Homework 9

Self grades are due at 11 PM on April 6, 2023.

1. Quadratic inequalities

Consider the set S defined by the following inequalities:

$$(x_1 \geq -x_2 + 1 \text{ and } x_1 \leq 0) \text{ or } (x_1 \leq -x_2 + 1 \text{ and } x_1 \geq 0)\,. \tag{1}$$

To be more precise,

$$S_1 = \{\vec{x} \in \mathbb{R}^2 \mid x_1 \geq -x_2 + 1, x_1 \leq 0\} \tag{2}$$

$$S_2 = \{\vec{x} \in \mathbb{R}^2 \mid x_1 \leq -x_2 + 1, x_1 \geq 0\} \tag{3}$$

$$S = S_1 \cup S_2. \tag{4}$$

(a) Draw the set S. Is it convex?

Solution:

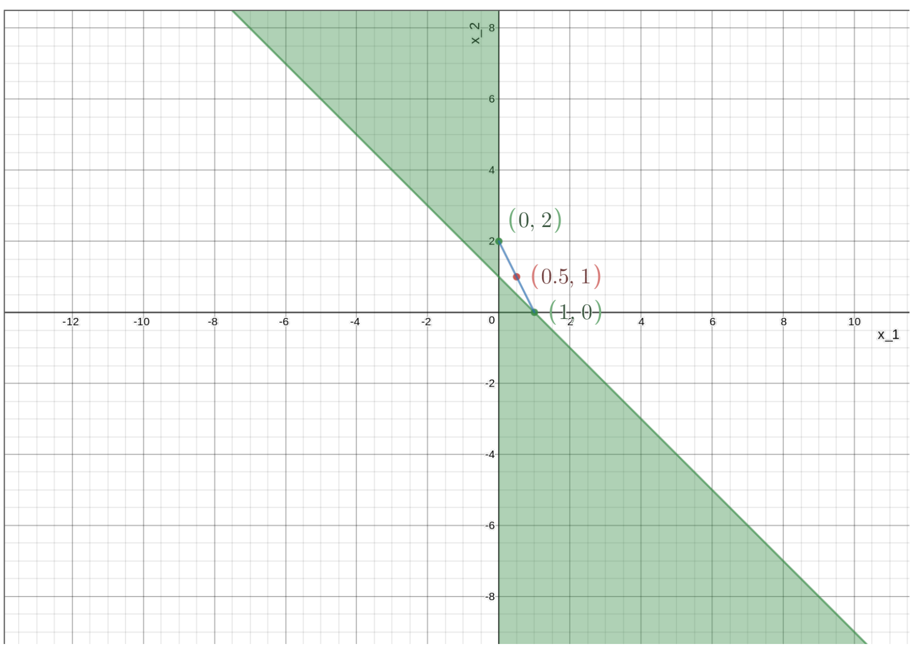

Figure 1: Set S

The set S as shown in Fig. 1 is not convex. We can prove this by counterexample. $(0, 2)$ and $(1, 0)$ both belong to the set, but the midpoint $(1/2, 1)$ does not.

(b) Show that the set S, can be described as a single quadratic inequality of the form $q(\vec{x}) = \vec{x}^\top A\vec{x} + 2\vec{b}^\top\vec{x} + c \leq 0$, for matrix $A = A^\top \in \mathbb{R}^{2\times 2}$, $\vec{b} \in \mathbb{R}^2$ and $c \in \mathbb{R}$ i.e S can be written as $S = \{\vec{x} \in \mathbb{R}^2 \mid q(\vec{x}) \leq 0\}$). Find $A, \vec{b}, c$.

Hint: Can you combine the constraints to make one quadratic constraint?

Solution: Within set S, $x_1 + x_2 - 1 \geq 0$ when $x_1 \leq 0$ and $x_1 + x_2 - 1 \leq 0$ when $x_1 \geq 0$. It follows that $q(\vec{x}) = x_1(x_1 + x_2 - 1) \leq 0$ if and only if it is in the set. Expressing $q(\vec{x})$ in the desired form:

$$q(\vec{x}) = x_1^2 + x_1 x_2 - x_1 = \vec{x}^\top A\vec{x} + 2\vec{b}^\top \vec{x} + c$$

where

$$A = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -1/2 \\ 0 \end{bmatrix}, \quad c = 0.$$

(c) Give the definition of the convex hull of a set. What is the convex hull of this set, i.e., S?

