Optimal transport mapping via input convex neural networks

Ashok Vardhan Makkuva   Amirhossein Taghvaei   Jason D. Lee
Sewoong Oh

April 22, 2021

- Introduction
- Formulation of 2-Wasserstein distance
- Minimax optimization over ICNNs
- Expriments

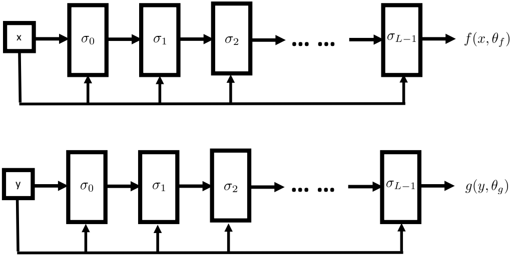
Notation

- $\mathcal{P}(\mathcal{X})$, the set of probability measures on a Polish space $\mathcal{X}$. $P, Q \in \mathcal{P}(\mathcal{X})$
- $\mathcal{B}(\mathcal{X})$, the Borel subsets of $\mathcal{X}$
- $T : \mathcal{X} \to \mathcal{Y}$, the measurable map, $(T\#Q)(A) = Q(T^{-1}(A)), \forall A \in \mathcal{B}(\mathcal{Y})$
- $L^1(P) := \{f \ is \ measurable \ \& \int f \ \mathrm{d}P < \infty\}$.
- $CVX(P)$, the set of all convex functions in $L^1(P)$.

Introduction

$$W_2^2(P,Q)$$

$$\sup_f \inf_g \mathcal{V}_{P,Q}(f,g) + C_{P,Q}$$



- Introduction
- Formulation of 2-Wasserstein distance
- Minimax optimization over ICNNs
- Expriments

Formulation of 2-Wasserstein distance

This part builds on the work in[1], which restricts the optimization problem to the variants of convex functions and leverages the input-convex neural networks to approximate 2-Wasserstein distance.

[1] Amirhossein Taghvaei and Amin Jalali. "2-wasserstein approximation via restricted convex potentials with application to improved training for gans". In: *arXiv preprint arXiv:1902.07197* (2019).

Formulation of 2-Wasserstein distance

$$W_2^2(P,Q) = \inf_{\pi \in \prod(P,Q)} \frac{1}{2} \mathbb{E}_{(X,Y)\sim\pi} \|X - Y\|^2 \tag{1}$$

where $\prod(P,Q)$ denotes the set of all joint probability distributions whose first and second marginals are P and Q.

Formulation of 2-Wasserstein distance

$$W_2^2(P,Q) = \inf_{\pi \in \prod(P,Q)} \frac{1}{2} \mathbb{E}_{(X,Y)\sim\pi} \|X - Y\|^2 \tag{1}$$

where $\prod(P,Q)$ denotes the set of all joint probability distributions whose first and second marginals are P and Q.

$$W_2^2(P,Q) = \sup_{(f,g)\in\Phi_c} \mathbb{E}_P[f(X)] + \mathbb{E}_Q[g(Y)] \tag{2}$$

where $\Phi_c := \{(f,g) \in L^1(P) \times L^1(Q) : f(x) + g(y) \leq \frac{1}{2}\|x-y\|_2^2, \forall (x,y) \mathrm{d}P \otimes \mathrm{d}Q\}$

Formulation of 2-Wasserstein distance

$$f(x) + g(y) \leq \frac{1}{2} \|x - y\|_2^2$$

Formulation of 2-Wasserstein distance

$$f(x) + g(y) \leq \frac{1}{2} \|x - y\|_2^2$$
$$\Longleftrightarrow \left[\frac{1}{2} \|x\|_2^2 - f(x)\right] + \left[\frac{1}{2} \|y\|_2^2 - g(y)\right] \geq \langle x, y \rangle$$

Formulation of 2-Wasserstein distance

$$
\begin{aligned}
& f(x)+g(y) \leq \frac{1}{2}\|x-y\|_2^2 \\
\Longleftrightarrow\ & \left[\frac{1}{2}\|x\|_2^2 - f(x)\right] + \left[\frac{1}{2}\|y\|_2^2 - g(y)\right] \geq \langle x, y\rangle \\
\Longleftrightarrow\ & f(x)+g(y) \geq \langle x, y\rangle
\end{aligned}
$$

Reparametrizing $\frac{1}{2}\|\cdot\|_2^2 - f(\cdot)$ and $\frac{1}{2}\|\cdot\|_2^2 - g(\cdot)$ by f and g,

