SLMath Exercise Sheet 2

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.

1. We proved the bound $\mathbb{E}\sup_{\nu\in\mathcal{P}(B_1)}|S_\varepsilon(\mu_n,\nu)-S_\varepsilon(\mu,\nu)| \lesssim_{s,d} \varepsilon^{-s+1}n^{-1/2}$ for any $s > d/2$ when $\mu \in \mathcal{P}(B_1)$ and $\varepsilon < 1$. Show that when $\varepsilon > 1$,
$$\mathbb{E}\sup_{\nu\in\mathcal{P}(B_1)}|S_\varepsilon(\mu_n,\nu)-S_\varepsilon(\mu,\nu)| \lesssim_{s,d} n^{-1/2}\,, \tag{1}$$
with no dependence on ε. (Hint: it suffices to show that $\|D^\alpha\varphi\|_\infty \lesssim_\alpha 1$ when $\varepsilon > 1$.)

2. Prove the Faà di Bruno formula
$$D^\alpha \log j(x) = \sum_{v=1}^{|\alpha|}\frac{(-1)^{v-1}}{v}\sum_{\beta_1+\cdots+\beta_v=\alpha}\prod_{\ell=1}^{v}\frac{D^{\beta_\ell}j(x)}{j(x)}\,. \tag{2}$$
in two ways:
   (a) By induction on α
   (b) By examination of the power series $\log(j(x)+y) = \log j(x) + \sum_{v=1}^\infty \frac{(-1)^{v-1}y^v}{vj(x)^v}$ and $j(x+z) = \sum_{\beta\geq 0}\frac{D^\beta j(x) z^\beta}{\beta!}$.

3. In the lecture, we claimed that the bound $\int(1+\|\xi\|^2)^s|\bar\Phi(\xi)|^2\,\mathrm{d}\xi \lesssim_s \varepsilon^{-2s+2}$ for all $s \geq 1$ follows from the bound where s is a positive integer by "interpolation." This exercise justifies this claim.
   (a) Let $t\in(0,1)$ and let $p,q\in(1,\infty)$. Via Hölder's inequality, show the L_p interpolation inequality
   $$\|h\|_{L_r} \leq \|h\|_{L_p}^t\|h\|_{L_q}^{1-t}\,,$$
   where $\frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}$.
   (b) Arguing as in part (a), show that for any integer k and $\delta\in(0,1)$,
   $$\int(1+\|\xi\|^2)^{k+\delta}|\bar\Phi(\xi)|^2\,\mathrm{d}\xi \leq \left(\int(1+\|\xi\|^2)^{k}|\bar\Phi(\xi)|^2\,\mathrm{d}\xi\right)^{1-\delta}\left(\int(1+\|\xi\|^2)^{k+1}|\bar\Phi(\xi)|^2\,\mathrm{d}\xi\right)^{\delta}\,. \tag{3}$$
   Conclude that the bound claimed in lecture holds.

4. This exercise establishes some important facts about Reproducing Kernel Hilbert Spaces (RKHS). A Hilbert space $\mathcal{H}$ of functions from a set $\mathcal{X}$ to $\mathbb{R}$ is an RKHS if, for each $x\in\mathcal{X}$, the evaluation functional $\delta_x:\mathcal{H}\to\mathbb{R}$ given by $f\mapsto f(x)$ is continuous.
   (a) Using the calculation given in the lecture, verify that the Sobolev space $H^s(\mathbb{R}^d)$ is an RKHS when $s > d/2$.
   (b) For each $x\in\mathcal{X}$, show the existence of a $k_x\in\mathcal{H}$ such that $f(x) = \langle k_x, f\rangle_{\mathcal{H}}$ for all $f\in\mathcal{H}$. (Hint: Riesz representation theorem.)

