# Homework 9

## Self grades are due at 11 PM on April 5, 2024.

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.

**Solution:** We first consider the quantity $\min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y})$, which can be viewed as a function of $\vec{x}$. We can write

$$F_0(\vec{x}, \vec{y}) \geq \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \tag{2}$$

$$\geq \min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \tag{3}$$

where both lines follow from the definition of a minimum. The inequality above holds for every $\vec{x} \in \mathcal{X}$, so it holds for the value $\vec{x}$ that minimizes this quantity, i.e.,

$$\min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}). \tag{4}$$

This inequality also holds for every $\vec{y} \in \mathcal{Y}$, so

$$\min_{\vec{y}\in\mathcal{Y}} \min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}). \tag{5}$$

By symmetry, we can reverse our treatment of $\vec{x}$ and $\vec{y}$ and arrive at the reversed inequality

$$\min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{y}\in\mathcal{Y}} \min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}). \tag{6}$$

Since both (5) and (6) must hold, the expressions must be equal, as desired.

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

$$p^\star \doteq \min_{\vec{x}\in\mathcal{X}} \max_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \tag{7}$$

$$d^\star \doteq \max_{\vec{y}\in\mathcal{Y}} \min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}). \tag{8}$$

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

**Solution:** By the definitions of minimization and maximization, we have that

$$L(\vec{y}) \doteq \min_{\vec{x}'} F_0(\vec{x}', \vec{y}) \leq F_0(\vec{x}, \vec{y}) \leq U(\vec{x}) \doteq \max_{\vec{y}'} F_0(\vec{x}, \vec{y}') \tag{9}$$

for every $\vec{x} \in \mathcal{X}$ and $\vec{y} \in \mathcal{Y}$, or more simply,

$$L(\vec{y}) \leq U(\vec{x}). \tag{10}$$

Since this inequality holds for all $\vec{x} \in \mathcal{X}$, it holds for the value of $\vec{x}$ that minimizes $U(\vec{x})$, and thus

$$p^{\star} = \min_{\vec{x} \in \mathcal{X}} U(\vec{x}) \geq L(\vec{y}). \tag{11}$$

Similarly, since the above holds for all $\vec{y} \in \mathcal{Y}$, it holds for the value of $\vec{y}$ that maximizes $L(\vec{y})$, and thus

$$p^{\star} \geq \max_{\vec{y} \in \mathcal{Y}} L(\vec{y}) = d^{\star} \tag{12}$$

as desired.

**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

$$f(\vec{x}) = \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x}. \tag{13}$$

First, we consider the unconstrained optimization problem

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) = \min_{\vec{x}\in\mathbb{R}^n} \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} \tag{14}$$

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 $p^\star$?*

**Solution:** If $\operatorname{rank}(A) = n$, then $A \succ 0$, and therefore the objective is strictly convex. Setting the gradient to 0 we obtain,

$$\nabla_{\vec{x}} f(\vec{x}) = 2A\vec{x} - 2\vec{b} = 0 \tag{15}$$
$$\implies A\vec{x} = \vec{b} \tag{16}$$
$$\implies \vec{x}^\star = A^{-1}\vec{b} \tag{17}$$

Where in the last step, we used that fact that a full rank square matrix is invertible. Plugging this back into our objective function we get,

$$f(\vec{x}^\star) = (\vec{b}^\top (A^{-1})^\top)A(A^{-1}\vec{b}) - 2\vec{b}^\top (A^{-1}\vec{b}) \tag{18}$$
$$= \vec{b}^\top (A^\top)^{-1} \cancel{AA^{-1}} \vec{b} - 2\vec{b}^\top A^{-1}\vec{b} \tag{19}$$
$$= \vec{b}^\top A^{-1}\vec{b} - 2\vec{b}^\top A^{-1}\vec{b} \tag{20}$$
$$p^\star = -\vec{b}^\top A^{-1}\vec{b} \tag{21}$$

(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 $\Lambda_r \in \mathbb{R}^{r\times r}$:

$$A = U_r \Lambda_r U_r^\top = \begin{bmatrix} U_r & U_1 \end{bmatrix} \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^\top \\ U_1^\top \end{bmatrix}. \tag{22}$$

Show that the minimizer $\vec{x}^\star$ of the optimization problem (14) 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 $\vec{\alpha}, \vec{\beta}$.*

**Solution:** Since $A \succeq 0$, $f(\vec{x})$ is convex and we can attempt to find the minimizer by setting the gradient to zero. Doing this we obtain,

$$A\vec{x} = b, \tag{23}$$

as in the part (a) of this problem.

