Homework 9

This homework is due at 11 PM on March 22, 2024.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned), as well as a printout of your completed Jupyter notebook(s).

1. **Optimizing Over Multiple Variables**

   In this exercise, we consider several problems in which we optimize over two variables, $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$, and a general (possibly nonconvex) objective function, $F_0(\vec{x}, \vec{y})$. Suppose also that $\vec{x}$ and $\vec{y}$ are constrained to different feasible sets $\mathcal{X}$ and $\mathcal{Y}$, respectively, which may or may not be convex.

   (a) Show that

   $$\min_{\vec{x} \in \mathcal{X}} \min_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}) = \min_{\vec{y} \in \mathcal{Y}} \min_{\vec{x} \in \mathcal{X}} F_0(\vec{x}, \vec{y}), \tag{1}$$

   i.e., if we minimize over both $\vec{x}$ and $\vec{y}$, then we can exchange the minimization order without altering the optimal value.

   (b) Show that $p^\star \geq d^\star$, where

   $$p^\star \doteq \min_{\vec{x} \in \mathcal{X}} \max_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}) \tag{2}$$

   $$d^\star \doteq \max_{\vec{y} \in \mathcal{Y}} \min_{\vec{x} \in \mathcal{X}} F_0(\vec{x}, \vec{y}). \tag{3}$$

   This statement is referred to as the *min-max theorem*.

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 questions that do not involve writing code, you can write solutions on separate paper or LaTeX PDF, just make sure to correctly mark the relevant pages when uploading to Gradescope.)

3. Duality

Consider the function

$$f(\vec{x}) = \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x}. \tag{4}$$

First, we consider the unconstrained optimization problem

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) = \min_{\vec{x}\in\mathbb{R}^n} \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} \tag{5}$$

for a real $n \times n$ symmetric matrix $A \in \mathbb{S}^n$ and $\vec{b} \in \mathbb{R}^n$. If the problem is unbounded below, then we say $p^\star = -\infty$. Let $\vec{x}^\star$ denote the minimizing argument of the optimization problem.

(a) Suppose $A \succeq 0$ (positive semidefinite) and $\vec{b} \in \mathcal{R}(A)$. Let $\operatorname{rank}(A) = n$. Find $p^\star$.

HINT: What does $A \succeq 0$ tell you about the function f? How can you leverage the rank of A to compute $p^\star$?

(b) Suppose $A \succeq 0$ (positive semidefinite) and $\vec{b} \in \mathcal{R}(A)$ as before. Let A be rank-deficient, i.e., $\operatorname{rank}(A) = r < n$. Let A have the compact/thin and full SVD as follows, with diagonal positive definite $\Lambda_r \in \mathbb{R}^{r\times r}$:

$$A = U_r\Lambda_r U_r^\top = \begin{bmatrix} U_r & U_1 \end{bmatrix} \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^\top \\ U_1^\top \end{bmatrix}. \tag{6}$$

Show that the minimizer $\vec{x}^\star$ of the optimization problem (5) is not unique by finding a general form for the family of solutions for $\vec{x}^\star$ in terms of $U_r, U_1, \Lambda_r, \vec{b}$.

HINT: As before, $A \succeq 0$ gives you some information about the objective function f. Can you use this information along with the fact that $b \in \mathcal{R}(A)$ to obtain a general form for the minimizers of f? Use the fact that any vector $\vec{x} \in \mathbb{R}^n$ can be written as $\vec{x} = U_r\vec{\alpha} + U_1\vec{\beta}$ for unique $\vec{\alpha}, \vec{\beta}$.

(c) If $A \not\succeq 0$ (A not positive semi-definite) show that $p^\star = -\infty$ by finding $\vec{v}$ such that $f(\alpha\vec{v}) \to -\infty$ as $\alpha \to \infty$.

HINT: $A \not\succeq 0$ means that there exists $\vec{v}$ such that $\vec{v}^\top A\vec{v} < 0$.

(d) Suppose $A \succeq 0$ (positive semidefinite) and $\vec{b} \notin \mathcal{R}(A)$. Find $p^\star$. Justify your answer mathematically.

HINT: From FTLA, we know that $\mathbb{R}^n = \mathcal{R}(A^\top) \oplus \mathcal{N}(A)$. Therefore, $\vec{b} = \vec{v}_1 + \vec{v}_2$ where $\vec{v}_1 \in \mathcal{R}(A) = \mathcal{R}(A^\top)$ and $\vec{v}_2 \in \mathcal{N}(A)$.

