\gamma_s^N=\hat\gamma_s$ for $s\le\tau^N(\hat\gamma)$ so that $\theta^N(\gamma^N)=\theta^N(\hat\gamma)$, where $\hat\gamma=(\hat X,\hat\mu)$. We show that

$$\theta^N(\gamma^N)=\lim_{i\to\infty}\hat\theta^{N,i}(\gamma^{N,i}).\tag{7.50}$$

By (7.1b) and (7.46),

$$\begin{aligned}\hat\theta^{N,i}(\gamma^{N,i})&=\int_0^{\hat\tau^{N,i}(\gamma^{N,i})}\biggl(\hat\lambda_s^i(X_s^{N,i})^T\biggl(\dot X_s^{N,i}-\int_{\mathbb{R}^l}A_s(X_s^{N,i},x)m_s^{N,i}(x)\,dx\biggr)\\&\quad-\frac12\|\hat\lambda_s^i(X_s^{N,i})\|^2_{\int_{\mathbb{R}^l}C_s(X_s^{N,i},x)m_s^{N,i}(x)\,dx}\\&\quad+\frac12\int_{\mathbb{R}^l}\|D\hat h_s^i(X_s^{N,i},x)\|^2_{c_s(X_s^{N,i},x)}m_s^{N,i}(x)\,dx\biggr)\,ds.\end{aligned}$$

Similarly,

$$\begin{aligned}\theta^N(\gamma^N)&=\int_0^{\tau^N(\gamma^N)}\biggl(\hat\lambda_s(X_s^N)^T\biggl(\dot X_s^N-\int_{\mathbb{R}^l}A_s(X_s^N,x)\hat m_s(x)\,dx\biggr)\\&\quad-\frac12\|\hat\lambda_s(X_s^N)\|^2_{\int_{\mathbb{R}^l}C_s(X_s^N,x)\hat m_s(x)\,dx}\\&\quad+\frac12\int_{\mathbb{R}^l}\|D\hat h_s(X_s^N,x)\|^2_{c_s(X_s^N,x)}\hat m_s(x)\,dx\biggr)\,ds.\end{aligned}$$

On recalling convergences (7.48a) and (7.48b) which are locally uniform in t, the fact that $\tilde{m}_s^{N,i}(x) = m_s^{N,i}(x)$ for $s \le \tau^{N,i}(\gamma^{N,i})$, and the convergences $\hat{\tau}^{N,i}(\gamma^{N,i}) \to \tau^N(\gamma^N)$, $\gamma^{N,i} \to \gamma^N$, for (7.50), it remains to check that

$$\lim_{i\to\infty} \int_0^{\hat{\tau}^N(\gamma^{N,i})} \hat{\lambda}_s^i(X_s^{N,i})^T \left(\dot{X}_s^{N,i} - \int_{\mathbb{R}^l} A_s(X_s^{N,i}, x) m_s^{N,i}(x)\,dx\right) ds = \int_0^{\tau^N(\gamma^N)} \hat{\lambda}_s(X_s^N)^T \left(\dot{X}_s^N - \int_{\mathbb{R}^l} A_s(X_s^N, x)\hat{m}_s(x)\,dx\right) ds. \tag{7.51}$$

The convergences $\hat{\tau}^{N,i}(\gamma^{N,i}) \to \tau^N(\gamma^N)$, $\gamma^{N,i} \to \gamma^N$, and $\dot{X}_s^{N,i} \to \dot{X}_s^N$ for almost all $s < \tau^N(\gamma^N)$, imply that the $\mathbf{1}_{\{s\le\hat{\tau}^{N,i}(\gamma^{N,i})\}}(s)\hat{\lambda}_s^i(X_s^{N,i})^T(\dot{X}_s^{N,i} - \int_{\mathbb{R}^l} A_s(X_s^{N,i}, x)m_s^{N,i}(x)\,dx)$ converge to $\mathbf{1}_{\{s\le\tau^N(\gamma^N)\}}(s)\hat{\lambda}_s(X_s^N)^T(\dot{X}_s^N - \int_{\mathbb{R}^l} A_s(X_s^N, x)\hat{m}_s(x)\,dx)$ as $i \to \infty$ for almost all s. Since the $\hat{\lambda}_s^i(u)$ are bounded uniformly in i and locally uniformly in (s, u), the uniform integrability needed to derive (7.51) follows by the bound $\sup_{\gamma\in K_\delta} \int_0^N |\dot{X}_s - \int_{\mathbb{R}^l} A_s(X_s, x) \times m_s(x)\,dx|^2\,ds < \infty$, which is a consequence of (6.18).

By (7.45), (7.50) and part 1 of Theorem 3.4, $\mathbf{I}^{**}(\gamma^N) = \theta^N(\gamma^N) = \tilde{\mathbf{I}}(\gamma^N)$. [Alternatively, one can follow the proof of part 1 of Theorem 3.4 by letting $i \to \infty$ in (7.45) to obtain that $\mathbf{I}^{**}(\gamma^N) \ge \theta^N(\gamma^N) \ge \tilde{\mathbf{I}}(\gamma^N)$.] Therefore, $\mathbf{I}^{**}(\hat{\gamma}) \ge \theta^N(\hat{\gamma}) = \theta^N(\gamma^N) = \tilde{\mathbf{I}}(\gamma^N)$. Let $\pi_t(\gamma)$, where $\gamma = (X, \mu)$, denote the projection $((X_{s\wedge t}, \mu_{s\wedge t}(\cdot)), s \in \mathbb{R}_+)$. We have that

$$\tilde{\mathbf{I}}(\gamma^N) \ge \inf_{\gamma:\pi_{\tau^N(\gamma^N)}(\gamma)=\pi_{\tau^N(\gamma^N)}(\gamma^N)} \tilde{\mathbf{I}}(\gamma) = \inf_{\gamma:\pi_{\tau^N(\hat{\gamma})}(\gamma)=\pi_{\tau^N(\hat{\gamma})}(\hat{\gamma})} \tilde{\mathbf{I}}(\gamma).$$

The sets $\pi^{-1}_{\tau^N(\hat{\gamma})}(\pi_{\tau^N(\hat{\gamma})}(\hat{\gamma}))$ are closed and decrease to $\hat{\gamma}$ as $N \to \infty$, so the rightmost side converges to $\tilde{\mathbf{I}}(\hat{\gamma})$, by $\tilde{\mathbf{I}}$ being lower compact. We conclude that $\mathbf{I}^{**}(\hat{\gamma}) \ge \tilde{\mathbf{I}}(\hat{\gamma})$, so $\mathbf{I}^{**}(\hat{\gamma}) = \tilde{\mathbf{I}}(\hat{\gamma})$.

8. Approximating the large deviation function.

By Theorem 3.4, in order to complete the proof of Theorem 2.1, it remains to establish an approximation theorem for $\mathbf{I}^{**}$ along the lines of part 2 of Theorem 3.4. We state it next.

THEOREM 8.1. *Suppose that Conditions 2.1–2.3, (2.4b) and (2.12d) hold. If $\mathbf{I}^{**}(X, \mu) < \infty$, then there exists sequence $(X^{(j)}, \mu^{(j)})$ whose members satisfy the requirements on $(\hat{X}, \hat{\mu})$ in the statement of Theorem 7.1 such that $(X^{(j)}, \mu^{(j)}) \to (X, \mu)$ and $\mathbf{I}^{**}(X^{(j)}, \mu^{(j)}) \to \mathbf{I}^{**}(X, \mu)$ as $j \to \infty$.*

PROOF. Let $\mu(ds, dx) = m_s(x)\,dx\,ds$ and

$$k_s(x) = \frac{1}{2m_s(x)} \operatorname{div}\bigl(c_s(X_s, x)m_s(x)\bigr) - a_s(X_s, x).$$

Since, by Theorem 6.1, $\int_0^t \int_{\mathbb{R}^l} |Dm_s(x)|^2/m_s(x)\,dx\,ds < \infty$, for all $t \in \mathbb{R}_+$, we have that $k_s(\cdot) \in \mathbb{L}^2_{\text{loc}}(\mathbb{R}^l, \mathbb{R}^l, m_s(x)\,dx)$ a.e.

