Final

Exam Location:

PRINT your student ID: ______________________

PRINT AND SIGN your name: ______________, ______________ ______________
(last) (first) (sign)

PRINT your discussion sections and (u)GSIs (the ones you attend): ______________________

Name and SID of the person to your left: ______________________

Name and SID of the person to your right: ______________________

Name and SID of the person in front of you: ______________________

Name and SID of the person behind you: ______________________

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.

2. **SID (2 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 (2 pts)**

   (a) (1 pts) What is your favorite song or piece of music?

   (b) (1 pts) What is your favorite hobby or pastime?

Do not turn the page until your proctor tells you to do so.

4. Singular Values (6 pts)

Suppose that $A \in \mathbb{R}^{4\times 3}$ has singular values 0, 1, $\sqrt{5}$, and 3. Let $B = \begin{bmatrix} A & 2I_4 & 3A \end{bmatrix} \in \mathbb{R}^{4\times 10}$, where $I_4 \in \mathbb{R}^{4\times 4}$ is the 4×4 identity matrix. **What are the nonzero singular values of B?** *Show your work and justify your answer(s).*

HINT: Consider the matrix $BB^\top \in \mathbb{R}^{4\times 4}$.

5. Hyperplanes (6 pts)

(a) (2 pts) Let $\vec{c}, \vec{x}_0 \in \mathbb{R}^n$, and let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n : \vec{c}^\top(\vec{x} - \vec{x}_0) = 0\}$ be a hyperplane. **Describe the set of all vectors normal to $\mathcal{H}$.** *You do not need to show your work for this subpart.*

(b) (4 pts) Let $\vec{c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2$, and define the hyperplane $\mathcal{H} = \{\vec{x} \in \mathbb{R}^2 : \vec{c}^\top \vec{x} = 0\}$. Let $\vec{y} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \in \mathbb{R}^2$. **Compute the minimum distance from $\vec{y}$ to $\mathcal{H}$.** *Show your work and justify your answer(s).*

6. Shadow Prices (13 pts)

You are opening a new boba tea shop, MooncoW Boba. You have some ingredients in stock, and are selling two different boba teas. The products and ingredient quantities for each tea, along with your stock, are listed below.

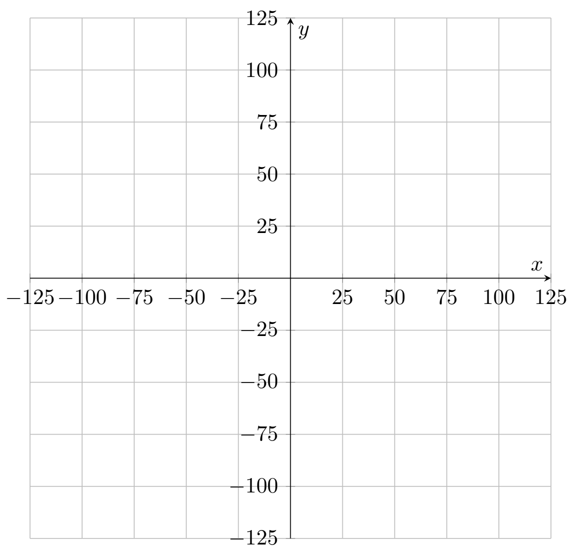
	Tea (Cups)	Milk (Cups)	Boba Pearls (Scoops)	Price ($)
Mango Tea, 1 Serving	3	0	1	5
Matcha Latte, 1 Serving	1	2	1	12
MooncoW Stock	300	200	150	N/A

You want to maximize the revenue of selling x servings of Mango Tea and y servings of Matcha Latte.

(a) (7 pts) The revenue maximization problem can be expressed as the following linear program (LP) $\mathcal{P}_0$:

$$\begin{aligned} \mathcal{P}_0: \qquad p^\star = \max_{x,y\in\mathbb{R}} \quad & 5x + 12y && (1) \\ \text{s.t.} \quad & 3x + y \leq 300, && (2) \\ & 2y \leq 200, && (3) \\ & x + y \leq 150, && (4) \\ & x \geq 0, && (5) \\ & y \geq 0. && (6) \end{aligned}$$

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 value $p^\star$ and optimal point $(x^\star, y^\star)$ of the LP $\mathcal{P}_0$.** *Justify your answer(s).*

(b) (6 pts) In order to earn more revenue, you change prices. Now, a serving of Mango Tea has a fixed price of a dollars, and a serving of Matcha Latte has a fixed price of b dollars. The new revenue maximization LP, $\mathcal{P}_1$, is as follows:

$$
\begin{aligned}
\mathcal{P}_1: \qquad p^\star = \max_{x,y\in\mathbb{R}} \quad & ax + by && (7)\\
\text{s.t.} \quad & 3x + y \leq 300, && (2)\\
& 2y \leq 200, && (3)\\
& x + y \leq 150, && (4)\\
& x \geq 0, && (5)\\
& y \geq 0. && (6)
\end{aligned}
$$

Consider the following dual variables corresponding to constraints in $\mathcal{P}_1$ and their optimal values:

Constraint	Dual Variable (λ_i)	Optimal Value ($\lambda_i^\star$)
$3x + y \leq 300$	λ_1	10
$2y \leq 200$	λ_2	0
$x + y \leq 150$	λ_3	0
$x \geq 0$	λ_4	0
$y \geq 0$	λ_5	5

i. **Based on the table above, list the constraints of $\mathcal{P}_1$ that must be active at optimum.** *Justify your answer(s).*

ii. **Find, in terms of a and b, the optimal value $p^\star$ and optimal point $(x^\star, y^\star)$ of $\mathcal{P}_1$.** *Justify your answer(s).*

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.

