EECS 127/227AT Optimization Models in Engineering

Spring 2019 **Final**

1. (1 Point) Tell us about a time that you succeeded this semester.

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

Do not turn this page until the proctor tells you to do so. You may work on the questions above.

PRINT your name and student ID: ____

3. (3 points) Convexity of functions

Let $g : \mathbb{R}^m \to \mathbb{R}$ be **any** convex function. For any $A \in \mathbb{R}^{m\times n}$, prove that $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(x) = g(Ax) \quad \forall x \in \mathbb{R}^n,$$

is also a convex function.

Solution: For every $\alpha \in [0, 1]$ and $x, y \in \mathbb{R}^n$, we have

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= g(A(\alpha x + (1-\alpha)y)) \\ &\le \alpha g(Ax) + (1-\alpha)g(Ay) \\ &= \alpha f(x) + (1-\alpha)f(y), \end{aligned}$$

which proves convexity of f.

PRINT your name and student ID: ______________________________

4. (2 points) Multiple Choice

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Consider the following optimization problems:

$$\begin{aligned} p_1^* = \min_{t\in\mathbb{R}, x\in\mathbb{R}^n} \quad & t \\ \text{s.t. } & \|x\|_2 = t, \\ & f(x) \le 0, \end{aligned} \tag{1}$$

$$\begin{aligned} p_2^* = \min_{t\in\mathbb{R}, x\in\mathbb{R}^n} \quad & t \\ \text{s.t. } & \|x\|_2 \le t, \\ & f(x) \le 0. \end{aligned} \tag{2}$$

Write the statement labels (A, B, C) corresponding to statements that are true in the box given below. More than one statement might be true; and you will get credit for this problem only if you write the labels corresponding to all statements that are true and do not write a label corresponding to any statement that is false. No justification is required.

(A) Problem (1) as written is a convex problem.

(B) Problem (2) as written is a convex problem.

(C) We necessarily have $p_1^* = p_2^*$.

Solution: Statement with labels B and C are true. Problem (1) as written is not a convex problem since we have an equality constraint which is not affine.
Problem (2) as written is a convex problem since both inequality constraints are convex.
Further $p_1^* = p_2^*$ due to the following argument.
Since the second problem is a relaxation of the first we have $p_1^* \ge p_2^*$. Next suppose we have (x^*, t^*) as candidate solutions for (2) with $\|x^*\|_2 = s < t^*$. Then we can decrease our objective value by using (x^*, s) which is feasible for both problems. Thus $p_1^* \le p_2^*$.

Print your name and student ID: ______________________________

5. (5 points) Connecting all Bears

A cellular service provider, BearT&T™, wants to place a base station in Berkeley so as to maximize the quality of service provided to its customers. Let $z_1, z_2, \ldots, z_m \in \mathbb{R}^2$ denote the fixed locations of the customers. The location for the new base station is given by the solution to:

$$\min_{x \in \mathbb{R}^2} \max_{i \in \{1,\ldots,m\}} \|x - z_i\|_2. \tag{3}$$

(a) (2 points) Explain why (3) is a convex problem.

Solution: The objective function

$$\max_{i \in \{1,\ldots,m\}} \|x - z_i\|_2$$

is pointwise maximum of convex functions of x; therefore, it is convex. Since there is no constraints, the problem is also convex.

(b) (3 points) Cast this problem in one of the standard convex optimization problems we have seen in class: LP, QP, QCQP or SOCP.

Solution: We can write this problem as an SOCP by introducing a slack variable:

$$\begin{aligned} \min_{x \in \mathbb{R}^2, t \in \mathbb{R}} \quad & t \\ \text{subject to} \quad & \|x - z_i\|_2 \leq t \quad \text{for} \quad i = 1, \ldots, m. \end{aligned}$$

PRINT your name and student ID: ____________________

6. (3 points) Newton's method

Given a symmetric positive definite matrix $Q \in \mathbb{S}^n_{++}$ and $b \in \mathbb{R}^n$, consider the minimization of the function $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(x) = \frac{1}{2} x^\top Q x - b^\top x.$$

Let x^* denote the point at which $f(x)$ is minimized, and define $\mathcal{B}(x^*)$ as the ball centered at x^* with unit ℓ_2-norm:

$$\mathcal{B}(x^*) = \{x \in \mathbb{R}^n : \|x - x^*\|_2 \leq 1\}.$$

Assume we use Newton's method to minimize f:

