EECS 127/227AT Optimization Models in Engineering UC Berkeley Spring 2023

Homework 11

Self grades are due at 11 PM on April 20, 2023.

1. About general optimization

In this exercise, we test your understanding of the general framework of optimization and its language. We consider an optimization problem in standard form:

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} f_0(\vec{x}) \;:\; f_i(\vec{x}) \le 0, \;\; i = 1, \ldots, m. \tag{1}$$

In the following we denote by $\mathcal{X}$ the feasible set. Note that the feasible set is a subset of $\mathbb{R}^n$ that satisfies the inequalities $f_i(\vec{x}) \le 0$, i.e $\mathcal{X} = \{\vec{x} \in \mathbb{R}^n \mid f_i(\vec{x}) \le 0, i = 1, \ldots, m\}$. We make no assumption about the convexity of $f_0(\vec{x})$ and $f_i(\vec{x}), \; i = 1, \ldots, m$. For the following statements, provide a proof or counter-example.

(a) A general optimization problem can be expressed as one with a linear objective.

Solution: The statement is true:

$$p^\star = \min_{\vec{x}\in\mathcal{X}, t} t \;:\; t \ge f_0(\vec{x}). \tag{2}$$

(b) A general optimization problem can be expressed as an unconstrained problem with a different objective function which could possibly take a value of ∞ for some values of $\vec{x}$.

Solution: Again the statement is true: let us define

$$g(\vec{x}) := \begin{cases} f_0(\vec{x}) & \text{if } \vec{x} \in \mathcal{X}, \\ +\infty & \text{otherwise} \end{cases} \tag{3}$$

Then

$$p^\star = \min_{\vec{x}} g(\vec{x}). \tag{4}$$

(c) A general optimization problem can be recast as a linear program (minimizing a linear objective subject to linear constraints), provided one allows for infinitely many constraints.

Solution: This is true again; we have

$$p^\star = \max_{t} t \;:\; t \le f_0(\vec{x}) \text{ for every } \vec{x} \in \mathcal{X}. \tag{5}$$

Alternatively,

$$p^\star = -\min_{t} -t \;:\; t \le f_0(\vec{x}) \text{ for every } \vec{x} \in \mathcal{X}. \tag{6}$$

It may not be convincing that the above constraints are linear. To see why, we consider the restricted case that $\mathcal{X} = \{1, 2, 3\}$; then our reformulation becomes

$$p^\star = \max_{t} \quad t \tag{7}$$

$$\text{s.t.} \quad t \le f_0(1) \tag{8}$$

$$t \le f_0(2) \tag{9}$$

$$t \leq f_0(3) \tag{10}$$

which are just 3 separate linear constraints. Adding more linear constraints in this way (even uncountably many of them) doesn't change the fact that each constraint is linear.

(d) If any of the constraint inequalities is strict (and therefore not active) at the optimum point, then we can remove the constraint from the original problem and obtain the same optimum value.

Note: Review the definition of active constraints from the textbooks: Boyd Section 4.1.1 and El Ghaoui Section 8.3.

Hint: Consider the problem

$$\min_x \; f(x) \; = \begin{cases} x^2 & \text{if } |x| \leq 1, \\ -1 & \text{otherwise} \end{cases} \tag{11}$$

$$\text{such that} \;\; |x| \leq 1 \tag{12}$$

Solution: This is *not* true in general. Consider the problem

$$p^\star := \min_x \; f_0(x) \; : \; |x| \leq 1, \tag{13}$$

where

$$f_0(x) = \begin{cases} x^2 & \text{if } |x| \leq 1, \\ -1 & \text{otherwise.} \end{cases} \tag{14}$$

The constraint $|x| \leq 1$ is not active at the optimum $x^\star = 0$, and $p^\star = 0$. However, if we remove it, the new optimal value becomes -1.

(e) Now, suppose for the formulation in (1), $f_0(\vec{x})$ is a convex function, $\mathcal{X}$ is a convex set, and all $f_i(\vec{x})$ are convex and continuous functions. Suppose $p^\star$ is achieved at a point $\vec{x}^\star$ where for some i, $f_i(\vec{x}^\star) < 0$. Prove that we can remove this inequality constraint and still retain the same optimum. In other words, show that

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \; f_0(\vec{x}) \; : \; f_j(\vec{x}) \leq 0, \;\; j = 1, \ldots i-1, i+1, \ldots, m. \tag{15}$$

HINT: Argue by contradiction that if by removing the inequality constraint $f_i(\vec{x}) \leq 0$, we achieve a different optimal for the problem in (1) at some point $\vec{\bar{x}}$ that satisfies $f_i(\vec{\bar{x}}) > 0$, then these exists a point $\vec{y}$ between $\vec{\bar{x}}$ and $\vec{x}^\star$ that is feasible to the original problem in (1). Use the continuity of f_i and the intermediate value theorem to come up with a $\vec{y}$ then show that it must be more optimal than $\vec{x}^\star$ in (1).

