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.

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 [37]. Results about classical solutions of linear second order PDE can be found in [35] 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 [28, 29, 30, 31, 32, 33]. We refer to [37] for the full account of the theory and further references. Some aspects of the theory for first order equations can also be found in [62].

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 [54]. A Bellman equation in the space of measures was also studied in [55]. 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 [46] 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 [63] which "lifts" the equation from the

Key words and phrases. Hamilton-Jacobi equations in the Wasserstein space, viscosity solutions, existence of solutions, comparison principle, Wasserstein space on graphs, manifold with boundary.

Wasserstein space to an Hilbert space of L^2 random variables and this approach was developed further in [50] (see also [20, 21] for more on the lifting procedure). We refer the readers to [5, 8, 9, 10, 11, 12, 13, 16, 17, 18, 19, 25, 36, 44, 45, 47, 48, 52, 56, 57, 58, 67, 71, 72, 73] for equations related to mean field control and optimal control/variational problems in spaces of probability measures. Equations related to control problems with partial observation were studied in [6] and equations related to differential games were investigated in [26, 59]. 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 [23, 24, 38, 39, 40, 41, 42, 60, 61]. 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, 47, 49, 53, 64, 65, 69, 70]. Uniqueness of appropriately defined viscosity solutions of first order HJB equations in the Wasserstein space was proved in [5, 58]. 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 [25]. 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 [73], 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 [22, 34, 66, 68]. 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 [51] 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 goal in this manuscript is to develop a mathematical setup for such equations, introduce a notion of a viscosity solution and study their well-posedness. Since the set of probability measures on a graph with n vertices is identified with a simplex in $\mathbb{R}^n$, the theory resembles the theory of viscosity solutions in finite dimension. Moreover, it can be seen that it can be recast in terms of viscosity solutions 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 stops here although in our case the manifold (the simplex) is flat. 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 a class of Hamilton-Jacobi equations with a convex and somehow coercive Hamiltonian and an individual noise type term. We prove two comparison results, one for an initial boundary value problem and the second for an initial value problem where the boundary is irrelevant. 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 [43], 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

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

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. [66] and [43]).

We fix $T > 0$ and assume that we are given $\mathcal{F}, \mathcal{G} \in C(\mathcal{P}(G))$ 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 shall keep our focus on the cases where

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

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

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

We list examples of g's satisfying (1.3) in (2.5) and (2.6). However, in most of this work, we do not need to assume that (1.3) holds.

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

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

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

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

Here

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

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. 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).

The above arguments relating (1.7) to the control problem (1.6) will not be pursued in this manuscript. This formalism was 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 $\bar{g}(s,t) \equiv 1$. One can check that the solution in this special case is given by

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

where

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

Here, $N(i) := \{j \in V : \ \omega_{ij} > 0\}$. In fact, for each $t \geq 0$, e^{tA} is a transition matrix and there exists a constant $C > 0$ such that if $\mu_i \geq \epsilon$ for all $i \in V$ then $(e^{tA}\mu)_i \geq C\epsilon t$ for all $i \in V$.

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 3.15, we introduce the so-called individual noise operator. In Section 4 we introduce the definition of viscosity solution and in Section 5 we prove two comparison results. Existence of viscosity solutions and some regularity results are discussed in Section 6. The Appendix (Section 7) contains proofs of some technical results.

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 $0 \le \epsilon < 1$, $\mathcal{P}_\epsilon(G) := \mathcal{P}(G) \cap (\epsilon, 1)^n$ 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 $\mathcal{C}(\rho^0, \rho^1)$ of paths connections probability measures. Given $\rho^0, \rho^1 \in \mathcal{P}(G)$, we denote as $\mathcal{C}(\rho^0, \rho^1)$ the set of pairs (σ, m) such that

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

and

$$\dot{\sigma}_i + \sum_{j\in N(i)} \sqrt{\omega_{ij}} m_{ji} = 0, \quad \text{in the weak sense on} \quad (0,1). \tag{2.1}$$

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 [43], 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

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

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 on $\mathcal{P}(G)$. We define the square 2-Monge-Kantorovich metric between $\rho^0, \rho^1 \in \mathcal{P}(G)$ by

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

Here the infimum is performed over the set of pairs (σ, v) such that $\sigma \in H^1(0,1;\mathbb{R}^n)$ 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 [43]). 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

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

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 [43]. By

Remark 6.5 and Theorem 7.5 [43], 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 $\epsilon > 0$ such that $\rho^0, \rho^1 \in \mathcal{P}_\epsilon(G)$, Theorem 7.3 [43] asserts that $\|\phi\|_{W^{1,1}(0,1)}$ is bounded by a constant which is independent of ρ^0 and ρ^1. Thus,

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

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

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, $\big\|\nabla_G\phi\big\|_\rho = \big\|\nabla_G\psi\big\|_\rho$.*
- (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 [43]), we use (2.8) to conclude. ☐

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

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

Proof. Setting

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

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

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

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),

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

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

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

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

$$\int_0^1 \big\|\nabla_G \psi_\infty\big\|_\sigma^2 dt \leq \|\bar{\rho} - \rho\|_{\ell_1} \sqrt{2\bar{\lambda}_\omega \epsilon^{-1}}\, n \sqrt{\int_0^1 \|\nabla_G \psi_\infty\|_\sigma^2 dt}.$$

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

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

This concludes the proof. □

Remark 3.6. *Let $\epsilon > 0$ and let $\rho \in \mathcal{P}(G)$ be such that $\rho_i \geq \epsilon$ 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*

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

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

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

Hence, if we denote by $|\sigma'|_{\mathcal{W}}$ the $\mathcal{W}$ metric derivative of σ, then $|\sigma'|_{\mathcal{W}} \leq \|v\|_\sigma$ 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*

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

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

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

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 $\epsilon > 0$ there exists $\delta > 0$ such that if $\bar{\rho} \in \mathcal{P}(G)$ and $\bar{v} \in T_\rho\mathcal{P}(G)$ then*