Solution: The convex hull of the set is the whole space, $\mathbb{R}^2$. To see this note than any point $z = (z_1, z_2) \in \mathbb{R}^2$ can be written as $z = \frac{x+y}{2}$ with $x, y \in S$ as follows:
$x = (2z_1, 1 - 2z_1), y = (0, 2(z_1 + z_2) - 1)$.

(d) We will now consider some convex optimization problems over S_1 that illustrate the role of the constraints in the optimization problem. For each of the following optimization problems find the optimal point, $\vec{x}^*$. Describe the constraints that are active in attaining the optimal value. *Hint: Suppose that there exists a point $\vec{x}$ such that $\nabla f(\vec{x}) = 0$. From the first order characterization of a convex function $\vec{x}$ would be an optimum value for f subject to no constraints. If $\vec{x}$ is not in the constraint set S_1, then the optimum point must be on the boundary of the set, i.e. it satisfies at least one of the constraints defining S_1 with equality.*

i. Minimize $f(\vec{x}) = (x_1 + 1)^2 + (x_2 - 3)^2$ subject to $\vec{x} \in S_1$.

Solution: We first compute the unconstrained optimal value of f. Notice that f is a convex function. Therefore, we can compute its optimal value by computing its gradient and setting it to 0. Doing so, we obtain the optimal value of f to be 0 attained at the point $\vec{x}^* = (-1, 3)$. Now, since $\vec{x}^* \in S_1$, $\vec{x}^*$ is the solution to the constrained optimization problem as well.

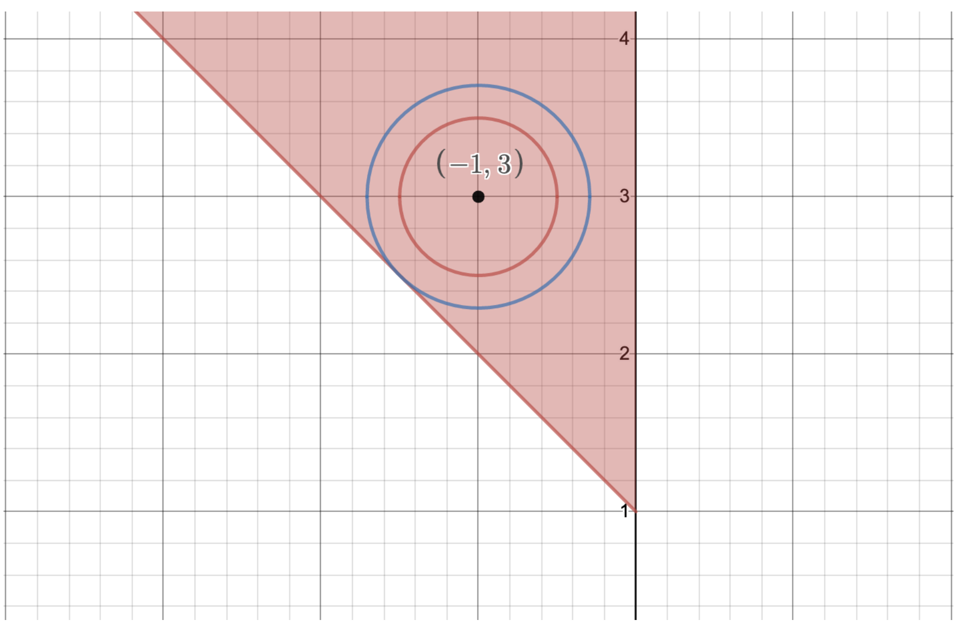

Figure 2: This figure illustrates the position of the optimum, $x^* = (-1, 3)$, and the level sets of the objective function, f, which are concentric circles around x^*.

ii. Minimize $f(\vec{x}) = (x_1+2)^2 + (x_2-2)^2$ subject to $\vec{x} \in S_1$.

