Final

1. **Honor Code (0 pts)**

   Please copy the following statement in the space provided below and sign your name.

   As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules and do this exam on my own.

   IF YOU DO NOT COPY THE HONOR CODE AND SIGN YOUR NAME,
   YOU WILL GET A 0 ON THE EXAM.

Solution:

2. **SID and Exam Location (2 pts)**

   BEFORE THE EXAM STARTS, WRITE YOUR SID AT THE TOP OF FIRST PAGE AND LAST PAGE.
   ALSO MENTION YOUR EXAM LOCATION AT THE TOP OF FIRST PAGE.
   No extra time will be given for this task.

3. **What is your favorite sport? Any answer, including no answer, will be given full credit. (2 pts)**

   Solution: Any answer is fine.

4. Descent Methods (12 pts)

Consider the least squares problem

$$\min_{\vec{x}\in\mathbb{R}^n} \frac{1}{2}\|A\vec{x}-\vec{b}\|_2^2,$$

for some $\vec{b}\in\mathbb{R}^n$, $A\in\mathbb{R}^{n\times n}$ such that $A\neq O_{n\times n}$.

(a) (4 pts) What is the most restrictive class of optimization problem that this problem belongs to from the following options? No justification is required.

NOTE: In this problem, choosing "I don't know" in a multiple choice question merits **1 point** to discourage guessing. Picking more than one of the choices will automatically get 0 points.

HINT: Recall that $LP \subseteq QP \subseteq QCQP \subseteq SOCP \subseteq$ *Convex optimization problems*.

(1) LP

(2) QP

(3) QCQP

(4) SOCP

(5) Convex Optimization Problem

(6) I don't know

Solution: QP.

(b) (4 pts) Suppose

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \vec{b} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}.$$

Consider running gradient descent for the least squares problem

$$\min_{\vec{x} \in \mathbb{R}^2} \frac{1}{2} \|A\vec{x} - \vec{b}\|_2^2,$$

with step size $\eta = 1$ and initialization $\vec{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. What is the next iterate $\vec{x}^{(1)}$? Please provide a numerical value for each entry of $\vec{x}^{(1)}$.

HINT: Note that $A^\top \vec{b} = \begin{bmatrix} 22 \\ 32 \end{bmatrix}$, $A^\top A = \begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix}$, and $A^{-1}\vec{b} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$.

Solution: The gradient is $A^\top(A\vec{x} - \vec{b})$. Evaluated at $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, this is $-A^\top \vec{b} = \begin{bmatrix} -22 \\ -32 \end{bmatrix}$. So, the next iterate is $\vec{0} - 1 \cdot \begin{bmatrix} -22 \\ -32 \end{bmatrix} = \begin{bmatrix} 22 \\ 32 \end{bmatrix}$.

(c) (4 pts) Suppose

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \vec{b} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}.$$

Consider running Newton's method for the least squares problem

$$\min_{\vec{x} \in \mathbb{R}^2} \frac{1}{2} \|A\vec{x} - \vec{b}\|_2^2,$$

with initialization $\vec{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. What is the next iterate $\vec{x}^{(1)}$? Please provide a numerical value for each entry of $\vec{x}^{(1)}$.

HINT: Note that $A^\top \vec{b} = \begin{bmatrix} 22 \\ 32 \end{bmatrix}$, $A^\top A = \begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix}$, *and* $A^{-1}\vec{b} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$.

Solution: Newton's Method converges in one iteration for this QP. So, the next iterate is $\vec{x}^{(1)} = A^{-1}\vec{b} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$, the minimizer.

5. Duality Trivia (20 pts)

Let $f_0, f_1, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ be given. Suppose we have the primal problem

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad f_0(\vec{x})$$
$$\text{s.t.} \quad f_i(\vec{x}) \le 0, \qquad \text{for each } i \in \{1, \cdots, m\},$$

where the dual function is $g(\vec{\lambda}) = \min_{\vec{x} \in \mathbb{R}^n} L(\vec{x}, \vec{\lambda})$ for the Lagrangian $L(\vec{x}, \vec{\lambda})$. For all subparts in this problem, no justification is required.

NOTE: In this problem, choosing "I don't know" in a multiple choice question merits **1 point** to discourage guessing. On the other hand, "None of the above" means that we can't say any of the above statements in generality (different scenarios could occur in different examples). In each subpart, picking more than one of the choices will automatically get 0 points.

(a) (4 pts) If $f_i(\vec{x})$ is a convex function for all $i = 0, \cdots, m$, then what can we say about $g(\vec{\lambda})$?

NOTE: The constant functions $-\infty$ and $+\infty$ are considered to be both convex and concave.