Solution: Note: The first version of this problem didn't have the assumption that the f_i are convex. If you tried to solve this problem without the convexity assumption, do not penalize yourself in self-grades, feel free to give yourself full points. For edification's sake, without the convexity assumption, the claim the problem asks you to prove is false.

Suppose for contradiction that by removing the inequality constraint $f_i(\vec{x}) \leq 0$, we achieve a more optimal result (we can not achieve a less optimal result by relaxing the feasible set). Concretely, let

$$\vec{\bar{x}} \in \operatorname*{argmin}_{\vec{x} \in \mathbb{R}^n} \; f_0(\vec{x}) \; : \; f_j(\vec{x}) \leq 0, \;\; j = 1, \ldots i-1, i+1, \ldots, m. \tag{16}$$

such that

$$\vec{\bar{x}} \notin \operatorname*{argmin}_{\vec{x} \in \mathbb{R}^n} \; f_0(\vec{x}) \; : \; f_i(\vec{x}) \leq 0, \;\; i = 1, \ldots, m \tag{17}$$

Then, we must have that $f_i(\vec{\bar{x}}) > 0$. Since $f_i(\vec{x}^\star) < 0$, by the continuity of f_i, there must exist some $\theta \in [0, 1]$ such that $f_i(\theta\vec{x}^\star + (1-\theta)\vec{\bar{x}}) = 0$.

By the convexity of f_0, we have that

$$f_0(\theta\vec{x}^\star + (1-\theta)\vec{\bar{x}}) \leq \theta f_0(\vec{x}^\star) + (1-\theta) f_0(\vec{\bar{x}}) \quad (18)$$

$$< f_0(\vec{x}^\star) \quad (19)$$

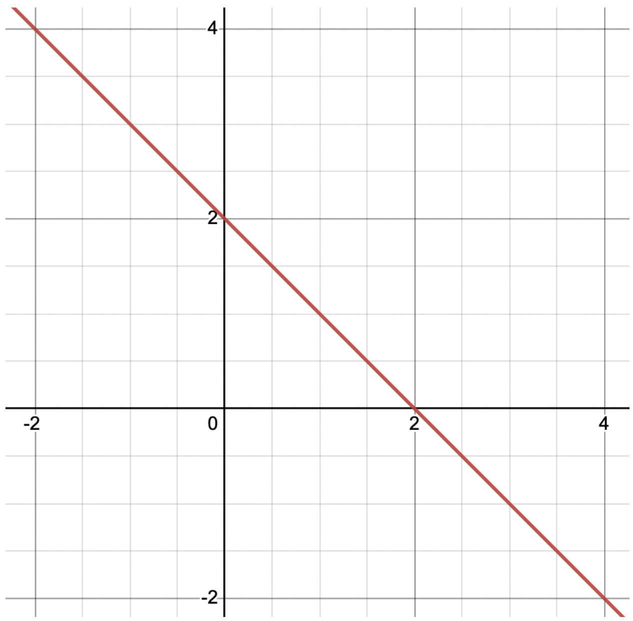
This contradicts our assumption that $p^\star$ is achieved at $\vec{x}^\star$ since $\theta\vec{x}^\star + (1-\theta)\vec{\bar{x}}$ is more optimal.

2. Fun with Hyperplanes

In this problem we work with hyperplanes, which are key components of linear programming as well as future topics such as support vector machines.

(a) Sketch the hyperplane $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^2 \mid \begin{bmatrix} 1 & 1 \end{bmatrix} \vec{x} = 2\}$.

Solution: See the following figure:



(b) Let $\vec{c} \in \mathbb{R}^n$ be nonzero, and let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top \vec{x} = 0\}$. Show that $\mathcal{H}$ is a linear subspace of $\mathbb{R}^n$. What is $\dim(\mathcal{H})$?

