COMPSCI 514: Final

Date: 12/19/2019, 10:30am-12:30pm (you may take the full 2 hours if desired).

Instructions:

- Please show your work/derive any answers as part of the solutions to receive full credit (and partial credit if you make a mistake).
- If you need extra space to show your work you can include additional pages. Clearly mark the top of the any additional page with your **name and problem number**. On the exam **indicate that the work is finished on an extra page.**
- If you have a question, raise your hand and we will come to you.
- If you need to use the restroom, **come to the front and turn in your cellphone before leaving**. You can pick up when you come back.

Optimization Warm Up (10 points)

1. (3 points) Consider two convex sets A and B. Prove that their intersection $A \cap B$ is also a convex set.

2. (3 points) Consider two convex functions $f(\vec{x})$ and $g(\vec{x})$. Prove that their sum $[f+g](\vec{x})$ is also a convex function.

3. (4 points) Consider optimizing the functions $f_1(x) = (x+1)^2$, $f_2(x) = x^2$, $f_3(x) = (x-1)^2$ in an online manner. What is the regret of the solution sequence $x^{(1)} = 0$, $x^{(2)} = 0$, $x^{(3)} = 1$?

Linear Algebra Warm Up (10 points)

1. (4 points) Indicate whether the following statements are *always*, *sometimes*, or *never* true. Justify your answers.

   (a) (2 points) When $\mathbf{V} \in \mathbb{R}^{n \times k}$ has orthonormal columns, for $\vec{x} \in \mathbb{R}^n$, $\|\mathbf{V}\mathbf{V}^T\vec{x}\|_2 \leq \|\vec{x}\|_2$.
   ALWAYS SOMETIMES NEVER

   (b) (2 points) Letting $\mathbf{U}_k \in \mathbb{R}^{n \times k}$ have columns equal to the top k left singular vectors of $\mathbf{X}$ and $\mathbf{V}_k \in \mathbb{R}^{d \times k}$ have columns equal to the top k right singular vectors of $\mathbf{X}$, $\mathbf{U}_k\mathbf{U}_k^T\mathbf{X} = \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T$. ALWAYS SOMETIMES NEVER

2. (4 points) Given $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\vec{y} \in \mathbb{R}^n$, consider the problem of projecting $\vec{y}$ onto the column span of $\mathbf{X}$.

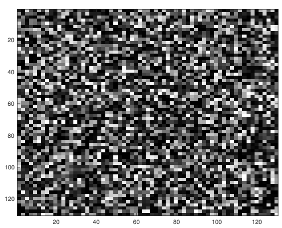
   (a) (2 points) What is this problem commonly known as in machine learning/data analysis?

   (b) (2 points) List three algorithms we have learned in class that can be used to solve this problem (or solve it approximately).

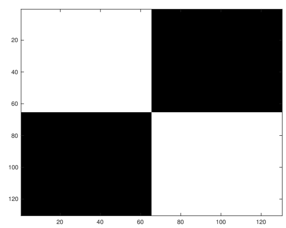
3. (2 points) $\mathbf{X} \in \mathbb{R}^{500\times 50}$ contains 500 well-clustered data points as its rows. In particular, there are ten cluster centers $\vec{y}_1, \ldots, \vec{y}_{10} \in \mathbb{R}^{50}$, such that each row $\vec{x}_i$ lies within Euclidean distance at most 1 of a center. Give an *upper bound* on $\min_{\mathbf{B}:\text{rank}(\mathbf{B})=10} \|\mathbf{X}-\mathbf{B}\|_F^2$.

Spectrum Matching (10 points)

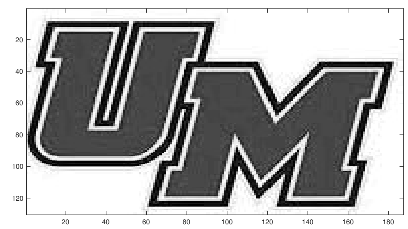
1. Consider the matrices (each with 130 rows) and the singular value spectrums pictured below.



