Homework 10

Self grades are due at 11 PM on November 16, 2023.

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

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

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

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

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

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

Solution: Since f is m-strongly convex, $g(\vec{x}) := f(\vec{x}) - \frac{m}{2}\|\vec{x}\|_2^2$ is a convex function. Since f is L-smooth, $h(\vec{x}) := \frac{L}{2}\|\vec{x}\|_2^2 - f(\vec{x})$ is a convex function. Since the sum of convex functions is convex, $g(\vec{x}) + h(\vec{x}) = \frac{L-m}{2}\|\vec{x}\|_2^2$ is a convex function. This implies that $L \ge m$.

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

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

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

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

Solution: From the hint we see that f is also M-smooth for all $M \ge L$. Hence, by setting $M := \frac{1}{\eta}$, we can simply use the result proved in class to conclude that for all $k \ge 0$ we have

$$\begin{aligned} f(\vec{x}^{(k)}) - p^* &\le \left(1 - \frac{m}{M}\right)^k (f(\vec{x}^{(0)}) - p^*) \\ &= (1 - \eta m)^k (f(\vec{x}^{(0)}) - p^*). \end{aligned}$$

2. Visualizing the Dual Problem

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

3. Maximizing a Sum of Logarithms

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

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

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

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

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

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

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

Solution: We have

$$\max_{\vec{x}\in\mathbb{R}^n} f_0(\vec{x}) = -\min_{\vec{x}\in\mathbb{R}^n}(-f_0(\vec{x})), \tag{5}$$

so

$$p^\star = -\min_{\vec{x}\in\mathbb{R}^n} \quad \sum_{i=1}^n -\alpha_i \log(x_i) \tag{6}$$

$$\text{s.t.} \quad \vec{x} \geq \vec{0}, \quad \vec{1}^\top \vec{x} = c. \tag{7}$$

The minimization problem we now consider is

$$p^\star_{\min} = \min_{\vec{x}\in\mathbb{R}^n} \quad \sum_{i=1}^n -\alpha_i \log(x_i) \tag{8}$$

$$\text{s.t.} \quad \vec{x} \geq \vec{0}, \quad \vec{1}^\top \vec{x} = c, \tag{9}$$

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

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

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

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

$$\vec{x} \doteq \begin{bmatrix} c - 1^\top \vec{x}^r + x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{bmatrix}. \tag{10}$$

Show that $\vec{x}$ is feasible for the original problem and has objective value strictly less than $p_r^\star$. 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 that of $\vec{x}$.

Solution: We want to show that the relaxed problem

$$p_r^\star = \min_{\vec{x}\in\mathbb{R}^n} \quad \sum_{i=1}^{n} -\alpha_i \log(x_i) \tag{11}$$

$$\text{s.t.} \quad \vec{x} \geq 0, \quad \vec{1}^\top \vec{x} \leq c, \tag{12}$$

has the same set of optimal 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:

$$\vec{x}^\star \doteq \begin{bmatrix} c - \vec{1}^\top \vec{x}^r + x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{bmatrix}. \tag{13}$$

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:

$$x_1^\star = \underbrace{(c - \vec{1}^\top \vec{x}^r)}_{>0} + x_1^r > x_1^r. \tag{14}$$

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:

$$\vec{1}^\top \vec{x}^\star = \sum_{i=1}^{n} x_i^\star \tag{15}$$

$$= x_1^\star + \sum_{i=2}^{n} x_i^\star \tag{16}$$

$$= c - \vec{1}^\top \vec{x}^r + x_1^r + \sum_{i=2}^{n} x_i^\star \tag{17}$$

$$= c - \vec{1}^\top \vec{x}^r + \sum_{i=1}^{n} x_i^r \tag{18}$$

$$= c. \tag{19}$$

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

$$\sum_{i=1}^{n} -\alpha_i \log(x_i^\star) = -\alpha_1 \log(x_1^\star) + \sum_{i=2}^{n} -\alpha_i \log(x_i^\star) \tag{20}$$

$$< -\alpha_1 \log(x_1^r) + \sum_{i=2}^{n} -\alpha_i \log(x_i^\star) \tag{21}$$

$$= -\alpha_1 \log(x_1^r) + \sum_{i=2}^{n} -\alpha_i \log(x_i^r) \tag{22}$$

$$= \sum_{i=1}^{n} -\alpha_i \log(x_i^r). \tag{23}$$

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 infer the correct form of $\vec{x}^\star$? 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.

