Homework 10

This homework is due at 11 PM on November 8, 2023.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

1. **Convergence of Gradient Descent for Different Step Sizes**

   Let $m > 0$ and $L > 0$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable, m-strongly convex and L-smooth function. We aim to find an optimal solution to the unconstrained minimization problem:

$$\min_{\vec{x} \in \mathbb{R}^n} \quad f(\vec{x}). \tag{1}$$

   For this, consider the gradient descent algorithm with fixed step size η, where $0 < \eta \leq \frac{1}{L}$. Namely, the algorithm starts at $\vec{x}^{(0)} \in \mathbb{R}^n$, and is at $\vec{x}^{(k)} \in \mathbb{R}^n$ at time $k \geq 0$, where

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \eta \nabla f(\vec{x}^{(k)}), \quad k \geq 0.$$

   (a) Show that $L \geq m$.

   (b) In class, and in the course reader, we have argued that, since f is m-strongly convex, the optimization problem (1) has a unique optimal solution, call it $\vec{x}^*$. (We proved this only under the assumption that f is differentiable, but it is true more generally, and you can either try to prove this for yourself or accept it as a fact.)

   Let $p^* := f(\vec{x}^*)$ denote the optimal value of the minimization problem (1). Show that for all $k \geq 0$ we have

$$f(\vec{x}^{(k)}) - p^* \leq (1 - \eta m)^k (f(\vec{x}^{(0)}) - p^*).$$

   HINT: If $M \geq L$, then $(M - L)I$ is a positive semidefinite matrix, where I denotes the $n \times n$ identity matrix.

2. Visualizing the Dual Problem

Download the Jupyter notebook `dual_visualize.ipynb`; complete the code where designated and answer the questions given in the space provided. (If you prefer, for the questions that do not involve writing code, you can write the solutions on a separate sheet of paper or LaTeX PDF; just make sure to correctly mark the relevant pages when uploading to Gradescope.)

3. Maximizing a Sum of Logarithms

Consider the following problem, which arises in the estimation of the transition probabilities of a discrete-time Markov chain:

$$p^\star = \max_{\vec{x}\in\mathbb{R}^n} \quad \sum_{i=1}^n \alpha_i \log(x_i) \tag{2}$$

$$\text{s.t.} \quad \vec{x} \geq \vec{0}, \quad \vec{1}^\top \vec{x} = c, \tag{3}$$

where $c > 0$ and $\alpha_i > 0$, $i = 1, \ldots, n$. (Recall that if $\vec{x}$ is a vector then by "$\vec{x} \geq \vec{0}$" we mean "$x_i \geq 0$ for each i.") We will determine in closed-form a minimizer, and show that the optimal objective value of this problem is

$$p^\star = \alpha \log(c/\alpha) + \sum_{i=1}^n \alpha_i \log(\alpha_i), \tag{4}$$

where $\alpha \doteq \sum_{i=1}^n \alpha_i$. We will show this in a series of steps. Note that the constraints $\vec{x} \geq \vec{0}$ are not really needed, since the domain of the objective function is $\mathbb{R}^n_{++} := \{\vec{x} \in \mathbb{R}^n | x_i > 0 \text{ for each } i \in \{1, \cdots, n\}\}$. Nevertheless, we will work with the formulation in which these constraints are stated explicitly.

(a) First, express the problem as a minimization problem which has optimal value $-p^\star$.

(b) In optimization, we often "relax" problems of the form $p^\star_{\min} = \min_{\vec{x}\in\mathcal{X}} f_0(\vec{x})$, by replacing the constraint set $\mathcal{X}$ with a larger constraint set $\mathcal{X}_r$, and instead solving $p^\star_r = \min_{\vec{x}\in\mathcal{X}_r} f_0(\vec{x})$, then showing a connection between $p^\star_{\min}$ and $p^\star_r$. In this problem, a particular relaxation we will use is to replace the equality constraint $\vec{1}^\top \vec{x} = c$ with an inequality constraint $\vec{1}^\top \vec{x} \leq c$.

Show that the relaxed problem has the same optimal value as the original problem, i.e., $p^\star_r = p^\star_{\min}$, and the two problems have the same solutions.

HINT: First argue that $p^\star_r \leq p^\star_{\min}$. Then, suppose for the sake of contradiction that $p^\star_r < p^\star_{\min}$. Let $\vec{x}^r$ be a solution to the relaxed minimization problem which has objective value $p^\star_r$. Consider the vector $\vec{x}$ given by

$$\vec{x} \doteq \begin{bmatrix} c - 1^\top \vec{x}^r + x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{bmatrix}. \tag{5}$$

Show that $\vec{x}$ is feasible for the original problem and has objective value strictly less than $p^\star_r$. Argue that this implies $p^\star_{\min} < p^\star_r$ and derive a contradiction. Finally, argue that any solution to the relaxed problem is a solution to the original problem, and vice-versa — you might need to use a construction similar to that of $\vec{x}$.

(c) After relaxing the equality constraint to an inequality constraint, form the Lagrangian $\mathcal{L}(\vec{x}, \vec{\lambda}, \mu)$ for the relaxed minimization problem, where λ_i is the dual variable corresponding to the inequality $x_i \geq 0$, and μ is the dual variable corresponding to the inequality constraint $\vec{1}^\top \vec{x} \leq c$. Note that the domain of the Lagrangian is $\mathbb{R}^n_{++} \times \mathbb{R}^n \times \mathbb{R}$, since the domain of the relaxed version of the problem is $\mathbb{R}^n_{++}$.

