ARMA manuscript No.
(will be inserted by the editor)

A Comparison Principle for Hamilton-Jacobi Equations Related to Controlled Gradient Flows in Infinite Dimensions

Jin Feng, Markos Katsoulakis

Abstract

We develop new comparison principles for viscosity solutions of Hamilton-Jacobi equations associated with controlled gradient flows in function spaces as well as the space of probability measures. Our examples are optimal control of Ginzburg-Landau and Fokker-Planck equations. They arise in limit considerations of externally forced nonequilibrium statistical mechanics models, or through the large deviation principle for interacting particle systems. Our approach is based on two key ingredients: an appropriate choice of geometric structure defining the gradient flow, and a free energy inequality resulting from such gradient flow structure. The approach allows us to handle Hamiltonians with singular state dependency in the nonlinear term, as well as Hamiltonians with a state space which does not satisfy the Radon-Nikodym property. In the case where the state space is a Hilbert

space, the method simplifies existing theories by avoiding the perturbed optimization principle.

1. Introduction

In this paper, we develop new comparison principles for viscosity solutions of Hamilton-Jacobi equations associated with controlled gradient flows in function spaces as well as the space of probability measures. The examples considered here are motivated by the optimal control of conservative as well as non-conservative Ginzburg-Landau models, and equations of Fokker-Planck type. Theses controlled PDEs arise as limits of externally forced nonequilibrium statistical mechanics models, or through the large deviation principle for interacting particle systems. Our approach is based on two key ingredients: an appropriate choice of geometric structure defining the gradient flow, and a free energy inequality resulting from the gradient flow structure. As a result, we can handle Hamiltonians with singular state dependency in the nonlinear term, as well as Hamiltonians with a state space which does not satisfy the Radon-Nikodym property. In the case where the state space is a Hilbert space, the method simplifies existing theories by avoiding the perturbed optimization principle.

The theory of Hamilton-Jacobi equations in infinite dimensions was first introduced and developed in a series of publications by Crandall and Lions [7]. Their analysis relies on the introduction of a suitable weak (*i.e. viscosity*) solution framework, guaranteeing existence and uniqueness of solutions under regularity assumptions. Here, we consider first order Hamilton-Jacobi equations of the type

$$f(\rho) - \alpha H\Big(\rho, \mathrm{grad} f(\rho)\Big) = h(\rho), \quad \rho \in E, \tag{1}$$

where $\alpha > 0$ is some constant, (E, r) is a metric space, ρ is a typical element in E and $\mathrm{grad} f$ is a suitably defined notion of gradient for a real valued function f defined on E. The above equation can be considered more generally from a functional analytic point of view. Using the operator notation $Hf(\rho) = H(\rho, \mathrm{grad} f(\rho))$, (1) becomes

$$(I - \alpha H)f = h, \tag{2}$$

where I is the identity map. Unlike the finite dimensional viscosity solution theory [9], available theories for general infinite dimensional cases (e.g. [7], [35] and [8]) require restrictive regularity assumptions on the solution f and the state space E. On the other hand, numerous important applications can be cast into such an infinite dimensional Hamilton-Jacobi theory framework. These applications range from the optimal control of PDEs to the proof of the probabilistic large deviation principle for general Markov processes.

This article develops applicable conditions for the Hamilton-Jacobi theory for a special class of optimally controlled PDEs (i.e. Examples 1, 2, and 3). These controlled PDEs can all be represented as an infinite dimensional gradient flow for a *free energy* functional – see (6). In the absence of control, some of these problems have been studied by [29], [23]. The controlled problems arise naturally from microscopic statistical mechanics considerations of interacting particle systems [20]. In this context, the control corresponds to an external field to the underlying microscopic system. Furthermore, similar control problems also arise in derivation of PDEs as scaling limit of stochastic interacting particle systems using the large deviation theory (Appendix B). In this case, the controlled problem is associated with the large deviation principle, and it gives us information about convergence to the most probable trajectory of the system as well as the rate of deviation from atypical trajectories. The connection between large deviation and control theory was first pointed out by [15], in the context of finite dimensional Markov processes, using a logarithmic transformation. See also [16] and the book [17]. Recently, a framework for handling general metric state spaces is proposed in [18]. A key step in this approach is the comparison principle for (2), which we prove in this paper for the examples in Appendix B. The topic of large deviation for interacting particle systems, especially under the hydrodynamic scaling, has been a primary point of focus in the recent probability literature. See for example [24], [36], and the review book [25]. In these works, the large deviation principles are established through pure probabilistic techniques, such as the super-exponential estimates for handling the multi-scale averaging effects, the Girsanov transformation for estimating deviation rate about regular trajectories, and an analytic approximation argument for transferring the estimates about regular trajectories to irregular ones. We expect, more generally, these problems can also be treated using multi-scale convergence of Hamilton-Jacobi equations for controlled

PDEs, and by establishing the corresponding comparison principles. When the comparison principle becomes available, the Hamilton-Jacobi equation approach provides an alternative method, which can be useful when the approximation step in the pure probabilistic approach is difficult to establish.

We next discuss the connections between our results and the existing viscosity solution theories. The definition of *viscosity solution* introduced by Crandall and Lions [6], in the setting of $E \subset R^d$, can be viewed as a way of extending H while preserving its dissipative property. The ideas involved in the generalization to infinite dimensions are along the same lines. However, the lack of local compactness on E creates substantial technical obstacles. The current PDE literature focuses mainly on the case where E is a Hilbert space. Crandall and Lions [7],[8] as well as Tataru [35] offered several definitions of viscosity solutions which can be consolidated in this case as follows:

Let $\alpha > 0$, (E, r) be a metric space, $M(E)$ be the space of measurable functions on E taking values in R, and $M(E, \overline{R})$ the space of measurable functions on E which take values in the extended real line $\overline{R} = R \cup \{\pm\infty\}$. By an operator H, we mean a map in $M(E, \overline{R})$ with its domain $\mathcal{D}(H) \subset M(E, \overline{R})$ and its range $\mathcal{R}(H) \subset M(E, \overline{R})$.

Definition 1. $\overline{f} \in M(E)$ is a *subsolution* of (2), denoted

$$(I - \alpha H)\overline{f} \leq h,$$

if and only if for each $f_0 \in \mathcal{D}(H)$ which is bounded from below, there exists $\rho_0 \in E$ such that

$$(\overline{f} - f_0)(\rho_0) = \sup_{\rho \in E}(\overline{f} - f_0)(\rho) \tag{3}$$

and

$$(\overline{f} - h)(\rho_0) \leq \alpha(Hf_0)^*(\rho_0), \tag{4}$$

where $(Hf_0)^*$ is the upper semicontinuous smoothing of Hf_0.

$\underline{f} \in M(E)$ is a *supersolution* of (2), denoted

$$(I - \alpha H)\underline{f} \geq h,$$

if and only if for each $f_1 \in \mathcal{D}(H)$ which is bounded from above, there exists $\rho_1 \in E$ such that

$$(f_1 - \underline{f})(\rho_1) = \sup_{\rho \in E}(f_1 - \underline{f})(\rho) \tag{5}$$

and

$$(\underline{f} - h)(\rho_1) \geq \alpha(Hf_1)_*(\rho_1),$$

where $(Hf_1)_*$ is the lower semicontinuous smoothing of Hf_1.

If f is both a subsolution and a supersolution, it is called a *viscosity solution.*

If for every upper-semicontinuous subsolution $\overline{f}$ and every lower semicontinuous supersolution $\underline{f}$, we have $\overline{f} \leq \underline{f}$, we say that *the comparison principle* holds.

In addition, f is a *strong subsolution* of (2) if for every ρ_0 satisfying (3), (4) holds. The concept of *strong supersolution* and *strong viscosity solution* are defined similarly.

The relationship between *strong* sub- (super-) solution and sub- (super-) solution is analogous to that between strong dissipativity and dissipativity of operators in Banach space $B(E)$. See Sato [31] and Sinestrari [32].

It is important to clarify a number of points regarding the above definition. First, one needs to exercise caution when choosing $\mathcal{D}(H)$. Having too many test functions in $\mathcal{D}(H)$ will make the existence proof harder, while having too few of them will make the uniqueness impossible. Therefore, it is best to keep only those test functions which are absolutely essential in uniqueness proofs. For instance, in the heuristic discussions motivating origins of Examples 1,2 and 3 below, we will first consider "obvious" test functions (13) which are very smooth. They are useful to reveal a hidden controlled gradient flow structure of H (see (10)) at an informal level. However, in later rigorous development for a uniqueness theory through the comparison principle, we will only need functions closely related to (40) and (41) in the domain, and discard the smooth test functions (13). See Conditions 1, 4 regarding general results in Theorems 3 and 5, and see (54) for $\mathcal{D}(H)$ for the examples treated in Theorems 6, 7 and 8 in Section 3. Test functions (13) are un-desirable in the rigorous development because they cause problem in applying Definition 1 – there is no a priori guarantee that the maximum and minimum points ρ_0, ρ_1 ever exist (see the next paragraph). There are ways of further extending definition of viscosity solution so that these smooth test functions can be used to play important roles in existence and convergence theory. We refer to Feng and Kurtz [18] for details on these issues as well as on how different notions of solutions relate to each other. Finally, we point out that when considering subsolution, the above definition excludes test functions which are not bounded from below. Therefore (3) and (4) are only applied to f_0s in (40), excluding f_1s in (41). Similarly, when considering supersolution, we only use the f_1s, excluding the f_0s.

While the supremum (or infimum) of any function can be approximated by a sequence of its values, it is generally not true that the extremal values can be attained in an infinite dimensional state space. Current proofs of the comparison principle for (2) rely on Definition 1 as the point of departure. Usually, by building test functions and by invoking variants of the *perturbed optimization principle* (Stegall [33] or Ekeland [13]), the extremal points in Definition 1 can be produced. See [7,8,35] for instance. Stegall's principle works on Banach state space which is Radon-Nikodym (e.g. Hilbert spaces). Ekeland's principle works in general metric space. However, the need of a norm-like differentiable functions for the regularity of H eventually limit application of the Ekeland principle approach to non-Banach space or Banach space without Radon-Nikodym property. When handling controlled PDEs, we could encounter state space which is L^1 or even the space of probability measures (with weak convergence of probability measure topology) as in

Example 3. L^1 space do not satisfy the Radon-Nikodym property, and the space of probability measure with the weak topology is not even a Banach space. In addition, the Hamiltonian H could contain a singularity in coefficients as in (27). In the case of (27), even if we view probability with Lebesgue density as a subspace of L^1, the Lipschitz in coefficient type condition in available literature is not satisfied. Both these difficulties cause the available comparison results in [7], [8], [35] to not apply. In this paper, we avoid invoking the perturbed optimization method. We carefully define the notion of gradient for functions and select a special class of semicontinuous test functions (40) and (41), by exploring the underlying geometry of the control problem (see Section 1.2.2) and by noting a free energy inequality (28) which holds for gradient systems (see Section 1.2.1). The use of free energy function as part of the test functions localize everything to compact set where the maximum/minimum points in Definition 1 always exists. This helps us consider problems in more general state spaces. We would like to mention that, in a Hilbert space setting, the use of similar type special test functions first appeared in Ishii [22]. Crandall and Lions, Cannarsa and Tessitore [4], Gozzi and Swiech [21] also have used similar techniques before.

Finally, in the case of Examples 1 and 2, the state spaces are Hilbert spaces and their Hamiltonians are non-singular. With appropriate extensions, existing theory [7,8,35] also apply. However, in the case of Example 3, the Hamiltonian contains singularity, and it is natural to consider the state space as the space of probability measures with the weak topology rather than a Banach space. The comparison principle in this case is no longer covered by any existing theory.

1.1. Examples of abstract controlled gradient flows in infinite dimensions

The discussion in this section is heuristic. Rigorous results will be developed in later sections. Our presentation here is inspired by Otto [29].

First of all, we introduce an abstract *controlled* gradient flow. Let E be a Riemannian manifold, and let the tangent space $T_\rho E$ at $\rho \in E$ be modeled by some Hilbert space with metric tensor $g_\rho(\cdot, \cdot)$ which is bilinear. We consider controlled gradient flows on E:

$$\dot{\rho} = -\mathrm{grad}\mathcal{E} + u(t), \tag{6}$$

where $\mathcal{E} : E \to (-\infty, +\infty]$ is a functional called the *free energy*, and $u(t) \in T_{\rho(t)}E$ is a control. For any function $f : E \to \overline{R}$, we use $df(\rho_0)$ to denote the differential of f at ρ_0, which belongs to the cotangent vector field; $\mathrm{grad}f(\rho_0)$ denotes the gradient at ρ_0, which belongs to the tangent vector field $T_{\rho_0}E$. They are related through the identity

$$\Big(df(\rho_0)\Big)v \equiv \lim_{t\to 0+} \frac{f(\rho^v(t)) - f(\rho_0)}{t} = g_{\rho_0}(v, \mathrm{grad}f(\rho_0)) \tag{7}$$

for a large dense class of $v \in T_{\rho_0}E$, where evolution $\rho^v(t)$ is defined by

$$\dot{\rho^v}(t) = v(t), \quad \rho^v(0) = \rho_0.$$

We consider optimal control of (6) with cost function

$$L(\rho, u) = \frac{1}{2} g_\rho(u, u). \tag{8}$$

That is, we consider the maximization problem

$$f(\rho_0) = \sup\{\int_0^\infty e^{-t}\Big(h(\rho(t)) - L(\rho(t), u(t))\Big)dt : (\rho, u) \text{ satisfies (6) with } \rho(0) = \rho_0\}. \tag{9}$$

where $h \in C_b(E)$. The f_0 is called the *value function* of the optimal control problem.

Next, we define Hamiltonian operator

$$\begin{aligned} Hf(\rho) &= \sup_{u \in T_\rho E}\{g_\rho(-\text{grad}\mathcal{E}(\rho) + u, \text{grad}f(\rho)) - \frac{1}{2}g_\rho(u,u)\} \qquad (10) \\ &= g_\rho(-\text{grad}\mathcal{E}(\rho), \text{grad}f(\rho)) + \frac{1}{2}g_\rho(\text{grad}f(\rho), \text{grad}f(\rho)). \end{aligned}$$

By the *dynamic programming principle*, the functional f defined in (9) is formally a solution of (2), see [9] for a rigorous proof in the finite dimensional case, as well as [8] for some infinite dimensional settings.

We discuss three examples of controlled PDEs which are special cases of (6). The form of the control variable in (6), as well as the cost (8) are suggested by a microscopic statistical mechanics perspective, where the control is the external field in the free energy $\mathcal{E}$.

Throughout the paper, we denote $\langle p, u\rangle = \int u(x)p(x)dx$ or its linear extension if u is a Schwartz distribution and p is a smooth function.

Example 1. [Controlled Allen-Cahn equation] Consider the domain $\mathcal{O} = [0,1]^d \subset R^d$ with periodic boundary condition. Let F be a function on R satisfying $F(r) \geq c_1 r^2 + c_2$ where $c_1 > 0, c_2 \in R$, and $F^{''} \in C_b(R)$. We define the free energy

$$\mathcal{E}(\rho) = \begin{cases} \frac{1}{2}\int_{\mathcal{O}} |\nabla\rho(x)|^2 dx + \int_{\mathcal{O}} F(\rho(x))dx, & \text{when } \rho \in H^1(\mathcal{O}) \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

Let us take $E = L^2(\mathcal{O})$ with the norm topology $\|\cdot\|_{L^2(\mathcal{O})}$, then $T_\rho E = L^2(\mathcal{O})$ with metric tensor $g_\rho(\cdot,\cdot) = \langle\cdot,\cdot\rangle$, where $\langle\cdot,\cdot\rangle$ is the usual inner product in $L^2(\mathcal{O})$. It follows that $\mathcal{E}$ has compact level set in E. In Section 3.3, we rigorously define the gradient according to (7) for a large class of smooth vector fields v. See Definition 2. We show that

$$\text{grad}\mathcal{E}(\rho) = -\Delta\rho + F^{'}(\rho),$$

and

$$\text{grad}f(\rho) = \frac{\delta f}{\delta\rho} \equiv \sum_{k=1}^{m} \partial_k\varphi(\langle\rho,\xi_1\rangle,\ldots,\langle\rho,\xi_m\rangle)\xi_k, \tag{12}$$

for test functions of the form

$$f(\rho) = \varphi(\langle\rho,\xi_1\rangle,\ldots,\langle\rho,\xi_m\rangle) = \varphi(\langle\rho,\xi\rangle), \tag{13}$$

where $\xi = (\xi_1,\ldots,\xi_m) \in C^\infty(\mathcal{O}), \varphi \in C^2(R^m)$, and $m = 1,2,\ldots$.

In this case, the controlled gradient flow (6) becomes the controlled *Allen-Cahn* equation

$$\frac{\partial}{\partial t}\rho - \Delta\rho + F^{'}(\rho(t,x)) - u(t,x) = 0, \quad x \in \mathcal{O}, \tag{14}$$

furthermore L in (8) reduces to

$$L(\rho,u) = \frac{1}{2}\|u\|^2_{L^2(\mathcal{O})},$$

and the H in (10) becomes

$$Hf(\rho) = \langle \Delta\rho - F^{'}(\rho), \frac{\delta f}{\delta\rho} \rangle + \frac{1}{2} \| \frac{\delta f}{\delta \rho} \|^2_{L^2(\mathcal{O})}, \tag{15}$$

for f of the form (13).