(c) Using the functionals constructed in part (b), define the *kernel* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by $k(x, y) := \langle k_x, k_y \rangle_{\mathcal{H}}$. Show that k is *symmetric* (i.e., $k(x, y) = k(y, x)$) and positive (i.e., $\sum_{i,j=1}^n a_i a_j k(x_i, x_j) \geq 0$ for all $a_1, \ldots, a_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in \mathcal{X}$).

(d) Prove the *representer property*: for each $x \in X$, the function $k(\cdot, x) \in \mathcal{H}$ and $\langle k(\cdot, x), f \rangle_{\mathcal{H}} = f(x)$.

(e) Show that if $\sup_{x \in \mathcal{X}} k(x, x) < \infty$, then all the functions in $\mathcal{H}$ are bounded. (Hint: $k(x, x) = \|k_x\|_{\mathcal{H}}^2$.)

(f) Show that $\mathcal{H} = \overline{\text{span}(k(\cdot, x) : x \in \mathcal{X})}$. (Hint: it suffices to show that if $f \perp k(\cdot, x)$ for all $x \in \mathcal{X}$, then $f = 0$.) This shows that the kernel k completely determines H. In fact, for every symmetric, positive $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ there exists a unique (up to isomorphism) RKHS with k as its kernel.

(g) Let $(\mathcal{Y}, \nu)$ be a measure space, and fix a function $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Assume that for each $x \in \mathcal{X}$ the function $y \mapsto \phi(x, y)$ lies in $L_2(Y, \nu)$. Show that the space

$$\mathcal{H} := \left\{ \int_{\mathcal{Y}} \phi(x, y) g(y) \, \mathrm{d}\nu(y) : g \in L_2(\mathcal{Y}, \nu) \right\} \tag{4}$$

equipped with the norm $\|f\|_{\mathcal{H}} := \inf\{\|g\|_{L_2} : f = \int_{\mathcal{Y}} \phi(\cdot, y) g(y) \, \mathrm{d}\nu(y)\}$ is an RKHS. Show that this generalizes the example of the Sobolev space considered in part (a).

5. (*Statistical Optimal Transport*, Section 2.8.2) This exercise establishes some statistical properties of RKHS's. We have already seen that $\sup_{\varphi : \|\varphi\|_{H^s} \leq 1} \int \varphi \, \mathrm{d}(\mu_n - \mu) \lesssim n^{-1/2}$. We will now show that this phenomenon is more general.

(a) Let $\mathcal{H}$ be an RKHS with kernel k which satisfies $\sup_{x \in \mathcal{X}} k(x, x) < \infty$. Show that for any $\mu \in \mathcal{P}(\mathcal{X})$, the function f_μ defined by

$$f_\mu(x) = \int k(x, x') \, \mathrm{d}\mu(x') \tag{5}$$

defines an element of $\mathcal{H}$, which satisfies $\langle f_\mu, g \rangle_{\mathcal{H}} = \int g \, \mathrm{d}\mu$ for any $g \in \mathcal{H}$. The function f_μ is sometimes called the "kernel mean embedding" of μ.

(b) Show that for any probability measures μ and ν,

$$\sup_{g \in \mathcal{H} : \|g\|_{\mathcal{H}} \leq 1} \int g \, \mathrm{d}(\mu - \nu) = \|f_\mu - f_\nu\|_{\mathcal{H}} \,, \tag{6}$$

where f_μ and f_ν are the kernel mean embeddings of μ and ν.

