COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.
Lecture 17

Logistics

- Problem Set 4 is due 4/22.
- Project progress report is due 4/16.
- We have no class on Tuesday – so the weekly quiz is due Wednesday night.

Summary

Last Class: Subspace embedding via sampling.

- Subspace embedding via sampling.
- The matrix leverage scores.
- Analysis via matrix concentration bounds.

Today:

- Intuition behind leverage scores
- Connection to effective resistances and spectral graph sparsifiers.

Subpace Embedding via Sampling

Theorem (Subspace Embedding via Leverage Score Sampling)

For any $A \in \mathbb{R}^{n \times d}$ with left singular vector matrix U, let $\tau_i = \|U_{i,:}\|_2^2$ and $p_i = \frac{\tau_i}{\sum \tau_i}$. Let $\mathbf{S} \in \mathbb{R}^{m \times n}$ have $\mathbf{S}_{:,j}$ independently set to $\frac{1}{\sqrt{m p_i}} \cdot e_i^T$ with probability p_i.

Then, if $m = O\left(\frac{d \log(d/\delta)}{\epsilon^2}\right)$, with probability $\geq 1 - \delta$, $\mathbf{S}$ is an ϵ-subspace embedding for A.

- Matches oblivious random projection up to the $\log d$ factor.
- Can sample according to the row norms of any orthonormal basis for $col(A)$.

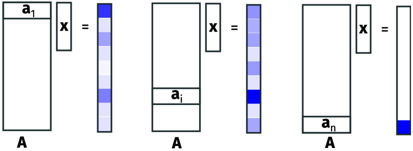
Leverage Score Intuition

Variational Characterization of Leverage Scores

For a matrix $A \in \mathbb{R}^{n \times d}$ with SVD $A = U\Sigma V^T$, the i^{th} leverage score is given by $\tau_i(A) = \|U_{i,:}\|_2^2$. Consider the maximization problem:

$$\tau_i(A) = \max_{x \in \mathbb{R}^d} \frac{[Ax](i)^2}{\|Ax\|_2^2}.$$

How much can a vector in A's column span 'spike' at position i.



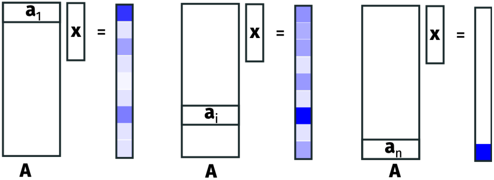
Can rewrite this problem as:

$$\max_{z:\|z\|_2=1} \frac{[Uz](i)^2}{\|Uz\|_2^2} = [Uz](i)^2.$$

What z maximizes this value?

Variational Characterization of Leverage Scores

$$\tau_i(A) = \max_{x \in \mathbb{R}^d} \frac{[Ax](i)^2}{\|Ax\|_2^2}.$$

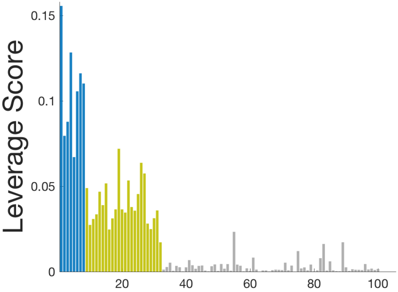


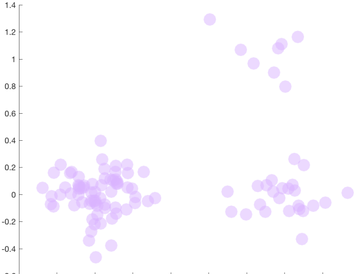
- Remember that we want $\|\mathbf{S}Ax\|_2^2 \approx \|Ax\|_2^2$ for all $x \in \mathbb{R}^d$.
- The leverage scores ensure that we sample each entry of Ax with high enough probability to well approximate $\|Ax\|_2^2$.
- In fact, could prove the subspace embedding theorem by showing that for a fixed $x \in \mathbb{R}^d$, $\|\mathbf{S}Ax\|_2^2 \approx \|Ax\|_2^2$, and then applying a net argument + union bound. Athough you would lose a factor d over the optimal bound.

Leverage Score Intuition

- When a_i is not spanned by the other rows of A, $\tau_i(A) = 1$.
- $\tau_i(A)$ is small when many rows are similar to a_i.