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k),$$

where the initial point is $x_0 \in \mathcal{B}(x^*)$. For any $k \in \mathbb{N}$, find

$$\max_{x_0 \in \mathcal{B}(x^*)} \|x_k - x^*\|_2.$$

Solution: Note that $\nabla^2 f(x_k) = Q$ for all $x_k \in \mathbb{R}^n$. Therefore, the update rule is:

$$x_{k+1} = x_k - Q^{-1} Q (x_k - x^*) = x^*,$$

so we have

$$x_k = x^* \quad \forall k \geq 1.$$

As a result,

$$\max_{x_0 \in \mathcal{B}(x^*)} \|x_k - x^*\|_2 = \begin{cases} 1 & k = 0, \\ 0 & \forall k \geq 1. \end{cases}$$

PRINT your name and student ID: ______________________________

7. (11 points) Linear algebra meets optimization

Let wide matrix $A \in \mathbb{R}^{m \times n}$ $(m < n)$ be full row rank.

(a) (2 points) Consider the ridge regression problem, where $b \in \mathbb{R}^m, x \in \mathbb{R}^n$ and the constant $\lambda > 0$ is given:

$$\min_{x} \quad ||Ax - b||_2^2 + \lambda ||x||_2^2 \tag{4}$$

Since this is a convex problem and the objective function is differentiable, the optimum can be found by setting the gradient to zero. Use this to find the optimal solution x^*.

Solution: Setting the gradient of the objective to 0 at optimum, we find that

$$2(A^\top A + \lambda I)x^* - 2A^\top b = 0$$

Since $A^\top A$ is PSD and λI is PD, it follows that $A^\top A + \lambda I$ is always invertible and hence

$$x^* = (A^\top A + \lambda I)^{-1} A^\top b$$

(b) (6 points) Now we rewrite the problem in (4) by adding a constraint

$$\min_{z = Ax - b} ||z||_2^2 + \lambda ||x||_2^2. \tag{5}$$

Let the Lagrangian corresponding to this problem be $\mathcal{L}(x, z, \nu)$, where ν is the dual variable corresponding to the equality constraint. Write out the dual function $g(\nu) = \inf_{x,z} \mathcal{L}(x, z, \nu)$ explicitly. Solve the dual problem to get ν^*. Find the corresponding values of $\tilde{x}, \tilde{z}$ such that $g(\nu^*) = \mathcal{L}(\tilde{x}, \tilde{z}, \nu^*)$.

Solution: The dual problem is

$$\max_{\nu} g(\nu)$$

where

$$\begin{aligned} g(\nu) &= \min_{x.z} \mathcal{L}(x, z, \nu) \\ &= \min_{x,z} \|z\|^2 + \lambda \|x\|^2 + \nu^\top (z - Ax + b) \end{aligned}$$

First we minimize over x. Setting the gradient to 0 we have that

$$2\lambda x^* - A^\top \nu = 0 \implies x^* = \frac{1}{2\lambda} A^\top \nu$$

Setting the gradient to 0 for z we have that

$$2z^* + \nu = 0 \implies z^* = -\frac{1}{2}\nu$$

Plugging back in and simplifying the expression we have that

$$g(\nu) = \nu^\top b - \frac{1}{4}\|\nu\|_2^2 - \frac{1}{4\lambda}\|A^\top \nu\|_2^2$$

Maximizing over ν again amounts to setting the gradient to 0 at optimum. Hence we have

$$b - \frac{1}{2}\nu^* - \frac{1}{2\lambda}AA^\top \nu^* = 0 \implies \nu^* = 2(\frac{1}{\lambda}AA^\top + I)^{-1}b = 2\lambda(AA^\top + \lambda I)^{-1}b$$

It follows that

$$\begin{aligned}\tilde{x} &= A^\top(AA^\top + \lambda I)^{-1}b \\ \tilde{z} &= A\tilde{x} - b \\ &= A(A^\top(AA^\top + \lambda I)^{-1}b) - b.\end{aligned}$$

(c) (3 points) Show that for every $\lambda > 0$,

$$(A^\top A + \lambda I)^{-1}A^\top b = A^\top(AA^\top + \lambda I)^{-1}b.$$

Hint: One approach is to start by considering $\lambda A^\top + A^\top AA^\top$. Another approach is to use the SVD of A.

Solution:

Method 1: Let A have the thin SVD,

$$A = U\Sigma V^\top,$$

where $U \in \mathbb{R}^{m\times m}, \Sigma \in \mathbb{R}^{m\times m}, V^\top \in \mathbb{R}^{m\times n}$.

