COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Spring 2026.

Lecture 19

Logistics

- Midterm 2 is this Thursday in class, 1-2:15pm.
- Closed book – no cheatsheets, calculators, or other aids allowed.
- Study guide is posted on the course webpage (under the Midterm 2 row in the schedule tab).
- Past finals, which cover mostly the same material as this midterm, are posted in Canvas.
- Ignore content on optimization, which we have not covered yet.
  convexity, gradient, Lipschitz
- Solutions to study guide questions also posted in Canvas.
- I will hold regular office hours today after class, focused on midterm review.

Studying

Suggested Studying Approach:

- Review the study guide to get a sense of what you need to know, and then mostly focus on doing practice questions from the past finals and the study guide.
- Review slides as needed.
- Do some practice exams, under time constraints, with no material in front of you.
- Note that some of the past exams were 2 hours long, but designed to be completed in 90 minutes. This exam will be a bit shorter.

Midterm Format

Rough Outline: (subject to changes)

- Question 1: 4-5 True/False questions. No justification needed.
- Question 2: 4-5 numerical answers, like quiz questions. No justification needed.
- Question 3: 4-5 part question on analyzing an algorithm. Similar in style to but easier than a homework question.
- Question 4: More challenging 4-5 part question on analyzing an algorithm – more similar to a homework question.
- Potentially some extra credit subquestions on Q3/Q4.

Questions

Content or Format Questions?

Questions

Questions

Questions

Spectral Graph Theory

Quiz Review

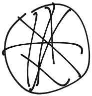
$\boxed{\lambda_{n-1}(\text{clique}) \geq \lambda_{n-1}(\text{cycle})}$

2 Multiple Choice 1 point

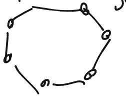
Let graph G consist of a clique on 10 nodes along with a cycle on 10 different nodes. There are no other edges. Let $L \in \mathbb{R}^{20\times 20}$ be the Laplacian matrix of G. What is the minimum eigenvalue of L?

- 0
- 10
- 20
- -10
- -100

10-node



10-node cycle



$v^T L v \geq 0 \ \forall v$

$= \lambda$ when v is an eigenvector

$[1\ 1\ 1\ 1]^T L \begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix} = 0$

$v^T L v = \sum_{(i,j)\in E} [v(i) - v(j)]^2$

$\lambda_n(L) = 0$

$Lv = (D-A)v$

$= Dv - Av$

$= \begin{bmatrix}d_1\\ d_2\\ \vdots\\ d_n\end{bmatrix} - \begin{bmatrix}d_1\\ \vdots\\ d_n\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}$

$\lambda_{n-1}(L)$

$v_{n-1}^T \mathbb{1} = 0$

$\sum_{i=1}^{20} v_{n-1}(i) = 0$

$v_{n-1} = \begin{bmatrix}1\\ \vdots\\ 1\\ -1\\ \vdots\\ -1\end{bmatrix} \begin{matrix}\}\ 10 \text{ clique nodes}\\ \}\ 10 \text{ cycle nodes}\end{matrix}$

$v_{n-1}^T L v_{n-1} = \sum_{(i,j)\in E} (v(i) - v(j))^2 = 0$

$v^T L v = v^T(\lambda v)$

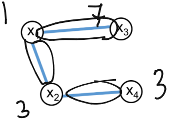
$= \lambda \|v\|_2^2$

$= \lambda$ if $\|v\|_2^2 = 1$

Quiz Review

3 Formula 1 point

Consider the unweighted graph G shown below and let L be its graph Laplacian. Let $\vec{x} = [1, 3, 7, 3]$. What is $\vec{x}^T L\vec{x}$?



$$(7-1)^2 + (3-1)^2 + (5-3)^2 = \underline{40}$$

Hint: You don't need to explicitly write down L or do any linear algebra.

Answer

$$V_{n-1} = \arg\min_{v} v^T L v \quad \text{s.t. } V_{n-1}^T 1 = 0, \ \|v\|_2 = 1$$

$$c = \arg\min c^T L c \quad \text{s.t. } c^T 1 = 0, \ c \in \{-1, 1\}$$

Quiz Review

4 Multiple Choice 1 point

$\|z\|_2 = 1 \quad z^T \mathbb{1} = 0$

Consider solving the optimization problem: $\min_{\{z\in\{-1,1\}^n:\ \text{not all entries of z are equal}\}} z^T L z.$

What is this optimization problem commonly known as?

- Computing the lowest eigenvalue of the graph Laplacian.
- ~~Computing~~ the second lowest eigenvalue of the graph Laplacian.
- Computing the minimum cut.
- ~~Computing the smallest node degree.~~
- Computing the maximum eigenvalue of the graph Laplacian.

$z \in \{-1,1\}^n \qquad z^T L z = 4 \cdot \text{CUT}$

$z^T \mathbb{1} = 0$

↳ balance

Quiz Review

5 Multiple Answer 1 point

Which of the following vectors can be used to recover the two communities in the stochastic block model? Check all that apply.

