Parameterized Wasserstein Geometric Flow

Haomin Zhou
School of Math, Georgia Tech

Collaborators:
Wuchen Li (South Carolina) Shu Liu (UCLA), Xiaojing Ye (Georgia State), Hao Wu (Wells Fargo), and Hongyuan Zha (CUHK-SZ)

Partially Supported by NSF
Table of Contents

Introduction to Wasserstein manifold

Wasserstein Geometric Flow

Parameterized Wasserstein geometric flow

Numerical Examples
Ultimate goal of our study

Design efficient methods to compute the solution of Wasserstein geometric flow (especially the dynamics of probability distribution $\rho_t$), so that one can efficiently generate samples from the distribution $\rho_t$ at different time $t$; and sample particle trajectories from the corresponding dynamical systems, especially in high dimensions, while providing rigorous error bounds.
Introduction to optimal transport

- Monge optimal map (1781).
- Kantorovich relaxation (1942).
- Benamou-Brenier formula (2000):

\[ W_2^2(\rho_0, \rho_1) = \min_{\rho, v} \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{1}{2} |v|^2 \rho \, dx \, dt \right\}, \]

subject to: \( \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad \rho(\cdot, 0) = \rho_0, \quad \rho(\cdot, 1) = \rho_1. \)

- \( W_2(\rho_0, \rho_1) \) defines a distance between two probability densities.
Euclidean metric vs. Wasserstein metric

An extensive literature: Otto, Villani, Kinderlehrer, McCann, Carlen, Lott, Strum, Gangbo, Yau, Evans, Caffarelli, Figalli, Chow, Osher, Engquist, Fathi and many more.

1 C. Villani, Optimal transport. Old and new. 2009
2 F. Santambrogio, Optimal transport for applied mathematicians. 2015
Optimal transport on graph

The notion of optimal transport on discrete structures (graphs or networks) was proposed independently.

- S.N. Chow, W. Huang, Y. Li, H.M. Zhou (2012):

$$\min_{\rho, \nu} \left\{ \int_0^1 \frac{1}{2} (\nu, \nu)_{\rho} \, dt \right\},$$

subject to: $$\frac{d\rho}{dt} + \text{div}_G(\rho \nu) = 0, \ \rho(\cdot, 0) = \rho_0, \ \rho(\cdot, 1) = \rho_1.$$ 

- A rich subject with many studies on interesting math questions and applications.
Wasserstein manifold

- The optimal transport distance induces a metric on probability density set $\mathcal{P}(\mathbb{R}^d)$, a bilinear form

$$g^W(\eta_1, \eta_2) = \int_{\mathbb{R}^d} \frac{1}{2}(\nabla \Phi_1, \nabla \Phi_2)\rho dx.$$ 

where $\Phi_i$ ($i = 1, 2$) solves

$$-\nabla \cdot (\rho \nabla \Phi_i) = \eta_i.$$

- $(\mathcal{P}(\mathbb{R}^d), g^W)$ form a Riemannian manifold known as Wasserstein manifold.

- Geometric flow can be constructed on $(\mathcal{P}(\mathbb{R}^d), g^W)$. 
Wasserstein gradient flow

The **Wasserstein gradient flow (WGF)** of $\mathcal{F}(\rho)$ is

$$
\frac{\partial \rho}{\partial t} = -\text{grad}_W \mathcal{F}(\rho) = \nabla \cdot (\rho \nabla \frac{\delta \mathcal{F}(\rho)}{\delta \rho}(x)).
$$

<table>
<thead>
<tr>
<th>WGF</th>
<th>Equation</th>
<th>Associated functional $\mathcal{F}(\rho)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fokker-Planck</strong></td>
<td>$\partial_t \rho = \nabla \cdot (\rho \nabla V) + D \Delta \rho$</td>
<td>$\int V \rho + D \rho \log \rho , dx$</td>
</tr>
<tr>
<td><strong>Porous Medium</strong></td>
<td>$\partial_t \rho = \Delta(\rho^m)$</td>
<td>$\frac{1}{m-1} \int \rho^m , dx$</td>
</tr>
<tr>
<td><strong>Aggregation</strong></td>
<td>$\partial_t \rho = \nabla \cdot (\rho \nabla (W * \rho))$</td>
<td>$\iint W(x - y)\rho(x)\rho(y) , dxdy$</td>
</tr>
</tbody>
</table>

**Table:** Examples of Wasserstein gradient flows.
Wasserstein Hamiltonian flow

The Wasserstein Hamiltonian flow (WHF)

\[
\frac{\partial_t \rho(x, t)}{\delta \Phi_t} = \frac{\delta}{\delta \Phi_t} \mathcal{H}(\rho_t, \Phi_t) = -\nabla \cdot (\rho_t \nabla \Phi(x, t))
\]

\[
\frac{\partial_t \Phi(x, t)}{\delta \rho_t} = -\frac{\delta}{\delta \rho_t} \mathcal{H}(\rho_t, \Phi_t) = -\frac{1}{2} |\nabla \Phi(x, t)|^2 - \frac{\delta \mathcal{F}}{\delta \rho}(\rho_t, x),
\]

with

\[
\mathcal{H}(\rho, \Phi) = \int_{\mathbb{R}^d} \frac{1}{2} |\nabla \Phi|^2 \rho dx + \mathcal{F}(\rho).
\]

The density of a Hamiltonian flow in sample space is a Hamiltonian flow in density manifold\(^1\).

