Math 612: Homework 2

Instructor: Geoffrey Schiebinger

November 18, 2019

This homework is due on December 9th at noon. Please submit your solutions electronically to geoff@math.ubc.ca

1 Optimization and Duality

The following two problems are adapted from Chapter 5 of *Convex Optimizaton* by Boyd and Vandenberghe.

Problem 1 (Relaxation of Boolean LP)**.** A Boolean linear program is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0,1\}, \quad i = 1, \ldots, n, \end{array} \tag{1}$$

and is, in general, very difficult to solve (even though the feasible set is finite, consisting of at most 2^n points).

In a general method called *relaxation*, the constraint that x_i be zero or one is replaced with the linear inequalities $0 \leq x_i \leq 1$:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \leq x_i \leq 1, \quad i = 1, \ldots, n. \end{array} \tag{2}$$

We refer to this as the LP relaxation of the Boolean LP. The LP relaxation is far easier to solve.

(a) Explain why the optimal value of the LP relaxation (2) is a lower bound on the optimal value of the Boolean LP (1). What can you say about the Boolean LP if the LP relaxation is infeasible?

(b) It sometimes happens that the LP relaxation has a solution with $x_i \in \{0, 1\}$. What can you say in this case?

Problem 2 (Lagrangian relaxation of Boolean LP)**.** In this exercise we study a different relaxation of the Boolean LP (1). It can be reformulated as the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i(1 - x_i) = 0, \quad i = 1, \ldots, n, \end{array}$$

which has quadratic equality constraints. While this problem is nonconvex, it has a convex dual.

(a) Find the Lagrange dual of this problem. The optimal value of the dual problem (which is convex) gives a lower bound on the optimal value of the Boolean LP. This method of finding a lower bound on the optimal value is called Lagrangian relaxation

(b) Show that the lower bounds obtained via Lagrangian relaxation, and via the LP relaxation, are the same. *Hint:* Derive the dual of the LP relaxation

2 Wasserstein Curves

Problem 3. In this problem we investigate constant speed geodesics in the space of probability distributions over the real numbers.

Let $\mu_0 = \mathcal{N}(0,1)$ be a normal distribution with mean 0 and variance 1, and let μ_1 denote the normal distribution with mean 10 and variance 1. Show that $\mu_t = \mathcal{N}(10t, 1)$ with $t \in (0,1)$ is a constant speed geodesic connecting μ_0 to μ_1.

For the next problem, it will help to familiarize yourself with the Python Optimal Transport package by looking at the following example code:

`https://pot.readthedocs.io/en/stable/auto_examples/plot_OT_2D_samples.html`

To compute an entropy regularized coupling, use the function

`ot.sinkhorn`

To compute the optimal transport coupling without entropy, use the function

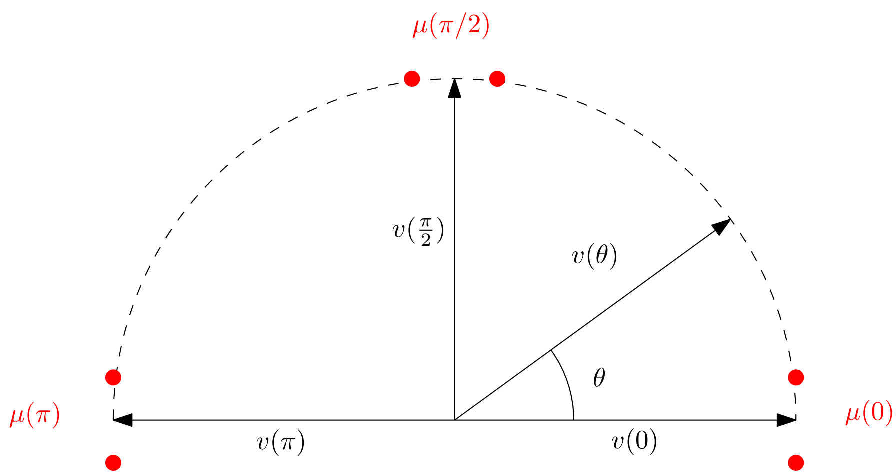
`ot.emd`

Problem 4. In this problem we investigate a curve in the space of probability distributions over a two-dimensional state space.