Example 2. [Controlled Cahn-Hilliard equation] Let $\mathcal{O}, F, \mathcal{E}$ be defined as in Example 1. We consider

$$E = \{\rho \in L^2(\mathcal{O}) : \int_{\mathcal{O}} \rho(x) dx = 0\}$$

with the norm metric $\| \cdot \|_{L^2(\mathcal{O})}$, and define $H^{-1}(\mathcal{O})$ as the completion of E under the norm

$$\|\rho\|^2_{-1} \equiv \sup_{p \in C^\infty(\mathcal{O})} \{2\langle \rho, p\rangle - \int_{\mathcal{O}} |\nabla p|^2 dx\}. \tag{16}$$

$H^{-1}(\mathcal{O})$ is a Hilbert space with inner product

$$(u_1, u_2)_{-1} \equiv \frac{\|u_1 + u_2\|^2_{-1} - \|u_1 - u_2\|^2_{-1}}{4}.$$

It follows that $\mathcal{E}$ has compact level set in E. We set $T_\rho E = H^{-1}(\mathcal{O})$ with metric tensor

$$g_\rho(u_1, u_2) = (u_1, u_2)_{-1} = \int_{\mathcal{O}} \nabla p_1 \nabla p_2 dx = \langle p_1, u_2 \rangle,$$

for all $u_1, u_2 \in H^{-1}(\mathcal{O})$, where $-\Delta p_i = u_i$. In Section 3.4, following a rigorous definition of gradient (Definition 3) motivated from (7), we show that

$$\text{grad}\mathcal{E}(\rho) = -\Delta(-\Delta\rho + F^{'}(\rho)),$$

and

$$\text{grad} f(\rho) = -\Delta \frac{\delta f}{\delta \rho}$$

for f of the form (13).

Therefore, (6) becomes the controlled *Cahn-Hilliard* equation

$$\frac{\partial}{\partial t}\rho = \Delta\Big(- \Delta\rho + F^{'}(\rho(t,x))\Big) + u(t,x),$$

or

$$\frac{\partial}{\partial t}\rho = \Delta\Big(- \Delta\rho + F^{'}(\rho(t,x)) - p(t,x)\Big) \tag{17}$$

where $u = -\Delta p$. Here, either u or p can be viewed as the control variable. The particular way the control enters into the equation follows from statistical mechanics derivations of (17) and related equations, where p corresponds to an external field to an underlying microscopic particle system. See Langer [26], Giacomin and Lebowitz [20]. The cost L in (8) reduces to

$$L(\rho, u) = \frac{1}{2}\|u\|_{-1}^2. \tag{18}$$

The choice of such L is partly motivated by the form of large deviation action functional derived from stochastic Ginzburg-Landau equations of the type (B.101). The connection between large deviations of (B.101) and the controlled problem (17) and (18) is discussed in Example 5 in Appendix B. See Feng and Kurtz [18] for details.

Finally, for f of the form (13), the Hamiltonian (10) becomes

$$Hf(\rho) = \langle \Delta(-\Delta\rho + F^{'}(\rho)), \frac{\delta f}{\delta\rho}\rangle + \frac{1}{2}\|\nabla\frac{\delta f}{\delta\rho}\|^2_{L^2(\mathcal{O})}. \tag{19}$$

Example 3. [Controlled Fokker-Planck equation] Unlike the previous two examples, we consider $\mathcal{O} = R^d$. Suppose $0 \leq \Psi \in C^2(R^d)$ is semi-convex in the sense that there exists $L_\Psi \in R$ satisfying

$$(\nabla\Psi(x) - \nabla\Psi(y)) \cdot (x - y) \geq L_\Psi |x - y|^2. \tag{20}$$

We also assume that

$$\lim_{|x|\to+\infty} \frac{\Psi(x)}{|x|^2} = +\infty. \tag{21}$$

We consider the (order)-2-Wasserstein metric space

$$E = \{\rho \in \mathcal{P}(R^d) : \int_{R^d} |x|^2 \rho(dx) < +\infty\},$$

with a metric d defined by (53). (E, d) is complete and separable. See Proposition 7.4.2 of Ambrosio, Gigli and Savaré [1]. A remark immediately after the proof of that proposition also shows that (E, d) is not locally compact. See also Chapter 7 of Villani [37] for properties of this space.

We denote the *Gibbs measure*

$$\rho_\infty(dx) = Z^{-1} e^{-2\Psi(x)} dx \in \mathcal{P}(R^d), \tag{22}$$

where

$$Z = \int_{R^d} e^{-2\Psi(x)} dx.$$

We define the *free energy* functional

$$\mathcal{E}(\rho) = \frac{1}{2} \int_{R^d} \log \frac{d\rho}{d\rho_\infty} d\rho \geq 0, \quad \rho \in \mathcal{P}(R^d) \tag{23}$$

where $d\rho/d\rho_\infty$ denotes the Radon-Nikodym derivative and $\mathcal{E}(\rho) = +\infty$ if ρ does not have a Lebesgue density. Suppose $\rho(dx) = \rho(x)dx$ has a Lebesgue density and $\rho \in E$,

$$\mathcal{E}(\rho) = \frac{1}{2} \log Z + \frac{1}{2} \int_{R^d} \rho(x) \log \rho(x) dx + \int_{R^d} \Psi(x) \rho(x) dx.$$

Because of (21), $\mathcal{E}$ has compact level set in E (Lemma 2).

Let $\rho \in E$. We introduce weighted Sobolev space $H^{-1}_\rho(R^d)$ as in Appendix A.2. This is a Hilbert space and its inner product and norm are denoted respectively by $(\cdot,\cdot)_{-1,\rho}$ and $\|\cdot\|_{-1,\rho}$. We set $T_\rho E = H^{-1}_\rho(R^d)$ with metric tensor

$$g_\rho(u_1, u_2) = (u_1, u_2)_{-1,\rho} = \int_{R^d} \nabla p_1(x) \nabla p_2(x) \rho(dx) = \langle p_1, u_2 \rangle \tag{24}$$

for every $u_1, u_2 \in T_\rho E$ with representation $-\nabla \cdot (\rho \nabla p_i) = u_i$ (see Lemma 4). In Section 3.5, following the formal suggestion of (7), we define gradient by Definition 4, then (see (63))

$$\mathrm{grad}\mathcal{E}(\rho) = -\Big(\frac{1}{2}\Delta\rho + \nabla \cdot (\rho \nabla \Psi)\Big)$$

for $\mathcal{E}(\rho) < +\infty$, and for f of the form (13),

$$\mathrm{grad} f(\rho) = -\nabla \cdot (\rho \nabla \frac{\delta f}{\delta \rho}).$$

Therefore, (6) yields the controlled *Fokker-Planck* equation

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

or

$$\frac{\partial}{\partial t}\rho = \frac{1}{2}\Delta\rho + \nabla \cdot \Big(\rho \nabla (\Psi + p)\Big) \tag{25}$$

where $u = -\nabla \cdot (\rho \nabla p)$. As in Example 2, we can view either u or p as the control. Furthermore, (8) becomes

$$L(\rho, u) = \frac{1}{2}\int_{R^d} |\nabla p(x)|^2 \rho(dx) = \frac{1}{2}\|u\|^2_{-1,\rho}.$$

Again, choice for such L is motivated by large deviation considerations of stochastic interacting particle systems. See example 6 in Appendix B. and detailed analysis in [18].

Finally, for f of the form (13) with $\xi_k \in C_c^\infty(R^d)$

$$Hf(\rho) = \int_{R^d} (\frac{1}{2}\Delta \frac{\delta f}{\delta \rho} - \nabla \Psi \cdot \nabla \frac{\delta f}{\delta \rho})\rho(dx) + \frac{1}{2}\int_{R^d} |\nabla \frac{\delta f}{\delta \rho}|^2 \rho(dx), \quad (26)$$

where

$$\frac{\delta f}{\delta \rho} = \sum_{k=1}^m \partial_k \varphi(\langle \rho, \xi_1 \rangle, \ldots, \langle \rho, \xi_m \rangle)\xi_k.$$

An additional feature of the Hamiltonian defined by (26) is that it is strongly singular in the following sense. Suppose $\rho(dx) = \rho(x)dx$ has a Lebesgue density, then

$$\begin{aligned} Hf(\rho) = \int_{R^d} (\Delta \frac{\delta f}{\delta \rho} - \nabla \Psi \cdot \nabla \frac{\delta f}{\delta \rho})(x)\sigma^2(\rho(x))dx \quad (27) \\ + \frac{1}{2}\int_{R^d} |\sigma(\rho(x))\nabla \frac{\delta f}{\delta \rho}(x)|^2 dx, \end{aligned}$$

where the state dependent coefficient term $\sigma(r) = \sqrt{r}$ is non-Lipschitz. Available comparison techniques in both finite dimensional [9] and Banach space settings [7,8,35] require σ to be Lipschitz continuous. Here, we prove a comparison principle for (26) in the presence of such a singularity, by exploiting the formal Riemannian structure (24) and properties given by the mass transport theory. See the discussion prior to Lemma 1 in Section 3.5.

1.2. Main ideas and sketch of the proofs

As discussed above, there are two main difficulties in studying the Hamilton-Jacobi equation in (2) with the H given by (10): first, the lack of local compactness in some infinite dimensional situations such as those in Examples 1, 2 and 3, and second, the presence of a strong singularity in the Hamiltonian, such as that in Example 3.

To address the first issue, we build test functions by exploring compact level set property of the free energy functionals, and by using the gradient flow structure of (6). We avoid the use of perturbed optimization principle arguments. The second issue is handled by introducing a component of the test function which exploits the Riemannian geometric structure of the state space. Below, we outline these two ideas.

1.2.1. Free energy and comparison principles A common feature for all the above examples is that, $\mathcal{E}(\rho)$ is nonnegative, lower semicontinuous with compact level sets, and at least formally

$$H\mathcal{E}(\rho) = -\frac{1}{2} g_\rho(\mathrm{grad}\mathcal{E}(\rho), \mathrm{grad}\mathcal{E}(\rho)) \le 0. \tag{28}$$

Furthermore, if $\overline{f}$ is a subsolution of (2), then heuristically we have

$$\overline{f}(\rho) - \alpha H(\rho, \mathrm{grad}\overline{f}(\rho)) \le h(\rho).$$

We can now define a perturbation $\overline{f}_\kappa$ of the subsolution by

$$\overline{f} = (1-\kappa)\overline{f}_\kappa + \kappa\mathcal{E}, \quad 0 < \kappa < 1.$$

By the convexity of H in the argument $\mathrm{grad} f$ we have

$$(1-\kappa)\overline{f}_\kappa \le \overline{f} \le \alpha H\overline{f} + h \le (1-\kappa)\alpha H\overline{f}_\kappa + \kappa\alpha H\mathcal{E} + h, \tag{29}$$

which in turn implies (due to (28)) that $\overline{f}_\kappa$ is a subsolution to the perturbed equation

$$(I - \alpha H)\overline{f}_\kappa \le (1-\kappa)^{-1} h. \tag{30}$$

The case of a supersolution $\underline{f}$ can be handled similarly. Comparison principle for the perturbed equation (29) is much easier to study than (2): The compact level set assumption on $\mathcal{E}$, and hence on $\overline{f}_\kappa$, guarantees the existence of a minimizing/maximizing point ρ_0 in Definition 1. In this case the comparison principle for the perturbed equations follows similarly to the finite dimensional case [9], allowing us to conclude $\overline{f}_\kappa \le \underline{f}_\kappa + o(\kappa)$. Then if the finite energy states $\{\rho : \mathcal{E}(\rho) < +\infty\}$ are dense in E, and if $\overline{f}, \underline{f}$ are continuous, we can pass ϵ to zero and obtain the comparison $\overline{f} \le \underline{f}$. The perturbed optimization principle is absent throughout such argument.

We formally consider a situation where a test function $\hat{f}_0$ is bounded and $H\hat{f}_0$ is bounded. We observe that, if $\overline{f}$ is bounded, letting ρ_0 attains the supreme of $\overline{f}_\kappa - \hat{f}_0$, then from (30) we expect $H\overline{f}_\kappa(\rho_0) \le H\hat{f}_0(\rho_0)$ (the subsolution property (4) is essentially a statement of this kind). Combined with (29), then $H\mathcal{E}(\rho_0) > -\infty$, giving important *a priori* regularity estimate on ρ_0. For instance, in Example 2, $H\mathcal{E}(\rho) > -\infty$ implies

$$\|\nabla(-\Delta\rho + F^{'}(\rho))\|_{L^2(\mathcal{O})} < +\infty,$$

hence $\rho \in H_3(\mathcal{O})$; and in Example 3, $H\mathcal{E}(\rho) > -\infty$ gives

$$I(\rho) = \int_{\mathcal{O}} \frac{|\nabla \frac{d\rho}{d\rho_\infty}|^2}{\frac{d\rho}{d\rho_\infty}} d\rho_\infty < +\infty.$$

I is known as the Fisher information, and plays an important role in mass transport inequalities such as the log-Sobolev inequalities. See [37], Appendix A. See also the Appendix of [18].

In rigorous formulation of the above argument, we would like to transfer the perturbation (the $+\kappa\mathcal{E}$ term) of subsolution $\overline{f}$ to that of test function $\hat{f}_0$, making the new test functions f_0 unbounded from above (for the subsolution case). Statements such as (30) will then be reformulated using the new unbounded test functions f_0 (instead the unbounded subsolution f_κ).

1.2.2. Distance functions and comparison principles In order to prove the comparison principle, standard viscosity solution literature requires the following regularity on the Hamiltonian: for each $m > 0$,

$$\Big(Hmd^2(\cdot,\gamma)\Big)(\rho) - \Big(H(-md^2(\rho,\cdot))\Big)(\gamma) \le \omega(md^2(\rho,\gamma) + d^2(\rho,\gamma)), \quad (31)$$

where function $\omega : [0,\infty) \to \overline{R}$ is continuous at 0, $\omega(0) = 0$ and d is a distance function on state space E. Such requirement is discussed in [9] for finite dimensional state spaces with d the Euclidean norm, and [7] for Banach state spaces with d the norm of the space. The need for a general distance function d was noted in Part I of [7] and in [27]. In [27], Lions commented that the general d is motivated for two reasons: first, in order to have a condition valid for manifolds; second, to have a condition invariant under change of variables. These papers then immediately require either $c_0|x-y| \le d(x,y)$ in the finite dimensional case or $c_0\|\rho-\gamma\| \le d(\rho,\gamma) \le c_1\|\rho-\gamma\|$ for the norm of a Banach space in the infinite dimensional case, for some $0 < c_0 < c_1$. Therefore, the results of these papers cannot be applied to singular Hamiltonians such as Example 26 or even its finite dimensional analog. Here, we refine (31) and claim that for Hamiltonians of the form (10), with or without singularity, it is good enough to choose

$$d^2(\rho_0, \gamma_0) = \inf\{\int_0^1 g_{\rho(r)}(\dot\rho(r), \dot\rho(r))dr : \rho(t) \in E, 0 \le t \le 1, \qquad (32)$$
$$\rho(0) = \rho_0, \rho(1) = \gamma_0\}.$$

Such d turns out to be the L^2 norm for Example 1, the $H^{-1}(\mathcal{O})$ norm (16) for Example 2, and the order-2-Wasserstein metric (53) for Examples 3.

We justify this claim by using the decomposition

$$H = H_l + H_q$$

where $H_l f = g_\rho(-\text{grad}\mathcal{E}, \text{grad} f)$ is the linear part, and $H_q f = \frac{1}{2} g_\rho(\text{grad} f, \text{grad} f)$ is the nonlinear quadratic part. A formal solution to the Cauchy problem

$$\frac{\partial}{\partial t} u(t, \rho_0) = H_q u(t, \rho_0), \quad u(0, \rho_0) = h(\rho_0) \qquad (33)$$

is given by the Hopf-Lax formula

$$u(t, \rho_0) = \sup\{h(\rho(t)) - \frac{1}{2}\int_0^t g_{\rho(s)}(\dot\rho(s), \dot\rho(s))ds\} \qquad (34)$$
$$= \sup_{\gamma_0 \in E}\{h(\gamma_0) - \frac{d^2(\rho_0, \gamma_0)}{2t}\}.$$

By setting

$$h(\rho) = -\infty I(\rho \ne \gamma) + 0 I(\rho = \gamma)$$

in (34), where $\gamma \in E$, we obtain $u(t, \rho) = -d^2(\rho, \gamma)/(2t), t > 0$. In particular, at time $t = 1$, we have

$$\frac{1}{2} d^2(\rho, \gamma) = \frac{\partial}{\partial t} u(t, \rho)\Big|_{t=1} = (H_q(-\frac{1}{2} d^2(\cdot, \gamma)))(\rho) = (H_q \frac{1}{2} d^2(\cdot, \gamma))(\rho).$$

Therefore

$$(H_q \frac{d^2(\cdot,\gamma)}{2})(\rho) = \frac{1}{2} d^2(\rho,\gamma) = (H_q \frac{d^2(\rho,\cdot)}{2})(\gamma). \tag{35}$$

Furthermore, if

$$d^2(\rho(t),\gamma(t)) \le e^{Lt} d^2(\rho(0),\gamma(0)), \quad \text{some } L \in R \tag{36}$$

holds for every

$$\dot\rho = -\text{grad}\mathcal{E}(\rho), \quad \dot\gamma = -\text{grad}\mathcal{E}(\gamma), \tag{37}$$

then

$$(H_l d^2(\cdot,\gamma))(\rho) - (H_l(-d^2(\rho,\cdot)))(\gamma) \le L d^2(\rho,\gamma). \tag{38}$$

Combining (38) with (35) gives

$$(Hmd^2(\cdot,\gamma))(\rho) - (H(-md^2(\rho,\cdot))(\gamma) \le Lmd^2(\rho,\gamma), \quad \forall m \ge 0. \tag{39}$$

The heuristic arguments in Sections 1.2.1 and 1.2.2 suggest that, having only test functions of the form (13) in the domain of H is not good enough, we would like to include test functions f_0, f_1 of the following form:

$$f_0(\rho) = (1-\kappa)md^2(\rho,\gamma) + \kappa\mathcal{E}(\rho) + \text{constant}, \tag{40}$$

and

$$f_1(\gamma) = -m(1+\kappa)d^2(\rho,\gamma) - \kappa\mathcal{E}(\gamma) + \text{constant}, \tag{41}$$

where $0 < \kappa < 1$ and $\gamma, \rho \in E$.

