
Scalable Computations of Wasserstein Barycenter via Input Convex Neural Networks

Jiaojiao Fan¹, Amirhossein Taghvaei², and Yongxin Chen¹

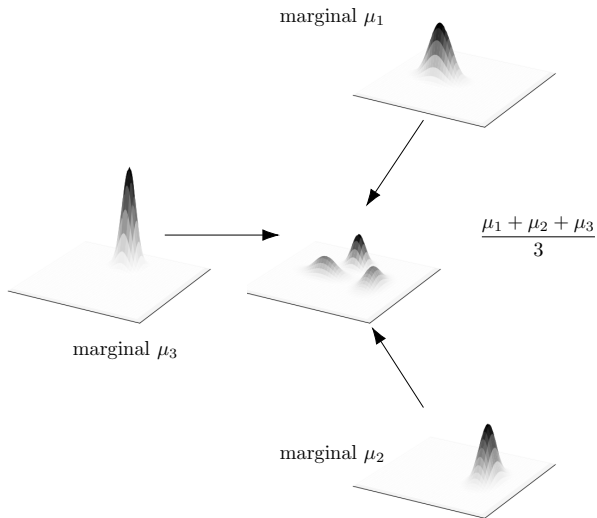


¹GEORGIA INSTITUTE OF TECHNOLOGY

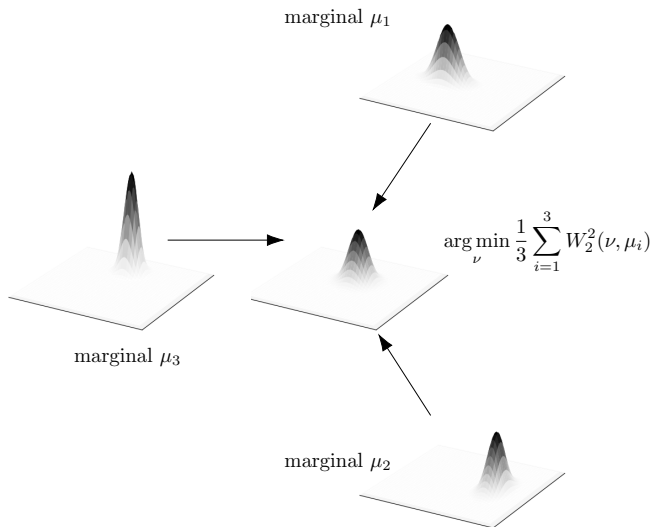
²UNIVERSITY OF CALIFORNIA, IRVINE

ICML 2021

Motivation



Motivation



Optimal transport

- Monge formulation (1781)

$$\inf_{T \# \nu = \mu} \int_{\mathbb{R}^n} c(x, T(x)) d\nu(x)$$

- Kantorovich formulation (1940)

$$\inf_{\pi \in \Pi(\nu, \mu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi(x, y)$$

$$T \# \nu = \mu \Leftrightarrow \nu(T^{-1}(A)) = \mu(A)$$

$\Pi(\nu, \mu)$: the set of joint distributions of ν, μ

Wasserstein distance

- Squared Wasserstein-2 distance between ν and μ :

$$W_2^2(\nu, \mu) := \min_{\pi \in \Pi(\nu, \mu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 d\pi(x, y)$$

- Kantorovich dual problem:

$$\frac{1}{2} W_2^2(\nu, \mu) = \sup_{(\phi, \psi) \in \Phi} \mathbb{E}_\nu[\phi(X)] + \mathbb{E}_\mu[\psi(Y)]$$

$$\Phi := \{(\phi, \psi) \in L^1(\nu) \times L^1(\mu); \phi(x) + \psi(y) \leq \frac{1}{2} \|x - y\|^2, \quad \forall x, y\}$$

Wasserstein distance

- Let $f(x) = \frac{\|x\|^2}{2} - \phi(x)$, $f'(y) = \frac{\|y\|^2}{2} - \psi(y)$

$$\frac{1}{2} W_2^2(\nu, \mu) = C_{\nu, \mu} - \inf_{(f, f') \in \Phi'} \{ \mathbb{E}_{\nu}[f(X)] + \mathbb{E}_{\mu}[f'(Y)] \}$$

