Notation

(1) (E, r). A complete, separable metric space.
(2) $\mathcal{B}(E)$. The σ-algebra of all Borel subsets of E.
(3) $A \subset M(E) \times M(E)$. An operator identified with its graph as a subset in $M(E) \times M(E)$.
(4) $B(E)$. The space of bounded, Borel measurable functions. Endowed with the norm $\|f\| = \sup_{x\in E} |f(x)|$, $(B(E), \|\cdot\|)$ is a Banach space.
(5) $B_{loc}(E)$. The space of locally bounded, Borel measurable functions, that is, functions in $M(E)$ that are bounded on each compact.
(6) $B_\epsilon(x) = \{y \in E : r(x, y) < \epsilon\}$. The ball of radius $\epsilon > 0$ and center $x \in E$.
(7) *buc*-convergence, *buc*-approximable, *buc*-closure, closed and dense, and *buc*-lim, See Definition A.6.
(8) $C(E)$. The space of continuous functions on E.
(9) $C_b(E) = C(E) \cap B(E)$.
(10) $C(E, \overline{R})$. The collection of functions that are continuous as mappings from E into $\overline{R}$ with the natural topology on $\overline{R}$.
(11) $C_c(E)$. For E locally compact, the functions that are continuous and have compact support.
(12) $C^k(\mathcal{O})$, for $\mathcal{O} \subset R^d$ open and $k = 1, 2, \ldots, \infty$. The space of functions whose derivatives up to kth order are continuous in $\mathcal{O}$.
(13) $C_c^k(\mathcal{O}) = C^k(\mathcal{O}) \cap C_c(\mathcal{O})$.
(14) $C^{k,\alpha}(\mathcal{O})$, for $\mathcal{O} \subset R^d$ open, $k = 1, 2, \ldots$, and $\alpha \in (0, 1]$. The space of functions $f \in C^k(\mathcal{O})$ satisfying

$$\|f\|_{k,\alpha} = \sup_{0\le\beta\le k} \sup_{\mathcal{O}} \left|\frac{\partial^\beta f}{\partial x^\beta}\right| + \sup_{|\beta|=k} \left[\frac{\partial^\beta f}{\partial x^\beta}\right]_\alpha < \infty,$$

where

$$\left[f\right]_\alpha = \sup_{x\ne y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}, \quad \alpha \in (0, 1].$$

(15) $C_{loc}^{k,\alpha}(\mathcal{O})$. The space of functions $f \in C(\mathcal{O})$ such that $f|_D \in C^{k,\alpha}(D)$ for every bounded open subset $D \subset \mathcal{O}$.
(16) $C_E[0, \infty)$. The space of E-valued, continuous functions on $[0, \infty)$.
(17) $\widehat{C}(U)$, for U locally compact. The space of continuous functions vanishing at infinity.
(18) $\mathcal{D}(A) = \{f : \exists (f, g) \in A\}$. The domain of an operator A.
(19) $\mathcal{D}^+(A) = \{f \in \mathcal{D}(A), f > 0\}$.
(20) $\mathcal{D}^{++}(A) = \{f \in \mathcal{D}(A), \inf_{y\in E} f(y) > 0\}$.
(21) $D_E[0, \infty)$. The space of E-valued, cadlag (right continuous with left limit) functions on $[0, \infty)$ with the Skorohod topology, unless another topology is specified. (See Ethier and Kurtz [**36**], Chapter 3).

(22) $\mathcal{D}(\mathcal{O})$, for $\mathcal{O} \subset R^d$ open. The space $C_c^\infty(\mathcal{O})$ with the topology giving the space of Schwartz test functions. (See D.1).
(23) $\mathcal{D}'(\mathcal{O})$. The space of continuous linear functionals on $\mathcal{D}(\mathcal{O})$, that is, the space of Schwartz distributions.
(24) g^* (respectively g_*). The upper semicontinuous (resp. lower semicontinuous) regularization of a function g on a metric space (E, r). The definition is given by (6.2) (resp. (6.3)).
(25) $\limsup_{n\to\infty} G_n$ and $\liminf_{n\to\infty} G_n$ for a sequence of sets G_n. Definition 2.4 in Section 2.3.
(26) $H_\rho^k(R^d)$, for $\rho \in \mathcal{P}(R^d)$. A weighted Sobolev space. See Appendix D.5.
(27) $K(E) \subset C_b(E)$. The collection of nonnegative, bounded, continuous functions.
(28) $K_0(E) \subset K(E)$. The collection of strictly positive, bounded, continuous functions.
(29) $K_1(E) \subset K_0(E)$. The collection of bounded continuous functions satisfying $\inf_{x\in E} f(x) > 0$.
(30) $M(E)$. The R-valued, Borel measurable functions on E.
(31) $M^u(E)$. The space of $f \in M(E)$ that are bounded above.
(32) $M^l(E)$. The space of $f \in M(E)$ that are bounded below.
(33) $M(E, \overline{R})$. The space of Borel measurable functions with values in $\overline{R}$ and $f(x) \in R$ for at least one $x \in E$.
(34) $M_E[0,\infty)$. The space of E-valued measurable functions on $[0,\infty)$.
(35) $M^{d\times d}$. The space of $d \times d$ matrices.
(36) $M^u(E,\overline{R}) \subset M(E,\overline{R})$ (respectively, $C^u(E,\overline{R}) \subset C(E,\overline{R})$). The collection of Borel measurable (respectively continuous) functions that are bounded above (that is, $f \in M^u(E,\overline{R})$ implies $\sup_{x\in E} f(x) < \infty$).
(37) $M^l(E,\overline{R}) \subset M(E,\overline{R})$ (respectively, $C^l(E,\overline{R}) \subset C(E,\overline{R})$). The collection of Borel measurable (respectively continuous) functions that are bounded below.
(38) $\mathcal{M}(E)$. The space of (positive) Borel measures on E.
(39) $\mathcal{M}_f(E)$. The space of finite (positive) Borel measures on E.
(40) $\mathcal{M}_m(U)$, U a metric space. The collection of $\mu \in \mathcal{M}(U\times[0,\infty))$ satisfying $\mu(U \times [0,t]) = t$ for all $t \geq 0$.
(41) $\mathcal{M}_m^T(U)$ $(T > 0)$. The collection of $\mu \in \mathcal{M}(U \times [0,T])$ satisfying $\mu(U \times [0,t]) = t$ for all $0 \leq t \leq T$.
(42) $\mathcal{P}(E) \subset \mathcal{M}_f(E)$. The space of probability measures on E.
(43) $\overline{R} = [-\infty,\infty]$.
(44) $\mathcal{R}(A) = \{g : \exists (f,g) \in A\}$. The range of an operator A.
(45) $T\#\rho = \gamma$. $\gamma \in \mathcal{P}(E)$ is the push-forward (Definition D.1) of $\rho \in \mathcal{P}(E)$ by the map T.

Introduction

CHAPTER 1

Introduction

The theory of large deviations is concerned with the asymptotic estimation of probabilities of rare events. In its basic form, the theory considers the limit of normalizations of $\log P(A_n)$ for a sequence of events with asymptotically vanishing probability. To be precise, for a sequence of random variables $\{X_n\}$ with values in a metric space (S, d), we are interested in the *large deviation principle* as formulated by Varadhan [**119**]

Definition 1.1. [Large Deviation Principle] $\{X_n\}$ satisfies a *large deviation principle* (LDP) if there exists a lower semicontinuous function $I : S \to [0, \infty]$ such that for each open set A,

$$\liminf_{n\to\infty} \frac{1}{n} \log P\{X_n \in A\} \geq -\inf_{x\in A} I(x), \tag{1.1}$$

and for each closed set B,

$$\limsup_{n\to\infty} \frac{1}{n} \log P\{X_n \in B\} \leq -\inf_{x\in B} I(x). \tag{1.2}$$

I is called the *rate function* for the large deviation principle. A rate function is *good* if for each $a \in [0, \infty)$, $\{x : I(x) \leq a\}$ is compact.

Beginning with the work of Cramér [**16**] and including the fundamental work on large deviations for stochastic processes by Freidlin and Wentzell [**52**] and Donsker and Varadhan [**33**], much of the analysis has been based on change of measure techniques. In this approach, a reference measure is identified under which the events of interest have high probability, and the probability of the event under the original measure is estimated in terms of the Radon-Nikodym derivative relating the two measures.

More recently Puhalskii [**97**], O'Brien and Vervaat [**91**], de Acosta [**24**] and others have developed an approach to large deviations analogous to the Prohorov compactness approach to weak convergence of probability measures.

Definition 1.2. $\{X_n\}$ *converges in distribution* to X (that is, the distributions $P\{X_n \in \cdot\}$ *converge weakly* to $P\{X \in \cdot\}$) if and only if $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for each $f \in C_b(S)$.

The analogy between the two theories becomes much clearer if we recall the following equivalent formulation of convergence in distribution.

Proposition 1.3. *$\{X_n\}$ converges in distribution to X if and only if for each open set A,*

$$\liminf_{n\to\infty} P\{X_n \in A\} \geq P\{X \in A\},$$

or equivalently, for each closed set B,

$$\limsup_{n\to\infty} P\{X_n \in B\} \le P\{X \in B\}.$$

Our main theme is the development of this approach to large deviation theory as it applies to sequences of cadlag stochastic processes. The proof of weak convergence results typically involves verification of relative compactness or *tightness* for the sequence and the unique characterization of the possible limit distribution. The analogous approach to large deviations involves verification of *exponential tightness* (Definition 3.2) and unique characterization of the possible rate function.

In Part 1, we present results on exponential tightness and give Puhalskii's analogue of the Prohorov compactness theorem. We also give complete exponential tightness analogues of standard tightness criteria for cadlag stochastic processes and show that the rate function for a sequence of processes is determined by the rate functions for their finite dimensional distributions, just as the limit distribution for a weakly convergent sequence of processes is determined by the limits of the finite dimensional distributions.

In Part 2, we focus on Markov processes and give general large deviation results based on the convergence of corresponding nonlinear semigroups. Again we have an analogy with the use of the convergence of linear semigroups in proofs of weak convergence of Markov processes. The success of this approach depends heavily on the idea of a viscosity solution of a nonlinear equation. We also exploit control theoretic representations of the limiting semigroups to obtain useful representations of the rate function.

In Part 3, we demonstrate the effectiveness of these methods in a wide range of large deviation results for Markov processes, including classical Freidlin-Wentzell theory, random evolutions, and infinite dimensional diffusions.

1.1. Basic methodology

1.1.1. Analogy with the theory of weak convergence. In both the weak convergence and large deviation settings, proofs consist of first verifying a compactness condition and then showing that there is only one possible limit. For weak convergence, the first step is usually accomplished by verifying *tightness* which, by Prohorov's theorem, implies relative compactness. The corresponding condition for large deviations is *exponential tightness* (Definition 3.2). Puhalskii [**97**] (and in more general settings, O'Brien and Vervaat [**91**] and de Acosta [**24**]) has shown that exponential tightness implies the existence of a subsequence along which the large deviation principle holds. (See Theorem 3.7.) For stochastic processes, the second step of these arguments can be accomplished by verifying weak convergence (or the large deviation principle) for the finite dimensional distributions and showing that the limiting finite dimensional distributions (finite dimensional rate functions) determine a unique limiting distribution (rate function) for the process distributions.

We extend this analogous development in a number of ways. For the Skorohod topology on the space of cadlag sample paths in a complete, separable metric space, we give a complete exponential tightness analogue of the tightness conditions of Kurtz [**71**] and Aldous [**2**] (Theorem 4.1). We then extend the characterization of the rate function in terms of the finite-dimensional rate functions (Puhalskii [**97**], Theorem 4.5, and de Acosta [**2**], Lemma 3.2) to allow the rate function to be finite on paths with discontinuities (Section 4.7). Finally, we apply these results to Markov processes, using the asymptotic behavior of generators and semigroups associated with the Markov processes to verify exponential tightness for the processes (Section 4.5) and the large deviation principle for the finite dimensional distributions (Chapters 5, 6, 7). These arguments are again analogous to the use of convergence of generators and semigroups to verify tightness and convergence of finite dimensional distributions in the weak convergence setting.

1.1.2. Nonlinear semigroups and viscosity methods. Results of Varadhan and Bryc (Proposition 3.8) relate large deviations for sequences of random variables to the asymptotic behavior of functionals of the form $\frac{1}{n}\log E[e^{nf(X_n)}]$. For Markov processes,

$$V_n(t)f(x) = \frac{1}{n}\log E[e^{nf(X_n(t))}|X(0) = x] \tag{1.3}$$

defines a nonlinear semigroup, and large deviations for sequences of Markov processes can be studied using the asymptotic behavior of the corresponding sequence of nonlinear semigroups. Viscosity methods for nonlinear equations play a central role. Fleming and others (cf. [**44, 45, 40, 48, 46**]) have used this approach to prove large deviation results for X_n at single time points and exit times. We develop these ideas further, showing how convergence of the semigroups and their generators H_n can be used to obtain both exponential tightness and the large deviation principle for the finite dimensional distributions. These results then imply the pathwise large deviation principle.

1.1.3. Control theory. The limiting semigroup usually admits a variational form known as the *Nisio semigroup* in control theory. Connections between control problems and large deviation results were first made by Fleming [**44, 45**] and developed further by Sheu [**108, 109**]. Dupuis and Ellis [**35**] systematically develop these connections showing that, in many situations, one can represent a large class of functionals of the processes as the minimal cost functions of stochastic control problems and then verify convergence of the functionals to the minimal cost functions of limiting deterministic control problems. This convergence can then be used to obtain the desired large deviation result.

Variational representations for the sequence of functionals can be difficult to obtain, and in the present work, we first verify convergence of the semigroups by methods that only require convergence of the corresponding generators and conditions on the limiting generator. Working only with the sequence of generators frequently provides conditions that are easier to verify than conditions that give a convergent sequence of variational representations. Variational representations for the limit are still important, however, as they provide methods for obtaining simple representations of the large deviation rate function. In Chapter 8, we discuss methods for obtaining such representations. In particular, we formulate a set of conditions at the infinitesimal level, so that by verifying a variational representation for the limit generator, we obtain the variational structure of the semigroup that gives the large deviation rate function.

