CS 15-759: Homework 3

Due: 2/21/2025, 11:59 PM on Canvas

Bonus points: If you find typos in my homework or lecture notes, please email me. You will earn +1 bonus points per typo found, and potentially more for especially egregious typos.

Hints: Hints are on the last page. It is recommended to think about the problem without hints for a while, and then look at the hints when stuck. The problems are meant to be difficult, so there is no shame in looking at the hints. If you make partial progress on problems (e.g., by following the hints) you will get partial points.

Problem 1: Smoothness of Softmax (15 Points)

Solve Exercise 9 in the Lecture Notes.

Problem 2: Bregman Divergence of ℓ_p for $p \in [1, 2]$ (15 Points)

Solve Exercise 10 in the Lecture Notes.

Problem 3: Practice with Hölder's Inequality (20 Points)

Hölder's inequality says the following. If $p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then:

$$a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \leq \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} \left(\sum_{i=1}^n |b_i|^q \right)^{1/q}.$$

(a) For $a_i, x_i, y_i > 0$ prove that:

$$\sum_{i=1}^n \frac{a_i^3}{x_i y_i} \geq \frac{(\sum_{i=1}^n a_i)^3}{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}.$$

(b) Prove that if $x \in \mathbb{R}^n$ and $p \leq q$ then $\|x\|_p \leq n^{1/p - 1/q} \|x\|_q$.

Problem 4: Packing Linear Programs (50 Points)

A packing linear program is the following. Given a matrix $A \in \mathbb{R}^{m \times n}_{>0}$ with positive entries, to find $\max_{x \in \mathbb{R}^n_{\geq 0}, Ax \leq 1} \sum_{i=1}^n x_i$. In other words, maximize the sum of x_i over all nonnegative vectors x

satisfying that $(Ax)_i \leq 1$ for all coordinates i. In this problem, we will describe how to design a multiplicative weights algorithm to solve a packing LP efficiently.

For this problem it will be nicer to work with the following decision version of a packing LP (which you should argue is equivalent): given a matrix A and μ, under the guarantee that

$$\min_{\sum_{i=1}^n x_i = 1} \max_{1 \leq j \leq m} (Ax)_j \leq \mu, \tag{0.1}$$

find an $x \geq 0$ satisfying $\sum_{i=1}^n x_i = 1$ and $(Ax)_j \leq (1+\varepsilon)\mu$ for all $1 \leq j \leq m$.

Problem: Given an algorithm which solves (0.1) up to $(1+\varepsilon)$-multiplicative accuracy in runtime $O(\mathsf{nnz}(A)\varepsilon^{-O(1)} \log(mn))$.

1 Hints

Problem 1: Directly verify that $z^\top \nabla^2 g_\lambda(x) z \le O(\lambda)\|z\|_\infty^2$ for all $x, z \in \mathbb{R}^n$.

Problem 2: Try to emulate the proof of Lemma 7.1 of the Lecture Notes.

Problem 4: Use the following setup for an MWU algorithm.

Consider the potential function $\Phi(x) = \sum_{j=1}^m e^{(Ax)_i}$, and initialize $x^{(0)} = 0$. Define $x^{(t+1)} = x^{(t)} + \delta\Delta^{(t)}$, where δ is a well-chosen parameter, and $\Delta^{(t)}$ satisfies that $\sum_{i=1}^n \Delta_i^{(t)} = 1$, and minimizes the first-order of the potential increase. More formally,

$$\Delta^{(t)} = \text{argmin}_{\sum_{i=1}^n \Delta_i = 1} \sum_{j=1}^m e^{(Ax)_i} (A\Delta)_i.$$

It's not hard to see that $\Delta^{(t)}$ will be one of the n vectors $(1, 0, \dots, 0)$, $(0, 1, \dots, 0)$, $\dots$.

Choices of parameters. Set $\delta = \frac{\varepsilon}{\mu}$, and run the iteration for $T = O(\varepsilon^{-2} \log m)$ steps. Finally, note that $\sum_{i=1}^n x_i^{(T)} = \delta T$, so outputs $\frac{1}{\delta T} x^{(T)}$.

Analysis. Compare $\Phi(x^{(t+1)})$ to $\Phi(x^{(t)})$.

References