J. Differential Equations 254 (2013) 2861–2876

Contents lists available at SciVerse ScienceDirect

Journal of Differential Equations

www.elsevier.com/locate/jde

Moreau–Yosida approximation and convergence of Hamiltonian systems on Wasserstein space

Hwa Kil Kim

Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA

ARTICLE INFO

Article history:
Received 4 June 2012
Available online 23 January 2013

Keywords:
Hamiltonian systems on Wasserstein space
Moreau–Yosida approximation
Stability

ABSTRACT

In this paper, we study the stability property of Hamiltonian systems on the Wasserstein space. Let H be a given Hamiltonian satisfying certain properties. We regularize H using the Moreau–Yosida approximation and denote it by H_τ. We show that solutions of the Hamiltonian system for H_τ converge to a solution of the Hamiltonian system for H as τ converges to zero. We provide sufficient conditions on H to carry out this process.

1. Introduction

Let $\mathcal{M}$ be the set of Borel probability measures on $\mathbb{R}^D$ with finite second moments equipped with the Wasserstein metric. We study a Hamiltonian type evolution problem in $\mathcal{M}$ of the following form:

$$\begin{cases} \dfrac{d}{dt}\mu_t + \nabla \cdot (\mathbb{J} v_t \mu_t) = 0, & t \in (0, T), \\ v_t \in \partial_- H(\mu_t) \cap T_{\mu_t}\mathcal{M}, \end{cases} \tag{1.1}$$

where the given function $H : \mathcal{M} \to (-\infty, \infty]$ is referred to as a Hamiltonian. Here $\mathbb{J} : \mathbb{R}^D \to \mathbb{R}^D$ is a matrix satisfying $\mathbb{J}v \perp v$ for all $v \in \mathbb{R}^D$. When $D = 2d$ then we can simply set $\mathbb{J}$ to be the $(2d) \times (2d)$ canonical symplectic matrix. Here, $\partial_- H(\mu)$ denotes the subdifferential of H at $\mu \in \mathcal{M}$ and $T_\mu \mathcal{M}$ is the tangent space at μ in $\mathcal{M}$ which will be defined below. There are various reasons for the terminology to be of *Hamiltonian* type. For example, (1.1) is, roughly speaking, a limit of finite dimensional Hamiltonian ODE [8]. Geometric justification was also made in [9].

The first systematic study addressing evolution problems on $\mathcal{M}$ of the Hamiltonian type was made by Ambrosio and Gangbo [1]. They studied the Hamiltonian system for locally subdifferentiable

E-mail address: hwakil@cims.nyu.edu.

0022-0396/$ – see front matter
http://dx.doi.org/10.1016/j.jde.2013.01.011

Hamiltonians and proved the existence of a solution. The theory in [1] covers a large class of systems which have recently generated a lot of interest, including the Vlasov–Poisson in one space dimension [4,12], the Vlasov–Monge–Ampère [5,8] and the semigeostrophic systems [3,6–8] are casted into the Hamiltonian type formalism.

We are interested in the stability property of Hamiltonian systems in the following sense. Let H be a given Hamiltonian. We ask ourselves whether there is any *regular* approximation H_τ of H such that solutions of (1.1) for H_τ exist and converge to a solution of the system (1.1) for H as the approximation parameter τ goes to zero.

Since the Wasserstein space is an infinite dimensional metric space, the existence of such an approximation is not a simple question. In this paper, we show that the Moreau–Yosida approximation is the one we are looking for. Let H be a Hamiltonian satisfying assumptions (H1) and (H2) whose statements will be given later. We first regularize the Hamiltonian H to obtain H_τ defined by

$$H_\tau(\mu) = \inf_{\nu\in\mathcal{M}}\left\{1/2\tau W(\mu,\nu)^2 + H(\nu)\right\}.$$

The new functional H_τ is $1/\tau$-concave even if H is not. Next, we apply the algorithm developed in [1] to solve

$$\begin{cases} \dfrac{d}{dt}\mu_t^\tau + \nabla\cdot\left(\mathbb{J}v_t^\tau\mu_t^\tau\right) = 0, & t\in(0,T), \\ v_t^\tau \in \partial^+ H_\tau\left(\mu_t^\tau\right)\cap T_{\mu_t^\tau}\mathcal{M}, \end{cases} \tag{1.2}$$

where $\partial^+ H_\tau(\mu_t^\tau)$ is the superdifferential of H_τ at μ_t^τ in the sense of [2]. Finally, we show, for any sequence τ_n converging to zero, μ^{τ_n} (up to subsequence) converges to μ which is a solution of (1.1).

Our assumptions on the Hamiltonian H allow H to be no locally subdifferentiable. Hence, our stability result allows us to construct solutions to the system (1.1) for Hamiltonians which are not everywhere subdifferentiable around the initial measure. This is not the case in [1]. At the end of this paper, we will discuss more about how the Moreau–Yosida approximation scheme is useful in the study of non-locally subdifferentiable Hamiltonians.

We briefly summarize the contents of each section. Section 2 is a preliminary on the Wasserstein space $\mathcal{M}$. In Section 3, we give an introduction to the Moreau–Yosida approximation of functionals defined on $\mathcal{M}$ and investigate some properties of it. The main feature in this section is Lemma 3.4 which is the key ingredient to prove Theorem 4.5. In Section 4, we prove our main stability result Theorem 4.5 under assumptions (H1) and (H2) on the Hamiltonian H. We show Hamiltonians considered in [1] satisfy (H1) and (H2), and so corresponding Hamiltonian systems are stable w.r.t. Moreau–Yosida approximation. Let us close this introduction by fixing notations and terminologies.

1.1. Notation and terminology

- $\mathcal{P}(\mathbb{R}^D) = \{\mu \mid \mu$ is a *Borel probability measure* on $\mathbb{R}^D\}$.
- Let $\mathcal{M}$ be the subspace of $\mathcal{P}(\mathbb{R}^D)$ with bounded second moment, *i.e.*

$$\mathcal{M} := \left\{\mu\in\mathcal{P}\left(\mathbb{R}^D\right):\ \mu\geqslant 0,\ \int_{\mathbb{R}^D} d\mu = 1,\ \int_{\mathbb{R}^D}|x|^2\,d\mu < \infty\right\}.$$

- Let $\mu\in\mathcal{P}(\mathbb{R}^D)$ and let $f:\mathbb{R}^D\to\mathbb{R}^k$! be a *Borel map*. Then $\nu := f_\#\mu$ is a *Borel measure* on $\mathbb{R}^k$ characterized by $\nu[B] = \mu[f^{-1}(B)]$ for all *Borel sets* $B\subset\mathbb{R}^k$. In this case, we say f *pushes* μ *forward to* ν.
- $C_c^\infty(\mathbb{R}^D)$ is the collection of all infinitely differentiable functions with compact support.
- We denote $C_b(\mathbb{R}^D)$ the collection of all continuous and bounded functions.

– Let $\mu_n, \mu \in \mathcal{P}(\mathbb{R}^D)$, we define μ_n converges *narrowly* to μ if

$$\int_{\mathbb{R}^D} f(x)\,d\mu_n(x) \to \int_{\mathbb{R}^D} f(x)\,d\mu(x) \quad \text{as } n \to \infty,$$

for any $f \in C_b(\mathbb{R}^D)$, *i.e.* μ_n weak* converges to μ.
– $Id : \mathbb{R}^D \to \mathbb{R}^D$ is the identity map, *i.e.* $Id(x) = x$ for all $x \in \mathbb{R}^D$.
– $\pi^i, \pi^{i,j} : \mathbb{R}^{nD} \to \mathbb{R}^D, \mathbb{R}^D \times \mathbb{R}^D$ are the standard projections, *i.e.*

$$\pi^i(x_1, x_2, \ldots, x_n) = x_i \quad \text{and} \quad \pi^{i,j}(x_1, x_2, \ldots, x_n) = (x_i, x_j).$$

– Let $\mu \in \mathcal{P}(\mathbb{R}^D)$ and let $f : \mathbb{R}^D \to \mathbb{R}^k$. We denote the L^2 norm of f by $\|f\|_\mu$, *i.e.*

$$\|f\|_\mu^2 := \|f\|_{L^2(\mu)}^2 = \int_{\mathbb{R}^D} \left|f(x)\right|^2 d\mu(x).$$

– Let $\mu \in \mathcal{P}(\mathbb{R}^D)$, we denote the support of μ by $\mathrm{supp}(\mu)$.
– Let $r > 0$ and $x \in \mathbb{R}^D$ then $B_x(r)$ denotes the open ball in $\mathbb{R}^D$ of center x and radius r.
– Let $x, y \in \mathbb{R}^D$, we denote the inner product of x and y by $\langle x, y\rangle$.

2. Wasserstein space

Recall that $\mathcal{M}$ is the subspace of $\mathcal{P}(\mathbb{R}^D)$ with bounded second moment. In this section, we show that $\mathcal{M}$ has a metric structure and we introduce a differentiable structure in $\mathcal{M}$. We refer to [2] and [11] for further details.

2.1. Wasserstein distance

Definition 2.1. Let $\mu, \nu \in \mathcal{M}$. Consider

$$W_2(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^D \times \mathbb{R}^D} |x-y|^2\,d\gamma(x,y) \right)^{1/2}. \tag{2.1}$$

Here, $\Gamma(\mu, \nu)$ denotes the set of Borel measures γ on $\mathbb{R}^D \times \mathbb{R}^D$ which have μ and ν as *marginals*, *i.e.* satisfying $\pi^1_\#(\gamma) = \mu$ and $\pi^2_\#(\gamma) = \nu$.

Eq. (2.1) defines a metric on $\mathcal{M}$ which is called the *Wasserstein distance*. It is known that the infimum in the right-hand side of Eq. (2.1) is always achieved. We will denote by $\Gamma_o(\mu, \nu)$ the set of γ which achieve the minimum in (2.1).

Definition 2.2. Let $\mu, \nu \in \mathcal{M}$ and $\gamma \in \Gamma_o(\mu, \nu)$. The *barycentric projection* $\bar{\gamma}^\nu_\mu : \mathbb{R}^D \to \mathbb{R}^D$ of γ with respect to the first marginal μ is characterized by

$$\int_{\mathbb{R}^D} \psi(x)\bar{\gamma}^\nu_\mu(x)\,d\mu(x) = \int_{\mathbb{R}^{2D}} \psi(x)\,y\,d\gamma(x,y), \quad \forall \psi \in C_b. \tag{2.2}$$

Similarly, the *barycentric projection* $\bar{\gamma}_\nu^\mu : \mathbb{R}^D \to \mathbb{R}^D$ of γ with respect to the second marginal ν is defined by

$$\int_{\mathbb{R}^D} \psi(y)\bar{\gamma}_\nu^\mu(y)\,d\nu(y) = \int_{\mathbb{R}^{2D}} \psi(y)x\,d\gamma(x,y), \quad \forall \psi \in C_b. \tag{2.3}$$

2.2. Differential structure on $\mathcal{M}$

Definition 2.3. Given $\mu \in \mathcal{M}$, let $T_\mu\mathcal{M}$ be the *tangent space* of $\mathcal{M}$ at μ defined as the closure of ∇C_c^∞ in $L^2(\mu)$, *i.e.*

$$T_\mu\mathcal{M} := \overline{\left\{\nabla\varphi\colon \varphi \in C_c^\infty\left(\mathbb{R}^D\right)\right\}}^{L^2(\mu)}.$$

For any $\mu \in \mathcal{M}$, there is an orthogonal decomposition

$$L^2(\mu) = T_\mu\mathcal{M} \oplus [T_\mu\mathcal{M}]^\perp, \tag{2.4}$$

where $[T_\mu\mathcal{M}]^\perp := \{w \in L^2(\mu)\colon\ \nabla\cdot(w\mu) = 0\}$. We will denote by $\pi_\mu : L^2(\mu) \to T_\mu\mathcal{M}$ the corresponding orthogonal projection.

As shown in [2], the tangent space enjoys many useful properties in analytic and geometric point of view. Here, we recall one of them which is related to absolutely continuous curves in $\mathcal{M}$. Let us first give the definition of absolutely continuous curves in metric spaces.

Definition 2.4. Let $(\mathbb{S}, \mathrm{dist})$ be a metric space. A curve $t \in (a,b) \mapsto \sigma_t \in \mathbb{S}$ is 2-*absolutely continuous* if there exists $\beta \in L^2(a,b)$ such that

$$\mathrm{dist}(\sigma_t, \sigma_s) \leqslant \int_s^t \beta(\tau)\,d\tau, \tag{2.5}$$

for all $a < s < t < b$. We then write $\sigma \in AC_2(a,b;\mathbb{S})$. For such curves the limit $|\sigma'|(t) := \lim_{s\to t} \mathrm{dist}(\sigma_t,\sigma_s)/|t-s|$ exists for $\mathcal{L}^1$-almost every $t \in (a,b)$. We call this limit the *metric derivative* of σ at t. It satisfies $|\sigma'| \leqslant \beta$ $\mathcal{L}^1$-almost everywhere.

We now recall Theorem 8.3.1 in [2], which says that the tangent space provides a canonical velocity field for the absolutely continuous curves in $\mathcal{M}$.

Proposition 2.5. *If $\mu \in AC_2(a,b;\mathcal{M})$ then there exists a Borel map $v : (a,b) \times \mathbb{R}^D \to \mathbb{R}^D$ such that $v_t \in L^2(\mu_t)$ for $\mathcal{L}^1$-almost every $t \in (a,b)$ and*

$$\frac{\partial\mu_t}{\partial t} + \nabla\cdot(v_t\mu_t) = 0.$$

We call v a velocity for μ. If w is another velocity for μ then $\pi_{\mu_t}(v_t) = \pi_{\mu_t}(w_t)$ for $\mathcal{L}^1$-almost every $t \in (a,b)$, where π_{μ_t} is defined in Definition 2.3. Moreover, one can choose v such that $v_t \in T_{\mu_t}\mathcal{M}$ and $\|v_t\|_{\mu_t} = |\mu'|(t)$ for $\mathcal{L}^1$-almost every $t \in (a,b)$. In that case, for $\mathcal{L}^1$-almost every $t \in (a,b)$, v_t is uniquely determined. We refer to v as the velocity of minimal norm, since if w is any other velocity associated to μ then $\|v_t\|_{\mu_t} \leqslant \|w_t\|_{\mu_t}$ for $\mathcal{L}^1$-almost every $t \in (a,b)$ and so $\mathrm{dist}(\mu_t,\mu_s) \leqslant \int_s^t \|v_\tau\|_{\mu_\tau}\,d\tau \leqslant \int_s^t \|w_\tau\|_{\mu_\tau}\,d\tau$ for all $a < s < t < b$.

Following [1], we give a notion of a differential and a definition of convex functions on $\mathcal{M}$.

Definition 2.6. Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper function on $\mathcal{M}$, *i.e.* the effective domain of H defined by $D(H) := \{\mu \in \mathcal{M}\colon\ H(\mu) < \infty\}$ is not empty. We say that $\xi \in L^2(\mu)$ belongs to the *subdifferential* $\partial_- H(\mu)$ if

$$H(\nu) \geqslant H(\mu) + \sup_{\gamma \in \Gamma_o(\mu,\nu)} \int_{\mathbb{R}^D \times \mathbb{R}^D} \langle \xi(x), y - x \rangle \, d\gamma(x, y) + o\big(W_2(\mu, \nu)\big),$$

as $\nu \to \mu$. We denote the domain of subdifferential by $D(\partial_- H) := \{\mu\colon\ \partial_- H(\mu) \neq \emptyset\}$.

If $-\xi \in \partial_-(-H)(\mu)$ then we say that ξ belongs to the *superdifferential* $\partial^+ H(\mu)$.

Remark 2.7. If $\partial_- H(\mu) \cap \partial^+ H(\mu) \neq \emptyset$ then we say that H is *differentiable* at μ. In this case, there is a unique vector in $\partial_- H(\mu) \cap \partial^+ H(\mu) \cap T_\mu \mathcal{M}$ and we define the *gradient vector* $\nabla_\mu H$ by the unique vector.

Definition 2.8. Let $H : \mathcal{M} \to (-\infty, \infty]$ be proper and let $\lambda \in \mathbb{R}$. We say that H is λ*-convex* if for every $\mu_0, \mu_1 \in \mathcal{M}$ and every optimal transport plan $\gamma \in \Gamma_o(\mu_0, \mu_1)$ we have

$$H(\mu_t) \leqslant (1-t)H(\mu_0) + tH(\mu_1) - \frac{\lambda}{2} t(1-t) W_2^2(\mu_0, \mu_1), \quad \forall t \in [0, 1], \tag{2.6}$$

where $\mu_t = ((1-t)\pi^1 + t\pi^2)_\# \gamma$. If $-H$ is $(-\lambda)$-convex then H is called λ*-concave*.

3. Moreau–Yosida approximation

In this section, we introduce the Moreau–Yosida approximation of functionals on $\mathcal{M}$.

Definition 3.1. Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper and coercive functional. For $\tau > 0$, the *Moreau–Yosida approximation* H_τ of H is defined as

$$H_\tau(\mu) = \inf_{\nu \in \mathcal{M}} \left\{ \frac{1}{2\tau} W_2^2(\mu, \nu) + H(\nu) \right\}. \tag{3.1}$$

Here, H is coercive means that there exist $\tau_* > 0$ and $\mu_* \in \mathcal{M}$ so that $H_{\tau_*}(\mu_*) > -\infty$. We also set

$$J_\tau[\mu] := \left\{ \mu_\tau\colon\ H_\tau(\mu) = \frac{1}{2\tau} W_2^2(\mu, \mu_\tau) + H(\mu_\tau) \right\}. \tag{3.2}$$

Lemma 3.2. *Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper and coercive functional, and H_τ be the Moreau–Yosida approximation of H. Then H_τ is $\frac{1}{\tau}$-concave.*

Proof. Let $\nu \in \mathcal{M}$ be fixed, then it is well known that $\mu \mapsto \frac{1}{2} W_2^2(\mu, \nu)$ is a 1-concave function on $\mathcal{M}$. This implies

$$\mu \to H_\tau(\mu) = \inf_{\nu \in \mathcal{M}} \left\{ \frac{1}{2\tau} W_2^2(\mu, \nu) + H(\nu) \right\},$$

is $\frac{1}{\tau}$-concave since it is an infimum of $\frac{1}{\tau}$-concave functionals. $\square$

We now introduce two lemmas which give relations between the subdifferential of H and the superdifferential of H_τ. They play the key role in the convergence of Hamiltonian systems.

Lemma 3.3. *Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper functional and H_τ be the Moreau–Yosida approximation of H. For $\mu_o \in \mathcal{M}$ given, if $\nu_o \in J_\tau[\mu_o]$ then H_τ is superdifferentiable at μ_o and H is subdifferentiable at ν_o, i.e. $\mu_o \in D(\partial^+ H_\tau)$ and $\nu_o \in D(\partial_- H)$. Furthermore, for any $\gamma \in \Gamma_o(\mu_o, \nu_o)$, we have*

$$\frac{Id - \bar{\gamma}_{\mu_o}^{\nu_o}}{\tau} \in \partial^+ H_\tau(\mu_o) \cap T_{\mu_o}\mathcal{M}, \qquad \frac{\bar{\gamma}_{\nu_o}^{\mu_o} - Id}{\tau} \in \partial_- H(\nu_o) \cap T_{\nu_o}\mathcal{M} \tag{3.3}$$

where $\bar{\gamma}_{\mu_o}^{\nu_o}$ ($\bar{\gamma}_{\nu_o}^{\mu_o}$) is the barycentric projection of γ with respect to the first (respectively, second) marginal as in Definition 2.2.

Proof. From the definition of $\nu_o \in J_\tau[\mu_o]$, we have

$$H_\tau(\mu) - H_\tau(\mu_o) \leqslant \frac{1}{2\tau} W_2^2(\mu, \nu_o) - \frac{1}{2\tau} W_2^2(\mu_o, \nu_o), \quad \forall \mu \in \mathcal{M}. \tag{3.4}$$

For a fixed μ, we choose $\eta \in \Gamma_o(\mu_o, \mu)$. Let $\eta = \int_{\mathbb{R}^D} \eta_x \, d\mu_o(x)$ and $\gamma = \int_{\mathbb{R}^D} \gamma_x \, d\mu_o(x)$ be the disintegration of η and γ w.r.t. μ_o. Define $\mathbf{u}_1 \in \mathcal{P}(\mathbb{R}^{3D})$ to be such that the disintegration of $\mathbf{u}_1$ w.r.t. μ_o is

$$\int_{\mathbb{R}^D} \eta_x \times \gamma_x \, d\mu_o(x). \tag{3.5}$$

We combine (3.4) and (3.5) to get

$$\begin{aligned}
H_\tau(\mu) - H_\tau(\mu_o) &\leqslant \frac{1}{2\tau} W_2^2(\mu, \nu_o) - \frac{1}{2\tau} W_2^2(\mu_o, \nu_o) \\
&\leqslant \frac{1}{\tau} \int_{\mathbb{R}^{3D}} \frac{|y-z|^2}{2} - \frac{|x-z|^2}{2} \, d\mathbf{u}_1(x, y, z) \\
&= \frac{1}{\tau} \int_{\mathbb{R}^{3D}} \langle x - z, y - x \rangle + \frac{|y-x|^2}{2} \, d\mathbf{u}_1(x, y, z) \\
&= \int_{\mathbb{R}^{2D}} \left\langle \frac{x - \bar{\gamma}_{\mu_o}^{\nu_o}(x)}{\tau}, y - x \right\rangle d\eta(x, y) + \frac{1}{2\tau} W_2^2(\mu_o, \mu)
\end{aligned} \tag{3.6}$$

which gives

$$\frac{Id - \bar{\gamma}_{\mu_o}^{\nu_o}}{\tau} \in \partial^+ H_\tau(\mu_o).$$

Furthermore, it is well known that $Id - \bar{\gamma}_{\mu_o}^{\nu_o} \in T_{\mu_o}\mathcal{M}$ (Proposition 4.3 of [1]). This concludes the first inclusion of (3.3).

To prove the second, we again exploit $\nu_o \in J_\tau[\mu_o]$ to get

$$\frac{1}{2\tau} W_2^2(\mu_o, \nu_o) + H(\nu_o) \leqslant \frac{1}{2\tau} W_2^2(\mu_o, \nu) + H(\nu), \quad \forall \nu \in \mathcal{M}.$$

For a fixed ν, let $\tilde{\eta} \in \Gamma_o(\nu_o, \nu)$ and define $\mathbf{u}_2 \in \mathcal{P}(\mathbb{R}^{3D})$ to be such that whose disintegration w.r.t. ν_o is

$$\int_{\mathbb{R}^D} \gamma_x \times \tilde{\eta}_x \, d\nu_o(x),$$

where, $\tilde{\eta} = \int_{\mathbb{R}^D} \tilde{\eta}_x \, d\nu_o(x)$ and $\gamma = \int_{\mathbb{R}^D} \gamma_x \, d\nu_o(x)$ are disintegrations of $\tilde{\eta}$ and γ w.r.t. ν_o. Computations as in (3.6) give

$$H_\tau(\nu) - H_\tau(\nu_o) \geqslant \int_{\mathbb{R}^{2D}} \left\langle \frac{\bar{\gamma}_{\nu_o}^{\mu_o}(x) - x}{\tau}, y - x \right\rangle d\tilde{\eta}(x, y) + \frac{1}{2\tau} W_2^2(\nu_o, \nu).$$

We conclude the second inclusion in (3.3) using the same argument as above. □

Lemma 3.4. *Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper functional and let H_τ be the Moreau–Yosida approximation of H. Given a sequence of measures μ_n and ν_n be such that $\nu_n \in J_{\tau_n}[\mu_n]$. Furthermore, suppose there is a constant C satisfying*

$$W_2(\mu_n, \nu_n) \leqslant C\tau_n, \tag{3.7}$$

for all n. If μ_n converges narrowly to μ as $\tau_n \to 0$, then ν_n also converges narrowly to μ. Furthermore, we have

$$\bigcap_{m=1}^{\infty} \overline{co}\left(\left\{ \frac{Id - \bar{\gamma}_{\mu_{\tau_n}}^{\nu_{\tau_n}}}{\tau_n} \mu_{\tau_n} \colon n \geqslant m \right\}\right) = \bigcap_{m=1}^{\infty} \overline{co}\left(\left\{ \frac{\bar{\gamma}_{\nu_{\tau_n}}^{\mu_{\tau_n}} - Id}{\tau_n} \nu_{\tau_n} \colon n \geqslant m \right\}\right). \tag{3.8}$$

Here $\overline{co}$ denotes the closed convex hull with respect to weak topology.*

Proof. Narrow convergence of ν_n to μ is trivial from the assumption (3.7) and the narrow convergence of μ_n to μ as $\tau_n \to 0$.

To prove (3.8), let us fix $\psi \in C_c^\infty(\mathbb{R}^D; \mathbb{R}^D)$. Then we have

$$\begin{aligned}
\int_{\mathbb{R}^D} \left\langle \psi(y), \frac{\bar{\gamma}_{\nu_{\tau_n}}^{\mu_{\tau_n}} - Id}{\tau_n}(y) \right\rangle d\nu_{\tau_n}(y) &= \int_{\mathbb{R}^{2D}} \left\langle \psi(y), \frac{x-y}{\tau_n} \right\rangle d\gamma_{\tau_n}(x, y) \\
&= \int_{\mathbb{R}^{2D}} \left\langle \psi(x) + \nabla\psi(\xi_{x,y}) \cdot (y - x), \frac{x-y}{\tau_n} \right\rangle d\gamma_{\tau_n}(x, y) \\
&= \int_{\mathbb{R}^D} \left\langle \psi(x), \frac{Id - \bar{\gamma}_{\mu_{\tau_n}}^{\nu_{\tau_n}}}{\tau_n}(x) \right\rangle d\mu_n(x) \\
&\quad + \int_{\mathbb{R}^{2D}} \left\langle \nabla\psi(\xi_{x,y}) \cdot (y - x), \frac{x-y}{\tau_n} \right\rangle d\gamma_{\tau_n}(x, y)
\end{aligned} \tag{3.9}$$

where $\gamma_{\tau_n} \in \Gamma_o(\mu_{\tau_n}, \nu_{\tau_n})$ and $\xi_{x,y}$ is a point on the line segment connecting x and y. Since $\psi \in C_c^\infty(\mathbb{R}^D; \mathbb{R}^D)$, we have

$$\left| \int_{\mathbb{R}^{2D}} \left\langle \nabla\psi(\xi_{x,y}) \cdot (y-x), \frac{x-y}{\tau_n} \right\rangle d\gamma_{\tau_n}(x,y) \right| \leqslant \frac{\|\nabla\psi\|_\infty}{\tau_n} W_2^2(\mu_{\tau_n}, \nu_{\tau_n})$$
$$\leqslant \|\nabla\psi\|_\infty C^2 \tau_n. \tag{3.10}$$

We combine Eqs. (3.9) and (3.10) to get

$$\lim_{n\to\infty} \int_{\mathbb{R}^D} \left\langle \psi(y), \frac{\bar{\gamma}_{\nu_{\tau_n}}^{\mu_{\tau_n}} - Id}{\tau_n}(y) \right\rangle d\nu_{\tau_n}(y) = \lim_{n\to\infty} \int_{\mathbb{R}^D} \left\langle \psi(x), \frac{Id - \bar{\gamma}_{\mu_{\tau_n}}^{\nu_{\tau_n}}}{\tau_n}(x) \right\rangle d\mu_n(x),$$

which concludes (3.8). $\square$

4. Convergence of Hamiltonian systems w.r.t. Moreau–Yosida approximation

Now we are ready to state our main result on the stability of Hamiltonian systems. More specifically, solutions of the approximated Hamiltonians converge to a solution of the original Hamiltonian system. Let us first be clear about the meaning of solution.

Definition 4.1. Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper and lower semi-continuous function. We say that a 2-absolutely continuous curve $\mu_t : [0, T] \to D(H)$ is a *solution of the Hamiltonian system* starting from $\bar{\mu} \in \mathcal{M}$, if there exists a vector field $v_t \in L^2(\mu_t)$ with $\|v_t\|_{\mu_t} \in L^1(0, T)$, such that

$$\begin{cases} \dfrac{d}{dt}\mu_t + \nabla \cdot (\mathbb{J} v_t \mu_t) = 0, \qquad \mu_0 = \bar{\mu}, \quad t \in (0, T), \\ v_t \in \partial_- H(\mu_t) \cap T_{\mu_t}\mathcal{M} \quad \text{a.e. } t \in (0, T). \end{cases} \tag{4.1}$$

Eq. (4.1) should be understood in the sense of a distribution: For any $\eta \in C_c^\infty(0, T)$ and $\zeta \in C_c^\infty(\mathbb{R}^D)$, we have

$$\int_0^T \int_{\mathbb{R}^D} \eta'(t)\zeta(x) + \eta(t)\langle \nabla\zeta(x) : \mathbb{J} v_t(x) \rangle \, d\mu_t(x)\, dt = 0.$$

To ensure the stability of Hamiltonian systems, we require two assumptions on the Hamiltonian.

(H1) There exist constants $C_o \in [0, \infty)$, $R_o \in (0, \infty]$ such that if $W_2(\mu, \bar{\mu}) < R_o$ then $\mu \in D(H)$ and, for each μ, there exists a unique $\nu \in J_\tau[\mu]$ satisfying:
- $\mu \mapsto \nu \in J_\tau[\mu]$ is continuous w.r.t. the topology induced by the Wasserstein distance and
$$\frac{W_2(\mu, \nu)}{\tau} \leqslant C_o. \tag{4.2}$$
- There exists a constant $k > 0$ such that
$$\text{if } \operatorname{supp}(\mu) \subset B_0(r) \text{ then } \operatorname{supp}(\nu) \subset B_0(kr), \tag{4.3}$$
for all $r > 0$ and μ. Recall, $B_0(r)$ is the ball around the origin with radius r in $\mathbb{R}^D$.

(H2) If $\mu_n \in D(\partial_- H)$ and μ_n converges narrowly to μ, then $\mu \in D(\partial_- H)$ and we have

$$\bigcap_{m=1}^{\infty} \overline{co}\big(\big\{w_n\mu_n\colon\ w_n \in \partial_- H(\mu_n) \cap T_{\mu_n}\mathcal{M},\ n \geqslant m\big\}\big) \subset \big\{w\mu\colon\ w \in \partial_- H(\mu) \cap T_\mu \mathcal{M}\big\}. \tag{4.4}$$

Remark 4.2.

1. Notice that our Hamiltonian H does not need to be subdifferentiable everywhere in a neighborhood of $\bar{\mu}$. We only assume that $D(\partial_- H)$ is closed in the weak* topology and (4.4) holds.
2. Concerning (H1), suppose H satisfies the following convexity condition for some $\lambda \in \mathbb{R}$:
   For all μ, ν_0 and ν_1 in $D(H)$, there exists a curve $\sigma : [0, 1] \to \mathcal{M}$ such that $\sigma_0 = \nu_0$, $\sigma_1 = \nu_1$ and

$$\mathcal{H}(\tau, \mu; \sigma_t) \leqslant (1-t)\mathcal{H}(\tau, \mu; \nu_0) + t\mathcal{H}(\tau, \mu; \nu_1) - \frac{1+\lambda\tau}{2\tau} t(1-t) W_2^2(\nu_o, \nu_1), \tag{4.5}$$

   for all $t \in [0, 1]$ and $0 < \tau < \frac{1}{\lambda^-}$. Here, $\mathcal{H}(\tau, \mu; \nu) := \frac{1}{2\tau} W_2^2(\nu, \mu) + H(\nu)$.
   Then, Theorem 4.1.2 in [2] says that if $\mu \in D(H)$ and $\lambda\tau > -1$ then there exists a unique $\mu_\tau \in J_\tau[\mu]$ and the map $\mu \in D(H) \mapsto \mu_\tau \in J_\tau[\mu]$ is continuous.

4.1. Existence of solutions of the regularized Hamiltonian systems

Lemma 4.3. *Let $H : \mathcal{M} \to (-\infty, \infty]$ be proper and lower semi-continuous, and satisfy* (H1)*. Let μ_n be a sequence satisfying $W_2(\mu_n, \bar{\mu}) < R_o$ and $\nu_n \in J_\tau[\mu_n]$. If μ_n converges to μ in the Wasserstein topology, then ν_n also converges to $\nu \in J_\tau[\mu]$ in the same topology. Furthermore, we have*

$$\mathcal{K}_o := \bigcap_{m=1}^{\infty} \overline{co}\left(\left\{\frac{Id - \bar{\gamma}_{\mu_n}^{\nu_n}}{\tau}\mu_n\colon\ n \geqslant m\right\}\right) \subset \left\{\frac{Id - \bar{\gamma}_{\mu}^{\nu}}{\tau}\mu\right\}, \tag{4.6}$$

where $\gamma_n \in \Gamma_o(\mu_n, \nu_n)$ and $\gamma \in \Gamma_o(\mu, \nu)$.

Proof. By (H1), there exists a $\nu \in J_\tau[\mu]$ such that $W_2(\nu_n, \nu) \to 0$.

Next, to prove (4.6), let us assume $\mathbf{u} \in \mathcal{K}_o$. Then, there exists a sequence $\{\Lambda_m\}_{m=1}^{\infty}$ such that

$$\Lambda_m = \sum_{i=m}^{l_m} \lambda_i^m \frac{Id - \bar{\gamma}_{\mu_i}^{\nu_i}}{\tau}\mu_i, \quad \sum_{i=m}^{l_m} \lambda_i^m = 1,\ 0 \leqslant \lambda_i^m \leqslant 1,\ m \leqslant l_m \in \mathbb{N}$$

and Λ_m weak* converges to $\mathbf{u}$. For any $F \in C_c(\mathbb{R}^D; \mathbb{R}^D)$, we have

$$\begin{aligned}\int_{\mathbb{R}^D} F \cdot d\mathbf{u} &= \lim_{m\to\infty} \sum_{i=m}^{l_m} \lambda_i^m \int_{\mathbb{R}^D} \left\langle F(x), \frac{x - \bar{\gamma}_{\mu_i}^{\nu_i}(x)}{\tau}\right\rangle d\mu_i(x) \\ &= \lim_{m\to\infty} \sum_{i=m}^{l_m} \lambda_i^m \int_{\mathbb{R}^D} \left\langle F(x), \frac{x-y}{\tau}\right\rangle d\gamma_i(x, y),\end{aligned} \tag{4.7}$$

where $\gamma_i \in \Gamma_o(\mu_i, \nu_i)$. Since $W_2(\mu_n, \nu), W_2(\nu_n, \nu) \to 0$ as $n \to \infty$, there exists $\gamma \in \Gamma_0(\mu, \nu)$ so that

$$\lim_{i\to\infty} W_2(\gamma_i, \gamma) = 0. \tag{4.8}$$

We combine (4.7) and (4.8), to get

$$\int_{\mathbb{R}^D} F \cdot d\mathbf{u} = \int_{\mathbb{R}^{2D}} \left\langle F(x), \frac{x-y}{\tau} \right\rangle d\gamma(x, y), \tag{4.9}$$

which proves (4.6). □

Now we generate a solution of the Hamiltonian system for H_τ. The proof of the following theorem is based on Theorem 7.4 in [1].

Theorem 4.4. *Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper and lower semi-continuous functional satisfying the assumption* (H1)*. Let C_o and R_o be constants in* (H1)*, and set $T = \frac{R_o}{C_o}$.*

If $\bar{\mu} \in \mathcal{M}$ has bounded support, then for each $\tau > 0$, there exists a solution of the following Hamiltonian system starting from $\bar{\mu}$

$$\begin{cases} \dfrac{d}{dt}\mu_t^\tau + \nabla \cdot \left(\mathbb{J} v_t^\tau \mu_t^\tau\right) = 0, \qquad \mu_0^\tau = \bar{\mu}, \quad t \in (0, T), \\[2ex] v_t^\tau = \dfrac{Id - \bar{\gamma}_{\mu_t^\tau}^{\nu_t^\tau}}{\tau} \in \partial^+ H_\tau\left(\mu_t^\tau\right) \cap T_{\mu_t^\tau} M, \quad a.e.\ t \in (0, T), \end{cases} \tag{4.10}$$

where $\nu_t^\tau \in J_\tau[\mu_t^\tau]$. Furthermore, $t \mapsto \mu_t^\tau$ is Lipschitz continuous w.r.t. the Wasserstein distance with Lipschitz constant C_o which is independent of τ.

Proof. *Step* 1. *Construction of a discrete solution.*

For given $\tau > 0$, we fix an integer N and divide $[0, T]$ in N equal subintervals of length $h = T/N$. We build discrete solutions $\mu_{t,\tau}^N$ satisfying:

(a) The Lipschitz constant of $t \mapsto \mu_{t,\tau}^N \in \mathcal{M}$ is less than C_o.
(b) For all $t \in [0, T]$, we have $\mathrm{supp}(\mu_{t,\tau}^N) \subset B_0(e^{\frac{(1+k)T}{\tau}} r)$ if $\mathrm{supp}(\bar{\mu}) \subset B_0(r)$. Here, $k > 0$ is same as in (H1).
(c) $\mu_{t,\tau}^N$ satisfies the discrete Hamiltonian equation

$$\frac{d}{dt}\mu_{t,\tau}^N + \nabla \cdot \left(w_{t,\tau}^N \mu_{t,\tau}^N\right) = 0, \quad t \in (0, T), \tag{4.11}$$

with

$$w_{t,\tau}^N = \mathbb{J}\frac{Id - \bar{\gamma}_{\mu_{t,\tau}^N}^{\nu_{t,\tau}^N}}{\tau} \quad \text{for } t = 0, h, 2h, \ldots, Nh, \tag{4.12}$$

where $\nu_{t,\tau}^N \in J_\tau[\mu_{t,\tau}^N]$ and $\gamma \in \Gamma_o(\mu_{t,\tau}^N, \nu_{t,\tau}^N)$.
Since N and τ are fixed, we use the notation $\mu_t := \mu_{t,\tau}^N$ for convenience.

(i) We build the solution in $[0,h]$.

Let us call $\mu_0 := \bar{\mu}$ and choose $\nu_0 \in J_\tau[\mu_0]$ by (H1). We fix $\gamma \in \Gamma_o(\mu_0,\nu_0)$ and set

$$w_0 := \mathbb{J}\frac{Id - \bar{\gamma}_{\mu_0}^{\nu_0}}{\tau}. \tag{4.13}$$

We define

$$\mu_t := (Id + tw_0)_\# \mu_0, \qquad w_t := \frac{(Id + tw_0)_\#(w_0\mu_0)}{\mu_t}, \quad t \in [0,h].$$

We claim that w_t is a velocity field for μ_t, that is,

$$\frac{d}{dt}\mu_t + \nabla\cdot(w_t\mu_t) = 0, \tag{4.14}$$

holds in the distribution sense in $(0,h)$. Indeed, for any $\phi \in C_c^\infty(\mathbb{R}^D)$, we have

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^D}\phi\,d\mu_t &= \frac{d}{dt}\int_{\mathbb{R}^D}\phi(Id + tw_0)\,d\mu_0 \\ &= \int_{\mathbb{R}^D}\big\langle\nabla\phi\big(x + tw_0(x)\big), w_0(x)\big\rangle\,d\mu_0(x) = \int_{\mathbb{R}^D}\langle\nabla\phi, w_t\rangle\,d\mu_t.\end{aligned} \tag{4.15}$$

Notice that Lemma 7.1 in [1] gives

$$\int_{\mathbb{R}^D}|w_t|^2\,d\mu_t \leqslant \int_{\mathbb{R}^D}|w_0|^2\,d\mu_0, \tag{4.16}$$

for all $t \in [0,h]$. Jensen's inequality with (H1) gives

$$\int_{\mathbb{R}^D}|w_0|^2\,d\mu_0 = \int_{\mathbb{R}^D}\left|\frac{Id - \bar{\gamma}_{\mu_0}^{\nu_0}}{\tau}\right|^2 d\mu_0 \leqslant \frac{1}{\tau^2}W_2^2(\mu_0,\nu_0) \leqslant C_o^2. \tag{4.17}$$

We exploit Proposition 2.5 with (4.16) and (4.17), to conclude that $t \mapsto \mu_t$ is Lipschitz continuous with a Lipschitz constant C_o.

Next we show the bound on the support. Since $\operatorname{supp}(\mu_0) \subset B_0(r)$, we have $\operatorname{supp}(\nu_0) \subset B_0(kr)$ by (H1). Hence, if $z \in \operatorname{supp}(\mu_t)$ then

$$|z| \leqslant \sup_{x\in\operatorname{supp}(\mu_0)}\left|x + t\mathbb{J}\frac{x - \bar{\gamma}_{\mu_0}^{\nu_0}(x)}{\tau}\right| \leqslant \left(r + h\frac{r + kr}{\tau}\right),$$

hence we have $\operatorname{supp}(\mu_t) \subset B_0((1 + \frac{1+k}{\tau}h)r)$.

(ii) We continue this process in $[h,2h]$.

Since we have $W_2(\mu_0,\mu_h) \leqslant hC_o \leqslant R_o$, we can choose $\nu_h \in J_\tau[\mu_h]$ and set

$$w_h := \mathbb{J}\frac{Id - \bar{\gamma}_{\mu_h}^{\nu_h}}{\tau}. \tag{4.18}$$

We introduce the following extension for $t \in (h, 2h]$,

$$\mu_t = \big(Id + (t-h)w_h\big)_{\#}(\mu_h), \qquad w_t = \frac{(Id + (t-h)w_h)_{\#}(w_h\mu_h)}{\mu_t}.$$

As in (i), we can check $t \mapsto \mu_t$ is Lipschitz continuous with a Lipschitz constant C_o in $[h, 2h]$. Furthermore, Eq. (4.15) holds and we have $\operatorname{supp}(\mu_t) \subset B_0((1 + \frac{1+k}{\tau}h)^2 r)$ for all $t \in [h, 2h]$.

(iii) We iterate the above process until we get a Lipschitz curve $t \mapsto \mu_t \in \mathcal{M}$ with Lipschitz constant C_o. The curve satisfies (4.15) for a.e. $t \in (0, T)$ and hence

$$\frac{d}{dt}\mu_t + \nabla \cdot (w_t\mu_t) = 0,$$

holds in the distribution sense. Furthermore, for all $t \in [0, T]$

$$\operatorname{supp}(\mu_t) \subset B_0\left(\left(1 + \frac{1+k}{\tau}\frac{T}{N}\right)^N r\right) \subset B_0\big(e^{\frac{(1+k)T}{\tau}} r\big).$$

Recalling that $\mu^N_{t,\tau} := \mu_t$ completes the proof.

Step 2. *Let N increase to* ∞.

From (a), $t \mapsto \mu^N_{t,\tau}$ is equi-bounded in $\mathcal{M}$ and equi-Lipschitz with Lipschitz constant C_o. Since $\mu^N_{t,\tau}$ has uniformly bounded supports, we may assume (up to a subsequence) that $\mu^N_{t,\tau}$ converges in the Wasserstein topology as $N \to \infty$. That is, there exists μ^τ_t such that $W_2(\mu^N_{t,\tau}, \mu^\tau_t) \to 0$ for any $t \in [0, T]$. Moreover, $t \mapsto \mu^\tau_t$ is Lipschitz continuous with Lipschitz constant C_o. As shown in [1], μ^τ_t solves $\frac{d}{dt}\mu^\tau_t + \nabla \cdot (w^\tau_t \mu^\tau_t) = 0$ with the following property

$$w^\tau_t \mu^\tau_t \in \bigcap_{M=1}^{\infty} \overline{co}\big\{w^N_{t,\tau}\mu^N_{t,\tau} \colon N \geqslant M\big\} \quad \text{a.e. } t \in (0, T).$$

Since

$$\begin{aligned} w^N_{t,\tau}\mu^N_{t,\tau} &= \big(Id + \big(t - [Nt]/N\big)w^N_{[Nt]/N,\tau}\big)_{\#}\big(w^N_{[Nt]/N,\tau}\mu^N_{[Nt]/N,\tau}\big) \\ &= \left(Id + \big(t - [Nt]/N\big)\mathbb{J}\frac{Id - \bar{\gamma}_{\mu^N_{[Nt]/N,\tau}}^{\nu^N_{[Nt]/N,\tau}}}{\tau}\right)_{\#}\left(\mathbb{J}\frac{Id - \bar{\gamma}_{\mu^N_{[Nt]/N,\tau}}^{\nu^N_{[Nt]/N,\tau}}}{\tau}\mu^N_{[Nt]N,\tau}\right). \end{aligned}$$

We also obtain

$$w^\tau_t \mu^\tau_t \in \bigcap_{M=1}^{\infty} \overline{co}\left\{\mathbb{J}\frac{Id - \bar{\gamma}_{\mu^N_{[Nt]/N,\tau}}^{\nu^N_{[Nt]/N,\tau}}}{\tau}\mu^N_{[Nt]N,\tau} \colon N \geqslant M\right\}. \tag{4.19}$$

Lemma 4.3 together with (4.19) gives

$$w^\tau_t \mu^\tau_t = \mathbb{J}v^\tau_t \mu^\tau_t, \quad v^\tau_t = \frac{Id - \bar{\gamma}_{\mu^\tau_t}^{\nu^\tau_t}}{\tau},$$

where $\nu^\tau_t \in J_\tau[\mu^\tau_t]$ for a.e. $t \in (0, T)$. This with Lemma 3.3 concludes the proof. $\square$

4.2. Stability of Hamiltonian flows

Theorem 4.5. *Let $H : \mathcal{M} \to (-\infty, \infty]$ be a proper and lower semi-continuous functional satisfying* (H1) *and* (H2)*. We assume that $\bar{\mu} \in \mathcal{M}$ has a bounded support. For each $\tau \in (0, 1)$, let μ^τ be the solution of the system* (4.10) *in Theorem* 4.4*. Then, $\{\mu^\tau\}_{\tau>0}$* (*up to a sequence*) *converges to a solution of the Hamiltonian system*

$$\begin{cases} \dfrac{d}{dt}\mu_t + \nabla \cdot (\mathbb{J} v_t \mu_t) = 0, \qquad \mu_0 = \bar{\mu}, \quad t \in (0, T), \\ v_t \in \partial_- H(\mu_t) \cap T_{\mu_t}\mathcal{M} \quad \textit{a.e. } t \in (0, T), \end{cases} \tag{4.20}$$

as τ converges to 0.

Proof. Since $t \mapsto \mu_t^\tau$ is equi-bounded in $\mathcal{M}$ and equi-Lipschitz continuous, we may assume that, for any $t \in [0, T]$, $\mu_t^{\tau_n}$ converges narrowly to μ_t as $\tau_n \to 0$, for some subsequence τ_n.

By the same reasoning as Step 2 in the proof of Theorem 4.4, μ_t solves

$$\begin{cases} \dfrac{d}{dt}\mu_t + \nabla \cdot (\mathbb{J} v_t \mu_t) = 0, \quad t \in (0, T), \\ \mu_0 = \bar{\mu}, \end{cases} \tag{4.21}$$

with

$$v_t \mu_t \in \bigcap_{M=1}^{\infty} \overline{co}\big\{v_t^{\tau_n} \mu_t^{\tau_n} \colon n \geqslant M\big\}, \tag{4.22}$$

for a.e. $t \in (0, T)$. Here

$$v_t^{\tau_n} = \frac{Id - \bar{\gamma}_{\mu_t^{\tau_n}}^{\nu_t^{\tau_n}}}{\tau_n} \in \partial^+ H\big(\mu_t^{\tau_n}\big) \cap T_{\mu_t^{\tau_n}}\mathcal{M}, \quad \nu_t^{\tau_n} \in J_{\tau_n}\big[\mu_t^{\tau_n}\big]. \tag{4.23}$$

From Lemma 3.4 together with (4.22) and (4.23), we know $\nu_t^{\tau_n} \to \mu_t$ narrowly as $\tau_n \to 0$ and

$$v_t \mu_t \in \bigcap_{M=1}^{\infty} \overline{co}\big\{\xi_t^{\tau_n} \nu_t^{\tau_n} \colon n \geqslant M\big\}, \tag{4.24}$$

where $\nu_t^{\tau_n} \in J_{\tau_n}[\mu_t^{\tau_n}]$ and $\xi_t^{\tau_n} = \frac{\bar{\gamma}_{\nu_t^{\tau_n}}^{\mu_t^{\tau_n}} - Id}{\tau_n} \in \partial_- H(\nu_t^{\tau_n}) \cap T_{\nu_t^{\tau_n}}\mathcal{M}$.

By (H2) and (4.24), we get

$$v_t \in \partial_- H(\mu_t) \cap T_{\mu_t}\mathcal{M}, \quad \text{a.e. } t \in (0, T),$$

which concludes the proof. □

4.3. Example

Let $\mu_o \in \mathcal{M}$ have a bounded support and $a > 0$, we define

$$H(\mu) := -\frac{a}{2}W_2^2(\mu,\mu_o) + \int V(x)\,d\mu(x) + \iint W(x,y)\,d\mu(x)\,d\mu(y), \tag{4.25}$$

where $V : \mathbb{R}^D \to \mathbb{R}$ is λ_V-convex for some $\lambda_V \in \mathbb{R}$, and $W : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is convex and even. Assume also that both are differentiable and have at most quadratic growth at infinity. Then, the function $H : \mathcal{M} \to (-\infty,\infty]$ as in (4.25) satisfies (H1) and (H2).

Proof. First, we notice that H defined as (4.25) is $(\lambda_V - a)$-convex and locally Lipschitz [1], *i.e.* there exist $R_{\bar\mu}, C_{\bar\mu} > 0$ such that

$$H(\mu_1) - H(\mu_2) \leqslant C_{\bar\mu} W_2(\mu_1,\mu_2), \tag{4.26}$$

for all μ_i with $W_2(\bar\mu,\mu_i) < R_{\bar\mu}$, $i = 1,2$.

Secondly, H satisfies the convexity condition (4.5) with $\lambda = \lambda_V - a$ (Chapter 9 in [2]). From Remark 4.2, it follows that for all sufficiently small $\tau > 0$ and $\forall\mu \in \mathcal{M}$, there exists a unique $\mu_\tau \in J_\tau[\mu]$. Furthermore, for fixed τ, $\mu \mapsto \mu_\tau \in J_\tau[\mu]$ is continuous.

We now show there is $0 < R_o < R_{\bar\mu}$ such that for all sufficiently small $\tau > 0$,

$$\text{if } W_2(\mu,\bar\mu) < R_o \text{ then } W_2(\bar\mu,\mu_\tau) < R_{\bar\mu}. \tag{4.27}$$

Once we have (4.27), (4.26) with $\mu_\tau \in J_\tau[\mu]$ gives

$$\frac{1}{2\tau}W_2^2(\mu,\mu_\tau) \leqslant H(\mu) - H(\mu_\tau) \leqslant C_\mu W_2(\mu,\mu_\tau),$$

for all μ with $W_2(\mu,\bar\mu) < R_o$. That is,

$$\frac{1}{\tau}W_2(\mu,\mu_\tau) \leqslant 2C_{\bar\mu}, \tag{4.28}$$

which implies that (4.2) in (H1) holds with $C_o = 2C_{\bar\mu}$.

Now, let us prove (4.27). We define $R_o := R_{\bar\mu}/2$. If $W_2(\mu,\bar\mu) < R_o$ then we have

$$\begin{aligned}\frac{1}{2\tau}W_2^2(\mu,\mu_\tau) + H(\mu_\tau) &\leqslant H(\mu)\\ &\leqslant H(\bar\mu) + C_{\bar\mu}W_2(\bar\mu,\mu)\\ &\leqslant H(\bar\mu) + C_{\bar\mu}R_o := C.\end{aligned} \tag{4.29}$$

We need to estimate $H(\mu_\tau)$ in (4.29). Since V and W have at most quadratic growth at infinity, there are constants c_1 and c_2 such that

$$\left|V(x)\right| \leqslant c_1|x|^2 + c_2 \quad\text{and}\quad \left|W(x,y)\right| \leqslant c_1\left(|x|^2+|y|^2\right) + c_2, \tag{4.30}$$

for all $x, y \in \mathbb{R}^D$. This implies

$$
\begin{aligned}
H(\mu_\tau) &= -\frac{a}{2}W_2^2(\mu_\tau,\mu_o) + \int V(x)\,d\mu_\tau(x) + \iint W(x,y)\,d\mu_\tau(x)\,d\mu_\tau(y) \\
&\geqslant -\frac{a}{2}W_2^2(\mu_\tau,\mu_o) - 3c_1\int |x|^2\,d\mu_\tau(x) - 2c_2 \\
&= -\frac{a}{2}W_2^2(\mu_\tau,\mu_o) - 3c_1 W_2^2(\mu_\tau,\delta_0) - 2c_2.
\end{aligned}
\tag{4.31}
$$

We combine (4.29) and (4.31), and get

$$
\frac{1}{2\tau}W_2^2(\mu,\mu_\tau) \leqslant \frac{a}{2}W_2^2(\mu_\tau,\mu_o) + 3c_1 W_2^2(\mu_\tau,\delta_0) + 2c_2 + C. \tag{4.32}
$$

Now let $\tau \to 0$ in (4.32). Since μ_o, δ_0 are fixed and $W_2(\bar{\mu},\mu) \leqslant R_o = R_{\bar{\mu}}/2$,

$$
W_2(\mu,\mu_\tau) \to 0 \quad \text{uniformly w.r.t. } \mu \text{ as } \tau \to 0, \tag{4.33}
$$

which implies (4.27).

To finish proving (H1), it remains to prove (4.3). Let $\tilde{H} : \mathcal{M} \to (-\infty,\infty]$ be defined by

$$
\tilde{H}(\mu) = -\frac{a}{2}W_2(\mu,\mu_o).
$$

For the Hamiltonian $\tilde{H}$, it was shown that (4.3) holds for some $k > 0$ (*refer* [10]). It is easy to see (4.3) holds with same $k > 0$ for the Hamiltonian H as in (4.25).

Hence, for the Hamiltonian H defined by (4.25), the assumption (H1) holds with $C_o = 2C_{\bar{\mu}}$ and $R_o = R_{\bar{\mu}}/2$. It was shown in [1] that the assumption (H2) also holds. $\square$

Comments. Suppose we want to solve the finite dimensional Hamiltonian system which consists of a single particle

$$
\begin{cases} x''(t) = -\nabla V\big(x(t)\big), \\ x'(0) = \bar{v}, \qquad x(0) = \bar{x}, \end{cases} \tag{4.34}
$$

where $V : \mathbb{R}^D \to \mathbb{R}$ is given and $\bar{v}, \bar{x} \in \mathbb{R}^D$. If V is not everywhere differentiable then we may try a regularization scheme as follows. We first solve the approximate system

$$
\begin{cases} x_\epsilon''(t) = -\nabla V_\epsilon\big(x_\epsilon(t)\big), \\ x_\epsilon'(0) = \bar{v}, \qquad x_\epsilon(0) = \bar{x}, \end{cases} \tag{4.35}
$$

where V_ϵ is any regular approximation of V. For example, we can define $V_\epsilon := \rho_\epsilon * V$ as the standard mollification of V. Next we check if the solution $x_\epsilon(t)$ of (4.35) converges to a solution $x(t)$ of (4.34) as ϵ goes to zero. Of course, we need certain properties on V to ensure this stability property hold. For example, if V is convex then the limiting solution $x(t)$ satisfies the differential inclusion $x''(t) \in -\partial_- V(x(t))$ instead of the first equation in (4.34).

Let us address the Hamiltonian system (1.1) in the Wasserstein space. As we saw in the previous sections, under certain conditions on the Hamiltonian H, the Hamiltonian system is stable with respect to the Moreau–Yosida approximation. Therefore, we may apply the Moreau–Yosida approximation scheme to study non-locally subdifferentiable Hamiltonians.

Let $\bar{\mu} \in \mathcal{M}$ and $\mathcal{B}$ be a neighborhood of $\bar{\mu}$ in the Wasserstein space. Suppose our Hamiltonian H is subdifferentiable only in a proper subset $\mathcal{D} \subset \mathcal{B}$, and $\bar{\mu} \in \mathcal{D}$. We want to solve the system (1.1) with the initial measure $\bar{\mu}$. To do this, we need an algorithm to construct solutions which stay in the subset $\mathcal{D}$.

In Theorem 4.4, we construct *approximate* solutions ν_τ for H as well as solutions μ_τ for H_τ. Notice that we have $\nu_\tau \in \mathcal{D}$ and we need only the assumption (H1) on H in Theorem 4.4. Note, the assumption (H1) has nothing to do with the subdifferentiability of H, which means that the construction of approximate solutions ν_τ relies entirely on the Moreau–Yosida approximation method. Next, in Theorem 4.5, we add the assumption (H2) on H which then implies the convergence of ν_τ to a solution of the system (1.1). The assumption (H2) does not require our Hamiltonian H to be subdifferentiable everywhere in the neighborhood $\mathcal{B}$ of $\bar{\mu}$. Instead, it requires $\mathcal{D}$ to be closed in the weak* topology and (4.4) hold.

Hence, as a direct result of our stability result, the Moreau–Yosida approximation scheme provides an algorithm to construct a solution of the system (1.1) for Hamiltonians which are subdifferentiable only in a proper subset $\mathcal{D}$ of a neighborhood of $\bar{\mu}$.

Acknowledgments

I would like to acknowledge my gratitude to Jacob Bedrossian and Ryan Hynd for valuable suggestions.

References

[1] L. Ambrosio, W. Gangbo, Hamiltonian ODEs in the Wasserstein space of probability measures, Comm. Pure Appl. Math. 61 (1) (2008) 18–53.
[2] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, second ed., Lectures Math. ETH Zürich, Birkhäuser Verlag, Basel, 2008.
[3] J.-D. Benamou, Y. Brenier, Weak existence for the semigeostrophic equations formulated as a coupled Monge–Ampère/transport problem, SIAM J. Appl. Math. 58 (5) (1998) 1450–1461 (electronic).
[4] Y. Brenier, Convergence of the Vlasov–Poisson system to the incompressible Euler equations, Comm. Partial Differential Equations 25 (3–4) (2000) 737–754.
[5] Y. Brenier, G. Loeper, A geometric approximation to the Euler equations: the Vlasov–Monge–Ampère system, Geom. Funct. Anal. 14 (6) (2004) 1182–1218.
[6] M. Cullen, M. Feldman, Lagrangian solutions of semigeostrophic equations in physical space, SIAM J. Math. Anal. 37 (5) (2006) 1371–1395 (electronic).
[7] M. Cullen, W. Gangbo, A variational approach for the 2-dimensional semi-geostrophic shallow water equations, Arch. Ration. Mech. Anal. 156 (3) (2001) 241–273.
[8] M. Cullen, W. Gangbo, G. Pisante, The semigeostrophic equations discretized in reference and dual variables, Arch. Ration. Mech. Anal. 185 (2) (2007) 341–363.
[9] W. Gangbo, H.K. Kim, T. Pacini, Differential forms on Wasserstein space and infinite dimensional Hamiltonian systems, Mem. Amer. Math. Soc. 211 (993) (2011).
[10] H.K. Kim, Hamiltonian systems and the calculus of differential forms on Wasserstein space, PhD thesis, Georgia Institute of Technology, 2009.
[11] C. Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, Amer. Math. Soc., Providence, RI, 2003.
[12] Y.X. Zheng, A. Majda, Existence of global weak solutions to one-component Vlasov–Poisson and Fokker–Planck–Poisson systems in one space dimension with measures as initial data, Comm. Pure Appl. Math. 47 (10) (1994) 1365–1401.