ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

J.M.P.A.

Journal de Mathématiques Pures et Appliquées 00 (2018) 1–47

On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations

Wilfrid Gangbo

Department of Mathematics, UCLA, Los Angeles, CA 90095; wgangbo@math.ucla.edu

Adrian Tudorascu

Department of Mathematics, West Virginia University, Morgantown, WV 26506; adriant@math.wvu.edu

Abstract

In this paper we elucidate the connection between various notions of differentiability in the Wasserstein space: some have been introduced intrinsically (in the Wasserstein space, by using typical objects from the theory of Optimal Transport) and used by various authors to study gradient flows, Hamiltonian flows, and Hamilton-Jacobi equations in this context. Another notion is extrinsic and arises from the identification of the Wasserstein space with the Hilbert space of square-integrable random variables on a non-atomic probability space. As a consequence, the classical theory of well-posedness for viscosity solutions for Hamilton-Jacobi equations in infinite-dimensional Hilbert spaces is brought to bear on well-posedness for Hamilton-Jacobi equations in the Wasserstein space.

Résumé. Dans cet article, nous élucidons le lien entre diverses notions de différentiabilité dans l'espace de Wasserstein: certaines ont été introduites intrinsèquement (dans l'espace de Wasserstein, en utilisant des objets typiques de la théorie du transport optimal) et utilisées par divers auteurs pour étudier les flots gradients, les flots Hamiltoniens ainsi que les équations de Hamilton–Jacobi dans ce contexte. Une notion alternative et extrinsèque, est basée sur l'identification de l'espace de Wasserstein avec l'espace de Hilbert des variables aléatoires de carré intégrables, sur un espace des mesures de probabilité non–atomique. Il s'avère que la théorie des équations de Hamilton-Jacobi dans l'espace de Wasserstein, repose sur la théorie classique des solutions de viscosité des équations de Hamilton–Jacobi dans les espaces de Hilbert de dimension infinie.

Keywords: Differentiability in the Wasserstein space, Hamilton-Jacobi equations in the Wasserstein space, Viscosity solutions
2010 MSC: 35R15, 46G05, 58D25

1. Introduction

This manuscript is a contribution to the theory of viscosity solutions for Hamilton-Jacobi equations in $\mathcal{P}_2(\mathbb{R}^d)$, the set of probability measure on $\mathbb{R}^d$ with finite second moments; it is endowed with the Wasserstein metric W_2. Many challenges already overcome on "flat" spaces, such as Banach spaces which satisfy the so-called

Radon-Nikodym property (cf. e.g. [11] [12]), have to be faced in the study of first order equations in $\mathcal{P}_2(\mathbb{R}^d)$. The latter set, with its special metric structure, allows for an Eulerian description of systems consisting of either finitely many or infinitely many particles. A Lagrangian description of these systems can be achieved via any fixed non-atomic probability space. For instance, as done in this manuscript, one can choose the probability space Ω to be the ball of unit volume in $\mathbb{R}^d$, centered at the origin. The probability measure here will be the the d-dimensional Lebesgue measure restricted to Ω (denoted by $\mathcal{L}^d_\Omega$) and the set of random variables is the Hilbert space

$$\mathbb{H} := L^2(\Omega; \mathbb{R}^d).$$

It is endowed with the inner product $\langle \cdot, \cdot \rangle$, defined for $X, Y \in \mathbb{H}$ by

$$\langle X, Y \rangle := \mathbb{E}(X \cdot Y) = \int_\Omega X(\omega) \cdot Y(\omega) d\omega.$$

We denote by $\|\cdot\|$ its associated norm, i.e.

$$\|X\|^2 := \langle X, X \rangle \text{ for all } X \in \mathbb{H}.$$

The set $\mathbb{H}$ is a Hilbert manifold with a single global chart given by the identity map on $\mathbb{H}$. The metric and the natural Levi-Civita connection on $\mathbb{H}$ are linked to the metric and the Levi-Civita connection [17] [18] on $\mathcal{P}_2(\mathbb{R}^d)$ (cf. also [10]). The push-forward operator $\sharp : \mathbb{H} \to \mathcal{P}_2(\mathbb{R}^d)$ associates to $X \in \mathbb{H}$ its law – the Borel measure $X_\sharp \mathcal{L}^d_\Omega$ – defined for any Borel subset B of $\mathbb{R}^d$ by

$$(X_\sharp \mathcal{L}^d_\Omega)(B) = \mathcal{L}^d_\Omega(X^{-1}(B)).$$

The map $\sharp$ yields an equivalence relation on $\mathbb{H}$: the class of equivalence of $X \in \mathbb{H}$ is denoted by $[X]_\sharp$, the set of $Y \in \mathbb{H}$ which have the same law as X, denoted by $\sharp(X)$. Note that $\sharp$ is surjective since the optimal mass transportation theory ensures [4] that any element of $\mathcal{P}_2(\mathbb{R}^d)$ is the law of the gradient of a convex function $\phi : \Omega \to \mathbb{R}$.

If μ is the law of $X \in \mathbb{H}$, ν is the law of $Y \in \mathbb{H}$ then (cf. e.g. [3])

$$W_2^2(\mu, \nu) = \inf_{\bar{X}, \bar{Y} \in \mathbb{H}} \Big\{ \mathbb{E}(\|\bar{X} - \bar{Y}\|^2) \ : \ \mu = \sharp(\bar{X}),\ \nu = \sharp(\bar{Y}) \Big\}. \tag{1.1}$$

Therefore, $\sharp$ is an isometry of the quotient space $\mathbb{H}/\sharp$ onto $\mathcal{P}_2(\mathbb{R}^d)$. This is the first hint that the intrinsic differential structure on $\mathcal{P}_2(\mathbb{R}^d)$, introduced in [3], may be inherited from the differential structure on the Hilbert space $\mathbb{H}$. One may also suspect that the Levi-Civita connection on both spaces may allow to link the Hessian of functions defined on $\mathcal{P}_2(\mathbb{R}^d)$ to functions defined on $\mathbb{H}/\sharp$.

There is a special non-commutative group related to the isometry $\sharp : \mathbb{H}/\sharp \to \mathcal{P}_2(\mathbb{R}^d)$, namely the set $\mathcal{G}(\Omega)$ of Borel maps $S : \Omega \to \Omega$ (they lie in $\mathbb{H}$) that are almost everywhere invertible and have the same law as the identity map **id**. The binary operation on $\mathcal{G}(\Omega)$ is the composition operation $\circ$ and the orbit of $X \in \mathbb{H}$ is the set $X \cdot \mathcal{G}(\Omega) = \{X \circ S \ : \ S \in \mathcal{G}(\Omega)\}$. The set of orbits of (points X in) $\mathbb{H}$ generated by the action of $\mathcal{G}(\Omega)$ form a partition of $\mathbb{H}$. We henceforth have another equivalence relation consistent with the right action of $\mathcal{G}(\Omega)$ on $\mathbb{H}$. This is very helpful for developing our intuition to better understand $\sharp$, but unfortunately the new equivalence relation differs from that induced by $\sharp$; indeed, $X \cdot \mathcal{G}(\Omega)$ is strictly contained in its closure $[X]_\sharp$. This conclusion follows from the fact that $\sharp(X) = \sharp(Y)$ if and only if there exists a sequence $(S_n)_n \subset \mathcal{G}(\Omega)$ such that (cf. e.g. Lemma 6.4 [9])

$$\lim_{n \to \infty} \|X - Y \circ S_n\| = 0.$$

In particular, this yields that the closure of $\mathcal{G}(\Omega)$ is the set $\mathcal{S}(\Omega)$ of Borel maps which have the same law as **id** (see Theorem 1.4 [5] for a refined result when $d \geq 2$).

We would like to understand how we can exploit the point of view that the quotient space $\mathbb{H}/\sharp$ is isometric to the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$ in order to make inferences on partial differential equations in the latter set. Partial results in this direction were obtained in the one-dimensional case for "mechanical" Hamiltonians on

$\mathcal{P}_2(\mathbb{R})$ [15]. In the current manuscript, we go beyond offering a new point of view on the differential structure on $\mathcal{P}_2(\mathbb{R}^d)$ and the concept of subdifferential of functions on this set; we were successful in substantially advancing the theory of existence and uniqueness of solutions to first order Hamilton-Jacobi equations on $\mathcal{P}_2(\mathbb{R}^d)$.

The notions of viscosity solutions to the differential equations (either on $\mathbb{H}$ or on $\mathcal{P}_2(\mathbb{R}^d)$) we are concerned with are expressed in terms of subdifferentials of functions. A few years ago, [3] introduced the concepts of sub and super-differential for real-valued functions on $\mathcal{P}_2(\mathbb{R}^d)$, intrinsic to the weak differential structure of this space. For practical reasons, a need later arose to modify these concepts (cf. e.g. [2]). It is proved in [2] that for λ-convex functions on $\mathcal{P}_2(\mathbb{R}^d)$ the definition of subdifferential in [3] is equivalent to that in [2]. A first task completed in this paper is to show that the definitions of subdifferential in [2] and [3] coincide (with no need for extra-assumptions, such as λ–convexity). A second task completed is to compare the extrinsic definitions of sub and super-differential of [9] with the intrinsic ones of [3]. For instance, we show that if $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $\tilde{U} : \mathbb{H} \to \mathbb{R}$ is its rearrangement invariant "lift" defined by

$$\tilde{U} = U \circ \sharp, \tag{1.2}$$

then the subdifferential of U at $\mu \in \mathcal{P}_2(\mathbb{M})$ is nonempty if and only if the subdifferential of $\tilde{U}$ at *some* $X_0 \in \mathbb{H}$ such that $\sharp(X_0) = \mu$ is nonempty. This is further equivalent to the fact that the subdifferential of $\tilde{U}$ at *any* $X \in \mathbb{H}$ whose law is μ is nonempty. Let $\partial_\cdot^o U(\mu)$ denote the element of minimal norm of $\partial_\cdot U(\mu)$ and let $\partial_\cdot^o \tilde{U}(X)$ denote the element of minimal $L^2(\mu;\mathbb{R}^d)$-norm of $\partial_\cdot \tilde{U}(X)$; we may express the L^2–subgradient in terms of the Wasserstein subgradient as

$$\partial_\cdot^o \tilde{U}(X) = \partial_\cdot^o U(\mu) \circ X. \tag{1.3}$$

The identity (1.3) is very subtle. First, it forces the level sets of X to be subsets of the level sets of $\partial_\cdot^o \tilde{U}(X)$. Secondly, it implies that gradients of rearrangement invariant functions defined on $\mathbb{H}$ have a special structure. Indeed, if, for instance, X is invertible, (1.3) implies that $\partial_\cdot^o \tilde{U}(X) \circ X^{-1}$ belongs to the closure of $\nabla C_c^1(\mathbb{R}^d)$ in $L^2(\mu;\mathbb{R}^d)$, where $\mu := \sharp(X)$. Thirdly, a more general attempt to express elements of the subdifferential of $\tilde{U}$ at X in terms of elements of the subdifferential of U at μ is doomed to fail even when $d = 1$. Example 3.20 provides us with a function U, its lift $\tilde{U}$ and ζ in the subdifferential of $\tilde{U}$ at X such that there is no ξ for which $\zeta = \xi \circ X$.

The conclusion reached in (1.3), which plays an instrumental role in our study, has been obtained by relying on two deep results: (i) Brenier [4] proved that for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ there is a convex function whose gradient pushes $\mathcal{L}^d_\Omega$ forward to μ. The existence of such a gradient map, which is referred to as an optimal map, is the essence of the theory of optimal transportation. (ii) A remarkable result by Caravenna and Daneri [8] ensures that given any convex function $\phi : \Omega \to \mathbb{R}$, one can disintegrate $\mathcal{L}^d_\Omega$ into probability measures $\{\nu_y \;:\; y \in \nabla\phi(\Omega)\}$ such that each ν_y is supported by the level set $\{\nabla\phi = y\}$ and is comparable to the Hausdorff measure $\mathcal{H}^{k(y)}$, where $k(y) \in \{0, 1, ..., d\}$.

The Levi-Civita connection on $\mathcal{P}_2(\mathbb{R}^d)$ [18], allows for a definition of the Hessian of U under appropriate conditions (cf. [10]). Indeed, assume that U is differentiable on $\mathcal{P}_2(\mathbb{R}^d)$ in the sense that for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ the sets $\partial_\cdot U(\mu)$ and $\partial^\cdot U(\mu)$ are simultaneously nonempty. As commonly done in convex analysis, we refer to the element of minimal norms in the subdifferential as the gradient of U at μ and denote it by $\nabla_w U(\mu)$. For each $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, let us define

$$\mathcal{T}_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\nabla C_c^1(\mathbb{R}^d)}^{L^2(\mu;\mathbb{R}^d)}.$$

We shall see that

$$\partial_\cdot U(\mu) + [\mathcal{T}_\mu \mathcal{P}_2(\mathbb{R}^d)]^\perp = \partial_\cdot U(\mu),$$

which implies $\nabla_w U(\mu) \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{R}^d)$.

We suppose that for any $\xi \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$ the directional derivative

$$\mu \mapsto \xi(U[\mu]) := \langle \nabla_w U(\mu), \xi \rangle_\mu$$

is differentiable on $\mathcal{P}_2(\mathbb{R}^d)$. (Here and thereon $\langle \cdot, \cdot \rangle_\mu$ denotes the standard inner product of $L^2(\mu;\mathbb{R}^d)$ and $\|\cdot\|_\mu$ the associated norm.) In particular, if $\xi_1, \xi_2 \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and we set $\bar{\nabla}_{\xi_1}\xi_2 = (\nabla\xi_2)\xi_1$, then

$$\text{Hess}\, U[\mu](\xi_1, \xi_2) = \xi_1\Big(\xi_2(U[\mu])\Big) - \Big(\bar{\nabla}_{\xi_1}\xi_2\Big)(U[\mu])$$

is meaningful. Assume also $\tilde{U}$ is twice continuously differentiable on $\mathbb{H}$, viewed as an infinite dimensional manifold. For $X \in \mathbb{H}$, denote by Hess $\tilde{U}(X) : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$ the Hessian of $\tilde{U}$ at X. One checks that, as a consequence of (1.3), if $\mu = \sharp(X)$, then

$$\text{Hess}\, U(\mu)(\xi_1, \xi_2) = \text{Hess}\, \tilde{U}(X)(\xi_1 \circ X, \xi_2 \circ X) \tag{1.4}$$

for any $\xi_1, \xi_2 \in \nabla C_c^\infty(\mathbb{R}^d)$.

Contrary to the fact that the real valued function $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ has a unique rearrangement invariant lift, given in (1.2), real-valued functions on

$$C\mathcal{P}_2(\mathbb{R}^d) := \{(\mu, \xi) \ : \ \mu \in \mathcal{P}_2(\mathbb{R}^d),\ \xi \in L^2(\mu; \mathbb{R}^d)\}$$

may have several distinct lifts $\tilde{H}$ on $\mathbb{H} \times \mathbb{H}$, satisfying the invariance property

$$\tilde{H}(X, \zeta) = \tilde{H}(X \circ S, \zeta \circ S) \tag{1.5}$$

for any $S \in \mathcal{S}(\Omega)$. For instance, the function on $C\mathcal{P}_2(\mathbb{R}^d)$ defined by

$$(\mu, \xi) \mapsto \mathcal{H}(\mu, \xi) := \frac{1}{2} \int_{\mathbb{R}^d} |\xi(x)|^2 \mu(dx),$$

admits infinitely many lifts on $\mathbb{H} \times \mathbb{H}$ satisfying (1.5), two of which are

$$\tilde{H}(X, \zeta) := \frac{1}{2} \|\text{proj}_{F[X]} \zeta\|^2, \quad \tilde{H}_1(X, \zeta) := \frac{1}{2} \|\zeta\|^2, \tag{1.6}$$

For $X \in \mathbb{H}$,

$$F[X] = \{\xi \circ X \ : \ \xi \in L^2(\mu; \mathbb{R}^d)\} := \overline{\{\phi \circ X \ : \ \phi \in C_c(\mathbb{R}^d; \mathbb{R}^d)\}}^{L^2(\Omega; \mathbb{R}^d)}, \tag{1.7}$$

where $\mu := \sharp(X)$. While $\tilde{H}_1$ is the unique continuous lift of $\mathcal{H}$, the natural (see below) lift $\tilde{H}$ comes with some serious handicaps since it is far from satisfying the sufficient conditions employed by [11], [12] in order to prove existence and uniqueness of continuous viscosity solutions in $\mathbb{H}$. Indeed, $\tilde{H}$ is not even continuous on $\mathbb{H} \times \mathbb{H}$. Thus, the extant proofs of existence and uniqueness do not work in this case.

We employ a workaround, based on studying some HJ equations with Hamiltonians that may, at a first glance, look like "linearized", "toy" Hamiltonians. These Hamiltonians depend on a functional parameter $\varphi \in C_c^\infty(\mathbb{R}^d)$ in the following way:

$$H_\varphi(\mu, \xi) := \langle \xi, \nabla\varphi \rangle_\mu - \frac{1}{2} \|\nabla\varphi\|_\mu^2 .$$

Define

$$\nabla F[X] := \{\xi \circ X \ : \ \xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{R}^d)\} := \overline{\{\nabla\phi \circ X \ : \ \phi \in C_c^1(\mathbb{R}^d)\}}^{L^2(\Omega; \mathbb{R}^d)}. \tag{1.8}$$

If $\zeta \in \mathbb{H}$ and $\xi \circ X := \text{proj}_{\nabla F[X]} \zeta$, then, thanks to the identities

$$\langle \xi, \nabla\varphi \rangle_\mu = \langle \xi \circ X, \nabla\varphi \circ X \rangle = \langle \zeta, \nabla\varphi \circ X \rangle,$$

the lift defined by (given a general Hamiltonian H on $\mathcal{T}\mathcal{P}_2(\mathbb{R}^d)$)

$$\tilde{H}(X, \zeta) := H\Big(\sharp(X), \text{proj}_{\nabla F[X]} \zeta\Big) \tag{1.9}$$

becomes

$$\tilde{H}_\varphi(X, \zeta) = \langle \zeta, \nabla\varphi \circ X \rangle - \frac{1}{2} \|\nabla\varphi \circ X\|^2 \tag{1.10}$$

when $H := H_\varphi$. One checks that the theory in [11], [12] can be applied to the smooth Hamiltonians in (1.10). The identity (note that this holds if and only if $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{R}^d)$)

$$\frac{1}{2} \|\xi\|_\mu^2 = \sup_{\varphi \in C_c^1(\mathbb{R}^d)} H_\varphi(\mu, \xi)$$

gives us our first glimpse that there may be ways to connect HJ equations on $\mathcal{P}_2(\mathbb{R}^d)$ to HJ equations on $\mathbb{H}$, even when the lift in (1.9) cannot be directly used. From our perspective, the lift in (1.9) is arguably the most natural lift for H satisfying (1.5). Indeed, for this lift, Theorem 4.4 shows that the identity (1.3) becomes directly useful in the study of viscosity solutions of Hamilton-Jacobi equations.

Other works [1], [16] on Hamilton-Jacobi equations in the Wasserstein space deal with more general metric settings and have various degrees of generality (either special types of Hamiltonians depending on metric derivatives and/or concepts of viscosity solutions based on special types of test functions). It will be interesting to investigate whether the measure-random variable duality approach used in this paper applies to such problems.

In this work, our focus is on genuine first-order Hamilton-Jacobi equations in the Wasserstein space of type (4.3) and (6.1).

2. Notation and Preliminaries

To emphasize the fact that most of the results proved in this manuscript are valid on spaces more general than $\mathbb{R}^d$, we shall denote $\mathbb{R}^d$ by $\mathbb{M}$. Throughout this manuscript, $\mathcal{P}_2(\mathbb{M})$ denotes the set Borel probability measures on $\mathbb{M}$, of finite second moments. This is a length space when endowed with W_2, the Wasserstein distance.

Given $\mu, \nu \in \mathcal{P}_2(\mathbb{M})$ we denote by $\Gamma(\mu, \nu)$ the set of Borel measures γ on $\mathbb{M} \times \mathbb{M}$, which have μ as first marginal, ν as second marginal. We denote by $\Gamma_o(\mu, \nu)$, the set of $\gamma \in \Gamma(\mu, \nu)$ such that

$$W_2^2(\mu, \nu) = \int_{\mathbb{M}\times\mathbb{M}} |x - y|^2 \gamma(dx, dy).$$

We denote the first (respectively, second) projection of $\mathbb{M} \times \mathbb{M}$ onto $\mathbb{M}$ by π^1 (respectively, π^2)

$$\pi^1(x, y) = x, \quad \pi^2(x, y) = y.$$

The space of uniformly continuous functions on $\mathbb{H}$ is $UC(\mathbb{H})$ while the subspace of bounded functions in $UC(\mathbb{H})$ is $BUC(\mathbb{H})$. The space $UC_s([0, T] \times V)$ consists of those functions $v : [0, T] \times V \to \mathbb{R}$ which are uniformly continuous in x uniformly with respect to t, and uniformly continuous on bounded sets.

Denote by $L^2(\mu)$ the set of Borel maps $\xi : \mathbb{M} \to \mathbb{M}$ such that $||\xi||_\mu^2 := \int_{\mathbb{M}} |\xi(x)|^2 \mu(dx) < \infty$. The union $\cup_{\mu\in\mathcal{P}_2(\mathbb{M})}\{\mu\} \times \mathcal{P}_2(\mathbb{M})$ is denoted by $C\mathcal{P}_2(\mathbb{M})$. The closure of $\nabla C_c^\infty(\mathbb{M})$ in $L^2(\mu)$ is denoted by $\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$.

We define $\sharp : \mathbb{H} \to \mathcal{P}_2(\mathbb{M})$ by

$$\sharp(X)(B) = \mathcal{L}^d_\Omega(X^{-1}(B)) \text{ for all } X \in \mathbb{H},\ B \subset \mathbb{M}$$

and, for any $X \in \mathbb{H}$,

$$[X]_\sharp := \{Y \in \mathbb{H} \ : \ \sharp(Y) = \sharp(X)\}.$$

Recall from the introduction that we denote

$$\mathcal{G}(\Omega) := \{S : \Omega \to \Omega \ : \ S \text{ is Borel, } \mathcal{L}^d\text{-a.e. invertible, and } S \in [\mathbf{id}_\Omega]_\sharp\},$$

where $\mathbf{id}_\Omega$ is the identity map of Ω. Also,

$$\mathcal{S}(\Omega) := \{S : \Omega \to \Omega \ : \ S \text{ is Borel and } S \in [\mathbf{id}_\Omega]_\sharp\}.$$

We say that a function $m : [0, \infty) \to [0, \infty)$ is a *modulus* if m is continuous, monotone, nondecreasing, sub-additive and $m(0) = 0$. A function $\sigma : [0, \infty) \times [0, \infty) \to [0, \infty)$ is a *local modulus* if $\sigma(\cdot, R)$ is a modulus for each $R \geq 0$ and σ is continuous and monotone nondecreasing in each variable.

In this manuscript, if $\mathcal{S}$ is a metric space $UC_s([0,T]\times\mathcal{S})$ denotes the vector space of all real valued functions U on $[0,T]\times\mathcal{S}$ which are uniformly continuous on bounded subsets of $[0,T]\times\mathcal{S}$ and such that $U(t,\cdot)$ is uniformly continuous uniformly with respect to $t\in[0,T]$.

If $X\in\mathbb{H}$, $F[X]$ is the closure of the set $\{f\circ X\ :\ f\in C_c(\mathbb{M};\mathbb{M})\}$ in $\mathbb{H}$. If $\zeta\in\mathbb{H}$, the projection of ζ onto $F[X]$ is denoted by $\mathrm{proj}_{F[X]}\zeta$. If the closure of the set $\{\nabla\varphi\circ X\ :\ \varphi\in\nabla C^1(\mathbb{M})\}$ in $\mathbb{H}$ is used instead and denoted by $\nabla F[X]$, then the projection of $\zeta\in\mathbb{H}$ onto this subspace is denoted by $\mathrm{proj}_{\nabla F[X]}\zeta$.

If $V:\mathbb{H}\to\mathbb{R}\cup\{\pm\infty\}$, $X,\zeta\in\mathbb{H}$ and $r>0$, we set

$$\epsilon[V](r,X,\zeta)=\inf_{H\in\mathbb{H}}\Big\{\frac{V(H+X)-V(X)-\langle\zeta,H\rangle}{\|H\|}\ :\ 0<\|H\|\le r\Big\}. \tag{2.1}$$

If $U:\mathcal{P}_2(\mathbb{M})\to\mathbb{R}\cup\{\pm\infty\}$, $\mu,\nu\in\mathcal{P}_2(\mathbb{M})$, and $\xi\in L^2(\mu)$, then for any $p\in\Gamma(\mu,\nu)$ we set

$$e_\mu(\nu,\xi,\gamma):=U(\nu)-U(\mu)-\int_{\mathbb{M}\times\mathbb{M}}\xi(x)\cdot(y-x)\gamma(dx,dx). \tag{2.2}$$

and then set

$$e^\mu[U](\xi,\nu):=\sup_{p\in\Gamma_0(\mu,\nu)}\Big\{\frac{e_\mu(\nu,\xi,\gamma)}{W_2(\mu,\nu)}\Big\},\qquad e_\mu[U](\xi,\nu):=\inf_{p\in\Gamma_0(\mu,\nu)}\Big\{\frac{e_\mu(\nu,\xi,\gamma)}{W_2(\mu,\nu)}\Big\}. \tag{2.3}$$

For $r>0$ we set

$$\epsilon^\mu[U](r,\xi):=\inf_\nu\Big\{e^\mu[U](\xi,\nu)\mid\nu\in\mathcal{P}_2(\mathbb{M}),\ 0<W_2(\mu,\nu)\le r\Big\}, \tag{2.4}$$

and

$$\epsilon_\mu[U](r,\xi):=\inf_\nu\Big\{e_\mu[U](\xi,\nu)\mid\nu\in\mathcal{P}_2(\mathbb{M}),\ 0<W_2(\mu,\nu)\le r\Big\}, \tag{2.5}$$

As a consequence of (1.1) we obtain the following remarks.

Remark 2.1. *Let $U:\mathcal{P}_2(\mathbb{M})\to\mathbb{R}\cup\{\pm\infty\}$ and set $\tilde U:=U\circ\sharp:\mathbb{H}\to\mathbb{R}\cup\{\pm\infty\}$. Then*

(i) U is continuous if and only if $\tilde U$ is continuous.

(ii) U is κ–Lipschitz if and only if $\tilde U$ is κ–Lipschitz.

Lemma 2.2. *If $X,\zeta\in\mathbb{H}$ and $\mu:=\sharp(X)$, then there exists a unique $\mathbf{v}_0\in L^2(\mu)$ such that*

$$\langle b\circ X,\zeta\rangle=\langle b,\mathbf{v}_0\rangle_\mu\ \textit{for any}\ b\in C_c^\infty(\mathbb{M},\mathbb{M}).$$

Proof. The linear map $L:L^2(\mu)\to\mathbb{R}$ defined by

$$L(\xi):=\langle\xi\circ X,\zeta\rangle$$

is continuous. By the Riesz representation theorem there exists a unique $\mathbf{v}_0\in L^2(\mu)$ such that

$$L(\xi)=\langle\xi,\mathbf{v}_0\rangle_\mu\ \text{for all}\ \xi\in L^2(\mu).$$

It is straight-forward to check that $\mathbf{v}_0$ is the unique map with the stated properties. □

It is easy to see that $\mathbf{v}_0$ is the projection of ζ onto $F[X]$ (defined in (1.7)), so we denote

$$\mathbf{v}_0=\mathrm{proj}_{F[X]}\zeta. \tag{2.6}$$

Remark 2.3. *We shall also consider the set of $\mathbf{v}\in L^2(\mu)$ such that*

$$\langle\nabla\varphi\circ X,\zeta\rangle=\langle\nabla\varphi,\mathbf{v}\rangle_\mu\ \textit{for all}\ \varphi\in C_c^\infty(\mathbb{M}), \tag{2.7}$$

which contains $\mathbf{v}_0$ found in Lemma 2.2. Observe that if $\mathbf{v}$ satisfies (2.7), so does $\mathbf{v}+\mathbf{w}$ for any $\mathbf{w}\in L^2(\mu)$ such that $\nabla\cdot(\mu\mathbf{w})=0$ in the sense of distributions.