The exam will continue on the following page.

7. Dual of QP (14 pts)

Let $Q \in \mathbb{R}^{n \times n}$ be a *symmetric positive definite matrix*, let $A \in \mathbb{R}^{n \times n}$ be an *invertible matrix*, and let $\vec{b} \in \mathbb{R}^n$. Consider the following quadratic program (QP):

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad \frac{1}{2}\vec{x}^\top Q \vec{x} \qquad \text{s.t.} \quad A\vec{x} \leq \vec{b}. \tag{8}$$

(a) (10 pts) **Prove that the dual problem of (8) is given by:**

$$d^\star = \max_{\vec{\lambda} \in \mathbb{R}^n} \quad -\frac{1}{2}\vec{\lambda}^\top A Q^{-1} A^\top \vec{\lambda} - \vec{\lambda}^\top \vec{b} \qquad \text{s.t.} \quad \vec{\lambda} \geq \vec{0}. \tag{9}$$

Show your work and justify your answer(s).

Recall the following information from the previous part. Let $Q \in \mathbb{R}^{n\times n}$ be a *symmetric positive definite* matrix, let $A \in \mathbb{R}^{n\times n}$ be an *invertible* matrix, and let $\vec{b} \in \mathbb{R}^n$. Recall the problem:

$$\begin{aligned} p^\star = \min_{\vec{x}\in\mathbb{R}^n} \quad & \frac{1}{2}\vec{x}^\top Q\vec{x} \\ \text{s.t.} \quad & A\vec{x} \leq \vec{b}. \end{aligned} \tag{8}$$

(b) (4 pts) If we take the *dual of the dual* of (8), we obtain the following problem (*no need to prove this*):

$$\begin{aligned} q^\star = \max_{\vec{\mu}\in\mathbb{R}^n} \quad & -\frac{1}{2}(\vec{\mu}-\vec{b})^\top (A^{-1})^\top Q A^{-1}(\vec{\mu}-\vec{b}) \\ \text{s.t.} \quad & \vec{\mu} \geq \vec{0}. \end{aligned} \tag{10}$$

Prove that the maximization problem (10) is equivalent to the original quadratic program (8).

HINT: Use the fact that A is invertible to utilize the substitution $\vec{\mu} = \vec{b} - A\vec{x}$.

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.

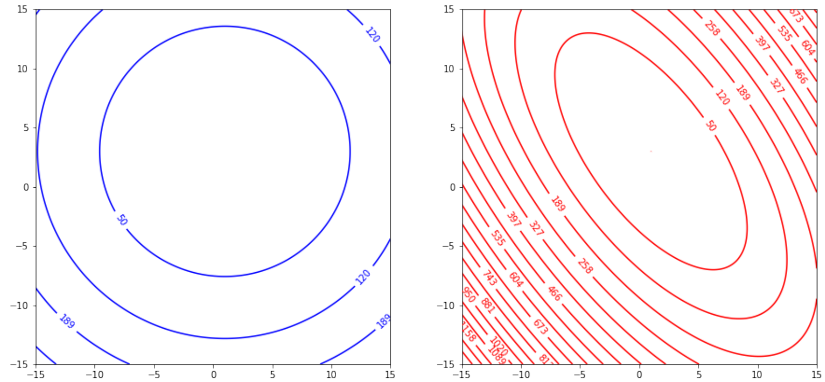
The exam will continue on the following page.

8. Quadratics and Gradient Descent (17 pts)

(a) (4 pts) Let $Q, R \in \mathbb{R}^{2\times 2}$ be symmetric positive definite matrices with the same *smallest* eigenvalue, i.e., $\lambda_2\{Q\} = \lambda_2\{R\}$, and different *largest* eigenvalues, i.e., $\lambda_1\{Q\} \neq \lambda_1\{R\}$. Let $\vec{b}, \vec{c} \in \mathbb{R}^2$. Let $f, g \colon \mathbb{R}^2 \to \mathbb{R}$ be such that

$$f(\vec{x}) \doteq \frac{1}{2}\vec{x}^\top Q\vec{x} + \vec{b}^\top \vec{x}, \qquad g(\vec{x}) \doteq \frac{1}{2}\vec{x}^\top R\vec{x} + \vec{c}^\top \vec{x}, \qquad \text{for all } \vec{x} \in \mathbb{R}^2. \tag{11}$$

Consider the following plots of the level sets of f (*left*) and g (*right*).