In Section 2, we will formulate the above informal arguments into a rigorous comparison theorem (Theorem 5). We will then apply it to Examples 1, 2 and 3, obtaining respective comparison principles in Theorems 6, 7, and 8. In the last example where d is the order-2-Wasserstein metric (53), our analysis will heavily rely upon mass transport techniques [28], [19] and [37]. We review the techniques in Appendix A.

2. A comparison principle

We identify operator with its graph. Let $H_0, H_1 \subset M(E, \overline{R}) \times M(E, \overline{R})$, $h_0, h_1 \in C_b(E)$ and $\alpha > 0$. Suppose $\overline{f} \in B(E)$ is a viscosity subsolution of

$$(I - \alpha H_0) f = h_0, \tag{42}$$

and $\underline{f} \in B(E)$ is a viscosity supersolution of

$$(I - \alpha H_1) f = h_1. \tag{43}$$

We are interested in the comparison between $\overline{f}$ and $\underline{f}$. We assume throughout this paper that H_0, H_1 are invariant under translation by a constant: $H_0(f + C) = H_0 f$ and $H_1(f + C) = H_1 f$ for $C \in R$. The most common situation is $H_1 = H_0 = H$ and $h_0 = h_1 = h$.

Condition 1 *Let (E, r) be a complete metric space.*

1. *There exists a function $\mathcal{E} \in M(E, \overline{R})$, $\mathcal{E} \not\equiv +\infty$, (hereafter referred to as the energy function) which is nonnegative, lower semicontinuous and has compact level sets; and a distance function $d(\rho, \gamma)$ which is lower*

semicontinuous (with respect to the product topology on $E \times E$ induced by $r \times r$) such that

$$(1-\kappa)md^2(\cdot,\gamma_0) + \kappa\mathcal{E}(\cdot) + C \in \mathcal{D}(H_0),$$

and

$$-(1+\kappa)md^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot) + C \in \mathcal{D}(H_1),$$

for every $0 < \kappa \le 1$, $m > 0$, $C \in R$ and $\rho_0, \gamma_0 \in E$ such that $\mathcal{E}(\rho_0) + \mathcal{E}(\gamma_0) < +\infty$;

2. *For each $0 < \kappa < 1$, there exists $\omega_\kappa : [0,\infty) \to [0,+\infty]$ satisfying*

$$\liminf_{\kappa\to 0+} \liminf_{r\to 0+} \omega_\kappa(r) \le 0.$$

Furthermore, for every $\rho_0, \gamma_0 \in E$ satisfying $\mathcal{E}(\rho_0) + \mathcal{E}(\gamma_0) < +\infty$,

$$\begin{aligned} \Big\{ \frac{1}{1-\kappa}\Big(H_0(m(1-\kappa)d^2(\cdot,\gamma_0) + \kappa\mathcal{E})\Big)^*(\rho_0) \qquad\qquad (44) \\ - \frac{1}{1+\kappa}\Big(H_1(-m(1+\kappa)d^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot))\Big)_*(\gamma_0)\Big\} \\ \le \omega_\kappa(md^2(\rho_0,\gamma_0) + d^2(\rho_0,\gamma_0)); \end{aligned}$$

Note that in general, d and r may induce different topologies on E. Throughout this paper, unless specified otherwise, convergence in E is always under the topology by r. We also require some regularity information on the $\overline{f}$ and $\underline{f}$.

Condition 2 *$\overline{f}$ and $\underline{f}$ are d-continuous in the sense that: for every $\rho_n, \rho_0 \in E$ satisfying $\lim_{n\to+\infty} d(\rho_n, \rho) = 0$, it follows that*

$$\lim_{n\to+\infty} \overline{f}(\rho_n) = \overline{f}(\rho_0) \text{ and } \lim_{n\to+\infty} \underline{f}(\rho_n) = \underline{f}(\rho_0).$$

Moreover, finite energy states are d-dense in the state space E: for any $\rho_0 \in E$, there exists $\rho_n \in E$,

$$\lim_{n\to+\infty} d(\rho_n, \rho_0) = 0, \quad \mathcal{E}(\rho_n) < +\infty.$$

Theorem 3. *Let $\overline{f}$ be a bounded upper semicontinuous strong subsolution of (42), and $\underline{f}$ be a bounded lower semicontinuous strong supersolution of (43), which both satisfy Condition 2.*

Assume that Conditions 1 hold, that $\alpha > 0$, and that one of the $h_0, h_1 \in C_b(E)$ is uniformly continuous with respect to d in the following sense

$$\lim_{\epsilon\to 0+} \sup_{(\rho,\gamma):d(\rho,\gamma)<\epsilon} |h_i(\rho) - h_i(\gamma)| = 0.$$

Then

$$\sup_{\rho\in E}(\overline{f}(\rho) - \underline{f}(\rho)) \le \sup_{\rho\in E}(h_0(\rho) - h_1(\rho)).$$

Proof. Let

$$\varPhi(\rho,\gamma) = \frac{1}{1-\kappa}\overline{f}(\rho) - \frac{1}{1+\kappa}\underline{f}(\gamma) - md^2(\rho,\gamma) - \frac{\kappa}{1-\kappa}\mathcal{E}(\rho) - \frac{\kappa}{1+\kappa}\mathcal{E}(\gamma).$$

Let $0 < \kappa < 1, m > 0$ be fixed. By the assumptions on $\overline{f}, \underline{f}$, and lower semicontinuity of d^2 and $\mathcal{E}$, and by the compact level set assumption on $\mathcal{E}$, there exists compact $K \subset E$ and $(\rho_0, \gamma_0) \in K \times K$ such that

$$\varPhi(\rho_0,\gamma_0) = \sup_{(\rho,\gamma)\in E\times E} \varPhi(\rho,\gamma) = \sup_{(\rho,\gamma)\in K\times K} \varPhi(\rho,\gamma). \tag{45}$$

Let

$$f_0(\rho) = (1-\kappa)md^2(\rho,\gamma_0) + \kappa\mathcal{E}(\rho).$$

Then, by (45) and the strong viscosity subsolution property of $\overline{f}$,

$$\overline{f}(\rho_0) - f_0(\rho_0) = \sup_{\rho\in E}(\overline{f}(\rho) - f_0(\rho)), \tag{46}$$

and

$$\alpha^{-1}(\overline{f} - h_0)(\rho_0) \le (H_0 f_0)^*(\rho_0) = \Big(H_0((1-\kappa)md^2(\cdot,\gamma_0) + \kappa\mathcal{E})\Big)^*(\rho_0) \tag{47}$$

Similarly, let

$$f_1(\gamma) = -(1+\kappa)md^2(\rho_0,\gamma) - \kappa\mathcal{E}(\gamma).$$

Then by the strong viscosity supersolution property,

$$f_1(\gamma_0) - \underline{f}(\gamma_0) = \sup_{\gamma\in E}(f_1(\gamma) - \underline{f}(\gamma)),$$

and

$$\begin{aligned} \alpha^{-1}(\underline{f} - h_1)(\gamma_0) &\ge (H_1 f_1)_*(\gamma_0) \\ &= \Big(H_1(-(1+\kappa)md^2(\rho_0,\cdot) - \kappa\mathcal{E})\Big)_*(\gamma_0). \end{aligned} \tag{48}$$

Therefore,

$$\begin{aligned} &\sup_{\rho\in E}(\frac{1}{1-\kappa}\overline{f}(\rho) - \frac{1}{1+\kappa}\underline{f}(\rho) - \frac{2\kappa}{1-\kappa^2}\mathcal{E}(\rho)) = \sup_{\rho\in E}\Phi(\rho,\rho) \\ &\le \sup_{(\rho,\gamma)\in E\times E}\Phi(\rho,\gamma) = \Phi(\rho_0,\gamma_0) \\ &\le \frac{1}{1-\kappa}\overline{f}(\rho_0) - \frac{1}{1+\kappa}\underline{f}(\gamma_0) \\ &\le \alpha[\frac{1}{1-\kappa}\Big(H_0((1-\kappa)md^2(\cdot,\gamma_0) + \kappa\mathcal{E}(\cdot))\Big)^*(\rho_0) \\ &\qquad -\frac{1}{1+\kappa}\Big(H_1(-(1+\kappa)md^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot))\Big)_*(\gamma_0)] \\ &\qquad +(h_0(\rho_0) - h_1(\gamma_0)) + (|\frac{1}{1-\kappa} - 1|\|h_0\| + |\frac{1}{1+\kappa} - 1|\|h_1\|) \\ &\le \alpha\omega_\kappa(md^2(\rho_0,\gamma_0) + d^2(\rho_0,\gamma_0)) + (h_0(\rho_0) - h_1(\gamma_0)) \\ &\qquad +\alpha(|\frac{1}{1-\kappa} - 1|\|h_0\| + |\frac{1}{1+\kappa} - 1|\|h_1\|), \end{aligned}$$

where the third inequality above follows by (47) and (48), and the fourth inequality follows by (44). By a well-known argument (e.g. Proposition 3.7 of [9]),

$$\lim_{m\to+\infty} md^2(\rho_0,\gamma_0)=0.$$

Therefore letting $m\to+\infty$, then $\kappa\to 0$,

$$\overline{f}(\rho)-\underline{f}(\rho)\le \sup_{\gamma\in E}(h_0(\gamma)-h_1(\gamma)),\quad \forall\rho \text{ such that } \mathcal{E}(\rho)<\infty$$

We conclude by Condition 2 and the continuities of $\overline{f}$ and $\underline{f}$.

Remark 1. The inequality (44) in Condition 1 is intimately related to assumption (H3) of Part I of Crandall and Lions [7], see also the discussion in Section 1.2.2. Suppose that $H_0=H_1=H$,

$$\mathcal{E}, md^2(\cdot,\gamma), -md^2(\rho,\cdot)\in\mathcal{D}(H),\quad \forall m>0,\rho,\gamma\in E; \tag{49}$$

and that H is convex in the sense that

$$H((1-\kappa)f+\kappa g)\le(1-\kappa)Hf+\kappa Hg,\quad \forall 0\le\kappa\le 1, f,g\in\mathcal{D}(H).$$

We also assume

$$C_0\equiv\sup_{\rho\in E}H\mathcal{E}(\rho)<\infty. \tag{50}$$

Then for the f_0 in (40)

$$\begin{aligned}Hf_0(\rho)&=H\Big((1-\kappa)md^2(\cdot,\gamma)+\kappa\mathcal{E}(\cdot)\Big)(\rho)\\&\le(1-\kappa)\Big(Hmd^2(\cdot,\gamma)\Big)(\rho)+\kappa H\mathcal{E}(\rho),\end{aligned}$$

and for the f_1 in (41),

$$\begin{aligned}
&H\Big(-md^2(\rho,\cdot)\Big)(\gamma) \\
&\le \frac{1}{1+\kappa}H\Big(-(1+\kappa)md^2(\rho,\cdot)-\kappa\mathcal{E}(\cdot)\Big)(\gamma)+\frac{\kappa}{1+\kappa}H\mathcal{E}(\gamma) \\
&= \frac{1}{1+\kappa}Hf_1(\gamma)+\frac{\kappa}{1+\kappa}H\mathcal{E}(\gamma).
\end{aligned}$$

If in addition, the Hf_0, Hf_1 are continuous functions, then (44) follows from the much simpler estimate (31). However, for the Hs in Examples 1, 2 and 3, it is too strong to require (49) and the continuity of $Hf_0(\rho), Hf_1(\gamma)$ for all $\rho, \gamma \in E$. The inclusion of the small perturbation term $\kappa\mathcal{E}$ allows us to essentially only consider (44) at points ρ, γ which are regular in the sense that $(\mathcal{E}-H\mathcal{E})(\rho)+(\mathcal{E}-H\mathcal{E})(\gamma)<+\infty$.

In Theorem 3, we required that $\overline{f}$ and $\underline{f}$ be respectively strong viscosity sub- and super- solutions. The following result (Chapter 9, Feng and Kurtz [18]) removes the "strongness" assumption.

Condition 4 *Let Condition 1 be satisfied.*

For each $0<\kappa\le 1$, $m=1,2,\dots$ and $\rho_0,\gamma_0\in E$ satisfying $\mathcal{E}(\rho_0)+\mathcal{E}(\gamma_0)<+\infty$, the following holds:

1. *There exists a lower semicontinuous g_0 such that*

$$(1-\kappa)md^2(\cdot,\gamma_0)+\kappa\mathcal{E}(\cdot)+g_0\in\mathcal{D}(H_0),$$

$g_0(\rho_0)=0$, $g_0(\rho)>0$ for $\rho\ne\rho_0$, and

$$\begin{aligned}
&\Big(H_0((1-\kappa)md^2(\cdot,\gamma_0)+\kappa\mathcal{E}(\cdot)+g_0)\Big)^*(\rho_0) \\
&\le\Big(H_0((1-\kappa)md^2(\cdot,\gamma_0)+\kappa\mathcal{E}(\cdot))\Big)^*(\rho_0).
\end{aligned} \tag{51}$$

2. *There exists a lower semicontinuous* g_1 *such that*

$$-(1+\kappa)md^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot) - g_1 \in \mathcal{D}(H_1),$$

$g_1(\gamma_0) = 0$, $g_1(\gamma) > 0$ *for* $\gamma \neq \gamma_0$, *and*

$$\begin{aligned} &\Big(H_1(-(1+\kappa)md^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot) - g_1)\Big)_*(\gamma_0) \\ \geq &\Big(H_1(-(1+\kappa)md^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot))\Big)_*(\gamma_0). \end{aligned} \tag{52}$$

Theorem 5. *Let* $\overline{f}$ *be a bounded upper semicontinuous subsolution of (42), and* $\underline{f}$ *be a bounded lower semicontinuous supersolution of (43), which both satisfy Condition 2.*

Assume that Condition 4 holds, that $\alpha > 0$, *and that one of the* $h_0, h_1 \in C_b(E)$ *is uniformly continuous with respect to d:*

$$\lim_{\epsilon\to 0+} \sup_{(\rho,\gamma):d(\rho,\gamma)<\epsilon} |h_i(\rho) - h_i(\gamma)| = 0.$$

Then

$$\sup_{\rho\in E}(\overline{f}(\rho) - \underline{f}(\rho)) \leq \sup_{\rho\in E}(h_0(\rho) - h_1(\rho)).$$

Proof. We only need to modify the proof of Theorem 3 slightly.

For the ρ_0, γ_0 in the proof of Theorem 3, let g_0 be the lower semicontinuous function in Condition 4, and let $\tilde{f}_0 = f_0 + g_0$. Then

$$\overline{f}(\rho_0) - \tilde{f}_0(\rho_0) = \overline{f}(\rho_0) - f_0(\rho_0) \geq \overline{f}(\rho) - f_0(\rho) > \overline{f}(\rho) - \tilde{f}_0(\rho), \quad \rho \neq \rho_0.$$

Hence ρ_0 is the unique maximum of $\overline{f} - \tilde{f}_0$. Therefore by viscosity subsolution property and Condition 4,

$$\alpha^{-1}(\overline{f} - h_0)(\rho_0) \leq (H_0\tilde{f}_0)^*(\rho_0) \leq (H_0 f_0)^*(\rho_0).$$

Hence we arrive at (47) again.

Similarly, using the super-solution property, we can obtain (48).

The rest of the proof follows that in Theorem 3.

3. Examples

We now apply Theorem 5 to establish comparison principles for Examples 1 (see Section 3.3), 2 (see Section 3.4) and 3 (see Section 3.5). The general program is as follows: First, we identify the d in Theorem 5 using (32), as well as the right notion of gradient for each example using informal geometrical argument. We then apply test functions f_0, f_1 (in (40) and (41)) to the operator H in (10). We find that Hf_0, Hf_1 are rigorously well defined functions taking values in the extended reals (i.e. they could be either $+\infty$ or $-\infty$ but not both). Additionally, they are semicontinuous functions. Finally, we resort Theorem 5 for a rigorous, but non-geometric proof of the respective comparison principles. It is important to take such a practical approach in the proof, because infinite dimensional geometry in the context we discuss here is usually problematic to justify rigorously.

