Sublinear Time Eigenvalue Approximation via Random Sampling

Cameron Musco

University of Massachusetts Amherst

Joint with: Rajarshi Bhattacharjee (UMass), Gregory Dexter (Purdue), Petros Drineas (Purdue), Archan Ray (UMass)

Eigenvalue Approximation

Basic linear algebraic primitive: Given symmetric $A \in \mathbb{R}^{n \times n}$, compute approximations to all of A's eigenvalues.

- Nearly exact computation of all eigenvalues in $O(n^\omega)$ time via full eigendecomposition/Schur decomposition — but this is prohibitive for large n.
- Accurate approximation to k largest magnitude eigenvalues using $\tilde{O}(k)$ matrix vector multiplications with A (power method, Krylov subspace methods, `eigs`). $\tilde{O}(n^2 \cdot k)$ time for dense matrices.

Eigenvalue Approximation

Basic linear algebraic primitive: Given symmetric $A \in \mathbb{R}^{n \times n}$, compute approximations to all of A's eigenvalues.

- Nearly exact computation of all eigenvalues in $O(n^\omega)$ time via full eigendecomposition/Schur decomposition — but this is prohibitive for large n.
- Accurate approximation to k largest magnitude eigenvalues using $\tilde{O}(k)$ matrix vector multiplications with A (power method, Krylov subspace methods, `eigs`). $\tilde{O}(n^2 \cdot k)$ time for dense matrices.

How well can we approximate the spectrum in sublinear time, i.e., $o(n^2)$ time for dense matrices?

Eigenvalue Approximation

Basic linear algebraic primitive: Given symmetric $A \in \mathbb{R}^{n \times n}$, compute approximations to all of A's eigenvalues.

- Nearly exact computation of all eigenvalues in $O(n^\omega)$ time via full eigendecomposition/Schur decomposition — but this is prohibitive for large n.
- Accurate approximation to k largest magnitude eigenvalues using $\tilde{O}(k)$ matrix vector multiplications with A (power method, Krylov subspace methods, `eigs`). $\tilde{O}(n^2 \cdot k)$ time for dense matrices.

How well can we approximate the spectrum in sublinear time, i.e., $o(n^2)$ time for dense matrices?

Need some assumptions — otherwise for a single pair (i, j), A_{ij} and A_{ji} can be arbitrarily large and dominate the top eigenvalues. Finding this single pair takes $\Omega(n^2)$ time.

Summary

Today:

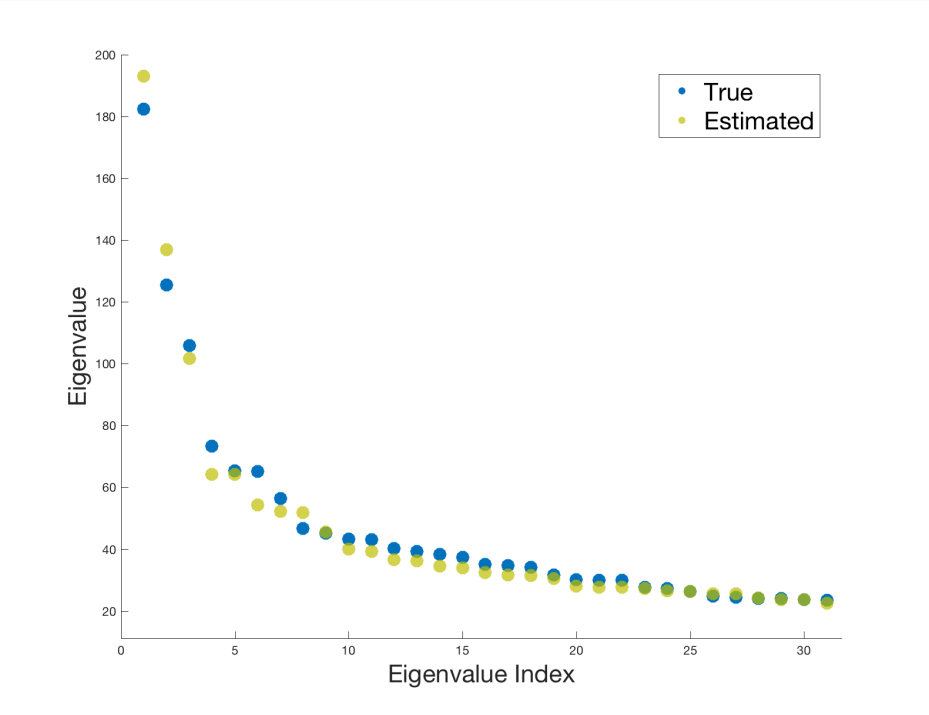
- Very simple sublinear time algorithm for approximating all eigenvalues of any symmetric bounded entry matrix.
- Just sample a uniform random principal submatrix and computes its eigenvalues.
- Improved algorithm for sparse matrices when you can sample rows/columns with probabilities proportional to their sparsity.
- Lots of open questions for bounded entry matrix computations.

