MA5232 Assignment 3

Nguyen Ngoc Khanh - A0275047B

May 20, 2025

Contents

1 QUESTION 1: GAUSSIAN MEASURES

Problem 1.1 (question 1: Gaussian measures)

Let $\Omega = \mathbb{R}^d$, cost function $c(x,y) = \|x-y\|^2$, and two Gaussian measures $\mu = \mathcal{N}(m_\mu, \Sigma_\mu)$ and $\nu = \mathcal{N}(m_\nu, \Sigma_\nu)$

(a) Prove that the map

$$T : x \mapsto m_\nu + A(x - m_\mu)$$

is optimal, that is, $T_\# \mu = \nu$ where $A = \Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}$

(b) Prove that

$$W_2^2(\mu,\nu) = \|m_\mu - m_\nu\|^2 + B(\Sigma_\mu, \Sigma_\nu)^2$$

where B is the Bures metric

$$B(\Sigma_\mu, \Sigma_\nu)^2 = \mathrm{tr}(\Sigma_\mu + \Sigma_\nu - 2(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2})$$

Lemma 1.2

If X is a random variable on $\mathbb{R}^d$, then for any linear map $A : \mathbb{R}^d \to \mathbb{R}^d$

$$\mathrm{Cov}(AX) = A\,\mathrm{Cov}(X)A^T$$

1.1 1.a

ν is the push-forward measure of μ under the map $T : \Omega \to \Omega$

Let $X \sim \mu$ and $Y = T(X)$, since T is affine, the distribution of Y is also Gaussian, we have

$$\mathbb{E}[Y] = m_\nu + A(\mathbb{E}[X] - m_\mu) = m_\nu$$

On the other hand, note that A is symmetric, hence

$$\begin{aligned}
\mathrm{Cov}(Y) &= A\,\mathrm{Cov}(X - m_\mu)A^T \\
&= A\Sigma_\mu A^T \\
&= A\Sigma_\mu A \\
&= \Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}\Sigma_\mu\Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2} \\
&= \Sigma_\nu
\end{aligned}$$

Hence, $T_\#\mu = \nu$

$T : \Omega \to \Omega$ is optimal

Let the convex function $\phi : \Omega \to \mathbb{R}$ be defined by

$$\phi(x) = m_\nu^T x + \frac{1}{2}(x - m_\mu)^T A(x - m_\mu)$$

We have that $(\nabla\phi)_\#\mu = \nu$ and $T = \nabla\phi$. By Brenier theorem, T is the unique optimal map.

1.2 1.b

Let $X \sim \mu$ and $Z = X - m_\mu$, since T is optimal, we have

$$
\begin{aligned}
W_2^2(\mu, \nu) &= \mathbb{E}[\|X - T(X)\|^2] \\
&= \mathbb{E}[\|X - m_\nu - A(X - m_\mu)\|^2] \\
&= \mathbb{E}[\|(m_\mu - m_\nu) + (I - A)Z\|^2] \\
&= \mathbb{E}[\|m_\mu - m_\nu\|^2 + \|(I - A)Z\|^2 + \langle m_\mu - m_\nu, (I - A)Z\rangle] \\
&= \|m_\mu - m_\nu\|^2 + \mathbb{E}[\|(I - A)Z\|^2] + \mathbb{E}[\langle m_\mu - m_\nu, (I - A)Z\rangle]
\end{aligned}
$$

Since Z is of Gaussian with mean zero, so is $(I - A)Z$, then

$$
\mathbb{E}[\langle m_\mu - m_\nu, (I - A)Z\rangle] = \langle m_\mu - m_\nu, \mathbb{E}[(I - A)Z]\rangle = 0
$$

