Optimal Transport HW 2

Bipin Koirala

April 9, 2023

Problem 1

Let, μ be a uniform distribution over the interval $[0,1]$, ν be a uniform distribution over the set $[-1/2,0]\cup[3/2,2]$. Obtain the displacement interpolation between μ and ν with respect to W_2 (Wasserstein-2) metric.

Solution: Here, the CDF of each distribution is given by

$$F_\mu(x)=\begin{cases}0, & \text{if } x<0\\ x, & \text{if } 0\le x\le 1\\ 1, & \text{if } x>1\end{cases}$$

and

$$F_\nu(x)=\begin{cases}0, & \text{if } x<-\frac{1}{2}\\ x+\frac{1}{2}, & \text{if } -\frac{1}{2}\le x\le 0\\ \frac{1}{2}, & \text{if } 0<x<\frac{3}{2}\\ x-1, & \text{if } \frac{3}{2}\le x\le 2\\ 1, & \text{if } x>2\end{cases}$$

The optimal transport map in Wassertein-2 metric is obtained via push-forward of CDF of source distribution μ through the inverse of the target distribution ν. Denote the inverse of CDF of target distribution as $F_\nu^{-1}(y)$. It is given by;

$$F_\nu^{-1}(y)=\begin{cases}y-\frac{1}{2}, & \text{if } 0\le y\le \frac{1}{2}\\ y+1, & \text{if } \frac{1}{2}<y\le 1\end{cases}$$

Now we can apply $F_\nu^{-1}(y)$ to $F_\mu(x)$ to obtain the optimal transport map $T(x)$:

$$T(x)=F_\nu^{-1}(F_\mu(x))=F_\nu^{-1}(x)=\begin{cases}x-\frac{1}{2}, & \text{if } 0\le x\le \frac{1}{2}\\ x+1, & \text{if } \frac{1}{2}<x\le 1\end{cases}$$

Now, the displacement interpolation is given by; $\mu_t=\big((1-t)x+tT(x)\big)_{\#}\mu$ i.e.

$$(1-t)x+tT(x)=\begin{cases}x-\frac{t}{2}, & \text{if } 0\le x\le \frac{1}{2}\\ x+t, & \text{if } \frac{1}{2}<x\le 1\end{cases}$$

When, $0 \leq x \leq \frac{1}{2}$, the displacement interpolation is given by;

$$\mu_t = (x - \frac{t}{2})_{\#}\mu$$

To get the PDF of μ_t, we can use change of variables. Let, $y = x - \frac{t}{2}$ then $\mu_t(y) = \mu_t(x)\left|\frac{dx}{dy}\right|$. Here, PDF of $\mu_t(x) = 1$ for $0 \leq x \leq 1$ and Jacobian evaluates to 1. Therefore;

$$\mu_t(y) = \begin{cases} 1, & \text{if } -t/2 \leq y \leq 1 - t/2 \\ 0, & \text{otherwise} \end{cases}$$

similarly, for $\frac{1}{2} < x \leq 1$, the displacement interpolation is given by;

$$\mu_t(y) = (x + t)_{\#}\mu = \begin{cases} 1, & \text{if } 1/2 + t \leq y \leq t + 1 \\ 0, & \text{otherwise} \end{cases}$$

Problem 2

Given any two sets $X \subset \mathbb{R}^n, Y \subset \mathbb{R}^n$, their Minkowski sum $X + Y$ is defined to be

$$X + Y = \{x + y; \quad x \in X, y \in Y\}$$

The Brunn-Minkowski inequality states that

$$|X + Y|^{1/n} \geq |X|^{1/n} + |Y|^{1/n}$$

where $|X|$ stands for the Lebesgue measure of X. Prove that the Brunn-Minkowski inequality using displacement convexity of the functional

$$\mathcal{U}(\rho) = -\int_{\mathbb{R}^n} \rho(x)^{1-1/n} \, dx$$

solution: Let ρ_0, ρ_1 be two probability distribution in $\mathbb{R}^n$; parameterized by $t \in [0, 1]$ such that $\rho_t(x) = (1 - t)\rho_0(x) + t\rho_1(x)$. Now,

$$\mathcal{U}(\rho_t) = -\int_{\mathbb{R}^n} \big\{(1 - t)\rho_0(x) + t\rho_1(x)\big\}^{1-1/n} dx$$

The function $(\cdot)^{1-1/n}$ is concave for $n \geq 1$. Therefore;

$$\big\{(1 - t)\rho_0(x) + t\rho_1(x)\big\}^{1-1/n} \geq (1 - t)\rho_0(x)^{1-1/n} + t\rho_1(x)^{1-1/n}$$