Solution: We have $\mathcal{H} = \mathcal{N}(\vec{c}^\top)$, where $\vec{c}^\top$ is interpreted as a $1 \times n$ matrix. Thus it is a linear subspace. By rank-nullity, we have $\dim(\mathcal{R}(\vec{c}^\top)) + \dim(\mathcal{N}(\vec{c}^\top)) = n$, and $\dim(\mathcal{R}(\vec{c}^\top)) = \dim(\mathcal{R}(\vec{c})) = 1$, so $\dim(\mathcal{H}) = \dim(\mathcal{N}(\vec{c}^\top)) = n - 1$.

Alternatively, you could show that $\mathcal{H}$ is closed under linear combination.

(c) Let $\vec{c} \in \mathbb{R}^n$ be nonzero, and let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top \vec{x} = 0\}$. Suppose $\vec{x}_\star \in \mathbb{R}^n$ is on one side of the hyperplane, i.e., $\vec{c}^\top \vec{x}_\star > 0$. Give any vector which is on the other side of the hyperplane but not on the hyperplane itself.

Solution: We propose the vector $-\vec{x}_\star$. Indeed, we have

$$\vec{c}^\top(-\vec{x}_\star) = -\vec{c}^\top \vec{x}_\star < 0. \tag{20}$$

Thus $-\vec{x}_\star$ is on the other side of the plane.

(d) Let $\vec{c} \in \mathbb{R}^n$ be nonzero, and let $\vec{x}_0 \in \mathbb{R}^n$ be arbitrary. Let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top(\vec{x} - \vec{x}_0) = 0\}$. Suppose $\vec{x}_\star \in \mathbb{R}^n$ is on one side of the hyperplane. Give any vector which is on the other side of the hyperplane but not on the hyperplane itself.

Solution: Suppose that without loss of generality we have $\vec{c}^\top(\vec{x}_\star - \vec{x}_0) > 0$. Then we have

$$\vec{c}^\top(\vec{x}_\star - \vec{x}_0) > 0 \implies \vec{c}^\top(\vec{x}_0 - \vec{x}_\star) < 0. \tag{21}$$

We want to find $\vec{z}$ such that $\vec{z} - \vec{x}_0 = \vec{x}_0 - \vec{x}_\star$. By algebra, $\vec{z} = 2\vec{x}_0 - \vec{x}_\star$. This gives $\vec{c}^\top(\vec{z} - \vec{x}_0) = \vec{c}^\top(\vec{x}_0 - \vec{x}_\star) < 0$, so $\vec{z}$ is the vector we want.

(e) Let $\vec{x}_0 \in \mathbb{R}^n$ be arbitrary. For a vector $\vec{c} \in \mathbb{R}^n$, let $\mathcal{H}(\vec{c}) \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top(\vec{x} - \vec{x}_0) = 0\}$. Show that $\vec{0} \in \mathcal{H}(\vec{c})$ for every $\vec{c} \in \mathbb{R}^n$ if and only if $\vec{x}_0 = \vec{0}$.

Solution: We first claim that, for a fixed $\vec{c} \in \mathbb{R}^n$, that $\vec{0} \in \mathcal{H}(\vec{c})$ if and only if $\vec{c}$ is orthogonal to $\vec{x}_0$. Indeed,

$$\begin{aligned} \vec{c}^\top \vec{x}_0 = 0 &\iff -\vec{c}^\top \vec{x}_0 = 0 && (22) \\ &\iff \vec{c}^\top \vec{0} - \vec{c}^\top \vec{x}_0 = 0 && (23) \\ &\iff \vec{c}^\top (\vec{0} - \vec{x}_0) = 0 && (24) \\ &\iff \vec{0} \in \mathcal{H}(\vec{c}). && (25) \end{aligned}$$

Thus $\vec{0} \in \mathcal{H}(\vec{c})$ for every $\vec{c} \in \mathbb{R}^n$ if and only if $\vec{x}_0$ is orthogonal to every $\vec{c} \in \mathbb{R}^n$. But this is equivalent to $\vec{x}_0 = \vec{0}$, and the claim is proved.