(d) Now derive the dual function $g(\vec{\lambda}, \mu)$ for the relaxed minimization problem, and solve the dual problem $d^\star_r = \max_{\substack{\vec{\lambda}\geq\vec{0} \\ \mu\geq 0}} g(\vec{\lambda}, \mu)$.

What are the optimal dual variables $\vec{\lambda}^\star, \mu^\star$?

(e) Show that strong duality holds for the relaxed problem, so $p^\star_r = d^\star_r$.

(f) From the $\vec{\lambda}^\star, \mu^\star$ obtained in the previous part, how do we obtain the optimal primal variable $x^\star$? What is the optimal objective function value $p^\star_r$? Finally, what is $p^\star$?

4. Quadratically Constrained Linear Program

Let $\vec{c}, \vec{x}_0 \in \mathbb{R}^n$ where $\vec{c} \neq \vec{0}$. Let $Q \in \mathbb{S}^n_{++}$ be a symmetric positive definite matrix. Let $\epsilon > 0$ be a positive scalar. Consider the following optimization problem

$$\begin{aligned} p^\star = \min_{\vec{x}} \quad & \vec{c}^\top \vec{x}, \\ \text{s.t.} \quad & \frac{1}{2}(\vec{x} - \vec{x}_0)^\top Q(\vec{x} - \vec{x}_0) \leq \epsilon. \end{aligned} \tag{6}$$

(a) Is this problem convex? Justify your answer.

(b) Prove that the dual function associated with the primal problem in (6) is

$$g(\lambda) = \begin{cases} -\infty & \text{if } \lambda = 0, \\ \vec{c}^\top \vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon & \text{if } \lambda > 0, \end{cases} \tag{7}$$

for $\lambda \geq 0$, where λ is the dual variable associated with the quadratic inequality constraint.

(c) Consider the dual problem of the primal problem in (6):

$$d^\star = \max_{\lambda \geq 0} g(\lambda), \tag{8}$$

where

$$g(\lambda) = \begin{cases} -\infty & \text{if } \lambda = 0, \\ \vec{c}^\top \vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon & \text{if } \lambda > 0. \end{cases} \tag{7}$$

Find the optimal dual variable $\lambda^\star$.

(d) Does strong duality hold for the optimization problem (6)? Justify your answer by directly computing $p^\star$ and $d^\star$, without appealing to constraint qualifications such as Slater's condition.

5. Linear Programs and Duality

Consider the following two linear programs:

$$\begin{aligned} \min_{x_1, x_2 \in \mathbb{R}} \quad & c_1 x_1 - 2x_2 && (9) \\ \text{s.t.} \quad & x_1 \geq 0, \\ & x_2 \geq 0, \\ & x_1 + x_2 \leq 1, \\ \max_{x_1 \in \mathbb{R}} \quad & c_2 x_1 && (10) \\ \text{s.t.} \quad & x_1 \geq 7. \end{aligned}$$

Find the values of c_1 and c_2 such that (10) is the Lagrangian dual problem of (9).

6. Does strong duality hold?

Consider

$$\min_{(x,y)\in\mathcal{D}} e^{-x} \tag{11}$$
$$\text{s.t. } x^2/y \le 0 \tag{12}$$

where $\mathcal{D} := \{(x,y) \mid y > 0\}$ is to be thought of as the domain of the constraint function $f_1(x,y) = x^2/y$.

(a) Prove that the problem is convex. Find the optimal value. *HINT: To prove the constraint function is convex, you will have to prove it is convex with respect to the vector* $\begin{bmatrix} x & y \end{bmatrix}^\top$. *Consider computing the Hessian of the constraint function, its determinant and trace, and show that it is PSD by analyzing signs of its eigenvalues.*

(b) Next, we will proceed to find an optimal solution and an optimal value for the dual problem. The Lagrangian dual function $g(\lambda)$, can be written as:

$$g(\lambda) = \inf_{(x,y)\in\mathcal{D}} \left(e^{-x} + \lambda \frac{x^2}{y} \right). \tag{13}$$

Explain why $g(\lambda)$ is lower bounded by 0 for $\lambda \ge 0$.

(c) Show that $g(\lambda) = 0$ for $\lambda \ge 0$. *HINT: To show that the infimum in Equation (13) is* 0, *we want to show there exist* (x,y) *such that both* e^{-x} *and* $\lambda\frac{x^2}{y}$ *can get arbitrarily close to* 0. *HINT: Consider a sequence* $\{x_k\}$ *going to* $+\infty$ *and a sequence* $\{y_k\}$ *also going to* $+\infty$ *such that* $\lim_{k\to\infty} \frac{x_k^2}{y_k} = 0$. *Simply put, we want to drive* x *to infinity in order to drive* e^{-x} *to* 0, *while having* y *grow faster than* x^2, *so that the second term also goes to* 0.

(d) Now, write the dual problem and find an optimal solution λ^* and an optimal value d^* for the dual problem using the results above. What is the duality gap?

(e) Does Slater's condition hold for this problem? Does strong duality hold?