ECE 6970 - Homework Assignment 2

October 7th 2019

Due to: Tuesday, October 22nd, 2019 (at the beginning of the lecture)

Instructions: Submission in pairs is allowed. Prove and explain every step in your answers.

1) **Distance cost in 1D:** We saw in class that if $\mu, \nu \in \mathcal{P}(\mathbb{R})$ are probability measures on the real line with cumulative distribution functions (CDFs) F and G, respectively, then

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} |x-y| \mathsf{d}\pi(x,y) = \int_0^1 \big|F^{-1}(t) - G^{-1}(t)\big| \mathsf{d}t = \int_{\mathbb{R}} \big|F(x) - G(x)\big| \mathsf{d}x. \tag{1}$$

Optimality above is achieved by the coupling π_H whose CDF is $H(x,y) = \min\big\{F(x), G(y)\big\}$. This is a guided exercise to prove the <u>second</u> equality in (1) and some related properties.

a) Show that π_H is indeed a coupling of μ and ν. It suffices to show that H has F and G as marginal CDFs.

b) Consider the set $\mathcal{A} \triangleq \Big\{(x,t) \;\Big|\; \min\big\{F(x), G(x)\big\} \le t \le \max\big\{F(x), G(x)\big\},\ x \in \mathbb{R}\Big\}$. Use Fubini's Theorem to show that

$$\int_{\mathbb{R}} \int_{\min\{F(x),G(y)\}}^{\max\{F(x),G(y)\}} \mathsf{d}t\ \mathsf{d}x = \int_0^1 \int_{\min\{F^{-1}(x),G^{-1}(y)\}}^{\max\{F^{-1}(x),G^{-1}(y)\}} \mathsf{d}x\ \mathsf{d}t.$$

c) Prove that $\max\{a,b\} - \min\{a,b\} = |a-b|$ and use this relation to conclude that

$$\int_0^1 \big|F^{-1}(t) - G^{-1}(t)\big| \mathsf{d}t = \int_{\mathbb{R}} \big|F(x) - G(x)\big| \mathsf{d}x.$$

Comment: F^{-1} and G^{-1} are the generalized inverses of F and G, respectively. It is preferable to treat them as such in your proofs. However, answers under the assumption that F and G are invertible will be accepted.

2) **Transport Maps in 1D:** Proceeding under the framework of Question 1 (although a distance costs is not necessary here), recall that if μ (whose CDF is F) has a density, then (1) is achieved by the transport plan $T^\star = G^{-1} \circ F$. Namely, we have

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} |x-y| \mathsf{d}\pi(x,y) = \int_{\mathbb{R}} |x - T^\star(x)| \mathsf{d}\mu(x). \tag{2}$$

This is a guided exercise to prove (2).

a) First, we will establish that $T^\star$ indeed pushes μ forward to ν. Prove that if there exists $\eta \in \mathcal{P}(\mathbb{R})$ such that $G^{-1}_{\#}\eta = \nu$ and $F_{\#}\mu = \eta$, then $T^\star_{\#}\mu = \nu$.

Hint: During our first encounter with transport maps in class we showed that $(S \circ T)_{\#}\mu = S_{\#}(T_{\#}\mu)$.

b) Show that $\eta = \mathsf{Unif}\big([0,1]\big)$ satisfies the above relations, and conclude $T^\star$ is a valid transport map.

c) Show that

$$\int_0^1 \left|F^{-1}(t) - G^{-1}(t)\right| \mathsf{d}t = \int_{\mathbb{R}} \Big|x - G^{-1}\big(F(x)\big)\Big| \mathsf{d}\mu(x).$$

Hint: Use Section (b), which established $F_{\#}\mu = \mathsf{Unif}\big([0,1]\big)$, and the change of variables formula from the lectures: If $\mu \in \mathcal{P}(\mathcal{X})$, $T : \mathcal{X} \to \mathcal{Y}$ and $f \in L^1(\mathcal{Y})$, then $\int_{\mathcal{Y}} f(y)\mathsf{d}(T_{\#}\mu)(y) = \int_{\mathcal{X}} f\big(T(x)\big)\mathsf{d}\mu(x)$.