Solution: Proceeding as in the proof for the previous problem, we first find the solution to the unconstrained optimization problem. We get that the unconstrained problem is minimized at the point $\vec{x}_u^* = (-2, 2)$. However, this point is not in the feasible set, S_1. Therefore, the true optimum, $\vec{x}^*$, has one or more constraints active. Now, we will attempt to solve the problem with one active constraint. Suppose the one active constraint is $x_1 \geq -x_2 + 1$. Since this constraint is active, we must try and minimize $f(\vec{x})$ subject to $\vec{x}$ satisfying $x_1 = -x_2 + 1$. Note that any point on this line can be written in the form $(0, 1) + \alpha(-1, 1)$. Now consider the function, $g(\alpha)$:

$$g(\alpha) = f((0,1) + \alpha(-1,1)) = (\alpha - 2)^2 + (\alpha - 1)^2.$$

Note that the function, $f(\alpha)$, is convex in α. Therefore, we can minimize $g(\alpha)$ by taking its derivative and setting it to 0. By doing this, we get that $\alpha = 3/2$ is the unique minimizer of $g(\alpha)$. Therefore, the minimizer of f subject to $x_1 = -x_2 + 1$ is the point $(-3/2, 5/2)$, and the function value is 0.5. Similarly, the minimizer of f assuming the second constraint, $x_1 \leq 0$, is active is obtained at the point $(0, 2)$, and the function value at this point is 4, which is higher than the value at $(-3/2, 5/2)$. The final possibility is that both constraints are active. However, the optimal value of f subject to both constraints being active will be greater than the value of f obtained at $(-3/2, 5/2)$ which is in S_1. Therefore, we get that $f(\vec{x})$ is minimized at the point $\vec{x}^* = (-3/2, 5/2)$ subject to $\vec{x} \in S_1$. There is one active constraint at $\vec{x}^*$.

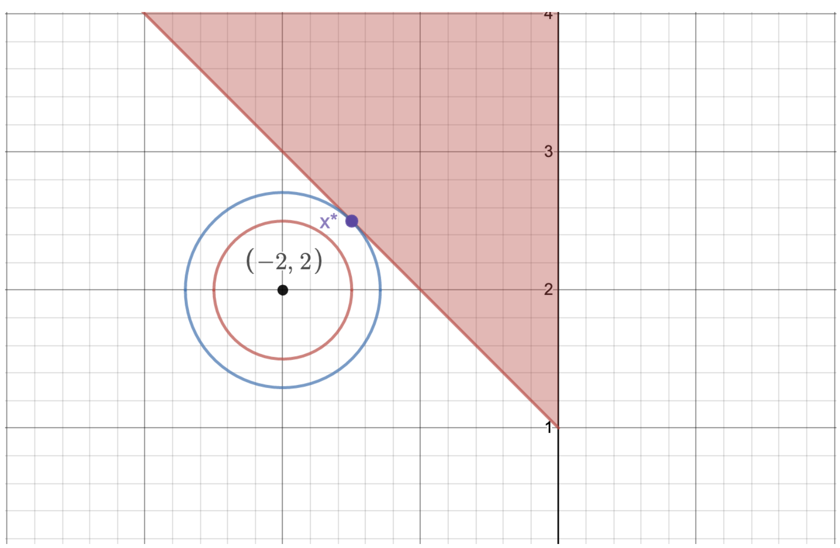

Figure 3: This figure illustrates the position of the optimum, $x^* = (-3/2, 5/2)$, and the level sets of the objective function, f, which are concentric circles around $(-2, 2)$. Note that in this case, the unconstrained optimum does not lie in the set, S_1 and the optimal point lies on the boundary of one of the constraints.

iii. Minimize $f(\vec{x}) = x_1^2 + x_2^2$ subject to $\vec{x} \in S_1$.

Solution: Proceeding as before, we first check the case where 0 constraints are active. However, the unconstrained minimizer of f is $(0, 0)$ which is not in S_1. Now, we check the cases where one of the constraints is active. Assume that the constraint $x_1 \leq 0$ is active. In this case the optimizer is again obtained at the point $(0, 0)$ which is not in S_1. We then consider the case where the constraint $x_1 \geq -x_2 + 1$ is active. As before, we define the function, $g(\alpha)$ as:

$$g(\alpha) = f((0,1) + \alpha(-1,1)) = \alpha^2 + (\alpha+1)^2.$$

By optimizing over α by setting its gradient with respect to α and setting it to 0, we get the optimal setting of α is $-1/2$. However, note that the point $(1/2, 1/2)$ does not belong to S_1 either. Therefore, the only remaining possibility is the possibility that both constraints are active. This can happen solely at the point $(0,1)$. At this point, the value of the function f is 1, the optimizer $\vec{x}^* = (0,1)$ and both constraints are active at $\vec{x}^*$.

