New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6763: Final Exam.

Friday, Dec. 22nd, 2023, 2:00 - 3:30pm
(90 minutes, 18 minutes per question)

Directions

- Show all of your work to receive full (and partial) credit.
- If more space is required, you may use extra sheets of paper, clearly marked with your name and the problem you are working on.
- There are 5 multipart questions worth **60 points total**.

1. Always, Sometimes, Never

(15 pts, 3 per question) Indicate whether each of the following statements is ALWAYS true, SOMETIMES true, or NEVER true. **Provide a short justification or example to explain your choice.**

(a) Given a matrix $\mathbf{V} \in \mathbb{R}^{n\times k}$ with orthonormal columns, for all $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{V}\mathbf{V}^T\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.

ALWAYS SOMETIMES NEVER

(b) The center-of-gravity method optimizes convex functions with fewer gradient oracle calls then gradient descent.

ALWAYS SOMETIMES NEVER

(c) If $\mathbf{A}$ has rank 1 then $\|\mathbf{A}\|_2 = \|\mathbf{A}\|_F$.

ALWAYS SOMETIMES NEVER

(d) Given two matrices $\mathbf{A}$ and $\mathbf{B}$ with top eigenvectors $\mathbf{a}_1$ and $\mathbf{b}_1$, if $\|\mathbf{A}-\mathbf{B}\|_2 \leq \epsilon$ then $\sin(\theta(\mathbf{a}_1, \mathbf{b}_1)) \leq \epsilon$.[1]

ALWAYS SOMETIMES NEVER

(e) Let $f_1(x), \ldots, f_n(x)$ be β-smooth convex functions and let $g(x) = \frac{1}{n}\sum_{i=1}^n f_i(x)$ be their average. g is β-smooth.

ALWAYS SOMETIMES NEVER

[1] Recall that $\theta(\mathbf{a}_1, \mathbf{b}_1)$ denotes the angle between the unit vectors $\mathbf{a}_1$ and $\mathbf{b}_1$.

2. Matrix Representations of Graphs and PSDness

(13 pts) Suppose we have an undirected graph G with n nodes. Let D denote the graph's diagonal degree matrix, let A denote its adjacency matrix, and let $L = D - A$ denote its Laplacian.

(a) (4pts) In class we proved that L is always positive semidefinite. Prove that the *normalized* Laplacian matrix, $D^{-1/2}LD^{-1/2}$ is also positive semidefinite for any G.

(b) (4pts) Consider the graph G' obtained by adding one new edge to G. Let L' be the Laplacian of the new graph. Prove that $L \preceq L'$.

(c) (5pts) Prove that, when when G has at least one edge and no self loops, A is *never* positive semidefinite.

3. ℓ_∞ Constrained Optimization Oracles

(12 pts) Consider the set $\mathcal{A} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_\infty \leq 1\}$, where $\|\mathbf{x}\|_\infty = \max_{i \in 1,\ldots d} |x_i|$.

(a) (3pts) Prove that $\mathcal{A}$ is convex.

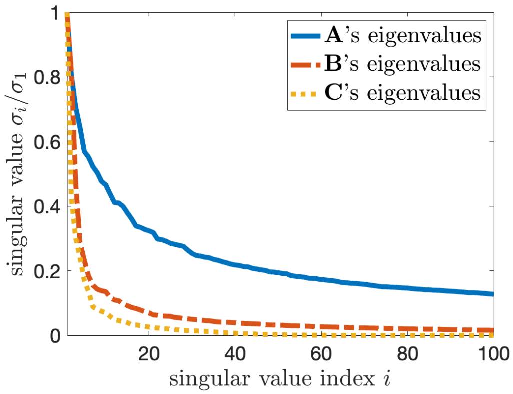
(b) (3pts) Write an equation or pseudocode to implement a projection oracle for $\mathcal{A}$. For any $\mathbf{x} \in \mathbb{R}^d$ your oracle should return the projection of $\mathbf{x}$ onto $\mathcal{A}$.

(c) (4pts) Write an equation or pseudocode to implement a separation oracle for $\mathcal{A}$. For any $\mathbf{x} \notin \mathcal{A}$, prove that your oracle returns a valid separating hyperplane.

(d) (2pts) Suppose f is a convex function and we want to minimize $f(x)$ subject to $x \in \mathcal{A}$. What is **one example** of an algorithm that requires a projection oracle to solve the minimization problem? What is **one example** of an algorithm that requires a separation oracle?