---

\(^1\)S. Chow, W. Li, H. Zhou, Wasserstein Hamiltonian flows, J. Differential Equations 268 (3) (2020)
Wasserstein Hamiltonian flow

<table>
<thead>
<tr>
<th>WHF</th>
<th>Equation system</th>
<th>Hamiltonian $\mathcal{H}(\rho, \Phi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Hamiltonian</td>
<td>$\partial_t \rho + \nabla \cdot (\rho \partial_p H(x, \nabla \Phi)) = 0$  \</td>
<td>$\int H(x, \nabla \Phi(x))\rho(x) , dx$</td>
</tr>
<tr>
<td>Hamiltonian system</td>
<td>$\partial_t \Phi + H(x, \nabla \Phi) = 0$</td>
<td>$H(x, \rho) = \frac{1}{2}</td>
</tr>
<tr>
<td>Wasserstein Geodesic</td>
<td>$\partial_t \rho + \nabla \cdot (\rho \nabla \Phi) = 0$</td>
<td>$\int \frac{1}{2}</td>
</tr>
<tr>
<td>Schrödinger equation$^1$</td>
<td>$\partial_t \rho + \nabla \cdot (\rho \nabla \Phi_t) = 0$</td>
<td>$\int \left( \frac{1}{2}</td>
</tr>
</tbody>
</table>

$^1$The corresponding Schrödinger equation of wave function $\psi = \sqrt{\rho}e^{i\Phi}$ is

$$i\partial_t \psi(x, t) = -\frac{1}{2} \Delta \psi(x, t) + V(x)\psi(x, t) + \frac{\delta}{\delta \rho} \mathcal{F}(|\psi(x, t)|^2, x)\psi(x, t).$$

Table: Examples of Wasserstein Hamiltonian flows.

There are many more examples like Schrödinger bridge problem and optimal density control.
Pushforward maps and parameterized space

• Consider a class of pushforward maps \( \{ T_\theta \}_\theta \in \Theta \) indexed by parameter \( \theta \in \Theta \subset \mathbb{R}^m \).

• A family of parametric distributions known as Generative Model

\[
\mathcal{P}_\Theta = \left\{ \rho_\theta = T_\theta^# \lambda \mid \theta \in \Theta \right\} \subset (\mathcal{P}, g^W),
\]

\( T_\theta^# \lambda \) defined by \( \rho_\theta(E) = \lambda(T_\theta^{-1}(E)) \).

• Generative Models by pushforward maps provide an easy-to-sample framework\(^1\) \(^2\).

---


Geometry on the parameter space

- Geometric structure on $\Theta$ by the pullback metric of $g^W$

$$G(\theta) = \int_{\mathbb{R}^d} \partial_\theta T_\theta(z)^\top \partial_\theta T_\theta(z) \, d\lambda(z).$$

- $\Theta$ and $G(\theta)$ form a Riemannian manifold.

- **Objective:** Compute $\{\theta_t\}$ s.t. $\rho_{\theta_t}$ serves as an approximation of $\rho_t$. Samples can be conveniently generated from $\rho_{\theta_t} \approx \rho_t$ by pushforward maps. Error bounds in Wasserstein metric can be established.
Parameterized Wasserstein geometric flows

- Parameterized WGF

\[ \dot{\theta}_t = -G(\theta_t) \nabla \theta F(\theta_t). \]

with \( F(\theta) = \mathcal{F}(\rho_\theta). \)

- Parameterized WHF

\[ \dot{\theta} = \nabla_p H(\theta, p) = G(\theta)^{-1} p, \]
\[ \dot{p} = -\nabla_\theta H(\theta, p) = \frac{1}{2} \left[ p^\top G(\theta)^{-\top} (\partial_{\theta_k} G(\theta)) G(\theta)^{-1} p \right]_{k=1}^m - \nabla_\theta F(\theta). \]

where

\[ H(\theta, p) = \frac{1}{2} p^\top G(\theta)^\dagger p + F(\theta). \]

- Rigorous error bounds are established for \( W_2(\rho_t, \rho_{\theta(t)}) \)
A 30-dim Parameteric Fokker-Planck equation

The 2-d projection of pushforward distribution of Fokker-Planck equation with potential

\[ V(x) = \frac{1}{50} \left( \sum_{i=1}^{30} x_i^4 - 16x_i^2 + 5x_i \right). \]
A 10-dim PWHF with quadratic potential

*Figure*: histogram: projection of pushforward distribution, curve: true distribution

The 1-d projection of pushforward distribution compared to the true density function.
Thank you!