Midterm

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 (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 your favorite restaurant in Berkeley?**

   (b) (1 pts) **What's some music that makes you happy?**

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

4. Rank 1 Approximation Error (7 pts)

Consider a 2×2 matrix

$$A = \begin{bmatrix} -4 & 8 \\ 7 & 1 \end{bmatrix}. \tag{1}$$

The eigenvalue-eigenvector pairs of $A^\top A$ are

$$(\lambda_1, \vec{v}_1) = \left(90, \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}\right), \qquad (\lambda_2, \vec{v}_2) = \left(40, \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}\right). \tag{2}$$

(a) (3 pts) **What are the singular values of the matrix A?** *You do not have to justify your answer.*

HINT: You should not have to do much computation here.

(b) (4 pts) For A as given in (1) with $A^\top A$ as given in (2), consider the following problem:

$$p^\star = \min_{\substack{B \in \mathbb{R}^{2\times 2} \\ \mathrm{rank}(B)=1}} \|A - B\|_F^2 . \tag{3}$$

What is the value of $p^\star$? *Justify your answer.*

HINT: You should not have to do any computation at all here.

5. ℓ^∞ Constraint (7 pts)

Consider the optimization problem

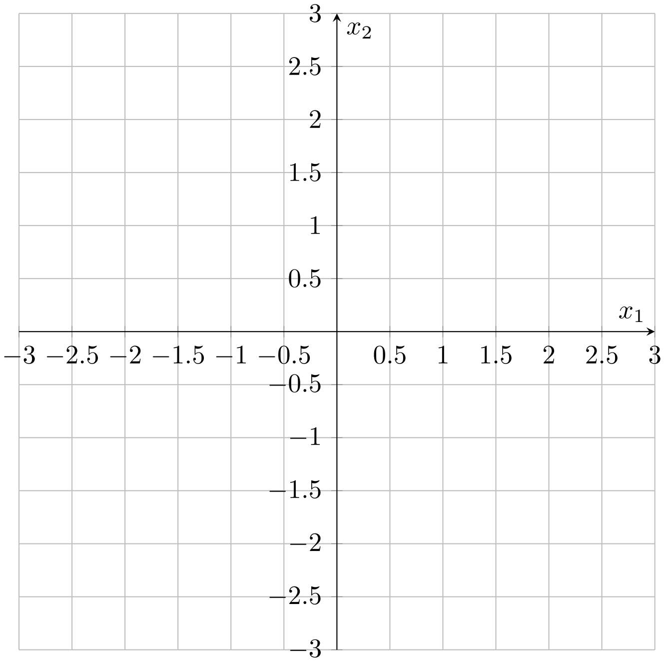
$$\min_{\vec{x}\in\mathbb{R}^2} \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \vec{x} \tag{4}$$

$$\text{s.t.} \quad \|\vec{x}\|_\infty \leq 1 \tag{5}$$

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix}^\top \vec{x} \leq -1, \tag{6}$$

where $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

(a) (4 pts) **Sketch the feasible region of the above optimization problem in the graph provided below. Label each constraint.**



Consider the optimization problem

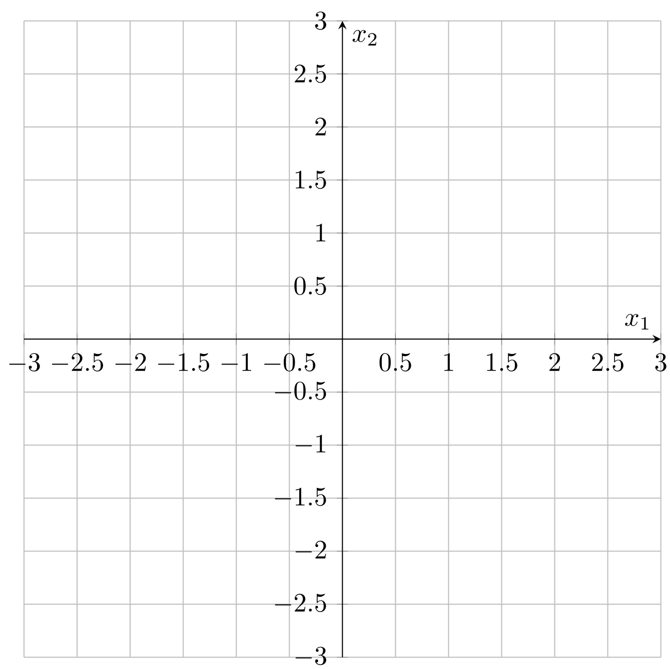
$$\min_{\vec{x} \in \mathbb{R}^2} \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \vec{x} \tag{4}$$