Using the SVD of A, $A^\top(AA^\top + \lambda I)^{-1}$ evaluates to,

$$\begin{aligned}A^\top(AA^\top + \lambda I_m)^{-1} &= V\Sigma U^\top(U\Sigma V^\top V\Sigma U^\top + \lambda I_m)^{-1} \\ &= V\Sigma U^\top(U\Sigma^2 U^\top + U\lambda I_m U^\top)^{-1} \\ &= V\Sigma U^\top(U(\Sigma^2 + \lambda I_m)^{-1}U^\top) \\ &= V\Sigma(\Sigma^2 + \lambda I_m)^{-1}U^\top.\end{aligned}$$

Note that $\Sigma^2 + \lambda I_m$ is invertible because A is full row rank and $\lambda > 0$.

Next we evaluate $(A^\top A + \lambda I)^{-1}A^\top$. We have,

$$\begin{aligned}(A^\top A + \lambda I_n)^{-1}A^\top &= (V\Sigma U^\top U\Sigma V^\top + \lambda I_n)^{-1}V\Sigma U^\top \\ &= (V\Sigma^2 V^\top + V\lambda I_m V^\top)^{-1}V\Sigma U^\top \\ &= (V(\Sigma^2 + \lambda I_m)V^\top)^{-1}V\Sigma U^\top \\ &= V(\Sigma^2 + \lambda I_m)^{-1}V^\top V\Sigma U^\top \\ &= V(\Sigma^2 + \lambda I_m)^{-1}\Sigma U^\top \\ &= V\Sigma(\Sigma^2 + \lambda I_m)^{-1}U^\top,\end{aligned}$$

where in the last equality we can interchange order of matrices since they are both diagonal.

Thus we have, $A^\top A + \lambda I)^{-1} A^\top = A^\top (AA^\top + \lambda I)^{-1}$,
which gives us,
$(A^\top A + \lambda I)^{-1} A^\top b = A^\top (AA^\top + \lambda I)^{-1} b$.
Method 2: Note

$$\lambda A^\top + A^\top A A^\top = A^\top(\lambda I + AA^\top) = (\lambda I + A^\top A)A^\top$$

Hence we have that

$$(\lambda I + A^\top A)^{-1} A^\top = A^\top(\lambda I + AA^\top)^{-1}$$

The result follows. Note that both $\lambda I + A^\top A$ and $\lambda I + AA^\top$ are invertible since they have strictly positive eigenvalues and are hence positive definite.

PRINT your name and student ID: ____________________

8. (9 points) A matrix optimization problem

Consider the following optimization problem

$$\begin{aligned} \min_{X \in \mathbb{R}^{n\times n}} \quad & \frac{1}{2}\|X\|_F^2 \\ \text{s.t.} \quad & X \in \mathcal{S}, \end{aligned}$$

where $\mathcal{S} = \{A \in \mathbb{R}^{n\times n} \mid \sigma_{\min}(A) \geq 2\}$ and $\sigma_{\min}(A)$ refers to the smallest singular value of A.

(a) (2 points) Is the objective function convex? Justify.

Solution:

Solution 1: Given $X, Y \in \mathbb{R}^{n\times n}$ and $\alpha \in [0,1]$, we have

$$\begin{aligned} \|\alpha X + (1-\alpha)Y\|_F^2 &= \langle \alpha X + (1-\alpha)Y, \alpha X + (1-\alpha)Y \rangle \\ &= \alpha^2\|X\|_F^2 + (1-\alpha)^2\|Y\|_F^2 + 2\alpha(1-\alpha)\langle X, Y\rangle \\ &\leq \alpha^2\|X\|_F^2 + (1-\alpha)^2\|Y\|_F^2 + 2\alpha(1-\alpha)\|X\|_F\|Y\|_F \\ &\leq \alpha^2\|X\|_F^2 + (1-\alpha)^2\|Y\|_F^2 + \alpha(1-\alpha)\left(\|X\|_F^2 + \|Y\|_F^2\right) \\ &= \alpha\|X\|_F^2 + (1-\alpha)\|Y\|_F^2, \end{aligned}$$

where the first inequality follows from Cauchy-Schwarz inequality and the second inequality follows from the fact that $ab \leq \frac{1}{2}(a^2+b^2)$ for all $a, b \in \mathbb{R}$. This shows that the objective function is convex.

Solution 2: Consider $g : \mathbb{R}^{n\times n} \to \mathbb{R}_+$ and $f : \mathbb{R}_+ \to \mathbb{R}$ defined as

$$g(X) = \|X\|_F,$$

$$f(z) = \frac{1}{2}z^2.$$

Then the objective function is

$$\frac{1}{2}\|X\|_F^2 = f \circ g(X).$$

Note that g is a convex function because it is a norm. In addition, f is a convex function and it is strictly increasing on its domain. Therefore, their composition $f \circ g$, which is the objective function, is also a convex function.