3. Duality

Consider the function

$$f(\vec{x}) = \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x}. \tag{26}$$

First, we consider the unconstrained optimization problem

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) = \min_{\vec{x}\in\mathbb{R}^n} \vec{x}^\top A\vec{x} - 2\vec{b}^\top\vec{x} \tag{27}$$

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,

$$\nabla_{\vec{x}} f(\vec{x}) = 2A\vec{x} - 2\vec{b} = 0 \tag{28}$$
$$\implies A\vec{x} = \vec{b} \tag{29}$$
$$\implies \vec{x}^\star = A^{-1}\vec{b} \tag{30}$$

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,

$$f(\vec{x}^\star) = (\vec{b}^\top (A^{-1})^\top)A(A^{-1}\vec{b}) - 2\vec{b}^\top (A^{-1}\vec{b}) \tag{31}$$
$$= \vec{b}^\top (A^\top)^{-1}\cancel{AA^{-1}}\vec{b} - 2\vec{b}^\top A^{-1}\vec{b} \tag{32}$$
$$= \vec{b}^\top A^{-1}\vec{b} - 2\vec{b}^\top A^{-1}\vec{b} \tag{33}$$
$$p^\star = -\vec{b}^\top A^{-1}\vec{b} \tag{34}$$

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

$$A = U_r\Lambda_r U_r^\top = \begin{bmatrix} U_r & U_1 \end{bmatrix} \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^\top \\ U_1^\top \end{bmatrix}. \tag{35}$$

Show that the minimizer $\vec{x}^\star$ of the optimization problem (27) 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,

$$A\vec{x} = b, \tag{36}$$

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 (36) and get another solution.

By the Fundamental Theorem of Linear Algebra we have,

$$\vec{x} = U_r\vec{\alpha} + U_1\vec{\beta} \tag{37}$$

$$\vec{b} = U_r\vec{\gamma}, \tag{38}$$

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

$$U_r\Lambda_r U_r^\top (U_r\vec{\alpha} + U_1\vec{\beta}) = U_r\vec{\gamma} \tag{39}$$

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

$$U_r\Lambda_r U_r^\top U_r\vec{\alpha} = U_r\vec{\gamma} \tag{40}$$

$$\implies \vec{\alpha} = \Lambda_r^{-1}\vec{\gamma} \tag{41}$$

$$= \Lambda_r^{-1}U_r^\top\vec{b}. \tag{42}$$

Thus any solution to Equation (36) and hence a minimizer to the optimization problem (27) can be written as,

$$\vec{x}^\star = U_r\Lambda_r^{-1}U_r^\top\vec{b} + U_1\vec{\beta}. \tag{43}$$

(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

$$\vec{v}^\top A\vec{v} = \mu < 0. \tag{44}$$

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,

$$f(\vec{x}) = f(\alpha\vec{v}) = \alpha^2\vec{v}^\top A\vec{v} + \alpha(-2\vec{b}^\top\vec{v}), \tag{45}$$

which goes to $-\infty$ as α goes to ∞ 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:

$$f(\alpha\vec{v}) = \alpha^2\vec{v}^\top A\vec{v} - 2\alpha(\vec{v}_1 + \vec{v}_2)^\top\vec{v}_2 = 0 - 2\alpha\left\|\vec{v}_2\right\|^2 \tag{46}$$

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 $0 \leq r \leq n$. Now we consider the constrained optimization problem

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \; & \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} \\ \text{s.t. } & \vec{x}^\top \vec{x} \geq 1. \end{aligned} \tag{47}$$

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

Solution:

$$\mathcal{L}(\vec{x}, \lambda) = \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} + \lambda(1 - \vec{x}^\top \vec{x}) \tag{48}$$

$$= \vec{x}^\top A\vec{x} - \vec{x}^\top \lambda\vec{x} - 2\vec{b}^\top \vec{x} + \lambda \tag{49}$$

$$= \vec{x}^\top (A - \lambda I)\vec{x} - 2\vec{b}^\top \vec{x} + \lambda \tag{50}$$

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

$$C = U_r\Lambda_r V_r^\top, \tag{51}$$

we define the pseudoinverse as

$$C^\dagger = V_r\Lambda_r^{-1}U_r^\top. \tag{52}$$

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 (47) can be written as,

$$d^\star = \max_{\substack{\lambda \geq 0 \\ A - \lambda I \succeq 0 \\ \vec{b} \in \mathcal{R}(A-\lambda I)}} \; -\vec{b}^{\,\top} (A - \lambda I)^\dagger \, \vec{b} + \lambda. \tag{53}$$

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:

$$g(\lambda) = \min_{\vec{x}} \mathcal{L}(\vec{x}, \lambda) = \min_{\vec{x}} \vec{x}^\top (A - \lambda I)\vec{x} - 2\vec{b}^\top \vec{x} + \lambda \tag{54}$$