A generator convergence approach directly based on the sequence of control problems and Girsanov transformations was discussed in Feng [**41**] in a more restrictive setting. The present work avoids explicit use of a Girsanov transformation, and it greatly generalizes [**41**] making the approach much more applicable.

1.2. The basic setting for Markov processes

1.2.1. Notation. Throughout, (E, r) will be a complete, separable metric space, $M(E)$ will denote the space of real-valued, Borel measurable functions on E, $B(E) \subset M(E)$, the space of bounded, Borel measurable function, $C_b(E) \subset B(E)$, the space of bounded continuous functions, and $\mathcal{P}(E)$, the space of probability measures on E. We identify an operator A with its graph and, for example, write $A \subset C_b(E) \times C_b(E)$ if the domain $\mathcal{D}(A)$ and range $\mathcal{R}(A)$ are contained in $C_b(E)$. An operator can be multi-valued and nonlinear.

The space of E-valued, cadlag functions on $[0, \infty)$ with the Skorohod topology will be denoted by $D_E[0, \infty)$. Define $q(x, y) = 1 \wedge r(x, y)$, and note that q is a metric on E that is equivalent to r. The following metric gives the Skorohod topology . (See Ethier and Kurtz [**36**], Chapter 3). Let Λ' be the collection of strictly increasing functions mapping $[0, \infty)$ onto $[0, \infty)$ and satisfying

$$\gamma(\lambda) \equiv \sup_{0 \le s < t} |\log \frac{\lambda(t) - \lambda(s)}{t - s}| < \infty. \tag{1.4}$$

For $x, y \in D_E[0, \infty)$, define

$$d(x, y) = \inf_{\lambda \in \Lambda'} \{\gamma(\lambda) \vee \int_0^\infty e^{-u} \sup_{t \ge 0} q(x(\lambda(t) \wedge u), y(t \wedge u)) du\}. \tag{1.5}$$

1.2.2. Continuous-time Markov processes. An E-valued Markov process X is usually characterized in terms of its *generator*, a linear operator $A \subset B(E) \times B(E)$. One approach to the characterization of X is to require that all processes of the form

$$g(X(t)) - g(X(0)) - \int_0^t Ag(X(s)) ds \tag{1.6}$$

be martingales with respect to some filtration $\{\mathcal{F}_t\}$ independent of g. This requirement defines the (linear) martingale problem of Stroock and Varadhan [**116**], that is, a process X satisfying this condition is said to be a solution of the *martingale problem* for A. If $\nu \in \mathcal{P}(E)$ and $X(0)$ has distribution ν, X is a solution of the martingale problem for (A, ν). We say that *uniqueness* holds for the martingale problem for (A, ν) if any two solutions have the same finite dimensional distributions. The martingale problem for A is *well-posed* if for each $\nu \in \mathcal{P}(E)$, there exists a unique solution of the martingale problem for (A, ν).

If g is bounded away from zero, it is easy to see that (1.6) is a martingale if and only if

$$g(X(t)) \exp\{-\int_0^t \frac{Ag(X(s))}{g(X(s))} ds\}$$

is a martingale. Consequently, if we define $\mathcal{D}(\mathcal{H}) = \{f \in B(E) : e^f \in \mathcal{D}(A)\}$ and set

$$\mathcal{H}f = e^{-f} A e^f, \tag{1.7}$$

then we can define the *exponential martingale problem* by requiring

$$\exp\{f(X(t)) - f(X(0)) - \int_0^t \mathcal{H}f(X(s)) ds\} \tag{1.8}$$

to be a martingale with respect to a filtration independent of f. One can always assume that the domain $\mathcal{D}(A)$ contains constants, and hence if $g \in \mathcal{D}(A)$, then there exists a constant c such that $g+c \in \mathcal{D}(A)$ is positive and bounded away from zero. Setting $f = \log(g+c)$, (1.8) is a martingale if and only if (1.6) is a martingale. It follows that X is a solution of the linear martingale problem for A if and only if it is a solution of the exponential martingale problem for $\mathcal{H}$.

Weak convergence results for a sequence $\{X_n\}$ of Markov processes can be based on convergence of the corresponding operator semigroups $\{T_n(t)\}$, where

$$T_n(t)g(x) = E[g(X_n(t))|X_n(0) = x].$$

At least formally, T_n satisfies

$$\frac{d}{dt}T_n(t)g = A_n T_n(t)g, \quad T_n(0)g = g,$$

where A_n is the generator for X_n. An analogous approach to large deviation results is suggested by Fleming [**45**] using the nonlinear contraction (in the sup norm) semigroup V_n defined in (1.3). Again, at least formally, V_n should satisfy

$$\frac{d}{dt}V_n(t)f = \frac{1}{n}\mathcal{H}_n(nV_n(t)f),$$

which leads us to define

$$H_n f = \frac{1}{n}\mathcal{H}_n(nf) = \frac{1}{n}e^{-nf}A_n e^{nf}. \tag{1.9}$$

Note that

$$\exp\{nf(X_n(t)) - nf(X(0)) - \int_0^t nH_n f(X_n(s))ds\} \tag{1.10}$$

is a martingale.

1.2.3. Discrete-time Markov processes. Consideration of discrete-time Markov processes leads to slightly different formulations of the two martingale problems. Let $\{Y_k, k \geq 0\}$ be a time homogeneous Markov chain with state space E and transition operator

$$Tf(x) = E[f(Y_{k+1})|Y_k = x].$$

For $\epsilon > 0$, define $X(t) = Y_{[t/\epsilon]}$. Then, setting $\mathcal{F}_t^X = \sigma(X(s) : s \leq t)$, for $g \in B(E)$,

$$\begin{aligned} g(X(t)) - g(X(0)) - \sum_{k=0}^{[t/\epsilon]-1} (Tg(Y_k) - g(Y_k)) \\ = g(X(t)) - g(X(0) - \int_0^{[t/\epsilon]\epsilon} \epsilon^{-1}(T-I)g(X(s))ds \end{aligned}$$

is an $\{\mathcal{F}_t^X\}$-martingale, and for $f \in B(E)$,

$$\exp\{f(X(t)) - f(X(0)) - \int_0^{[t/\epsilon]\epsilon} \epsilon^{-1} \log e^{-f}Te^f(X(s))ds\}$$

is an $\{\mathcal{F}_t^X\}$-martingale. For a sequence, $X_n(t) = Y^n_{[t/\epsilon_n]}$, we define

$$A_n g = \epsilon_n^{-1}(T_n - I)g \tag{1.11}$$

and

$$H_n f = \frac{1}{n\epsilon_n}\log e^{-nf}T_n e^{nf} = \frac{1}{n\epsilon_n}\log(e^{-nf}(T_n - I)e^{nf} + 1), \tag{1.12}$$

so that

$$g(X_n(t)) - g(X_n(0)) - \int_0^{[t/\epsilon_n]\epsilon_n} A_n g(X_n(s))ds$$

and

$$\exp\{nf(X_n(t)) - nf(X(0)) - \int_0^{[t/\epsilon_n]\epsilon_n} nH_n f(X_n(s))ds\}$$

are martingales.

In Chapter 3, we give conditions for exponential tightness that are typically easy to check in the Markov process setting using the convergence (actually, only the boundedness) of H_n. In particular, $\sup_n \|H_n f\| < \infty$ for a sufficiently large class of f implies exponential tightness, at least if the state space is compact. In Section 4.7, we give conditions under which exponential tightness and the large deviation principle for finite dimensional distributions imply the large deviation principle in the Skorohod topology. In Chapters 5, 6 and 7, we give conditions under which convergence of the sequence of semigroups $\{V_n\}$ implies the large deviation principle for the finite dimensional distributions and conditions under which convergence of $\{H_n\}$ implies convergence of $\{V_n\}$. Consequently, convergence of $\{H_n\}$ is an essential ingredient in the results.

1.3. Related approaches

There are several approaches in the literature that are closely related to the methods we develop here. We have already mentioned the control theoretic methods developed in detail by Dupuis and Ellis [**35**] that play a key role in the rate function representation results of Chapter 8. Two other approaches share important characteristics with our methods.

1.3.1. Exponential martingale method of de Acosta.

de Acosta [**26, 27**] develops a general approach to large deviation results that, in the case of Markov processes, exploits the exponential martingale (1.10). Let λ be a measure on $[0, 1]$, and assume that $e^{af} \in \mathcal{D}(A_n)$ for all $a \in R$. Then the fact that (1.10) is a martingale implies that

$$\exp\{\int_0^1 f(X_n(s))\lambda(ds) - \lambda[0,1]f(X_n(0)) - \int_0^1 \mathcal{H}_n e^{\lambda(s,1]f}(X_n(s))ds\}$$

has expectation 1. Setting

$$\Phi_n(x, \lambda, f) = \lambda[0,1]f(x(0)) + \int_0^1 \mathcal{H}_n e^{\lambda(s,1]f}(x(s))ds,$$

convergence of H_n implies (at least formally) that

$$\frac{1}{n}\Phi_n(x, n\lambda, f) \to \Phi(x, \lambda, f) = \lambda[0,1]f(x(0)) + \int_0^1 He^{\lambda(s,1]f}(x(s))ds.$$

Comparing this convergence to the conditions of Theorems 2.1 and 3.1 of [**26**] indicates a close relationship between the methods developed here and those developed by de Acosta, at least at the formal computational level.

1.3.2. The maxingale method of Puhalskii. Assume that the semigroups V_n converge to V. Then V is nonlinear in the sense that, in general,

$$V(t)(af + bg) \neq aV(t)f + bV(t)g.$$

However, defining

$$a \oplus b = \max\{a, b\}, \quad a \odot b = a + b,$$

and using $(\oplus, \odot)$ as operations on R in place of the usual $(+, \times)$, V becomes "linear"

$$V(t)((a \odot f) \oplus (b \odot g)) = (a \odot V(t)f) \oplus (b \odot V(t)g),$$

and defining

$$a \oplus_n b = \frac{1}{n} \log(e^{na} + e^{nb}), \quad a \odot_n b = a + b,$$

V_n is linear under $(\oplus_n, \odot_n)$:

$$V_n(t)((a \odot_n f) \oplus_n (b \odot_n g)) = (a \odot_n V_n(t)f) \oplus_n (b \odot_n V_n(t)g).$$

Furthermore,

$$\lim_{n\to\infty} a \oplus_n b = a \oplus b, \quad a \odot_n b = a \odot b.$$

The change of algebra (semi-ring, to be precise) on R produces a linear structure for V (or the V_n). Results analogous to those of linear analysis, such as the Riesz representation theorem, hold and can be used to study the properties of V. The counterpart of a measure is an *idempotent measure* . (See Puhalskii [**100**].) Taking this view point and mimicking the martingale problem approach in weak convergence theory for stochastic processes, Puhalskii [**99, 100**] defines a *maxingale problem* and develops large deviation theory for semimartingales.

The main result of [**100**], Theorem 5.4.1, can be stated roughly as follows: let $\{X_n\}$ be a sequence of Markov processes and H_n be the operators so that

$$M^f_{s,n}(t) = e^{nf(X_n(t)) - nf(X_n(s)) - \int_s^t H_n f(X_n(r))dr}$$

is a mean one martingale for each $f \in \mathcal{D}(H_n)$, each $s \geq 0$ and $t \geq s$, as in (1.10). Assuming that $H = \lim_{n\to\infty} H_n$ in an appropriate sense, under additional conditions on H and the convergence, $\{X_n\}$ is exponentially tight and along any subsequence of $\{X_n\}$, there exists a further subsequence such that the large deviation principle holds with some rate function $I(x)$. Then $\Pi(A) = \sup_{x\in A} e^{-I(x)}$ defines an idempotent probability measure on the trajectory space of the processes. (See Chapter 1 of [**100**]. Note that an idempotent measure is not a measure in the usual sense.) Π then solves a *maxingale problem* for H:

$$e^{f(x(t)) - f(x(0)) - \int_0^t Hf(x(s))ds}$$

is a maxingale under $\Pi(dx)$ for every $f \in \mathcal{D}(H)$ in the sense that

(1.13)
$$\frac{\sup_{\{x:x(r)=z(r),0\leq r\leq s\}} \exp\{f(x(t)) - f(x(s)) - \int_s^t Hf(x(r))dr - I(x(\cdot))\}}{\sup_{\{x:x(r)=z(r),0\leq r\leq s\}} \exp\{-I(x(\cdot))\}} = 1,$$

for every $0 \leq s \leq t$ and every trajectory z. If the maxingale problem for H has a unique idempotent probability measure Π_0 as solution, then $\Pi_0 = \Pi$ and the large deviation principle holds for $\{X_n\}$ with rate function I.

1.4. Examples

As explained in Section 1.2, an essential ingredient of our result is the convergence of $\{H_n\}$. As the following examples show, at least formally, the basic calculations for verifying this convergence may be quite simple. The calculations given here are frequently heuristic; however, we give rigorous results for these and other examples in Part 3.

