WELL-POSEDNESS FOR HAMILTON-JACOBI EQUATIONS ON THE WASSERSTEIN SPACE ON GRAPHS

WILFRID GANGBO, CHENCHEN MOU, AND ANDRZEJ ŚWIĘCH

Abstract. In this manuscript, given a metric tensor on the probability simplex, we define differential operators on the Wasserstein space of probability measures on a graph. This allows us to propose a notion of graph individual noise operator and investigate Hamilton-Jacobi equations on this Wasserstein space. We prove comparison principles for viscosity solutions of such Hamilton-Jacobi equations and show existence of viscosity solutions by Perron's method. We also discuss a model optimal control problem and show that the value function is the unique viscosity solution of the associated Hamilton-Jacobi-Bellman equation.

MSC: 35D40, 35F21, 35R15, 49L25, 49Q20

Keywords: Hamilton-Jacobi equations in the Wasserstein space, viscosity solutions, existence of solutions, comparison principle, Wasserstein space on graphs, manifold with boundary

1. Introduction

Partial differential equations (PDE) in infinite dimensional and abstract spaces have been studied steadily over the last several decades. The main interest has always been in Hamilton-Jacobi-Bellman (HJB) equations related to deterministic and stochastic optimal control problems for control of PDE and stochastic PDE and other abstract differential equations. Recently there has been a renewed interest in such equations in spaces of probability measures due to their connection to mean field control and mean field game problems. The theory of first and second order PDE in Hilbert spaces has been developed the most. A complete overview of various approaches, classical solutions, viscosity solutions, mild solutions, L^2-solutions, solutions using backward stochastic differential equations methods can be found in [43]. Results about classical solutions of linear second order PDE can be found in [39] and earlier results about mild solutions for first order PDE and solutions using convex regularization procedures can be found in [4]. Viscosity solutions in Hilbert spaces have been originally introduced by Crandall and P. L. Lions in [32, 33, 34, 35, 36, 37]. We refer to [43] for the full account of the theory and further references. Some aspects of the theory for first order equations can also be found in [68].

The original interest in the PDE in spaces of probability measures came from partially observed optimal control problems through the study of fully observable so called separated problems where one controls a new measure valued state process (unnormalized conditional density of the original state with respect to the observation process) which satisfies the so-called Duncan-Mortensen-Zakai equation. Early attempts to look at HJB equations in the space of measures for such a problem was made in [60]. A Bellman equation in the space of measures was also studied in [61]. A renewed interest in HJB equations in spaces of probability measures started with the development of the theory of mass transport and a calculus in the Wasserstein space of probability measures and later the study of mean field control and mean field game problems. The first definition of a viscosity solution using sub- and super-differentials in the Wasserstein space appeared in [52] and later different notions of viscosity solutions were introduced of equations in the space of probability measures and more abstract metric spaces in various contexts. In particular a notion of the so-called L-viscosity solution was introduced in [70] which "lifts" the equation from the Wasserstein space to an Hilbert space of L^2 random variables and this approach was developed further in [56] (see also [22, 23] for more on the lifting procedure). We refer the readers to [5, 8, 10, 9, 11, 12, 13, 16, 17, 18, 20, 21, 29, 40, 41, 42, 50, 51, 53, 54, 58, 62, 63, 64, 74, 79, 80, 81] for equations related to mean field control and optimal control/variational problems in spaces of probability measures. In particular convergence problems for particle approximations have been studied using PDE methods in [18, 20, 21, 40, 41, 50, 58, 74]. Equations related to control problems with partial observation were studied in [6] and equations related to differential games were investigated in [30, 65]. HJB equations in the Wasserstein and metric spaces with formal Riemannian structure as well as completely regular spaces, mostly related to control of gradient flows, large deviations and fluid dynamics were studied by different techniques in [27, 28, 44, 45, 46, 47, 48, 66, 67]. Various comparison theorems and uniqueness results for appropriately defined viscosity solutions were proved in these papers. HJB equations in abstract metric spaces were studied by various techniques in [1, 14, 15, 53, 55, 59, 71, 72, 77, 78]. Uniqueness of appropriately defined viscosity solutions of first order HJB equations in the Wasserstein space was proved in [5, 64]. Uniqueness of viscosity solutions of a second order Bellman master equation in the Wasserstein space arising in stochastic optimal control problems for McKean-Vlasov diffusion processes was established in [29]. In [9, 41] general comparison results for viscosity solutions of second-order parabolic partial differential equations in the Wasserstein space were proved. Other papers containing uniqueness results are [17], where a uniqueness result for a notion of viscosity solution for a class of integro-differential Bellman equations of a special type was shown, and [81], where well-posedness of viscosity solutions of parabolic master equations, including HJB master equations associated with control problems for McKean-Vlasov stochastic differential equations was established. There is also vast literature on master equations of mean field games which are integro-differential PDE in the space of probability measures. We do not discuss them here since they are not HJB equations.

In this manuscript we investigate Hamilton-Jacobi equations on the Wasserstein space of probability measures on graphs. Discrete optimal transport calculus, in the space of probability measures on graphs and gradient and Hamiltonian like flows on graphs, have been studied in many papers; we refer for instance to [25, 38, 73, 75]. In particular, finite state mean field games have received significant attention in recent years. Master equation for finite state mean field games with Wright–Fisher common noise have been studied in [7] and [57] derived master equations from finite state Hamilton-Jacobi equation which appear in potential games. However very little is known about Hamilton-Jacobi equations in such spaces. The only results in this direction are in [24] about Hamilton-Jacobi equations on complete graphs (every pair of distinct vertices is connected by a unique edge). Therefore, the analysis in [24] does not involve a graph structure and the underlying probability measure space is endowed with the flat Euclidean metric ℓ_2. Note that the ℓ_2 differential structure is not comparable to the differential structures considered in this manuscript. Indeed in our set up, each point $\mu \in \mathcal{P}(G)$ comes with a metric tensor $g(\mu)$, which naturally leads us to consider the Wasserstein space of probability measures on general connected graphs. Our goal is to introduce a notion of viscosity solution and develop a well-posedness theory. Since the set of probability measures on a graph with n vertices is identified with a simplex in $\mathbb{R}^n$, one may be tempted to recast our work within the theory of viscosity solutions in finite dimension on Riemannian manifolds with boundary (see Remark 4.4). We refer for instance to [3] for the theory of viscosity solutions on Riemannian manifolds. The analogy we point out in Remark 4.4 does not facilitate our work even if in our case the manifold (the simplex) is flat. Indeed, we have to deal with Hamiltonians which vanish near the boundary of the simplex since we are working on the Wasserstein space. This makes our study different from the classical theory of viscosity solutions. Hence, we present everything from the beginning and with details.

We focus on initial value problems for a class of Hamilton-Jacobi-Bellman equations with a convex and somehow coercive Hamiltonian which degenerates close to the boundary, which also involves a linear operator obtained by discretizing the so–called individual noise operator in Mean Field Games (cf. e.g. [26]). Of course different types of equations can be considered and we expect the theory to be developed in various directions. It is certainly also interesting to study initial boundary value problems on open subsets of the set of probability measures, however in this paper we only consider equations on the whole space. We prove two comparison results, the main one for the initial value problem where the boundary is irrelevant and a version of it for the initial boundary value problem. We also study the optimal control problem associated with a model Hamilton-Jacobi-Bellman equation and we prove that the value function is continuous on the whole space and it is the unique viscosity solution of the HJB equation. For our model control problem, the value function, and hence the unique viscosity solution of the HJB equation which is continuous up to the boundary of the set of probability measures, is predetermined on the boundary and cannot be prescribed there. Our viscosity solutions are only defined on the interior of the set of probability measures and our comparison theorem does not need any information about the behavior of viscosity sub/supersolutions on the boundary. However, it may be possible to consider viscosity solutions to such problems on the whole space or treat them as constrained viscosity solutions (solutions to state constraint problems). This is left for future research. Finally, we also discuss the existence of viscosity solutions by Perron's method. Even though Perron's method here is a rather straightforward adaptation of the classical Perron's method, we present full details for the sake of completeness.

Throughout this manuscript, we fix an undirected graph $G = (V, E, \omega)$, where $V = \{1, \cdots, n\}$ is the set of vertices and $E \subset V^2$ is the set of edges. The weight $\omega = (\omega_{ij})$ is a n by n symmetric matrix with nonnegative entries such that $\omega_{ij} > 0$ if $(i,j) \in E$. As in [49], we assume for simplicity that the graph is connected, simple, with no self-loops or multiple edges. We denote by $\mathcal{P}(G)$ the probability simplex

$$\Big\{\rho \in [0,1]^n \;\Big|\; \sum_{i=1}^n \rho_i = 1\Big\}.$$

We use a symmetric function $g : [0,1]^2 \to [0,\infty)$, to induce an equivalence relation on $\mathbb{S}^{n\times n}$, the set of n by n skew-symmetric matrices: if $\rho \in \mathcal{P}(G)$, we say that $v, \tilde{v} \in \mathbb{S}^{n\times n}$ are ρ-equivalent if $(v_{ij} - \tilde{v}_{ij})g_{ij}(\rho) = 0$ for all $(i,j) \in E$. We denote the quotient space by $\mathbb{H}_\rho$. Under appropriate conditions which will later be specified, g is used to define a metric tensor on $\mathcal{P}(G)$ and endow $\mathbb{H}_\rho$ with an *inner product* and a *discrete norm* as follows:

$$(v, \tilde{v})_\rho := \frac{1}{2} \sum_{(i,j)\in E} v_{ij}\tilde{v}_{ij} g_{ij}(\rho) \quad \text{and} \quad \|v\|_\rho := \sqrt{(v,v)_\rho}, \quad \forall\, v, \tilde{v} \in \mathbb{S}^{n\times n}. \tag{1.1}$$

Here the coefficient $1/2$ accounts for the fact that whenever $(i,j) \in E$ then $(j,i) \in E$.

If $\phi : V \to \mathbb{R}^n$, its graph gradient denoted $\nabla_G \phi$ is defined as

$$\nabla_G \phi := \sqrt{\omega_{ij}}(\phi_i - \phi_j)_{(i,j)\in E}.$$

The adjoint of ∇_G for the $(\cdot,\cdot)_\rho$ inner product is $-\mathrm{div}_\rho : \mathbb{H}_\rho \to \mathbb{R}^n$ given by

$$\mathrm{div}_\rho(v) = \left(\sum_{j=1}^n \sqrt{\omega_{ij}} v_{ji} g_{ij}(\rho)\right)_{i=1}^n, \quad \forall\, v \in \mathbb{S}^{n\times n}.$$

We call div_ρ the divergence operator. In this manuscript, we impose that

$$\int_0^1 \frac{dr}{\sqrt{g(r, 1-r)}} < +\infty, \tag{1.2}$$

to ensure that the expression $\mathcal{W}$, defined below in (2.7), is a metric on $\mathcal{P}(G)$ (cf. [73] and [49]).

We fix $T > 0$ and assume that we are given $\mathcal{F}$, $\mathcal{U}_0 \in C(\mathcal{P}(G))$ and $\mathcal{H} \in C(\mathcal{P}(G) \times \mathbb{S}^{n\times n})$. We denote by $\mathcal{L}(\rho,\cdot)$ the Legendre transform of $\mathcal{H}(\rho,\cdot)$ with respect to the inner product $(\cdot,\cdot)_\rho$. Setting

$$\bar{g}(s,t) := \frac{\log s - \log t}{s - t} g(s,t),$$

for $s \neq t$ such that $s, t > 0$, in this introduction, we will keep our focus on the cases where g satisfies (2.5), or more generally when

$$\bar{g} \quad \text{has a unique continuous extension to} \quad [0,1]^2. \tag{1.3}$$

As a consequence of (1.3), as a function a-priori defined on a subset of $(0,1)^n$,

$$\rho \to \mathrm{div}_\rho\big(\nabla_G \log \rho\big) \quad \text{has a unique continuous extension to} \quad [0,1]^n. \tag{1.4}$$

In light of (1.4), standard ODEs theory ensures that given $\bar{v} \in L^1(0,T; \mathbb{S}^{n\times n})$ and $\hbar \geq 0$, the system of equations

$$\dot{\sigma} + \mathrm{div}_\sigma\Big(\bar{v} + \hbar \nabla_G \log \sigma\Big) = 0 \tag{1.5}$$

has a distributional solution $\sigma : [0,T] \to \mathbb{R}^n$, of class $W^{1,1}$.

When the range of σ is contained in $\mathcal{P}(G)$, we call $\bar{v}$ a control for σ on $[0,T]$. For $t \in (0,T]$ we consider

$$\mathcal{U}(t,\mu) = \inf_{(\sigma,\bar{v})} \left\{ \mathcal{U}_0(\sigma_0) + \int_0^t \Big(\mathcal{L}(\sigma,\bar{v})ds - \mathcal{F}(\sigma)\Big)ds : \quad \sigma_t = \mu \right\}, \tag{1.6}$$

where the infimum is performed over the set of $(\sigma,\bar{v})$ such that $\bar{v}$ is a control for σ over $[0,t]$. Formally at least, we expect $\mathcal{U}$ to satisfy a Hamilton–Jacobi equation, after defining a suitable notion of Wasserstein gradient operator on the set of functions on $\mathcal{P}(G)$. More precisely, we expect that $\mathcal{U}$ would satisfy, in a sense which remains to be specified, the equation

$$\partial_t \mathcal{U}(t,\mu) + \mathcal{H}\big(\mu, \nabla_{\mathcal{W}} \mathcal{U}(t,\mu)\big) + \mathcal{F}(\mu) = \hbar \Delta_{\mathrm{ind}} \mathcal{U}(t,\mu). \tag{1.7}$$

Here

$$\Delta_{\mathrm{ind}} \mathcal{U}(t,\mu) := \Big(\mathrm{div}_\mu\big(\nabla_{\mathcal{W}} \mathcal{U}(\mu)\big), \log \mu\Big) = -\mathcal{O}_\mu\big(\nabla_{\mathcal{W}} \mathcal{U}(\mu)\big)$$

and we have set

$$\mathcal{O}_\mu(p) := -\big(p, \nabla_G \log \mu\big)_\mu, \qquad \forall (p,\mu) \in \mathcal{P}(G) \times \mathbb{S}^{n\times n}.$$

We call Δ_{ind}, the graph individual noise operator (see Subsection 3.4 for comments on how Δ_{ind} could be associated to stochastic processes which are time continuous Markov chains on V).

The assumption (1.3) ensures that $\mathcal{O}_\mu(p)$ satisfies (6.1), an essential condition in the application of Perron's method to obtain the existence of a solution to (1.7). Note that $\mathcal{O}_\mu(p)$ cannot be incorporated into the Hamiltonian since the modified Hamiltonian would fail to satisfy (A-v) and so, the conditions imposed on $\mathcal{H}(\mu,p)$ and $\mathcal{O}_\mu(p)$ are of different types.

In this manuscript, the existence of a solution to (1.7) will not rely on the control problem (1.6), brought up here only to motivate the study of (1.7).

Observe that (1.7) is linear in $\mathcal{U}$, when $\mathcal{F}\equiv 0$, $\mathcal{H}\equiv 0$ and g is given by Example 2.5, which means $\bar g(s,t)\equiv 1$. When $\hbar=1$, the solution in to (1.7) case is given by (see subsection 3.4)

$$\mathcal{U}(t,\mu):=\mathcal{U}_0\big(e^{At}\mu\big),$$

where

$$A_{ij}=\begin{cases}\omega_{ij}, & \text{if} \quad j\in N(i);\\ 0, & \text{if} \quad j\notin N(i), j\neq i;\\ -\sum_{k\in N(i)}\omega_{ik}, & \text{if} \quad j=i.\end{cases} \tag{1.8}$$

Here, $N(i):=\{j\in V:\ \omega_{ij}>0\}$. For each $t\geq 0$, e^{At} is known to be a transition matrix and A is a Q–matrix. Therefore, as we will explain in Subsection 3.4, there are Markov chains associated to the paths $(t,\mu)\to e^{At}\mu$.

The plan of paper is the following. In Section 2 we present the definitions, notation and the mathematical setup for the Wasserstein space of probability measures on a finite graph. Section 3 collects preliminary material about calculus on the Wasserstein space on a graph and in Definition 3.18, we introduce the so-called individual noise operator. In Section 4 we introduce the definition of viscosity solution and in Section 5 we prove comparison results. Existence of viscosity solutions by Perron's method and some regularity results are presented in Section 6. In Section 7 we discuss a model optimal control problem and show that the value function is the unique viscosity solution of the associated HJB equation.

2. Definitions and Notation

We denote the set of skew–symmetric $n\times n$ matrices as $\mathbb{S}^{n\times n}$. Let $G=(V,E,\omega)$ denote an undirected graph of vertices $V=\{1,...,n\}$ and edges E, with a weighted metric $\omega=(\omega_{ij})$ given by an n by n symmetric matrix with nonnegative entries ω_{ij} and such that $\omega_{ij}>0$ if $(i,j)\in E$. For simplicity, assume that the graph is connected and simple, with no self–loops or multiple edges. We set

$$\bar\lambda_\omega:=\sup_{(i,j)\in E}\omega_{ij}^{-1}\quad\text{and}\quad C_\omega:=\sup_{(i,j)\in E}\sqrt{\omega_{ij}}.$$

The range and kernel of the gradient operator. It is customary to identify a function $\phi: V\to\mathbb{R}$ with a vector $\phi=(\phi_i)_{i=1}^n\in\mathbb{R}^n$. We use the standard *inner product* and *norm* on $\mathbb{R}^n$:

$$(\phi,\tilde\phi):=\sum_{i=1}^n\phi_i\tilde\phi_i\quad\text{and}\quad\|\phi\|=\sqrt{(\phi,\phi)},\quad\forall\,\phi,\tilde\phi\in\mathbb{R}^n.$$

We denote by $R(\nabla_G)$ the range of ∇_G (defined in the introduction) and by $\mathbf{1}\in\mathbb{R}^n$ the vector whose entries are all equal to 1. Since G is connected, the kernel of ∇_G is the one dimensional space spanned by $\mathbf{1}$. The orthogonal complement in $\mathbb{R}^n$ of the latter space is $\ker(\nabla_G)^\perp$, the set of $h\in\mathbb{R}^n$ such that $\sum_{i=1}^n h_i=0$.

G-Divergence of vector field. The *divergence* operator associates to any vector field m on G a function on V defined by

$$\nabla_G \cdot (m) = \mathrm{div}_G(m) := \Big(\sum_{j \in N(i)} \sqrt{\omega_{ij}} m_{ji} \Big)_{i=1}^n .$$

Set of probability measures and its boundary. We identify $\mathcal{P}(G)$, the set of probability measures on V, with the simplex

$$\mathcal{P}(G) = \Big\{ \rho = (\rho_i)_{i=1}^n \subset [0,1]^n \ \Big| \ \sum_{i=1}^n \rho_i = 1 \Big\}.$$

We denote for $0 \le \varepsilon < 1$, $\mathcal{P}_\varepsilon(G) := \mathcal{P}(G) \cap (\varepsilon, 1)^n$ so that $\mathcal{P}_0(G)$ is the interior of $\mathcal{P}(G)$. The boundary of $\mathcal{P}(G)$ is $\mathcal{P}(G) \setminus \mathcal{P}_0(G)$.

The set $\mathcal{C}_s^t(\rho^0, \rho^1)$ of paths connecting probability measures. Given $\rho^0, \rho^1 \in \mathcal{P}(G)$ and $0 \le s < t$, we denote by $\mathcal{C}_s^t(\rho^0, \rho^1)$ the set of pairs (σ, m) such that

$$\sigma \in H^1(s,t;\mathcal{P}(G)),\ m \in L^2(s,t;\mathbb{S}^{n\times n}), \quad (\sigma(s), \sigma(t)) = (\rho^0, \rho^1)$$

and for $i = 1, ..., n$,

$$\dot{\sigma}_i + \sum_{j \in N(i)} \sqrt{\omega_{ij}} m_{ji} = 0, \quad \text{in the weak sense on} \quad (0,t). \tag{2.1}$$

Throughout this manuscript $g : [0,\infty) \times [0,\infty) \to [0,\infty)$ satisfies the following assumptions:

- (H-i) g is continuous on $[0,\infty) \times [0,\infty)$ and is of class C^∞ on $(0,\infty) \times (0,\infty)$;
- (H-ii) $g(r,s) = g(s,r)$ for any $s, r \in [0,\infty)$;
- (H-iii) $\min\{r,s\} \le g(r,s) \le \max\{r,s\}$ for any $r, s \in [0,\infty)$;
- (H-iv) $g(\lambda r, \lambda s) = \lambda g(r,s)$ for any $\lambda, s, r \in [0,\infty)$;
- (H-v) g is concave.

We set

$$g_{ij}(\rho) = g(\rho_i, \rho_j), \quad \forall\, \rho \in \mathbb{R}^n, \quad \forall\, i, j \in V.$$

The Hilbert spaces $\mathbb{H}_\rho$ and integration by parts. If $\rho \in \mathcal{P}(G)$, we shall use the inner product defined in (1.1). Similarly, if $m, \tilde{m} \in \mathbb{S}^{n\times n}$, we set

$$(m, \tilde{m}) := \frac{1}{2} \sum_{(i,j) \in E} m_{ij} \tilde{m}_{ij} \quad \text{and} \quad \|m\| := \sqrt{(m,m)}.$$

If $\phi \in \mathbb{R}^n$ and $v \in \mathbb{S}^{n\times n}$, we have the integration by parts formula

$$(\nabla_G \phi, v)_\rho = -(\phi, \mathrm{div}_\rho(v)). \tag{2.2}$$

Using the notation from [49], we denote by $T_\rho\mathcal{P}(G)$ the closure of the range of ∇_G in $\mathbb{H}_\rho$. We refer to $T_\rho\mathcal{P}(G)$ as the tangent space to $\mathcal{P}(G)$. We denote by π_ρ the projection onto $T_\rho\mathcal{P}(G)$.

Using the fact that by (H-iii) $g_{ij}(\rho) \le \rho_i + \rho_j$, one shows that

$$\|\mathrm{div}_\rho(v)\|_{\ell_2} \le \sqrt{2n} C_\omega \, \|v\|_\rho, \quad \text{and so,} \quad \|\mathrm{div}_\rho(v)\|_{\ell_1} \le \sqrt{2} n C_\omega \, \|v\|_\rho. \tag{2.3}$$

Connected components. Let $\rho \in \mathcal{P}(G)$. We say that $i, j \in V$ are g-connected if either $i = j$

or $i \neq j$ but there are $i_1, i_2, ..., i_k \in V$ such that $i_1 = i$, $i_k = j$, $(i_l, i_{l+1}) \in E$ for $l = 1, ..., k-1$ and

$$\prod_{l=2}^{k} g_{i_{l-1} i_l}(\rho) > 0.$$

Example 2.1. *Examples of g satisfying* (H-i)-(H-v) *and* (1.2) *include*

$$g(r,s) = \frac{r+s}{2}, \tag{2.4}$$

$$g(r,s) = \int_0^1 r^{1-t} s^t dt = \begin{cases} \frac{r-s}{\log r - \log s}, & \text{if } r \neq s; \\ 0, & \text{if } r = 0 \text{ or } s = 0; \\ r, & \text{if } r = s, \end{cases} \tag{2.5}$$

and

$$g(r,s) = \begin{cases} 0, & \text{if } r = 0 \text{ or } s = 0; \\ \frac{2}{\frac{1}{r} + \frac{1}{s}}, & \text{otherwise.} \end{cases} \tag{2.6}$$

One can generate more examples by taking convex combinations of the g's in (2.4)-(2.6).