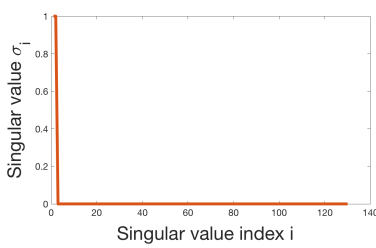
Matrix A



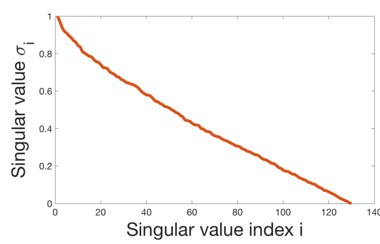
Matrix B



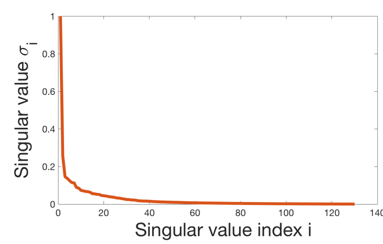
Matrix C



Spectrum 1



Spectrum 2

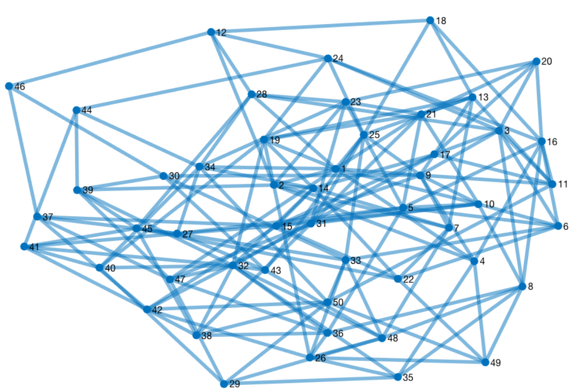


Spectrum 3

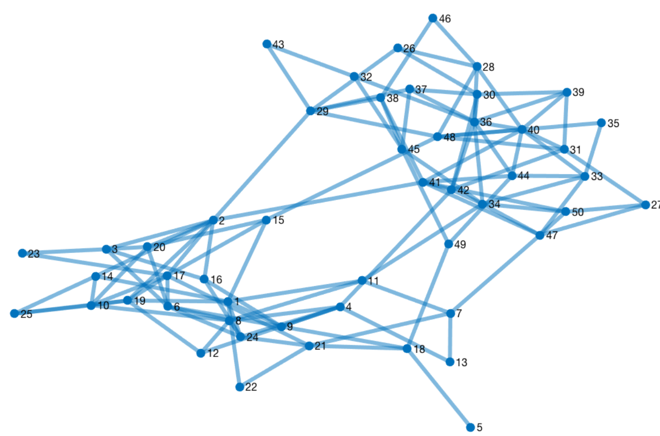
(a) (3 points) Match each of the matrices to its corresponding singular value spectrum. Explain in a few sentences why you picked the matches you did.

(b) (2 points) List the matrices in order of how well they would be approximated by a rank-20 approximation. Justify your answer.

2. (3 points) Which of the graphs pictured below has the lowest second smallest Laplacian eigenvalue? Give a sentence or two justifying your answer.



Graph A



Graph B

3. (2 points) Would partitioning Graph B above using its minimum cut give a useful separation of the nodes into two communities? Give a sentence or two justifying your answer.

The Power of Gradient Descent (10 points)

For any $\mathbf{X} \in \mathbb{R}^{n \times d}$ consider the optimization problem: $\vec{z}^\star = \underset{\vec{z} \in \mathbb{R}^d : \|\vec{z}\|_2 \leq 1}{\arg\max} f(\vec{z})$ where $f(\vec{z}) = \|\mathbf{X}\vec{z}\|_2^2$.

1. (2 points) What is the optimal solution $\vec{z}^\star$? What is the optimal value $\|\mathbf{X}\vec{z}^\star\|_2^2$?

2. (2 points) What is $\vec{\nabla} f(\vec{z})$?