EXAMPLE 1.4. [**Freidlin-Wentzell Theory - I**] The best-known example of large deviations for Markov processes comes from the work of Freidlin and Wentzell on diffusions with small diffusion coefficient. (See Chapters 3 and 4 in [**52**].) Consider a sequence of d-dimensional diffusion processes with X_n satisfying the Itô equation

$$X_n(t) = x + \frac{1}{\sqrt{n}} \int_0^t \sigma(X_n(s))dW(s) + \int_0^t b(X_n(s))ds.$$

Let $a(x) = \sigma(x) \cdot \sigma^T(x)$. Then the (linear) generator is

$$A_n g(x) = \frac{1}{2n} \sum_{ij} a_{ij}(x) \partial_i \partial_j g(x) + \sum_i b_i(x) \partial_i g(x),$$

where we can take $\mathcal{D}(A_n)$ to be the collection of functions of the form $c + f$ where $c \in R$ and $f \in C_c^2(R^d)$, the space of twice continuously differentiable functions with compact support in R^d.

$$\mathcal{H}_n f(x) = \frac{1}{2n} \sum_{ij} a_{ij}(x) \partial_i \partial_j f(x) + \frac{1}{2n} \sum_{ij} a_{ij}(x) \partial_i f(x) \partial_j f(x) + \sum_i b_i(x) \partial_i f(x)$$

and

$$H_n f(x) = \frac{1}{2n} \sum_{ij} a_{ij}(x) \partial_i \partial_j f(x) + \frac{1}{2} \sum_{ij} a_{ij}(x) \partial_i f(x) \partial_j f(x) + \sum_i b_i(x) \partial_i f(x).$$

Consequently, $Hf = \lim_{n \to \infty} H_n f$ is

$$Hf(x) = \frac{1}{2} (\nabla f(x))^T \cdot a(x) \cdot \nabla f(x) + b(x) \cdot \nabla f(x). \tag{1.14}$$

We can identify the rate function I in a simple form by finding a variational representation of H. First, we introduce a pair of functions on $R^d \times R^d$:

$$H(x,p) \equiv \frac{1}{2} p^T \cdot a(x) \cdot p + b(x) \cdot p = \frac{1}{2} |\sigma^T(x) \cdot p|^2 + b(x) \cdot p, \tag{1.15}$$

$$L(x,q) \equiv \sup_{p \in R^d} \{p \cdot q - H(x,p)\}.$$

$H(x,p)$ is convex in p, and H and L are dual Fenchel-Legendre transforms. In particular,

$$H(x,p) = \sup_{q \in R^d} \{p \cdot q - L(x,q)\}.$$

Therefore

$$Hf(x) = H(x, \nabla f(x)) = \mathbf{H} f(x) \equiv \sup_{u \in R^d} \{Af(x,u) - L(x,u)\},$$

where $Af(x,u) = u \nabla f(x)$, for $f \in C_c^2(R^d)$. This implies that H is a Nisio semigroup generator studied in control theory (e.g. Chapter 8). Applying Corollary 8.28

or Corollary 8.29, we arrive at the rate function for $\{X_n\}$ represented in a variational form:

$$I(x) = I_0(x_0) + \inf_{\{u:(x,u)\in\mathcal{J}\}} \int_0^\infty L(x(s), u(s))ds,$$

where I_0 is the rate function for $\{X_n(0)\}$ and $\mathcal{J}$ is the collection of solutions of the functional equation

$$f(x(t)) - f(x(0)) = \int_0^t Af(x(s), u(s))ds, \quad \forall f \in C_c^2(R^d).$$

In the present setting, this equation reduces to $\dot{x}(t) = u(t)$, so

$$(1.16) \quad I(x) = \begin{cases} I_0(x(0)) + \int_0^\infty L(x(s), \dot{x}(s))ds & \text{if } x \text{ is absolutely continuous} \\ \infty & \text{otherwise} \end{cases}$$

One can use other variational representations of the operator H and arrive at different expressions for the rate function I. For example, if we choose

$$Af(x,u) = u \cdot (\sigma^T(x)\nabla f(x)) + b(x) \cdot \nabla f(x), \quad f \in C_0^2(R^d),$$

$$L(x,u) = \frac{1}{2}|u|^2,$$

and define

$$\mathbf{H}f(x) = \sup_{u\in R^d}\{Af(x,u) - L(x,u)\},$$

then $H = \mathbf{H}$ and the rate function can be expressed as

$$I(x) = I_0(x(0)) + \frac{1}{2}\inf\{\int_0^\infty |u(s)|^2 ds : u \in L^2[0,\infty)$$
$$\text{and } x(t) = x(0) + \int_0^t b(x(s))ds + \int_0^t \sigma(x(s))u(s)ds\}.$$

Example 1.5. [**Freidlin-Wentzell theory – II**]Wentzell [**128, 129**] has studied the jump process analogue of the small diffusion problem. Again, let $E = R^d$, and consider, for example,

$$(1.17) \qquad A_n g(x) = n\int_{R^d}(g(x + \frac{1}{n}z) - g(x))\eta(x, dz),$$

where for each $x \in R^d$, $\eta(x,\cdot)$ is a measure on $\mathcal{B}(R^d)$ and for each $B \in \mathcal{B}(R^d)$, $\eta(\cdot, B)$ is a Borel-measurable function. (For the moment, ignore the problem of defining an appropriate domain.) Then

$$\mathcal{H}_n f(x) = n\int_{R^d}(e^{f(x+\frac{1}{n}z)-f(x)} - 1)\eta(x, dz)$$

and

$$H_n f(x) = \int_{R^d}(e^{n(f(x+\frac{1}{n}z)-f(x))} - 1)\eta(x, dz).$$

Assuming $\int_E e^{\alpha\cdot z}\eta(x, dz) < \infty$, for each $\alpha \in R^d$, $Hf = \lim_{n\to\infty} H_n f$ is given by

$$Hf(x) = \int_{R^d}(e^{\nabla f(x)\cdot z} - 1)\eta(x, dz).$$

More generally, one can consider generators of the form

$$A_n g(x) = n\int_{R^d}(g(x+\frac{1}{n}z)-g(x)-\frac{1}{n}z\cdot\nabla g(x))\eta(x,dz) + \frac{1}{2n}\sum_{ij}a_{ij}(x)\partial_i\partial_j g(x)+b(x)\cdot\nabla g(x).$$

Then

$$\begin{aligned}\mathcal{H}_n f(x) &= n\int_{R^d}(e^{f(x+\frac{1}{n}z)-f(x)}-1-\frac{1}{n}z\cdot\nabla f(x))\eta(x,dz)+b(x)\cdot\nabla f(x)\\ &\quad +\frac{1}{2n}\sum_{ij}a_{ij}(x)\partial_i f(x)\partial_j f(x+\frac{1}{2n}\sum_{ij}a_{ij}(x)\partial_i\partial_j f(x))\end{aligned}$$

and

$$\begin{aligned}H_n f(x) &= \int_{R^d}(e^{n(f(x+\frac{1}{n}z)-f(x))}-1-z\cdot\nabla f(x))\eta(x,dz)+b(x)\cdot\nabla f(x)\\ &\quad +\frac{1}{2}\sum_{ij}a_{ij}(x)\partial_i f(x)\partial_j f(x)+\frac{1}{2n}\sum_{ij}a_{ij}(x)\partial_i\partial_j f(x).\end{aligned}$$

Now, assuming $\int_{R^d}(e^{\alpha\cdot z}-1-\alpha\cdot z)\eta(x,dz)<\infty$, for each $\alpha\in R^d$, $Hf = \lim_{n\to\infty}H_n f$ is given by

$$Hf(x)=\int_{R^d}(e^{\nabla f(x)\cdot z}-1-z\cdot\nabla f(x))\eta(x,dz)+\frac{1}{2}\sum_{ij}a_{ij}(x)\partial_i f(x)\partial_j f(x)+b(x)\cdot\nabla f(x).$$

A variational representation of H can be constructed as in Example 1.4. Define

$$H(x,p)=\int_{R^d}(e^{p\cdot z}-1-z\cdot p)\eta(x,dz)+\frac{1}{2}|\sigma^T(x)p|^2+b(x)\cdot p$$

and

$$L(x,q)=\sup_{p\in R^d}\{p\cdot q-H(x,p)\}.$$

Then

$$Hf(x)=H(x,\nabla f(x))=\mathbf{H}f(x)\equiv\sup_{u\in R^d}\{Af(x,u)-L(x,u)\},$$

where

$$Af(x,u)=u\cdot\nabla f(x),\quad f\in C_c^2(R^d),$$

and the rate function is given by (1.16).

EXAMPLE 1.6. [**Partial sum processes**] Functional versions of the classical large deviation results of Cramér [**16**] have been considered by a number of authors, including Borovkov [**11**] and Mogulskii [**87**]. Let $\xi_1,\xi_2,\ldots$ be independent and identically distributed R^d-valued random variables with distribution ν. Define

$$X_n(t)=\beta_n^{-1}\sum_{k=1}^{[t/\epsilon_n]}\xi_k.$$

Then, as in (1.11),

$$A_n f(x)=\epsilon_n^{-1}(\int_{R^d}f(x+\beta_n^{-1}z)\nu(dz)-f(x))$$

and

$$H_n f(x) = \frac{1}{n\epsilon_n} \log \int_{R^d} e^{n(f(x+\beta_n^{-1}z)-f(x))}\nu(dz).$$

If $\int_{R^d} e^{\alpha|z|}\nu(dz) < \infty$, for each $\alpha > 0$, $\epsilon_n = n^{-1}$, and $\beta_n = n$, then $H_n f \to Hf$ given by

$$Hf(x) = \log \int_{R^d} e^{\nabla f(x)\cdot z}\nu(dz).$$

If $\int_{R^d} z\nu(dz) = 0$, $\int_{R^d} e^{\alpha|z|}\nu(dz) < \infty$ for some $\alpha > 0$, $\epsilon_n\beta_n^2 = n$, and $\epsilon_n n \to 0$, then $H_n f \to Hf$ given by

$$Hf(x) = \frac{1}{2}\int_{R^d} (\nabla f(x)\cdot z)^2\nu(dz).$$

Once again, a variational representation of H and the rate function I can be constructed using Fenchel-Legendre transforms. Define

$$H(p) = \frac{1}{2}\int_{R^d} (z\cdot p)^2\nu(dz)$$

and

$$L(q) = \sup_{p\in R^d}\{p\cdot q - H(p)\}.$$

Then

$$I(x) = \begin{cases} \int_0^\infty L(\dot{x}(s))ds & \text{if } x \text{ is absolutely continuous} \\ \infty & \text{otherwise.} \end{cases}$$

EXAMPLE 1.7. [**Levy processes**] If η in Example 1.5 does not depend on x, then the process is a Lévy process, that is, a process with stationary, independent increments. Lynch and Sethuraman [**82**] and Mogulskii [**88**] consider the real-valued case. de Acosta [**25**] considers Banach space-valued processes.

In de Acosta's setting, let E be a Banach space and E^* its dual. Recall that for $f \in C^1(E)$, the gradient $\nabla f(x) = \nabla_x f(x)$ is the unique element in E^* satisfying

$$f(x+z) - f(x) = \langle z, \nabla f(x)\rangle + o(|z|), \quad \forall z \in E.$$

Let $F \subset E^*$ separate points, and let $\mathcal{D}(A)$ be the collection of functions of the form

$$f(x) = g(\langle x,\xi_1\rangle,\ldots,\langle x,\xi_m\rangle),$$

where $\xi_1,\ldots,\xi_m \in F$ and $g \in C_c^2(R^m)$. Then,

$$\nabla f(x) = \sum_{k=1}^m \partial_k g(\langle x,\xi_1\rangle,\ldots,\langle x,\xi_m\rangle)\xi_k \in E^*.$$

Assuming $\int_E 1 \wedge \langle z, \xi\rangle^2 \eta(dz) < \infty$, for every $\xi \in F$, we have

$$\begin{aligned}A_n g(x) &= n\int_E \Big(g(\langle x + \frac{1}{n}z, \xi_1\rangle, \ldots, \langle x + \frac{1}{n}z, \xi_m\rangle) - g(\langle x, \xi_1\rangle, \ldots, \langle x, \xi_m\rangle) \\ &\qquad -\frac{1}{n}I_{\{\|z\|\leq 1\}}\sum_{k=1}^m \langle z, \xi_k\rangle \partial_k g(\langle x, \xi_1\rangle, \ldots, \langle x, \xi_m\rangle)\Big)\eta(dz) \\ &\qquad +\sum_{k=1}^m \langle b, \xi_k\rangle \partial_k g(\langle x, \xi_1\rangle, \ldots, \langle x, \xi_m\rangle), \\ &= n\int_E \Big(g(\langle x + \frac{1}{n}z, \xi_1\rangle, \ldots, \langle x + \frac{1}{n}z, \xi_m\rangle) - g(\langle x, \xi_1\rangle, \ldots, \langle x, \xi_m\rangle) \\ &\qquad -\frac{1}{n}I_{\{\|z\|\leq 1\}}\langle z, \nabla f(x)\rangle\Big)\eta(dz) + \langle b, \nabla f(x)\rangle,\end{aligned}$$

$$\mathcal{H}_n f(x) = n\int_E (e^{f(x+\frac{1}{n}z)-f(x)} - 1 - \frac{1}{n}I_{\{\|z\|\leq 1\}}\langle z, \nabla f(x)\rangle)\eta(dz) + \langle b, \nabla f(x)\rangle,$$

and

$$H_n f(x) = \int_E (e^{n(f(x+\frac{1}{n}z)-f(x))} - 1 - \frac{1}{n}I_{\{\|z\|\leq 1\}}\langle z, \nabla f(x)\rangle)\eta(dz) + \langle b, \nabla f(x)\rangle.$$

Now, assuming $\int_E (e^{\langle \alpha, z\rangle} - 1 - I_{\{|z|\leq 1\}}\langle \alpha, z\rangle)\eta(dz) < \infty$, for each $\alpha \in E^*$, $Hf = \lim_{n\to\infty} H_n f$ is given by

$$Hf(x) = \int_E (e^{\langle z, \nabla f(x)\rangle} - 1 - I_{\{\|z\|\leq 1\}}\langle z, \nabla f(x)\rangle)\eta(dz) + \langle b, \nabla f(x)\rangle.$$

Define

$$H(p) = \int_E (e^{\langle z, p\rangle} - 1 - I_{|z|\leq 1}\langle z, p\rangle)\eta(dz) + \langle b, z\rangle, \quad p \in E^*,$$

and

$$L(q) = \sup_{p\in E^*}\{\langle q, p\rangle - H(p)\}, \quad q \in E.$$

H is a convex function and L is its Fenchel-Legendre transform, and hence,

$$H(p) = \sup_{q\in E}\{\langle q, p\rangle - L(q)\}.$$

Consequently

$$Hf(x) = \mathbf{H}f(x) \equiv \sup_{u\in E}\{Af(x, u) - L(u)\},$$

where

$$Af(x, u) = \langle u, \nabla f(x)\rangle,$$

and the rate function is again given by

$$I(x) = \begin{cases} I_0(x(0)) + \int_0^\infty L(\dot{x}(s))ds & \text{if } x \text{ is absolutely continuous} \\ \infty & \text{otherwise.}\end{cases}$$

EXAMPLE 1.8. [**Random evolutions - I**] Let B be the generator for a Markov process Y with state space E_0, and let $Y_n(t) = Y(nt)$. Let X_n satisfy

$$\dot{X}_n(t) = F(X_n(t), Y_n(t)),$$

where $F : R^d \times E_0 \to R^d$. If Y is ergodic with stationary distribution π, then

$$\lim_{n\to\infty} X_n = X,$$

where X satisfies $\dot{X} = \overline{F}(X)$, with $\overline{F}(x) = \int F(x,y)\pi(dy)$. Averaging results of this type date back at least to Hasminskii [**66**]. Corresponding large deviation theorems have been considered by Freidlin [**49, 50**]. (See Freidlin and Wentzell [**52**], Chapter 7.)