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:

$$\nabla_{\vec{x}}\mathcal{L} = 2(A - \lambda I)\vec{x} - 2\vec{b} = 0 \tag{55}$$

$$(A - \lambda I)\vec{x} = \vec{b} \tag{56}$$

$$\vec{x}^\star = (A - \lambda I)^\dagger\vec{b} \tag{57}$$

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:

$$\begin{aligned}
\mathcal{L}(\vec{x}^\star, \lambda) &= \vec{b}^\top ((A - \lambda I)^\dagger)^\top (A - \lambda I)(A - \lambda I)^\dagger \vec{b} - 2\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda && (58)\\
&= \vec{b}^\top (A - \lambda I)^\dagger (A - \lambda I)(A - \lambda I)^\dagger \vec{b} - 2\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda && (59)\\
&= \vec{b}^\top (A - \lambda I)^\dagger \vec{b} - 2\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda && (60)\\
&= -\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda && (61)
\end{aligned}$$

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

$$(A - \lambda I)^\dagger (A - \lambda I)(A - \lambda I)^\dagger = (A - \lambda I)^\dagger. \tag{62}$$

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

$$g(\lambda) = \begin{cases} -b^\top (A - \lambda I)^\dagger b + \lambda & \text{if } A - \lambda I \succeq 0, b \in \mathcal{R}(A - \lambda I) \\ -\infty & \text{else} \end{cases} \tag{63}$$

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

$$d^\star = \max_{\lambda \geq 0} g(\lambda) \tag{64}$$

4. A Slalom Problem

A skier must slide from left to right by going through n parallel gates of known position (x_i, y_i) and width c_i, $i = 1, \ldots, n$. The initial position (x_0, y_0) is given, as well as the final one, (x_{n+1}, y_{n+1}). Before reaching the final position, the skier must go through gate i by passing between the points $(x_i, y_i - c_i/2)$ and $(x_i, y_i + c_i/2)$ for each $i \in \{1, \ldots, n\}$.

Figure 1 is an example and does not have the right value of n nor show the true (x_i, y_i, c_i) values. Use values for (x_i, y_i, c_i) from Table 1.

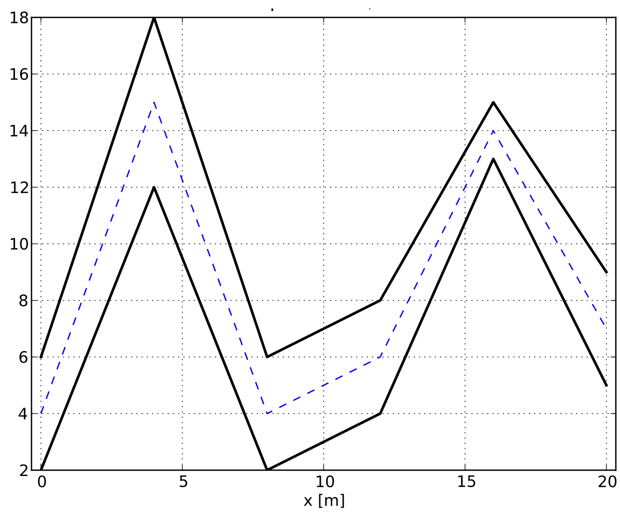


Figure 1: Slalom problem with $n = 6$ gates. The initial and final positions are fixed and not included in the figure. The skier slides from left to right. The middle path is dashed and connects the center points of gates.

Table 1: Problem data for Problem 2. Here $n = 5$.

i	x_i	y_i	c_i
0	0	4	N/A
1	4	5	3
2	8	4	2
3	12	6	2
4	16	8	1
5	20	7	2
6	24	4	N/A

(a) Given the data $\{(x_i, y_i, c_i)\}_{i=0}^{n+1}$, write an optimization problem that minimizes the total length of the path. Your answer should come in the form of an SOCP.