In the remainder of this section we consider

$$\bar{L} \in C([0,T] \times \mathbb{H} \times \mathbb{H}), \quad L \in C([0,T] \times C\mathcal{P}_2(\mathbb{M}))$$

bounded below and such that

$$\bar{L}(t, X, \xi \circ X) = L(t, \mu, \xi)$$

for all $t \in [0,T]$, $X \in \mathbb{H}$, $\mu = \sharp(X)$ (recall that this means $\mu = X_\sharp \mathcal{L}^d_\Omega$) and $\xi \in L^2(\mu)$. We assume that there exist $\kappa_0 > 0$ and $\kappa_1 \in \mathbb{R}$ such that

$$\kappa_0 \|B\|^2 - \kappa_1 \leq \bar{L}(t, X, B) \text{ for all } (t, X, B) \in [0,T] \times \mathbb{H} \times \mathbb{H}. \tag{2.8}$$

Lemma 2.4. *Assume $\bar{L}(\bar{t}, \bar{X}, \cdot)$ is strictly convex for any $(\bar{t}, \bar{X}) \in [0,T] \times \mathbb{H}$. Then for any $(t, X) \in [0,T] \times \mathbb{H}$ there exists a unique $\mathbf{v}_0 \in L^2(\sharp(X))$ such that $\mathbf{v}_0$ minimizes*

$$i_0 := \inf_{\mathbf{v}} \{\bar{L}(t, X, \mathbf{v} \circ X) \ : \ \mathbf{v} \ \textit{ satisfies } \ (2.7)\}\,. \tag{2.9}$$

Proof. Since the set of $\mathbf{v}$ satisfying (2.7) is nonempty and $\bar{L}$ is bounded below, $i_0 \in \mathbb{R}$. By (2.8), any minimizing sequence $(\mathbf{v}_n)_n$ is bounded in $L^2(\mu)$ and so, is weakly pre–compact in $L^2(\mu)$. We may assume without loss of generality that the whole sequence weakly converges to some $\mathbf{v}_0 \in L^2(\mu)$. Choose

$$\mathbf{w}_n := \sum_{k=1}^n c_k^n \mathbf{v}_k, \quad c_k^n \geq 0, \quad \sum_{k=1}^n c_k^n = 1$$

such that $(\mathbf{w}_n)_n$ converges strongly to $\mathbf{v}_0$. Observe that $\mathbf{v}_0$ satisfies (2.7). It remains to show that $\bar{L}(t, X, \mathbf{v}_0 \circ X) \leq i_0$. Let $\epsilon > 0$ be arbitrary. Assume without loss of generality that

$$\bar{L}(t, X, \mathbf{v}_n \circ X) \leq i_0 + \epsilon$$

for every $n \geq 1$. We have

$$\bar{L}(t, X, \mathbf{w}_n \circ X) \leq \sum_{k=1}^n c_k^n \bar{L}(t, X, \mathbf{v}_k \circ X) \leq \sum_{k=1}^n c_k^n (i_0 + \epsilon) = i_0 + \epsilon \tag{2.10}$$

and

$$\lim_n \|\mathbf{w}_n \circ X - \mathbf{v}_0 \circ X\| = \lim_n \|\mathbf{w}_n - \mathbf{v}_0\|_\mu = 0.$$

By the continuity of $\bar{L}$ and (2.10)

$$\bar{L}(t, X, \mathbf{v}_0 \circ X) = \limsup_n \bar{L}(t, X, \mathbf{w}_n \circ X) \leq i_0 + \epsilon.$$

This proves that $\mathbf{v}$ minimizes (2.9). The strict convexity of $\bar{L}(t, X, \cdot)$ ensures the uniqueness of $\mathbf{v}_0 \circ X$, which in turn ensures the uniqueness of $\mathbf{v}_0$. □

Definition 2.5. *Assume $\bar{L}(\bar{t}, \bar{X}, \cdot)$ is strictly convex for any $(\bar{t}, \bar{X}) \in [0,T] \times \mathbb{H}$. Let $(t, X, \zeta) \in [0,T] \times \mathbb{H} \times \mathbb{H}$ and set $\mu := \sharp(X)$. Using the notation in Lemma 2.4,*

(i) we refer to $\bar{\zeta} := \mathbf{v} \circ X \in \mathbb{H}$ as the $\bar{L}$–projection of ζ onto $\nabla F[X]$ (see (1.8)*) and write $\bar{\zeta} = \mathrm{proj}_{\nabla F[X], \bar{L}} \zeta$.*

(ii) When $\bar{L}(B) \equiv \|B\|^2$ for $B \in \mathbb{H}$, the $\bar{L}$–projection is simply denoted $\mathrm{proj}_{\nabla F[X]} \zeta$ and referred to as the projection of ζ onto $\nabla F[X]$ (since it is easy to see that it coincides with the orthogonal projection of ζ onto the closed subspace defined in (1.8)*).*

Proposition 2.6. *Assume both $\bar{L}(\bar{t}, \bar{X}, \cdot)$ and $\bar{H}(\bar{t}, \bar{X}, \cdot)$ are strictly convex and of class C^1 for any $(\bar{t}, \bar{X}) \in [0,T] \times \mathbb{H}$. Let $(t, X, \zeta) \in [0,T] \times \mathbb{H} \times \mathbb{H}$. Then $\mathbf{v} \circ X = \mathrm{proj}_{\nabla F[X], \bar{L}} \zeta$ for some $\mathbf{v}$ satisfying* (2.7) *if and only if*

$$\xi := \nabla_b L(t, \mu, \mathbf{v}) \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M}).$$

Proof. Set $\mu := \sharp(X)$. Assume $\mathbf{v} \circ X = \mathrm{proj}_{\nabla F[X],\bar{L}}\zeta$ and let $\mathbf{w} \in L^2(\mu)$ be such that $\nabla \cdot (\mu \mathbf{w}) = 0$ in the sense of distributions. For every $r \in \mathbb{R}$, (2.7) holds if we replace $\mathbf{v}$ by $\mathbf{v} + r\mathbf{w}$. Thus,

$$0 = \frac{d}{dr}\bar{L}(t, X, \mathbf{v} \circ X + r\mathbf{w} \circ X)\Big|_{r=0} = \langle \nabla_B \bar{L}(t, X, \mathbf{v} \circ X), \mathbf{w} \circ X\rangle$$

If $\bar{L}(t, X, \cdot)$ is differentiable on $\mathbb{H}$ and $L(t, \mu, \cdot)$ is differentiable on $L^2(\mu)$ then

$$\nabla_B \bar{L}(t, X, \mathbf{v} \circ X) = \nabla_b L(t, \mu, \mathbf{v}). \tag{2.11}$$

From this we deduce

$$0 = \langle \nabla_b L(t, \mu, \mathbf{v}), \mathbf{w}\rangle.$$

Since $\mathbf{w}$ is an arbitrary vector such that $\nabla \cdot (\mu \mathbf{w}) = 0$ in the sense of distributions, we conclude that $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M})$.

Conversely, suppose $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M})$. Let $\bar{\mathbf{v}} \in L^2(\mu)$ such that (2.7) holds. Then $\nabla \cdot (\mu(\bar{\mathbf{v}} - \mathbf{v})) = 0$ in the sense of distribution. We have

$$\bar{L}(t, X, \bar{\mathbf{v}} \circ X) \geq \bar{L}(t, X, \mathbf{v} \circ X) + \langle \xi \circ X, \bar{\mathbf{v}} \circ X - \mathbf{v} \circ X\rangle = \bar{L}(t, X, \mathbf{v} \circ X).$$

This proves that $\mathbf{v}$ minimizes (2.9). By Lemma 2.4, $\mathbf{v} \circ X = \mathrm{proj}_{\nabla F[X],\bar{L}}\zeta$. □

Remark 2.7. *Assume both $\bar{L}(\bar{t}, \bar{X}, \cdot)$ and $\bar{H}(\bar{t}, \bar{X}, \cdot)$ are strictly convex and of class C^1 for any $(\bar{t}, \bar{X}) \in [0, T] \times \mathbb{H}$. Let $(t, X, \zeta) \in [0, T] \times \mathbb{H} \times \mathbb{H}$. Using the notation in Definition 2.5 we have*

(i)
$$\bar{L}(t, X, \mathrm{proj}_{\nabla F[X],\bar{L}}\zeta) \leq \bar{L}(t, X, \zeta), \quad \bar{H}(t, X, \mathrm{proj}_{\nabla F[X],\bar{L}}\zeta) \leq \bar{H}(t, X, \zeta). \tag{2.12}$$

(ii) We do not know if we can replace $\mathrm{proj}_{\nabla F[X],\bar{L}}\zeta$ by $\mathrm{proj}_{\nabla F[X]}\zeta$ in (2.12), which is why we sometimes impose the condition in this form when necessary.

We recall the following result:

Proposition 2.8. *$X, X_0 \in \mathbb{H}$ have the same law if and only if for each positive integer n there exists $S_n \in \mathcal{G}(\Omega)$ such that*

$$\|X_0 \circ S_n - X\| \leq \frac{1}{n}. \tag{2.13}$$

Proof. Clearly (2.13) implies that X and X_0 have the same law. Conversely, if $X \in [X_0]_\sharp$, then Lemma 6.4 [9] implies (2.13). □

As a consequence, the following holds:

Corollary 2.9. *Let $X, X_0, \zeta \in \mathbb{H}$ and let $\{S_n\}_n$ be as in Proposition 2.8 such that $\{\bar{\zeta}_n\}_n := \{\zeta \circ S_n\}_n$ converges weakly to $\bar{\zeta}$ in $\mathbb{H}$. Let $\xi \in L^2(\mu)$ be uniquely defined by $\xi \circ X_0 := \mathrm{proj}_{F[X_0]}\zeta$, where $\mu := \sharp(X_0)$. Then $\xi \circ X = \mathrm{proj}_{F[X]}\bar{\zeta}$.*

Proof. Let $\bar{\xi}$ be uniquely defined by $\bar{\xi} \circ X := \mathrm{proj}_{F[X]}\bar{\zeta}$. For any $\phi \in C_c(\mathbb{M}; \mathbb{M})$ we have

$$\langle \xi, \phi\rangle_\mu = \langle \zeta, \phi \circ X_0\rangle = \langle \zeta \circ S_n, \phi \circ X_0 \circ S_n\rangle.$$

As $\{\zeta \circ S_n\}_n$ converges weakly to $\bar{\zeta}$ and $\{\phi \circ X_0 \circ S_n\}_n$ convergences strongly to $\phi \circ X$ we conclude that

$$\langle \xi, \phi\rangle_\mu = \langle \bar{\zeta}, \phi \circ X\rangle =: \langle \bar{\xi}, \phi\rangle_\mu.$$

Since ϕ is arbitrary, this concludes the proof of the Lemma. □

Corollary 2.10. *Let $E : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$ be continuous and satisfy the invariance property*

$$E(X, \zeta) = E(X \circ S, \zeta \circ S) \tag{2.14}$$

for any map $S \in \mathcal{S}(\Omega)$ and any $X, \zeta \in \mathbb{H}$. Then

$$E(X_0, \xi \circ X_0) = E(X, \xi \circ X)$$

for any $X_0, X \in \mathbb{H}$ such that $\sharp(X_0) = \sharp(X) =: \mu$ and any $\xi \in L^2(\mu)$.

Proof. Using the notation in Proposition 2.8, we choose $S_n \in \mathcal{G}(\Omega)$ such that (2.13) holds. By approximating ξ in $L^2(\mu)$ with functions in $C_c(\mathbb{M}; \mathbb{M})$ we get that $(\xi \circ X_0 \circ S_n)_n$ converges strongly to $\xi \circ X$ in $\mathbb{H}$. Thus, by the continuity of E and the invariance property (2.14) we get

$$E(X, \xi \circ X) = \lim_{n \to \infty} E(X_0 \circ S_n, \xi \circ X_0 \circ S_n) = E(X_0, \xi \circ X_0),$$

which concludes the proof. □

3. Differentiability of Rearrangement Invariant maps

Throughout this section, we assume that $U : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R} \cup \{\pm\infty\}$.

Definition 3.1. *Let $\mu \in \mathrm{dom}(U)$.*

(i) The weak–inf subgradient of U at $\mu \in \mathcal{P}_2(\mathbb{M})$ is the set of all $\xi \in L^2(\mu)$ such that

$$U(\nu) - U(\mu) \geq \inf_{\gamma \in \Gamma_o(\mu,\nu)} \int_{\mathbb{M} \times \mathbb{M}} \xi(x) \cdot (y - x) \gamma(dx, dx) + o(W_2(\mu, \nu)),$$

for $\nu \in \mathcal{P}_2(\mathbb{M})$. We denote this set by $\partial^-_{inf} U(\mu)$.

(ii) We define the weak-sup subgradient by replacing "inf" with "sup" in the inequality above. We denote this set by $\partial^-_{sup} U(\mu)$.

For $\mu, \nu \in \mathcal{P}_2(\mathbb{M})$ and $\gamma \in \Gamma(\mu, \nu)$, let be as defined in (2.2). We reformulate the above definition in terms of optimal bounds on $o(W_2(\mu, \nu))$. First, if $\mu \neq \nu$, we introduce the expressions

$$H^-(\nu, \xi) := \sup_{p \in \Gamma_0(\mu,\nu)} \frac{e_\mu(\nu, \xi, p)}{W_2(\mu, \nu)}, \quad H^+(\nu, \xi) := \inf_{p \in \Gamma_0(\mu,\nu)} \frac{e_\mu(\nu, \xi, p)}{W_2(\mu, \nu)}. \tag{3.1}$$

Secondly, for any $r > 0$ we set

$$h^-(r, \xi) := \inf_\nu \{H^-(\nu, \xi) \ : \ 0 < W_2(\mu, \nu) \leq r\},$$

$$h^+(r, \xi) := \inf_\nu \{H^+(\nu, \xi) \ : \ 0 < W_2(\mu, \nu) \leq r\}.$$

Remark 3.2. *For $\mu \in \mathcal{P}_2(\mathbb{M})$ and $\xi \in L^2(\mu)$ the following hold:*

(i) $\xi \in \partial^-_{inf} U(\mu)$ if and only if $\lim_{r \to 0^+} h^-(r, \xi) \geq 0$.
(ii) $\xi \in \partial^-_{sup} U(\mu)$ if and only if $\lim_{r \to 0^+} h^+(r, \xi) \geq 0$.
(iii) For any $\nu \in \mathcal{P}_2(\mathbb{M})$ we have $h^\pm(W_2(\mu, \nu), \xi) \leq H^\pm(\nu, \xi)$.

Lemma 3.3. *For $\mu, \nu \in \mathcal{P}_2(\mathbb{M})$, any $p, \hat{p} \in \Gamma(\mu, \nu)$ and $\zeta \in C^2_c(\mathbb{M})$, we have*

$$\Big| \int_{\mathbb{M} \times \mathbb{M}} \nabla \zeta(x) \cdot (y - x) [\hat{p}(dx, dy) - p(dx, dy)] \Big| \leq \frac{1}{2} \|\nabla^2 \zeta\|_\infty \Big[\|\pi^1 - \pi^2\|^2_p + \|\pi^1 - \pi^2\|^2_{\hat{p}} \Big].$$

Proof. Since $p, \hat{p} \in \Gamma(\mu, \nu)$,

$$\int_{\mathbb{M}\times\mathbb{M}} [\zeta(y) - \zeta(x)][\hat{p}(dx, dy) - p(dx, dy)] = 0. \tag{3.2}$$

For any $x, y \in \mathbb{M}$ there exists $r \in [-1, 1]$ depending on x and y such that

$$\zeta(y) - \zeta(x) - \nabla\zeta(x) \cdot (y - x) = \frac{r}{2}\|\nabla^2\zeta\|_\infty |x - y|^2. \tag{3.3}$$

Since

$$\Big|\int_{\mathbb{M}\times\mathbb{M}} r\,\|\nabla^2\zeta\|_\infty |y - x|^2\big(\hat{p}(dx, dy) - p(dx, dy)\big)\Big| \leq \|\nabla^2\zeta\|_\infty\Big(\|\pi^1 - \pi^2\|_p^2 + \|\pi^1 - \pi^2\|_{\hat{p}}^2\Big),$$

we use (3.2), (3.3) to conclude the proof of the Lemma. □

Proposition 3.4. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{M})$, $\xi \in L^2(\mu)$ and $p, \hat{p} \in \Gamma_o(\mu, \nu)$. If $\zeta \in C_c^2(\mathbb{M})$, then*

$$\Big|e_\mu(\nu, \xi, p) - e_\mu(\nu, \xi, \hat{p})\Big| \leq \Big(\|\nabla^2\zeta\|_\infty W_2(\mu, \nu) + 2\|\xi - \nabla\zeta\|_\mu\Big) W_2(\mu, \nu)$$

Proof. First, we use the decomposition

$$e_\mu(\nu, \xi, p) - e_\mu(\nu, \xi, \hat{p}) = \int_{\mathbb{M}\times\mathbb{M}} \nabla\zeta(x) \cdot (x - y)(\hat{p}(dx, dy) - p(dx, dy)) \tag{3.4}$$

$$+ \int_{\mathbb{M}\times\mathbb{M}} (\xi(x) - \nabla\zeta(x)) \cdot (x - y)(\hat{p}(dx, dy) - p(dx, dy)). \tag{3.5}$$

Next we apply Cauchy–Schwarz inequality to obtain

$$\Big|\int_{\mathbb{M}\times\mathbb{M}} (\xi(x) - \nabla\zeta(x)) \cdot (x - y)\hat{p}(dx, dy)\Big| \leq \|\xi - \nabla\zeta\|_\mu \|\pi^1 - \pi^2\|_{\hat{p}}. \tag{3.6}$$

Similarly,

$$\Big|\int_{\mathbb{M}\times\mathbb{M}} (\xi(x) - \nabla\zeta(x)) \cdot (x - y)p(dx, dy)\Big| \leq \|\xi - \nabla\zeta\|_\mu \|\pi^1 - \pi^2\|_p. \tag{3.7}$$

We use Lemma 3.3 to control the expression in the right hand side of (3.4) and combine this with (3.4-3.7) to obtain the thesis. □

Corollary 3.5. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{M})$, let $\xi \in L^2(\mu)$ and let $\zeta \in C_c^2(\mathbb{M})$. Then*

$$\Big|H^+(\nu, \xi) - H^-(\nu, \xi)\Big| \leq \|\nabla^2\zeta\|_\infty W_2(\mu, \nu) + 2\|\xi - \nabla\zeta\|_\mu.$$

Theorem 3.6. *For any $\mu \in \mathcal{P}_2(\mathbb{M})$ we have*

$$\partial^-_{inf} U(\mu) \cap \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}) = \partial^-_{sup} U(\mu) \cap \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}).$$

Proof. Let $\xi \in L^2(\mu)$. By Corollary 3.5, for any $\zeta \in C_c^2(\mathbb{M})$ we have

$$\begin{aligned} \inf_{0<W_2(\mu,\nu)\leq r} H^+(\nu, \xi) &\geq \inf_{0<W_2(\mu,\nu)\leq r}\Big(H^-(\nu, \xi) - \|\nabla^2\zeta\|\|_\infty W_2(\mu, \nu) - 2\|\xi - \nabla\zeta\|_\mu\Big) \\ &\geq \inf_{0<W_2(\mu,\nu)\leq r}\Big(H^-(\nu, \xi) - \|\nabla^2\zeta\|_\infty r - 2\|\xi - \nabla\zeta\|_\mu\Big). \end{aligned}$$

In terms of $h^\pm$, this reads

$$h^+(r, \xi) \geq h^-(r, \xi) - \|\nabla^2\zeta\|_\infty r - 2\|\xi - \nabla\zeta\|_\mu,$$

which, together with the fact that $h^+ \leq h^-$, implies

$$\lim_{r\to 0^+} h^-(r, \xi) \geq \lim_{r\to 0^+} h^+(r, \xi) \geq \lim_{r\to 0^+} h^-(r, \xi) - 2\|\xi - \nabla\zeta\|_\mu.$$

If, in addition, $\xi \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$, then we can choose ζ so that $\|\xi - \nabla\zeta\|_\mu$ is arbitrarily small and conclude that

$$\lim_{r\to 0^+} h^-(r,\xi) = \lim_{r\to 0^+} h^+(r,\xi).$$

We use Remark 3.2 to conclude the proof of the Theorem. □

If $\gamma \in \Gamma_o(\mu,\nu)$ then (cf. [3]) the barycentric projection of γ, of base μ, belongs to $\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$ and so,

$$\int_{\mathbb{M}\times\mathbb{M}} w(x)\cdot(y-x)\gamma(dx,dy) = 0$$

for any $w \in \left[\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})\right]^\perp$. Thus,

$$\partial^-_{inf}U(\mu)\cap\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}) + \left[\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})\right]^\perp = \partial^-_{inf}U(\mu) \tag{3.8}$$

and

$$\partial^-_{sup}U(\mu)\cap\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}) + \left[\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})\right]^\perp = \partial^-_{sup}U(\mu). \tag{3.9}$$

Thanks to Theorem 3.6, we can use (3.8) and (3.9) to obtain the following proposition.

Proposition 3.7. *If $U : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}\cup\{\pm\infty\}$ and $\mu \in \mathrm{dom}(U)$, then $\partial^-_{inf}U(\mu) = \partial^-_{sup}U(\mu)$. We define*

$$\partial^- U(\mu) := \partial^-_{inf}U(\mu) = \partial^-_{sup}U(\mu).$$

Definition 3.8. *(i) We define the weak-sup supergradient of a function $U : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ at $\mu \in \mathcal{P}_2(\mathbb{M})$ as the set of all $\xi \in L^2(\mu)$ such that*

$$U(\nu) - U(\mu) \le \sup_{\gamma\in\Gamma_o(\mu,\nu)} \int\!\!\int \xi(x)\cdot(y-x)\,\gamma(dx,dy) + o(W_2(\mu,\nu))$$

for $\nu \in \mathcal{P}_2(\mathbb{M})$. We denote this set by $\partial^+_{sup}U(\mu)$.

(ii) We define the weak-sup supergradient by replacing "sup" with "inf" in the inequality above. We denote this set by $\partial^+_{inf}U(\mu)$.

Since

$$\partial^+_{inf}U(\mu) = -\partial^-_{sup}(-U)(\mu) \quad \text{and} \quad \partial^+_{sup}U(\mu) = -\partial^-_{inf}(-U)(\mu),$$

Proposition 3.7 implies:

Proposition 3.9. *For any $U : \mathcal{P}_2(\mathbb{M}) \to [-\infty,\infty]$ and any $\mu \in \mathcal{P}_2(\mathbb{M})$ we have $\partial^+_{inf}U(\mu) = \partial^+_{sup}U(\mu)$, which will be denoted by $\partial^+U(\mu)$. Also, $\partial^+U(\mu) = -\partial^-(-U)(\mu)$.*

We define

$$\partial_\bullet U(\mu) := \partial^- U(\mu)\cap\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}) \quad \text{and} \quad \partial^\bullet U(\mu) := \partial^+ U(\mu)\cap\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$$

Theorem 3.10. *The set $\partial_\bullet U(\mu)\cap\partial^\bullet U(\mu)$ has at most one element.*

Proof. If $\xi,\ \zeta \in \partial^- U(\mu)\cap\partial^+U(\mu)\cap\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$, then for any sequence $\{\mu_n\}_n \subset \mathcal{P}_2(\mathbb{M})$ such that $W_2(\mu,\mu_n)\to 0$ and any sequence of plans $\{\gamma_n\}_n$ such that $\gamma_n \in \Gamma_o(\mu,\mu_n)$ for all n, we have

$$\lim_{n\to\infty} \frac{\int\!\!\int [\xi(x)-\zeta(x)]\cdot(y-x)\,\gamma_n(dx,dy)}{W_2(\mu,\mu_n)} = 0. \tag{3.10}$$

Choose $\mu_n := [\mathrm{Id} + n^{-1}\nabla\phi]_\sharp\mu$ for some non-identically zero $\phi \in C^1_c(\mathbb{R}^d)$. For sufficiently large n we have that $\Gamma_o(\mu,\mu_n) = \{[\mathrm{Id}\times(\mathrm{Id}+n^{-1}\nabla\phi)]_\sharp\mu\}$, so (3.10) yields $\langle\xi-\zeta,\nabla\phi\rangle_{L^2(\mu;\mathbb{R}^d)} = 0$. Since ϕ is arbitrary, we are done. □

Definition 3.11. *We say that $U : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R} \cup \{\pm\infty\}$ is differentiable at $\mu \in \mathrm{dom}(U)$ if $\partial^- U(\mu) \cap \partial^+ U(\mu) \neq \emptyset$. In this case, according to Theorem 3.10, there exists a unique $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M})$, which we now denote by $\nabla_w U(\mu)$, such that (using the notation in (2.2))*

$$\lim_{n\to\infty} \frac{e_\mu(\mu_n, \nabla_w U(\mu), \gamma)}{W_2(\mu, \mu_n)} = 0 \tag{3.11}$$

for any sequence $\{\mu_n\}_n \subset \mathcal{P}_2(\mathbb{M})$ such that $W_2(\mu, \mu_n) \to 0$ and any sequence of plans $\{\gamma_n\}_n$ such that $\gamma_n \in \Gamma_o(\mu, \mu_n)$ for all n.

Lasry and Lyons have introduced another, less intrinsic, notion of differentiability by associating to each function $U : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ a map $\tilde{U} : \mathbb{H} \to \mathbb{R}$ given by

$$\tilde{U}(X) := U(X_\sharp P),$$

where $(\Omega, \mathcal{B}(\Omega), P)$ is a non-atomic probability space. Each $\mu \in \mathcal{P}_2(\mathbb{M})$ corresponds to a random variable X on the probability space $(\Omega, \mathcal{B}(\Omega), P)$ via $\mu = X_\sharp P$. The range of the operator $U \mapsto \tilde{U}$ is the set of all functionals $V : \mathbb{H} \to \mathbb{R}$ such that

$$X,\ Y \in \mathbb{H},\quad X_\sharp P = Y_\sharp P \quad \Longrightarrow \quad V(X) = V(Y).$$

We are particular about this probability space, and recall that we have chosen Ω to be the ball in $\mathbb{R}^d$ of unit volume and centered at the origin, while $P := \mathcal{L}^d_\Omega$. This will give us access to the powerful tools of Optimal Transport; in particular, to any $\mu \in \mathcal{P}_2(\mathbb{M})$ there corresponds a unique random variable which is Lebesgue a.e. equal to the gradient of a convex function. This fact will be of crucial importance in the sequel. Also, recall that we have called functionals such as $\tilde{U}$ above *rearrangement invariant*, R.I. for short. In the sequel we shall denote by $\partial^\pm V(X)$ the super (sub, respectively) Frechét gradient at $X \in \mathbb{H}$. As in (3.1), for $X, Y, \zeta \in \mathbb{H}$ we define

$$\tilde{H}[X](Y, \zeta) := \frac{V(Y) - V(X) - \langle \zeta, Y - X \rangle}{\|Y - X\|} \tag{3.12}$$

and

$$\tilde{h}[X](r, \zeta) := \inf_Y \{\tilde{H}[X](Y, \zeta) \ :\ 0 < \|Y - X\| \le r\}.$$

Remark 3.12. *For $X, \zeta \in \mathbb{H}$ the following hold.*

(i) $\zeta \in \partial^- V(X)$ *if and only if* $\lim_{r\to 0^+} \tilde{h}[X](r, \zeta) \ge 0$.