Since $(I - A)Z$ is of mean zero and A is symmetric, then

$$
\begin{aligned}
\mathbb{E}[\|(I - A)Z\|^2] &= \operatorname{tr}\operatorname{Cov}((I - A)Z) \\
&= \operatorname{tr}((I - A)\Sigma_\mu(I - A)^T) \\
&= \operatorname{tr}((I - A)\Sigma_\mu(I - A)) \\
&= \operatorname{tr}((I - A)^2\Sigma_\mu) \\
&= \operatorname{tr}\Sigma_\mu + \operatorname{tr}(A^2\Sigma_\mu) - 2\operatorname{tr}(A\Sigma_\mu) \\
&= \operatorname{tr}\Sigma_\mu + \operatorname{tr}(A\Sigma_\mu A^T) - 2\operatorname{tr}(A\Sigma_\mu) \\
&= \operatorname{tr}\Sigma_\mu + \operatorname{tr}\Sigma_\nu - 2\operatorname{tr}(A\Sigma_\mu) && \text{(from previous part)}
\end{aligned}
$$

We have

$$
\begin{aligned}
\operatorname{tr}(A\Sigma_\mu) &= \operatorname{tr}(\Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}\Sigma_\mu) \\
&= \operatorname{tr}((\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}\Sigma_\mu\Sigma_\mu^{-1/2}) \\
&= \operatorname{tr}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}
\end{aligned}
$$

Hence,

$$
W_2^2(\mu, \nu) = \|m_\mu - m_\nu\|^2 + B(\Sigma_\mu, \Sigma_\nu)^2
$$

2 QUESTION 2: NON-UNIQUENESS

Problem 2.1 (question 2: non-uniqueness)
Find two distribution P and Q such that the optimal transport map between P and Q is not unique

In $\mathbb{R}^4$, let $A = (1, 0, 0, 0)$, $B = (0, 1, 0, 0)$, $C = (0, 0, 1, 0)$, $D = (0, 0, 0, 1)$, then $\|A - C\| = \|A - D\| = \|B - C\| =$

$\|B - D\| = \sqrt{2}$, let

$$P = \frac{1}{2}(\delta_A + \delta_B)$$
$$Q = \frac{1}{2}(\delta_C + \delta_D)$$

Then, under the cost $c(x, y) = \|x - y\|^2$, then there are two maps transporting P to Q as follows:

$$T_1 : A \mapsto C, B \mapsto D$$
$$T_2 : A \mapsto D, B \mapsto C$$

Both are optimal and of the same cost.

3 QUESTION 3: OPTIMAL OR NOT

Problem 3.1 (question 3: optimal or not)
Assume that $X, Y \in \mathbb{R}^2$ such that

$$X \sim \mathcal{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \text{ and } Y \sim \mathcal{N}_2\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}\right)$$

For each of the following maps, say whether it is (or not) an optimal transport between X and Y and why

(a) $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T_1(x) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \cdot x$$

(b) $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T_2(x) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \cdot x$$

(c) $T_3 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T_3(x) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

(d) $T_4 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T_4(x) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} \cdot x$$

3.1 NECESSARY CONDITION

A necessary condition for a map $T(x) = b + Ax$ being optimal is $T_{\#}\mu = \nu$. Since both X and Y are Gaussians, it sufficient to check that $Y = \mathbb{E}[T(X)]$ and $\mathrm{Cov}(Y) = \mathrm{Cov}(T(X))$, we have

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \mathbb{E}[Y] = \mathbb{E}[T(X)] = b + A\mathbb{E}[X] = b$$
$$\begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} = \mathrm{Cov}(Y) = \mathrm{Cov}(T(X)) = A\,\mathrm{Cov}(X)A^T = AA^T$$

We can easily see that T_1 is the only map satisfying the necessary condition.

