

Assignment due for February 27th

You should answer the theoretical questions on paper. For the numerical part, you should write a Jupyter notebook that you will deposit on the dropbox folder whose link has been communicated by email.

Your notebook should include the output of the code cells and the file should be named as
YourFamilyName_dm2324.ipynb

Consider two probability measures μ, ν on \mathbf{R}^d with supports \mathcal{X}, \mathcal{Y} respectively.

We assume that \mathcal{X} is compact and that \mathcal{Y} is finite, $\mathcal{Y} = \{y_1, \dots, y_J\}$.

Let $q = (\nu(\{y_j\}))_{1 \leq j \leq J} \in \mathbf{R}^J$.

Let $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbf{R}$ be continuous. For any $y \in \mathcal{Y}$, let $c_y = c(\cdot, y)$ be the function $x \mapsto c(x, y)$.

We assume that the groundcost c satisfies the assumption

$$\forall y, z \in \mathcal{Y}, \quad \forall t \in \mathbf{R}, \quad \mu((c_y - c_z)^{-1}(\{t\})) = 0. \quad (\text{GC})$$

For any $v \in \mathbf{R}^J$, we define the c -transform $v^c : \mathcal{X} \rightarrow \mathbf{R}$ by $v^c(x) = \min_j (c(x, y_j) - v_j)$, and we set

$$H(v) = \int_{\mathcal{X}} v^c(x) d\mu(x) + \sum_{j=1}^J v_j q_j.$$

For $v \in \mathbf{R}^J$, we also define the open Laguerre cells

$$L_j(v) = \{x \in \mathcal{X} \mid \forall k \neq j, c(x, y_j) - v_j < c(x, y_k) - v_k\}.$$

PART A: SEMI-DUAL OPTIMAL TRANSPORT

1. Explain (briefly) the relation between H and the optimal transport cost $W(\mu, \nu)$.
2. Let $f_i : \mathbf{R}^J \rightarrow \mathbf{R}, i \in I$ be a collection of convex functions.
 - a. Is $\sup_{i \in I} f_i$ convex? concave? (Justify with a proof, or give a counter-example).
 - b. Deduce that H is a concave function.
3. For any $v \in \mathbf{R}^J$, let $m(v) = (\mu(L_j(v)))_{1 \leq j \leq J} \in \mathbf{R}^J$. We denote by $\langle \cdot, \cdot \rangle$ the Euclidean product. The goal of this question is to show that H is \mathcal{C}^1 . We will admit that, since H is concave, it is sufficient to show that there exists a continuous selection of supergradients.
 - a. Prove that $\forall u, v \in \mathbf{R}^J, \quad H(u) \leq H(v) + \langle q - m(v), u - v \rangle$.
 - b. Show that $m : \mathbf{R}^J \rightarrow \mathbf{R}$ is a continuous function.
 - c. Assume that $v \in \mathbf{R}^J$ is such that $m(v) = q$. What can be said about $H(v)$?
4. Here, we use the Averaged Stochastic Gradient Descent (ASGD) algorithm to minimize $-H$.
 - a. Write the pseudo-code of the ASGD algorithm. Explain the required gradient computation.
 - b. Is the convergence of ASGD guaranteed in this case, in terms of the values of H ? (Justify briefly, with a reference.)

PART B: SEMI-DISCRETE WGAN

In this numerical part, we consider the quadratic cost $c(x, y) = \|x - y\|_2^2$ on \mathbf{R}^d .

5. Here $d = 2$. Let $\Theta := (\mathbf{R}^2)^2$. For $\theta = (a, b) \in \Theta$, we define $g_\theta : [0, 1] \rightarrow \mathbf{R}^2$ by

$$\forall z \in [0, 1], \quad g_\theta(z) = za + (1 - z)b.$$

We denote by μ_θ the distribution of $g_\theta(Z)$ where Z is uniform in $[0, 1]$.

Consider also the $\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$ with $y_1 = (-1, 0)$ and $y_2 = (1, 0)$.

The above function H related to (μ_θ, ν) will be denoted by H_θ .

The goal of this question is to numerically solve $\min_{\theta \in \Theta} W(\mu_\theta, \nu)$.

- a. Can you guess a solution θ of this problem?
 - b. For a fixed $\theta \in \Theta$, implement in Python the ASGD algorithm that approaches $v \in \text{Argmax } H_\theta$.
 - c. Compare your ASGD algorithm with the PyTorch implementation of ASGD.
 - d. In Python, implement an algorithm that approaches the minimum of $\theta \rightarrow W(\mu_\theta, \nu)$.
 - e. Illustrate the behavior of the algorithm by displaying the configurations obtained along the algorithm (by drawing samples of μ_θ). Is the behavior stable if you vary the optimization parameters?
 - f. Is there some configuration of μ_θ for which the optimization of v may be singular?
6. The goal of this last question is to learn a generative model μ_θ for the MNIST dataset.
- You may download the MNIST dataset as in Practical Session 3. If you need, you may work on a subsampled version of MNIST (e.g. with 6000 points, 600 for each class).
- a. Adapt the algorithm of the previous question in order to learn a generative model for MNIST.
 - b. For a fixed configuration of μ_θ , illustrate the behavior of ASGD to estimate $W(\mu_\theta, \nu)$.
 - c. Can you propose a better parameterization of the dual variable v for which the learning would be faster?