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

(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 for any $\bar{\rho}$ in a small enough neighborhood of ρ in $\mathcal{P}(G)$, $\|\cdot\|_{\bar{\rho}}$ 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 $\mathcal{P}_0(G)$.*

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 3.14 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 $\mathcal{P}_0(G)$.*

Lemma 3.12. *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 $\epsilon > 0$ 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

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

and so,

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

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,

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

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

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

Since $\epsilon > 0$ is arbitrary, we conclude that $|v_{ij} - \tilde v_{ij}| = 0$. ☐

Lemma 3.13. *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 3.14. *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

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

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 $\epsilon > 0$ and choose $\delta \in (0, \delta_0)$ such that

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

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). □

Definition 3.15. *If $u : \mathcal{P}(G) \to \mathbb{R}$ is differentiable at $\rho \in \mathcal{P}_0(G)$, the graph individual noise operator $\triangle_{\mathrm{ind}}$ is defined by*

$$\triangle_{\mathrm{ind}}u(\rho) := \Big(\mathrm{div}_\rho\big(\nabla_{\mathcal{W}}u(\rho)\big), \log\rho\Big). \tag{3.16}$$

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

$$\triangle_{\mathrm{ind}}u(\rho) = -\Big(\nabla_{\mathcal{W}}u(\rho), \nabla_G\log\rho\Big)_\rho. \tag{3.17}$$

Remark 3.16. *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.*

Lemma 3.17. *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*

$$\dot\sigma(t_0) + \mathrm{div}_{\sigma(t_0)}(v(t_0)) = 0, \tag{3.18}$$

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 $\epsilon>0$ 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

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

Hence,

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

which proves the first statement of the lemma, as $\epsilon>0$ is arbitrary. In light of Lemma 3.14, we now conclude that the second statement of the lemma holds. □

Corollary 3.18. *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 3.12, except that now, we can choose $\delta>0$ such that $\sigma:[-\delta,\delta]\to\mathcal{P}_0(G)$. By Lemma 3.17 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 $\frac{\delta\mathcal{F}}{\delta\rho}(\rho)=0$. □

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

In this section we introduce a notion of viscosity solution. We assume that (1.2) hold. 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 3.14 (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

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

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 contained in the restrictive class of the discrete analogue of the so-called "separable Hamiltonians".

Given $u_0:\mathcal{P}(G)\to\mathbb{R}$, we consider the Hamilton-Jacobi equation

$$\partial_t u(t,\mu)+\mathcal{H}\big(\mu,\nabla_{\mathcal{W}}u(t,\mu)\big)+\mathcal{F}(\mu)=\mathcal{O}_\mu\big(\nabla_{\mathcal{W}}u(t,\mu)\big),\qquad u(0,\cdot)=u_0 \tag{4.3}$$

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 $u(0,\cdot)\le u_0$ 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*

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

(ii) *A function $u\in\mathrm{LSC}([0,T)\times\mathcal{P}_0(G))$ is a viscosity supersolution to (4.3) if $u(0,\cdot)\ge u_0$ 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*

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

(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 3.18, 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 $1 \le i < j \le n$, 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 jth 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:*

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

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

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

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

In this section we show comparison principles for viscosity solutions to equation (4.3). We will consider two cases, a boundary value problem for (4.3) and a case when the boundary $\mathcal{P}(G)\backslash\mathcal{P}_0(G)$ is irrelevant.

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

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

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

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

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

(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 $0<\epsilon<1$ a constant C_ϵ such that for every $b_1, b_2 \geq 0$ (if $\mathcal{J}$ is as in (4.1))

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

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 $\epsilon>0$, there exists $\theta_\epsilon>0$ such that $a\geq\theta_\epsilon$ when $\mu\in\mathcal{P}_\epsilon(G)$. Setting $\mathcal{H}(\mu,p):=a(\mu)\|p\|_\mu^\kappa$, we have*

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

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

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

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

$$\frac{\delta\mathcal{I}_j}{\delta\mu}(\mu)=\frac{1}{\mu_j^2}\Big(\frac{1}{n},\cdots,\frac{1}{n},\frac{1}{n}-1,\frac{1}{n},\cdots,\frac{1}{n}\Big)^T, \quad \nabla_G\Big(\frac{\delta\mathcal{I}_j}{\delta\mu}\Big)(\mu)=\begin{cases}0 & \textit{if } k,l\neq j \textit{ or } k=l=j,\\ -\sqrt{\omega_{jl}}\mu_j^{-2} & \textit{if } k=j, l\neq j,\\ \sqrt{\omega_{jk}}\mu_j^{-2} & \textit{if } k\neq j, l=j.\end{cases} \tag{5.3}$$

Hence,

$$
\begin{aligned}
\mathcal{O}_\mu\Big(\nabla_{\mathcal{W}}\mathcal{I}(\mu)\Big) &= \sum_{(j,l)\in E} \omega_{jl} g_{jl}(\mu) \frac{1}{\mu_j^2}\big(\log\mu_j - \log\mu_l\big) \\
&= \sum_{(j,l)\in E, j<l} \omega_{jl} g_{jl}(\mu) \Big(\frac{1}{\mu_j^2} - \frac{1}{\mu_l^2}\Big)\big(\log\mu_j - \log\mu_l\big) \\
&= -\sum_{(j,l)\in E, j<l} \omega_{jl} g_{jl}(\mu) \Big(\frac{\mu_l+\mu_j}{\mu_j^2\mu_l^2}\Big)\big(\log\mu_j - \log\mu_l\big)(\mu_j-\mu_l) \le 0, \qquad (5.4)
\end{aligned}
$$

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*

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

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

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

we conclude that

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

where

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

Hence,

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

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* $\ell'<0$. *We would need to impose an additional condition that* $\lim_{t\to 0^+}\ell(t)=+\infty$, *to use* $\sum_{i\in V}\ell(\mu_i)$ *in place of* $\sum_{i\in V}1/\mu_i$ *in the proof of the comparison principle.*