- [ ] Highest eigenvector of the adjacency matrix **A.**
- [ ] Second highest eigenvector of the adjacency matrix **A.**
- [ ] Second highest eigenvector of the Laplacian matrix **L.**
- [ ] Lowest eigenvector of the Laplacian matrix **L.**
- [ ] Second lowest eigenvector of the Laplacian matrix **L.**

More Spectral Graph Theory Review

$L = D - A = (n-1)I - A = (n-1)I - (J - I) = nI - J$

3. Let G be a fully connected graph (a complete graph), with self-loops. What are the eigenvalues of its corresponding adjacency matrix $\mathbf{A}$ and Laplacian $\mathbf{L}$? What if there are no self-loops?

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} n \\ n \\ n \\ n \\ n \end{pmatrix}$$

$\lambda_1 = n$, $\vec{v}_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

$\lambda_2 = 0$, v_2

$\vdots = 0$, $\vdots$

$\lambda_n = 0$, v_n $\Big\}$ any basis for the null space

$L = D - A$
$= nI - A$

$$tr(A) = \sum_{i=1}^{n} A_{ii} = \sum_{i=1}^{n} \lambda_i$$

$= n = n + ? + ? + ?$

null(A) has dimension $n - rank(A) = n - 1$

$\{x : Ax = 0\}$

$A = \begin{bmatrix} 0 & & 1 \\ & 0 & \\ 1 & & 0 \end{bmatrix}$ $\lambda_1 = n-1$, $\lambda_2 = -1$, $\lambda_n = -1$

$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$

$(J - I)v = Av = Jv - v$
$= \lambda v - v$
$(\lambda - 1)v$

More Spectral Graph Theory Review

- -

1. Consider a graph G with Laplacian matrix $\mathbf{L}$. Consider the problem: $x_* = \arg\min_{x:\|x\|=1} x^T \mathbf{L} x$. What is x_*? What value of $x_*^T \mathbf{L} x_*$ does it achieve?

General Linear Algebra

Views of Matrix Multiplication

(column combination)

$$\begin{bmatrix} & A & \end{bmatrix} \overset{V}{\begin{bmatrix} v(1) \\ \vdots \\ v(n) \end{bmatrix}}$$

$$A_1 \cdot v(1) + A_2 \cdot v(2) + \dots A_n v(n)$$

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} = \begin{bmatrix} | & | & \\ d_1 A_1 & d_2 A_2 & \dots \\ | & | & \end{bmatrix}$$

(row inner products)

$$\begin{bmatrix} [\ a_1^T\ ] \\ \vdots\ A \\ [\ a_n^T\ ] \end{bmatrix} \begin{bmatrix} v(1) \\ \vdots \\ v(n) \end{bmatrix} = \begin{bmatrix} a_1^T v \\ a_2^T v \\ \vdots \\ a_n^T v \end{bmatrix}$$

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a_1^T 1 \\ \vdots \\ a_n^T 1 \end{bmatrix}$$

Views of Matrix Multiplication

$n \times d \quad n \times p \quad p \times d$

Let $X = AB$. Fill in the blank:

X's columns are spanned by ... A's columns

$$X_i = \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} B_i \end{bmatrix}$$

$\mathrm{col}(X) \subseteq \mathrm{col}(A) \qquad \mathrm{rank}(X) \leq \mathrm{rank}(A)$

X's rows are spanned by ... rows B

$$\begin{bmatrix} x_i^T \end{bmatrix} = \begin{bmatrix} a_i^T \end{bmatrix} \begin{bmatrix} B \end{bmatrix}$$

$$X = \begin{array}{c} \\ n \end{array} \begin{array}{c} k \\ \begin{bmatrix} C \end{bmatrix} \end{array} \begin{array}{c} d \\ \begin{bmatrix} V^T \end{bmatrix} \end{array} k$$

$\mathrm{row}(X) \subseteq \mathrm{row}(B) \qquad \mathrm{rank}(X) \leq \mathrm{rank}(B)$

Row Span/Column Span

Let $\mathbf{X} \in \mathbb{R}^{n \times n}$ be a symmetric matrix with $\text{rank}(\mathbf{X}) = k$.

What is the dimension of $\mathbf{X}$'s row span?

k

$X = CV^T \leftarrow$ basis for row span

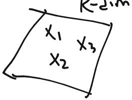
What is the dimension of $\mathbf{X}$'s column span?

k

What is the dimension of $\mathbf{X}$'s null space?

$n-k$

$$\begin{bmatrix} X \end{bmatrix}\begin{bmatrix} z \end{bmatrix} = 0 = \begin{bmatrix} x_1^T z \\ \vdots \\ x_n^T z \end{bmatrix}$$



How many of $\mathbf{X}$'s eigenvalues are equal to 0?

$\underline{n-k}$

Let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the SVD of $\mathbf{X}$. Prove that the i^{th} column of $\mathbf{V}$, $\mathbf{v}_i$, is an eigenvector of $\mathbf{X}^T\mathbf{X}$ with eigenvalue σ_i^2.

