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

Ultimate goal of our study

Design efficient methods to compute the solution of Wasserstein geometric flow (especially the dynamics of probability distribution ρ_t), so that one can *efficiently* generate samples from the distribution ρ_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

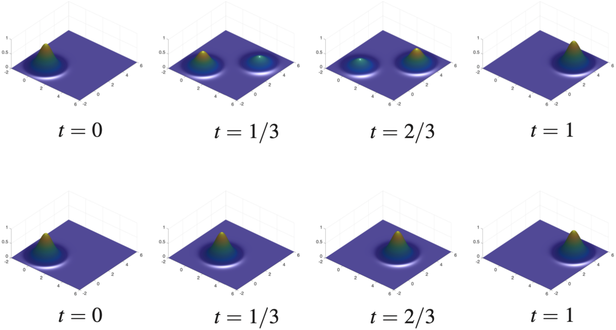
- Monge optimal map (1781).
- Kantorovich relaxation (1942).
- Benamou-Brenier formula (2000) :

$$W_2^2(\rho_0, \rho_1) = \min_{\rho, v} \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{1}{2}|v|^2 \ \rho \ dxdt \right\}, \tag{1}$$

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

- $W_2(\rho_0, \rho_1)$ 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.

- A. Mielke (2011).
- J. Maas (2011).
- S.N. Chow, W. Huang. Y. Li, H.M. Zhou (2012):

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

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

- 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 Φ_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

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

WGF	Equation	Associated functional $\mathcal{F}(\rho)$
Fokker-Planck equation	$\partial_t \rho = \nabla \cdot (\rho \nabla V) + D\Delta \rho$	$\int V\rho + D\rho \log \rho \ dx$
Porous Medium equation	$\partial_t \rho = \Delta(\rho^m)$	$\frac{1}{m-1} \int \rho^m \ dx$
Aggregation equation	$\partial_t \rho = \nabla \cdot (\rho \nabla (W * \rho))$	$\iint W(x-y)\rho(x)\rho(y) \ dxdy$

Table: Examples of Wasserstein gradient flows.

Wasserstein Hamiltonian flow

The **Wasserstein Hamiltonian flow (WHF)**

$$\begin{aligned} \partial_t\rho(x,t) &= \frac{\delta}{\delta\Phi_t}\mathcal{H}(\rho_t,\Phi_t) = -\nabla\cdot(\rho_t\nabla\Phi(x,t)) \\ \partial_t\Phi(x,t) &= -\frac{\delta}{\delta\rho_t}\mathcal{H}(\rho_t,\Phi_t) = -\frac{1}{2}|\nabla\Phi(x,t)|^2 - \frac{\delta\mathcal{F}}{\delta\rho}(\rho_t,x), \end{aligned} \tag{3}$$

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

WHF	Equation system	Hamiltonian $\mathcal{H}(\rho,\Phi)$
Classical Hamiltonian system	$\partial_t\rho + \nabla\cdot(\rho\partial_p H(x,\nabla\Phi)) = 0$ $\partial_t\Phi + H(x,\nabla\Phi) = 0$	$\int H(x,\nabla\Phi(x))\rho(x)\,dx$ $H(x,p) = \frac{1}{2}\lvert p\rvert^2 + V(x)$
Wasserstein Geodesic	$\partial_t\rho + \nabla\cdot(\rho\nabla\Phi) = 0$ $\partial_t\Phi + \frac{1}{2}\lvert\nabla\Phi\rvert^2 = 0$	$\int \frac{1}{2}\lvert\nabla\Phi(x)\rvert^2\rho(x)\,dx$
Schrödinger equation[1]	$\partial_t\rho + \nabla\cdot(\rho_t\nabla\Phi_t) = 0$ $\partial_t\Phi + \frac{1}{2}\lvert\nabla\Phi_t\rvert^2 = -V - \frac{\delta}{\delta\rho}\mathcal{F}(\rho,\cdot) + \frac{1}{2}\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}$	$\int(\frac{1}{2}\lvert\nabla\Phi\rvert^2 + V)\rho\,dx + \mathcal{F}(\rho) + \mathcal{I}(\rho)$

Table: Examples of Wasserstein Hamiltonian flows.

There are many more examples like Schrödinger bridge problem and optimal density control.

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

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

Pushforward maps and parameterized space

- Consider a class of pushforward maps $\{T_\theta\}_{\theta\in\Theta}$ ($T_\theta : \mathbb{R}^d \to \mathbb{R}^d$) 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].

[1] Ian J. Goodfellow et al. Generative Adversarial Nets. NIPS, 2014.

[2] Martin Arjovsky et al. Wasserstein Generative Adversarial Networks. ICML, 2016.

Geometry on the parameter space

- Geometric structure on Θ by the pullback metric of g^W

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

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

- Objective: Compute $\{\theta_t\}$ s.t. ρ_{θ_t} serves as an approximation of ρ_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

$$\dot{\theta}_t = -G(\theta_t)^\dagger \nabla_\theta F(\theta_t).$$

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

- Parameterized WHF

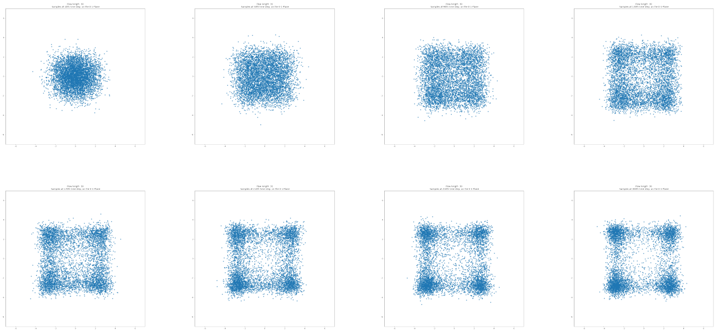
$$\begin{aligned} \dot{\theta} &= \nabla_p H(\theta, p) = G(\theta)^{-1} p, \\ \dot{p} &= -\nabla_\theta H(\theta, p) = \frac{1}{2}[p^\top G(\theta)^{-\top}(\partial_{\theta_k} G(\theta)) G(\theta)^{-1} p]_{k=1}^m - \nabla_\theta F(\theta). \end{aligned} \tag{4}$$

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

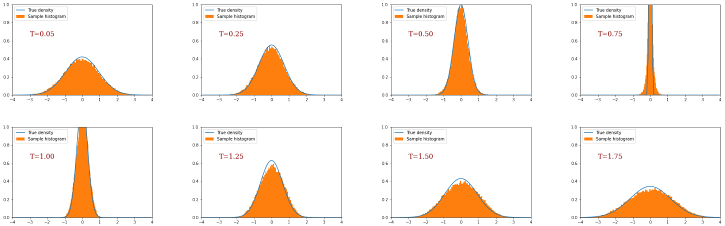


Figure: histogram: projection of pushforward distribution, curve: true distribution

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

Thank you!