Available online at www.sciencedirect.com
SCIENCE DIRECT®

Nonlinear Analysis 63 (2005) e2067 – e2078

Nonlinear Analysis

www.elsevier.com/locate/na

Microscopic derivations of several Hamilton–Jacobi equations in infinite dimensions, and large deviation of stochastic systems

Jin Feng[1]

Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01002, USA

Abstract

We consider Hamilton–Jacobi equations which characterize optimal controlled partial differential equations of the following types: the Allen–Cahn equation, the Cahn–Hilliard equation, a nonlinear Fokker–Planck equation, and a Vlasov–Fokker–Planck equation. In each of the examples, the optimal control problem and its associated cost functional can be derived as limit from a microscopically defined stochastic system, using the probabilistic theory of large deviation. The physical context here makes it natural to derive a free energy inequality, which is very useful in proving the well-posedness of the Hamilton–Jacobi equation. The article is written using informal arguments. Rigorous results will appear elsewhere.

Keywords: Hamilton–Jacobi equations in infinite dimensions; Large deviations; Viscosity solutions; Stochastic systems

1. Introduction

1.1. Examples of optimal controlled partial differential equations

We consider the following optimal control problems:

1. The controlled Allen–Cahn equation

$$\frac{\partial}{\partial t}\rho(t,x) = \Delta_x \rho(t,x) - F'(\rho(t,x)) + u(t,x) \tag{1}$$

[1] A large portion of the material presented here is taken from collaborative research with Markos Katsoulakis and with Thomas G. Kurtz in the last several years.

E-mail address: feng@math.umass.edu.

0362-546X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2005.03.011

with running cost $L(\rho, u) = \frac{1}{2}\int_{\mathcal{O}} |u(x)|^2\, dx = \frac{1}{2}\|u\|^2_{L^2(\mathcal{O})}$. In the above, $\mathcal{O} = [0, 1)^d$ with periodic boundary condition; we consider solutions $\rho = \rho(t, x)$, where $\rho(t, \cdot) \in E \equiv L^2(\mathcal{O})$ and controls $u = u(t, x)$, where $u(t, \cdot) \in L^2(\mathcal{O})$.

2. The controlled Cahn–Hilliard equation

$$\frac{\partial}{\partial t}\rho(t, x) = \Delta_x(-\Delta_x \rho(t, x) + F'(\rho(t, x))) + u(t, x) \tag{2}$$

with running cost $L(\rho, u) = \frac{1}{2}\|u\|^2_{-1} \equiv \sup_{p \in C^\infty(\mathcal{O})} \{\langle u, p\rangle - \frac{1}{2}\int_{\mathcal{O}} |\nabla p|^2\, dx\}$, where $\mathcal{O} = [0, 1)^d$ with periodic boundary condition. We consider solutions $\rho = \rho(t, x)$, where $\rho(t, \cdot) \in E \equiv L^2(\mathcal{O})$. The controls $u = u(t, x)$ are taken to be such that $u(t, \cdot) \in H^{-1}(\mathcal{O}) \equiv \{u : \|u\|^2_{-1} < +\infty\}$.

3. The controlled nonlinear Fokker–Planck equation

$$\frac{\partial}{\partial t}\rho(t, x) = \frac{1}{2}\Delta\rho + \nabla \cdot (\rho\nabla(\Psi + \Phi * \rho)) + u \tag{3}$$

with running cost $L(\rho, u) = \frac{1}{2}\|u\|^2_{-1,\rho} \equiv \sup_{p \in C_c^\infty(R^d)} \{\langle u, p\rangle - \frac{1}{2}\int_{R^d} |\nabla p|^2\, d\rho\}$. We consider $\rho(t, \cdot) \in E \equiv \mathcal{P}_2(R^d)$ (which is the space of probability measures with finite second moment). $\Phi, \Psi \in C^2(R^d)$ and Φ is even. We consider controls u satisfying $u(t) \in H^{-1}_{\rho(t)}(R^d) = \{u : \|u\|_{-1,\rho(t)} < +\infty\}$ for $t \geqslant 0$. At least formally, $u = -\nabla \cdot (\rho\nabla p)$ for some $p(t, x)$ and $L(\rho, u) = \frac{1}{2}\int_{R^d} |\nabla p|^2\, d\rho$.

4. The controlled nonlinear Vlasov–Fokker–Planck equation

$$\begin{aligned}\frac{\partial}{\partial t}\rho(t, x, v) = &-v \cdot \nabla_x \rho + \nabla_v \cdot (\rho\nabla_x(\rho * \Phi + \Psi)) \\ &+ \left(\beta\nabla_v \cdot (\rho v) + \frac{\sigma^2}{2}\Delta_{vv}\rho\right) + u,\end{aligned} \tag{4}$$

where $(\rho * \Phi)(x) = \int_{R^d \times R^d} \Phi(x - y)\rho(t, dy, dv) = \int_{R^d} \Phi(x - y)\rho(t, dx, R^d)$, $\sigma > 0$ and $u = -\sigma\nabla_v(\rho\nabla_v p)$ for some $p(t, x)$ and the running cost $L(\rho, u) = \frac{1}{2}\int_{R^d \times R^d} |\nabla_v p|^2\, d\rho$. For each $t \geqslant 0$, $\rho(t) = \rho(t, dx, dv) \in E \equiv \mathcal{P}_2(R^d \times R^d)$ the space of probability measures on $R^d \times R^d$ with finite second moments. $\Phi(x)$, $\Psi(x)$ are the same as in the last example.

