# EECS 127/227AT Optimization Models in Engineering

## Spring 2020 Final

**Write your Student ID and full name on the first page of your submission. Start every subpart of every problem on a new page**, except for the Honor Code and the first two questions, which can all be written on the first page. **The top of every page should clearly state the problem and subpart being solved.**

There are a total of 107 points on this exam.

**HONOR CODE**
Copy the following statements below and sign your name.

*I will respect my classmates and the integrity of this exam by following this honor code. I affirm:*

- *I have read the instructions for this exam in the associated Piazza posts and understand them.*
- *All of the work submitted here is my original work.*
- *I did not reference any sources on the internet other than the course textbooks, notes, homework and discussion sheets.*
- *I did not collaborate with any other human being on this exam.*

1. **(2 Points) What are you looking forward to over summer break?**

2. **(2 Points) If you had a superpower, what would it be?**

Start every subpart of every problem on a new page.

**3. (8 points) Newton's method**

Consider the function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = x^4.$$

(a) (2 points) Find the optimal value $x^* = \operatorname{argmin}_x f(x)$.

**Solution:** Note that $f(x) = x^4 \geq 0$ and is equal to 0 if and only if $x = 0$. Thus $x^* = 0$

(b) (6 points) Now, we analyze the performance of Newton's method on this problem. Starting from $x_0$, for $k \geq 0$ we take Newton steps of the form

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}.$$

**Find the minimum number of Newton steps that are required to be within a distance of $\epsilon > 0$ from optimum $x^*$.** You may use the optimum value of $x^*$ from part (a). Formally find $k^* \in \mathbb{N}$ (where $\mathbb{N}$ is the set of positive integers) which is the smallest $k$ for which $|x_k - x^*| \leq \epsilon$, i.e.,

$$k^* = \min_{k \in \mathbb{N} : |x_k - x^*| \leq \epsilon} k.$$

Assume that $x_0 > \epsilon > 0$. Your answer should be in terms of $\epsilon$ and $x_0$.

**Solution:** First we calculate the derivatives of $f(x)$.

$$\begin{aligned} f'(x) &= 4x^3 \\ f''(x) &= 12x^2. \end{aligned}$$

For $k \geq 0$ we have,

$$\begin{aligned} x_{k+1} &= x_k - \frac{f'(x_k)}{f''(x_k)} \\ &= x_k - \frac{4x_k^3}{12x_k^2} \\ &= x_k - \frac{x_k}{3} \\ &= \frac{2}{3} x_k. \end{aligned}$$

Thus,

$$x_k = \left(\frac{2}{3}\right)^k x_0.$$

Since $x^* = 0$,

$$\begin{aligned}
&|x_k - x^*| \leq \epsilon \\
\implies &\left(\frac{2}{3}\right)^k |x_0| \leq \epsilon \\
\implies &\left(\frac{2}{3}\right)^k \leq \frac{\epsilon}{|x_0|} \\
\implies &k \log\left(\frac{2}{3}\right) \leq \log\left(\frac{\epsilon}{|x_0|}\right) \\
\implies &k \geq \frac{\log\left(\frac{\epsilon}{|x_0|}\right)}{\log\left(\frac{2}{3}\right)}.
\end{aligned}$$

We switched the sign of inequality since $\log\left(\frac{2}{3}\right) < 0$. The smallest natural number $k^*$ for which this occurs is,

$$k^* = \left\lceil \frac{\log\left(\frac{\epsilon}{|x_0|}\right)}{\log\left(\frac{2}{3}\right)} \right\rceil = \left\lceil \frac{\log\left(\frac{\epsilon}{x_0}\right)}{\log\left(\frac{2}{3}\right)} \right\rceil.$$

