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 (3 pts)**

   When the exam starts, write your SID at the top of every page. *No extra time will be given for this task.*

3. **Favorites. Any answer, as long as you write it down, will be given full credit. (2 pts)**

   (a) (1 pts) **What's a movie you are looking forward to watching this summer?**

   Solution: Any answer is fine.

   (b) (1 pts) **If you could have any animal as a pet, what animal would you choose?**

   Solution: Any answer is fine.

4. Convex Functions (7 pts)

(a) (3 pts) **Prove that the function** $f\colon \mathbb{R}_{++} \to \mathbb{R}$ **given by** $f(x) \doteq \log(1/x)$ **is a strictly convex function**, where $\mathbb{R}_{++}$ is the set of strictly positive real numbers.

Solution: The domain $\mathbb{R}_{++}$ is convex. To show that f is convex we see that $f(x) = -\log(x)$ and compute its Hessian $\nabla^2 f(x) = \frac{1}{x^2}$, which is strictly positive over $\mathbb{R}_{++}$.

(b) (4 pts) **Is the function** $g\colon \mathbb{R}_{++} \to \mathbb{R}$ **given by** $g(x) \doteq \max\{(ax-b)^2, \log(1/x)\}$ **a convex function?** *Justify your answer.*

You may use the fact that the function $f(x) \doteq \log(1/x)$ is convex for $x \in \mathbb{R}_{++}$.

Solution: The domain $\mathbb{R}_{++}$ is convex. We know from part (a) that $\log(1/x)$ is a (strictly) convex function. Since the Hessian $\nabla_x^2 (ax-b)^2 = 2a^2 \geq 0$, we know that the function $x \mapsto (ax-b)^2$ is convex. We know that the maximum of convex functions is convex, so g is convex.

5. Hyperplanes (7 pts)

(a) (3 pts) **Give a hyperplane of the form $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top(\vec{x} - \vec{x}_0) = 0\}$ which goes through the point** $(2,3)$ **and is orthogonal to the vector** $(1,1)$. *No justification is necessary.*

Solution: One candidate hyperplane is

$$\vec{c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \vec{x}_0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}. \tag{1}$$

(b) (4 pts) Let $\vec{c}_1 \neq \vec{0}$ and $\vec{c}_2 \neq \vec{0}$ be two vectors in $\mathbb{R}^n$. Define the two hyperplanes $\mathcal{H}_1 = \{\vec{x} \in \mathbb{R}^n \mid \vec{c}_1^\top \vec{x} = 0\}$ and $\mathcal{H}_2 = \{\vec{x} \in \mathbb{R}^n \mid \vec{c}_2^\top \vec{x} = 0\}$ where $\vec{c}_1^\top \vec{c}_2 = 0$. **Give any point $\vec{x}^\star$ in terms of $\vec{c}_1$ and $\vec{c}_2$ such that** $\vec{c}_1^\top \vec{x}^\star > 0$ **and** $\vec{c}_2^\top \vec{x}^\star > 0$.

HINT: Draw a picture of the two hyperplanes in 2D.

Solution: The point $\vec{x}^\star = \vec{c}_1 + \vec{c}_2$ satisfies these two conditions. We have

$$\begin{aligned}
\vec{c}_1^\top \vec{x}^\star &= \vec{c}_1^\top(\vec{c}_1 + \vec{c}_2) && (2)\\
&= \|\vec{c}_1\|_2^2 + 0 && (3)\\
&> 0, \qquad \text{for } \vec{c}_1 \neq 0. && (4)
\end{aligned}$$

We also have

$$\begin{aligned}
\vec{c}_2^\top \vec{x}^\star &= \vec{c}_2^\top(\vec{c}_1 + \vec{c}_2) && (5)\\
&= 0 + \|\vec{c}_2\|_2^2 && (6)\\
&> 0, \qquad \text{for } \vec{c}_2 \neq 0. && (7)
\end{aligned}$$

6. Newton's Method (6 pts)

(a) (3 pts) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function that we are attempting to minimize using Newton's method. Suppose that at the k^{th} iterate $\vec{x}_k \in \mathbb{R}^n$ we have $\nabla^2 f(\vec{x}_k) = \alpha_k I_n$, where $\alpha_k > 0$ is some positive constant and $I_n \in \mathbb{R}^{n \times n}$ is the identity matrix. **Write the Newton's method step for $\vec{x}_{k+1}$ in terms of $\vec{x}_k$, α_k, and $\nabla f(\vec{x}_k)$.**

Solution: Since $[\alpha_k I_n]^{-1} = \frac{1}{\alpha_k} I_n$, the Newton's method step is

$$\vec{x}_{k+1} = \vec{x}_k - \frac{1}{\alpha_k} \nabla f(\vec{x}_k). \tag{8}$$

(b) (3 pts) Now suppose we are trying to minimize the same function f via gradient descent. **Write the gradient descent step for $\vec{x}_{k+1}$ in terms of $\vec{x}_k$ and $\nabla f(\vec{x}_k)$, with some arbitrary step size $\eta_k > 0$ at time k. For what value of η_k is the gradient descent update equation the same as the Newton's update equation from part (a)?**

Solution: The gradient descent step is

$$\vec{x}_{k+1} = \vec{x}_k - \eta_k \nabla f(\vec{x}_k). \tag{9}$$

The two descent steps are equivalent when $\eta_k = \frac{1}{\alpha_k}$.