(1) $g(\vec{\lambda})$ is convex in $\vec{\lambda}$.
(2) $g(\vec{\lambda})$ is concave in $\vec{\lambda}$.
(3) $g(\vec{\lambda})$ is neither convex nor concave in $\vec{\lambda}$.
(4) None of the above.
(5) I don't know.

Solution: $g(\vec{\lambda})$ is concave in $\vec{\lambda}$.

(b) (4 pts) If $f_i(\vec{x})$ is neither convex nor concave, for all $i = 0, \cdots, m$, then what can we say about $g(\vec{\lambda})$?

NOTE: The constant functions $-\infty$ and $+\infty$ are considered to be both convex and concave.

(1) $g(\vec{\lambda})$ is convex in $\vec{\lambda}$.
(2) $g(\vec{\lambda})$ is concave in $\vec{\lambda}$.
(3) $g(\vec{\lambda})$ is neither convex nor concave in $\vec{\lambda}$.
(4) None of the above.
(5) I don't know.

Solution: $g(\vec{\lambda})$ is concave in $\vec{\lambda}$.

(c) (4 pts) Suppose $p^\star = 0$, and $f_0(\vec{x})$ is linear, and $f_i(\vec{x})$ are affine for $i = 1, \cdots, m$. What is the strongest statement we can make about $d^\star$?

(1) $d^\star = 0$.
(2) $d^\star \le 0$.
(3) $d^\star \ge 0$.
(4) None of the above.
(5) I don't know.

Solution: $d^\star = 0$ since it's a feasible LP.

(d) (4 pts) Suppose $p^\star = 0$, and $f_i(\vec{x})$ is a convex function for each $i \in \{0, 1, \cdots, m\}$. Moreover, there exists a vector $\vec{c} \in \mathbb{R}^n$ that satisfies $\max_{i \in \{1, \cdots, m\}} f_i(\vec{c}) = -3$. What is the strongest statement we can make about $d^\star$?

(1) $d^\star = 0$.

(2) $d^\star \leq 0$.

(3) $d^\star \geq 0$.

(4) None of the above.

(5) I don't know.

Solution: $d^\star = 0$ since Slater's condition holds for $\vec{c}$.

(e) (4 pts) Suppose $p^\star = 0$, and $f_i(\vec{x})$ is a convex quadratic function of $\vec{x}$ for each $i \in \{0, 1, \cdots, m\}$. What is the strongest statement we can make about $d^\star$?

(1) $d^\star = 0$.

(2) $d^\star \leq 0$.

(3) $d^\star \geq 0$.

(4) None of the above.

(5) I don't know.

Solution: $d^\star \leq 0$ by weak duality. It is a QCQP, and we cannot ascertain strong duality.

6. Norms and Linear Algebra (12 pts)

Let $A \in \mathbb{R}^{m \times n}$ with $m \leq n$, with columns $\vec{a}_1$ through $\vec{a}_n$, i.e., $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix}$. Suppose $AA^\top = I_{m \times m}$, where $I_{m \times m}$ denotes the $m \times m$ identity matrix.

(a) (4 pts) Find the average of the squared norms of the columns of A, i.e. $\frac{1}{n} \sum_{i=1}^{n} \|\vec{a}_i\|_2^2$.

Solution: Note that $\sum_i \|\vec{a}_i\|_2^2 = \|A\|_F^2 = \text{tr}(AA^\top) = m$. Thus, the average of the squared norms is $\dfrac{m}{n}$.

(b) (8 pts) Show that the norm of every column of A is at most 1, i.e. $\|\vec{a}_i\|_2 \leq 1$ for all $i \in \{1, \cdots, n\}$.

HINT: Recall that the spectral norm of a matrix equals its largest singular value.

Solution: There are at least two solutions.

First solution: Note that A has m orthonormal rows since $AA^\top = I_m$. By the Basis Extension Theorem, there exist $n-m$ row vectors such that, together with the m rows of A, form an orthonormal basis for $\mathbb{R}^n$. Let $A' \in \mathbb{R}^{n\times n}$ denote the resulting matrix, which can be written in the form:

$$A' = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \\ \vec{q}_1 & \vec{q}_2 & \dots & \vec{q}_n \end{bmatrix}$$

where $\vec{q}_1, \cdots, \vec{q}_n \in \mathbb{R}^{n-m}$. Since A' is an orthogonal matrix, the columns of A' must have norm 1 since they are orthonormal. So, $1 = \|\vec{a}_i\|_2^2 + \|\vec{q}_i\|_2^2$, so $\|\vec{a}_i\|_2^2 \leq 1$.