Leverage Score Intuition

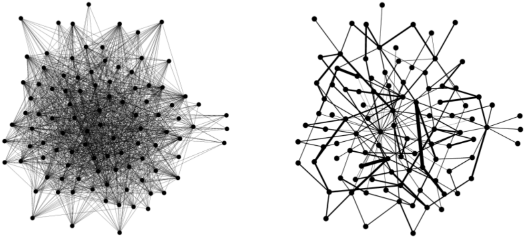




Spectral Graph Sparsification

Graph Sparsification

Given a graph $G = (V, E)$, find a (weighted) subgraph G' with many fewer edges that approximates various properties of G.[1]



Cut Sparsifier: (Karger) For any set of nodes S,

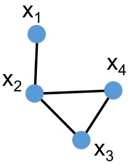
$$CUT'(S, V \setminus S) \approx_\epsilon CUT(S, V \setminus S).$$

---

[1] Image taken from Nick Harvey's notes `https://www.cs.ubc.ca/~nickhar/W15/Lecture11Notes.pdf`.

The Graph Laplacian

For a graph with adjacency matrix $A \in \{0,1\}^{n\times n}$ and diagonal degree matrix $D \in \mathbb{R}^{n\times n}$, $L = D - A$ is the graph Laplacian.



$$\overset{\mathbf{D}}{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}} - \overset{\mathbf{A}}{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}} = \overset{\mathbf{L}}{\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}}$$

L can be written as $L = \sum_{(u,v)\in E} L_{u,v}$ where $L_{u,v}$ is an 'edge Laplacian'

$$\overset{\mathbf{L}}{\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}} = \overset{\mathbf{L}_{1,2}}{\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}} + \overset{\mathbf{L}_{2,4}}{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}} + \cdots$$

Laplacian Smoothness

Observation 1: For any $z \in \mathbb{R}^d$,

$$z^T L z = \sum_{(u,v)\in E} z^T L_{u,v} z = \sum_{(u,v)\in E} (z(i) - z(j))^2.$$

$$\begin{bmatrix} v(1) & v(2) & v(3) & v(4) \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v(1) \\ v(2) \\ v(3) \\ v(4) \end{bmatrix}$$

- $z^T L z$ measures how smoothly z varies across the graph.
- If $z \in \{-1, 1\}^n$ is a cut indicator vector with $z(i) = 1$ for $i \in S$ and $z(i) = -1$ otherwise, then $z^T L z = 4 \cdot CUT(S, V \setminus S)$.
- So G' with (weighted) Laplacian $L' \approx_\epsilon L$ will be a cut sparsifier, with $CUT'(S, V \setminus S) \approx_\epsilon CUT(S, V \setminus S)$ for all S.
- Such a G' is called an ϵ-spectral sparsifier of G.

Laplacian Factorization

Observation 2: $L_{u,v} = b_{u,v} b_{u,v}^T$. So $L = \sum_{(u,v)\in E} b_{u,v} b_{u,v}^T$.

$$\underset{\mathbf{L_{2,4}}}{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}} = \underset{\mathbf{b_{2,4}}}{\begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}} \underset{\mathbf{b_{2,4}}^T}{\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}}$$

$$\begin{matrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{matrix}$$

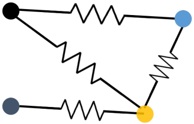
vertex-edge
incidence matrix B

That is, letting $B \in \mathbb{R}^{m\times n}$ have rows $\{b_{u,v}^T : (u,v) \in E\}$, $L = B^T B$.

- So if a sampling matrix $\mathbf{S}$ is a subspace embedding for B, then $B^T \mathbf{S}^T \mathbf{S} B \approx_\epsilon B^T B \approx_\epsilon L$. I.e., $\mathbf{S}B$ is the weighted vertex-edge incidence matrix of an ϵ-spectral sparsifier of G.
- By our results on subspace embedding, every graph G has an ϵ-spectral sparsifier with just $O(n \log n/\epsilon^2)$ edges.

Leverage Scores and Effective Resistance

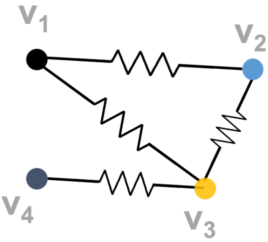
A spectral sparsifier G' of G with $O(n \log n/\epsilon^2)$ edges can be constructed by sampling rows of the vertex-edge incidence matrix via their leverage scores. What are these leverage scores?