The Monge-Kantorovich metric In $\mathcal{P}(G)$. For $\rho^0, \rho^1 \in \mathcal{P}(G)$, we define the 2-Monge-Kantorovich metric by

$$\mathcal{W}(\rho^0, \rho^1) := \left(\inf_{(\sigma, v)} \left\{ \int_0^1 (v, v)_\sigma dt \;\middle|\; \dot\sigma + \mathrm{div}_\sigma(v) = 0,\ \sigma(0) = \rho^0,\ \sigma(1) = \rho^1 \right\} \right)^{\frac{1}{2}}. \tag{2.7}$$

Here the infimum is performed over the set of pairs (σ, v) such that $\sigma \in H^1(0,1; \mathcal{P}(G))$ and $v : [0,1] \to \mathbb{S}^{n \times n}$ is measurable. Recall that if $C_g < +\infty$, then $\mathcal{W}(\rho^0, \rho^1) < +\infty$ for any $\rho^0, \rho^1 \in \mathcal{P}(G)$ (see Proposition 3.7 [49]). There exists a minimizer (σ, v) in (2.7) such that $\|v\|_\sigma = \mathcal{W}(\rho^0, \rho^1)$ almost everywhere on $(0,1)$. Using the continuity equation and the second identity in (2.3), we conclude that

$$\|\dot\sigma(t)\|_{\ell_\infty} \leq \sqrt{2} n C_\omega \mathcal{W}(\rho^0, \rho^1). \tag{2.8}$$

This proves that the $W^{1,\infty}$-norm of σ is bounded by a constant depending only on n, g, G, ω. Further assume that $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$, where γ_P is the Poincaré function on G given in [49]. By Remark 6.5 and Theorem 7.5 [49], we can find a Borel map $\phi \equiv \phi[\rho^0, \rho^1] : [0,1] \to \mathbb{R}^n$ such that $v = \nabla_G \phi$ and

$$v_{ij} = \nabla_G \phi \quad \text{is uniquely determined on} \quad \{t \in (0,1) :\ g_{ij}(\sigma(t)) > 0\}. \tag{2.9}$$

Under the stringent assumption that there exists $\varepsilon > 0$ such that $\rho^0, \rho^1 \in \mathcal{P}_\varepsilon(G)$, Theorem 7.3 [49] asserts that $\|\phi\|_{W^{1,1}(0,1)}$ is bounded by a constant which is independent of ρ^0 and ρ^1, but depends on ε. Thus,

$$(\rho^0, \rho^1) \to \phi[\rho^0, \rho^1](1) \text{ is continuous for the metric } \ell_1 \text{ on } \mathcal{P}_\varepsilon(G) \times \mathcal{P}_\varepsilon(G). \tag{2.10}$$

Remark 2.2. *We recall that the $(\mathcal{P}(G), \mathcal{W})$ topology is the same as the $(\mathcal{P}(G), \ell_1)$ topology (cf.* [73]*) and thus it is also the same as the ℓ_2–topology. Therefore, $\mathcal{P}(G)$ is a compact set and the notion of a continuous function is the same for all these three topologies. In particular, $\mathcal{P}_0(G)$ is a dense subset of $\mathcal{P}(G)$ for the $\mathcal{W}$-topology. Since $\mathcal{P}(G)$ is a compact set, it has a finite diameter.*

Throughout the paper, for any $r > 0$ and $\mu \in \mathcal{P}(G)$, we denote the open ball with radius r centered at μ in $(\mathcal{P}(G), \|\cdot\|_{\ell_2})$ by $B_r(\mu)$. By Remark 2.2, $B_r(\mu)$ is also an open neighborhood of μ in $(\mathcal{P}(G), \mathcal{W})$ and in $(\mathcal{P}(G), \|\cdot\|_{\ell_1})$. Similarly, for any $t \in [0,T], r > 0, \mu \in \mathcal{P}(G)$, we use $B_r(t,\mu)$ to denote the open ball with radius r centered at (t,μ) in $[0,T] \times (\mathcal{P}(G), \|\cdot\|_{\ell_2})$.

3. Preliminaries

Throughout the section, we use the same notation as in Section 2 and assume that (H-i)-(H-v) and (1.2) hold. For $\rho \in \mathcal{P}(G)$, we set

$$\lambda_g(\rho) = \sup_{(i,j)\in E} \left\{ \frac{\sqrt{2}}{\sqrt{\omega_{ij}}} \frac{n}{\sqrt{g_{ij}(\rho)}} : \ g_{ij}(\rho) > 0 \right\}. \tag{3.1}$$

Note that $\lambda_g(\rho) < \infty$ if ρ has a g-connected component of cardinality greater than or equal to 2.

Remark 3.1. *If $\varepsilon > 0$ and $\rho \in \mathcal{P}(G)$ is such that $\rho_i \geq \varepsilon$ for all $i \in V$ then $\lambda_g(\rho) \leq \sqrt{2\bar{\lambda}_\omega \varepsilon^{-1}}n$.*

3.1. Further properties of tangent vectors and tangent spaces.

For $\rho \in \mathcal{P}(G)$ and $v \in T_\rho\mathcal{P}(G)$, denote by $[v]_\rho$ the set of $\tilde{v} \in T_\rho\mathcal{P}(G)$ such that v and $\tilde{v}$ are ρ–equivalent.

Lemma 3.2. *For any $\rho \in \mathcal{P}(G)$ such that $\lambda_g(\rho) < \infty$, there exists $P_\rho : T_\rho\mathcal{P}(G) \to \mathbb{R}^n$ such that if $\phi \in \mathbb{R}^n$ and we set $\psi := P_\rho\big([\nabla_G\phi]_\rho\big)$ then*

- (i) *$\nabla_G\psi$ and $\nabla_G\phi$ are ρ-equivalent and so, $\|\nabla_G\phi\|_\rho = \|\nabla_G\psi\|_\rho$.*
- (ii) *$|\psi_i| \leq \lambda_g(\rho)\Big\|\nabla_G\phi\Big\|_\rho$ for all $i \in V$.*

Proof. Let $C_1(\rho), \cdots, C_N(\rho)$ be all the g-connected components of $\rho \in \mathcal{P}(G)$ and for $l \in \{1, \cdots, N\}$, set

$$k_l := \min_{k\in C_l(\rho)} k.$$

Given $\phi : V \to \mathbb{R}$, we define

$$\psi_i := \phi_i - \phi_{k_l}, \ \ \forall i \in C_l(\rho).$$

Note that if $i, j \in C_l(\rho)$ then

$$\psi_{k_l} = 0 \quad \text{and} \quad (\nabla_G\psi)_{ij} = (\nabla_G\phi)_{ij}. \tag{3.2}$$

This is enough to conclude that $\nabla_G\psi$ and $\nabla_G\phi$ are ρ-equivalent.

If $i \in C_l(\rho)$ and $i \neq k_l$, we can find $l_1 = k_l, \cdots, l_{\alpha_i} = i$ such that $g_{l_1 l_2}, \cdots, g_{l_{\alpha_i - 1} l_{\alpha_i}} > 0$. The identity

$$\psi_{l_m} = \psi_{l_{m-1}} + \big(\nabla_G\phi\big)_{l_m l_{m-1}}, \qquad \forall m \geq 2$$

and $\psi_{l_1} = 0$ implies that the sequence $\big(\psi_{l_m}\big)_{m=1}^{\alpha_i}$ is uniquely determined by $\nabla_G\phi$. This is enough to conclude that the map P_ρ is well–defined.

Let E_l be the set of (i,j) in E such that $i, j \in C_l(\rho)$. We use the first identity in (3.2) to conclude that

$$2\Big\|\nabla_G\phi\Big\|_\rho^2 = \sum_{l=1}^{N} \sum_{(i,j)\in E_l} \big(\nabla_G\psi\big)_{ij}^2 \, g_{ij}(\rho).$$

If $i \in C_l(\rho)$ and $i \neq k_l$, using the above notation, we have

$$2\Big\|\nabla_G \phi\Big\|_\rho^2 \geq \omega_{l_1 l_2}\ \psi_{l_2}^2\ g_{l_1 l_2}(\rho) + \sum_{m=3}^{\alpha_i} \omega_{l_{m-1} l_m}\ \big(\psi_{l_{m-1}} - \psi_{l_m}\big)^2\ g_{l_{m-1} l_m}(\rho).$$

One checks that

$$\big|\psi_i\big| \leq \sum_{m=2}^{\alpha_i} \frac{\sqrt{2}}{\sqrt{\omega_{l_{m-1} l_m}}} \frac{1}{\sqrt{g_{l_{m-1} l_m}(\rho)}} \Big\|\nabla_G \phi\Big\|_\rho.$$

We conclude that (ii) holds for i in the union of the sets $C_l(\rho)$ of a cardinality greater than or equal to 2. It is obvious that (ii) continues to hold for i in the union of the sets $C_l(\rho)$ with cardinality 1. The proof of (iii) follows from the fact that $\psi_i = \phi_i - \phi_1$ and $\omega_{1i}\ |\psi_i|^2\ g_{1i}(\rho) \leq \big\|\nabla_G \phi\big\|_\rho^2$. □

Corollary 3.3. *By Lemma 3.2, if $\rho \in \mathcal{P}(G)$ and $\lambda_g(\rho) < \infty$, then for any $v \in T_\rho \mathcal{P}(G)$ there exists $\psi \in \mathbb{R}^n$ such that $v = \nabla_G \psi$ and $|\psi_i| \leq \lambda_g(\rho) \|v\|_\rho$ for all $i \in V$.*

3.2. **The Wasserstein metric and the space of absolutely continuous paths on $(\mathcal{P}(G), \mathcal{W})$.**

Lemma 3.4. *For any $\rho, \bar\rho \in \mathcal{P}(G)$, we have $\|\bar\rho - \rho\|_{\ell_1} \leq 2\sqrt{n} C_\omega\ \mathcal{W}(\rho, \bar\rho)$.*

Proof. Since there exists a $\mathcal{W}$ geodesic connecting ρ to $\bar\rho$, (cf. Theorem 4.5-(i) in [49]), we use (2.8) to conclude. □

Lemma 3.5. *If $\varepsilon > 0$ and $\rho, \bar\rho \in \mathcal{P}(G)$ are such that $\rho_i, \bar\rho_i \geq \varepsilon$ for all $i \in V$ then*

$$\sqrt{\varepsilon}\ \mathcal{W}(\rho, \bar\rho) \leq \sqrt{2\bar\lambda_\omega} n\ \|\bar\rho - \rho\|_{\ell_1}.$$

Proof. Setting

$$\sigma(t) = (1-t)\rho + t\bar\rho, \qquad \forall t \in [0,1],$$

we have $\sigma_i(t) \geq \varepsilon$ for $i \in V$ and $t \in [0,1]$. We then use Remark 3.1 to conclude that

$$\lambda_g(\sigma(t))\sqrt{\varepsilon} \leq \sqrt{2\bar\lambda_\omega} n. \tag{3.3}$$

We define

$$E(\phi) := \int_0^1 \Big(\frac{1}{2}\|\nabla_G \phi\|_{\sigma(t)}^2 - (\phi, \bar\rho - \rho)\Big) dt, \qquad \forall \phi \in L^2(0,1;\mathbb{R}^n).$$

For $\phi \in L^2(0,1;\mathbb{R}^n)$, using the operator $P_{\sigma(t)}$ from Lemma 3.2 and setting $\psi(t) = \phi(t) - \phi_1(t)$, we have

$$\psi \in L^2(0,1;\mathbb{R}^n), \quad \psi = P_\sigma\big([\nabla_G \phi(t)]_\sigma\big), \quad E(\phi) = E(\psi).$$

By (3.3),

$$E(\psi) \geq \int_0^1 \Big(\frac{\varepsilon}{4\bar\lambda_\omega n^3} \|\psi\|_{\ell_2}^2 - \|\psi\|_{\ell_2}\ \|\bar\rho - \rho\|_{\ell_2}\Big) dt.$$

This proves that E is bounded from below and if $(\psi_k)_k$ is a sequence in the range of P_σ such that $\big(E(\psi_k)\big)_k$ decreases to the infimum of E over $L^2(0,1;\mathbb{R}^n)$ then $(\psi_k)_k$ is bounded in $L^2(0,1;\mathbb{R}^n)$. Hence, $(\psi_k)_k$ admits a point of accumulation ψ_∞ for the weak topology. Since $\phi \to E(\phi)$ is a quadratic and convex function, we conclude that

$$\liminf_{k \to +\infty} E(\psi_k) \geq E(\psi_\infty).$$

We can assume without loss of generality that $\psi_\infty = P_\sigma\big([\nabla_G\psi_\infty]_\sigma\big)$. The Euler-Lagrange equation satisfied by ψ_∞ is

$$\int_0^1 \Big(\big(\nabla_G\psi_\infty, \nabla_G\phi\big)_\sigma - (\bar\rho-\rho,\phi)\Big)dt = 0, \qquad \forall \phi \in L^2(0,1;\mathbb{R}^n). \tag{3.4}$$

This means that

$$\dot\sigma + \mathrm{div}_\sigma(\nabla_G\psi_\infty) = 0. \tag{3.5}$$

Using $\phi = \psi_\infty$ in (3.4), we obtain

$$\int_0^1 \big\|\nabla_G\psi_\infty\big\|_\sigma^2 dt = \int_0^1 (\bar\rho-\rho,\psi_\infty)dt \le \|\bar\rho-\rho\|_{\ell_1}\int_0^1 \|\psi_\infty\|_{\ell_\infty}dt \le \|\bar\rho-\rho\|_{\ell_1}\int_0^1 \lambda_g(\sigma)\|\nabla_G\psi_\infty\|_\sigma dt.$$

We first use (3.3) and then use Hölder's inequality to conclude that

$$\int_0^1 \big\|\nabla_G\psi_\infty\big\|_\sigma^2 dt \le \|\bar\rho-\rho\|_{\ell_1}\ \sqrt{2\bar\lambda_\omega\varepsilon^{-1}}n\sqrt{\int_0^1 \|\nabla_G\psi_\infty\|_\sigma^2 dt}.$$

We simplify the previous identity and use the fact that, by (3.5), $\nabla_G\psi_\infty$ is a velocity for σ to obtain

$$\mathcal{W}\big(\sigma(0),\sigma(1)\big) \le \int_0^1 \|\nabla_G\psi_\infty\|_\sigma dt \le \sqrt{\int_0^1 \|\nabla_G\psi_\infty\|_\sigma^2 dt} \le \|\bar\rho-\rho\|_{\ell_1}\ \sqrt{2\bar\lambda_\omega\varepsilon^{-1}}n.$$

This concludes the proof. □

Remark 3.6. *Let $\varepsilon>0$ and let $\rho\in\mathcal{P}(G)$ be such that $\rho_i\ge\varepsilon$ for all $i\in V$. Suppose $f\in\mathbb{R}^n$ is such that $\sum_{i=1}^n f_i = 0$. As done in Lemma 3.5, one can show that there exists $\phi\in\mathbb{R}^n$ such that*

$$f + \mathrm{div}_\rho(\nabla_G\phi) = 0, \quad \|\nabla_G\phi\|_\rho \le \|f\|_{\ell_1}\sqrt{2\bar\lambda_\omega\varepsilon^{-1}}n.$$

Remark 3.7. *Suppose that $\sigma:[0,1]\to\mathcal{P}(G)$ and $v:[0,1]\to\mathbb{R}^n$ is a Borel map such that*

$$\dot\sigma + \mathrm{div}_\sigma(v) = 0 \quad \textit{in the weak sense in } (0,1) \quad \textit{and} \quad \int_0^1 \|v(t)\|_{\sigma(t)}^2 dt < +\infty.$$

By definition of $\mathcal{W}$, we have that σ is an absolutely continuous curve on $(\mathcal{P}(G),\mathcal{W})$ since

$$\mathcal{W}(\sigma(t),\sigma(s)) \le \int_s^t \|v(\tau)\|_{\sigma(\tau)}d\tau, \qquad \forall 0\le s<t\le 1.$$

Hence, if we denote by $|\sigma'|_\mathcal{W}$ the $\mathcal{W}$ metric derivative of σ, then $|\sigma'|_\mathcal{W}\le\|v\|_\sigma$ a.e. on $(0,1)$.

We next show that v can be chosen in an optimal way.

Proposition 3.8. *Suppose that $\sigma:[0,1]\to\mathcal{P}(G)$ such that*

$$\mathcal{W}(\sigma(t),\sigma(s)) \le \int_t^s \beta(\tau)d\tau \quad \textit{and} \quad \beta\in L^2(0,1). \tag{3.6}$$

Then there exists $v:(0,1)\to\mathbb{S}^{n\times n}$ Borel such that $v(t)\in T_{\sigma(t)}\mathcal{P}(G)$ for almost every t,

$$\dot\sigma + \mathrm{div}_\sigma(v) = 0 \quad \textit{in the weak sense in } (0,1) \tag{3.7}$$

and

$$\|v\|_\sigma \le |\sigma'|_\mathcal{W} \le \beta, \quad |\dot\sigma| \le \sqrt{2}nC_\omega|\sigma'|_\mathcal{W} \quad \textit{a.e. on} \quad [0,1]. \tag{3.8}$$

Proof. We skip the proof since it is similar to the proof of Theorem 8.3.1 of [2]. □

3.3. The Wasserstein gradient on $\mathcal{P}(G)$.

Definition 3.9 (Wasserstein gradient). *Let $\mathcal{F} : \mathcal{P}(G) \to \mathbb{R}$ and $\rho \in \mathcal{P}(G)$.*

(i) *We say that $\mathcal{F}$ is $\mathcal{W}$-differentiable at ρ if there exist $v \in T_\rho\mathcal{P}(G)$ and $C > 0$ such that: for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $\bar\rho \in \mathcal{P}(G)$ and $\bar v \in T_\rho\mathcal{P}(G)$ then*

$$\|\bar\rho - \rho\|_{\ell_1} \leq \delta \quad \Longrightarrow \quad \left|\mathcal{F}(\bar\rho) - \mathcal{F}(\rho) - (\bar v, v)_\rho\right| \leq \varepsilon\mathcal{W}(\bar\rho, \rho) + C\left\|\bar\rho - \rho + \mathrm{div}_\rho(\bar v)\right\|_{\ell_1}. \tag{3.9}$$

(ii) *We write $\mathcal{F} \in C^1(\mathcal{P}_0(G), \mathcal{W})$ if $\mathcal{F}$ is $\mathcal{W}$-differentiable everywhere on $\mathcal{P}_0(G)$ and its Wasserstein gradient $\nabla_\mathcal{W}\mathcal{F}$ is continuous on $\mathcal{P}_0(G)$.*

Remark 3.10. *Let $\mathcal{F}$ and ρ be as in Definition 3.9.*

(i) *We will later show that when there exists v as in Definition 3.9, it is uniquely determined. If this is the case, we use the notation $v = \nabla_\mathcal{W}\mathcal{F}(\rho)$ and call v the Wasserstein gradient of $\mathcal{F}$ at ρ. One similarly defines Wasserstein sub and super gradients.*

(ii) *Observe that if $\rho \in \mathcal{P}_0(G)$ then $\|\cdot\|_\rho$ and $\|\cdot\|_{\ell_2}$ are equivalent. Therefore in Definition 3.9, there is no confusion about what it means that $\nabla_\mathcal{W}\mathcal{F}$ is continuous on $\mathcal{P}_0(G)$. However, if $\rho \in \partial\mathcal{P}(G)$, we may have $\|p\|_\rho = 0$ while we have $\|p\|_{\ell_2} > 0$.*

Definition 3.11 (Fréchet derivative). *Let $\mathcal{F} : \mathcal{P}(G) \to \mathbb{R}$ and let $\rho \in \mathcal{P}(G)$.*

(i) *We say that $\mathcal{F}$ has a Fréchet derivative at ρ if there exists $p \in \mathbb{R}^n$ such that*

$$\sum_{i=1}^n p_i = 0, \quad \textit{and} \quad \lim_{s\to 0^+} \frac{\mathcal{F}((1-s)\rho + s\bar\rho) - \mathcal{F}(\rho)}{s} = (p, \bar\rho - \rho),\ \forall \bar\rho \in \mathcal{P}(G). \tag{3.10}$$

We will later show that there is at most one $p \in \mathbb{R}^n$ satisfying (3.10). When such p exists, we write $p = \frac{\delta\mathcal{F}}{\delta\rho}(\rho)$ and call it the Fréchet derivative at ρ. Lemma 3.15 shows a relation between $\frac{\delta\mathcal{F}}{\delta\rho}$ and $\nabla_\mathcal{W}\mathcal{F}$. One similarly defines Fréchet sub and super differentials.

(ii) *We write that $\mathcal{F} \in C^1(\mathcal{P}_0(G), \ell_2)$ if $\mathcal{F}$ has a continuous Fréchet derivative everywhere on $\mathcal{P}_0(G)$.*

Remark 3.12. *Note that the Fréchet derivative is independent of the graph structure, i.e. the edges E of the graph. However, the Wasserstein gradient depends on E and the metric tensor g.*

Lemma 3.13. *If $\nabla_\mathcal{W}\mathcal{F}(\rho)$ exists for some $\rho \in \mathcal{P}(G)$, then it is uniquely determined as an element of the quotient space $T_\rho\mathcal{P}(G)$.*

Proof. Assume $v, \tilde v \in T_\rho\mathcal{P}(G)$ are Wasserstein gradients of $\mathcal{F}$ at ρ. We are to show that if $(i,j) \in E$ and $g_{ij}(\rho) > 0$ then $v_{ij} = \tilde v_{ij}$. We assume without loss of generality that $\rho_i \geq \rho_j$. Since by (H-iii) we have $(\rho_i, \rho_j) \neq (0,0)$, we conclude that $\rho_i > 0$. For $0 < a << 1$, we set $v^a_{kl} = 0$ except that

$$v^a_{ij} = -v^a_{ji} = -\frac{\sqrt{\omega_{ij}}}{g_{ij}(\rho)}a. \tag{3.11}$$

Note that $\mathrm{div}_\rho(v^a)_k = 0$ when $k \neq i, j$ and

$$\mathrm{div}_\rho(v^a)_i = \omega_{ij}a = -\mathrm{div}_\rho(v^a)_j.$$

We set

$$\sigma(s) = \rho - s\,\mathrm{div}_\rho(v^a), \quad \bar\rho = \sigma(1), \quad \bar v^a(s) = v^a\frac{g_{ij}(\rho)}{g_{ij}(\sigma(s))}, \qquad \forall s \in [0,1]. \tag{3.12}$$

Since $0 < a << 1$, the range of σ is contained in $\mathcal{P}(G)$ and the range of $g_{ij} \circ \sigma$ lies in $(0, \infty)$.