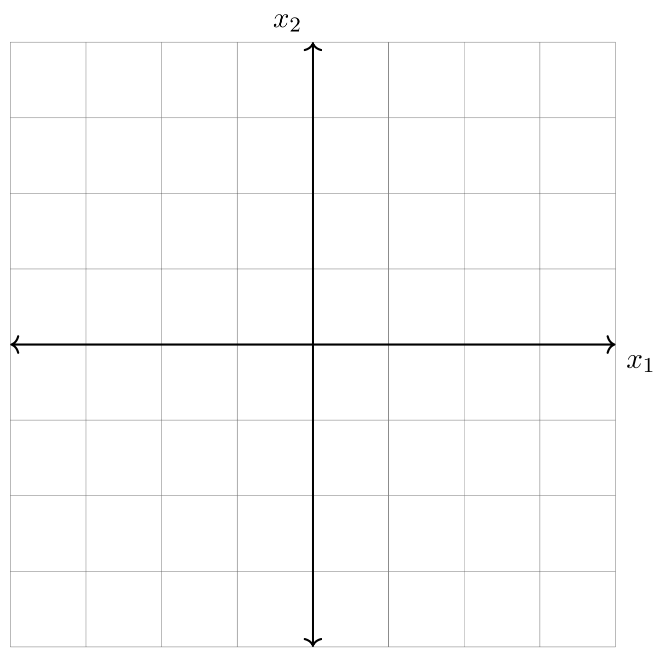
Start every subpart of every problem on a new page.

**4. (9 points) A Linear Program**

Consider the following LP:

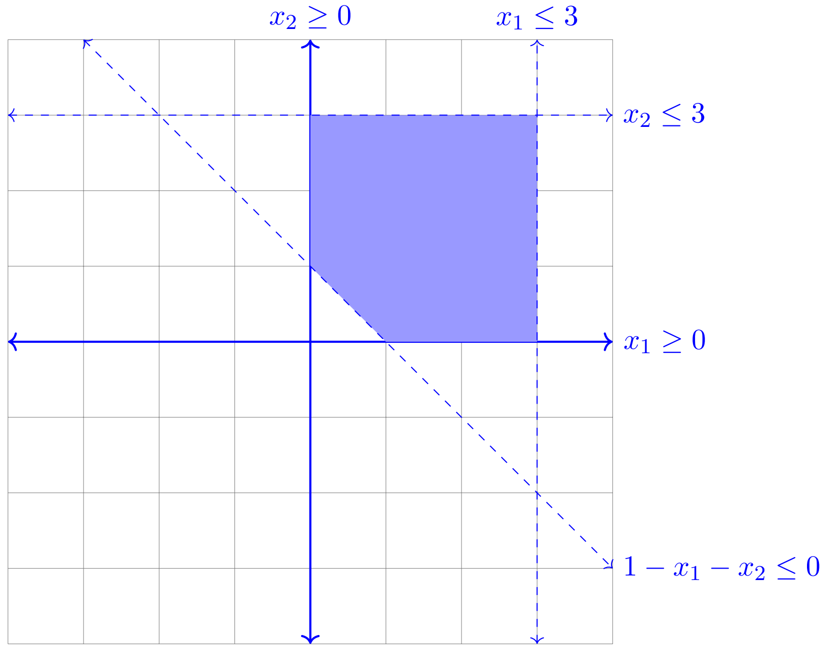
$$
\begin{aligned}
p^* = \min_{\vec{x} \in \mathbb{R}^2} \; & x_1 + x_2 \\
\text{s.t. } & x_1 \geq 0 \\
& x_2 \geq 0 \\
& 1 - x_1 - x_2 \leq 0 \\
& x_1 \leq 3 \\
& x_2 \leq 3.
\end{aligned}
$$

(a) (5 points) Copy the axes below onto your answer sheet. **Plot the feasible region for the optimization problem above. Label your constraints.**




**Solution:**

(b) (4 points) **Find** $p^*$. Justify your answer. *Hint: You don't have to find the dual to solve this part.* **Solution:** To solve any LP, we can just look at the vertices of the polytope defined by constraints, evaluate the objective at each of these points, and find the minimum. In this problem, there are a total of 5 vertices. At $(0,3),(3,0)$, the objective value is 3. At $(3,3)$, the objective value is 6. At $(0,1),(1,0)$, the value is 1. Therefore, $p^* = 1$, which is attained anywhere on the line connecting $(0,1)$ and $(1,0)$.




Start every subpart of every problem on a new page.