For $\theta \in [0,\pi]$, let $v(\theta)$ denote the vector $v(\theta) = [\cos(\theta), \sin(\theta)]$. For each angle θ, we construct a measure $\mu(\theta)$ consisting of two point-masses:

$$\mu(\theta) = \frac{1}{2}\delta_{x_1(\theta)} + \frac{1}{2}\delta_{x_2(\theta)},$$

where $x_1(\theta) = v(\theta - \frac{\pi}{32})$ and $x_2(\theta) = v(\theta + \frac{\pi}{32})$. This set-up is illustrated in the Figure below.



As θ varies from 0 to π, the measure $\mu(\theta)$ describes a curve in the space of probability distributions. In this problem we investigate piecewise-geodesic approximations to this curve for different entropy parameters.

Denote the entropic coupling between $\mu(\theta_1)$ and $\mu(\theta_2)$ with entropy parameter $\lambda > 0$ by $\Gamma_\lambda(\theta_1, \theta_2, \lambda)$:

$$\begin{aligned}\Gamma_\lambda(\theta_1,\theta_2) = \text{argmin}_\gamma \ & \sum_{i=1}^{2}\sum_{j=1}^{2} \|x_i(\theta_1) - x_j(\theta_2)\|^2 \gamma_{i,j} + \lambda \sum_{i,j} \gamma_{i,j}\log\gamma_{i,j} \\ \text{s.t} \ & \sum_i \gamma_{i,j} = \frac{1}{2} \\ & \sum_j \gamma_{i,j} = \frac{1}{2}.\end{aligned} \tag{3}$$

(a) We begin by computing a single geodesic connecting $\mu(0)$ to $\mu(1)$. Compute $\Gamma_\lambda(0, \pi)$ for $\lambda = 0$ and $\lambda = 0.1$. Express the answers as 2×2 matrices with entry i, j denoting the mass transported from $x_i(0)$ to $x_j(\pi)$.

(b) If we imagine the measure $\mu(\theta)$ slowly rotating around the circle from $\theta = 0$ to $\theta = \pi$, we see that the top point rotates around and become the bottom point for $\theta = \pi$. Does $\Gamma_\lambda(0, \pi)$ capture this behavior for any λ?

(c) We capture this rotation by composing multiple couplings. Let G^N_λ denote the coupling obtained by the composition

$$G^N_\lambda = \Gamma_\lambda(0, \frac{\pi}{N}) \circ \Gamma_\lambda(\frac{\pi}{N}, \frac{2\pi}{N}) \circ \ldots \circ \Gamma_\lambda(\frac{N-1}{N}\pi, \pi),$$

for an integer $N > 0$. Compute numerical values of G^N_λ for $N = 1, 2, 3, 4, 5, 10, 20, 100$ with the setting $\lambda = \frac{0.1}{N}$. Plot the (1,2) and (1,1) entries of G^N_λ as a function of N.

The composition can be computed with matrix multiplication if you first normalize the rows of Γ_λ to sum to 1 instead of $\frac{1}{2}$ (otherwise the product will shrink with N).

(d) We now examine how well geodesic interpolation works from $\mu(0)$ to $\mu(\theta)$, for various angles θ. For $\lambda = 0$, construct a constant speed geodesic connecting $\mu(0)$ to $\mu(\theta)$. Denote the measure at the mid-point along this geodesic by $\hat{\mu}(\theta/2)$. Compute the Wasserstein distance

$$I(\theta) = W_2(\hat{\mu}(\theta/2), \mu(\theta/2))$$

for $\theta = \frac{\pi}{32}, \frac{\pi}{16}, \frac{\pi}{8}, \frac{\pi}{4}, \frac{\pi}{2}, \pi$. Let $\tilde{\mu}(\theta/2)$ denote the interpolating distribution coming from the random coupling (instead of the OT coupling). Compute the Wasserstein distance

$$R(\theta) = W_2(\tilde{\mu}(\theta/2), \mu(\theta/2)).$$

Plot $I(\theta)$ and $R(\theta)$ as a function of θ.