4. Matrix Spectra

(8 pts) Suppose we have 100×100 matrices $\mathbf{A}_1, \mathbf{A}_2$, and $\mathbf{A}_3$ with singular value matrices $\mathbf{\Sigma_A}$, $\mathbf{\Sigma_B}$, and $\mathbf{\Sigma_C}$, respectively. The diagonal values in these matrices are plotted and some specific numbers are shown below.



$$\mathbf{\Sigma_A} = [1, .81, .71, .64, \ldots, .21, .2]$$
$$\mathbf{\Sigma_B} = [1, .82, .47, .29, \ldots, .02, .01]$$
$$\mathbf{\Sigma_C} = [1, .42, .30, .24, \ldots, .001, .001]$$

(a) (2pts) Rank the matrices in order of which would be best approximated by a rank-40 approximation. Justify your ranking.

(b) (3pts) We want to find the top right singular vector of $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ using the **power method**. Rank the matrices in order of which one you expect the method to converge faster for. Justify your ranking.

(c) (3pts) We want to solve the linear systems $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2, \min_{\mathbf{x}} \|\mathbf{Bx} - \mathbf{y}\|_2$, and $\min_{\mathbf{x}} \|\mathbf{Cx} - \mathbf{y}\|_2$ using **gradient descent**. Rank the matrices in order of which one you expect the method to converge faster for. Justify your ranking.

Problem 5: Block Sparse Recovery

(12 pts) In addition to being sparse, many vectors in practice (images, Fourier transforms, etc.) exhibit *block structure*, where non-zeros are contained in contiguous blocks. We call a vector $\mathbf{x} \in \mathbb{R}^n$ (k, q)-*block sparse* if 1) $\mathbf{x}$ has at most kq non-zero entries and 2) all of those non-zeros are contained in at most q blocks of at most k consecutive indices. For example, the following two vectors are $(4, 1)$-block sparse:

$$\mathbf{x}_1 = [0, 0, \underbrace{1, -2, 3, .5,}_{\text{block 1}} 0, 0, 0] \qquad \mathbf{x}_2 = [\underbrace{.1, .3, -4, 2,}_{\text{block 1}} 0, 0, 0, 0, 0]$$

The following vectors are $(3, 2)$-block sparse:

$$\mathbf{x}_3 = [\underbrace{3, -1, 2,}_{\text{block 1}} 0, \underbrace{.1, -9, 1,}_{\text{block 2}} 0, 0] \qquad \mathbf{x}_2 = [0, \underbrace{-4, 2,}_{\text{block 1}} 0, 0, 0, \underbrace{.5, -6, 7}_{\text{block 2}}] \qquad \mathbf{x}_5 = [0, \underbrace{1, -1, 2,}_{\text{block 1}} \underbrace{5, 3,}_{\text{block 2}} 0, 0, 0, 0]$$

(a) (2 pts) We say a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ satisfies the (k, q, ϵ)-block-RIP property if for *all* (k, q)-block sparse vectors $\mathbf{x}$,

$$(1 - \epsilon)\|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \epsilon)\|\mathbf{x}\|_2^2.$$

Since any (k, q)-block sparse vector is kq sparse, a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ satisfies the (k, q, ϵ)-block-RIP property if it satisfies the (kq, ϵ)-RIP property. If $\mathbf{A}$ is a random Johnson-Lindenstrauss matrix, how many rows m does it need to satisfy the (kq, ϵ)-RIP property with probability $9/10$? **Your answer should be in big-Oh notation.**

(b) (6 pts) Improve on your bound above by proving that, with probability 9/10), a random Johnson-Lindenstrauss matrix with $O\left(\frac{kq + q\log(n/q)}{\epsilon^2}\right)$ rows satisfies the (k, q, ϵ)-block-RIP property. **Hint:** Use the Subspace Embedding theorem.

(c) (4 pts) Prove that if $\mathbf{A}$ satisfies the $(k, 2q, \epsilon)$-block RIP property for any $\epsilon < 1$, then $\mathbf{x}$ is the unique (k, q)-block sparse vector satisfying $\mathbf{b} = \mathbf{Ax}$. In other words, $\mathbf{x}$ can be recovered from the m linear measurements in $\mathbf{b}$.