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

Solution: The Lagrangian for this problem is

$$\mathcal{L}(\vec{x}, \vec{\lambda}, \mu) = \sum_{i=1}^{n} \alpha_i \log(1/x_i) + \sum_{i=1}^{n} \lambda_i(-x_i) + \mu(\vec{1}^\top \vec{x} - c) \tag{24}$$

$$= \sum_{i=1}^{n} (\alpha_i \log(1/x_i) + (\mu - \lambda_i)x_i) - \mu c, \tag{25}$$

with the domain $\mathbb{R}^n_{++} \times \mathbb{R}^n \times \mathbb{R}$.

(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

$$g(\vec{\lambda}, \mu) = \min_{\vec{x} \in \mathbb{R}^n_{++}} \mathcal{L}(\vec{x}, \vec{\lambda}, \mu) = -\mu c + \sum_{i=1}^{n} \min_{x_i > 0} (\alpha_i \log(1/x_i) + (\mu - \lambda_i)x_i) \tag{26}$$

$$= -\mu c + \sum_{i=1}^{n} \begin{cases} (\alpha_i \log((\mu - \lambda_i)/\alpha_i) + \alpha_i), & \mu - \lambda_i > 0 \\ -\infty, & \mu - \lambda_i \leq 0 \end{cases} \tag{27}$$

$$= \begin{cases} -\mu c + \sum_{i=1}^{n} (\alpha_i \log((\mu - \lambda_i)/\alpha_i) + \alpha_i), & \forall i\colon \mu - \lambda_i > 0 \\ -\infty, & \exists i\colon \mu - \lambda_i \leq 0 \end{cases} \tag{28}$$

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

$$\mu^\star = \frac{\sum_{i=1}^{n} \alpha_i}{c} = \frac{\alpha}{c}. \tag{29}$$

(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. Alternatively, since the only inequality constraints are affine constraints, strong duality should hold by the refined Slater's condition (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 by minimizing over $\vec{x} \in \mathbb{R}^n_{++}$, obtaining

$$x_i^\star = \frac{\alpha_i}{\mu^\star} = \frac{c\alpha_i}{\alpha}, \quad i = 1, \ldots, n. \tag{30}$$

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

$$\begin{aligned}
p_r^\star &= \sum_{i=1}^n -\alpha_i \log\left(\frac{c\alpha_i}{\alpha}\right) && (31)\\
&= \sum_{i=1}^n -\left(\alpha_i \log\left(\frac{c}{\alpha}\right) + \alpha_i \log(\alpha_i)\right) && (32)\\
&= -\alpha \log\left(\frac{c}{\alpha}\right) - \sum_{i=1}^n \alpha_i \log(\alpha_i). && (33)\\
p^\star &= -p^\star_{\min} && (34)\\
&= -p_r^\star && (35)\\
&= \alpha \log\left(\frac{c}{\alpha}\right) + \sum_{i=1}^n \alpha_i \log(\alpha_i). && (36)
\end{aligned}$$

4. Quadratically Constrained Linear Program

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

$$\begin{aligned} p^\star = \min_{\vec{x}} \quad & \vec{c}^\top \vec{x}, \\ \text{s.t.} \quad & \frac{1}{2}(\vec{x} - \vec{x}_0)^\top Q(\vec{x} - \vec{x}_0) \leq \epsilon. \end{aligned} \tag{37}$$

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

Solution: Since $Q \succ 0$, it follows that $\frac{1}{2}(\vec{x} - \vec{x}_0)^\top Q(\vec{x} - \vec{x}_0) - \epsilon$ is a convex function so the feasible set is convex. Moreover, the objective function is linear (thus convex) so the optimization problem is convex.

