SLMath Exercise Sheet 1

Jonathan Niles-Weed

This is a long list of exercises, of very different flavors. I encourage you to attempt all or most of them, but only complete those that speak to you.

0. Install the POT (Python Optimal Transport) package (`https://pythonot.github.io/`). Follow the example code in the documentation (or use an LLM) to write simulations to empirically explore the following phenomena:

   (a) $\mathbb{E}W_1(\mu_n, \mu) \leq \frac{D}{2}\sqrt{\frac{k}{n}}$ if $\mu \in \mathcal{P}(\mathcal{X})$ where $|\mathcal{X}| \leq k$, $\operatorname{diam}(\mathcal{X}) \leq D$. Explore finite metrics beyond those given in lecture where this bound is very loose or very tight.

   (b) $\mathbb{E}W_1(\mu_n, \mu) \lesssim n^{-1/d}$ if $\mu \in \mathcal{P}([0,1]^d)$ for $d \geq 3$. What happens if μ is concentrated on a k-dimensional subspace? How about a k-dimensional manifold?

   (c) The empirical fluctuations of $W_1(\mu_n, \mu)$ are much smaller than its mean value, when d is large.

   (d) $\mathbb{E}W_1(\mu_n, \mu) \asymp n^{-1/2}$ if $\mu \in \mathcal{P}([0,1])$.

   (e) Let μ_n be an empirical measure corresponding to samples from $\mathcal{N}(0, I)$, and let ν_n be an empirical measure corresponding to samples from $\mathcal{N}(0, cI)$ for some $c > 1$. How large must c be for $W_1(\mu_n, \nu_n)$ to be consistently larger than $W_1(\mu_n, \mu_n')$, where μ_n' is an independent copy of μ_n.

1. This exercise establishes that $W_p(\mu_n, \mu) \overset{a.s.}{\to} 0$ under general conditions.

   (a) Let $(\mathcal{X}, d)$ be a compact metric space of diameter D. Show that for any $\mu, \nu \in \mathcal{P}(\mathcal{X})$, the inequality

   $$W_p^p(\mu, \nu) \leq D^{p-1} W_1(\mu, \nu) \tag{1}$$

   holds.

   (b) Fix a point $x_0 \in \mathcal{X}$. Using the Arzelà–Ascoli theorem, show that the set $\mathcal{L}$ of 1-Lipschitz functions on $\mathcal{X}$ which vanish at x_0 is compact in the uniform topology. Conclude that, for any $\varepsilon > 0$, there exists a finite set of Lipschitz functions $\{f_1, \ldots, f_M\}$ such that, for any $f \in \mathcal{L}$,

   $$\min_{m \in [M]} \|f - f_m\|_\infty \leq \varepsilon\,. \tag{2}$$

   (c) Using the previous part, show that $\sup_{f \in \mathcal{L}} \int f(\mathrm{d}\mu_n - \mathrm{d}\mu) \to 0$ almost surely. Conclude that $W_1(\mu_n, \mu) \overset{a.s.}{\to} 0$ and, by part (a), that $W_p(\mu_n, \mu) \overset{a.s.}{\to} 0$ for all $p \geq 1$ as well.

   (d) Extend this argument to non-compact $\mathcal{X}$, under the assumption that μ has finite pth moment. (Hint: for any $\varepsilon > 0$ and $x_0 \in \mathcal{X}$ there exists a compact K such that

   $$\limsup_{n \to \infty} \int_{\mathcal{X} \setminus K} d(x_0, x)^p \,\mathrm{d}\mu_n(x) \leq \varepsilon \tag{3}$$

   almost surely.)

2. The goal of this exercise is to show that the basic estimate $\mathbb{E}W_p(\mu_n, \mu) \lesssim D(k/n)^{1/2p}$ for metric spaces $\mathcal{X}$ with cardinality k and diameter D is not improvable in general.

(a) Suppose that $\mathcal{X}$ is equipped with the discrete metric $d(x,y) = D\mathbb{1}_{x\neq y}$. Show that $W_p^p(\mu_n,\nu) = \frac{D^p}{2}\|\mu_n - \mu\|_1$, and conclude that if μ is the uniform measure, then $\mathbb{E}W_p(\mu_n,\mu) \gtrsim D(k/n)^{1/2p}$ as long as $k \leq n$.