Integrating both sides $\mathbb{R}^n$ we get;

$$\int_{\mathbb{R}^n} \big\{(1 - t)\rho_0(x) + t\rho_1(x)\big\}^{1-1/n} \geq \int_{\mathbb{R}^n} (1 - t)\rho_0(x)^{1-1/n} + \int_{\mathbb{R}^n} t\rho_1(x)^{1-1/n}$$

$$-\mathcal{U}(\rho_t) \geq -(1 - t)\mathcal{U}(\rho_0) - t\mathcal{U}(\rho_1)$$

$$\boxed{\mathcal{U}(\rho_t) \leq (1 - t)\mathcal{U}(\rho_0) + t\mathcal{U}(\rho_1)}$$

$\mathcal{U}(\rho)$ is indeed displacement convex. Now, let $\rho_0(x) = \frac{\chi_X(x)}{|X|}$, $\rho_1(y) = \frac{\chi_Y(y)}{|Y|}$ and $\rho_t(x) = \frac{\chi_{(1-t)X+tY}(x)}{|(1-t)X+tY|}$ be characteristic densities.

From displacement convexity,

$$
\begin{aligned}
\int_{\mathbb{R}^n} \frac{\chi_{(1-t)X+tY}(x)}{|(1-t)X+tY|^{1-1/n}} dx &\geq \int_{\mathbb{R}^n} \frac{\chi_X(x)}{|X|^{1-1/n}} dx + \int_{\mathbb{R}^n} \frac{\chi_Y(y)}{|Y|^{1-1/n}} dy \\
\int_{x \in \chi_{(1-t)X+tY}} \frac{1}{|(1-t)X+tY|^{1-1/n}} dx &\geq \int_{x \in \chi_X} \frac{1}{|X|^{1-1/n}} dx + \int_{x \in \chi_Y} \frac{1}{|Y|^{1-1/n}} dy \\
\frac{|(1-t)X+tY|}{|(1-t)X+tY|^{1-1/n}} &\geq \frac{|X|}{|X|^{1-1/n}} + \frac{|Y|^{1-1/n}}{|Y|^{1-1/n}} \\
|(1-t)X+tY|^{1/n} &\geq |X|^{1/n} + |Y|^{1/n}
\end{aligned}
$$

The above inequality is valid for all $t \in [0,1]$, so taking $t = 1/2$

$$
\left(\frac{1}{2}\right)^{1/n} |X+Y|^{1/n} \geq \frac{1}{2}\Big(|X|^{1/n} + |Y|^{1/n}\Big)
$$

Since $(\frac{1}{2})^{1/n} \geq \frac{1}{2} \ \forall \ n \geq 1$. Therefore;

$$
|X+Y|^{1/n} \geq |X|^{1/n} + |Y|^{1/n}
$$

Problem 3

Let μ be a Gaussian distribution with mean m_0 and covariance $\Sigma_0 > 0$. Let ν be a Gaussian distribution with mean m_1 and covariance $\Sigma_1 > 0$. Both distributions are over the Euclidean space $\mathbb{R}^n$. Obtain the displacement interpolation between μ and ν with respect to W_2 (Wasserstein-2) metric.

Solution:

Wasserstein-2 distance between two distributions is given by $W_2(\mu,\nu) = \sqrt{\min_{\pi \in \prod(\mu,\nu)} \int ||x-y||^2 d\pi}$. This W_2 behaves like a Riemannian metric. The geodesic between μ and ν. is given by

$$
\min_{\rho,\mu} \int_0^1 \langle \nabla\varphi, \nabla\varphi \rangle_{\rho_t} \, dt
$$

s.t. $\frac{\partial \rho_t}{\partial t} + \nabla(\rho \nabla \varphi) = 0$ and $\rho_0 = \mu$ and $\rho_1 = \nu$. Where $\rho_t = ((1-t)I_d + tT^*)_{\#}\mu$ is the push-forward of μ at any time t. This fluid-dynamic formulation of optimal transport gives an upper bound on the length (geodesic).

This geodesic gives the displacement interpolation.