Second solution: Since $AA^\top = I_{m\times m}$, the matrix A has m singular values, each of which equals 1. Thus, for each $i \in \{1, \cdots, n\}$:

$$\|\vec{a}_i\|_2 \leq \frac{\|A\vec{e}_i\|_2}{\|\vec{e}_i\|_2} \leq \max_{\substack{\vec{v}\in\mathbb{R}^n: \\ \|\vec{v}\|_2 \leq 1}} \frac{\|A\vec{v}\|_2}{\|\vec{v}\|_2} = \|A\|_2 = 1.$$

7. Slater's Condition (12 pts)

Consider the following optimization problem

$$\begin{aligned} \min_{x,y\in\mathbb{R}} \quad & 2x+y \\ \text{s.t.} \quad & x^2+y^2 \le 1, \\ & (x-2)^2+y^2 \le c, \end{aligned}$$

for some parameter $c > 0$.

(a) (4 pts) If $c = 4$, does this problem satisfy Slater's condition? Justify your answer.

Solution: Yes, e.g., the point $\left(\frac{1}{2}, 0\right)$ is strictly feasible.

(b) (4 pts) If $c = 1$, what is the feasible set of the problem?

HINT: For each constraint, try drawing the set of all $(x, y) \in \mathbb{R}^2$ that satisfy the constraint.

Solution: In this case, the two inequality constraints correpsond to closed circles of radius 1 centered at $(0, 0)$ and $(2, 0)$, respectively, so their intersection contains only the point $(1, 0)$, i.e., the feasible set is $\{(1, 0)\}$.

(c) (4 pts) If $c = 1$, does the problem satisfy Slater's condition? Justify your answer.

Solution: No, since the feasible set has empty relative interior.

8. An Optimization Problem over Matrices (8 pts)

Let $I_{n\times n}$ denote the $n \times n$ identity matrix. Is the following a convex optimization problem?

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times n}} \quad & \|X - 2I_{n\times n}\|_F^2 \\ \text{s.t.} \quad & \vec{e}_i^\top X \vec{e}_i \geq 0, \qquad \text{for each } i \in \{1, \cdots, n\}, \\ & \text{tr}(X) = n, \end{aligned}$$

where tr(X) denotes the trace of the matrix X, $\|X - 2I_{n\times n}\|_F$ denotes the Frobenius norm of the matrix $X - 2I_{n\times n}$, and for each $i \in \{1, \cdots, n\}$ we have used $\vec{e}_i \in \mathbb{R}^n$ to denote the vector with 1 in the i-th coordinate and 0 elsewhere. Justify your answer.

Solution: The objective function can be written as:

$$\|X - 2I_{n\times n}\|_F^2 = \sum_{i=1}^{n}(X_{ii} - 2)^2 + \sum_{i=1}^{n} \sum_{j\in\{1,\cdots,n\}\setminus\{i\}} X_{ij}^2$$

which is convex in the components of X. The first constraint equation states that the diagonal entries of X are all non-negative, which is an affine inequality constraint in the components of X. Meanwhile, the second constraint is an affine equality in the components of X. Thus, the constraint set is the intersection of two convex sets, and is therefore a convex set as well.

9. Cones (12 pts)

Let $K := \{\vec{x} = (x_1, x_2) \in \mathbb{R}^2 \; : \; x_2 \geq 2x_1 \geq 0\}$.

(a) (4 pts) Prove that K is a cone.

HINT: A set K is a cone *if and only if for all $\vec{x} \in K$ and $\alpha \in \mathbb{R}$, $\alpha \geq 0$, we have $\alpha\vec{x} \in K$.*

Solution: To prove that K is a cone we need to show that for all $\vec{x} \in K$ and $\alpha \in \mathbb{R}$, $\alpha \geq 0$, we have $\alpha\vec{x} \in K$. If $\vec{x} = (x_1, x_2)$ then $\alpha\vec{x} = (\alpha x_1, \alpha x_2)$. If $\vec{x} \in K$ then $x_2 \geq 2x_1 \geq 0$, which implies that $\alpha x_2 \geq 2\alpha x_1 \geq 0$, because $\alpha \geq 0$. But this implies that $\alpha\vec{x} \in K$, which is what was to be shown.

(b) (8 pts) Let K^* denote the dual cone of K which is defined as follows

$$K^* = \{\vec{y} = (y_1, y_2) \in \mathbb{R}^2 : \vec{y}^\top \vec{x} \geq 0 \text{ for all } \vec{x} \in K\}.$$

Compute an expression for K^* of the form

$$K^* = \{\vec{y} = (y_1, y_2) \in \mathbb{R}^2 : a_1 y_1 + a_2 y_2 \geq 0, b_1 y_1 + b_2 y_2 \geq 0\},$$

for some scalars $a_1, a_2, b_1, b_2 \in \mathbb{R}$.