7. Solving a Quadratic Program (10 pts)

Consider the quadratic program

$$p^\star = \min_{\vec{x}\in\mathbb{R}^3} \left(\vec{x}^\top M\vec{x} - 2\vec{b}^\top \vec{x}\right), \tag{10}$$

where the matrix $M \in \mathbb{R}^{3\times 3}$ is defined as follows

$$M = 4\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}^\top + 3\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}^\top. \tag{11}$$

(a) (5 pts) **If $\vec{b} = \begin{bmatrix}1\\2\\3\end{bmatrix}$, then is $p^\star = -\infty$ or is it finite?** You do not need to calculate $p^\star$. *Justify your answer.*

Solution: $p^\star = -\infty$ because the matrix M is not full rank and the vector $\vec{b}$ is not in the range of M.

Additional justification The nullspace of matrix M is span $\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$ and $\vec{b}$ is not orthogonal to the nullspace. We can take $\vec{x} = \alpha\begin{bmatrix}0\\1\\0\end{bmatrix}$ for some fixed α so that $\vec{x}^\top M\vec{x} = 0$ and $\vec{b}^\top\vec{x} = 2\alpha$. Hence taking $\alpha \to \infty$, we get that the objective value tends to $-\infty$.

(b) (5 pts) **If $\vec{b} = \begin{bmatrix}1\\0\\1\end{bmatrix}$, then is $p^\star = -\infty$ or is it finite?** You do not need to calculate $p^\star$. *Justify your answer.*

Solution: $p^\star$ is finite since $\vec{b}$ is in the range of matrix M.

Additional justification The solution to the optimization problem can be obtained by setting the gradient to zero:

$$\nabla\left(\vec{x}^\top M\vec{x} - 2\vec{b}^\top\vec{x}\right) = 2(M\vec{x} - \vec{b}) = 0 \tag{12}$$

So $\vec{x}^\star$ is the set of solutions of $M\vec{x} = \vec{b}$, which take the form

$$\vec{x}^\star = \begin{bmatrix}\frac{1}{4}\\0\\\frac{1}{3}\end{bmatrix} + \alpha\begin{bmatrix}0\\1\\0\end{bmatrix}, \qquad \text{for some } \alpha \in \mathbb{R}. \tag{13}$$

The corresponding value $p^\star$ is

$$\begin{aligned} p^\star &= \vec{x}^{\star\top} M\vec{x}^\star - 2\vec{b}^\top\vec{x}^\star && (14)\\ &= \vec{b}^\top\vec{x}^\star - 2\vec{b}^\top\vec{x}^\star && (15)\\ &= -\vec{b}^\top\vec{x} && (16)\\ &= -(1/3 + 1/4) = -1/7 && (17)\end{aligned}$$

8. Quadratically Constrained Linear Program (16 pts)

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

$$\begin{aligned} p^\star = \min_{\vec{x}} \quad & \vec{c}^\top \vec{x} \\ \text{s.t.} \quad & \frac{1}{2}(\vec{x} - \vec{x}_0)^\top Q(\vec{x} - \vec{x}_0) \leq \epsilon. \end{aligned} \tag{18}$$

(a) (4 pts) **Is this problem convex? Does strong duality hold here?** *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. 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.

(b) (8 pts) **Show that the dual function associated with the primal problem in (18) is**

$$g(\lambda) = \begin{cases} -\infty & \text{if } \lambda = 0 \\ \vec{c}^\top \vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon & \text{if } \lambda > 0, \end{cases} \tag{19}$$

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

Solution:

The Lagrangian associated with the problem is

$$L(\vec{x}, \lambda) = \vec{c}^\top \vec{x} + \lambda \left(\frac{1}{2}(\vec{x} - \vec{x}_0)^\top Q(\vec{x} - \vec{x}_0) - \epsilon \right) \tag{20}$$

Thus, the dual function is given by

$$g(\vec{\lambda}) = \min_{\vec{x}} L(\vec{x}, \lambda). \tag{21}$$

For $\lambda = 0$:

$$L(\vec{x}, 0) = \vec{c}^\top \vec{x}, \tag{22}$$

and

$$\begin{aligned} g(0) &= \min_{\vec{x}} L(\vec{x}, 0) && (23) \\ &= \min_{\vec{x}} \vec{c}^\top \vec{x} && (24) \\ &= -\infty. && (25) \end{aligned}$$

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

$$\begin{aligned} \nabla_{\vec{x}} L(\vec{x}^\star(\lambda), \lambda) &= 0 && (26) \\ \vec{c} + \lambda Q(\vec{x}^\star(\lambda) - \vec{x}_0) &= 0 && (27) \\ \implies \vec{x}^\star(\lambda) &= \vec{x}_0 - \frac{1}{\lambda} Q^{-1}\vec{c}. && (28) \end{aligned}$$