For simplicity, assume B is a bounded operator

$$Bg(y) = \lambda(y) \int (g(z) - g(y))\mu(y, dz). \tag{1.18}$$

Then

$$A_n g(x,y) = F(x,y) \cdot \nabla_x g(x,y) + n\lambda(y) \int (g(x,z) - g(x,y))\mu(y,dz),$$

$$\mathcal{H}_n f(x,y) = F(x,y) \cdot \nabla_x f(x,y) + n\lambda(y) \int (e^{f(x,z)-f(x,y)} - 1)\mu(y,dz)$$

and

$$H_n f(x,y) = F(x,y) \cdot \nabla_x f(x,y) + \lambda(y) \int (e^{n(f(x,z)-f(x,y))} - 1)\mu(y,dz).$$

Let $f_n(x,y) = f(x) + \frac{1}{n}h(x,y)$. Then

$$\lim_{n\to\infty} H_n f_n(x,y) = F(x,y) \cdot \nabla_x f(x) + \lambda(y) \int (e^{h(x,z)-h(x,y)} - 1)\mu(y,dz).$$

Consequently, H is multivalued.

One approach to dealing with this limit is to select h so that the limit is independent of y, that is, to find functions $h(x,y)$ and $g(x)$ such that

$$F(x,y) \cdot \nabla_x f(x) + \lambda(y) \int (e^{h(x,z)-h(x,y)} - 1)\mu(y,dz) = g(x). \tag{1.19}$$

Multiply both sides of the equation by $e^{h(x,y)}$, and fix x. One needs to solve an equation of the form

$$\alpha(y)\overline{h}(y) + \lambda(y) \int (\overline{h}(z) - \overline{h}(y))\mu(y,dz) = \gamma \overline{h}(y), \tag{1.20}$$

where $\overline{h}$ plays the role of $e^{h(x,y)}$ and hence must be positive. Of course, (1.20) is just an eigenvalue problem for the operator

$$Q = (\alpha I + B). \tag{1.21}$$

If E_0 is finite, that is, Y is a finite Markov chain, and B is irreducible, then the Perron-Froebenius theorem implies the existence of a unique (up to multiplication by a constant) strictly positive $\overline{h}$ which is the eigenfunction corresponding to the largest eigenvalue for Q. Note that in (1.19), it is only the ratio $\overline{h}(z)/\overline{h}(y)$ that is relevant, so the arbitrary constant cancels.

For linear semigroups, this approach to convergence of generators was introduced in Kurtz [**71**] and has been developed extensively by Papanicolaou and Varadhan [**93**] and Kushner (for example, [**79**]). For nonlinear semigroups, Kurtz [**72**] and Evans [**37**] have used similar arguments.

If we use $\gamma(\alpha(\cdot))$ to denote the largest eigenvalue for the Q in (1.21), then we arrive at a single valued limit operator

$$Hf(x) = \gamma(F(x,\cdot) \cdot \nabla f(x)).$$

Define

$$H(x,p) = \gamma(F(x,\cdot) \cdot p)$$

and

$$L(x,q) = \sup_{p\in R^d}\{p\cdot q - H(x,p)\},$$

for $x, p, q \in R^d$. There are various probabilistic representations for the principle eigenvalue γ. (See Appendix B.) It follows from these representations that H is convex in p, so

$$Hf(x) = H(x, \nabla f(x)) = \mathbf{H}f(x) \equiv \sup_{u\in R^d}\{Af(x,u) - L(x,u)\},$$

where $Af(x,u) = u\cdot\nabla f(x)$, for $f \in C_c^1(R^d)$.

By the results in Chapter 8, (1.16) gives a representation of the rate function. Another representation of the rate function can be of special interest. Donsker and Varadhan [**32**] generalize the classical Rayleigh-Ritz formula and obtain the following variational representation of the principal eigenvalue for Q when B is a general operator satisfying the maximum principle:

$$\gamma(\alpha) = \sup_{\mu\in\mathcal{P}(E_0)}\{\int_{E_0}\alpha d\mu - I_B(\mu)\},$$

where

$$I_B(\mu) = -\inf_{f\in\mathcal{D}(B),\inf_y f(y)>0}\int_{E_0}\frac{Bf}{f}d\mu. \tag{1.22}$$

Therefore, setting

$$Af(x,\mu) = \int_{E_0}F(x,y)d\mu(dy)\cdot\nabla f(x),$$

H satisfies

$$Hf(x) = \mathbf{H}f(x) \equiv \sup_{\mu\in\mathcal{P}(E_0)}\{Af(x,\mu) - I_B(\mu)\}.$$

By the methods in Chapter 8, the rate function can also be represented as

$$I(x) = I_0(x(0)) + \inf\{\int_0^\infty I_B(\mu(s))ds : \dot{x}(t) = \int_{E_0}F(x(t),\cdot)d\mu(t)\}, \mu(t)\in\mathcal{P}(E_0), t\geq 0\},$$

if x is absolutely continuous, and $I(x) = \infty$ otherwise.

Example 1.9. [**Random evolutions - II**] Let B be as in (1.18), let Y be the corresponding Markov process, and let $Y_n(t) = Y(n^3t)$. Let X_n satisfy

$$\dot{X}_n(t) = nF(X_n(t), Y_n(t)).$$

Suppose that Y is ergodic with stationary distribution π and that

$$\int_{E_0}F(x,y)\pi(dy) = 0. \tag{1.23}$$

Then

$$A_ng(x,y) = nF(x,y)\cdot\nabla_x g(x,y) + n^3\lambda(y)\int(g(x,z) - g(x,y))\mu(y,dz),$$

$$\mathcal{H}_nf(x,y) = nF(x,y)\cdot\nabla_x f(x,y) + n^3\lambda(y)\int(e^{f(x,z)-f(x,y)} - 1)\mu(y,dz),$$

and

$$H_n f(x,y) = nF(x,y)\cdot\nabla_x f(x,y) + n^2\lambda(y)\int (e^{n(f(x,z)-f(x,y))}-1)\mu(y,dz).$$

Let $f_n(x,y) = f(x) + \frac{1}{n^2}h_1(x,y) + \frac{1}{n^3}h_2(x,y)$, and assume that

$$Bh_1(x,\cdot)(y) = -F(x,y)\cdot\nabla_x f(x).$$

If E_0 is finite and B is irreducible, then (1.23) ensures the existence of h_1. Note that $h_1(x,y)$ will be of the form $\alpha(x,y)\cdot\nabla_x f(x)$, where α is a vector function satisfying

$$B\alpha(x,\cdot)(y) = -F(x,y).$$

It follows that

$$\begin{aligned}\lim_{n\to\infty} H_n f_n(x,y) &= \frac{\lambda(y)}{2}\int_{E_0}(h_1(x,z)-h_1(x,y))^2\mu(y,dz)\\ &\qquad +\lambda(y)\int_{E_0}(h_2(x,z)-h_2(x,y))\mu(y,dz),\end{aligned}$$

and again we have a multivalued limit.

As before, one approach is to select h_2 so that the limit is independent of y. If E_0 is finite, (1.23) ensures that h_2 exists. In any case, if h_2 exists, the limit will be

$$Hf(x) = \int_{E_0}\frac{\lambda(y)}{2}\int_{E_0}((\alpha(x,z)-\alpha(x,y))\cdot\nabla_x f(x))^2\mu(y,dz)\pi(dy).$$

Note that H is of the same form as H in Example 1.4, indicating that this "random evolution" behaves like the "small diffusion" process. If $Y_n(t) = Y(n^3 t)$ is replaced by $Y_n(t) = Y(n^2 t)$, then X_n will converge in distribution to a diffusion. (See, for example, Ethier and Kurtz [**36**], Chapter 12.)

The results in Example 1.4 give variational representations of the rate function.

EXAMPLE 1.10. [**Periodic diffusions**] Baldi [**6**] has considered large deviations for models of the following type. Let σ be a periodic $d\times d$-matrix-valued function (for each $1\le i\le d$, there is a period $p_i>0$ such that $\sigma(y)=\sigma(y+p_ie_i)$ for all $y\in R^d$), and let X_n satisfy the Itô equation

$$dX_n(t) = \frac{1}{\sqrt{n}}\sigma(\alpha_n X_n(t))dW(t),$$

where $\alpha_n>0$ and $\lim_{n\to\infty} n^{-1}\alpha_n=\infty$. Let $a=\sigma\sigma^T$. Then

$$A_n f(x) = \frac{1}{2n}\sum_{ij} a_{ij}(\alpha_n x)\frac{\partial^2}{\partial x_i\partial x_j}f(x),$$

$$\mathcal{H}_n f(x) = \frac{1}{2n}\sum_{ij} a_{ij}(\alpha_n x)\partial_i\partial_j f(x) + \frac{1}{2n}\sum_{ij} a_{ij}(\alpha_n x)\partial_i f(x)\partial_j f(x),$$

and

$$H_n f(x) = \frac{1}{2n}\sum_{ij} a_{ij}(\alpha_n x)\partial_i\partial_j f(x) + \frac{1}{2}\sum_{ij} a_{ij}(\alpha_n x)\partial_i f(x)\partial_j f(x).$$

Let $f_n(x) = f(x)+\epsilon_n h(x,\alpha_n x)$, where $\epsilon_n = n\alpha_n^{-2}$. Noting that, by assumption, $\epsilon_n\alpha_n = n\alpha_n^{-1}\to 0$, if we select h with the same periods in y as the a_{ij} so that

$$\frac{1}{2}\sum_{ij} a_{ij}(y)\left(\frac{\partial^2}{\partial y_i\partial y_j}h(x,y) + \partial_i f(x)\partial_j f(x)\right) = g(x), \tag{1.24}$$

for some g independent of y, then

$$\lim_{n\to\infty} H_n f_n(x,y) = g(x).$$

For each x, the desired $h(x,\cdot)$ is the solution of the linear partial differential equation

$$\frac{1}{2}\sum_{ij} a_{ij}(y)\frac{\partial^2}{\partial y_i \partial y_j} h(x,y) = g(x) - \frac{1}{2}\sum_{i,j} a_{ij}(y)\partial_i f(x)\partial_j f(x) \tag{1.25}$$

on $[0,p_1]\times\cdots\times[0,p_d]$ with periodic boundary conditions extended periodically to all of R^d. For h to exist, we must have

$$g(x) = \frac{1}{2}\sum_{ij} \overline{a}_{ij}\partial_i f(x)\partial_j f(x), \tag{1.26}$$

where $\overline{a}_{ij}$ is the average of a_{ij} with respect to the stationary distribution for the diffusion on $[0,p_1]\times\cdots\times[0,p_d]$ whose generator is

$$A_0 f(y) = \frac{1}{2}\sum_{i,j} a_{ij}(y)\frac{\partial^2}{\partial y_i \partial y_j} f(y),$$

with periodic boundary conditions. To see that (1.26) must hold, simply integrate both sides of (1.25) by the stationary distribution. In particular,

$$h(x,y) = \frac{1}{2}\sum_{ij} h_{ij}(y)\partial_i f(x)\partial_j f(x),$$

where h_{ij} satisfies

$$A_0 h_{ij}(y) = \overline{a}_{ij} - a_{ij}(y).$$

If A_0 is uniformly elliptic and f is C^3, then, with sufficient regularity on the a_{ij}, $h \in C^2$ will exist.

By (1.26), the limit Hf is a special case of Hf in Example 1.4. Consequently, we can identify the rate function as in that example.

Example 1.11. [**Donsker-Varadhan theory for occupation measures**] Large deviation theory for occupation measures (see Donsker and Varadhan [**33**]) is closely related to Example 1.8. Let Y be a Markov process with generator B and state space E_0, and for $n = 1,2,\ldots$, define

$$\Gamma_n(C,t) = \frac{1}{n}\int_0^{nt} I_C(Y(s))ds.$$

Then Z_n defined by $Z_n(t) = (Y(nt),\Gamma_n(\cdot,t))$ is a Markov process with state space $E = E_0\times\mathcal{M}_F(E_0)$, where $\mathcal{M}_F(E_0)$ is the space of finite measures on E_0. Let $h\in C_b(E_0\times R^m)$ be differentiable in the real variables and satisfy $h(\cdot,x)\in\mathcal{D}(B)$ for $x\in R^m$. For $\alpha_i\in C_b(E_0)$, $i=1,\ldots,m$, let

$$g(y,z) = h(y,\langle\alpha_1,z\rangle,\ldots,\langle\alpha_m,z\rangle),\quad (y,z)\in E_0\times\mathcal{M}_F(E_0),$$

where $\langle\alpha,z\rangle = \int_{E_0}\alpha dz$. The generator for Z_n satisfies

$$A_n g(y,z) = nBh(y,\langle\alpha_1,z\rangle,\ldots,\langle\alpha_m,z\rangle) + \sum_{i=1}^m \alpha_i(y)\partial_i h(y,\langle\alpha_1,z\rangle,\ldots,\langle\alpha_m,z\rangle),$$

where $\partial_i h(y, x_1, \ldots, x_m) = \frac{\partial}{\partial x_i} h(y, x_1, \ldots, x_m)$. For definiteness, let B be a diffusion operator

$$Bg(y) = \frac{1}{2}\sum_{i,j} a_{ij}(y)\frac{\partial^2}{\partial y_i \partial y_j} g(y) + \sum_i b_i(y)\frac{\partial}{\partial y_i} g(y).$$

Then for $f(y,z) = h(y, \langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle)$,

$$\mathcal{H}_n f(y,z) = A_n f(y,z) + \frac{n}{2}\sum_{i,j} a_{ij}(y)\frac{\partial}{\partial y_i} f(y,z)\frac{\partial}{\partial y_j} f(y,z)$$

and

$$H_n f(y,z) = A_n f(y,z) + \frac{n^2}{2}\sum_{i,j} a_{ij}(y)\frac{\partial}{\partial y_i} f(y,z)\frac{\partial}{\partial y_j} f(y,z).$$

If we let $f(z) = h(\langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle)$ and

$$f_n(y,z) = h(\langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle) + \frac{1}{n} h_0(y, \langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle),$$

then

$$\begin{aligned} &\lim_{n\to\infty} H_n f_n(y,z) \\ &= \sum_{i=1}^m \alpha_i(y)\partial_i h(\langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle) + Bh_0(y, \langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle) \\ &\quad + \frac{1}{2}\sum_{i,j} a_{ij}(y)\frac{\partial}{\partial y_i} h_0(y, \langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle)\frac{\partial}{\partial y_j} h_0(y, \langle \alpha_1, z\rangle, \ldots, \langle \alpha_m, z\rangle). \end{aligned} \tag{1.27}$$

As in Example 1.8, we have a multivalued limit, and the rate function representation is also similar to the representation in Example 1.8.