Let function η be as in Condition 2.3. We introduce $\eta_r(x) = \eta(|x|/r)$ and

$$k_s^r(x) = \frac{1}{2\eta_r^2(x) m_s(x)} \operatorname{div}\bigl(c_s(X_s, x)\eta_r^2(x) m_s(x)\bigr) - a_s(X_s, x), \tag{8.1}$$

where $x \in \mathbb{R}^l$ and $r > 0$. We also let S_r represent the open ball in $\mathbb{R}^l$ of radius r centered at the origin.

We first prove that one can choose $(X^{(j)}, \mu^{(j)})$ of the required form that converge to (X, μ) as $j \to \infty$ and are such that $\mathbf{I}_t^{**}(X^{(j)}, \mu^{(j)}) \to \mathbf{I}_t^{**}(X, \mu)$ for all t, where $\mathbf{I}_t^{**}$ is defined by (6.21).

Let us begin with the case where $C_t(u, x) = 0$ for all (t, u, x) and $A_t(u, x)$ is Lipschitz continuous in u locally uniformly in t and uniformly in x. By Theorem 6.1, $\dot{X}_s = \int_{\mathbb{R}^l} A_s(X_s, x) m_s(x)\,dx$ a.e., the latter equation having a unique solution. Let $\rho_\kappa(x) = (1/\kappa^l)\rho(x/\kappa)$ for $\kappa > 0$, where $\rho(x)$ is a mollifier on $\mathbb{R}^l$. We define, for $i, j, j' \in \mathbb{N}$ and $\alpha > 0$,

$$m_s^{i,j,j'}(x) = M_s^{i,j,j'}\bigl(\hat{m}_s^{i,j'}(x)\eta_j^2(x) + e^{-\alpha|x|}\bigl(1 - \eta_j^2(x)\bigr)\bigr) \tag{8.2a}$$

and

$$M_s^{i,j,j'} = \left(\int_{\mathbb{R}^l} \bigl(\hat{m}_s^{i,j'}(x)\eta_j^2(x) + e^{-\alpha|x|}\bigl(1 - \eta_j^2(x)\bigr)\bigr)\,dx\right)^{-1}, \tag{8.2b}$$

where

$$\hat{m}_s^{i,j'}(x) = \int_{\mathbb{R}^l} \rho_{1/i}(\tilde{x})\hat{m}_s^{j'}(x - \tilde{x})\,d\tilde{x}, \qquad \hat{m}_s^{j'}(x) = m_s(x) \wedge j' \vee \frac{1}{j'}. \tag{8.3}$$

We note that, thanks to Theorem 6.1, $\hat{m}_s^{j'} \in \mathbb{W}^{1,2}_{\text{loc}}(\mathbb{R}^l)$.

We use Lemma 6.7 to define $X^{i,j,j'}$ as the solution of the equation

$$\dot{X}_s^{i,j,j'} = \int_{\mathbb{R}^l} A_s\bigl(X_s^{i,j,j'}, x\bigr) m_s^{i,j,j'}(x)\,dx,$$

with $X_0^{i,j,j'} = X_0$. The densities $m_s^{i,j,j'}(x)$ are of class $\mathbb{C}^1$ in x, with bounded derivatives, and are locally bounded away from zero, and the $X^{i,j,j'}$ are locally Lipschitz continuous by Lemma 6.7.

We introduce further

$$M_s^{j,j'} = \left(\int_{\mathbb{R}^l} \bigl(\hat{m}_s^{j'}(x)\eta_j^2(x) + e^{-\alpha|x|}\bigl(1 - \eta_j^2(x)\bigr)\bigr)\,dx\right)^{-1}, \tag{8.4a}$$

$$m_s^{j,j'}(x) = M_s^{j,j'}\bigl(\hat{m}_s^{j'}(x)\eta_j^2(x) + e^{-\alpha|x|}\bigl(1 - \eta_j^2(x)\bigr)\bigr) \tag{8.4b}$$

and

$$\dot{X}_s^{j,j'} = \int_{\mathbb{R}^l} A_s\bigl(X_s^{j,j'}, x\bigr) m_s^{j,j'}(x)\,dx, \qquad X_0^{j,j'} = X_0. \tag{8.4c}$$

Let also

$$M_s^j = \left(\int_{\mathbb{R}^l} \big(m_s(x) \eta_j^2(x) + e^{-\alpha|x|} \big(1 - \eta_j^2(x)\big)\big)\, dx \right)^{-1}, \tag{8.5a}$$

$$m_s^j(x) = M_s^j \big(m_s(x) \eta_j^2(x) + e^{-\alpha|x|} \big(1 - \eta_j^2(x)\big)\big) \tag{8.5b}$$

and

$$\dot{X}_s^j = \int_{\mathbb{R}^l} A_s(X_s^j, x) m_s^j(x)\, dx, \qquad X_0^j = X_0.$$

We have that

$$\begin{aligned} &\lim_{i\to\infty} M_s^{i,j,j'} = M_s^{j,j'}, \qquad \lim_{i\to\infty} \int_{\mathbb{R}^l} \big| m_s^{i,j,j'}(x) - m_s^{j,j'}(x) \big|\, dx = 0, \\ &\lim_{i\to\infty} X_s^{i,j,j'} = X_s^{j,j'}, \end{aligned} \tag{8.6a}$$

$$\begin{aligned} &\lim_{j'\to\infty} M_s^{j,j'} = M_s^j, \qquad \lim_{j'\to\infty} \int_{\mathbb{R}^l} \big| m_s^{j,j'}(x) - m_s^j(x) \big|\, dx = 0, \\ &\lim_{j'\to\infty} X_s^{j,j'} = X_s^j, \end{aligned} \tag{8.6b}$$

and

$$\begin{aligned} &\lim_{j\to\infty} M_s^j = 1, \qquad \lim_{j\to\infty} \int_{\mathbb{R}^l} \big| m_s^j(x) - m_s(x) \big|\, dx = 0, \\ &\lim_{j\to\infty} X_s^j = X_s. \end{aligned} \tag{8.6c}$$

The third convergence on each line is proved by a similar compactness argument to the one used in the proof of Lemma 6.7.

By (6.22),

$$\begin{aligned} \mathbf{I}_t^{**}(X, \mu) = \int_0^t \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \bigg(\int_{\mathbb{R}^l} \bigg(Dh(x)^T \bigg(\frac{1}{2} \operatorname{div}\big(c_s(X_s, x) m_s(x)\big) \\ - a_s(X_s, x) m_s(x) \bigg) - \frac{1}{2} \| Dh(x) \|^2_{c_s(X_s,x)} m_s(x) \bigg)\, dx \bigg)\, ds \end{aligned} \tag{8.7}$$

and

$$\begin{aligned} &\mathbf{I}_t^{**}\big(X^{i,j,j'}, \mu^{i,j,j'}\big) \\ &\quad = \int_0^t \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \bigg(\int_{\mathbb{R}^l} \bigg(Dh(x)^T \bigg(\frac{1}{2} \operatorname{div}\big(c_s(X_s^{i,j,j'}, x) m_s^{i,j,j'}(x)\big) \\ &\qquad - a_s\big(X_s^{i,j,j'}, x\big) m_s^{i,j,j'}(x) \bigg) - \frac{1}{2} \| Dh(x) \|^2_{c_s(X_s^{i,j,j'},x)} m_s^{i,j,j'}(x) \bigg)\, dx \bigg)\, ds. \end{aligned}$$