HINT: One can equivalently write $K = \{\vec{x} = (x_1, x_2) \in \mathbb{R}^2 \ : \ x_2 \geq 2x_1 \geq 0\}$ *as* $K = \{(x_1, 2x_1 + u) \in \mathbb{R}^2 : x_1 \geq 0, u \geq 0\}$.

Solution: By definition

$$\begin{aligned} K^* &= \{\vec{y} \in \mathbb{R}^2 : \vec{y}^T \vec{x} \geq 0 \text{ for all } \vec{x} \in K\} \\ &= \{\vec{y} \in \mathbb{R}^2 : y_1 x_1 + y_2 x_2 \geq 0 \text{ for all } x_2 \geq 2x_1 \geq 0\} \\ &\stackrel{(a)}{=} \{\vec{y} \in \mathbb{R}^2 : y_1 x_1 + y_2(2x_1 + u) \geq 0 \text{ for all } x_1 \geq 0, u \geq 0\} \\ &= \{\vec{y} \in \mathbb{R}^2 : (y_1 + 2y_2)x_1 + y_2 u \geq 0 \text{ for all } x_1 \geq 0, u \geq 0\} \\ &\stackrel{(v)}{=} \{\vec{y} \in \mathbb{R}^2 : y_1 + 2y_2 \geq 0, y_2 \geq 0\}. \end{aligned}$$

Here, in step (a) we have written x_2 as $2x_1 + u$. In step (b) we have observed that if $y_1 + 2y_2 < 0$ then we can choose $x_1 \geq 0$ so large that we have $(y_1 + 2y_2)x_1 + y_2 u < 0$, while if $y_2 < 0$, then we can choose $u \geq 0$ so large that we have $(y_1 + 2y_2)x_1 + y_2 u < 0$.

10. Constrained Optimization (8 pts)

Consider the following optimization problem for fixed scalars $\alpha, \beta \in \mathbb{R}$:

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & \frac{1}{2}(x-\alpha)^2 + \beta x \\ \text{s.t.} \quad & x \geq 0. \end{aligned}$$

Show that the optimal solution of this optimization problem is given by $x^* = \max\{0, \alpha - \beta\}$.

HINT: Try first to solve the unconstrained problem $\min_{x \in \mathbb{R}} \frac{1}{2}(x-\alpha)^2 + \beta x$.

Solution: Define $f(x) = \frac{1}{2}(x-\alpha)^2 + \beta x$. There are many ways to do it. One is to use KKT conditions. Other is to use geometric arguments. We will follow the latter solution method here.

There are only two possible outcomes, either $x^* > 0$ or $x^* = 0$.

Consider a scenario when $\alpha - \beta \leq 0$. Then the gradient of objective function is

$$\frac{df(x)}{dx} = x - (\alpha - \beta),$$

which is always non-negative over the domain of the optimization problem $x \geq 0$. Therefore the minimum value would occur at 0. That is, $x^* = 0$.

Next, we consider the scenario when $\alpha - \beta > 0$. We claim that $x^* = \alpha - \beta$. Observe that $x^* > 0$ therefore from necessary and sufficient conditions of optimization posits that $\frac{df(x^*)}{dx} = 0$ which is indeed the case.

This establishes that $x^* = \max\{0, \alpha - \beta\}$.

11. Sum of Even Polynomials (12 pts)

Let $a_1, a_2 \in \mathbb{R}$. Consider the optimization problem:

$$\begin{aligned} \min_{x_1, x_2 \in \mathbb{R}} \quad & (x_1 + a_1)^4 + (x_2 + a_2)^6 \\ \text{s.t.:} \quad & x_1 \geq 0, \\ & x_2 \geq 0, \\ & x_1 + x_2 = 1. \end{aligned}$$

The objective function and all the constraint functions are assumed to have domain $\mathbb{R}^2$. Also note that the constraints are defined by affine functions.

(a) (4 pts) Show that the objective function is convex.

Solution: The feasible set is the intersection of sets defined by linear inequality constraints and linear equality constraints, so it is a polyhedron, and is therefore a convex set. The Hessian of the objective function is a diagonal matrix with diagonal entries $12(x_1 + a_1)^2$ and $30(x_2 + a_2)^4$ respectively, both of which are non-negative, so the Hessian is symmetric positive semidefinite at all $\vec{x} \in \mathbb{R}^2$. Hence the objective function is a convex function.

The problem is therefore a problem of minimizing a convex function over a polyhedron. This is a convex optimization problem.

(b) (8 pts) We can write a Lagrangian for the problem in the form

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2, \nu) = (x_1 + a_1)^4 + (x_2 + a_2)^6 - \lambda_1 x_1 - \lambda_2 x_2 + \nu(x_1 + x_2 - 1).$$

Here λ_1 and λ_2 are the dual variables corresponding to the two inequality constraints respectively, and ν is the dual variable corresponding to the equality constraint. With this Lagrangian in mind, write down the KKT equations for the problem.