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 $a,\beta,\epsilon,\delta\in(0,1]$, $\lambda\in(\frac{1}{2},1]$, we define

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

and

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

We set

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

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

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

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

Since for every $\beta, a, \epsilon, \delta \in (0,1]$ and $\frac{1}{2} \le \lambda \le 1$, $M_{a,\epsilon,\delta} \le M_*$ for some constant M_*, it is easy to see (see e.g. [27], Proposition 3.7 for such argument) that

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

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

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

Theorem 5.5 (Comparison Principle, No Boundary Condition)**.** *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 $\Psi_{a,\epsilon,\delta}$. We will use the notation Ψ in place of $\Psi_{a,\epsilon,\delta}$ and to alleviate the notation, we simply denote a maximizer of $\Psi_{a,\epsilon,\delta}$ by $(\bar{t},\bar{s},\bar{\mu},\bar{\nu})$, without displaying the dependence in $\beta, a, \epsilon, \delta$. 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

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

and

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

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

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

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

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

Similarly, since

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

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

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

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,

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

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

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

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 $\omega(a,\epsilon,\delta)$ such that $\lim_{a\to0}\limsup_{\epsilon\to0}\limsup_{\delta\to0}\omega(a,\epsilon,\delta)=0$ and

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

Now (A-ii) and (5.15) imply

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

Similarly,

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

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

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

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

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

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

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

We combine this with (5.17) to conclude that

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

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

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

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

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

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

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

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

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

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

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

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

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 $0<\lambda<1,\beta,\epsilon,\delta>0$ we consider the function

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

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 $\epsilon,\delta>0$, 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)

$$\frac{\beta}{T^2}+\frac{\bar t-\bar s}{\delta}+\gamma(\frac1\lambda)\mathcal H\Big(\bar\mu,\bar p\Big)+\mathcal F(\bar\mu)-\mathcal O_{\bar\mu}(\bar p)\leq(1-\lambda)\mathcal F_\infty,$$

where

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

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

$$-\frac{\beta}{T^2}-\frac{\bar s-\bar t}{\delta}+\mathcal H(\bar\nu,\bar p)+\mathcal F(\bar\nu)-\mathcal O_{\bar\nu}(\bar p)\geq0.$$

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

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

This allows us to conclude as in Step 3 of the proof of Theorem 5.5 by taking $\lim_{\epsilon\to 0}\limsup_{\delta\to 0}$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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) *$v \geq u$ 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*

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

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

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

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

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

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, No Boundary Condition)**.** *Let the assumptions of Theorem 5.5 be satisfied, let* (6.1) *hold and let* $u_0 \in C(\mathcal{P}_0(G))$. *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* $\underline{u}_*(0,\mu) = \bar{u}^*(0,\mu) = u_0(\mu)$ *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 $u_0(\mu) = \underline{u}_*(0,\mu) \leq u_*(0,\mu) \leq u^*(0,\mu) \leq \bar{u}^*(0,\mu) =: u_0(\mu)$ and so, $u_*(0,\mu) = u^*(0,\mu) = u_0(\mu)$ 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 $v(0,\cdot) = u_0(\cdot)$. 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 the same way we can prove Perron's method theorem for boundary value problems.

Theorem 6.4 (Perron's Method, Boundary Condition)**.** *Let the assumptions of Theorem 5.6 be satisfied, let* (6.1) *hold and let* $u_0 \in C(\mathcal{P}_0(G))$, $h \in C([0,T) \times (\mathcal{P}(G) \setminus \mathcal{P}_0(G)))$ *be such that* $h = u_0$ *on* $\{0\} \times (\mathcal{P}(G) \setminus \mathcal{P}_0(G))$. *Suppose that* $\underline{u} \in USC([0,T) \times \mathcal{P}(G))$ *is a bounded viscosity subsolution to* (4.3), $\bar{u} \in LSC([0,T)\times\mathcal{P}(G))$ *is a bounded viscosity supersolution to* (4.3). *Suppose in addition that* $\underline{u}_*(0,\mu) = \bar{u}^*(0,\mu) = u_0(\mu)$ *for all* $\mu \in \mathcal{P}(G)$ *and* $\underline{u}_*(t,\mu) = \bar{u}^*(t,\mu) = h(t,\mu)$ *for all* $(t,\mu) \in (0,T) \times (\mathcal{P}(G) \setminus \mathcal{P}_0(G))$. *Then, setting*

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

the function $u := \sup_{w\in\mathcal{S}} w$ *is a viscosity solution to* (4.3) *such that* $u = h$ *on* $[0,T) \times (\mathcal{P}(G) \setminus \mathcal{P}_0(G))$.

In light of Theorems 5.5, 5.6, 6.3 and 6.4, 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 achieve this goal in the next proposition under the assumptions of Theorem 5.5 for the problem without a boundary condition, but we are unable do so when a boundary value is prescribed, see the comments after the proof of Proposition 6.5.