$$\text{s.t.} \quad \|\vec{x}\|_\infty \leq 1 \tag{5}$$

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix}^\top \vec{x} \leq -1, \tag{6}$$

where $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

(b) (3 pts) **Sketch the 2-level set, the 0-level set, and the -2-level set of the objective function in (4). Label each set.**



6. Convexity of the Probability Simplex (5 pts)

Let n be a positive integer. The *probability simplex* on $\mathbb{R}^n$, denoted $\mathcal{P}_n$, is the set

$$\mathcal{P}_n = \left\{ \vec{x} \in \mathbb{R}^n \;\middle|\; x_i \geq 0 \; \forall i, \; \sum_{i=1}^{n} x_i = 1 \right\} \qquad \text{where} \qquad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \tag{7}$$

Is $\mathcal{P}_n$ convex? If yes, prove it. If no, justify your answer using an example.

7. Vector Calculus (12 pts)

(a) (6 pts) Let $A \in \mathbb{S}^n$ be an $n \times n$ symmetric matrix. **Compute the gradient with respect to $\vec{x}$ of the function $f \colon \mathbb{R}^n \setminus \{\vec{0}\} \to \mathbb{R}$ given by**

$$f(\vec{x}) \doteq \frac{\vec{x}^\top A \vec{x}}{\vec{x}^\top \vec{x}}. \tag{8}$$

Show your work.

HINT: Recall the quotient rule for finding the gradient of $h(\vec{x}) = \frac{n(\vec{x})}{d(\vec{x})}$ where n and d are scalar-valued functions:

$$\nabla h(\vec{x}) = \frac{d(\vec{x})\nabla n(\vec{x}) - n(\vec{x})\nabla d(\vec{x})}{(d(\vec{x}))^2}. \tag{9}$$

(b) (6 pts) Let $\vec{u} \in \mathbb{R}^n$. **Compute the Jacobian with respect to $\vec{x}$ of the function $\vec{g} \colon \mathbb{R}^n \to \mathbb{R}^n$ given by**

$$\vec{g}(\vec{x}) \doteq \vec{x}(\vec{x}^\top \vec{u}). \tag{10}$$

Show your work.

8. Gradient Descent (18 pts)

Let $\vec{y} \in \mathbb{R}^n$ be a fixed and known vector. In this problem we will use gradient descent to solve the following problem:

$$\min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) \tag{11}$$

$$\text{where} \quad f_0(\vec{x}) \doteq \frac{1}{2} \|\vec{x} - \vec{y}\|_2^2 . \tag{12}$$

(a) (4 pts) **Is $f_0(\vec{x})$ a convex function?** *Justify your answer.*

We run gradient descent on f_0 with step size $\eta > 0$ and initialization $\vec{x}_0$, obtaining the iterates $\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots$.

(b) (5 pts) **Prove that for each $t \geq 0$ we have**

$$\vec{x}_t - \vec{y} = (1 - \eta)^t (\vec{x}_0 - \vec{y}). \tag{13}$$

Recall that we are attempting to solve the problem

$$\min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) \tag{11}$$

$$\text{where} \quad f_0(\vec{x}) \doteq \frac{1}{2} \|\vec{x} - \vec{y}\|_2^2 . \tag{12}$$