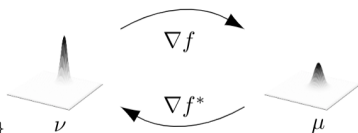
- Semi-dual formulation of optimal transport

$$\frac{1}{2} W_2^2(\nu, \mu) = C_{\nu, \mu} - \inf_{f \in \mathbf{CVX}} \{ \mathbb{E}_{\nu}[f(X)] + \mathbb{E}_{\mu}[f^*(Y)] \}$$

CVX stands for the set of convex functions, f^* is the convex conjugate function of f

$$C_{\nu, \mu} := (1/2) \{ \mathbb{E}_{\nu}[\|X\|^2] + \mathbb{E}_{\mu}[\|Y\|^2] \}$$

$$\Phi' := \{ (f, f') \in L^1(\nu) \times L^1(\mu); f(x) + f'(y) \geq xy, \forall x, y \}$$



Dual problem over convex functions

- Explicit form of f^*

$$f^*(y) = \sup_{g \in \mathbf{CVX}} \langle y, \nabla g(y) \rangle - f(\nabla g(y)) = \langle y, \nabla f^*(y) \rangle - f(\nabla f^*(y))$$

- Dual form of Wasserstein-2 distance¹

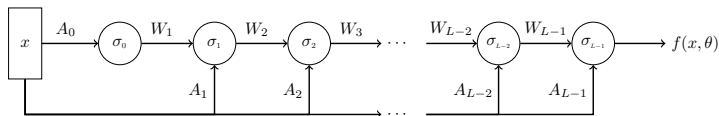
$$\frac{1}{2} W_2^2(\nu, \mu) = \sup_{f \in \mathbf{CVX}} \inf_{g \in \mathbf{CVX}} \mathcal{V}_{\nu, \mu}(f, g) + C_{\nu, \mu}$$

where

$$\mathcal{V}_{\nu, \mu}(f, g) = -\mathbb{E}_{\nu}[f(X)] - \mathbb{E}_{\mu}[\langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))]$$

1. Makuva, A., Taghvaei, A., Oh, S., & Lee, J. "Optimal transport mapping via input convex neural networks." ICML 2020

Input Convex Neural Networks



- This network defines a convex map $x \mapsto f(x; \theta) = z_L$ and

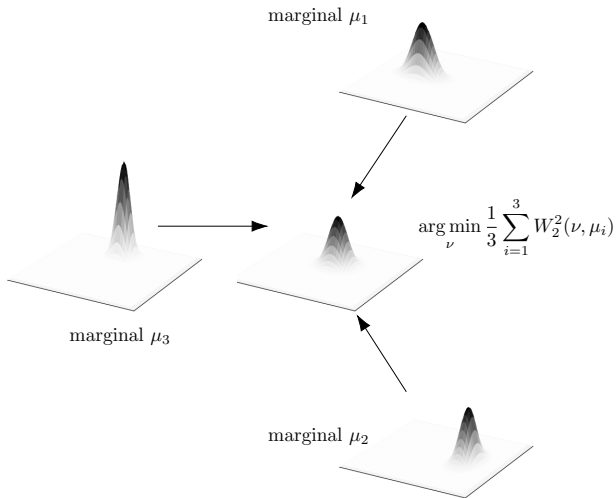
$$z_{\ell+1} = \sigma_{\ell}(W_{\ell}z_{\ell} + A_{\ell}x + b_{\ell}).$$

- This map $f(x; \theta)$ is convex if 1) $W_{1:L-1}$ are non-negative; 2) $\sigma_{0:L-1}$ are convex; 3) $\sigma_{1:L-1}$ are non-decreasing.

Amos, Brandon, Lei Xu, and J. Zico Kolter. "Input convex neural networks." ICML 2017.

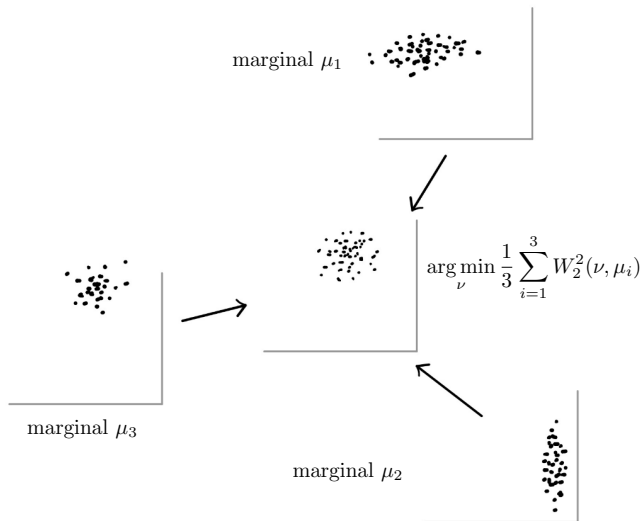
Wasserstein Barycenter

Wasserstein barycenter $\tilde{\nu}$ is the minimizer of $\sum_{i=1}^N a_i W_2^2(\nu, \mu_i)$.