3.2 SUFFICIENT CONDITION

Assuming the cost function is $c(x, y) = \|x - y\|^2$, the unique optimal transport map from $\mathcal{N}(m_\mu, \Sigma_\mu)$ into $\mathcal{N}(m_\nu, \Sigma_\nu)$ is $T : \mathbb{R}^2 \to \mathbb{R}^2$

$$T(x) = m_\nu + A(x - m_\mu) = (m_\nu - Am_\mu) + Ax$$

where $A = \Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}$. When μ and ν are as in the question, we have

$$A = I^{-1/2}(I^{1/2}\Sigma_\nu I^{1/2})^{1/2}I^{-1/2} = (\Sigma_\nu)^{1/2} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

Hence,

$$T(x) = (m_\nu - Am_\mu) + Ax = m_\nu + Ax = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} x = T_1(x)$$

Hence, T_1 is optimal with respect to the cost function $c(x, y) = \|x - y\|^2$

4 QUESTION 4: BASICS OF THE POT LIBRARY, CODING

See next page

code

April 12, 2025

1 MA5232 ASSIGNMENT 3

Nguyen Ngoc Khanh - A0275047B

[1]:
```python
import math # import a package
import numpy as np # import a package under an alias
from sklearn import linear_model # import a submodule
from os import mkdir # import a specific function
```

[2]:
```python
from sklearn.datasets import make_blobs
n_samples = 50
n_features = 2
offset = 0.5
seed = 42
centers = np.array ([
    [0 , 0],
    [offset, 0],
])
Z, c = make_blobs(n_samples=n_samples, centers=centers, n_features=n_features,␣
↪random_state=seed, cluster_std=0.05, shuffle=False)
X = Z[c == 0] # source distribution
Y = Z[c == 1] # target distribution
```

[3]:
```python
def plot2D_samples_mat(xs, xt, G, ax, thr=1e-8, **kwargs):
    if ("color" not in kwargs) and ("c" not in kwargs):
        kwargs["color"] = "k"
    mx = G.max()
    if "alpha" in kwargs:
        scale = kwargs["alpha"]
        del kwargs["alpha"]
    else:
        scale = 1

    for i in range(xs.shape[0]):
        for j in range(xt.shape[0]):
            if G[i, j] / mx > thr:
                ax.plot([xs[i, 0], xt[j, 0]],
                        [xs[i, 1], xt[j, 1]],
```

```
                    alpha=G[i, j] / mx*scale, **kwargs,
            )
```

2 ANSWER START FROM HERE

[4]:
```python
# import POT, plt
import ot
from matplotlib import pyplot as plt
# eps
eps = 1e-6
```

[5]:
```python
# get data shape - move n points into m points
n, d = X.shape
m, d = Y.shape
m, n, d
```

[5]: (25, 25, 2)

2.1 (a) Compute the two cost matrices between samples:

- $d_1 = (||x_i - y_i||_2)_{ij}$
- $d_2 = (||x_i - y_i||_2^2)_{ij}$

[6]:
```python
d1 = ot.dist(X, Y, metric="euclidean")
d2 = ot.dist(X, Y, metric="sqeuclidean")

assert tuple(d1.shape) == tuple(d1.shape) == (n, m)
assert np.max(np.abs(d1**2 - d2)) < eps
```

2.2 (b) Computer the weights of each point in the distribution (use uniform weights)

