New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6763: Final Exam Review.

Logistics

- **Time/place:** 2:00pm - 3:40pm on Tuesday 12/21 in our usual classroom. It will be a 1.5 hour exam.
- **Material allowed:** One double sided sheet of paper with whatever you want on it is allowed.

Concepts to Know

Random variables and concentration.

- This was the first section of the course, but the tools we learned never stopped being important. You will not be tested on them in isolation, but don't be surprised if problems require applying techniques like computing expectations/variances or random variables, applying the basic concentration bounds, applying union bounds, etc.

Optimization.

- Meaning of condition number for convex optimization problem. Meaning in the special case of $f(x) = \|Ax - b\|_2$.
- Stationary point definition.
- What is the center of gravity method? How many iterations does it take to minimize a convex function over a convex set? Why isn't it used in practice?
- Grunbaum's Theorem, and why it is relevant to analyzing the center of gravity method.
- Be able to recognize a linear program (linear constraints, linear objective).
- Definition of a separation oracle. Be able to describe a separation oracle for simple convex sets (e.g. for the unit ball).

Singular Value Decomposition

- Basic properties of matrices with orthonormal columns and orthonormal rows. When does multiplying by such a matrix V preserve the norm of a vector?
- Singular value decomposition.
- Relation to norms: $\|A\|_F^2 = \sum_{i=1}^{d} \sigma_i^2$. $\|A\|_2^2 = \max_{x:\|x\|_2=1} \|Ax\|_2^2 = \sigma_1^2$.
- Optimal low-rank approximation.
- What general properties of matrices make them low-rank?
- Power method and its analysis. What properties of the target matrix A control the convergence rate?
- Semantic embeddings from similarity measure.

Spectral Graph Theory

- Matrix representations of graphs (adjacency matrix A, Laplacian matrix L, edge-vertex incidence matrix B).
- Relationship between edge-vertex incidence matrix and Laplacian matrix.
- How to compute the value of cuts in a "linear algebraic way" using L or B.
- Intuitively what is the meaning of the smallest Laplacian eigenvector.
- Stochastic block model.
- Basic strategy for recovering the partition in a stochastic block model using a spectral method.
- Matrix concentration inequalities (you don't need to know the one stated in class, but know what it means to write down the expectation of a random matrix and bound its distance from that expectation).
- Davis-Kahan eigenvector perturbation theorem.

Randomized Numerical Linear Algebra

- What is a matrix sketch.
- Subspace embedding theorem.
- How subspace embeddings can be used to approximately solve least squares regression.
- ϵ-net arguments. Why was one needed for subspace embedding proof? Even if you can't reproduce the proof yourself, know the main components.
- Size of ϵ-net for the unit ball in $\mathbb{R}^d$.
- Definition of Randomized Hadamard matrices and how they are used to compute JL sketches faster.

Compressed Sensing + High Dimensional Geometry

- What is the compressed sensing problem and why is it important.
- Restricted Isometry Property, why it allows for sparse recovery.
- What matrices satisfy RIP? With what parameters? How many rows?
- Basis pursuit optimization problem for efficient recovery (you don't need to be able to reproduce the analysis).

Practice Problems

From *Foundations of Data Science* book (https://www.cs.cornell.edu/jeh/book.pdf). Note that this book uses $\|x\|$ to denote $\|x\|_2$ for a vector x.

- **Exercises:** 3.6, 3.7, 3.8, 3.10, 3.11, 3.12, 3.13, 3.18 (we showed this in class, make sure you can prove), 3.20, 3.21, 3.22, 3.26, 7.27, 12.31, 12.36
- **Exercises:** 12.33, 12.38 (try to figure out what the eigenvectors and eigenvalues of A are by hand)
- **Exercises:** 6.16. Understand the second equation on page 194: $\mathbb{E}[X] = AB$.
- **Exercises:** 2.2
    1. For $V \in \mathbb{R}^{n\times d}$ with orthonormal columns and vector $x \in \mathbb{R}^n$, when is $\|V^T x\|_2 = \|x\|_2$? Always, sometimes, or never?

2. Let A_k be the optimal k-rank approximation for A obtained via an SVD. prove that $\|A - A_k\|_2 = \sigma_{k+1}(A)$.
3. For any $V \in \mathbb{R}^{n \times d}$ with orthonormal columns, VV^T is the projection matrix onto the subspace spanned by the columns of V (V's column span). We used this fact many times when discussing low-rank approximation. Show that $VV^T = (VV^T)(VV^T)$. Why does this property make intuitive sense if VV^T is a projection?
4. Let A_k be the optimal k-rank approximation for A obtained via an SVD. Prove that $\|A - A_k\|_F^2 = \|A\|_F^2 - \|A_k\|_F^2$. Is the same statement true if we switch these to spectral norms $\|\cdot\|_2$?
5. Let $X \in R^{n \times 900}$ have random entries drawn independently as $\{0, 1\}$, each with probability $1/2$. Let $Y \in R^{n \times 900}$ have rows corresponding to 30×30 pixel black and white images of handwritten digits. All entries of Y are in $\{0, 1\}$. How do you expect $\sum_{i=11}^{900} \sigma_i(X)^2$ and $\sum_{i=11}^{900} \sigma_i(Y)^2$ to compare?
6. You have a data matrix $X \in \mathbb{R}^{n \times d}$ where each row corresponds to a student and each column corresponds to their grade on an assignment. The final column is their cumulative grade. What do you expect the rank of this matrix to be? Do you think it is well approximated by an even lower rank matrix?
7. Let $X \in \mathbb{R}^{n \times d}$ have SVD $U\Sigma V^T$ with singular values $\sigma_1(X), ...\sigma_d(X)$. What are the eigenvalues of $(X^T X)^q$ What are its eigenvectors. How about $(XX^T)^q$. What is the runtime required to apply either of these two matrices to a vector?
8. In the stochastic block model, why is clustering with the second largest eigenvector of the expected adjacency matrix equivalent to clustering with the second smallest eigenvector of the expected Laplacian? Are these two approaches identical when clustering using the actual rather than the expected matrices? Describe a natural variant of the stochastic block model where these two algorithms would not be equivalent even on the expected matrices.