Thus,

$$g(\vec{\lambda}) = L(\vec{x}^\star(\lambda), \lambda) \tag{29}$$

$$= \vec{c}^\top \vec{x}^\star(\lambda) + \lambda \left(\frac{1}{2}(\vec{x}^\star(\lambda) - \vec{x}_0)^\top Q(\vec{x}^\star(\lambda) - \vec{x}_0) - \epsilon \right) \tag{30}$$

$$= \vec{c}^\top \left(\vec{x}_0 - \frac{1}{\lambda} Q^{-1}\vec{c} \right) + \lambda \left(\frac{1}{2}(\vec{x}_0 - \frac{1}{\lambda}Q^{-1}\vec{c} - \vec{x}_0)^\top Q(\vec{x}_0 - \frac{1}{\lambda}Q^{-1}\vec{c} - \vec{x}_0) - \epsilon \right) \tag{31}$$

$$= \vec{c}^\top \vec{x}_0 - \frac{1}{\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda(\frac{1}{2}(\frac{1}{\lambda}Q^{-1}\vec{c})^\top Q(\frac{1}{\lambda}Q^{-1}\vec{c}) - \epsilon) \tag{32}$$

$$= \vec{c}^\top \vec{x}_0 - \frac{1}{\lambda}\vec{c}^\top Q^{-1}\vec{c} + \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon \tag{33}$$

$$= \vec{c}^\top \vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon. \tag{34}$$

(c) (4 pts) Consider the dual problem of the primal problem in (18):

$$d^\star = \max_{\lambda \geq 0} g(\lambda), \tag{35}$$

where

$$g(\lambda) = \begin{cases} -\infty & \text{if } \lambda = 0 \\ \vec{c}^\top \vec{x}_0 - \frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} - \lambda\epsilon & \text{if } \lambda > 0. \end{cases} \tag{19}$$

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

Solution: We know a priori that the dual problem

$$\max_{\lambda \geq 0} \vec{c}^\top \vec{x}_0 - \left(\frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda\epsilon \right) = \max_{\lambda \geq 0} - \left(\frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda\epsilon \right) \tag{36}$$

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

$$\frac{d}{d\lambda}\left(\frac{1}{2\lambda}\vec{c}^\top Q^{-1}\vec{c} + \lambda\epsilon \right)\Big|_{\lambda=\lambda^\star} = 0 \tag{37}$$

$$\implies -\frac{1}{2(\lambda^\star)^2}\vec{c}^\top Q^{-1}\vec{c} + \epsilon = 0 \tag{38}$$

$$\implies \lambda^\star = \sqrt{\frac{\vec{c}^\top Q^{-1}\vec{c}}{2\epsilon}} > 0 \qquad \text{since } \vec{c} \neq 0. \tag{39}$$

9. Support Vector Regression (11 pts)

Let $\epsilon > 0$. Let $X \in \mathbb{R}^{n\times d}$ be a data matrix and $\vec{y} \in \mathbb{R}^n$ be a label vector, such that

$$X = \begin{bmatrix} \vec{x}_1^\top \\ \vdots \\ \vec{x}_n^\top \end{bmatrix}, \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \tag{40}$$

where $\vec{x}_i^\top$ are the rows of X and y_i are the entries of $\vec{y}$. Consider the problem

$$\begin{aligned} \min_{\vec{w}\in\mathbb{R}^d} \quad & \frac{1}{2}\|\vec{w}\|_2^2 \\ \text{s.t.} \quad & \|X\vec{w}-\vec{y}\|_\infty \le \epsilon. \end{aligned} \tag{41}$$

(a) (5 pts) **Rewrite this problem as an equivalent quadratic program (i.e. with quadratic objective function and finitely many linear constraints).**

Solution: We have

$$\|X\vec{w}-\vec{y}\|_\infty \le \epsilon \tag{42}$$

$$\iff -\epsilon\vec{1} \le X\vec{w}-\vec{y} \le \epsilon\vec{1}. \tag{43}$$

Thus the problem becomes

$$\begin{aligned} p^\star = \min_{\vec{w}\in\mathbb{R}^d} \quad & \frac{1}{2}\|\vec{w}\|_2^2 \\ \text{s.t.} \quad & X\vec{w}-\vec{y} \le \epsilon\vec{1} \\ & X\vec{w}-\vec{y} \ge -\epsilon\vec{1}. \end{aligned} \tag{44}$$

(b) (6 pts) Now consider the problem:

$$\min_{\vec{w}\in\mathbb{R}^d} \left\{ \frac{1}{2}\|\vec{w}\|_2^2 + \lambda \sum_{i=1}^{n} \max\{0, |\vec{x}_i^\top\vec{w}-y_i| - \epsilon\} \right\}. \tag{45}$$

Rewrite this problem as an equivalent quadratic program (i.e. with quadratic objective function and finitely many linear constraints).

HINT: Introduce a new variable $\vec{z}$.