Let $\varepsilon > 0$ and let $\delta > 0$ be such that (3.9) holds for v and $\tilde{v}$. Assuming $2\omega_{ij} a \leq \delta$ we get $\|\bar{\rho} - \rho\|_{\ell_1} \leq \delta$. Since $\bar{\rho} - \rho + \mathrm{div}_\rho(\bar{v}) = 0$, we conclude that

$$\big|\mathcal{F}(\bar{\rho}) - \mathcal{F}(\rho) - (v^a, v)_\rho\big|, \; \big|\mathcal{F}(\bar{\rho}) - \mathcal{F}(\rho) - (v^a, \tilde{v})_\rho\big| \leq \varepsilon \mathcal{W}(\bar{\rho}, \rho)$$

and so,

$$\big|(v^a, v - \tilde{v})_\rho\big| \leq 2\varepsilon \mathcal{W}(\bar{\rho}, \rho). \tag{3.13}$$

But,

$$\big|(v^a, v - \tilde{v})_\rho\big| = \sqrt{\omega_{ij}}\, a |v_{ij} - \tilde{v}_{ij}| \quad \text{and} \quad \mathrm{div}_\rho(v^a) = \mathrm{div}_\sigma(\bar{v}^a). \tag{3.14}$$

The first identity in (3.12) and the last identity in (3.14) yield $\dot{\sigma} + \mathrm{div}_\sigma(\bar{v}^a) = 0$. Thus,

$$\mathcal{W}^2(\bar{\rho}, \rho) \leq \int_0^1 \|\bar{v}^a(s)\|^2_{\sigma(s)} ds = a^2 \omega_{ij} \int_0^1 \frac{1}{g(\rho_i - \omega_{ij} as, \rho_j + \omega_{ij} as)} ds.$$

We conclude that for a sufficiently small, we have

$$\mathcal{W}^2(\bar{\rho}, \rho) \leq \int_0^1 \|\bar{v}^a(s)\|^2_{\sigma(s)} ds = a^2 C^2 \omega_{ij}, \qquad C^2 := \frac{2}{g_{ij}(\rho)}. \tag{3.15}$$

This, together with (3.13) and the first identity in (3.14), implies

$$\sqrt{\omega_{ij}}\, a |v_{ij} - \tilde{v}_{ij}| \leq 2\sqrt{\omega_{ij}}\, \varepsilon a C.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $|v_{ij} - \tilde{v}_{ij}| = 0$. □

Lemma 3.14. *If $\frac{\delta \mathcal{F}}{\delta \rho}(\rho)$ exists for $\rho \in \mathcal{P}(G)$, then it is uniquely determined.*

Proof. Suppose $\xi, \tilde{\xi} \in \mathbb{R}^n$ are Fréchet derivatives of $\mathcal{F}$ at ρ. The second identity in (3.10) implies that $(\tilde{\xi} - \xi, \bar{\rho} - \rho) = 0$ for all $\bar{\rho} \in \mathcal{P}(G)$. This means that $\tilde{\xi} - \xi$ is parallel to $\mathbf{1} := (1, \cdots, 1)$. The first identity in (3.10) implies that $\tilde{\xi} - \xi$ is perpendicular to $\mathbf{1}$. Consequently, $\tilde{\xi} - \xi = 0$. □

Lemma 3.15. *Let $\mathcal{F} : \mathcal{P}(G) \to \mathbb{R}$ and $\rho \in \mathcal{P}(G)$.*

(i) *If $\mathcal{F}$ has both the Fréchet derivative and the Wasserstein gradient at ρ then $\nabla_{\mathcal{W}} \mathcal{F}(\rho) = \nabla_G(\delta \mathcal{F}/\delta \rho)(\rho)$.*
(ii) *If $\mathcal{F}$ has the Fréchet derivative in an ℓ_1-neighborhood of ρ and if $\delta \mathcal{F}/\delta \rho$ is continuous at ρ for the ℓ_1 metric, then $\mathcal{F}$ has the Wasserstein gradient at ρ and $v := \nabla_{\mathcal{W}} \mathcal{F}(\rho) = \nabla_G(\delta \mathcal{F}/\delta \rho)(\rho)$.*

Proof. (i) Suppose that $\mathcal{F}$ has both the Fréchet derivative and the Wasserstein gradient at ρ and set $v^1 = \nabla_G(\delta \mathcal{F}/\delta \rho)(\rho)$, $v^2 = \nabla_{\mathcal{W}} \mathcal{F}(\rho)$. We are to show that whenever $(i,j) \in E$ is such that $g_{ij}(\rho) > 0$, we have $v^1_{ij} = v^2_{ij}$. We can assume without loss of generality that $\rho_i \geq \rho_j$. For $0 < a << 1$, let v^a be as in (3.11) and let $\sigma^a(s) \in \mathcal{P}(G)$ be as in (3.12). We first use the fact that $\mathcal{F}$ has the Wasserstein gradient at ρ and then use that $\mathcal{F}$ has the Fréchet derivative at ρ to obtain

$$\big(v^a, v^2\big)_\rho = \lim_{s \to 0^+} \frac{\mathcal{F}(\sigma^a(s)) - \mathcal{F}(\rho)}{s} = -\bigg(\frac{\delta \mathcal{F}}{\delta \rho}(\rho), \mathrm{div}_\rho(v^a)\bigg) = \big(v^a, v^1\big)_\rho.$$

This means

$$-a \frac{\sqrt{\omega}_{ij}}{g_{ij}(\rho)} v^2_{ij} = -a \frac{\sqrt{\omega}_{ij}}{g_{ij}(\rho)} v^1_{ij}, \qquad \forall 0 < a << 1$$

and so, $v^1_{ij} = v^2_{ij}$.

(ii) Assume that $\mathcal{F}$ has the Fréchet derivative in an ℓ_1-neighborhood of ρ and $\delta\mathcal{F}/\delta\rho$ is continuous at ρ for the ℓ_1 metric. Thanks to Lemma 3.4, we may choose a constant $c \equiv c(G, g)$ such that $\| \cdot - \cdot \|_{\ell_1} \leq c\mathcal{W}(\cdot, \cdot)$. Let $\delta_0 > 0$ be such that $\mathcal{F}$ has the Fréchet derivative in B, the closed ℓ_1-ball of radius δ_0 and centered at ρ. Let $\varepsilon > 0$ and choose $\delta \in (0, \delta_0)$ such that

$$2c \sup_{\eta\in B} \Big\| \frac{\delta\mathcal{F}}{\delta\rho}(\eta) - \frac{\delta\mathcal{F}}{\delta\rho}(\rho) \Big\|_{\ell_\infty} \leq \varepsilon.$$

Assume

$$\bar\rho \in \mathcal{P}(G) \quad \text{and} \quad \|\bar\rho - \rho\|_{\ell_1} \leq \delta_0, \quad \bar v \in T_\rho\mathcal{P}(G).$$

Set $\rho_t := \rho + t(\bar\rho - \rho)$. If $t \in (0, 1)$ and $|h|$ is small enough, since $\rho_{t+h} = \rho_t + h(\bar\rho - \rho_t)$, $t \to \mathcal{F}(\rho_t)$ is differentiable on $(0, 1)$ and its Fréchet derivative is $\big(\delta\mathcal{F}/\delta\rho(\rho_t), \bar\rho - \rho\big)$. Since $\delta\mathcal{F}/\delta\rho$ is continuous at ρ, its absolute value is bounded by a constant M on B. Thus, $t \to \mathcal{F}(\rho_t)$ is Lipschitz and so,

$$\mathcal{F}(\rho_1) - \mathcal{F}(\rho_0) = \Big(\frac{\delta\mathcal{F}}{\delta\rho}(\rho), \bar\rho - \rho\Big) + \int_0^1 \Big(\frac{\delta\mathcal{F}}{\delta\rho}(\rho_t) - \frac{\delta\mathcal{F}}{\delta\rho}(\rho), \bar\rho - \rho\Big) dt.$$

Thus,

$$\begin{aligned}\mathcal{F}(\rho_1) - \mathcal{F}(\rho_0) =& \Big(\nabla_G \frac{\delta\mathcal{F}}{\delta\rho}(\rho), \bar v\Big)_\rho + \Big(\frac{\delta\mathcal{F}}{\delta\rho}(\rho), \bar\rho - \rho + \mathrm{div}_\rho(\bar v)\Big) \\ &+ \int_0^1 \Big(\frac{\delta\mathcal{F}}{\delta\rho}(\rho_t) - \frac{\delta\mathcal{F}}{\delta\rho}(\rho), \bar\rho - \rho\Big) dt.\end{aligned}$$

Hence,

$$\Big|\mathcal{F}(\bar\rho) - \mathcal{F}(\rho) - \big(v, \bar v\big)_\rho\Big| \leq \Big\|\frac{\delta\mathcal{F}}{\delta\rho}(\rho)\Big\|_{\ell_\infty} \|\bar\rho - \rho + \mathrm{div}_\rho(\bar v)\|_{\ell_1} + \sup_{\eta\in B} \Big\|\frac{\delta\mathcal{F}}{\delta\rho}(\eta) - \frac{\delta\mathcal{F}}{\delta\rho}(\rho)\Big\|_{\ell_\infty} \|\bar\rho - \rho\|_{\ell_1}.$$

We bound the ℓ_1 norm by the $\mathcal{W}$-metric and use the condition on ε to conclude (ii). ☐

Lemma 3.16. *Let $T > 0$ and $\sigma \in \mathrm{AC}_2((0, T); (\mathcal{P}(G), \mathcal{W}))$ and let v be the velocity given by Proposition 3.8. The proposition asserts that $\mathcal{T}$, the set of $t_0 \in (0, T)$ such that the metric derivative of σ at t_0 exists, $v(t_0) \in T_{\sigma(t_0)}\mathcal{P}(G)$, σ is differentiable at t_0 and*

$$\dot\sigma(t_0) + \mathrm{div}_{\sigma(t_0)}(v(t_0)) = 0, \tag{3.16}$$

is of full measure in $(0, T)$. If $\mathcal{F} : \mathcal{P}(G) \to \mathbb{R}$ has the Wasserstein gradient at $\sigma(t_0)$ and $t_0 \in \mathcal{T}$ then

$$\frac{d}{dt}\mathcal{F}(\sigma(t))\Big|_{t=t_0} = \Big(\nabla_{\mathcal{W}}\mathcal{F}(\sigma(t_0)), v(t_0)\Big)_{\sigma(t_0)}.$$

If we further assume that $\frac{\delta\mathcal{F}}{\delta\rho}(\sigma(t_0))$ exists, then

$$\frac{d}{dt}\mathcal{F}(\sigma(t))\Big|_{t=t_0} = \Big(\frac{\delta\mathcal{F}}{\delta\sigma}(\sigma(t_0)), \dot\sigma(t_0)\Big).$$

Proof. Let $t_0 \in \mathcal{T}$ and let $C > 0$ be such that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $\rho \equiv \sigma(t_0)$ and $\bar v \in T_{\sigma(t_0)}\mathcal{P}(G)$ then (3.9) holds. Let $\bar o : (-1, 1) \to \mathbb{R}$ be a function continuous at 0 and such that $\bar o(0) = 0$ and

$$\sigma(t) - \sigma(t_0) + (t - t_0)\mathrm{div}_{\sigma(t_0)}(v(t_0)) = (t - t_0)\bar o(t - t_0).$$

For $\|\sigma(t)-\sigma(t_0)\|_{\ell_1} << 1$, we use (3.9) to infer

$$\left|\frac{\mathcal{F}(\sigma(t))-\mathcal{F}(\rho))}{t-t_0} - \big(\nabla_{\mathcal{W}}\mathcal{F}(\rho), v(t_0)\big)_\rho\right| \le \varepsilon \frac{\mathcal{W}\big(\sigma(t),\rho\big)}{|t-t_0|} + C\|\bar{\sigma}(t-t_0)\|_{\ell_1}.$$

Hence,

$$\limsup_{t\to t_0}\left|\frac{\mathcal{F}(\sigma(t))-\mathcal{F}(\rho))}{t-t_0} - \big(\nabla_{\mathcal{W}}\mathcal{F}(\rho), v(t_0)\big)_\rho\right| \le \varepsilon|\sigma'|(t_0),$$

which proves the first statement of the lemma, as $\varepsilon > 0$ is arbitrary. In light of Lemma 3.15, we now conclude that the second statement of the lemma holds. □

Corollary 3.17. *Assume that $\mathcal{F}:\mathcal{P}_0(G)\to\mathbb{R}$ has a local minimum at $\rho\in\mathcal{P}_0(G)$.*

(i) *If $\mathcal{F}\in C^1\big(\mathcal{P}_0(G),\mathcal{W}\big)$ then $\nabla_{\mathcal{W}}\mathcal{F}(\rho)=0$.*
(ii) *If $\mathcal{F}\in C^1\big(\mathcal{P}_0(G),\ell_2\big)$ then $\frac{\delta\mathcal{F}}{\delta\rho}(\rho)=0$.*

Proof. (i) Assume that $\mathcal{F}\in C^1\big(\mathcal{P}_0(G),\mathcal{W}\big)$. Let $(\sigma,\bar{v}^a)$ be as in the proof of Lemma 3.13, except that now, we can choose $\delta>0$ such that $\sigma:[-\delta,\delta]\to\mathcal{P}_0(G)$. Recall the weighted metric satisfies $\omega_{ij}>0$ for any $(i,j)\in E$. By Lemma 3.16 and the minimality property of $\mathcal{F}$ and ρ, the following proves (i):

$$0=\frac{\mathcal{F}(\sigma(t))-\mathcal{F}(\rho)}{t}=\Big(\nabla_{\mathcal{W}}\mathcal{F}(\rho),\bar{v}^a(0)\Big)_\rho = a\frac{\big(\nabla_{\mathcal{W}}\mathcal{F}(\rho)\big)_{ij}\omega_{ij}}{g_{ij}(\rho)}.$$

(ii) Assume that $\mathcal{F}\in C^1\big(\mathcal{P}_0(G),\ell_2\big)$. For any $f\in\mathbb{R}^n$ such that $\sum_{i=1}^n f_i=0$, $t\to\mathcal{F}(\rho+tf)$ achieves its minimum at $t=0$ and so, its derivative at $t=0$ is null, which means $(f,\frac{\delta\mathcal{F}}{\delta\rho}(\rho))=0$. We choose $f=\frac{\delta\mathcal{F}}{\delta\rho}(\rho)$ to conclude that $\frac{\delta\mathcal{F}}{\delta\rho}(\rho)=0$. □

Definition 3.18. *If $u:\mathcal{P}(G)\to\mathbb{R}$ is differentiable at $\rho\in\mathcal{P}_0(G)$, the graph individual noise operator $\triangle_{\rm ind}$ is defined by*

$$\triangle_{\rm ind}u(\rho) := \Big(\mathrm{div}_\rho\big(\nabla_{\mathcal{W}}u(\rho)\big),\log\rho\Big). \tag{3.17}$$

When (1.3) holds, we can extend the definition of $\triangle_{\rm ind}u(\rho)$ up to the boundary of $\mathcal{P}(G)$. Integrating by parts (cf. (2.2)), we conclude that

$$\triangle_{\rm ind}u(\rho) = -\Big(\nabla_{\mathcal{W}}u(\rho),\nabla_G\log\rho\Big)_\rho. \tag{3.18}$$

Remark 3.19. *In the continuum setting, the individual noise operator is known to be a second order differential operator, obtained by differentiating Wasserstein derivatives with respect to spatial derivatives. However, in the discrete setting, the individual noise operator is obtained just as a special combination of first order Wasserstein derivatives. Here, the spatial graph gradient exists for every function since there is no notion of smoothness with respect to the graph gradient.*

3.4. **The individual noise operator $\triangle_{\rm ind}$.** The goal this section is to comment on the relation between the individual noise operator $\triangle_{\rm ind}$ and some continuous time discrete state Markov chains. For the sake of illustration, we keep our focus on the case where g satisfies (2.5). Let A be the matrix given in (1.8). It satisfies the following properties:

(a) $A_{ij}\ge 0$ for all $(i,j)\in V^2$ such that $i\neq j$;

(b) $A_{ii} = -\sum_{j\neq i} A_{ij}$ for all $i \in V$,

which, according to standard terminology in probability theory, makes A a rate–matrix (or a Q–matrix). Therefore (cf. e.g. [69]), there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $\mu \in \mathcal{P}(G)$, we can find a Markov chain $S : [0, T] \times \Omega \to V$ such that $\mathbb{P}\big(S(0, \cdot) = i\big) = \mu_i$ and

$$\mathbb{P}\Big(S(t+h, \cdot) = i \,|\, S(t, \cdot) = j\Big) = \big(e^{hA}\big)_{ji}, \qquad \forall t, h \geq 0,$$

for all $i, j \in V$ such that $\mathbb{P}(S(t, \cdot) = j) \neq 0$. Setting

$$\sigma_i(t) = \mathbb{P}\big(S(t, \cdot) = i\big), \qquad \forall i \in V,$$

it is apparent that

$$\sigma_i(t+h) = \sum_{j=1}^{n} \big(e^{hA}\big)_{ji} \sigma_j(t) = \Big(1 + A_{ii}h + o(h)\Big)\sigma_i(t) + \sum_{j\neq i}^{n} \Big(A_{ji}h + o(h)\Big)\sigma_j(t).$$

Hence, if A is symmetric, using (b), we conclude that

$$\frac{\sigma_i(t+h) - \sigma_i(t)}{h} = \sum_{j\neq i}^{n} A_{ji}\Big(\sigma_j(t) - \sigma_i(t)\Big) + \frac{o(h)}{h},$$

and so, if t is a point of differentiability for σ then

$$\dot{\sigma}_i(t) = \sum_{j\neq i}^{n} A_{ji}\big(\sigma_j(t) - \sigma_i(t)\big) \tag{3.19}$$

for all $i \in V$. By (1.8), (3.19) is equivalent to

$$\dot{\sigma}(t) = \mathrm{div}_{\sigma(t)}\big(\nabla_G(\log\sigma(t))\big). \tag{3.20}$$

Thus, the unique solution to (3.19), or equivalently the unique solution to (3.20), is given by

$$\sigma(t) = e^{tA}\mu.$$

Given a sufficiently smooth function $\mathcal{U}_0 : \mathcal{P}(G) \to \mathbb{R}$, we define $\mathcal{U} : [0, +\infty) \times \mathcal{P}(G) \to \mathbb{R}$ by

$$\mathcal{U}(t, \mu) := \mathcal{U}_0\big(\sigma(t)\big).$$

In the introduction, we recalled that for each $t \geq 0$, e^{At} is known to be a transition matrix. One checks that there exists a continuous function $t \to C_t \in (0, +\infty)$ such that if $\mu_i \geq \varepsilon$ for all $i \in V$ then $(e^{-At}\mu)_i \geq C_t\varepsilon$ for all $i \in V$. Therefore, if $\sigma(t) \in \mathcal{P}_0(G)$ then for $h > 0$ small enough, the path $h \to \nu(h) := e^{-Ah}\sigma(t)$ belongs to $\mathcal{P}_0(G)$ and satisfies the identity

$$\dot{\nu}(h) + \mathrm{div}_{\nu(h)}\big(\nabla_G(\log\nu(h))\big) = 0.$$

Since $\mathcal{U}(t+h, \nu(h)) = \mathcal{U}_0(\sigma(t))$, we use Lemma 3.16 to infer

$$0 = \frac{d}{dh}\mathcal{U}(t+h, \nu(h)) = \partial_t\mathcal{U}(t+h, \nu(h)) + \Big(\nabla_{\mathcal{W}}\mathcal{U}(t+h, \nu(h)), \nabla_G \log\nu(h)\Big)_{\nu(h)}.$$

Setting $h = 0$, we conclude that

$$0 = \partial_t\mathcal{U}(t, \sigma(t)) - \Delta_{\mathrm{ind}}\mathcal{U}(t, \sigma(t)).$$

This links the laws of the Markov chains $(S_t)_{t\geq 0}$ to the PDE

$$\partial_t\mathcal{U} = \Delta_{\mathrm{ind}}\mathcal{U}, \text{ on } (0, +\infty) \times \mathcal{P}(G), \qquad \mathcal{U}(0, \cdot) = \mathcal{U}_0. \tag{3.21}$$

4. Viscosity solutions on $\mathcal{P}(G)$.

In this section we introduce a notion of viscosity solution. We assume that (1.2) holds. We fix $T > 0$ and assume that $\mathcal{F} \in C(\mathcal{P}(G))$ and $\mathcal{H} \in C(\mathcal{P}(G) \times \mathbb{S}^{n\times n})$.

Recall that we denote by $C^1(\mathcal{P}_0(G), \ell_2)$ the set of real valued functions on $\mathcal{P}_0(G)$ which have a continuous Fréchet derivative and we denote by $C^1(\mathcal{P}_0(G), \mathcal{W})$ the set of real valued functions on $\mathcal{P}_0(G)$ which have a continuous Wasserstein gradient. By Lemma 3.15 (ii),

$$C^1\big(\mathcal{P}_0(G), \ell_2\big) \subset C^1\big(\mathcal{P}_0(G), \mathcal{W}\big).$$

Note that for $\nu \in \mathcal{P}(G)$, the function

$$\mu \to \mathcal{J}(\mu, \nu) := 1/2\|\mu - \nu\|^2_{\ell_2} \tag{4.1}$$

is of class $C^1(\mathcal{P}_0(G), \ell_2)$. Similarly, $\mathcal{J}(\mu, \cdot)$ is of class $C^1(\mathcal{P}_0(G), \ell_2)$ and we have

$$\nabla_{\mathcal{W}}\mathcal{J}(\cdot, \nu)(\mu) \equiv \nabla_G(\mu - \nu) \quad \text{and} \quad \nabla_{\mathcal{W}}\mathcal{J}(\mu, \cdot)(\nu) \equiv \nabla_G(\nu - \mu).$$

We also consider the function

$$\mu \to \mathcal{I}(\mu) := \sum_{i=1}^n \frac{1}{\mu_i} = \sum_{i=1}^n \mathcal{I}_i(\mu), \qquad \forall \mu \in \mathcal{P}_0(G), \tag{4.2}$$

which is of class $C^1(\mathcal{P}_0(G), \ell_2)$.

For each $\mu \in \mathcal{P}(G)$, we assume to be given a linear functional

$$\mathcal{O}_\mu : \mathbb{S}^{n\times n} \to \mathbb{R}$$

such that $\mu \to \mathcal{O}_\mu(p)$ is continuous for all $p \in \mathbb{S}^{n\times n}$.

Remark 4.1. *Any $\bar{\mathcal{H}} : \mathcal{P}(G) \times \mathbb{S}^{n\times n} \to \mathbb{R}$, can be written as $\bar{\mathcal{H}}(\mu, p) = \mathcal{H}(\mu, p) + \mathcal{F}(\mu)$, where*

$$\mathcal{H}(\mu, p) := \bar{\mathcal{H}}(\mu, p) - \bar{\mathcal{H}}(\mu, 0), \quad \mathcal{F}(\mu) := \bar{\mathcal{H}}(\mu, 0).$$

In the sequel, we chose to adopt the notation $\mathcal{H}(\mu, p) + \mathcal{F}(\mu)$ only to emphasize the fact that we will impose assumptions on $\bar{\mathcal{H}}(\mu, p) - \bar{\mathcal{H}}(\mu, 0)$. Therefore, $\mathcal{H}(\mu, p) + \mathcal{F}(\mu)$ represents a large class of Hamiltonians and is not limited to the class of the discrete analogue of the so-called "separable Hamiltonians". Observe that the separable Hamiltonians are widely used in the mean field control and mean field game literature, see e.g. [19, 23, 51, 76]. *In the sequel, we adopt the notation $\mathcal{H}(\mu, p) + \mathcal{F}(\mu)$ only to emphasize the fact that we are making assumptions on $\bar{\mathcal{H}}(\mu, p) - \bar{\mathcal{H}}(\mu, 0)$.*

Given $\mathcal{U}_0 : \mathcal{P}(G) \to \mathbb{R}$, we consider the Hamilton-Jacobi equation

$$\partial_t u(t, \mu) + \mathcal{H}\big(\mu, \nabla_{\mathcal{W}} u(t, \mu)\big) + \mathcal{F}(\mu) = \mathcal{O}_\mu\big(\nabla_{\mathcal{W}} u(t, \mu)\big), \qquad u(0, \cdot) = \mathcal{U}_0 \tag{4.3}$$

for a class of Hamiltonian functions $\mathcal{H}$ which will be specified later.