(iii) For any $Y \in \mathbb{H}$ using the notation of $e[V]$ in (2.1) we have

$$e[V](X, \|Y - X\|, \zeta) \le \tilde{H}[X](Y, \zeta).$$

Lemma 3.13. *Let $X \in \mathbb{H}$, $\mu := \sharp(X)$, and let $\xi \in L^2(\mu)$. Let $\epsilon > 0$, $\nu \in \mathcal{P}_2(\mathbb{M})$ and let $\gamma \in \Gamma_o(\mu, \nu)$. Then there exists $Z \in \mathbb{H}$ such that the following hold:*

(i) $\nu = \sharp(Z)$;

(ii) $\|X - Z\| \le \epsilon + W_2(\mu, \nu)$;

(iii) For any $\phi \in C^1_c(\mathbb{M}; \mathbb{M})$ we have

$$\begin{aligned} &\left| \int_{\mathbb{M}\times\mathbb{M}} \xi(x) \cdot (y - x)(\gamma - p)(dx, dy) \right| \\ \le\ & W_2(\mu, \nu)\Big[(3 + \epsilon)\|\xi - \phi\|_\mu + \epsilon \|\nabla\phi\|_\infty W_2(\mu, \nu) + \epsilon \|\xi\|_\mu\Big], \end{aligned}$$

where $\nabla\phi$ denotes the Jacobian matrix of ϕ and we have set

$$p := \sharp(X \times Z) \in \Gamma(\mu, \nu). \tag{3.13}$$

Proof. Choose a random variable $A = (S, T) \in L^2(\Omega; \mathbb{R}^d \times \mathbb{R}^d)$ such that $\sharp(A) = \gamma$. Due to the marginal properties of γ, we have $\mu = \sharp(S)$ and $\nu = \sharp(T)$. Also, by the optimality of γ we have that

$$W_2(\mu, \nu) = \|S - T\|. \tag{3.14}$$

By Lemma 6.4 [9], there exists a measure preserving (preserves $\mathcal{L}^d_\Omega$) map τ such that

$$\|X - S \circ \tau\| \leq W_2(\mu, \nu)\epsilon. \tag{3.15}$$

Thus,

$$\|X - T \circ \tau\| \leq \|X - S \circ \tau\| + \|S \circ \tau - T \circ \tau\| \leq \epsilon W_2(\mu, \nu) + W_2(\mu, \nu). \tag{3.16}$$

Set

$$Z := T \circ \tau, \quad Y := S \circ \tau.$$

Note that (i) and (ii) are satisfied.

Let $\phi \in C^1_c(\mathbb{M}; \mathbb{M})$. We have

$$\begin{aligned}
\Big|\langle \xi \circ Y, Z - Y\rangle - \langle \xi \circ X, Z - X\rangle\Big| \quad &\leq \quad \Big|\langle \xi \circ Y - \phi \circ Y, Z - Y\rangle\Big| \\
&+ \quad \Big|\langle \phi \circ Y, Z - Y\rangle - \langle \phi \circ X, Z - X\rangle\Big| \\
&+ \quad \Big|\langle \phi \circ X - \xi \circ X, Z - X\rangle\Big| \\
&\leq \quad \Big|\langle \xi \circ Y - \phi \circ Y, Z - Y\rangle\Big| \\
&+ \quad \Big|\langle \phi \circ Y - \phi \circ X, Z - Y\rangle\Big| + \Big|\langle \phi \circ X, X - Y\rangle \\
&+ \quad \Big|\langle \phi \circ X - \xi \circ X, Z - X\rangle\Big|.
\end{aligned}$$

We conclude that

$$\begin{aligned}
\Big|\langle \xi \circ Y, Z - Y\rangle - \langle \xi \circ X, Z - X\rangle\Big| \quad &\leq \quad \|\xi - \phi\|_\mu \|Z - Y\| \\
&+ \quad \|\nabla\phi\|_\infty \|Y - X\| \|Z - Y\| + \|\phi\|_\mu \|Y - X\| \\
&+ \quad \|\xi - \phi\|_\mu \|Z - X\|.
\end{aligned}$$

This, together with (3.14 – 3.16) implies

$$\Big|\langle \xi \circ Y, Z - Y\rangle - \langle \xi \circ X, Z - X\rangle\Big| \leq W_2(\mu, \nu)\Big[(2 + \epsilon)\|\xi - \phi\|_\mu + \epsilon\|\nabla\phi\|_\infty W_2(\mu, \nu) + \epsilon\|\phi\|_\mu\Big].$$

We use the inequality $\|\phi\|_\mu \leq \|\xi - \phi\|_\mu + \|\xi\|_\mu$ to conclude the proof. □

The above lemma is useful for proving:

Theorem 3.14. *Let $X \in \mathbb{H}$, $\mu := \sharp(X)$ and $\xi \in L^2(\mu)$. Then:*

(i) $\xi \circ X \in \partial^- \tilde{U}(X)$ implies $\xi \in \partial^- U(\mu)$. The converse holds if $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M})$.
(ii) $\xi \circ X \in \partial^+ \tilde{U}(X)$ implies $\xi \in \partial^+ U(\mu)$. The converse holds if $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M})$.

Proof. We will only prove the first statement as one can obtain the second one by duality.

Part I. Assume $\xi \in L^2(\mu)$. Let $r > 0$ and $\nu \in \mathcal{P}_2(\mathbb{M})$ be such that $0 < W_2(\mu, \nu) \leq r$. Let $\gamma \in \Gamma_o(\mu, \nu)$. By Lemma 3.13, we can find $Z \in \mathbb{H}$ such that (i) – (iii) in the Lemma hold for $\epsilon = rW_2(\mu, \nu)$. In terms of the bistochastic measure p defined in (3.13), we have

$$e_\mu(\nu, \xi, \gamma) = \tilde{U}(Z) - \tilde{U}(X) - \langle \xi \circ X, Z - X\rangle - \int_{\mathbb{M}\times\mathbb{M}} \xi(x) \cdot (y - x)(\gamma - p)(dx, dy).$$

Thus, by Lemma 3.13 (iii), we have, for any $\phi \in C_c^1(\mathbb{M};\mathbb{M})$,

$$\frac{e_\mu(\nu,\xi,\gamma)}{W_2(\mu,\nu)} \geq \frac{\tilde{U}(Z)-\tilde{U}(X)-\langle \xi\circ X, Z-X\rangle}{W_2(\mu,\nu)} - (3+\epsilon)||\xi-\phi||_\mu - \epsilon||\nabla\phi||_\infty W_2(\mu,\nu) - \epsilon||\xi||_\mu,$$

and so,

$$\frac{e_\mu(\nu,\xi,\gamma)}{W_2(\mu,\nu)} \geq e[\tilde{U}](2r,X,\xi\circ X)\frac{||Z-X||}{W_2(\mu,\nu)} - (3+\epsilon)||\xi-\phi||_\mu - \epsilon||\nabla\phi||_\infty W_2(\mu,\nu) - \epsilon||\xi||_\mu.$$

If $\xi\circ X\in\partial^-\tilde{U}(X)$, then for $\delta>0$ we can find $r_0>0$ such that $e[\tilde{U}](2r,X,\xi\circ X)\geq -\delta$ for any $r\in(0,r_0)$. We conclude that for such r, due Lemma 3.13 (ii), we have

$$\begin{aligned}\frac{e_\mu(\nu,\xi,\gamma)}{W_2(\mu,\nu)} &\geq -\delta\frac{||Z-X||}{W_2(\mu,\nu)} - (3+\epsilon)||\xi-\phi||_\mu - r||\nabla\phi||_\infty W_2(\mu,\nu) - \epsilon||\xi||_\mu\\ &\geq -\delta(1+r)-(3+r^2)||\xi-\phi||_\mu - r^2||\nabla\phi||_\infty - r^2||\xi||_\mu.\end{aligned}$$

We first minimize over (ν,γ) to conclude that

$$h^+(r,\xi)\geq -\delta(1+r)-(3+r^2)||\xi-\phi||_\mu - r^2||\nabla\phi||_\infty - r^2||\xi||_\mu.$$

Hence,

$$\lim_{r\to 0^+} h^+(r,\xi)\geq -\delta - 3||\xi-\phi||_\mu.$$

Since $\delta>0$ and $\phi\in C_c^1(\mathbb{M};\mathbb{M})$ are arbitrary and $\xi\in L^2(\mu)$, we have $\lim_{r\to 0^+} h^+(r,\xi)\geq 0$. Hence, $\xi\in\partial^- U(\mu)$.

Part II. Conversely, assume $\xi\in\partial_\bullet U(\mu)$. Let $r>0$ and $Y\in\mathbb{H}$ be such that $0<||Y-X||\leq r$. Set

$$\nu := \sharp(Y),\quad \gamma:=\sharp(X\times Y),$$

and pick any $\gamma_o\in\Gamma_o(\mu,\nu)$ and any $\varphi\in C_c^1(\mathbb{M})$. We write the decomposition

$$\begin{aligned}\tilde{U}(Y)-\tilde{U}(X)-\langle\xi\circ X,Y-X\rangle &= U(\eta)-U(\mu)-\int_{\mathbb{M}\times\mathbb{M}}\xi(x)\cdot(y-x)\gamma_0(dx,dy)\\ &- \int_{\mathbb{M}\times\mathbb{M}}[\xi(x)-\nabla\varphi(x)]\cdot(y-x)(\gamma-\gamma_0)(dx,dy)\\ &- \int_{\mathbb{M}\times\mathbb{M}}\nabla\varphi(x)\cdot(y-x)(\gamma-\gamma_0)(dx,dy).\end{aligned}$$

We combine Lemma 3.3 with (3.6 – 3.7) to obtain

$$\begin{aligned}\tilde{U}(Y)-\tilde{U}(X)-\langle\xi\circ X,Y-X\rangle &\geq U(\eta)-U(\mu)-\int_{\mathbb{M}\times\mathbb{M}}\xi(x)\cdot(y-x)\gamma_0(dx,dy)\\ &- ||\xi-\nabla\varphi||_\mu\Big(||Y-X||+W_2(\mu,\nu)\Big)\\ &- \frac{1}{2}||\nabla^2\varphi||_\infty\Big(||Y-X||^2+W_2^2(\mu,\nu)\Big).\end{aligned}$$

We divide the above identities by $||Y-X||$ and use the fact that $W_2(\mu,\nu)\leq ||Y-X||\leq r$ to obtain

$$\frac{\tilde{U}(Y)-\tilde{U}(X)-\langle\xi\circ X,Y-X\rangle}{||Y-X||}\geq h^+(r,\xi)\frac{W_2(\mu,\nu)}{||Y-X||} - 2||\xi-\nabla\varphi||_\mu - r||\nabla^2\varphi||_\infty$$

For every $\epsilon>0$, there is $r_0>0$ such that for $r\in(0,r_0)$, $h^+(r,\xi)\geq-\epsilon$. For such r we have

$$\frac{\tilde{U}(Y)-\tilde{U}(X)-\langle\xi\circ X,Y-X\rangle}{||Y-X||}\geq -\epsilon - 2||\xi-\nabla\varphi||_\mu - r||\nabla^2\varphi||_\infty.$$

We use the fact that Y is arbitrary to conclude that

$$e[\tilde{U}](r, X, \xi \circ X) \geq -\epsilon - 2\|\xi - \nabla\varphi\|_\mu - r\|\nabla^2\varphi\|_\infty,$$

and so,

$$\lim_{r \to 0^+} e[\tilde{U}](2r, X, \xi \circ X) \geq -\epsilon - 2\|\xi - \nabla\varphi\|_\mu.$$

By the fact that $\epsilon > 0$, $\varphi \in C_c^1(\mathbb{M})$ are arbitrary and $\xi \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$, we conclude that $\lim_{r\to 0^+} e[\tilde{U}](r, X, \xi \circ X) \geq 0$, i.e. $\xi \circ X \in \partial^- \tilde{U}(X)$. □

As an immediate consequence of the above theorem and the definitions of $\partial_\bullet$ and $\partial^\bullet$, we deduce:

Corollary 3.15. *Let $X \in \mathbb{H}$, $\mu := \sharp(X)$ and $\xi \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$. Then the following hold:*

(i) $\xi \in \partial_\bullet U(\mu)$ if and only if $\xi \circ X \in \partial^- \tilde{U}(X)$;
(ii) $\xi \in \partial^\bullet U(\mu)$ if and only if $\xi \circ X \in \partial^+ \tilde{U}(X)$.

3.1. Subdifferential of R.I. functions

If $X \in \mathbb{H}$ is a Borel map, we denote $\mu := \sharp(X)$. Then we know $\mathcal{L}^d_\Omega$ disintegrates with respect to μ as

$$\int_\Omega \xi(x)\, dx = \int_{\mathbb{R}^d} \int_{X^{-1}(y)} \xi(z)\, \mu_y(dz)\, \mu(dy), \tag{3.17}$$

where $\{\mu_y\}_{y \in \mathbb{R}^d}$ is a family of Borel probability measures such that:
(i) $\mathbb{R}^d \ni y \mapsto \mu_y(B)$ is Borel for any Borel $B \subset \Omega$;
(ii) $\mu_y(\Omega \backslash X^{-1}(y)) = 0$ for μ-a.e. $y \in \mathbb{R}^d$.
Using $\mu = X_\sharp \mathcal{L}^d_\Omega$, we rewrite (3.17) as

$$\int_\Omega \xi(x)\, dx = \int_\Omega \int_{X^{-1}(X(x))} \xi(z)\, \mu_{X(x)}(dz)\, dx. \tag{3.18}$$

We denote $\nu_x := \mu_{X(x)}$ and note that (i) above, along with the Borel measurability of X, implies $\Omega \ni x \mapsto \nu_x(B)$ is a Borel map for any Borel set $B \subset \Omega$ (as a composition of Borel maps). Note also that $\nu_x \equiv \nu_z$ for all $z \in \Omega$ such that $X(z) = X(x)$ (or, equivalently, for all $z \in X^{-1}(X(x))$).

For any integer $m \geq 1$ denote by $\mathbb{B}_m$ the open ball centered at the origin in $\mathbb{R}^m$ such that $\mathcal{L}^m(\mathbb{B}_m) = 1$. Theorem 3.16 below is a collection of deep results, proved in [8].

Theorem 3.16. *Let $\Phi : \Omega \to \mathbb{R}$ be convex, set $X := \nabla\Phi$ and let* (3.18) *be the disintegration of $\mathcal{L}^d_\Omega$ with respect to the level sets of X. Then:*
(i) For every $x \in \Omega$ at which Φ is differentiable $X^{-1}(X(x))$ is a convex set of Hausdorff dimension $k(x) \in \{0, 1, ..., d\}$.
(ii) If for each $k \in \{0, 1, ..., d\}$ we denote

$$\Omega_k := \{x \in \Omega \ :\ X^{-1}(X(x)) \text{ has Hausdorff dimension } k\},$$

we have that Ω_k is a Borel set for each $k \in \{0, 1, ..., d\}$.
(iii) For $\mathcal{L}^d$-a.e. $x \in \Omega$ at which Φ is differentiable we have that

$$\mathcal{H}^{k(x)} \ll \nu_x \ll \mathcal{H}^{k(x)}.$$

Moreover, for each $k \in \{1, ..., d\}$ for which $\Omega_k \neq \emptyset$ there exist Borel maps

$$\sigma_k : \mathbb{B}_k \times \Omega_k \to \Omega_k \quad \text{and} \quad \alpha_k : \mathbb{B}_k \times \Omega_k \to [0, \infty)$$

such that $\sigma_k(\cdot,x)$ is invertible, $(z,x)\mapsto\sigma_k(\cdot,x)^{-1}(z)$ is Borel, and $\alpha_k(\cdot,x)$ is a positive probability density on $X^{-1}(X(x))$ for all $x\in\Omega_k$. Furthermore,

$$\int_{X^{-1}(X(x))}\varphi(z)\nu_x(dz)=\int_{\mathbb{B}_k}\varphi(\sigma_k(s,x))\alpha_k(s,x)\,ds\ \textit{for all}\ \varphi\in C(\bar\Omega),\tag{3.19}$$

and

$$\int_{X^{-1}(X(x))}\psi(\sigma_k(\cdot,x)^{-1}(z))\nu_x(dz)=\int_{\mathbb{B}_k}\psi(s)\alpha_k(s,x)\,ds\ \textit{for all}\ \psi\in C(\bar{\mathbb{B}}_k).\tag{3.20}$$

Note that the map $\sigma_k(\cdot,x)$ is a reparametrization of the k-dimensional convex set $X^{-1}(X(x))$, which maps $\mathbb{B}_k$ onto $X^{-1}(X(x))$ and pushes $\alpha_k(\cdot,x)$ forward to $\nu_{X(x)}$. For the Borel measurability of σ_k and α_k see Proposition 4.17 and Theorem 4.18 [8] (notation is different from ours).

Theorem 3.17. *Let $\Phi:\Omega\to\mathbb{R}$ be convex such that $\nabla\Phi=:X\in\mathbb{H}$. Let $\tilde U:\mathbb{H}\to\mathbb{R}$ be R.I. and assume there exists a Borel map $\zeta\in\mathbb{H}$ such that $\zeta\in\partial^-\tilde U(X)$. Then*

$$\mathrm{proj}_{F[X]}\zeta\in\partial^-\tilde U(X)\ \textit{and}\ \mathrm{proj}_{\nabla F[X]}\zeta\in\partial^-\tilde U(X).$$

Proof. Let $\{X_n\}_n\subset\mathbb{H}$ be a sequence of Borel maps such that $0<\|X_n\|\to 0$. For any integer $n\ge 2$ let $\{B_n^m\}_m$ be a countable partition of Borel subsets of $\mathbb{R}^d$ of diameter at most $\|X_n\|^2$. Fix $1\le k\le d$ and $x\in\Omega_k$. Let $A_k(\cdot,x)$ be the optimal map that pushes $\mathcal{L}^k|_{\mathbb{B}_k}$ forward to $\alpha_k(\cdot,x)\mathcal{L}^k|_{\mathbb{B}_k}$; we also know that its inverse $A_k(\cdot,x)^{-1}$ pushes $\alpha_k(\cdot,x)\mathcal{L}^k|_{\mathbb{B}_k}$ forward to $\mathcal{L}^k|_{\mathbb{B}_k}$. Since $(s,x)\mapsto\alpha_k(s,x)$ is Borel, we use Appendix 9, Corollary 9.8 to conclude that, after possibly redefining them on negligible sets, the maps $(s,x)\mapsto A_k(s,x)$ and $(s,x)\mapsto A_k(\cdot,x)^{-1}(s)$ are also Borel. So, the map $(s,x)\mapsto\sigma_k(A_k(s,x),x)=:\Lambda_k(s,x)$ is a Borel map with the property $\Lambda_k(\cdot,x)_\sharp\mathcal{L}^k|_{\mathbb{B}_k}=\nu_x$, while $\Lambda_k(\cdot,x)^{-1}(z):=A_k(\cdot,x)^{-1}\circ\sigma_k(\cdot,x)^{-1}(z)$ is a Borel map such that $\Lambda_k(\cdot,x)^{-1}_\sharp\nu_x=\mathcal{L}^k|_{\mathbb{B}_k}$. Let $G_k:\mathbb{B}_k\to\mathbb{B}_k\times\mathbb{B}_k$ be invertible such that $G_k,\ G_k^{-1}$ are Borel maps and $G_{k\sharp}\mathcal{L}^k|_{\mathbb{B}_k}=\mathcal{L}^k|_{\mathbb{B}_k}\otimes\mathcal{L}^k|_{\mathbb{B}_k}$. Let

$$E_k(z,x):=(\Lambda_k(G_k^1(\Lambda_k(\cdot,x)^{-1}(z)),x),\Lambda_k(G_k^2(\Lambda_k(\cdot,x)^{-1}(z)),x)).$$

Note that the maps

$$S_{k,n}(z,x):=X_n(E_k^1(z,x)),\ T_{k,n}(z,x):=\zeta(E_k^2(z,x))$$

are Borel and satisfy

$$[S_{k,n}(\cdot,x)\times T_{k,n}(\cdot,x)]_\sharp\nu_x=\vartheta_{n,x}\otimes\eta_x=:\gamma_{n,x},$$

where $\vartheta_{n,x}:=X_{n\sharp}\nu_x$ and $\eta_x:=\zeta_\sharp\nu_x$. This implies

$$\begin{aligned}\int_{X^{-1}(X(x))}X_n(z)\nu_x(dz)\cdot\int_{X^{-1}(X(x))}\zeta(z)\nu_x(dz)&=\int_{\mathbb{M}\times\mathbb{M}}x_1\cdot x_2\,\gamma_{n,x}(dx_1,dx_2)\\&=\langle S_{k,n}(\cdot,x),T_{k,n}(\cdot,x)\rangle_{\nu_x}.\end{aligned}\tag{3.21}$$

We have that $T_{k,n}(\cdot,x)_\sharp\nu_x=\zeta_\sharp\nu_x$ implies

$$\mathcal{L}^k(\Lambda_k(\cdot,x)^{-1}(\zeta^{-1}(B_n^m)))=\mathcal{L}^k((G_k^2)^{-1}\circ\Lambda_k(\cdot,x)^{-1}(\zeta^{-1}(B_n^m)))=:r.$$

We restrict our attention to the set M of all m for which $r>0$ and we consider the optimal map $\tilde\tau_{n,m}^k(\cdot,x)$ which pushes $\mathcal{L}^k|_{\Lambda_k(\cdot,x)^{-1}(\zeta^{-1}(B_n^m))}$ forward to $\mathcal{L}^k|_{(G_k^2)^{-1}\circ\Lambda_k(\cdot,x)^{-1}(\zeta^{-1}(B_n^m))}$. Since $(z,x)\mapsto\Lambda_k(\cdot,x)^{-1}(z)$ and $(z,x)\mapsto(G_k^2)^{-1}\circ\Lambda_k(\cdot,x)^{-1}(z)$ are Borel maps and $\zeta^{-1}(B_n^m)$ is a Borel set, it follows by Appendix 9, Corollary 9.9 that $(z,x)\mapsto\tilde\tau_{n,m}^k(z,x)$ can also be taken to be Borel, which further implies $(z,x)\mapsto\tilde\tau_{n,m}^k(\cdot,x)\circ\Lambda_k(\cdot,x)^{-1}(z)=:\tau_{n,m}^k(z,x)$ is a Borel map. This map satisfies $\tau_{n,m}^k(\cdot,x)_\sharp\nu_x|_{T_{k,n}(\cdot,x)^{-1}(B_n^m)}=\nu_x|_{\zeta^{-1}(B_n^m)}$.

Let $\tau_n^k(\cdot,x):=\sum_{m\in M}\tau_{n,m}^k(\cdot,x)\mathbf{1}_{T_{k,n}(\cdot,x)^{-1}(B_n^m)}$ defined on $\Omega_k\times\Omega_k$ to see that

$$\tau_n^k(\cdot,x)_\sharp\nu_x=\nu_x\quad\text{and}\quad\|T_{k,n}(\cdot,x)\circ\tau_n^k(\cdot,x)-\zeta\|_{\nu_x}\le\|X_n\|^2.$$

This yields

$$\langle \bar{Z}_{k,n}(\cdot,x),\zeta\rangle_{\nu_x} \geq \langle S_{k,n}(\cdot,x),T_{k,n}(\cdot,x)\rangle_{\nu_x} - \|X_n\|^2\|X_n\|_{\nu_x},$$

if we set $\bar{Z}_{k,n}(z,x) := S_{k,n}(\tau_n^k(z,x),x)$ for $z \in \Omega_k$ and $x \in \Omega_k$. By (3.21), we infer

$$\int_{X^{-1}(X(x))} X_n(z)\nu_x(dz)\cdot\int_{X^{-1}(X(x))}\zeta(z)\nu_x(dz) \leq \langle \bar{Z}_{k,n}(\cdot,x),\zeta\rangle_{\nu_x} + \|X_n\|^2\|X_n\|_{\nu_x} \tag{3.22}$$

for every $x \in \Omega_k$. Let $\tilde{Z}_{k,n}(\cdot,x)$ denote the extension by zero of $\bar{Z}_{k,n}(\cdot,x)$ outside $X^{-1}(X(x))$. Next let us define $Z_{k,n} : \Omega \to \mathbb{R}^d$ by $Z_{k,n}(x) := \tilde{Z}_{k,n}(x,x)$ if $x \in \Omega_k$ and $Z_{k,n}(x) := 0$ otherwise. Since $(z,x) \mapsto \tilde{Z}_{k,n}(z,x)$ is a Borel map from $\Omega_k \times \Omega_k$ into $\mathbb{R}^d$, we deduce $Z_{k,n}$ is a Borel map from Ω into $\mathbb{R}^d$. Let $Z_{0,n} := X_n\mathbf{1}_{\Omega_0}$ (a Borel map) and define $Z_n : \Omega \to \mathbb{R}^d$ by $Z_n := \sum_{k=0}^d Z_{k,n}$. Thus, Z_n is Borel. In fact, $Z_n \in \mathbb{H}$ and the property $\tilde{Z}_{k,n}(z,z) = \tilde{Z}_{k,n}(z,x)$ for all $x \in \Omega_k$ and all $z \in X^{-1}(X(x)) \cap \Omega_k$ ensures that

$$Z_{n\sharp}\mathcal{L}^d_\Omega = X_{n\sharp}\mathcal{L}^d_\Omega \quad \text{and} \quad (X+Z_n)_\sharp\mathcal{L}^d_\Omega = (X+X_n)_\sharp\mathcal{L}^d_\Omega$$

for all n. So, by the rearrangement invariance of $\tilde{U}$ and the fact that $\zeta \in \partial^-\tilde{U}(X)$ we have

$$\liminf_{n\to\infty}\frac{\tilde{U}(X+X_n)-\tilde{U}(X)-\langle\zeta,Z_n\rangle}{\|X_n\|} \geq 0. \tag{3.23}$$

Next let

$$\xi(y) := \int_{X^{-1}(y)}\zeta(z)\mu_y(dz) \text{ for } \mu-\text{a.e. } y \in \mathbb{M}$$

so that

$$\xi\circ X(x) := \int_{X^{-1}(X(x))}\zeta(z)\nu_x(dz) \text{ for } \mathcal{L}^d-\text{a.e. } x \in \Omega.$$

Note that $\langle \xi\circ X - \zeta, \varphi\circ X\rangle = 0$ for all $\varphi \in C_c(\mathbb{M};\mathbb{M})$, which is equivalent to the fact that $\xi\circ X = \mathrm{proj}_{F[X]}\zeta$. By (3.22) we get, after integrating in x with respect to the measure $\mathcal{L}^d|_\Omega$ and using the fact that

$$\begin{aligned}\int_\Omega \|X_n\|_{\nu_x}dx &< 1+\int_\Omega\|X_n\|^2_{\nu_x}dx\\ &= 1+\int_\Omega\int_{X^{-1}(X(x))}|X_n(z)|^2\nu_x(dz)dx\\ &= 1+\|X_n\|^2,\end{aligned}$$

the inequality

$$\langle\zeta,Z_n\rangle + \|X_n\|^2(1+\|X_n\|^2) \geq \langle\xi\circ X, X_n\rangle,$$

which, in light of (3.23), yields

$$\mathrm{proj}_{F[X]}\zeta = \xi\circ X \in \partial^-\tilde{U}(X), \tag{3.24}$$

which, by Theorem 3.14, is equivalent to $\xi \in \partial^-U(\mu)$ (where $\mu := \sharp(X)$). This yields

$$\bar{\xi} := \mathrm{proj}_{\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})}\xi \in \partial_\bullet U(\mu) \subset \partial^-U(\mu).$$

Theorem 3.14 now implies $\bar{\xi}\circ X \in \partial^-\tilde{U}(X)$. However, since

$$\langle\xi-\bar{\xi},f\rangle_\mu = 0 \text{ for all } f \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}),$$

we conclude

$$\langle\zeta, f\circ X\rangle = \langle\bar{\xi},f\rangle_\mu \text{ for all } f \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}).$$

But $\bar{\xi} \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$, so (2.7) and Proposition 2.6 yield $\bar{\xi}\circ X = \mathrm{proj}_{\nabla F[X]}\zeta$. The proof is completed. □

Remark 3.18. *The Borel measurability of the map Z_n in the above proof was obtained from the (joint) Borel measurability of the maps A_k, Λ_k etc. This is the object of Appendix 9. In fact, there we proved something more general, which may have its own appeal to the reader; we showed that if one has two one-parameter families of probability densities which are Borel measurable (jointly with respect to their variables and the parameter), then the optimal maps between these densities are jointly Borel measurable as well.*

Theorem 3.19. *Let X, $X_0 \in \mathbb{H}$ be such that $X \in [X_0]_\sharp$ (recall that this means X and X_0 have the same law). Let $\{S_n\}_n \subset \mathcal{G}(\Omega)$ be a sequence as given by Proposition 2.8. If $\tilde{U} : \mathbb{H} \to \mathbb{R}$ is R.I., then the following hold:*

(i) If $\zeta \in \partial^- \tilde{U}(X_0)$, then every point of accumulation $\bar{\zeta}$ of $\{\zeta \circ S_n\}_n$ for the weak topology satisfies $\bar{\zeta} \in \partial^- \tilde{U}(X)$.
(ii) $\partial^- \tilde{U}(X_0) \neq \emptyset$ if and only if $\partial^- \tilde{U}(X) \neq \emptyset$.
(iii) If $\zeta \in \partial^- \tilde{U}(X_0)$, then $\mathrm{proj}_{F[X_0]}\zeta \in \partial^- \tilde{U}(X_0)$ and $\mathrm{proj}_{\nabla F[X_0]}\zeta \in \partial^- \tilde{U}(X_0)$.