W_2 distance between two Gaussian measures has a closed form expression and is given by:

$$
W_2^2(\mu,\nu) = ||m_0 - m_1||^2 + \text{trace}\left(\Sigma_0 + \Sigma_1 - 2\sqrt{\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2}}\right)
$$

The displacement interpolation between any two Gaussian measure is also a Gaussian measure. Based on McCann's theorem, the linear map T between two centered Gaussian distribution $\mathcal{N}(0,\Sigma_0)$ and $\mathcal{N}(1,\Sigma_1)$ is given by;

$$
T = \Sigma_1^{1/2}\left(\Sigma_1^{1/2}\Sigma_0\Sigma_1^{1/2}\right)^{-1/2}\Sigma_1^{1/2}
$$

and $y = Tx$ push forward $\mathcal{N}(0, \Sigma_0)$ to $\mathcal{N}(0, \Sigma_1)$. Therefore, the optimal transport plan is $[I_d \times y]_\# \mathcal{N}(0, \Sigma_0)$. Now, let $A(t) = [(1-t)T + tT] \quad \forall t \in [0,1]$. With this the geodesic from $\mathcal{N}(0, \Sigma_0)$ to $\mathcal{N}(0, \Sigma_1)$ is given by

$$\mathcal{N}(0, A(t)\Sigma_1 A(t))$$

Furthermore, since this is a linear interpolation, the final geodesic is given by

$$\mathcal{N}\Big((1-t)m_0 + t\ m_1, A(t)\Sigma_1 A(t)\Big)$$

Problem 4

Let μ, ν be two probability vectors, and π be a joint distribution of these two vectors. Let H denote the entropy. Prove

$$H(\pi) \leq H(\mu) + H(\nu)$$

and

$$H(\pi) \geq \max(H(\mu), H(\nu))$$

Solution: From the definition of entropy we have the following expressions

$$\begin{aligned}
H(\pi) &= -\sum_{i,j} \pi(i,j) \log \pi(i,j) \\
H(\mu) &= -\sum_{i} \mu(i) \log \mu(i) \\
H(\nu) &= -\sum_{j} \nu(j) \log \nu(j)
\end{aligned}$$

The K.L. divergence between the joint distribution (π) and the product distribution of μ and ν is given by:

$$D_{KL}(\pi || \mu \times \nu) = \sum_{i,j} \pi(i,j) \log \frac{\pi(i,j)}{\mu(i)\nu(j)}$$

Since, $D_{KL} \geq 0$ we have the following;

$$\begin{aligned}
-\sum_{i,j} \pi(i,j) \log \pi(i,j) &\leq -\sum_{i,j} \pi(i,j) \log \mu(i) - \sum_{i,j} \pi(i,j) \log \nu(j) \\
&= -\sum_{i} (\sum_{j} \pi(i,j)) \log \mu(i) - \sum_{j} (\sum_{i} \pi(i,j)) \log \nu(j) \\
&= -\sum_{i} \mu(i) \log \mu(i) - \sum_{j} \nu(j) \log \nu(j)
\end{aligned}$$

By definition of entropy we get;

$$\boxed{H(\pi) \leq H(\mu) + H(\nu)}$$

Now, without loss of generality take $H(\mu) \geq H(\nu)$ then;

$$H(\pi) = H(\mu) + H(\nu|\mu) \geq H(\mu)$$

Similarly, if $H(\nu) \geq H(\mu)$ then;

$$H(\pi) = H(\nu) + H(\mu|\nu) \geq H(\nu)$$

In either case, we get $\boxed{H(\pi) \geq \max(H(\mu), H(\nu))}$

Problem 5

Define two sets $X, Y \subset \mathbb{R}^2$ as

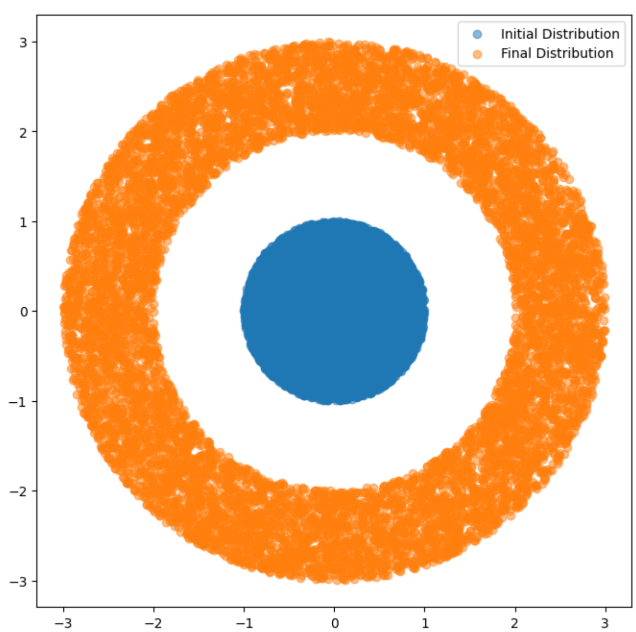
$$X = \{x \mid ||x|| \leq 1\}, \quad Y = \{y \mid 2 \leq ||y|| \leq 3\}$$