For parts (e) and (f), consider real $n \times n$ symmetric matrix $A \in \mathbb{S}^n$ and $\vec{b} \in \mathbb{R}^n$. Let $\operatorname{rank}(A) = r$, where $0 \le r \le n$. Now we consider the constrained optimization problem

$$\begin{aligned} p^\star = \min_{\vec{x}\in\mathbb{R}^n} &\ \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} \\ \text{s.t. } &\ \vec{x}^\top \vec{x} \ge 1. \end{aligned} \tag{7}$$

(e) Write the Lagrangian $\mathcal{L}(\vec{x}, \lambda)$, where λ is the dual variable corresponding to the inequality constraint.

(f) For any matrix $C \in \mathbb{R}^{n\times n}$ with $\operatorname{rank}(C) = r \le n$ and compact SVD

$$C = U_r\Lambda_r V_r^\top, \tag{8}$$

we define the pseudoinverse as

$$C^\dagger = V_r\Lambda_r^{-1}U_r^\top. \tag{9}$$

We use the "dagger" operator to represent this. If $\vec{d}$ lies in the range of C, then a solution to the equation $C\vec{x} = \vec{d}$, can be written as $\vec{x} = C^\dagger \vec{d}$, even when C is not full rank. Show that the dual problem to the primal problem (7) can be written as,

$$d^\star = \max_{\substack{\lambda \geq 0 \\ A - \lambda I \succeq 0 \\ \vec{b} \in \mathcal{R}(A - \lambda I)}} -\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda. \tag{10}$$

HINT: To show this, first argue that when the constraints are not satisfied $\min_{\vec{x}} \mathcal{L}(\vec{x}, \lambda) = -\infty$. *Then show that when the constraints are satisfied,* $\min_{\vec{x}} \mathcal{L}(\vec{x}, \lambda) = -\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda$.

HINT: Compute $g(\lambda)$ *and explore its behavior under the constraints.*

4. Sensitivity and Dual Variables

In this problem, we explore the interpretation of dual variables as sensitivity parameters of the primal problem. Recall the canonical convex primal problem

$$\min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{11}$$

$$\text{s.t.} \quad f_i(\vec{x}) \le 0, \quad i = 1,\dots,m \tag{12}$$

$$h_j(\vec{x}) = 0, \quad j = 1,\dots,p \tag{13}$$

where f_i are convex for all $i = 0,\dots,m$ and h_j are affine for all $j = 1,\dots,p$.

For $\vec{u} = (u_1,\dots,u_m)^\top \in \mathbb{R}^m$ and $\vec{v} = (v_1,\dots,v_p)^\top \in \mathbb{R}^p$, we consider the *perturbed* problem

$$p^\star(\vec{u},\vec{v}) = \min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{14}$$

$$\text{s.t.} \quad f_i(\vec{x}) \le u_i, \quad i = 1,\dots,m \tag{15}$$

$$h_j(\vec{x}) = v_j, \quad j = 1,\dots,p \tag{16}$$

In other words, $p^\star(\vec{u},\vec{v})$ is a function of $\vec{u}$ and $\vec{v}$ that gives the optimal value for the perturbed problem (if it is feasible). If the problem is infeasible (i.e. no points exist that satisfy the constraints), we say that $p^\star(\vec{u},\vec{v}) = +\infty$ otherwise. Note that $p^\star(\vec{0},\vec{0})$ is the value of the original problem.

(a) Prove that $p^\star(\vec{u},\vec{v})$ is jointly convex[1] in $(\vec{u}\in\mathbb{R}^m,\ \vec{v}\in\mathbb{R}^p)$.

HINT: Let

$$\mathcal{D} \doteq \{(\vec{x}\in\mathbb{R}^n, \vec{u}\in\mathbb{R}^m, \vec{v}\in\mathbb{R}^p) \mid f_i(\vec{x}) \le u_i \quad \forall i, \quad h_j(\vec{x}) = v_j \quad \forall j\}, \tag{17}$$

which is the set of all triples $(\vec{x},\vec{u},\vec{v})$ such that $\vec{x}$ is a feasible point for the perturbed problem with the perturbations $(\vec{u},\vec{v})$. Show that $\mathcal{D}$ is convex. Now define $F(\vec{x},\vec{u},\vec{v})$ to be a function that is equal to $f_0(\vec{x})$ on $\mathcal{D}$ and $+\infty$ otherwise. Prove that $F(\vec{x},\vec{u},\vec{v})$ is jointly convex in $(\vec{x},\vec{u},\vec{v})$, and then observe that

$$p^\star(\vec{u},\vec{v}) = \min_{\vec{x}} F(\vec{x},\vec{u},\vec{v}). \tag{18}$$

From here, to show that $p^\star(\vec{u},\vec{v})$ is jointly convex in $(\vec{u},\vec{v})$, you may prove and use the following lemma: Let S_1, S_2 be convex sets with a function $f\colon S_1 \times S_2 \to \mathbb{R}$ which is jointly convex in both arguments. Define $g(\vec{y}) = \min_{\vec{x}\in S_1} f(\vec{x},\vec{y})$. Then $g(\vec{y})$ is convex in $\vec{y}\in S_2$.