1.2. *The associated Hamilton–Jacobi equation in infinite dimensions*

For each of the four controlled problem above, we introduce a bivariate function $H(\rho, \varphi)$:

$$H(\rho, \varphi) = \langle \Delta\rho - F'(\rho), \varphi\rangle + \frac{1}{2}\|\varphi\|^2_{L^2(\mathcal{O})}, \quad \rho \in E, \ \ \varphi \in C^\infty(\mathcal{O}), \tag{5}$$

$$H(\rho, \varphi) = \langle \Delta(-\Delta\rho + F'(\rho)), \varphi\rangle + \frac{1}{2}\|\nabla\varphi\|^2_{L^2(\mathcal{O})}, \quad \rho \in E, \ \ \varphi \in C^\infty(\mathcal{O}), \tag{6}$$

$$\begin{aligned}H(\rho, \varphi) = &\left\langle \frac{1}{2}\Delta\rho + \nabla(\rho\nabla(\Psi + \Phi * \rho)), \varphi \right\rangle \\ &+ \frac{1}{2}\int_{R^d} |\nabla\varphi|^2\, d\rho, \quad \rho \in E, \ \ \varphi \in C_c^\infty(R^d),\end{aligned} \tag{7}$$

$$H(\rho, \varphi) = \left\langle \frac{\sigma^2}{2} \Delta_{vv}\rho + \beta\nabla_v \cdot (\rho v) - v \cdot \nabla_x \rho - \nabla_v \cdot (\rho \nabla_x(\rho * \Phi + \Psi)), \varphi \right\rangle + \frac{\sigma^2}{2} \int_{R^d \times R^d} |\nabla_v \varphi|^2 \, d\rho, \quad \rho \in E, \ \varphi \in C_c^\infty(R^d \times R^d). \tag{8}$$

We consider test functions of the form

$$f(\rho) = \psi(\langle \rho, \varphi_1 \rangle, \ldots, \langle \rho, \varphi_m \rangle) = \psi(\langle \rho, \vec{\varphi} \rangle), \tag{9}$$

where $\varphi_1, \ldots, \varphi_m \in C_c^\infty(\mathscr{O})$, $\psi \in C^2(R^m)$ and $m = 1, 2, \ldots$. We define a Hamiltonian operator in each case by setting $Hf(\rho) = H(\rho, \delta f/\delta\rho)$, with the unconstrained variational derivative defined by

$$\frac{\delta f}{\delta \rho} \equiv \sum_{k=1}^{m} \partial_k \psi(\langle \rho, \varphi_1 \rangle, \ldots, \langle \rho, \varphi_m \rangle)\varphi_k. \tag{10}$$

Many questions regarding the control problems (1)–(4) can be characterized through Hamilton–Jacobi equation

$$(I - \alpha H) f = h, \quad \alpha > 0, \ h \in C_b(E). \tag{11}$$

By the *dynamic programming principle*, a formal solution is given by the value function

$$f(\rho_0) = R_\alpha h(\rho_0) \equiv \sup \left\{ \int_0^\infty e^{-\alpha^{-1}t} (\alpha^{-1} h(\rho(t)) - L(\rho(t), u(t))) \, dt : (\rho(\cdot), u(\cdot)), \rho(0) = \rho_0 \right\}. \tag{12}$$

Solution for (11) should be interpreted in the *viscosity solution* sense. The definition we use here follows [4], which generalizes that in [1]. Heuristically, $\overline{f}$ is called a *sub*- (viscosity) solution to (11), if $(I - \alpha H)\overline{f} \leqslant h$; and $\underline{f}$ is a *super*- (viscosity) solution, if $(I - \alpha H)\underline{f} \geqslant h$. f is a solution if it is both sub- and super-solution. We say that the *comparison principle* holds for (11) if for every upper semicontinuous bounded subsolution $\overline{f}$ and for every lower semicontinuous bounded supersolution $\underline{f}$, we have $\overline{f} \leqslant \underline{f}$. It follows that such comparison implies uniqueness of bounded continuous viscosity solution.