Based on these plots, which matrix, Q or R, has the larger condition number? *Justify your answer(s).*

(b) (6 pts) Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix, and let $\vec{b} \in \mathbb{R}^n$. Consider the optimization problem

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \qquad \text{where} \qquad f(\vec{x}) \doteq \frac{1}{2}\vec{x}^\top Q \vec{x} + \vec{b}^\top \vec{x}. \tag{12}$$

Let $\vec{x}^\star$ solve (12). Let $(\vec{x}_t)_{t=0}^\infty$ be gradient descent iterates, with step size $\eta > 0$, for the problem (12). **Write the update rule for $\vec{x}_{t+1}$ in terms of $\vec{x}_t$. Then, use this update rule to prove that**

$$\vec{x}_t - \vec{x}^\star = (I_n - \eta Q)^t (\vec{x}_0 - \vec{x}^\star), \qquad \textbf{for all } t \geq 0. \tag{13}$$

Here $I_n \in \mathbb{R}^{n \times n}$ is the $n \times n$ identity matrix.

HINT: It may be useful to first compute $\vec{x}^\star$ in closed form.

Recall the following information from the previous part. Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix, and let $\vec{b} \in \mathbb{R}^n$. Consider the optimization problem

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \qquad \text{where} \qquad f(\vec{x}) \doteq \frac{1}{2}\vec{x}^\top Q \vec{x} + \vec{b}^\top \vec{x}. \tag{12}$$

Let $\vec{x}^\star$ solve (12). We run gradient descent on (12) with step size $\eta > 0$ to get iterates $(\vec{x}_t)_{t=0}^\infty$. In part 8(b) we showed that

$$\vec{x}_t - \vec{x}^\star = (I_n - \eta Q)^t (\vec{x}_0 - \vec{x}^\star), \qquad \text{for all } t \geq 0, \tag{13}$$

where $I_n \in \mathbb{R}^{n \times n}$ is the $n \times n$ identity matrix.

(c) (4 pts) We now analyze the convergence of gradient descent in an eigenvector basis. Let $Q = U \Lambda U^\top$, where U is orthonormal and Λ is diagonal with entries $\lambda_1 \geq \cdots \geq \lambda_n$. Let $\vec{z}_t = U^\top \vec{x}_t$ and $\vec{z}^\star = U^\top \vec{x}^\star$. **Using (13), prove that**

$$(\vec{z}_t - \vec{z}^\star)_i = (1 - \eta \lambda_i)^t (\vec{z}_0 - \vec{z}^\star)_i, \qquad \textbf{for all } t \geq 0 \quad \textbf{and} \quad i \in \{1, \ldots, n\}, \tag{14}$$

where $(\vec{z})_i$ is the i^{th} entry of $\vec{z}$ for any vector $\vec{z}$.

Recall the following information from previous parts. Let $Q \in \mathbb{R}^{n\times n}$ be a *symmetric positive definite* matrix such that $Q = U\Lambda U^\top$, where U is orthonormal and Λ is diagonal with entries $\lambda_1 \geq \cdots \geq \lambda_n$, and we let $\vec{b} \in \mathbb{R}^n$. Consider the optimization problem

$$\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) \qquad \text{where} \qquad f(\vec{x}) \doteq \frac{1}{2}\vec{x}^\top Q\vec{x} + \vec{b}^\top\vec{x}. \tag{12}$$

Let $\vec{x}^\star$ solve (12). We run gradient descent on (12) with step size $\eta > 0$ to get iterates $(\vec{x}_t)_{t=0}^{\infty}$. Let $\vec{z}_t = U^\top\vec{x}_t$ and $\vec{z}^\star = U^\top\vec{x}^\star$. In parts 8(b) and 8(c) we showed that

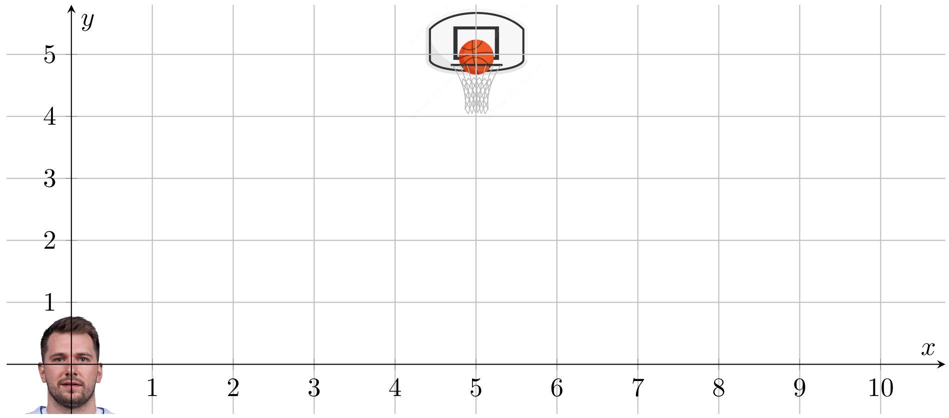
$$(\vec{z}_t - \vec{z}^\star)_i = (1-\eta\lambda_i)^t(\vec{z}_0 - \vec{z}^\star)_i, \qquad \text{for all } t \geq 0 \quad \text{and} \quad i \in \{1,\ldots,n\}, \tag{14}$$

where $(\vec{z})_i$ is the i^{th} entry of $\vec{z}$ for any vector $\vec{z}$.