By (8.2a),

$$\mathbf{I}_t^{**}(X^{i,j,j'},\mu^{i,j,j'}) \le M_s^{i,j,j'}\Big(\int_0^t I_1^j(X_s^{i,j,j'},\hat{m}_s^{i,j'},s)\,ds+\int_0^t I_2^j(X_s^{i,j,j'},s)\,ds\Big), \tag{8.8}$$

where, for generic $\tilde{X}_s$ and $\tilde{m}_s$,

$$I_1^j(\tilde{X}_s,\tilde{m}_s,s)=\sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\Big(Dh(x)^T\Big(\frac{1}{2}\operatorname{div}\big(c_s(\tilde{X}_s,x)\eta_j^2(x)\tilde{m}_s(x)\big) - a_s(\tilde{X}_s,x)\eta_j^2(x)\tilde{m}_s(x)\Big)-\frac{1}{2}\|Dh(x)\|^2_{c_s(\tilde{X}_s,x)}\eta_j^2(x)\tilde{m}_s(x)\Big)\,dx \tag{8.9a}$$

and

$$I_2^j(\tilde{X}_s,s)=\sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}(Dh(x)^T\Big(\frac{1}{2}\operatorname{div}\big(c_s(\tilde{X}_s,x)e^{-\alpha|x|}\big(1-\eta_j^2(x)\big)\big) - a_s(\tilde{X}_s,x)e^{-\alpha|x|}\big(1-\eta_j^2(x)\big)) - \frac{1}{2}\|Dh(x)\|^2_{c_s(\tilde{X}_s,x)}e^{-\alpha|x|}\big(1-\eta_j^2(x)\big)\Big)\,dx. \tag{8.9b}$$

We prove that

$$\lim_{i\to\infty} I_1^j(X_s^{i,j,j'},\hat{m}_s^{i,j'},s)=I_1^j(X_s^{j,j'},\hat{m}_s^{j'},s). \tag{8.10}$$

Let, in analogy with (8.1),

$$k_s^{i,j,j'}(x)=\frac{1}{2\eta_j^2(x)\hat{m}_s^{i,j'}(x)}\operatorname{div}\big(c_s(X_s^{i,j,j'},x)\eta_j^2(x)\hat{m}_s^{i,j'}(x)\big) - a_s(X_s^{i,j,j'},x). \tag{8.11}$$

This function is an element of $\mathbb{L}^2(\mathbb{R}^l,\mathbb{R}^l,\eta_j^2(x)\hat{m}_s^{i,j'}(x)\,dx)$.

The supremum in $I_1^j(X_s^{i,j,j'},\hat{m}_s^{i,j'},s)$ is attained at a unique element $g_s^{i,j,j'}$ of $\mathbb{L}_0^{1,2}(\mathbb{R}^l,\mathbb{R}^l,\eta_j^2(x)\hat{m}_s^{i,j'}(x)\,dx)$ such that

$$\int_{\mathbb{R}^l}Dp(x)^Tk_s^{i,j,j'}(x)\eta_j^2(x)\hat{m}_s^{i,j'}(x)\,dx = \int_{\mathbb{R}^l}Dp(x)^Tc_s\big(X_s^{i,j,j'},x\big)g_s^{i,j,j'}(x)\eta_j^2(x)\hat{m}_s^{i,j'}(x)\,dx \tag{8.12}$$

for all $p\in\mathbb{C}_0^1(\mathbb{R}^l)$ and

$$I_1^j\big(X_s^{i,j,j'},\hat{m}_s^{i,j'},s\big)=\int_{\mathbb{R}^l}\frac{1}{2}\|g_s^{i,j,j'}(x)\|^2_{c_s(X_s^{i,j,j'},x)}\hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx. \tag{8.13}$$

Similarly, the supremum in $I_1^j(X_s^{j,j'}, \hat{m}_s^{j'}, s)$ is attained at a unique element $g_s^{j,j'}$ of $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, \eta_j^2(x)\hat{m}_s^{j'}(x)\,dx)$ such that

$$\begin{aligned}\int_{\mathbb{R}^l} Dp(x)^T k_s^{j,j'}(x) m_s^{j'}(x)\eta_j^2(x)\,dx \\ = \int_{\mathbb{R}^l} Dp(x)^T c_s(X_s^j, x) g_s^{j,j'}(x) m_s^{j'}(x)\eta_j^2(x)\,dx\end{aligned} \tag{8.14}$$

and

$$I_1^j(X_s^{j,j'}, \hat{m}_s^{j'}, s) = \int_{\mathbb{R}^l} \frac{1}{2}\|g_s^{j,j'}(x)\|^2_{c_s(X_s^{j,j'},x)} \hat{m}_s^{j'}(x)\eta_j^2(x)\,dx, \tag{8.15}$$

where

$$k_s^{j,j'}(x) = \frac{1}{2\eta_j^2(x)\hat{m}_s^{j'}(x)} \operatorname{div}\bigl(c_s(X_s^{j,j'}, x)\eta_j^2(x)\hat{m}_s^{j'}(x)\bigr) - a_s(X_s^{j,j'}, x). \tag{8.16}$$

Let

$$Q_1 = \int_{\mathbb{R}^l} \|Dp(x)\|^2_{c_s(X_s^{i,j,j'},x)} \hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx,$$

$$Q_2 = \int_{\mathbb{R}^l} \|k_s^{i,j,j'}(x)\hat{m}_s^{i,j'}(x) - k_s^{j,j'}(x)\hat{m}_s^{j'}(x)\|^2_{c_s(X_s^{i,j,j'},x)^{-1}} \frac{\eta_j^2(x)}{\hat{m}_s^{i,j'}(x)}\,dx$$

and

$$Q_3 = \int_{\mathbb{R}^l} \|c_s(X_s^{j,j'}, x) g_s^{j,j'}(x)\|^2_{c_s(X_s^{i,j,j'},x)^{-1}} \frac{\hat{m}_s^{j'}(x)^2\eta_j^2(x)}{\hat{m}_s^{i,j'}(x)}\,dx.$$

By (8.12) and (8.14), we have that

$$\begin{aligned}&\int_{\mathbb{R}^l} Dp(x)^T c_s(X_s^{i,j,j'}, x) g_s^{i,j,j'}(x)\hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx \\ &\quad = \int_{\mathbb{R}^l} Dp(x)^T\bigl(k_s^{i,j,j'}(x)\hat{m}_s^{i,j'}(x) - k_s^{j,j'}(x)\hat{m}_s^{j'}(x)\bigr)\eta_j^2(x)\,dx \\ &\qquad + \int_{\mathbb{R}^l} Dp(x)^T c_s(X_s^{j,j'}, x) g_s^{j,j'}(x)\hat{m}_s^{j'}(x)\eta_j^2(x)\,dx \\ &\quad \le \sqrt{Q_1}\sqrt{Q_2} + \sqrt{Q_1}\sqrt{Q_3}.\end{aligned}$$

Hence,

$$\begin{aligned}&\sqrt{\int_{\mathbb{R}^l} \|g_s^{i,j,j'}(x)\|^2_{c_s(X_s^{i,j,j'},x)} \hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx} \\ &\quad = \sup_{p\in\mathbb{C}_0^1(\mathbb{R}^l): Q_1\le 1} \int_{\mathbb{R}^l} Dp(x) c_s(X_s^{i,j,j'}, x) g_s^{i,j,j'}(x)\hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx \\ &\quad \le \sqrt{Q_2} + \sqrt{Q_3}.\end{aligned}$$

By (8.13), for arbitrary $\kappa > 0$,

$$I_1^j\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big) \le \frac{1}{2}\bigg(1+\frac{1}{\kappa}\bigg)Q_2 + \frac{1}{2}(1+\kappa)Q_3.$$

By (8.3), $\|\hat{m}_s^{i,j'} - \hat{m}_s^{j'}\|_{\mathbb{W}^{1,2}(S_{2j})} \to 0$ as $i \to \infty$ (see, e.g., Lemma 3.16 on page 66 in Adams and Fournier [1]), so, on recalling (8.11), (8.16) and Condition 2.2, we have that $Q_2 \to 0$ as $i \to \infty$. The integrand in Q_3 tends to $\|g_s^{j,j'}(x)\|^2_{c_s(X_s^{j,j'},x)} \hat{m}_s^{j'}(x)\eta_j^2(x)$ in Lebesgue measure; see (8.6a). Since the function $\hat{m}_s^{j'}(x)/\hat{m}_s^{i,j'}(x)$ is bounded in x and i, by dominated convergence, Q_3 converges to $\int_{\mathbb{R}^l} \|g_s^{j,j'}(x)\|^2_{c_s(X_s^{j,j'},x)} \hat{m}_s^{j'}(x)\eta_j^2(x)\,dx$ as $i \to \infty$ so that, on recalling (8.15),

$$\limsup_{i\to\infty} I_1^j\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big) \le I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big).$$