$X = V \Lambda V^T$

$X \in \mathbb{R}^{n \times d}$

$\begin{bmatrix} U \end{bmatrix} \begin{bmatrix} \Sigma \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}$

V has ortho columns

orthonormal cols

diag

$\boxed{\begin{matrix} X^T X \\ X X^T \end{matrix}}$

$X^T X = \underbrace{(U \Sigma V^T)^T}_{X^T} \underbrace{(U \Sigma V^T)}_{X}$

$V \Sigma U^T U \Sigma V^T$

$X^T X = \begin{bmatrix} V \Sigma^2 V^T \end{bmatrix} \begin{bmatrix} v_i \end{bmatrix}$

$\boxed{\begin{matrix} (u_i, \sigma_i^2) \\ X X^T \end{matrix}}$

$\begin{bmatrix} v_1^T \\ \vdots \\ v_d \end{bmatrix} \begin{bmatrix} v_i \end{bmatrix}$

$V \Sigma^2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$

$V \begin{bmatrix} 0 \\ 0 \\ \sigma_i^2 \\ 0 \\ 0 \end{bmatrix} = \sigma_i^2 \cdot v_i$

SVD And Eigendecomposition

Let $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ be the SVD of $\mathbf{X}$. Prove that the i^{th} column of $\mathbf{U}$, $\mathbf{u}_i$, is an eigenvector of $\mathbf{X}\mathbf{X}^T$ with eigenvalue σ_i^2.

Low Rank Approximation Example

5. $\mathbf{X} \in \mathbb{R}^{500 \times 50}$ 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 $\min_{\mathbf{B}:\text{rank}(\mathbf{B})=10} \|\mathbf{X} - \mathbf{B}\|_F^2$.

6. Consider two matrices $\mathbf{A} = \begin{bmatrix} 1.01 & 0 \\ 0 & 1 \end{bmatrix}$ or $\mathbf{B} = \begin{bmatrix} 1.1 & 0 \\ 0 & 1 \end{bmatrix}$.

   (a) What are their eigenvalues and eigenvectors?

   (b) On which matrix will power method converge more quickly?

3. Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ have SVD $\mathbf{X} = \mathbf{U\Sigma V}^T$ with singular values $\sigma_1(\mathbf{X}), \ldots, \sigma_d(\mathbf{X})$.

   (a) What are the eigenvalues of the matrix $(\mathbf{X}^T\mathbf{X})^2 + (\mathbf{X}^T\mathbf{X})^3$? What are its eigenvectors? How about the matrix $(\mathbf{X}\mathbf{X}^T)^2 + (\mathbf{X}\mathbf{X}^T)^3$?

   (b) What is the runtime required to compute $\left[(\mathbf{X}^T\mathbf{X})^2 + (\mathbf{X}^T\mathbf{X})^3\right]\vec{v}$ for any $\vec{v} \in \mathbb{R}^d$.

   (c) Name one method discussed in class which relies on efficiently applying a polynomial in $\mathbf{X}^T\mathbf{X}$ to a vector (or more generally, applying a polynomial in a matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ to a vector).

Johnson-Lindenstrauss

Low Distortion Embedding Definition and JL Statement

$x_1 \ldots x_n \in \mathbb{R}^d \to \tilde{x}_1, \ldots \tilde{x}_n \in \mathbb{R}^m \quad m << d$

$(1-\varepsilon)\|x_i - x_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1+\varepsilon)\|x_i - x_j\|_2$

$\|X - XVV^T\|_F$

<u>JL Lemma:</u> always LDE into

$m = O\left(\frac{\log(n)}{\varepsilon^2}\right)$ dimensions

(random liner map works)

Let $\boldsymbol{\pi} \in \{-1, 1\}^n$ have uniform random ± 1 entries. Let $\mathbf{y} \in \mathbb{R}^n$ be an arbitrary vector. What is

$$\mathbb{E}[(\boldsymbol{\pi}^T \mathbf{y})^2]? = Var(\pi^T y)$$

$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$

$\boxed{\mathbb{E}\|\Pi y\|_2^2 = \|y\|_2^2}$

$$\begin{array}{c} \pi \\ \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix} \end{array} \begin{bmatrix} y \end{bmatrix}$$

$$\mathbb{E}[(\pi^T y)] = \mathbb{E} \sum_{i=1}^{d} \pi(i)\, y(i) = \sum_{i=1}^{d} \mathbb{E}\, \pi(i)\, y(i)$$

$$= 0$$

$$Var(\pi^T y) = Var\left(\sum_{i=1}^{d} \pi(i)\, y(i)\right) = \sum_{i=1}^{d} Var(\pi(i) y(i))$$

$$= \sum_{i=1}^{d} y(i)^2 \cdot Var(\pi(i))$$

$$= \sum_{i=1}^{d} y(i)^2 = \|y\|_2^2$$

$\tilde{x}_i = \Pi x_i$

$\forall y \quad \|\Pi y\|_2 \approx \|y\|_2$ distributional JL

set $y_{ij} = x_i - x_j$