(c) Show that if μ_n is an empirical measure, then $\mathbb{E}\|f_\mu - f_{\mu_n}\| \lesssim n^{-1/2}$. Conclude that for any RKHS whose kernel is bounded,

$$\mathbb{E} \sup_{g \in \mathcal{H} : \|g\|_{\mathcal{H}} \leq 1} \int g \, \mathrm{d}(\mu_n - \mu) \lesssim n^{-1/2} \,. \tag{7}$$

6. This exercise verifies some basic facts about dual solutions to the entropic optimal transport problem. Let (f_n, g_n) be optimal solutions to the dual entropic OT problem

$$\sup_{f,g} \int f \, \mathrm{d}\mu_n + \int g \, \mathrm{d}\nu - \varepsilon \int e^{\frac{f(x) + g(y) - \frac{1}{2}\|x - y\|^2}{\varepsilon}} \, \mathrm{d}\mu_n(x) \, \mathrm{d}\nu(y) \,, \tag{8}$$

where as usual $\mu, \nu \in \mathcal{P}(B_1)$ and μ_n is an empirical measure. Write $\varphi_n(x) = \frac{1}{2}\|x\|^2 - f_n(x)$.

(a) Show that if (f_n, g_n) are optimal solutions, then so are $(f_n + c, g_n - c)$ for any constant c, so there is no uniqueness in the solution to the dual problem.

(b) Show that the normalization $\int f_n \, \mathrm{d}\mu = 0$ is enough to ensure that (f_n, g_n) are unique.

(c) Show that if the solutions are normalized such that $\int f_n \, \mathrm{d}\mu = 0$, then $\sup_{x \in B_2} \varphi_n(x) \leq 3$.

7. This exercise gives practice with "big-Oh in probability" notation. Recall that we say that $X_n = o_p(a_n)$ for some positive sequence a_n if $a_n^{-1} X_n \xrightarrow{p} 0$, and $X_n = O_p(a_n)$ if for any $\varepsilon > 0$ there exists a positive R such that $\mathbb{P}\{|X_n| \leq a_n R\} \leq \varepsilon$ for all n.

   (a) Show that if $X_n = o_p(a_n)$ and $Y_n = O_p(b_n)$, then $X_n Y_n = o_p(a_n b_n)$.

   (b) Conclude that the quantity Δ_n as defined in the lecture is $o_p(n^{-1/2})$, as claimed.

8. This exercise completes the proof of the limit theorem for the Wasserstein distances on finite metric spaces.

   (a) Prove the continuous mapping theorem, if you don't know it already: if $Z_n \xrightarrow{d} Z$ and g is continuous, then $g(Z_n) \xrightarrow{d} g(Z)$. (Hint: recall that $Z_n \xrightarrow{d} Z$ means that $\mathbb{E}h(Z_n) \to \mathbb{E}h(Z)$ for all continuous, bounded h.)

   (b) Use this to show that

$$\sqrt{n} \sup_{f^* \in \mathcal{D}^*} \langle f^*, \mu_n - \mu \rangle \xrightarrow{d} \sup_{f^* \in \mathcal{D}^*} \langle f^*, G \rangle , \tag{9}$$

   where $G \sim \mathcal{N}(0, \Sigma)$, with $\Sigma_{ii} = \mu_i(1 - \mu_i)$.

9. Since $\sqrt{n} W_p^p(\mu_n, \nu) \geq 0$, the same property must hold for its distributional limit. Use the expression developed in lecture to verify that this is the case.

   (a) Show that, when $\mu = \nu$, the set $\mathcal{D}^*$ satisfies

$$\mathcal{D}^* = \{f \in \mathbb{R}^k : f(x) - f(x') \leq d^p(x, x')\} \tag{10}$$

   (b) Conclude that, in this case, $\sup_{f^* \in \mathcal{D}^*} \langle f^*, G \rangle \geq 0$, and that it is almost surely positive if μ is supported on at least 2 points.

10. Mimic the analysis of the entropic optimal transport problem to show that if $f_n \in \mathcal{D}_n$ is a sequence of optimal dual solutions for (μ_n, ν) and if f_∞ is any limit point, then $f_\infty \in \mathcal{D}^*$ almost surely.