Formulation of 2-Wasserstein distance

$$W_2^2(P,Q) = \sup_{(f,g)\in\tilde{\Phi}_c} \mathbb{E}_P\left[\frac{1}{2}\|X\|_2^2 - f(X)\right] + \mathbb{E}_Q\left[\frac{1}{2}\|Y\|_2^2 - g(Y)\right] \quad (3)$$

where
$\tilde{\Phi}_c := \{(f,g) \in L^1(P) \times L^1(Q) : f(x) + g(y) \geq \langle x, y\rangle, \forall (x,y) \mathrm{d}P \otimes \mathrm{d}Q\}$

Formulation of 2-Wasserstein distance

$$W_2^2(P,Q) = \sup_{(f,g)\in\tilde{\Phi}_c} \mathbb{E}_P\left[\frac{1}{2}\|X\|_2^2 - f(X)\right] + \mathbb{E}_Q\left[\frac{1}{2}\|Y\|_2^2 - g(Y)\right] \quad (3)$$

where
$\tilde{\Phi}_c := \{(f,g) \in L^1(P) \times L^1(Q) : f(x) + g(y) \geq \langle x, y\rangle, \forall (x,y) \mathrm{d}P \otimes \mathrm{d}Q\}$

$$W_2^2(P,Q) = \frac{1}{2}\mathbb{E}[\|X\|_2^2 + \|Y\|_2^2] + \sup_{(f,g)\in\tilde{\Phi}_c} [-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[g(Y)]]$$

Formulation of 2-Wasserstein distance

Theroem 2.9 (Existence of an optimal pair of convex conjugate functions)[2] Let P, Q be two probability measures on $\mathbb{R}^d$, with finite second order moments. There exists a pair (f, f^*) of lower semi-continuous proper conjugate convex functions on $\mathbb{R}^d$, then we can get

$$W_2^2(P, Q) = C_{P,Q} + \sup_{f \in CVX(P)} [-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f^*(Y)]] \tag{4}$$

where $C_{P,Q} = \frac{1}{2}\mathbb{E}[\|X\|_2^2 + \|Y\|_2^2]$, and $f^*(y) = \sup_x \langle x, y \rangle - f(x)$ is the convex conjugate of $f(\cdot)$.

[2] Cédric Villani. *Topics in optimal transportation*. 58. American Mathematical Soc., 2003.

- Introduction
- Formulation of 2-Wasserstein distance
- Minimax optimization over ICNNs
- Expriments

Minimax formulation

$$W_2^2(P,Q) = C_{P,Q} + \sup_{f \in CVX(P)} \left[-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f^*(Y)] \right]$$

Minimax formulation

$$W_2^2(P,Q) = C_{P,Q} + \sup_{f \in CVX(P)} \left[-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f^*(Y)]\right]$$

Using a minimax formulation,

$$W_2^2(P,Q) = C_{P,Q} + \sup_{f \in CVX(P)} \inf_{g \in CVX(Q)} \mathcal{V}_{P,Q}(f,g) \tag{5}$$

where

$$\mathcal{V}_{P,Q} = -\mathbb{E}_P[f(X)] - \mathbb{E}_Q[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))] \tag{6}$$

Minimax formulation

$$f^*(Y) \;\geq\; \langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))$$

Minimax formulation

$$f^*(Y) \ \geq \ \langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))$$

$$\mathbb{E}_Q[f^*(Y)] \ \geq \ \mathbb{E}_Q[\langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))]$$

Minimax formulation

$$
\begin{aligned}
f^*(Y) \;&\geq\; \langle Y, \nabla g(Y)\rangle - f(\nabla g(Y)) \\
\mathbb{E}_Q[f^*(Y)] \;&\geq\; \mathbb{E}_Q[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))] \\
-\mathbb{E}_Q[f^*(Y)] \;&\leq\; -\mathbb{E}_Q[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))]
\end{aligned}
$$

Minimax formulation

$$
\begin{aligned}
f^*(Y) &\geq \langle Y, \nabla g(Y) \rangle - f(\nabla g(Y)) \\
\mathbb{E}_Q[f^*(Y)] &\geq \mathbb{E}_Q[\langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))] \\
-\mathbb{E}_Q[f^*(Y)] &\leq -\mathbb{E}_Q[\langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))] \\
-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f^*(Y)] &\leq -\mathbb{E}_P[f(X)] - \mathbb{E}_Q[\langle Y, \nabla g(Y) \rangle - f(\nabla g(Y))]
\end{aligned}
$$

Minimax formulation

$$
\begin{aligned}
f^*(Y) &\geq \langle Y, \nabla g(Y)\rangle - f(\nabla g(Y)) \\
\mathbb{E}_Q[f^*(Y)] &\geq \mathbb{E}_Q[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))] \\
-\mathbb{E}_Q[f^*(Y)] &\leq -\mathbb{E}_Q[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))] \\
-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f^*(Y)] &\leq -\mathbb{E}_P[f(X)] - \mathbb{E}_Q[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))] \\
-\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f^*(Y)] &= \inf_{g \in CVX(Q)} \mathcal{V}_{P,Q}(f, g)
\end{aligned}
$$

Minimax formulation

$$
\begin{aligned}
\nabla g(y) &= \nabla(\frac{1}{2}\left\|y\right\|_2^2 - g_o(y)) \\
&= y - \nabla g_o(y)
\end{aligned}
$$

Suppose T is the optimal transport map, then $\nabla g_o(y) = \nabla_y \frac{1}{2}\left\|x-y\right\|_2^2 = y - x$, plugging it into above, we can get $\nabla g(y) = x$.