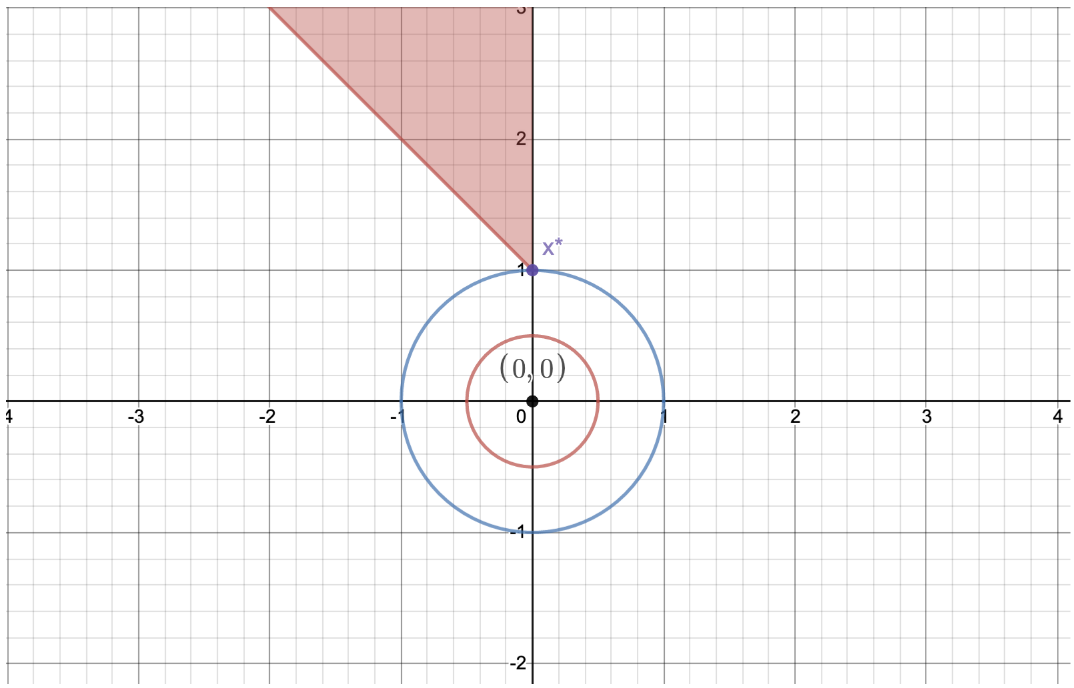

Figure 4: This figure illustrates the position of the optimum, $x^* = (0,1)$, and the level sets of the objective function, f, which are concentric circles around $(0,0)$. Note that in this case, the unconstrained optimum does not lie in the set, S_1 and the optimal point lies on the boundary of *both* of the constraints.

2. 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

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

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

$$\begin{aligned} F_0(\vec{x}, \vec{y}) &\geq \min_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}) && (6) \\ &\geq \min_{\vec{x} \in \mathcal{X}} \min_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}) && (7) \end{aligned}$$

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.,

$$\min_{\vec{x} \in \mathcal{X}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{x} \in \mathcal{X}} \min_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}). \tag{8}$$

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

$$\min_{\vec{y} \in \mathcal{Y}} \min_{\vec{x} \in \mathcal{X}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{x} \in \mathcal{X}} \min_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}). \tag{9}$$

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

$$\min_{\vec{x} \in \mathcal{X}} \min_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{y} \in \mathcal{Y}} \min_{\vec{x} \in \mathcal{X}} F_0(\vec{x}, \vec{y}). \tag{10}$$

Since both (9) and (10) must hold, the expressions must be equal, as desired.

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

$$\begin{aligned} p^\star &\doteq \min_{\vec{x} \in \mathcal{X}} \max_{\vec{y} \in \mathcal{Y}} F_0(\vec{x}, \vec{y}) && (11) \\ d^\star &\doteq \max_{\vec{y} \in \mathcal{Y}} \min_{\vec{x} \in \mathcal{X}} F_0(\vec{x}, \vec{y}). && (12) \end{aligned}$$

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

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

$$L(\vec{y}) \doteq \min_{\vec{x}'} F_0(\vec{x}', \vec{y}) \leq F_0(\vec{x}, \vec{y}) \leq U(\vec{x}) \doteq \max_{\vec{y}'} F_0(\vec{x}, \vec{y}') \tag{13}$$

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

$$L(\vec{y}) \leq U(\vec{x}). \tag{14}$$

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

$$p^\star = \min_{\vec{x} \in \mathcal{X}} U(\vec{x}) \geq L(\vec{y}). \tag{15}$$

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

$$p^\star \geq \max_{\vec{y} \in \mathcal{Y}} L(\vec{y}) = d^\star \tag{16}$$

as desired.