(b) Assume that strong duality holds, and that the dual optimum is attained, for the unperturbed primal problem (11). Let $(\vec{\lambda}^\star,\vec{\sigma}^\star)$ be the optimal dual variables for the dual of (11). Show that for any point $\vec{z}$ that is feasible for the perturbed problem (14), we have

$$f_0(\vec{z}) \ge p^\star(\vec{0},\vec{0}) - \vec{u}^\top\vec{\lambda}^\star - \vec{v}^\top\vec{\sigma}^\star \tag{19}$$

HINT: Let L and g be the Lagrangian and dual function, respectively, of the unperturbed primal problem (11). Use strong duality to relate $p^\star(\vec{0},\vec{0})$ to $g(\vec{\lambda}^\star,\vec{\sigma}^\star)$. Upper-bound the value of $g(\vec{\lambda}^\star,\vec{\sigma}^\star)$ by the value of $L(\vec{z},\vec{\lambda}^\star,\vec{\sigma}^\star)$. Then, noting that $\vec{z}$ is feasible for the perturbed problem (14), apply bounds on $f_i(\vec{z})$ and $h_j(\vec{z})$ to bound $L(\vec{z},\vec{\lambda}^\star,\vec{\sigma}^\star)$.

[1] Recall that a function $f\colon A\times B \to \mathbb{R}$ is jointly convex in $(\vec{a}\in A,\ \vec{b}\in B)$ if $A\times B$ is convex, and for all $\theta\in[0,1]$, and for all $\vec{a}_1,\vec{a}_2\in A,\ \vec{b}_1,\vec{b}_2\in B$, we have that $f(\theta\vec{a}_1 + (1-\theta)\vec{a}_2,\ \theta\vec{b}_1 + (1-\theta)\vec{b}_2) \le \theta f(\vec{a}_1,\vec{b}_1) + (1-\theta)f(\vec{a}_2,\vec{b}_2)$.

(c) Using the result of part (b), show that for all $\vec{u}, \vec{v}$, we have

$$p^\star(\vec{u}, \vec{v}) \geq p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\sigma}^\star. \tag{20}$$

(d) Suppose we only have 1 equality and 1 inequality constraint (that is u, v are scalars). For each (u, v), exactly one of the following three cases must apply: either we must have $p^\star(u, v) > p^\star(0, 0)$, we must have $p^\star(u, v) < p^\star(0, 0)$, or it is impossible to conclude which one is greater. For each of the following situations, argue which case applies.

i. $\lambda^\star$ is large (as compared with $\sigma^\star$) and $u < 0$.

ii. $\lambda^\star$ is large (as compared with $\sigma^\star$) and $u > 0$.

iii. $\sigma^\star$ is large (as compared with $\lambda^\star$) and positive and $v < 0$.

iv. $\sigma^\star$ is large (as compared with $\lambda^\star$) and negative and $v > 0$.

HINT: Use the bound you computed in part (c).

Note that we can think of u and v as variables we choose — by examining how the solution to our original primal problem changes, we can describe how "sensitive" our problem is to its different constraints!

5. KKT with Circles

Consider the problem

$$\min_{\vec{x}\in\mathbb{R}^2} \quad x_1^2 + x_2^2 \tag{21}$$

$$\text{s.t.} \quad (x_1-1)^2 + (x_2-1)^2 \leq 2 \tag{22}$$

$$(x_1-1)^2 + (x_2+1)^2 \leq 2 \tag{23}$$

where $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top \in \mathbb{R}^2$.

(a) Sketch the feasible region and the level sets of the objective function. Find the optimal point $\vec{x}^*$ and the optimal value p^*.

(b) Does strong duality hold?

(c) Write the KKT conditions for this optimization problem. Do there exist Lagrange multipliers λ_1^* and λ_2^* that prove the optimality of $\vec{x}^*$?

6. Homework Process

With whom did you work on this homework? List the names and SIDs of your group members.

NOTE: If you didn't work with anyone, you can put "none" as your answer.