**5. (10 points) Reformulate**

(a) (5 points) **Reformulate** the following problem as an SOCP:

$$\min_{\vec{x}} \max_{i=1,2,\ldots,m} \|A\vec{x} - B\vec{y}_i\|_2 .$$

**Solution:** Introducing slack variable $t \in \mathbb{R}$, we can rewrite the problem as,

$$\begin{aligned} &\min_{\vec{x},t} t \\ &\text{s.t. } \|A\vec{x} - B\vec{y}_i\|_2 \leq t, \ i = 1, 2, \ldots, m. \end{aligned}$$

This is an SOCP in standard form.

(b) (5 points) The optimization problem below is not a convex QP because $A$ is not a positive semidefinite symmetric matrix. **Reformulate the following problem as a convex QP:**

$$\begin{aligned} &\min_{\vec{x} \in \mathbb{R}^2} \vec{x}^\top A\vec{x} \\ &\text{s.t. } \vec{c}^\top \vec{x} \geq 1, \end{aligned}$$

where $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ and $\vec{c} \in \mathbb{R}^2$.

*Hint: You can write $\vec{x}^\top A\vec{x}$ as $\vec{x}^\top B\vec{x}$, where $B$ is symmetric.*

**Solution:**

Note that:

$$\vec{x}^\top A\vec{x} = \vec{x}^\top B\vec{x}$$

With $B = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$.
Because $\lambda(B) = \{\frac{3}{2}, \frac{1}{2}\}$, $B$ is PD, so the following problem is a convex QP since the objective being minimized is convex while the constraint is linear:

$$\min_{x \in \mathbb{R}^n} \vec{x}^\top B \vec{x}$$
$$\text{s.t. } \vec{c}^\top \vec{x} \geq 1.$$

Start every subpart of every problem on a new page.

**6. (14 points) Orthogonal lines**

(a) (4 points) **Finding the equation of a line.** Suppose we have $n$ noisy samples from a line in the 2D plane $\{(x_i, y_i)\}_{i=1}^{n}$, which is governed by the equation:

$$\beta x + y = \gamma_1 \qquad \text{(line } A\text{)}.$$

We would like to use the samples to recover the parameters of the line $\vec{z} = [\beta, \gamma_1]^\top$. We can formulate this as a least squares problem:

$$\min_{\vec{z}} \left\| G\vec{z} - \vec{h} \right\|_2^2,$$

for matrix $G \in \mathbb{R}^{n\times 2}$ and vector $\vec{h} \in \mathbb{R}^n$. **Find $G$ and $\vec{h}$.**

**Solution:** The least squares problem involves minimizing the sum of residuals i.e

$$\min_{\beta,\gamma_1} \sum_{i=1}^{n} (y_i + \beta x_i - \gamma_1)^2.$$

Thus,

$$h = \begin{bmatrix} -y_1 \\ \vdots \\ -y_n \end{bmatrix}$$

$$G = \begin{bmatrix} x_1 & -1 \\ \vdots & \vdots \\ x_n & -1 \end{bmatrix}.$$

(b) (4 points) Suppose now, we have two lines in the 2D plane:

$$\beta x + y = \gamma_1 \qquad \text{(line } A\text{)}$$
$$-x + \beta y = \gamma_2 \qquad \text{(line } B\text{)}$$

**Show that lines $A$ and $B$ are orthogonal to one another.**

**Solution:** The inner product of the normal directions to the line $[\beta, 1]$ and $[-1, \beta]$ is 0 – hence they are orthogonal lines.

(c) (6 points) Consider lines $A$ and $B$ as in part (b). Suppose we have $n$ noisy samples from each line: $\{(x_{A_i}, y_{A_i})\}_{i=1}^{n}$ and $\{(x_{B_i}, y_{B_i})\}_{i=1}^{n}$, where $x_{A_i}$ represents the $x$–coordinate of the $i^{\text{th}}$ data point generated from line $A$, and so on.

We are interested in recovering the equations of both lines from these noisy samples. We could estimate the parameters of the two lines independently as in part (a), but then our recovered lines would not necessarily be orthogonal.