Definition 4.2.

(i) *A function $u \in \mathrm{USC}([0, T) \times \mathcal{P}_0(G))$ is a viscosity subsolution to* (4.3) *if $u(0, \cdot) \leq \mathcal{U}_0$ and for every $(t_0, \rho_0) \in (0, T) \times \mathcal{P}_0(G)$ and every $\varphi \in C^1\big((0, T) \times \mathcal{P}_0(G), \ell_2\big)$ such that $u - \varphi$ has a local maximum at (t_0, ρ_0), we have*

$$\partial_t \varphi(t_0, \rho_0) + \mathcal{H}(\rho_0, \nabla_{\mathcal{W}}\varphi(t_0, \rho_0)) + \mathcal{F}(\rho_0) \leq \mathcal{O}_{\rho_0}\big(\nabla_{\mathcal{W}}\varphi(t_0, \rho_0)\big).$$

(ii) *A function* $u \in \mathrm{LSC}([0,T) \times \mathcal{P}_0(G))$ *is a viscosity supersolution to* (4.3) *if* $u(0,\cdot) \geq \mathcal{U}_0$ *and for every* $(t_0, \rho_0) \in (0,T) \times \mathcal{P}_0(G)$ *and every* $\varphi \in C^1\big((0,T) \times \mathcal{P}_0(G), \ell_2\big)$ *such that* $u - \varphi$ *has a local minimum at* (t_0, ρ_0)*, we have*

$$\partial_t \varphi(t_0, \rho_0) + \mathcal{H}(\rho_0, \nabla_{\mathcal{W}} \varphi(t_0, \rho_0)) + \mathcal{F}(\rho_0) \geq \mathcal{O}_{\rho_0}\big(\nabla_{\mathcal{W}} \varphi(t_0, \rho_0)\big).$$

(iii) *A function* u *is a viscosity solution of* (4.3) *if it is both a viscosity subsolution and a viscosity supersolution.*

Remark 4.3. *By Corollary 3.17, every* $\varphi \in C^1\big((0,T) \times \mathcal{P}_0(G), \ell_2\big)$ *which achieves a local maximum at* $(t, \mu) \in (0,T) \times \mathcal{P}_0(G)$*, satisfies* $\partial_t \varphi(t,\mu) = 0$ *and* $\nabla_{\mathcal{W}} \varphi(t,\mu) = 0$*. Hence, every smooth function for which* (4.3) *holds pointwise on* $(0,T) \times \mathcal{P}_0(G)$*, is also a viscosity solution. An analogous conclusion can be drawn for viscosity subsolutions and supersolutions.*

Remark 4.4. *For any* $(i,j) \in E$ *such that* $1 \leq i < j \leq n$*, we define* $e_{ij} \in \mathbb{R}^n$ *to be such that all its entries are null, except that the* i*-th entry is* -1 *and the* j*th entry is* 1*. If* $u : \mathcal{P}(G) \to \mathbb{R}$ *and its Fréchet derivative exists at* $\rho \in \mathcal{P}_0(G)$*, we can define the following limit when it exists:*

$$\nabla^{e_{ij}} u(\rho) := \lim_{t \to 0} \frac{u(\rho + t e_{ij}) - u(\rho)}{t}.$$

When the Fréchet derivative of u *exists in a neighborhood of* ρ *and is continuous at* ρ*, then*

$$\nabla_{\mathcal{W}} u(\rho) = \nabla_G \left(\frac{\delta u}{\delta \rho} \right)(\rho)$$

and so, $\sqrt{\omega_{ij}} \nabla^{e_{ij}} u(\rho)$ *are the entries of* $\nabla_{\mathcal{W}} u(\rho)$*.*

Thus, if we consider $\mathcal{P}_0(G)$ *to be a flat Riemannian manifold,* $\nabla_{\mathcal{W}} u(\rho)$ *only depends on the derivatives of* u *in the directions that span the tangent space. Hence, we can conclude that if* u *is a Wasserstein-viscosity solution to*

$$\partial_t u(t,\rho) + \mathcal{H}(\rho, \nabla_{\mathcal{W}} u(t,\rho)) + \mathcal{F}(\rho) = \mathcal{O}_\rho\big(\nabla_{\mathcal{W}} u(t,\rho)\big)$$

then at least formally, u *is a viscosity solution to*

$$\partial_t u(t,\rho) + \mathcal{H}\Big(\rho, \big(\sqrt{\omega_{ij}} \nabla^{e_{ij}} u(t,\rho)\big)\Big) + \mathcal{F}(\rho) = \mathcal{O}_\rho\Big(\big(\sqrt{\omega_{ij}} \nabla^{e_{ij}} u(t,\rho)\big)\Big)$$

which we can consider to be a PDE on a flat Riemannian manifold. Moreover, after a change of coordinates, the equation can be transformed into an equation on $(0,T) \times \Omega$*, where* Ω *is an open subset of* $\mathbb{R}^{n-1}$*.*

5. Comparison principles

The goal of this section is to show a comparison principle for viscosity solutions to equation (4.3) and its version for a boundary value problem.

We now introduce the assumptions on $\mathcal{H}$ and $\mathcal{O}$. We fix $\kappa > 1$ and assume that and there exist positive constants $t_* > 1$ and non-negative functions $\gamma, \bar{\gamma}, \omega_* \in C([0,\infty))$ such that for any $\mu, \nu \in \mathcal{P}_0(G)$, and $p, q \in \mathbb{S}^{n \times n}$, the following hold:

(A-i) $\mathcal{H} \in C\big(\mathcal{P}_0(G) \times \mathbb{S}^{n \times n}\big)$ and $\mathcal{H}(\mu, \cdot)$ is convex.

(A-ii) $\lim_{t \to 1^+} \bar{\gamma}(t) = 1$, $\gamma(t) > 1$ for any $t \in (1, t_*)$ and we have

$$t\gamma(t)\mathcal{H}(\mu, p) \leq \mathcal{H}(\mu, tp) \leq \bar{\gamma}(t)\, \mathcal{H}(\mu, p), \quad \forall t > 0.$$

(A-iii) For every $0 < \varepsilon < 1$ there exists $\theta_\varepsilon > 0$ such that $\theta_\varepsilon \|p\|_\mu^\kappa \leq \mathcal{H}(\mu, p)$ for all $\mu \in \mathcal{P}_\varepsilon(G)$.

(A-iv) We have $\mathcal{H}(\mu,0)=0$ and there are moduli ω_ε and constants C_ε for $0<\varepsilon<1$ such that

$$\mathcal{H}(\mu,p)-\mathcal{H}(\nu,p)\geq -\omega_\varepsilon(\|\mu-\nu\|_{\ell_2})\|p\|_\mu^\kappa - C_\varepsilon\big|\|p\|_\mu-\|p\|_\nu\big|\Big(\|p\|_\mu^{\kappa-1}+\|p\|_\nu^{\kappa-1}\Big),\ \ \forall\mu\in\mathcal{P}_\varepsilon(G).$$

(A-v) If $\mathcal{I}$ is as in (4.2) then

$$|\mathcal{H}(\mu,p)|\leq C\|p\|_\mu^\kappa \mathcal{I}(\mu)^{-\kappa},\quad \forall(\mu,p)\in\mathcal{P}_0(G)\times\mathbb{S}^{n\times n}.$$

($\mathcal{O}$) There exist a constant $C\geq 0$ and for every $0<\varepsilon<1$ a constant C_ε such that for every $b_1,b_2\geq 0$ (if $\mathcal{J}$ is as in (4.1))

$$\begin{aligned}&\mathcal{O}_\mu\Big(b_1\nabla_{\mathcal{W}}\mathcal{J}(\cdot,\nu)(\mu)+b_2\nabla_{\mathcal{W}}\mathcal{I}(\mu)\Big)+\mathcal{O}_\nu\Big(b_1\nabla_{\mathcal{W}}\mathcal{J}(\mu,\cdot)(\nu)+b_2\nabla_{\mathcal{W}}\mathcal{I}(\nu)\Big)\\ &\leq C_\varepsilon b_1\|\mu-\nu\|_{\ell_2}^2+Cb_2(\|\nabla_{\mathcal{W}}\mathcal{I}(\mu)\|_\mu\mathcal{I}(\mu)^{-1}+\|\nabla_{\mathcal{W}}\mathcal{I}(\nu)\|_\nu\mathcal{I}(\nu)^{-1}),\quad \forall\mu,\nu\in\mathcal{P}_\varepsilon(G).\end{aligned}\tag{5.1}$$

Example 5.1. *Let $a\in C(\mathcal{P}(G))$ be non-negative such that $a\mathcal{I}^\kappa$ is bounded from above and for every $\varepsilon>0$, there exists $\theta_\varepsilon>0$ such that $a\geq\theta_\varepsilon$ when $\mu\in\mathcal{P}_\varepsilon(G)$. Setting $\mathcal{H}(\mu,p):=a(\mu)\|p\|_\mu^\kappa$, we have*

$$\mathcal{H}(\mu,p)=\mathcal{H}(\nu,q)+\big(a(\mu)-a(\nu)\big)\|p\|_\mu^\kappa+a(\nu)\big(\|p\|_\mu^\kappa-\|q\|_\nu^\kappa\big).$$

We choose ω_* to be the modulus of continuity of a and we use the fact that

$$\Big|\|p\|_\mu^\kappa-\|q\|_\nu^\kappa\Big|\leq\kappa\big|\|p\|_\mu-\|q\|_\nu\big|\left(\|p\|_\mu^{\kappa-1}+\|q\|_\nu^{\kappa-1}\right),$$

to conclude that (A-i)-(A-v) hold.

Observe that the ℓ_2-Lipschitz constant of the function $J:=\mathcal{I}^{-1}$ on $\mathcal{P}_0(G)$ is less than or equal to 1 and so, J admits a unique Lipschitz extension on $\mathcal{P}(G)$ which we continue to denote by J. Since on $\mathcal{P}_0(G)$, $J(\mu)\leq\mu_i$ for all $i\in V$, one concludes that $nJ\leq\sum_{i\in V}\mu_i=1$ on $\mathcal{P}(G)$, and J vanishes on the boundary of $\mathcal{P}(G)$. Therefore, (A-i)-(A-v) hold for

$$a(\mu):=C_0J^\kappa(\mu),\quad \theta_\varepsilon=C_0\varepsilon^\kappa n^{-\kappa},\quad C_\varepsilon:=\kappa C_0 n^{-\kappa}.$$

Remark 5.2. *Since $\mathcal{I}^{-1}$ is bounded from above by n, (A-v) implies that*

$$|\mathcal{H}(\mu,p)|\leq Cn^{-\kappa}\|p\|_\mu^\kappa,\quad\forall(\mu,p)\in\mathcal{P}(G)\times\mathbb{S}^{n\times n}.\tag{5.2}$$

Example 5.3. *Assume that $\mathcal{O}_\mu$ is the graph individual noise operator so that*

$$\mathcal{O}_\mu(p)=-\big(p,\nabla_G\log\mu\big)_\mu.$$

We have

$$\begin{aligned}\mathcal{O}_\mu\Big(\nabla_{\mathcal{W}}\mathcal{I}(\mu)\Big)&=-\frac{1}{2}\sum_{(k,l)\in E}\big(\nabla_{\mathcal{W}}\mathcal{I}(\mu)\big)_{kl}g_{kl}(\mu)\big(\nabla_G\log\mu\big)_{kl}\\ &=-\frac{1}{2}\sum_{(k,l)\in E}\big(\sum_{j=1}^n\nabla_{\mathcal{W}}\mathcal{I}_j(\mu)\big)_{kl}g_{kl}(\mu)\big(\nabla_G\log\mu\big)_{kl}.\end{aligned}$$

One checks that

(5.3)

$$\frac{\delta\mathcal{I}_j}{\delta\mu}(\mu)=\frac{1}{\mu_j^2}\Big(\frac{1}{n},\cdots,\frac{1}{n},\frac{1}{n}-1,\frac{1}{n},\cdots,\frac{1}{n}\Big)^T,\quad \nabla_G\Big(\frac{\delta\mathcal{I}_j}{\delta\mu}\Big)(\mu)=\begin{cases}0 & \textit{if } k,l\neq j \textit{ or } k=l=j,\\ -\sqrt{\omega_{jl}}\mu_j^{-2} & \textit{if } k=j,l\neq j,\\ \sqrt{\omega_{jk}}\mu_j^{-2} & \textit{if } k\neq j,l=j.\end{cases}$$

Hence,

$$\mathcal{O}_\mu\Big(\nabla_{\mathcal{W}}\mathcal{I}(\mu)\Big) = \sum_{(j,l)\in E}\omega_{jl}g_{jl}(\mu)\frac{1}{\mu_j^2}\big(\log\mu_j - \log\mu_l\big)$$

$$= \sum_{(j,l)\in E, j<l}\omega_{jl}g_{jl}(\mu)\bigg(\frac{1}{\mu_j^2}-\frac{1}{\mu_l^2}\bigg)\big(\log\mu_j - \log\mu_l\big)$$

$$= -\sum_{(j,l)\in E, j<l}\omega_{jl}g_{jl}(\mu)\bigg(\frac{\mu_l+\mu_j}{\mu_j^2\mu_l^2}\bigg)\big(\log\mu_j - \log\mu_l\big)(\mu_j-\mu_l) \le 0, \tag{5.4}$$

where we have used the fact that $\big(\log\mu_j - \log\mu_l\big)(\mu_j-\mu_l)\ge 0$.

Note that

$$\mathcal{O}_\mu\Big(\nabla_{\mathcal{W}}\mathcal{J}(\cdot,\nu)(\mu)\Big) = -\frac{1}{2}\sum_{(i,j)\in E}\omega_{ij}\big((\mu_i-\nu_i)-(\mu_j-\nu_j)\big)\big(\log\mu_i-\log\mu_j\big)g_{ij}(\mu).$$

We similarly compute $\mathcal{O}_\nu\Big(\nabla_{\mathcal{W}}\mathcal{J}(\mu,\cdot)(\nu)\Big)$ *to conclude that*

$$\mathcal{O}_\mu\Big(\nabla_{\mathcal{W}}\mathcal{J}(\cdot,\nu)(\mu)\Big) + \mathcal{O}_\nu\Big(\nabla_{\mathcal{W}}\mathcal{J}(\mu,\cdot)(\nu)\Big)$$
$$= -\frac{1}{2}\sum_{(i,j)\in E}\omega_{ij}\big((\mu_i-\nu_i)-(\mu_j-\nu_j)\big)\Big(\big(\log\mu_i-\log\mu_j\big)g_{ij}(\mu) - \big(\log\nu_i-\log\nu_j\big)g_{ij}(\nu)\Big).$$

We denote by E_{ij} *each one of the expressions in the above sum. Since*

$$E_{ij} = -\frac{1}{2}\omega_{ij}\big((\mu_i-\nu_i)-(\mu_j-\nu_j)\big)\Big(\big(\log\mu_i-\log\nu_i\big)+\big(\log\nu_j-\log\mu_j\big)\Big)g_{ij}(\mu)$$
$$-\frac{1}{2}\omega_{ij}\big((\mu_i-\nu_i)-(\mu_j-\nu_j)\big)\big(\log\nu_i-\log\nu_j\big)(g_{ij}(\mu)-g_{ij}(\nu)),$$

we conclude that

$$E_{ij}\le C_\varepsilon\|\mu-\nu\|_{\ell_2}^2$$

where

$$C_\varepsilon := 2C_\omega\log\Big(\frac{1}{\varepsilon}\Big)\ \mathrm{Lip}(g|_{[\varepsilon,1]^2}) + \frac{2C_\omega}{\varepsilon}.$$

Hence,

$$\mathcal{O}_\mu\Big(\nabla_{\mathcal{W}}\mathcal{J}(\cdot,\nu)(\mu)\Big) + \mathcal{O}_\nu\Big(\nabla_{\mathcal{W}}\mathcal{J}(\mu,\cdot)(\nu)\Big) \le n^2C_\varepsilon\|\mu-\nu\|_{\ell_2}^2.$$

This concludes the proof of (5.1).

Remark 5.4. *The conclusion* (5.4) *in Example 5.3 continues to hold if instead of* $\mathcal{I}(\mu)=\sum_{i\in V}1/\mu_i$, *we take* $\mathcal{I}(\mu)=\sum_{i\in V}\ell(\mu_i)$ *for any positive function* $\ell\in C^\infty(0,+\infty)$ *such that* $\ell'<0$.

Let u be a viscosity subsolution and v be a viscosity supersolution to (4.3) such that u and $-v$ are bounded above. For any $a,\beta,\varepsilon,\delta\in(0,1]$, $\lambda\in(\frac{1}{2},1]$, we define

$$\Psi_0(t,s,\mu,\nu) := \lambda u(t,\mu) - v(s,\nu) - \frac{\beta}{T-t} - \frac{\beta}{T-s}$$

and

$$\Psi_{a,\varepsilon,\delta}(t,s,\mu,\nu) := \Psi_0(t,s,\mu,\nu) - \frac{\|\mu-\nu\|_{\ell_2}^2}{2\varepsilon} - \frac{(t-s)^2}{2\delta} - a\sum_{i=1}^n\Big(\frac{1}{\mu_i}+\frac{1}{\nu_i}\Big).$$

We set

$$M := \sup_{[0,T)\times\mathcal{P}_0(G)} \Psi_0(t,t,\mu,\mu),$$

$$M_a := \sup_{[0,T)\times\mathcal{P}_0(G)} \left(\Psi_0(t,t,\mu,\mu) - 2a\sum_{i=1}^n \frac{1}{\mu_i}\right),$$

$$M_{a,\varepsilon} := \sup_{[0,T)\times\mathcal{P}_0(G)^2} \left(\Psi_0(t,t,\mu,\mu) - \frac{\|\mu-\nu\|_{\ell_2}^2}{2\varepsilon} - a\sum_{i=1}^n \left(\frac{1}{\mu_i}+\frac{1}{\nu_i}\right)\right),$$

$$M_{a,\varepsilon,\delta} := \sup_{[0,T)^2\times\mathcal{P}_0(G)^2} \Psi_{a,\varepsilon,\delta}.$$

Since for every $\beta, a, \varepsilon, \delta \in (0,1]$ and $\frac{1}{2} \le \lambda \le 1$, $M_{a,\varepsilon,\delta} \le M_*$ for some constant M_*, it is easy to see (see e.g. [31], Proposition 3.7 for such argument) that

$$\lim_{\delta\to 0} M_{a,\varepsilon,\delta} = M_{a,\varepsilon}, \tag{5.5}$$

$$\lim_{\delta\to 0} M_{a,\varepsilon} = M_a, \tag{5.6}$$

$$\lim_{\delta\to 0} M_a = M. \tag{5.7}$$

Theorem 5.5 (Comparison Principle)**.** *Assume that $\mathcal{H}$ satisfies (A-i)-(A-v) and $\mathcal{F} \in C(\mathcal{P}(G))$. Assume further that $\mathcal{O}$ is as above and satisfies ($\mathcal{O}$). If u is a viscosity subsolution to (4.3), v is a viscosity supersolution to (4.3), $u, -v$ are bounded above and $u(0,\cdot) \le v(0,\cdot)$ on $\mathcal{P}_0(G)$, then $u \le v$ in $[0,T) \times \mathcal{P}_0(G)$.*

Proof. Suppose on the contrary that $u \le v$ in $[0,T)\times\mathcal{P}_0(G)$ fails. Let $(\tilde{t},\tilde{\mu}) \in (0,T)\times\mathcal{P}_0(G)$ be such that $3e := u(\tilde{t},\tilde{\mu}) - v(\tilde{t},\tilde{\mu}) > 0$.

Step 1. Properties of maximizer of $\Psi_{a,\varepsilon,\delta}$. We will use the notation Ψ in place of $\Psi_{a,\varepsilon,\delta}$ and to alleviate the notation, we simply denote a maximizer of $\Psi_{a,\varepsilon,\delta}$ by $(\bar{t},\bar{s},\bar{\mu},\bar{\nu})$, without displaying the dependence in $\beta, a, \varepsilon, \delta$. It is clear that there exist $0 < \lambda_0 < 1, \beta_0 > 0, a_0 > 0$ such that if $\lambda_0 < \lambda < 1, 0 < \beta < \beta_0$ and $0 < a < a_0$, then $\Psi(\bar{t},\bar{s},\bar{\mu},\bar{\nu}) > 2e$ and $\lambda u(0,\bar{\mu}) - v(0,\bar{\mu}) < e$. Moreover, we always have

$$\bar{\mu}_i,\ \bar{\nu}_i \ge c_1 a, \qquad \forall i \in V \tag{5.8}$$

for some independent constant c_1.

We start by observing that

$$M_{a,\varepsilon,\delta} + \frac{(\bar{t}-\bar{s})^2}{4\delta} = \Psi(\bar{t},\bar{s},\bar{\mu},\bar{\nu}) + \frac{(\bar{t}-\bar{s})^2}{4\delta} \le M_{a,\varepsilon,2\delta} \tag{5.9}$$

and

$$M_{a,\varepsilon,\delta} + \frac{\|\bar{\mu}-\bar{\nu}\|^2}{4\varepsilon} + \frac{(\bar{t}-\bar{s})^2}{4\delta} \le M_{a,2\varepsilon,2\delta}. \tag{5.10}$$

Thus, (5.9), together with (5.5), implies that

$$\lim_{\delta\to 0} \frac{(\bar{t}-\bar{s})^2}{\delta} = 0, \qquad \forall a,\ \varepsilon > 0. \tag{5.11}$$

Now (5.5), (5.6) and (5.10) give us

$$\lim_{\varepsilon\to 0}\limsup_{\delta\to 0}\frac{\|\bar\mu-\bar\nu\|^2_{\ell_2}}{\varepsilon}=0. \tag{5.12}$$

Similarly, since

$$M_{a,\varepsilon,\delta}+\frac{a}{2}(\mathcal{I}(\bar\mu)+\mathcal{I}(\bar\nu))+\frac{\|\bar\mu-\bar\nu\|^2}{4\varepsilon}+\frac{(\bar t-\bar s)^2}{4\delta}\le M_{a/2,2\varepsilon,2\delta}, \tag{5.13}$$

(5.5), (5.6) and (5.7) yield

$$\lim_{a\to 0}\limsup_{\varepsilon\to 0}\limsup_{\delta\to 0} a(\mathcal{I}(\bar\mu)+\mathcal{I}(\bar\nu))=0. \tag{5.14}$$

Since Ψ is upper semicontinuous, in particular it follows from (5.8), (5.11) and (5.12) (even though the full conclusions of (5.8), (5.11), (5.12) are not necessary) that for $\lambda_0<\lambda<1, 0<\beta<\beta_0, 0<a<a_0$ and for sufficiently small ε,δ, we must have $\bar t,\bar s>0$.

Step 2. Control on gradients of C^1 functions which touch u from above or touch v from below.