3.1. Identification of d.

We consider Examples 1 and 2 first. For any given point $\rho \in E$, the tangent spaces are respectively $T_\rho E = (L^2(\mathcal{O}), \langle \cdot, \cdot \rangle)$, and $T_\rho E \equiv (H^{-1}(\mathcal{O}), (\cdot, \cdot)_{-1})$. The tangent spaces are indeed state independent, therefore d should coincide with the local distance induced by the norm of the respective tangent space. That is, $d(\rho,\gamma) = \|\rho-\gamma\|_{L^2(\mathcal{O})}$ for Example 1 and $d(\rho,\gamma) = \|\rho-\gamma\|_{-1}$ for Example 2.

The case of Example 3 is more involved, and we only provide a reference here. The informal geometric structure we used in that example is introduced by Otto [29], where the d satisfying (32) is identified as the order-2-Wasserstein metric:

$$d^2(\rho,\gamma) = \inf\{\int_{R^d\times R^d} |x-y|^2 d\pi : \pi \in \mathcal{P}(R^d\times R^d), \quad \pi(\cdot, R^d) = \rho(\cdot), \pi(R^d,\cdot) = \gamma(\cdot)\}. \tag{53}$$

3.2. The domain of H and a convexity structure

As emphasized by the discussions following Definition 1, the choice of $\mathcal{D}(H)$ can be subtle in any infinite dimensional context. Here, we explicitely specify $\mathcal{D}(H)$ for the examples coming up next.

$$\mathcal{D}(H) = \{f_0, f_1, f_0 + g_0, f_1 - g_1\} \tag{54}$$

where f_0, f_1 are defined respectively by (40) and (41) with $d, \mathcal{E}$ specified on case specific basis in Sections 3.3, 3.4 and 3.5. The g_0, g_1 are defined by (60) and (61) in the Allen-Cahn case as well as the Cahn-Hilliard case, and by (72) and (73) in the Fokker-Planck case. We will show that $Hf \in M(E, \overline{R})$ is well defined for each $f \in D(H)$.

We observe the following identities and convex inequalities.

By the bilinear property of $g_\rho(\cdot,\cdot)$,

$$Hf_0(\rho) = (1-\kappa)m g_\rho(-\text{grad}\mathcal{E}(\rho), \text{grad}d^2(\cdot,\gamma)(\rho)) \tag{55}$$

$$+\frac{m^2(1-\kappa)^2}{2} g_\rho(\mathrm{grad} d^2(\cdot,\gamma)(\rho), \mathrm{grad} d^2(\cdot,\gamma)(\rho))$$
$$-(\kappa - \frac{\kappa^2}{2}) g_\rho(\mathrm{grad}\mathcal{E}(\rho), \mathrm{grad}\mathcal{E}(\rho))$$
$$+m\kappa(1-\kappa) g_\rho(\mathrm{grad} d^2(\cdot,\gamma)(\rho), \mathrm{grad}\mathcal{E}(\rho)).$$

By convexity,

$$\begin{aligned} Hf_0(\rho) \le (1-\kappa)\Big(& m g_\rho(-\mathrm{grad}\mathcal{E}(\rho), \mathrm{grad} d^2(\cdot,\gamma)(\rho)) \\ & + \frac{m^2}{2} g_\rho(\mathrm{grad} d^2(\cdot,\gamma)(\rho), \mathrm{grad} d^2(\cdot,\gamma)(\rho))\Big) \\ & + \kappa\Big(-\frac{1}{2} g_\rho(\mathrm{grad}\mathcal{E}(\rho), \mathrm{grad}\mathcal{E}(\rho))\Big). \end{aligned} \tag{56}$$

Similarly,

$$\begin{aligned} Hf_1(\gamma) = (1+\kappa) & m g_\gamma(-\mathrm{grad}\mathcal{E}(\gamma), -\mathrm{grad} d^2(\rho,\cdot)(\gamma)) \\ & + \frac{m^2}{2} g_\gamma(\mathrm{grad} d^2(\rho,\cdot)(\gamma), \mathrm{grad} d^2(\rho,\cdot)(\gamma)) \\ & + (\kappa + \frac{\kappa^2}{2}) g_\gamma(\mathrm{grad}\mathcal{E}(\gamma), \mathrm{grad}\mathcal{E}(\gamma)) \\ & + m\kappa(1+\kappa) g_\gamma(\mathrm{grad} d^2(\rho,\cdot)(\gamma), \mathrm{grad}\mathcal{E}(\gamma)); \end{aligned} \tag{57}$$

and

$$\begin{aligned} Hf_1(\gamma) \ge (1+\kappa)\Big(& m g_\gamma(\mathrm{grad}\mathcal{E}(\gamma), \mathrm{grad} d^2(\rho,\cdot)(\gamma)) \\ & + \frac{m^2}{2} g_\gamma(\mathrm{grad} d^2(\rho,\cdot)(\gamma), \mathrm{grad} d^2(\rho,\cdot)(\gamma))\Big) \\ & + \frac{\kappa}{2} g_\gamma(\mathrm{grad}\mathcal{E}(\gamma), \mathrm{grad}\mathcal{E}(\gamma)). \end{aligned} \tag{58}$$

3.3. Optimal controlled Allen-Cahn equation

Let $\mathcal{O}, F, \mathcal{E}$ be defined according to Example 1, $E = (L^2(\mathcal{O}), \|\cdot\|_{L^2(\mathcal{O})})$. We use $\langle \cdot,\cdot\rangle$ to denote either the usual inner product in $L^2(\mathcal{O})$, or its extension as a dual pairing if Schwartz distributions are involved. Following the geometric discussions in Example 1, we choose $g_\rho(u,v) = \langle u, v\rangle$ and define the gradient for a function on E as follows:

Definition 2 (Gradient). Let f be a function which maps E into the extended reals, and $\rho_0 \in E$. For any $p \in C^\infty(\mathcal{O}) \subset L^2(\mathcal{O})$ and any t, let $\rho(t) \equiv \rho_0 + tp$ be a flow generated by p starting at position ρ_0. We say that the gradient $\mathrm{grad} f(\rho_0)$ exists at ρ_0, if and only if it can be identified as the unique element in the distribution space $\mathcal{D}^{'}(\mathcal{O})$ through identity

$$\lim_{t\to 0} \frac{1}{t}\Big(f(\rho(t)) - f(\rho_0)\Big) = \langle \mathrm{grad} f(\rho_0), p\rangle \quad \forall p \in C^\infty(\mathcal{O}).$$

It follows that, for the free energy $\mathcal{E}$ defined in (11), and ρ satisfying $\mathcal{E}(\rho) < +\infty$, $\mathrm{grad}\mathcal{E}(\rho) = -\Delta\rho + F^{'}(\rho)$ in the distribution sense; and that for f of the form (13),

$$\mathrm{grad} f(\rho) = \frac{\delta f}{\delta \rho};$$

and that for $f(\rho) = \|\rho - \gamma\|^2_{L^2(\mathcal{O})}$,

$$\mathrm{grad} f(\rho) = 2(\rho - \gamma), \quad \rho, \gamma \in E.$$

We choose $d(\rho,\gamma) = \|\rho - \gamma\|_{L^2(\mathcal{O})}$, and $\mathcal{E}$ by (1). Then (55) and (57) are both well-defined. We illustrate (55) only: let

$$f_0(\rho) = (1-\kappa) m \|\rho - \gamma\|^2_{L^2(\mathcal{O})} + \kappa\mathcal{E}(\rho) + c, \tag{59}$$

where $m > 0, 0 < \kappa \leq 1, c \in R$, $\gamma \in L^2(\mathcal{O})$, then

$$Hf_0(\rho) = \begin{cases} (1-\kappa)2m\langle \Delta\rho - F^{'}(\rho), \rho - \gamma\rangle & \\ \quad +(1-\kappa)^2 2m^2 \|\rho-\gamma\|^2_{L^2(\mathcal{O})} & \\ \quad -\kappa(1-\frac{\kappa}{2})\|\Delta\rho - F^{'}(\rho)\|^2_{L^2(\mathcal{O})} & \\ +2m\kappa(1-\kappa)\langle -\Delta\rho + F^{'}(\rho), \rho-\gamma\rangle & \text{if } \rho \in H^2(\mathcal{O}) \ , \\ -\infty & \text{otherwise;} \end{cases}$$
$$\leq (1-\kappa)\Big(2m\langle -\Delta\rho + F^{'}(\rho), \rho-\gamma\rangle + 2m^2\|\rho-\gamma\|^2_{L^2(\mathcal{O})}\Big) - \frac{\kappa}{2}\|\Delta\rho - F^{'}(\rho)\|^2_{L^2(\mathcal{O})}.$$

In the above, the right hand side of the inequality is understood as $-\infty$ when $\rho \notin H^2(\mathcal{O})$. The perturbation by $\kappa > 0$ introduces a small but important higher order term $-\frac{\kappa}{2}\|\Delta\rho - F^{'}(\rho)\|^2_{L^2(\mathcal{O})}$. It is because of this term, that $Hf_0 \in M(E, \overline{R})$ becomes well defined and is upper semicontinuous in (E, r). Similarly, the Hf_1 in (57) is well defined and is lower semicontinuous in (E, r).

We verify Condition 4 next. Let

$$\{\alpha_0(x) = 1, \alpha_{2j-1}(x) = \sqrt{2}\cos(2\pi jx), \alpha_{2j}(x) = \sqrt{2}\sin(2\pi jx) : j = 1, 2, \ldots\}$$

and

$$\lambda_0 = 0, \lambda_{2j-1} = \lambda_{2j} = 4\pi^2 j^2.$$

For $k = (k_1, \ldots, k_d)$, $k_i = 1, 2, \ldots$, let

$$e_k(x) = \alpha_{k_1}(x_1)\cdots\alpha_{k_d}(x_d), \quad \lambda_k = \lambda_{k_1} + \ldots + \lambda_{k_d}.$$

Then $\{e_1, \ldots, e_k, \ldots\} \subset C^\infty(\mathcal{O})$ is an orthonormal basis of eigenfunctions for $-\Delta$ in $E = L^2(\mathcal{O})$ with λ_k as the corresponding eigenvalues:

$$-\Delta e_k = \lambda_k e_k.$$

For each $\rho_0 \in E$, let

$$g_0(\rho) = \sum_k 2^{-k} \langle \rho - \rho_0, e_k \rangle^2. \tag{60}$$

Then

$$\text{grad} g_0(\rho) = \sum_k 2^{-k+1} \langle \rho - \rho_0, e_k \rangle e_k.$$

$\text{grad} g_0(\rho_0) = 0$. Let f_0 be defined as in (59), then $H(f_0 + g_0)$ is well defined as in the case of Hf_0. Furthermore, $(H(f_0 + g_0))^*(\rho_0) = (Hf_0)^*(\rho_0)$ so (51) holds. Similarly, for each $\gamma_0 \in E$, we define

$$g_1(\gamma) = \sum_k 2^{-k} \langle \gamma - \gamma_0, e_k \rangle^2. \tag{61}$$

Then (52) holds.

Theorem 6. *Suppose $h_0, h_1 \in C_b(E)$ and one of them is uniformly continuous. Let $\overline{f}, \underline{f} \in C_b(E)$ be respectively sub- and super-solution of (42) and (43) with $H_0 = H_1 = H$, where the H is defined as above with $\mathcal{D}(H)$ consists of test functions $f_0, f_0 + g_0$ and $f_1, f_1 - g_1$. Then*

$$\sup_{\rho \in E} (\overline{f}(\rho) - \underline{f}(\rho)) \leq \sup_{\rho \in E} (h_0(\rho) - h_1(\rho)).$$

Proof. We choose

$$d^2(\rho, \gamma) = \|\rho - \gamma\|^2_{L^2(\mathcal{O})},$$

and verify (44). Let $\rho_0, \gamma_0 \in E$, then (56), (58) and the semicontinuities of Hf_0, Hf_1 mentioned prior to Theorem 6 imply that

$$\begin{aligned}
&\frac{1}{1-\kappa}\Big(H(m(1-\kappa)d^2(\cdot,\gamma_0)+\kappa\mathcal{E}(\cdot))\Big)^*(\rho_0) \\
&\qquad\qquad -\frac{1}{1+\kappa}\Big(H(-m(1+\kappa)d^2(\rho_0,\cdot)-\kappa\mathcal{E}(\cdot))\Big)_*(\gamma_0) \\
&= \frac{1}{1-\kappa}H(m(1-\kappa)d^2(\cdot,\gamma_0)+\kappa\mathcal{E}(\cdot))(\rho_0) \\
&\qquad -\frac{1}{1+\kappa}H(-m(1+\kappa)d^2(\rho_0,\cdot)-\kappa\mathcal{E}(\cdot))(\gamma_0) \\
&\le 2m\langle(\Delta\rho_0 - F^{'}(\rho_0)) - (\Delta\gamma_0 - F^{'}(\gamma_0)), \rho_0-\gamma_0\rangle \\
&\qquad -\frac{\kappa(1-\kappa)^{-1}}{2}\|\Delta\rho_0 - F^{'}(\rho_0)\|^2_{L^2(\mathcal{O})} - \frac{\kappa(1+\kappa)^{-1}}{2}\|\Delta\gamma_0 - F^{'}(\gamma_0)\|^2_{L^2(\mathcal{O})} \\
&\le 2L_F m\|\rho_0-\gamma_0\|^2_{L^2(\mathcal{O})},
\end{aligned}$$

where L_F is the Lipschitz constant of F. As before, the right hand side on the first inequality above is understood as $-\infty$ when either $\rho_0 \notin H^2(\mathcal{O})$ or $\gamma_0 \notin H^2(\mathcal{O})$.

By Theorem 5, the conclusion follows.

3.4. Optimal controlled Cahn-Hilliard equation

Following the notation and assumptions of Example 2, we consider the comparison principle for the controlled Cahn-Hilliard equation. With the geometrical intuitions in Example 2 in mind, we choose the metric tensor $g_\rho(u,v) = (u,v)_{-1}$, and next define the gradient of a function on E:

Definition 3. Let $f : E \to \overline{R}$. Let $\rho_0 \in E$. For every $p \in C^\infty(\mathcal{O})$ and $u = -\Delta p$, we generate a flow $\rho(t)$ such that $\rho(0) = \rho_0$ and $\dot{\rho} = u_0 = -\Delta p$. We say the gradient of f at ρ_0 exists, and denoted by $\text{grad} f(\rho_0)$, if and only if it can be identified as the unique element in distribution space $\mathcal{D}^{'}(\mathcal{O})$ through identity

$$\lim_{t\to 0+}\frac{1}{t}(f(\rho(t))-f(\rho(0)))=\langle \mathrm{grad} f(\rho_0),p\rangle,\quad \forall p\in C^{\infty}(\mathcal{O}).$$

In this notation,

$$\mathrm{grad}\ \mathcal{E}(\rho)=-\Delta\Big(-\Delta\rho+F^{'}(\rho)\Big),\quad \mathcal{E}(\rho)<\infty$$

and

$$\mathrm{grad} f(\rho)=-\Delta\frac{\delta f}{\delta\rho}$$

for f of the form (13).

Following the heuristics of Example 2, we choose $d(\rho,\gamma)=\|\rho-\gamma\|_{-1}$. Then (55) and (57) are both well defined. We only illustrate (55). First, note that the definition of $\|\cdot\|_{-1}$ in (16), and note that

$$\|\rho\|_{-1}^2=\langle(-\Delta)^{-1}\rho,\rho\rangle,\quad \forall\rho\in E.$$

Let

$$f_0(\rho)=(1-\kappa)m\|\rho-\gamma\|_{-1}^2+\kappa\mathcal{E}(\rho)+c,\quad m\geq 0, 0<\kappa\leq 1,\gamma\in E,$$

and consider $\mathcal{E}(\rho)<+\infty$. Then

$$\mathrm{grad} d^2(\cdot,\gamma)(\rho)=2(\rho-\gamma),\quad -\mathrm{grad}\mathcal{E}(\rho)=\Delta(-\Delta\rho+F^{'}(\rho)),$$

$$g_\rho(\mathrm{grad} d^2(\cdot,\gamma)(\rho),\mathrm{grad} d^2(\cdot,\gamma)(\rho))=4\|\rho-\gamma\|_{-1}^2.$$

In addition,

$$g_\rho(\mathrm{grad}\mathcal{E}(\rho),\mathrm{grad}\mathcal{E}(\rho))=\|-\Delta(-\Delta\rho+F^{'}(\rho))\|_{-1}^2,$$

and when the above term is bounded, the following is finite

$$g_\rho(\mathrm{grad} d^2(\cdot,\gamma)(\rho), \mathrm{grad}\mathcal{E}(\rho)) = 2\langle -\Delta\rho + F^{'}(\rho), \rho-\gamma\rangle.$$

Therefore the following is well defined and takes values in $[-\infty, +\infty)$

$$\begin{aligned} Hf_0(\rho) &= -2m(1-\kappa)^2\langle -\Delta\rho + F^{'}(\rho), \rho-\gamma\rangle + 2(1-\kappa)^2 m^2\|\rho-\gamma\|_{-1}^2 \\ &\qquad -(\kappa - \frac{\kappa^2}{2})\| -\Delta(-\Delta\rho + F^{'}(\rho))\|_{-1}^2 \\ &\le (1-\kappa)\Big(-2m\langle -\Delta\rho + F^{'}(\rho), \rho-\gamma\rangle + 2m^2\|\rho-\gamma\|_{-1}^2\Big) \\ &\qquad -\frac{\kappa}{2}\| -\Delta(-\Delta\rho + F^{'}(\rho))\|_{-1}^2. \end{aligned}$$

where the right hand side is understood as $-\infty$ when $\|\nabla(-\Delta\rho)\|_{L^2(\mathcal{O})}^2 = \|-\Delta(-\Delta\rho)\|_{-1}^2 = +\infty$. Once again, because of the higher order perturbation term, for any $0 < \kappa \le 1$, Hf_0 is well defined and is upper semicontinuous in E. Recall that we use the L^2-norm to induce the metric r in (E, r).