Solution: We have

$$\max\{0, |\vec{x}_i^\top\vec{w}-y_i| - \epsilon\} = \max\{0, \max\{\vec{x}_i^\top\vec{w}-y_i, -(\vec{x}_i^\top\vec{w}-y_i)\} - \epsilon\} \tag{46}$$

$$= \max\{0, \max\{\vec{x}_i^\top\vec{w}-y_i-\epsilon, -(\vec{x}_i^\top\vec{w}-y_i)-\epsilon\}\} \tag{47}$$

$$= \max\{0, \vec{x}_i^\top\vec{w}-y_i-\epsilon, -(\vec{x}_i^\top\vec{w}-y_i)-\epsilon\}. \tag{48}$$

Use slack variable z_i which is $\ge$ all three terms. Thus we have the following QP formulation

$$\begin{aligned} \min_{\substack{\vec{w}\in\mathbb{R}^d \\ \vec{z}\in\mathbb{R}^n}} \quad & \frac{1}{2}\|\vec{w}\|_2^2 + \lambda\vec{1}^\top\vec{z} \\ \text{s.t.} \quad & \vec{z} \ge X\vec{w}-\vec{y}-\epsilon\vec{1} \\ & \vec{z} \ge -(X\vec{w}-\vec{y})-\epsilon\vec{1} \\ & \vec{z} \ge \vec{0}. \end{aligned} \tag{49}$$

10. Candidate Solution of Linear Programs (12 pts)

(a) (5 pts) Consider the linear program

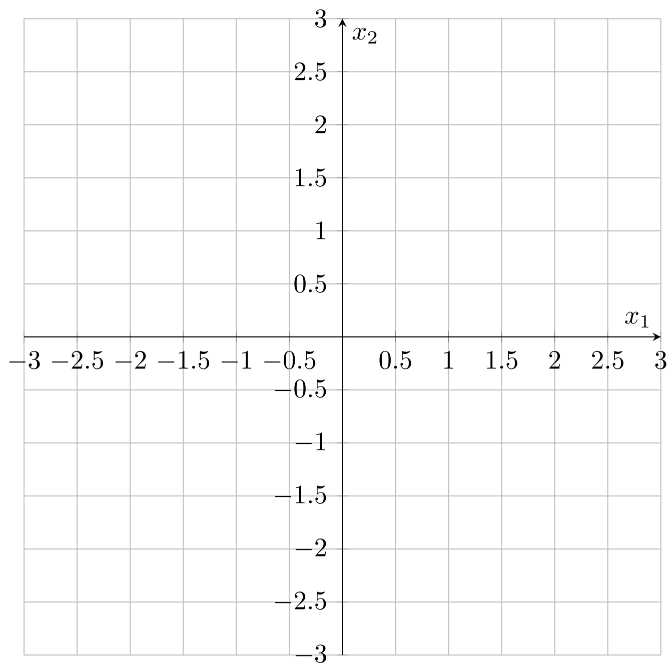
$$\min_{\vec{x}\in\mathbb{R}^2} \quad \begin{bmatrix}2\\1\end{bmatrix}^\top \vec{x} \tag{50}$$

$$\text{s.t.} \quad \vec{x} \geq 0 \tag{51}$$

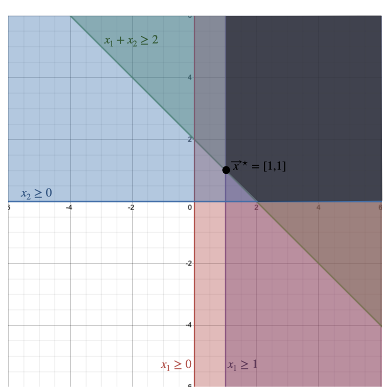
$$\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\vec{x} \geq \begin{bmatrix}2\\1\end{bmatrix}. \tag{52}$$

i. **Sketch and shade the feasible region of the above optimization problem in the graph provided below.**

ii. **Use your sketch to identify the optimal solution $\vec{x}^\star$ and write it in the box below.**



Solution:



The feasible region is the area shaded in black. The optimal solution is achieved at $\vec{x}^\star = \begin{bmatrix}1\\1\end{bmatrix}$.

(b) (7 pts) Let $A \in \mathbb{S}^n$ be a *symmetric* matrix and $\vec{c} \neq \vec{0}$ be a vector in $\mathbb{R}^n$. Consider the linear program

$$\min_{\vec{x} \in \mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{53}$$

$$\text{s.t.} \quad A\vec{x} \geq \vec{c} \tag{54}$$

$$\vec{x} \geq 0. \tag{55}$$

Consider a point $\vec{x}^\star > 0$ such that $A\vec{x}^\star = \vec{c}$. **Prove that $\vec{x}^\star$ is a minimizer of the above optimization problem.**

HINT: Let $\vec{\lambda}$ be the dual variables associated with the constraints $A\vec{x} \geq \vec{c}$ and $\vec{\mu}$ be the dual variables associated with the constraints $\vec{x} \geq 0$. Use the KKT conditions.

Solution: Write the Lagrangian associated with the problem

$$L(\vec{x}, \vec{\lambda}, \vec{\nu}) = \vec{c}^\top \vec{x} + \vec{\lambda}^\top (c - A\vec{x}) - \vec{\nu}^\top \vec{x} \tag{56}$$

And compute its gradient:

$$\nabla_{\vec{x}} L(\vec{x}, \vec{\lambda}, \vec{\nu}) = \vec{c} - A\vec{\lambda} - \vec{\nu} \tag{57}$$

Now we write the KKT conditions

$$\text{(Lagrangian stationarity)} \quad \vec{c} - A\vec{\lambda} - \vec{\nu} = 0 \tag{58}$$

$$\text{(Complementary slackness)} \quad \vec{\lambda}^\top (c - A\vec{x}) = 0 \tag{59}$$

$$\text{(Complementary slackness)} \quad \vec{\nu}^\top \vec{x} = 0 \tag{60}$$

$$\text{(Primal feasibility)} \quad A\vec{x} \geq \vec{c} \quad \text{and} \quad \vec{x} \geq 0 \tag{61}$$

$$\text{(Dual feasibility)} \quad \vec{\lambda} \geq 0 \quad \text{and} \quad \vec{\nu} \geq 0 \tag{62}$$