However, now this equation has infinite solutions since $\vec{b}$ lies in the range of $A$ and $A$ is rank-deficient. Indeed we can add any vector from the (non-trivial) nullspace of $A$ to any particular solution $\vec{x}_0$ of Equation (23) and get another solution.

By the Fundamental Theorem of Linear Algebra we have,

$$\vec{x} = U_r\vec{\alpha} + U_1\vec{\beta} \tag{24}$$

$$\vec{b} = U_r\vec{\gamma}, \tag{25}$$

where we used the fact that $\vec{b} \in \mathcal{R}(A)$. Using this we obtain,

$$U_r\Lambda_r U_r^\top(U_r\vec{\alpha} + U_1\vec{\beta}) = U_r\vec{\gamma} \tag{26}$$

Since the columns of $U_1$ and $U_r$ are orthogonal to each other and because $U_r^\top U_r = I, \Lambda_r$ is invertible we have,

$$U_r\Lambda_r U_r^\top U_r\vec{\alpha} = U_r\vec{\gamma} \tag{27}$$

$$\implies \vec{\alpha} = \Lambda_r^{-1}\vec{\gamma} \tag{28}$$

$$= \Lambda_r^{-1}U_r^\top\vec{b}. \tag{29}$$

Thus any solution to Equation (23) and hence a minimizer to the optimization problem (14) can be written as,

$$\vec{x}^\star = U_r\Lambda_r^{-1}U_r^\top\vec{b} + U_1\vec{\beta}. \tag{30}$$

(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 $\vec{v}^\top A\vec{v} < 0$.*

**Solution:** Since $A \not\succeq 0$ there exists an eigenvalue, eigenvector pair $(\mu, \vec{v})$ such that

$$\vec{v}^\top A\vec{v} = \mu < 0. \tag{31}$$

Assuming without loss of generality that $-2\vec{b}^\top\vec{v} \leq 0$ (If it is positive then multiply $\vec{v}$ by $-1$) we can take $\vec{x} = \alpha\vec{v}$ to obtain,

$$f(\vec{x}) = f(\alpha\vec{v}) = \alpha^2\vec{v}^\top A\vec{v} + \alpha(-2\vec{b}^\top\vec{v}), \tag{32}$$

which goes to $-\infty$ as $\alpha$ goes to $\infty$ since $\vec{v}^\top A\vec{v} < 0$ and $-2\vec{b}^\top\vec{v} \leq 0$.

(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 $\vec{v}_2 \in \mathcal{N}(A)$.*

**Solution:** First, note that since $A$ is symmetric, we have $\mathcal{R}(A) = \mathcal{R}(A^\top)$. We have $\vec{b} = \vec{v}_1 + \vec{v}_2$ with $\vec{v}_1 \in \mathcal{R}(A) = \mathcal{R}(A^\top)$ and $\vec{v}_2 \in \mathcal{N}(A)$ as $\mathbb{R}^n = \mathcal{R}(A) \oplus \mathcal{N}(A)$ from the Fundamental Theorem of Linear Algebra. We cannot have $\vec{v}_2 = 0$ as otherwise we'd get $\vec{b} = \vec{v}_1 \in \mathcal{R}(A)$ which is a contradiction. Now, let $\vec{v} = \vec{v}_2$. We get from this:

$$f(\alpha\vec{v}) = \alpha^2\vec{v}^\top A\vec{v} - 2\alpha(\vec{v}_1 + \vec{v}_2)^\top\vec{v}_2 = 0 - 2\alpha\left\|\vec{v}_2\right\|^2 \tag{33}$$

where we used the fact that $\vec{v}_2 \in \mathcal{N}(A)$ and $\vec{v}_1 \in \mathcal{R}(A)$. As $\alpha \to \infty$, we get that $f(\alpha\vec{v}) \to -\infty$ from which we conclude that $p^\star = -\infty$.