3. 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.)

4. Dual of the dual of a linear program

Consider a standard linear program P:

$$\begin{aligned} \min_{\vec{x}} \quad & \vec{c}^\top \vec{x} && (17)\\ \text{s.t.} \quad & A\vec{x} = \vec{b} && (18)\\ & x \geq 0. && (19) \end{aligned}$$

where $\vec{x}, \vec{c} \in \mathbb{R}^n, \vec{b} \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$.

(a) Formulate the Lagrangian of the problem P, and write the dual problem.

Note: The dual problem should not have the variable $\vec{x}$.

Solution: Denote the dual variable associated with the equality constraint in P by $\vec{\nu} \in \mathbb{R}^m$, and the dual variable associated with the inequality by $\vec{\lambda} \in \mathbb{R}^n$. The Lagrangian of P is given by

$$\mathcal{L}(\vec{x}, \vec{\nu}, \vec{\lambda}) = \vec{c}^\top \vec{x} + \vec{\nu}^\top (A\vec{x} - \vec{b}) + \lambda^\top(-\vec{x}) = -\vec{b}^\top \vec{\nu} + (\vec{c} + A^\top \vec{\nu} - \vec{\lambda})^\top \vec{x}. \tag{20}$$

To get the dual problem, we take the minimum of the Lagrangian with respect to our primal variable $\vec{x}$, i.e.,

$$g(\vec{\nu}) = \min_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\nu}, \vec{\lambda}) = \begin{cases} -\vec{b}^\top \vec{\nu} & \text{if } \vec{c} + A^\top \vec{\nu} - \vec{\lambda} = \vec{0} \\ -\infty & \text{otherwise.} \end{cases} \tag{21}$$

The dual problem is then given by D:

$$\begin{aligned} \max_{\vec{\nu}, \vec{\lambda}} \quad & -\vec{b}^\top \vec{\nu} && (22)\\ \text{s.t.} \quad & \vec{c} + A^\top \vec{\nu} - \vec{\lambda} = \vec{0} && (23)\\ & \vec{\lambda} \geq \vec{0}. && (24) \end{aligned}$$

We can simplify this further by noting that $\vec{\lambda}$ is constrained to be $\vec{\lambda} \geq \vec{0}$, since it is the dual variable associated with an inequality constraint. Then, $\vec{c} + A^\top \vec{\nu} - \vec{\lambda} = \vec{0} \iff \vec{c} + A^\top \vec{\nu} = \vec{\lambda} \geq \vec{0}$, and we can eliminate $\vec{\lambda}$ to get

$$\begin{aligned} \max_{\vec{\nu}} \quad & -\vec{b}^\top \vec{\nu} && (25)\\ \text{s.t.} \quad & \vec{c} + A^\top \vec{\nu} \geq \vec{0}. && (26) \end{aligned}$$

(b) Express the dual problem as an equivalent minimization problem. Find the dual of this minimization problem, i.e., the dual of the dual. Compare it to the original linear program formulation.

Solution: Say the optimal value for the dual problem is given by

$$d^* = \max_{\vec{\nu}} \quad -\vec{b}^\top \vec{\nu} \quad \text{s.t.} \quad \vec{c} + A^\top \vec{\nu} \geq \vec{0}. \tag{27}$$

The dual problem can be expressed by an equivalent minimization problem:

$$\begin{aligned} d^* = -\min_{\vec{\nu}} \quad & \vec{b}^\top \vec{\nu} && (28)\\ \text{s.t.} \quad & \vec{c} + A^\top \vec{\nu} \geq 0. && (29) \end{aligned}$$