(b) Now, let $\mathcal{X}$ be *any* metric space with at least 2 points. Show that $\mathbb{E}W_p(\mu_n,\mu)$ can decrease no faster than $n^{-1/2p}$ in general. (Hint: consider $\mu = \frac{1}{2}\delta_x + \frac{1}{2}\delta_y$ for two distinct $x, y \in \mathcal{X}$.)

3. This exercise generalizes the example of P_N considered in the lecture. Say that a finite metric space $(\mathcal{X}, d)$ is a *tree metric* if there exists a positive-weighted tree $T = (V, E)$ (i.e., a connected, cycle-free graph along with a function $w : E \to \mathbb{R}_{>0}$) and a injection $\phi : \mathcal{X} \to V$ such that

$$d(x,y) = d_{T,w}(\phi(x), \phi(y))\,, \tag{4}$$

where $d_{T,w}$ denotes the graph metric on T, i.e., for any $u, v \in V$, $d_{T,w}(u,v)$ is defined to be the sum of the edge weights on the unique path in T between u and v. For notational convenience, we assume that $\mathcal{X} = V$ and that ϕ is the identity map.

(a) Check your understanding of the definition by verifying that the metric space P_N considered in the lecture is a tree metric.

(b) We will give two proofs of the following identity: for any $\mu, \nu \in \mathcal{P}(\mathcal{X})$,

$$W_1(\mu,\nu) = \sum_{e\in E} w(e)|\mu(T_e) - \nu(T_e)|\,, \tag{5}$$

where for each edge $e \in E$, we denote by T_e either of the two connected components that arise when the edge e is removed from T, and where

$$\mu(T_e) := \sum_{v\in V(T_e)} \mu(\{v\})\,, \tag{6}$$

and analogously for $\nu(T_e)$.

i. Check that this agrees with the formula for W_1 on P_N given in the lecture.

ii. Primal proof of upper bound: build a coupling by induction, removing one leaf of T at a time.

iii. Dual proof: let f be an arbitrary function on $\mathcal{X}$. Orient the edges of T, and for each $e = (u, v)$, define $\Delta_e = f(u) - f(v)$. Show that

$$\langle f, \mu - \nu\rangle = \sum_{e\in E} \Delta_e(\mu(T_e) - \nu(T_e))\,, \tag{7}$$

for an appropriate choice of T_e. Conclude by showing that it is possible to construct a Lipschitz f such that $\Delta_e = w(e)\,\mathrm{sign}(\mu(T_e) - \nu(T_e))$.

(c) Use part (b) to show that for any $\mu \in \mathcal{P}(X)$, we have the bound $\mathbb{E}W_1(\mu_n,\mu) \leq \Phi \cdot n^{-1/2}$, where $\Phi := \sum_{e\in E} w(e)\sqrt{\mu(T_e)(1-\mu(T_e))}$, and that this bound is essentially tight for large n.

4. Via duality and integration by parts, show that for any $\mu, \nu \in \mathcal{P}([0,1])$,

$$W_1(\mu,\nu) = \int_0^1 |(\mu-\nu)([0,t])|\,\mathrm{d}t\,. \tag{8}$$

Use this fact to show that for any $\mu \in \mathcal{P}([0,1])$,

$$\mathbb{E}W_1(\mu,\mu_n) \leq n^{-1/2}\,. \tag{9}$$

5. (*Statistical Optimal Transport*, Section 2.4) This exercise will outline a proof of the "AKT" (Ajtai, Komós, Tusnády) theorem, due to Bobkov and Ledoux: for any $\mu \in \mathcal{P}([0,1]^2)$,

$$\mathbb{E}W_1(\mu_n, \mu) \lesssim \sqrt{\frac{\log n}{n}}\,. \tag{10}$$

Like the other bounds we have seen, it is based on a deterministic estimate. Given a probability measure μ, we denote by ϕ_μ its characteristic function $\phi_\mu(m) := \int e^{\mathbf{i}\langle m,x\rangle}\mu(\mathrm{d}x)$. We will first show for any $\mu, \nu \in \mathcal{P}([0,1]^2)$,

$$W_1(\mu,\nu)^2 \leq \sum_{m\in\mathbb{Z}^2 : m\neq 0} \|m\|^{-2}\|\phi_\mu(m) - \phi_\nu(m)|^2\,. \tag{11}$$

(a) Denote by $\widetilde{\mathrm{Lip}}$ the set of 2π-periodic, 1-Lipschitz, smooth functions on $\mathbb{R}^2$. Show that if $\mu,\nu \in \mathcal{P}([0,1]^2)$ then $W_1(\mu,\nu) = \tilde{W}_1(\mu,\nu)$, where $\widetilde{W_1}(\mu,\nu) := \sup_{f\in\widetilde{\mathrm{Lip}}} \int f(\mathrm{d}\mu - \mathrm{d}\nu)$.