Observe that,

$$\varphi:(t,\mu)\to\frac{\beta}{\lambda(T-t)}+\frac{\mathcal{J}(\mu,\bar\nu)}{\lambda\varepsilon}+\frac{(t-\bar s)^2}{2\lambda\delta}+\frac{a}{\lambda}\sum_{i=1}^n\frac{1}{\mu_i}$$

belongs to $C^1\big((0,T)\times\mathcal{P}_0(G),\ell_2\big)$ and is such that $u-\varphi$ achieves its maximum at $(\bar t,\bar\mu)$ in $(0,T)\times\mathcal{P}_0(G)$. Since u is a viscosity subsolution, we infer

$$\frac{\beta}{(T-\bar t)^2}+\frac{\bar t-\bar s}{\delta}+\lambda\mathcal{H}\Big(\bar\mu,\frac{\bar p}{\lambda}\Big)+\lambda\mathcal{F}(\bar\mu)\le\lambda\mathcal{O}_{\bar\mu}\Big(\frac{\bar p}{\lambda}\Big),$$

where we have set

$$\bar p:=\frac{\nabla_{\mathcal{W}}\mathcal{J}(\cdot,\bar\nu)(\bar\mu)}{\varepsilon}+a\nabla_{\mathcal{W}}\mathcal{I}(\bar\mu)=:\bar p_1+\bar p_2.$$

Let $\mathcal{F}_\infty\in\mathbb{R}$ be such that $|\mathcal{F}|\le\mathcal{F}_\infty$. We have

$$\frac{\beta}{T^2}+\frac{\bar t-\bar s}{\delta}+\lambda\mathcal{H}\Big(\bar\mu,\frac{\bar p}{\lambda}\Big)+\mathcal{F}(\bar\mu)-\mathcal{O}_{\bar\mu}(\bar p)\le(1-\lambda)\mathcal{F}_\infty. \tag{5.15}$$

By (5.3), we can find a constant C independent of μ such that

$$\|\nabla_{\mathcal{W}}\mathcal{I}(\bar\mu)\|_{\bar\mu}\le C\sum_{i=1}^n\frac{1}{\bar\mu_i^2}. \tag{5.16}$$

Since $\mathcal{H}(\bar\mu,\cdot)$ is a convex function and $\eta:=(1+\lambda)/2$ is between 0 and 1, we have

$$\lambda\mathcal{H}\Big(\bar\mu,\frac{\bar p}{\lambda}\Big)\ge\frac{\lambda}{\eta}\mathcal{H}\Big(\bar\mu,\eta\frac{\bar p_1}{\lambda}\Big)-\frac{\lambda(1-\eta)}{\eta}\mathcal{H}\Big(\bar\mu,\frac{\eta}{1-\eta}\frac{\bar p_2}{\lambda}\Big).$$

Using (5.16) and (A-v), we obtain for a constant $\bar C>C$ independent of a,ε,δ such that

$$\lambda\mathcal{H}\Big(\bar\mu,\frac{\bar p}{\lambda}\Big)\ge\frac{\lambda}{\eta}\mathcal{H}\Big(\bar\mu,\eta\frac{\bar p_1}{\lambda}\Big)-\bar C\left|\frac{\eta}{(1-\eta)\lambda}\right|^\kappa\left(a^\kappa\sum_{i=1}^n\frac{1}{\bar\mu_i^{2\kappa}}\right)\frac{1}{\mathcal{I}(\bar\mu)^\kappa}.$$

By (5.14), we can find $\omega(a,\varepsilon,\delta)$ such that $\lim_{a\to 0}\limsup_{\varepsilon\to 0}\limsup_{\delta\to 0}\omega(a,\varepsilon,\delta)=0$ and

$$\lambda\mathcal{H}\Big(\bar\mu,\frac{\bar p}{\lambda}\Big)\ge\frac{\lambda}{\eta}\mathcal{H}\Big(\bar\mu,\eta\frac{\bar p_1}{\lambda}\Big)-\omega(a,\varepsilon,\delta).$$

Now (A-ii) and (5.15) imply

$$\frac{\beta}{T^2} + \frac{\bar t - \bar s}{\delta} + \gamma\Big(\frac{\eta}{\lambda}\Big)\mathcal{H}(\bar\mu, \bar p_1) + \mathcal{F}(\bar\mu) - \mathcal{O}_{\bar\mu}(\bar p) \le (1-\lambda)\mathcal{F}_\infty + \omega(a,\varepsilon,\delta). \tag{5.17}$$

Similarly,

$$\tilde\varphi : (s,\nu) \to \frac{\beta}{T-s} + \frac{\mathcal{J}(\bar\mu,\nu)}{\varepsilon} + \frac{(\bar t - s)^2}{2\delta} + a\sum_{i=1}^n \frac{1}{\nu_i}$$

belongs to $C^1\big((0,T)\times\mathcal{P}_0(G),\ell_2\big)$ and is such that $v + \tilde\varphi$ achieves its minimum at $(\bar s,\bar\nu)$ in $(0,T)\times\mathcal{P}_0(G)$. Using the fact that v is a viscosity supersolution, we infer

$$-\frac{\beta}{T^2} - \frac{\bar s - \bar t}{\delta} + \mathcal{H}(\bar\nu,\bar q) + \mathcal{F}(\bar\nu) - \mathcal{O}_{\bar\nu}(\bar q) \ge 0. \tag{5.18}$$

Here, we have set

$$\bar q := -\frac{1}{\varepsilon}\nabla_{\mathcal{W}}\mathcal{J}(\bar\mu,\cdot)(\bar\nu) - a\nabla_{\mathcal{W}} I(\bar\nu) =: -\bar q_1 - \bar q_2.$$

We notice that $-\bar q_1 = \bar p_1$.

Since $\eta > \lambda$, in light of (A-ii), for $\tau < 1$ sufficiently close to 1 we have

$$r := \gamma\Big(\frac{\eta}{\lambda}\Big) - \tau\bar\gamma\Big(\frac{1}{\tau}\Big) > 0.$$

Similarly as before, we use the convexity of $\mathcal{H}(\bar\nu,\cdot)$, (A-ii) and (A-v), to obtain

$$\mathcal{H}\Big(\bar\nu,\bar q\Big) \le \tau\mathcal{H}\Big(\bar\nu,\frac{\bar p_1}{\tau}\Big) + (1-\tau)\mathcal{H}\Big(\bar\nu, -\frac{1}{1-\tau}\bar q_2\Big) \le \tau\bar\gamma\Big(\frac{1}{\tau}\Big)\mathcal{H}\Big(\bar\nu,\bar p_1\Big) + \omega(a,\varepsilon,\delta),$$

where ω is as before. This, together with (5.18) implies that

$$-\frac{\beta}{T^2} - \frac{\bar s - \bar t}{\delta} + \tau\bar\gamma\Big(\frac{1}{\tau}\Big)\mathcal{H}(\bar\nu,\bar p_1) + \mathcal{F}(\bar\nu) - \mathcal{O}_{\bar\nu}(\bar q) + \omega(a,\varepsilon,\delta) \ge 0.$$

We combine this with (5.17) to conclude that

$$\begin{aligned}&\gamma\Big(\frac{\eta}{\lambda}\Big)\mathcal{H}(\bar\mu,\bar p_1) - \tau\bar\gamma\Big(\frac{1}{\tau}\Big)\mathcal{H}(\bar\nu,\bar p_1) + \mathcal{F}(\bar\mu) - \mathcal{F}(\bar\nu)\\ &\le (1-\lambda)\mathcal{F}_\infty - 2\beta T^{-2} + \mathcal{O}_{\bar\mu}(\bar p) - \mathcal{O}_{\bar\nu}(\bar q) + \omega(a,\varepsilon,\delta).\end{aligned}$$

By (5.1), (5.12), (5.14) and (5.16),

$$\gamma\Big(\frac{\eta}{\lambda}\Big)\mathcal{H}(\bar\mu,\bar p_1) - \tau\bar\gamma\Big(\frac{1}{\tau}\Big)\mathcal{H}(\bar\nu,\bar p_1) + \mathcal{F}(\bar\mu) - \mathcal{F}(\bar\nu) \le (1-\lambda)\mathcal{F}_\infty - 2\beta T^{-2} + \omega(a,\varepsilon,\delta)$$

(for a different $\omega(a,\varepsilon,\delta)$ satisfying the same properties) and hence, using (A-iii),

$$\tau\bar\gamma\Big(\frac{1}{\tau}\Big)\Big(\mathcal{H}\big(\bar\mu,\bar p_1\big) - \mathcal{H}\big(\bar\nu,\bar p_1\big)\Big) + \mathcal{F}(\bar\mu) - \mathcal{F}(\bar\nu) + r\theta_{ac_1}\|\bar p_1\|_{\bar\mu}^\kappa \le (1-\lambda)\mathcal{F}_\infty - 2\beta T^{-2} + \omega(a,\varepsilon,\delta).$$

Thanks to (A-iv), we conclude that if $\omega_{\mathcal{F}}$ is the ℓ_2-modulus of continuity of $\mathcal{F}$ then

$$\begin{aligned}&-\tau\bar\gamma\Big(\frac{1}{\tau}\Big)\Big(\omega_{ac_1}(\|\bar\mu-\bar\nu\|_{\ell_2})\|\bar p_1\|_{\bar\mu}^\kappa + C_{ac_1}\big|\|\bar p_1\|_{\bar\mu} - \|\bar p_1\|_{\bar\nu}\big|\ \Big(\|\bar p_1\|_{\bar\mu}^{\kappa-1} + \|\bar p_1\|_{\bar\nu}^{\kappa-1}\Big)\Big) + r\theta_{ac_1}\|\bar p_1\|_{\bar\mu}^\kappa\\ &\le (1-\lambda)\mathcal{F}_\infty - 2\beta T^{-2} + \omega_{\mathcal{F}}\big(\|\bar\mu-\bar\nu\|_{\ell_2}\big) + \omega(a,\varepsilon,\delta).\end{aligned} \tag{5.19}$$

Step 3. Relative smallness of $\big|\|\bar p_1\|_{\bar\mu} - \|\bar p_1\|_{\bar\nu}\big|$. Using the fact that $\mu_i, \nu_i \geq ac_1$ for all $i = 1, ..., n$, we easily have

$$\begin{aligned}\big|\|\bar p_1\|_{\bar\mu} - \|\bar p_1\|_{\bar\nu}\big| &\leq \big|\|\bar p_1\|^2_{\bar\mu} - \|\bar p_1\|^2_{\bar\nu}\big|^{\frac12} \\ &= \left(\frac12\Big|\sum_{(i,j)\in E}(\bar p_1)^2_{ij}\Big(g_{ij}(\bar\mu) - g_{ij}(\bar\nu)\Big)\Big|\right)^{\frac12} \leq K_a\|\bar p_1\|_{\bar\mu}\|\|\bar\mu-\bar\nu\|^{\frac12}_{\ell_2}\end{aligned}$$

and

$$\|\bar p_1\|_{\bar\nu} \leq K_a\|\bar p_1\|_{\bar\mu}$$

for some constant K_a.

Putting it all together in (5.19) we obtain that for some constant K_a

$$\begin{aligned}&-K_a\Big(\omega_{ac_1}(\|\bar\mu-\bar\nu\|_{\ell_2}) + \|\bar\mu-\bar\nu\|^{\frac12}_{\ell_2}\Big)\|\bar p_1\|^\kappa_{\bar\mu} + r\theta_{ac_1}\|\bar p_1\|^\kappa_{\bar\mu} \\ \leq &(1-\lambda)\mathcal{F}_\infty - 2\beta T^{-2} + \omega_{\mathcal{F}}\big(\|\bar\mu-\bar\nu\|_{\ell_2}\big) + \omega(a,\varepsilon,\delta).\end{aligned}$$

We now take λ so that $(1-\lambda)\mathcal{F}_\infty < \beta T^{-2}$ and then take $\lim_{a\to0}\limsup_{\varepsilon\to0}\limsup_{\delta\to0}$ of both sides of the above and use (5.12) to obtain a contradiction. □

We next show that a comparison principle still holds even if we weaken the assumptions on $\mathcal{H}$ and $\mathcal{O}_\mu$, provided we have additional information about how u and v behave on $[0,T)\times\partial\mathcal{P}(G)$.

Theorem 5.6 (Comparison Principle, Boundary Condition)**.** *Let the assumptions of Theorem 5.5 be satisfied except that we now only require* $\mathcal{H}$ *to satisfy (A-i)-(A-iv) and* $\mathcal{O}_\mu$ *to satisfy (*$\mathcal{O}$*) with* $b_2 = 0$*. If* $u \in USC([0,T)\times\mathcal{P}(G))$ *is a viscosity subsolution to* (4.3)*,* $v \in LSC([0,T)\times\mathcal{P}(G))$ *is a viscosity supersolution to* (4.3)*,* $u, -v$ *are bounded above,* $u(0,\cdot) \leq v(0,\cdot)$ *on* $\mathcal{P}(G)$ *and* $u \leq v$ *on* $[0,T)\times\partial\mathcal{P}(G)$*, then* $u \leq v$ *in* $[0,T)\times\mathcal{P}(G)$*.*

Proof. Since the arguments here are similar to those of the proof of Theorem 5.5, we just sketch the necessary adjustments. Suppose that $u \not\leq v$ on $[0,T)\times\mathcal{P}(G)$. For $0<\lambda<1, \beta, \varepsilon, \delta > 0$ we consider the function

$$\Psi_{\varepsilon,\delta}(t,s,\mu,\nu) := \lambda u(t,\mu) - v(s,\nu) - \frac{\|\mu-\nu\|^2_{\ell_2}}{2\varepsilon} - \frac{(t-s)^2}{2\delta} - \frac{\beta}{T-t} - \frac{\beta}{T-s}$$

and we denote its maximizer by $(\bar t, \bar s, \bar\mu, \bar\nu)$. It is easy to see that there exist $0 < \lambda_0 < 1, \beta_0 > 0$ such that for every $\lambda_0 < \lambda < 1, 0 < \beta < \beta_0$ there is $\eta > 0$ (depending only on λ, β) such that for sufficiently small $\varepsilon, \delta > 0$, we have $\eta < \bar t, \bar s < T - \eta, \bar\mu, \bar\nu \in \mathcal{P}_\eta$. The proof now repeats the lines of the proof of Theorem 5.5 and is easier since we do not need to deal with terms coming from the functions $\mathcal{I}(\mu)$ and $\mathcal{I}(\nu)$. We have in place of (5.15)

$$\frac{\beta}{T^2} + \frac{\bar t - \bar s}{\delta} + \gamma(\frac{1}{\lambda})\mathcal{H}\Big(\bar\mu, \bar p\Big) + \mathcal{F}(\bar\mu) - \mathcal{O}_{\bar\mu}(\bar p) \leq (1-\lambda)\mathcal{F}_\infty,$$

where

$$\bar p := \frac{\nabla_{\mathcal{W}}\mathcal{J}(\cdot,\bar\nu)(\bar\mu)}{\varepsilon}.$$

The part from (5.15) to (5.17) is skipped and we have in place of (5.18)

$$-\frac{\beta}{T^2} - \frac{\bar s - \bar t}{\delta} + \mathcal{H}(\bar\nu, \bar p) + \mathcal{F}(\bar\nu) - \mathcal{O}_{\bar\nu}(\bar p) \geq 0.$$

We set $r = \gamma(\frac{1}{\lambda}) - 1 > 0$ and we obtain instead of (5.19),

$$\begin{aligned} & -\omega_\eta(\|\bar\mu - \bar\nu\|_{\ell_2})\|\bar p\|^\kappa_{\bar\mu} - C_\eta\big|\|\bar p_1\|_{\bar\mu} - \|\bar p_1\|_{\bar\nu}\big| \left(\|\bar p_1\|^{\kappa-1}_{\bar\mu} + \|\bar p_1\|^{\kappa-1}_{\bar\nu}\right) + r\theta_\eta\|\bar p_1\|^\kappa_{\bar\mu} \\ & \le (1-\lambda)\mathcal{F}_\infty - 2\beta T^{-2} + \omega_{\mathcal{F}}\big(\|\bar\mu - \bar\nu\|_{\ell_2}\big) + C_\eta \frac{\|\bar\mu - \bar\nu\|^2_{\ell_2}}{\varepsilon}. \end{aligned} \tag{5.20}$$

This allows us to conclude as in Step 3 of the proof of Theorem 5.5 by taking $\lim_{\varepsilon\to 0}\limsup_{\delta\to 0}$ of both sides of the above. □

6. Perron's method

The goal of this section is to use Perron's method to show the existence of a viscosity solution to (4.3). Throughout the section, we assume that $\mathcal{F} \in C(\mathcal{P}(G))$, $\mathcal{H}$ is continuous on $\mathcal{P}_0(G) \times \mathbb{S}^{n\times n}$ and $\mathcal{O}_\mu : \mathbb{S}^{n\times n} \to \mathbb{R}$ is linear, $\mu \to \mathcal{O}_\mu(p)$ is continuous for all $p \in \mathbb{S}^{n\times n}$ and there exists a constant $C_{\mathcal{O}}$ such that

$$|\mathcal{O}_\mu(p)| \le C_{\mathcal{O}}\|p\|_{\ell_2}, \qquad \forall (\mu, p) \in \mathcal{P}_0(G) \times \mathbb{S}^{n\times n}. \tag{6.1}$$

For example when (1.3) holds, the individual noise operator satisfies (6.1).

When $\mathbb{S}$ is a topological space, for a function f defined on a subset of $Q \subset \mathbb{S}$, we will write f^* to denote its upper semicontinuous envelope and f_* to denote its lower semicontinuous envelope, i.e.

$$f^*(y) = \limsup_{z\to y} f(z) \quad \text{and} \quad f_*(y) = \liminf_{z\to y} f(z).$$

In Lemma 6.1 we do not consider the initial condition to be part of the definition of viscosity subsolution and we consider viscosity subsolutions to be functions on $(0, T) \times \mathcal{P}_0(G)$.

Lemma 6.1. *Let $\mathcal{S}$ be a family of viscosity subsolutions to* (4.3). *Let $v := \sup\{w\,;\, w \in \mathcal{S}\}$ and assume that $v^* < +\infty$ on $(0, T) \times \mathcal{P}_0(G)$. Then v^* is a viscosity subsolution to* (4.3).

Proof. Suppose that $\varphi \in C^1\big((0, T) \times \mathcal{P}_0(G), \ell_2\big)$ and there exists $r > 0$ and $(t^0, \mu^0) \in (0, T) \times \mathcal{P}_0(G)$ such that $v^* - \varphi$ achieves its maximum on $\bar B_r(t^0, \mu^0)$ at (t^0, μ^0). We may assume without loss of generality that $\bar B_r(t^0, \mu^0) \subset (0, T) \times \mathcal{P}_0(G)$. By the definition of v^*, there exists (t^n, μ^n) and $w_n \in \mathcal{S}$ such that

$$(t^n, \mu^n) \to (t^0, \mu^0) \quad \text{and} \quad w_n(t^n, \mu^n) \to v^*(t^0, \mu^0) \quad \text{as } n \to +\infty. \tag{6.2}$$

Set

$$\varphi_\delta(t, \mu) := \varphi(t, \mu) + \delta|t - t^0|^2 + \delta\|\mu - \mu^0\|^2_{\ell_2} \quad \text{on } (0, T) \times \mathcal{P}_0(G).$$

Note that φ_δ is of class $C^1\big((0, T) \times \mathcal{P}_0(G), \ell_2\big)$. Furthermore, (t^0, μ^0) is a strict maximizer for $v^*(t, \mu) - \varphi_\delta(t, \mu)$ on $\bar B_r(t^0, \mu^0)$. For any $n \in \mathbb{N}$, let $(\hat t^n, \hat\mu^n)$ be a maximizer of $w_n - \varphi_\delta$ over $\bar B_r(t^0, \mu^0)$. Observe that

$$w_n(t^n, \mu^n) - \varphi_\delta(t^n, \mu^n) \le w_n(\hat t^n, \hat\mu^n) - \varphi_\delta(\hat t^n, \hat\mu^n) \le v^*(\hat t^n, \hat\mu^n) - \varphi_\delta(\hat t^n, \hat\mu^n).$$

Thus, if (t^∞, w^∞) is a point of accumulation for $\big((\hat t^n, \hat\mu^n)\big)_n$ then by (6.2), we have

$$v^*(t^0, \mu^0) - \varphi_\delta(t^0, \mu^0) = \limsup_{n\to+\infty}(w_n(t^n, \mu^n) - \varphi_\delta(t^n, \mu^n)) \le \limsup_{n\to+\infty}(v^*(\hat t^n, \hat\mu^n) - \varphi_\delta(\hat t^n, \hat\mu^n)).$$

We use the fact that v^* is upper semicontinuous to conclude that

$$v^*(t^0, \mu^0) - \varphi_\delta(t^0, \mu^0) \le v^*(t^\infty, \mu^\infty) - \varphi_\delta(t^\infty, \mu^\infty).$$

Since (t^0,μ^0) is the unique maximizer of $v^*-\varphi_\delta$ over $\bar{B}_r(t^0,\mu^0)$, we conclude that $(t^0,\mu^0)=(t^\infty,w^\infty)$ and so, (t^0,μ^0) is the unique point of accumulation of $\big((\hat{t}^n,\hat{\mu}^n)\big)_n$. Thus, the whole sequence $\big((\hat{t}^n,\hat{\mu}^n)\big)_n$ converges to (t^0,μ^0) and so, for n large enough, $(\hat{t}^n,\hat{\mu}^n)$ belongs to $B_r(t^0,\mu^0)$. Note that

$$\partial_t\varphi_\delta(t,\mu)=\partial_t\varphi(t,\mu)+2\delta(t-t^0)\quad\text{and}\quad\nabla_{\mathcal{W}}\varphi_\delta(t,\mu)=\nabla_{\mathcal{W}}\varphi(t,\mu)+2\delta\nabla_G(\mu-\mu^0).$$

Since $w_n\in\mathcal{S}$ and $(\hat{t}^n,\hat{\mu}^n)$ maximizes $w_n-\varphi_\delta$ over $\bar{B}_r(t^0,\mu^0)$, we obtain that

$$\begin{aligned}&\partial_t\varphi(\hat{t}^n,\hat{\mu}^n)+2\delta(\hat{t}^n-t^0)+\mathcal{H}\big(\hat{\mu}^n,\nabla_{\mathcal{W}}\varphi(\hat{t}^n,\hat{\mu}^n)+2\delta\nabla_G(\hat{\mu}^n-\mu^0)\big)+\mathcal{F}(\hat{\mu}^n)\\ \le&\mathcal{O}_{\hat{\mu}^n}\big(\nabla_{\mathcal{W}}\varphi(\hat{t}^n,\hat{\mu}^n)\big)+2\delta\mathcal{O}_{\hat{\mu}^n}\big(\nabla_{\mathcal{W}}\nabla_G(\hat{\mu}^n-\mu^0)\big).\end{aligned}$$

Observe that since $\mu^0\in\mathcal{P}_0(G)$, $\|\cdot\|_{\hat{\mu}^n}$ and $\|\cdot\|_{\ell_2}$ are equivalent.