Let g_1, g_0 be defined according to (60) and (61) in the previous example. Then similar calculations show that Condition 4 holds.

Theorem 7. *Let $h_1, h_2 \in C_b(E)$, and let one of them be uniformly continuous in d (i.e. the $\|\cdot\|_{-1}$ norm):*

$$\lim_{\epsilon\to 0+} \sup_{\rho,\gamma\in E:\|\rho-\gamma\|_{-1}<\epsilon} |h_i(\rho) - h_i(\gamma)| = 0.$$

Suppose $\overline{f}$ is a bounded upper-semicontinuous viscosity subsolution of (42) on E, and that $\underline{f}$ is a bounded lower semi-continuous supersolutions of (43) on E. $H_0 = H_1 = H$, where H is defined as above with $\mathcal{D}(H)$ consists of test functions $f_0, f_1, f_0 + g_0, f_0 - g_1$.

Assume that both $\overline{f}$ and $\underline{f}$ are continuous in d. Then

$$\sup_{\rho \in E}(\overline{f}(\rho) - \underline{f}(\rho)) \leq \sup_{\rho \in E}(h_0(\rho) - h_1(\rho)).$$

Remark 2. Section 3.6 outlines the construction of $\overline{f}, \underline{f}$ satisfying the above requirements.

Proof. We only need to verify (44). Let $\rho_0, \gamma_0 \in E$, then as in the proof of Theorem 6,

$$\begin{aligned}
&\frac{1}{1-\kappa}\Big(H(m(1-\kappa)d^2(\cdot,\gamma_0) + \kappa\mathcal{E}(\cdot))\Big)^*(\rho_0) \\
&\qquad\qquad - \frac{1}{1+\kappa}\Big(H(-m(1+\kappa)d^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot))\Big)_*(\gamma_0) \\
&= \frac{1}{1-\kappa}H(m(1-\kappa)d^2(\cdot,\gamma_0) + \kappa\mathcal{E}(\cdot))(\rho_0) \\
&\qquad - \frac{1}{1+\kappa}H(-m(1+\kappa)d^2(\rho_0,\cdot) - \kappa\mathcal{E}(\cdot))(\gamma_0) \\
&\leq 2m\langle(\Delta\rho_0 - F^{'}(\rho_0)) - (\Delta\gamma_0 - F^{'}(\gamma_0)), \rho_0 - \gamma_0\rangle \\
&\qquad - \frac{\kappa}{2(1-\kappa)}\|(-\Delta)(-\Delta\rho_0 + F^{'}(\rho_0))\|_{-1}^2 \\
&\qquad\quad - \frac{\kappa}{2(1+\kappa)}\|(-\Delta)(-\Delta\gamma_0 + F^{'}(\gamma_0))\|_{-1}^2 \\
&\leq -2m\|\nabla(\rho_0 - \gamma_0)\|_{L^2(\mathcal{O})}^2 + 2mL_F\|\rho_0 - \gamma_0\|_{L^2(\mathcal{O})}^2 \\
&\leq -2m\|\nabla(\rho_0 - \gamma_0)\|_{L^2(\mathcal{O})}^2 + 2mL_F\|\nabla(\rho_0 - \gamma_0)\|_{L^2(\mathcal{O})}\|\rho_0 - \gamma_0\|_{-1} \\
&\leq \frac{1}{2}L_F^2 m\|\rho_0 - \gamma_0\|_{-1}^2,
\end{aligned}$$

where

$$L_F \equiv \sup_{r \neq s}\frac{|F^{'}(s) - F^{'}(r)|}{|s-r|} < +\infty.$$

The right hand side of the first inequality above is understood as $-\infty$ when either $\rho_0 \notin H^3(\mathcal{O})$ or $\gamma_0 \notin H^3(\mathcal{O})$.

3.5. Optimal controlled Fokker-Planck equations

Following the notations of Example 3, we prove the comparison principle for the controlled Fokker-Planck equation. Throughout this section, d is the Wasserstein metric in (53) and $\mathcal{E}$ is defined in (23).

We define a notion of gradient which is motivated by the geometric perspectives in Example 3. Formally, we would like to take $g_\rho(u,v) = (u,v)_{-1,\rho}$ where the latter is defined in Appendix A.2. We recall that, for each $\rho_0 \in E$ and $p \in C_c^\infty(R^d)$, there exists a unique weak solution $\rho(t) \in E$ for

$$\dot{\rho} = -\nabla \cdot (\rho \nabla p), \quad \rho(0) = \rho_0. \tag{62}$$

Definition 4 (Gradient). Let $f : E \to \overline{R}$. We say the gradient of f at ρ_0 exists, which is denoted by $\mathrm{grad} f(\rho_0)$, if and only if it can be identified as a unique element in the Schwartz space of distributions $\mathcal{D}'(R^d)$ through the identity

$$\lim_{t\to 0+} \frac{f(\rho(t)) - f(\rho(0))}{t} = \langle \mathrm{grad}\ f(\rho_0), p\rangle, \quad \forall p \in C_c^\infty(R^d),$$

where the $\rho(t)$ on the left hand side is the solution to (62).

We explicitly compute the gradient for test functions $\mathcal{E}$ and $d^2(\cdot,\gamma)$ next.

According to (37) and (38) of [23], if $\mathcal{E}(\rho_0) < +\infty$ and $\rho(t)$ is a solution to (62), then for each $p \in C_c^\infty(R^d)$,

$$\lim_{t\to 0+} \frac{\mathcal{E}(\rho(t)) - \mathcal{E}(\rho_0)}{t} = -\int_{R^d} \frac{1}{2}\Delta p(x)\rho_0(dx) + \int_{R^d} \nabla\Psi(x)\cdot\nabla p(x)\rho_0(dx),$$

therefore

$$\mathrm{grad}\ \mathcal{E}(\rho_0) = -\frac{1}{2}\Delta\rho_0 - \nabla\cdot(\rho_0\nabla\Psi) \in \mathcal{D}'(R^d). \tag{63}$$

Let $\rho_0, \gamma_0 \in E$, and ρ_0 have a Lebesgue density. According to mass transport theory (Theorem 9 in Appendix A), there exists a convex lower semi-continuous $\varphi_{\rho_0,\gamma_0} : R^d \mapsto R \cup \{+\infty\}$ such that $\pi_0(dx, dy) = \rho_0(dx)\delta_{\nabla\varphi_{\rho_0,\gamma_0}(x)}(dy)$ satisfies

$$d^2(\rho_0, \gamma_0) = \int |x-y|^2 d\pi_0 = \int |\nabla p_{\rho_0,\gamma_0}|^2 d\rho_0,$$

where $p_{\rho_0,\gamma_0}(x) = \frac{1}{2}|x|^2 - \varphi_{\rho_0,\gamma_0}(x)$. Let $p \in C_c^\infty(R^d)$ and $\rho(t)$ be the solution to (62). Following computations on page 11 of [23],

$$\begin{aligned}
&\lim_{t\to 0+} \frac{\frac{1}{2}d^2(\rho(t),\gamma) - \frac{1}{2}d^2(\rho_0,\gamma)}{t} \qquad (64)\\
&= \int (x-y)\cdot \nabla p(x)\pi_0(dx,dy)\\
&= \int (x - \nabla\varphi_{\rho_0,\gamma_0})(x)\cdot\nabla p(x)\rho_0(dx) = \int \nabla p_{\rho_0,\gamma_0}(x)\cdot\nabla p(x)\rho_0(dx).
\end{aligned}$$

Therefore

$$\text{grad}\frac{1}{2}d^2(\cdot,\gamma_0)(\rho_0) = -\nabla\cdot(\rho_0\nabla p_{\rho_0,\gamma_0}). \qquad (65)$$

Next, we define the *Fisher information* functional. Let ρ_∞ be the Gibbs measure in (22), for $\rho \in E$, we define

$$I(\rho) \equiv \begin{cases} \int \frac{|\nabla \frac{d\rho}{d\rho_\infty}|^2}{\frac{d\rho}{d\rho_\infty}} d\rho_\infty & \text{when } \nabla\frac{d\rho}{d\rho_\infty} \in L^1_{loc}(R^d), \\ +\infty & \text{otherwise;} \end{cases} \qquad (66)$$

where the convention $0/0 = 0$ is used.

We will prove the following in Appendix A. Let ρ have Lebesgue density, then (by (A.84))

$$g_{\rho_0}(\text{grad}d^2(\cdot,\gamma_0)(\rho_0), \text{grad}d^2(\cdot,\gamma_0)(\rho_0)) = 4\int_{R^d} |\nabla p_{\rho_0,\gamma_0}|^2 d\rho_0 = 4d^2(\rho_0,\gamma_0).$$

Let $\mathcal{E}(\rho) + I(\rho) < +\infty$, then (by (A.88))

$$g_\rho(\mathrm{grad}\mathcal{E}(\rho), \mathrm{grad}\mathcal{E}(\rho)) = \frac{1}{4} I(\rho);$$

moreover, $\nabla\rho \in L^1_{loc}(R^d)$ (Theorem 10) and by Lemma 6,

$$g_\rho(\mathrm{grad}\mathcal{E}(\rho), \mathrm{grad}d^2(\cdot,\gamma)(\rho)) = -\int_{R^d} \nabla p_{\rho,\gamma}(x) \cdot \frac{\nabla\rho + \rho\nabla(2\Psi)}{\rho}(x)\rho(x)dx.$$

As in the previous two examples, we now consider test functions

$$f_0(\rho) = (1-\kappa)md^2(\rho,\gamma_0) + \kappa\mathcal{E}(\rho) + c, \tag{67}$$

and

$$f_1(\gamma) = -m(1+\kappa)d^2(\rho_0,\gamma) - \kappa\mathcal{E}(\gamma) + c, \tag{68}$$

where $m \geq 0, 0 < \kappa \leq 1, \rho_0, \gamma_0 \in E$ and $c \in R$.

We now rigorously define Hf_0, Hf_1 (in view of the formal definition of H in (10) and Appendix A.2) as follows:

$$Hf_0(\rho) = \begin{cases} -m(1-\kappa)^2 \int_{R^d} \nabla p_{\rho,\gamma_0} \cdot \frac{\nabla\rho+\rho\nabla(2\Psi)}{\rho} d\rho \\ \quad +(1-\kappa)^2 2m^2 d^2(\rho,\gamma_0) - \kappa(1-\frac{\kappa}{2})\frac{1}{4}I(\rho), & \text{if } I(\rho) < +\infty \\ -\infty & \text{otherwise.} \end{cases} \tag{69}$$

and

$$Hf_1(\gamma) = \begin{cases} m(1+\kappa)^2 \int_{R^d} \nabla q_{\gamma,\rho_0} \cdot \frac{\nabla\gamma+\gamma\nabla(2\Psi)}{\gamma} d\gamma \\ \quad +(1+\kappa)^2 2m^2 d^2(\rho_0,\gamma) + \kappa(1+\frac{\kappa}{2})\frac{1}{4}I(\gamma), & \text{if } I(\gamma) < +\infty \\ +\infty & \text{otherwise.} \end{cases} \tag{70}$$

where

$$p_{\rho,\gamma_0}(x) = \frac{1}{2}|x|^2 - \varphi_{\rho,\gamma_0}(x), \quad q_{\gamma,\rho_0}(y) = \frac{1}{2}|y|^2 - \psi_{\gamma,\rho_0}(y)$$

and $\varphi_{\rho,\gamma_0}, \psi_{\gamma,\rho_0}$ are the lower semicontinuous convex functions

$$\nabla\varphi_{\rho,\gamma_0}\#\rho = \gamma_0, \quad \nabla\psi_{\gamma,\rho_0}\#\gamma = \rho_0.$$

See Definition 5 for the notation $\#$.

As before, by convexity in $\kappa \in (0,1)$,

$$Hf_0(\rho) \leq (1-\kappa)\Big(- m\int_{R^d} \nabla p_{\rho,\gamma_0} \cdot \frac{\nabla\rho + \rho\nabla(2\Psi)}{\rho} d\rho + 2m^2 d^2(\rho,\gamma_0)\Big) - \frac{\kappa}{8} I(\rho), \tag{71}$$

and

$$Hf_1(\gamma) \geq (1+\kappa)\Big(m\int_{R^d} \nabla q_{\gamma,\rho_0} \cdot \frac{\nabla\gamma + \gamma\nabla(2\Psi)}{\gamma} d\gamma + 2m^2 d^2(\rho_0,\gamma)\Big) + \frac{\kappa}{8} I(\gamma).$$

Like the previous two examples, the small perturbation $\kappa > 0$ introduces an extra higher order term I in Hf_0, Hf_1, and we expect it ensures semicontinuity property for Hf_0, Hf_1. This is made precise below.

Lemma 1. *Hf_0 in (69) is upper semicontinuous, and Hf_1 in (70) is lower semicontinuous in (E,d).*

Proof. The conclusion follows from Lemma 7 of Appendix A.

Let $\mathcal{E}(\rho_0) + \mathcal{E}(\gamma_0) < +\infty$. By mass transport inequality (A.98),

$$(1-\kappa)^{-1} Hf_0(\rho_0) - (1+\kappa)^{-1} Hf_1(\gamma_0) \leq -2L_\Psi m d^2(\rho_0,\gamma_0).$$

We note that if either $I(\rho_0) = +\infty$ or $I(\gamma_0) = +\infty$, the inequality is trivial. This verifies (44).

We now verify (51) and (52) in Condition 4. Let $\{e_1, \ldots, e_k, \ldots\} \subset C_c^\infty(R^d)$ be a dense subset of $C_b(R^d)$ in the bounded uniform convergence on compact topology. Let

$$m_k = (\sup_x(|e_k(x)| + |\Delta e_k(x)| + |\nabla e_k(x)|))^2.$$

We denote $\langle e_k, \rho\rangle = \int e_k(x)\rho(dx)$. For each $\rho_0 \in E$, we define

$$g_0(\rho) = \sum_k 2^{-k}(1+m_k)^{-1}\langle e_k, \rho - \rho_0\rangle^2. \tag{72}$$

Then

$$\text{grad} g_0(\rho) = -\nabla \cdot (\rho \nabla \frac{\delta g_0}{\delta \rho})$$

where

$$\frac{\delta g_0}{\delta \rho} = \sum_k 2^{-k+1}(1+m_k)^{-1}\langle e_k, \rho - \rho_0\rangle e_k.$$

In particular,

$$\Delta \frac{\delta g_0}{\delta \rho}, \nabla \frac{\delta g_0}{\delta \rho}, \frac{\delta g_0}{\delta \rho} \in C_c(R^d).$$

Therefore, the $H(f_0 + g_0)$ (where the H is given by (10)) is well defined. Moreover, (51) holds:

$$(H(f_0+g_0))^*(\rho_0) = (Hf_0)^*(\rho_0).$$

Similarly, for each $\gamma_0 \in E$, we can define

$$g_1(\gamma) = \sum_k \frac{1}{2^k}\langle \gamma - \gamma_0, e_k\rangle^2, \tag{73}$$

and verify that (52) holds.

In order to apply Theorem 5, we still need $\mathcal{E}$ to have compact level set in (E, d).

Lemma 2. *Let (21) hold. Then $\mathcal{E}$ has compact level set in (E, d).*

Proof. (21) implies the existence of a measurable function $G : [0, +\infty) \mapsto [0, \infty)$ and $\epsilon > 0$ such that $\lim_{t\to+\infty} t^{-1}G(t) = +\infty$ and $\int_{R^d} e^{\epsilon G(|x|^2)-2\Psi(x)}dx < +\infty$.

By a well-known variational representation of relative entropy (e.g. Corollary 6.2.3 and 6.2.13 of [11])

$$2\mathcal{E}(\rho) = \sup_{\varphi\in B(R^d)} \{\int_{R^d} \varphi d\rho - \log \int_{R^d} e^{\varphi} d\rho_\infty\}.$$

Hence for each $M > 0$,

$$\int_{R^d} \epsilon G(|x|^2) \wedge M d\rho \leq 2\mathcal{E}(\rho) + \log \int_{R^d} e^{\epsilon G(|x|^2)\wedge M} d\rho_\infty.$$

This implies

$$\sup_{\rho\in E, \mathcal{E}(\rho)\leq C} \int G(|x|^2)d\rho < +\infty.$$

Therefore, $\{\rho \in E : \mathcal{E}(\rho) \leq C\}$ is relatively compact in E for each $C < +\infty$.

Combine the above conclusions, we verified all the requirements of Theorem 5, and arrive at the following conclusion.