We run gradient descent on f_0 with step size $\eta > 0$ and initialization $\vec{x}_0$, obtaining the iterates $\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots$. Assume that from part (b), for every $t \geq 0$ we have

$$\vec{x}_t - \vec{y} = (1 - \eta)^t (\vec{x}_0 - \vec{y}). \tag{13}$$

(c) (4 pts) **Determine the range of $\eta \in \mathbb{R}$ such that, for *all* initializations $\vec{x}_0$, we have $\vec{x}_1 = \vec{y}$.** *Justify your answer.*

Recall that we are attempting to solve the problem

$$\min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) \tag{11}$$

$$\text{where} \quad f_0(\vec{x}) \doteq \frac{1}{2} \|\vec{x} - \vec{y}\|_2^2 . \tag{12}$$

We run gradient descent on f_0 with step size $\eta > 0$ and initialization $\vec{x}_0$, obtaining the iterates $\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots$. Assume that from part (b), for every $t \geq 0$ we have

$$\vec{x}_t - \vec{y} = (1 - \eta)^t (\vec{x}_0 - \vec{y}). \tag{13}$$

(d) (5 pts) **Determine the range of $\eta \in \mathbb{R}$ such that, for *all* initializations $\vec{x}_0$ and *all* $t \geq 0$, we have that $\vec{x}_t$ is a *convex* combination of $\vec{x}_0$ and $\vec{y}$.** *Justify your answer.*

HINT: What happens if we add $\vec{y}$ to both sides of (13)?

9. Shift Matrix (10 pts)

Let $V \in \mathbb{R}^{n\times n}$ be a square orthonormal matrix, i.e., its columns are orthogonal and have norm 1:

$$V = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_{n-1} & \vec{v}_n \\ \downarrow & \downarrow & \dots & \downarrow & \downarrow \end{bmatrix}. \tag{14}$$

Now, we define the shifted matrix $W \in \mathbb{R}^{n\times n}$, which is composed of the columns of V shifted to the left by 1 index and padded by a zero vector:

$$W = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow & \uparrow \\ \vec{v}_2 & \vec{v}_3 & \dots & \vec{v}_n & \vec{0} \\ \downarrow & \downarrow & \dots & \downarrow & \downarrow \end{bmatrix}. \tag{15}$$

(a) (4 pts) **What is** $\operatorname{rank}(V)$**? What about** $\operatorname{rank}(W)$**?** *You do not need to justify your answers.*

Recall our definitions of $V \in \mathbb{R}^{n\times n}$, which is a square orthonormal matrix, and $W \in \mathbb{R}^{n\times n}$ which is a shifted copy of V:

$$V = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n-1} & \vec{v}_n \\ \downarrow & \downarrow & \cdots & \downarrow & \downarrow \end{bmatrix}, \qquad W = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow & \uparrow \\ \vec{v}_2 & \vec{v}_3 & \cdots & \vec{v}_n & \vec{0} \\ \downarrow & \downarrow & \cdots & \downarrow & \downarrow \end{bmatrix}. \tag{14, 15}$$

(b) (6 pts) **Find a basis for the null space of** $V - W$ **and compute** $\operatorname{rank}(V-W)$**.** *Show your work.*

HINT: Write out the definition of null space for $V - W$*.*

10. Symmetric Matrices (10 pts)

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix.

(a) (4 pts) **Prove that if A is symmetric then A^{2k} is symmetric positive semidefinite for all integers $k > 1$.**

(b) (6 pts) **Prove that if A is symmetric then its *matrix exponential*, defined as $\mathrm{e}^{A} \in \mathbb{R}^{n\times n}$ given by**

$$\mathrm{e}^{A} = I + A + \frac{1}{2}A^2 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k \tag{16}$$

is symmetric positive definite.

HINT: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \mathrm{e}^{x}$ has the series definition

$$\mathrm{e}^{x} = 1 + x + \frac{1}{2}x^2 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}x^k. \tag{17}$$

[Extra page for scratch work that will not be graded unless you tell us in the original problem space.]