(b) Prove that the dual function associated with the primal problem in (37) is

$$g(\lambda) = \begin{cases} -\infty & \text{if } \lambda = 0, \\ \vec{c}^\top \vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon & \text{if } \lambda > 0, \end{cases} \tag{38}$$

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

Solution: The Lagrangian associated with the problem is

$$L(\vec{x}, \lambda) = \vec{c}^\top \vec{x} + \lambda \left(\frac{1}{2}(\vec{x} - \vec{x}_0)^\top Q(\vec{x} - \vec{x}_0) - \epsilon \right) \tag{39}$$

Thus, the dual function is given by

$$g(\lambda) = \min_{\vec{x}} L(\vec{x}, \lambda). \tag{40}$$

For $\lambda < 0$, we have $g(\lambda) = -\infty$, so we consider only $\lambda \geq 0$ below.

For $\lambda = 0$:

$$L(\vec{x}, 0) = \vec{c}^\top \vec{x}, \tag{41}$$

and

$$\begin{aligned} g(0) &= \min_{\vec{x}} L(\vec{x}, 0) && (42) \\ &= \min_{\vec{x}} \vec{c}^\top \vec{x} && (43) \\ &= -\infty. && (44) \end{aligned}$$

For $\lambda > 0$: note that $L(\vec{x}, \lambda)$ is a strictly convex in $\vec{x}$. Thus, the minimum is obtained when

$$\begin{aligned} & \nabla_{\vec{x}} L(\vec{x}^\star(\lambda), \lambda) = 0, \text{ i.e., } \vec{c} + \lambda Q(\vec{x}^\star(\lambda) - \vec{x}_0) = 0 && (45) \\ \implies & \vec{x}^\star(\lambda) = \vec{x}_0 - \frac{1}{\lambda} Q^{-1}\vec{c}. && (46) \end{aligned}$$

Thus,

$$\begin{aligned} g(\lambda) &= L(\vec{x}^\star(\lambda), \lambda) && (47) \\ &= \vec{c}^\top \vec{x}^\star(\lambda) + \lambda \left(\frac{1}{2}(\vec{x}^\star(\lambda) - \vec{x}_0)^\top Q(\vec{x}^\star(\lambda) - \vec{x}_0) - \epsilon \right) && (48) \end{aligned}$$

$$= \vec{c}^\top \left(\vec{x}_0 - \frac{1}{\lambda}Q^{-1}\vec{c}\right) + \lambda\left(\frac{1}{2}(\vec{x}_0 - \frac{1}{\lambda}Q^{-1}\vec{c} - \vec{x}_0)^\top Q(\vec{x}_0 - \frac{1}{\lambda}Q^{-1}\vec{c} - \vec{x}_0) - \epsilon\right) \tag{49}$$

$$= \vec{c}^\top\vec{x}_0 - \frac{1}{\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda(\frac{1}{2}(\frac{1}{\lambda}Q^{-1}\vec{c})^\top Q(\frac{1}{\lambda}Q^{-1}\vec{c}) - \epsilon) \tag{50}$$

$$= \vec{c}^\top\vec{x}_0 - \frac{1}{\lambda}\vec{c}^\top Q^{-1}\vec{c} + \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon \tag{51}$$

$$= \vec{c}^\top\vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon. \tag{52}$$

(c) Consider the dual problem of the primal problem in (37):

$$d^\star = \max_{\lambda\geq 0} g(\lambda), \tag{53}$$

where

$$g(\lambda) = \begin{cases} -\infty & \text{if } \lambda = 0, \\ \vec{c}^\top\vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon & \text{if } \lambda > 0. \end{cases} \tag{38}$$

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

Solution: We know a priori that the dual problem

$$\max_{\lambda\geq 0} \vec{c}^\top\vec{x}_0 - \left(\frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda\epsilon\right) = \left(\max_{\lambda\geq 0} - \left(\frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda\epsilon\right)\right) + \vec{c}^\top\vec{x}_0 \tag{54}$$

is convex. We can solve it by setting the derivative $\frac{d}{d\lambda}g(\lambda^\star) = 0$. That is,

$$\begin{aligned} &\frac{d}{d\lambda}\left(\frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda\epsilon\right)\Big|_{\lambda=\lambda^\star} = 0 \\ \implies & -\frac{1}{2(\lambda^\star)^2}\vec{c}^\top Q^{-1}\vec{c} + \epsilon = 0 \\ \implies & \lambda^\star = \sqrt{\frac{\vec{c}^\top Q^{-1}\vec{c}}{2\epsilon}} > 0 \qquad \text{since } \vec{c} \neq 0. \end{aligned} \tag{55}$$

(d) Does strong duality hold for the optimization problem (37)? Justify your answer by directly computing $p^\star$ and $d^\star$, without appealing to constraint qualifications such as Slater's condition.