The solution $\vec{x}^\star$ satisfies the complementary slackness condition $\vec{\lambda}^\top (c - A\vec{x}^\star) = 0$ for any $\vec{\lambda}$ and the primal feasibility conditions. Further, since $\vec{x}^\star$ is not necessarily zero then we need $\vec{\nu}^\star = \vec{0}$ for the complimentary slackness condition $\vec{\nu}^\top \vec{x} = 0$ to hold. Finally we note that setting $\vec{\lambda}^\star = \vec{x}^\star$ satisfies the Lagrangian stationarity and the dual feasibility conditions.

So the candidate solution $\vec{x}^\star$ satisfies KKT conditions. Since KKT conditions are sufficient for optimality in linear programs then $\vec{x}^\star$ is an optimal solution of the optimization problem.

11. Inscribed Box in a Polyhedron (8 pts)

Let $\mathcal{P} := \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} \leq \vec{b}\}$ be a bounded polyhedron with matrix $A \in \mathbb{R}^{m\times n}$ and vector and $\vec{b} \in \mathbb{R}^m$ such that

$$A = \begin{bmatrix} \vec{a}_1^\top \\ \vdots \\ \vec{a}_m^\top \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \tag{63}$$

where $\vec{a}_i^\top$ are the rows of A and b_i are the entries of $\vec{b}$. In this problem, we wish to find the maximal radius box

$$\mathcal{B}(\vec{x}_0, r) = \{\vec{x}_0 + r\vec{u} \mid \|\vec{u}\|_\infty \leq 1\}, \tag{64}$$

such that $\mathcal{B}(\vec{x}_0, r) \subseteq \mathcal{P}$. In other words, we want to solve

$$\max_{\vec{x}_0, r} \quad r \tag{65}$$

$$\text{s.t.} \quad \mathcal{B}(\vec{x}_0, r) \subseteq \mathcal{P}. \tag{66}$$

Express this problem as a linear program with at most m constraints. *Justify your answer.*

HINT: Recall the ℓ^1—ℓ^∞ duality:

$$\|\vec{v}\|_1 = \max_{\vec{u}\colon \|\vec{u}\|_\infty \leq 1} \vec{u}^\top \vec{v}. \tag{67}$$

Solution: Given the definition of the box:

$$\mathcal{B}(\vec{x}_0, r) = \{\vec{x}_0 + r\vec{u} \mid \|\vec{u}\|_\infty \leq 1\}, \tag{68}$$

we can express the the constraint $B(\vec{x}_0, r) \subseteq \mathcal{P}$ as

$$A(\vec{x}_0 + r\vec{u}) \leq \vec{b} \quad \forall \vec{u}\colon \|\vec{u}\|_\infty \leq 1. \tag{69}$$

Conditioning on each element of the vector and using the maximum operator, we can equivalently express the above constraints as

$$\max_{\|\vec{u}\|_\infty \leq 1} \{\vec{a}_i^\top \vec{x}_0 + r\vec{a}_i^\top \vec{u}\} \leq b_i, \qquad \forall i. \tag{70}$$

Invoking ℓ^1–ℓ^∞ duality, this condition can be expressed as

$$\vec{a}_i^\top \vec{x}_0 + r\|\vec{a}_i\|_1 \leq b_i, \qquad \forall i. \tag{71}$$

Putting this all together, we can now write our program with m constraints

$$\max_{\vec{x}_0, r} \quad r \tag{72}$$

$$\text{s.t.} \quad \vec{a}_i^\top \vec{x}_0 + r\|\vec{a}_i\|_1 \leq b_i \ \forall i. \tag{73}$$

This is a linear program since the objective is affine and each constraint is affine. (Since A is a constant, we can treat $\|\vec{a}_i\|_1$ as a constant.)

12. Gambler's Destiny (16 pts)

Aekus and Aditya notice that their local casino is offering great odds for betting on a marathon with n athletes. They have $c > 0$ dollars with which they can place bets. For each athlete i, they can bet b_i dollars. If athlete i wins they receive $b_i r_i > 0$ dollars and receive nothing for all other athletes. Athlete i wins with probability p_i. Exactly one athlete wins each race.

(a) (10 pts) Suppose Aekus and Aditya wish to *maximize* their expected profit after one bet in which they can bet up to c dollars. Formally, they want to solve the optimization problem:

$$\begin{aligned} \min_{b_1,\ldots,b_n} \quad & -\left(\sum_{i=1}^n p_i b_i r_i + \left(c - \sum_{i=1}^n b_i\right)\right) \\ \text{s.t.} \quad & \sum_{i=1}^n b_i \leq c, \\ & b_i \geq 0 \; \forall i. \end{aligned} \tag{74}$$

Write the Lagrangian and KKT conditions for (74) where λ is the dual variable corresponding to the constraint $\sum_{i=1}^n b_i \leq c$ and μ_i is the dual variable corresponding to the constraint $b_i \geq 0$. Then, show that for the primal and dual optimizers $b_i^\star$, $\mu_i^\star$, and $\lambda^\star$, the following relations hold:

i. if for any athlete j we have that $p_j r_j > 1$, then $\sum_{i=1}^n b_i^\star = c$.
ii. if for any athlete j we have that $p_j r_j < 1$, then $b_j^\star = 0$.
iii. if $p_i r_i > p_j r_j$, then $\mu_j^\star > \mu_i^\star$.