We call the minimization problem our new primal problem P'. We can obtain the dual problem of the problem in 27 by dualizing P' while ignoring the negative sign, and then adding it back in at the end. Denote the dual variable associated with the inequality constraint $\vec{c} + A^\top \vec{\nu} \geq 0$ by $\vec{z} \in \mathbb{R}^n$. The Lagrangian for P' is given by

$$\mathcal{L}(\vec{\nu}, \vec{z}) = \vec{b}^\top \vec{\nu} + \vec{z}^\top(-\vec{c} - A^\top \vec{\nu}) = -\vec{c}^\top \vec{z} + (-A\vec{z} + \vec{b})^\top \vec{\nu}. \tag{30}$$

The dual is given by taking the minimum of the Lagrangian over the *primal* variable of the problem we are dualizing, i.e., $\vec{\nu}$.

$$h(\vec{z}) = \min_{\vec{\nu}} \mathcal{L}(\vec{\nu}, \vec{z}) = \begin{cases} -\vec{c}^\top \vec{z} & \text{if } A\vec{z} - \vec{b} = 0 \\ -\infty & \text{otherwise.} \end{cases} \tag{31}$$

The dual problem for P' is then given by

$$\begin{aligned} \max_{\vec{z}} \quad & -\vec{c}^\top \vec{z} && (32)\\ \text{s.t.} \quad & A\vec{z} = \vec{b} && (33)\\ & \vec{z} \geq \vec{0}. && (34) \end{aligned}$$

Then the dual of the problem obtained in part a) is given by

$$\begin{array}{ccc} -\max_{\vec{z}} & -\vec{c}^\top \vec{z} \\ \text{s.t.} & A\vec{z} = \vec{b} \\ & \vec{z} \geq \vec{0} \end{array} \quad = \quad \begin{array}{ccc} \min_{\vec{z}} & \vec{c}^\top \vec{z} \\ \text{s.t.} & A\vec{z} = \vec{b} \\ & \vec{z} \geq \vec{0} \end{array} \tag{35}$$

which is the linear problem in (17). This tells us that the dual of the dual of a linear program is the same linear program.

5. Minimizing a Sum of Logarithms

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

$$p^\star = \max_{\vec{x} \in \mathbb{R}^n} \quad \sum_{i=1}^n \alpha_i \log(x_i) \tag{36}$$
$$\text{s.t.} \quad \vec{x} \geq 0, \quad \vec{1}^\top \vec{x} = c, \tag{37}$$

where $c > 0$ and $\alpha_i > 0$, $i = 1, \ldots, n$. (Recall that if $\vec{x}$ is a vector then by "$\vec{x} \geq 0$" 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

$$p^\star = \alpha \log(c/\alpha) + \sum_{i=1}^n \alpha_i \log(\alpha_i), \tag{38}$$

where $\alpha \doteq \sum_{i=1}^n \alpha_i$. We will show this in a series of steps.

(a) First, express the problem as a minimization problem which has optimal value $p^\star_{\min}$.

Solution: We have

$$\max_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) = -\min_{\vec{x} \in \mathbb{R}^n} (-f_0(\vec{x})), \tag{39}$$

so

$$p^\star = -\min_{\vec{x} \in \mathbb{R}^n} \quad \sum_{i=1}^n -\alpha_i \log(x_i) \tag{40}$$
$$\text{s.t.} \quad \vec{x} \geq 0, \quad \vec{1}^\top \vec{x} = c. \tag{41}$$

The minimization problem we now consider is

$$p^\star_{\min} = \min_{\vec{x} \in \mathbb{R}^n} \quad \sum_{i=1}^n -\alpha_i \log(x_i) \tag{42}$$
$$\text{s.t.} \quad \vec{x} \geq 0, \quad \vec{1}^\top \vec{x} = c, \tag{43}$$

so that $p^\star_{\min} = -p^\star$.

(b) In optimization, we often "relax" problems of the form $p^\star_{\min} = \min_{\vec{x} \in \mathcal{X}} f_0(\vec{x})$, i.e., 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

$$\vec{x} \doteq \begin{bmatrix} c - 1^\top \vec{x}^r + x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{bmatrix}. \tag{44}$$

Show that $\vec{x}$ is feasible for the original problem and has objective value $< p_r^\star$. Argue that this implies $p_{\min}^\star < p_r^\star$ 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 $\vec{x}$.