Solution: By substituting the optimal dual variable (55) into the expression (46) for the optimal primal variable, we can compute the optimal primal variable as follows:

$$\vec{x}^\star(\lambda^\star) = \vec{x}_0 - \frac{1}{\lambda^\star}Q^{-1}\vec{c} = \vec{x}_0 - \sqrt{\frac{2\epsilon}{\vec{c}^\top Q^{-1}\vec{c}}}Q^{-1}\vec{c}.$$

Thus, $p^\star$ can be computed as follows:

$$\begin{aligned} p^\star &= \vec{c}^\top\vec{x}^\star(\lambda) \\ &= \vec{c}^\top\left(\vec{x}_0 - \sqrt{\frac{2\epsilon}{\vec{c}^\top Q^{-1}\vec{c}}}Q^{-1}\vec{c}\right) \\ &= \vec{c}^\top\vec{x}_0 - \sqrt{\frac{2\epsilon}{\vec{c}^\top Q^{-1}\vec{c}}}\cdot\vec{c}^\top Q^{-1}\vec{c} \\ &= \vec{c}^\top\vec{x}_0 - \sqrt{2\epsilon\vec{c}^\top Q^{-1}\vec{c}}. \end{aligned}$$

Meanwhile, $d^\star$ can be computed as follows:

$$\begin{aligned} d^\star &= g(\lambda^\star) \\ &= \vec{c}^\top \vec{x}_0 - \frac{1}{2}\sqrt{\frac{2\epsilon}{\vec{c}^\top Q^{-1}\vec{c}}} \cdot \vec{c}^\top Q^{-1}\vec{c} - \sqrt{\frac{\vec{c}^\top Q^{-1}\vec{c}}{2\epsilon}} \cdot \epsilon \\ &= \vec{c}^\top \vec{x}_0 - \sqrt{2\epsilon\vec{c}^\top Q^{-1}\vec{c}}. \end{aligned}$$

Thus, $p^\star = d^\star$, so strong duality holds.

Note: The feasible set contains points in its relative interior (for example, $\vec{x} = \vec{x}_0$) so Slater's condition implies that strong duality holds for the given problem.

5. Linear Programs and Duality

Consider the following two linear programs:

$$\begin{aligned} \min_{x_1, x_2 \in \mathbb{R}} \quad & c_1 x_1 - 2x_2 && (56)\\ \text{s.t.} \quad & x_1 \geq 0, \\ & x_2 \geq 0, \\ & x_1 + x_2 \leq 1, \\ \max_{x_1 \in \mathbb{R}} \quad & c_2 x_1 && (57)\\ \text{s.t.} \quad & x_1 \geq 7. \end{aligned}$$

Find the values of c_1 and c_2 such that (57) is the Lagrangian dual problem of (56).

Solution: We write the Lagrangian:

$$\begin{aligned} \mathcal{L}(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) &= c_1 x_1 - 2x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1 + x_2 - 1) \\ &= x_1(c_1 - \lambda_1 + \lambda_3) + x_2(\lambda_3 - \lambda_2 - 2)x_2 - \lambda_3. \end{aligned}$$

Then we see that minimizing the Lagrangian over $x_1, x_2 \in \mathbb{R}$ to find the dual objective function yields a finite value iff $c_1 - \lambda_1 + \lambda_3 = 0$ and $\lambda_3 - \lambda_2 - 2 = 0$, in which case the objective of the dual problem becomes $-\lambda_3$. We can eliminate λ_2 by using the condition $\lambda_3 - \lambda_2 - 2 = 0$ to set λ_2 to equal $\lambda_3 - 2$, and the condition $\lambda_2 \geq 0$ would then strengthen the condition $\lambda_3 \geq 0$ to $\lambda_3 \geq 2$. We can eliminate λ_1 by using the condition $c_1 - \lambda_1 + \lambda_3 = 0$ to set λ_1 to $c_1 + \lambda_3$. The condition $\lambda_1 \geq 0$ would now be replaced by the condition $\lambda_3 \geq -c_1$.