(d) (3 pts) Define the interval $(\alpha,\beta) \subseteq \mathbb{R}$ such that $\lim_{t\to\infty}\vec{z}_t = \vec{z}^\star$ for any choice of initialization $\vec{z}_0$ if and only if the step size $\eta \in (\alpha,\beta)$. **Using (14), identify α and β.** *Show your work and justify your answer(s).*

9. Luka Navigates a Grid (10 pts)

Luka is playing basketball, and wants to get to the basket for a dunk. He is currently at position $\vec{v}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and will take 10 steps through positions $\vec{v}_1, \ldots, \vec{v}_9$ to get to position $\vec{v}_{10} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$, which is where the basket is.



(a) (4 pts) Luka has two defenders, located at points $\vec{x}, \vec{y} \in \mathbb{R}^2$, and he wants to take a path such that:

- The path maximizes the smallest distance from any location on the path, i.e., the $\vec{v}_t$, to any defender, i.e., $\vec{x}$ or $\vec{y}$.
- The path does not go out of bounds, i.e., outside the rectangle with corners $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 10 \\ 5 \end{bmatrix}$.

Formulate the optimization problem to find Luka's steps by filling in choices for (1), (2), and (3) below. *You do not need to show your work for this subpart.*

NOTE: Your problem doesn't have to be written in standard form, nor does it have to be convex.

$$\max_{\substack{q \in \mathbb{R} \\ \vec{v}_0, \vec{v}_1 \ldots \vec{v}_9, \vec{v}_{10} \in \mathbb{R}^2}} \quad q \tag{15}$$

$$\text{s.t.} \quad q \leq \|\vec{x} - \vec{v}_t\|_2, \quad \text{for all } t \in \{0, \ldots, 10\}, \tag{16}$$

$$q \leq \|\boxed{\quad (1) \quad}\|_2, \quad \text{for all } t \in \{0, \ldots, 10\}, \tag{17}$$

$$\boxed{\quad (2) \quad} \leq \vec{v}_t, \quad \text{for all } t \in \{0, \ldots, 10\}, \tag{18}$$

$$\vec{v}_t \leq \boxed{\quad (3) \quad}, \quad \text{for all } t \in \{0, \ldots, 10\}, \tag{19}$$

$$\vec{v}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \vec{v}_{10} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}. \tag{20}$$

Recall the following information from the previous part. Luka is currently at position $\vec{v}_0$, and wants to travel through positions $\vec{v}_1, \ldots, \vec{v}_9$ to get to position $\vec{v}_{10}$ subject to some constraints. In part 9(a), we formulated an optimization problem which outputs an optimal path for Luka.

(b) (3 pts) Now, suppose each step Luka makes must be exactly one unit long in the ℓ^2-norm sense. **Write one or more constraints that capture this condition.** *You do not need to show your work for this subpart.*

NOTE: Your constraint(s) do not have to be convex.

(c) (3 pts) Suppose $p_{\mathrm{a}}^{\star}$ is the optimal value of our problem from part 9(a). Then, suppose $p_{\mathrm{b}}^{\star}$ is the optimal value of this problem after we add our constraint from part 9(b). **Which of the following is guaranteed to be true?**

(A) $p_{\mathrm{a}}^{\star} = p_{\mathrm{b}}^{\star}$;

(B) $p_{\mathrm{a}}^{\star} \leq p_{\mathrm{b}}^{\star}$;

(C) $p_{\mathrm{a}}^{\star} \geq p_{\mathrm{b}}^{\star}$;

(D) None of the above.

Justify your answer.

10. Bandwidth Allocation (20 pts)

(a) (6 pts) **Consider the functions f_1 and f_2 defined below. Is f_1 convex? Is f_2 convex? For each function, if it is convex, prove it; if it is not, explain why.** *Justify your answer(s).*

i. $f_1 : S_1 \to \mathbb{R}$, where $S_1 \doteq \{x \in \mathbb{R} : x > 0\}$, and $f_1(x) \doteq 1/x$ for all $x \in S_1$.

ii. $f_2 : S_2 \to \mathbb{R}$, where $S_2 \doteq \{x \in \mathbb{R} : x \neq 0\}$, and $f_2(x) \doteq 1/x$ for all $x \in S_2$.

(b) (6 pts) In a network bandwidth allocation scenario, a total of n clients share a communication channel to send signals. The channel has an available bandwidth that has been normalized to 1. The fraction of bandwidth allocated to the i^{th} client is represented by x_i. Since the channel bandwidth is shared among n clients and every client must be allocated a non-zero fraction, $\sum_{i=1}^{n} x_i \leq 1$ and $x_i > 0$. If a client gets more bandwidth, they have better performance and their error decreases. Let the error coefficient associated with the i^{th} client be $c_i > 0$, so that the error incurred with x_i fraction of the bandwidth is c_i/x_i. The goal is to minimize the sum of incurred error among n clients. The optimization problem is formulated as follows.