To this end, we can write an optimization problem to perform joint estimation of the parameters $\vec{w} = [\beta, \gamma_1, \gamma_2]^\top$:

$$\min_{\vec{w}} \left\| C\vec{w} - \vec{d} \right\|_2^2,$$

for $C \in \mathbb{R}^{2n\times 3}$ and $\vec{d} \in \mathbb{R}^{2n}$. **Find $C$ and $\vec{d}$.**

**Solution:** Similar to part (a) we would like to minimize the sum of the residuals,

$$\min_{\beta,\gamma_1,\gamma_2} \left( \sum_{i=1}^{n} (y_{A_i} + \beta x_{A_i} - \gamma_1)^2 + \sum_{i=1}^{n} (\beta y_{B_i} - x_{B_i} - \gamma_2)^2 \right).$$

Rewriting this in matrix form we have

$$C = \begin{bmatrix} x_{A_1} & -1 & 0 \\ \vdots & \vdots & \vdots \\ x_{A_n} & -1 & 0 \\ y_{B_1} & 0 & -1 \\ \vdots & \vdots & \vdots \\ y_{B_n} & 0 & -1 \end{bmatrix}, \quad \vec{d} = \begin{bmatrix} -y_{A_1} \\ \vdots \\ -y_{A_n} \\ x_{B_1} \\ \vdots \\ x_{B_n} \end{bmatrix}.$$

Start every subpart of every problem on a new page.

**7. (11 points) 1D regularization**

Consider the following regularized one-dimensional regression problem for $x \in \mathbb{R}$ :

$$x^* = \arg\min_x \; f(x),$$

for

$$f(x) = \frac{1}{2}(\alpha - x)^2 + \lambda \mathbb{I}_{x \neq 0}.$$

Here, $\alpha \in \mathbb{R}, \alpha \neq 0$ is known and fixed and

$$\mathbb{I}_{x \neq 0} = \begin{cases} 1, & x \neq 0 \\ 0, & x = 0 \end{cases},$$

so the regularizer penalizes the objective function when $x \neq 0$.

(a) (8 points) **Find** $x^*$. Show your work.

**Solution:** First, and most importantly, note that because of the indicator function in the objective $f(x)$ is not differentiable. We must break the problem into cases to be able to solve it. The indicator function in the objective leads us to break up the problem into 2 cases – the case when $x = 0$ and when $x \neq 0$. If $x = 0$ the objective always reduces to $\frac{1}{2}\alpha^2$ whereas is $x \neq 0$ the objective reduces to $\frac{1}{2}(\alpha - x)^2 + \lambda$. For this case it is optimal to choose $x = \alpha \neq 0$ since $\alpha \neq 0$ and the resultant objective value is $\lambda$. From this we see that $x^* = \alpha$, since $(\alpha - x)^2$ is minimized at $x = \alpha$. To see this, we take the derivative and set it equal to zero

$$0 = \frac{d}{dx}\left(\frac{1}{2}(\alpha - x)^2 + \lambda\right) = \alpha - x$$

which shows $x^* = \alpha$ when $x \neq 0$. (For full credit, you are not required to show the above calculation of the derivative, you may infer $x^* = \alpha$ as the minimum of the quadratic directly). Now the two objective values in the two cases are $\lambda$ and $\frac{1}{2}\alpha^2$. We compare these to arrive at the final optimal solution: When $\lambda \leq \frac{1}{2}\alpha^2$ then $x^* = \alpha$ is optimal, else $x^* = 0$ is optimal (if $\lambda = \frac{1}{2}\alpha^2$ then $x^* = 0$ or $x^* = \alpha$).

**Alternate solution:** Because of the the indicator function in the objective we break the problem into 2 cases – the case when $x = 0$ and when $x \neq 0$. If $x = 0$ the objective always reduces to $\frac{1}{2}\alpha^2$ whereas is $x \neq 0$ the objective reduces to $\frac{1}{2}(\alpha - x)^2 + \lambda$.

$$\min_x \min \; \left(\frac{1}{2}\alpha^2, \frac{1}{2}(\alpha - x)^2 + \lambda\right)$$

Since we can take minimums in any order, we can swap the mins to arrive at

$$\min\left(\min_x \frac{1}{2}\alpha^2, \min_x \frac{1}{2}(\alpha - x)^2 + \lambda\right)$$

where we bring the min inside each component of the vector $(\frac{1}{2}\alpha^2, \frac{1}{2}(\alpha - x)^2 + \lambda) \in \mathbb{R}^2$. Solving the inner minimizations as in the solution above we have

$$\min \; \left(\frac{1}{2}\alpha^2, \lambda\right)$$

Hence if $\frac{1}{2}\alpha^2 \leq \lambda$ this implies that $x^* = 0$ and if $\lambda \leq \frac{1}{2}\alpha^2$ then $x^* = \alpha$ is optimal. When $\lambda = \frac{1}{2}\alpha^2$ both $x = 0$ and $x = \alpha$ are optimal.