(b) Convince yourself that if f is smooth and 2π-periodic, then $\int f\,\mathrm{d}\mu = \sum_{m\in\mathbb{Z}^2} \hat{f}_m \phi_\mu(m)$, where $\hat{f}_m := \frac{1}{(2\pi)^2}\int_{[0,2\pi]^2} e^{-\mathbf{i}\langle m,x\rangle} f(x)\,\mathrm{d}x$.

(c) Show that if f is 1-Lipschitz, then $\sum_{m\in\mathbb{Z}^2} \|m\|^2 |\hat{f}_m|^2 \leq 1$.

(d) Prove (11).

(e) Let $\nu \star \gamma_\varepsilon$ denote the convolution of ν with the Gaussian $\mathcal{N}(0, \varepsilon I)$. Show that $\widetilde{W_1}(\nu, \nu\star\gamma_\varepsilon) \leq \sqrt{2\varepsilon}$ and that $\phi_{\nu\star\gamma_\varepsilon}(m) = \phi_\nu(m)e^{-\varepsilon\|m\|^2/2}$.

(f) Conclude that $W_1(\mu_n,\mu) \leq \widetilde{W_1}(\mu_n\star\gamma_\varepsilon, \mu\star\gamma_\varepsilon) + 2\sqrt{2\varepsilon}$ for any $\varepsilon > 0$, and use (11) with a good choice of ε to conclude that $\mathbb{E}W_1(\mu_n,\mu) \lesssim \sqrt{\log n/n}$.

6. (*Statistical Optimal Transport*, Exercise 2.1) This exercise proves that the rate proved in lecture for $d \geq 3$ is tight.

(a) Show that there exists a positive universal constant c such that for any $n \geq 1$, the Lebesgue measure of a ball in $\mathbb{R}^d$ with radius $c\sqrt{d}\,n^{-1/d}$ is at most $(2n)^{-1}$. (Hint: recall that the unit ball in $\mathbb{R}^d$ has Lebesgue measure $\pi^{d/2}/\Gamma(\frac{d}{2}+1)$.)

(b) Let μ be the uniform measure on $[0,1]^d$, and let $\tilde{\mu}_n$ be any measure supported on n points $x_1,\ldots,x_n$. If we denote by $B(x_i,\epsilon)$ a ball of radius ϵ around x_i, show that $\mu(\bigcup_{i=1}^n B(x_i,\epsilon)) \leq \frac{1}{2}$ if $\epsilon = c\sqrt{d}\,n^{-1/d}$, where c is the constant from part (a).

(c) Conclude that if $\gamma \in \Gamma(\mu,\tilde{\mu}_n)$, then $\int \|x-y\|\,\mathrm{d}\gamma(x,y) \geq \frac{1}{2}\cdot c\sqrt{d}\,n^{-1/d}$.

7. Show by example that the dyadic partitioning bound from the lecture can be arbitrarily inaccurate: for any $\varepsilon > 0$, there $\mu,\nu \in \mathcal{P}([0,1]^d)$ such that $W_1(\mu,\nu) \leq \varepsilon$ but for all $J \geq 0$, the dyadic partitioning upper bound satisfies

$$\sqrt{d}\sum_{j=0}^{J-1} 2^{-j} \sum_{Q\in\mathcal{Q}_{j+1}} |\mu(Q) - \nu(Q)| + \sqrt{d}2^{-J} \geq \sqrt{d}\,. \tag{12}$$

8. Mimic the proof of the dyadic partitioning bound to show that for any $\mu,\nu \in \mathcal{P}[0,1]^d$,

$$W_p^p(\mu,\nu) \leq d^{p/2}\sum_{j=0}^{J-1} 2^{-jp} \sum_{Q\in\mathcal{Q}_{j+1}} |\mu(Q) - \nu(Q)| + d^{p/2}2^{-Jp}\,. \tag{13}$$

Conclude that $\mathbb{E}W_p(\mu_n,\mu) \lesssim \sqrt{d}n^{-1/d}$ as long as $d > 2p$.