(b) (3 points) Is the constraint set convex? Justify.

Solution: Assume $X \in \mathcal{S}$. Then $\sigma_{\min}(X) = \sigma_{\min}(-X) \geq 2$. Consider the point $Y = \frac{1}{2}X + \frac{1}{2}(-X) = \mathbf{0}$. Clearly Y is a convex combination of two points in $\mathcal{S}$, but $\sigma_{\min}(Y) = 0$, and therefore, $Y \notin \mathcal{S}$. As a result, the constraint set $\mathcal{S}$ is not convex.

Common mistake:

i. Claim that $\sigma_{\min}(X) = \min_{\|u\|_2^2 \leq 1, \|v\|_2^2 \leq 1} u^\top X v$. This is not true, as the right hand side is minimized to the negative of the largest singular value.

ii. Claim that $\sigma_{\min}$ is convex or concave. This is not true, as the smallest singular value of a matrix is actually neither convex nor concave.

(c) (4 points) By using the singular value decomposition of X, rewrite the objective function and constraints in terms of the singular values of X and find a solution X^*.

Solution:

Let $X = U\Sigma V^\top$ denote the SVD decomposition of X. Then,

$$\|X\|_F^2 = \text{trace}(V\Sigma U^\top U\Sigma V^\top) = \text{trace}(\Sigma^2) = \sum_{i=1}^n \sigma_i^2.$$

Then we can write the optimization problem as

$$\begin{aligned} \min_{U,\Sigma,V} \quad & \sqrt{\sum_{i=1}^n \sigma_i^2} \\ \text{s.t.} \quad & \sigma_{\min} \geq 2. \end{aligned}$$

Any matrix in $\mathbb{R}^{n\times n}$ which has $\sigma_1 = \cdots = \sigma_n = 2$ is a solution for this problem. In other words, the solution set for the original problem is given as

$$\left\{2UV^\top : U, V \in \mathbb{R}^{n\times n},\ UU^\top = U^\top U = VV^\top = V^\top V = I\right\}.$$

PRINT your name and student ID: ______________________________

9. (8 points + 5 bonus points) Energy functions for linear systems

Given $A \in \mathbb{R}^{n \times n}$ and $x_0 \in \mathbb{R}^n$, consider the linear-time-invariant system with no input:

$$x_{k+1} = Ax_k \quad \forall k \in \mathbb{N}. \tag{6}$$

Let P be a symmetric positive semidefinite matrix, and let Q be a symmetric positive definite matrix in $\mathbb{R}^{n \times n}$; that is, $P \in \mathbb{S}^n_+$ and $Q \in \mathbb{S}^n_{++}$. Assume P and Q satisfy

$$A^\top PA - P \preceq -Q, \tag{7}$$

and define $V : \mathbb{R}^n \to \mathbb{R}$ as

$$V(x_k) = x_k^\top Px_k \quad \forall x_k \in \mathbb{R}^n.$$

(a) (2 points) Show that $V(x_{k+1}) - V(x_k) \leq 0$ for all $k \in \mathbb{N}$.
Hint: Use (7)*.*

Solution: For every $x_k \in \mathbb{R}^n$, we have

$$\begin{aligned} V(x_{k+1}) - V(x_k) &= x_{k+1}^\top Px_{k+1} - x_k^\top Px_k \\ &= x_k^\top (A^\top PA - P)x_k \\ &\leq x_k^\top(-Q)x_k \\ &\leq 0, \end{aligned}$$

where the last inequality follows from the fact that $Q \in \mathbb{S}^n_{++}$.

(b) (4 points) Find a constant $\beta \in (0, \infty)$ in terms of eigenvalues of P and Q such that

$$V(x_{k+1}) \leq (1 - \beta)V(x_k) \quad \forall k \in \mathbb{N}.$$

Hint: If you can find $\alpha, \gamma \in (0, \infty)$ *such that*

$$x_k^\top Qx_k \geq \alpha \|x_k\|_2^2 \quad \forall x_k \in \mathbb{R}^n,$$
$$x_k^\top Px_k \leq \gamma \|x_k\|_2^2 \quad \forall x_k \in \mathbb{R}^n,$$

then

$$-x_k^\top Qx_k \leq -\frac{\alpha}{\gamma} x_k^\top Px_k = -\frac{\alpha}{\gamma} V(x_k) \quad \forall x_k \in \mathbb{R}^n.$$