On the other hand, by (8.9a) and integration by parts,

$$\begin{aligned} &I_1^j\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big) \\ &\quad = \sup_{h\in\mathbb{C}_0^2(\mathbb{R}^l)} \int_{\mathbb{R}^l}\bigg(-\frac{1}{2}\operatorname{tr}\big(c_s\big(X_s^{i,j,j'},x\big)D^2h(x)\big) \\ &\qquad - Dh(x)^T a_s\big(X_s^{i,j,j'},x\big) - \frac{1}{2}\|Dh(x)\|^2_{c_s(X_s^{i,j,j'},x)}\bigg)\hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx \end{aligned}$$

and a similar representation holds for $I_1^j(X_s^{j,j'}, \hat{m}_s^{j'}, s)$, which facts imply, in view of (8.6a) and the continuity properties in Condition 2.1, that

$$\liminf_{i\to\infty} I_1^j\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big) \ge I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big). \tag{8.17}$$

We have proved (8.10). We now show that integrals with respect to s converge too. Let us note that, by (8.12) and (8.13),

$$\begin{aligned} &I_1^j\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big) \\ &\qquad \le \int_{\mathbb{R}^l} \frac{1}{2}\|k_s^{i,j,j'}(x)\|^2_{c_s(X_s^{i,j,j'},x)^{-1}} \hat{m}_s^{i,j'}(x)\eta_j^2(x)\,dx, \end{aligned}$$

so, by (8.11), and Conditions 2.1 and 2.2 there exists $M > 0$ such that

$$\begin{aligned} &I_1^{i,j,j'}\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big) \\ &\qquad \le M\bigg(1 + \int_{\mathbb{R}^l} \frac{|D\hat{m}_s^{i,j'}(x)|^2}{\hat{m}_s^{i,j'}(x)}\eta_j^2(x)\,dx\bigg). \end{aligned}$$

Accounting for (8.3), we have that

$$\begin{aligned}
&\frac{1}{2}\frac{|D\hat{m}_s^{i,j'}(x)|^2}{\hat{m}_s^{i,j'}(x)} \\
&\quad= \sup_{y\in\mathbb{R}^l}\Big(y^T D\hat{m}_s^{i,j'}(x) - \frac{1}{2}|y|^2\hat{m}_s^{i,j'}(x)\Big) \\
&\quad\le \int_{\mathbb{R}^l}\rho_{1/i}(\tilde{x})\sup_{y\in\mathbb{R}^l}\Big(y^T D\hat{m}_s^{j'}(x-\tilde{x}) - \frac{1}{2}|y|^2\hat{m}_s^{j'}(x-\tilde{x})\Big)\,d\tilde{x} \\
&\quad= \frac{1}{2}\int_{\mathbb{R}^l}\rho_{1/i}(\tilde{x})\frac{|D\hat{m}_s^{j'}(x-\tilde{x})|^2}{\hat{m}_s^{j'}(x-\tilde{x})}\,d\tilde{x}.
\end{aligned}$$

Therefore, recalling that $\int_{\mathbb{R}^l}\rho_{1/i}(x)\,dx = 1$ and the definition of $\hat{m}_s^{j'}(x)$ in (8.3),

$$\int_{\mathbb{R}^l}\frac{|D\hat{m}_s^{i,j'}(x)|^2}{\hat{m}_s^{i,j'}(x)}\eta_j^2(x)\,dx \le \int_{\mathbb{R}^l}\frac{|D\hat{m}_s^{j'}(x)|^2}{\hat{m}_s^{j'}(x)}\,dx \le \int_{\mathbb{R}^l}\frac{|Dm_s(x)|^2}{m_s(x)}\,dx. \tag{8.18}$$

Since $\int_0^t\int_{\mathbb{R}^l}|Dm_s(x)|^2/m_s(x)\,dx\,ds < \infty$ by Theorem 6.1, (8.10) and the dominated convergence theorem yield the convergence

$$\lim_{i\to\infty}\int_0^t I_1^j\big(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s\big)\,ds = \int_0^t I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big)\,ds. \tag{8.19}$$

Let us show that

$$\lim_{j'\to\infty}\int_0^t I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big)\,ds = \int_0^t I_1^j\big(X_s^{j}, m_s, s\big)\,ds. \tag{8.20}$$

We have that

$$\begin{aligned}
&\big|I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big) - I_1^j\big(X_s^{j,j'}, m_s, s\big)\big| \\
&\quad\le \frac{1}{2}\int_{\mathbb{R}^l}\Bigg(\bigg\|\frac{1}{2\eta_j^2(x)m_s(x)}\operatorname{div}\big(c_s\big(X_s^{j,j'},x\big)\eta_j^2(x)m_s(x)\big) \\
&\qquad - a_s\big(X_s^{j,j'},x\big)\bigg\|^2_{c_s(X_s^{j,j'},x)^{-1}} \\
&\qquad + \bigg\|\frac{1}{2\eta_j^2(x)}\operatorname{div}\big(c_s\big(X_s^{j,j'},x\big)\eta_j^2(x)\big) - a_s\big(X_s^{j,j'},x\big)\bigg\|^2_{c_s(X_s^{j,j'},x)^{-1}}\Bigg) \\
&\qquad \times \eta_j^2(x)m_s(x)\big(1 - \mathbf{1}_{[1/j',j']}\big(m_s(x)\big)\big)\,dx,
\end{aligned}$$

so, by dominated convergence,

$$\lim_{j'\to\infty}\int_0^t\big|I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big) - I_1^j\big(X_s^{j,j'}, m_s, s\big)\big|\,ds = 0. \tag{8.21}$$

Let $\vartheta > 0$ be such that $\|y\|^2_{c_s(u,x)} \ge \vartheta |y|^2$, for all $s \in [0,t]$, for all u from a large enough ball, all x and all y. By the convergence of $X^{j,j'}$ to X^j as $j' \to \infty$, the continuity of $c_s(u,x)$ in u locally uniformly in s and uniformly in x, and by $c_s(u,x)$ being positive definite uniformly in x and locally uniformly in (s,u), given arbitrary $\delta \in (0,1)$ and $\kappa \in (0,1)$, for all j' great enough, locally uniformly in s,

$$
\begin{aligned}
I_1^j&\big(X_s^{j,j'}, m_s, s\big) \\
&\le \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \Big(Dh(x)^T k_s^j(x) \\
&\quad - \frac{1}{2}(1-\delta)(1-\kappa) \| Dh(x)\|^2_{c_s(X_s^j,x)} \Big) m_s(x) \eta_j^2(x)\, dx \\
&\quad + \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \bigg(\int_{\mathbb{R}^l} \Big(Dh(x)^T \Big(\frac{1}{2} \operatorname{div}\big((c_s(X_s^{j,j'},x) - c_s(X_s^j,x)) m_s(x) \eta_j^2(x)\big) \\
&\quad - \big(a_s(X_s^{j,j'},x) - a_s(X_s^j,x)\big) m_s(x) \eta_j^2(x) \Big) \\
&\quad - \frac{1}{2}\delta(1-\kappa) \| Dh(x)\|^2_{c_s(X_s^j,x)} m_s(x)\eta_j^2(x) \Big)\, dx \bigg) \\
&\le (1-\delta)^{-1}(1-\kappa)^{-1} \\
&\quad \times \sup_{h \in \mathbb{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \Big(Dh(x)^T k_s^j(x) - \frac{1}{2} \| Dh(x)\|^2_{c_s(X_s^j,x)} \Big) m_s(x) \eta_j^2(x)\, dx \\
&\quad + \delta^{-1}(1-\kappa)^{-1} \frac{\vartheta^{-1}}{2} \int_{\mathbb{R}^l} \bigg| \frac{1}{2} \frac{\operatorname{div}((c_s(X_s^{j,j'},x) - c_s(X_s^j,x)) m_s(x)\eta_j^2(x))}{m_s(x)\eta_j^2(x)} \\
&\quad - \big(a_s(X_s^{j,j'},x) - a_s(X_s^j,x)\big) \bigg|^2 m_s(x) \eta_j^2(x)\, dx.
\end{aligned} \tag{8.22}
$$