Proposition 6.5. *Let the assumptions of Theorem 5.5 be satisfied (recall that we assume* (6.1) *in this section). Suppose that* $u_0 : \mathcal{P}_0(G) \to \mathbb{R}$ *is a function such that one of the following two conditions holds:*

(i) u_0 *is* ℓ_2*-Lipschitz;*
(ii) $\mathcal{O} \equiv 0$ *and* u_0 *is* $\mathcal{W}$*-Lipschitz.*

Then there exists a constant $C_0 > 0$ *which depends only on* $u_0, \mathcal{H}, \mathcal{F}$ *such that the functions*

$$\underline{u}(t,\mu) = -C_0 t + u_0(\mu), \quad \overline{u}(t,\mu) = C_0 t + u_0(\mu)$$

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* $\epsilon > 0$ *there is a constant* K_ϵ *such that*

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

Proof. In the case (i), we assume l_0 is the ℓ_2–Lipschitz constant of u_0. We fix $C_0 > C > 0$ whose value will be specified later and set $\underline{u}(t,\mu) \equiv -C_0 t + u_0(\mu)$. 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 $\big\|\frac{\delta\varphi}{\delta\mu}(t^0,\mu^0)\big\|_{\ell_2} \leq l_0$ and so,

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

Set

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

We have

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

This proves that $\underline{u}$ is a viscosity subsolution to (4.3) such that $\underline{u}(0,\cdot) = u_0$. In a similar manner, we construct a viscosity supersolution $\bar{u}$ to (4.3), which is such that $\bar{u}(0,\cdot) = u_0$. 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 $\|\nabla_{\mathcal{W}}\varphi(t^0,\mu^0)\|_{\mu^0} \leq n l_0 C$. 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

$$-C_0 t + u_0(\mu) \leq u(t,\mu) \leq C_0 t + u_0(\mu) = C_0 t + u_0(\mu) \tag{6.5}$$

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

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

By the comparison principle,

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

Thanks to (6.5), we conclude that

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

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

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

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

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

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 $\mu\in\mathcal{P}_\epsilon(G)$ and δ is sufficiently small so that $C_\delta < \frac{\epsilon}{2}$. 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) $\|\cdot\|_{\bar\rho} \ge \sqrt{\epsilon}\|\cdot\|_{\ell_2}$ on $\mathcal{P}_\epsilon(G)$, to deduce that

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

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

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

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

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

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

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

Thus, (6.8) and (6.9) imply

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

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

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

which, for large n, will have maxima close to (t_0, μ_0). We thus conclude from (6.10) that for every $(t, \mu) \in (0, T) \times \mathcal{P}_\epsilon(G)$ 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

$$\left\| \frac{\bar{\rho}_n - \mu_n}{\delta} \right\|_{\ell_2} \leq K_\epsilon.$$

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

Let now $t \in (0, T), \mu, \nu \in \mathcal{P}_\epsilon(G)$. 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 $|h'(\bar{s})| \leq K_\epsilon \|\nu - \mu\|_{\ell_2}$. We now conclude that

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

It remains to send $\delta \to 0$. $\square$

If $u_0 \in C(\mathcal{P}(G))$ (and hence u_0 is uniformly continuous), let u_0^δ for $0 < \delta < 1$ be the sup-convolution of u_0 defined as in the proof of Proposition 6.5. Then u_0^δ is ℓ_2-Lipschitz and $u_0 \leq u_0^\delta \leq u_0 + a_\delta$, 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 $\overline{u}(0, \mu) = u_0(\mu)$ for all $\mu \in \mathcal{P}_0(G)$. We can construct a bounded continuous viscosity subsolution $\underline{u}$ in the same way by approximating u_0 by its inf-convolutions.

Unfortunately in general it does not seem possible to construct viscosity subsolutions $\underline{u}$ and viscosity supersolutions $\overline{u}$ to (4.3) when a boundary condition is prescribed. Even if we assume in (A-iii) that there is $\theta > 0$ such that $\theta \|p\|_\mu^\kappa \leq \mathcal{H}(\mu, p)$ for all $\mu \in \mathcal{P}_0(G)$ (for instance if $\mathcal{H}(\mu, p) \equiv \|p\|_\mu^\kappa$), the Hamiltonian $\mathcal{H}$ may still degenerate near $\partial\mathcal{P}(G)$ since $\|p\|_\mu^\kappa$ may become small even when $\|p\|_{\ell_2}^\kappa$ is large and μ is near $\partial\mathcal{P}(G)$. This prevents typical constructions of supersolution barriers. Moreover, even if the Hamiltonian were $\|\cdot\|_{\ell_2}$-coercive near $\partial\mathcal{P}(G)$, it is not clear how one would produce a viscosity subsolution unless some special compatibility conditions on the data were satisfied. We are also not able to use the individual noise type operator to produce barriers near $\partial\mathcal{P}(G)$.

7. Appendix: Differentiability properties of $\mathcal{W}^2(\rho^*, \cdot)$

Using the same terminology as in [43], we denote by $\gamma_P(\rho)$ the Poincaré constant of $\rho \in \mathcal{P}(G)$. We fix $\rho^* \in \mathcal{P}(G)$ and define

$$\mathcal{F}_0(\rho) = \frac{1}{2}\mathcal{W}^2(\rho^*, \cdot), \qquad \forall \rho \in \mathcal{P}(G).$$

We set

$$H(a,b) = \sup_{\rho \in \mathcal{P}(G)} \Big\{ (a,\rho) + \frac{1}{2}\|b\|_\rho^2 \Big\}, \quad \forall (a,b) \in \mathbb{R}^n \times \mathbb{S}^{n\times n}. \tag{7.1}$$

Given $\lambda \in \mathrm{BV}_{\mathrm{loc}}(0,1;\mathbb{R}^n)$, we denote by $\dot\lambda^{\mathrm{abs}}\mathcal{L}^1$ the absolutely continuous part of distributional derivative $\dot\lambda$ and denote by $\dot\lambda^{\mathrm{sing}}$ its singular part. As in [43], B_* stands for the set of $\lambda \in \mathrm{BV}_{\mathrm{loc}}(0,1;\mathbb{R}^n)$ such that $H\big(\dot\lambda, \nabla_G \lambda\big) = 0$. In convex analysis this means that, for any non-negative Borel regular measure ν such that $-\dot\lambda_i^{\mathrm{sing}} << \nu$ and ν and $\mathcal{L}^1$ are mutually singular, we have

$$H\left(\dot\lambda^{\mathrm{abs}}, \nabla_G\lambda\right) = 0 \quad \mathcal{L}^1 \quad \text{a.e. in } (0,1), \qquad \max_{i=1 \le i \le n}\Big\{\frac{d\dot\lambda_i^{\mathrm{sing}}}{d\nu}\Big\} = 0, \quad \nu \quad \text{a.e. in } (0,1). \tag{7.2}$$