Solution: For every $x_k \in \mathbb{R}^n$, we have

$$x_k Qx_k \geq \lambda_{\min}(Q)\|x_k\|_2^2,$$
$$\lambda_{\max}(P)\|x_k\|_2^2 \geq x_k Px_k,$$

which gives

$$x_k^\top Qx_k \geq \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)} x_k^\top Px_k = \beta V(x_k),$$

where we have defined $\beta := \lambda_{\min}(Q)/\lambda_{\max}(P)$. Then, for every $x_k \in \mathbb{R}^n$:

$$\begin{aligned} A^\top PA - P \preceq Q &\implies x_k^\top A^\top PAx_k - x_k^\top Px_k \leq -x_k^\top Qx_k \\ &\implies V(x_{k+1}) - V(x_k) \leq -\beta V(x_k). \end{aligned}$$

(c) (2 points) Along with the nonnegativity of $V(x_k)$, part (b) shows that $V(x_k)$ converges to zero for every initialization. In analysis of dynamical systems, functions like V are used to represent the energy in the system, and its convergence to zero indicates the dissipation of this energy. For this reason, finding V is an important problem in the study of dynamical systems.

Given $A \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{S}^n_{++}$ for system (6), we can use the following optimization problem to find the matrix P defining the function V:

$$\begin{aligned} \min_{P\in\mathbb{S}^n_+} \quad & 0 \\ \text{subject to} \quad & A^\top PA - P + Q \preceq 0. \end{aligned} \tag{8}$$

Only for this part of the question, consider the scalar version of problem (8):

$$\begin{aligned} \min_{p\in\mathbb{R}} \quad & 0 \\ \text{subject to} \quad & p \geq 0 \\ & a^2p - p + q \leq 0, \end{aligned} \tag{9}$$

where $a, q \in \mathbb{R}$ are some fixed constants and $q > 0$. Under what conditions on a and q, is the problem (9) feasible?

Solution: For feasibility of the problem, we need constraint set to be nonempty, i.e.,

$$\mathcal{S} = \{p \in \mathbb{R} : p \geq 0,\ (a^2-1)p \leq -q\}$$

must be nonempty. Since $q > 0$, the term $a^2 - 1$ must be negative, or equivalently, we must have $a^2 < 1$. Note that this condition corresponds to the stability of the dynamical system.

(d) (**Bonus**: 5 points) Consider the non-scalar problem (8) again. Find the dual problem corresponding to (8) by explicitly deriving the dual function and the feasibility constraints of the dual problem. Do **not** dualize the constraint $P \in \mathbb{S}^n_+$.

Hint: For any $Y \in \mathbb{S}^n$, we have

$$\max_{\Lambda\succeq 0} \langle Y, \Lambda\rangle = \begin{cases} 0 & \textit{if } Y \preceq 0 \\ +\infty & \textit{otherwise.} \end{cases}$$

Solution: We know that an optimization problem given as

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{C} \end{aligned}$$

is equivalent to

$$\min_x \quad f(x) + h(x),$$

where

$$h(x) = \begin{cases} 0 & x \in \mathcal{C}, \\ +\infty & x \notin \mathcal{C}. \end{cases}$$

Therefore, as suggested by the hint, we can write

$$p^* = \min_{P\in\mathbb{S}^n_+} \max_{\Lambda\in\mathbb{S}^n_+} \langle A^\top PA - P + Q, \Lambda\rangle.$$

Then the dual problem is given by

$$d^* = \max_{\Lambda \in \mathbb{S}^n_+} \; g(\Lambda),$$

where the dual function $g(\Lambda)$ is

$$g(\Lambda) = \min_{P \in \mathbb{S}^n_+} \; \langle A^\top P A - P + Q, \Lambda \rangle.$$

By rearranging terms:

$$\begin{aligned} g(\Lambda) &= \min_{P \in \mathbb{S}^n_+} \; \langle P, A \Lambda A^\top - \Lambda \rangle + \langle Q, \Lambda \rangle \\ &= \langle Q, \Lambda \rangle - \max_{P \in \mathbb{S}^n_+} \; \langle P, -A \Lambda A^\top + \Lambda \rangle \\ &= \begin{cases} \langle Q, \Lambda \rangle & \text{if } \; \Lambda - A \Lambda A^\top \preceq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

where the last equality follows from the hint. Then the dual problem can be written as

$$\begin{aligned} \max_{\Lambda \in \mathbb{S}^n_+} & \quad \langle Q, \Lambda \rangle \\ \text{subject to} & \quad \Lambda - A \Lambda A^\top \preceq 0. \end{aligned}$$

PRINT your name and student ID: ____

10. (6 points) Minimizing quadratics

Consider the following optimization problem:

$$p^* = \inf_{x \in \mathbb{R}^2} x^\top A x + b^\top x,$$

where $A \in \mathbb{S}^2_+$ and $b \in \mathbb{R}^2$.