Proof. Since we can replace $\|X_0 \circ S_n - X\|$ by $\|X \circ S_n^{-1} - X_0\|$ in (2.13), (i) readily implies (ii).

Thus, it is enough to prove (i). For that, fix $r > 0$ and pick $H \in \mathbb{H}$ arbitrary such that $0 < \|H\| \leq r$. Set

$$H_n := X \circ S_n^{-1} + H \circ S_n^{-1} - X_0$$

so that

$$H_n \circ S_n := X + H - X_0 \circ S_n. \tag{3.25}$$

We have

$$\tilde{U}(X + H) - \tilde{U}(X) = \tilde{U}(X \circ S_n^{-1} + h \circ S_n^{-1}) - \tilde{U}(X) = \tilde{U}(X_0 + H_n) - \tilde{U}(X_0),$$

and so,

$$\tilde{U}(X + H) - \tilde{U}(X) - \langle \zeta_n, H\rangle = \tilde{U}(X_0 + H_n) - \tilde{U}(X_0) - \langle \zeta, H_n - (X - S_n^{-1} - X_0)\rangle.$$

Thus,

$$\tilde{U}(X + H) - \tilde{U}(X) - \langle \zeta_n, H\rangle \geq e[\tilde{U}]\Big(\|H_n - (X - S_n^{-1} - X_0)\|, \zeta\Big). \tag{3.26}$$

Observe that

$$\|H_n\|^2 \leq 2\Big(\|X \circ S_n^{-1} - X_0\|^2 + \|H \circ S_n^{-1}\|^2\Big)$$

and

$$\|X \circ S_n^{-1} - X_0\| \leq \frac{1}{n}, \quad \|H \circ S_n^{-1}\| = \|H\|.$$

Thus, for $n > r^{-1}$ we have

$$\|H_n\|^2 \leq 2\Big(\frac{1}{n^2} + \|H\|^2\Big) \leq 4r^2.$$

This, together with (3.26) yields

$$\tilde{U}(X + H) - \tilde{U}(X) - \langle \zeta_n, H\rangle \geq e[\tilde{U}](2r, \zeta). \tag{3.27}$$

Since $\|\zeta \circ S_n\| = \|\zeta\|$ for all n, we may assume, without loss of generality, that the sequence

$$\{\zeta_n := \zeta \circ S_n\}_n \rightharpoonup \bar{\zeta} \text{ weakly in } \mathbb{H}. \tag{3.28}$$

By letting n tend to ∞ in (3.27) we obtain

$$\tilde{U}(X + H) - \tilde{U}(X) - \langle \bar{\zeta}, h\rangle \geq e[\tilde{U}](2r, \zeta).$$

Consequently, $e[\tilde{U}](r, \bar{\zeta}) \geq e[\tilde{U}](2r, \zeta)$. This proves that if $\zeta \in \partial^- \tilde{U}(X_0)$, then $\bar{\zeta} \in \partial^- \tilde{U}(X)$.

To prove (iii) let $\mu := \sharp(X_0)$ and set $X := \mathcal{M}(X_0)$, where $\mathcal{M}(X_0)$ pushes $\mathcal{L}^d_\Omega$ forward to μ and is the gradient of a real valued convex function defined on Ω. By (ii), $\bar{\zeta} \in \partial^- \tilde{U}(\mathcal{M}(X_0))$, which, in light of Corollary 2.9 and Theorem 3.17, implies

$$\xi \circ \mathcal{M}(X_0) = \mathrm{proj}_{F[\mathcal{M}(X_0)]}\bar{\zeta} \in \partial^- \tilde{U}(\mathcal{M}(X_0)) \text{ and } \xi \in \partial^- U(\mu).$$

Theorem 3.14 and Corollary 2.9 then show $\mathrm{proj}_{F[X_0]}\zeta := \xi \circ X_0 \in \partial^- \tilde{U}(X_0)$, and, as in the proof of Theorem 3.17, this further implies that $\mathrm{proj}_{\nabla F[X_0]}\zeta \in \partial^- \tilde{U}(X_0)$. □

It is natural to ask whether any element in the sub or superdifferential of $\tilde{U}$ at X is in $F[X]$ (so that $\zeta = \text{proj}_{F[X]}\zeta$). The following example shows that that is not the case.

Example 3.20. *Let $d = 1$ and $X_0 \equiv 0$ in $\Omega = [0, 1]$. Define the map $\tilde{U} : \mathbb{H} \to \mathbb{R}$ by $\tilde{U}(X) := \langle \text{Id}, \mathcal{M}(X)\rangle$, so that $\tilde{U}(X_0) = 0$. Since*

$$||\text{Id} - \mathcal{M}(X)|| = W_2(\mathcal{L}^d_\Omega, \sharp(X)) \leq ||\text{Id} - X||,$$

we deduce

$$\tilde{U}(X) - \tilde{U}(X_0) \geq \langle \text{Id}, X - X_0\rangle \text{ for all } X \in \mathbb{H}.$$

Thus, Id $\in \partial^- \tilde{U}(X_0)$ *even though* Id $\notin F[X_0]$ *(as $F[X_0]$ is the subspace of $L^2(\Omega)$ consisting of functions which are Lebesgue a.e. equal to constant functions).*

We combine Theorems 3.14 and 3.19 to conclude:

Theorem 3.21. *Let $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and set $\tilde{U} := U \circ \sharp$. Then the following are equivalent:*

(i) $\partial^\pm U(\mu) \neq \emptyset$;

(ii) $\partial^\pm \tilde{U}(X) \neq \emptyset$ for some *$X \in \mathbb{H}$ such that $\mu = \sharp(X)$;*

(iii) $\partial^\pm \tilde{U}(X) \neq \emptyset$ for all *$X \in \mathbb{H}$ such that $\mu = \sharp(X)$.*

Corollary 3.22. *Let $X \in \mathbb{H}$ and let $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. Set $\mu = \sharp(X)$ and set $\tilde{U} := U \circ \sharp$. If ξ is the element of minimal norm of $\partial^\pm U(\mu)$ and ζ is the element of minimal norm of $\partial^\pm \tilde{U}(X)$ then $\zeta = \xi \circ X$. In particular, U is differentiable at μ if and only if $\tilde{U}$ is differentiable at X. In this case, $\nabla_{L^2} \tilde{U}(X) = \nabla_w U(\mu_0) \circ X$.*

Proof. It suffices to prove the corollary in the case of the subdifferential. Suppose ξ is the element of minimal norm in $\partial^- U(\mu)$ (which must necessarily lie in $\partial_\bullet U(\mu)$) and ζ is the element of minimal norm in $\partial^- \tilde{U}(X)$ Since, by Theorem 3.19, $\text{proj}_{\nabla F[X]}\zeta$ belongs to $\partial^- \tilde{U}(X)$, we obtain $\text{proj}_{\nabla F[X]}\zeta = \zeta$ and so, there exists $\bar{\xi} \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{M})$ such that $\zeta = \bar{\xi} \circ X$. But $\xi \in \partial^- U(\mu)$ implies, by Theorem 3.14, $\xi \circ X \in \partial^- \tilde{U}(X)$, and so the minimality property of ζ implies

$$||\bar{\xi}||_\mu = ||\zeta|| \leq ||\xi \circ X|| = ||\xi||_\mu \tag{3.29}$$

with a strict inequality unless $\zeta \equiv \xi \circ X$, $\mathcal{L}^d_\Omega$ a.e. Since $\bar{\xi} \circ X = \zeta \in \partial_\bullet \tilde{U}(X)$, we use Theorem 3.14 again to see that $\bar{\xi} \in \partial^- U(\mu)$ and so, by the norm-minimality property of ξ

$$||\xi||_\mu \leq ||\bar{\xi}||_\mu = ||\zeta|| \tag{3.30}$$

with a strict inequality unless $\xi \equiv \bar{\xi}$, μ-a.e. We combine (3.29) and (3.30) to infer $\zeta \equiv \xi \circ X$, $\mathcal{L}^d_\Omega$ a.e.

Recall that U is differentiable at μ if and only if both sets $\partial^\pm U(\mu)$ are nonempty. This is equivalent to saying that both sets $\partial^\pm \tilde{U}(X)$ are nonempty, which in turn is equivalent to saying that $\tilde{U}$ is differentiable at X. The identity $\nabla_{L^2} \tilde{U}(X) = \nabla_w U(\mu) \circ X$ follows. □

4. Hamilton-Jacobi equations in the Wasserstein space

Throughout this section, we are given $H : [0, T] \times \cup_{\mu \in \mathcal{P}_2(\mathbb{M})}\{\{\mu\} \times \mathbb{R} \times L^2(\mu)\} \to \mathbb{R}$. We define $\tilde{H} : [0, T] \times \mathbb{H} \times \mathbb{R} \times \mathbb{H} \to \mathbb{R}$ by

$$\tilde{H}(t, X, r, \zeta) := H(t, \sharp(X), r, \xi), \text{ where } \xi \circ X = \text{proj}_{\nabla F[X]}\zeta. \tag{4.1}$$

One important invariance property of $\tilde{H}$ is stated below.

Lemma 4.1. *For every $S \in \mathcal{S}(\Omega)$ we have*

$$\tilde{H}(t, X \circ S, r, \zeta \circ S) = \tilde{H}(t, X, r, \zeta) \tag{4.2}$$

for all $(t, X, r, \zeta) \in [0, T] \times \mathbb{H} \times \mathbb{R} \times \mathbb{H}$.

Proof. It is easy to see that $\mathrm{proj}_{\nabla F[X]}\zeta = \mathrm{proj}_{\nabla F[X\circ S]}(\zeta \circ S)$, which yields the desired thesis in view of (4.1). □

Consider the Hamilton-Jacobi equation

$$\partial_t U(t,\mu) + H(t,\mu,U(t,\mu),\nabla_w U(t,\mu)) = 0 \text{ for } (t,\mu) \in [0,T)\times\mathcal{P}_2(\mathbb{M}), \tag{4.3}$$

together with its counterpart in $\mathbb{H}$

$$\partial_t \tilde{U}(t,X) + \tilde{H}(t,X,\tilde{U}(t,X),\nabla\tilde{U}(t,X)) = 0 \text{ for } (t,X) \in [0,T)\times\mathbb{H}. \tag{4.4}$$

4.1. Definition for time dependent HJE

Definition 4.2. *Let $U : [0,T)\times\mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ be locally bounded and $U_0 : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$.*

(1) We say that U is a viscosity subsolution *for* (4.3) *with initial data U_0 if U is upper semicontinuous and*

$$U(0,\cdot) \le U_0 \text{ in } \mathcal{P}_2(\mathbb{M}) \text{ and } \theta + H(t,\mu,U(t,\mu),\xi) \le 0. \tag{4.5}$$

for any $(t,\mu) \in [0,T)\times\mathcal{P}_2(\mathbb{M})$ and $(\theta,\xi) \in \partial^\bullet U(t,\mu)$.

(2) We say that U is a viscosity supersolution *for* (4.3) *with initial data U_0 if U is lower semicontinuous and*

$$U(0,\cdot) \ge U_0 \text{ in } \mathcal{P}_2(\mathbb{M}) \text{ and } \theta + H(t,\mu,U(t,\mu),\xi) \ge 0 \tag{4.6}$$

for any $(t,\mu) \in [0,T)\times\mathcal{P}_2(\mathbb{M})$ and $(\theta,\xi) \in \partial_\bullet U(t,\mu)$.

(3) We say that U is a viscosity solution *for* (4.3) *with initial data $U_0 : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ if it is both a viscosity subsolution and supersolution.*

Similarly, the corresponding objects are defined in $\mathbb{H}$:

Definition 4.3. *Let $\tilde{U} : [0,T)\times\mathbb{H} \to \mathbb{R}$ be locally bounded and $\tilde{U}_0 : \mathbb{H} \to \mathbb{R}$.*

(1) We say that $\tilde{U}$ is a viscosity subsolution *for* (4.4) *with initial data $\tilde{U}_0 : \mathbb{H} \to \mathbb{R}$ if $\tilde{U}$ is upper semicontinuous and*

$$\tilde{U}(0,\cdot) \le \tilde{U}_0 \text{ in } \mathbb{H} \text{ and } \theta + \tilde{H}(t,X,\tilde{U}(t,X),\zeta) \le 0. \tag{4.7}$$

for any $(t,X) \in [0,T)\times\mathbb{H}$ and $(\theta,\zeta) \in \partial^+\tilde{U}(t,X)$.

(2) We say that $\tilde{U}$ is a viscosity supersolution *for* (4.4) *with initial data $\tilde{U}_0 : \mathbb{H} \to \mathbb{R}$ if $\tilde{U}$ is lower semicontinuous and*

$$\tilde{U}(0,\cdot) \ge \tilde{U}_0 \text{ in } \mathbb{H} \text{ and } \theta + \tilde{H}(t,X,\tilde{U}(t,X),\zeta) \ge 0. \tag{4.8}$$

for any $(t,X) \in [0,T)\times\mathbb{H}$ and $(\theta,\zeta) \in \partial^-\tilde{U}(t,X)$.

(3) We say that $\tilde{U}$ is a viscosity solution *for* (4.4) *with initial data $\tilde{U}_0 : \mathbb{H} \to \mathbb{R}$ if it is both a viscosity subsolution and supersolution.*

The equivalence between these notions is given by:

Theorem 4.4. *Let $U_0 : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ be given and define $\tilde{U}_0 : \mathbb{H} \to \mathbb{R}$ by $\tilde{U}_0(X) = U_0(\sharp(X))$. Then the following hold:*

(1) If $U : [0,T)\times\mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ is a viscosity subsolution (resp. supersolution) for (4.3) *with initial data U_0, then $\tilde{U} : [0,T)\times\mathbb{H} \to \mathbb{R}$ given by $\tilde{U}(t,X) = U(t,\sharp(X))$ is a viscosity subsolution (resp. supersolution) for* (4.4) *with initial data $\tilde{U}_0$.*

(2) If $\tilde{U} : [0,T)\times\mathbb{H} \to \mathbb{R}$ is an R.I. viscosity subsolution (resp. supersolution) for (4.4) *with initial data $\tilde{U}_0$, then $U : [0,T)\times\mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ given by $U(t,\sharp(X)) = \tilde{U}(t,X)$ is a viscosity subsolution (resp. supersolution) for* (4.3) *with initial data U_0.*

Proof. We will only analyze the case of subsolutions below; the same argument settles the case of viscosity supersolutions.

1. Let $(\theta,\zeta) \in \partial^+\tilde{U}(t,X)$ and denote $\mu := \sharp(X)$. From the rearrangement invariance of $\tilde{U}(t,\cdot)$ we deduce, by Theorem 3.19 (iii), that $(\theta,\xi\circ X) \in \partial^+\tilde{U}(t,X)$, where $\xi\circ X := \mathrm{proj}_{\nabla F[X]}\zeta$. Now we use Theorem 3.14 once again to deduce that $(\theta,\xi) \in \partial^\bullet U(t,\mu)$, and so $\theta + H(t,\mu,U(t,\mu),\xi) \leq 0$. We use (4.1) to infer that $\tilde{U}$ is a viscosity subsolution for (4.4).

2. If $(\theta,\xi) \in \partial^\bullet U(t,\mu)$, Theorem 3.14 implies that for any $X \in \mathbb{H}$ such that $\mu = \sharp(X)$ we have $(\theta,\xi\circ X) \in \partial^+\tilde{U}(t,X)$; we deduce $\theta + \tilde{H}(t,X,\tilde{U}(t,X),\xi\circ X) \leq 0$. But $\mu = \sharp(X)$ implies $\tilde{H}(t,X,\tilde{U}(t,X),\xi\circ X) = H(t,\mu,U(t,\mu),\xi)$, so if $\tilde{U}$ is a viscosity subsolution for (4.4), then U is a viscosity subsolution for (4.3). □

When are we guaranteed that viscosity solutions to (4.4) are R.I.?

Proposition 4.5. *Assume $\mathcal{H} : [0,T]\times\mathbb{H}\times\mathbb{R}\times\mathbb{H} \to \mathbb{R}$ satisfies* (4.2). *Assume that* (4.4) *for $\mathcal{H}$ has a unique continuous viscosity solution $\tilde{U}$ with $\tilde{U}(0,\cdot) = \tilde{U}_0$, where $\tilde{U}_0$ is continuous and R.I. Then $\tilde{U}(t,\cdot)$ is R.I. for all $t \in [0,T]$.*

Proof. First we shall prove that for any $S \in \mathcal{G}(\Omega)$ we have that $\tilde{V}(t,X) := \tilde{U}(t,X\circ S)$ is a continuous viscosity solution for (4.4) with initial data $\tilde{V}_0(X) := \tilde{U}_0(X\circ S)$. For that, let $(\theta,\zeta) \in \partial^+\tilde{V}(t,X)$, so that

$$\tilde{U}(s,Y\circ S) - \tilde{U}(t,X\circ S) \leq \theta(s-t) + \langle\zeta, Y-X\rangle + o(|s-t| + \|Y-X\|).$$

But S is invertible (with S^{-1} measure preserving), so the inequality above is equivalent to

$$\tilde{U}(s,Y) - \tilde{U}(t,X\circ S) \leq \theta(s-t) + \langle\zeta, Y\circ S^{-1}-X\rangle + o(|s-t| + \|Y\circ S^{-1}-X\|),$$

or, further, to

$$\tilde{U}(s,Y) - \tilde{U}(t,X\circ S) \leq \theta(s-t) + \langle\zeta\circ S, Y-X\circ S\rangle + o(|s-t| + \|Y-X\circ S\|),$$

Clearly, the inequality in the last display is equivalent to $(\theta,\zeta\circ S) \in \partial^+\tilde{U}(t,X\circ S)$. By the hypothesis, we infer $\theta + \mathcal{H}(t,X\circ S,\tilde{U}(t,X\circ S),\zeta\circ S) \leq 0$, which, by the invariance property (4.2), implies $\tilde{V}(t,X) = \tilde{U}(t,X\circ S)$ is a continuous viscosity subsolution for (4.4) with initial data $\tilde{V}_0(X) := \tilde{U}_0(X\circ S)$. Likewise for continuous supersolutions, so $\tilde{V}$ is a continuous viscosity solution for (4.4) with initial data $\tilde{V}_0$. But $\tilde{V}_0 \equiv \tilde{U}_0$, so by the assumed uniqueness of continuous viscosity solutions, we infer $\tilde{U}(t,X) = \tilde{U}(t,X\circ S)$ for any $(t,X) \in [0,T]\times\mathbb{H}$ and any invertible, measure-preserving map S. We conclude by using Proposition 2.8 and the continuity of $\tilde{U}(t,\cdot)$. □

We are now ready to formulate:

Corollary 4.6. *Let $U_0 : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ be continuous and assume further that for $\tilde{H} : [0,T]\times\mathbb{H}\times\mathbb{R}\times\mathbb{H} \to \mathbb{R}$ given by* (4.1) *the problem* (4.4) *possesses a unique continuous viscosity solution $\tilde{U}$ for the initial data $\tilde{U}_0(X) := U_0(\sharp(X))$. Then $\tilde{U}$ is R.I. and the map $U(t,\sharp(X)) := \tilde{U}(t,X)$ is the unique continuous viscosity solution for* (4.3) *with initial data U_0.*

Proof. Existence for (4.3) follows from the existence assumption on (4.4), by Proposition 4.5 and Theorem 4.4 (2). Uniqueness for (4.3) follows from the uniqueness assumption on (4.4), by Theorem 4.4 (1). □

Remark 4.7. *Note that we only need continuity for the map $X \mapsto \tilde{U}(t,X)$ for all $t \in (0,T]$ in order to prove Proposition 4.5 (and, consequently, Corollary 4.6).*

At this point it is tempting to seek conditions on H which guarantee that (4.4) has a unique solution. Following [11] and [12], we would like $\tilde{H}$ to satisfy the conditions listed below.

There is a local modulus σ such that

$$|\tilde{H}(t_1, X_1, r_1, \zeta_1) - \tilde{H}(t_2, X_2, r_2, \zeta_2)| \leq \sigma(|t_1 - t_2| + |r_1 - r_2| + \|X_1 - X_2\| + \|\zeta_1 - \zeta_2\|, R) \tag{4.9}$$

for all $(t_i, X_i, r_i, \zeta_i) \in [0, T] \times \mathbb{H} \times \mathbb{R} \times \mathbb{H}$ such that $|r_i|, \|X_i\|,\ \|\zeta_i\| \leq R$ for $i = 1, 2$.

There is $\alpha > 0$ such that

$$r \mapsto \tilde{H}(t, X, r, \zeta) + \alpha\, r \text{ is nondecreasing} \tag{4.10}$$

for all $(t, X, \zeta) \in [0, T] \times \mathbb{H} \times \mathbb{H}$.

There is a local modulus σ_H such that

$$\tilde{H}(t, X, r, \zeta) - \tilde{H}(t, X, r, \zeta + \lambda\vartheta(X)) \leq \sigma_H(\lambda, \lambda + \|\zeta\|) \tag{4.11}$$

whenever $\lambda \geq 0$, and $(t, X, r, \zeta) \in [0, T] \times \mathbb{H} \times \mathbb{R} \times \mathbb{H}$ such that $\|X\| \geq K$ (for some $K > 0$). Here $\vartheta(X) = X/\|X\|$ for $X \neq 0$.

Finally, there is a modulus m_H such that

$$\tilde{H}(t, Y, r, \lambda\vartheta(X - Y)) - \tilde{H}(t, X, r, \lambda\vartheta(X - Y)) \leq m_H(\lambda\|X - Y\| + \|X - Y\|) \tag{4.12}$$

for all $X \neq Y \in \mathbb{H}$, $(t, r) \in [0, T] \times \mathbb{R}$ and $\lambda \geq 0$.

4.2. Affine Hamiltonians as cornerstone cases for convex Hamiltonians

Let

$$b : [0, T] \times \mathbb{M} \times \mathcal{P}_2(\mathbb{M}) \to \mathbb{M}$$

be continuous and such that

$$\text{for any } (t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{M}) \text{ the map } b(t, \cdot, \mu) \in \mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}). \tag{$b1$}$$

Let

$$H(t, \mu, r, \xi) := \langle b(t, \cdot, \mu), \xi\rangle_\mu + \mathcal{F}(t, \mu, r).$$

Remark 4.8. *By* (4.1)*,*

$$\tilde{H}(t, X, r, \zeta) = \langle b(t, \cdot, \sharp(X)) \circ X, \zeta\rangle + \tilde{\mathcal{F}}(t, X, r),$$

because, due to the composition $b(t, \cdot, \sharp(X)) \circ X$, we do not need to replace ζ by $\mathrm{proj}_{\nabla F[X]}\zeta$ *in the expression for $\tilde{H}$.*

Let us from now use the notation $b(t, X, \mu)$ to denote $b(t, \cdot, \mu) \circ X$.

Assume the following:

$$\text{Either } b \text{ is bounded, or } (t, y) \mapsto b(t, y, \mu) \cdot y \text{ is bounded and } |b(t, y, \mu)| \leq A|y| + B \tag{$b2$}$$

for some $A,\ B \geq 0$ and all $(t, y, \mu) \in [0, T] \times \mathbb{M} \times \mathcal{P}_2(\mathbb{M})$.

There exists a modulus σ_b such that

$$\|b(t_1, X_1, \sharp(X_1)) - b(t_2, X_2, \sharp(X_2)\| \leq \sigma_b(|t_1 - t_2| + \|X_1 - X_2\|) \tag{$b3$}$$

for all $(t_i, X_i) \in [0, T] \times \mathbb{H}$.

On $\mathcal{F}$ we assume the following: there exists a local modulus σ_f such that

$$\left|\mathcal{F}(t_1, \mu_1, r_1) - \mathcal{F}(t_2, \mu_2, r_2)\right| \leq \sigma_f(|t_1 - t_2| + |r_1 - r_2| + W_2(\mu_1, \mu_2), R) \tag{$\mathcal{F}1$}$$

for any $(t_i, r_i, \mu_i) \in [0, T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{M})$ such that $|r_i|,\ W_2(\mu_i, \delta_0) \leq R$ for $i = 1, 2$.

There exists $\beta \geq 0$ such that for any $(t,\mu) \in [0,T] \times \mathcal{P}_2(\mathbb{M})$

$$r \mapsto \mathcal{F}(t,\mu,r) + \beta r \text{ is nondecreasing.} \tag{$\mathcal{F}2$}$$

There exists a modulus m_f such that

$$\left|\mathcal{F}(t,\mu_1,r) - \mathcal{F}(t,\mu_2,r)\right| \leq m_f(W_2(\mu_1,\mu_2)) \tag{$\mathcal{F}3$}$$

for any $(t,r,\mu_i) \in [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{M})$ for $i = 1,2$.

Lemma 4.9. *Let $\tilde{H}$ be as above, where b satisfies* $(b1)$–$(b3)$ *and $\mathcal{F}$ satisfies* $(\mathcal{F}1)$, $(\mathcal{F}2)$. *Let $X_1, X_2, \zeta_1, \zeta_2 \in \mathbb{H}$ and $t_1, t_2 \in [0,T]$.*

(i) If $\|X_i\|,\ \|\zeta_i\| \leq R$, then

$$\left|\langle b(t_1,X_1,\mu_1),\zeta_1\rangle - \langle b(t_2,X_2,\mu_2),\zeta_2\rangle\right| \leq (AR+B)\|\zeta_1-\zeta_2\| + R\,\sigma_b(|t_1-t_2| + \|X_1-X_2\|).$$

(ii) If $K > 0$ then for any $t \in [0,T]$, $r \in \mathbb{R}$ and $X \in \mathbb{H}$ such that $\|X\| \geq K$ we have

$$\tilde{H}(t,X,r,\zeta) - \tilde{H}(t,X,r,\zeta + \lambda\vartheta(X)) \leq \lambda \min\{K^{-1}\|c\|_\infty, \|b\|_\infty\}$$

where $c(t,y,\mu) := b(t,y,\mu)\cdot y$.

(iii) We have

$$\int_\Omega [b(t,X_1(x),\mu_1) - b(t,X_2(x),\mu_2)] \cdot \lambda\vartheta(X_1-X_2)(x)\,dx \leq \lambda\sigma_b(\|X_1-X_2\|).$$

Proof. (i) We have

$$\begin{aligned}
\left|\langle b(t_1,X_1,\mu_1),\zeta_1\rangle \right. & - \left.\langle b(t_2,X_2,\mu_2),\zeta_2\rangle\right| \\
&= \left|\int_\Omega [b(t_1,X_1(x),\mu_1)\cdot\zeta_1(x) - b(t_2,X_2(x),\mu_2)\cdot\zeta_2(x)]\,dx\right| \\
&\leq \|b(t_1,\cdot,\mu_1)\|_{\mu_1}\|\zeta_1-\zeta_2\| + \|b(t_1,X_1,\mu_1) - b(t_2,X_2,\mu_2)\|\|\zeta_2\| \\
&\leq (AR+B)\|\zeta_1-\zeta_2\| + R\,\sigma_b(|t_1-t_2| + \|X_1-X_2\|),
\end{aligned}$$

where we used $(b1)$, $(b2)$.

(ii) We have, if $\|X\| \geq K$,

$$\tilde{H}(t,X,r,\zeta) - \tilde{H}(t,X,r,\zeta+\lambda\vartheta(X)) = \lambda\int_\Omega b(t,X(x),\sharp(X))\cdot\frac{X(x)}{\|X\|}\,dx,$$

so we get the bound $\lambda\|b\|_\infty$ if b is bounded, or $\lambda K^{-1}\|c\|_\infty$ is finite.

(iii) By means of $(b2)$, we readily estimate

$$\begin{aligned}
\int_\Omega [b(t,X_1(x),\mu_1) - b(t,X_2(x),\mu_2)]\cdot\lambda\vartheta(X_1-X_2)(x)\,dx &\leq \lambda\|b(t,X_1,\mu_1) - b(t,X_2,\mu_2)\| \\
&\leq \lambda\sigma_b(\|X_1-X_2\|),
\end{aligned}$$

which finishes the proof. □

In what follows, $UC_s([0,T]\times\mathcal{P}_2(\mathbb{M}))$ denotes the vector space of all functions which are uniformly continuous in μ uniformly with respect to t and uniformly continuous on bounded sets; $UC_s([0,T]\times\mathbb{H})$ is defined similarly.