(b) (3 points) **What happens to $x^*$ if $\lambda \to \infty$?** You may solve this using the previous part or without using it. **Solution:**

From the previous problem, if $\lambda \to \infty$ that $x^* = 0$, since As $\lambda \to \infty$ we will have $\lambda > \frac{1}{2}\alpha^2$, which will set $x$ to 0. Alternatively, to see this directly, we may argue that there is an increasing penalty on $x$ being away from 0 as $\lambda \to \infty$, and the regularization penalty with outweigh the quadratic part of the function.

Start every subpart of every problem on a new page.

**8. (21 points) Minimizing a Quadratic Form**

Let $A \in S^n$ and $B \in S^n_{++}$, i.e. $A$ is symmetric and $B$ is positive definite. Consider the following optimization problem:

$$\begin{aligned} p^* = \quad & \min_{\vec{x} \in \mathbb{R}^n} && \vec{x}^\top A \vec{x} \\ & \text{s.t.} && \vec{x}^\top B \vec{x} \geq 1. \end{aligned}$$

(a) (3 points) **Is this optimization problem convex as stated?** Provide justification.

**Solution:** No; the problem is not convex since the constraint

$$\vec{x}^\top B \vec{x} \geq 1$$

is not convex (Note that $B$ is positive definite while for the constraint to be convex in given form we would require $B$ to be negative semi definite.) Also the objective function is not necessarily convex since we do not know if $A$ is positive semi definite.

(b) (6 points) **Find the Lagrangian dual problem using $\lambda \in \mathbb{R}$ as the dual variable corresponding to the constraint.**

**Solution:** The Lagrangian dual is

$$\begin{aligned} d^* &= \max_{\lambda \geq 0} \min_{\vec{x} \in \mathbb{R}^n} && \vec{x}^\top A \vec{x} + \lambda(1 - \vec{x}^\top B \vec{x}) \\ &= \max_{\lambda \geq 0} \min_{\vec{x} \in \mathbb{R}^n} && \vec{x}^\top (A - \lambda B)\vec{x} + \lambda \\ &= \max_{\lambda \geq 0} && \lambda \\ & \quad \text{s.t.} && A - \lambda B \succeq 0. \end{aligned}$$

In the second line, we use

$$\min_{\vec{x} \in \mathbb{R}^n} \vec{x}^\top (A - \lambda B)\vec{x} = \begin{cases} 0 & A - \lambda B \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

It turns out that for this problem strong duality holds.

(c) (6 points) Assume $A \succeq 0$. Let $\vec{x}^* \in \mathbb{R}^n$ and $\lambda^* \geq 0$ be optimal solutions to the primal and dual programs, respectively. **Show that**

$$A\vec{x}^* = \lambda^* B \vec{x}^*.$$

**Solution:**

Since strong duality holds and the problem is differentiable, the KKT conditions are necessary for optimality. Note, the problem is non-convex, and so the KKT conditions are not sufficient, but since we are dealing with optimal primal and dual points, we can use the necessary conditions. Thus,

$$\nabla_{\vec{x}} \mathcal{L}(\vec{x}^*, \lambda^*) = \nabla_{\vec{x}} (\vec{x}^\top (A - \lambda^* B)\vec{x}^* + \lambda^*) = (A - \lambda^* B)\vec{x}^* = 0$$

from which it follows

$$A\vec{x}^* = \lambda^* B \vec{x}^*.$$

(d) (6 points) Again, assume $A \succeq 0$ and let $\vec{x}^* \in \mathbb{R}^n$ and $\lambda^* \geq 0$ be optimal solutions to the primal and dual programs. **Show that**