Letting $n\to+\infty$ and using the continuity of $\mathcal{F},\mathcal{H},\mathcal{O}_\mu$, and (6.1), we obtain

$$\partial_t\varphi(t^0,\mu^0)+\mathcal{H}(\mu^0,\nabla_{\mathcal{W}}\varphi(t^0,\mu^0))+\mathcal{F}(\mu^0)\le\mathcal{O}_\mu\big(\nabla_{\mathcal{W}}\varphi(t^0,\mu^0)\big).$$

This concludes the proof of the lemma. □

Lemma 6.2. *Suppose that u is a viscosity subsolution to* (4.3) *such that u_* is not a viscosity supersolution to* (4.3). *Then, there exist $(t^0,\mu^0)\in(0,T)\times\mathcal{P}_0(G)$, $\delta,r>0$, such that $B_{2r}(t^0,\mu^0)\subset(0,T)\times\mathcal{P}_0(G)$ and a viscosity subsolution v to* (4.3) *such that the following hold.*

(i) *$v\ge u$ on $[0,T)\times\mathcal{P}_0(G)$ and $v=u$ on $([0,T)\times\mathcal{P}_0(G))\setminus B_r(t^0,\mu^0)$.*
(ii) *There exists a sequence $\big((t^n,\mu^n)\big)_n\subset(0,T)\times\mathcal{P}_0(G)$ such that*

$$(t^n,\mu^n)\to(t^0,\mu^0),\quad u(t^n,\mu^n)\to u_*(t^0,\mu^0),\quad v(t^n,\mu^n)-u(t^n,\mu^n)\to\delta\quad\textit{as } n\to+\infty.$$

Proof. Since u_* is not a supersolution to (4.3), there exists $\varphi\in C^1\big((0,T)\times\mathcal{P}_0(G),\ell_2\big)$, $r>0$ and $(t^0,\mu^0)\in(0,T)\times\mathcal{P}_0(G)$ such that $u_*-\varphi$ attains the minimum value 0 at $(t^0,\mu^0)\in(0,T)\times\mathcal{P}_0(G)$ on $B_{2r}(t^0,\mu^0)\subset(0,T)\times\mathcal{P}_0(G)$ and

$$\partial_t\varphi(t^0,\mu^0)+\mathcal{H}(\mu^0,\nabla_{\mathcal{W}}\varphi(t^0,\mu^0))+\mathcal{F}(\mu^0)<\mathcal{O}_\mu\big(\nabla_{\mathcal{W}}\varphi(t^0,\mu^0)\big).$$

By a continuity argument, if $\delta,\gamma>0$ are sufficiently small, reducing the value of r if necessary, we obtain that

$$(t,\mu)\to\varphi_{\delta,\gamma}(t,\mu):=\varphi(t,\mu)+\delta-\gamma\|\mu-\mu^0\|^2_{\ell_2}-\gamma|t-t^0|^2$$

is a classical subsolution to (4.3) on $B_r(t^0,\mu^0)\subset(0,T)\times\mathcal{P}_0(G)$. Thus, by Remark 4.3, $\varphi_{\delta,\gamma}$ is a viscosity subsolution to (4.3) on $B_r(t^0,\mu^0)$. Observe that

$$u(t,x)\ge u_*(t,x)\ge\varphi(t,x)\quad\text{on}\quad B_r(t^0,\mu^0).$$

If we choose $\delta=\frac{r^2\gamma}{8}$, then

$$u(t,\mu)>\varphi_{\delta,r}(t,\mu)\quad\text{on}\quad B_r(t^0,\mu^0)\setminus\bar{B}_{\frac{r}{2}}(t^0,\mu^0).$$

Setting

$$v(t,\mu)=\begin{cases}\max\{u(t,\mu),\varphi_{\delta,\gamma}(t,\mu)\}, & \text{on}\quad B_r(t^0,\mu^0),\\ u(t,\mu), & \text{otherwise},\end{cases}\tag{6.3}$$

we conclude that $v=u$ on the open set

$$\Omega:=(0,T)\times\mathcal{P}_0(G)\setminus\bar{B}_{\frac{r}{2}}(t^0,\mu^0).$$

Thus, v is a viscosity subsolution to (4.3) on Ω. Since, by Lemma 6.1, $v = \max\{u, \varphi_{\delta,\gamma}\}$ is a viscosity subsolution to (4.3) on $B_r(t^0, \mu^0)$ and since the union of the open sets Ω and $B_r(t^0, \mu^0)$ is $(0, T) \times \mathcal{P}_0(G)$, we conclude that v is a viscosity subsolution to (4.3) on $[0, T) \times \mathcal{P}_0(G)$.

Let $\{(t^n, \mu^n)\}_{n\in\mathbb{N}} \subset (0, T) \times \mathcal{P}_0(G)$ be such that

$$\lim_{n\to+\infty}(t^n, \mu^n) = (t^0, \mu^0) \quad \text{and} \quad \lim_{n\to+\infty} u(t^n, \mu^n) = u_*(t^0, \mu^0).$$

We have

$$\lim_{n\to+\infty}(v(t^n, \mu^n) - u(t^n, \mu^n)) \geq \varphi_{\delta,\gamma}(t^0, \mu^0) - u_*(t^0, \mu^0) = u_*(t^0, \mu^0) + \delta - u_*(t^0, \mu^0) = \delta,$$

which completes the proof of (ii). □

Theorem 6.3 (Perron's Method)**.** *Let the assumptions of Theorem 5.5 be satisfied, let* (6.1) *hold and let $\mathcal{U}_0 \in C(\mathcal{P}_0(G))$. Suppose that $\underline{u}$ is a bounded viscosity subsolution to* (4.3)*, $\bar{u}$ is a bounded viscosity supersolution to* (4.3) *and in addition $\underline{u}_*(0, \mu) = \bar{u}^*(0, \mu) = \mathcal{U}_0(\mu)$ for all $\mu \in \mathcal{P}_0(G)$. Then, setting*

$$\mathcal{S} := \Big\{ w \, : \, \underline{u} \leq w \leq \bar{u} \text{ on } [0, T) \times \mathcal{P}_0(G) \text{ and } w \text{ is a viscosity subsolution to } (4.3) \Big\},$$

the function $u := \sup_{w\in\mathcal{S}} w$ is a viscosity solution to (4.3)*.*

Proof. By Lemma 6.1, u^* is a viscosity subsolution to (4.3). Since $\underline{u} \leq u \leq \bar{u}$, we have $\underline{u} \leq u^* \leq \bar{u}$ and $\mathcal{U}_0(\mu) = \underline{u}_*(0, \mu) \leq u_*(0, \mu) \leq u^*(0, \mu) \leq \bar{u}^*(0, \mu) =: \mathcal{U}_0(\mu)$ and so, $u_*(0, \mu) = u^*(0, \mu) = \mathcal{U}_0(\mu)$ for $\mu \in \mathcal{P}_0(G)$. By the maximality property of u, this implies that $u = u^*$ and so, u is a viscosity subsolution to (4.3). If u_* fails to be a viscosity supersolution to (4.3), let v be the viscosity subsolution to (4.3) provided by Lemma 6.2. Observe that $v(0, \cdot) = \mathcal{U}_0(\cdot)$. By the comparison principle, $v \leq \bar{u}$ on $[0, T) \times \mathcal{P}_0(G)$. Also $\underline{u} \leq u \leq v$ by the construction of v. Hence $v \in \mathcal{S}$ and so, by the maximality property of u, we have $v \leq u$, which contradicts (ii) of Lemma 6.2. Thus, u_* is also a viscosity supersolution to (4.3) and then comparison yields $u^* \leq u_*$. Therefore $u = u^* = u_*$ is a viscosity solution to (4.3). □

In light of Theorems 5.5 and 6.3, to show that (4.3) has a unique viscosity solution, it suffices to construct a viscosity subsolution $\underline{u}$ and a viscosity supersolution $\bar{u}$ to (4.3). We show how this can be done in the rest of this section.

Proposition 6.4. *Let the assumptions of Theorem 5.5 be satisfied (recall that we assume* (6.1) *in this section). Suppose that $\mathcal{U}_0 : \mathcal{P}_0(G) \to \mathbb{R}$ is a function such that one of the following two conditions holds:*

(i) *$\mathcal{U}_0$ is ℓ_2-Lipschitz;*
(ii) *$\mathcal{O} \equiv 0$ and $\mathcal{U}_0$ is $\mathcal{W}$-Lipschitz.*

Then there exists a constant $C_0 > 0$ which depends only on $\mathcal{U}_0, \mathcal{H}, \mathcal{F}$ such that the functions

$$\underline{u}(t, \mu) = -C_0 t + \mathcal{U}_0(\mu), \quad \overline{u}(t, \mu) = C_0 t + \mathcal{U}_0(\mu)$$

are respectively a viscosity subsolution and a viscosity supersolution to (4.3)*. Moreover, if u is a bounded viscosity solution to* (4.3) *then $u(\cdot, \mu)$ is C_0-Lipschitz on $[0, T)$ for every $\mu \in \mathcal{P}_0(G)$ and for every $\varepsilon > 0$ there is a constant K_ε such that*

$$|u(t, \mu) - u(t, \nu)| \leq K_\varepsilon \|\mu - \nu\|_{\ell_2} \quad \text{for all } t \in [0, T], \mu, \nu \in \mathcal{P}_\varepsilon(G). \tag{6.4}$$

Proof. In the case (i), we assume l_0 is the ℓ_2–Lipschitz constant of $\mathcal{U}_0$. We fix $C_0 > C > 0$ whose value will be specified later and set $\underline{u}(t,\mu) \equiv -C_0 t + \mathcal{U}_0(\mu)$. Let $\varphi \in C^1\big((0,T)\times\mathcal{P}_0(G),\ell_2\big)$ be such that there are $r>0$ and (t^0,ρ^0) such that $\bar{B}_r(t^0,\rho^0) \subset (0,T)\times\mathcal{P}_0(G)$ and $\underline{u}-\varphi$ achieves its maximum on $\bar{B}_r(t^0,\rho^0)$ at (t^0,ρ^0). Note that $\partial_t\varphi(t^0,\rho^0) = -C_0$ and $\big\|\frac{\delta\varphi}{\delta\mu}(t^0,\mu^0)\big\|_{\ell_2} \le l_0$ and so,

$$\|\nabla_{\mathcal{W}}\varphi(t^0,\mu^0)\|_{\mu^0} \le 2n^2 l_0 C_\omega.$$

Set

$$C := C_{\mathcal{O}} l_0 + \sup_{(\mu,p)}\Big\{\big|\mathcal{H}(\mu,p)+\mathcal{F}(\mu)\big| : \mu\in\mathcal{P}_0(G), p\in\mathbb{S}^{n\times n}, \|p\|_\mu \le 2n^2 l_0 C_\omega\Big\}.$$

We have

$$\partial_t\varphi(t^0,\rho^0) + \mathcal{H}(\rho_0,\nabla_{\mathcal{W}}\varphi(t^0,\rho_0)) + \mathcal{F}(\rho^0) - \mathcal{O}_{\rho^0}\big(\nabla_{\mathcal{W}} u(t^0,\rho^0)\big) \le -C_0 + C.$$

This proves that $\underline{u}$ is a viscosity subsolution to (4.3) such that $\underline{u}(0,\cdot)=\mathcal{U}_0$. In a similar manner, we construct a viscosity supersolution $\bar{u}$ to (4.3), which is such that $\bar{u}(0,\cdot)=\mathcal{U}_0$. We apply Theorems 5.5 and 6.3 to conclude the proof in case (i).

In the case (ii), one shows that if $\underline{u}-\varphi$ achieves a local maximum at $(t^0,\rho^0)\in(0,T)\times\mathcal{P}_0(G)$, then $\|\nabla_{\mathcal{W}}\varphi(t^0,\mu^0)\|_{\mu^0} \le n l_0 C$. We follow the same lines of arguments to conclude the proof in the case (ii) when $C_{\mathcal{O}}=0$.

To show Lipschitz continuity in t, we notice that comparison principle gives us

$$-C_0 t + \mathcal{U}_0(\mu) \le u(t,\mu) \le C_0 t + \mathcal{U}_0(\mu) = C_0 t + \mathcal{U}_0(\mu) \tag{6.5}$$

for any $t\in[0,T)$ and $\mu\in\mathcal{P}_0(G)$. Let $s>0$ and define $v(t,\mu)=u(t+s,\mu)$. Since $\mathcal{H}$ is time independent, v is a viscosity solution to (4.3) such that $v(0,\cdot)=u(s,\cdot)$. We have

$$v(0,\cdot) - \|v(0,\cdot)-u(0,\cdot)\|_\infty \le u(0,\cdot) \le v(0,\cdot) + \|v(0,\cdot)-u(0,\cdot)\|_\infty.$$

By the comparison principle,

$$v(t,\cdot) - \|v(0,\cdot)-u(0,\cdot)\|_\infty \le u(t,\cdot) \le v(t,\cdot) + \|v(0,\cdot)-u(0,\cdot)\|_\infty \quad \text{on } (0,T-s)\times\mathcal{P}_0(G).$$

Thanks to (6.5), we conclude that

$$-C_0 s \le -\|u(s,\cdot)-u(0,\cdot)\|_\infty \le u(t+s,\cdot)-u(t,\cdot) \le \|u(s,\cdot)-u(0,\cdot)\|_\infty \le C_0 s \quad \text{on } (0,T-s)\times\mathcal{P}_0(G).$$

Thus, $u(\cdot,\mu)$ is C_0-Lipschitz on $[0,T)$ for $\mu\in\mathcal{P}_0(G)$.

To prove (6.4), for every $\delta>0$ we define the sup-convolution of u in the μ variable by

$$u^\delta(t,\mu) = \sup_{\rho\in\mathcal{P}_0(G)}\left\{u(t,\rho) - \frac{\|\mu-\rho\|_{\ell_2}^2}{2\delta}\right\}.$$

Let $\bar\rho$ be a maximizing point. It is easy to see that we must have

$$\|\mu-\bar\rho\|_{\ell_2} \le 2\sqrt{\|u\|_\infty\delta} =: C_\delta.$$

Let now $0<t<T, \mu\in\mathcal{P}_{C_\delta}(G)$. Then $\bar\rho\in\mathcal{P}_0(G)$. Suppose $u^\delta-\varphi$ has a maximum at (t,μ). Then

$$u(t,\bar\rho) - \frac{\|\mu-\bar\rho\|_{\ell_2}^2}{2\delta} - \varphi(t,\mu) \ge u(s,\rho) - \frac{\|\nu-\rho\|_{\ell_2}^2}{2\delta} - \varphi(s,\nu) \tag{6.6}$$

for all s,ν,ρ. If we set $\nu=\rho+(\mu-\bar\rho)$ we thus have

$$u(t,\bar\rho)-\varphi(t,\mu) \ge u(s,\rho) - \varphi(s,\rho+(\mu-\bar\rho))$$

so $u-\varphi(\cdot,\cdot+(\mu-\bar\rho))$ has a maximum at $(t,\bar\rho)$. Thus, using the definition of viscosity subsolution,

$$\partial_t\varphi(t,\mu) + \mathcal{H}(\bar\rho, \nabla_{\mathcal{W}}\varphi(t,\mu)) + \mathcal{F}(\bar\rho) \le \mathcal{O}_{\bar\rho}\big(\nabla_{\mathcal{W}}\varphi(t,\mu)\big) \le C_{\mathcal{O}}\|\nabla_{\mathcal{W}}\varphi(t,\mu)\|_{\ell_2}. \tag{6.7}$$

Assume in the sequel that $\mu \in \mathcal{P}_\varepsilon(G)$ and δ is sufficiently small so that $C_\delta < \frac{\varepsilon}{2}$. Since $u(\cdot,\mu)$ is C_0-Lipschitz, $|\partial_t\varphi(t,\mu)| \le C_0$. We use in (6.7), (A-iii) and the fact that by (H-iii) $\|\cdot\|_{\bar\rho} \ge \sqrt{\varepsilon}\|\cdot\|_{\ell_2}$ on $\mathcal{P}_\varepsilon(G)$, to deduce that

$$\theta_{\frac{\varepsilon}{2}}\varepsilon^{\frac{\kappa}{2}}\|\nabla_{\mathcal{W}}\varphi(t,\mu))\|_{\ell_2}^{\kappa} \le C_{\mathcal{O}}\|\nabla_{\mathcal{W}}\varphi(t,\mu)\|_{\ell_2} + C_0 + \mathcal{F}_\infty,$$

where $|\mathcal{F}| \le \mathcal{F}_\infty$. Thus, some constant K_ε independent of δ we have

$$\|\nabla_{\mathcal{W}}\varphi(t,\mu)\|_{\ell_2} \le K_\varepsilon. \tag{6.8}$$

Setting $s = t, \rho = \bar\rho$ in (6.6) we also see that the function

$$\nu \to -\frac{\|\nu - \bar\rho\|_{\ell_2}^2}{2\delta} - \varphi(t,\nu)$$

has a maximum at μ so

$$\frac{\delta\varphi}{\delta\rho}(t,\mu) = \frac{\bar\rho - \mu}{\delta}. \tag{6.9}$$

Since G is connected $\nabla_G p = 0$ if and only if $p_i = p_j = 0$ for all i, j and thus, on the set of null average p, $\|\nabla_G p\|_{\ell_2}$ and $\|p\|_{\ell_2}$ are two equivalent norms. Hence, since $\nabla_{\mathcal{W}}\varphi(t,\mu) = \nabla_G(\frac{\delta\varphi}{\delta\rho})(t,\mu)$, there is a constant C such that

$$\Big\|\frac{\delta\varphi}{\delta\rho}(t,\mu)\Big\|_{\ell_2} \le C_\varepsilon\|\nabla_{\mathcal{W}}\varphi(t,\mu)\|_{\ell_2}.$$

Thus, (6.8) and (6.9) imply

$$\|\frac{\bar\rho - \mu}{\delta}\|_{\ell_2} \le K_\varepsilon \tag{6.10}$$

for some constant K_ε.

The set of points (t,μ) such that $u^\delta - \varphi$ has a maximum at (t,μ) for a smooth function φ is dense in $(0,T) \times \mathcal{P}_0(G)$ (where in $\mathcal{P}_0(G)$ we take the $\|\cdot\|_{\ell_2}$ norm). This can be seen by considering for every $(t_0,\mu_0) \in (0,T) \times \mathcal{P}_0(G), n = 1, 2, ...$, the functions

$$u^\delta(t,\mu) - n((t-t_0)^2 + \|\mu - \mu_0\|_{\ell_2}^2)$$

which, for large n, will have maxima close to (t_0,μ_0). We thus conclude from (6.10) that for every $(t,\mu) \in (0,T) \times \mathcal{P}_\varepsilon(G)$ there is a sequence (t_n,μ_n) such that if $\bar\rho_n$ is the maximizing point for $u^\delta(t_n,\mu_n)$, then

$$\Big\|\frac{\bar\rho_n - \mu_n}{\delta}\Big\|_{\ell_2} \le K_\varepsilon.$$

Thus, by passing to a subsequence, we obtain that for every $(t,\mu) \in (0,T) \times \mathcal{P}_\varepsilon(G)$, there exists a maximizing point $\bar\rho$ for $u^\delta(t,\mu)$ such that (6.10) holds.

Let now $t \in (0,T), \mu, \nu \in \mathcal{P}_\varepsilon(G)$. We define the function

$$\psi_\delta(s) = u^\delta(t, \mu + s(\nu - \mu)), \qquad \forall s \in [0,1].$$

The function ψ_δ is Lipschitz and hence differentiable a.e. Let $0<\bar{s}<1$ be a point of differentiability of ψ_δ and let $h\in C^1(\mathbb{R})$ be a function such that $\psi_\delta - h$ has a maximum at $\bar{s}$. Let $\bar{\rho}$ be a maximizing point for $u^\delta(t,\mu+s(\nu-\mu))$ satisfying (6.10). Then the function

$$s\to u(t,\bar{\rho})-\frac{\|\mu+s(\nu-\mu)-\bar{\rho}\|_{\ell_2}^2}{2\delta}-h(s)$$

has a maximum at $\bar{s}$. Therefore

$$h'(\bar{s})=\left(\frac{\bar{\rho}-(\mu+s(\nu-\mu))}{\delta},\nu-\mu\right)$$

and thus $|h'(\bar{s})|\le K_\varepsilon\|\nu-\mu\|_{\ell_2}$. We now conclude that

$$|u^\delta(t,\nu)-u^\delta(t,\mu)|=|\psi_\delta(1)-\psi_\delta(0)|\le K_\varepsilon\|\nu-\mu\|_{\ell_2}.$$

It remains to send $\delta\to 0$. □

If $\mathcal{U}_0\in C(\mathcal{P}(G))$ (and hence $\mathcal{U}_0$ is uniformly continuous), let u_0^δ for $0<\delta<1$ be the sup-convolution of $\mathcal{U}_0$ defined as in the proof of Proposition 6.4. Then u_0^δ is ℓ_2-Lipschitz and $\mathcal{U}_0\le u_0^\delta\le \mathcal{U}_0+a_\delta$, where $a_\delta\to 0$ as $\delta\to 0$. Therefore for every $0<\delta<1$ there is a constant $C_\delta>0$ such that

$$\overline{u}_\delta(t,\mu):=C_\delta t+u_0^\delta(\mu)$$

is a viscosity supersolution to (4.3). Then the function

$$\overline{u}:=\inf_{0<\delta<1}\overline{u}_\delta$$

is a bounded continuous viscosity supersolution to (4.3) such that $\overline{u}(0,\mu)=\mathcal{U}_0(\mu)$ for all $\mu\in\mathcal{P}_0(G)$. We can construct a bounded continuous viscosity subsolution $\underline{u}$ in the same way by approximating $\mathcal{U}_0$ by its inf-convolutions.

7. Optimal control problem

In this section we apply our results to a model optimal control problem and show that the value function is a unique viscosity solution of the associated Hamilton-Jacobi equation. The Hamiltonian for our model problem is of the type from Example 5.1 and $\mathcal{O}_\mu=0$. Throughout this section we assume that

$$\mathcal{U}_0,\ \mathcal{F}\in C\big(\mathcal{P}(G)\big),$$

and $c>0$ is such that $|\mathcal{U}_0|,|\mathcal{F}|\le c$.

We define the function $\bar{\mathcal{L}}:\mathcal{P}(G)\times\mathbb{S}^{n\times n}\to[0,+\infty]$ by

$$\bar{\mathcal{L}}(\mu,m)=\begin{cases}0, & \text{if } \mu\in\partial\mathcal{P}(G),\ m=0;\\ +\infty, & \text{if } \mu\in\partial\mathcal{P}(G),\ m\neq 0;\\ \frac{1}{2a(\mu)}\sum_{(i,j)\in E}\frac{m_{ij}^2}{g_{ij}(\mu)}, & \text{if } \mu\in\mathcal{P}_0(G),\end{cases}\tag{7.1}$$

where

$$a(\mu):=\frac{1}{\mathcal{I}^2(\mu)}.$$

It is easy to see that if $\mu\in\mathcal{P}_0(G)$ then

$$\sup_{m\in\mathbb{S}^{n\times n}}\Big\{(p,m)-\bar{\mathcal{L}}(\mu,m)\Big\}=\frac{1}{2}a(\mu)\|p\|_\mu^2=:\bar{\mathcal{H}}(\mu,p),\qquad \forall p\in\mathbb{S}^{n\times n}.$$

Moreover, if $\mu \in \partial\mathcal{P}(G)$

$$\sup_{m\in\mathbb{S}^{n\times n}} \Big\{(p,m) - \bar{\mathcal{L}}(\mu,m)\Big\} = \sup_{m\in\mathbb{S}^{n\times n}} (p,m) = \begin{cases} 0, & \text{if } p=0; \\ +\infty, & \text{if } p\neq 0. \end{cases}$$

Recall that, given $\rho^0,\rho^1 \in \mathcal{P}(G)$, we denote by $\mathcal{C}_0^t(\rho^0,\rho^1)$ the set of pairs (σ,m) such that

$$\sigma \in H^1\big(0,t;\mathcal{P}(G)\big),\ m \in L^2\big(0,t;\mathbb{S}^{n\times n}\big), \quad (\sigma(0),\sigma(t)) = (\rho^0,\rho^1)$$

and

$$\dot\sigma + \mathrm{div}_G(m) = 0, \quad \text{in the weak sense on} \quad (0,t).$$

Given $\rho \in \mathcal{P}(G)$, we define $\mathcal{C}_0^t(\cdot,\rho)$ to be the union of all $\mathcal{C}_0^t(\rho^0,\rho)$ such that $\rho^0 \in \mathcal{P}(G)$, and similarly we define $\mathcal{C}_s^t(\cdot,\rho)$ for $0<s<t$.