Recall (c.f. Theorem 7.4 [43]) that if $\rho^0, \rho^1 \in \mathcal{P}(G)$ and $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$, then

$$\min_{(\rho,m)}\Big\{\mathcal{A}(\rho,m) \;\Big|\; (\rho,m) \in \mathcal{C}(\rho^0,\rho^1)\Big\} = \max_{\lambda}\Big\{\big(\lambda(1),\rho^1\big) - \big(\lambda(0),\rho^0\big) \;\Big|\; \lambda \in \mathrm{B}_*\Big\}. \tag{7.3}$$

Remark 7.1. *Assume that $\rho^* \in \mathcal{P}(G)$ with $\gamma_P(\rho^*) > 0$ and $\rho \in \mathcal{P}_0(G)$. In [43], we obtained a $\mathcal{W}$-geodesic μ of constant speed such that $\mu(0) = \rho^*$ and $\mu(1) = \rho$. We combine Remark 6.3 and Theorem 7.3 of [43] to conclude that v, the velocity of minimal norm of μ, is such that v is continuous near 1 and $v(t) = \nabla_G\lambda(t)$ for all t near 1, where λ is a maximizer in (7.3). Furthermore, we have $\|v(t)\|_{\mu(t)} = \mathcal{W}(\rho^*,\rho)$ for every $t < 1$ close to 1. By Corollary 3.3, we can assume without loss of generality that $\|\lambda(1)\|_{\ell_\infty} \le \lambda_g(\rho)\|v(1)\|_\rho$.*

Lemma 7.2. *For any $\rho \in \mathcal{P}_0(G)$ there exists $\lambda^\infty \in \mathbb{R}^n$ such that the following hold.*

(i) *Whenever $\bar\rho \in \mathcal{P}(G)$ and its Poincaré constant satisfies $\gamma_P(\bar\rho) > 0$ then*

$$\mathcal{F}_0(\bar\rho) \ge \mathcal{F}_0(\rho) + (\lambda^\infty, \bar\rho - \rho). \tag{7.4}$$

(ii) *$\nabla_G\lambda^\infty$ is a Wasserstein subgradient of $\mathcal{F}_0$ at ρ, $\|\nabla\lambda^\infty\|_\rho \le \mathcal{W}(\rho^*,\rho)$ and $\|\lambda^\infty\|_{\ell_\infty} \le \lambda_g(\rho)\mathcal{W}(\rho^*,\rho)$.*

Proof. By Remark 2.2, there is a sequence $(\rho^{*,k})_k \subset \mathcal{P}_0(G)$ which converges to ρ^* in the $\mathcal{W}$ metric. By (7.3) there exists $\lambda^k \in B_*$ such that

$$\frac{1}{2}\mathcal{W}^2\big(\rho^{*,k},\rho\big) = \big(\lambda^k(1),\rho\big) - \big(\lambda^k(0),\rho^{*,k}\big). \tag{7.5}$$

By Remark 7.1,

$$\|\nabla_G\lambda^k(1)\|_\rho \le \mathcal{W}(\rho^{*,k},\rho) \quad \text{and} \quad \|\lambda^k(1)\|_{\ell_\infty} \le \lambda_g(\rho)\mathcal{W}(\rho^{*,k},\rho). \tag{7.6}$$

If $\bar\rho \in \mathcal{P}(G)$ and $\gamma_P(\bar\rho) > 0$, due to the maximality property of λ^k expressed (7.5), we infer

$$\frac{1}{2}\mathcal{W}^2(\rho^{*,k},\bar\rho) - \frac{1}{2}\mathcal{W}^2(\rho^{*,k},\rho) \ge \big(\lambda^k(1), \bar\rho - \rho\big).$$

Thus, letting $k \to \infty$, we obtain (7.4).

If $\bar{v} \in T_\rho \mathcal{P}(G)$, (7.4) implies

$$\begin{aligned}\mathcal{F}_0(\bar{\rho}) \geq& \mathcal{F}_0(\rho) - \big(\lambda^\infty, \mathrm{div}_\rho(\bar{v})\big) + \big(\lambda^\infty, \bar{\rho} - \rho + \mathrm{div}_\rho(\bar{v})\big) \\ \geq& \mathcal{F}_0(\rho) + \big(\nabla_G \lambda^\infty, \bar{v}\big)_\rho - \|\lambda^\infty\|_{\ell_\infty} \big\|\bar{\rho} - \rho + \mathrm{div}_\rho(\bar{v})\big\|_{\ell_1},\end{aligned}$$

which concludes the proof. ☐

Let $\rho \in \mathcal{P}(G)$ and let $\mu : [0,1] \to \mathbb{R}^n$ be a geodesic of constant speed connecting ρ^* to ρ and assume that the range of μ is entirely contained in $(0,1]^n$. Since the range is a compact set, there exists $\epsilon > 0$ such that the range of μ is contained in $[2\epsilon, 1-2\epsilon]^n$. The geodesic μ is uniquely characterized by the Euler-Lagrange equations

$$\dot{\rho} = \nabla_\phi \mathcal{H}_g(\rho, \phi), \quad \dot{\phi} = -\nabla_\rho \mathcal{H}_g(\rho, \phi), \tag{7.7}$$

where,

$$\mathcal{H}_g(\rho, \phi) = \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} g(\rho_i, \rho_j)(\phi_i - \phi_j)^2.$$

This means

$$\dot{\mu}_i + \sum_{j \in N(i)} \omega_{ij}(\phi_j - \phi_i) g(\mu_i, \mu_j) = 0, \qquad \dot{\phi}_i + \frac{1}{2} \sum_{j \in N(i)} \omega_{ij} \partial_1 g(\mu_i, \mu_j)(\phi_i - \phi_j)^2 = 0, \tag{7.8}$$

along with the boundary conditions $\mu(0) = \rho^*$ and $\mu(1) = \rho$.