(a) (3 points) Suppose $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Find a vector b with $\|b\|_2 = 1$ such that $p^* > -\infty$.
Hint: Is A invertible?

Solution: A has following eigenvalue, eigenvector pairs:

$$\lambda_1 = 2, v_1 = [1, 1]^\top$$
$$\lambda_2 = 0, v_2 = [1, -1]^\top.$$

For p^* to be finite we need b to be orthogonal to v_2, the eigenvector corresponding to 0 eigenvalue. This along with the condition that $\|b\|_2 = 1$ gives $b = \frac{1}{\sqrt{2}}[1, 1]^\top$.

(b) (3 points) Now assume A is a symmetric positive definite matrix, i.e. $A \in \mathbb{S}^2_{++}$ and $b = [0, 0]^\top$. Suppose we add a ℓ_∞-norm regularizer term to the objective to get the following optimization problem:

$$p^* = \inf_{x \in \mathbb{R}^2} x^\top A x + \|x\|_\infty .$$

Write the corresponding dual problem as

$$\begin{aligned} d^* = \sup_{y \in \mathbb{R}^2} \quad & g(y) \\ \text{subject to} \quad & \|y\|_c \leq 1, \end{aligned}$$

where you will determine $g(y)$ and c.
Hint: For every $x \in \mathbb{R}^2$, we have

$$\sup_{y \in \mathbb{R}^2 \,:\, \|y\|_1 \leq 1} x^\top y = \|x\|_\infty .$$

Solution:

Consider the Lagrangian,

$$\mathcal{L}(x, y) = x^\top A x + x^\top y.$$

Then,

$$p^* = \inf_{x \in \mathbb{R}^2} \sup_{y \in \mathbb{R}^2 \,:\, \|y\|_1 \leq 1} x^\top A x + x^\top y,$$

where we used the hint,

$$\sup_{y \in \mathbb{R}^2 \,:\, \|y\|_1 \leq 1} x^\top y = \|x\|_\infty .$$

Switching the order of min and max we get the dual problem,

$$d^* = \sup_{y \in \mathbb{R}^2 \,:\, \|y\|_1 \leq 1} \inf_{x \in \mathbb{R}^2} x^\top A x + x^\top y.$$

Consider the inner minimization problem. The objective function is strictly convex since A is positive definite and since the problem is unconstrained we can find the optimizer by setting the derivative to zero. Setting derivative to zero we get,

$$\begin{aligned} &2Ax + y = 0 \\ &\implies x = -\frac{1}{2}A^{-1}y. \end{aligned}$$

A is invertible so A^{-1} exists.
Substituting this value of x we get,

$$d^* = \sup_{y \in \mathbb{R}^2 \,:\, \|y\|_1 \leq 1} -\frac{1}{4}y^\top A^{-1} y.$$

We can put this in the form asked in question as,

$$\begin{aligned} d^* = &\sup_{y \in \mathbb{R}^2} -\frac{1}{4}y^\top A^{-1} y \\ &\text{s.t. } \|y\|_1 \leq 1. \end{aligned}$$

PRINT your name and student ID: ____________________

11. (13 points) A matrix game

Let $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$ be a payoff matrix for two games as described in the parts below. Suppose row player, R, chooses action x and column player, C, chooses action y, then both players get the payoff $s = x^\top Ay$. R wishes to minimize payoff, while C wishes to maximize payoff.

(a) Suppose $x \in \mathcal{E}, y \in \mathcal{E}$, where $\mathcal{E} = \{[0,1]^\top, [1,0]^\top\}$.

i. (3 points) Suppose R chooses x first and then C chooses y. The optimal payoff s_R^* is given by

$$s_R^* = \min_{x\in\mathcal{E}} \max_{y\in\mathcal{E}} x^\top Ay.$$

For the given matrix A, $s_R^* = 3$ achieved for $x^* = [1,0]^\top, y^* = [0,1]^\top$.

Now suppose C chooses y first and then R chooses x. The optimal payoff s_C^* is given by,

$$s_C^* = \max_{y\in\mathcal{E}} \min_{x\in\mathcal{E}} x^\top Ay.$$

Find s_C^* for the given matrix A. Justify your answer.