Wasserstein Barycenter

- Only require samples from marginal distributions



Cost function

Minimizing summed W -2 distances:

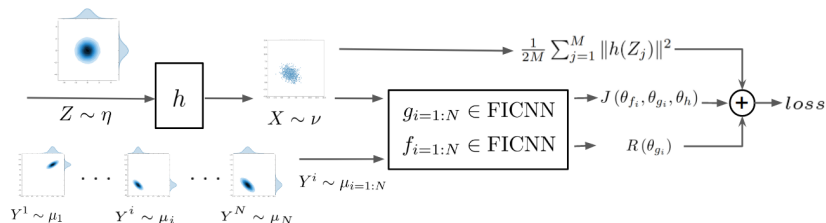
$$\min_h \sup_{f_i \in \text{ICNN}} \inf_{g_i \in \text{ICNN}} \frac{1}{2} \mathbb{E}_\eta [\|h(Z)\|^2] + \sum_{i=1}^N a_i \bar{\mathcal{V}}_{\eta, \mu_i}(f_i, g_i)$$

where

$$\bar{\mathcal{V}}_{\eta, \mu_i}(f, g) = -\mathbb{E}_\eta[f_i(h(Z))] - \mathbb{E}_{\mu_i}[\langle Y^i, \nabla g_i(Y^i) \rangle - f_i(\nabla g_i(Y^i))]$$

$$Z \sim \mathcal{N}(\mathbf{0}, I)$$

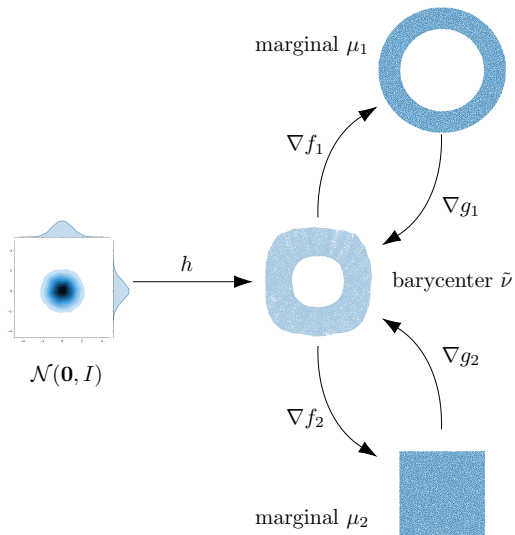
Algorithm



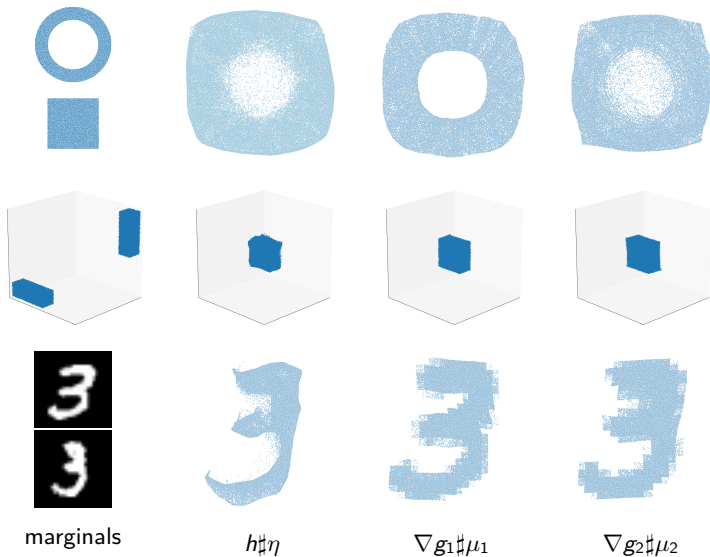
$$J(\theta_{f_i}, \theta_{g_i}, \theta_h) = \frac{1}{M} \sum_{j=1}^M f_i \left(\nabla g_i \left(Y_j^i \right) \right) - \left\langle Y_j^i, \nabla g_i \left(Y_j^i \right) \right\rangle - f_i \left(h(Z_j) \right)$$