Proposition 4.10. *Let H as in Example 4.2, where b satisfies* $(b1)$–$(b3)$ *and $\mathcal{F}$ satisfies* $(\mathcal{F}1)$, $(\mathcal{F}2)$. *Let $U_0 : \mathcal{P}_2(\mathbb{M}) \to \mathbb{R}$ be uniformly continuous. Then there exists a unique $U \in UC_s([0,T]\times\mathcal{P}_2(\mathbb{M}))$ which is a viscosity solution for* (4.3) *with $U(0,\cdot) \equiv U_0$.*

Proof. Let us begin by checking that $\tilde{H}$ satisfies (4.9)–(4.12). By $(\mathcal{F}1)$, we have

$$\begin{aligned}\left|\tilde{\mathcal{F}}(t_1, X_1, r_1) - \tilde{\mathcal{F}}(t_2, X_2, r_2)\right| &= \left|\mathcal{F}(t_1, \sharp(X_1), r_1) - \mathcal{F}(t_2, \sharp(X_2), r_2)\right| \\ &\le \sigma_f(|t_1 - t_2| + |r_1 - r_2| + W_2(\sharp(X_1), \sharp(X_2)), R) \\ &\le \sigma_f(|t_1 - t_2| + |r_1 - r_2| + \|X_1 - X_2\|, R),\end{aligned}$$

whenever $|r_i|$, $\|X_i\| \le R$ (the latter being equivalent to $W_2(\mu_i, \delta_0) \le R$) where $\mu_i := \sharp(X_i)$. This, together with Lemma 4.9 shows that $\tilde{H}$ satisfies (4.9).

Condition (4.10) follows trivially from $(\mathcal{F}2)$. Then note that (4.11) follows from Lemma 4.9 (ii). We use (iii) of the same Lemma and $(\mathcal{F}3)$ to see that (4.12) holds with $m_H := \sigma_b + m_f$.

The assumptions on U_0 yield the uniform continuity of $\tilde{U}_0$. According to Existence Theorem 1.1 [12], for any uniformly continuous $\tilde{U}_0$, there exists a unique $\tilde{U} \in UC_s([0,T] \times \mathbb{H})$ which is a viscosity solution for (4.4) with $\tilde{U}(0,\cdot) \equiv \tilde{U}_0$. From the properties of $\tilde{H}$, we also infer $\tilde{U}(t,\cdot)$ is R.I. for all $t \in [0,T]$ (as in the proof of Proposition 4.5). Then a version of Corollary 4.6 applies, first to give existence of a continuous viscosity solution U (given by $U(t, \sharp(X)) := \tilde{U}(t, X)$) for (4.3) with $U(0,\cdot) \equiv U_0$. From $\tilde{U} \in UC_s([0,T] \times \mathbb{H})$ we also get $U \in UC_s([0,T] \times \mathcal{P}_2(\mathbb{M}))$. As for uniqueness, if $U \in UC_s([0,T] \times \mathcal{P}_2(\mathbb{M}))$ is a viscosity solution in this class for (4.3) with $U(0,\cdot) \equiv U_0$, then it is easy to see that $\tilde{U}(t, X) := U(t, \sharp(X))$ belongs to $UC_s([0,T] \times \mathbb{H})$ and, by Theorem 4.4 (2), is viscosity solution for (4.4) with $\tilde{U}(0,\cdot) \equiv \tilde{U}_0$. □

Remark 4.11. *We have thus proved that if b and $\mathcal{F}$ satisfy $(b1)$–$(b3)$ and $(\mathcal{F}1)$, $(\mathcal{F}2)$, then for any $U_0 \in UC(\mathcal{P}_2(\mathbb{M}))$ there exists a unique $U \in UC_s([0,T] \times \mathcal{P}_2(\mathbb{M}))$ which is a viscosity solution for the semilinear transport equation*

$$\partial_t U(t,\mu) + \langle b(t,\cdot,\mu), \nabla_w U(t,\mu)\rangle_\mu + \mathcal{F}(t,\mu,U(t,\mu)) = 0 \textit{ for } (t,\mu) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \tag{4.13}$$

and such that $U(0,\cdot) \equiv U_0$ in $\mathcal{P}_2(\mathbb{M})$.

5. Time dependent HJE and convex hamiltonians

In Section 8, we provide examples of Hamiltonians satisfying the assumptions we impose in this section.

5.1. Assumptions

Suppose

$$\bar{H} \in C([0,T] \times \mathbb{H} \times \mathbb{H}) \tag{5.1}$$

is such that

$$\bar{H}(t, X, \cdot) \text{ is convex for any } (t,X) \in [0,T] \times \mathbb{H} \tag{5.2}$$

and

$$\bar{H}(0,\cdot,0) \quad \text{is bounded.} \tag{5.3}$$

We denote by $\bar{L}(t,X,\cdot)$ the Legendre transform of $\bar{H}(t,X,\cdot)$, and for $b \in \nabla_x C_b^1([0,T] \times \mathbb{M})$ (i.e. the set of spatial gradients of functions in $C^1([0,T] \times \mathbb{M})$ with bounded spatial gradients) we define

$$\bar{H}_b(t,X,\zeta) := \langle b(t,\cdot) \circ X, \zeta\rangle - \bar{L}(t,X,b(t,\cdot) \circ X) = \langle b(t,\cdot) \circ X, \zeta\rangle + \bar{F}(t,X)$$

for $(t,X,\zeta) \in [0,T] \times \mathbb{H} \times \mathbb{H}$.

We strengthen assumption (4.9) by imposing that there exist a monotone nondecreasing function e on $[0,\infty)$ and a local modulus of continuity σ such that

$$|\bar{H}(t_1,X_1,\zeta_1) - \bar{H}(t_2,X_2,\zeta_2)| \le \sigma(|t_1 - t_2| + \|\zeta_1 - \zeta_2\|, R) + e(R)\|X_1 - X_2\| \tag{5.4}$$

for all $(t_i, X_i, \zeta_i) \in [0,T] \times \mathbb{H} \times \mathbb{R} \times \mathbb{H}$ such that $\|\zeta_i\| \leq R$ for $i = 1, 2$.

We assume that there are monotone nondecreasing functions $\theta_l, \theta_h : [0,\infty) \to \mathbb{R}$ such that

$$\lim_{u\to\infty} \frac{\theta_h(u)}{u} = \lim_{u\to\infty} \frac{\theta_l(u)}{u} = \infty, \tag{5.5}$$

$$\bar{H}(t,X,\zeta) \geq \theta_h(\|\zeta\|) \quad \text{and} \quad \bar{L}(t,X,B) \geq \theta_l(\|B\|) \tag{5.6}$$

for any $X, \zeta, B \in \mathbb{H}$ and $t \in [0,T]$.

We assume that $\bar{H}$ satisfies the invariance property (4.2) and

$$\bar{H}(t,X,\zeta) \geq \tilde{H}(t,X,\zeta), \tag{5.7}$$

where we have set

$$\tilde{H}(t,X,\zeta) = \bar{H}(t,X,\mathrm{proj}_{\nabla F[X]}\zeta).$$

For $t \in (0,T]$ and $X, Y \in \mathbb{H}$ we define

$$\bar{C}_t(Y,X) := \inf_{\Sigma}\Big\{ \int_0^t \bar{L}(\tau, \Sigma_\tau, \dot{\Sigma}_\tau) d\tau \ : \quad \Sigma_0 = Y,\ \Sigma_t = X,\ \ \Sigma \in W^{1,1}(0,t;\mathbb{H}) \Big\}.$$

Similarly, we define $\bar{C}_t^b$ by replacing $\bar{L}$ by $\bar{L}_b$ in the above definition, where $\bar{L}_b$ is the Legendre transform of $\bar{H}_b$.

Additional assumption 5.1. *We say that $\bar{H}$ satisfies (A) if there is a set $\mathbb{D}$ dense in $\mathbb{H}$, such that for any $X, Y \in \mathbb{D}$ and $t \in (0,T]$ there exist sequences*

$$(b^n)_n \subset W^{1,\infty}((0,t)\times\mathbb{M};\mathbb{M}) \cap \nabla_x C_b^1((0,t)\times\mathbb{M}), \quad (\Sigma^n)_n \subset W^{1,1}((0,t);\mathbb{H})$$

such that

$$\partial_s\Sigma_s^n =: \dot{\Sigma}_s^n = b_s^n \circ \Sigma_s^n \qquad on\ \ (0,t),$$
$$\lim_n W_2\big(\sharp(Y), \sharp(\Sigma_0^n)\big) = \lim_n W_2\big(\sharp(X), \sharp(\Sigma_t^n)\big) = 0 \tag{5.8}$$

and

$$\bar{C}_t(Y,X) \geq \liminf_{n\to\infty} \int_0^t \bar{L}(s, \Sigma_s^n, \dot{\Sigma}_s^n) ds\,.$$

5.2. Properties of $\bar{H}$ and $\bar{L}$

Given $(t_1, X_1, B_1) \in [0,T] \times \mathbb{H} \times \mathbb{H}$, since θ_h is superlinear, we use (5.6) to obtain $\zeta_1 \in \mathbb{H}$ such that

$$\bar{H}(t_1,X_1,\zeta_1) + \bar{L}(t_1,X_1,B_1) = \langle \zeta_1, B_1 \rangle. \tag{5.9}$$

Lemma 5.2. *Suppose $\bar{H}$ satisfies (4.10 – 4.12), (5.4) and (5.6) hold. There are monotone nondecreasing functions $\bar{e}, \tilde{e} : [0,\infty) \to [0,\infty)$ such that the following hold for any $(t_i, X_i, B_i) \in [0,T] \times \mathbb{H} \times \mathbb{H}$ such tha $\|B_1\|, \|B_2\| \leq R$.*

- *(i) If $\zeta_1 \in \mathbb{H}$ is as in (5.9) then $\|\zeta_1\| \leq \bar{e}(R)$. After replacing $\bar{e}(R)$ by $\max\{R, \bar{e}(R)\}$ we may assume that $\bar{e}(R) \geq R$.*
- *(ii) We have*

$$|\bar{L}(t_1,X_1,B_1) - \bar{L}(t_2,X_2,B_2)| \leq \sigma\big(|t_1-t_2|, \bar{e}(R)\big) + \tilde{e}(R)\big(\|X_1-X_2\| + \|B_1-B_2\|\big).$$

Proof. Let us assume without loss of generality that $\bar{H}(0,0,0) = 0$.

(i) By the maximality property of ζ_1 and (5.4),

$$\langle \zeta_1, B_1 \rangle - \bar{H}(t_1, X_1, \zeta_1) \geq -\bar{H}(t_1, X_1, 0) \geq -\sigma(t_1, 0) - e(0)R.$$

By (5.6)

$$[\|\zeta_1\| + e(0)]R + \sigma(T, 0) \geq \theta_h(\|\zeta_1\|).$$

Since θ_h is superlinear (see (5.5)), we conclude the proof of (i).

(ii) Suppose, without loss of generality, that

$$\bar{L}(t_1, X_1, B_1) \geq \bar{L}(t_2, X_2, B_2).$$

We have

$$\begin{aligned} \bar{L}(t_1, X_1, B_1) - \bar{L}(t_2, X_2, B_2) &\leq \langle \zeta_1, B_1 - B_2 \rangle + \bar{H}(t_2, X_2, \zeta_1) - \bar{H}(t_1, X_1, \zeta_1) \\ &\leq \bar{e}(R)\|B_1 - B_2\| + \sigma(|t_2 - t_1|, \bar{e}(R)) + e(\bar{e}(R))\|X_1 - X_2\|. \end{aligned}$$

We conclude the proof of (ii) by interchanging the roles of (t_1, X_1, B_1) and (t_2, X_2, B_2) . □

Remark 5.3. *Assume $\bar{H}$ satisfies (4.2), (4.10 – 4.12) and (5.3 – 5.7). If $B_R \subset \mathbb{H}$ is the ball of radius R then $\bar{H}$ is bounded on* $[0,T] \times \mathbb{H} \times B_R$.

Proof. Suppose $(t, X, B) \in [0,T] \times \mathbb{H} \times B_R$. Then

$$|\bar{L}(t, X, B)| \leq |\bar{L}(t, X, B) - \bar{L}(t, X, 0)| + |\bar{L}(t, X, 0) - \bar{L}(0, X, 0)| + |\bar{L}(0, X, 0)|.$$

We apply Lemma 5.2 (ii) to obtain

$$|\bar{L}(t, X, B)| \leq \tilde{e}(R)R + \sigma(T, \tilde{e}(0)) + |\bar{L}(0, X, 0)|.$$

Since $-\sup_X |H(0, X, 0)| \leq \bar{L}(0, X, 0) \leq -\theta_h(0)$, we conclude the proof. □

Lemma 5.4. *If $t \in [0,T]$, $X, B \in \mathbb{H}$ and $S \in \mathcal{S}(\Omega)$, then*

$$\bar{L}(t, X \circ S, B \circ S) = \bar{L}(t, X, B).$$

Proof. We have

$$\bar{L}(t, X \circ S, B \circ S) = \sup_{\tilde{\zeta} \in \mathbb{H}} \Big\{ \langle \tilde{\zeta}, B \circ S \rangle - \bar{H}(t, X \circ S, \tilde{\zeta}) \Big\} \geq \sup_{\zeta \in \mathbb{H}} \Big\{ \langle \zeta \circ S, B \circ S \rangle - \bar{H}(t, X \circ S, \zeta \circ S) \Big\}$$

and so (4.2) and $\langle \zeta \circ S, B \circ S \rangle = \langle \zeta, B \rangle$ imply

$$\bar{L}(t, X \circ S, B \circ S) \geq \bar{L}(t, X, B). \tag{5.10}$$

We apply Proposition 2.8 with $X_0 = \mathbf{id}$ and $X = S$ to obtain a sequence of maps $S_n \in \mathcal{G}(\Omega)$ such that S_n converges to S. We apply (5.10) to obtain

$$\bar{L}(t, X, B) = \bar{L}(t, X \circ S_n \circ S_n^{-1}, B \circ S_n \circ S_n^{-1}) \geq \bar{L}(t, X \circ S_n, B \circ S_n).$$

Since $\{X \circ S_n\}_n$ converges to $X \circ S$ and $\{B \circ S_n\}_n$ converges to $B \circ S$, by Lemma 5.2 (ii) (which implies that $\bar{L}$ is continuous) we conclude that $\bar{L}(t, X, B) \geq \bar{L}(t, X \circ S, B \circ S)$. This, together with (5.10), completes the proof. □

By Corollary 2.10, the following functions are well–defined:

$$H(t,\mu,\xi) := \bar{H}(t,X,\xi\circ X) \text{ for all } (t,\mu)\in[0,T]\times\mathcal{P}_2(\mathbb{M}) \text{ and all } \xi\in\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M}),$$

where $X\in\mathbb{H}$ is such that $\mu=\sharp(X)$. Similarly, if $b\in\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$ then

$$L(t,\mu,b) := \bar{L}(t,X,b\circ X).$$

Remark 5.5. *Let $t\in[0,T]$, $\mu\in\mathcal{P}_2(\mathbb{M})$, $\zeta\in\mathbb{H}$ and $X\in\mathbb{H}$ be such that $\mu=\sharp(X)$.*

(i) The Legendre transform of $L(t,\mu,\cdot)$ is $H(t,\mu,\cdot)$.

(ii) If $\xi\in\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$ and $\mathrm{proj}_{\nabla F[X]}\zeta=\xi\circ X$ then $H(t,\mu,\xi)=\bar{H}(t,X,\mathrm{proj}_{\nabla F[X]}\zeta)$.

Proof. One obtains (ii) by using the definition of H. Therefore, only (i) needs to be proved. Observe that if $\xi\in\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$ then

$$\langle b,\xi\rangle_\mu - H(t,\mu,\xi) = \langle b\circ X,\xi\circ X\rangle - \bar{H}(t,X,\xi\circ X) \le \bar{L}(t,X,b\circ X) = L(t,\mu,b).$$

Maximizing over $\xi\in\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$ we obtain (this is how we define the left-hand-side below)

$$(H(t,\mu,\cdot))^*(b) \le L(t,\mu,b).$$

If $\zeta\in\mathbb{H}$ then, by (5.7),

$$\langle\zeta,b\circ X\rangle - \bar{H}(t,X,\zeta) \le \langle\mathrm{proj}_{\nabla F[X]}\zeta,b\circ X\rangle - \bar{H}(t,X,\mathrm{proj}_{\nabla F[X]}\zeta).$$

Writing $\mathrm{proj}_{\nabla F[X]}\zeta=\xi\circ X$ for $\xi\in\mathcal{T}_\mu\mathcal{P}_2(\mathbb{M})$ we conclude that

$$\langle\zeta,b\circ X\rangle - \bar{H}(t,X,\zeta) \le \langle\xi\circ X,b\circ X\rangle - \bar{H}(t,X,\xi\circ X) = \langle\xi,b\rangle_\mu - H(t,\mu,\xi) \le \big(H(t,\mu,\cdot)\big)^*(b).$$

Maximizing over $\zeta\in\mathbb{H}$ we conclude that

$$L(t,\mu,b) = \bar{L}(t,X,b\circ X) \le (H(t,\mu,\cdot))^*(b).$$

□

Lemma 5.6. *Suppose $\bar{H}$ satisfies (A), (4.10 – 4.12), (5.4) and (5.6) hold. If $b\in W^{1,\infty}(\mathbb{M};\mathbb{M})\cap C_b(\mathbb{M};\mathbb{M})$, then $\bar{H}_b$ satisfies (4.9 – 4.12) for appropriate moduli of continuity.*

Proof. Apply Lemma 4.9 when $\tilde{\mathcal{F}}\equiv 0$, to obtain that for appropriate moduli of continuity,

$$(t,X,\zeta)\to\langle b(t,\cdot)\circ X,\zeta\rangle$$

satisfies (4.9 – 4.12). It remains to check that $\bar{F}$ satisfies (4.9 – 4.12) for appropriate local modulus of continuity. By Lemma 5.2 (ii), $\bar{F}$ satisfies (4.9). Since $\bar{F}$ is independent of $r\in\mathbb{R}$ and $\zeta\in\mathbb{H}$, it satisfies (4.10 – 4.11). By Lemma 5.2 (ii), if $t\in[0,T]$ and $X\in\mathbb{H}$, then

$$\begin{aligned}\bar{F}(t_2,Y)-\bar{F}(t_1,X) &= \bar{L}(t_1,X,b(t_1,\cdot)\circ X)-\bar{L}(t_2,Y,b(t_2,\cdot)\circ Y)\\ &\le \sigma\big(|t_2-t_1|,\tilde{e}(\|b\|_\infty)\big)+\tilde{e}(\|b\|_\infty)\big)\Big(\|X-Y\|+\|b(t_1,X)-b(t_2,Y)\|\Big)\\ &\le \sigma\big(|t_2-t_1|,\tilde{e}(\|b\|_\infty)\big)+\tilde{e}(\|b\|_\infty)\big)(\mathrm{Lip}(b)+1)(\|X-Y\|+|t_1-t_2|).\end{aligned}$$

This concludes the proof. □

For $t\in[0,T]$, $\mu\in\mathcal{P}_2(\mathbb{M})$ and $\xi\in L^2(\mu)$, set

$$H_b(t,\mu,\xi) := \bar{H}_b(t,X,\xi\circ X),$$

where $X\in\mathbb{H}$ is such that $\mu=\sharp(X)$. Using the definition of $\tilde{H}$ given in (4.1) and applying Remark 4.8, we get

$$\tilde{H}_b(t,X,\zeta)=\bar{H}_b(t,X,\zeta) \text{ for all } (t,X,\zeta)\in[0,T]\times\mathbb{H}\times\mathbb{H}.$$

5.3. Comparison principle

We define

$$\bar{U}(t,X) = \inf_Y \Big\{\bar{C}_t(Y,X) + \tilde{U}_0(Y)\Big\}, \qquad \bar{U}_b(t,X) = \inf_Y \Big\{\bar{C}^b_t(Y,X) + \tilde{U}_0(Y)\Big\}$$

for $t \in (0,T]$ and $X \in \mathbb{H}$.

Theorem 5.7 (Comparison principle)**.** *Assume $\bar{H}$ satisfies (4.2), (4.10 – 4.12) and (5.3 – 5.7). Suppose $U_0 \in UC(\mathcal{P}_2(\mathbb{M}))$ and $\bar{H}(0,\cdot,0)$ are bounded and let $U^{\pm} \in C([0,T]\times\mathcal{P}_2(\mathbb{M}))$.*

- *(i) If $U^+(t,\cdot) \in UC(\mathcal{P}_2(\mathbb{M}))$ uniformly in $t \in [0,T]$ and U^+ is a* viscosity supersolution *for* (4.3) *with initial condition U_0, then $\bar{U} \le \widetilde{U^+}$.*
- *(ii) $\bar{U}(t,\cdot)$ is R.I. and so, $U(t,\mu) := \bar{U}(t,X)$ is well–defined for $\mu \in \mathcal{P}_2(\mathbb{M})$, where $X \in \mathbb{H}$ and $\mu = \sharp(X)$*
- *(iii) If $U^-(t,\cdot) \in UC(\mathcal{P}_2(\mathbb{M}))$ uniformly in $t \in [0,T]$ and U^- is a* viscosity subsolution *for* (4.3) *with initial condition U_0, then $\bar{U} \ge \widetilde{U^-}$.*
- *(iv) There is at most one $U \in C([0,T]\times\mathcal{P}_2(\mathbb{M}))$* viscosity solution *for* (4.3) *with initial condition U_0 such that $U(t,\cdot) \in UC(\mathcal{P}_2(\mathbb{M}))$ uniformly in $t \in [0,T]$*

Proof. (i) Since $\bar{H}$ satisfies (4.10 – 4.12) and (5.4), by the theory of viscosity solution in Banach spaces (cf. e.g. [11] [12] [13]), $\bar{U}$ is the unique viscosity solution for

$$\partial_t \bar{U}(t,X) + \bar{H}(t,X,\bar{U}(t,X),\nabla\bar{U}(t,X)) = 0 \text{ for } (t,X) \in [0,T)\times\mathbb{H},$$

with initial data $\tilde{U}_0$. Furthermore, $\bar{U} \in BUC([0,T]\times\mathbb{H})$. Since U^+ is a viscosity supersolution for (4.3) with initial condition U_0, applying Theorem 4.4, we obtain that $\widetilde{U^+}$ is a viscosity supersolution for (4.4) with initial data $\tilde{U}_0$. By the fact that $\tilde{H} \le \bar{H}$, we conclude that $\widetilde{U^+}$ is a viscosity supersolution for

$$\partial_t \widetilde{U^+}(t,X) + \bar{H}(t,X,\widetilde{U^+}(t,X),\nabla\widetilde{U^+}(t,X)) = 0 \text{ for } (t,X) \in [0,T)\times\mathbb{H}$$

with initial data $\tilde{U}_0$. We can compare it then to the viscosity solution $\bar{U}$ (which is thus a subsolution); we invoke the comparison principle [11] [12] (using $\bar{H}$ as our Hamiltonian), to conclude the proof of (i).

(ii) Since $\bar{H}$ satisfies the invariance property (4.2), we may use the uniqueness property of viscosity solution on Banach spaces (cf. e.g. [11] [12] [13]) to conclude that $\bar{U}(t,X\circ S) = U(t,X)$ for any $(t,X) \in [0,T]\times\mathbb{H}$ and any $S \in \mathcal{S}(\Omega)$. Since $\bar{U}$ is continuous, we use Corollary 2.10 to infer that $\bar{U}(t,\cdot)$ is R.I.

(iii) Under the additional assumptions imposed in (iii), fix $\bar{t} \in (0,T]$ and $\bar{X}, Y \in \mathbb{H}$. We are to prove that

$$C_{\bar{t}}(Y,\bar{X}) + \tilde{U}_0(Y) \ge \widetilde{U^-}(\bar{t},\bar{X}). \tag{5.11}$$

Let $\{(b^n,\Sigma^n)\}_n$ be the sequence from the assumption 5.1. Given $\epsilon > 0$ arbitrary, we choose n such that

$$\bar{C}_{\bar{t}}(\bar{X},Y) \ge -\epsilon + \int_0^{\bar{t}} \bar{L}(s,\Sigma^n_s,\dot{\Sigma}^n_s)ds. \tag{5.12}$$

Observe that the Legendre transform of $\bar{H}_{b_n}(t,X,\cdot)$ is $\bar{L}_{b_n}(t,X,\cdot)$ given by

$$\bar{L}_{b_n}(t,X,B) = \begin{cases} \bar{L}(t,X,b_n(t,\cdot)\circ X) & \text{if} \quad B = b_n(t,\cdot)\circ X \quad \text{a.e.} \\ \infty & \text{if} \quad B \ne b_n(t,\cdot)\circ X \quad \text{a.e.} \end{cases}$$

The (unique) viscosity solution $\tilde{\mathcal{V}}$ of

$$\partial_t \tilde{\mathcal{V}}(t,X) + \tilde{H}_{b_n}(t,X,\nabla\tilde{\mathcal{V}}(t,X)) = 0 \quad \text{for} \quad (t,X) \in [0,T)\times\mathbb{H} \tag{5.13}$$

with initial data $\tilde{U}_0$ is therefore

$$\tilde{\mathcal{V}}(t,X) := \inf_\Sigma \left\{\tilde{U}_0(\Sigma(0)) + \int_0^t \bar{L}(s,\Sigma(s),\dot{\Sigma}(s))ds \ : \ \dot{\Sigma} = b_n(\cdot,\Sigma), \ \ \Sigma(t) = X\right\}. \tag{5.14}$$

Remark 5.5 gives that the Legendre transform on $L(t,\mu,\cdot)$ is $H(t,\mu,\cdot)$ and so, $H_{b_n} \le H$. We conclude that $\widetilde{U^-}$ is a viscosity subsolution for

$$\partial_t \widetilde{U^-}(t,\mu) + H_{b_n}(t,\mu,\nabla_w \widetilde{U^-}(t,\mu)) = 0 \quad \text{for} \quad (t,\mu) \in [0,T)\times \mathcal{P}_2(\mathbb{M}),$$

with initial data U_0. Thanks to Theorem 4.4, $\widetilde{U^-}$ is a viscosity subsolution for (5.13) with initial data $\tilde{U}_0$. Viscosity solutions being also a viscosity supersolutions, we use the comparison principle [11] for (5.13) to conclude that $\widetilde{U^-} \le \tilde{\mathcal{V}}$. In particular,

$$\widetilde{U^-}(\bar{t},\Sigma^n_{\bar{t}}) \le \tilde{\mathcal{V}}(\bar{t},\Sigma^n_{\bar{t}}) \le \tilde{U}_0(\Sigma^n_0) + \int_0^{\bar{t}} \bar{L}(s,\Sigma^n_s,\dot{\Sigma}^n_s)ds \le \tilde{U}_0(\Sigma^n_0) + \bar{C}_{\bar{t}}(Y,\bar{X}) + \epsilon,$$

where we have used (5.12). Since (5.8) holds and U^- and U_0 are continuous, we conclude that

$$U^-(\bar{t},\sharp(\bar{X}_t)) \le U_0(\sharp(Y_0)) + \bar{C}_{\bar{t}}(Y,\bar{X}) + \epsilon.$$

By the fact that $\epsilon > 0$ is arbitrary, we conclude the proof of (iii).