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

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} \quad (34)$$
$$\text{s.t. } \vec{x}^\top \vec{x} \ge 1.$$

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

**Solution:**

$$\mathcal{L}(\vec{x}, \lambda) = \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} + \lambda(1 - \vec{x}^\top \vec{x}) \quad (35)$$
$$= \vec{x}^\top A\vec{x} - \vec{x}^\top \lambda\vec{x} - 2\vec{b}^\top \vec{x} + \lambda \quad (36)$$
$$= \vec{x}^\top (A - \lambda I)\vec{x} - 2\vec{b}^\top \vec{x} + \lambda \quad (37)$$

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

$$C = U_r \Lambda_r V_r^\top, \quad (38)$$

we define the pseudoinverse as

$$C^\dagger = V_r \Lambda_r^{-1} U_r^\top. \quad (39)$$

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 (34) can be written as,

$$d^\star = \max_{\substack{\lambda \ge 0 \\ A - \lambda I \succeq 0 \\ \vec{b} \in \mathcal{R}(A - \lambda I)}} -\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda. \quad (40)$$

*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.*

**Solution:**

$$g(\lambda) = \min_{\vec{x}} \mathcal{L}(\vec{x}, \lambda) = \min_{\vec{x}} \vec{x}^\top (A - \lambda I)\vec{x} - 2\vec{b}^\top \vec{x} + \lambda \quad (41)$$

Drawing from parts (c) and (d), we can see that if $A - \lambda I \not\succeq 0$ or if $A - \lambda I \succeq 0, \vec{b} \notin \mathcal{R}(A - \lambda I)$, then we can choose $\vec{x}$ to drive the Lagrangian to $-\infty$. If the constraints are satisfied, however, then we can proceed like in part (b) by taking the gradient:

$$\nabla_{\vec{x}} \mathcal{L} = 2(A - \lambda I)\vec{x} - 2\vec{b} = 0 \quad (42)$$
$$(A - \lambda I)\vec{x} = \vec{b} \quad (43)$$
$$\vec{x}^\star = (A - \lambda I)^\dagger \vec{b} \quad (44)$$

where in the last step, we used the fact that the PSD contraint on $A - \lambda I$ is satisfied and $\vec{b}$ lies in the range of $A - \lambda I$, so we can use the pseudoinverse and the gradient-zero point is indeed the minimum. Plugging this back into the Lagrangian, we get:

$$\begin{aligned}
\mathcal{L}(\vec{x}^\star, \lambda) &= \vec{b}^\top((A-\lambda I)^\dagger)^\top(A-\lambda I)(A-\lambda I)^\dagger\vec{b} - 2\vec{b}^\top(A-\lambda I)^\dagger\vec{b} + \lambda && (45)\\
&= \vec{b}^\top(A-\lambda I)^\dagger(A-\lambda I)(A-\lambda I)^\dagger\vec{b} - 2\vec{b}^\top(A-\lambda I)^\dagger\vec{b} + \lambda && (46)\\
&= \vec{b}^\top(A-\lambda I)^\dagger\vec{b} - 2\vec{b}^\top(A-\lambda I)^\dagger\vec{b} + \lambda && (47)\\
&= -\vec{b}^\top(A-\lambda I)^\dagger\vec{b} + \lambda && (48)
\end{aligned}$$

where we used the fact that $(A - \lambda I)^\dagger$ is symmetric and by properties of pseudo inverse,

$$(A-\lambda I)^\dagger(A-\lambda I)(A-\lambda I)^\dagger = (A-\lambda I)^\dagger. \tag{49}$$

Now, we have a full expression for our dual function:

$$g(\lambda) = \begin{cases} -b^\top(A-\lambda I)^\dagger b + \lambda & \text{if } A - \lambda I \succeq 0, b \in \mathcal{R}(A-\lambda I) \\ -\infty & \text{else} \end{cases} \tag{50}$$

The dual problem follows, as it is just a maximization of the dual function:

$$d^\star = \max_{\lambda \geq 0} g(\lambda) \tag{51}$$

**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

$$\min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{52}$$

$$\text{s.t.} \quad f_i(\vec{x}) \leq 0, \quad i = 1, \ldots, m \tag{53}$$

$$h_j(\vec{x}) = 0, \quad j = 1, \ldots, p \tag{54}$$

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

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

$$p^\star(\vec{u}, \vec{v}) = \min_{\vec{x}\in\mathbb{R}^n} \quad f_0(\vec{x}) \tag{55}$$

$$\text{s.t.} \quad f_i(\vec{x}) \leq u_i, \quad i = 1, \ldots, m \tag{56}$$

$$h_j(\vec{x}) = v_j, \quad j = 1, \ldots, p \tag{57}$$

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 $p^\star(\vec{u}, \vec{v}) = +\infty$. 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*

$$\mathcal{D} \doteq \{(\vec{x} \in \mathbb{R}^n, \vec{u} \in \mathbb{R}^m, \vec{v} \in \mathbb{R}^p) \mid f_i(\vec{x}) \leq u_i \quad \forall i, \quad h_j(\vec{x}) = v_j \quad \forall j\}, \tag{58}$$