Solution: We want to show that the relaxed problem

$$p_r^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad \sum_{i=1}^{n} -\alpha_i \log(x_i) \tag{45}$$

$$\text{s.t.} \quad \vec{x} \geq 0, \quad \vec{1}^\top \vec{x} \leq c, \tag{46}$$

has the same set of solutions as the original minimization problem. We begin by showing that $p_{\min}^\star = p_r^\star$. Indeed, since the relaxed problem minimizes the same objective function over a larger feasible set, $p_r^\star \leq p_{\min}^\star$. We now show that $p_r^\star \geq p_{\min}^\star$.

Suppose for the sake of contradiction that $\vec{x}^r$ is an optimal solution to the relaxed problem which achieves objective value $p_r^\star < p_{\min}^\star$. If $\vec{x}^r$ were feasible for the original minimization problem, then it would be a better solution than the solutions which achieve $p_{\min}^\star$, which is already a contradiction. Thus, suppose $\vec{x}^r$ is infeasible for the original minimization problem, i.e., $\vec{1}^\top \vec{x}^r < c$. Then consider the following solution vector:

$$\vec{x}^\star \doteq \begin{bmatrix} c - \vec{1}^\top \vec{x}^r + x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{bmatrix}. \tag{47}$$

We claim that this choice of solution vector both fulfills all the constraints of the original problem, and achieves a better optimal value. In both parts, we use a crucial inequality:

$$x_1^\star = \underbrace{(c - \vec{1}^\top \vec{x}^r)}_{>0} + x_1^r > x_1^r. \tag{48}$$

To show that $\vec{x}^\star \geq 0$, we just need to show that the first entry $x_1^\star$ is non-negative. This is given by $x_1^\star > x_1^r \geq 0$.

To show that $\vec{1}^\top \vec{x}^\star = c$, we calculate:

$$\vec{1}^\top \vec{x}^\star = \sum_{i=1}^{n} x_i^\star \tag{49}$$

$$= x_1^\star + \sum_{i=2}^{n} x_i^\star \tag{50}$$

$$= c - \vec{1}^\top \vec{x}^r + x_1^r + \sum_{i=2}^{n} x_i^\star \tag{51}$$

$$= c - \vec{1}^\top \vec{x}^r + \sum_{i=1}^{n} x_i^\star \tag{52}$$

$$= c. \tag{53}$$

Finally, we show that the objective value is strictly improved:

$$\sum_{i=1}^{n} -\alpha_i \log(x_i^\star) = -\alpha_1 \log(x_1^\star) + \sum_{i=2}^{n} -\alpha_i \log(x_i^\star) \tag{54}$$

$$< -\alpha_1 \log(x_1^r) + \sum_{i=2}^{n} -\alpha_i \log(x_i^\star) \tag{55}$$

$$= -\alpha_1 \log(x_1^r) + \sum_{i=2}^{n} -\alpha_i \log(x_i^r) \tag{56}$$

$$= \sum_{i=1}^{n} -\alpha_i \log(x_i^r). \tag{57}$$

Thus $\vec{x}^r$ could not be a solution to the relaxed problem, a contradiction.

This establishes that $p_r^\star \geq p_{\min}^\star$ and thus $p_r^\star = p_{\min}^\star$. This argument also shows that all solutions for the relaxed problem are feasible for the original problem, and since $p_r^\star = p_{\min}^\star$, they are the same set of solutions.

How did we cook up $\vec{x}^\star$? The main idea is that since the objective function considered each x_i independently, one can come up with a "better" point for any suboptimal point x^r just by moving one of the x_i, in our case x_1. And since it's monotonically decreasing in each x_i, we can make the x_i larger to get the desired result. Finally, also because the objective is monotonically decreasing in each x_i, we should make the x_i as large as possible subject to the constraints; this is how we came up with the fact that $x_1^\star$ needs to be large. The precise value of $x_1^\star$ is just bookkeeping to ensure that the constraint hits equality for $\vec{x}^\star$.

(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 $\vec{1}^\top \vec{x} \leq c$.

Solution: The Lagrangian for this problem is

$$\mathcal{L}(\vec{x}, \mu) = \sum_{i=1}^{n} \alpha_i \log(1/x_i) + \sum_{i=1}^{n} \lambda_i(-x_i) + \mu(\vec{1}^\top \vec{x} - c) \tag{58}$$

$$= \sum_{i=1}^{n} (\alpha_i \log(1/x_i) + (\mu - \lambda_i)x_i) - \mu c, \tag{59}$$

(d) Now derive the dual function $g(\vec{\lambda}, \mu)$ for the relaxed minimization problem, and solve the dual problem $d_r^\star = \max_{\substack{\vec{\lambda} \geq \vec{0} \\ \mu \geq 0}} g(\vec{\lambda}, \mu)$. What are the optimal dual variables $\vec{\lambda}^\star, \mu^\star$?