We are interested in the well-posedness of (11) and its connection with the probabilistic *large deviation theory*. In Section 2, we introduce four stochastic models originated from physical literature. In Section 3, we reveal connections between controlled partial differential equations (1)–(4) and the stochastic models, using the theory of large deviation for Markov processes. The connection can be roughly explained as follows: By building a random term in each microscopic model, we introduce a selection of preference criteria on the system's path of evolution. The nonlinear scaling in the large deviation limit captures such selection behavior at the macroscopic level, giving rise to the deterministic control problem with a certain cost structure. In Section 4, we outline a program for proving the comparison principle of (11), and for constructing solution using a convergence of viscosity solution scheme motivated by large deviation considerations. Various simplified versions of such a program are well known; if the state space for the control and/or the large deviation problem is finite dimensional. However, there have been counter-examples showing that native generalizations to the infinite dimensional setup may fail. Our generalization relies on a probabilistic property (23) regarding concentration of measures on compact sets for the rescaled stochastic systems, and a free energy inequality (28) for the limiting control equations. They are natural for the examples we consider here.

2. Microscopic level stochastic models

In this section, we first introduce two stochastic Ginzburg–Landau models which are associated with the controlled Allen–Cahn and Cahn–Hilliard equations. Then, we discuss two stochastic particle systems which are linked with the controlled Fokker–Planck and Vlasov–Fokker–Planck equations.

2.1. Stochastic Ginzburg–Landau equations

The stochastic Ginzburg–Landau equations are phenomonological models used by physicists to study phase transition behaviors. See [8]. Let $\rho(x)$ be the order parameter at location $x \in \mathcal{O}$. Depending on applications, such parameter could model the local density of a material, or strength of magnetization, etc. Let $F : R \mapsto R$ model a potential field. We define the *Ginzburg–Landau–Wilson free energy* functional

$$\mathscr{E}(\rho) = \frac{1}{2}\int_{\mathcal{O}} |\nabla \rho|^2 \,\mathrm{d}x + \int_{\mathcal{O}} F(\rho(x)). \tag{13}$$

First, we consider $d=1$. Let $W(t,x)$ be a time-space Brownian sheet. By the Ginzburg–Landau model, we refer to the following stochastic partial differential equation:

$$\partial_t \rho_n = -A\,\frac{\delta \mathscr{E}}{\delta \rho} + \frac{1}{\sqrt{n}}\sqrt{A}\partial_t \partial_x W(t,x), \tag{14}$$

where n is a scaling parameter and A is a positive operator. The form of A in front of $\delta\mathscr{E}/\delta\rho$ but $\sqrt{A}$ in front of the random term is motivated by the *fluctuation-dissipation* theorem in statistical mechanics.

By integration by parts, the first (un-constrained) variation of this functional is $\delta\mathscr{E}/\delta\rho = -\Delta\rho + F'(\rho)$. If we let $A=I$ be the identity operator, we arrive at the *stochastic Allen–Cahn equation* which is a randomly perturbed reaction-diffusion equation

$$\partial_t \rho_n = \Delta \rho_n - F'(\rho_n) + \frac{1}{\sqrt{n}}\partial_t \partial_x W(t,x). \tag{15}$$

On the other hand, if we take $A = -\varDelta$, we arrive at the *stochastic Cahn–Hilliard equation*

$$\begin{aligned}\partial_t \rho = \varDelta_x \frac{\delta \mathscr{E}}{\delta \rho} + \nabla_x \cdot \partial_t \partial_x \mathbf{W}(t,x) &= \varDelta_x(-\varDelta_x \rho(t,x) + F'(\rho(t,x))) \\ &\quad + \nabla_x \cdot \partial_t \partial_x \mathbf{W}(t,x).\end{aligned} \tag{16}$$

In the above, we have written the equation in a form allowing the possibility of $d > 1$. The $\mathbf{W} = (W_1, \ldots, W_d)$ is a vector valued Brownian sheet consisting independent Brownian sheets $W_k(t, x)s$. Unfortunately, when $d > 1$, the above equations do not have function valued solutions. For our practical purpose here, we may discretize $\mathcal{O}$ and discretize the stochastic equations on $\mathcal{O}$ to a well-posed finite stochastic system. We will let the discretization mesh size go to zero slowly as $n \to +\infty$ [4].

2.2. Interacting stochastic particle systems

First, we consider a finite system

$$\mathrm{d}X_i(t) = -\nabla \Psi(X_i(t)) - \frac{1}{n} \sum_{j=1}^{n} \nabla \Phi(X_i(t) - X_j(t))\,\mathrm{d}t + \mathrm{d}W_i(t), \tag{17}$$

where $W_i(t),\ i = 1, 2, \ldots, n$ are R^d-valued independent standard Brownian motions. By symmetry, we can equivalently consider the process of empirical measures

$$\rho_n(t, \mathrm{d}x) = n^{-1} \sum_{k=1}^{n} \delta_{X_k(t)}(\mathrm{d}x).$$

$\{\rho_n(t) : t \geqslant 0\}$ is a probability-measure-valued Markov process. Choose $\Phi(x) = \theta|x|^2, \theta > 0$ and let $\Psi(x)$ be a double well function such as $|x|^4/4 - |x|^2/2$, the (17) in such situation is known as the Curie–Weiss model.