Theorem 8. *Let $h_0, h_1 \in C_b(E)$ be uniformly continuous with respect to d. Suppose (21) is satisfied.*

We assume $\overline{f}, \underline{f} \in C_b(E)$ is respectively viscosity subsolution of (42) and supersolution of (43), where $H_0 = H_1 = H$ and H is defined as above with

$\mathcal{D}(H)$ consists of test functions of the form $f_0, f_1, f_0 + g_0, f_1 - g_1$. Then

$$\sup_{\rho \in E}(\overline{f}(\rho) - \underline{f}(\rho)) \leq \sup_{\rho \in E}(h_0(\rho) - h_1(\rho)).$$

3.6. Existence of viscosity solutions

The remaining open question is the existence of sub and super solutions as well as viscosity solutions for the equation (2) which satisfy Condition 2, i.e. d-continuity on metric space (E, r). For all three model problems treated in Section 3, a complete detailed construction of sub- and super- solutions can be found in Feng and Kurtz [18]. The method there relies on a connection of the H operator with the probabilistic large deviation theory for Markov processes, on a generalization of the Barles-Perthame [2],[3] technique, and on operator extension techniques adapted to the viscosity solution context. It is important to point our that, in general infinite dimensional context, there are known examples where the Barles-Perthame procedure fails. The key ingredient to the generalization in [18] relies on a special property regarding resolvents for some convergent sequence of Hamilton-Jacobi equations. Using the probabilistic connection, such property can be translated into a uniform concentration of probability measures on compact set property (i.e. the exponential tightness) for the associated sequence of Markov processes. Such probabilistic property can be verified using a type of Lyapunov function technique for the three examples in this article.

Another direct approach to the existence issue can be based on dynamic programming principle for the optimal control problem (see Section 1.1) associated with the H operator. In finite dimensional optimal controlled problems, such method is standard (e.g. [17]). In the infinite dimensional context, certain important technical a priori estimates are needed. We discuss the details of such construction in a separate future publication.

Appendix A. A summary of mass transport theory with quadratic cost

We provide some technical estimates used in Example 3.5. Most of the material regarding mass transport theory are taken from Villani [37]. See also [19,28,14,1], and Appendix D of [18].

Throughout, E denote the space of probability measures on R^d with finite second moment.

Definition 5 (Push-forward of a probability measure). Let $T : R^d \to R^d$ be a measurable map, and $\rho_0 \in \mathcal{P}(R^d)$. The image measure of ρ_0 by T, denoted as $(T\#\rho_0)(A) \equiv \rho_0(T^{-1}(A))$ for all Borel measurable $A \subset R^d$, is called the *push-forward* of ρ_0.

Appendix A.1. The Kantorovich duality and mass transport maps

For each $\varphi : R^d \mapsto R \cup \{+\infty\}$, we introduce the Legendre transform

$$\varphi^*(y) = \sup_{x \in R^d} \{x \cdot y - \varphi(x)\}.$$

If $\varphi \not\equiv +\infty$, then φ^* is convex and lower semicontinuous. We also denote $\varphi^{**} = (\varphi^*)^*$.

Let $\rho_0, \gamma_0 \in \mathcal{P}(R^d)$, we define

$$\Phi(\rho_0, \gamma_0) \equiv \{(p, q) \in L^1(d\rho_0) \times L^1(d\gamma_0) : p(x) + q(y) \leq \frac{|x-y|^2}{2}\},$$

and

$$\Pi(\rho_0, \gamma_0) \equiv \{\pi(dx, dy) \in \mathcal{P}(R^d \times R^d), \pi(\cdot, R^d) = \rho_0(\cdot), \pi(R^d, \cdot) = \gamma_0(\cdot)\}.$$

Let d be defined as in (53):

$$d^2(\rho_0, \gamma_0) = \inf\{\int_{R^d \times R^d} |x-y|^2 \pi(dx, dy) : \pi \in \Pi(\rho_0, \gamma_0)\}.$$

Theorem 9. *Suppose $\rho_0, \gamma_0 \in E$. Then*

a). There exists $\pi_0 \in \Pi(\rho_0, \gamma_0)$ and a lower semicontinuous convex function $\varphi_{\rho_0,\gamma_0} \in M(R^d, \overline{R})$ such that

$$p_{\rho_0,\gamma_0}(x) = \frac{1}{2}|x|^2 - \varphi_{\rho_0,\gamma_0}(x), \quad q_{\rho_0,\gamma_0}(y) = \frac{1}{2}|y|^2 - \varphi^*_{\rho_0,\gamma_0}(y),$$

$(p_{\rho_0,\gamma_0}, q_{\rho_0,\gamma_0}) \in \Phi(\rho_0, \gamma_0)$, and that

$$\begin{aligned} \frac{1}{2}d^2(\rho_0, \gamma_0) &= \sup\{\int p d\rho + \int q d\gamma : (p, q) \in \Phi(\rho_0, \gamma_0)\} \qquad \text{(A.74)} \\ &= \int_{R^d} p_{\rho_0,\gamma_0}(x)\rho_0(dx) + \int_{R^d} q_{\rho_0,\gamma_0}(y)\gamma_0(dy) \\ &= \frac{1}{2}\int_{R^d \times R^d} |x-y|^2 \pi_0(dx, dy). \end{aligned}$$

b). Suppose moreover that ρ_0 has a Lebesgue density. The π_0 in part a) is unique. The π_0 and the $\varphi_{\rho_0,\gamma_0}$ in part a). have to satisfy

$$\pi_0(dx, dy) = \rho_0(dx)\delta_{\nabla\varphi_{\rho_0,\gamma_0}(x)}(dy), \qquad \text{(A.75)}$$

which implies $\int_{R^d} |\nabla p_{\rho_0,\gamma_0}(x)|^2 \rho_0(dx) = \int |x-y|^2 \pi_0(dx,dy) = d^2(\rho_0,\gamma_0)$*. Moreover,*

$$\nabla \varphi_{\rho_0,\gamma_0} \# \rho_0 = \gamma_0. \tag{A.76}$$

Furthermore, $\nabla \varphi_{\rho_0,\gamma_0}$ *is the unique* ($d\rho_0$*-almost everywhere) gradient of a convex function satisfying (A.76).*

Proof. The existence of π_0 and the first and the last equalities in (A.74) are known as the Kantorovich duality. See Theorem 1.3 of Villani [37]. The existence of $p_{\rho_0,\gamma_0}, q_{\rho_0,\gamma_0}$ and the second equality in (A.74) follows from Theorem 2.9 of [37].

Part b) follows from Theorem 2.12 of [37].

Appendix A.2. Weighted Sobolev spaces $H^1_\mu(R^d)$ *and* $H^{-1}_\mu(R^d)$

Let $\mu \in \mathcal{P}(R^d)$, we define

$$\|p\|^2_{1,\mu} = \int_{R^d} |\nabla p|^2 d\mu, \quad \forall p \in C_c^\infty(R^d)$$

and

$$H^1_\mu(R^d) = \text{the completion of } C_c^\infty(R^d) \text{ under } \|\cdot\|_{1,\mu}.$$

For each $u \in \mathcal{D}'(R^d)$, we define

$$\|u\|^2_{-1,\mu} \equiv \sup_{p \in C_c^\infty(R^d)} \{2\langle u,p\rangle - \|p\|^2_{1,\mu}\}, \tag{A.77}$$

we also introduce space

$$H^{-1}_\mu(R^d) \equiv \{\text{equivalent class of } u \in \mathcal{D}'(R^d) : \|u\|^2_{-1,\mu} < +\infty\}.$$

Direct verification shows that $(H^1_\mu(\mathcal{O}),\ \|\cdot\|_{1,\mu})$ and $(H^{-1}_\mu(\mathcal{O}),\ \|\cdot\|_{-1,\mu})$ are Hilbert spaces. The inner product and norm satisfy

$$(u,v)_{k,\mu} = \frac{1}{4}\{\|u+v\|^2_{k,\mu} - \|u-v\|^2_{k,\mu}\}, \quad k=-1,1. \tag{A.78}$$

Let $p,q \in H^1_\mu(R^d)$. For each q fixed, $l_q(p) = (p,q)_{1,\mu}$ is a linear functional in p satisfying

$$|(p,q)_{1,\mu}| \le \|p\|_{1,\mu}\|q\|_{1,\mu} = \|q\|_{1,\mu}\|\nabla p\|_{L^2(\mu)}, \quad \forall p \in C^\infty_c(R^d). \tag{A.79}$$

Next, we use this functional to extend the nation of ∇q from $q \in C^\infty_c(R^d)$ to all $q \in H^1_\mu(R^d)$. We denote such extension $\hat\nabla q$. The argument is extracted from Dawson and Gärtner [10].

Let L be the collection of maps $x \mapsto \nabla p(x)$ for $p \in C^\infty_c(R^d)$. L is a linear subset of $L^2(\mu)$. Let $L^2_{\mu,\nabla}(R^d)$ be the closure in $L^2(\mu)$ of L (elements in this space are equivalent classes). Since $p \leftrightarrow \nabla p$ is a one to one correspondence between $C^\infty_c(R^d)$ and L, l_q can be viewed as a linear functional on L as well. By (A.79), such functional can be linearly extended to all $p \in H^1_\mu(R^d)$ and the extension is bounded by $\|q\|_{1,\mu}$. By the Riesz representation theorem, there exists a unique $\xi \in L^2_{\mu,\nabla}(R^d)$ such that

$$l_q(p) = \int_{R^d} \xi \cdot \nabla p d\mu, \quad \forall p \in C^\infty_c(R^d). \tag{A.80}$$

We define such $\xi = \hat\nabla q$. We have

$$\|q\|^2_{1,\mu} = \sup_{p\in C^\infty_c(R^d)} \{2l_q(p) - \int_{R^d} |\nabla p|^2 d\mu\} = \|\hat\nabla q\|^2_{L^2(\mu)}.$$

The above construction gives us the following property

Lemma 3. *There is a one to one correspondence between $H^1_\mu(R^d)$ and $L^2_{\mu,\nabla}(R^d)$. For each $p, q \in H^1_\mu(R^d)$, we have representation*

$$(p, q)_{1,\mu} = \int_{R^d} \hat{\nabla} p \hat{\nabla} q d\mu = \int_{R^d} \hat{\nabla} p \hat{\nabla} q d\mu.$$

We have the following concerning $H^1_\mu(R^d)$ and $H^{-1}_\mu(R^d)$.

Lemma 4.

1. *For each $u \in H^{-1}_\mu(R^d)$, there exists a unique $p \in H_{1,\mu}(R^d)$ such that*

$$u = -\nabla \cdot (\mu \hat{\nabla} p) \in \mathcal{D}'(R^d).$$

2. *$H^{-1}_\mu(R^d)$ and $H^1_\mu(R^d)$ are dual space of each other.*
3. *For each*

$$u = -\nabla \cdot (\mu \hat{\nabla} p), \quad v = -\nabla \cdot (\mu \hat{\nabla} q) \tag{A.81}$$

where $p, q \in H^1_\mu(R^d)$,

$$(u, v)_{-1,\mu} = \int_{R^d} \hat{\nabla} p \cdot \hat{\nabla} q d\mu = \langle u, q \rangle = \langle p, v \rangle. \tag{A.82}$$

Proof. Let $u \in H^{-1}_\mu(R^d)$ and $\varphi \in C^\infty(R^d)$. By the definition of $\| \cdot \|_{-1,\mu}$,

$$2t\langle u, \varphi \rangle - t^2 \int_{R^d} |\nabla \varphi|^2 d\mu \le \|u\|^2_{-1,\mu} \quad t > 0,$$

implying

$$\langle u, \varphi \rangle \le \|\varphi\|_{1,\mu} \|u\|_{-1,\mu}. \tag{A.83}$$

Define $l(\varphi) \equiv \langle u, \varphi \rangle$, then l is a bounded linear functional on $H^1_\mu(R^d)$. By (A.83) and the Riesz representation theorem using the $H^1_\mu(R^d)$ inner product, there exists $p_0 \in H^1_\mu(R^d)$ such that

$$l(\varphi) = \int \nabla \varphi \hat{\nabla} p_0 d\mu, \quad \forall \varphi \in C^\infty_c(R^d).$$

Therefore

$$u = -\nabla \cdot (\mu \hat{\nabla} p_0).$$

Because of (A.83), $H^1_\mu(R^d) \subset \Big(H^{-1}_\mu(R^d)\Big)^*$ and $H^{-1}_\mu(R^d) \subset \Big(H^1_\mu(R^d)\Big)^*$. Since Hilbert space is reflexive,

$$\Big(H^1_\mu(R^d)\Big)^* = H^{-1}_\mu(R^d), \qquad \Big(H^{-1}_\mu(R^d)\Big)^* = H^1_\mu(R^d).$$

Assuming representation (A.81) holds, using approximation, the supreme in A.77 is given by

$$\|u\|^2_{-1,\mu} = \int_{R^d} |\hat{\nabla} p|^2 d\mu.$$

By (A.78), then (A.82) holds.

Appendix D of [18] explicitly identifies a number of interesting class of functions p where $\hat{\nabla} p = \nabla p$. ∇p denote the Schwartz distributional derivative. In particular, let $\rho(dx) = \rho(x)dx \in E$ have a Lebesgue density and $\gamma \in E$. Let φ be convex so that $\nabla\varphi\#\rho = \gamma$ (part b). of Theorem 9). Let $p(x) = \frac{1}{2}|x|^2 - \varphi(x)$. Then p is locally Lipschitz (by convexity) in the interior of the support of ρ and by the above discussions, $p \in H^1_\rho(R^d)$ and $\hat{\nabla} p = \nabla p$ and

$$\|p\|^2_{1,\rho} = \int_{R^d} |\nabla p|^2 d\rho = d^2(\rho, \gamma). \tag{A.84}$$

Appendix A.3. The Fisher information functional

Let ρ_∞ be defined as in (22) $\rho_\infty(dx) = Z^{-1} e^{-2\Psi(x)} dx$, and the *Fisher information* function I be defined by (66). We additionally assume $\Psi \geq 0$.

When $\rho(dx) = \rho(x)dx$, $d\rho/d\rho_\infty(x) = Z\rho(x)e^{2\Psi(x)}$. Hence, provided either $\nabla\rho \in L^1_{loc}(R^d)$ or $\nabla(\rho e^{2\Psi}) \in L^1_{loc}(R^d)$, local integrability of the other also holds and

$$\nabla(\rho e^{2\Psi}) = e^{2\Psi}\nabla\rho + e^{2\Psi}\rho\nabla(2\Psi) \in L^1_{loc}(R^d). \tag{A.85}$$

Therefore, we also have

$$I(\rho) \equiv \begin{cases} \int_{R^d} \frac{|\nabla\rho(x)+2\rho(x)\nabla\Psi(x)|^2}{\rho(x)}dx & \text{when } \rho(dx) = \rho(x)dx, \nabla\rho \in L^1_{loc}(R^d) \\ \infty & \text{otherwise.} \end{cases}$$

At least formally, there are other equivalent representation of I. Let

$$I_1(\rho) \equiv \begin{cases} 4\int_{R^d} |\nabla\sqrt{\frac{d\rho}{d\rho_\infty}}|^2 d\mu_\infty & \text{when } \rho << \rho_\infty \text{ and } \nabla\sqrt{\frac{d\rho}{d\rho_\infty}} \in L^1_{loc}(R^d), \\ \infty & \text{otherwise;} \end{cases}$$

Then similar to the derivation of (A.85), $\nabla(\sqrt{\rho}e^{\Psi}) \in L^1_{loc}(R^d)$ is equivalent to $\nabla\sqrt{\rho} \in L^1_{loc}(R^d)$ and

$$\nabla(\sqrt{\rho(x)}e^{\Psi(x)}) = e^{\Psi(x)}\nabla\sqrt{\rho(x)} + \nabla\Psi(x)$$

therefore

$$I_1(\rho) \equiv \begin{cases} 4\int_{R^d} |\nabla\sqrt{\rho(x)} + \nabla\Psi(x)|^2 dx & \text{when } \rho(dx) = \rho(x)dx, \nabla\sqrt{\rho} \in L^2_{loc}(R^d), \\ \infty & \text{otherwise;} \end{cases}$$

The functional $4^{-1}I_1 : E \mapsto [0, +\infty]$ is a Dirichlet form associated with a reversible Markov process. More specifically, let Y be the solution to stochastic differential equation

$$dY(t) = -2\nabla\Psi(Y(t))dt + \sqrt{2}dW(t).$$

Then ρ_∞ is the stationary distribution for Y, and Y is reversible with respect to ρ_∞ in the sense that

$$\int fT(t)g d\rho_\infty = \int gT(t)f d\rho_\infty, \quad \forall f, g \in C_b(R^d)$$

where $T(t)f(x) = E[f(Y(t))|Y(0) = x]$ for $f \in B(R^d)$. Let B be the weak infinitesimal generator of the continuous time process Y. That is, for $f \in \mathcal{D}(B)$,

$$\sup_{t>0} t^{-1} \sup_{x\in R^d} |T(t)f(x) - f(x)| < +\infty, \quad Bf(x) = \lim_{t\to 0+} t^{-1}(T(t)f(x) - f(x)), x \in R^d.$$

We can identify

$$B\varphi(x) = \Delta\varphi(x) - 2\nabla\Psi(x)\nabla\varphi(x), \quad \forall \varphi \in C_c^\infty(R^d).$$

We define

$$I_B(\rho) = -\inf_{\varphi\in\mathcal{D}(B), \inf_y \varphi(y)>0} \int \frac{B\varphi}{\varphi} d\rho. \tag{A.86}$$

By Theorem 7.44 of Stroock [34],

$$I_1(\rho) = 4 I_B(\rho). \tag{A.87}$$

The above I_1 is defined for all $\rho \in \mathcal{P}(R^d)$, I_1 is lower semicontinuous in the weak convergence of probability measure topology, and it is convex.