(iv) is an obvious consequence of (i) and (iii). □

5.4. Lipschitz property of subsolutions to the Eikonal equation

The eikonal equations studied here will later be used to show that subsolutions for Hamilton–Jacobi equations are Lipschitz when the Hamiltonian is coercive. Consider the equation

$$\|\nabla_w V(\mu)\|_\mu = \lambda \text{ in } \mathcal{P}_2(\mathbb{M}). \tag{5.15}$$

Proposition 5.8. *For any real constant $\lambda \ge 0$, any viscosity subsolution $U \in BUC(\mathcal{P}_2(\mathbb{M}))$ of* (5.15) *is λ–Lipschitz.*

Proof. Note that any viscosity subsolution U for (5.15) is also a viscosity subsolution for

$$\partial_t W(t,\mu) + \frac{1}{2}\|\nabla_w W(t,\mu))\|^2 - \frac{1}{2}\lambda^2 = 0, \quad W(0,\cdot) = U. \tag{5.16}$$

By [19], a viscosity solution for (5.16) is given by the Hopf-Lax type formula,

$$V(t,\mu) = \inf_{\nu\in\mathcal{P}_2(\mathbb{M})} \Big\{ U(\nu) + \frac{1}{2t} W_2^2(\mu,\nu) \Big\} + \frac{\lambda^2}{2} t.$$

In fact, if m is a modulus a continuity for U and $|U| \le M$, the proof of Proposition 4.7 [16] shows that

$$\bar{m}(\delta) := 2m(\sqrt{(2M+1)\delta}) + 2(1+\lambda^2)\delta$$

is a modulus of continuity for $V(t,\cdot)$. Thanks to the comparison principle in Theorem 5.7 applied to (5.16), since V is a viscosity supersolution, we infer

$$\frac{1}{2t} W^2(\mu,\nu) + \frac{\lambda^2}{2} t \ge U(\mu) - U(\nu) \text{ for all } t > 0,\ \mu,\ \nu \in \mathcal{P}_2(\mathbb{M}).$$

The desired Lipschitz continuity follows by minimizing the left hand side with respect to t. □

6. A stationary HJE

Consider the following problem

$$U(\mu) + H(\mu, U(\mu), \nabla_w U(\mu)) = 0 \text{ in } \mathcal{P}_2(\mathbb{M}) \tag{6.1}$$

and let us attempt to identify sufficient conditions on H which yield uniqueness of viscosity solutions in $BC(\mathcal{P}_2(\mathbb{M}))$ (the set of all bounded, continuous functions on $\mathcal{P}_2(\mathbb{M})$). First, let us assume there exists $\tilde{H} : \mathbb{H} \times \mathbb{R} \times \mathbb{H} \to \mathbb{R}$ continuous such that

$$\tilde{H}(X, r, \xi \circ X) = H(\sharp(X), r, \xi) \text{ for all } X \in \mathbb{H},\ r \in \mathbb{R},\ \xi \in L^2(\sharp(X)).$$

Let us further assume that

$$\tilde{H}(X, r, \zeta) \geq \tilde{H}(X, r, \mathrm{proj}_{\nabla F[X]}\zeta) \text{ for all } (r, X, \zeta) \in \mathbb{R} \times \mathbb{H} \times \mathbb{H}, \tag{6.2}$$

$$\lim_{\|\zeta\| \to \infty} \tilde{H}(X, r, \zeta) = \infty \text{ uniformly with respect to } (r, X) \in \mathbb{R} \times \mathbb{H}, \tag{6.3}$$

$$\zeta \mapsto \tilde{H}(X, r, \zeta) \text{ is } \mathbb{H}\text{–weakly l.s.c. for each } (r, X) \in \mathbb{R} \times \mathbb{H}, \tag{6.4}$$

$$\zeta \mapsto \tilde{H}(X, r, \zeta) \text{ is quasiconvex for any } (r, X) \in \mathbb{R} \times \mathbb{H}, \tag{6.5}$$

i.e. the sublevel sets $\{\zeta : \tilde{H}(X, r, \zeta) \leq \alpha\}$ are convex for all $\alpha \in \mathbb{R}$ and all $(r, X) \in \mathbb{R} \times \mathbb{H}$.

Furthermore, (4.10) is replaced by:

$$r \mapsto \tilde{H}(X, r, \zeta) \text{ is nondecreasing for all } (X, \zeta) \in \mathbb{H} \times \mathbb{H}. \tag{6.6}$$

For example, if $H(\mu, r, \xi) = \frac{1}{2}\|\xi\|^2 + \mathcal{F}(\mu)$, we see that $\tilde{H}$ is uniquely determined as

$$\tilde{H}(X, r, \zeta) := \frac{1}{2}\|\zeta\|^2 + \mathcal{F}(X_\sharp \chi) = \frac{1}{2}\|\zeta\|^2 + \tilde{\mathcal{F}}(X).$$

If $\mathcal{F}$ is continuous and bounded, then $\tilde{H}$ satisfies the conditions for existence and uniqueness (4.9)-(4.12) (also, see [11] and [12]), plus (6.2)-(6.5).

Our strategy is the following:

(1) First note that any viscosity subsolution U of (6.1) is also a viscosity subsolution for (5.15), where $\lambda \in \mathbb{R}$ is determined by $\tilde{H}(X, r, \zeta) \leq \|U\|_\infty$; indeed, if $\xi \in \partial^\bullet U(\mu)$, then $U(\mu) + H(\mu, U(\mu), \xi) \leq 0$, which by (6.3) implies the existence of $\lambda \in \mathbb{R}$ (depending only on $\|U\|_\infty$) such that $\|\xi\|_\mu \leq \lambda$.

(2) For any real constant $\lambda \geq 0$, we know by Proposition 5.8 that any viscosity subsolution for (5.15) lying in $BC(\mathcal{P}_2(\mathbb{M}))$ is also Lipschitz continuous on $\mathcal{P}_2(\mathbb{M})$. (Note that this is known to be true in the finite-dimensional and also the L^2 settings [13]).

(3) From (1) and (2) it follows that any viscosity subsolution U (in $BC(\mathcal{P}_2(\mathbb{M}))$; we always discuss such solutions only) of (6.1) is Lipschitz continuous on $\mathcal{P}_2(\mathbb{M})$. Thus, $\tilde{U}$ is Lipschitz continuous in $\mathbb{H}$. So, in particular, if a viscosity subsolution lies in $BC(\mathcal{P}_2(\mathbb{M}))$, then it, in fact, lies in $BUC(\mathcal{P}_2(\mathbb{M}))$ (bounded, uniformly continuous real valued functions on $\mathcal{P}_2(\mathbb{M})$).

(4) Finally, we show that if $U \in BC(\mathcal{P}_2(\mathbb{M}))$ is a viscosity solution for (6.1), then $\tilde{U}$ is a viscosity solution for

$$\tilde{U}(X) + \tilde{H}(X, \tilde{U}(X), \nabla \tilde{U}(X)) = 0 \text{ in } \mathbb{H}. \tag{6.7}$$

Since [11] guarantees a unique viscosity solution for (6.7) in $BUC(\mathbb{H})$, we obtain uniqueness for (6.1) in light of the equivalence between $U \in BUC(\mathcal{P}_2(\mathbb{M}))$ and $\tilde{U} \in BUC(\mathbb{H})$. The local Lipschitz continuity of the viscosity solution $\tilde{U}$ will benefit our analysis as follows: it is known [20] that a locally Lipschitz continuous map $\mathcal{V}$ on

$\mathbb{H}$ is differentiable at any X in a dense subset of $\mathbb{H}$. Furthermore, the superdifferential set at any point X satisfies (see, e.g., [6] page 49)

$$\partial^+\mathcal{V}(X) \subset \overline{\mathrm{conv}\{weak - \lim_{k\to\infty} \nabla\mathcal{V}(X_k)\ :\ X_k \to X\}}^{\mathbb{H}-weak}, \tag{6.8}$$

i.e. for any $\zeta \in \partial^+\mathcal{V}(X)$ there exists a sequence $\{\zeta_n\}_n \in \mathbb{H}$ which converges weakly to ζ and such that

$$\zeta_n = \sum_{i=1}^{m_n} \lambda_n^i \vartheta_n^i, \text{ where } \lambda_n^i \geq 0, \ \sum_{i=1}^{m_n} \lambda_n^i = 1$$

and there exist sequences $\{X_k^{n,i}\}_k$ converging strongly to X such that $\mathcal{V}$ is differentiable at each $X_k^{n,i}$ and

$$\nabla\mathcal{V}(X_k^{n,i}) \rightharpoonup \vartheta_n^i \text{ weakly as } k \to \infty.$$

We use (6.4) to conclude

$$\liminf_{k\to\infty} \tilde{H}(X_k^{n,i}, \mathcal{V}(X_k^{n,i}), \nabla\mathcal{V}(X_k^{n,i})) \geq \tilde{H}(X, \mathcal{V}(X), \vartheta_n^i) \text{ for each } i = 1...m_n$$

and so (6.5) implies

$$\liminf_{k\to\infty} \max_{i=1...m_n} \tilde{H}(X_k^{n,i}, \mathcal{V}(X_k^{n,i}), \nabla\mathcal{V}(X_k^{n,i})) \geq \max_{i=1...m_n} \tilde{H}(X, \mathcal{V}(X), \vartheta_n^i) \geq \tilde{H}(X, \mathcal{V}(X), \zeta_n).$$

It follows

$$\liminf_{n\to\infty} \liminf_{k\to\infty} \max_{i=1...m_n} \tilde{H}(X_k^{n,i}, \mathcal{V}(X_k^{n,i}), \nabla\mathcal{V}(X_k^{n,i})) \geq \tilde{H}(X, \mathcal{V}(X), \zeta). \tag{6.9}$$

But, if $\mathcal{V} := \tilde{U}$ (where U is a viscosity solution for (6.1)), we know that at all points of differentiability $X_k^{n,i}$ we have that U is also differentiable at $\mu_k^{n,i} := \sharp(X_k^{n,i})$ and so

$$\tilde{U}(X_k^{n,i}) + \tilde{H}(X_k^{n,i}, \tilde{U}(X_k^{n,i}), \nabla\tilde{U}(X_k^{n,i})) = U(\mu_k^{n,i}) + H(\mu_k^{n,i}, U(\mu_k^{n,i}), \nabla_w U(\mu_k^{n,i})) = 0.$$

In light of the strong convergence of $X_k^{n,i}$ to X as $k \to \infty$, the above equality and (6.9) imply

$$\tilde{U}(X) + \tilde{H}(X, \tilde{U}(X), \zeta) \leq 0.$$

Thus, $\tilde{U}$ is a viscosity subsolution for (6.7). Just as in the previous section, it is easy to show that $\tilde{U}$ is a viscosity supersolution: indeed, if $\zeta \in \partial^-\tilde{U}(X)$, we know $\mathrm{proj}_{\nabla F[X]}\zeta =: \xi \circ X \in \partial^-\tilde{U}(X)$, which is equivalent to $\xi \in \partial_{\bullet}U(\sharp(X))$. Thus, by virtue of (6.2), we have (for $\mu := \sharp(X)$)

$$\tilde{U}(X) + \tilde{H}(X, \tilde{U}(X), \zeta) \geq \tilde{U}(X) + \tilde{H}(X, \tilde{U}(X), \mathrm{proj}_{\nabla F[X]}\zeta) = U(\mu) + H(\mu, U(\mu), \xi) \geq 0.$$

We conclude that $\tilde{U}$ is, indeed, a viscosity solution for (6.7).

We have thus proved:

Theorem 6.1. *If $\tilde{H}$ satisfies* (4.9), (4.11), (4.12), *and* (6.2)–(6.6), *then* (6.1) *admits a unique viscosity solution in $BC(\mathcal{P}_2(\mathbb{M}))$.*

Remark 6.2. *Thus, we have a comparison principle for locally Lipschitz viscosity solutions $U \in BUC(\mathcal{P}_2(\mathbb{M}))$ for* (6.1). *Furthermore, note that the same argument will work if $\tilde{H}$ satisfies the conditions for existence and uniqueness* (4.9)-(4.12) *(also, see [11] and [12]), plus*

$$\tilde{H}(X, r, \zeta) \leq \tilde{H}(X, r, \mathrm{proj}_{\nabla F[X]}\zeta) \text{ for all } (r, X, \zeta) \in \mathbb{R} \times \mathbb{H} \times \mathbb{H}, \tag{6.10}$$

$$\lim_{\|\zeta\|\to\infty} \tilde{H}(X, r, \zeta) = -\infty \text{ uniformly with respect to } (r, X) \in \mathbb{R} \times \mathbb{H}, \tag{6.11}$$

$$\zeta \mapsto \tilde{H}(X, r, \zeta) \text{ is } L^2(\Omega; \mathbb{R}^d)\text{–weakly u.s.c. for each } (r, X) \in \mathbb{R} \times \mathbb{H}, \tag{6.12}$$

$$\zeta \mapsto \tilde{H}(X, r, \zeta) \text{ is quasiconcave for any } (r, X) \in \mathbb{R} \times \mathbb{H}. \tag{6.13}$$

Indeed, note that $\tilde{H}$ satisfies (4.9)-(4.12) *and* (6.11)-(6.13) *if and only if $-\tilde{H}$ satisfies* (4.9)-(4.12) *and* (6.3)-(6.5).

Such a Hamiltonian is provided by

$$\tilde{H}(X, \zeta) := -\frac{1}{2}\|\zeta\|^2 + \tilde{\mathcal{F}}(X),$$

which corresponds to $H(\mu, \xi) = -\|\xi\|_\mu^2/2 + \mathcal{F}(\mu)$.

7. Appendix I: A particular stationary problem revisited

We will now specialize to a particular equation of type (7.1), namely take $\lambda \in \mathbb{R}$ and consider

$$U(\mu) + \frac{1}{2}\|\nabla_w U(\mu)\|_\mu^2 = \lambda. \tag{7.1}$$

We will show uniqueness of viscosity solutions without the need to a priori prove that the solution is Lipschitz continuous (so that we will not need the deep result (6.8) by [20]).

Theorem 7.1. *For any $\lambda \in \mathbb{R}$ the problem* (7.1) *has the comparison principle in $C(\mathcal{P}_2(\mathbb{M}))$.*

Proof. Note that any viscosity subsolution U of (7.1) is a viscosity subsolution for

$$V(\mu) + \langle \nabla_w V(\mu), b \rangle_\mu - \frac{1}{2}\|b\|_\mu^2 = \lambda. \tag{7.2}$$

for any $b \in C_c^1(\mathbb{R}^d; \mathbb{R}^d)$. The Hamiltonian for (7.2) is

$$H(\mu, \zeta) := \langle \zeta, b \rangle_\mu - \frac{1}{2}\|b\|_\mu^2 - \lambda,$$

so the corresponding Lagrangian is $L(\mu, \xi) = \lambda + \|b\|_\mu^2/2$ if $\xi \equiv b$ μ-a.e. and $L(\mu, \xi) = +\infty$ else. Since H is, again, a Hamiltonian as in Subsection 4.2 (time-independent), and b is sufficiently regular for conditions ($b1$)-($b2$), ($\mathcal{F}1$), ($\mathcal{F}2$) to be satisfied, we have that (7.2) is equivalent to

$$\tilde{V}(X) + \langle \nabla \tilde{V}(X), b \circ X \rangle - \frac{1}{2}\|b \circ X\|^2 = \lambda, \; X \in \mathbb{H}. \tag{7.3}$$

So, $\tilde{U}$ is a viscosity subsolution for (7.3). But

$$\tilde{\mathcal{U}}[b](X) := \inf_{\sigma(0)=X} \frac{1}{2}\int_0^\infty e^{-s}\tilde{L}(\sigma(s), \dot{\sigma}(s))\, ds = \lambda + \frac{1}{2}\int_0^\infty e^{-s}\|\dot{\tilde{\sigma}}(s) \circ X\|^2\, ds,$$

is the unique viscosity solution for (7.3), where $\partial_s \tilde{\sigma}(s; y) = b(\tilde{\sigma}(s; y))$, $\tilde{\sigma}(0; y) = y$, $s \in [0, \infty)$, $y \in \mathbb{R}^d$. By the comparison principle for (7.3), we get

$$\tilde{\mathcal{U}}[b](X) \geq \tilde{U}(X) \text{ for all } X \in \mathbb{H} \text{ and any } b \in C_c^1(\mathbb{R}^d; \mathbb{R}^d).$$

In particular, for $b \equiv 0$ we get

$$\lambda \geq \tilde{U}(X) \text{ for all } X \in \mathbb{H}.$$

But we know that any continuous viscosity supersolution V of (7.1) yields a continuous viscosity supersolution $\tilde{V}$ for

$$\tilde{U}(X) + \frac{1}{2}\|\nabla \tilde{U}(X)\|^2 = \lambda.$$

The comparison principle for this problem applied to the continuous viscosity supersolution $\tilde{V}$ and the (obvious) continuous viscosity solution (and, thus, subsolution) $\tilde{W} \equiv \lambda$ yields

$$\tilde{V}(X) \geq \lambda \text{ for all } X \in \mathbb{H}.$$

So, $V(\mu) \geq \lambda \geq U(\mu)$ for all $\mu \in \mathcal{P}_2(\mathbb{M})$, for all continuous viscosity subsolutions U and all continuous viscosity supersolutions V. □

8. Appendix II: Examples

Throughout this section, $\mathcal{F} \in C_b(\mathcal{P}_2(\mathbb{M}))$ (continuous and bounded) and satisfies ($\mathcal{F}1$), ($\mathcal{F}2$) and ($\mathcal{F}3$). Let $\bar{\mathcal{F}} : \mathbb{H} \to \mathbb{R}$ be defined by

$$\bar{\mathcal{F}}(X) := \mathcal{F}(\sharp(X)).$$

Let

$$f \in W^{1,\infty}((0,T) \times \mathbb{M}; \mathbb{M}) \cap C_b([0,T] \times \mathbb{M}; \mathbb{M}), \quad \bar{\mathcal{F}} \in C^1(\mathbb{H}) \cap C_b(\mathbb{H}) \cap \mathrm{Lip}(\mathbb{H}).$$

Let

$$\bar{l} \in C^2([0,T] \times \mathbb{M})$$

and set

$$l(t,x,v) = \bar{l}(x,v) + \langle f(t,x), v\rangle_x.$$

We suppose that

$$\bar{l}(x,\cdot) \text{ is a convex function for all } x \in \mathbb{M}$$

and there are constants $\kappa_0, \kappa_2, \kappa_4 > 0, \kappa_1, \kappa_3$ such that

$$\kappa_2 l(\bar{s}, \bar{x}, \bar{v}) + \kappa_3 \geq l(s,x,v) \geq \kappa_0 |v|^2 - \kappa_1 \tag{8.1}$$

for any $s, \bar{s} \in [0,T]$ and any $x, v, \bar{x}, \bar{v} \in \mathbb{M}$ such that $|v| \leq |\bar{v}|$.

Suppose here exist a local modulus of continuity σ and a monotone nondecreasing function $\tilde{e} : [0,\infty) \to [0,\infty)$ such that

$$l(\bar{s}, \bar{x}, \bar{v}) - l(s,x,v) \leq \tilde{e}(R)(|s - \bar{s}| + |\bar{x} - x|) + \sigma(|\bar{v} - v|, \tilde{e}(R)) \tag{8.2}$$

if $s, \bar{s} \in [0,T]$, $x, \bar{x}, v, \bar{v} \in \mathbb{M}$ and $|v|, |\bar{v}| \leq R$. For instance, if $p \geq 2$, $a \in C^2(\bar{\Omega})$ is positive, and $f \equiv 0$, then

$$\bar{l}(x,v) := a(x)\frac{|v|^p}{p}$$

satisfies (8.1 – 8.2) with

$$\kappa_2 = \frac{\max a}{\min a}, \quad \kappa_3 = 0, \quad \tilde{e}(R) = \mathrm{Lip}(a)\frac{R^p}{p}, \quad \sigma(u,\tau) = 2\|a\|_\infty \Big(\frac{p\tau}{\mathrm{Lip}(a)}\Big)^{\frac{p-1}{p}} u.$$

Any positive linear combination of functions satisfying (8.1 – 8.2) also satisfies (8.1 – 8.2).

Let h be the Legendre transform of l. We define

$$\bar{H}(t,X,\zeta) := \int_\Omega h(t, X(\omega), \zeta(\omega))d\omega + \bar{\mathcal{F}}(X), \quad \bar{L}(t,X,B) := \int_\Omega l(t, X(\omega), B(\omega))d\omega - \bar{\mathcal{F}}(X)$$

for $t \in \mathbb{R}$ and $X, \zeta, B \in \mathbb{H}$. We obtain that $\bar{L}(t,X,\cdot)$ is the Legendre transform of $\bar{H}(t,X,\cdot)$.

Similarly, we obtain that $L(t,\mu,\cdot)$ is the Legendre transform of $H(t,\mu,\cdot)$ if we define

$$H(t,\mu,\xi) := \int_{\mathbb{M}} h(t,x,\xi)\mu(dx) + \mathcal{F}(\mu), \quad L(t,\mu,b) := \int_{\mathbb{M}} l(t,x,b)\mu(dx) - \mathcal{F}(\mu)$$

for $t \in \mathbb{R}$, $\mu \in \mathcal{P}_2(\mathbb{M})$ and $\xi, b \in L^2(\mu)$.

For $X, Y \in \mathbb{H}$ and $T > 0$, we define

$$\bar{C}_0^T(X,Y) := \inf_\Sigma \Big\{ \int_0^T \bar{L}(t, \Sigma_t, \dot{\Sigma}_t)dt \;\; : \;\; \Sigma_0 = X,\ \Sigma_T = Y;\ \ \Sigma \in AC_2(0,T;\mathbb{H}) \Big\}.$$

For $\mu, \nu \in \mathcal{P}_2(\mathbb{M})$ we define

$$C_0^T(\mu, \nu) := \inf_{(\sigma, \mathbf{v})} \Big\{ \int_0^T L(t, \sigma_t, \mathbf{v}_t) dt \quad : \quad \sigma_0 = \mu, \ \sigma_T = \nu; \quad \sigma \in AC_2\big(0, T; \mathcal{P}_2(\mathbb{M})\big) \Big\},$$

where the infimum is performed over the set of $(\sigma, \mathbf{v})$ such that $\mathbf{v} : (0, T) \times \mathbb{M} \to \mathbb{M}$ is a Borel field such that

$$\partial_t \sigma + \nabla \cdot (\sigma \mathbf{v}) = 0 \text{ in } \mathcal{D}'((0, T) \times M). \tag{8.3}$$

Lemma 8.1. *Suppose $(\Sigma^n)_n \subset W^{1,2}(0, T; \mathbb{H})$ converges to Σ in $L^2(0, T; \mathbb{H})$ and $(\dot\Sigma^n)_n$ converges weakly to $\dot\Sigma$ in $L^2(0, T; \mathbb{H})$.*

(i) If $\bar l \equiv 0$ (meaning we drop $\bar l$ from the definition of $\bar L$) then

$$\lim_{n\to\infty} \int_0^T \bar L(t, \Sigma_t^n, \dot\Sigma_t^n) dt = \int_0^T \bar L(t, \Sigma_t, \dot\Sigma_t) dt. \tag{8.4}$$

(ii) If we further assume that $(\dot\Sigma^n)_n$ converges pointwise to $\dot\Sigma$ and there exists a nonnegative function $g \in L^1(0, T; \mathbb{H})$ such that

$$\int_{(0,T)\times\Omega} \bar l(\Sigma_t^n, \dot\Sigma_t^n) dt d\omega \le \int_{(0,T)\times\Omega} g(t, \omega) dt d\omega, \tag{8.5}$$

then (8.4) continues to hold for any $\bar l$ which satisfies (8.1 – 8.2).

Proof. (i) Since $\bar{\mathcal{F}}$ is Lipschitz, applying Jensen's inequality, we have

$$|\bar{\mathcal{F}}(\Sigma_t^n) - \bar{\mathcal{F}}(\Sigma_t)| \le \mathrm{Lip}(\bar{\mathcal{F}}) \|\Sigma(t, \cdot) - \Sigma^n(t, \cdot)\|.$$

Hence,

$$\lim_{n\to\infty} \int_0^T |\bar{\mathcal{F}}(\Sigma_t^n) - \bar{\mathcal{F}}(\Sigma_t)| dt \le \mathrm{Lip}(\bar{\mathcal{F}}) \lim_{n\to\infty} \Big(\int_0^T \|\Sigma^n(t, \cdot) - \Sigma(t, \cdot)\|^2 dt \Big)^{\frac{1}{2}} = 0. \tag{8.6}$$

We have

$$\begin{aligned} \langle f(t, \Sigma^n(t,\cdot), \dot\Sigma^n(t,\cdot)\rangle - \langle f(t, \Sigma(t,\cdot), \dot\Sigma(t,\cdot)\rangle \quad &= \quad \langle f(t, \Sigma^n(t,\cdot), \dot\Sigma^n(t,\cdot)\rangle - \langle f(t, \Sigma(t,\cdot), \dot\Sigma^n(t,\cdot)\rangle \\ &+ \quad \langle f(t, \Sigma(t,\cdot), \dot\Sigma^n(t,\cdot)\rangle - \langle f(t, \Sigma(t,\cdot), \dot\Sigma(t,\cdot)\rangle. \end{aligned}$$

Since $(\dot\Sigma^n)_n$ is weakly pre–compact, its norm is bounded by a finite number, say, M. We have

$$\begin{aligned} & \Big| \int_0^T \Big(\langle f(t, \Sigma^n(t,\cdot), \dot\Sigma^n(t,\cdot)\rangle - \langle f(t, \Sigma(t,\cdot), \dot\Sigma(t,\cdot)\rangle \Big) dt \Big| \\ \le \quad & M \,\mathrm{Lip}(f) \Big(\int_0^T \|\Sigma^n(t,\cdot) - \Sigma(t,\cdot)\|^2 dt \Big)^{\frac{1}{2}} \\ + \quad & \Big| \int_{(0,T)\times\Omega} f(t, \Sigma(t,\omega)) \cdot (\dot\Sigma^n(t,\omega) - \dot\Sigma(t,\omega)) dt d\omega \Big|. \end{aligned}$$

By the fact that $f(t, \Sigma(t, \omega)) \in L^2((0, T) \times \Omega, \mathbb{M})$, the last expression tends to 0 as n tends to ∞. We use (8.6) to conclude the proof of (i).

(ii) Further assume that $(\dot\Sigma^n)_n$ converges pointwise to $\dot\Sigma$ and (8.5). We apply the monotone convergence theorem to obtain that

$$\lim_{n\to\infty} \int_{(0,T)\times\Omega} \bar l(\Sigma_t^n, \dot\Sigma_t^n) dt d\omega = \int_{(0,T)\times\Omega} \bar l(\Sigma_t, \dot\Sigma_t) dt d\omega.$$

This, together with (i), yields the proof of (ii). □

Observe that if $\Sigma \in W^{1,1}(0,T;\mathbb{H})$ is such that

$$\int_0^T ds \int_\Omega \bar{L}(s,\Sigma_s,\dot{\Sigma}_s)d\omega \leq \bar{C}_0 < \infty, \tag{8.7}$$

then, by (8.1), $\Sigma \in W^{1,2}(0,T;\mathbb{H})$.