Solution: We have

$$g(\vec{\lambda}, \mu) = \min_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}, \mu) = -\mu c + \sum_{i=1}^{n} \min_{x_i \geq 0} (\alpha_i \log(1/x_i) + (\mu - \lambda_i)x_i) \tag{60}$$

$$= -\mu c + \sum_{i=1}^{n} \begin{cases} (\alpha_i \log((\mu - \lambda_i)/\alpha_i) + \alpha_i), & \mu - \lambda_i > 0 \\ -\infty, & \mu - \lambda_i \leq 0 \end{cases} \tag{61}$$

$$= \begin{cases} -\mu c + \sum_{i=1}^{n} (\alpha_i \log((\mu - \lambda_i)/\alpha_i) + \alpha_i), & \forall i\colon \mu - \lambda_i > 0 \\ -\infty, & \exists i\colon \mu - \lambda_i \leq 0 \end{cases} \tag{62}$$

The minimum with respect to x_i in the first expression is attained at the unique point $x_i = \alpha_i/(\mu - \lambda_i)$, which we obtain by verifying that the expression is convex with respect to $\vec{x}$ and setting the gradient to 0. The dual is thus $d_r^\star = \max_{\substack{\vec{\lambda} \geq \vec{0} \\ \mu \geq 0}} g(\vec{\lambda}, \mu)$. To solve for the optimal dual variables, we solve for $\vec{\lambda}^\star$ first and then $\mu^\star$. For *every* choice of μ, it is optimal to pick $\vec{\lambda}^\star = \vec{0}$ so as to increase the quantity in the logarithm (because $\vec{\lambda} \geq \vec{0}$). Setting $\vec{\lambda}^\star = \vec{0}$, taking the gradient of $g(\vec{0}, \mu)$ with respect to μ, and setting it to 0, we obtain the optimal

$$\mu^\star = \frac{\sum_{i=1}^n \alpha_i}{c} = \frac{\alpha}{c}. \tag{63}$$

(e) Show that strong duality holds for the relaxed problem, so $p_r^\star = d_r^\star$.

Solution: We want to apply Slater's condition. We can verify that the objective and constraint functions are convex by taking the Hessian of each and verifying that they are positive semidefinite. For a strictly feasible point, we need to find an $\vec{x} \in \mathbb{R}^n$ such that each $x_i > 0$ and $\sum_{i=1}^n x_i < c$. There are many such $\vec{x}$, but one way to find them is to suppose that all x_i are the same, say χ, and find χ such that $n\chi < c$. This is achieved at $\chi = \frac{c}{2n}$, so $\vec{x} = \frac{c}{2n}\vec{1}$. Thus Slater's condition holds and strong duality holds.

The only inequality constraints are affine constraints, and hence by refined Slater's condition, strong duality should hold (no strictly feasible point is necessary since there are no non-affine inequalities).

(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 $p_r^\star$? Finally, what is $p^\star$?

Solution: We obtain the optimal primal solution as

$$x_i^\star = \frac{\alpha_i}{\mu^\star} = \frac{c\alpha_i}{\alpha}, \quad i = 1, \ldots, n. \tag{64}$$

The expression for the optimal objective value follows by substituting this optimal solution back into the objective:

$$\begin{aligned}
p_r^\star &= \sum_{i=1}^n -\alpha_i \log\left(\frac{c\alpha_i}{\alpha}\right) && (65)\\
&= \sum_{i=1}^n -\left(\alpha_i \log\left(\frac{c}{\alpha}\right) + \alpha_i \log(\alpha_i)\right) && (66)\\
&= -\alpha \log\left(\frac{c}{\alpha}\right) - \sum_{i=1}^n \alpha_i \log(\alpha_i) && (67)\\
p^\star &= -p_{\min}^\star && (68)\\
&= -p_r^\star && (69)\\
&= \alpha \log\left(\frac{c}{\alpha}\right) + \sum_{i=1}^n \alpha_i \log(\alpha_i). && (70)
\end{aligned}$$

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.