The above model is closely related to a stochastic system studied in the kinetic theory of gas in a situation where the gas molecules are kept in contact with an environment having constant temperature $T \geqslant 0$. Let $X_i \in R^d,\ i = 1, \ldots, n$ denote the position of n indistinguishable particles with unit mass. They interact with each other as well as with an outside environment. The movement of each particle is dictated by four types of force: (1) that induced by a pairwise interaction potential Φ which depends on the particle distance $|X_i - X_j|$; (2) that due to an external potential Ψ which only dependents on the position of the ith particle; (3) that due to friction with the environment which acts on each individual particle and is proportional to its speed; and (4) a random force acting on each particle due to collisions received from the environment. Introducing velocity vector $V_i \in R^d$ for the ith particle, then by the Newtonian laws,

$$\dot{X}_i(t) = V_i(t),$$

$$\mathrm{d}V_i(t) = -\frac{1}{n} \sum_{j=1}^{n} \nabla \Phi(X_i(t) - X_j(t)) - \beta V_i(t)\,\mathrm{d}t - \nabla \Psi(X_i(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W_i(t), \tag{18}$$

$i = 1, 2, \ldots, n$. By the *fluctuation-dissipation* theorem, $\sigma^2 = 2\beta kT$ where $k > 0$ is the Boltzmann constant.

In this context, we consider measure-valued Markov process $\rho_n(t, \mathrm{d}x, \mathrm{d}v) = n^{-1} \sum_{k=1}^{n} \delta_{(X_i(t), V_i(t))}(\mathrm{d}x, \mathrm{d}v)$.

3. Large deviations for Markov processes and Hamilton–Jacobi equations

3.1. The large deviation principle

Let X_n be a sequence of metric space S-valued random variables. The *large deviation principle* for $\{X_n : n = 1, 2, \ldots\}$ refers to the following asymptotic

$$P(X_n \in A) \sim \mathrm{e}^{-nv(A)} = \mathrm{e}^{-n \inf_{\rho \in A} I(\rho)}, \quad \text{for every Borel set } A \subset S, \tag{19}$$

where the set function v is given by the *rate function* $I : S \mapsto [0, +\infty]$ such that $v(A) = \inf_{\rho \in A} I(\rho)$. Intuitively, I measures the rate of deviation from the most likely event(s). As in the theory of weak convergence of probability measures, the large deviation principle has an equivalent characterization using moment functionals (due to Varadhan and Brycs): Suppose that the X_n's are sufficiently concentrated on compact sets (i.e. the *exponential tightness* condition holds) then $\{X_n : n = 1, 2, \ldots\}$ satisfies the large deviation principle if and only if $V(f) = \lim_{n \to +\infty} V_n(f)$ for every $f \in C_b(S)$, where $V_n(f) \equiv (1/n) \log E[\mathrm{e}^{nf(X_n)}]$. Furthermore,

$$I(x) = \sup_{f \in C_b(S)} \{f(x) - V(f)\}, \quad V(f) = \sup_{x \in S} \{f(x) - I(x)\}. \tag{20}$$

3.2. The case of Markov process and Hamilton–Jacobi equations

Let n be fixed. Each of the above stochastic process $\{\rho_n(t) : t \geqslant 0\}$ is Markov with continuous (in time) path in the space $S \equiv C_E[0, +\infty)$. The state space E is either function- or probability-measure valued. Let

$$V_n(t) f(\rho) = \frac{1}{n} \log E[\mathrm{e}^{nf(\rho_n(t))} | \rho_n(0) = \rho], \quad \forall f \in C_b(E). \tag{21}$$

Following the functional convergence characterization of the large deviation principle, we expect existence of the following limit plays an important role for the large deviation of the process to hold: $V(t)f = \lim_{n \to +\infty} V_n(t) f$ for $f \in C_b(E)$ and $t \geqslant 0$. It is interesting to note that $V_n(t)$ actually forms a nonlinear semigroup $V_n(s+t) = V_n(s) V_n(t)$. Therefore, we further expect that large deviation is implied by convergence of generators $H_n \to H$. Formal computation reveals that

$$H_n f(x) = \lim_{t \to 0+} \frac{1}{t} (V_n(t) f(x) - f(x)) = \frac{1}{n} \mathrm{e}^{-nf(x)} A_n \mathrm{e}^{nf(x)}, \tag{22}$$

where A_n is the usual linear generator of the transition semigroup induced by the Markov process $\rho_n(t)$. Feng and Kurtz [4] rigorously develops the above program to prove large deviation for metric-space-valued Markov processes. The four essential steps in such a program are:

1. Identifying generator A_n and H_n, and verifying the existence of limit $H_n \to H$.
2. Verifying *exponential compact containment condition*: for each compact set $K_0 \subset E$ and $T > 0$, $a > 0$, there exists another compact set $K_1 \subset E$ such that

$$\sup_{\rho_0 \in K_0} P(\exists t \in [0, T], \text{ such that } \rho_n(t) \notin K_1 | \rho_n(0) = \rho_0) \leqslant e^{-na}. \tag{23}$$

3. For a large class of functions $D \subset C_b(E)$, and $h \in D$, $\alpha > 0$, we verify the *comparison principle* for (11).
4. Verifying a variational Nisio-semigroup representation holds for H, hence giving the Hamilton–Jacobi equation a control problem interpretation. The running cost structure L is derived from such representation.