*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*

$$p^\star(\vec{u}, \vec{v}) = \min_{\vec{x}} F(\vec{x}, \vec{u}, \vec{v}). \tag{59}$$

*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 $\vec{y} \in S_2$.*

**Solution:** First, we show that $\mathcal{D}$ is convex. Indeed, let $(\vec{x}_1, \vec{u}_1, \vec{v}_1), (\vec{x}_2, \vec{u}_2, \vec{v}_2) \in \mathcal{D}$ and $\lambda \in [0, 1]$. We want to show that

$$\lambda(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)(\vec{x}_2, \vec{u}_2, \vec{v}_2) = (\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2, \lambda\vec{u}_1 + (1-\lambda)\vec{u}_2, \lambda\vec{v}_1 + (1-\lambda)\vec{v}_2) \in \mathcal{D}. \tag{60}$$

Indeed, for each $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, p\}$, we have

$$f_i(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) \leq \lambda f_i(\vec{x}_1) + (1-\lambda)f_i(\vec{x}_2) \tag{61}$$

$$\leq \lambda(\vec{u}_1)_i + (1-\lambda)(\vec{u}_2)_i \tag{62}$$

$$= (\lambda\vec{u}_1 + (1-\lambda)\vec{u}_2)_i. \tag{63}$$

$$h_j(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) = \lambda h_j(\vec{x}_1) + (1-\lambda)h_j(\vec{x}_2) \tag{64}$$

[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 $f(\theta\vec{a}_1 + (1-\theta)\vec{a}_2,\ \theta\vec{b}_1 + (1-\theta)\vec{b}_2) \leq \theta f(\vec{a}_1, \vec{b}_1) + (1-\theta)f(\vec{a}_2, \vec{b}_2)$.

$$= \lambda(\vec{v}_1)_j + (1-\lambda)(\vec{v}_2)_j \quad (65)$$
$$= (\lambda\vec{v}_1 + (1-\lambda)\vec{v}_2)_j. \quad (66)$$

This shows that $\mathcal{D}$ is convex.

Now we show that $F$ is convex. Let $(\vec{x}_1, \vec{u}_1, \vec{v}_1), (\vec{x}_2, \vec{u}_2, \vec{v}_2) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ and $\lambda \in (0, 1)$ (the endpoints $\lambda \in \{0, 1\}$ can be dealt with separately). Then there are two cases:

(i) If $(\vec{x}_1, \vec{u}_1, \vec{v}_1) \notin \mathcal{D}$ or $(\vec{x}_2, \vec{u}_2, \vec{v}_2) \notin \mathcal{D}$, then we must have either $F(\vec{x}_1, \vec{u}_1, \vec{v}_1) = \infty$ or $F(\vec{x}_2, \vec{u}_2, \vec{v}_2) = \infty$. Thus we must have

$$F(\lambda(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)(\vec{x}_2, \vec{u}_2, \vec{v}_2)) \leq \lambda F(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)F(\vec{x}_2, \vec{u}_2, \vec{v}_2) = \infty. \quad (67)$$

(ii) If both $(\vec{x}_1, \vec{u}_1, \vec{v}_1) \in \mathcal{D}$ and $(\vec{x}_2, \vec{u}_2, \vec{v}_2) \in \mathcal{D}$, then because $\mathcal{D}$ is convex, we have $\lambda(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)(\vec{x}_2, \vec{u}_2, \vec{v}_2) \in \mathcal{D}$. Thus we have

$$F(\lambda(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)(\vec{x}_2, \vec{u}_2, \vec{v}_2)) = f_0(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) \quad (68)$$
$$\leq \lambda f_0(\vec{x}_1) + (1-\lambda) f_0(\vec{x}_2) \quad (69)$$
$$= \lambda F(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)F(\vec{x}_2, \vec{u}_2, \vec{v}_2). \quad (70)$$

Thus in every case we have

$$F(\lambda(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)(\vec{x}_2, \vec{u}_2, \vec{v}_2)) \leq \lambda F(\vec{x}_1, \vec{u}_1, \vec{v}_1) + (1-\lambda)F(\vec{x}_2, \vec{u}_2, \vec{v}_2), \quad (71)$$

so that $F$ is convex.