We introduce two more functionals: First, let

$$\begin{aligned} I_2(\rho) &= 4 \sup_{\psi\in C_c^\infty(R^d)} \langle -\Delta\psi + \nabla(2\Psi)\cdot\nabla\psi - |\nabla\psi|^2, \rho\rangle \\ &= 2 \sup_{\varphi\in C_c^\infty(R^d)} \langle -\Delta\varphi + \nabla\varphi\cdot\nabla(2\Psi) - \frac{1}{2}|\nabla\varphi|^2, \rho\rangle \\ &= \|\Delta\rho + \nabla(\rho\nabla(2\Psi))\|_{-1,\rho}^2. \end{aligned}$$

The last identity follows from the definition of $\| \cdot \|_{-1,\rho}$ norm in (A.77). By definition $I_2 \leq I_1$, $I_2 \geq 0$, I_2 is convex and is lower semicontinuous in the weak convergence of probability measure topology.

We also define

$$I_3(\rho) = 2 \sup_{\xi \in C_c^\infty(R^d)} \langle -\text{div}\xi + 2\xi \cdot \nabla \Psi - \frac{1}{2}|\xi|^2, \rho \rangle,$$

where ξ is an R^d-valued function. Clearly, $I_3 \geq I_2$.

Appendix D.6 of Feng and Kurtz [18] proves that all these functionals are equivalent.

Lemma 5.

$$I(\rho) = I_1(\rho) = I_2(\rho) = I_3(\rho), \quad \rho \in E.$$

Let $\mathcal{E}$ be the free energy function defined by (23) and let gradient be defined according to Definition 4 . Then by (63) and the definition of I_2,

$$I(\rho) = 4\|\text{grad}\mathcal{E}(\rho)\|_{-1,\rho}^2 \quad \text{whenever } \mathcal{E}(\rho) < \infty. \tag{A.88}$$

We summarize the above results.

Theorem 10. *(A.88) hold. I is lower semicontinuous in E. If $I(\rho) < \infty$, then $\rho(dx) = \rho(x)dx$ has a Lebesgue density and $\nabla\sqrt{\rho} \in L^2_{loc}(R^d)$ and $\nabla\rho \in L^1_{loc}(dx)$. Furthermore*

$$\nabla\sqrt{\rho} = \frac{1}{2}\frac{\nabla\rho}{\sqrt{\rho}} \in L^2_{loc}(R^d). \tag{A.89}$$

In view of the explicit expression of grad $\mathcal{E}$ in (63 and grad $d^2(\cdot,\gamma)$ in (65), and in view of Lemma 4, we expect

$$\langle -\text{grad}\mathcal{E}(\rho), \text{grad}\frac{1}{2}d^2(\cdot,\gamma)(\rho)\rangle_{-1,\rho} \tag{A.90}$$

$$= -\int_{R^d} \nabla p_{\rho,\gamma}(x) \cdot (\frac{1}{2}\frac{\nabla \rho(x)}{\rho(x)} + \nabla \Psi)\rho(x)dx,$$

where $p_{\rho,\gamma}(x) = |x|^2/2 - \varphi_{\rho,\gamma}(x)$ and $\varphi_{\rho,\gamma}$ is the lower semicontinuous convex function $\nabla \varphi_{\rho,\gamma} \# \rho = \gamma$ (Theorem 9). We prove this next.

Lemma 6. *Let $\mathcal{E}(\rho) + I(\rho) < \infty$ and $\gamma \in E$, then (A.90) holds.*

Proof. By approximation, there exists a sequence of $p_n \in C_c^\infty(R^d)$ such that $\lim_{n\to\infty} \|p_n - p_{\rho,\gamma}\|_{1,\rho} = 0$, or equivalently (by (A.82))

$$\lim_{n\to\infty} \| - \nabla(\rho \nabla p_n) - \text{grad}\frac{1}{2} d^2(\rho, \gamma)\|_{-1,\rho} = 0.$$

Since $-\nabla(\rho\nabla p_n) \in H_\rho^{-1}(R^d)$, replacing the Ψ by $\Psi + p_n$ and apply (A.88), using the definition of I,

$$-\int_{R^d} |\frac{1}{2}\frac{\nabla \rho(x)}{\rho(x)} + \nabla(\Psi + p_n)|^2 \rho(x)dx = \| - \text{grad}\mathcal{E}(\rho) + (-\nabla(\rho\nabla p_n))\|_{-1,\rho}^2.$$

Similarly, replacing Ψ by $\Psi - p_n$,

$$-\int_{R^d} |\frac{1}{2}\frac{\nabla \rho(x)}{\rho(x)} + \nabla(\Psi - p_n)|^2 \rho(x)dx = \| - \text{grad}\mathcal{E}(\rho) + (-\nabla(\rho\nabla p_n))\|_{-1,\rho}^2.$$

Therefore, by the polarization identity ($H_\rho^{-1}(R^d)$ is a Hilbert space)

$$\langle -\text{grad}\mathcal{E}(\rho), -\nabla(\rho\nabla p_n)\rangle_{-1,\rho} = -\int_{R^d} \nabla p_n(x) \cdot (\frac{1}{2}\frac{\nabla \rho(x)}{\rho(x)} + \nabla \Psi)\rho(x)dx.$$

Sending $n \to \infty$, we arrive at (A.90).

Appendix A.4. Mass transport inequalities and regularities

Lemma 7. *Suppose that $\sup_{n=1,2,\ldots} I(\rho_n) < +\infty$ and that $\lim_{n\to+\infty} d(\rho_n, \rho_0) = 0$ where $\rho_0 \in E$. (By lower semicontinuity of I, this implies $I(\rho_0) < +\infty$, by* *definition of I, this implies that $\rho_n(dx) = \rho_n(x)dx$ and $\rho_0(dx) = \rho_0(x)dx$ have Lebesgue densities). Further suppose that $\gamma_n, \gamma_0 \in E$ have Lebesgue densities and $\lim_{n\to+\infty} d(\gamma_n, \gamma_0) = 0$.*

Then

$$\lim_{n\to\infty} \int_{R^d} \frac{\nabla\rho_n + \rho_n\nabla(2\Psi)}{\sqrt{\rho_n}} \cdot (\nabla p_{\rho_n,\gamma_n}\sqrt{\rho_n})dx \tag{A.91}$$
$$= \int_{R^d} \frac{\nabla\rho_0 + \rho_0\nabla(2\Psi)}{\sqrt{\rho_0}} (\nabla p_{\rho_0,\gamma_0}\sqrt{\rho_0})dx,$$

where $p_{\rho_n,\gamma_n}, p_{\rho_0,\gamma_0}$ are defined following the notational convention in Theorem 9.

Note that by $\sup_{n=0,1,\dots} I(\rho_n) < +\infty$ and by (A.84), the integrals in (A.91) are well defined finite quantities.

Proof. To simplify notation, let

$$\xi_n = \frac{\nabla\rho_n + \rho_n\nabla(2\Psi)}{\sqrt{\rho_n}} = 2(\nabla\sqrt{\rho_n} + \sqrt{\rho_n}\nabla\Psi) \in L^2(R^d)$$

and

$$\xi_0 = \frac{\nabla\rho_0 + \rho_0\nabla(2\Psi)}{\sqrt{\rho_0}} = 2(\nabla\sqrt{\rho_0} + \sqrt{\rho_0}\nabla\Psi) \in L^2(R^d).$$

Since

$$|\int_{R^d} \xi_n(\nabla p_{\rho_n,\gamma_n}\sqrt{\rho_n})dx - \int_{R^d} \xi_0(\nabla p_{\rho_0,\gamma_0}\rho_0)dx|$$
$$\le |\int_{R^d} \xi_n(\nabla p_{\rho_n,\gamma_n}\sqrt{\rho_n} - \nabla p_{\rho_0,\gamma_0}\sqrt{\rho_0})dx| + |\int_{R^d}(\xi_n - \xi_0)\nabla p_{\rho_0,\gamma_0}\sqrt{\rho_0}dx|,$$

and since

$$\sup_{n=0,1,\dots}\int_{R^d}|\xi_n|^2 dx = \sup_{n=0,1,\dots} I(\rho_n) < \infty,$$
$$\sup_{n=0,1,\dots}\int_{R^d}|\nabla p_{\rho_n,\gamma_n}\sqrt{\rho_n}|^2 dx = \sup_{n=0,1,\dots} d^2(\rho_n,\gamma_n) < \infty,$$

(A.91) holds if $\{\xi_n\}$ converges weakly to ξ_0 in $L^2(R^d)$:

$$\langle \xi_n, q\rangle \to \langle \xi_0, q\rangle, \quad q \in L^2(R^d), \tag{A.92}$$

and the following norm convergence in $L^2(dx)$ holds:

$$\nabla p_{\rho_n,\gamma_n}\sqrt{\rho_n} \to \nabla p_{\rho_0,\gamma_0}\sqrt{\rho_0}. \tag{A.93}$$

Indeed, because

$$\int_{R^d} |\sqrt{\rho_n}\nabla p_{\rho_n,\gamma_n}|^2 dx = d^2(\rho_n,\gamma_n) \to d^2(\rho_0,\gamma_0) = \int_{R^d} |\sqrt{\rho_0}\nabla p_{\rho_0,\gamma_0}|^2 dx,$$

even (A.93) is implied by the seemingly weaker statement:

$$\sqrt{\rho_n}\nabla p_{\rho_n,\gamma_n} \rightharpoonup^w \sqrt{\rho_0}\nabla p_{\rho_0,\gamma_0} \quad \text{in } L^2(dx). \tag{A.94}$$

By

$$\begin{aligned}\sup_n \int_{R^d} (|\sqrt{\rho_n}|^2 + |\xi_n|^2) dx &= \sup_n \int_{R^d} (|\sqrt{\rho_n}|^2 + 4|\sqrt{\rho_n} + \sqrt{\rho_n}\nabla\Psi|^2) dx \\ &= \sup_n (1 + I(\rho_n)) < \infty,\end{aligned}$$

if we denote $\mathcal{O}_m = \{x \in R^d : |x| < m\}$ for $m = 1, 2, \ldots$, then for each m fixed,

$$\sup_n \int_{\mathcal{O}_m} (|\sqrt{\rho_n}|^2 + |\nabla\sqrt{\rho_n}|^2) dx < \infty.$$

That is, $\{\sqrt{\rho_n}\}$ is bounded in $H^1(\mathcal{O}_m)$. Therefore, for the m fixed, by selecting subsequences if necessary, we can find $\eta \in L^2(\mathcal{O}_m)$ such that

$$\lim_{n\to\infty} \|\sqrt{\rho_n} - \eta\|_{L^2(\mathcal{O}_m)} = 0, \quad \nabla\sqrt{\rho_n} \rightharpoonup^w \nabla\eta \text{ in } L^2(\mathcal{O}_m).$$

In particular, this implies

$$\int_{\mathcal{O}_m} |\rho_n - \eta^2| dx \le \|\sqrt{\rho_n} - \eta\|_{L^2(\mathcal{O}_m)} \|\sqrt{\rho_n} + \eta\|_{L^2(\mathcal{O}_m)} \to 0.$$

By assumption, ρ_n converges to ρ_0 in the 2-Wasserstein metric, we can identify that $\eta^2(x) = \rho_0(x)$ for $x \in \mathcal{O}_m$ a.e-dx. Hence

$$\nabla\sqrt{\rho_n} \rightharpoonup^w \nabla\sqrt{\rho_0} \text{ in } L^2(\mathcal{O}_m).$$

By the arbitrariness of m, (A.92) follows. Furthermore,

$$\lim_{n\to\infty} \|\sqrt{\rho_n} - \sqrt{\rho_0}\|_{L^2(\mathcal{O}_m)} = 0, \quad \forall m = 1, 2, \ldots. \tag{A.95}$$

We will need this to verify (A.94) next.

Since

$$\|(\nabla p_{\rho_n,\gamma_n})\sqrt{\rho_n}\|^2_{L^2(R^d)} = \int_{R^d} |\nabla p_{\rho_n,\gamma_n}|^2 d\rho_n = d^2(\rho_n, \gamma_n) \le C < \infty,$$

$\{\sqrt{\rho_n}\nabla p_{\rho_n,\gamma_n}\}$ is uniformly bounded in $L^2(R^d)$. For any $L^2(R^d)$-weakly convergent limit point $\xi \in L^2(R^d)$, at least along subsequences

$$\lim_{n\to\infty} \int_{R^d} (\nabla p_{\rho_n,\gamma_n})\sqrt{\rho_n}\eta dx = \int_{R^d} \xi\eta dx, \quad \eta \in L^2(R^d).$$

In particular, since $\sqrt{\rho_0} \in L^2(R^d)$, for each $\chi \in C_c^\infty(R^d)$,

$$\lim_{n\to\infty} \int_{R^d} (\nabla p_{\rho_n,\gamma_n})\sqrt{\rho_n}(\sqrt{\rho_0} \cdot \chi)dx = \int_{R^d} \xi \cdot (\sqrt{\rho_0}\chi)dx.$$

Combined with (A.95), this implies

$$\lim_{n\to\infty} \int_{R^d} (\nabla p_{\rho_n,\gamma_n})\sqrt{\rho_n} \cdot (\sqrt{\rho_n}\chi)dx = \int_{R^d} \xi\sqrt{\rho_0} \cdot \chi dx, \quad \chi \in C_c^\infty(R^d). \tag{A.96}$$

This is because for m large enough so that $\mathcal{O}_m \supset \text{supp}(\chi)$,

$$\begin{aligned} &|\int_{R^d} (\nabla p_{\rho_n,\gamma_n})\sqrt{\rho_n}(\sqrt{\rho_n} - \sqrt{\rho_0}) \cdot \chi dx| \\ &\le \Big(\int_{R^d} |\nabla p_{\rho_n,\gamma_n}|^2 d\rho_n\Big)^{1/2} \|\sqrt{\rho_n} - \sqrt{\rho_0}\|_{L^2(\mathcal{O}_m)} \sup_x |\chi(x)| \\ &= d(\rho_n, \gamma_n)\|\sqrt{\rho_n} - \sqrt{\rho_0}\|_{L^2(\mathcal{O}_m)} \sup_x |\chi(x)| \to 0, \end{aligned}$$

where the limit is taken as $n \to \infty$.

Following the notations of Theorem 9, we define

$$\pi_n(dx,dy) = \delta_{\nabla\varphi_{\rho_n,\gamma_n}(x)}(dy)\rho_n(x)dx, \quad \pi_0(dx,dy) = \delta_{\nabla\varphi_{\rho_0,\gamma_0}(x)}(dy)\rho_0(x)dx.$$

Since $d(\rho_n,\rho_0) + d(\gamma_n,\gamma_0) \to 0$, $\sup_n \int_{R^d\times R^d}(|x|^2 + |y|^2)d\pi_n < \infty$, hence π_n is tight. By Lemma 9, Proposition 10 of McCann [28], and Rockafellar's theorem [30] on convex function characterization of cyclically monotone set in $R^d \times R^d$, $\pi_n \Rightarrow \pi_0$ in the topology of weak convergence of probability measures. Therefore,

$$\begin{aligned}
&\int_{R^d} \nabla p_{\rho_0,\gamma_0}(x)\cdot\chi(x)\rho_0(x)dx = \int_{R^d\times R^d}(x-y)\cdot\chi(x)d\pi_0(dx,dy) \\
&= \lim_{n\to\infty}\int_{R^d\times R^d}(x-y)\cdot\chi(x)d\pi_n(dx,dy) = \lim_{n\to\infty}\int_{R^d}\nabla p_{\rho_n,\gamma_n}(x)\cdot\chi(x)\rho_n(x)dx \\
&= \int_{R^d}\xi(x)\sqrt{\rho_0(x)}\cdot\chi(x)dx, \qquad \chi\in C_c^\infty(R^d),
\end{aligned}$$

where the last step follows from (A.96). Hence $\xi = \nabla p_{\rho_0,\gamma_0}\sqrt{\rho_0}$, giving (A.94).

By interpolation inequality (Theorem 9.17 of [37]),

$$\mathcal{E}(\rho) \le \mathcal{E}(\gamma) + \frac{1}{2}d(\rho,\gamma)\sqrt{I(\rho)} - \frac{L_\Psi}{2}d^2(\rho,\gamma),$$

where $L_\Psi \in R$ is the constant in (20). In particular, the finiteness of $I(\rho)$ and $\mathcal{E}(\gamma)$ implies the finiteness of $\mathcal{E}(\rho)$. The following is a refinement of the above inequality.

Lemma 8. *Let $\rho,\gamma \in E$ both have Lebesgue density*

$$\rho(dx) = \rho(x)dx, \quad \gamma(dy) = \gamma(y)dy.$$

We denote $p_{\rho,\gamma}, q_{\rho,\gamma}$ following the notations in Theorem 9. Then

1. if $I(\rho) < +\infty$,

$$\mathcal{E}(\gamma) \geq \mathcal{E}(\rho) + \int_{R^d} \nabla p_{\rho,\gamma}(x) \cdot (\frac{1}{2}\frac{\nabla \rho(x)}{\rho(x)} + \nabla \Psi(x))\rho(x)dx + \frac{L_\Psi}{2} d^2(\rho,\gamma). \tag{A.97}$$

2. if $\mathcal{E}(\rho) + \mathcal{E}(\gamma) + I(\rho) + I(\gamma) < +\infty$, then

$$\int_{R^d} \nabla p_{\rho,\gamma}(x) \frac{\nabla \rho + \rho \nabla(2\Psi)}{\rho}(x)\rho(x)dx + \int_{R^d} \nabla q_{\rho,\gamma}(y) \frac{\nabla \gamma + \gamma \nabla(2\Psi)}{\gamma}(y)\gamma(y)dy \geq 2L_\Psi d^2(\rho,\gamma). \tag{A.98}$$

Proof. By symmetry, we only need to prove (A.97).