- View each edge as a 1-Ohm resistor.
- If we fix a current of 1 between u, v, the voltage drop across the nodes is known as the effective resistance between u and v.
- We will show that the leverage score of each edge is exactly equal to its effective resistance.
- Intuitively, to form a spectral sparsifier, we should sample high resistance edges with high probability, since they are 'bottlenecks'.

Electrical Flows

For a flow $f \in \mathbb{R}^m$, the currents going into each node are given by $B^T f$.



$$\underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}}_{\mathbf{B^T}} \underbrace{\begin{pmatrix} 3 \\ 0 \\ -1 \\ -1 \end{pmatrix}}_{\mathbf{f}} = \begin{pmatrix} 3 \\ -4 \\ 0 \\ 1 \end{pmatrix}$$

The electrical flow when one unit of current is sent from u to v is:

$$f^e = \underset{f : B^T f = b_{u,v}}{\arg\min} \|f\|_2 .$$

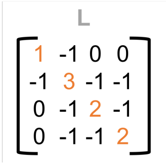
Since power (energy/time) is given by $P = I^2 \cdot R$.

Leverage Scores and Effective Resistance

$$f^e = \underset{f: B^T f = b_{u,v}}{\arg\min} \|f\|_2.$$

By Ohm's law, the voltage drop across (u, v) (i.e., the effective resistance) is simply the entry $f^e_{u,v}$ (since u, v is a unit resistor).

- To solve for f, note that we can assume that f is in the column span of B. Otherwise, it would not have minimal norm. So $f = B\phi$ for some vector $\phi \in \mathbb{R}^n$.
- Then need to solve $B^T B\phi = b_{u,v}$. I.e., $L\phi = b_{u,v}$. ϕ is unique up to its component in the null-space of L.



L

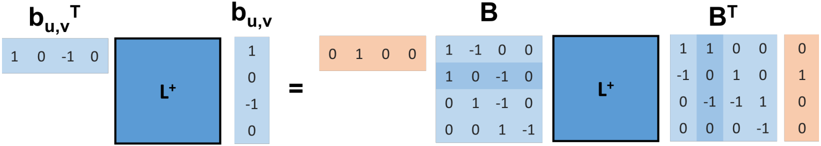
$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}$$

- $\phi = L^+ b_{u,v}$.

Leverage Scores and Effective Resistance

The effective resistance across edge (u, v) is given by

$$b_{u,v}(B^T B)^+ b_{u,v} = e_{u,v}^T B (B^T B)^+ B^T e_{u,v}.$$



$$\overset{\mathbf{b_{u,v}^T}}{\begin{bmatrix}1 & 0 & -1 & 0\end{bmatrix}} \; \mathbf{L^+} \; \overset{\mathbf{b_{u,v}}}{\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}} = \begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix} \overset{\mathbf{B}}{\begin{bmatrix}1 & -1 & 0 & 0\\1 & 0 & -1 & 0\\0 & 1 & -1 & 0\\0 & 0 & 1 & -1\end{bmatrix}} \; \mathbf{L^+} \; \overset{\mathbf{B^T}}{\begin{bmatrix}1 & 1 & 0 & 0\\-1 & 0 & 1 & 0\\0 & -1 & -1 & 1\\0 & 0 & 0 & -1\end{bmatrix}} \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

Write $B = U\Sigma V^T$ in its SVD.

$$\begin{aligned} e_{u,v}^T B (B^T B)^+ B^T e_{u,v} &= e_{u,v}^T U \Sigma V^T (V \Sigma^{-2} V^T) V \Sigma U^T e_{u,v} \\ &= e_{u,v}^T U U^T e_{u,v} \\ &= U_{u,v}^T U_{u,v} = \|U_{u,v}\|_2^2. \end{aligned}$$

I.e., the effective resistance is exactly the leverage score of the corresponding row in B.

Some History

- The concept of spectral sparsification was first introduced by Spielman and Teng '04 in their seminal work on fast system solvers for graph Laplacians. In this work, sparsifiers are used as preconditioners (like in Problem Set 3).
- Spielman and Srivastava '08 showed how to construct sparsifiers with $O(n \log n/\epsilon^2)$ edges via effective resistance (leverage score) sampling.
- Batson, Spielman, and Srivastava '08 showed how to achieve $O(n/\epsilon^2)$ edges with a deterministic algorithm.
- Marcus, Spielman, and Srivastava '13 built on this work to give optimal bipartite expanders with any degree and to resolve the famous Kadison-Singer problem in functional analysis.