$$\vec{x}^* \notin \mathcal{N}(A) \implies \vec{x}^{*\top} B \vec{x}^* = 1.$$

**Solution:** Since strong duality holds and the problem is differentiable, the KKT conditions are necessary for optimality. Note, the problem is non-convex, and so the KKT conditions are not sufficient, but since we are dealing with optimal primal and dual points, we can use the necessary conditions. Now, by complementary slackness,

$$\lambda^*(1 - \vec{x}^{*\top} B \vec{x}^*) = 0.$$

We assume $A\vec{x}^* \neq 0$ and, from the previous part,

$$A\vec{x}^* = \lambda^* B \vec{x}^*.$$

Thus, $\lambda^* \neq 0$ so

$$\vec{x}^{*\top} B \vec{x}^* = 1.$$

Start every subpart of every problem on a new page.

**9. (30 points) Duality**

We aim to solve the following problem with convex duality:

$$p^* = \min_{\vec{x} \in \mathbb{R}^d} \|\vec{y} - \vec{x}\|_\infty \quad \text{s.t } \|\vec{x}\|_1 \le \mu. \tag{Primal}$$

for some $\mu > 0$. Here $\vec{x} \in \mathbb{R}^d$ is the variable we optimize over and $\vec{y} \in \mathbb{R}^d$ is a fixed and known vector. Assume that $\|\vec{y}\|_1 > \mu$ (otherwise we could set $\vec{x} = \vec{y}$ and obtain a optimal value of 0).

(a) (2 points) First, let us consider a simple example, to gain intuition. Let $\vec{y} = [2, 1]^\top$, and let $\mu = 1$. Here $p^* = 1$. **Find the optimal solution $\vec{x}^*$ that achieves this value.**

**Solution:** We choose $\vec{x}^* = [1, 0]^\top$. Check that $\vec{x}^*$ is feasible since $1 + 0 \le 1$, and so $\vec{x}$ satisfies the constraint and achieves $p^* = \min(2 - 1, 1 - 0) = 1$.

(b) (2 points) Now, consider a second example. Let $\vec{y} = [2, 1]^\top$, and let $\mu = 2$.
Here $p^* = 0.5$. **Find the optimal solution $\vec{x}^*$ that achieves this value.**

**Solution:** We choose $\vec{x}^* = [1.5, 0.5]^\top$. Check that $\vec{x}^*$ is feasible since $1.5 + 0.5 \le 2$ and hence satisfies the constraint and achieves $p^* = \min(2 - 1.5, 1 - 0.5) = 0.5$.

Now, assume the entries of $\vec{y}$ are ordered as $y_1 \ge y_2 \cdots \ge y_d \ge 0$ and suppose $y_1 \ge \tau^* \ge 0$ satisfies:

$$\sum_{i=1}^{d} (y_i - \min(y_i, \tau^*)) = \mu.$$

We will show that $\tau^*$ is the optimal solution to the problem (**??**), i.e $\tau^* = p^*$ (It can be shown that such a $\tau^*$ always exists but you will not be required to prove this).

(c) (6 points) Next define $\vec{w}$ such that,

$$w_i = y_i - \min(y_i, \tau^*).$$

**Prove that $\|\vec{y} - \vec{w}\|_\infty = \tau^*$ and $\|\vec{w}\|_1 = \mu$. Justify why this implies that $p^* \le \tau^*$.**

**Solution:** We first prove that $\|\vec{y} - \vec{x}^*\|_\infty = \tau^*$. We have for all $i$:

$$|y_i - x_i^*| = |y_i - y_i + \min(y_1, \tau^*)| = \min(\tau^*, y_i).$$

And furthermore, since $y_1 \ge \tau^*$, we have $|y_1 - w_1| = \tau^*$ and this proves the first claim. For the second claim, we have:

$$\|\vec{x}^*\|_1 = \sum_{i=1}^{d} |y_i - \min(y_i, \tau^*)| = \sum_{i=1}^{d} y_i - \min(y_i, \tau^*) = \mu.$$

We can remove the absolute value sign because $\min(y_i, \tau^*) \le y_i$ by definition. This proves the second claim.

Finally, since $\vec{w}$ is a feasible point which attains objective value $\tau^*$, we must have $p^* \le \tau^*$.