Now by definition we have

$$p^\star(\vec{u}, \vec{v}) = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ (\vec{x}, \vec{u}, \vec{v}) \in \mathcal{D}}} f_0(\vec{x}) = \min_{\substack{\vec{x} \in \mathbb{R}^n \\ (\vec{x}, \vec{u}, \vec{v}) \in \mathcal{D}}} F(\vec{x}, \vec{u}, \vec{v}) = \min_{\vec{x} \in \mathbb{R}^n} F(\vec{x}, \vec{u}, \vec{v}), \quad (72)$$

because $F$ takes values $+\infty$ outside of $\mathcal{D}$, so the minimum will never be achieved outside of $\mathcal{D}$.

To show that $p^\star(\vec{u}, \vec{v})$ is convex, we prove the lemma in the hint. 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 we show $g$ is convex in $\vec{y} \in S_2$. Indeed, let $\vec{y}_1, \vec{y}_2 \in S_2$ and $\lambda \in [0, 1]$. Then by definition of the pointwise minimum, we have

$$g(\lambda\vec{y}_1 + (1-\lambda)\vec{y}_2) \leq f(\vec{x}, \lambda\vec{y}_1 + (1-\lambda)\vec{y}_2), \qquad \forall \vec{x} \in S_1. \quad (73)$$

Because $S_1$ is a convex set, it also holds if we write $\vec{x}$ as a convex combination of points in $S_1$, i.e.,

$$g(\lambda\vec{y}_1 + (1-\lambda)\vec{y}_2) \leq f(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2, \lambda\vec{y}_1 + (1-\lambda)\vec{y}_2) \quad (74)$$
$$= f(\lambda(\vec{x}_1, \vec{y}_1) + (1-\lambda)(\vec{x}_2, \vec{y}_2)), \qquad \forall \vec{x}_1, \vec{x}_2 \in S_1. \quad (75)$$

Now we use the convexity of $f$ to obtain

$$g(\lambda\vec{y}_1 + (1-\lambda)\vec{y}_2) \leq \lambda f(\vec{x}_1, \vec{y}_1) + (1-\lambda) f(\vec{x}_2, \vec{y}_2), \qquad \forall \vec{x}_1, \vec{x}_2 \in S_1. \quad (76)$$

Since this holds for all $\vec{x}_1$ and $\vec{x}_2$, it must hold for the minimizing $\vec{x}_1, \vec{x}_2$ (which, to be clear, are chosen as a function of $\vec{y}_1, \vec{y}_2$), so this gives

$$g(\lambda\vec{y}_1 + (1-\lambda)\vec{y}_2) \leq \min_{\vec{x}_1, \vec{x}_2 \in S_1} [\lambda f(\vec{x}_1, \vec{y}_1) + (1-\lambda) f(\vec{x}_2, \vec{y}_2)] \quad (77)$$

$$= \lambda \min_{\vec{x}_1 \in S_1} f(\vec{x}_1, \vec{y}_1) + (1-\lambda) \min_{\vec{x}_2 \in S_1} f(\vec{x}_2, \vec{y}_2) \tag{78}$$

$$= \lambda g(\vec{y}_1) + (1-\lambda) g(\vec{y}_2). \tag{79}$$

Thus $g$ is convex in $\vec{y} \in S_2$.

Applying it to this case, we observe that

$$p^\star(\vec{u}, \vec{v}) = \min_{\vec{x} \in \mathbb{R}^n} F(\vec{x}, \vec{u}, \vec{v}) \tag{80}$$

and the latter function is jointly convex in all three of its input variables, so $p^\star(\vec{u}, \vec{v})$ is jointly convex in $(\vec{u}, \vec{v})$.