Transformation of type (22) can be traced back to the works of Fleming and Sheu on large deviation of exit probabilities for diffusions, in the late 1970s. See [5] for a review.

To illustrate the applicability of the above program, we verify the first and the last steps for the weakly interacting particle system (17) which was first studied in [2]. Let f be of the form (9). By Ito's formula,

$$\begin{aligned} H_n f(\rho) &= \left\langle \Delta\rho + \nabla \cdot (\rho \nabla(\Psi + \rho * \Phi)), \frac{\delta f}{\delta \rho} \right\rangle + \frac{1}{2} \int \left| \nabla \frac{\delta f}{\delta \rho} \right|^2 d\rho \\ &\quad + \frac{1}{2n} \sum_{i,j=1}^{l} \partial^2_{ij} \psi(\langle \vec{\varphi}, \rho \rangle) \int \nabla \varphi_i \cdot \nabla \varphi_j \, d\rho \to H\left(\rho, \frac{\delta f}{\delta \rho}\right) = Hf(\rho). \end{aligned}$$

The variational formula

$$\begin{aligned} H(\rho, \varphi) &= \sup_{u \in H^{-1}_{\rho}(R^d)} \left\{ \langle \Delta\rho + \nabla \cdot (\rho \nabla(\Psi + \rho * \Phi)) + u, \varphi \rangle - \frac{1}{2} \|u\|^2_{-1,\rho} \right\}, \\ &\rho \in E, \quad \varphi \in C_c^\infty(R^d) \end{aligned}$$

then gives the Nisio semigroup representation and the optimal control problem (3). More precisely, convergence of $V_n(t)$ to $V(t)$ implies the large deviation holds for $\{\rho_n(t) : n = 1, 2, \ldots\}$ conditioning on $\rho_n(0) = \rho_0$. Let $I_t(\cdot|\rho_0) : E \mapsto [0, +\infty]$ denote the rate function. By the comparison principle and by the dynamic programming principle, we can show that

$$\begin{aligned} V(t) f(\rho) &= \sup_{u(\cdot)} \left\{ f(\rho(t)) - \int_0^t L(\rho(s), u(s)) \, ds : (\rho(\cdot), u(\cdot)) \text{ satisfies } (3) \right\} \\ &= \sup_{\gamma \in E} \{ f(\gamma) - I_t(\gamma|\rho) \}, \end{aligned}$$

where the last equality follows from the second part of (20). By the first part of (20),

$$\begin{aligned} I_t(\gamma_0|\rho_0) &\equiv \sup_{f \in C_b(E)} \{ f(\gamma_0) - V(t) f(\rho_0) \} \\ &= \inf_u \left\{ \int_0^t L(\rho(s), u(s)) \, ds : \rho(0) = \rho_0, \rho(t) = \gamma_0 \right\}. \end{aligned} \tag{24}$$

By the Markovian property of $\rho_n(\cdot)$ and by a projective limit argument, $\{\rho_n(\cdot) : n=1, 2, \ldots\}$ satisfies large deviation with rate function

$$\begin{aligned} I(\rho(\cdot)) &= \sup_{0 \leqslant t_1 < \cdots < t_k} \sum_{k=1}^{k} I_{t_i - t_{i-1}}(\rho(t_i)|\rho(t_{i-1})) \\ &= \inf_{u(\cdot)} \left\{ \int_0^\infty L(\rho(s), u(s))\, \mathrm{d}s : (\rho(\cdot), u(\cdot)) \text{ satisfies } (3) \right\}. \end{aligned}$$

We use the convention that inf over empty set gives $+\infty$.

We remark that although all the four large deviation examples considered here are infinite-dimensional versions of the single scale type random perturbation problems considered in the Freidlin–Wentzell theory [6], the general H_n operator convergence method in [4] can also be effectively applied to other more complex problems with multi-scales.

4. Existence, uniqueness and convergence of the Hamilton–Jacobi equations

We claimed that convergence $H_n \to H$ implies $V_n \to V$. As in the classical nonlinear semigroup theory, this can be proved by verifying convergence of resolvents in the viscosity sense. Next, let

$$(I - \alpha H_n) f_n = h_n, \tag{25}$$

where $h_n \to h$ in appropriate sense. We show $f_n \to f$. We will also need the proof of comparison principle for (11), and discuss the verification of (23). We choose to do all these through examples.

4.1. Controlled gradient flows and free energy inequalities

The first three control problems in (1)–(3) can all be written as controlled gradient flows

$$\dot{\rho} = -\mathrm{grad}\, \mathscr{E}(\rho) + u. \tag{26}$$

In addition, their corresponding H operator admits a common structure

$$Hf(\rho) = \langle -\mathrm{grad}\, \mathscr{E}(\rho), \mathrm{grad}\, f(\rho) \rangle_\rho + \tfrac{1}{2} \|\mathrm{grad}\, f(\rho)\|_\rho^2. \tag{27}$$

See [3,9] for detail and a history of such formulation.