3) **Kantorovich Duality Regularizer:** In class, we proved the Kantorovich duality for optimal transport:

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathsf{d}\pi(x,y) = \sup_{(\varphi,\psi)\in\Phi_c} \int_{\mathcal{X}} \varphi(x)\mathsf{d}\mu(x) + \int_{\mathcal{Y}} \psi(y)\mathsf{d}\nu(y), \tag{3}$$

where Φ_c is the set of all pairs of potentials (measurable functions) (φ, ψ) satisfying $\varphi(x) + \psi(y) \le c(x, y)$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. The first step in our proof was to introduce a regularizer $\Gamma(\pi)$ to the left-hand side (LHS) of (3), and extend the infimum to $\mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$, the set of all nonnegative Borel measures on $\mathcal{X} \times \mathcal{Y}$. Our regularizer was

$$\Gamma(\pi) = \sup_{(\varphi,\psi)\in\mathcal{C}_b^0(\mathcal{X})\times\mathcal{C}_b^0(\mathcal{Y})} \int_{\mathcal{X}} \varphi(x)\mathsf{d}(\mu - \pi_{\mathcal{X}})(x) + \int_{\mathcal{Y}} \psi(y)\mathsf{d}(\nu - \pi_{\mathcal{Y}})(y),$$

where $\pi_{\mathcal{X}}$ and $\pi_{\mathcal{Y}}$ are the $\mathcal{X}$- and $\mathcal{Y}$-marginals of $\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$.

a) Show that

$$\Gamma(\pi) = \begin{cases} 0, & \pi \in \Pi(\mu,\nu) \\ +\infty, & \text{otherwise} \end{cases}.$$

b) Denoting $\mathsf{K}(\pi) \triangleq \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathsf{d}\pi(x,y)$, use the previous section to justify that

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathsf{K}(\pi) = \inf_{\pi\in\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \mathsf{K}(\pi) + \Gamma(\pi).$$

4) **Gluing Lemma:** The Gluing Lemma is used for proving the triangle inequality of Wasserstein Distances. It reads as follows: Let $\mu_j \in \mathcal{P}(\mathcal{X}_j)$, for $j = 1, 2, 3$, be three probability measures on their corresponding spaces. Let $\pi_{12} \in \Pi(\mu_1, \mu_2)$ and $\pi_{23} \in \Pi(\mu_2, \mu_3)$. Then there exists $\pi \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ such that $\pi_{\mathcal{X}_1,\mathcal{X}_2} = \pi_{12}$ and $\pi_{\mathcal{X}_2,\mathcal{X}_3} = \pi_{23}$, where $\pi_{\mathcal{X},\mathcal{Y}}$ is the marginal of π on $\mathcal{X} \times \mathcal{Y}$.

Prove the Gluing Lemma for densities. Namely, assume $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_2 = \mathbb{R}^d$ and let f_j, $j = 1, 2, 3$, be the probability density functions (PDFs) of μ_j. Let $g_{12}(x, y)$ be a PDF on $\mathbb{R}^d \times \mathbb{R}^d$ that has $f_1(x)$ and $f_2(y)$ as its marginals. Similarly, $g_{23}(y, z)$ is a PDF with marginals $f_2(y)$ and $f_3(z)$. Construct a PDF $g(x, y, z)$ on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ such that $\int_{\mathbb{R}^d} g(x, y, z)\mathsf{d}z = g_{12}(x, y)$ and $\int_{\mathbb{R}^d} g(x, y, z)\mathsf{d}x = g_{23}(y, z)$.

5) **Triangle Inequality for Wasserstein Distances:** The p-Wasserstein distance, for $p \in [1, +\infty)$, between $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$, where $\mathcal{P}_p(\mathcal{X})$ is the set of probability measures over $\mathcal{X}$ with $\int_{\mathcal{X}} \|x\|^p \mathsf{d}\mu(x) < \infty$, is:

$$\mathsf{W}_p(\mu,\nu) \triangleq \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x - y\|^p \mathsf{d}\pi(x,y) \right)^{1/p}. \tag{4}$$

We saw in class that $\Big(\mathcal{P}_p(\mathcal{X}), \mathsf{W}_p\Big)$ is a metric space. In particular, W_p satisfies the triangle inequality: For any $\mu, \nu, \eta \in \mathcal{P}_p(\mathcal{X})$, we have

$$\mathsf{W}_p(\mu,\eta) \le \mathsf{W}_p(\mu,\nu) + \mathsf{W}_p(\nu,\eta)$$

Use the Gluing Lemma to prove the triangle inequality for W_p.

Hint: Cast π_{12} and π_{23} from the Gluing Lemma as optimal coupling for (μ, ν) and (ν, η), respectively.