NOTE: It is not necessary to attempt to compute the dual objective function. Also, you are not asked to solve the KKT equations, just to write them down.

Solution: The KKT conditions are comprised of four groups: (1) primal feasibility; (2) dual feasibility; (3) complementary slackness; and (4) Lagrangian stationarity. Since the Lagrangian is convex in $\vec{x}$ for each fixed choice of the dual variables $(\lambda_1, \lambda_2, \nu)$, the Lagrangian stationary condition can be written in the form

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2, \nu) = (x_1 + a_1)^4 + (x_2 + a_2)^6 - \lambda_1 x_1 - \lambda_2 x_2 + \nu(x_1 + x_2 - 1),$$

where $\vec{x}^*$ and $(\lambda_1^*, \lambda_2^*, \nu^*)$ are the primal and dual variables being solved for in the KKT conditions.
The equations given by the KKT conditions are the following:
(1) From primal feasibility:

$$\begin{aligned} x_1^* &\geq 0, \\ x_2^* &\geq 0, \\ x_1^* + x_2^* &= 1. \end{aligned}$$

(2) From dual feasibility:

$$\begin{aligned} \lambda_1^* &\geq 0, \\ \lambda_2^* &\geq 0. \end{aligned}$$

(3) From complementary slackness:

$$\begin{aligned} \lambda_1^* x_1^* &= 0, \\ \lambda_2^* x_2^* &= 0. \end{aligned}$$

(4) From Lagrange stationarity:

$$\begin{aligned} 4(a_1 + x_1^*)^3 - \lambda_1^* + \nu^* &= 0, \\ 6(a_2 + x_2^*)^5 - \lambda_2^* + \nu^* &= 0. \end{aligned}$$

12. Zero-Sum Games (22 pts)

Consider a matrix $A \in \mathbb{R}^{m \times n}$. Let

$$p^* = \min_{\vec{x} \in \mathcal{P}_m} \max_{\vec{y} \in \mathcal{P}_n} \vec{x}^\top A \vec{y},$$

where $\mathcal{P}_m$ and $\mathcal{P}_n$ are unit simplices in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, given by

$$\mathcal{P}_m = \left\{ \vec{x} \in \mathbb{R}^m : \sum_{i=1}^{m} x_i = 1, x_i \geq 0 \text{ for each } i \in \{1, 2, ..., m\} \right\},$$
$$\mathcal{P}_n = \left\{ \vec{y} \in \mathbb{R}^n : \sum_{j=1}^{n} y_j = 1, y_j \geq 0 \text{ for each } j \in \{1, 2, ..., n\} \right\}.$$

Furthermore, for each $j \in \{1, 2, .., n\}$, let $\vec{e}_j$ denote the vector with 1 in its j-th coordinate and 0, otherwise. Define

$$\mathcal{E}_n = \{\vec{e}_j \; : \; j \in \{1, 2, ..., n\}\}.$$

(a) (6 pts) Show that for any $\vec{x} \in \mathcal{P}_m$, there exists $\vec{e}_{j^*} \in \mathcal{E}_n$ such that

$$\vec{x}^\top A \vec{e}_{j^*} = \max_{\vec{y} \in \mathcal{P}_n} \vec{x}^\top A \vec{y}.$$

HINT: What kind of optimization problem is $\max_{\vec{y} \in \mathcal{P}_n} \vec{x}^\top A \vec{y}$*?*

Solution: Note that for any fixed $\vec{x} \in \mathcal{P}_m$ the optimization problem $\max_{\vec{y} \in \mathcal{P}_n} \vec{x}^\top A \vec{y}$ is a linear program over a simplex. Since simplex is a polytope, we know that the optimal solution will be attained at one of the vertices and the set of vertices of $\mathcal{P}_n$ coincide with $\mathcal{E}_n$.

(b) (8 pts) Show that

$$\min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{y}\in\mathcal{P}_n} \vec{x}^\top A\vec{y} = \min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{e}_j\in\mathcal{E}_n} \vec{x}^\top A\vec{e}_j.$$

HINT: Try to show that $\min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{y}\in\mathcal{P}_n} \vec{x}^\top A\vec{y} \leq \min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{e}_j\in\mathcal{E}_n} \vec{x}^\top A\vec{e}_j$*, and* $\min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{y}\in\mathcal{P}_n} \vec{x}^\top A\vec{y} \geq \min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{e}_j\in\mathcal{E}_n} \vec{x}^\top A\vec{e}_j$.