To show the first two control examples can be written in (26), we take $\mathscr{E}$ to be (13); for the third example, however, we need to define a new functional:

$$\begin{aligned} \mathscr{E}(\rho) = \frac{1}{2} \int_{R^d} \rho(x) \log \rho(x)\, \mathrm{d}x + \int_{R^d} \Psi(x) \rho(\mathrm{d}x) \\ + \frac{1}{2} \int_{R^d \times R^d} \Phi(x - y) \rho(x) \rho(y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

if $\rho(\mathrm{d}x) = \rho(x)\, \mathrm{d}x$ has a Lebesgue density, and $\mathscr{E}(\rho) = +\infty$ if ρ does not have a Lebesgue density. The definition of the gradient in each case is better explained using geometric intuitions [3]. The idea is that there are hidden conserved quantities in the control dynamic, and we need a notion of gradient to respect such constrain. Specifically, the controlled Allen–Cahn equation does not have any conserved quantity, the controlled Cahn–Hilliard equation satisfies $\int_{\mathcal{O}} \rho(t,x)\,dx = \text{constant}$, and the controlled Fokker–Planck equation satisfies that $\int_x \rho(t,dx) = 1$ and $\rho(t,dx) \geqslant 0$. Let $\delta f/\delta\rho$ denote the unconstrained first variational derivative of *f*, then computationally, grad $f(\rho) = \delta f/\delta\rho$ in the Allen–Cahn case; grad $f(\rho) = -\nabla\cdot\nabla(\delta f/\delta\rho)$ in the Cahn–Hilliard case; and grad $f(\rho) = -\nabla\cdot(\rho\nabla(\delta f/\delta\rho))$ in the Fokker–Planck case. The $\langle\cdot,\cdot\rangle_\rho$ is respectively the inner product of the spaces $L^2(\mathcal{O})$, $H^{-1}(\mathcal{O})$ and $H^{-1}_{\rho(t)}(R^d)$. These are the tangent spaces where the control lives.

From (27), we obtain an important inequality regarding the free energy functional

$$H\mathscr{E}(\rho) = -\tfrac{1}{2}\|\text{grad}\,\mathscr{E}(\rho)\|^2_\rho \equiv -\tfrac{1}{2}\imath(\rho) \leqslant 0. \tag{28}$$

In the Allen–Cahn case, $\imath(\rho) = \|-\Delta\rho + F'(\rho)\|^2$; in the Cahn–Hilliard case, $\imath(\rho) = \|\nabla(-\Delta\rho + F'(\rho)\|^2$; and in the Fokker–Planck case, $\imath(\rho) = \int_{R^d} |\tfrac{1}{2}\nabla\rho/\rho + \nabla\Psi + \nabla(\rho * \Phi)|^2 \rho(dx)$ is the Fisher information functional. Consider the most likely trajectory of the rescaled stochastic systems. It is the solution to (26) in the absence of control (i.e. $u = 0$). In this case, $\imath(\rho(t))$ gives the rate of dissipation for free energy at *t*. There is a rich literature studying the speed of convergence of $\rho(t)$ to equilibrium, using $\mathscr{E}$, $\imath$ and inequalities relating them (Sobolev, log-Sobolev and the HWI-type mass transport inequalities).

In finite-dimensional settings, gradient operator is monotone and gradient flows has a contraction property. Complete analogy can be proved for the infinite-dimensional case (26) as well: Let $\rho(\cdot)$ and $\gamma(\cdot)$ be two solutions with the same control $u(\cdot)$ but different initial values $\rho(0) \neq \gamma(0)$. Then for each $T > 0$,

$$\exists C_T = C_T(\Psi, \Phi) \in R, \quad \text{such that } d(\rho(t),\gamma(t)) \leqslant e^{C_T t} d(\rho(0),\gamma(0)). \tag{29}$$

The metric *d* is chosen to be compatible with the inner product $\langle\cdot,\cdot\rangle_\rho$. For the cases (1)–(3), such *d* is respectively the metric induced by the L^2 norm, the H^{-1} norm, and the order-2-Wasserstein metric. See [3] for motivation and the history of such derivation. We note that under mild assumptions, $\mathscr{E}$ has compact level sets in the metric space (E, d). Using the *free energy inequality* (28) and using the contraction property (29), we can prove a very useful property for the control problems: Let $T > 0$. Suppose $(\rho_n(\cdot), u_n(\cdot))$ satisfy the control equations and are such that $a \equiv \sup_n \int_0^T L(\rho_n(s), u_n(s))\,ds < +\infty$, and suppose that $\rho_n(0) \in K_0$ for some compact set K_0, then

$$\exists \text{ compact } K_1 = K_1(a, T, K_0), \quad \text{such that } \rho_n(t) \in K_1, \quad 0 \leqslant t \leqslant T. \tag{30}$$