11. Second Principal Component (8 pts)

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix with eigenvalue-eigenvector pairs given by $(\lambda_1, \vec{v}_1), \ldots, (\lambda_n, \vec{v}_n)$, where $\lambda_1 > \cdots > \lambda_n$. Consider the problem

$$p^\star = \max_{\vec{x} \in \mathbb{R}^n} \quad \vec{x}^\top A \vec{x} \qquad (18)$$
$$\text{s.t.} \quad \|\vec{x}\|_2^2 = 1$$
$$\vec{x}^\top \vec{v}_1 = 0.$$

Show that $p^\star = \lambda_2$**.** *Prove your answer.*

HINT: First find an $\vec{x}$ *which is feasible and* $\vec{x}^\top A \vec{x} = \lambda_2$. *Then show that* $p^\star \leq \lambda_2$.

[Extra page for scratch work that will not be graded unless you tell us in the original problem space.]

12. Block Ridge Regression (13 pts)

In this problem, we consider a certain generalization of ridge regression. For $d > 0$, let $A \in \mathbb{R}^{n\times(3d)}$ and $y \in \mathbb{R}^n$. Let $\vec{x}_1, \vec{x}_2, \vec{x}_3 \in \mathbb{R}^d$ be three vectors, each of dimension d. We associate regularization parameter λ_i^2 to each vector $\vec{x}_i$. We stack the $\vec{x}_i$ up to get a long $3d$-dimensional vector $\vec{x} \in \mathbb{R}^{3d}$:

$$\vec{x} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \end{bmatrix}. \tag{19}$$

With this notation, the block ridge regression problem is

$$\vec{x}_{\mathrm{BRR}} = \operatorname*{argmin}_{\vec{x}\in\mathbb{R}^{3d}} f_0(\vec{x}) \tag{20}$$

$$\text{where} \qquad f_0(\vec{x}) \doteq \|A\vec{x} - \vec{y}\|_2^2 + \sum_{i=1}^{3} \lambda_i^2 \|\vec{x}_i\|_2^2 \tag{21}$$

$$= \|A\vec{x} - \vec{y}\|_2^2 + \|D\vec{x}\|_2^2 \quad \text{for} \quad D = \begin{bmatrix} \lambda_1 I_d & 0_{d\times d} & 0_{d\times d} \\ 0_{d\times d} & \lambda_2 I_d & 0_{d\times d} \\ 0_{d\times d} & 0_{d\times d} & \lambda_3 I_d \end{bmatrix}. \tag{22}$$

(a) (6 pts) **Compute $\nabla f_0(\vec{x})$.** *Show your work.*

(b) (7 pts) Recall the solution to the ridge regression problem given in class, i.e., $\vec{x}_{\mathrm{RR}} = (A^\top A + \lambda I)^{-1} A^\top \vec{y}$. **Give an expression for $\vec{x}_{\mathrm{BRR}}$ that is similar in structure to the expression for $\vec{x}_{\mathrm{RR}}$.** *Justify your answer.*

13. Low-Rank Matrix Completion (28 pts)

Consider a matrix $A \in \mathbb{R}^{m\times n}$. If some entries are corrupted, one principled way to identify A is to find the matrix $B \in \mathbb{R}^{m\times n}$ of minimal rank that agrees with A on all known entries. This can be formulated as an optimization problem whose objective function is $\mathrm{rank}(B)$. Because the $\mathrm{rank}(\cdot)$ function is not continuous, we use the intuition that a low-rank matrix will only have a few nonzero singular values, and instead use the sum-of-singular-values function as the objective:

$$f(B) \doteq \sum_{i=1}^{\mathrm{rank}(B)} \sigma_i\{B\} \tag{23}$$

where $\sigma_i\{B\}$ is the i^{th} largest singular value of B. In this problem we will explore some properties of f.

(a) (8 pts) **Prove that**

$$f(B) \le \max_{\substack{C\in\mathbb{R}^{m\times n} \\ \|C\|_2 \le 1}} \mathrm{tr}\left(C^\top B\right). \tag{24}$$

Here $\mathrm{tr}(\cdot)$ is the trace, which for a matrix $X \in \mathbb{R}^{m\times n}$ with entries X_{ij} is $\mathrm{tr}(X) = \sum_{i=1}^{\min\{m,n\}} X_{ii}$.