Solution: Observe that $\mathcal{E}_n \subseteq \mathcal{P}_n$. Therefore, for every $\vec{x}\in\mathcal{P}_m$,

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

Moreover from part (a) we know that for every $\vec{x}\in\mathcal{P}_m$ there exists $j^*\in\{1,2,...,n\}$ such that

$$\max_{\vec{y}\in\mathcal{P}_n} \vec{x}^\top A\vec{y} = \vec{x}^\top A\vec{e}_{j^*} \leq \max_{\vec{y}\in\mathcal{E}_n} \vec{x}^\top A\vec{y},$$

where last inequality holds because $\vec{e}_{j^*}\in\mathcal{E}_n$.

We can therefore conclude that for any $\vec{x}\in\mathcal{P}_m$,

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

The claim in the problem follows by taking a minimization over $\vec{x}$ on both sides of the above equation.

(c) (8 pts) Formulate the optimization problem $\min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{e}_j\in\mathcal{E}_n} \vec{x}^\top A\vec{e}_j$ as a linear program.

HINT: The set $\mathcal{E}_n$ contains only finitely many vectors.

Solution: Note that we can represent the optimization problem $\min_{\vec{x}\in\mathcal{P}_m} \max_{\vec{e}_j\in\mathcal{E}_n} \vec{x}^\top A\vec{e}_j$ as

$$\min_{\vec{x}\in\mathbb{R}^m, v\in\mathbb{R}} \quad v \tag{1}$$

$$v \geq \sum_{i=1}^{m} x_i A_{ij} \quad \forall\, j \in \{1, 2, 3.., n\}, \tag{C1}$$

$$\sum_{i=1}^{d} x_i = 1, \tag{C2}$$

$$x_i \geq 0, \quad \forall\, i \in \{1, 2, 3.., m\}. \tag{C3}$$

13. Gradient Descent for Convex Functions (12 pts)

Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function, with $\mathrm{dom}(f) = \mathbb{R}$. We carry out one step of gradient descent on f, with step size $\frac{1}{4}$, starting at $x_0 \in \mathbb{R}$. This leads us to

$$x_1 = x_0 - \frac{1}{4}\frac{df}{dx}(x_0).$$

Suppose the function $g(x) := 2x^2 - f(x)$ is convex. Show that

$$f(x_1) \leq f(x_0) - \frac{1}{8}\left(\frac{df}{dx}(x_0)\right)^2.$$

NOTE: You may not assume that either f or g satisfies any property other than the ones mentioned above.

Solution: We have $\frac{dg}{dx}(x_0) = 4x_0 - \frac{df}{dx}(x_0)$. Since g is convex, we have

$$g(x_1) \geq g(x_0) + (x_1 - x_0)\frac{dg}{dx}(x_0),$$

which reads

$$2x_1^2 - f(x_1) \geq 2x_0^2 - f(x_0) - \frac{1}{4}(4x_0 - \frac{df}{dx}(x_0))\frac{df}{dx}(x_0),$$

i.e.

$$2(x_0 - \frac{1}{4}\frac{df}{dx}(x_0))^2 - f(x_1) \geq 2x_0^2 - f(x_0) - \frac{1}{4}(4x_0 - \frac{df}{dx}(x_0))\frac{df}{dx}(x_0),$$

i.e.

$$\frac{1}{8}(\frac{df}{dx}(x_0))^2 - f(x_1) \geq -f(x_0) + \frac{1}{4}(\frac{df}{dx}(x_0))^2,$$

i.e.

$$f(x_1) \leq f(x_0) - \frac{1}{8}(\frac{df}{dx}(x_0))^2,$$

which is what was to be shown.

14. Second-Order Cone Programming (14 pts)

Fix $A \in \mathbb{R}^{m\times n}$, $\vec{y} \in \mathbb{R}^m$, and $\lambda, \mu > 0$. Consider the following modified version of the LASSO problem with both an ℓ^2 regularization term and an ℓ^1 regularization term, i.e.,

$$\min_{\vec{x}\in\mathbb{R}^n} \|A\vec{x}-\vec{y}\|_2^2 + \mu\|\vec{x}\|_2 + \lambda\|\vec{x}\|_1.$$

We will show that this problem can be equivalently written as an SOCP.

NOTE: The ℓ^2 regularization term is given as $\mu\|\vec{x}\|_2$, rather than $\mu\|\vec{x}\|_2^2$.