EXAMPLE 1.12. [**Stochastic reaction-diffusion equations**] We consider a stochastic reaction-diffusion equation on a rescaled lattice, which is a Ginzburg-Landau model of non-conservative type. Let $F \in C^2(R)$, $\sup_r |F''(r)| < \infty$, and $\mathcal{O} = [0,1)^d$ with periodic boundary. We discretize $\mathcal{O}$ into

$$\Lambda_m \equiv \{0, \frac{1}{m}, \ldots, \frac{k}{m}, \ldots, 1 - \frac{1}{m}\}^d \subset \mathcal{O},$$

also with periodic boundary conditions (that is, $0 = 1$). Let $m = m(n)$ depend on n so that $\lim_{n\to\infty} m(n) = \infty$. When there is no possibility of confusion, we simply write m in place of $m(n)$. For $\rho \in R^{|\Lambda_m|}$, define

$$\begin{aligned} \nabla_m^i \rho(x) &= \frac{m}{2}(\rho(x + \frac{1}{m}e_k) - \rho(x - \frac{1}{m}e_k)); \\ \nabla_m \rho(x) &= (\nabla_m^1 \rho(x), \ldots, \nabla_m^d \rho(x)) \\ \Delta_m \rho(x) &= \nabla_m \cdot \nabla_m \rho(x) \\ &= (\frac{m}{2})^2 \sum_{k=1}^d \Big(\rho(x + \frac{2}{m}e_k) - 2\rho(x) + \rho(x - \frac{2}{m}e_k)\Big), \end{aligned} \tag{1.28}$$

and for a vector-valued function,

$$\xi(x) = \Big(\xi_1(x), \ldots, \xi_d(x)\Big),$$

define

$$\text{div}_m\xi(x) = \nabla_m \cdot \xi(x) = \sum_{k=1}^{d} \nabla_m^k \rho_k(x).$$

We consider a finite system of stochastic differential equations

$$dY_n(t,x) = (\Delta_m Y_n)(t,x)dt - F'(Y_n(t,x))dt + \frac{m^{d/2}}{\sqrt{n}} dB(t,x), \tag{1.29}$$

where $x \in \Lambda_m$ and $\{B(t,x) : x \in \Lambda_m\}$ are independent standard Brownian motions.

Let $E_n = R^{|\Lambda_m|}$. For $p \in C^\infty(\mathcal{O})$ and $\rho \in E_n$, define

$$\langle \rho, p \rangle_m \equiv \sum_{x \in \Lambda_m} p(x)\rho(x)m^{-d},$$

and

$$\langle \rho, p \rangle \equiv \int_{\mathcal{O}} p(x)\rho(x)dx, \qquad \rho \in L^2(\mathcal{O}).$$

For $p_1, \ldots, p_l \in C^\infty(\mathcal{O})$ and $\varphi \in C_c^\infty(R^l)$, let $\vec{\gamma} = (\gamma_1, \ldots, \gamma_l)$ and

$$\begin{aligned} f_n(\rho) &= \varphi(\langle \rho, \gamma_1 \rangle_m, \ldots, \langle \rho, \gamma_l \rangle_m) = \varphi(\langle \rho, \vec{\gamma} \rangle_m), \\ f(\rho) &= \varphi(\langle \rho, \vec{\gamma} \rangle), \end{aligned} \tag{1.30}$$

and

$$\begin{aligned} A_n f_n(\rho) &= \sum_{i=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\sum_{x \in \Lambda_m} \gamma_i(x)(\Delta_m \rho(x) - F'(\rho(x)))m^{-d} \Big) \\ &\quad + \frac{1}{2n} \sum_{i,j=1}^{l} \partial_{ij}^2 \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\sum_{x \in \Lambda_m} \gamma_i(x)\gamma_j(x)m^{-d} \Big). \end{aligned}$$

Then, by Itô's formula, Y_n is an E_n-valued solution to the martingale problem for A_n and

$$\begin{aligned} H_n f_n(\rho) &\equiv \frac{1}{n} e^{-nf} A_n e^{nf}(\rho) \\ &= \sum_{i=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\sum_{x \in \Lambda_m} \gamma_i(x)(\Delta_m \rho(x) - F'(\rho(x)))m^{-d} \Big) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle_m) \partial_j \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\sum_{x \in \Lambda_m} \gamma_i(x)\gamma_j(x)m^{-d} \Big) \\ &\quad + \frac{1}{2n} \sum_{i,j=1}^{l} \partial_{ij}^2 \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\sum_{x \in \Lambda_m} \gamma_i(x)\gamma_j(x)m^{-d} \Big). \end{aligned}$$

Define $\eta_n : E_n \to E = L^2(\mathcal{O})$ by

$$\eta_n(\rho_n)(x_1, \ldots, x_d) \equiv \sum_{i_1, \ldots, i_d = 0}^{m-1} \rho_n(\frac{i_1}{m}, \ldots, \frac{i_d}{m}) \prod_{j=1}^{d} \mathbf{1}_{[i_j/m, (i_j+1)/m)}(x_j). \tag{1.31}$$

For $\rho_n \in E_n$ and $\rho \in E$ satisfying $\|\eta_n(\rho_n) - \rho\|_{L^2(\mathcal{O})} \to 0$, we have

$$
\begin{aligned}
(1.32) \quad H_n f_n(\rho_n) &\to Hf(\rho) \\
&\equiv \sum_{i=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle)\Big(\langle \rho, \Delta\gamma_i\rangle - \langle F'(\rho), \gamma_i\rangle\Big) \\
&\quad + \frac{1}{2}\sum_{i,j=1}^{l} \partial_i\varphi(\langle\rho,\vec{\gamma}\rangle)\partial_j\varphi(\langle\rho,\vec{\gamma}\rangle)\int_{\mathcal{O}} \gamma_i(x)\gamma_j(x)dx \\
&= \langle \rho, \Delta\frac{\delta f}{\delta\rho}\rangle - \langle F'(\rho), \frac{\delta f}{\delta \rho}\rangle + \frac{1}{2}\|\frac{\delta f}{\delta\rho}\|^2_{L^2(\mathcal{O})}, \\
&= H(\rho, \frac{\delta f}{\delta\rho}),
\end{aligned}
$$

where

$$H(\rho, p) = \langle \Delta\rho + F^{'}(\rho), p\rangle + \frac{1}{2}\|p\|^2_{L^2(\mathcal{O})}, \quad p \in C^\infty(\mathcal{O}),$$

and

$$\frac{\delta f}{\delta \rho} = \sum_{i=1}^{l} \partial_i\varphi(\langle\rho,\vec{\gamma}\rangle)\gamma_i. \tag{1.33}$$

By analogy with Example 1.4, we expect $\{\eta_n(Y_n)\}$ to satisfy a large deviation principle in $L^2(\mathcal{O})$. Let

$$L(\rho, u) = \sup_{p \in C^\infty(\mathcal{O})} \{\langle u, p\rangle - H(\rho, p)\}.$$

The rate function is analogous to that of (1.16) and should be given by

$$I(\rho) = I_0(\rho(0)) + \int_0^\infty L(\rho(s), \dot{\rho}(s))ds.$$

A rigorous proof of this statement is given in Chapter 13, Theorem 13.7.

The equation for Y_n is a discretized version of the stochastic partial differential equation given in weak form by

$$
\begin{aligned}
\int_{\mathcal{O}} \varphi(x)X_n(t,x)dx &= \int_{\mathcal{O}}\varphi(x)X_n(0,x)dx + \int_{\mathcal{O}\times[0,t]} \Delta\varphi(x)X_n(s,x)dxds \\
&\quad - \int_{\mathcal{O}\times[0,t]} \varphi(x)F'(X_n(s,x))dxds \\
&\quad + \frac{1}{\sqrt{n}}\int_{\mathcal{O}\times[0,t]}\varphi(x)W(dx\times ds),
\end{aligned}
\tag{1.34}
$$

where $W(dx \times ds)$ is a space-time Gaussian white noise. Unfortunately, (1.34) does not have a function-valued solution when $d \geq 2$. To study the large deviation problem in $L^2(\mathcal{O})$, as described above, even for the discretized equation (1.29), requires that m go to infinity slowly enough with n. (See Theorem 13.7.)

Equation (1.34) is also known as a stochastically perturbed Allen-Cahn equation. It has been used in material science to study time evolution of material density when the total mass is not conserved. More generally, in statistical physics, when F is non-convex, the equation is used to model phase transition (Spohn [**113**]). When $d = 1$, (1.34) is well-defined and Sowers [**112**] gives the large deviation principle.

EXAMPLE 1.13. [**Stochastic Cahn-Hilliard equations**] Spohn [**113**] introduces a type of Ginzburg-Landau model with a conserved quantity motivated by the study of phase transitions. Let $\mathcal{O}$ and F be as in Example 1.12. Formally, the equation can be written

$$\partial_t X_n(t,x) = \nabla \cdot \Big(\nabla(-\Delta X_n(t,x) + F'(X_n(t,x))) + \frac{1}{\sqrt{n}} \partial_t \partial_x \mathbf{W}(t,x) \Big) \tag{1.35}$$

or in a more easily interpretable, but still formal, weak form

$$\begin{aligned} \int_{\mathcal{O}} \varphi(x) X_n(t,x) dx &= \int_{\mathcal{O}} \varphi(x) X_n(0,x) dx - \int_{\mathcal{O}\times[0,t]} \Delta^2 \varphi(x) X_n(s,x) dx ds \\ &\quad + \int_{\mathcal{O}\times[0,t]} \Delta\varphi(x) F'(X_n(s,x)) dx ds \\ &\quad + \frac{1}{\sqrt{n}} \sum_{k=1}^{d} \int_{\mathcal{O}\times[0,t]} \partial_{x_k} \varphi(x) W_k(dx \times ds), \end{aligned} \tag{1.36}$$

where the W_k are independent Gaussian white noises and $\mathbf{W} = (W_1, \ldots, W_d)$. Without the stochastic term, (1.35) is known as the Cahn-Hilliard equation and originally arose in material science where the solution gives the evolution of a material density.

In these applications, (1.35) is derived as a phenomenological model using rough and heuristic arguments. Frequently, only asymptotic properties are of practical interest. Indeed, as in the previous example, (1.35) does not have a function-valued solution. Therefore, for the large deviation problem here, we consider a discretized version.

Let ∇_m and Δ_m be defined according to (1.28), and consider the system

$$dY_n(t;x) = \mathrm{div}_m \Big(\nabla_m(-\Delta_m Y_n(t,x) + F'(Y_n(t,x)))dt + \frac{m^{d/2}}{\sqrt{n}} d\mathbf{B}(t,x) \Big), \quad x \in \Lambda_m, \tag{1.37}$$

where

$$\mathbf{B}(t,x) = \Big(B_1(t,x), \cdots, B_d(t,x) \Big), \quad x \in \Lambda_m,$$

with the $B_j(\cdot, x)$ independent, R^d-valued standard Brownian motions. Note that the solution satisfies

$$\sum_{x\in\Lambda_m} Y_n(t,x) = \sum_{x\in\Lambda_m} Y_n(0,x), \quad t \geq 0,$$

so we may as well choose the state spaces

$$E_n \equiv \{\rho \in R^{|\Lambda_m|} : \sum_{x\in\Lambda_m} \rho(x) = 0\}, \quad E \equiv \{\rho \in L^2(\mathcal{O}) : \int_{\mathcal{O}} \rho(x) dx = 0\}. \tag{1.38}$$

The map η_n in (1.31) embeds E_n into E. For functions of the form (1.30), by Itô's formula,

$$\begin{aligned} A_n f(\rho) &= \sum_{i=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\langle \gamma_i, \Delta_m(-\Delta_m \rho + F'(\rho)) \rangle_m \Big) \\ &\quad + \frac{1}{2n} \sum_{i,j=1}^{l} \partial_{ij}^2 \varphi(\langle \rho, \vec{\gamma} \rangle_m) \sum_{z\in\Lambda_m} \nabla_m \gamma_i(z) \cdot \nabla_m \gamma_j(z) m^{-d}. \end{aligned} \tag{1.39}$$

Therefore

$$\begin{aligned} H_n f(\rho) &\equiv \frac{1}{n} e^{-nf} A_n e^{nf}(\rho) \\ &= \sum_{i=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle_m) \Big(\langle \gamma_i, \Delta_m(-\Delta_n \rho + F'(\rho)) \rangle_m \Big) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle_m) \partial_j \varphi(\langle \rho, \vec{\gamma} \rangle_m) \big(\sum_{x \in \Lambda_m} \nabla_m \gamma_i(x) \cdot \nabla_m \gamma_j(x) m^{-d} \big) \\ &\quad + \frac{1}{2n} \sum_{i,j=1}^{l} \partial^2_{ij} \varphi(\langle \rho, \vec{\gamma} \rangle_m) \sum_{z \in \Lambda_m} \Big(\nabla_m \gamma_i(z) \cdot \nabla_m \gamma_j(z) m^{-d} \Big). \end{aligned}$$

As in (1.33), let

$$\frac{\delta f}{\delta \rho} = \sum_{i=1}^{l} \partial_i \varphi(\langle \rho, \vec{\gamma} \rangle) \gamma_i.$$

Then for every $\rho_n \in E_n$ satisfying $\lim_{n\to\infty} \|\eta_n(\rho_n) - \rho\|^2_{L^2(\mathcal{O})} = 0$,

$$\lim_{n\to\infty} H_n f(\rho_n) = Hf(\rho) = H(\rho, \frac{\delta f}{\delta \rho}), \tag{1.40}$$

where

$$H(\rho, p) = \langle \Delta(-\Delta\rho + F'(\rho)), p \rangle + \frac{1}{2} \|\nabla p\|^2_{L^2(\mathcal{O})}, \quad \rho \in L^2(\mathcal{O}), p \in C^\infty(\mathcal{O}). \tag{1.41}$$

Following the method of rate function identification in the finite dimensional Example 1.4, we expect that the rate function can be identified using the same arguments as in Example 1.12. The proof of this result is given in Chapter 13. See Theorem 13.13.