By the convergence of $X_s^{j,j'}$ to X_s^j as $j' \to \infty$, Condition 2.1 and the convergence of $\int_0^t \int_{\mathbb{R}^l} |Dm_s(x)|^2 / m_s(x)\, dx\, ds$, the integral from 0 to t of the second integral on the rightmost side of (8.22) tends to zero as $j' \to \infty$. Therefore, by (8.21), (8.1), (8.7) and (8.9a),

$$
\limsup_{j' \to \infty} \int_0^t I_1^j\big(X_s^{j,j'}, \hat{m}_s^{j'}, s\big)\, ds \le \int_0^t I_1^j\big(X_s^j, m_s, s\big)\, ds
$$

and by an analogue of (8.17), we obtain (8.20).

We now take a limit as $j \to \infty$. By a similar reasoning to the one used in (8.22), given arbitrary $\delta \in (0,1)$ and $\kappa \in (0,1)$, for all j great enough, locally uniformly in s,

$$
\begin{aligned}
&I_1^j(X_s^j, m_s, s)\\
&\quad\le \delta^{-1}(1-\delta)^{-1}(1-\kappa)^{-1}\frac{1}{2}\int_{\mathbb{R}^l}\|D\eta_j(x)\|^2_{c_s(X_s,x)}m_s(x)dx\\
&\qquad + (1-\delta)^{-2}(1-\kappa)^{-1}\\
&\qquad \times \sup_{h\in\mathbb{C}_0^1(\mathbb{R}^l)}\int_{\mathbb{R}^l}\Big(Dh(x)^T k_s(x) - \frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)}\Big)m_s(x)\eta_j^2(x)\,dx\\
&\qquad + \delta^{-1}(1-\kappa)^{-1}\frac{\vartheta^{-1}}{2}\int_{\mathbb{R}^l}\bigg|\frac{1}{2}\frac{\operatorname{div}((c_s(X_s^j,x)-c_s(X_s,x))m_s(x)\eta_j^2(x))}{m_s(x)\eta_j^2(x)}\\
&\qquad - \big(a_s(X_s^j,x)-a_s(X_s,x)\big)\bigg|^2 m_s(x)\eta_j^2(x)\,dx
\end{aligned}
$$

so that, by Condition 2.3 (with $\lambda = 0$) and Condition 2.1, we have, on recalling (8.7), that

$$
\lim_{j\to\infty}\int_0^t I_1^j(X_s^j, m_s, s)\,ds = \mathbf{I}_t^{**}(X,\mu). \tag{8.23}
$$

Putting together (8.19), (8.20) and (8.23) yields the convergence

$$
\lim_{j\to\infty}\lim_{j'\to\infty}\lim_{i\to\infty}\int_0^t I_1^j(X_s^{i,j,j'}, \hat{m}_s^{i,j'}, s)\,ds = \mathbf{I}_t^{**}(X,\mu). \tag{8.24}
$$

We now show that the term I_2^j is inconsequential. On recalling that $|a_s(u,x)|$ grows at most linearly in $|x|$ and $|\operatorname{div} c_s(u,x)|$ and $\|c_s(u,x)\|$ are bounded in x locally uniformly in (s,u), we have that, for some $L > 0$, all (i,j), and all $s \le t$, according to (8.9b),

$$
\begin{aligned}
I_2^j(X_s^{i,j,j'}, s) &\le \int_{x\in\mathbb{R}^l:|x|\ge j}\sup_{y\in\mathbb{R}^l}\bigg(y^T\bigg(\frac{1}{2}\frac{\operatorname{div}(c_s(X_s^{i,j,j'},x)(1-\eta_j^2(x)))}{1-\eta_j^2(x)}\\
&\qquad - \alpha c_s(X_s^{i,j,j'},x)\frac{x}{2|x|} - a_s(X_s^{i,j,j'},x)\bigg) - \frac{1}{2}\|y\|^2_{c_s(X_s^{i,j,j'},x)}\bigg)\\
&\qquad \times \big(1-\eta_j^2(x)\big)e^{-\alpha|x|}\,dx\\
&\le \int_{x\in\mathbb{R}^l:|x|\ge j} L\bigg(1+\alpha^2+|x|^2+\frac{1}{j^2}\frac{|D\eta(|x|/j)|^2}{1-\eta^2(|x|/j)}\bigg)e^{-\alpha|x|}\,dx.
\end{aligned}
$$

Since $\eta(y) = 0$ for $y \ge 2$ and (2.6) holds, the latter integral tends to 0 as $j \to \infty$, so,

$$
\lim_{j\to\infty}\limsup_{j'\to\infty}\limsup_{i\to\infty}\int_0^t I_2^j(X_s^{i,j,j'}, s)\,ds = 0. \tag{8.25}
$$

By (8.6a), (8.6b), (8.6c), (8.8), (8.24) and (8.25),

$$\limsup_{j\to\infty}\limsup_{j'\to\infty}\limsup_{i\to\infty}\mathbf{I}_t^{**}(X^{i,j,j'},\mu^{i,j,j'})\le\mathbf{I}_t^{**}(X,\mu).$$

Thus, there exist sequences $j'(j)\to\infty$ and $i(j)\to\infty$ as $j\to\infty$ such that $(X^{i(j),j,j'(j)},\mu^{i(j),j,j'(j)})\to(X,\mu)$ and

$$\limsup_{j\to\infty}\mathbf{I}_t^{**}(X^{i(j),j,j'(j)},\mu^{i(j),j,j'(j)})\le\mathbf{I}_t^{**}(X,\mu).$$

The reverse inequality follows from the lower semicontinuity of $\mathbf{I}_t^{**}(X,\mu)$ [see (6.22), where we let $\mathbf{I}_t^{**}(X,\mu)=\infty$ if $\mathbf{I}^{**}(X,\mu)=\infty$], so

$$\lim_{j\to\infty}\mathbf{I}_t^{**}(X^{i(j),j,j'(j)},\mu^{i(j),j,j'(j)})=\mathbf{I}_t^{**}(X,\mu), \tag{8.26}$$

and one can take $(X^{(j)},\mu^{(j)})=(X^{i(j),j,j'(j)},\mu^{i(j),j,j'(j)})$.

Suppose now that $C_t(u,x)-G_t(u,x)c_t(u,x)^{-1}G_t(u,x)^T$ is positive definite uniformly in x and locally uniformly in (t,u). We proceed similar to the case where $C_t(u,x)=0$ and define $m_s^{i,j,j'}(x)$, $M_s^{i,j,j'}$, $M_s^{j,j'}$, M_s^j, and $m_s^j(x)$ by the respective relations (8.2a), (8.2b), (8.3), (8.4a)–(8.4c), (8.5a) and (8.5b). We let

$$\dot X_s^{i,j,j'}=\dot X_s^{j,j'}=\dot X_s^j=\dot X_s\mathbf{1}_{\{|\dot X_s|\le j\}}(s),\qquad X_0^{i,j,j'}=X_0^{j,j'}=X_0^j=X_0.$$

The convergences in (8.6a), (8.6b) and (8.6c) still hold.