Solution:

For $y = [0,1]^\top$, the inner minimization is given by,

$$\begin{aligned}\min_{x\in\mathcal{E}} x^\top A[0,1]^\top &= \min_{x\in\mathcal{E}} x^\top [3,2]^\top \\ &= \min(3,2) \\ &= 2,\end{aligned}$$

achieved for $x^* = [0,1]^\top$.

For $y = [1,0]^\top$, the inner minimization is given by,

$$\begin{aligned}\min_{x\in\mathcal{E}} x^\top A[1,0]^\top &= \min_{x\in\mathcal{E}} x^\top [1,4]^\top \\ &= \min(1,4) \\ &= 1,\end{aligned}$$

achieved for $x^* = [1,0]^\top$. Thus $s_C^* = 2$ achieved for $x^* = [0,1]^\top, y^* = [0,1]^\top$.

ii. (1 point) Compare s_R^* to s_C^*. Who is better off — the first player or the second player?

Solution: We have $s_R^* = 3 > 2 = s_C^*$. If R goes first both players get a larger payoff than if R goes second so R prefers to go second. Similarly C prefers to go second. Thus it is better to go second in this game and react to other player's action.

iii. (2 points) Now suppose A was unknown. Does your choice of whether to go first or second remain the same? Justify.

Solution: By weak duality we have,

$$s_R^* = \min_{x\in\mathcal{E}} \max_{y\in\mathcal{E}} x^\top Ay \geq \max_{y\in\mathcal{E}} \min_{x\in\mathcal{E}} x^\top Ay = s_C^*.$$

Thus it is better for a player to go second irrespective of what the payoff matrix A is.

(b) Suppose $x \in \mathcal{P}, y \in \mathcal{P}$ where $\mathcal{P} = \{z = [z_1, z_2]^\top \in \mathbb{R}^2 \mid z_1 \geq 0, z_2 \geq 0, z_1 + z_2 = 1\}$. Suppose R chooses x first and then C chooses y. Let p_R^* denote the optimal payoff in this case given by,

$$p_R^* = \min_{x \in \mathcal{P}} \max_{y \in \mathcal{P}} \ x^\top A y.$$

i. (3 points) For a given $x \in \mathcal{P}$, show that $\max_{y \in \mathcal{P}} x^\top A y = \max_{y \in \mathcal{E}} x^\top A y$.
Hint: Show that

$$\max_{y \in \mathcal{P}} \ x^\top A y \leq \max_{y \in \mathcal{E}} \ x^\top A y,$$
$$\max_{y \in \mathcal{P}} \ x^\top A y \geq \max_{y \in \mathcal{E}} \ x^\top A y.$$

Solution: Let $(x^\top A)_i$ refer to i^{th} entry of row vector $x^\top A$. Then we have for every $y \in \mathcal{P}$,

$$\begin{aligned} x^\top A y &= (x^\top A)_1 y_1 + (x^\top A)_2 y_2 \\ &\leq \max((x^\top A)_1, (x^\top A)_2)(y_1 + y_2) \\ &= \max((x^\top A)_1, (x^\top A)_2) \\ &= \max_{y \in \mathcal{E}} x^\top A y, \end{aligned}$$

where the first inequality follows since $y_1, y_2 \geq 0$ and the second inequality follows since $y_1 + y_2 = 1$. Since this is true for every $y \in \mathcal{P}$ we have,

$$\max_{y \in \mathcal{P}} x^\top A y \leq \max_{y \in \mathcal{E}} x^\top A y. \tag{10}$$

Since $\mathcal{E} \subset \mathcal{P}$ we also have,

$$\max_{y \in \mathcal{P}} x^\top A y \geq \max_{y \in \mathcal{E}} x^\top A y. \tag{11}$$

From (10), (11), we have

$$\max_{y \in \mathcal{P}} x^\top A y = \max_{y \in \mathcal{E}} x^\top A y.$$

ii. (4 points) Formulate a Linear Program with finitely many constraints to find p_R^*, which is equivalent to

$$p_R^* = \min_{x \in \mathcal{P}} \max_{y \in \mathcal{E}} \ x^\top A y$$

due to result of part (i).
Solution: We have,

$$\begin{aligned} p_R^* &= \min_{x \in \mathcal{P}} \max_{y \in \mathcal{P}} x^\top A y \\ &= \min_{x \in \mathcal{P}} \max_{y \in \mathcal{E}} x^\top A y, \end{aligned}$$

using the result of the previous part.We first introduce a slack variable as follows,

$$
\begin{aligned}
p_R^* = &\min_{x\in\mathcal{P}, v\in\mathbb{R}} \ v \\
\text{s.t.} \ \ & x^\top A \begin{bmatrix} 0 \\ 1 \end{bmatrix} \le v \\
& x^\top A \begin{bmatrix} 1 \\ 0 \end{bmatrix} \le v.
\end{aligned}
$$