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} \frac{c_i}{x_i} \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i \leq 1, \\ & x_i > 0, \qquad \text{for all } i \in \{1, \dots, n\}. \end{aligned} \tag{21}$$

Here, as usual, x_i denotes the i^{th} entry of $\vec{x}$. **Prove that if $\vec{x}^\star$ is an optimal solution to (21) then $\sum_{i=1}^{n} x_i^\star = 1$.**

HINT: Let $\vec{z}$ be optimal for (21) yet such that $\sum_{i=1}^{n} z_i < 1$. Let $s \doteq 1 - \sum_{i=1}^{n} z_i$ and consider $\vec{y} \doteq \vec{z} + s\vec{e}_1$, where $\vec{e}_1 \in \mathbb{R}^n$ is the first column of the identity matrix.

(c) (8 pts) Now consider the following optimization problem (again with $c_1, \ldots, c_n > 0$):

$$\begin{aligned} p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} \frac{c_i}{x_i} \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1, \\ & x_i > 0, \qquad \text{for all } i \in \{1, \ldots, n\}. \end{aligned} \tag{22}$$

Again, as usual, x_i denotes the i^{th} entry of $\vec{x}$. In part 10(b), we proved that the earlier problem (21) is a relaxation of (22). We now solve the problem (22).

i. **First, state the Cauchy-Schwarz inequality and its equality condition.**

ii. **Next, for any $\vec{x}$ feasible for (22), find vectors $\vec{u}, \vec{w}$ such that $\|\vec{u}\|_2^2 = \sum_{i=1}^{n} c_i/x_i$ and $\|\vec{w}\|_2^2 = \sum_{i=1}^{n} x_i = 1$, and apply Cauchy-Schwarz to get a lower bound on $\sum_{i=1}^{n} c_i/x_i$.**

iii. **Finally, find the optimal $\vec{x}^\star$ which achieves this lower bound via the equality condition of Cauchy-Schwarz.**

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.

The exam will continue on the following page.

11. Maximum Entropy (17 pts)

We aim to solve the following optimization problem:

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad f(\vec{x}) \qquad \text{where} \qquad f(\vec{x}) \doteq \sum_{i=1}^{n} x_i \log(x_i) \tag{23}$$

$$\text{s.t.} \quad x_i \geq 0, \qquad \text{for all } i \in \{1, \dots, n\}, \tag{24}$$

$$\sum_{i=1}^{n} x_i = 1, \tag{25}$$

where we adopt the convention that $0 \log(0) = 0$ so as to define f on the entire feasible set. This problem determines the probability distribution on $\{1, \dots, n\}$ with the maximal *Shannon entropy*, although this is irrelevant to solving the problem itself.

(a) (3 pts) Let $L(\vec{x}, \vec{\lambda}, \nu)$ be the Lagrangian of the problem, where $\vec{\lambda}$ are the dual variables corresponding to the constraints in (24), and ν is the dual variable corresponding to the constraint in (25). **Prove that**

$$L(\vec{x}, \vec{\lambda}, \nu) = f(\vec{x}) - \vec{x}^\top(\vec{\lambda} - \nu \vec{1}_n) - \nu, \tag{26}$$

where $\vec{1}_n \in \mathbb{R}^n$ is the vector of ones in $\mathbb{R}^n$. *Show your work and justify your answer(s).*

Recall the following information from the previous part. We aim to solve the following optimization problem:

$$p^{\star} = \min_{\vec{x} \in \mathbb{R}^n} \quad f(\vec{x}) \qquad \text{where} \qquad f(\vec{x}) \doteq \sum_{i=1}^{n} x_i \log(x_i) \tag{23}$$

$$\text{s.t.} \quad x_i \geq 0, \qquad \text{for all } i \in \{1, \ldots, n\}, \tag{24}$$

$$\sum_{i=1}^{n} x_i = 1, \tag{25}$$

where we adopt the convention that $0 \log(0) = 0$ so as to define f on the entire feasible set. Let $\vec{\lambda} \in \mathbb{R}^n$ be the dual variables corresponding to the constraint (24), and $\nu \in \mathbb{R}$ be the dual variable corresponding to the constraint (25).

(b) (4 pts) **Prove that strong duality holds for this problem.**

NOTE: You may assume (*i.e., do not need to prove*) that f is convex.

Recall the following information from previous parts. We aim to solve the following optimization problem:

$$p^\star = \min_{\vec{x} \in \mathbb{R}^n} \quad f(\vec{x}) \qquad \text{where} \qquad f(\vec{x}) \doteq \sum_{i=1}^{n} x_i \log(x_i) \tag{23}$$

$$\text{s.t.} \quad x_i \geq 0, \qquad \text{for all } i \in \{1, \ldots, n\}, \tag{24}$$

$$\sum_{i=1}^{n} x_i = 1, \tag{25}$$

where we adopt the convention that $0 \log(0) = 0$ so as to define f on the entire feasible set. Let $\vec{\lambda} \in \mathbb{R}^n$ be the dual variables corresponding to the constraints in (24), and let $\nu \in \mathbb{R}$ be the dual variable corresponding to the constraint in (25). In part 11(a) we showed that, if we define $\vec{1}_n \in \mathbb{R}^n$ to be the vector of ones in $\mathbb{R}^n$, the Lagrangian is

$$L(\vec{x}, \vec{\lambda}, \nu) = f(\vec{x}) - \vec{x}^\top (\vec{\lambda} - \nu \vec{1}_n) - \nu. \tag{26}$$