This is of course a limit version of the exponential compact containment condition in (23). They are very closely connected. Indeed, in the case of stochastic Ginzburg–Landau models, (23) can be proved by observing a sequence version of (28) and (29) holds: First, if $\mathscr{E}(\rho_n(0)) < +\infty$,

$$\exp\left\{\mathscr{E}(\rho_n(t)) - \mathscr{E}(\rho_n(0)) - \int_0^t H_n\mathscr{E}(\rho_n(s))\,ds\right\} = \text{super-martingale.}$$

Second, for each $T > 0$, there exists $C_T \in R$ such that $d(\rho_n(t),\gamma_n(t)) \leqslant e^{C_T t} d(\rho_n(0),\gamma_n(0))$ for all $n = 1, 2, \ldots,$ where $\rho_n(\cdot)$ and $\gamma_n(\cdot)$ are two solutions of the same stochastic equation in each models in Section 2 driven by the same noise term.

We discuss how to use the above properties to study the solution of (11) next.

4.2. Uniqueness of viscosity solution—the comparison principle

Current techniques for proving the comparison principle all rely upon the maximum principle satisfied by H, and upon approximating sub- and super-solutions by "smooth" test functions (Chapter 9 of [4]). For equations of the type here, Lions [7] formulated a key condition which can be informally stated as follows: there exists a distance function d and a $\omega : [0, +\infty) \mapsto R$ which is continuous at 0, $\omega(0) = 0$, such that

$$H(\mu d^2(\cdot,\gamma))(\rho) - H(-\mu d^2(\rho,\cdot))(\gamma) \leqslant \omega(\mu d^2(\rho,\gamma)), \quad \mu > 0. \tag{31}$$

It is interesting to observe that, for the optimal controlled equations (1)–(3), we can choose d exactly as those previous ones which give (29). In each case, (31) follows either from Sobolev inequalities or the mass transport version. See [3,4]. From (31), we see that it is important to include $\pm\mu d^2(\cdot,\gamma)$ as test functions for H. We can either include them in the beginning when computing the limit of H_ns, or use operator extension techniques [4].

Let $\overline{f}$ be a subsolution to (11) and f_0 be a test function. Available techniques for rigorous proof of the comparison principle also require us to know that $\overline{f} - f_0$ attains the maximum at some point ρ. When E is non-locally-compact, in the absence of strong assumptions on $\overline{f}$ and f_0, this poses a serious technical problem. Similar situation also occurs for super-solutions. To address the problem, [1] proposes to build elaborate test functions by using Ekeland's perturbed optimization principle. The approach requires E be a Banach space satisfying the Radon–Nikodym property, which exclude our third and fourth control examples (3) and (4).

We explore a different perturbation argument utilizing (28). We assume without loss of generality that $\mathscr{E} \geqslant 0$. For each test function f and $0 < \kappa < 1$, let f_κ be defined by $f = (1-\kappa) f_\kappa + \kappa\mathscr{E}$. By convexity in $H(\rho,\cdot)$, $Hf(\rho) \leqslant (1-\kappa)Hf_\kappa(\rho) - \frac{1}{2}\kappa\iota(\rho) \leqslant (1-\kappa)Hf_\kappa(\rho)$. Hence if $\overline{f}$ is a subsolution, then at least formally $\overline{f}_\kappa$ is a subsolution for a perturbed equation $(1-\kappa)(I-\alpha H)\overline{f}_\kappa \leqslant (I-\alpha H)\overline{f} \leqslant h$. By the compact level set property of $\mathscr{E}$, if $\overline{f}$ is bounded upper semicontinuous and f_0 is lower semicontinuous and bounded from below, then $\overline{f}_\kappa - f_0$ attains a maximum. Similarly perturbing the super-solution $\underline{f}_\kappa$, we can then prove the comparison principle for the perturbed equation $\overline{f}_\kappa \leqslant \underline{f}_\kappa$. Using the density of $\{\rho : \mathscr{E}(\rho) < +\infty\}$ in E, if $\overline{f}$ and $\underline{f}$ are continuous, then we have $\overline{f} \leqslant \underline{f}$. Continuities of $\overline{f}$, $\underline{f}$ can usually be recovered from the structure of H [4].

4.3. Convergence of solutions—the Barles–Perthame procedure

Let $C_b(E) \ni h_n \to h \in C_b(E)$ and $H_n \to H$. We next discuss the convergence of f_n in (25) to the f in (11). We assume the comparison principle for (11) is already verified, therefore if f is a bounded continuous viscosity solution, it is the unique one.