When both ρ, γ are compactly supported, (A.97) is Theorem 4.1 of Cordero-Erausquin, Gangbo and Houdré [5]. The general case follows by approximating general probability density functions by densities with compact support and by using the convergence result of Lemma 7, and by the lower semicontinuity of $\mathcal{E}$.

Appendix B. Connections with large deviation theory

The H operators in Examples 1, 2 and 3 also arise naturally from large deviation of stochastic interacting particle systems. Indeed, such large deviation connection gives a direct approach to the construction of sub and super viscosity solution of (2). For a rigorous exposition on this approach in the context of general Markov processes, see [18].

Let $\rho_n(t)$, $0 \le t < +\infty$ be an E-valued Markov process with trajectories in space $C([0,\infty); E)$. By *the large deviation principle* for the $\{\rho_n : n = 1, 2, \ldots\}$, we mean the following probabilistic asymptotic is satisfied:

$$\lim_{n \to +\infty} \frac{1}{n} \log P(\rho_n(\cdot) \in B) = - \inf_{\rho \in B} I(\rho) \tag{B.99}$$

for all "nice" measurable set $B \subset C([0,\infty), E)$. Or heuristically,

$$P(\rho_n(\cdot) \in B) \sim Z_n^{-1} \exp\{-n \inf_{\rho \in B} I(\rho)\},$$

where Z_n is some normalizing constant growing at sub-exponential rate of n. $I(\cdot) : C([0,\infty); E) \to [0, +\infty]$ is called *the action functional*.

The connection between large deviation (B.99) and Hamilton-Jacobi equation (2) can roughly be described as follows: Let A_n be the linear generator [12] of the process ρ_n. We introduce transformed nonlinear operator $H_n f \equiv \frac{1}{n} e^{-nf} A_n e^{nf}$. Suppose that $H_n \to H$ in a certain sense and that $\rho_n(t)$ is concentrated in compact set of E with high probability (see [18]), then there exists bounded upper semicontinuous subsolution and bounded lower semicontinuous supersolution to

$$(I - \alpha H) f = h,$$

where $\alpha > 0, h \in C_b(E)$. Moreover, if the comparison principle holds for "sufficiently many" hs, then the large deviation principle (B.99) holds. Furthermore, if the H corresponds to a control problem, then the action functional I can be identified as the running cost of the control structure.

Example 4. Let $\mathcal{O}, F$ be defined according to Example 1, and let ρ_n be the solution to stochastic PDE

$$\frac{\partial}{\partial t}\rho_n = \Delta\rho_n - F^{'}(\rho_n(t,x)) + \frac{1}{\sqrt{n}}\frac{W(\partial t, \partial x)}{\partial t \partial x}, \tag{B.100}$$

where $W(t,x), (t,x) \in [0,\infty) \times \mathcal{O}$ is a scalar valued time-space Brownian sheet.

If we view $\rho_n(t)$ as $E = L^2(\mathcal{O})$-valued diffusions, and take test functions of the form (13), then by Ito's formula

$$H_n f(\rho) = \sum_{i=1}^m \partial_i \varphi(\langle \rho, \xi \rangle)\langle \Delta\rho - F^{'}(\rho), \xi_i\rangle + \frac{1}{2}\sum_{i,j=1}^m \partial_i\varphi(\langle\rho,\xi\rangle)\partial_j\varphi(\langle\rho,\xi\rangle)\langle\xi_i,\xi_j\rangle$$
$$+\frac{1}{2n}\sum_{i,j=1}^m \partial^2_{ij}\varphi(\langle\rho,\xi\rangle)\langle\xi_i,\xi_j\rangle.$$

Consequently, $\lim_{n\to+\infty} H_n f(\rho_n) = Hf(\rho)$ whenever $\lim_{n\to+\infty}\|\rho_n - \rho\|_{L^2(\mathcal{O})} = 0$, where H is defined by (15).

Example 5. Now we consider a stochastically perturbed Cahn-Hilliard equation, which models the evolution of conserved quantity $\int \rho_n(t,x)dx =$ constant:

$$\frac{\partial}{\partial t}\rho_n = \nabla\cdot\Big(\nabla(-\Delta\rho_n + F^{'}(\rho_n)) + \frac{1}{\sqrt{n}}\frac{\mathbf{W}(\partial t,\partial x)}{\partial t\partial x}\Big) \tag{B.101}$$

where $\mathbf{W}(t,x) = (W_1(t,x),\ldots,W_d(t,x))$ is an R^d-valued Brownian sheet and F is as in Example 2.

Similar to Example 4, taking test functions of the form (13) and apply Ito's formula,

$$\begin{aligned} H_n f(\rho) = &\sum_{i=1}^m \partial_i \varphi(\langle \rho, \xi \rangle) \langle \Delta(-\Delta\rho + F^{'}(\rho(x))), \xi_i \rangle \\ &+ \frac{1}{2} \sum_{i,j=1}^m \partial_i \varphi(\langle \rho, \xi \rangle) \partial_j \varphi(\langle \rho, \xi \rangle) \int_{\mathcal{O}} \nabla \xi_i \cdot \nabla \xi_j dx \\ &+ \frac{1}{2n} \sum_{i,j=1}^m \partial_{ij}^2 \varphi(\langle \rho, \xi \rangle) \int_{\mathcal{O}} \nabla \xi_i \cdot \nabla \xi_j dx. \end{aligned}$$

Therefore $\lim_{n\to+\infty} H_n f(\rho_n) = Hf(\rho)$ whenever $\lim_{n\to+\infty} \|\rho_n - \rho\|_{L^2(\mathcal{O})} = 0$, where H is given by (19).

In general, stochastic partial differential equations (B.100) and (B.101) do not always have function valued solutions, the calculations above are therefore only formal. To make the argument rigorous, for the practical purpose here, we can discretize the space and the Brownian sheet on a finite lattice. Then, as n goes to infinity, provided the lattice mesh size go to zero slowly enough, the same H can still be derived as a limit from the H_ns corresponding to the discrete system.

Example 6. Let $\Phi, \Psi \in C^2(R^d)$ and $\Phi(z) = \Phi(-z)$. We consider a system of weakly interacting particles governed by the stochastic differential equations:

$$dX_{n,i}(t) = -\nabla\Psi\Big(X_{n,i}(t)\Big) - \frac{1}{n}\sum_{j=1}^n \nabla\Phi\Big(X_{n,i}(t) - X_{n,j}(t)\Big)dt + dW_i(t), \tag{B.102}$$

where $W_i(t), i = 1, 2, \ldots, n$ are R^3-valued independent standard Brownian motions. Because each particle plays a symmetric role in (B.102), instead of the vector valued process $(X_{n,1}(t),\dots,X_{n,n}(t))$, we can equivalently consider the process of empirical measures $\rho_n(t,dx) = n^{-1}\sum_{k=1}^n \delta_{X_{n,k}(t)}(dx)$. Apply Ito's formula, for the f in (13) with $\xi_k \in C_c^\infty(R^d)$,

$$H_n f(\rho) = \sum_{i=1}^m \partial_i \varphi(\langle \rho, \xi\rangle)\langle \rho, B(\rho)\xi_i\rangle + \frac{1}{2}\sum_{ij=1}^m \partial_i\varphi(\langle\rho,\xi\rangle)\partial_j\varphi(\langle\rho,\xi\rangle)\langle\rho,\nabla\xi_i\cdot\nabla\xi_j\rangle$$
$$+\frac{1}{2n}\sum_{i,j=1}^m \partial_{ij}^2\varphi(\langle\rho,\xi\rangle)\langle\rho,\nabla\xi_i\cdot\nabla\xi_j\rangle,$$

where

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

Let d be the order-2-Wasserstein metric. Suppose $\lim_n d(\rho_n,\rho) = 0$ in the topology of weak convergence of probability measures, and that $\nabla\Phi(z)$ has sub-quadratic growth as z approaches infinity. Then $H_n f(\rho_n) \to Hf(\rho)$ where

$$Hf(\rho) = \sum_{i=1}^k \partial_i\varphi(\langle \mathbf{p},\rho\rangle)\langle\rho, B(\rho)p_i\rangle + \frac{1}{2}\sum_{i,j=1}^k \partial_i\varphi(\langle\mathbf{p},\rho\rangle)\partial_j\varphi(\langle\mathbf{p},\rho\rangle)\langle\rho,(\nabla p_i)^T\nabla p_j\rangle$$
$$= \langle\rho, B(\rho)\frac{\delta f}{\delta\rho}\rangle + \frac{1}{2}\int_{R^d}|\nabla\frac{\delta f}{\delta\rho}|^2 d\rho = \langle B^*(\rho)\rho, \frac{\delta f}{\delta\rho}\rangle + \frac{1}{2}\int_{R^d}|\nabla\frac{\delta f}{\delta\rho}|^2 d\rho.$$

Setting $\Phi = 0$, we arrive at the H in (26) as a special case. In such situation, each particle in (B.102) evolve independently of the other, with transition probability density governed by the Fokker-Planck equation given by generator $B^*\rho = \frac{1}{2}\Delta\rho + \nabla(\rho\nabla\Psi)$.

We assumed $\Phi = 0$ in the proof of the comparison Theorem 8. The general case, however, follows from minor modifications of the same proof.

A careful comparison between the above Hs and those in the comparison Theorems 6, 7 and 8 reveals a gap. The H in the comparison theorem is always "larger" than that we derived here, in the sense that the larger H contains a richer class of test functions (40) and (41). These test functions played key roles in the proof of the comparison results, and they are semi-continuous but not continuous. Indeed, Definition 1 of viscosity solution is inappropriate to use for equation defined by the "smaller" H, because there is no a priori guarantee that those extremal points in the definition exists any more. Such a gap can be filled either by directly working with test functions (40) and (41) when deriving the limit of H_n, or by a *viscosity extension* technique in [18]. By this technique, we prove that any sub- (respectively super-) viscosity solution of (2) for the smaller H in a different sense, is also a sub (respectively super-) solution of the equation for the larger H in the sense of Definition 1. See Chapter 13 of [18] for detail. Viscosity extension for the case of the probability-measure-valued state space example is so involved that a whole section 13.3.3 in [18]is devoted for the proof.

Acknowledgements. We thank Djivede Kelome and Thomas G. Kurtz for useful discussions and two referees for constructive comments. The reference of Ishii [22] was pointed out to us by one of the referees.

References

1. L. AMBROSIO, N. GIGLI AND G. SAVARÉ: *Gradient flows in metric spaces and in the space of probability measures.* Birkhäuser Verlag, Basel Boston Berlin, 2005.

2. G. BARLES, AND B. PERTHAME: Discontinuous solutions of deterministic optimal stopping time problems. *Math. Methods Num. Anal.* **21** 557-579, (1987).

3. G. Barles, and B. Perthame: Exit time problems in optimal control and vanishing viscosity method. *SIAM J. Control Optim.* **26** 1133-1148, (1988).

4. P. Cannarsa, and M. E. Tessitore: Infinite dimensional Hamilton-Jacobi equations and Dirichlet boundary control problems in parabolic type. *SIAM J. Control Optim.* **34** 1831–1847, (1996).

5. D. Cordero-Erausquin, W. Gangbo, and C. Houdre: Inequalities for generalized entropy and optimal transportation. *Recent Advances in the theory and applications of mass transport*, 73-94 *Contemporary Mathematics*, 353. AMS. Providence, RI 2004.

6. M.G. Crandall, and P-L Lions: Viscosity solutions of Hamilton-Jacobi Equations. *Trans. A. M. S.* **277** 1-42, (1983).

7. M.G. Crandall, and P-L Lions: Hamilton-Jacobi Equations in Infinite Dimensions, PART I *J. Funct. Anal.* **62** 3. 379-396, (1985). PART II *J. Funct. Anal.* **65** 3. 368-405, (1986). PART III *J. Funct. Anal.* **68** 2. 214-247, (1986). PART IV *J. Funct. Anal.* **90** 2. 237-283, (1990). PART V *J. Funct. Anal.* **97** 2. 417-465, (1991). PART VII *J. Funct. Anal.* **125** 1. 111-148, (1994).

8. M. G. Crandall, and P-L Lions: Hamilton-Jacobi Equations in Infinite Dimensions, PART VI: Nonlinear A and Tataru's Method Refined. *Evolution equations, control theory, and biomathematics (Han sur Lesse, 1991), 51–89, Lecture Notes in Pure and Appl. Math.* **155** Dekker, New York, 1994.

9. M. G. Crandall, H. Ishii, and P-L Lions: User's guide to viscosity solutions of second order partial differential equations. *Bulletin A.M.S., N.S.* **27** 1-67, (1992).

10. D.A. Dawson and J. Gärtner: Large deviation from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* **20** 247-308, (1987).

11. A. Dembo and O. Zeitouni: *Large Deviations Techniques Techniques and Applications*, Second Edition, Springer, New York, 1998.

12. S. N. Ethier, and T. G. Kurtz; *Markov Processes.* Wiley, New York, 1986.

13. I. EKELAND: Nonconvex minimization problems. *Bulletin A.M.S. (New Series)* **Vol I** no. 3, 443-474, (1979).

14. L.C. EVANS: Partial differential equations and Monge-Kantorovich mass transfer. *Curr. Develop. in Maths.* 1997.

15. W. H. FLEMING: Exit probabilities and optimal stochastic control. *Applied Math. Optimiz.* **4** 329-346, (1978).

16. W. H. FLEMING: A stochastic control approach to some large deviation problems *Recent mathematical methods in dynamic programming (Rome 1984) Lecture Notes in Math.* **1119**, Springer, Berlin-New York, 1985, 52-66.

17. W. H. FLEMING, AND H. M. SONER: *Controlled Markov processes and viscosity solutions.* Springer-Verlag, New York, 1991.

18. J. FENG, AND T. G. KURTZ: Large deviations of stochastic processes. *submitted, American Mathematical Society book series.* 2005.

19. W. GANGBO, AND J. R. MCCANN: The geometry of optimal transportation. *Acta Mathematica* **177** 113-161, (1996).

20. G. GIACOMIN, AND J. L. LEBOWITZ: Phase segregation dynamics in particle systems with long range interactions. I. Macroscopic limits, *J. Stat. Phys.* **87**, 37, (1997).

21. F. GOZZI, AND A. SWIECH: Hamilton-Jacobi-Bellman equations for the optimal control of the Duncan-Mortensen-Zakai equation. *J. Funct. Anal.* **172**, no. 2, 466–510, (2000).

22. H. ISHII: Viscosity solutions for a class of Hamilton-Jacobi equations in Hilbert spaces, *J. Funct. Anal.* **105**, 2, 301-341, (1992).

23. R. JORDAN, D. KINDERLEHRER, AND F. OTTO: The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **Vol. 29** No.1, 1-17, (1998).

24. C. KIPNIS, S. OLLA AND S.R.S. VARADHAN: Hydrodynamical limits and large deviations for simple exclusion processes. *Comm. Pure Appl, Math* XLII,

115-137 (1989)

25. C. Kipnis and C Landim: *Scaling limits of interacting particle systems* Springer, New York 1999.

26. J.S. Langer: Theory of spinodal decomposition in alloys, *Ann. Phys.* **65** 53, (1971).

27. P.L. Lions: Some properties of the viscosity semigroups for Hamilton-Jacobi equations. *Nonlinear differential equations.* Edited by J. K. Hale and P Martinez-Amores, Pitman Research Notes in Mathematics 132. 1988.

28. J. R. McCann: Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **Vol. 80** No.2, 309-323, (1995).

29. F. Otto: The geometry of dissipative evolution equations: the porous medium equation. *Comm. P.D.Es* **26** no. 1-2, 101-174, (2001).

30. R.T. Rockafellar: Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, **17** 497-510, (1966).

31. K. Sato: On the generators of non-negative contraction semi-groups in Banach lattices. *J. Math. Soc. Japan* **20** No. 3, 423-436, (1968).

32. E. Sinestrari: Accretive differential operators. *Bolletino U.M.I.,* **13**-B, (5), 19-31, (1976).

33. C. Stegall: Optimization of functions on certain subsets of Banach spaces. *Math. Ann.* **236**, 171-176, (1978).

34. D. Stroock: *An introduction to the theory of large deviations* Springer-Verlag, Berlin, 1984.

35. D. Tataru: Viscosity solutions of Hamilton-Jacobi equations with unbounded nonlinear terms. *J. Math. Anal. Appl.* **163** no. 2, 345–392, (1992).

36. S.R.S. Varadhan: Lectures on hydrodynamic scaling *Hydrodynamic limits and related topics* Fields Institute Communications. Editors: S. Feng A.T. Lawniczak, S.R.S. Varadhan, Published by AMS. 2000.

37. C. Villani: *Topics in mass transportation.* American Mathematical Society, Providence, Rhode Island. Graduate Studies in Mathematics. Vol 58. 2003.

Jin Feng
Department of Mathematics and Statistics
University of Massachusetts at Amherst
Amherst, MA 01002
email:feng@math.umass.edu

and

Markos Katsoulakis
Department of Mathematics and Statistics
University of Massachusetts at Amherst
Amherst, MA 01002
email:markos@math.umass.edu