(d) (10 points) Derive a dual of the problem by introducing a vector valued dual variable, $\vec{z}$ for the $\ell_\infty$ norm in the objective and a scalar valued dual variable, $\lambda$, for the constraint. **Formally, prove that a dual of the above program is:**

$$\begin{aligned} \max_{\vec{z},\lambda} \; & \vec{z}^\top \vec{y} - \lambda\mu \\ \text{s.t } & \|\vec{z}\|_1 \le 1 \\ & \|\vec{z}\|_\infty \le \lambda \\ & \lambda \ge 0. \end{aligned} \tag{Dual}$$

*Hint: Write out the $\ell_\infty$ norm in the primal objective as $\max_{\|\vec{z}\|_1 \le 1} \vec{z}^\top(\vec{y} - \vec{x})$.*

**Solution:** We first re-write the objective as follows:

$$\max_{\|\vec{z}\|_1 \le 1} \vec{z}^\top(\vec{y} - \vec{x})$$

which means we can re-write the optimization problem as follows:

$$p^* = \min_{\vec{x}} \max_{\|\vec{z}\|_1 \le 1, \lambda \ge 0} \vec{z}^\top(\vec{y} - \vec{x}) + \lambda(\|\vec{x}\|_1 - \mu).$$

From the above equation, we may write a Lagrangian of the above problem as:

$$\mathcal{L}(\vec{x}, \vec{z}, \lambda) = \vec{z}^\top(\vec{y} - \vec{x}) + \lambda(\|\vec{x}\|_1 - \mu).$$

We will now prove the dual function is defined for $\lambda \ge 0$ as:

$$g(\vec{z}, \lambda) = \min_{\vec{x}} \vec{z}^\top \vec{y} - \lambda\mu + (\lambda \|\vec{x}\|_1 - \vec{z}^\top \vec{x}) = \begin{cases} -\infty & \text{if } \|\vec{z}\|_\infty > \lambda \\ \vec{z}^\top \vec{y} - \lambda\mu & \text{otherwise} \end{cases}$$

For the first case, let $\|\vec{u}\|_1 = 1$ be such that $\vec{u}^\top \vec{z} = \|\vec{z}\|_\infty$ from the fact that the $\ell_1$ and $\ell_\infty$ norms are dual norms. Then picking $\vec{x} = \alpha\vec{u}$, we get:

$$\vec{z}^\top \vec{y} - \lambda\mu + (\lambda \|\vec{x}\|_1 - \vec{z}^\top \vec{x}) = \vec{z}^\top \vec{y} - \lambda\mu + \alpha(\lambda - \|\vec{z}\|_\infty).$$

Taking $\alpha \to \infty$, we get the first case in the definition of the dual function. For, the second case, we have by Hölders inequality:

$$\vec{z}^\top \vec{y} - \lambda\mu + (\lambda \|\vec{x}\| - \vec{z}^\top \vec{x}) \ge \vec{z}^\top \vec{y} - \lambda\mu + (\lambda \|\vec{x}\|_1 - \|\vec{z}\|_\infty \|\vec{x}\|_1) \ge \vec{z}^\top \vec{y} - \lambda\mu$$

with equality achieved when $\vec{x} = 0$. This completes the definition of the dual function. Therefore, we may write the dual as follows:

$$\begin{aligned} \max_{\vec{z},\lambda} \; & \vec{z}^\top \vec{y} - \lambda\mu \\ \text{s.t } & \|\vec{z}\|_1 \le 1 \\ & \|\vec{z}\|_\infty \le \lambda \\ & \lambda \ge 0. \end{aligned}$$

**Alternative Solution**

We first write out the optimization problem by using the dual characterization of the $\ell_\infty$ norm:

$$\min_{\|\vec{x}\|_1 \le \mu} \max_{\|\vec{z}\|_1 \le 1} \vec{z}^\top (\vec{y} - \vec{x}).$$