By the defination of convex conjugate, $f^*(y) = \sup_x \langle x, y\rangle - f(x)$, then we can get $f^*(y) = \langle y, \nabla g(y)\rangle - f(\nabla g(y))$

ICNN

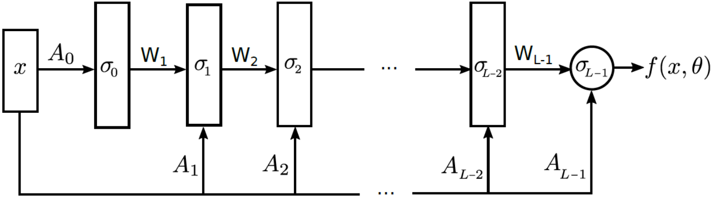


Figure 1: The input convex neural network (ICNN) architecture

$$z_{l+1} = \sigma_l(W_l z_l + A_l x + b_l), \;\; f(x;\theta) = z_L \tag{7}$$

where $\{W_l\}$, $\{A_l\}$ are weight matrices, and $\{b_l\}$ are the bias terms, and $\theta = (\{W_l\}, \{A_l\}, \{b_l\})$.

ICNN

$$z_{l+1} = \sigma_l(W_l z_l + A_l x + b_l),\ \ f(x;\theta) = z_L \tag{8}$$

To ensure that $f(x;\theta)$ is convex,

- all entries of the weights W_l are non-negative
- activation function σ_0 is convex
- σ_l is convex and non-decreasing, for $l = 1, \ldots, L-1$.

Minimax optimization over ICNNs

$$\max_{\theta_f} \min_{\theta_g} J(\theta_f, \theta_g) + R(\theta) \tag{9}$$

where $R(\cdot)$ denotes the regularization term, and $J(\theta_f, \theta_g) = \frac{1}{M}\sum_{i=1}^{M} -f(X_i) - \langle Y_i, \nabla g(Y_i)\rangle + f(\nabla g(Y_i))$ corresponding to

$$\mathcal{V}_{P,Q} = -\mathbb{E}_P[f(X)] - \mathbb{E}_Q\left[\langle Y, \nabla g(Y)\rangle - f(\nabla g(Y))\right]$$

Minimax optimization over ICNNs

Algorithm 1 The numerical procedure to solve the optimization problem (9).

Input: Source dist. Q, Target dist. P, Batch size M, Generator iterations K, Total iterations T
for $t = 1, \ldots, T$ **do**
  Sample batch $\{X_i\}_{i=1}^{M} \sim P$
  for $k = 1, \ldots, K$ **do**
    Sample batch $\{Y_i\}_{i=1}^{M} \sim Q$
    Update θ_g to minimize (9) using Adam method
  end for
  Update θ_f to maximize (9) using Adam method
  Projection: $w \leftarrow \max(w, 0)$, for all $w \in \{W^l\} \in \theta_f$
end for

- Introduction
- Formulation of 2-Wasserstein distance
- Minimax optimization over ICNNs
- Expriments

Minimax optimization over ICNNs

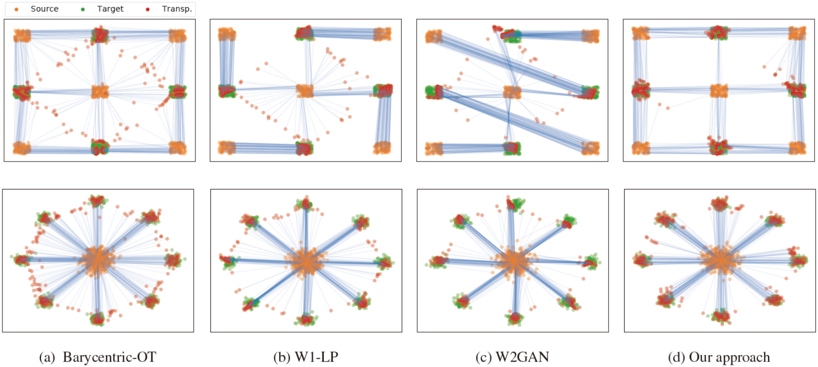


Figure 2: The transport maps learned by various approaches on 'Checker board' and 'mixture of eight Gaussians' datasets.