The dual is now expressed in terms of λ_3 as

$$\begin{aligned} \max_{\lambda_3 \geq 0}. \quad & -\lambda_3 \\ \text{s.t.} \quad & \lambda_3 \geq 2, \\ & \lambda_3 \geq -c_1. \end{aligned}$$

With this we see that $c_2 = -1$ and $c_1 = -7$.

6. Does strong duality hold?

Consider

$$\min_{(x,y)\in\mathcal{D}} e^{-x} \tag{58}$$

$$\text{s.t. } x^2/y \leq 0 \tag{59}$$

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

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

Solution: The second derivative of the objective function is e^{-x}, which is non-negative. Thus, the objective is a convex function. Furthermore, the constraint is jointly convex, as can be verified by showing the Hessian is PSD. The Hessian for $g(x,y) = \frac{x^2}{y}$ with domain $\mathcal{D}$ is given by

$$\nabla^2 g(x,y) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix}. \tag{60}$$

Suppose λ_1, λ_2 are the eigenvalues of $\nabla^2 g(x,y)$. The determinant of the Hessian is 0, which gives us $\lambda_1\lambda_2 = 0$. Moreover, the trace of the Hessian is $\frac{2}{y} + \frac{2x^2}{y^3} > 0$ (since $y > 0$), which gives $\lambda_1 + \lambda_2 > 0$. Thus one eigenvalue must be positive and the other must be 0, which shows that the Hessian is positive semidefinite. Hence the problem is convex. Furthermore, at all feasible (x,y), we have $x = 0$. Hence the optimal value is $e^{-0} = 1$.

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

$$g(\lambda) = \inf_{(x,y)\in\mathcal{D}} \left(e^{-x} + \lambda\frac{x^2}{y}\right). \tag{61}$$

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

Solution: This is true since both terms in the sum are non-negative because $y > 0$.

(c) Show that $g(\lambda) = 0$ for $\lambda \geq 0$. *HINT: To show that the infimum in Equation (61) is 0, we want to show there exist (x,y) such that both e^{-x} and $\lambda\frac{x^2}{y}$ can get arbitrarily close to 0. HINT: Consider a sequence $\{x_k\}$ going to $+\infty$ and a sequence $\{y_k\}$ also going to $+\infty$ such that $\lim_{k\to\infty}\frac{x_k^2}{y_k} = 0$. Simply put, we want to drive x to infinity in order to drive e^{-x} to 0, while having y grow faster than x^2, so that the second term also goes to 0.*

Solution: To show that the infimum is 0, we pick any sequence $\{x_k\}$ going to $+\infty$ and pick a sequence $\{y_k\}$ also going to $+\infty$ such that $\lim_{k\to\infty}\frac{x_k^2}{y_k} = 0$. One example of such sequence pair is $y_k = x_k^4$ and $x_k = 2k$.

This gives $\lim_{k\to\infty} e^{-x_k} + \frac{x_k^2}{y_k} = 0$, so $g(\lambda) = \inf_{(x,y)\in\mathcal{D}} e^{-x} + \lambda\frac{x^2}{y} = 0$.

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

Solution: The Lagrange dual problem is

$$d^* = \sup_{\lambda\geq 0} g(\lambda), \tag{62}$$

where $g(\lambda) = \inf_{(x,y)\in\mathcal{D}} \ e^{-x} + \lambda x^2/y$. Note that $g(\lambda) = 0$ for all $\lambda \geq 0$. Thus, $d^* = 0$ and any $\lambda \geq 0$ is optimal. The duality gap is 1.

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

Solution: While the primal problem is convex, we cannot find a point that is strictly in the interior of the domain and satisfies the constraint as needed for Slater's condition. Specifically, for Slater's condition to hold we need the existence of an (x, y) pair such that $x^2/y < 0$. Note there is no such pair (x, y) since $y > 0$ and $x^2 \geq 0$. Hence Slater's condition does not hold for this problem.

From the previous parts we saw that $p^* \neq d^*$, and thus strong duality does not hold. Furthermore, the problem is convex. For convex problems we know that if Slater's condition holds then we must have strong duality (i.e Slater's is a sufficient condition). However since strong duality does not hold it implies that Slater's does not hold.

Note that this problem is an example illustrating that **convexity alone is not enough to guarantee strong duality for an optimization problem**.