We also know that $\nabla_G \phi$ is uniquely determined. Replacing ϕ_i by $\phi_i - 1/n \sum_{j=1}^n \phi_i(1)$, one checks that (7.8) still holds and

$$\sum_{i=1}^n \phi_i(1) = 0. \tag{7.9}$$

Since the second identity in (7.8) depends only on $\nabla_G \phi$ and μ, we conclude that $\dot{\phi}$ is uniquely determined and so, if $\bar{\phi}$ is another solution, we must have that $\bar{\phi} - \phi = c$ is a constant. But $\sum_{i=1}^n (\bar{\phi}_i - \phi_i)(1) = 0$ implies that $\sum_{i=1}^n c_i = 0$, while $\nabla_G(\bar{\phi} - \phi) = 0$ implies that $c_i = c_j$ for all $(i,j) \in E$. Since (G, V, ω) is a connected graph, we obtain that $c_i = c_j$ for all $i, j \in V$ and so, $c_i = 0$ for all $i \in V$. This shows that the curve ϕ is then uniquely determined under the convention (7.9), which we impose in the sequel. We define

$$\mu \equiv \mu[\rho^*, \rho] \quad \text{and} \quad \phi \equiv \phi[\rho^*, \rho].$$

We recall that the velocity v of μ satisfies the identity

$$v(t) = \nabla_G \phi(t), \qquad \|v(t)\|_{\mu(t)} = \mathcal{W}(\mu(0), \mu(1)), \qquad \forall t \in [0,1]. \tag{7.10}$$

Let $\delta > 0$ be sufficiently small such that for any $\bar{\rho} \in B_\delta(\rho) \subset \mathcal{P}_0(G)$ there is a unique geodesic connecting ρ^* to $\bar{\rho}$ and the geodesic is contained in $[\epsilon, 1-\epsilon]^n$. One readily concludes that $(t, \bar{\rho}) \to \phi[\rho^*, \bar{\rho}](t)$ is continuous on $[0,1] \times B_\delta(\rho)$.

Proposition 7.3. *Let $\rho, \rho^* \in \mathcal{P}_0(G)$ and let $\delta > 0$ be small enough such that for all $\bar{\rho} \in B_\delta(\rho) \subset \mathcal{P}_0(G)$ there is a unique geodesic connecting ρ^* to $\bar{\rho}$. We further assume that there exists $\epsilon > 0$ such that these geodesics are contained in $[\epsilon, 1-\epsilon]^n$. Then:*

(i) *$\mathcal{F}_0$ has the Fréchet derivative at each $\bar{\rho} \in B_\delta(\rho)$ and its derivative $\delta \mathcal{F}_0 / \delta \rho(\bar{\rho}) = \phi[\rho^*, \bar{\rho}](1)$ is ℓ_1-continuous in $B_\delta(\rho)$.*

(ii) *$\mathcal{F}_0$ has the Wasserstein gradient at each $\bar{\rho} \in B_\delta(\rho)$ and its gradient $\nabla_{\mathcal{W}} \mathcal{F}_0(\bar{\rho}) = \nabla_G\big(\phi[\rho^*, \bar{\rho}](1)\big)$ is ℓ_1-continuous in $B_\delta(\rho)$.*

Proof. (i) Let $\bar\rho \in B_\delta(\rho)$. For all $s \in [0,1]$, the geodesic of constant speed connecting ρ^* to $\rho + s(\bar\rho - \rho) = \rho + sf$ is also contained in $[\epsilon, 1-\epsilon]^n$. We have

$$\begin{aligned}\mathcal{F}_0(\rho+sf) - \mathcal{F}_0(\rho) &\geq \big(\phi[\rho^*,\rho](1), \rho+sf\big) - \big(\phi[\rho^*,\rho](0), \rho^*\big) - \big(\phi[\rho^*,\rho](1), \rho\big) + \big(\phi[\rho^*,\rho](0), \rho^*\big)\\ &= s\big(\phi[\rho^*,\rho](1), f\big).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathcal{F}_0(\rho+sf) - \mathcal{F}_0(\rho) &\leq \Big(\phi[\rho^*,\rho+sf](1), \rho+sf\Big) - \Big(\phi[\rho^*,\rho+sf](0), \rho^*\Big)\\ &\quad -\Big(\phi[\rho^*,\rho+sf](1), \rho\Big) + \Big(\phi[\rho^*,\rho+sf](0), \rho^*\Big)\\ &= s\Big(\phi[\rho^*,\rho+sf](1), f\Big).\end{aligned}$$

Hence,

$$\begin{aligned}\limsup_{s\to 1^-} \frac{\mathcal{F}_0(\rho+sf) - \mathcal{F}_0(\rho)}{s} &\leq \limsup_{s\to 1^-} \Big(\phi[\rho^*,\rho+sf](1), f\Big)\\ &= \Big(\phi[\rho^*,\rho](1), f\Big) \leq \liminf_{s\to 1^-} \frac{\mathcal{F}_0(\rho+sf) - \mathcal{F}_0(\rho)}{s}.\end{aligned}$$

This shows that

$$\lim_{s\to 1^-} \frac{\mathcal{F}_0(\rho+sf) - \mathcal{F}_0(\rho)}{s} = \Big(\phi[\rho^*,\rho](1), f\Big).$$

Hence, $\mathcal{F}_0$ has the Fréchet derivative at ρ and $\delta\mathcal{F}_0/\delta\rho(\rho) = \phi[\rho^*,\rho](1)$. Since we may replace ρ by any $\bar\rho \in B_\delta(\rho)$ in the above arguments, we have

$$\frac{\delta\mathcal{F}_0}{\delta\rho}(\bar\rho) = \phi[\rho^*,\bar\rho](1), \qquad \forall \bar\rho \in B_\delta(\rho)$$

and so, since $\phi[\rho^*,\cdot](1)$ is ℓ_1-continuous on $B_\delta(\rho)$, we conclude the proof of (ii).