Bertini, Landam, and Olla [**10**] proved a large deviation principle for a variant of (1.37) in which the discrete version of $\nabla \mathbf{W}$ is replaced by a discrete version of ΔW, where W is a scalar Brownian sheet. The current form of (1.37) was treated using a slightly different technique in Feng [**42**].

EXAMPLE 1.14. [**Weakly interacting stochastic particles**] For $n = 1, 2, \ldots$, we consider the finite system of stochastic differential equations

$$dX_{n,i}(t) = -\nabla \Psi\Big(X_{n,i}(t)\Big) - \frac{1}{n} \sum_{j=1}^{n} \nabla \Phi\Big(X_{n,i}(t) - X_{n,j}(t)\Big) dt + dW_i(t), \tag{1.42}$$

where $W_i(t)$, $i = 1, 2, \ldots, n$, are independent, R^d-valued, standard Brownian motions. Define the empirical measure-valued processes

$$\rho_n(t) = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_{n,k}(t)}. \tag{1.43}$$

Then under appropriate growth conditions on Ψ and Φ, a law of large numbers (the McKean-Vlasov limit) holds. Specifically, ρ_n converges to a probability measure-valued solution of

$$\frac{\partial}{\partial t} \rho = \frac{1}{2} \Delta \rho + \nabla \cdot (\rho \nabla \Psi) + \nabla \cdot (\rho \nabla (\rho * \Phi)), \tag{1.44}$$

where $\rho * \Phi(x) = \int_{R^d} \Phi(x-y)\rho(dy)$. We consider the corresponding large deviation principle.

When $\Phi(z) = \theta|z|^2/2$, $\theta > 0$, (1.42) is known as the ferromagnetic Curie-Weiss model

$$dX_{n,k}(t) = -\nabla\Psi(X_{n,k}(t)) - \frac{\theta}{n}\sum_{j=1}^{n}(X_{n,k}(t) - X_{n,j}(t))dt + dW_k(dt). \tag{1.45}$$

The large deviation problem for this model, as well as for a larger class of models, has been considered by Dawson and Gärtner in a series of publications. See [**23**] and [**22**] and the references therein. In connection with gas kinetics in statistical mechanics, the system (1.42) with a general semi-convex Φ gives a microscopic statistical foundation for certain deterministic models of the evolution of a spatially homogeneous gas in a granular media. In the infinite system limit, $n \to \infty$, the law of large numbers gives a nonlinear partial differential equation modeling the evolution. See Section 9.6 of Villani [**127**] for a discussion of these models.

Our methods applied to this problem seem more involved than those of Dawson and Gärtner [**23**]; however, our interest is not only in the large deviation problem, but also in the well-posedness of an associated nonlinear equation (1.49) it induces. This nonlinear equation is a special case of a Hamilton-Jacobi equation in the space of probability measures and is closely related to the Hamilton-Jacobi equations in Banach spaces studied by Crandall and Lions [**21, 20**] and Tataru [**117**]. The previous work, however, requires the state space to be a subset of a Banach space satisfying the Radon-Nikodym property. We note that even though the space of probability measures with Lebesgue density is a bounded subset of $L^1(R^d)$ (i.e. $\int_{R^d} \rho(x)dx = 1$), $L^1(R^d)$ does not satisfy the Radon-Nikodym property. (See, for example, [**31**], page 31.)

Assume that $\Phi, \Psi \in C^2(R^d)$ and that $|\nabla\Phi(z)|$ has sub-quadratic growth as $|z| \to \infty$. More conditions on Φ, Ψ will be imposed later. To simplify, we only consider the case where Φ is an even function (i.e. $\Phi(x) = \Phi(-x)$).

Let d be the (order) 2-Wasserstein metric on E (i.e. (D.14) with $p = 2$). Then (E, d) is a complete separable metric space and $\rho_n \to \rho_0$ in (E, d) if and only if $\rho_n \Rightarrow \rho_0$, in the sense of weak convergence of probability measures, and $\int_{R^d} |x|^2 d\rho_n \to \int_{R^d} |x|^2 d\rho_0$ (Lemma D.17). Define

$$E_n \equiv \{\rho(dx) \equiv \frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}(dx), x_k \in R^d, k = 1, 2, \ldots\} \subset E.$$

Let r_n be the restriction of d to E_n. For fixed n, the corresponding topology is just the topology of weak convergence.

For each n, $\rho_n(\cdot)$ in (1.43) is a probability-measure-valued Markov process. We calculate its generator next. Define

$$B(\rho)p(x) = \frac{1}{2}\Delta p(x) - \nabla(\Psi + \Phi * \rho)(x) \cdot \nabla p(x), \quad \forall p \in C^2(R^d), \rho \in \mathcal{P}(R^d).$$

Let $p_1, \ldots, p_l \in C_c^2(R^d)$, and define

$$\mathbf{p} = (p_1, \ldots, p_l), \qquad \langle \mathbf{p}, \rho \rangle = (\int_{R^d} p_1 d\rho, \ldots, \int_{R^d} p_l d\rho).$$

Then for

$$f(\rho) = \varphi(\langle \mathbf{p}, \rho \rangle), \qquad \varphi \in C^2(R^l), \tag{1.46}$$

$$A_n f(\rho) = \sum_{i=1}^{l} \partial_i \varphi(\langle \mathbf{p}, \rho \rangle) \langle \rho, B(\rho) p_i \rangle + \frac{1}{2n} \sum_{i,j=1}^{l} \partial_{ij}^2 \varphi(\langle \mathbf{p}, \rho \rangle) \langle \rho, (\nabla p_i)^T \nabla p_j \rangle$$

Therefore

$$\begin{aligned} H_n f(\rho) &\equiv \frac{1}{n} e^{-nf} A_n e^{nf}(\rho) \\ &= \sum_{i=1}^{l} \partial_i \varphi(\langle \mathbf{p}, \rho \rangle) \langle \rho, B(\rho) p_i \rangle \\ &\qquad + \frac{1}{2} \sum_{i,j=1}^{l} \partial_i \varphi(\langle \mathbf{p}, \rho \rangle) \partial_j \varphi(\langle \mathbf{p}, \rho \rangle) \langle \rho, (\nabla p_i)^T \nabla p_j \rangle \\ &\qquad\quad + \frac{1}{2n} \sum_{i,j=1}^{l} \partial_{ij}^2 \varphi(\langle \mathbf{p}, \rho \rangle) \langle \rho, (\nabla p_i)^T \nabla p_j \rangle. \end{aligned} \tag{1.47}$$

Setting

$$\frac{\delta f}{\delta \rho} = \sum_{i=1}^{l} \partial_i \varphi(\langle \mathbf{p}, \rho \rangle) p_i,$$

for $\rho_n \in E_n, \rho \in E$ satisfying $d(\rho_n, \rho) \to 0$,

$$\lim_{n \to \infty} H_n f(\rho_n) = H f(\rho) = H(\rho, \frac{\delta f}{\delta \rho}),$$

where

$$H(\rho, p) = \langle \rho, B(\rho) p \rangle + \frac{1}{2} \int |\nabla p|^2 d\rho = \langle B^*(\rho)\rho, p \rangle + \frac{1}{2} \int |\nabla p|^2 d\rho, \tag{1.48}$$

and for each $\gamma \in E$, $B^*(\gamma) : \rho \in \mathcal{P}(R^d) \to \mathcal{D}'(R^d)$ is defined by

$$B^*(\gamma)\rho = \frac{1}{2}\Delta \rho + \nabla \cdot (\rho \nabla (\Psi + \gamma * \Phi)).$$

Large deviation results for this example are discussed in Chapter 13 (Theorem 13.37). The key here is to establish a uniqueness result for weak solutions of

$$(I - \alpha H) f = h, \quad \alpha > 0, \tag{1.49}$$

for sufficiently many $h \in C_b(E)$. This is achieved in two steps in Example 9.35 and in Section 13.3.4. See Theorem 13.32.

1.5. An outline of the major obstacles

From the discussion in the previous sections, we see that one of our main assumptions should be the convergence of $\{H_n\}$ to a limit operator H in the sense that for each $f \in \mathcal{D}(H)$, there exist $f_n \in \mathcal{D}(H_n)$ such that

$$f_n \to f, \quad H_n f_n \to H f, \tag{1.50}$$

where the type of convergence may depend on the particular problem. To rigorously formulate a large deviation program using generator convergence and nonlinear semigroup theory requires some work, but it is straightforward (Chapter 5) provided we assume that the limit operator H satisfies a *range condition*. Specifically, we need existence of solutions of

$$(I - \alpha H)f = h, \tag{1.51}$$

for all sufficiently small $\alpha > 0$ and a large enough collection of functions h. Unfortunately, there are very few examples for which this condition can actually be verified in the classical sense, that is, with solutions f satisfying the differentiability requirements assumed in the examples above. We overcome this difficulty by using the weak (viscosity) solution theory introduced by Crandall and Lions [**19**]. The range condition is replaced by the requirement that a *comparison principle* (Definitions 6.4 and 7.3) holds for (1.51). Assuming that

$$(I - \alpha H_n)f_n = h_n, \tag{1.52}$$

for a sequence h_n converging to h, we adapt a technique of Barles and Perthame [**7**] [**8**] to prove that the comparison principle implies the convergence of f_n. This technique does not require *a priori* estimates on the regularity of f.

There is a well developed, powerful theory for verifying the comparison principle, at least when the state space is a subset of R^d or a compact metric space (Chapter 6). Indeed, the approach has been applied in ad hoc ways to various large deviation examples (e.g. [**48**], [**40**]) for quite some time. When the state space is not compact, however, major obstacles arise. First, we need a good definition of viscosity solution in this context. The definition should be weak enough to allow extension of the Barles-Perthame arguments to noncompact metric spaces, yet strong enough so that the comparison principle can be proved for interesting examples. Our choice here is a definition (Definition 7.1) that preserves a nonlinear version of the maximum principle for H. This definition is different from existing definitions [**21**], [**117**], [**20**] in the literature on partial differential equations, where structural information about the equation is usually built into the definition. Second, the Barles and Perthame limiting procedure may not always work when E is not compact. This procedure is compatible with a variant of uniform convergence over compacts, but without other information, the usual estimates needed to derive convergence of solutions of (1.52) require the convergence in (1.50) to be *uniform* over the whole space. In general, uniformity will not hold for a sufficiently large collection of f. The uniformity requirement can be relaxed if we can verify the *exponential compact containment condition* (Condition 2.8). In applications, this condition can frequently be verified by a stochastic Lyapunov function technique (Section 4.6).

The general versions of our large deviation theorems are given in Chapter 7. A short version is summarized in Chapter 2. Further generalizations of these results are also given. For instance, we discuss situations where certain functionals of the Markov processes satisfy a large deviation principle, while the full processes may not. We also discuss the use of a general notion of convergence of test functions and operators, to handle processes with multiple scales.

In most of the examples, the most difficult technical argument comes in verifying the *comparison principle* (Definitions 6.4 and 7.3). Verification is an analytic issue and often gives the impression of being rather involved and disconnected from the probabilistic large deviation problems. This disconnect is not always the case. Specific probabilistic structures can give insight into the solution of the analytic problem. For instance, the large deviation theory for the interacting particle system in Example 1.14 and stochastic equations in Examples 1.12 and 1.13) leads to Hamilton-Jacobi operators, (1.48), (1.32), (1.41), that describe the evolution of optimally controlled partial differential equations. The corresponding comparison theories in the literature ([**21**], [**117**], [**20**]) are technical and limited. Moreover, these theories do not apply for Example 1.14. We solve the problem by devising simple, new comparison techniques exploiting the special structure in these problems. As a result, we not only arrive at large deviation principles, but also obtain simpler proofs of analytic viscosity solution results.

CHAPTER 2

An overview

The purpose of this chapter is to provide a rough "road map" for reading the general theory that follows. We collect versions of the main results regarding large deviations for Markov processes and describe some motivation for our approach. Proofs of these results will be given in later chapters. The results collected here may not be the sharpest or the fullest versions in the paper. Indeed, for most of the results, further generalizations are given later. For example, the majority of the material in Chapters 3 and 4 applies to general processes which may be non-Markovian; Chapter 5 contains a pure semigroup formulation (as opposed to the viscosity solution approach given here) of the large deviation results; Chapters 6 and 7 have a number of new results regarding viscosity solutions and their convergence, many of which may have application in areas besides large deviations.

A typical application requires the following steps:

(1) *Verify convergence of the sequence of operators H_n and derive the limit operator H.* Many examples of this convergence are given in the Introduction. In general, convergence will be in the extended limit or graph sense. (See Definition A.12.) In some examples, the limit is described in terms of a pair of operators, $(H_\dagger, H_\ddagger)$, where, roughly, $H_\dagger$ is the $\limsup$ of $\{H_n\}$ and $H_\ddagger$ is the $\liminf$.

(2) *Verify exponential tightness.* The convergence of H_n (in fact, boundedness of $\{H_n f_n\}$ for an appropriate collection of sequences $\{f_n\}$) typically gives exponential tightness, provided one can verify the exponential compact containment condition (4.10). Lyapunov function arguments can be used to verify this condition (Section 4.6). Alternatively, one can avoid verifying the compact containment condition by compactifying the state space and verifying the large deviation principle in the compactified space. If the rate function is infinite for every path that hits a point added in the compactification, then the large deviation principle holds for the original space (Theorem 4.11).

(3) *Verify the comparison principle for the limiting operator H (or the pair $(H_\dagger, H_\ddagger)$).* The comparison principle asserts a strong form of uniqueness for the equation $(I - \alpha H)f = h$. If the comparison principle holds for a sufficiently large class of functions h, then one can conclude that the nonlinear semigroups $\{V_n\}$ converge. Convergence of the semigroups implies the large deviation principle for the finite dimensional distributions of the sequence of processes, which by exponential tightness, then gives the large deviation principle in $D_E[0, \infty)$. The rate function is characterized by the limiting semigroup. Chapter 9 discusses concrete techniques for verifying the comparison principle in a number of situations. These techniques are general, yet practical enough to cover every example in the Introduction.

