Homework 4

Toffalini Leonardo

Exercise 1

Prove that for all $p \in \Delta_n$, $D_{\text{KL}}(p, p^1) \leq \log n$. Here p^1 is the uniform probability distribution with $p_i^1 = \frac{1}{n}$ for $i = 1, ..., n$.

Solution: Let $p = (p_1, ..., p_n)$.

$$\begin{aligned} D_{\text{KL}}(p, p^1) &= \sum_{i=1}^{n} p_i \log \frac{p_i}{1/n} = \sum_{i=1}^{n} p_i \log(p_i n) = \sum_{i=1}^{n} p_i (\log p_i + \log n) = \sum_{i=1}^{n} p_i \log p_i + \sum_{i=1}^{n} p_i \log n \\ &= \sum_{i=1}^{n} p_i \log p_i + \log n \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p_i \log p_i + \log n. \end{aligned}$$

Since $p_i \leq 1$ for all $i = 1, ...n$, we have $\log p_i \leq 0$ for all $i = 1, ...n$. Thus, we get the final result:

$$D_{\text{KL}}(p, p^1) \leq \log n.$$

□

Exercise 2

Verify that in the exponential gradient descent algorithm, p^{t+1} is the projection of w^{t+1} onto Δ_n with respect to D_H (i.e. show that $p^{t+1} = \text{argmin}\ \{D_H(x, w^{t+1}) : x \in \Delta_n\}$).

Solution: Let's recall the definitions:

$$\Delta_n := \left\{ p \in [0,1]^n : \sum_{i=1}^{n} p_i = 1 \right\}$$

$$p^{t+1} := \text{argmin}_{p \in \Delta_n} \{D_{\text{KL}}(p, p^t) + \alpha \nabla f(p^t) \cdot p\}$$

$$w_i^{t+1} := p_i^t e^{-\alpha \nabla f(p^t)_i}$$

$$p_i^{t+1} := \frac{w_i^{t+1}}{\sum_{j=1}^{n} w_j^{t+1}}$$

Let $z = w^{t+1}$,thus $p^{t+1} = z/\|z\|_1$. First we derive $D_H(z/\|z\|_1), z)$, then show that this is a lower bound for any other point on Δ_n.

$$\begin{aligned} D_H\left(\frac{z}{\|z\|_1}, z\right) &= -\sum \frac{z_i}{\|z\|_1} \log \frac{z_i}{\frac{z_i}{\|z\|_1}} + \sum \left(z_i - \frac{z_i}{\|z\|_1}\right) \\ &= -\frac{\log \|z\|_1}{\|z\|_1} \sum z_i + \sum z_i - \frac{1}{\|z\|_1} \sum z_i \\ &= -\log \|z\|_1 + \|z\|_1 - 1. \end{aligned}$$

$$\begin{aligned} D_H(x, z) &= -\sum x_i \log \frac{z_i}{x_i} + \sum z_i - \sum x_i \\ &= -\sum x_i \log \frac{z_i}{x_i} + \|z\|_1 - 1. \end{aligned}$$

Thus, we need only prove that

$$\sum x_i \log \frac{z_i}{x_i} \leq \log \|z\|_1.$$

We know that $x \log \frac{1}{x}$ is concave, thus we need only minimize $-\sum x_i \log \frac{z_i}{x_i}$. However, we need to keep in mind that $x \in \Delta_n$, which is an additional constraint. Hence, the following convex problem needs to be solved:

$$\min_{x \in \Delta_n} \quad -\sum x_i \log \frac{z_i}{x_i}.$$

With a Lagrange multiplier the above is equivalent to the following:

$$\min_{x} \quad -\sum x_i \log \frac{z_i}{x_i} + \mu\left(\sum x_i - 1\right).$$

Taking the gradient of the objective function we get that

$$\log x_i - \log z_i - 1 + \mu = 0.$$

Solving this we get that

$$x_i = z_i \cdot e^{1-\mu}.$$

To accomodate for the constraint $x \in \Delta_1 \iff \|x\|_1 = 1$, we have

$$1 = \|x\|_1 = \sum x_i = \sum z_i \cdot e^{1-\mu} = e^{1-\mu}\|z\|_1.$$

Thus $e^{1-\mu} = \frac{1}{\|z\|_1}$.