We now exchange the min and the max in the previous formulation and obtain:

$$\begin{aligned}\max_{\|\vec{z}\|_1 \le 1} \min_{\|\vec{x}\|_1 \le \mu} \vec{z}^\top (\vec{y} - \vec{x}) &= \max_{\|\vec{z}\|_1 \le 1} \left( \vec{z}^\top \vec{y} - \max_{\|\vec{x}\|_1 \le \mu} \vec{z}^\top \vec{x} \right) = \max_{\|\vec{z}\|_1 \le 1} \left( \vec{z}^\top \vec{y} - \max_{\|\vec{u}\|_1 \le 1} \mu \vec{z}^\top \vec{u} \right) \\ &= \max_{\|\vec{z}\|_1 \le 1} \vec{z}^\top \vec{y} - \mu \|\vec{z}\|_\infty\end{aligned}$$

where the last equality follows from the dual characterization of the $\ell_\infty$ norm. By introducing a slack variable, $\lambda$, for the $\ell_\infty$ term, we have:

$$\begin{aligned}&\max_{\vec{z},\lambda} \vec{z}^\top \vec{y} - \lambda\mu \\ &\text{s.t } \|\vec{z}\|_1 \le 1 \\ &\|\vec{z}\|_\infty \le \lambda \\ &\lambda \ge 0.\end{aligned}$$

Now, let $k = \max\{i : y_i \ge \tau^*\}$. Because the $y_i$ are in decreasing order, we know that $y_1 \ge y_2 \ge \cdots \ge y_k \ge \tau^*$, and $k$ is the number of $y_i$ that are greater than or equal to $\tau^*$. Note that $k > 0$ as $y_1 > \tau^*$.

(e) (8 points) We will now design a dual feasible point, $(\vec{z}^*, \lambda^*)$ with dual objective value $\tau^*$ for the dual problem given in (**??**). Consider $\vec{z}^* = [z_1^*, z_2^*, \ldots, z_d^*]^\top$ defined as:

$$z_i^* = \begin{cases} \frac{1}{k} & i = 1, 2, \ldots, k \\ 0 & i = k+1, \ldots, d. \end{cases}$$

- First, **verify that** $\|\vec{z}^*\|_1 \le 1$**.**
- Next, **find a** $\lambda^*$ **such that** $(\vec{z}^*, \lambda^*)$ **is feasible for** (**??**) **and the value of the dual objective is** $\tau^*$**.** Highlight your choice for $\lambda^*$ by drawing a box around it. Justify your answer.

**Solution:** We have,

$$z_i^* = \begin{cases} \frac{1}{k} & i \in [k] \\ 0 & \text{otherwise} \end{cases}.$$

The notation $i \in [k]$ means $i = 1, 2, ..., k$. Note that $\|\vec{z}\|_1 = \sum_{i=1}^{k} 1/k = 1 \le 1$.
Further $\|\vec{z}\|_\infty = 1/k$. Thus for dual feasibility we require $\lambda \ge \frac{1}{k}$. Intuitively we see that for a given $\vec{z}$, the dual objective is maximized for the smallest possible feasible $\lambda$. This motivates us to check if $\lambda = 1/k$ indeed acheives a dual objective value of $\tau^*$. Note that this is a crucial step in the proof since the candidate $\vec{z}^*$ given to us is not known apriori to be optimal.

We have,

$$(\vec{z}^{*})^{\top}\vec{y} - \lambda^{*}\mu = k^{-1}(\sum_{i=1}^{k} y_i - \mu) = \tau^{*} + k^{-1}(\sum_{i=1}^{k}(y_i - \min(\tau^{*}, y_i)) - \mu) = \tau^{*}$$

where the second equality follows because for all $i \in [k]$, $\min(y_i, \tau^{*}) = \tau^{*}$ and the final equality follows because $\forall j \notin [k]$, $\min(y_i, \tau^{*}) = y_i$ which implies,

$$\mu = \sum_{i=1}^{d} y_i - \min(y_i, \tau^{*}) = \sum_{i=1}^{k} y_i - \min(y_i, \tau^{*}).$$

Therefore, we have a dual feasible point with objective value, $\tau^{*}$.

(f) (2 points) Using the results from the previous parts, **justify why $\tau^{*}$ is the optimal value of (??).**

**Solution:** Since, we have found primal and dual feasible points with the same objective value, we have:

$$\tau^{*} \leq d^{*} \leq p^{*} \leq \tau^{*}$$

which concludes the proof of the result.