(4) *Construct a variational representation for H.* Typically we can identify the limiting semigroup as the Nisio semigroup for a control problem. The control problem then gives an alternative and more explicit representation of the rate function (Chapter 8).

In Chapters 10 to 13, we apply these methods to the examples in the Introduction.

2.1. Basic setup

Let (E_n, r_n), $n = 1, 2, \ldots$ and (E, r) be complete separable metric spaces.

a) (Continuous time Markov processes) Let $A_n \subset B(E_n) \times B(E_n)$, and assume the following: For each $n = 1, 2, \ldots$, existence and uniqueness hold for the $D_{E_n}[0, \infty)$-martingale problem for (A_n, μ) for each initial distribution $\mu \in \mathcal{P}(E_n)$. Letting $P^n_y \in \mathcal{P}(D_{E_n}[0, \infty))$ denote the distribution of the solution of the martingale problem for A_n starting from $y \in E_n$, the mapping $y \mapsto P^n_y$ is Borel measurable taking the weak topology on $\mathcal{P}(D_{E_n}[0, \infty))$. For each n, Y_n is a solution of the martingale problem for A_n.

b) (Discrete time processes) Let $\{\widetilde{Y}_n(k), k = 0, 1, \cdots\}$ be a time homogeneous Markov chain with state space E_n and transition operator T_n on $B(E_n)$:

$$T_n f(y) = E[f(\widetilde{Y}_n(k+1))|\widetilde{Y}_n(k) = y], \quad f \in B(E_n).$$

Let $\epsilon_n > 0$, and define

$$Y_n(t) \equiv \widetilde{Y}_n([\frac{t}{\epsilon_n}]).$$

Then

$$E[f(Y_n(t))|Y_n(0) = y] = T_n^{[t/\epsilon_n]} f(y), \quad f \in B(E_n).$$

In either case, we suppose $\eta_n : E_n \to E$ is Borel measurable and assume $X_n \equiv \eta_n(Y_n) \in D_E[0, \infty)$. We are interested in establishing the large deviation principle for X_n. For $f \in M(E)$, we set $\eta_n f = f \circ \eta_n$.

REMARK 2.1. In many applications, it suffices to take $E_n = E$ and $\eta_n(y) = y$. More generally, however, E_n may be a discrete set that is asymptotically dense in E or a higher dimensional space in which the large deviation principle is only verified for certain components.

2.2. Compact state spaces

We first assume that E is compact which simplifies a number of technical issues. This assumption is not as restrictive as it may first appear. Many results in R^d can be obtained by first verifying the large deviation principle in the one-point compactification $E = R^d \cup \{\infty\}$.

The notion of a viscosity solution for a nonlinear equation is central to the results that follow.

DEFINITION 2.2 (Viscosity Solution). Let E be a compact metric space, and $H \subset C(E) \times B(E)$. Fix $h \in C(E)$ and $\alpha > 0$. Let $f \in B(E)$ and define $g = \alpha^{-1}(f - h)$, that is, $f - \alpha g = h$. Then

a) f is a *viscosity subsolution* of $(I - \alpha H)f = h$ if and only if f is upper semicontinuous and for each $(f_0, g_0) \in H$ such that $\sup_x (f - f_0)(x) = \|f - f_0\|$, there exists $x_0 \in E$ satisfying

$$(f - f_0)(x_0) = \|f - f_0\| \tag{2.1}$$

and

$$\alpha^{-1}(f(x_0) - h(x_0)) = g(x_0) \le (g_0)^*(x_0). \tag{2.2}$$

b) f is a *viscosity supersolution* of $(I - \alpha H)f = h$ if and only if f is lower semicontinuous and for each $(f_0, g_0) \in H$ such that $\sup_x (f_0 - f)(x) = \|f_0 - f\|$, there exists $x_0 \in E$ satisfying

$$(f_0 - f)(x_0) = \|f - f_0\| \tag{2.3}$$

and

$$\alpha^{-1}(f(x_0) - h(x_0)) = g(x_0) \ge (g_0)_*(x_0). \tag{2.4}$$

The *comparison principle* holds for h if for each subsolution $\overline{f}$ and each supersolution $\underline{f}$, $\overline{f} \le \underline{f}$.

Theorem 2.3. *Suppose E is compact.*

Step 1: *(Convergence of H_n) In the continuous time case, let*

$$H_n f \equiv \frac{1}{n} e^{-nf} A_n e^{nf}, \quad e^{nf} \in \mathcal{D}(A_n).$$

In the discrete time case, let

$$H_n f \equiv \frac{1}{n\epsilon_n} \log e^{-nf} T_n e^{nf} = \frac{1}{n\epsilon_n} \log(1 + e^{-nf}(T_n - I)e^{nf}).$$

Suppose $H \subset C(E) \times B(E)$ with $\mathcal{D}(H)$ dense in $C(E)$, and

$$H \subset ex\text{-}\lim_{n\to\infty} H_n$$

in the sense of Definition A.12. That is, for each $(f, g) \in H$, there exists $(f_n, g_n) \in H_n$ such that

$$\lim_{n\to\infty} \|\eta_n f - f_n\| + \|\eta_n g - g_n\| = 0.$$

Step 2: *(Exponential tightness) Under the convergence assumptions on $\{H_n\}$, exponential tightness of $\{X_n\}$ holds.*

Step 3: *(The comparison principle) Let $\alpha_0 > 0$, and assume that for each $0 < \alpha < \alpha_0$, there exists a subset $D_\alpha \subset C(E)$ such that D_α is dense in $C(E)$ and that, for each $h \in D_\alpha$, the comparison principle (Definition 6.4) holds for viscosity sub and super solutions of*

$$(I - \alpha H)f = h. \tag{2.5}$$

Then a viscosity solution $R_\alpha h$ exists for each $h \in D_\alpha$, and R_α extends continuously to all of $C(E)$.

In addition, suppose that $\{X_n(0)\}$ satisfies a large deviation principle with a good rate function I_0.

Then

a) *$\{X_n\}$ satisfies the large deviation principle with a good rate function I.*

b) *The limit*

$$V(t)h \equiv \lim_{m\to\infty} R^m_{t/m} h$$

exists and defines a strongly continuous operator semigroup on $C(E)$*, and*

$$I(x) = \sup_{\{t_1,\cdots,t_k\}\subset\Delta_x^c} \sup_{f_1,\cdots,f_k\in C_b(E)} \Big(I_0(x(0)) + \sum_{i=1}^k f_i(x(t_i)) - V(t_i - t_{i-1})f_i(x(t_{i-1}))\Big), \tag{2.6}$$

for $x \in D_E[0,\infty)$*, where* Δ_x *denotes the set of discontinuities of* x*.*

This theorem is proved in Chapter 6.

Consider the classical Freidlin and Wentzell diffusion problem (Example 1.4). Then

$$A_n g(x) = \frac{1}{2n}\sum_{ij} a_{ij}(x)\partial_i\partial_j g(x) + \sum_i b_i(x)\partial_i g(x),$$

and

$$H_n f(x) = \frac{1}{2n}\sum_{ij} a_{ij}(x)\partial_i\partial_j f(x) + \frac{1}{2}\sum_{ij} a_{ij}(x)\partial_i f(x)\partial_j f(x) + \sum_i b_i(x)\partial_i f(x).$$

Assuming a and b are bounded on bounded subsets, the convergence

$$Hf = \lim_{n\to\infty} H_n f$$

for

$$Hf(x) = \frac{1}{2}(\nabla f(x))^T \cdot a(x)\cdot\nabla f(x) + b(x)\cdot\nabla f(x)$$

is immediate for all $f \in D_d = \{f \in C(E) : f|_{R^d} - f(\infty) \in C_c^2(R^d)\}$. (Taking $E = R^d \cup \{\infty\}$, $H_n f(\infty) = Hf(\infty) = 0$.) Exponential tightness then holds in $D_E[0,\infty)$. Assuming the large deviation principle holds for $\{X_n(0)\}$, the conditions of Theorem 2.3 are then satisfied for this example provided we can verify the comparison principle. Note that H can be written $Hf(x) = H(x,\nabla f(x))$. Chapter 9 gives a detailed discussion of conditions under which the comparison principle holds for H of this form. In particular, if a and b are bounded and continuous and $a(x)$ is positive definite for all x, then Lemmas 9.15 and 9.16 give the desired result.

The conclusion of Theorem 2.3 is less than satisfactory, since $\{V(t)\}$ is only determined implicitly and the rate function is expressed in terms of $\{V(t)\}$. For our example, however, we can write

$$Hf(x) = \sup_u (u\cdot\sigma^T(x)\nabla f(x) + b(x)\nabla f(x) - |u|^2/2),$$

where $\sigma\sigma^T = a$. This representation suggests that V is the Nisio semigroup for a control problem. In particular, we should have

$$V(t)f(x_0) = \sup(f(x(t)) - \int_0^t \frac{1}{2}|u(s)|^2 ds),$$

where the supremum is over x and u satisfying

$$x(t) = x(0) + \int_0^t \sigma(x(s))u(s)ds + \int_0^t b(x(s))ds, \quad x(0) = x_0. \tag{2.7}$$

It then follows, at least under boundedness and continuity assumptions on σ and b, that

$$I(x) = \inf\{I_0(x(0)) + \int_0^t \frac{1}{2}|u(s)|^2 ds\},$$

where the infimum is over all u such that (x,u) satisfies (2.7).

More generally, we look for a linear operator $A \subset C(E) \times C(E \times U)$, where U is another complete, separable metric space, and a lower semicontinuous $L : E \times U \to [0,\infty)$ such that

$$Hf(x) = \sup_{u \in U}(Af(x,u) - L(x,u)).$$

The corresponding control problem may use relaxed controls, that is measures λ on $U \times [0,\infty)$ where $\lambda(U \times [0,t]) = t$. The control problem requires

$$f(x(t)) = f(x(0)) + \int_{U\times[0,t]} Af(x(s),u)\lambda(du \times ds), \tag{2.8}$$

for all $f \in \mathcal{D}(A)$, and V is given by

$$V(t)f(x_0) = \sup_{(x,\lambda)}(f(x(t)) - \int_{U\times[0,t]} L(x(s),u)\lambda(du \times ds),$$

where the supremum is over (x,λ) satisfying (2.8) with $x(0) = x_0$. Some additional constraint on λ is also possible. The rate function then satisfies

$$I(x) = \inf\{I_0(x(0)) + \int_{U\times[0,t]} L(x(s),u)\lambda(du \times ds)\}.$$

See Theorem 8.14.

2.3. General state spaces

We now allow E to be an arbitrary complete, separable metric space. In this more general setting, it is often difficult to verify uniform convergence of $H_n f_n$ and, consequently, weaker notions of convergence are useful. One such notion is convergence of bounded sequences of functions, uniformly on compact sets, which we refer to as *buc*-convergence (see Definition A.6) and denote *buc*- lim.

With reference to Examples 1.8 - 1.11, we also see that the natural limit of H_n may have a range in functions defined on a larger space than E (call it E'). Consequently, we have mappings $\eta_n : E_n \mapsto E$, $\eta'_n : E_n \mapsto E'$, and $\gamma : E' \mapsto E$ which are consistent in the sense that $\eta_n = \gamma \circ \eta'_n$. We require that η_n, η'_n are Borel measurable and γ is continuous and onto (i.e. $E = \gamma(E')$). We also work with a notion of convergence which is essentially *buc*-convergence for our sequence of spaces $\{E_n\}$.

Recall the definitions of set convergence.

DEFINITION 2.4. Define

$$\liminf_{n\to\infty} G_n = \{x \in E : \exists x_n \in G_n, \ni \lim_{n\to\infty} x_n = x\},$$

and

$$\limsup_{n\to\infty} G_n = \{x \in E : \exists n_1 < n_2 < \cdots \text{ and } x_{n_k} \in G_{n_k} \ni \lim_{k\to\infty} x_{n_k} = x\}.$$

If $G \equiv \limsup_n G_n = \liminf_n G_n$, then we say G_n converges to G and write $G = \lim_{n\to\infty} G_n$.

The following notion of convergence will be central to our main results.

DEFINITION 2.5. [LIM convergence induced by an index set $\mathcal{Q}$] Let $\mathcal{Q}$ be an index set. For $n = 1, 2, \ldots$ and $q \in \mathcal{Q}$, let $K_n^q \subset E_n$ satisfy the following:

(1) For $q_1, q_2 \in \mathcal{Q}$, there exists $q_3 \in \mathcal{Q}$ such that $K_n^{q_1} \cup K_n^{q_2} \subset K_n^{q_3}$.
(2) For each $x \in E$, there exists $q \in \mathcal{Q}$ such that $x \in \liminf_{n\to\infty} \eta_n(K_n^q)$.
(3) For each $q \in \mathcal{Q}$, there exist a compact $\widehat{K}^q \subset E$

$$\lim_{n\to\infty} \sup_{y\in K_n^q} \inf_{x\in\widehat{K}^q} r(\eta_n(y), x) = 0,$$

and a compact $\widehat{K}'^q \subset E'$ such that

$$\lim_{n\to\infty} \sup_{y\in K_n^q} \inf_{x\in\widehat{K}'^q} r'(\eta_n'(y), x) = 0. \tag{2.9}$$

(4) For each compact $K \subset E$, there exists $q \in \mathcal{Q}$, such that $K \subset K^q \equiv \liminf_{n\to\infty} \eta_n(K_n^q)$.

For $f_n \in B(E_n)$ and $f \in B(E)$, define $f = \text{LIM} f_n$ if and only if $\sup_n \|f_n\| < \infty$ and

$$\lim_{n\to\infty} \sup_{y\in K_n^q} |f_n(y) - \eta_n f(y)| = 0,$$

where $\eta_n f = f \circ \eta_n$.

If the requirements in the above definition hold, the compact sets $\widehat{K}^q$ and $\widehat{K}'^q$ can always be chosen to be $\widehat{K}^q = \limsup_{n\to\infty} \eta_n(K_n^q)$ and $\widehat{K}'^q = \limsup_{n\to\infty} \eta_n'(K_n^q)$. Note that the definition of LIM given here is a special case of the abstract definition given in Section A.2. (See Condition A.13.) We give two useful examples of LIM.