$$R(\theta_{g_i}) = \lambda \sum_{W_l \in \theta_{g_i}} \|\max(-W_l, 0)\|_F^2$$

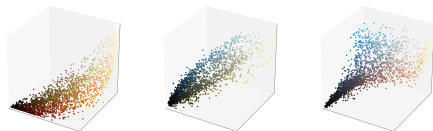
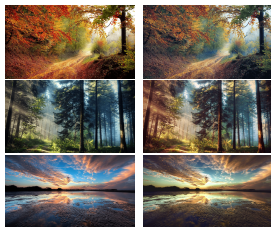
Two methods to recover the barycenter



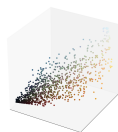
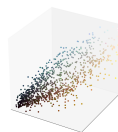
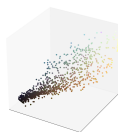
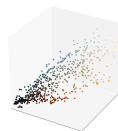
2d and 3d



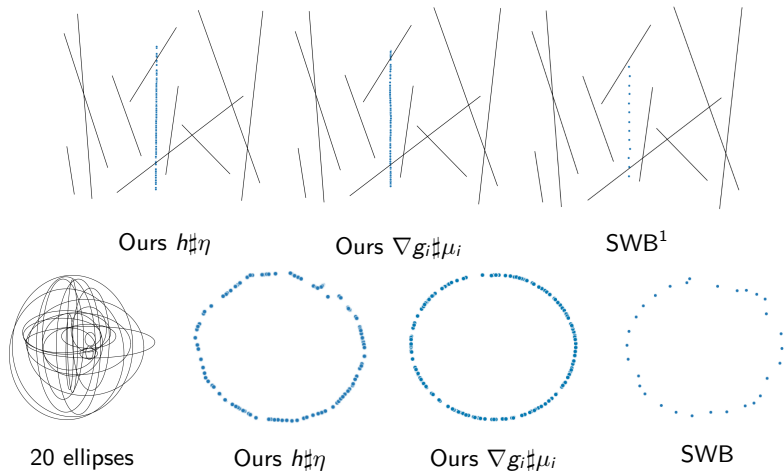
Color palette averaging

Source colors $\{I_i\}$ 

Left: Source images
Right: Pushforward images

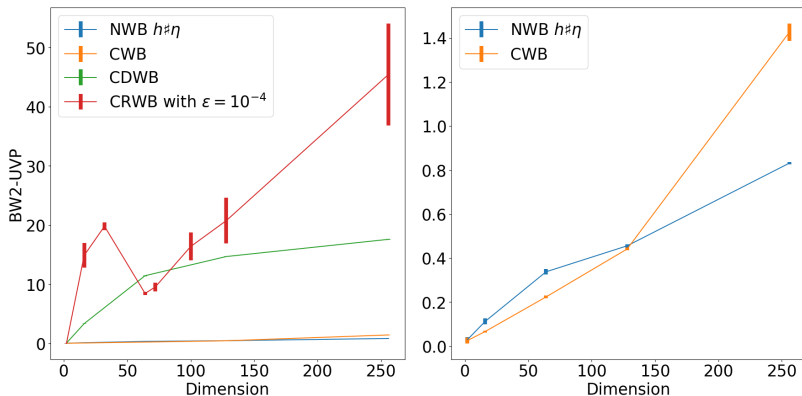
 $\nabla g_1 \# \mu(I_1)$  $\nabla g_2 \# \mu(I_2)$  $\nabla g_3 \# \mu(I_3)$  $h \# \eta$

Sharp geometries



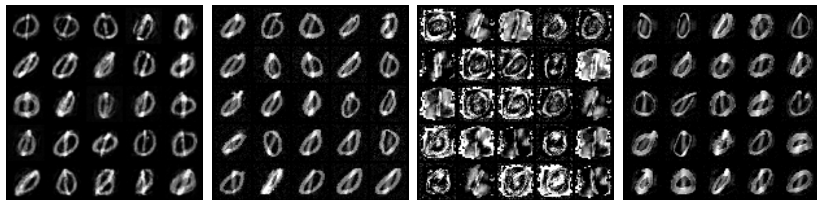
Claici, S., Chien, E., & Solomon, J. "Stochastic wasserstein barycenters." ICML 2018