The following reasoning is sketchy out of necessity. Replacing $a_s(X_s,x)$ with $a_s(X_s,x)+G_s(X_s,x)^T\lambda$ in the proof above, one can see in analogy with (8.26), that there exist sequences $i(j)\to\infty$ and $j'(j)\to\infty$ as $j\to\infty$ such that, for all $\lambda\in\mathbb{R}^n$ with rational components and for almost all $s\in[0,t]$,

$$\begin{aligned}\lim_{j\to\infty}\int_0^t\int_{\mathbb{R}^l}\Bigg\|&\frac{Dm_s^{(j)}(x)}{2m_s^{(j)}(x)}-\Phi_{s,m_s^{(j)}(\cdot),X_s^{(j)}}(x)\\&-\Psi_{s,m_s^{(j)}(\cdot),X_s^{(j)}}(x)\lambda\Bigg\|^2_{c_s(X_s^{(j)},x)}m_s^{(j)}(x)\,dx\,ds\\=\int_0^t\int_{\mathbb{R}^l}\Bigg\|&\frac{Dm_s(x)}{2m_s(x)}-\Phi_{s,m_s(\cdot),X_s}(x)\\&-\Psi_{s,m_s(\cdot),X_s}(x)\lambda\Bigg\|^2_{c_s(X_s,x)}m_s(x)\,dx\,ds\end{aligned}\tag{8.27a}$$

and

$$\begin{aligned}&\lim_{j\to\infty}\int_{\mathbb{R}^l}\Bigg\|\frac{Dm_s^{(j)}(x)}{2m_s^{(j)}(x)}-\Phi_{s,m_s^{(j)}(\cdot),X_s^{(j)}}(x)-\Psi_{s,m_s^{(j)}(\cdot),X_s^{(j)}}(x)\lambda\Bigg\|^2_{c_s(X_s^{(j)},x)}m_s^{(j)}(x)\,dx\\&\quad=\int_{\mathbb{R}^l}\Bigg\|\frac{Dm_s(x)}{2m_s(x)}-\Phi_{s,m_s(\cdot),X_s}(x)-\Psi_{s,m_s(\cdot),X_s}(x)\lambda\Bigg\|^2_{c_s(X_s,x)}m_s(x)\,dx,\end{aligned}$$

respectively, where $m^{i(j),j,j'(j)}$ is relabeled as $m^{(j)}$ and X_s^j, as $X_s^{(j)}$.

It follows that

$$\lim_{j\to\infty}\int_{\mathbb{R}^l} G_s\big(X_s^{(j)},x\big)\biggl(\frac{Dm_s^{(j)}(x)}{2m_s^{(j)}(x)}-\Phi_{s,m_s^{(j)}(\cdot),X_s^{(j)}}(x)\biggr)m_s^{(j)}(x)\,dx$$
$$=\int_{\mathbb{R}^l} G_s(X_s,x)\biggl(\frac{Dm_s(x)}{2m_s(x)}-\Phi_{s,m_s(\cdot),X_s}(x)\biggr)m_s(x)\,dx$$

and

$$\lim_{j\to\infty}\int_{\mathbb{R}^l}\big\|\Psi_{s,m^{(j)}(\cdot),X_s^{(j)}}(x)\big\|^2_{c_s(X_s^{(j)},x)}m_s^{(j)}(x)\,dx$$
$$=\int_{\mathbb{R}^l}\big\|\Psi_{s,m(\cdot),X_s}(x)\big\|^2_{c_s(X_s,x)}m_s(x)\,dx,$$

so, by (2.14),

$$\lim_{j\to\infty}\int_{\mathbb{R}^l} Q_{s,m_s^{(j)}(\cdot)}\big(X_s^{(j)},x\big)m_s^{(j)}(x)\,dx=\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx.$$

By dominated convergence,

$$\lim_{j\to\infty}\int_0^t\biggl\|\dot X_s^{(j)}-\int_{\mathbb{R}^l} A_s\big(X_s^{(j)},x\big)m_s^{(j)}(x)\,dx-\int_{\mathbb{R}^l} G_s\big(X_s^{(j)},x\big)\biggl(\frac{Dm_s^{(j)}(x)}{2m_s^{(j)}(x)}$$
$$-\Phi_{s,m_s^{(j)}(\cdot),X_s^{(j)}}(x)\biggr)m_s^{(j)}(x)\,dx\biggr\|^2_{(\int_{\mathbb{R}^l} Q_{s,m_s^{(j)}(\cdot)}(X_s^{(j)},x)m_s^{(j)}(x)\,dx)^{-1}}ds$$
$$=\int_0^t\biggl\|\dot X_s-\int_{\mathbb{R}^l} A_s(X_s,x)m_s(x)\,dx-\int_{\mathbb{R}^l} G_s(X_s,x)\biggl(\frac{Dm_s(x)}{2m_s(x)}$$
$$-\Phi_{s,m_s(\cdot),X_s}(x)\biggr)m_s(x)\,dx\biggr\|^2_{(\int_{\mathbb{R}^l} Q_{s,m_s(\cdot)}(X_s,x)m_s(x)\,dx)^{-1}}ds,$$

which completes the proof by (6.21) and (8.27a).

We have thus proved that in both cases there exist $(X^{(j)},\mu^{(j)})$ with needed regularity properties that converge to (X,μ) and are such that $\mathbf{I}_t^{**}(X^{(j)},\mu^{(j)})\to\mathbf{I}_t^{**}(X,\mu)$ for all $t\in\mathbb{R}_+$. Picking a suitable subsequence, we can assume that $\mathbf{I}_j^{**}(X^{(j)},\mu^{(j)})\le\mathbf{I}_j^{**}(X,\mu)+1/j$. We redefine the subsequence $(X_t^{(j)},\mu_t^{(j)})$ for $t\ge j$ such that $\mathbf{I}_j^{**}(X^{(j)},\mu^{(j)})=\mathbf{I}^{**}(X^{(j)},\mu^{(j)})$, thanks to Lemma 6.7. The resulting sequence will still converge to (X,μ). In addition, $\limsup_{j\to\infty}\mathbf{I}^{**}(X^{(j)},\mu^{(j)})\le\mathbf{I}^{**}(X,\mu)$, which yields the assertion of Theorem 8.1 by the lower semicontinuity of $\mathbf{I}^{**}$. $\square$

9. Proof of Theorem 2.1. Suppose that $\lim_{\varepsilon\to0}\mathbf{P}^\varepsilon(|X_0^\varepsilon-\hat u|>\kappa)^\varepsilon=0$, for arbitrary $\kappa>0$. Then any large deviation limit point $\tilde{\mathbf{I}}$ of $\mathbf{P}^\varepsilon$ is such that $\tilde{\mathbf{I}}(X,\mu)=\infty$ unless $X_0=\hat u$. If (X,μ) is such that $X_0=\hat u$ and $\mathbf{I}^{**}(X,\mu)<\infty$, by

Theorems 5.1, 7.1 and 8.1, there exist (X^i, μ^i), which satisfy the hypotheses on $(\hat{X}, \hat{\mu})$ in Theorem 7.1, such that $\mathbf{I}^{**}(X^i, \mu^i) = \tilde{\mathbf{I}}(X^i, \mu^i)$, $(X^i, \mu^i) \to (X, \mu)$ as $i \to \infty$, and $\mathbf{I}^{**}(X^i, \mu^i) \to \mathbf{I}^{**}(X, \mu)$ as $i \to \infty$. By Theorem 3.4 (with the role of $\mathcal{U}$ being played by the set of functions $U^{\lambda(\cdot), f}_{t\wedge\tau}$ in Theorem 5.1 and with the role of $\tilde{\mathcal{U}}$ being played by the set of functions θ^N in Lemma 7.2) and Theorem 6.1, $\tilde{\mathbf{I}}(X, \mu) = \mathbf{I}^{**}(X, \mu) = \mathbf{I}(X, \mu)$ for all (X, μ).