Solution: The Lagrangian can be written as

$$\mathcal{L}(\vec{b}, \lambda, \vec{\mu}) = -\sum_{i=1}^n b_i(p_i r_i - 1) + c + \lambda(\sum_{i=1}^n b_i - c) - \sum_{i=1}^n \mu_i b_i.$$

The KKT conditions say that:

i. Stationarity: $1 - p_i r_i + \lambda^\star - \mu_i^\star = 0 \; \forall i$.
ii. Primal Feasibility: $\sum_{i=1}^n b_i^\star \leq c$ and $b_i^\star \geq 0 \; \forall i$.
iii. Dual Feasibility: $\lambda^\star \geq 0$ and $\mu^\star \geq 0 \; \forall i$.
iv. Complementary Slackness: $\lambda(\sum_{i=1}^n b_i^\star - c) = 0$ and $\mu_i^\star b_i^\star = 0 \; \forall i$.

Now, because the problem is a feasible linear program (and thus the problem is convex and Slater's condition holds), the KKT conditions are necessary and sufficient for optimality. (We only need necessary conditions in this case, i.e., $(\vec{b}^\star, \vec{\mu}^\star, \vec{\lambda}^\star)$ are optimal $\implies$ they satisfy KKT.) We now show the three conditions:

i. if $p_j r_j > 1$, then from stationarity:

$$\begin{aligned} 0 &= 1 - p_j r_j + \lambda^\star - \mu_j^\star && (75) \\ &< \lambda^\star - \mu_j^\star && (76) \\ &\leq \lambda^\star && (77) \end{aligned}$$

Hence, $\lambda^\star > 0$. Thus, by complementary slackness, $\sum_{i=1}^n b_i^\star = c$.

ii. if $p_j r_j < 1$, from from Stationarity:

$$0 = 1 - p_j r_j + \lambda^\star - \mu_j^\star \tag{78}$$

$$> \lambda^\star - \mu_j^\star \tag{79}$$
$$\geq -\mu_j^\star \tag{80}$$

Hence, $\mu_j^\star > 0$. Thus, by complementary slackness, $b_j^\star = 0$.

iii. The stationarity condition tells us that $1 - p_i r_i + \lambda^\star - \mu_i^\star = 1 - p_j r_j + \lambda^\star - \mu_j^\star$. If $p_i r_i > p_j r_j$ then $\mu_j^\star > \mu_i^\star$.

(b) (6 pts) **Use these relations to argue that if there exists athlete j such that $p_j r_j > 1$ and $p_j r_j > p_i r_i$ for all other athletes $i \neq j$, then $b_j^\star = c$.** In other words, Aekus and Aditya should allocate all their money to the most profitable bet in expectation, if one exists.

Solution: From the previous part, we know that $\mu_j^\star < \mu_i^\star$ for any $i \neq j$ and $\sum_{i=1}^n b_i^\star = c$. This in turn implies that there exists some k such that $b_k^\star > 0$. By complementary slackness, we have that $\mu_k^\star = 0$, but since $\mu_j^\star \geq 0$, we must have that $\mu_j^\star = 0$ and $j = k$. Now for all $i \neq j$, $\mu_i^\star > 0$ so $b_i^\star = 0$. This implies that $\sum_{i=1}^n b_i^\star = b_j^\star = c$.

13. Conjugate Gradient Method (39 pts)

In this problem we explore a new descent method, called the conjugate gradient method[1], to solve the problem $A\vec{x} = \vec{b}$ where $A \in \mathbb{S}^n_{++}$ is a symmetric positive definite matrix, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^n$.

(a) (5 pts) Consider a set of vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$, all of which are in $\mathbb{R}^n$. Suppose that for some $\vec{v} \in \mathbb{R}^n$ we have that $\vec{v}^\top A\vec{u}_i = 0$ for all $1 \le i \le k$. **Show that $\vec{v}^\top A\vec{w} = 0$ for any vector $\vec{w} \in \operatorname{span}(\vec{u}_1, \ldots, \vec{u}_k)$.**

Solution: We can express any vector $\vec{w} \in \operatorname{span}\{\vec{u}_1, \ldots, \vec{u}_k\}$ as $\vec{w} = \alpha_1\vec{u}_1 + \alpha_2\vec{u}_2 + \ldots + \alpha_k\vec{u}_k$. Hence,

$$\vec{v}^\top A\vec{w} = \alpha_1\vec{v}^\top\vec{u}_1 + \alpha_1\vec{v}^\top\vec{u}_2 + \ldots + \vec{v}^\top\vec{u}_k \tag{81}$$

$$= 0. \tag{82}$$

So $\vec{v}$ is conjugate in A with any vector $\vec{w} \in \operatorname{span}\{\vec{u}_1, \ldots, \vec{u}_k\}$.