3. (2 points) Prove that the constraint set $\mathcal{S} = \{\vec{z} \in \mathbb{R}^d : \|\vec{z}\|_2 \leq 1\}$ is convex. Give the projection function $P_{\mathcal{S}}(\vec{z})$ onto this set.

4. (1 point) What is the projected gradient descent update step for this problem, with step size η? **Hint 1:** We are *maximizing* $f(\vec{z})$, so the step should be in opposite direction than if we were minimizing. **Hint 2:** The update step will consist of two separate operations.

5. (3 points) In general, projected gradient descent is only guaranteed to converge when *minimizing* a convex function. Nevertheless, prove that for any $\eta > 0$, projected gradient descent initialized with a random start vector will converge to $\vec{z}^\star$ when applied to this maximization problem. **Hint:** You may use that for any matrix $\mathbf{A}$, the power method initialized with a random start vector and applied to $\mathbf{A}$ will converge to $\mathbf{A}$'s top right singular vector.

High Dimensional Space and Random Projection (EXTRA CREDIT: 8 points)

We learned that high-dimensional space behaves very differently than low-dimensional space. How then are dimensionality reduction results like the Johnson-Lindenstrauss lemma possible? In this problem we explore this question and see how the geometry of high-dimensional space can be used to prove bounds on the effectiveness of dimensionality reduction. Recall that the JL Lemma is:

Lemma 1 (Johnson-Lindenstrauss). *Let $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$ have each entry chosen i.i.d. as $\frac{1}{\sqrt{m}} \cdot \mathcal{N}(0,1)$. If we set $m = O\left(\frac{\log n}{\epsilon^2}\right)$ then for any set of n vectors $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$, with high probability, for all i, j: $(1-\epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2 \leq \|\mathbf{\Pi}\vec{x}_i - \mathbf{\Pi}\vec{x}_j\|_2^2 \leq (1+\epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2$.*

1. (2 points) Consider n unit vectors in d dimensions: $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ with $\|\vec{x}_i\|_2^2 = 1$ for all i. Consider random projection matrix $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$ with each entry chosen i.i.d. as $\frac{1}{\sqrt{m}} \cdot \mathcal{N}(0,1)$. Show that, for $m = O\left(\frac{\log n}{\epsilon^2}\right)$, with high probability, letting $\tilde{x}_i = \mathbf{\Pi}\vec{x}_i$, for all i:

$$(1-\epsilon) \leq \|\tilde{x}_i\|_2^2 \leq (1+\epsilon)$$

and for all i, j: $(1-\epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2^2 \leq (1+\epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2$.

Hint: Add a vector to the set $\vec{x}_1, \ldots, \vec{x}_n$ and apply the JL lemma.

2. (3 points) Let $\tilde{x}_i$ be given as above and let $\bar{x}_i = \tilde{x}_i \cdot \frac{1}{\|\tilde{x}_i\|_2}$, so $\bar{x}_i$ is a unit vector. Show that if $\vec{x}_1, \ldots, \vec{x}_n$ are nearly orthonormal, with $|\langle \tilde{x}_i, \tilde{x}_j \rangle| \leq \epsilon$ for all i, j, then with high probability, $\bar{x}_1, \ldots, \bar{x}_n$ are nearly orthonormal too, with $|\langle \bar{x}_i, \bar{x}_j \rangle| \leq c \cdot \epsilon$ for all i, j and some constant c. **Hint:** Consider the squared distance between each pair $\tilde{x}_i, \tilde{x}_j$ and each pair $\vec{x}_i, \vec{x}_j$. You may assume $\epsilon < 1$.

3. (3 points) In d dimensional space, there are at most $2^{O(\epsilon^2 d)}$ nearly orthonormal unit vectors with pairwise dot products all bounded by ϵ. Use this fact, along with part (2) to show that the JL lemma cannot be improved up to constants: to preserve the pairwise distances between n points, we must project into at least $\Omega\left(\frac{\log n}{\epsilon^2}\right)$ dimensions. **Hint:** You can solve this problem using the conclusion of part (2), even if you did not solve that part.