In the general setting of Theorem 2.1, let $\mathcal{L}^\varepsilon_u$ denote the regular conditional distribution of $(X^\varepsilon, \mu^\varepsilon)$ given that $X^\varepsilon_0 = u$, where $u \in \mathbb{R}^n$ and is otherwise arbitrary. By what has been proved, if $u^\varepsilon \to \hat{u}$ as $\varepsilon \to 0$, then the $\mathcal{L}^\varepsilon_{u^\varepsilon}$ obey the LDP in $\mathbb{C}(\mathbb{R}_+, \mathbb{R}^n) \times \mathbb{C}_\uparrow(\mathbb{R}_+, \mathbb{M}(\mathbb{R}^l))$ with the large deviation function $\breve{\mathbf{I}}_{\hat{u}}$ as defined in the statement of Theorem 2.1, where $\mathbf{I}_0(\hat{u}) = 0$ and $\mathbf{I}_0(u) = \infty$ if $u \neq \hat{u}$. Since by the hypotheses of Theorem 2.1 the distributions of X^ε_0 obey the LDP with a large deviation function $\mathbf{I}_0$, it follows that the distributions of $(X^\varepsilon, \mu^\varepsilon)$ obey the LDP with $\mathbf{I}(X, \mu) = \mathbf{I}_0(X_0) + \breve{\mathbf{I}}_{X_0}(X, \mu)$; see, for example, Chaganty [9], Puhalskii [38]. Theorem 2.1 has been proved.

APPENDIX

PROOF OF LEMMA 2.1. By Theorem 6.1, if $\mathbf{I}'(X, \mu) < \infty$, then

$$\begin{aligned}\int_0^t \sup_{h \in \mathbb{C}^1_0(\mathbb{R}^l)} \int_{\mathbb{R}^l} \biggl(& Dh(x)^T \Bigl(\frac{1}{2} \operatorname{div}_x c_s(X_s, x) - a_s(X_s, x)\Bigr) \\ & - \frac{1}{2} \|Dh(x)\|^2_{c_s(X_s, x)}\biggr) m_s(x)\, dx\, ds < \infty. \end{aligned} \tag{A.1}$$

Suppose (2.8) holds and let L denote an upper bound on the left-hand side of (A.1). By (6.6) in the statement of Lemma 6.4 and Condition 2.1, L is also an upper bound on the integrals on the left of (A.1) for $h \in \mathbb{W}^{1,q}_0(S)$, where S is an open ball in $\mathbb{R}^l$, $q \geq 2$, and $q > l$. On taking $h(x) = \kappa(|x|^2 \vee r_1^2 \wedge r_2^2 - r_2^2)$, where $\kappa > 0$ and $0 < r_1 < r_2$, we have that

$$\begin{aligned}\int_0^t \int_{x \in \mathbb{R}^l : r_1 \le |x| \le r_2} \biggl(& \kappa x^T \Bigl(\frac{1}{2} \operatorname{div}_x c_s(X_s, x) - a_s(X_s, x)\Bigr) \\ & - \kappa^2 \|c_s(X_s, x)\| |x|^2 \biggr) m_s(x)\, dx\, ds \le L. \end{aligned}$$

If r_1 is great enough, there exists $\delta > 0$ such that $x^T a_s(X_s, x) \le -\delta |x|^2$ if $|x| \ge r_1$. Therefore, for small enough $\kappa > 0$, great enough r_1, and all $r_2 > r_1$,

$$\frac{\kappa\delta}{2} \int_0^t \int_{x \in \mathbb{R}^l : r_1 \le |x| \le r_2} |x|^2 m_s(x)\, dx\, ds \le L.$$

The square integrability of $a_s(X_s, x)$ now follows by it growing no faster than linearly in x; see Condition 2.1.

Suppose now (2.9) holds. We take, for given s, $\delta > 0$, and $r > 0$,

$$h(x) = -\big(\hat{a}_s(X_s, x) \vee (-\delta) \wedge \delta\big)\eta_r(x),$$

where $\eta_r(x) = \eta(|x|/r)$ and $\eta(y)$ satisfies the requirements of Condition 2.3. Then $h \in \mathbb{W}_0^{1,q}(S)$, for large enough ball S, and, for $\kappa \in (0, 1)$,

$$\begin{aligned}
&\int_{\mathbb{R}^l}\bigg(Dh(x)^T\bigg(\frac{1}{2}\operatorname{div}_x c_s(X_s, x) - a_s(X_s, x)\bigg) - \frac{1}{2}\|Dh(x)\|^2_{c_s(X_s,x)}\bigg)m_s(x)\,dx \\
&\quad\geq \frac{1-\kappa}{2}\int_{\mathbb{R}^l}\|D_x\hat{a}_s(X_s, x)\|^2_{c_s(X_s,x)}\mathbf{1}_{\{|\hat{a}_s(X_s,x)|\leq\delta\}}(s, x)\eta_r(x)^2 m_s(x)\,dx \\
&\qquad - \int_{\mathbb{R}^l}\big(\hat{a}_s(X_s, x) \vee (-\delta) \wedge \delta\big)\frac{1}{r}D\eta\bigg(\frac{|x|}{r}\bigg)\frac{x^T}{|x|} \\
&\qquad \times c_s(X_s, x)D_x\hat{a}_s(X_s, x)m_s(x)\,dx \\
&\qquad - \frac{1}{2r^2}\bigg(1 + \frac{1}{\kappa}\bigg)\int_{\mathbb{R}^l}\big(\hat{a}_s(X_s, x) \vee (-\delta) \wedge \delta\big)^2\bigg\|D\eta\bigg(\frac{|x|}{r}\bigg)\bigg\|^2_{c_s(X_s,x)} m_s(x)\,dx.
\end{aligned}$$

As $r \to \infty$, the integrals from 0 to t of the latter two integrals converge to zero [we recall that by Theorem 6.1, $\int_0^t\int_{\mathbb{R}^l}|x^T a_s(X_s, x)|/|x|m_s(x)\,dx\,ds < \infty$, so $\int_0^t\int_{\mathbb{R}^l}|x^T c_s(X_s, x)D_x\hat{a}_s(X_s, x)|/|x|m_s(x)\,dx\,ds < \infty$]. Therefore,

$$\frac{1}{2}\int_0^t\int_{\mathbb{R}^l}\bigg\|a_s(X_s, x) - \frac{1}{2}\operatorname{div}_x c_s(X_s, x)\bigg\|^2_{c_s(X_s,x)^{-1}} m_s(x)\,dx\,ds \leq L,$$

which implies the square integrability of $a_s(X_s, x)$ thanks to Conditions 2.1 and 2.2. □

REFERENCES

[1] ADAMS, R. A. and FOURNIER, J. J. F. (2003). *Sobolev Spaces*, 2nd ed. *Pure and Applied Mathematics* (*Amsterdam*) **140**. Elsevier/Academic Press, Amsterdam. MR2424078

[2] AGMON, S. (1959). The L_p approach to the Dirichlet problem. I. Regularity theorems. *Ann. Scuola Norm. Sup. Pisa* (3) **13** 405–448. MR0125306

[3] AUBIN, J.-P. and EKELAND, I. (1984). *Applied Nonlinear Analysis*. Wiley, New York. MR0749753

[4] AUBIN, J.-P. and FRANKOWSKA, H. (1990). *Set-Valued Analysis*. *Systems & Control*: *Foundations & Applications* **2**. Birkhäuser, Boston, MA. MR1048347

[5] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[6] BOGACHEV, V. I., KRYLOV, N. and RÖCKNER, M. (1996). Regularity of invariant measures: The case of non-constant diffusion part. *J. Funct. Anal.* **138** 223–242. MR1391637

[7] BOGACHEV, V. I., KRYLOV, N. V. and RËKNER, M. (2009). Elliptic and parabolic equations for measures. *Uspekhi Mat. Nauk* **64** 5–116. MR2640966

[8] BOGACHEV, V. I., KRYLOV, N. V. and RÖCKNER, M. (2001). On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions. *Comm. Partial Differential Equations* **26** 2037–2080. MR1876411

[9] CHAGANTY, N. R. (1997). Large deviations for joint distributions and statistical applications. *Sankhyā Ser. A* **59** 147–166. MR1665683

[10] DEUSCHEL, J. D. and STROOCK, D. W. (2001). *Large Deviations*, 2nd ed. Amer. Math. Soc., Providence.