(a) (4 pts) Find vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$ and scalars $u, v \in \mathbb{R}$ (each of them in terms of μ and λ) such that the optimization problem is equivalent to solving

$$\begin{aligned} \min_{\substack{\vec{x},\vec{r}\in\mathbb{R}^n \\ t,p\in\mathbb{R}}} \quad & \vec{a}^\top\vec{x} + ut + vp + \vec{b}^\top\vec{r} \\ \text{s.t.} \quad & \|A\vec{x}-\vec{y}\|_2^2 \le t, \\ & \|\vec{x}\|_2 \le p, \\ & |x_i| \le r_i, \quad \forall i \in \{1,\ldots,n\}. \end{aligned}$$

Solution: Set $\vec{a} = \vec{0}$, $u = 1$, $v = \mu$, $\vec{b} = \lambda\vec{1}$.

(b) (6 pts) Find a second-order cone constraint over the variables $(\vec{x}, t)$ that is equivalent to the constraint $\|A\vec{x} - \vec{y}\|_2^2 \leq t$.

HINT: You may use without proof the fact that for any $\vec{z} \in \mathbb{R}^n$ and $t \in \mathbb{R}$

$$\|\vec{z}\|_2^2 \leq t \text{ if and only if } \left\| \begin{bmatrix} 2\sqrt{2}\vec{z} \\ 1 - 2t \end{bmatrix} \right\|_2 \leq 1 + 2t.$$

Solution: Using the hint, we can write

$$\begin{aligned} \{(\vec{x}, t) \in &\mathbb{R}^{n+1} : \|A\vec{x} - \vec{y}\|_2^2 \leq t\} \\ &= \left\{ (\vec{x}, t) \in \mathbb{R}^{n+1} : \left\| \begin{bmatrix} 2\sqrt{2}(A\vec{x} - \vec{y}) \\ 1 - 2t \end{bmatrix} \right\|_2 \leq 1 + 2t \right\} \\ &= \left\{ (\vec{x}, t) \in \mathbb{R}^{n+1} : \left\| \begin{bmatrix} 2\sqrt{2}A & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} \vec{x} \\ t \end{bmatrix} + \begin{bmatrix} -2\sqrt{2}\vec{y} \\ 1 \end{bmatrix} \right\|_2 \leq 1 + 2t \right\}, \end{aligned}$$

which is a cone over the variables $(\vec{x}, t)$.

(c) (4 pts) For each $i \in \{1, \ldots, n\}$, find a second-order cone constraint over the variables (x_i, r_i) that is equivalent to the constraint $|x_i| \leq r_i$.

NOTE. We know that $\|x\|_2 \leq t$ is a second-order cone constraint. Using this and parts (a)-(c), we can write the optimization problem

$$\min_{\vec{x} \in \mathbb{R}^n} \quad \|A\vec{x} - \vec{y}\|_2^2 + \mu\|\vec{x}\|_2 + \lambda\|\vec{x}\|_1,$$

as an SOCP. You are not expected to do this.

Solution: The set $\{(x_i, r_i) \in \mathbb{R}^2 : |x_i| \leq r_i\}$ is a cone over the variables (x_i, r_i) because it can also be written as $\{(x_i, r_i) \in \mathbb{R}^2 : \|x_i\|_2 \leq r_i\}$.

15. Support Vector Machines (12 pts)

Recall the maximum margin support vector machine problem:

$$\min_{\vec{w}\in\mathbb{R}^k, b\in\mathbb{R}} \quad \frac{1}{2}\|\vec{w}\|_2^2$$
$$\text{s.t.} \quad y_i(\vec{w}^\top \vec{x}_i + b) \geq 1 \quad \forall i \in \{1,\dots,n\},$$

where the data points $(\vec{x}_i, y_i)$, with features $\vec{x}_i \in \mathbb{R}^k$ and labels $y_i \in \{+1,-1\}$ for $i \in \{1,\dots,n\}$, are given.

(a) (8 pts) Consider the pairs of features $\vec{x}_i \in \mathbb{R}^2$ and labels $y_i \in \{+1,-1\}$ given in Figure 1. The maximum margin hyperplane for this data along with the support vectors are depicted in Figure 2. Find the vector $\vec{w}$ and scalar b that solve this problem.

Index i	Features $(x_{i1}, x_{i2}) \in \mathbb{R}^2$	Label $y_i \in \{+1,-1\}$
1	$(1,1)$	$+1$
2	$(3,4)$	$+1$
3	$(3,5)$	$+1$
4	$(4,0)$	-1
5	$(5,1)$	-1
6	$(6,6)$	-1