Lemma 7.1. *Let $\rho \in \mathcal{P}(G)$ and fix $i \in \{1,\cdots,n\}$. Suppose that $(\sigma,m) \in \mathcal{C}_0^T(\cdot,\rho)$ is such that $\bar{\mathcal{L}}(\sigma,m) \in L^1(0,T)$. Then there exists a positive constant C independent of σ such that the following hold.*

(i) *We have $\|m\|^2_{L^2(0,T)} \le 2n^2\|\bar{\mathcal{L}}(\sigma,m)\|_{L^1(0,T)}$ and $\|\dot\sigma\|_{L^2(0,T)} \le \sqrt{C}\|m\|_{L^2(0,T)}$.*

(ii) *If there exist $t_0,t_1 \in [0,T]$ such that $t_0<t_1$ and $\sigma_i([t_0,t_1]) \subset (0,+\infty)$ then*

$$2C(t_1-t_0)\int_{t_0}^{t_1} \bar{\mathcal{L}}(\sigma,m)ds \ge \Big(\log\big(\sigma_i(t_1)\big) - \log\big(\sigma_i(t_0)\big)\Big)^2.$$

(iii) *Either $\sigma_{i_0}([0,T]) \subset (0,+\infty)$ or $\sigma_{i_0}([0,T]) = \{0\}$.*

Proof. **1.** We use the fact that $a \le n^2$ and $g_{ij} \le 1$ to obtain

$$\bar{\mathcal{L}}(\sigma,m) \ge \frac{1}{2n^2}\|m\|^2. \tag{7.2}$$

Furthermore, the identity

$$\dot\sigma_i + \sum_{j\in N(i)} \sqrt{\omega_{ij}}\, m_{ij} = 0, \tag{7.3}$$

implies that for some positive constant C independent of σ, we have

$$|\dot\sigma_i|^2 \le C\sum_{j\in N(i)} m_{ij}^2. \tag{7.4}$$

This concludes the proof of (i).

2. Suppose that $t_0,t_1 \in [0,T]$ are such that $t_0<t_1$ and $\sigma_i([t_0,t_1]) \subset (0,+\infty)$. Then

$$\bar{\mathcal{L}}(\sigma,m) = \frac12\left(\sum_{k\in V}\frac{1}{\sigma_k}\right)^2 \sum_{(k,j)\in E}\frac{m_{kj}^2}{g_{jk}(\sigma)} \ge \frac12\left(\sum_{k\in V}\frac{1}{\sigma_k}\right)^2 \sum_{j\in N(i)}\frac{m_{ij}^2}{g_{ij}(\sigma)} \ge \frac{1}{2C}\frac{\dot\sigma_i^2}{\sigma_i^2}. \tag{7.5}$$

Thanks to Hölder's inequality, we have

$$(t_1-t_0)\int_{t_0}^{t_1}\frac{\dot\sigma_i^2}{\sigma_i^2}ds \ge \left(\int_{t_0}^{t_1}\frac{|\dot\sigma_i|}{\sigma_i}ds\right)^2 \ge \left(\int_{t_0}^{t_1}\frac{\dot\sigma_i}{\sigma_i}ds\right)^2 = \Big(\log\big(\sigma_i(t_1)\big) - \log\big(\sigma_i(t_0)\big)\Big)^2. \tag{7.6}$$

Combining (7.5) and (7.6) we thus obtain (ii).

3. Suppose now that there exists $\bar t \in [0,T]$ such that $\sigma_i(\bar t) = 0$. Assume to the contrary that we can find $\bar s$ such that $\sigma_i(\bar s) > 0$. The open set $\{s \in (0,T):\ \sigma_i(s) > 0\}$ has a connected component I which contains $\bar s$. We have that I is an open interval of the form (a,b) such that either $\sigma_i(a) = 0$ or $\sigma_i(b) = 0$. Suppose for instance that $\sigma_i(b) = 0$. Then by (ii), whenever $0 < r < b - \bar{s}$, we have

$$2CT \int_0^T \bar{\mathcal{L}}(\sigma, m) ds \geq \Big(\log\big(\sigma_i(b-r)\big) - \log\big(\sigma_i(\bar{s})\big) \Big)^2.$$

Letting $r \to 0^+$, we obtain a contradiction. □

If $t \geq 0$ and $(\sigma, m) \in \mathcal{C}_0^t(\cdot, \rho)$ for some $\rho \in \mathcal{P}(G)$, we set

$$\mathcal{A}_0^t(\sigma, m) := \mathcal{U}_0(\sigma(0)) + \int_0^t \big(\bar{\mathcal{L}}(\sigma, m) - \mathcal{F}(\sigma)\big) ds.$$

We define the value function

$$\mathcal{U}(t, \rho) := \inf_{(\sigma,m)} \Big\{ \mathcal{A}_0^t(\sigma, m) : \ (\sigma, m) \in \mathcal{C}_0^t(\cdot, \rho) \Big\}.$$

Setting

$$\sigma(s) := \rho, \quad m(s) := 0, \qquad \forall s \in [0, t],$$

we have $(\sigma, m) \in \mathcal{C}_0^t(\cdot, \rho)$ and so,

$$-c(t+1) \leq \mathcal{U}(t, \rho) \leq t\Big(\bar{\mathcal{L}}(\rho, 0) - \mathcal{F}(\rho)\Big) + \mathcal{U}_0(\rho).$$

Since $\bar{\mathcal{L}}(\rho, 0) = 0$, we conclude that

$$|\mathcal{U}(t, \rho)| \leq (t+1)c. \tag{7.7}$$

Thus, if $(\sigma, m) \in \mathcal{C}_0^t(\cdot, \rho)$ is such that

$$\mathcal{U}_0(\sigma(0)) + \int_0^t \big(\bar{\mathcal{L}}(\sigma, m) - \mathcal{F}(\sigma)\big) ds \leq \mathcal{U}(t, \rho) + 1, \tag{7.8}$$

we have

$$\int_0^t \bar{\mathcal{L}}(\sigma, m) ds \leq \mathcal{U}(t, \rho) + 1 + \int_0^t \mathcal{F}(\sigma) ds - \mathcal{U}_0(\sigma(0)) \leq 2(t+1)c + 1.$$

and so by (7.2), we have

$$\frac{1}{2n^2} \int_0^t \|m\|^2 ds \leq 2(t+1)c + 1. \tag{7.9}$$

Theorem 7.2. *For every $t \in [0, T], \rho \in \mathcal{P}(G)$, there exists $(\sigma^*, m^*) \in \mathcal{C}_0^t(\cdot, \rho)$ such that*

$$\mathcal{U}(t, \rho) = \mathcal{A}_0^t(\sigma^*, m^*).$$

Proof. If $\rho \in \partial\mathcal{P}(G)$, in light of Lemma 7.1, we have that the only pair $(\sigma, m) \in \mathcal{C}_0^t(\cdot, \rho)$ for which $\mathcal{A}_0^t(\sigma, m) < +\infty$ is the trivial pair $\sigma(s) = \rho, m(s) = 0$ for $s \in [0, t]$ so we are done.

Assume in the sequel that $\rho \in \mathcal{P}_0(G)$ and let $(\sigma_k, m_k) \in \mathcal{C}_0^t(\cdot, \rho)$ be such that $\lim_{k\to\infty} \mathcal{A}_0^t(\sigma_k, m_k) = \mathcal{U}(t, \rho)$. We use Lemma 7.1 to conclude that $(\sigma_k)_k$ is bounded in $H^1(0, t; \mathbb{R}^n)$ and $(m_k)_k$ is bounded in $L^2(0, t; \mathbb{R}^{n\times n})$. Passing to a subsequence if necessary, we can assume without loss of generality that there is $(\sigma^*, m^*) \in H^1(0, t; \mathbb{R}^n) \times L^2(0, t; \mathbb{S}^{n\times n})$ such that $\sigma_k \to \sigma^*$ uniformly and $m_k \rightharpoonup m^*$ weakly in $L^2(0, t; \mathbb{S}^{n\times n})$. We use Lemma 7.1 to conclude that the range of each σ_k is contained on $\mathcal{P}_0(G)$. Due to the uniform convergence property of $(\sigma_k)_k$ and the fact that each $\sigma_k([0, t])$ is a compact set we can assume that that there exists $\varepsilon > 0$ such that

$$\sigma_k([0, t]) \subset \mathcal{P}_\varepsilon(G), \qquad \forall k \in \mathbb{N}. \tag{7.10}$$

One checks that $\sigma^*([0,t]) \in \mathcal{P}_{\varepsilon/2}(G)$ and $(\sigma^*, m^*) \in \mathcal{C}_0^t(\cdot, \rho)$. Since (7.10) expresses the fact that the range of σ_k is uniformly aways from $\partial\mathcal{P}(G)$, one uses standard method of the calculus of variations to conclude, since $\bar{\mathcal{L}}$ is continuous on $\mathcal{P}_0(G) \times \mathbb{S}^{n\times n}$ and $\bar{\mathcal{L}}(\mu, \cdot)$ is convex, that

$$\liminf_{k\to+\infty} \int_0^t \bar{\mathcal{L}}(\sigma_k, m_k) ds \geq \int_0^t \bar{\mathcal{L}}(\sigma^*, m^*) ds.$$

Since $(\sigma_k)_k$ converges uniformly and $\mathcal{F}$ and $\mathcal{U}_0$ are continuous, we deduce that

$$\mathcal{A}_0^t(\sigma^*, m^*) \leq \liminf_{k\to+\infty} \mathcal{A}_0^t(\sigma_k, m_k).$$

□

Theorem 7.3. *The value function $\mathcal{U}$ is continuous on $[0,T] \times \mathcal{P}(G)$.*

Proof. Let $t_0 \in [0,T], \rho_0 \in \mathcal{P}(G)$. Let $\{(t_k, \rho_k)\}_{k=1}^{+\infty}$ be an arbitrary sequence in $[0,T] \times \mathcal{P}(G)$ such that $|t_k - t_0| \to 0$ and $\mathcal{W}(\rho_0, \rho_k) \to 0$ as $k \to +\infty$. By Remark 2.2, this is equivalent to $\|\rho_k - \rho_0\|_{\ell_2} \to 0$.

Lower semicontinuity of $\mathcal{U}$. To simplify the argument we assume that $\lim_{k\to+\infty} \mathcal{U}(t_k, \rho_k) = \liminf_{(t,\rho)\to(t_0,\rho_0)} \mathcal{U}(t, \rho)$. We fix $\delta > 0$ and suppose that $t_k \leq t_0 + \delta$ for all $k \in \mathbb{N}$. Let (σ_k^*, m_k^*) be optimal paths for $\mathcal{U}(t_k, \rho_k)$. We consider the extensions to $[0, t_0 + \delta]$ and still use the same notation to denote them, that is we set

$$\sigma_k^*(t) := \begin{cases} \sigma_k^*(t), & t \in [0, t_k]; \\ \rho_k, & t \in [t_k, t_0 + \delta], \end{cases} \qquad m_k^*(t) := \begin{cases} m_k^*(t), & t \in [0, t_k]; \\ 0, & t \in [t_k, t_0 + \delta], \end{cases} \tag{7.11}$$

By Lemma 7.1,

$$\int_0^{t_0+\delta} \|m_k^*\|^2 ds \leq 2n^2 \int_0^{t_0+\delta} \bar{\mathcal{L}}(\sigma_k^*, m_k^*) ds = 2n^2 \mathcal{U}(t_k, \rho_k) + 2n^2 \int_0^{t_0+\delta} \mathcal{F}(\sigma_k^*) ds - \mathcal{U}_0(\sigma_k^*(0))$$

and so, by (7.7), $(m_k)_k$ is bounded in $L^2(0, t_0+\delta; \mathbb{S}^{n\times n})$. As it was done in the proof of Theorem 7.2, we may assume without loss of generality that there is a pair $(\bar{\sigma}, \bar{m})$ such that

$$\sigma_k \to \bar{\sigma} \text{ in } C([0, t_0+\delta]; \mathbb{R}^n), \quad m_k \rightharpoonup \bar{m} \text{ weakly in } L^2(0, t_0+\delta; \mathbb{S}^{n\times n}), \quad \bar{\sigma} \in H^1(0, t_0+\delta; \mathbb{R}^n).$$

We have

$$(\bar{\sigma}, \bar{m}) \equiv (\rho_0, 0) \text{ on } [0, t_0 + \delta] \qquad \text{and} \qquad (\bar{\sigma}, \bar{m}) \in \mathcal{C}_0^{t_0}(\cdot, \rho_0).$$

Note that

$$\lim_{k\to+\infty} \mathcal{U}(t_k, \rho_k) = \lim_{k\to+\infty} \mathcal{A}_0^{t_k}(\sigma_k^*, m_k^*) = \lim_{k\to+\infty} \Big(\mathcal{A}_0^{t_0+\delta}(\sigma_k^*, m_k^*) + (t_0 + \delta - t_k)\mathcal{F}(\rho_k) \Big)$$

Case 1. If $\rho_0 \in \mathcal{P}_0(G)$ we now argue as in the proof of Theorem 7.2 to conclude that

$$\lim_{k\to+\infty} \mathcal{U}(t_k, \rho_k) \geq \mathcal{A}_0^{t_0+\delta}(\bar{\sigma}, \bar{m}) + (t_0 + \delta - t_0)\mathcal{F}(\rho) = \mathcal{A}_0^{t_0}(\bar{\sigma}, \bar{m}) + \delta\mathcal{F}(\rho) \geq \mathcal{U}(t_0, \rho_0) + \delta\mathcal{F}(\rho).$$

We use the fact that $\delta > 0$ is arbitrary to conclude that $\lim_{k\to+\infty} \mathcal{U}(t_k, \rho_k) \geq \mathcal{U}(t_0, \rho_0)$.

Case 2. Suppose $\rho_0 \in \partial\mathcal{P}(G)$ and $(\rho_0)_i = 0$. If $\rho_k \in \partial\mathcal{P}(G)$ then by Lemma 7.1(iii), we have $(\sigma_k^*, m_k^*) \equiv (\rho_k, 0)$. If $\rho_k \in \mathcal{P}_0(G)$, then, again by Lemma 7.1, we must have for every $t \in [0, t_k]$

$$C_1 \geq C_2 \int_0^{t_k} \bar{\mathcal{L}}(\sigma_k^*, m_k^*) ds \geq \Big(\log\big((\rho_k)_i\big) - \log\big((\sigma_k^*)_i(t)\big) \Big)^2$$

for some absolute constants $C_1, C_2 > 0$. This implies that $\max\{(\sigma_k^*)_i(t) : t \in [0, t_k]\} \to 0$ and hence

$$\int_0^{t_k} \|m_k^*\|^2 ds \to 0.$$

We thus conclude that $(\bar{\sigma}, \bar{m}) \equiv (\rho_0, 0)$ on $[0, t_0 + \delta]$. It now easily follows that

$$\lim_{k \to +\infty} \mathcal{U}(t_k, \rho_k) \geq \mathcal{U}_0(\rho_0) - t_0 \mathcal{F}(\rho_0) = \mathcal{U}(t_0, \rho_0).$$

Upper semicontinuity of $\mathcal{U}$. Let us assume now that $\lim_{k \to +\infty} \mathcal{U}(t_k, \rho_k) = \limsup_{(t,\rho) \to (t_0, \rho_0)} \mathcal{U}(t, \rho)$. In the argument below, we distinguish between the case $t_0 = 0$ and the case $t_0 > 0$. Setting

$$m_k \equiv 0, \quad \sigma_k \equiv \rho_k \text{ on } [0, t_k], \tag{7.12}$$

we have $(\sigma_k, m_k) \in \mathcal{C}_0^{t_k}(\cdot, \rho_k)$ and so,

$$\mathcal{U}(t_k, \rho_k) \leq \mathcal{A}_0^{t_k}(\sigma_k, m_k) = -\int_0^{t_k} \mathcal{F}(\rho_k) ds + \mathcal{U}_0(\rho_k), \tag{7.13}$$

When $t_0 = 0$, since $\mathcal{F}$ and $\mathcal{U}_0$ are continuous, (7.13) implies that $\lim_{k \to +\infty} \mathcal{U}(t_k, \rho_k) \leq \mathcal{U}_0(\rho)$. Thus $\mathcal{U}$ is upper semicontinuous at $(0, \rho_0)$. In the sequel, we assume that $t_0 > 0$ and fix an optimal couple (σ^*, m^*) in $\mathcal{U}(t_0, \rho_0)$.

Case 1. Suppose that $\rho_0 \in \partial \mathcal{P}(G)$. Let i be such that $(\rho_0)_i = 0$. Since by (7.7) $\bar{\mathcal{L}}(\sigma^*, m^*) \in L^1(0, t_0)$, we use Lemma 7.1 to conclude that $\sigma_i^*([0, t_0]) = \{0\}$ and so, $\sigma^*([0, t_0]) \subset \partial \mathcal{P}(G)$. Thus $\bar{\mathcal{L}}(\sigma^*, m^*) \equiv 0$ on $(0, t_0)$ and so, $m^* \equiv 0$ on $(0, t_0)$. This proves that $(\sigma^*, m^*) \equiv (\rho_0, 0)$ on $(0, t_0)$. We choose (σ_k, m_k) as in (7.12) and apply (7.13) to conclude that

$$\lim_{k \to +\infty} \mathcal{U}(t_k, \rho_k) \leq -\int_0^{t_0} \mathcal{F}(\rho_0) ds + \mathcal{U}_0(\rho_0) = \mathcal{A}_0^{t_0}(\sigma^*, m^*) = \mathcal{U}(t_0, \rho_0).$$

Case 2. Suppose that $\rho_0 \in \mathcal{P}_0(G)$. By Lemma 7.1, there is $\varepsilon > 0$ such that $\sigma^*([0, t_0]) \subset \mathcal{P}_\varepsilon(G)$. Choose $\delta \in (0, t_0)$ and assume without loss of generality that $-\delta/2 \leq t_k - t_0 \leq \delta$ for all k so that

$$\delta/2 \leq t_k - t_0 + \delta \leq 2\delta, \qquad \forall k \in \mathbb{N}. \tag{7.14}$$

We first integrate (7.3) and use (7.4) and (7.2) to conclude that

$$\|\sigma^*(t_0) - \sigma^*(t_0 - \delta)\|^2 \leq 2C\delta \int_{t_0 - \delta}^{t_0} \|m^*\|^2 ds \leq 4n^2 C\delta \int_{t_0 - \delta}^{t_0} \bar{\mathcal{L}}(\sigma^*, m^*) ds =: 4n^2 C\delta\omega(\delta). \tag{7.15}$$

We define

$$\sigma^k(t) := \begin{cases} \sigma^*(t), & t \in [0, t_0 - \delta]; \\ \left(1 - \frac{t_k - t}{t_k - t_0 + \delta}\right)\rho_k + \frac{t_k - t}{t_k - t_0 + \delta}\sigma^*(t_0 - \delta), & t \in [t_0 - \delta, t_k]. \end{cases} \tag{7.16}$$

We note that $\sigma^k([0, t_k]) \subset \mathcal{P}_\varepsilon(G)$ and

$$\dot{\sigma}^k = \frac{\rho_k - \sigma^*(t_0 - \delta)}{t_k - t_0 + \delta}, \qquad \text{on} \quad (t_0 - \delta, t_k).$$

We use Remark 3.6 to find $\phi \in \mathbb{R}^n$ such that

$$\dot{\sigma}^k + \mathrm{div}_\rho(\nabla_G \phi) = 0 \qquad \text{and} \qquad \|\nabla_G \phi\|_{\rho_0}^2 \leq \|\dot{\sigma}^k\|_{\ell_1}^2 \frac{2n\bar{\lambda}_\omega}{\varepsilon}.$$

Setting

$$m_{ij}^k = g_{ij}(\rho_0)\big(\nabla_G \phi\big)_{ij},$$

we conclude that $(\sigma^k, m^k) \in \mathcal{C}(\cdot, \rho_k)$ and

$$\frac{1}{2}\sum_{(i,j)\in E}\frac{(m_{ij}^k)^2}{g_{ij}(\rho_0)} \le \frac{2\bar{\lambda}_\omega n}{\varepsilon}\|\dot{\sigma}^k\|_{\ell_1}^2 = \frac{2\bar{\lambda}_\omega n}{\varepsilon}\frac{\|\rho_k - \sigma^*(t_0-\delta)\|_{\ell_1}^2}{(t_k - t_0 + \delta)^2}.$$

Since $g_{ij}(\rho_0) \le 1$, we infer

$$\begin{aligned}\|m^k\|^2 \le& \frac{4\bar{\lambda}_\omega n}{\varepsilon}\frac{\|\rho_k - \sigma^*(t_0)\|_{\ell_1}^2 + \|\sigma^*(t_0) - \sigma^*(t_0-\delta)\|_{\ell_1}^2}{(t_k - t_0 + \delta)^2}\\ \le& \frac{4\bar{\lambda}_\omega n^2}{\varepsilon}\frac{\|\rho_k - \sigma^*(t_0)\|_{\ell_2}^2 + \|\sigma^*(t_0) - \sigma^*(t_0-\delta)\|_{\ell_2}^2}{(t_k - t_0 + \delta)^2}.\end{aligned}$$

This, together with (7.14) and (7.15) implies

$$\|m^k\|^2 \le \frac{16\bar{\lambda}_\omega n^2}{\varepsilon\delta^2}\Big(\|\rho_k - \sigma^*(t_0)\|_{\ell_2}^2 + 4n^2 C\delta\omega(\delta)\Big). \tag{7.17}$$

Since $(\sigma^k, m^k) \in \mathcal{C}_0^{t_k}(\cdot, \rho_k)$, we have

$$\mathcal{U}(t_k, \rho_k) \le \mathcal{A}_0^{t_k}(\sigma^k, m^k) = \mathcal{A}_0^{t_0-\delta}(\sigma^*, m^*) + \int_{t_0-\delta}^{t_k}\big(\bar{\mathcal{L}}(\sigma^k, m^k) - \mathcal{F}(\sigma^k)\big)ds$$

We use the fact that $\sigma^k([0, t_k]) \subset \mathcal{P}_\varepsilon(G)$ to infer $g_{ij}(\sigma^k) \ge \varepsilon$ and $a(\sigma^k) \ge \varepsilon^2/n^2$ and so,

$$\bar{\mathcal{L}}(\sigma^k, m^k) \le \frac{n^2}{\varepsilon^3}\|m^k\|^2$$

Since m^k is a constant on $[t_0-\delta, t_k]$ and $|\mathcal{F}| \le c$, we have

$$\mathcal{U}(t_k, \rho_k) \le \mathcal{A}_0^{t_0-\delta}(\sigma^*, m^*) + (t_k - t_0)\Big(\frac{n^2}{\varepsilon^3}\|m^k\|^2 + c\Big).$$

We now use (7.17) and the fact that $|t_k - t_0| \le \delta$ to obtain

$$\mathcal{U}(t_k, \rho_k) \le \mathcal{A}_0^{t_0-\delta}(\sigma^*, m^*) + c\delta + \frac{16\bar{\lambda}_\omega n^4}{\varepsilon^4}\left(\frac{\|\rho_k - \rho_0\|_{\ell_2}^2}{\delta} + 4n^2 C\omega(\delta)\right).$$

We first let $k \to +\infty$ and then $\delta \to 0^+$ to infer

$$\varlimsup_{k\to+\infty}\mathcal{U}(t_k, \rho_k) \le \mathcal{A}_0^{t_0}(\sigma^*, m^*) = \mathcal{U}(t_0, \rho_0).$$

Thus, $\mathcal{U}$ is upper semicontinuous at (t_0, ρ_0). □

Theorem 7.4. *$\mathcal{U}(t,\rho)$ satisfies the Dynamic Programming Principle (DPP), i.e. for any $(t_0, \rho_0) \in [0,T] \times \mathcal{P}(G)$ and $t \in (0, t_0]$*

$$\mathcal{U}(t_0, \rho_0) = \inf_{(\sigma,m)}\left\{\int_t^{t_0}\Big(\bar{\mathcal{L}}(\sigma(s), m(s)) - \mathcal{F}(\sigma(s))\Big)ds + \mathcal{U}(t, \sigma(t)) : (\sigma, m) \in \mathcal{C}_t^{t_0}(\cdot, \rho_0)\right\}. \tag{7.18}$$

Theorem 7.5. *$\mathcal{U}$ is the unique bounded viscosity solution to* (4.3)*.*

Proof. The uniqueness part follows directly from Theorem 5.5. We only need to show that $\mathcal{U}$ is a viscosity solution to (4.3). It is obvious that $\mathcal{U}(0,\mu)=\mathcal{U}_0(\mu)$.