(b) Assume that strong duality holds, and that the dual optimum is attained, for the unperturbed primal problem (52). Let $(\vec{\lambda}^\star, \vec{\sigma}^\star)$ be the optimal dual variables for the dual of (52). Show that for any point $\vec{z}$ that is feasible for the perturbed problem (55), we have

$$f_0(\vec{z}) \geq p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\sigma}^\star \tag{81}$$

*HINT: Let $L$ and $g$ be the Lagrangian and dual function, respectively, of the unperturbed primal problem (52). 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 (55), apply bounds on $f_i(\vec{z})$ and $h_j(\vec{z})$ to bound $L(\vec{z}, \vec{\lambda}^\star, \vec{\sigma}^\star)$.*

**Solution:** We have by strong duality and the bounds in (55) on $f_i$, $h_j$ that

$$p^\star(\vec{0}, \vec{0}) = d^\star \tag{82}$$

$$= g(\vec{\lambda}^\star, \vec{\sigma}^\star) \tag{83}$$

$$= \min_{\vec{x} \in \mathbb{R}^n} L(\vec{x}, \vec{\lambda}^\star, \vec{\sigma}^\star) \tag{84}$$

$$\leq L(\vec{z}, \vec{\lambda}^\star, \vec{\sigma}^\star) \tag{85}$$

$$= f_0(\vec{z}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\vec{z}) + \sum_{j=1}^{p} \sigma_j^\star h_j(\vec{z}) \tag{86}$$

$$\leq f_0(\vec{z}) + \sum_{i=1}^{m} \lambda_i^\star u_i + \sum_{j=1}^{p} \sigma_j^\star v_j \tag{87}$$

$$= f_0(\vec{z}) + \vec{u}^\top \vec{\lambda}^\star + \vec{v}^\top \vec{\sigma}^\star. \tag{88}$$

Rearranging obtains the desired equality.

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

$$p^\star(\vec{u}, \vec{v}) \geq p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\sigma}^\star. \tag{89}$$

**Solution:** Let $\mathcal{F}(\vec{u}, \vec{v}) \subseteq \mathbb{R}^n$ be the feasible set for the perturbed problem (55), so that $\min_{\vec{z} \in \mathcal{F}(\vec{u}, \vec{v})} f_0(\vec{z}) = p^\star(\vec{u}, \vec{v})$. From part (b), for any $\vec{z}$ feasible for (55), we have

$$f_0(\vec{z}) \geq p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\sigma}^\star. \tag{90}$$

We take minimums over $\vec{z} \in \mathcal{F}(\vec{u}, \vec{v})$ on both sides, and note that the right-hand side does not depend on $\vec{z}$, to obtain

$$p^\star(\vec{u}, \vec{v}) = \min_{\vec{z} \in \mathcal{F}(\vec{u}, \vec{v})} f_0(\vec{z}) \geq \min_{\vec{z} \in \mathcal{F}(\vec{u}, \vec{v})} [p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\sigma}^\star] = p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\sigma}^\star, \tag{91}$$

as desired.

(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, assume that $|u| \approx |v| \approx 1$, the small Lagrange multiplier has absolute value $\approx 1$, the large Lagrange multiplier has absolute value $\approx 10$, and 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!

**Solution:** For the scalar case, we can restate the relationship proved in part (b) as

$$p^\star(u, v) \geq p^\star(0, 0) - \lambda^\star u - \sigma^\star v. \tag{92}$$

For $u = v = 0$, this bound is (trivially) tight: it just says that $p^\star(0, 0) = p^\star(0, 0)$. This means that when we increase the value of $u$ and $v$ incrementally, we can use this bound to determine whether $p^\star(u, v)$ increases or decreases as compared with $p^\star(0, 0)$.

i. Since $\lambda^\star$ is large and $u$ is negative, we know that $-\lambda^\star u \geq 0$, so our lower bound on $p^\star(u, v)$ increases; thus, $p^\star(u, v)$ must increase accordingly.

ii. In this case, $-\lambda^\star u \leq 0$, so the lower bound on $p^\star(u, v)$ decreases; this doesn't allow us to say anything about whether $p^\star(u, v)$ increases, decreases, or remains the same, since it could still obey this bound for any of the three cases.

iii. As in case **4.**((d))i, $-\sigma^\star v \geq 0$, so $p^\star(u, v)$ increases.

iv. As in case **4.**((d))iii, $-\sigma^\star v \geq 0$, so $p^\star(u, v)$ increases.