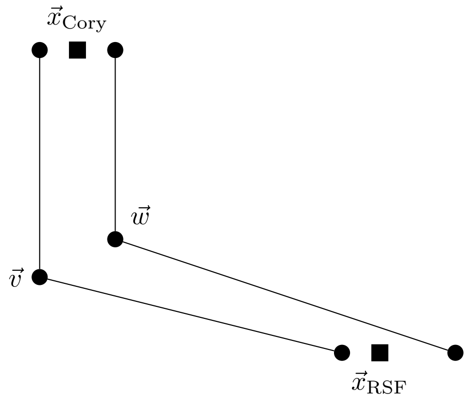
In part 11(b) we showed that strong duality holds.

(c) (10 pts) The results of part 11(b), along with the fact that the problem is convex, mean that the KKT conditions are necessary and sufficient for optimality (*you do not need to prove this*). **State the KKT conditions and use them to solve for optimal primal and dual variables** $(\vec{x}^\star, \vec{\lambda}^\star, \nu^\star)$**.** *Show your work and justify your answer(s).*

HINT: First, prove that $x_i^\star = \mathrm{e}^{\lambda_i^\star - (1+\nu^\star)}$ *for each* $i \in \{1, \ldots, n\}$.

12. Path-Planning and SOCPs (10 pts)

Anish wants to find the best way to carry final exams from Cory Hall ($\vec{x}_{\text{Cory}}$) to RSF ($\vec{x}_{\text{RSF}}$). He identifies that the best possible route, going around the edge of campus, must stay within the following path. *NOTE*: *The diagram below is not drawn to scale.*



Here $\vec{x}_{\text{Cory}}, \vec{x}_{\text{RSF}}, \vec{v}, \vec{w}$ are points in $\mathbb{R}^2$. Anish wants to find the optimal $\vec{x}$, given by the solution to the following problem $\mathcal{P}_0$, to find the shortest path:

$$\mathcal{P}_0: \quad \min_{\substack{\vec{x}\in\mathbb{R}^2 \\ \theta\in\mathbb{R}}} \quad \|\vec{x} - \vec{x}_{\text{Cory}}\|_2 + \|\vec{x} - \vec{x}_{\text{RSF}}\|_2 \tag{27}$$

$$\text{s.t.} \quad \vec{x} = \vec{w} + \theta(\vec{v} - \vec{w}), \tag{28}$$

$$0 \leq \theta \leq 1. \tag{29}$$

We restrict $\vec{x}$ to be a convex combination of $\vec{v}$ and $\vec{w}$. The objective function is the total length of the path.

(a) (4 pts) **Is the problem $\mathcal{P}_0$, as written (i.e., without any reformulations), a SOCP?** *Justify your answer(s).*

Recall the following information from the previous part. Consider the problem

$$\mathcal{P}_0: \quad \min_{\substack{\vec{x}\in\mathbb{R}^2 \\ \theta\in\mathbb{R}}} \quad \|\vec{x}-\vec{x}_{\text{Cory}}\|_2 + \|\vec{x}-\vec{x}_{\text{RSF}}\|_2 \tag{27}$$

$$\text{s.t.} \quad \vec{x} = \vec{w} + \theta(\vec{v}-\vec{w}), \tag{28}$$

$$0 \le \theta \le 1. \tag{29}$$

(b) (6 pts) We now reformulate $\mathcal{P}_0$. **Prove that the optimal $\theta^\star$ for the problem $\mathcal{P}_0$ is the same as the optimal $\theta^\star$ for the following problem $\mathcal{P}_1$:**

$$\mathcal{P}_1: \quad \min_{\theta,s,t\in\mathbb{R}} \quad s+t \tag{30}$$

$$\text{s.t.} \quad 0 \le \theta \le 1, \tag{31}$$

$$\|\theta(\vec{v}-\vec{w}) + \vec{w} - \vec{x}_{\text{Cory}}\|_2 \le s, \tag{32}$$

$$\|\theta(\vec{v}-\vec{w}) + \vec{w} - \vec{x}_{\text{RSF}}\|_2 \le t. \tag{33}$$

HINT: If you introduce inequality constraints relating to slack variables, remember to justify why they must achieve equality at the optimal solution.

13. Equivariance of Newton's Method (9 pts)

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a twice-continuously-differentiable function with invertible Hessian. Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix. Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be defined as

$$g(\vec{y}) \doteq f(A\vec{y}), \qquad \text{for all } \vec{y} \in \mathbb{R}^n. \tag{34}$$

Let $(\vec{x}_t)_{t=0}^{\infty}$ be a sequence of Newton's method iterates on f starting from $\vec{x}_0 \in \mathbb{R}^n$, and let $(\vec{y}_t)_{t=0}^{\infty}$ be a sequence of Newton's method iterates on g starting from $\vec{y}_0 = A^{-1}\vec{x}_0$. **Prove that $\vec{y}_t = A^{-1}\vec{x}_t$ for all t.**

HINT: Use induction on t, with base case $t = 0$.