(ii) In light of Lemma 3.14, (i) implies (ii). □

Remark 7.4. *Let $\epsilon > 0$ and let $\rho, \rho^e \in \mathcal{P}(G)$ be such that $\rho_i \geq 10\epsilon$ for all $i \in V$ and $\sqrt{2}nC_\omega\mathcal{W}(\rho,\rho^e) \leq \epsilon$. Let μ be the geodesic of constant speed connecting ρ^e to ρ. We use Lemma 3.4 to conclude that*

$$\|\mu(t) - \rho\|_{\ell_1} \leq \sqrt{2}nC_\omega\mathcal{W}(\mu(t),\rho) = \sqrt{2}n(1-t)C_\omega\mathcal{W}(\rho^e,\rho) \leq \epsilon.$$

Hence, $\mu_i(t) \geq 9\epsilon$ for all $i \in V$.

If $\bar\rho \in \mathcal{P}(G)$ is such that $\sqrt{2}nC_\omega\mathcal{W}(\rho,\bar\rho) \leq \epsilon$ and $\bar\mu$ is a geodesic of constant speed connecting ρ^e to $\bar\rho$ then

$$\|\mu(t) - \rho^e\|_{\ell_1} \leq \sqrt{2}n(1-t)C_\omega\mathcal{W}(\rho^e,\bar\rho) \leq \sqrt{2}n(1-t)C_\omega\Big(\mathcal{W}(\rho^e,\rho) + \mathcal{W}(\rho,\bar\rho)\Big) \leq 2\epsilon.$$

Hence, $\bar\mu_i(t) \geq 8\epsilon$ for all $i \in V$. In conclusion, we have proven that if we set $\delta := \epsilon/(\sqrt{2}nC_\omega)$ then whenever $\bar\rho \in B_\delta(\rho)$ then every geodesic connecting ρ^e to ρ is contained in $[\epsilon, 1-\epsilon]^n$ and so, it is uniquely determined.

Corollary 7.5. *Assume that $\gamma_P(\rho^*) > 0$. Then:*

(i) *$\mathcal{F}_0$ has the Fréchet derivative and the Wasserstein gradient on $\mathcal{P}_0(G)$, and for any $\rho \in \mathcal{P}_0(G)$ we have $\|\nabla_\mathcal{W}\mathcal{F}_0(\rho)\|_\rho = \mathcal{W}(\rho,\rho^*)$. If ρ^e is on a geodesic connecting ρ^* to ρ and ρ^e is sufficiently close to ρ then*

$$\frac{\delta\mathcal{F}_0}{\delta\rho}(\rho) = \phi[\rho^e,\rho](1) \quad \textit{and} \quad \nabla_\mathcal{W}\mathcal{F}_0(\rho) = \nabla_G\phi[\rho^e,\rho](1).$$

In particular, $\nabla_G \phi[\rho^e, \rho](1)$ is independent of the geodesic.
(ii) *If $\rho^* \in \mathcal{P}_0(G)$ then both $\frac{\delta \mathcal{F}_0}{\delta \rho}$ and $\nabla_{\mathcal{W}} \mathcal{F}_0$ are ℓ_1–continuous at ρ.*

Proof. (i) Assume $\rho \in \mathcal{P}_0(G)$ and there exists $\epsilon > 0$ such that $\rho_i \geq 10\epsilon$ for all $i \in V$. Let $\lambda \in B_*$ be the maximizer in the dual formulation of $1/2\mathcal{W}^2(\rho^*, \rho)$ as chosen in the proof of Lemma 7.2 so that

$$\frac{1}{2}\mathcal{W}^2\big(\rho^*, \rho\big) = \big(\lambda(1), \rho\big) - \big(\lambda(0), \rho^*\big)$$

holds. Let μ be a geodesic of constant speed connecting ρ^* to ρ. Choose $r \in (0,1)$ so that $\mathcal{W}\big(\rho, \mu(1-r)\big) \leq \epsilon/(2\sqrt{n}C_\omega)$ and set $\rho^e := \mu(1-r)$. We have

$$\frac{1}{2}\mathcal{W}^2\big(\rho^e, \rho\big) = (\phi[\rho^e, \rho](1), \rho) - \big(\phi[\rho^e, \rho](0), \rho^e\big).$$

Let $\bar{\mu} : [1-r, 1] \to \mathcal{P}(G)$ be the geodesic connecting $\rho^e := \mu(1-r)$ to $\bar{\rho}$ and extend $\bar{\mu}$ to $[0, 1-r]$ to be the restriction of μ to $[0, 1-r]$. If v is the velocity of μ and $\bar{v}$ is the velocity of $\bar{\mu}$ on $[1-r, 1]$, we have

$$\mathcal{F}_0(\bar{\rho}) \leq \frac{1}{2}\int_0^{1-r} \|v\|_\mu^2 dt + \frac{1}{2}\int_{1-r}^1 \|\bar{v}\|_{\bar{\mu}}^2 dt = \mathcal{F}_0(\rho) + \frac{1}{2}\Big(\mathcal{W}^2\big(\rho^e, \bar{\rho}\big) - \mathcal{W}^2\big(\rho^e, \rho\big)\Big). \tag{7.11}$$

By Remark 7.4 and Proposition 7.3, $\mathcal{W}^2\big(\rho^e, \cdot\big)$ has the Fréchet derivative and the Wasserstein gradient at ρ. Furthermore,

$$\frac{\delta \mathcal{W}^2\big(\rho^e, \cdot\big)}{\delta \rho}(\rho) = \phi[\rho^e, \rho](1) \quad \text{and} \quad \nabla_{\mathcal{W}} \mathcal{W}^2\big(\rho^e, \cdot\big)(\rho) = \nabla_G \phi[\rho^e, \rho](1). \tag{7.12}$$