Solution: Assume that (x_i, z_i) is the crossing point of gate i, the path length minimization problem is thus

$$\min_{\vec{z}} \quad \sum_{i=1}^{n+1} \left\| \begin{bmatrix} x_i \\ z_i \end{bmatrix} - \begin{bmatrix} x_{i-1} \\ z_{i-1} \end{bmatrix} \right\|_2 \tag{65}$$

$$\text{s.t.} \quad y_i - c_i/2 \le z_i \le y_i + c_i/2, \text{ for } i = 1, \dots, n \tag{66}$$

$$z_0 = y_0, z_{n+1} = y_{n+1}, \tag{67}$$

which is equivalent to

$$\min_{\vec{z}, \vec{t}} \quad \sum_{i=1}^{n+1} t_i \tag{68}$$

$$\text{s.t.} \quad y_i - c_i/2 \le z_i \le y_i + c_i/2, \text{ for } i = 0, \dots, n+1 \tag{69}$$

$$\left\| \begin{bmatrix} x_i \\ z_i \end{bmatrix} - \begin{bmatrix} x_{i-1} \\ z_{i-1} \end{bmatrix} \right\|_2 \le t_i, \text{ for } i = 1, \dots, n+1. \tag{70}$$

with the convention $c_0 = c_{n+1} = 0$. Hence, the problem is an SOCP.

(b) Solve the problem numerically with the data given in Table 1. *HINT: You should be able to use packages such as `cvxpy` and `numpy`.*

Solution: The code can be found in the corresponding Jupyter notebook.

5. **Formulating Optimization problems**

(a) **Linear Separability.** Let $(\vec{x}_i, y_i)$ be given data points with $\vec{x}_i \in \mathbb{R}^n$ and binary labels $y_i \in \{-1, 1\}$. We want to know if it is possible to find a hyperplane $\mathcal{L} = \{\vec{x} \in \mathbb{R}^n : \vec{h}^\top \vec{x} + b = 0\}$ that separates all the points with labels $y_i = -1$ from all the points with labels $y_i = 1$. In other words, can we find a vector $\vec{h} \in \mathbb{R}^n$ and a scalar $b \in \mathbb{R}$ such that $\vec{h}^\top \vec{x}_i + b \leq 0$ for all i such that $y_i = 1$ and $\vec{h}^\top \vec{x}_i + b > 0$ for all i such that $y_i = -1$. We want to cast this task as the following LP

$$p^\star = \min_{\vec{h}, b, z} \quad f_0(\vec{h}, b, z) \tag{71}$$
$$s.t. \quad \vec{h}^\top \vec{x}_i + b \leq 0 \qquad \forall i : y_i = 1 \tag{72}$$
$$\vec{h}^\top \vec{x}_i + b \geq z \qquad \forall i : y_i = -1 \tag{73}$$

Complete this formulation by specifying a linear objective function f_0. What does the solution $p^\star$ say about the existence of the separating hyperplane?

Solution: With the choice of the objective function

$$f_0(\vec{h}, b, z) = -z, \tag{74}$$

the separating hyperplane exists if $p^\star < 0$.

To see this note that $p^\star < 0$ if and only if the optimal solution to the problem $(\vec{h}^\star, b^\star, z^\star)$ is such that $z^\star > 0$. Now consider the hyperplane $\mathcal{L}^\star = \{\vec{x} \in \mathbb{R}^n : \vec{h}^{\star\top} \vec{x} + b^\star = 0\}$. From feasibility of the optimal solution, we can see that

$$\vec{h}^{\star\top} \vec{x}_i + b^\star \leq 0 \qquad \forall i : y_i = 1 \tag{75}$$
$$\vec{h}^{\star\top} \vec{x}_i + b^\star \geq z^\star > 0 \qquad \forall i : y_i = -1. \tag{76}$$

which is the definition of linear separability. Thus $\mathcal{L}^\star$ is indeed a separating hyperplane for this data.

(b) **Chebyshev Center.** Let $\mathcal{P} \subset \mathbb{R}^n$ be a non-empty polyhedron defined as the intersection of m hyperplanes $\mathcal{P} = \{\vec{x} : \vec{a}_i^\top \vec{x} \leq b_i \; \forall i = 1, 2, \ldots, m\}$. We define the Euclidean ball in $\mathbb{R}^n$ with radius R and center $\vec{x}_0$ as the set $\mathcal{B}(\vec{x}_0, R) = \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{x}_0\|_2 \leq R\}$. We want to find a point $\vec{x}_0 \in \mathcal{P}$ that is the center of the largest Euclidean ball contained in $\mathcal{P}$. Cast this problem as an LP.