With $x^* = \frac{z}{\|z\|_1}$ we have

$$\sum x_i^* \log \frac{z_i}{x_i^*} = \log \|z\|_1.$$

Circling back to what we were hoping to show, we have

$$\sum x_i \log \frac{z_i}{x_i} \leq \log \|z\|_1.$$

Furthermore, the maximum is attained at $x = \frac{z}{\|z\|_1}$, which is exactly what we were hoping to show.

$\square$

Exercise 3

Given a sequence of conex and differentiable function $f^1, f^2, \ldots : K \to \mathbb{R}$ and a sequence of points $x^1, x^2, \ldots \in K$, define the **regret** up to time T to be

$$\text{Regret}_T := \sum_{t=1}^{T} f^t(x^t) - \min_{x \in K} \sum_{t=1}^{T} f^t(x).$$

Consider the following strategy:

$$x^{t+1} := \text{argmin}_{x \in K} \sum_{i=1}^{t} f^i(x) + R(x)$$

for a convex regularizer $R : K \to \mathbb{R}$ and $x^1 := \text{argmin}_{x \in K} R(x)$. Assume that the gradient of each f_i is bounded everywhere by G and that the diameter of K is bounded by D. Prove the following:

1. $\text{Regret}_T \leq \sum_{t=1}^{T} (f^t(x^t) - f^t(x^{t+1})) - R(x^1) + R(x^*)$ for every $T = 1, \ldots$, where

$$x^* := \text{argmin}_{x \in K} \sum_{t=1}^{T} f^t(x)$$

.

2. Given $\varepsilon > 0$, use this method for $R(x) := \frac{1}{\eta}\|x\|_2^2$ for an appropriate choice of η and T to get $\frac{1}{T}\text{Regret}_T \leq \varepsilon$.

Solution: Solution □

Exercise 4

Let $G = (V, E)$ be an undirected graph with n vertices and m edges, and let $s, t \in V$ distinct vertices. Let $B \in \mathbb{R}^{n \times m}$ denote the vertex-edge incidence matrix of the graph, that is we pick an arbitrary orientation of the graph and set the column corresponding to each arc uv as $e_v - e_u$. Consider the s–t maximum flow problem:

$$\begin{aligned} \max_{x \in \mathbb{R}^m, F \geq 0} \quad & F \\ \text{s.t.} \quad & Bx = Fb, \\ & \|x\|_\infty \leq 1, \end{aligned}$$

where $b = e_s - e_t$.

1. Prove that the dual of this formulation is equvalent to the following:

$$\begin{aligned} \min_{y \in \mathbb{R}^n} \quad & \sum_{uv \in E} |y_u - y_v| \\ \text{s.t.} \quad & y_s - y_t = 1. \end{aligned}$$

2. Prove that the optimal value of the above problem is equal to $\text{MinCut}_{s,t}(G)$, the minimum number of edges one needs to remove from

G to disconnect s from t. This latter problem is known as the s–t minimum cut problem.

3. Reformulate the dual as follows:

$$\begin{aligned} \min_{x \in \mathbb{R}^m} \quad & \|x\|_1 \\ \text{s.t.} \quad & x \in \operatorname{Im}(B^T) \\ & x \cdot z = 1 \end{aligned}$$

for some $z \in \mathbb{R}^m$ that depends on G and s, t. Write an explicit formula for z.

Solution:

1. Notice that this problem is just an LP, so we can use the simple LP dual.

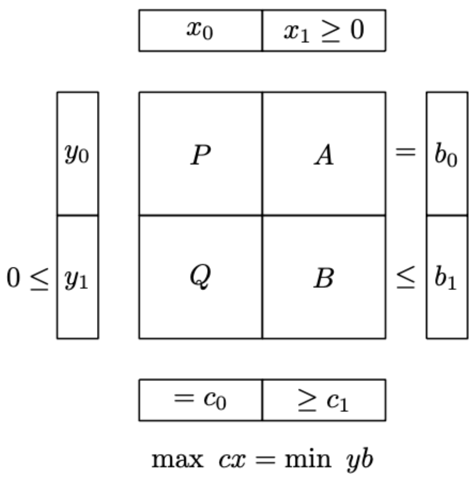

Figure 1: General form of LP duality theorem.