Let $\overline{f}(\rho) = \sup\{\limsup_{n\to+\infty} f_n(\rho_n) : \rho_n \to \rho\}$, and $\underline{f}(\rho) = \inf\{\liminf_{n\to+\infty} f_n(\rho_n) : \rho_n \to \rho\}$. We show that the $\overline{f}$ and $\underline{f}$ are, respectively, bounded upper semicontinuous sub-solution and lower semicontinuous super-solution to (11) in the viscosity sense. Suppose that this is true and that the comparison principle holds, then we can also conclude $f_n \to f$ in the sense of uniform convergence on compact sets. By the explicit probabilistic representation of V_n in (21), a solution f_n for (25) can usually be constructed. Indeed, replacing A_n by its Yosida approximation, [4] showed that an asymptotically equivalent equation to (25) always have a solution in the classical sense. Therefore the convergence of $f_n \to f$ also gives a constructive proof of the existence of solution for (11).

Barles and Perthame first introduced the above mentioned procedure in the case when E is finite dimensional. We mention two key steps in the generalization provided by Feng and Kurtz [4]. First, convergence of operators $H_n \to H$ can be formulated in terms of graph convergence: $\exists f_n$, $f_n \to f$ and $H_n f_n \to Hf$. Such generality is not needed for the examples here but becomes absolutely essential when dealing with multi-scale problems. Function convergence $f_n \to f$ is in the sense of a modified version of uniform convergence on compacts. The Barles–Perthame procedure is compatible with such notion of convergence, but without other information, the usual estimates needed to derive convergence of f_ns requires uniform convergence on the whole space. Such a gap can be bridged if we have the exponential compact containment condition (23). Let K_0 be any given compact set, $h_n^1, h_n^2 \in C_b(E)$ and $\varepsilon > 0$. Specifically, using the dynamic programming principle, such condition implies $\sup_{\rho \in K_0} ((I - \alpha H_n)^{-1} h_n^1(\rho) - (I - \alpha H_n)^{-1} h_n^2(\rho)) \leqslant \varepsilon + \sup_{\rho \in K_1(\varepsilon, T, K_0)} (h_n^1(\rho) - h_n^2(\rho))$ where T is a large number depending on $\varepsilon > 0$.

Similar ideas used in such generalization can also be applied to show that the value function f (12) is a viscosity solution to (11). We need a priori estimate (30) in that case.

5. The case of controlled Vlasov–Fokker–Planck equation

Neither the large deviation principle for (18) nor the comparison principle for Hamilton–Jacobi equation associated with it has been rigorously proved yet.

The controlled Vlasov–Fokker–Planck equation (4) may not be a controlled gradient flow, however, it is as a controlled J-gradient flow in the following sense. We can find a free energy functional $\mathscr{E}$, and a special matrix J. By defining $J\,\mathrm{grad}$ appropriately, it becomes $\dot{\rho} = -J\,\mathrm{grad}\,\mathscr{E}(\rho) + u$ (personal communication from W. Gangbo. See also Chapter 8.3.2. of [9]). At least formally, all arguments in this article can be modified to go through. In particular, a free energy inequality similar to (28) still holds and the contraction property (29) is valid if d is taken to be the order-2-Wasserstein metric.

References

[1] M.G. Crandall, P.L. Lions, Hamilton–Jacobi equations in infinite dimensions, Part I–VII, J. Funct. Anal. 62 (3) (1985) 379–396;
M.G. Crandall, P.L. Lions, Hamilton–Jacobi equations in infinite dimensions, Part I–VII, J. Funct. Anal. 65 (3) (1986) 368–405;
M.G. Crandall, P.L. Lions, Hamilton–Jacobi equations in infinite dimensions, Part I–VII, J. Funct. Anal. 68 (2) (1986) 214–247;
M.G. Crandall, P.L. Lions, Hamilton–Jacobi equations in infinite dimensions, Part I–VII, J. Funct. Anal. 90 (2) (1990) 237–283;
M.G. Crandall, P.L. Lions, Hamilton–Jacobi equations in infinite dimensions, Part I–VII, J. Funct. Anal. 97 (2) (1991) 417–465;
M.G. Crandall, P.L. Lions, Hamilton–Jacobi equations in infinite dimensions, Part I–VII, J. Funct. Anal. 125 (11) (1994) 111–148.

[2] D.A. Dawson, J. Gärtner, Large deviation from the McKean–Vlasov limit for weakly interacting diffusions, Stochastics 20 (1987) 247–308.
[3] J. Feng, M. Katsoulakis, A Hamilton–Jacobi theory for controlled gradient flows in infinite dimensions, 2004, under review.
[4] J. Feng, T.G. Kurtz, Large deviations of stochastic processes, 2000, under revision.
[5] W.H. Fleming, H.M. Soner, Controlled Markov Processes and Viscosity Solutions, Springer, New York, 1991.
[6] M.I. Freidlin, A.D. Wentzell, Random Perturbations of Dynamical Systems, Springer, New York, 1983.
[7] P.L. Lions, Some properties of the viscosity semigroups for Hamilton–Jacobi equations, in: J.K. Hale, P. Martinez-Amores (Eds.), Nonlinear Differential Equations, Research Notes in Mathematics, vol. 132, Pitman, London, 1988.
[8] H. Spohn, Large Scale Dynamics of Interacting Particles, Texts and Monographs in Physics, Springer, New York.
[9] C. Villani, Topics in Mass Transportation, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.