Solution: Any point $\vec{x} \in \mathcal{B}(\vec{x}_0, R)$ can we expressed as $\vec{x} = \vec{x}_0 + \vec{u}$ where $\|\vec{u}\|_2 \leq R$. To satisfy the condition that $\mathcal{B}(\vec{x}_0, R) \subset \mathcal{P}$ we need for all $\vec{u} \in \mathbb{R}^n$ with norm $\|\vec{u}\|_2 \leq R$:

$$\vec{a}_i^\top (\vec{x}_0 + \vec{u}) \leq b_i \qquad \forall i = 1, 2, \ldots, m \tag{77}$$

We take the maximum over $\vec{u}$ of both sides to get the equivalent condition

$$\max_{\|\vec{u}\|_2 \leq R} \left(\vec{a}_i^\top (\vec{x}_0 + \vec{u})\right) \leq b_i \tag{78}$$
$$\vec{a}_i^\top \vec{x}_0 + \max_{\|\vec{u}\|_2 \leq R} \left(\vec{a}_i^\top \vec{u}\right) \leq b_i \qquad \forall i = 1, 2, \ldots, m \tag{79}$$

The inner product $\vec{a}_i^\top \vec{u}$ is maximized when $\vec{u}$ is the longest possible vector along the direction of $\vec{a}_i$, thus

$$\max_{\|\vec{u}\|_2 \leq R} \left(\vec{a}_i^\top \vec{u}\right) = \vec{a}_i^\top \left(\frac{R}{\|\vec{a}_i\|_2} \vec{a}_i\right) \tag{80}$$

$$= R \|\vec{a}_i\|_2 \tag{81}$$

This gives the following conditions

$$\vec{a}_i^\top \vec{x}_0 + R \|\vec{a}_i\|_2 \leq b_i \qquad \forall i = 1, 2, \ldots, m \tag{82}$$

Now we can write the problem of finding the largest ball enclosed in $\mathcal{P}$ as

$$\min_{\vec{x}_0, R} -R \tag{83}$$

$$s.t. \quad \vec{a}_i^\top \vec{x}_0 + R \|\vec{a}_i\|_2 \leq b_i \qquad \forall i = 1, 2, \ldots, m \tag{84}$$

6. Dual Norms and SOCP

Consider the problem

$$p^\star = \min_{\vec{x}\in\mathbb{R}^n} \|A\vec{x}-\vec{y}\|_1 + \mu\|\vec{x}\|_2, \tag{85}$$

where $A \in \mathbb{R}^{m\times n}$, $\vec{y}\in\mathbb{R}^m$, and $\mu > 0$.

(a) Express this (primal) problem in standard SOCP form.

Solution: Introducing slack variables $\vec{z}\in\mathbb{R}^m$, $t\in\mathbb{R}$, we can write

$$\min_{\vec{x},\vec{z},t} \vec{z}^\top\vec{1} + \mu t \tag{86}$$

$$\text{s.t. } |(A\vec{x}-\vec{y})_i| \le z_i, \quad i = 1,\ldots,m \tag{87}$$

$$\|\vec{x}\|_2 \le t, \tag{88}$$

which can be written as

$$\min_{\vec{x},\vec{z},t} \vec{z}^\top\vec{1} + \mu t \tag{89}$$

$$\text{s.t. } \left|\vec{a}_i^\top\vec{x} - y_i\right| \le z_i, \quad i = 1,\ldots,m \tag{90}$$

$$\|\vec{x}\|_2 \le t, \tag{91}$$

where $\vec{a}_i$ are the rows of A and y_i are the entries of $\vec{y}$.

This expression now satisfies our definition of an SOCP: the objective is linear, and all constraints are SOC constraints, of the form

$$\left\|\begin{bmatrix} A_i & B_i & C_i\end{bmatrix}\begin{bmatrix}\vec{x}\\ \vec{z}\\ t\end{bmatrix} - \vec{d}_i\right\|_2 \le \begin{bmatrix}\vec{f}_i\\ \vec{g}_i\\ h_i\end{bmatrix}^\top\begin{bmatrix}\vec{x}\\ \vec{z}\\ t\end{bmatrix} + k_i. \tag{92}$$

For the first m constraints, we just take $A_i = \vec{a}_i^\top$, $B_i = 0$, $C_i = 0$, $\vec{d}_i = y_i$, $\vec{f}_i = \vec{0}$, $\vec{g}_i = \vec{e}_i$ (i.e., the i^{th} standard basis vector), $h_i = 0$, and $k_i = 0$. For the last constraint, we take $A_i = I$, $B_i = 0$, $C_i = 0$, $\vec{d}_i = \vec{0}$, $\vec{f}_i = \vec{0}$, $\vec{g}_i = \vec{0}$, $h_i = 1$, and $k_i = 0$.