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.
The exam will continue on the following page.

14. Error of LASSO Regression (24 pts)

Let $X \in \mathbb{R}^{n \times d}$ be a matrix, let $\vec{\alpha}_0 \in \mathbb{R}^d$ and $\vec{w} \in \mathbb{R}^n$ be vectors, and let $\lambda > 0$ be a scalar. Suppose that $\vec{y} \in \mathbb{R}^n$ is defined as

$$\vec{y} \doteq X\vec{\alpha}_0 + \vec{w}. \tag{35}$$

Given X and $\vec{y}$, we want to recover $\vec{\alpha}_0$ via LASSO regression. We find an estimate $\vec{\alpha}^\star$ which solves the LASSO problem:

$$\vec{\alpha}^\star = \operatorname*{argmin}_{\vec{\alpha} \in \mathbb{R}^d} f(\vec{\alpha}), \qquad \text{where} \qquad f(\vec{\alpha}) \doteq \frac{1}{2n}\|\vec{y} - X\vec{\alpha}\|_2^2 + \lambda\|\vec{\alpha}\|_1. \tag{36}$$

Define the error of the estimate $\vec{\alpha}^\star$ to be the vector $\vec{\delta} \in \mathbb{R}^d$, i.e.,

$$\vec{\delta} \doteq \vec{\alpha}^\star - \vec{\alpha}_0. \tag{37}$$

In this problem, we will derive and interpret an upper bound on the squared prediction error $\|X\vec{\delta}\|_2^2 = \|X\vec{\alpha}^\star - X\vec{\alpha}_0\|_2^2$.

(a) (5 pts) **Prove that**

$$\|\vec{y} - X\vec{\alpha}^\star\|_2^2 = \|\vec{w}\|_2^2 + \|X\vec{\delta}\|_2^2 - 2(X^\top \vec{w})^\top \vec{\delta}. \tag{38}$$

HINT: Consider adding and subtracting $X\vec{\alpha}_0$ from the term inside the norm on the left-hand side of (38).

Recall the following information from the previous part. We have $\vec{y} = X\vec{\alpha}_0 + \vec{w}$, where $X \in \mathbb{R}^{n\times d}$, $\vec{y} \in \mathbb{R}^n$, $\vec{w} \in \mathbb{R}^n$, and $\vec{\alpha}_0 \in \mathbb{R}^d$. Given X and $\vec{y}$, we want to recover $\vec{\alpha}_0$ via LASSO regression, and we find an estimate $\vec{\alpha}^\star \in \mathbb{R}^d$ such that:

$$\vec{\alpha}^\star = \operatorname*{argmin}_{\vec{\alpha}\in\mathbb{R}^d} f(\vec{\alpha}), \qquad \text{where} \qquad f(\vec{\alpha}) \doteq \frac{1}{2n}\|\vec{y} - X\vec{\alpha}\|_2^2 + \lambda\|\vec{\alpha}\|_1, \tag{36}$$

where $\lambda > 0$. We define $\vec{\delta} \doteq \vec{\alpha}^\star - \vec{\alpha}_0$. In part 14(a), we derived the equality

$$\|\vec{y} - X\vec{\alpha}^\star\|_2^2 = \|\vec{w}\|_2^2 + \|X\vec{\delta}\|_2^2 - 2(X^\top\vec{w})^\top\vec{\delta}. \tag{38}$$

(b) (8 pts) **Using (38), prove that**

$$\frac{1}{2n}\|X\vec{\delta}\|_2^2 \le \left(\frac{X^\top\vec{w}}{n}\right)^\top \vec{\delta} + \lambda(\|\vec{\alpha}_0\|_1 - \|\vec{\alpha}^\star\|_1). \tag{39}$$

HINT: Start by arguing that $f(\vec{\alpha}^\star) \le f(\vec{\alpha}_0)$.

Recall the following information from previous parts. We have $\vec{y} = X\vec{\alpha}_0 + \vec{w}$, where $X \in \mathbb{R}^{n\times d}$, $\vec{y} \in \mathbb{R}^n$, $\vec{w} \in \mathbb{R}^n$, and $\vec{\alpha}_0 \in \mathbb{R}^d$. Given X and $\vec{y}$, we want to recover $\vec{\alpha}_0$ via LASSO regression, and we find an estimate $\vec{\alpha}^\star \in \mathbb{R}^d$ which solves the LASSO problem with regularizer $\lambda > 0$. We define $\vec{\delta} \doteq \vec{\alpha}^\star - \vec{\alpha}_0$. In parts 14(a) and 14(b), we derived the inequality

$$\frac{1}{2n}\|X\vec{\delta}\|_2^2 \leq \left(\frac{X^\top \vec{w}}{n}\right)^\top \vec{\delta} + \lambda(\|\vec{\alpha}_0\|_1 - \|\vec{\alpha}^\star\|_1). \tag{39}$$

(c) (6 pts) From now on, assume that λ is chosen so that[1]

$$2\left\|\frac{X^\top \vec{w}}{n}\right\|_\infty \leq \lambda. \tag{40}$$

Using (39) and (40), prove that

$$\frac{1}{2n}\|X\vec{\delta}\|_2^2 \leq \lambda\left(\frac{1}{2}\|\vec{\delta}\|_1 + \|\vec{\alpha}_0\|_1 - \|\vec{\alpha}^\star\|_1\right). \tag{41}$$

HINT: Use Hölder's inequality.