Lemma 8.2 (Approximation by time–Lipschitz/Space–bounded functions)**.** *Let $X, Y \in \mathbb{H}$ be such that $|X|, |Y| \leq C$ for a constant $C > 0$. Let $\Sigma \in W^{1,2}(0,T;\mathbb{H})$ be such that $\Sigma_0 = Y$, $\Sigma_T = X$ and (8.7) holds. For any $\epsilon > 0$ and $\delta > 0$, there exists*

$$\bar{\Sigma} \in W^{1,\infty}(0,T;\mathbb{H}) \cap L^\infty((0,T)\times\Omega;\mathbb{M})$$

such that $\bar{\Sigma}_0 = Y$, $\|\bar{\Sigma}_T - X\| \leq \delta$ and

$$\int_0^T \bar{L}(s,\bar{\Sigma}_s,\dot{\bar{\Sigma}}_s)ds \leq \epsilon + \int_0^T \bar{L}(s,\Sigma_s,\dot{\Sigma}_s)ds.$$

Proof. For $r > 0$ we define

$$\phi^r(t) = \begin{cases} -r & \text{if} \quad t \leq -r \\ t & \text{if} \quad r \leq t \leq r \\ r & \text{if} \quad t \geq r \end{cases}, \quad \forall\ s \in \mathbb{R}, \qquad \Phi^r(x) = (\phi^r(x_1), \cdots, \phi^r(x_d)) \quad \forall\ s \in \mathbb{M}\,.$$

Set

$$\Sigma_t^r(\omega) := \Sigma_0(\omega) + \int_0^t \Phi^r(\dot{\Sigma}_s(\omega))ds\,.$$

We have

$$\Sigma_0^r = Y, \quad |\Sigma_t^r| \leq \|\Sigma_0\|_\infty + Tr, \quad |\dot{\Sigma}_t^r| \leq |\dot{\Sigma}_t|, \quad |\dot{\Sigma}_t^r| \leq r. \tag{8.8}$$

For each $\omega \in \Omega$ we define

$$E^r(\omega) := \bigcup_{i=1}^d \Big\{s \in [0,T] \ : \ |\dot{\Sigma}_s^i(\omega)| \geq r\Big\}$$

and

$$E^r := \bigcup_{i=1}^d \Big\{(s,\omega) \in [0,T]\times\Omega \ : \ |\dot{\Sigma}_s^i(\omega)| \geq r\Big\}.$$

Since $|\dot{\Sigma}|$ is square integrable, we have

$$\lim_{r\to\infty}(\mathcal{L}^1 \otimes \mathcal{L}^d)(E^r) = \lim_{r\to\infty}\int_{E^r} \|\dot{\Sigma}_t(\omega)\|^2 dt d\omega = 0. \tag{8.9}$$

By the fact that

$$|\Sigma_t^r(\omega) - \Sigma_t(\omega)| \leq \int_{[0,t]\cap E^r(\omega)} |\dot{\Sigma}_s(\omega)|ds,$$

we obtain the time pointwise estimate

$$\|\Sigma_t^r - \Sigma_t\|^2 \leq (\mathcal{L}^1 \otimes \mathcal{L}^d)(E^r)\int_{E^r} \|\dot{\Sigma}_t(\omega)\|^2 dt d\omega. \tag{8.10}$$

Thanks to (8.9) we conclude that

$$\lim_{r\to\infty}\|\Sigma_t^r - \Sigma_t\|^2 = 0 \quad \forall\ t \in [0,T]. \tag{8.11}$$

Furthermore,

$$\int_0^T \|\Sigma_t^r - \Sigma_t\|^2 dt \leq T(\mathcal{L}^1 \otimes \mathcal{L}^d)(E^r)\int_{E^r} \|\dot{\Sigma}_t(\omega)\|^2 dt d\omega. \tag{8.12}$$

Similarly,

$$\int_0^T \|\dot{\Sigma}_t^r - \dot{\Sigma}_t\|^2 dt \leq \int_{E^r} |\dot{\Sigma}_s(\omega)|^2 ds d\omega. \tag{8.13}$$

Thus, $\{\Sigma^r\}_r$ converges to Σ in $W^{1,2}(0,T;\mathbb{H})$. Thanks to (8.1) and the second inequality in (8.8), we infer

$$\bar{l}(\Sigma_t^r, \dot{\Sigma}_t^r) \leq \kappa_2 \bar{l}(\Sigma_t, \dot{\Sigma}_t) + \kappa_3.$$

We apply Lemma 8.1 to conclude the proof. □

Lemma 8.3 (Approximation by time-C^1/space-C^∞ bounded functions)**.** *Let $X, Y \in \mathbb{H}$ and let $\Sigma \in W^{1,2}(0,T;\mathbb{H})$ be such that Σ, $\dot{\Sigma} \in L^\infty((0,T)\times\Omega;\mathbb{M})$, $\Sigma_0 = Y$, $\Sigma_T = X$ and (8.7) holds. Suppose $X, Y \in L^\infty(\Omega;\mathbb{M})$. Then for any $\epsilon > 0$ and $\bar{\delta} > 0$, there exist*

$$\bar{\Sigma} \in W^{1,\infty}(0,T;\mathbb{H}) \cap L^\infty((0,T)\times\Omega;\mathbb{M}) \text{ and } \Sigma^* \in C^\infty([0,T];\mathbb{H}) \cap L^\infty((0,T)\times\Omega;\mathbb{M})$$

such that

(i) $\bar{\Sigma}_0 = Y$, $\bar{\Sigma}_T = X$, $|\bar{\Sigma}| \leq \|\Sigma\|_\infty$, *and* $|\dot{\bar{\Sigma}}| \leq \|\dot{\Sigma}\|_\infty$,

$$\int_0^T \bar{L}(s, \bar{\Sigma}_s, \dot{\bar{\Sigma}}_s) ds < \epsilon + \int_0^T \bar{L}(s, \Sigma_s, \dot{\Sigma}_s) ds.$$

(ii) $\|\bar{\Sigma}_0 - Y\|, \|\bar{\Sigma}_T - X\| \leq \bar{\delta}$, $|\bar{\Sigma}| \leq \|\Sigma\|_\infty$, $|\dot{\bar{\Sigma}}| \leq \|\dot{\Sigma}\|_\infty$ *and*

$$\int_0^T \bar{L}(s, \Sigma_s^*, \dot{\Sigma}_s^*) ds \leq \epsilon + \int_0^T \bar{L}(s, \Sigma_s, \dot{\Sigma}_s) ds. \tag{8.14}$$

Proof. (i) Assume $|\Sigma| \leq R_1$ and $|\dot{\Sigma}| \leq R_2$. Choose $n > 1$ integer and set

$$\delta := \frac{T}{n}, \quad t_i := \delta i, \quad \Sigma_s^\delta := \Big(1 - \frac{s - t_i}{\delta}\Big)\Sigma_{t_i} + \frac{s - t_i}{\delta}\Sigma_{t_{i+1}} \quad \forall s \in [t_i, t_{i+1}] \quad \forall\, i = 0, \cdots, n.$$

Clearly, $|\Sigma^\delta| \leq R_1$. If $s \in [t_i, t_{i+1}]$ then

$$\|\Sigma_{t_i} - \Sigma_s\|^2 = \Big\|\int_{t_i}^{t_{i+1}} \dot{\Sigma}_\tau d\tau\Big\|^2 \leq \delta \int_\Omega \int_{t_i}^{t_{i+1}} |\dot{\Sigma}_\tau|^2 d\tau d\omega \leq \delta \int_0^T \|\dot{\Sigma}_t\|^2 dt.$$

Thus,

$$\|\Sigma_s^\delta - \Sigma_s\|^2 \leq \delta \int_0^T \|\dot{\Sigma}_t\|^2 dt\,. \tag{8.15}$$

If $s \in [t_i, t_{i+1}]$ then

$$|\dot{\bar{\Sigma}}_s| = \frac{|\Sigma_{t_{i+1}} - \Sigma_{t_i}|}{\delta} = \frac{\int_{t_i}^{t_{i+1}} |\dot{\Sigma}_t| dt}{\delta} \leq \frac{\delta R_2}{\delta} = R_2.$$

This, together with (8.15), implies that $\{\Sigma^\delta\}_\delta$ converges to Σ in $L^2(0,T;\mathbb{H})$ and $\{\Sigma^\delta\}_\delta$ is weakly pre–compact in $W^{1,2}(0,T;\mathbb{H})$. Hence, $\{\dot{\Sigma}^\delta\}_\delta$ weakly converges to $\dot{\Sigma}$ in $L^2(0,T;\mathbb{H})$. By Lemma 8.2 (i)

$$\lim_{\delta\to 0} \int_0^T \Big(\langle f(t, \Sigma^\delta), \dot{\Sigma}^\delta\rangle - \bar{\mathcal{F}}(\Sigma^\delta)\Big) dt = \int_0^T \Big(\langle f(t, \Sigma), \dot{\Sigma}\rangle - \bar{\mathcal{F}}(\Sigma)\Big) dt. \tag{8.16}$$

Since $\bar{l}(\Sigma_{t_i}^*, \cdot)$ is convex, by Jensen's inequality

$$\int_{t_i}^{t_{i+1}} \bar{l}(\Sigma_{t_i}, \dot{\Sigma}_s) ds \geq \delta \bar{l}(\Sigma_{t_i}, \dot{\Sigma}_s^\delta) = \int_{t_i}^{t_{i+1}} \bar{l}(\Sigma_{t_i}, \dot{\Sigma}_s^\delta) ds. \tag{8.17}$$

Note that

$$\begin{aligned}\int_{t_i}^{t_{i+1}} \bar{l}(\Sigma_s,\dot{\Sigma}_s)ds - \int_{t_i}^{t_{i+1}} \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)ds &= \int_{t_i}^{t_{i+1}} \Big(\bar{l}(\Sigma_s,\dot{\Sigma}_s) - \bar{l}(\Sigma_{t_i},\dot{\Sigma}_s)\Big)ds \\ &+ \int_{t_i}^{t_{i+1}} \Big(\bar{l}(\Sigma_{t_i},\dot{\Sigma}_s) - \bar{l}(\Sigma_{t_i},\dot{\Sigma}^\delta_s)\Big)ds \\ &+ \int_{t_i}^{t_{i+1}} \Big(\bar{l}(\Sigma_{t_i},\dot{\Sigma}^\delta_s) - \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)\Big)ds.\end{aligned}$$

Thanks to (8.17) we conclude that

$$\begin{aligned}\int_{t_i}^{t_{i+1}} \bar{l}(\Sigma_s,\dot{\Sigma}_s)ds - \int_{t_i}^{t_{i+1}} \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)ds &\geq \int_{t_i}^{t_{i+1}} \Big(\bar{l}(\Sigma_s,\dot{\Sigma}_s) - \bar{l}(\Sigma_{t_i},\dot{\Sigma}_s)\Big)ds \\ &+ \int_{t_i}^{t_{i+1}} \Big(\bar{l}(\Sigma_{t_i},\dot{\Sigma}^\delta_s) - \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)\Big)ds. \qquad (8.18)\end{aligned}$$

We use (8.2) and the fact that $|\dot{\Sigma}| \leq R_2$ to conclude that

$$\Big|\bar{l}(\Sigma_{t_i},\dot{\Sigma}^\delta_s) - \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)\Big| \leq \tilde{e}(R_2)|\Sigma_{t_i} - \Sigma^\delta_s| \leq \tilde{e}(R)|\Sigma_{t_i} - \Sigma_{t_{i+1}}| \leq \tilde{e}(R_2)\int_{t_i}^{t_{i+1}} |\dot{\Sigma}_\tau| d\tau$$

for any $s \in [t_i, t_{i+1}]$. Thus,

$$\int_{t_i}^{t_{i+1}} |\bar{l}(\Sigma_{t_i},\dot{\Sigma}^\delta_s) - \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)|ds \leq \delta\tilde{e}(R_2)\int_{t_i}^{t_{i+1}} |\dot{\Sigma}_\tau| d\tau. \qquad (8.19)$$

Similarly,

$$\int_{t_i}^{t_{i+1}} |\bar{l}(\Sigma_s,\dot{\Sigma}_s) - \bar{l}(\Sigma_{t_i},\dot{\Sigma}_s)|ds \leq \tilde{e}(R_2)|\Sigma_s - \Sigma_{t_i}|dt \leq \delta\tilde{e}(R_2)\int_{t_i}^{t_{i+1}} |\dot{\Sigma}_\tau| d\tau. \qquad (8.20)$$

We combine (8.18 – 8.20) to obtain

$$\int_0^T ds\int_\Omega \Big(\bar{l}(\Sigma_s,\dot{\Sigma}_s) + 2\delta\tilde{e}(R_2)|\dot{\Sigma}_s|\Big)d\omega \geq \int_0^T ds\int_\Omega \bar{l}(\Sigma^\delta_s,\dot{\Sigma}^\delta_s)d\omega, \qquad (8.21)$$

which, together with (8.16), concludes the proof of (i).

(ii) By a first approximation, we may assume without loss of generality that Σ satisfies the same properties as $\bar{\Sigma}$ found in (i). First, extend $\Sigma(\cdot,\omega)$ by setting

$$\Sigma_t(\omega) := \Sigma(t,\omega) := \begin{cases} \Sigma_0(\omega) & \text{if} \quad t < 0 \\ \Sigma_T(\omega) & \text{if} \quad t > T. \end{cases}$$

Let $\varrho \in C^\infty_c(\mathbb{R})$ be a non negative probability density supported in $[-1,1]$. Set

$$\bar{\varrho}^\epsilon(t) := \epsilon^{-1}\bar{\varrho}(\epsilon^{-1}t), \qquad \Sigma^\epsilon_t(\omega) := \bar{\varrho}^\epsilon * \Sigma(\cdot,\omega), \quad \forall (t,\omega) \in \mathbb{R}\times\Omega.$$

We have

$$C^\infty([0,T];\mathbb{H}) \cap L^\infty((0,T)\times\Omega;\mathbb{M}),$$

and

$$|\Sigma^\epsilon| \leq ||\Sigma||_\infty, \quad |\dot{\Sigma}^\epsilon| \leq ||\dot{\Sigma}||_\infty, \quad |\Sigma^\epsilon(t,\omega) - \Sigma(t,\omega)| \leq \epsilon||\dot{\Sigma}||_\infty \int_{\mathbb{R}} |s|\bar{\varrho}(s)ds\,. \qquad (8.22)$$

Furthermore,

$$\lim_{\epsilon\to 0^+} ||\dot{\Sigma}^\epsilon - \dot{\Sigma}|| = 0. \qquad (8.23)$$

By (8.22), (8.23) and Lemma 8.1 we can choose ϵ small enough and set $\Sigma^* := \Sigma^\epsilon$ to see that (8.14) holds. □

Lemma 8.4 (Existence of Eulerian coordinates)**.** *Let $X, Y \in \mathbb{H} \cap L^\infty(\Omega; \mathbb{M})$ and let $\Sigma \in W^{1,2}(0, T; \mathbb{H})$ be such that $\Sigma,\ \dot\Sigma \in L^\infty((0,T)\times\Omega; \mathbb{M})$ and $\Sigma_0 = Y, \Sigma_T = X$ and (8.7) holds. Then, for any $\epsilon > 0$ and $\delta > 0$, there exist*

$$\Sigma^\epsilon \in C^\infty([0,T]\times\Omega; \mathbb{M}) \text{ such that } \dot\Sigma^\epsilon \in L^\infty((0,T)\times\Omega; \mathbb{M}) \tag{8.24}$$

and the pair $(\sigma^\epsilon, \mathbf{v}^\epsilon)$ consisting of the path of Borel probability densities $\sigma^\epsilon(t,\cdot),\ t \in [0,T]$ and its corresponding velocity field $\mathbf{v}^\epsilon$ such that

$$0 < \sigma^\epsilon \in W^{1,\infty}([0,T]; C^\infty(\mathbb{M})), \quad \mathbf{v}^\epsilon \in C^\infty([0,T]\times\mathbb{M}; \mathbb{M}) \tag{8.25}$$

and

$$\partial_t\sigma^\epsilon + \nabla\cdot(\sigma^\epsilon\mathbf{v}^\epsilon) = 0 \text{ in } \mathcal{D}'((0,T)\times M). \tag{8.26}$$

We also have that for any ball $B \subset \mathbb{M}$ there exists a positive R such that

$$\int_0^T \Big(\sup_B |v_t^\epsilon| + \mathrm{Lip}(\mathbf{v}_t^\epsilon, B)\Big)dt, \ \int_0^T L(t, \sigma_t^\epsilon, \mathbf{v}_t^\epsilon)dt < \infty, \quad |\mathbf{v}^\epsilon| \le R. \tag{8.27}$$

Furthermore, there exists a unique solution to the initial value differential equation

$$\dot S_t^\epsilon = \mathbf{v}_t^\epsilon \circ S_t^\epsilon, \quad S_0 = \mathbf{id} \tag{8.28}$$

and, if we set $\Sigma_t^\epsilon := S_t^\epsilon \circ \Sigma_0^\epsilon$, we have

$$\int_0^T \bar L(t, \Sigma_t^\epsilon, \dot\Sigma_t^\epsilon)dt \le \int_0^T \bar L(t, \Sigma_t, \dot\Sigma_t)dt + \epsilon c_1 MT + T m_f(\epsilon\|\mathbf{id}\|_\varrho) \tag{8.29}$$

for some positive M, and

$$W_2(\sharp(X), \sigma_T^\epsilon), \quad W_2(\sharp(Y), \sigma_0^\epsilon) \le \delta. \tag{8.30}$$

Proof. By a first approximation argument, thanks to Lemma 8.3, we may assume that Σ equals the Σ^* found there and so, in particular,

$$\Sigma \in C^\infty([0,T]; \mathbb{H}) \cap L^\infty((0,T)\times\mathbb{M}; \mathbb{M}), \quad |\Sigma|,\ |\dot\Sigma| \le R$$

for some $R > 0$. Set

$$\varrho^\epsilon(z) := (2\pi\epsilon)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2\epsilon}}$$

and define

$$\sigma^\epsilon(t,x) := \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega))d\omega$$

and

$$E^\epsilon(t,x) := \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega))\dot\Sigma_t(\omega)d\omega \quad \text{and} \quad \mathbf{v}^\epsilon(t,x) := \frac{E^\epsilon(t,x)}{\sigma^\epsilon(t,x)}.$$

Observe that (8.24) holds and

$$\sigma^\epsilon \in C^\infty([0,T]\times\mathbb{M}), \quad E^\epsilon \in C^\infty([0,T]\times\mathbb{M}; \mathbb{M}), \tag{8.31}$$

and so, since $\sigma^\epsilon > 0$, we reach the second assertion in (8.25). We also obtain the first inequality in (8.27). Since $|\dot\Sigma| \le R$, the third inequality in (8.27) holds. Direct computations give (8.26). We combine the latter property together with the first and third inequalities in (8.27) and apply Lemma 8.1.4 [3] to conclude that the differential equation (8.28) admits a unique solution S^ϵ. Set

$$\bar\sigma_t^\epsilon := S_{t\sharp}^\epsilon \sigma_0^\epsilon$$

to see, by (8.28), that

$$\partial_t\bar\sigma^\epsilon + \nabla\cdot(\bar\sigma^\epsilon\mathbf{v}^\epsilon) \text{ in } \mathcal{D}'((0,T)\times M).$$

Since $\bar{\sigma}_0^\epsilon = \sigma_0^\epsilon$, thanks to (8.27) we may apply Proposition 8.1.7 [3] to infer that $\bar{\sigma}_t^\epsilon = \sigma_t^\epsilon$ for any $t \in [0, 1]$.

Let $\Sigma_0^\epsilon \in \mathbb{H}$ be such that $\sharp(\Sigma_0^\epsilon) = \sigma_0^\epsilon$. Although it does not matter here, in fact, the optimal transportation theory ensures that we can choose Σ_0^ϵ to be the gradient of a convex function. Set

$$\Sigma_t^\epsilon := S_t^\epsilon \circ \Sigma_0^\epsilon.$$

Observe that

$$\sharp(\Sigma_t^\epsilon) = \sigma_t^\epsilon \quad \text{and} \quad \dot{\Sigma}_t^\epsilon = \mathbf{v}_t^\epsilon \circ \Sigma_t^\epsilon \quad \forall\, t \in [0, T]. \tag{8.32}$$

Since by (8.25) and (8.28), $S^\epsilon \in C^\infty([0, T] \times \mathbb{M}; \mathbb{M})$, we obtain the first assertion in (8.24). The second one follows from the inequality in (8.27) and the first identity in (8.28).

In order to estimate $\int_0^T L(t, \sigma_t^\epsilon, \mathbf{v}_t^\epsilon)dt$ we introduce the function

$$l_0(t, x, m, \rho) = \begin{cases} l(t, x, \dfrac{m}{\rho})\rho & \text{if} \quad \rho > 0 \\ \quad 0 & \text{if} \quad \rho = 0, m = 0 \\ \quad \infty & \text{if} \quad (\rho = 0, m \neq 0) \quad \text{or} \quad \rho < 0\,. \end{cases}$$

One checks that for any $t \in \mathbb{R}$ and $x \in \mathbb{M}$, the bi–Legendre transform of $l_0(t, x, \cdot, \cdot)$ equals $l_0(t, x, \cdot, \cdot)$. Hence, $l_0(t, x, \cdot, \cdot)$ is a convex lower semicontinuous function. Furthermore, it is 1–homogeneous. We have

$$l_0(t, x, E^\epsilon(t, x), \sigma^\epsilon(t, x)) = l_0\Big(t, x, \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega))(\dot{\Sigma}_t(\omega), 1)d\omega\Big). \tag{8.33}$$

We use Jensen's inequality to conclude that

$$l_0(t, x, E^\epsilon(t, x), \sigma^\epsilon(t, x)) \leq \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega)) l_0(t, x, (\dot{\Sigma}_t(\omega), 1))d\omega\,. \tag{8.34}$$

We combine (8.33) and (8.34) to conclude that

$$l(t, x, \mathbf{v}^\epsilon(t, x))\sigma^\epsilon(t, x) \leq \int_\Omega \varrho^\epsilon(t, x - \Sigma_t(\omega)) l(t, x, \dot{\Sigma}_t(\omega))d\omega.$$

We exploit (8.1) and the third inequality in (8.27) to obtain

$$\begin{aligned} l(t, x, \mathbf{v}^\epsilon(t, x))\sigma^\epsilon(t, x) \quad &\leq \quad \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega)) l(t, \Sigma_t(\omega), \dot{\Sigma}_t(\omega))d\omega \\ &+ \quad M \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega))|x - \Sigma_t(\omega)|d\omega, \end{aligned} \tag{8.35}$$

where

$$M := \kappa_3 + \kappa_2 \sup_{|v| \leq R} \tilde{l}(0, 0, v).$$

Observe that if we set $\sigma_t := \sharp(\Sigma_t)$, then

$$\int_\mathbb{M} dx \int_\Omega \varrho^\epsilon(x - \Sigma_t(\omega))|x - \Sigma_t(\omega)|d\omega = \int_\mathbb{M} dx \int_\mathbb{M} \varrho^\epsilon(x - y)|x - y|\sigma_t(dy) = \epsilon c_1, \tag{8.36}$$

where $c_1 := \int_\mathbb{M} |x|\varrho_1(x)dx$. We combine (8.35 – 8.36) to conclude that

$$\int_{(0,T)\times\mathbb{M}} l(t, x, \mathbf{v}^\epsilon(t, x))\sigma^\epsilon(t, x)dtdx \leq \epsilon M T c_1 + \int_{(0,T)\times\Omega} l(t, \Sigma_t(\omega), \dot{\Sigma}_t(\omega))dtd\omega\,. \tag{8.37}$$

Since $\sigma_t^\epsilon = \varrho^\epsilon * \sigma_t$, we have (cf. e.g. Lemma 5.19 [14])

$$W_2^2(\sigma_t, \sigma_t^\epsilon) \leq \epsilon^2 \|\mathbf{id}\|_\varrho^2,$$

which proves (8.30). Furthermore, since $\mathcal{F}$ satisfies ($\mathcal{F}3$) ,

$$\Big|\int_0^T \bar{\mathcal{F}}(\Sigma_t^\epsilon)dt - \int_0^T \mathcal{F}(\sigma_t)dt\Big| = \Big|\int_0^T \mathcal{F}(\sigma_t^\epsilon)dt - \int_0^T \mathcal{F}(\sigma_t)dt\Big| \leq T m_f(\epsilon\|\mathbf{id}\|_\varrho).$$

This, together with (8.37) implies

$$\int_0^T L(t,\sigma_t^\epsilon,\mathbf{v}_t^\epsilon)dt \leq \int_0^T \bar{L}(t,\Sigma_t,\dot{\Sigma}_t)dt + \epsilon c_1 MT + T m_f(\epsilon\|\mathbf{id}\|_\varrho). \tag{8.38}$$

By (8.32) we have

$$\int_0^T L(t,\sigma_t^\epsilon,\mathbf{v}_t^\epsilon)dt = \int_0^T \bar{L}(t,\Sigma_t^\epsilon,\Sigma_t^\epsilon)dt.$$

This, together with (8.38), is all we need to conclude the proof. □

Theorem 8.5 (Eulerian coordinates with Lipschitz velocity)**.** *Let $X, Y \in \mathbb{H}\cap L^\infty(\Omega;\mathbb{M})$ and let $\Sigma \in W^{1,2}(0,T;\mathbb{H})$ be such that $\Sigma_0 = Y$, $\Sigma_T = X$. Then for any $\epsilon > 0$ and $\delta > 0$, there exists*

$$\Sigma^* \in W^{1,2}(0,T;\mathbb{H}) \text{ such that } \dot{\Sigma}^* \in C([0,T]\times\Omega;\mathbb{M}), \tag{8.39}$$

and

$$\mathbf{v}^* \in W^{1,\infty}((0,T)\times\mathbb{M};\mathbb{M})$$

such that

$$\dot{\Sigma}_t^* = \mathbf{v}_t^* \circ \Sigma_t^*$$

and

$$\partial_t\sigma^* + \nabla\cdot(\sigma^*\mathbf{v}^*) = 0 \text{ in } \mathcal{D}'((0,T)\times\mathbb{M}).$$

Furthermore,

$$W_2(\sharp(Y),\sigma_0^*), \quad W_2(\sharp(X),\sigma_T^*) \leq \delta$$

$$\int_0^T L(s,\sigma_s^*,\mathbf{v}_s^*)ds = \int_0^T \bar{L}(s,\Sigma_s^*,\dot{\Sigma}_s^*)ds \leq \epsilon + \int_0^T ds\int_\Omega \bar{L}(s,\Sigma_s,\dot{\Sigma}_s)ds.$$

Proof. If (8.7) fails, there is nothing to prove. Assume in the sequel that (8.7) holds. We apply the successive approximation results in Lemmas 8.2, 8.3 and 8.4 to assume, without loss of generality, that Σ satisfies the same properties as Σ^ϵ exhibited in Lemma 8.4. More precisely,

$$\Sigma \in W^{1,\infty}(0,T;\mathbb{H}),$$

$$\Sigma \in C^\infty([0,T];\mathbb{H}), \quad \dot{\Sigma} \in L^\infty((0,T)\times\Omega;\mathbb{M}). \tag{8.40}$$

Also, there are

$$0 < \sigma \in C^\infty([0,T]\times\mathbb{M}), \quad \mathbf{v} \in C^\infty([0,T]\times\mathbb{M};\mathbb{M}) \tag{8.41}$$

such that

$$\partial_t\sigma + \nabla\cdot(\sigma\mathbf{v}) = 0 \text{ in } \mathcal{D}'((0,T)\times\mathbb{M}), \tag{8.42}$$

and for any ball $B \subset \mathbb{M}$ there exists $0 < R < \infty$ such that

$$\int_0^T \Big(\sup_B |v_t| + \mathrm{Lip}(\mathbf{v}_t, B)\Big)dt, \int_0^T L(t,\sigma_t,\mathbf{v}_t)dt < \infty, \quad |\mathbf{v}| \leq R. \tag{8.43}$$

Furthermore, there exists a unique solution to the initial value differential equation

$$\dot{S}_t = \mathbf{v}_t \circ S_t, \quad S_0 = \mathbf{id} \tag{8.44}$$

and $\Sigma_t = S_t \circ \Sigma_0$.

Let $\Phi^r \in C^\infty(\mathbb{M};\mathbb{M})$ be such that $\mathrm{Lip}(\Phi^r) \leq 2$, $|\Phi^r| \leq r+2$, and

$$\Phi^r(x) = \begin{cases} x & \text{if} \quad |x| \leq r \\ (r+2)\dfrac{x}{|x|} & \text{if} \quad |x| \geq r+2 \end{cases}$$

and set

$$\mathbf{v}^r(t,x) := \mathbf{v}\big(t, \Phi^r(x)\big).$$

Observe that

$$\mathbf{v}^r \in C^\infty([0,T]\times\mathbb{M};\mathbb{M}), \quad |\mathbf{v}^r| \leq \|\mathbf{v}\|_\infty, \quad \mathrm{Lip}(\mathbf{v}^r) \leq 2\mathrm{Lip}(\mathbf{v}|_{[0,T]\times B_{r+2}}), \tag{8.45}$$

where B_{r+2} is the closed ball of radius $r+2$ centered at the origin. Let $S^r \in C^\infty([0,T]\times\mathbb{M};\mathbb{M})$ be the unique solution to the differential equation

$$\dot S^r_t = \mathbf{v}^r_t \circ S^r_t, \quad S^r_0 = \mathbf{id} \tag{8.46}$$

and set $\Sigma^r_t := S^r_t \circ \Sigma_0$. Observe that S^r_t maps B_{r-TR} into B_r and so, $S^r_t(x) = S_t(x)$ for all $x \in B_{r-TR}$. Thus, if we set

$$\mathbb{M} \setminus E^r := \{\omega \in \Omega \ : \ |\Sigma_0(\omega)| \leq r - TR\},$$

we have

$$\Sigma^r_t(\omega) = \Sigma_t(\omega) \quad \forall \omega \in \mathbb{M}\setminus E^r, \tag{8.47}$$

and also, since $|\Sigma_0| \in L^1(\Omega)$,

$$\lim_{r\to\infty} \mathcal{L}^d(E^r) = 0. \tag{8.48}$$

Using the last inequality in (8.43) and the first one in (8.45) we have

$$\|\Sigma^r_t - \Sigma_t\|^2 = \int_{E^r} \Big| \int_0^t [\dot\Sigma^r_t(\omega) - \dot\Sigma_t(\omega)] d\tau \Big|^2 d\omega \leq 4R^2 T \mathcal{L}^d(E^r).$$

This, together with (8.48) proves that $\{\Sigma^r\}_r$ converges to Σ in $L^2(0,T;\mathbb{H})$ and

$$W_2\Big(\sharp(\Sigma^r), \sharp(\Sigma_t)\Big) = 0, \quad \forall\ t \in [0,T].$$

Similarly, since

$$\|\dot\Sigma^r_t - \dot\Sigma_t\|^2 = \int_{E^r} \Big|\dot\Sigma^r_t(\omega) - \dot\Sigma_t(\omega)\Big|^2 d\omega \leq 4R^2 T \mathcal{L}^d(E^r),$$

$\{\dot\Sigma^r\}_r$ converges to $\dot\Sigma$ in $L^2(0,T;\mathbb{H})$. By (8.1),

$$l(t, \Sigma^r_t, \dot\Sigma^r_t) \leq \kappa_2 l(0,0,\dot\Sigma^r_t) + \kappa_3 \leq \kappa_3 + \kappa_2 \sup_{|v|\leq R} l(0,0,v).$$

We use the fact that $\dot\Sigma^r_t = \mathbf{v}^r_t \circ \Sigma^r_t$ and apply Lemma 8.1 to conclude the proof of the theorem. □

Remark 8.6. *Let $t \in [0,T]$ and $X \in \mathbb{H}$ and assume $f(t,\cdot) = \nabla g(t,\cdot)$ for some Lipschitz function $g(t,\cdot) \in C^1(\mathbb{M}) \cap C_b(\mathbb{M})$. If $\zeta \in \mathbb{H}$, then*

$$\bar H(t, X, \mathrm{proj}_{\nabla F[X]}\zeta) \leq \bar H(t, X, \zeta).$$

Proof. Since $\nabla g(t,\cdot)\circ X \in F[X]$, we infer that for any $\zeta \in \mathbb{H}$,

$$\mathrm{proj}_{\nabla F[X]}\big(\zeta - \nabla g(t,\cdot)\circ X\big) = \mathrm{proj}_{\nabla F[X]}\zeta - \nabla g(t,\cdot)\circ X.$$

This concludes the proof. □

9. Appendix III: Joint measurability of parameter-dependent optimal maps

For any integer $m \geq 1$ let $\mathbb{B}_m$ denote the ball centered at the origin with $\mathcal{L}^m(\mathbb{B}_m) = 1$.