Scalability with Gaussian marginals



- CWB: continuous WB without minimax (Korotin et al., 2021)
- CDWB: fast free-support WB (Cuturi & Doucet, 2014)
- CRWB: continuous regularized WB (Li et al., 2020)

MNIST 0 and 1

Ours $h\#\eta$ Ours $\nabla g_i\#\mu_i$

CRWB

CWB

- CRWB: continuous regularized WB (Li et al., 2020)
- CWB: continuous WB without minimax (Korotin et al., 2021)

MNIST and USPS


 μ_1 : MNIST

 μ_2 : USPS
Ours $h\#\eta$ Ours $\nabla g_i\#\eta(\mu_i)$

Barycenter serving as GAN

- One marginal: the barycenter $\tilde{\nu} =$ the marginal μ_1

$$h(Z) \sim \mu_1, \quad Z \sim \mathcal{N}(\mathbf{0}, I)$$

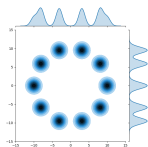
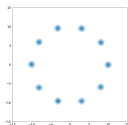
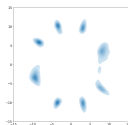
Similar to Wasserstein GAN but with Wasserstein-2 metric

- Multiple marginals:

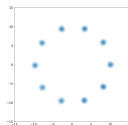
$$h(Z) \sim \tilde{\nu} \Rightarrow \nabla f_i(h(Z)) \sim \mu_i, \quad Z \sim \mathcal{N}(\mathbf{0}, I)$$

Generate multiple marginal distributions after one training

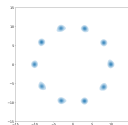
One marginal: Gaussian mixture

marginal μ_1 Ours $h_{\sharp}^{\eta}$ 

WGAN



WGAN-GP

W2GN¹

1. Korotin A, Egiazarian V, Asadulaev A, Safin A, Burnaev E. "Wasserstein-2 generative networks." ICLR 2021

One marginal: MNIST

Ours $h\#eta$

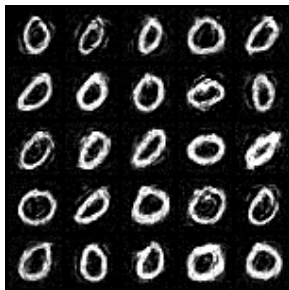
WGAN

WGAN-GP

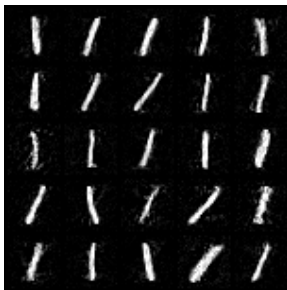
W2GN¹

1. Korotin A, Egiazarian V, Asadulaev A, Safin A, Burnaev E. "Wasserstein-2 generative networks." ICLR 2021

Multiple marginals: 0 and 1



generated '0'



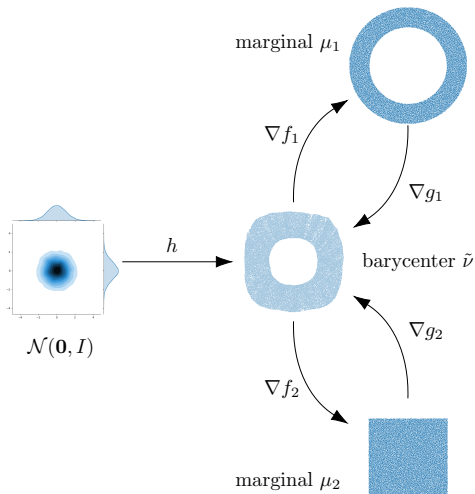
generated '1'

Multiple marginals: MNIST and USPS

 μ_1 : MNIST μ_2 : USPSgenerated μ_1 generated μ_2

Takeaway

- Require only samples from marginal distributions
- Recover as many barycenter samples as one wish
- Generate multiple marginals after one training





Thank you!

Contact:

`jiaojiaofan@gatech.edu`

ArXiv:

`https://arxiv.org/abs/2007.04462`

Code:

`https://github.com/sbyebs/Scalable-Wasserstein-Barycenter`