11. Let $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ be an absolutely continuous probability measure such that ν has positive density on $\operatorname{int} \operatorname{supp} \nu$ and such that $\operatorname{supp} \nu$ is convex, and let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ be arbitrary. Show that the optimal solution to the dual problem

$$\sup_{\psi} \int \|x\|^2 - 2\psi^*(x) \, \mathrm{d}\mu + \int \|y\|^2 - 2\psi(y) \, \mathrm{d}\nu \tag{11}$$

    is unique up to a constant shift. Defining ϕ to be the convex conjugate of the optimal ψ yields a canonical pair of optimal potentials, up to constant shift. (Hint: Brenier's theorem implies that $\nabla\psi$ is well defined and unique almost everywhere on $\operatorname{int} \operatorname{supp} \nu$.)

12. This exercise justifies the variance bounds presented in Lecture 4.

    (a) Prove the Efron–Stein inequality, if you don't know it already: let $X_1, \dots, X_n$ be i.i.d. random variables, let $Z = f(X_1, \dots, X_n)$ for some function f, and write $Z_i = f(X_1, \dots, X_i', \dots, X_n)$ for a version of Z where X_i is replaced by an independent copy X_i'. Then

$$\operatorname{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}(Z - Z_i)_+^2 . \tag{12}$$

    (Hint: Let $\Delta_i = \mathbb{E}[Z \mid X_1, \dots, X_i] - \mathbb{E}[Z \mid X_1, \dots, X_{i-1}]$. Show that $\operatorname{Var}(Z) = \sum_{i=1}^n \mathbb{E}\Delta_i^2$. Then, show that $\mathbb{E}\Delta_i^2 \leq \mathbb{E}(Z - \mathbb{E}[Z \mid \{X_j : j \neq i\}])^2$. Finally, use the identity $\operatorname{Var}(W) = (W - W')_+^2$, whenever W' is an i.i.d. copy of W.)

(b) Let $(\mathcal{X}, d)$ be *any* compact metric space, and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$ be arbitrary. Use the Efron–Stein inequality to prove that $\mathrm{Var}(W_2^2(\mu_n, \nu)) \lesssim n^{-1}$. (Hint: view $W_2^2(\mu_n, \nu)$ as a function of n i.i.d. samples from μ.)

(c) Let ν satisfy the assumptions of the previous problem, so that there exist unique dual solutions for the 2-Wasserstein distance ν and any element of $\mathcal{P}_2(\mathbb{R}^d)$. Assume that μ and ν are compactly supported. Let μ_n be the empirical measure corresponding to $(X_1, \ldots, X_n)$ and let μ_n' correspond to $(X_1', X_2, \ldots, X_n)$. We aim to show that $\mathrm{Var}\left(W_2^2(\mu_n, \nu) - \int (\|\cdot\|^2 - 2\varphi^*)\, \mathrm{d}\mu_n\right) = o(n^{-1})$, where φ^* is the (unique) optimal dual solution for (μ, ν)

i. Let φ_n be the unique optimal dual solution for (μ_n, ν). Show that, up to a constant shift, $\varphi_n \to \varphi^*$ uniformly on the support of μ almost surely.

ii. Using the Efron–Stein inequality, argue that it suffices to prove that $n^2 \mathbb{E}(R_n - R_n')_+^2 \to 0$, where $R_n = W_2(\mu_n, \nu) - \int (\|\cdot\|^2 - 2\varphi^*)\, \mathrm{d}\mu_n$ and $R_n' = W_2(\mu_n', \nu) - \int (\|\cdot\|^2 - 2\varphi^*)\, \mathrm{d}\mu_n'$.

iii. Show that $R_n - R_n' \leq \frac{2}{n}((\varphi^*(X_1) - \varphi_n(X_1)) - (\varphi^*(X_1') - \varphi_n(X_1')))$. Conclude that $n^2(R_n - R_n')_+^2 \lesssim \|\varphi^* - \varphi_n\|_{L_\infty(\mu)}^2$.

iv. Argue that $\mathbb{E}\|\varphi^* - \varphi_n\|_{L_\infty(\mu)}^2 \to 0$.