We can write this as an LP as follows:

$$
\begin{aligned}
p_R^* = &\min_{x\in\mathbb{R}^2, v\in\mathbb{R}} \ v \\
\text{s.t.} \ \ & x^\top A \begin{bmatrix} 0 \\ 1 \end{bmatrix} \le v \\
& x^\top A \begin{bmatrix} 1 \\ 0 \end{bmatrix} \le v \\
& x \ge 0 \\
& x^\top \mathbf{1} = 1.
\end{aligned}
$$

,

PRINT your name and student ID: ______________________________

12. (5 points + 3 bonus points) Soft-margin SVM

Consider the soft-margin SVM problem,

$$p^*(C) = \min_{w\in\mathbb{R}^m, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \xi_i \tag{12}$$
$$\text{s.t. } 1-\xi_i - y_i(x_i^\top w - b) \le 0, \quad i = 1,2,\ldots,n$$
$$-\xi_i \le 0, \quad i = 1,2,\ldots,n,$$

where $x_i \in \mathbb{R}^m$ refers to the i^{th} training data point, $y_i \in \{-1,1\}$ is its label, and $C \in \mathbb{R}_+$ (i.e. $C > 0$) is a hyperparameter.

Let α_i denote the dual variable corresponding to the inequality $1-\xi_i - y_i(x_i^\top w - b) \le 0$ and let β_i denote the dual variable corresponding to the inequality $-\xi_i \le 0$.

The Lagrangian is then given by

$$\mathcal{L}(w,b,\xi,\alpha,\beta) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i(1-\xi_i - y_i(x_i^\top w - b)) - \sum_{i=1}^n \beta_i\xi_i.$$

Suppose $w^*, b^*, \xi^*, \alpha^*, \beta^*$ satisfy the KKT conditions.

Classify the following statements as true or false. **Justify your answers mathematically**. A correct answer with missing or incorrect justification will be given 0 points.

(a) (3 points) Suppose the optimal solution w^*, b^* changes when the training point x_i is removed. Then originally, we necessarily have $y_i(x_i^\top w^* - b^*) = 1 - \xi_i^*$.

Solution: True. Since optimal w^* changes if we remove point x_i we have $\alpha_i^* \neq 0$. By complementary slackness we have,

$$\alpha_i^*(1-\xi_i^* - y_i(x_i^\top w^* - b^*)) = 0,$$

which gives,

$$1-\xi_i^* - y_i(x_i^\top w^* - b^*) = 0$$
$$\implies y_i(x_i^\top w^* - b^*) = 1-\xi_i^*.$$

(b) (2 points) Suppose the optimal solution w^*, b^* changes when the training point x_i is removed. Then originally, we necessarily have $\alpha_i^* > 0$.

Solution: True. Since optimal w^* changes if we remove point x_i we have $\alpha_i^* \neq 0$. Further by dual feasibility we have $\alpha_i^* \geq 0$ which together gives $\alpha_i^* > 0$.

(c) (**Bonus**: 3 points) Suppose the data points are strictly linearly separable, i.e. there exist $\tilde{w}$ and $\tilde{b}$ such that for all i,

$$y_i(x_i^\top \tilde{w} - \tilde{b}) > 0.$$

Then $p^*(C) \to \infty$ as $C \to \infty$.

Solution: False.
Since

$$y_i(x_i^\top \tilde{w} - \tilde{b}) > 0.$$

we have for sufficiently small $\epsilon > 0$,

$$\begin{aligned} & y_i(x_i^\top \tilde{w} - \tilde{b}) \geq \epsilon \\ \implies & y_i \left(x_i^\top \frac{\tilde{w}}{\epsilon} - \frac{\tilde{b}}{\epsilon} \right) \geq 1. \end{aligned}$$

Thus, $\bar{w} = \frac{\tilde{w}}{\epsilon}$, $\bar{b} = \frac{\tilde{b}}{\epsilon}, \bar{\xi} = 0$ is a feasible point with objective value $\frac{1}{2} \|\bar{w}\|_2^2 < \infty$ irrespective of value of C.