By (7.11) and in light of the first identity in (7.12), $\phi[\rho^e, \rho](1)$ is in the super-differential of $\mathcal{F}_0$ at ρ. But by Lemma 7.2, there exists $\lambda^\infty \in \mathbb{R}^n$ in the sub-differential of $\mathcal{F}_0$ at ρ. Since $\rho \in \mathcal{P}_0(G)$, we have that $\phi[\rho^e, \rho](1) - \lambda^\infty$ is orthogonal to any $f \in \mathbb{R}^n$ such that $\sum_{i=1}^n f_i = 0$. In other words, $\phi[\rho^e, \rho](1) - \lambda^\infty$ is parallel to $\mathbf{1} := (1, \cdots, 1)$. Since we have imposed the normalization property that both $\phi[\rho^e, \rho](1)$ and λ^∞ are perpendicular to $\mathbf{1}$, we conclude that $\phi[\rho^e, \rho](1) = \lambda^\infty$.

By (7.11) and in light of the second identity in (7.12), there exists a constant $C > 0$ such that, for every $\bar{\epsilon} > 0$, there exists $\bar{\delta} > 0$ satisfying for any $\bar{v} \in T_\rho \mathcal{P}(G)$

$$\|\bar{\rho} - \rho\|_{\ell_1} \leq \bar{\delta} \quad \Longrightarrow \quad \mathcal{F}_0(\bar{\rho}) \leq \mathcal{F}_0(\rho) + \big(\nabla_G \phi[\rho^e, \rho](1), \bar{v}\big)_\rho + C\|\bar{\rho} - \rho + \operatorname{div}_\rho(\bar{v})\|_{\ell_1} + \bar{\epsilon}\mathcal{W}(\rho, \bar{\rho}).$$

Hence, $\mathcal{F}_0$ has a Wasserstein super-gradient at ρ which is $\nabla_G \phi[\rho^e, \rho](1)$. Since in light of Lemma 7.2, $\mathcal{F}_0$ has a Wasserstein sub-gradient at ρ, we conclude that $\nabla_G \phi[\rho^e, \rho](1)$ is the gradient of $\mathcal{F}_0$ at ρ. We use (7.10) to obtain the identity $\|\nabla_{\mathcal{W}} \mathcal{F}_0(\rho)\|_\rho = \mathcal{W}(\rho, \rho^*)$.

(ii) Further assume that $\rho^* \in \mathcal{P}_0(G)$. We observe that $\phi[\rho^e, \rho](l) = \phi[\rho^*, \rho](lr + 1 - r)$ for $l \in [0,1]$. We use (2.10) when $l = 1$ to conclude that $\nabla_{\mathcal{W}} \mathcal{F}_0$ is ℓ_1-continuous at ρ. $\square$

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 9048215 and Hong Kong RGC Grant GRF 9043379. WG would like to thank Y. Gao, W. Li and J-G. Liu for fruitful discussions.

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] 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.

[10] 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.

[11] 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.

[12] A. Bensoussan, P.J. Graber and S.C.P. Yam, *Control on Hilbert spaces and application to mean field type control theory*, ArXiv preprint, 2020, ArXiv 2005.10770v2, https://urldefense.com/v3/__https://arxiv.org/abs/2005.10770__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP-BpI6BhQ$.

[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, https://arxiv.org/pdf/2203.14554.pdf.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] G. Conforti, R.C. Kraaij and D. Tonon, *Hamilton–Jacobi equations for controlled gradient flows: the comparison principle*, ArXiV preprint, 2023, arXiv:2111.13258, https://urldefense.com/v3/__https://arxiv.org/abs/2111.13258__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP9OaivIwg$.

[24] G. Conforti, R.C. Kraaij and D. Tonon, *Hamilton–Jacobi equations for controlled gradient flows: cylindrical test functions*, ArXiV preprint, 2023, arXiv:2302.06571, https://urldefense.com/v3/__https://arxiv.org/abs/2302.06571__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP8fpGomNA$.

[25] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham and M. Rosestolato, *Master Bellman equation in the Wasserstein space: Uniqueness of viscosity solutions*, ArXiV preprint, 2022, arXiv:2107.10535v2, https://urldefense.com/v3/__https://arxiv.org/abs/2107.10535__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP-5VGSexw$.

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

[27] 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.

[28] 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.

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

[34] J. Cui, S. Liu and H. Zhou, *Wasserstein Hamiltonian flow with common noise on graph*, ArXiV preprint, 2022, arXiv:2204.01185, https://urldefense.com/v3/__https://arxiv.org/pdf/2204.01185.pdf__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP9sP1FZOw$.

[35] 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

[36] T. Dominguez and J.-C. Mourrat, *Infinite-dimensional Hamilton-Jacobi equations for statistical inference on sparse graphs*, ArXiV preprint, 2022, arXiv:2209.04516, https://urldefense.com/v3/__https://arxiv.org/pdf/2209.04516.pdf__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP-k0qNnhw$.

[37] 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.

[38] 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.

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

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

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

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

[48] 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.

[49] 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.

[50] 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.

[51] 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, `https://urldefense.com/v3/__https://arxiv.org/pdf/2212.05675.pdf__;!!KjDnqvtInNPT!iIO2byVWz5OiuuErhIoj12L_Fb4VvEi_5vENgY-EpMOX4C7-uQEszeY6ArSnIRPx7SiSj_UJr81Xi4OFpP-xreHoSQ$`.

[52] 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.

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

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

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

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

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

[58] 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.

[59] 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.

[60] 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.

[61] 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.

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

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

[64] 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.

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

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

[67] 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.

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

[69] A. Nakayasu, *Metric viscosity solutions for Hamilton-Jacobi equations of evolution type*, Adv. Math. Sci. Appl. 24 (2014), no. 2, 333–351.

[70] 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.

[71] 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.

[72] 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.

[73] 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