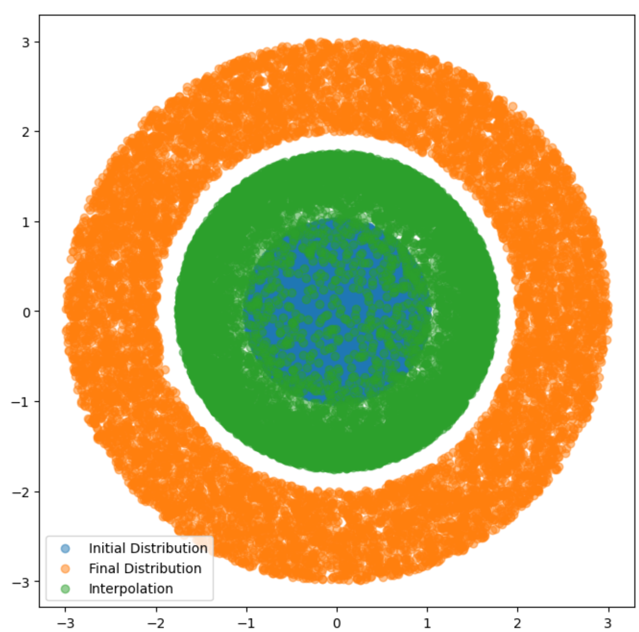
Let μ be the uniform probability distribution on X and ν be the uniform probability distribution on Y. Obtain the displacement interpolation between μ and ν with respect to W_2 (Wasserstein-2) metric.

Solution: Given X and Y as defined, X represents a unit disk centered at the origin and Y represents an annulus (ring-shaped region) with an inner radius of 2 units and an outer radius of 3. Let $t \in [0, 1]$ be the interpolation parameter. The displacement interpolation μ_t will be a distribution over a shape that smoothly transitions from X to Y as t progresses from 0 to 1.

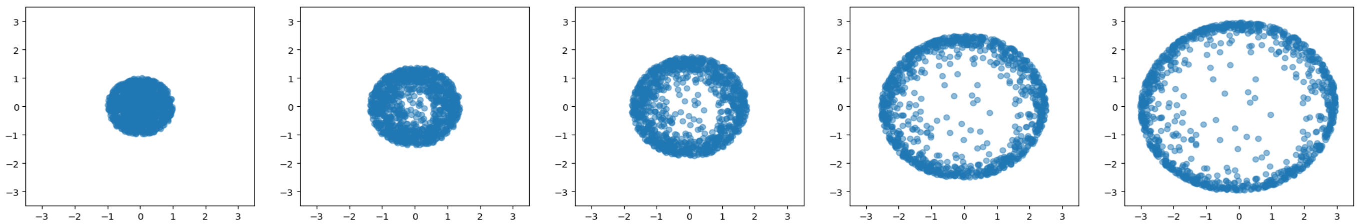
The displacement interpolation between two uniform probability distributions on non-convex sets like X and Y is not analytically tractable so to compute the Wasserstein-2 distance or interpolate the distribution numerically, we would need to use an optimal transport algorithm, such as Sinkhorn's iterative method for Wasserstein computation.

Below is a result of interpolation using Sinkhorn algorithm. Code is provided at the last page.





The figure on the left shows the initial and final distribution. And the figure on the right shows the displacement interpolation between these distributions.



Visualization of interpolation between the two distributions at different time period.

Problem 6

Let,

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^n} \Big(\frac{1}{2}||\nabla \rho||^2 + 4\sqrt{\rho}\Big)\ dx$$

be a functional over the space of probability densities. Derive the gradient flow of $\mathcal{F}$ with respect to W_2.

Solution: Given a very small variation $\delta\rho(x)$ in $\rho(x)$; the change in $\mathcal{F}$ due to this variation is