Viscosity subsolution. Let $\varphi \in C^1\big((0,T)\times \mathcal{P}_0(G),\ell_2\big)$ be such that $u-\varphi$ has a local maximum at $(t_0,\rho_0)\in(0,T)\times\mathcal{P}_0(G)$. Let $\psi\in\mathbb{R}^n$. We denote $v=\nabla_G\psi$. Since $\rho_0\in\mathcal{P}_0(G)$, there exists a constant $r\in[0,t_0]$ and $\sigma\in C^1\big([t_0-r,t_0];(\mathcal{P}_0(G),\ell_2)\big)$ which solves

$$\dot\sigma(s)+\mathrm{div}_{\sigma(s)}(v)=0,\quad \sigma(t_0)=\rho_0.$$

Thus, for any $t\in[t_0-r,t_0]$, we have by Theorem 7.4

$$\begin{aligned} 0 &\le \frac{\mathcal{U}(t_0,\rho_0)-\varphi(t_0,\rho_0)-\mathcal{U}(t,\sigma(t))+\varphi(t,\sigma(t))}{t_0-t}\\ &\le \frac{1}{t_0-t}\left(\int_t^{t_0}\Big(\bar{\mathcal{L}}(\sigma(s),\mathrm{div}_{\sigma(s)}(v))-\mathcal{F}(\sigma(s))\Big)ds-\varphi(t_0,\rho_0)+\varphi(t,\sigma(t))\right). \end{aligned}\tag{7.19}$$

Letting $t\to t_0^-$ in (7.19) and using Lemma 3.16, we now have

$$0\le \bar{\mathcal{L}}(\rho_0,\mathrm{div}_{\rho_0}(v))-\mathcal{F}(\rho_0)-\Big(\mathrm{div}_{\rho_0}(v),\nabla_{\mathcal{W}}\varphi(t_0,\rho_0)\Big)-\partial_t\varphi(t_0,\rho_0)$$

Therefore, taking the infimum above over all $v=\nabla_G\psi$ and using the fact that $\nabla_{\mathcal{W}}\varphi(t_0,\rho_0)\in T_{\rho_0}\mathcal{P}(G)$, we obtain

$$\begin{aligned} 0 &\le -\partial_t\varphi(t_0,\rho_0)-\mathcal{F}(\rho_0)+\inf\Big\{\bar{\mathcal{L}}(\rho_0,\mathrm{div}_{\rho_0}(v))-\Big(\mathrm{div}_{\rho_0}(v),\nabla_{\mathcal{W}}\varphi(t_0,\rho_0)\Big):v=\nabla_G\psi,\psi\in\mathbb{R}^n\Big\}\\ &= -\partial_t\varphi(t_0,\rho_0)-\mathcal{F}(\rho_0)+\inf\Big\{\bar{\mathcal{L}}(\rho_0,m)-\Big(m,\nabla_{\mathcal{W}}\varphi(t_0,\rho_0)\Big):m\in\mathbb{S}^{n\times n}\Big\}\\ &= -\partial_t\varphi(t_0,\rho_0)-\mathcal{F}(\rho_0)-\bar{\mathcal{H}}(\rho_0,\nabla_{\mathcal{W}}\varphi(t_0,\rho_0)). \end{aligned}$$

Viscosity supersolution. Let $\varphi\in C^1\big((0,T)\times\mathcal{P}_0(G),\ell_2\big)$ be such that $u-\varphi$ has a local minimum at $(t_0,\rho_0)\in(0,T)\times\mathcal{P}_0(G)$. Then, for any sufficiently small $\varepsilon>0$ and $r\in(0,t_0]$, there exists $(\sigma,m)\in\mathcal{C}_{t_0-r}^{t_0}(\cdot,\rho_0)$ such that

$$\begin{aligned} 0 &\ge \mathcal{U}(t_0,\rho_0)-\varphi(t_0,\rho_0)-\mathcal{U}(t_0-r,\sigma(t_0-r))+\varphi(t_0-r,\sigma(t_0-r))\\ &\ge -\varepsilon_0 r+\int_{t_0-r}^{t_0}\Big(\bar{\mathcal{L}}(\sigma(s),m(s))-\mathcal{F}(\sigma(s))\Big)ds-\varphi(t_0,\rho_0)+\varphi(t_0-r,\sigma(t_0-r)). \end{aligned}\tag{7.20}$$

Using Lemma 7.1, we have $\sigma(t)\in\mathcal{P}_0(G)$ for any $t\in[t_0-r,t_0]$. Dividing by r on (7.20), we can get by Lemma 3.16

$$\begin{aligned} \varepsilon &\ge \frac{1}{r}\left(\int_{t_0-r}^{t_0}\Big(\bar{\mathcal{L}}(\sigma(s),m(s))-\mathcal{F}(\sigma(s))\Big)ds-\varphi(t_0,\rho_0)+\varphi(t_0-r,\sigma(t_0-r))\right)\\ &= \frac{1}{r}\int_{t_0-r}^{t_0}\Big(\bar{\mathcal{L}}(\sigma(s),m(s))-\mathcal{F}(\sigma(s))-\partial_t\varphi(s,\sigma(s))-\Big(\nabla_{\mathcal{W}}\varphi(s,\sigma(s)),m(s)\Big)\Big)\,ds\\ &\ge \frac{1}{r}\int_{t_0-r}^{t_0}\Big(-\partial_t\varphi(s,\varphi(s))-\bar{\mathcal{H}}(\sigma(s),\nabla_{\mathcal{W}}\varphi(s,\sigma(s)))-\mathcal{F}(\sigma(s))\Big)ds. \end{aligned}$$

Sending $r\to0^+$ and then $\varepsilon\to0^+$, we obtain

$$\partial_t\varphi(t_0,\rho_0)+\bar{\mathcal{H}}(\rho_0,\nabla_{\mathcal{W}}\varphi(t_0,\rho_0))+\mathcal{F}(\rho_0)\ge0.$$

□

Acknowledgements. The research of WG was supported by NSF grant DMS–21 54 578 and AFOSR grant FA9550-18-1-0502. The research project started when CM was a visiting faculty at UCLA. CM acknowledges generous support of the AFOSR through grant FA9550-18-1-0502, Hong Kong RGC Grant ECS 21302521, Hong Kong RGC Grant GRF 11311422 and Hong Kong RGC Grant GRF 11303223. WG would like to thank Y. Gao, W. Li and J-G. Liu for fruitful discussions.

Data Availability. All data generated or analyzed during this study are included in the main article file.

References

[1] L. Ambrosio and J. Feng, *On a class of first order Hamilton-Jacobi equations in metric spaces*, J. Differential Equations 256 (2014), no. 7, 2194–2245.

[2] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd Ed., Lectures in Mathematics ETH Zürich, Birkhaüser Verlag, Basel, 2008.

[3] D. Azagra, J. Ferrera and B. Sanz, *Viscosity solutions to second order partial differential equations on Riemannian manifolds*, J. Differential Equations 245 (2008), no. 2, 307–336.

[4] V. Barbu and G. Da Prato, *Hamilton-Jacobi equations in Hilbert spaces*, Research Notes in Mathematics, 86. Pitman (Advanced Publishing Program), Boston, MA, 1983.

[5] Z. Badreddine and H. Frankowska, *Solutions to Hamilton-Jacobi equation on a Wasserstein space*, Calc. Var. Partial Differential Equations 61 (2022), no. 9, 41 pp.

[6] E. Bandini, A. Cosso, M. Fuhrman and H. Pham, *Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem*, Stochastic Process. Appl. 129 (2019), no. 2, 674–711.

[7] E. Bayraktar, A. Cecchin, A. Cohen and F. Delarue, *Finite state mean field games with Wright–Fisher common noise*, J. Math. Pures Appl. 147 (2021), 98–162.

[8] E. Bayraktar, A. Cosso and H. Pham, *Randomized dynamic programming principle and Feynman-Kac representation for optimal control of McKean-Vlasov dynamics*, Trans. Amer. Math. Soc., 370 (2018), no. 3, 2115–2160.

[9] E. Bayraktar, I. Ekren and X. Zhang, *Comparison of viscosity solutions for a class of second order PDEs on the Wasserstein space*, ArXiV preprint, 2023, arXiv:2309.05040.

[10] A. Bensoussan, H. Cheung and S.C.P. Yam, *Control in Hilbert space and first-order mean field type problem*, Stochastic analysis, filtering, and stochastic optimization, 1–32, Springer, Cham, 2022.

[11] A. Bensoussan, J. Frehse and S.C.P. Yam, *The master equation in mean field theory*, J. Math. Pures Appl. (9) 103 (2015), no. 6, 1441–1474.

[12] A. Bensoussan, J. Frehse and S.C.P. Yam, *On the interpretation of the Master Equation*, Stochastic Process, Appl. 127 (2017), no. 7, 2093–2137.

[13] A. Bensoussan and S.C.P. Yam, *Control problem on space of random variables and master equation*, ESAIM Control Optim. Calc. Var. 25 (2019), Art. 10, 36 pp.

[14] U. Bessi, *The stochastic value function on metric measure spaces*, Discrete Contin. Dyn. Syst. 37 (2017), no. 4, 1819–1839.

[15] U. Bessi, *Hamilton-Jacobi in metric spaces with a homological term*, Comm. Partial Differential Equations 45 (2020), no. 7, 776–819.

[16] R. Buckdahn, J. Li, S. Peng and C. Rainer, *Mean-field stochastic differential equations and associated PDEs*, Ann. Probab. 45 (2017), no. 2, 824–878.

[17] M. Burzoni, V. Ignazio, M. Reppen and H.M. Soner, *Viscosity solutions for controlled McKean-Vlasov jump diffusions*, SIAM J. Control Optim. 58 (2020), no. 3, 1676–1699.

[18] P. Cardaliaguet, S. Daudin, J. Jackson and P.E. Souganidis, *An algebraic convergence rate for the optimal control of McKean-Vlasov dynamics*, ArXiV preprint, 2023, arXiv:2203.14554.

[19] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions, *The master equation and the convergence problem in mean field games*, Annals of Mathematics Studies, 201. Princeton University Press, Princeton, NJ, (2019). x+212 pp.

[20] P. Cardaliaguet, J. Jackson, N. Mimikos-Stamatopoulos and P.E. Souganidis *Sharp convergence rates for mean field control in the region of strong regularity*, ArXiV preprint, 2023, arXiv:2312.11373.

[21] P. Cardaliaguet and P.E. Souganidis, *Regularity of the value function and quantitative propagation of chaos for mean field control problems*, NoDEA Nonlinear Differential Equations Appl. 30 (2023), no. 2, Paper No. 25, 37 pp.

[22] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications I: Mean field FBSDEs, control, and games*, Probability Theory and Stochastic Modelling, 83. Springer, Cham, 2018.

[23] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications II: Mean field games with common noise and master equations*, Probability Theory and Stochastic Modelling, 84. Springer, Cham, 2018.

[24] A. Cecchin, *Finite state N-agent and mean field control problems*, ESAIM Control Optim. Calc. Var. 27 (2021), Paper No. 31, 33 pp.

[25] S.N. Chow, W. Huang, Y. Li and H. Zhou. *Fokker–Planck Equations for a Free Energy Functional or Markov Process on a Graph*, Arch. Ration. Mech. Anal. 203 (2012), no. 3, 969–1008.

[26] Y.T. Chow and W. Gangbo, *A partial Laplacian as an infinitesimal generator on the Wasserstein space*, J. Differential Equations 267 (2019), no. 10, 6065–6117.

[27] G. Conforti, R.C. Kraaij and D. Tonon, *Hamilton–Jacobi equations for controlled gradient flows: the comparison principle*, J. Funct. Anal. 284 (2023), no. 9, Paper No. 109853, 53 pp.

[28] G. Conforti, R.C. Kraaij and D. Tonon, *Hamilton–Jacobi equations for controlled gradient flows: cylindrical test functions*, ArXiV preprint, 2023, arXiv:2302.06571.

[29] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham and M. Rosestolato, *Master Bellman equation in the Wasserstein space: Uniqueness of viscosity solutions*, Trans. Amer. Math. Soc. 377 (2024), no. 1, 31–83.

[30] A. Cosso and H. Pham, *Zero-sum stochastic differential games of generalized McKean-Vlasov type*, J. Math. Pures Appl. (9) 129 (2019), 180–212.

[31] M.G. Crandall, H. Ishii and P.L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. 27 (1992), 1–67.

[32] M.G. Crandall and P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions. I. Uniqueness of viscosity solutions*, J. Funct. Anal. 62 (1985), no. 3, 379–396.

[33] M.G. Crandall and P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions. II. Existence of viscosity solutions*, J. Funct. Anal. 65 (1986), no. 3, 368–405.

[34] M.G. Crandall and P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions. III.*, J. Funct. Anal. 68 (1986), no. 2, 214–247.

[35] M.G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions. IV. Hamiltonians with unbounded linear terms*, J. Funct. Anal. 90 (1990), no. 2, 237–283.

[36] M.G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions. V. Unbounded linear terms and B-continuous solutions*, J. Funct. Anal. 97 (1991), no. 2, 417–465.

[37] M.G. Crandall and P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions. 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.

[38] J. Cui, S. Liu and H. Zhou,*Wasserstein Hamiltonian flow with common noise on graph*, SIAM J. Appl. Math. 83 (2023), no. 2, 484–509.

[39] G. Da Prato and J. Zabczyk, *Second order partial differential equations in Hilbert spaces*, London Mathematical Society Lecture Note Series, 293. Cambridge University Press, Cambridge, 2002.

[40] S. Daudin, F. Delarue and J. Jackson, *On the Optimal Rate for the Convergence Problem in Mean Field Control*, ArXiV preprint, 2023, arXiv:2305.08423.

[41] S. Daudin, J. Jackson and B. Seeger, *Well-posedness of Hamilton-Jacobi equations in the Wasserstein space: non-convex Hamiltonians and common noise*, ArXiV preprint, 2023, arXiv:2312.02324.

[42] T. Dominguez and J.-C. Mourrat, *Infinite-dimensional Hamilton-Jacobi equations for statistical inference on sparse graphs*, ArXiV preprint, 2022, arXiv:2209.04516.

[43] G. Fabbri, F. Gozzi and A. Święch, *Stochastic Optimal Control in Infinite Dimension: Dynamic Programming and HJB Equations, with a contribution by M. Fuhrman and G. Tessitore*, Probability Theory and Stochastic Modelling 82, Springer, Cham, 2017.

[44] J. Feng and M. Katsoulakis, *A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions*, Arch. Ration. Mech. Anal. 192 (2009), no. 2, 275–310.

[45] J. Feng and T.G. Kurtz, *Large deviations for stochastic processes*, Mathematical Surveys and Monographs, 131. American Mathematical Society, Providence, RI, 2006.

[46] J. Feng, T. Mikami and J. Zimmer, *A Hamilton-Jacobi PDE associated with hydrodynamic fluctuations from a nonlinear diffusion equation*, Comm. Math. Phys. 385 (2021), no. 1, 1–54.

[47] J. Feng and T. Nguyen, *Hamilton-Jacobi equations in space of measures associated with a system of conservation laws*, J. Math. Pures Appl. (9) 97 (2012), no. 4, 318–390.

[48] J. Feng and A. Święch, *Optimal control for a mixed flow of Hamiltonian and gradient type in space of probability measures*, with an appendix by Atanas Stefanov, Trans. Amer. Math. Soc. 365 (2013), no. 8, 3987–4039.

[49] W. Gangbo, W. Li and C. Mou, *Geodesics of minimal length in the set of probability measures on graphs*, ESAIM Control Optim. Calc. Var. 25 (2019), Paper No. 78, 36 pp.

[50] W. Gangbo, S. Mayorga and A. Święch, *Finite dimensional approximations of Hamilton-Jacobi-Bellman equations in spaces of probability measures*, SIAM J. Math. Anal. 53 (2021), no. 2, 1320–1356.

[51] W. Gangbo and A. Mészáros, *Global well-posedness of Master equations for deterministic displacement convex potential mean field games*, Comm. Pure Appl. Math., 75 (12) (2022), 2685–2801.

[52] W. Gangbo, T. Nguyen and A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space*, Methods Appl. Anal. 15 (2008), no. 2, 155–183.

[53] W. Gangbo and A. Święch, *Optimal transport and large number of particles*, Discrete Contin. Dyn. Syst. 34 (2014), no. 4, 1397–1441.

[54] W. Gangbo and A. Święch, *Existence of a solution to an equation arising from the theory of mean field games*, J. Differential Equations 259 (2015), no. 11, 6573–6643.

[55] W. Gangbo and A. Święch, *Metric viscosity solutions of Hamilton-Jacobi equations depending on local slopes*, Calc. Var. Partial Differential Equations 54 (2015), no. 1, 1183–1218.

[56] W. Gangbo and A. Tudorascu, *On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations*, J. Math. Pures Appl. (9) 125 (2019), 119–174.

[57] Y. Gao, W. Li and J-G. Liu *Master equations for finite state mean field games with nonlinear activations*, ArXiV preprint, 2022, arXiv:2212.05675.

[58] M. Germain, H. Pham and X. Warin, *Rate of convergence for particle approximation of PDEs in Wasserstein space*, J. Appl. Probab. 59 (2022), no. 4, 992–1008.

[59] Y. Giga, N. Hamamuki and A. Nakayasu, *Eikonal equations in metric spaces*, Trans. Amer. Math. Soc. 367 (2015), no. 1, 49–66.

[60] O. Hijab, *Partially observed control of Markov processes. III.*, Ann. Probab. 18 (1990), no. 3, 1099–1125.

[61] O. Hijab, *Infinite-dimensional Hamilton-Jacobi equations with large zeroth-order coefficient*, J. Funct. Anal. 97 (1991), no. 2, 311–326.

[62] R. Hynd and H.K. Kim, *Infinite horizon value functions in the Wasserstein spaces*, J. Differential Equations 258 (2015), no. 6, 1933–1966.

[63] R. Hynd and H.K. Kim, *Value functions in the Wasserstein spaces: finite time horizons*, J. Funct. Anal. 269 (2015), no. 4, 968–997.

[64] C. Jimenez, A. Marigonda and M. Quincampoix, *Optimal control of multiagent systems in the Wasserstein space*, Calc. Var. Partial Differential Equations 59 (2020), no. 2, Paper No. 58, 45 pp.

[65] C. Jimenez and M. Quincampoix, *Hamilton Jacobi Isaacs equations for differential games with asymmetric information on probabilistic initial condition*, J. Math. Anal. Appl. 457 (2018), no. 2, 1422–1451.

[66] R.C. Kraaij, *A general convergence result for viscosity solutions of Hamilton-Jacobi equations and non-linear semigroups*, J. Funct. Anal. 282 (2022), no. 5, Paper No. 109346, 55 pp.

[67] R.C. Kraaij, *The exponential resolvent of a Markov process and large deviations for Markov processes via Hamilton-Jacobi equations*, Electron. J. Probab. 25 (2020), Paper No. 134, 39 pp.

[68] X. J. Li, J. M. Yong, *Optimal Control Theory for Infinite-Dimensional Systems.* Systems and Control: Foundations and Applications (Birkhäuser, Boston, 1995).

[69] T. Liggett, *Continuous Time Markov Processes: An Introduction*, Graduate Texts in Mathematics, vol. 113 AMS.

[70] P.L. Lions, *Mean Field Games*, Cours au Collège de France, 2007–2011.

[71] Q. Liu and A. Nakayasu, *Convexity preserving properties for Hamilton-Jacobi equations in geodesic spaces*, Discrete Contin. Dyn. Syst. 39 (2019), no. 1, 157–183.

[72] Q. Liu, N. Shanmugalingam and X. Zhou, *Equivalence of solutions of eikonal equation in metric spaces*, J. Differential Equations 272 (2021), 979–1014.

[73] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal. 261 (2011), no. 8, 2250–2292.

[74] S. Mayorga and A. Święch, *Finite dimensional approximations of Hamilton-Jacobi-Bellman equations for stochastic particle systems with common noise*, to appear in SIAM J. Control Optim.

[75] A. Mielke, *A gradient structure for reaction diffusion systems and for energy-drift-diffusion systems*, Nonlinearity 24 (2011), no. 4, 1329–1346.

[76] C. Mou and J. Zhang, *Wellposedness of second order master equations for mean field games with nonsmooth data*, to appear in Mem. Amer. Math. Soc., arXiv: arXiv:1903.09907.
[77] A. Nakayasu, *Metric viscosity solutions for Hamilton-Jacobi equations of evolution type*, Adv. Math. Sci. Appl. 24 (2014), no. 2, 333–351.
[78] A. Nakayasu, and T. Namba, *Stability properties and large time behavior of viscosity solutions of Hamilton-Jacobi equations on metric spaces*, Nonlinearity 31 (2018), no. 11, 5147–5161.
[79] H. Pham and X. Wei, *Dynamic programming for optimal control of stochastic McKean-Vlasov dynamics*, SIAM J. Control Optim. 55 (2017), no. 2, 1069–1101.
[80] H. Pham and X. Wei, *Bellman equation and viscosity solutions for mean-field stochastic control problem*, ESAIM Control Optim. Calc. Var. 24 (2018), no. 1, 437–461.
[81] C. Wu and J. Zhang, *Viscosity solutions to parabolic master equations and McKean-Vlasov SDEs with closed-loop controls*, Ann. Appl. Probab. 30 (2020), no. 2, 936–986.

Email address: wgangbo@math.ucla.edu

Department of Mathematics, UCLA, Los Angeles, CA 90095, USA

Email address: chencmou@cityu.edu.hk

Department of Mathematics, City University of Hong Kong, Hong Kong SAR, China

Email address: swiech@math.gatech.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA