A NOTE ON HAMILTONIAN ODES IN THE WASSERSTEIN SPACE OF PROBABILITY MEASURES

Liu Yang
Division of Applied Mathematics
Brown University
Providence, RI 02912, USA
liu_yang@brown.edu

ABSTRACT

This is a note on Hamiltonian ODEs in the Wasserstein space of probability measures, based on Ambrosio & Gangbo (2008). It's more focused on the introduction of the preliminaries, including the tangent space of Wasserstein space and convex analysis. There might be some mistakes, due to author's limited knowledge.

1 INTRODUCTION

The motivation of this reading this paper is to learn the dynamics in W_2 space, and its connection to stochastic differential equations (SDEs), trying to apply to generative models, in specific, introduce dynamics to flow-based models. Later I found it's only related to ODEs, i.e. deterministic flow models, instead of SODEs. But the perspective and some mathematical facts are pretty interesting, and might be helpful in the future research.

The paper starts from the differential structure of $P_2(R^D)$, then goes to convex analysis on $P_2(R^D)$, and finally builds the theory of Hamiltonian ODE's in the infinite-dimensional space $P_2(R^D)$.

I am going to present the definitions and important results, as well as my intuitive understandings, but skip the proofs.

2 DIFFERENTIAL STRUCTURE OF THE WASSERSTEIN SPACE $P_2(R^D)$

$P_2(R^D)$ space is the metric space of probability measures on R^D endowed with Wasserstein-2 (W_2) distance as metric. Lets first review W_2 distance.

2.1 WASSERSTEIN-2 DISTANCE

Consider the Wasserstein space $P_2(R^D)$ of the probability measures with finite quadratic moments in R^D, endowed with the Wasserstein distance W_2, defined as

$$W_2^2(\mu,\nu) = \min_{\gamma}\{\int_{R^D\times R^D} |x-y|^2 d\gamma(x,y) : \gamma \in \Gamma(\mu,\nu)\} \tag{1}$$

where $\Gamma(\mu,\nu)$ is the set of Borel probability measures on $R^D \times R^D$ which have μ and ν as their marginals.

Actually we could have the following topology induced by W_2 distance:

Theorem 1 *A sequence $\{\mu_n\}_{n=1}^{\infty}$ converges to μ in $P_2(R^D)$ if and only if μ_n narrowly converges to μ, i.e. weak convergence in the duality with $C_b(R^D)$ (continuous and bounded functions), and $M_2(\mu_n)$ converges to $M_2(\mu)$ as n goes to infinity.*

If R^D is replaced with a compact set in R^D, then the condition of the second moment convergence could be removed.

For more details, see `https://www.math.u-psud.fr/~filippo/Wp.pdf`.

2.2 Tangent space to $P_2(R^D)$

Theorem 2 *For any absolute continuity curve $\mu_t : [a, b] \to P_2(R^D)$ there exist $v_t \in L^2(\mu_t; R^D)$ for which the following two equations holds:*

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \tag{2}$$

$$\lim_{h \to 0} \frac{1}{|h|} W_2(\mu_{t+h}, \mu_t) = \|v_t\|_{L^2(\mu_t)} \quad \textit{for a.e. } t \tag{3}$$

Given the continuity equation 2 holds, the asymptotic equality 3 holds if and only if

$$v_t \in T_{\mu_t} P_2(R^D) = \overline{\{\nabla \phi : \phi \in C_c^\infty(R^D)\}}^{L^2(\mu_t; R^D)} \quad \textit{for a.e. } t \tag{4}$$

Finally, the map $t \to v_t \in L^2(\mu_t, R^D)$ is uniquely determined up to L^1-negligible sets.

The continuity equation 2 is in the sense of distributions in $[a, b] \times R^D$, i.e.

$$\begin{gathered}\int_a^b \int_{R^D} (\partial_t \phi(x,t) + \langle v_t(x), \nabla_x \phi(x,t) \rangle) d\mu_t(x) dt = 0 \\ \forall \phi \in C_c^\infty(R^D \times (a, b)).\end{gathered} \tag{5}$$

The above theorem tells that $T_\mu P_2(R^D)$ defines the tangent space to $P_2(R^D)$ at μ.

Essentially, the optimal plans between μ_{t+h} and μ_t asymptotically behave as the plans induced by the transport maps $(id + hv_t)$. Actually, when $\mu_t \in P_2^a(R^D)$ (a.c. w.r.t $\mathcal{L}^D$), where the optimal plans are unique and induced by maps, we have

$$\frac{t_h - id}{h} \to v_t \text{ in } L^2(\mu_t; R^D) \text{ as } h \to 0 \tag{6}$$

Also, a duality argument gives

$$[T_\mu P_2(R^D)]^\perp = \{v \in L^2(\mu, R^D) : \nabla \cdot (v\mu) = 0\} \tag{7}$$

The existence of the "potential flow" velocity field is important to me, since I was exploring flow-based generators that would drive the distribution along a prescribed trajectory (or "curve"). This theorem guarantees the existence of the generator as long as the trajectory is absolutely continuous in W_2 sense.

2.3 Frechet subdifferential

Recall that in Euclidean space, the subdifferential of a continuous function (not necessarily differentiable) $f : R^n \to R$ at $x \in R^n$ is defined as

$$\partial f(x) = \{g \in R^n : f(y) \geq f(x) + g \cdot (y - x), \forall y \in R^n\} \tag{8}$$

We can define the subdifferential in $P_2(R^D)$ in a similar way:

Let $H : P_2(R^D) \to (-\infty, +\infty]$ be a proper, lower semicontinuous function and let $H(\mu) < \infty$. We define Frechet subdifferential as

$$\begin{aligned}\partial H(\mu) = \{v \in L^2(\mu, R^D) : H(\nu) \geq H(\mu) + \sup_{\gamma \in \Gamma_o(\mu,\nu)} \int_{R^D \times R^D} \langle v(x), y - x \rangle d\gamma(x, y) \\ + o(W_2(\mu, \nu)) \text{ as } \nu \to \mu\},\end{aligned} \tag{9}$$

where $\sup_{\gamma \in \Gamma_o(\mu,\nu)} \int_{R^D \times R^D} \langle v(x), y - x \rangle d\gamma(x, y)$ could be viewed as the inner product of v and optimal transport displacement.

3 CONVEX ANALYSIS ON $P_2(R^D)$

3.1 CONSTANT SPEED GEODESIC IN $P_2(R^D)$

Let $\mu_0, \mu_1 \in P_2(R^D)$ and let $\gamma \in \Gamma_o(\mu_0, \mu_1)$ be an optimal transport plan. Let $\pi_1 : R^D \times R^D : (z, w) \to z$ and $\pi_2 : R^D \times R^D : (z, w) \to w$ be the first and second projections of $R^D \times R^D$ onto R^D.

Now we define the interpolation

$$\mu_t = ((1-t)\pi_1 + t\pi_2)_{\#}\gamma \tag{10}$$

then $t \to \mu_t$ is a geodesic in $P_2(R^D)$ of constant speed, i.e.

$$W_2(\mu_s, \mu_t) = |t-s| W_2(\mu_0, \mu_1) \tag{11}$$

Equipped with the constant speed geodesic, we could then define the convexity of functions in $P_2(R^D)$.

3.2 λ-CONVEXITY

Let $H : P_2(R^D) \to (-\infty, +\infty]$ be proper and let $\lambda \in R$. We say that H is λ-convex if for every $\mu_0, \mu_1 \in P_2(R^D)$ and every optimal transport plan $\gamma \in \Gamma_o(\mu_0, \mu_1)$, we have

$$H(\mu_t) \leq (1-t)H(\mu_0) + tH(\mu_1) - \lambda t(1-t) W_2^2(\mu_0, \mu_1) \quad \forall t \in [0, 1] \tag{12}$$

Here $\mu_t = ((1-t)\pi_1 + t\pi_2)_{\#}\gamma$, where π_1 and π_2 are the above projections.

Now we give an example as follows:

Let $\mu \in P_2(R^D)$ and define

$$H(\mu) = -\frac{1}{2} W_2^2(\mu, \nu) \quad \mu \in P_2(R^D). \tag{13}$$

Then H is (-1)-convex. Furthermore,

$$\partial H(\mu) \cap T_\mu P_2(R^D) = \{\bar{\gamma} - id : \gamma \in \Gamma_o(\mu, \nu)\} \tag{14}$$

where $\bar{\gamma}$ is the barycentric projection, characterized by

$$\int_{R^D} \phi(x)\bar{\gamma}(x) d\mu(x) = \int_{R^D \times R^D} \phi(x) y d\gamma(x, y), \quad \forall \phi \in C_b(R^D). \tag{15}$$

In particular:

$$\partial H(\mu) \cap T_\mu P_2(R^D) = \{t_\mu^\nu - id\}, \quad \forall \mu \in P_2^a(R^D). \tag{16}$$

where $P_2^a(R^D)$ is the subset of $P_2(R^D)$ consisting of probability measures absolutely continuous w.r.t Lebesgue measure.

4 HAMILTONIAN ODES

Definition 1 *Let $J : R^D \to R^D$ be linear map and $Jv \perp v$ for all $v \in R^D$. Let $H : P_2(R^D) \to (-\infty, +\infty]$ be a proper and lower semicontinuous function. We say an absolutely continuous curve $\mu_t : [0, T] \to P_2(R^D)$, $H(\mu_t) < \infty$, is a Hamiltonian ODE relative to H, starting from $\bar{\mu} \in P_2(R^D)$, if there exist $v_t \in L^2(\mu_t; R^D)$ with $\|v_t\|_{L^2(\mu_t)} \in L^1(0, T)$, such that*

$$\begin{cases} \dfrac{d}{dt}\mu_t + \nabla \cdot (J v_t \mu_t) = 0, \quad \mu_0 = \bar{\mu}, \quad t \in (0, t) \\ v_t \in T_{\mu_t} P_2(R^D) \cap \partial H(\mu_t) \quad \textit{for a.e. } t. \end{cases} \tag{17}$$

Theorem 3 *Let μ_t be a Hamiltonian ODE relative to H above, with $\|v_t\|_{L^2(\mu_t)} \in L^\infty(0, T)$. If H is λ-convex for some $\lambda \in R$, then $t \to H(\mu_t)$ is constant.*

The existence of solutions can be established if one imposes a growth condition on the gradient and a "continuity property" of the gradients.

5 DISCUSSION

The dynamics in W_2 space, and its connection to stochastic differential equations (SDEs), trying to apply to generative models, in specific, introduce dynamics to flow-based models, could be very important.

REFERENCES

Luigi Ambrosio and Wilfrid Gangbo. Hamiltonian odes in the wasserstein space of probability measures. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(1):18–53, 2008.