(b) (5 pts) Let $\vec{v} \in \mathbb{R}^n$ and $\vec{w} \in \mathbb{R}^n$ with $\vec{v} \ne \vec{0}$ and $\vec{w} \ne \vec{0}$. **Show that if $\vec{v}$ and $\vec{w}$ are such that $\vec{v}^\top A\vec{w} = 0$, they must be linearly independent.**

Solution: Suppose $\vec{v}$ and $\vec{w}$ are linearly dependent. Then, $\vec{v} = \alpha\vec{w}$ for $\alpha \ne 0$, so $\vec{v}^\top A\vec{w} = \alpha(\vec{w}^\top A\vec{w}) \ne 0$ since $A \succ 0$, which contradicts the claim that $\vec{v}$ and $\vec{w}$ are conjugate in A. Thus, if $\vec{v}$ and $\vec{w}$ are conjugate in A, they must be linearly independent.

(c) (5 pts) Recall that $A \in \mathbb{S}^n_{++}$ is a symmetric positive definite matrix. Suppose $\{\vec{u}_1, \ldots, \vec{u}_n\}$ are a set of vectors in $\mathbb{R}^n$ such that $\vec{u}_i^\top A\vec{u}_j = 0$ for all i, j where $i \ne j$. Consider the matrix,

$$U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix}. \tag{83}$$

Show that $\operatorname{rank}\left(U^\top AU\right) = n$.

This implies that U has full rank (you do *not* need to show this), and hence $\{\vec{u}_1, \ldots, \vec{u}_n\}$ forms a basis for $\mathbb{R}^n$. We call this a *conjugate basis* in A for $\mathbb{R}^n$.

Solution:

$$U^\top AU = \begin{bmatrix} \vec{u}_1^\top A\vec{u}_1 & \vec{u}_1^\top A\vec{u}_2 & \ldots & \vec{u}_1^\top A\vec{u}_n \\ \vec{u}_2^\top A\vec{u}_1 & \vec{u}_2^\top A\vec{u}_2 & \ldots & \vec{u}_2^\top A\vec{u}_n \\ \vdots & & \ddots & \vdots \\ \vec{u}_n^\top A\vec{u}_1 & \vec{u}_n^\top A\vec{u}_2 & \ldots & \vec{u}_n^\top A\vec{u}_n \end{bmatrix} = \begin{bmatrix} \vec{u}_1^\top A\vec{u}_1 & 0 & \ldots & 0 \\ 0 & \vec{u}_2^\top A\vec{u}_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \vec{u}_n^\top A\vec{u}_n \end{bmatrix} \tag{84}$$

Since A is PD we have $\vec{u}_i^\top A\vec{u}_i > 0$, so all the diagonal entries of the diagonal matrix $U^\top AU$ are non-zero. Thus $\operatorname{rank}\left(U^\top AU\right) = n$.

(d) (7 pts) Recall that $A \in \mathbb{S}^n_{++}$ is a symmetric positive definite matrix, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^n$. Consider

$$\vec{x}^\star = \sum_{i=1}^{n} \frac{\vec{u}_i^\top\vec{b}}{\vec{u}_i^\top A\vec{u}_i}\vec{u}_i, \tag{85}$$

where $\{\vec{u}_1, \ldots, \vec{u}_n\}$ form a conjugate basis in A for $\mathbb{R}^n$ (i.e. $\vec{u}_i^\top A\vec{u}_j = 0$ for all i, j where $i \ne j$) as defined in part (c). **Show that $\vec{x}^\star$ is a solution to $A\vec{x} = \vec{b}$.** *Show your work.*

HINT: For any vectors $\vec{c} \in \mathbb{R}^n$ and $\vec{d} \in \mathbb{R}^n$ and any basis $\{\vec{z}_1, \ldots, \vec{z}_n\}$ of $\mathbb{R}^n$,

$$\vec{c} = \vec{d} \iff \textit{for all } 1 \le i \le n, \textit{ we have } \vec{z}_i^\top\vec{c} = \vec{z}_i^\top\vec{d}. \tag{86}$$

[1] Two vectors $\vec{v}$ and $\vec{w}$ are "*conjugate in A*" if $\vec{v}^\top A\vec{w} = 0$, and such pairs of vectors are useful in the algorithm, giving it the name.

Solution: Iff $\vec{x}^\star$ is a solution, then $A\vec{x}^\star = \vec{b}$.

$$A\vec{x}^\star = \sum_{i=1}^{n} \frac{\vec{u}_i^\top \vec{b}}{\vec{u}_i^\top A\vec{u}_i} A\vec{u}_i \tag{87}$$

We check that $A\vec{x}^\star = \vec{b}$ using the hint on our conjugate basis $\{\vec{u}_1, \ldots, \vec{u}_n\}$:

$$\begin{aligned}\vec{u}_i^\top A\vec{x}^\star &= \frac{\vec{u}_i^\top \vec{b}}{\vec{u}_i^\top A\vec{u}_i} \vec{u}_i^\top A\vec{u}_i & (88)\\ &= \vec{u}_i^\top \vec{b} & (89)\end{aligned}$$