Using the general LP form, we can write our problem as

$$
\begin{gathered}
x_0 = x, \quad x_1 = F \in \mathbb{R} \\
P = B, \quad A = -b \\
Q = \begin{bmatrix} I_m \\ -I_m \end{bmatrix} \\
b_0 = \underline{0}_n, \quad b_1 = \underline{1}_{2m} \\
c_0 = \underline{0}_m, \quad c_1 = 1.
\end{gathered}
$$

With this notation, our primal problem becomes

$$
\begin{gathered}
\max x_0 c_1 + x_1 c_1 \\
Bx - Fb = Px_0 + Ax_1 = b_0 \\
\|x\|_1 \leq 1 \iff \begin{bmatrix} I_m \\ -I_m \end{bmatrix} x_0 \leq 1 \iff Qx_0 \leq 1 \\
F \geq 0 \iff x_1 \geq 0.
\end{gathered}
$$

The dual of this general LP is the following:

$$
\begin{gathered}
y_0 B = 0 \implies y_u^0 - y_v^0 = 0 \quad \forall (uv) \in E \\
\implies y_u^0 = y_v^0 \quad \forall (uv) \in E \\
-y_0 b \geq 1 \iff y_0 b \leq -1 \iff y_s^0 - y_t^0 \leq 1
\end{gathered}
$$

These were the parts that come from P, A and c_0, c_1. Now the part that comes from Q and c_0. Let y_e^1 denote the coordinate of y_1 corresponding to edge $e \in E$, and $\left(y_e^1\right)'$ the other coordinate corresponding to edge e, since $y_1 \in \mathbb{R}^{2m}$.

$$
\begin{gathered}
\begin{bmatrix} I_m \\ -I_m \end{bmatrix} y_1 = 0 \iff y_e^1 - \left(y_e^1\right)' = 0 \quad \forall e \in E \\
\iff y_e^1 = \left(y_e^1\right)' \quad \forall e \in E
\end{gathered}
$$

The dual objective function then becomes:

$$
y_0 b_0 + y_1 b_1 \iff y_1 \cdot \underline{1}_{2m} = \sum_{i=1}^{2m} y_i^1 = \sum_{e \in E} 2y_e.
$$

2. Let S and T denote the following set of vertices:

$$
\begin{gathered}
S = \{u : y_u = 1\} \\
T = \{v : y_v = 0\}.
\end{gathered}
$$

From this we can see that from the constraint $y_s - y_t = 1$ it follows that $s \in S$ and $t \in T$. Moreover, we can see that $|y_u - y_v| = 0$ if $u, v \in S$ or $u, v \in T$. Thus, the only set of edges that remain in the sum are the ones that run between S and T. This is precisely the number of edges that when removed disconnect S from T. Hence, this dual is equivalent to $\mathrm{MinCut}_{s,t}(G)$.

3. Since $x \in \mathrm{Im}(B^T)$, this means that there exists a $y \in \mathbb{R}^n$ such that $yB = x$. Notice that this y can be precisely what we defined in the second section of this problem, meaning that $x_e = x_{uv} = y_u - y_v$.

With this the objective functions match, we only need to show that there exists a $z \in \mathbb{R}^m$ such that

$$
x \cdot z = y_s - y_t.
$$

We can see that $x \cdot z = y_s - y_t$ means that we need to find a z_e for all edges such that

$$\sum_{e \in \text{ in}(u)} x_e z_e = \sum_{e \in \text{ out}(u)} x_e z_e$$

for all $u \in V - \{s, s, t\}$.

Meaning, that the weighted sum of the in-edges to a vertex u must equal the weighted sum of the out-edges from a vertex for all inner vertices (so exluding the source and the sink).

The extra constraint is that the weighted sum of the out-edges from s must equal 1. Likewise that weighted sum of the in-edges to t must equal 1.

This is basically a linear system of equations that is easily solvable.

$\square$