[11] DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. *Comm. Pure Appl. Math.* **28** 1–47. MR0386024

[12] DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461. MR0428471

[13] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358

[14] EKELAND, I. and TEMAM, R. (1976). *Convex Analysis and Variational Problems*. North-Holland, Amsterdam. MR0463994

[15] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

[16] FENG, J. and KURTZ, T. G. (2006). *Large Deviations for Stochastic Processes*. *Mathematical Surveys and Monographs* **131**. Amer. Math. Soc., Providence, RI. MR2260560

[17] FREĬDLIN, M. I. (1978). The averaging principle and theorems on large deviations. *Uspekhi Mat. Nauk* **33** 107–160, 238. MR0511884

[18] FREĬDLIN, M. I. and WENTZELL, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **260**. Springer, New York. MR1652127

[19] GËRTNER, J. (1977). On large deviations from an invariant measure. *Teor. Verojatnost. i Primenen.* **22** 27–42. MR0471040

[20] GIKHMAN, I. I. and SKOROKHOD, A. V. (1982). *Stokhasticheskie Differentsialnye Uravneniya i Ikh Prilozheniya*. "Naukova Dumka", Kiev. MR0678374

[21] GILBARG, D. and TRUDINGER, N. S. (1983). *Elliptic Partial Differential Equations of Second Order*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **224**. Springer, Berlin. MR0737190

[22] HAS'MINSKIĬ, R. Z. (1969). *Ustoichivost Sistem Differentsialnykh Uravnenii Pri Sluchainykh Vozmushcheniyakh Ikh Parametrov*. Izdat. "Nauka", Moscow. MR0259283

[23] HEINONEN, J., KILPELÄINEN, T. and MARTIO, O. (1993). *Nonlinear Potential Theory of Degenerate Elliptic Equations*. Oxford Univ. Press, New York. MR1207810

[24] IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. *North-Holland Mathematical Library* **24**. North-Holland, Amsterdam. MR1011252

[25] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR0959133

[26] KRYLOV, N. V. (1980). *Controlled Diffusion Processes*. *Applications of Mathematics* **14**. Springer, New York. MR0601776

[27] KUFNER, A., JOHN, O. and FUČÍK, S. (1977). *Function Spaces*. Noordhoff International Publishing, Leyden. MR0482102

[28] LIPTSER, R. (1996). Large deviations for two scaled diffusions. *Probab. Theory Related Fields* **106** 71–104. MR1408417

[29] LIPTSER, R. SH. and SHIRYAYEV, A. N. (1989). *Theory of Martingales*. *Mathematics and Its Applications* (*Soviet Series*) **49**. Kluwer Academic, Dordrecht. MR1022664

[30] MALYSHKIN, M. N. (2000). Subexponential estimates for the rate of convergence to the invariant measure for stochastic differential equations. *Teor. Veroyatnost. i Primenen.* **45** 489–504. MR1967786

[31] METAFUNE, G., PALLARA, D. and RHANDI, A. (2005). Global properties of invariant measures. *J. Funct. Anal.* **223** 396–424. MR2142345

[32] MORREY, C. B. JR. (2008). *Multiple Integrals in the Calculus of Variations*. Springer, Berlin. MR2492985

[33] NEČAS, J. (2012). *Direct Methods in the Theory of Elliptic Equations*. Springer, Heidelberg. MR3014461

[34] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco. MR0198505

[35] PARDOUX, E. and VERETENNIKOV, A. YU. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736

[36] PUHALSKII, A. (1991). On functional principle of large deviations. In *New Trends in Probability and Statistics*, *Vol.* 1 (*Bakuriani*, 1990) (V. Sazonov and T. Shervashidze, eds.). 198–218. VSP, Utrecht. MR1200917

[37] PUHALSKII, A. (1992). Weak convergence theory approach to large deviations. In *Large Deviations and Applications*. Oberwolfach.

[38] PUHALSKII, A. (1995). Large deviation analysis of the single server queue. *Queueing Systems Theory Appl.* **21** 5–66. MR1372048

[39] PUHALSKII, A. (1997). Large deviations of semimartingales: A maxingale problem approach. I. Limits as solutions to a maxingale problem. *Stochastics Stochastics Rep.* **61** 141–243. MR1488137

[40] PUHALSKII, A. (2001). *Large Deviations and Idempotent Probability*. *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics* **119**. Chapman & Hall/CRC, Boca Raton, FL. MR1851048

[41] PUHALSKII, A. A. (2015). On large deviations of coupled diffusions with time scale separation. Available at arXiv:1306.5446.

[42] PUHALSKII, A. A. (2005). Stochastic processes in random graphs. *Ann. Probab.* **33** 337–412. MR2118868

[43] PUHALSKII, A. A. and VLADIMIROV, A. A. (2007). A large deviation principle for join the shortest queue. *Math. Oper. Res.* **32** 700–710. MR2348243

[44] PUKHAL'SKIĬ, A. A. (1993). On the theory of large deviations. *Teor. Veroyatnost. i Primenen.* **38** 553–562. MR1404664

[45] RÖCKNER, M. and ZHANG, T. S. (1992). Uniqueness of generalized Schrödinger operators and applications. *J. Funct. Anal.* **105** 187–231. MR1156676

[46] RÖCKNER, M. and ZHANG, T. S. (1994). Uniqueness of generalized Schrödinger operators. II. *J. Funct. Anal.* **119** 455–467. MR1261099

[47] SHAPOSHNIKOV, S. V. (2008). On interior estimates for the Sobolev norms of solutions of elliptic equations. *Mat. Zametki* **83** 316–320. MR2431591

[48] SOHR, H. (2001). *The Navier–Stokes Equations: An Elementary Functional Analytic Approach*. Birkhäuser, Basel. MR1928881

[49] STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **233**. Springer, Berlin. MR0532498

[50] TOPSØE, F. (1970). *Topology and Measure*. *Lecture Notes in Mathematics* **133**. Springer, Berlin. MR0422560

[51] VERETENNIKOV, A. (2013). On large deviations in the averaging principle for SDE's with a "full dependence", revisited [MR1681106]. *Discrete Contin. Dyn. Syst. Ser. B* **18** 523–549. MR2999091

[52] VERETENNIKOV, A. YU. (1992). On large deviations for ergodic process empirical measures. In *Topics in Nonparametric Estimation*. *Adv. Soviet Math.* **12** 125–133. Amer. Math. Soc., Providence, RI. MR1191695

[53] VERETENNIKOV, A. YU. (1994). Large deviations in averaging principle for stochastic differential equation systems (noncompact case). *Stochastics Stochastics Rep*. **48** 83–96. MR1786193

[54] VERETENNIKOV, A. YU. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process*. *Appl*. **70** 115–127. MR1472961

[55] VERETENNIKOV, A. YU. (1998). Letter to the editors: "On large deviations for stochastic differential equations with small diffusion and averaging." *Teor. Veroyatnost*. *i Primenen*. **43** 819. MR1692448

[56] VERETENNIKOV, A. YU. (1998). On large deviations for stochastic differential equations with small diffusion and averaging. *Teor. Veroyatnost*. *i Primenen*. **43** 349–351. MR1679007

[57] VERETENNIKOV, A. YU. (1998). On large deviations in the averaging principle for stochastic differential equations with "complete dependence." *Teor. Veroyatnost*. *i Primenen*. **43** 765–767. MR1692401

[58] VERETENNIKOV, A. YU. (1999). On large deviations in the averaging principle for SDEs with a "full dependence." *Ann*. *Probab*. **27** 284–296. MR1681106

[59] VERETENNIKOV, A. YU. (2000). On large deviations for SDEs with small diffusion and averaging. *Stochastic Process*. *Appl*. **89** 69–79. MR1775227

IITP RAS
BOLSHOY KARETNY PER. 19
MOSCOW 127994
RUSSIA
E-MAIL: puhalski@mailfrom.ru