Hence, $A\vec{x}^\star = \vec{b}$

(e) (9 pts) Recall that $A \in \mathbb{S}^n_{++}$ is a symmetric positive definite matrix, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^n$. Consider a conjugate basis in A given by $\{\vec{u}_1, \cdots, \vec{u}_n\}$. This implies that $\vec{u}_i^\top A\vec{u}_j = 0$ for all i, j where $i \neq j$. Then the $(k+1)$-th iterate of conjugate gradient descent, $\vec{x}_{k+1}$, is calculated as:

$$\vec{x}_{k+1} = \sum_{i=1}^{k} \frac{\vec{u}_i^\top \vec{r}_i}{\vec{u}_i^\top A\vec{u}_i} \vec{u}_i, \tag{90}$$

where $\vec{r}_i = \vec{b} - A\vec{x}_i$. Recall from part (d) that

$$\vec{x}^\star = \sum_{i=1}^{n} \frac{\vec{u}_i^\top \vec{b}}{\vec{u}_i^\top A\vec{u}_i} \vec{u}_i. \tag{91}$$

Show that $\vec{x}_{n+1} = \vec{x}^\star$. This means that the conjugate gradient method converges to the solution of $A\vec{x} = \vec{b}$ in n iterations.

HINT: Prove and use the fact that $\vec{u}_k^\top \vec{r}_k = \vec{u}_k^\top \vec{b}$ for every $1 \leq k \leq n$.

HINT: Use part (a).

Solution: First we prove that $\vec{u}_k^\top \vec{r}_k = \vec{u}_k^\top \vec{b}$ for every $1 \leq k \leq n$:

$$\vec{u}_k^\top \vec{r}_k = \vec{u}_k^\top \vec{b} - \vec{u}_k^\top A\vec{x}_k \tag{92}$$

Since $\vec{x}_k \in \operatorname{span}(\vec{u}_1, \ldots, \vec{u}_{k-1})$, by part (a), $\vec{u}_k^\top A\vec{x}_k = 0$. Hence, we have that $\vec{u}_k^\top \vec{r}_k = \vec{u}_k^\top \vec{b}$. Now, for the main claim, we express $\vec{x}_{n+1}$ in summation form:

$$\begin{aligned}\vec{x}_{n+1} &= \sum_{i=1}^{n} \frac{\vec{u}_i^\top \vec{r}_i}{\vec{u}_i^\top A\vec{u}_i} \vec{u}_i & (93)\\ &= \sum_{i=1}^{n} \frac{\vec{u}_i^\top b}{\vec{u}_i^\top A\vec{u}_i} \vec{u}_i & (94)\\ &= \vec{x}^\star. & (95)\end{aligned}$$

(f) (8 pts) Recall that $A \in \mathbb{S}^n_{++}$ is a symmetric positive definite matrix, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^n$. Consider $f(\vec{x}) = \frac{1}{2}\vec{x}^\top A\vec{x} - \vec{b}^\top \vec{x}$. Note that the conjugate gradient iterates in (90) can be recursively expressed as

$$\vec{x}_{k+1} = \vec{x}_k + \frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k} \vec{u}_k, \tag{96}$$

where $\vec{r}_k = \vec{b} - A\vec{x}_k$.

Show that for all $1 \leq k \leq n$ **we have that** $f(\vec{x}_{k+1}) \leq f(\vec{x}_k)$**.** Thus, if we use the conjugate gradient method to minimize the the function $f(\vec{x})$, the objective function will be non-increasing in every iteration.

HINT: Use the first order condition of convexity and part (a).

HINT: Use the fact that $\vec{u}_k^\top \vec{r}_k = \vec{u}_k^\top \vec{b}$ for every $1 \leq k \leq n$.

Solution: $f(x)$ is a convex function, so we can use the first order condition of convexity. Writing the taylor expansion around x_{k+1}:

$$
\begin{aligned}
f(\vec{x}_k) &\geq f(\vec{x}_{k+1}) + \nabla_x f(\vec{x}_{k+1})^\top (\vec{x}_k - \vec{x}_{k+1}) && (97)\\
&= f(\vec{x}_{k+1}) - \left(A\vec{x}_{k+1} - \vec{b}\right)^\top \left(\frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\vec{u}_k\right) && (98)\\
&= f(\vec{x}_{k+1}) - \left(A\left(\vec{x}_k + \frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\vec{u}_k\right) - \vec{b}\right)^\top \left(\frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\vec{u}_k\right) && (99)\\
&= f(\vec{x}_{k+1}) - \left(\frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\right)\left(\vec{x}_k^\top A\vec{u}_k + \frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\vec{u}_k^\top A\vec{u}_k - \vec{b}^\top \vec{u}_k\right) && (100)\\
&= f(\vec{x}_{k+1}) - \left(\frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\right)\left(\vec{x}_k^\top A\vec{u}_k + \vec{u}_k^\top \vec{b} - \vec{u}_k^\top \vec{b}\right) && (101)\\
&= f(\vec{x}_{k+1}) - \left(\frac{\vec{u}_k^\top \vec{r}_k}{\vec{u}_k^\top A\vec{u}_k}\right)(\vec{x}_k^\top A\vec{u}_k) && (102)\\
&= f(\vec{x}_{k+1}). && (103)
\end{aligned}
$$

The last equality comes from part (a), since $\vec{x}_k \in \operatorname{span}(\vec{u}_1, \ldots, \vec{u}_{k-1})$ so $\vec{x}_k^\top A\vec{u}_k = 0$.