(b) Find a dual to the problem and express it in standard SOCP form. *HINT: Recall that for every vector $\vec{z}$, the following dual norm equalities hold:*

$$\|\vec{z}\|_2 = \max_{\vec{u}\,:\,\|\vec{u}\|_2\le 1} \vec{u}^\top\vec{z}, \qquad \|\vec{z}\|_1 = \max_{\vec{u}\,:\,\|\vec{u}\|_\infty\le 1}\vec{u}^\top\vec{z}. \tag{93}$$

Solution: Using the hint, we can rewrite the objective function of the original problem as

$$\|A\vec{x}-\vec{y}\|_1 + \mu\|\vec{x}\|_2 = \max_{\vec{u}:\|\vec{u}\|_\infty\le 1}\vec{u}^\top(A\vec{x}-\vec{y}) + \mu\max_{\vec{v}:\|\vec{v}\|_2\le 1}\vec{v}^\top\vec{x}. \tag{94}$$

We can then express the original (primal) problem as

$$p^\star = \min_{\vec{x}}\max_{\vec{u},\vec{v}:\|\vec{u}\|_\infty\le 1,\|\vec{v}\|_2\le 1}\vec{u}^\top(A\vec{x}-\vec{y}) + \mu\vec{v}^\top\vec{x}. \tag{95}$$

To form the dual, we reverse the order of min and max:

$$d^\star = \max_{\vec{u},\vec{v}:\|\vec{u}\|_\infty\le 1,\|\vec{v}\|_2\le 1}\min_{\vec{x}}\vec{u}^\top(A\vec{x}-\vec{y}) + \mu\vec{v}^\top\vec{x} \tag{96}$$

$$\doteq \max_{\vec{u},\vec{v}:\|\vec{u}\|_\infty \le 1, \|\vec{v}\|_2 \le 1} g(\vec{u},\vec{v}), \tag{97}$$

where g is defined as

$$g(\vec{u},\vec{v}) \doteq \min_{\vec{x}} \vec{u}^\top (A\vec{x} - \vec{y}) + \mu \vec{v}^\top \vec{x} \tag{98}$$

$$= \min_{\vec{x}} (\vec{u}^\top A + \mu \vec{v}^\top)\vec{x} - \vec{u}^\top \vec{y} \tag{99}$$

$$= \begin{cases} -\vec{u}^\top \vec{y} & \text{if } A^\top \vec{u} + \mu \vec{v} = \vec{0} \\ -\infty & \text{otherwise.} \end{cases} \tag{100}$$

We can thus rewrite the dual problem as

$$d^\star = \max_{\vec{u},\vec{v}} \quad -\vec{u}^\top \vec{y} \tag{101}$$

$$\text{s.t.} \quad A^\top \vec{u} + \mu \vec{v} = \vec{0} \tag{102}$$

$$\|\vec{u}\|_\infty \le 1, \quad \|\vec{v}\|_2 \le 1. \tag{103}$$

Noting that the first constraint fully restricts the value of $\vec{v}$ — rewriting it, $\vec{v} = -\frac{A^\top \vec{u}}{\mu}$ — we can plug this value into the third constraint and eliminate $\vec{v}$ from our optimization altogether:

$$d^\star = \max_{\vec{u}} \quad -\vec{u}^\top \vec{y} \tag{104}$$

$$\text{s.t.} \quad \left\|A^\top \vec{u}\right\|_2 \le \mu \tag{105}$$

$$\|\vec{u}\|_\infty \le 1, \tag{106}$$

generating our final SOCP dual. If desired, we can further rewrite the final constraint as $\|\vec{u}\|_\infty = \max_i |u_i| \le 1 \Leftrightarrow |u_i| \le 1, \ i = 1, \ldots, m \Leftrightarrow u_i \le 1$ and $u_i \ge -1, \ i = 1, \ldots, m$ to make the linearity of that constraint more explicit.

(c) Assume strong duality holds[1] and that $m = 100$ and $n = 10^6$, i.e., A is 100×10^6. Which problem would you choose to solve using a numerical solver: the primal or the dual? Justify your answer.

Solution: To determine the rough computational complexity of each problem, we examine the number of variables and the number of constraints in each problem. The primal SOCP has $\sim 10^6$ variables and 201 constraints, while the dual has 100 variables and 201 constraints. The dual problem is thus much more efficient to solve.

[1]In fact, you can show that strong duality holds using Sion's theorem, a generalization of the minimax theorem that is beyond the scope of this class.

7. 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.