Figure 1: Data points and their labels

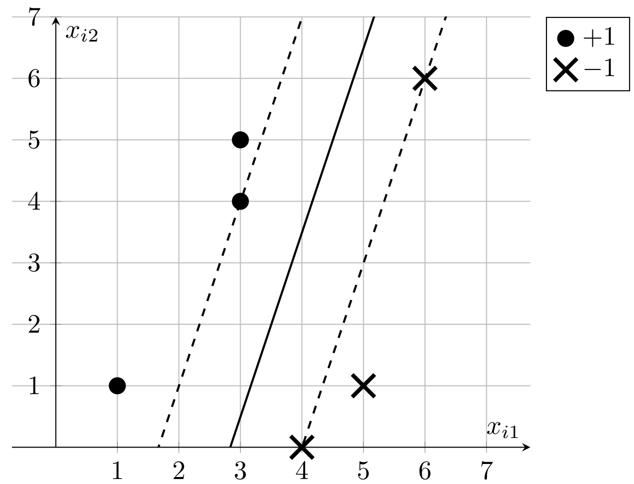

Figure 2: Maximum margin hyperplane and support vectors

HINT: Note that the constraints in the maximum margin support vector machine problem must be satisfied with equality at the support vectors.

HINT: You are likely to find at least one of these two calculations to be useful:

$$\begin{bmatrix} 3 & 4 & 1 \\ 4 & 0 & 1 \\ 6 & 6 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -3/7 & 1/7 & 2/7 \\ 1/7 & -3/14 & 1/14 \\ 12/7 & 3/7 & -8/7 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 1 \\ 5 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -1/4 & 0 & 1/4 \\ -1/8 & 1/4 & -1/8 \\ 11/8 & -1/4 & -1/8 \end{bmatrix}.$$

Solution: Using the hint, we note that the constraints in the maximum margin support vector machine problem are satisfied with equality at the support vectors. The support vectors are given in Figure 2 by $(3,4)^\top$, which is classified as $+1$, and $(4,0)^\top$, $(6,6)^\top$, which are classified as -1. This gives rise to the following equations in terms of the variables w, b:

$$1((3,4)^\top w + b) = 1,$$
$$-1((4,0)^\top w + b) = 1,$$
$$-1((6,6)^\top w + b) = 1.$$

Putting these equations in matrix form gives us:

$$\begin{bmatrix} 3 & 4 & 1 \\ 4 & 0 & 1 \\ 6 & 6 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

The inverse of the matrix on the left hand side is provided to us in the hint, which gives the solution:

$$\begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} = \begin{bmatrix} 3 & 4 & 1 \\ 4 & 0 & 1 \\ 6 & 6 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -3/7 & 1/7 & 2/7 \\ 1/7 & -3/14 & 1/14 \\ 12/7 & 3/7 & -8/7 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

This gives $w_1^* = \frac{-6}{7}$, $w_2^* = \frac{2}{7}$, $b^* = \frac{17}{7}$.

(b) (4 pts) Now, consider the pairs of features $\vec{x}_i \in \mathbb{R}^2$ and labels $y_i \in \{+1, -1\}$ given in Figure 3, and depicted visually in Figure 4:

Index i	Features $(x_{i1}, x_{i2}) \in \mathbb{R}^2$	Label $y_i \in \{+1, -1\}$
1	$(1, 1)$	$+1$
2	$(4.5, 1)$	$+1$
3	$(4, 6)$	$+1$
4	$(4, 0)$	-1
5	$(4, 2)$	-1
6	$(5, 1)$	-1

Figure 3: Data points and their labels

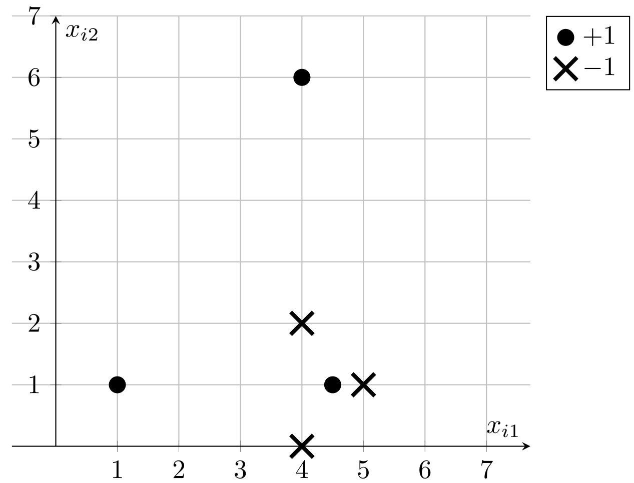

Figure 4: Visual depiction of data points and labels

If possible, find a separating hyperplane that solves the maximum margin support vector machine problem with this data, or provide a justification why such a hyperplane cannot be found.

Solution: Such a hyperplane cannot be found because the data are not linearly separable. This is because the point $(4.5, 1)$, which is classified as $+1$, can be written as a convex combination of the points classified as -1:

$$\begin{bmatrix} 4.5 \\ 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 4 \\ 0 \end{bmatrix} + \frac{1}{4}\begin{bmatrix} 4 \\ 2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