HINT: Expand B into its SVD. Try to find a $D \in \mathbb{R}^{m\times n}$ such that $\|D\|_2 = 1$ and $\mathrm{tr}\left(D^\top B\right) = f(B)$.

HINT: You may use the cyclic property of traces without proof. If XYZ and ZXY are valid matrix products then $\mathrm{tr}(XYZ) = \mathrm{tr}(ZXY)$.

Recall that we defined

$$f(B) \doteq \sum_{i=1}^{\mathrm{rank}(B)} \sigma_i\{B\}, \tag{23}$$

where $\sigma_i\{B\}$ is the i^{th} largest singular value of B.

(b) (9 pts) **Prove that**

$$f(B) \geq \max_{\substack{C \in \mathbb{R}^{m \times n} \\ \|C\|_2 \leq 1}} \mathrm{tr}\big(C^\top B\big). \tag{25}$$

Here $\mathrm{tr}(\cdot)$ is the trace, which for a matrix $X \in \mathbb{R}^{m \times n}$ with entries X_{ij} is $\mathrm{tr}(X) = \sum_{i=1}^{\min\{m,n\}} X_{ii}$.

HINT: Let $r \doteq \mathrm{rank}(B)$ and expand B into its outer product SVD, i.e., $B = \sum_{i=1}^{r} \sigma_i\{B\}\vec{u}_i\vec{v}_i^\top$.

HINT: You may use the cyclic and linearity properties of traces without proof. If XYZ and ZXY are valid matrix products then $\mathrm{tr}(XYZ) = \mathrm{tr}(ZXY)$. Also, $\mathrm{tr}(\alpha X + \beta Y) = \alpha\,\mathrm{tr}(X) + \beta\,\mathrm{tr}(Y)$ for $\alpha, \beta \in \mathbb{R}$.

From parts (a) and (b) together, we can conclude that

$$f(B) = \max_{\substack{C \in \mathbb{R}^{m \times n} \\ \|C\|_2 \leq 1}} \operatorname{tr}\left(C^\top B\right). \tag{26}$$

Here $\operatorname{tr}(\cdot)$ is the trace, which for a matrix $X \in \mathbb{R}^{m \times n}$ with entries X_{ij} is $\operatorname{tr}(X) = \sum_{i=1}^{\min\{m,n\}} X_{ii}$.

(c) (6 pts) **Show that for all $B_1, B_2 \in \mathbb{R}^{m \times n}$ we have**

$$f(B_1 + B_2) \leq f(B_1) + f(B_2) \tag{27}$$

i.e., the function f satisfies the triangle inequality.

HINT: Use the characterization of f given by (26). Also, you may use the linearity property of traces without proof, i.e., $\operatorname{tr}(\alpha X + \beta Y) = \alpha \operatorname{tr}(X) + \beta \operatorname{tr}(Y)$ for $\alpha, \beta \in \mathbb{R}$.

Recall that we defined the function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ by

$$f(B) = \sum_{i=1}^{\operatorname{rank}(B)} \sigma_i\{B\} = \max_{\substack{C\in\mathbb{R}^{m\times n} \\ \|C\|_2 \le 1}} \operatorname{tr}\left(C^\top B\right). \tag{23, 26}$$

You may assume from part (c) that we know that f satisfies the triangle inequality, i.e., for all $B_1, B_2 \in \mathbb{R}^{m\times n}$ we have

$$f(B_1 + B_2) \le f(B_1) + f(B_2). \tag{27}$$

(d) (5 pts) **Is f a convex function? If yes, prove it. If no, justify your answer using an example.**

HINT: You can use without proof the fact that $f(\alpha B) = |\alpha|\, f(B)$ for all $\alpha \in \mathbb{R}$ and $B \in \mathbb{R}^{m\times n}$.

[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.]

Midterm Instructions

Read the following instructions before the exam.

There are 13 problems of varying numbers of points. You have 120 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 26 pages on the exam, so there should be 13 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 ONE 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.