$$w_x, w_y$$

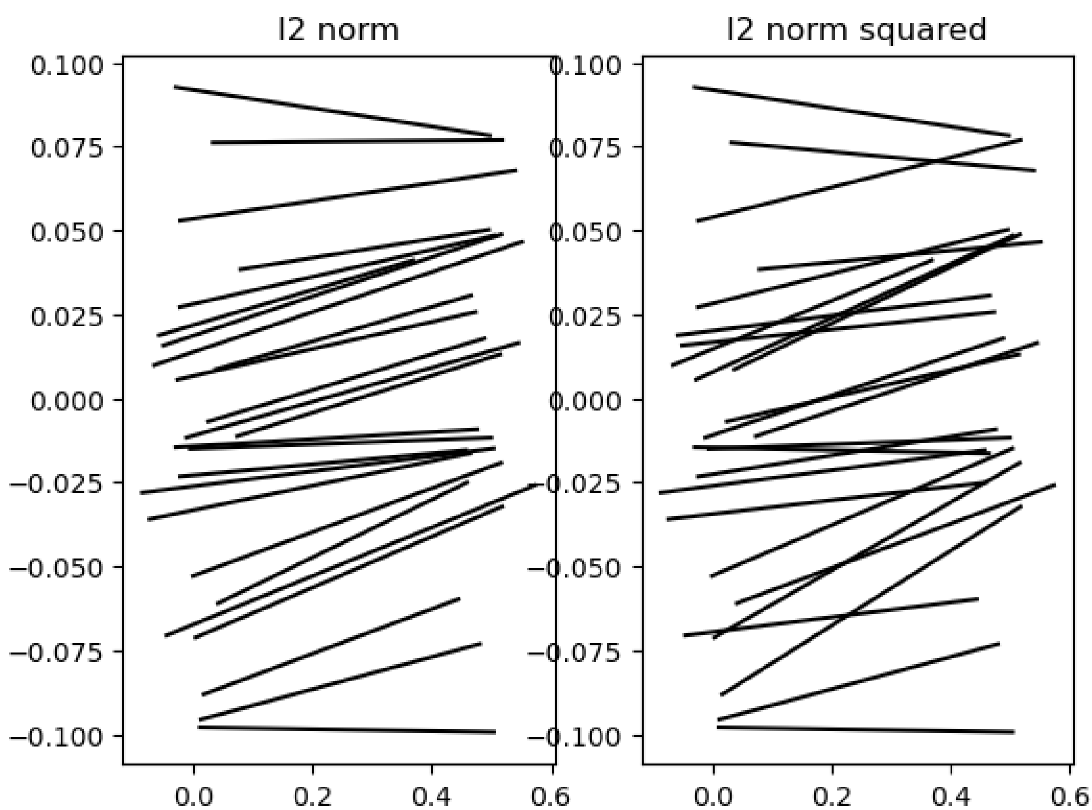
[7]:
```python
wx = np.ones(shape=(n,)) / n
wy = np.ones(shape=(m,)) / m

assert np.sum(wx) - 1 < eps
assert np.sum(wy) - 1 < eps
```

2.3 (c) For each cost matrix, compute the optimal transport plan between the source and target distributions

[8]:
```python
T1 = ot.emd(wx, wy, d1)
T2 = ot.emd(wx, wy, d2)
```

2.4 (d) Plot the two optimal transport plan along with the source and target samples

[9]:
```
fig, axes = plt.subplots(ncols=2)
ax1, ax2 = axes
ax1.set_title("l2 norm")
ax2.set_title("l2 norm squared")
plot2D_samples_mat(X, Y, T1, ax=ax1)
plot2D_samples_mat(X, Y, T2, ax=ax2)
plt.show()
```



[10]:
```
float(np.max(d1))
```

[10]: 0.660144262623403

2.5 (e) What is the major difference between the two optimal transport plans?

There are many *crossings* in l_2-norm squared distance.

Informally speaking, transporting between two uniform distributions supported on finite sets X, Y of the same size $|X| = n = m = |Y|$ consists of *basic transport*, that is, how much Earth to move from $x \in X$ into $y \in Y$. When plotting, each *basic transport* corresponds to a line segment connecting X to Y.

Since the maximum distance between any two points $x \in X$ and $y \in Y$ is less than 1, informally speaking, l_2-norm squared distance put less penalty, so it enables the transport to travel further with respect to Euclidean distance to achieve optimal transport. Hence, there are more *crossings*

2.6 (f) What are the Wasserstein distances in these two cases?

[11]:
```
W1 = ot.emd2(wx, wy, d1)
W2 = ot.emd2(wx, wy, d2)

assert W1 - np.sum(T1 * d1) < eps
assert W2 - np.sum(T2 * d2) < eps

W1, W2
```

[11]: (0.5049892452380658, 0.2557974949833624)