[1] Intuitively, this means that λ is chosen to be larger than the maximum correlation of $\vec{w}$ with any one column ("feature") of X.

Recall the following information from previous parts. We have $\vec{y} = X\vec{\alpha}_0 + \vec{w}$, where $X \in \mathbb{R}^{n\times d}$, $\vec{y} \in \mathbb{R}^n$, $\vec{w} \in \mathbb{R}^n$, and $\vec{\alpha}_0 \in \mathbb{R}^d$. Given X and $\vec{y}$, we want to recover $\vec{\alpha}_0$ via LASSO regression, and we find an estimate $\vec{\alpha}^\star \in \mathbb{R}^d$ which solves the LASSO problem with regularizer $\lambda > 0$. We define $\vec{\delta} \doteq \vec{\alpha}^\star - \vec{\alpha}_0$.

(d) (5 pts) The results of parts 14(a), 14(b), and 14(c) can be used to show that if $\lambda \geq 2\|X^\top \vec{w}/n\|_\infty$ then

$$\frac{1}{n}\|X\vec{\delta}\|_2^2 \leq 12\lambda\|\vec{\alpha}_0\|_1. \tag{42}$$

In this subpart, assume (42) holds — *you do not need to prove it*. Suppose that $\vec{\alpha}_0$ has $s > 0$ nonzero entries, i.e., $\|\vec{\alpha}_0\|_0 = s$. (Here $\|\cdot\|_0$ is the ℓ^0 "norm" which counts the number of nonzero entries of its input).

i. **Using (42), prove that**

$$\frac{1}{n}\|X\vec{\delta}\|_2^2 \leq 12\lambda\sqrt{s}\|\vec{\alpha}_0\|_2. \tag{43}$$

HINT: Recall the inequality (proved in homework, and thus usable without proof) that

$$\|\vec{x}\|_1 \leq \sqrt{\|\vec{x}\|_0} \cdot \|\vec{x}\|_2, \qquad \textit{for all } \vec{x} \in \mathbb{R}^d. \tag{44}$$

ii. Suppose that $\|\vec{\alpha}_0\|_2 = 1$ and $\lambda = 1$ are fixed. **How does the right-hand side of (43) grow in the high-dimensional limit $d \to \infty$ if $s = \sqrt{d}$?**

iii. Suppose that $\|\vec{\alpha}_0\|_2 = 1$ and $\lambda = 1$ are fixed. **How does the right-hand side of (43) grow in the high-dimensional limit $d \to \infty$ if $s = 1$?**

The result of this problem shows that the prediction error $\|X\vec{\delta}\|_2^2$ is bounded, even when arbitrarily many spurious features are added (i.e., $d \to \infty$), when using a LASSO estimator.

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.

This is an extra page for scratch work that will not be graded unless you tell us in the original problem space.

Doodle page!

Draw us something if you want, or give us suggestions or complaints.

You can also use this page to report anything suspicious that you might have noticed.

You can also use this page to write solutions if you need the space, but please tell us in the original problem space.

Final Instructions

Read the following instructions before the exam.

There are 14 problems of varying numbers of points. There are 150 points on the exam. You have 180 minutes for the exam. The problems are of varying difficulty, so pace yourself accordingly, do easier problems first, and avoid spending too much time on any one question until you have gotten all of the other points you can. Problems are not necessarily ordered in terms of difficulty, so be sure to read all the problems.

There are 36 pages on the exam, so there should be 18 sheets of paper in the exam. The exam is printed double-sided. Do not forget the problems on the back sides of the pages! Notify a proctor immediately if a page is missing. **Do not tear out or remove any of the pages. Do not remove the exam from the exam room.**

No collaboration is allowed, and do not attempt to cheat in any way. Cheating will not be tolerated.

Write your student ID on each page. If a page is found without a student ID, and some pages from your exam go missing, we will have no way of giving you credit for those pages. All exam pages will be separated during scanning.

You may consult TWO handwritten 8.5” $\times$ 11” note sheet(s) (front and back). No phones, calculators, tablets, computers, other electronic devices, or scratch paper are allowed.

Please write your answers legibly in the boxed spaces provided on the exam. The space provided should be adequate. **If you still run out of space, please use a blank page and clearly tell us in the original problem space where to look for your solution.**

Unless otherwise specified, show all of your work in order to receive full credit. Partial credit will be given for substantial progress on each problem.

We will not be able to answer most questions or offer clarifications during the exam.

If you need to use the restrooms during the exam, bring your student ID card, your phone, and your exam to a proctor. You can collect them once you return from the restrooms.

Our advice to you: if you can’t solve the problem, state and solve a simpler one that captures at least some of its essence. You might get some partial credit, and more importantly, you will perhaps find yourself on a path to the solution.

Good luck!

Do not turn the page until your proctor tells you to do so.