$$\begin{aligned}
\delta\mathcal{F} &= \mathcal{F}(\rho+\delta\rho) - \mathcal{F}(\rho) \\
&= \int_{\mathbb{R}^n} \Big(\frac{1}{2}||\nabla(\rho+\delta\rho)||^2 + 4\sqrt{\rho+\delta\rho}\Big)\ dx - \int_{\mathbb{R}^n} \Big(\frac{1}{2}||\nabla\rho||^2 + 4\sqrt{\rho}\Big)\ dx \\
&= \int_{\mathbb{R}^n} \frac{1}{2}||\nabla(\rho+\delta\rho)||^2\ dx + \int_{\mathbb{R}^n} 4\sqrt{\rho+\delta\rho}\ dx - \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\rho||^2\ dx - \int_{\mathbb{R}^n} 4\sqrt{\rho}\ dx \\
&\approx \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\rho + \nabla\delta\rho||^2\ dx + \int_{\mathbb{R}^n} \big(4\sqrt{\rho} + 2\rho^{-1/2}\delta\rho\big)\ dx - \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\rho||^2\ dx - \int_{\mathbb{R}^n} 4\sqrt{\rho}\ dx \\
&= \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\rho||^2\ dx + \int_{\mathbb{R}^n} \nabla\rho\cdot\nabla\delta\rho\ dx + \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\delta\rho||^2\ dx + \int_{\mathbb{R}^n} 4\sqrt{\rho}\ dx + \int_{\mathbb{R}^n} \frac{2}{\sqrt{\rho}}\delta\rho\ dx\ \cdots \\
&\cdots - \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\rho||^2\ dx - \int_{\mathbb{R}^n} 4\sqrt{\rho}\ dx \\
&= \int_{\mathbb{R}^n} (\nabla\rho\cdot\nabla\delta\rho)\ dx + \int_{\mathbb{R}^n} \frac{2}{\sqrt{\rho}}\delta\rho\ dx + \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\delta\rho||^2\ dx
\end{aligned}$$

Using integration by parts, we have $\int_{\mathbb{R}^n}(\nabla\rho\cdot\nabla\delta\rho)\ dx = -\int_{\mathbb{R}^n}\delta\rho\Delta\rho\ dx$. With this it follows that;

$$\delta\mathcal{F} = -\int_{\mathbb{R}^n} \delta\rho\Delta\rho\ dx + \int_{\mathbb{R}^n} \frac{2}{\sqrt{\rho}}\delta\rho\ dx + \int_{\mathbb{R}^n} \frac{1}{2}||\nabla\delta\rho||^2\ dx$$

Now, we need to compute the first variation of the functional $\mathcal{F}$ with respect to a small perturbation in probability density $\rho(x)$.

$$\begin{aligned}
\frac{\delta\mathcal{F}}{\delta\rho} &= -\frac{1}{\delta\rho}\int_{\mathbb{R}^n} \delta\rho\Delta\rho\ dx + \frac{1}{\delta\rho}\int_{\mathbb{R}^n} \frac{2}{\sqrt{\rho}}\delta\rho\ dx + \frac{1}{\delta\rho}\int_{\mathbb{R}^n} \frac{1}{2}||\nabla\delta\rho||^2\ dx \\
&= -\Delta\rho + \frac{2}{\sqrt{\rho}} + 0
\end{aligned}$$

Then, the gradient flow is given by;

$$\boxed{\frac{\partial\rho}{\partial t} = \nabla\cdot\Big(\rho\nabla\frac{\delta\mathcal{F}}{\delta\rho}\Big) = \nabla\cdot\Big(\rho\nabla\big(-\Delta\rho + \frac{2}{\sqrt{\rho}}\big)\Big)}$$

Code for Problem 5

```
import numpy as np
import matplotlib.pyplot as plt
import ot

# Generate points for the two distributions
def generate_points(n_points, shape):
    if shape == 'disk':
        angles = np.random.uniform(0, 2 * np.pi, n_points)
        radii = np.sqrt(np.random.uniform(0, 1, n_points))
        x = radii * np.cos(angles)
        y = radii * np.sin(angles)
    elif shape == 'annulus':
        angles = np.random.uniform(0, 2 * np.pi, n_points)
        radii = np.sqrt(np.random.uniform(4, 9, n_points))
        x = radii * np.cos(angles)
        y = radii * np.sin(angles)
    return np.column_stack((x, y))

n_points = 10000
X = generate_points(n_points, 'disk')
Y = generate_points(n_points, 'annulus')

# Compute cost matrix
M = ot.dist(X, Y)
M /= M.max()

# Compute Sinkhorn distance and coupling
reg = 0.03
P = ot.sinkhorn(np.ones(n_points) / n_points, np.ones(n_points) / n_points, M, reg)
W2 = np.sum(P * M)

print(f"Wasserstein-2 distance: {W2:.4f}")

# Interpolate between the distributions
t = 0.5
Z = np.dot(P, Y) / np.sum(P, axis=1, keepdims=True)
interp_points = (1 - t) * X + t * Z

# Plot the results
plt.figure(figsize=(6,6))
plt.scatter(X[:, 0], X[:, 1], label='Initial Distribution', alpha=0.5)
plt.scatter(Y[:, 0], Y[:, 1], label='Final Distribution', alpha=0.5)
plt.scatter(interp_points[:, 0], interp_points[:, 1], label='Interpolation', alpha=0.5)
plt.legend()
plt.show()
```