9.1. Some preliminaries

This subsection contains a Lemma and its corollary, both considered as part of the folklore in optimal transport theory.

Lemma 9.1. *Let μ_0 and μ_1 be Borel probability measures on $\mathbb{B}_k$, absolutely continuous with respect to $\mathcal{L}^k_{\mathbb{B}_k}$. Then there exist a convex function $\phi : \mathbb{R}^k \to \mathbb{R}$ and a Borel set $B \subset \mathbb{B}_k$ such that*

(i) ϕ *is Lipschitz and the range of* $\partial_i\phi$ *is contained in* $[0,1]$ *for every* $i \in \{1, \cdots, k\}$.
(ii) $\nabla\phi_\sharp\mu_0 = \mu_1$.
(iii) $\mu_0(\mathbb{B}_k \setminus B) = 0$.
(iv) *If* $\bar\phi : \mathbb{R}^k \to \mathbb{R}$ *is any convex function such that* $\bar\nabla\phi_\sharp\mu_0 = \mu_1$, *then* $\{\nabla\phi \neq \nabla\bar\phi\} \cap B = \emptyset$.

As a direct application of Lemma 9.1, we obtain the following corollary.

Corollary 9.2. *Assume $\{\mu_0^n\}_n = \{\lambda_0^n\mathcal{L}^d_{\mathbb{B}_k}\}_n$ and $\{\mu_1^n\}_n = \{\lambda_1^n\mathcal{L}^d_{\mathbb{B}_k}\}_n$ are sequences of Borel probability measures on $\mathbb{B}_k$, absolutely continuous with respect to $\mathcal{L}^k_{\mathbb{B}_k}$. Assume $\{\lambda_0^n\}_n$ converges to λ_0 in $L^1(\mathbb{B}_k)$, $\mu_0 = \lambda_0\mathcal{L}^d_{\mathbb{B}_k}$ and $\{\lambda_1^n\}_n$ converges narrowly to λ_1 in $L^1(\mathbb{B}_k)$, $\mu_1 = \lambda_1\mathcal{L}^d_{\mathbb{B}_k}$. Let $\phi_n : \mathbb{R}^k \to \mathbb{R}$ be a convex function such that $\nabla\phi_{n\sharp}\mu_0^n = \mu_1^n$ and $\partial.\phi_n(x)$ is contained in $\mathbb{B}_k$ for every $x \in \mathbb{B}_d$. Then*

$$\lim_n \|\nabla\phi_n - \nabla\phi\|_{L^2(B;\mathbb{R}^k)} = 0,$$

where ϕ and B are as in Lemma 9.1. If we further assume that $\lambda_0^n, \lambda_1^n, \lambda_0, \lambda_1$ are positive and continuous on the closure of $\mathbb{B}_k$ and ϕ_n and ϕ are differentiable on $\mathbb{B}_k$, then

$$\lim_n \nabla\phi_n(x) = \nabla\phi(x) \text{ for all } x \in \mathbb{B}_k.$$

9.2. Measurability

We skip the proof of the first proposition below, as it can be obtained by standard mollification and renormalization arguments.

Proposition 9.3. *Let $d,\ k \geq 1$ be integers and let $\lambda : \mathbb{B}_k \times \mathbb{B}_d \to [0,\infty)$ be Lebesgue measurable such that $\lambda(\cdot, x)$ is a probability density for a.e. $x \in \mathbb{B}_d$. Then there exists a sequence of strictly positive functions $\{\lambda^n\}_n \subset C^\infty(\bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d)$ and a Borel set $T \subset \mathbb{B}_d$ such that $\mathcal{L}^d(\mathbb{B}_d \backslash T) = 0$ and*

$$\lim_{n\to\infty} \|\lambda^n(\cdot, x) - \lambda(\cdot, x)\|_{L^1(\mathbb{B}_k)} = 0 \text{ for all } x \in T \tag{9.1}$$

and,

$$\lim_{n\to\infty} \|\lambda^n - \lambda\|_{L^1(\mathbb{B}_k\times\mathbb{B}_d)} = 0. \tag{9.2}$$

Proposition 9.4. *For $i = 1, 2$ let $\lambda_i \in C^\infty(\bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d)$ such that for all $x \in \bar{\mathbb{B}}_d$ the functions $\lambda_i(\cdot, x)$ are positive probability densities. For each $x \in \bar{\mathbb{B}}_d$ denote by $\Lambda(\cdot, x)$ the optimal map (with respect to the quadratic cost) pushing $\lambda_1(\cdot, x)$ forward to $\lambda_2(\cdot, x)$, and by $\tilde\Lambda(\cdot, x)$ the optimal map pushing $\lambda_2(\cdot, x)$ forward to $\lambda_1(\cdot, x)$. Then*

(i) $\Lambda(\cdot, x),\ \tilde\Lambda(\cdot, x) \in C^\infty(\bar{\mathbb{B}}_k; \bar{\mathbb{B}}_k)$ for each $x \in \bar{\mathbb{B}}_d$.

(ii) Moreover, there exists $\theta \in (0, 1)$ such that

$$\theta \mathrm{I}_k \leq \nabla_s\Lambda(s, x) \leq \theta^{-1}\mathrm{I}_k,\ \ \theta \mathrm{I}_k \leq \nabla_s\tilde\Lambda(s, x) \leq \theta^{-1}\mathrm{I}_k$$

for all $(s, x) \in \mathbb{B}_k \times \mathbb{B}_d$, where I_k denotes the $k \times k$ identity matrix.

(iii) The maps $\Lambda,\ \tilde\Lambda$ belong to $C(\bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d; \bar{\mathbb{B}}_k)$.

Proof. The first claim follows readily from a celebrated result by Caffarelli [7]. Obviously, the same regularity is enjoyed by $\tilde{\Lambda}(\cdot, x)$, which we use to denote the optimal map pushing $\lambda_2(\cdot, x)$ forward to $\lambda_1(\cdot, x)$. Next, note that $\lambda_i \in C^\infty(\bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d)$ for $i = 1, 2$ implies the existence of constants $m, C \in (0, 1)$ such that

$$m \le \lambda_i(\cdot, x) \le m^{-1}, \;\; \|\lambda_i(\cdot, x)\|_{C^{0,1}(\bar{\mathbb{B}}_k)} \le C^{-1} \text{ for all } x \in \bar{\mathbb{B}}_d \text{ and } i = 1, 2.$$

Another fundamental result on the regularity of optimal transport maps from [7] yields (*ii*) for some $\theta \in (0, 1)$ depending solely on m, C. To prove (*iii*), fix $(s_0, x_0) \in \bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d$. Since

$$|\Lambda(s, x) - \Lambda(s_0, x_0)| \le |\Lambda(s, x) - \Lambda(s_0, x)| + |\Lambda(s_0, x) - \Lambda(s_0, x_0)|,$$

we see that it suffices to prove that the continuity of $\Lambda(\cdot, x)$ at s_0 is uniform in x and $\Lambda(s_0, \cdot)$ is continuous at x_0. As far as the former is concerned, we have

$$\begin{aligned}(t - t_0) \cdot [\tilde{\Lambda}(t, y) - \tilde{\Lambda}(t_0, y)] &= \int_0^1 \nabla_s \tilde{\Lambda}((1 - \tau)t_0 + \tau t)(t - t_0) \cdot (t - t_0)\, d\tau \\ &\ge \theta |t - t_0|^2 \text{ for all } t_0,\ t \in \bar{\mathbb{B}}_k,\ y \in \bar{\mathbb{B}}_d.\end{aligned}$$

Set $y := x$ and $t := \Lambda(s, x)$, $t_0 := \Lambda(s_0, x)$ to get

$$(s - s_0) \cdot \left[\Lambda(s, x) - \Lambda(s_0, x)\right] \ge \theta \left|\Lambda(s, x) - \Lambda(s_0, x)\right|^2,$$

which implies

$$\theta^{-1}|s - s_0| \ge \left|\Lambda(s, x) - \Lambda(s_0, x)\right|,$$

so $\Lambda(\cdot, x)$ is Lipschitz uniformly with respect to x.

In order to show that $\Lambda(s_0, \cdot)$ is continuous at x_0, let $\{x_m\}_m \in \bar{\mathbb{B}}_d$ converge to x_0. Note that $\lambda(\cdot, x_m)$ converges to $\lambda(\cdot, x_0)$ uniformly. We apply Corollary 9.2 to conclude that $\{\Lambda(s_0, x_m)\}_m$ converges to $\Lambda(s_0, x_0)$. □

Assume $A \subset \mathbb{B}_d$ is a Borel set with $\mathcal{L}^d(A) > 0$. Let

$$\lambda_i : \mathbb{B}_k \times A \to [0, \infty) \qquad i = 1, 2$$

be Borel maps such that $\lambda_i(\cdot, x)$ are Borel probability densities for a.e. $x \in A$. We extend λ_i to $\mathbb{B}_k \times \mathbb{B}_d$ by setting $\lambda_i(s, x) = 1$ if $s \in \mathbb{B}_k$ and $x \in \mathbb{B}_d \setminus A$. Set

$$\mu^i = \lambda_i \mathcal{L}^{k+d}_{\mathbb{B}_k \times \mathbb{B}_d}.$$

We obtain the disintegration (μ^i_x) of μ^i given by

$$\mu^i_x = \lambda_i(\cdot, x)\mathcal{L}^k_{\mathbb{B}_k}$$

and so, $\lambda_i(\cdot, x)$ is a well–defined function for almost every $x \in \mathbb{B}_d$, say $x \in \Omega^*_d$.

Let

$$K_i := \{\lambda_i > 0\}, \qquad K_i^x = \{s \in \mathbb{B}_k \;:\; (s, x) \in K_i\} \quad (\forall\ x \in \mathbb{B}_d),$$

and denote by proj the orthogonal projection of $\mathbb{R}^{k+d}$ onto $\mathbb{R}^k$.

Lemma 9.5. *There exist a F_σ–set F_i (in fact, a countable union of compact sets) and a null measure Borel set F_i^∞ such that*

(i) $K_i = F_i \cup F_i^\infty$.
(ii) $K_i^x \times \{x\} = (F_i^x \times \{x\}) \cup ((F_i^\infty)^x \times \{x\})$, $K_i^x = F_i^x \cup (F_i^\infty)^x$.
(iii) For any $x \in \mathbb{B}_d$, F_i^x is a Borel set.
(iv) For $\mathcal{L}^d$–almost every $x \in \mathbb{B}_d$, $(F_i^\infty)^x$ is of null measure in $\mathbb{R}^k$ and so, for these x, $N_i(x) := K_i^x \setminus F_i^x$ is of null measure.
(v) $\cup_{x \in \mathbb{B}_d}(F_i^x \times \{x\}) = F_i$ *and so, it is a Borel set of full measure in K_i.*

Proof. (i) For each integer n, there exists a compact set $F_i^n \subset K_i$ such that $\mathcal{L}^{k+d}(K_i \setminus F_i^n) \leq n^{-1}$ and $F_i^n \subset K_i$. Set

$$F_i := \bigcup_{n\geq 1} F_i^n, \quad F_i^\infty := K_i \setminus F_i. \tag{9.3}$$

Note that F_i and F_i^∞ are Borel sets such that the latter set is of null measure.

(ii) We have

$$K_i^x \times \{x\} = K_i \cap (\mathbb{B}_k \times \{x\}) = (F_i \cup F_I^\infty) \cap (\mathbb{B}_k \times \{x\}) = (F_i \cap (\mathbb{B}_k \times \{x\})) \cup (F_i^\infty \cap (\mathbb{B}_k \times \{x\})).$$

This is enough to conclude the proof of (ii).

(iii) Observe that

$$F_i^x = \mathrm{proj}\,(F_i^x \times \{x\}) = \bigcup_{n\geq 1} \mathrm{proj}\,(F_i^n \cap (\mathbb{B}_k \times \{x\})).$$

Since the projection of any compact set is a compact set, we obtain that F_i^x is a Borel set as a countable union of compact sets.

(iv) By Fubini's Theorem,

$$0 = \mathcal{L}^{k+d}(F_i^\infty) = \int_{\mathbb{B}_d} \mathcal{L}^k((F_i^\infty)^x)dx$$

and so, for $\mathcal{L}^d$–almost every $x \in \mathbb{B}_d$, $(F_i^\infty)^x$ is of null measure in $\mathbb{R}^k$. Since $K_i^x = F_i^x \cup (F_i^\infty)^x$, if $s \in N_i(x) = K_i^x \setminus F_i^x$ then $x \notin F_i^x$ and so, $x \in (F_i^\infty)^x$. In other words, $N_i(x) \subset (F_i^\infty)^x$. Thus, $N_i(x)$ is of null measure in $\mathbb{R}^k$.

(v) The proof of (v) is obvious. □

Remark 9.6. *Since F_i^∞ is a set of null measure, $\int_{F_i} \lambda_i ds dx = 1$ and so, modifying λ_i on a set of null measure, we may assume that $F_i^\infty = \emptyset$.*

Theorem 9.7. *There exists a Borel map $\bar{\Lambda} : F_1 \to F_2$ such that for $\mathcal{L}^d$–almost every x, $\bar{\Lambda}(\cdot, x)$ pushes μ_x^1 forward to μ_x^2 and*

$$W_2^2(\mu_x^1, \mu_x^2) = \int_{F_1^x} |\bar{\Lambda}(s, x) - s|^2 \mu_x^1(ds).$$

Proof. By Proposition 9.3 there exist $\lambda_i^n \in C^\infty(\bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d)$ positive functions such that $\lambda_i^n(\cdot, x)$ is a probability density for every $x \in \mathbb{B}_d$ and

$$\lim_n ||\lambda_i^n - \lambda_i||_{L^1(\mathbb{B}_k \times \mathbb{B}_d)} = 0. \tag{9.4}$$

From the proof of said proposition we also see that there are Borel sets $T \subset \mathbb{B}_d$ of full measure in $\mathbb{B}_d$ and $J \subset \mathbb{B}_k \times \mathbb{B}_d$ of full measure in $\mathbb{B}_k \times \mathbb{B}_d$ such that

$$\lim_{n\to\infty} \lambda_i^n(s, x) = \lambda_i(s, x) \qquad \forall \quad (s, x) \in J, \tag{9.5}$$

and

$$||\lambda_i^n(\cdot, x) - \lambda_i(\cdot, x)||_{L^1(\mathbb{B}_k)} = 0 \quad \text{if} \quad x \in T\,. \tag{9.6}$$

For each $x \in T$, let $\Lambda_n(\cdot, x)$ be the unique optimal map that pushes $\lambda_1^n(\cdot, x)\mathcal{L}^k_{\mathbb{B}_k}$ forward to $\lambda_2^n(\cdot, x)\mathcal{L}^k_{\mathbb{B}_k}$. Then, by Proposition 9.4, Λ_n is continuous on $\bar{\mathbb{B}}_k \times \bar{\mathbb{B}}_d$ and

$$|\Lambda^n| \leq \mathrm{diam}(\mathbb{B}_k)\,. \tag{9.7}$$

For $x \in \Omega_d^* \cap T$, let $\Lambda(\cdot, x)$ be the unique optimal map that pushes μ_x^1 forward to μ_x^2. By Corollary 9.2, there exists a Borel set $B_x \subset \mathbb{B}_k$ such that $\mu_x^1(\mathbb{B}_k \setminus B_x) = 0$ and

$$\lim_n \int_{B^x} |\Lambda_n(\cdot, x) - \Lambda(\cdot, x)|^2 dx = 0, \quad \forall\ x \in \Omega_d^* \cap T. \tag{9.8}$$

Since

$$\int_{\mathbb{B}_d} \mu_x^1(\mathbb{B}_k)dx = \int_{\mathbb{B}_k \times \mathbb{B}_d} \lambda_1(s,x)dsdx = \int_{F_1} \lambda_1(s,x)dsdx = \int_{\mathbb{B}_d} \mu_x^1(F_1^x)dx,$$

we conclude that for almost every $x \in \mathbb{B}_d$, say without loss of generality $x \in \Omega_d^* \cap T$,

$$1 = \mu_x^1(\mathbb{B}_k) = \mu_x^1(F_1^x).$$

Thus, $\mu_x^1(F_1^x \triangle B^x) = 0$ for almost every $x \in \mathbb{B}_d$. This, together with (9.8), implies that for a.e. $x \in \mathbb{B}_d$

$$\lim_n \int_{F_1^x} |\Lambda_n(\cdot,x) - \Lambda(\cdot,x)|^2 dx = 0, \quad \forall\, x \in \Omega_d^* \cap T\,. \tag{9.9}$$

Set

$$f_{n,m}(x) := \|\Lambda_n(\cdot,x) - \Lambda_m(\cdot,x)\|_{L^2(F_1^x)}, \quad f_n(x) := \|\Lambda_n(\cdot,x) - \Lambda(\cdot,x)\|_{L^2(F_1^x)}.$$

Claim 1. $f_{n,m}$ is a Borel function.

Proof 1. Let F_1^l be the compact sets in (9.3) so that the sets $G_1^l = \mathbb{B}_k \setminus F_1^l$ are open. Since $\chi_{G_1^l}$ is lower semicontinuous, so is $|\Lambda_n(s,\cdot) - \Lambda_m(s,\cdot)|^2 \chi_{G_1^l}(s,\cdot)$, and so we use Fatou's Lemma to conclude that

$$x \to g_{n,m,l}^2(x) := \int_{\mathbb{B}_k} |\Lambda_n(s,x) - \Lambda_m(s,x)|^2 \chi_{G_1^l}(s,x)ds$$

is lower semicontinuous. Since

$$x \to h_{n,m}^2(x) := \int_{\mathbb{B}_k} |\Lambda_n(s,x) - \Lambda_m(s,x)|^2 ds = f_{n,m,l}^2(x) + g_{n,m,l}^2(x)$$

is continuous, $f_{n,m,l}^2$ is upper semicontinuous. By the monotone convergence theorem

$$f_{n,m}(x) = \sup_{l \geq 1} f_{n,m,l}(x)$$

and so, $f_{n,m}$ is a Borel function as a supremum of Borel functions.

Claim 2. $\chi_{T \cap \Omega_d^*} f_n$ is a Borel function.

Proof 2. We use the triangle inequality to obtain

$$|f_{n,m}(x) - f_n(x)| \leq f_m(x) \quad \text{if} \quad x \in T \cap \Omega_d^* \tag{9.10}$$

and so, by (9.9),

$$\lim_n f_{n,m}(x) = f_m(x) \quad \text{if} \quad x \in T \cap \Omega_d^*\,. \tag{9.11}$$

Thus, $\chi_{T \cap \Omega_d^*} f_n$ is a Borel function as a limit of Borel functions.

Claim 3. $(\Lambda_n)_n$ is a Cauchy sequence in $L^2(F_1)$.

Proof 3. We use again (9.9) to conclude, in view of (9.7), that $(f_n)_n$ converges to 0 in $L^2(\mathbb{B}_d)$. Since

$$\|f_{n,m}\|_{L^2(F_1)}^2 \leq 2\|f_n\|_{L^2(F_1)}^2 + 2\|f_m\|_{L^2(F_1)}^2,$$

we conclude that $(\Lambda_n)_n$ is a Cauchy sequence in $L^2(\mathbb{B}_k \times \mathbb{B}_d)$.

Thus, there exists a Borel map $\bar{\Lambda} \in L^2(F_1)$ such that $(\Lambda_n)_n$ converges to $\bar{\Lambda}$ in $L^2(F_1)$. By (9.9), there exists a Borel set $T_0 \subset T \cap \Omega_d^*$ of full measure such that, if $x \in T_0$, then

$$\int_{F_1^x} |\bar{\Lambda}(s,x) - \Lambda(s,x)|^2 ds = 0.$$

Thus, if $x \in T_0$

$$\mathcal{L}^k(\{s \in F_1^x \mid \bar{\Lambda}(s,x) \neq \Lambda(s,x)\}) = 0. \tag{9.12}$$

We use (9.12) to conclude that for $x \in T_0$ and $F \in C_b(\mathbb{R}^k)$ we have

$$\int_{F_1^x} F(\bar{\Lambda}(s,x))\lambda_1(s,x)ds = \int_{F_1^x} F(\Lambda(s,x))\lambda_1(s,x)ds = \int_{F_2^x} F(s)\lambda_2(s,x)ds.$$

Thus,

$$\bar{\Lambda}(\cdot,x)_\sharp \mu_1^x = \mu_2^x.$$

We conclude by noting that

$$\int_{F_1^x} |\bar{\Lambda}(s,x) - s|^2 \lambda_1(s,x)ds = \int_{F_1^x} |\Lambda(s,x) - s|^2 \lambda_1(s,x)ds = W_2^2(\mu_1^x, \mu_2^x).$$

□

An immediate consequence of Theorem 9.7 is:

Corollary 9.8. *Let $A \subset \mathbb{B}_d$ be a Borel set of positive volume and $\lambda_1, \lambda_2 : \mathbb{B}_k \times A \to [0,\infty)$ be Lebesgue measurable such that $\lambda_i(\cdot,x)$ is a probability density for all $x \in A$, $i = 1,2$. Then there exists a Borel map $\Lambda : \mathbb{B}_k \times A \to \mathbb{R}^k$ such that for $\mathcal{L}^d$-a.e. $x \in A$ the map $\Lambda(\cdot,x)$ pushes $\lambda_1(\cdot,x)$ forward to $\lambda_2(\cdot,x)$ optimally.*

Corollary 9.9. *Let $A, B \subset \mathbb{B}_d$ be Borel sets of positive volume. Let $f_1, f_2 : \mathbb{B}_k \times A \to A$ be Borel functions such that for all $x \in A$ we have*

$$\mathcal{L}^k(f_1(\cdot,x)^{-1}(B)) = \mathcal{L}^k(f_2(\cdot,x)^{-1}(B)) > 0.$$

Then there exists an optimal map pushing $\mathcal{L}^k|_{f_1(\cdot,x)^{-1}(B)}$ forward to $\mathcal{L}^k|_{f_2(\cdot,x)^{-1}(B)}$ which is jointly Borel on $\mathbb{B}_k \times A$.

Proof. Since $s \in f_i(\cdot,x)^{-1}(B)$ is equivalent to $(s,x) \in f_i^{-1}(B)$, we infer

$$g_i(s,x) := \mathbf{1}_{f_i(\cdot,x)^{-1}(B)}(s) = \mathbf{1}_{f_i^{-1}(B)}(s,x),\ i = 1,2,$$

are Borel in (s,x); obviously, so are

$$\lambda_i(s,x) := \frac{g_i(s,x)}{\mathcal{L}^k(f_i(\cdot,x)^{-1}(B))} = \frac{g_i(s,x)}{\int_0^1 g_i(\tau,x)d\tau},\ i = 1,2.$$

It follows (from previous theorem) that there exists a version $\Lambda(\cdot,x)$ of the optimal map pushing $\lambda_1(\cdot,x)$ forward to $\lambda_2(\cdot,x)$ such that Λ is jointly Borel. But

$$\int_0^1 g_1(\tau,x)d\tau = \int_0^1 g_2(\tau,x)d\tau$$

implies that $\Lambda(\cdot,x)$ is also the optimal map that pushes $g_1(\cdot,x)$ forward to $g_2(\cdot,x)$, so we are done. □

Acknowledgements

The research of WG was supported by NSF grants DMS–11 60 939 and DMS–17 00 202. The research of AT was supported by NSF Grant DMS–16 00 272 and by Grant #246063 from the Simons Foundation.

References

[1] L. Ambrosio, J. Feng, *On a class of first-order Hamilton-Jacobi equations in metric spaces*, J. Diff. Eq. **256** (2014), No. 7, pp. 2194–2245.

[2] L. Ambrosio, W. Gangbo, *Hamiltonian ODE's in the Wasserstein space of probability measures*, Comm. Pure Appl. Math. **61** (2008), No. 2, pp. 18–53.

[3] L. Ambrosio, N. Gigli, G. Savaré, Gradient flows in metric spaces and the Wasserstein spaces of probability measures. *Lectures in Mathematics*, ETH Zurich, Birkhäuser, 2005.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **XLIV** (1991), 375–417.

[5] Y. Brenier, W. Gangbo, *L^p approximation of maps by diffeomorphisms*, Calc. Var. Partial Differential Equations **16**, no. 2 (2003), 147–164.

[6] J.M. Borwein, Q.J. Zhu, *A survey of subdifferential calculus with applications*, Nonlin. Anal. **38** (1999), 687–773.

[7] L. Caffarelli, *Boundary regularity of maps with convex potentials*, Ann. Math. (Second Series), **144**, No. 3 (1996), pp. 453–496.

[8] L. Caravenna, S. Daneri, *The disintegration of the Lebesgue measure on the faces of a convex function*, J. Funct. Anal. **258** (2010), 3604–3661.

[9] P. Cardaliaguet, *Notes on Mean-Field Games*, lectures by P.L. Lions, Collège de France, 2010.

[10] Y T. Chow, W. Gangbo, *A partial Laplacian as an infinitesimal generator on the Wasserstein space*, Submitted.

[11] M.G. Crandall, P.-L. Lions, *Hamilton-Jacobi Equations in Infinite Dimensions I. Uniqueness of Viscosity Solutions*, J. Funct. Anal. **62** (1985), 379–396.

[12] M.G. Crandall, P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions II. Existence of viscosity solutions*, J. Funct. Anal. **65** (1986), 368–405.

[13] M.G. Crandall, P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions III*, J. Funct. Anal. **68** (1986), 214–247.

[14] W. Gangbo, H.K. Kim, T. Pacini, *Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems*, Memoirs of the AMS, vol 211 (2011), no 993 (3 of 5) 1–77.

[15] W. Gangbo, T. Nguyen, A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space*, Meth. Appl. Anal. **15**, No. 2 (2008), pp. 155–184.

[16] W. Gangbo, A. Swiech, *Metric viscosity solutions of Hamilton–Jacobi equations depending on local slopes*, Calc. Var. (2015) 54:1183–1218

[17] N. Gigli, *Second order analysis on $(\mathcal{P}_2(M), W_2)$*, Memoirs of the AMS, vol 216 (2012), no 1018 (End of volume).

[18] J. Lott, *Some Geometric Calculations on Wasserstein Space*, Commun. Math. Phys. (2008) 277, 423–437.

[19] R. Hynd, H.K. Kim, *Value functions in the Wasserstein spaces: finite time horizons*, J. Funct. Anal. **269** (2015), 968–997.

[20] D. Preiss, *Differentiability of Lipschitz functions on Banach spaces*, J. Funct. Anal. **91** (1990), 312–345.