EXAMPLE 2.6. If $E_n = E$, $\mathcal{Q}$ is the class of compact sets in E, and $K_n^K = K$ for every $n = 1, 2, \ldots$ and $K \in \mathcal{Q}$, then LIM is just *buc*-convergence.

EXAMPLE 2.7. With reference to Examples 1.12 and 1.13, let η_n be given by (1.31). For $m = m(n)$, let $E = E' = L^2(\mathcal{O})$, and let $\mathcal{Q}$ be the class of compact sets in E. For each $\rho \in E$, define its projection into $E_n = R^{|\Lambda_m|}$ by

$$(\pi_n\rho)(x) = m^d \int_{y_1=\frac{i_1}{m}}^{\frac{i_1+1}{m}} \cdots \int_{y_d=\frac{i_d}{m}}^{\frac{i_d+1}{m}} \rho(y_1, \ldots, y_d) dy_d \ldots dy_1, \quad x = (\frac{i_1}{m}, \ldots, \frac{i_d}{m}) \in \Lambda_m. \tag{2.10}$$

Then $\eta_n(\pi_n(\rho)) \in E$, and assuming $m(n) \to \infty$, for every compact set $K \subset L^2(\mathcal{O})$,

$$\lim_{n\to\infty} \sup_{\rho\in K} \|\eta_n(\pi_n(\rho)) - \rho\|_{L^2(\mathcal{O})} = 0. \tag{2.11}$$

If, for each $K \in \mathcal{Q}$, we define

$$K_n^K = \pi_n(K) \equiv \{\pi_n(\rho) : \rho \in K \subset E\} \subset E_n = R^{|\Lambda_{m(n)}|},$$

then all the requirements for $\{K_n^K : K \in \mathcal{Q}\}$ in Definition 2.5 are satisfied. In particular, for each $K \in \mathcal{Q}$, by (2.11) and the above definition of K_n^K,

$$\lim_{n\to\infty} \sup_{\gamma_n\in K_n^K} \inf_{\gamma\in K} \|\eta_n(\gamma_n) - \gamma\|_{L^2(\mathcal{O})} \le \lim_{n\to\infty} \sup_{\rho\in K} \|\eta_n(\pi_n(\rho)) - \rho\|_{L^2(\mathcal{O})} = 0.$$

Therefore, we can choose $\widehat{K}^K = K$ to satisfy the third requirement in Definition 2.5. Finally, since

$$\sup_{\gamma\in K_n^K} |f_n(\gamma) - \eta_n f(\gamma)| = \sup_{\rho\in K} |f_n(\pi_n(\rho)) - f(\eta_n(\pi_n(\rho)))|,$$

if $f \in C(E)$, $f = \mathrm{LIM} f_n$ is implied by the following two conditions:

(1) $\sup_n \|f_n\| < \infty$,

(2) for $\gamma_n \in R^{|\Lambda_{m(n)}|}$ satisfying $\eta_n(\gamma_n) \to \rho$ in $\|\cdot\|_{L^2(\mathcal{O})}$, $\lim_{n\to\infty} f_n(\gamma_n) = f(\rho)$.

Dropping compactness and generalizing the notion of convergence requires extra work and part of the extra work comes in verifying the following condition.

CONDITION 2.8. *For each $q \in \mathcal{Q}$, $T > 0$, and $a > 0$, there exists $\widehat{q}(q, a, T) \in \mathcal{Q}$ such that*

(2.12)

$$\limsup_{n\to\infty} \sup_{y \in K_n^q} \frac{1}{n} \log P\{Y_n(t) \notin K_n^{\widehat{q}(q,a,T)}, \text{ for some } 0 \le t \le T | Y_n(0) = y\} \le -a.$$

The Lyapunov function techniques discussed in Section 4.6 are frequently useful for verifying this condition.

EXAMPLE 2.9. We illustrate the Lyapunov function approach for Example 1.12. Definitions and properties of the projection $\pi_m : E \to R^{|\Lambda_m|} \equiv E_n$ and of LIM convergence are discussed in Example 2.7. In this context, Condition 2.8 becomes: for each compact $K_0 \subset L^2(\mathcal{O})$, $T > 0$, and $a > 0$, there exists another compact $K_1 \subset L^2(\mathcal{O})$ such that

(2.13)

$$\limsup_{n\to\infty} \sup_{\rho_0 \in \pi_m(K_0)} \frac{1}{n} \log P(\exists t, 0 \le t \le T, \ni Y_n(t) \notin \pi_m(K_1) | Y_n(0) = \rho_0) \le -a.$$

Example 1.12 is motivated by a physical context. We can associate with the model (1.29) a *free energy functional* of the form

$$\mathcal{E}_m(\rho) \equiv \frac{1}{2} \sum_{x \in \Lambda_m} |\nabla_m \rho(x)|^2 m^{-d} + \sum_{x \in \Lambda_m} F(\rho(x)) m^{-d}, \qquad \rho \in E_n \equiv R^{|\Lambda_m|}, \tag{2.14}$$

where $\nabla_m g(x)$ is the vector with components

$$\nabla_m^k \rho(x) = \frac{m}{2}(\rho(x + m^{-1}e_k) - \rho(x - m^{-1}e_k)).$$

$\mathcal{E}_m$ is the desired Lyapunov function.

Define

$$\begin{aligned} H_n \mathcal{E}_m(\rho) &\equiv \frac{1}{n} e^{-n\mathcal{E}_m} A_n e^{n\mathcal{E}_m}(\rho) \qquad (2.15)\\ &= \langle \Delta_m \rho - F'(\rho), -\Delta_m \rho + F'(\rho) \rangle_m + \frac{1}{2} \|\Delta_m \rho - F'(\rho)\|^2_{L^2(\Lambda_m)} \\ &\quad + \frac{m^d}{2n} \sum_{x \in \Lambda_m} (\frac{d}{2} m^{2-d} + F''(\rho(x)) m^{-d}) \\ &= -\frac{1}{2} \|\Delta_m \rho - F'(\rho)\|^2_{L^2(\Lambda_m)} + \frac{1}{2n} \Big(\frac{1}{2} d m^{2+d} + m^d \sum_{x \in \Lambda_m} F''(\rho(x)) m^{-d} \Big). \end{aligned}$$

Then by Itô's formula,

$$\exp\{n\mathcal{E}_m(Y_n(t)) - n\mathcal{E}_m(Y_n(0)) - \int_0^t n H_n \mathcal{E}_m(Y_n(s)) ds\}$$

is a positive local martingale (hence a supermartingale), and we can apply Lemma 4.20 to obtain (2.13).

If $m^{2+d} = O(n)$, then

$$\sup_n \sup_{\rho \in R^{|\Lambda_m|}} H\mathcal{E}_m(\rho) < \infty. \tag{2.16}$$

Let $C > C_0 > 0$ and $\rho_0 \in R^{|\Lambda_m|}$ be such that $\mathcal{E}_m(\rho_0) \le C_0 < \infty$. By the finiteness of Λ_m, $\{\rho \in R^{|\Lambda_m|} : \mathcal{E}_m(\rho) < C\}$ is open in E_n. By Lemma 4.20,

$$P(\exists t \in [0,T], \mathcal{E}_m(Y_n(t)) \ge C | Y_n(0) = \rho_0) \le e^{-n(C-C_0)+nT \sup_n \sup_{\rho \in L^2(\Lambda_m)} H_n \mathcal{E}_m(\rho)}.$$

If

$$F(r) \ge c_1 + c_2 r^2 \text{ for some } c_1 \in R, c_2 > 0, \tag{2.17}$$

then

$$K_C \equiv \text{cl } (\cup_n \{\eta_n(\rho) : \rho \in E_n, \mathcal{E}_m(\rho) \le C < \infty\})$$

is a compact subset of $E = L^2(\mathcal{O})$. Therefore, for every $a, T, C_0 > 0$, by selecting C large enough, we can find a compact set $K_{a,T,C_0} \subset L^2(\mathcal{O})$ such that

$$\sup_{\{\rho \in R^{|\Lambda_m|} : \mathcal{E}_m(\rho) \le C_0\}} P(\exists t, 0 \le t \le T, \eta_n(Y_n(t)) \notin K_{a,T,C_0}, | Y_n(0) = \rho) \le e^{-na}. \tag{2.18}$$

Next, we extend (2.18) to (2.13) by relaxing the requirements on the initial data. For this purpose, we need to make use of a stability result regarding equation (1.29) which is proved in Lemma 13.4: For $T > 0$, there exists $C_T > 0$ such that if Y_n and Z_n are solutions of (1.29),

$$\sup_{0 \le t \le T} \|Y_n(t) - Z_n(t)\|_{L^2(\Lambda_m)} \le C_T \|Y_n(0) - Z_n(0)\|_{L^2(\Lambda_m)} \quad a.s. \tag{2.19}$$

Noting that the collection of $\rho \in L^2(\mathcal{O})$ such that

$$\sup_n \mathcal{E}_m(\pi_m(\rho)) < \infty \tag{2.20}$$

is dense in $L^2(\mathcal{O})$, for compact $K_0 \subset L^2(\mathcal{O})$ and $\epsilon > 0$, there exist $\rho_0^{(1)}, \ldots, \rho_0^{(N)}$ such that $K_0 \subset \cup_{k=1}^N B(\rho_0^{(k)}, C_1^{-1}\epsilon)$ and

$$C_0^\epsilon \equiv \max_{1 \le k \le N} \sup_m \mathcal{E}_m(\pi_m(\rho_0^{(k)})) < \infty. \tag{2.21}$$

Therefore, by (2.18) and (2.19), there exists a compact K_{a,T,C_0^ϵ} such that

$$\begin{aligned} &\sup_{\{\rho \in \pi_m(K_0)\}} P(\exists t, 0 \le t \le T, \eta_n(Y_n(t)) \notin K^\epsilon_{a,T,C_0^\epsilon}, | Y_n(0) = \rho) \\ &\le \sup_{\{\rho \in R^{|\Lambda_m|} : \mathcal{E}_m(\rho) \le C_0^\epsilon\}} P(\exists t, 0 \le t \le T, \eta_n(Y_n(t)) \notin K_{a,T,C_0}, | Y_n(0) = \rho) \\ &\le e^{-na}, \end{aligned}$$

where

$$K^\epsilon_{a,T,C_0^\epsilon} = \{\rho \in E : \inf_{\gamma \in K_{a,T,C_0^\epsilon}} \|\rho - \gamma\|_{L^2(\mathcal{O})} < \epsilon\}.$$

For a sequence $\epsilon_l \to 0$, let

$$K_1 = \text{cl} \cap_l K^{\epsilon_l}_{a+l,T,C_0^{\epsilon_l}}.$$

Then K_1 is compact, and since $\rho = \pi_m(\eta_n(\rho))$ for every $\rho \in E_n$, $\{\eta_n(Y_n(t)) \in K_1\} \subset \{Y_n(t) \in \pi_m(K_1)\}$ and (2.13) holds.

For most φ, (1.32) will be unbounded and the convergence of $H_n f_n$ will not be uniform over E_n. Examples of this type lead us to introduce a pair of limiting operators: $H_\dagger$, corresponding (roughly) to the limit when $\{H_n f_n\}$ is bounded above and $H_\ddagger$ corresponding to $\{H_n f_n\}$ being bounded below. (See the Convergence Condition 7.11.)

We have the following counterpart of Theorem 2.3. Note that since we are no longer assured that the extrema in Definition 2.2 are achieved, more general definitions of viscosity solution and the comparison principle are also required. (See Definitions 7.1 and 7.3.)

THEOREM 2.10. *Let (E_n, r_n), (E', r'), (E, r), η_n, η'_n, and γ be as above, and let* LIM *be given by Definition 2.5.*

Step 1: *(Convergence of H_n) In the continuous time case, let*

$$H_n f \equiv \frac{1}{n} e^{-nf} A_n e^{nf}, \quad e^{nf} \in \mathcal{D}(A_n).$$

In the discrete time case, let

$$H_n f \equiv \frac{1}{n\epsilon_n} \log e^{-nf} T_n e^{nf} = \frac{1}{n\epsilon_n} \log(1 + e^{-nf}(T_n - I)e^{nf}).$$

Let $H_\dagger \subset C^l(E, \overline{R}) \times M^u(E', \overline{R})$ and $H_\ddagger \subset C^u(E, \overline{R}) \times M^l(E', \overline{R})$, and assume that $\{H_n\}$ satisfies the Convergence Condition 7.11.

Step 2: *(Exponential tightness) Suppose $\mathcal{D}(H_\dagger) \cap \mathcal{D}(H_\ddagger) \cap C_b(E)$ is* buc-*dense in $C_b(E)$ and Condition 2.8 is satisfied. Then $\{X_n\}$ is exponentially tight.*

Step 3: *(The comparison principle) Let $\alpha_0 > 0$, and assume that for each $0 < \alpha < \alpha_0$, there exists a subset $D_\alpha \subset C_b(E)$ such that $C_b(E)$ is the buc-closure (Definition A.6) of D_α and that for each $h \in D_\alpha$, the comparison principle holds for viscosity subsolutions of $(I - \alpha H_\dagger)f = h$ and supersolutions of $(I - \alpha H_\ddagger)f = h$. Then a viscosity solution $R_\alpha h$ exists for each $h \in D_\alpha$, and R_α extends continuously (in terms of* buc-*convergence) to all of $C_b(E)$.*

In addition, we suppose that $\{X_n(0)\}$ satisfies a large deviation principle with a good rate function I_0.

Then

a) *$\{X_n\}$ satisfies the large deviation principle with a good rate function I.*

b) *The limit*

$$V(t)h \equiv \lim_{m\to\infty} R^m_{t/m} h$$

exists and defines an operator semigroup on $C_b(E)$, and the rate function is given by (2.6).

The above result is a special case of Theorem 7.18. The assumption that $\limsup_{n\to\infty} \eta'_n(K^q_n)$ is compact in E' can be replaced by weaker conditions. See Theorem 7.24. Theorem A.8 gives conditions under which $C_b(E)$ is the buc-closure of a subset D.

Finally, the control representation method for simplifying the rate function (Theorem 8.14 and Corollary 8.28) continues to apply for non-compact E.