**5. KKT with Circles**

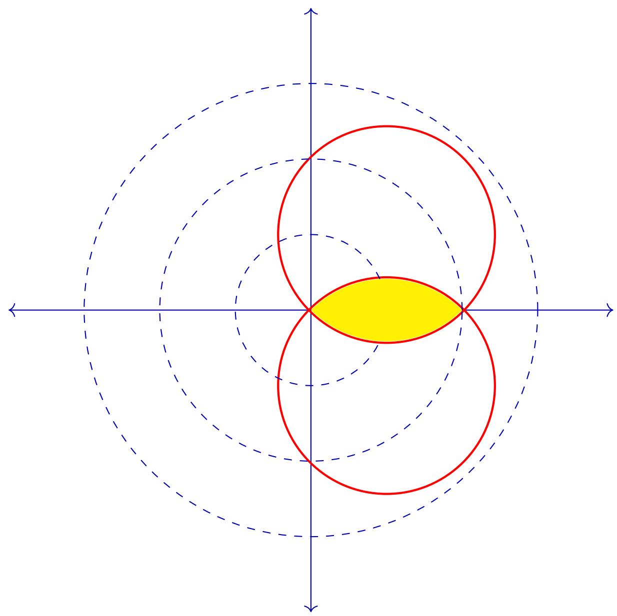
Consider the problem

$$\min_{\vec{x}\in\mathbb{R}^2} \quad x_1^2 + x_2^2 \tag{93}$$
$$\text{s.t.} \quad (x_1-1)^2 + (x_2-1)^2 \leq 2 \tag{94}$$
$$(x_1-1)^2 + (x_2+1)^2 \leq 2 \tag{95}$$

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 $p^*$.

**Solution:**



The feasible region is given by the yellow area in the graphic above. The optimal solution is the closest point to the origin inside the feasible region. Since the origin is an element of this feasible region, we have $\vec{x}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$, and $p^* = 0$.

(b) Does strong duality hold?

**Solution:** The problem is convex (i.e., the objective function and the feasible set are both convex). The feasible set contains interior points (e.g., $\vec{x} = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$), so Slater's condition is satisfied and thus strong duality holds.

(c) Write the KKT conditions for this optimization problem. Do there exist Lagrange multipliers $\lambda_1^*$ and $\lambda_2^*$ that prove the optimality of $\vec{x}^*$?

**Solution:** The Lagrangian is given by

$$\mathcal{L}(x,\lambda) = x_1^2 + x_2^2 + \lambda_1[(x_1-1)^2 + (x_2-1)^2 - 2] + \lambda_2[(x_1-1)^2 + (x_2+1)^2 - 2]. \tag{96}$$

We can write the KKT conditions as follows:

i. Stationarity:

$$x_1^* + (\lambda_1^* + \lambda_2^*)(x_1^* - 1) = 0, \tag{97}$$

$$x_2^* + \lambda_1^*(x_2^* - 1) + \lambda_2^*(x_2^* + 1) = 0. \tag{98}$$

ii. Primal feasibility:

$$(x_1^* - 1)^2 + (x_2^* - 1)^2 - 2 \leq 0, \tag{99}$$

$$(x_1^* - 1)^2 + (x_2^* + 1)^2 - 2 \leq 0. \tag{100}$$

iii. Dual feasibility:

$$\lambda_1^* \geq 0,\ \lambda_2^* \geq 0. \tag{101}$$

iv. Complementary slackness:

$$\lambda_1^*[(x_1^* - 1)^2 + (x_2^* - 1)^2 - 2] = 0, \tag{102}$$

$$\lambda_2^*[(x_1^* - 1)^2 + (x_2^* + 1)^2 - 2] = 0. \tag{103}$$

From the stationarity conditions (along with dual feasibility), we can conclude that

$$x_1^* = 0,\ x_2^* = 0 \Rightarrow \lambda_1^* = \lambda_2^* = 0. \tag{104}$$

Since in the previous part we already show this is a convex problem with differentiable objective and constraint functions, and that Slater's condition holds, we know KKT conditions provide necessary and sufficient conditions for optimality. Since these values for $\vec{x}^*$ and $\vec{\lambda}^*$ satisfy the KKT conditions and strong duality holds, we can conclude that $\vec{x}^*$ is primal optimal (and, additionally, that $\vec{\lambda}^*$ is dual optimal).

**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.