Our Main Result

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with entries bounded in magnitude by 1, and eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

Main Result: There is an algorithm that reads $O\left(\frac{\log^6 n}{\epsilon^6}\right)$ entries of A and outputs $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that, for all $i \in [n]$,

$$|\lambda_i - \tilde{\lambda}_i| \leq \epsilon \cdot n.$$



Some Remarks

How good are $\pm\epsilon n$ additive error approximations to each of A's eigenvalues?

- $|\lambda_i| \leq \|A\|_F \leq n$ for all i.
- $\sum \lambda_i^2 = \|A\|_F^2 \leq n^2$. So there are at most $1/\epsilon^2$ outlying eigenvalues with $|\lambda_i| \geq \epsilon \cdot n$.
- These are the only eigenvalues for which we give a non-trivial approximation.

Some Remarks

How good are $\pm\epsilon n$ additive error approximations to each of A's eigenvalues?

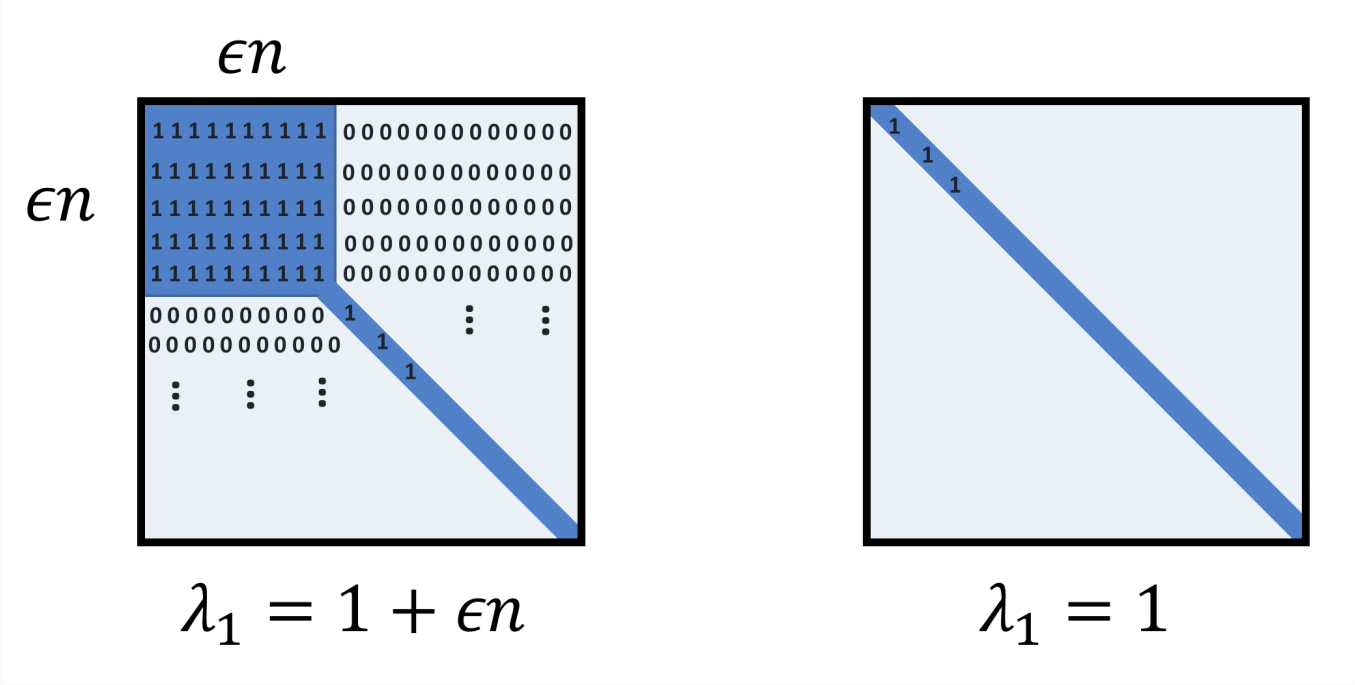
- $|\lambda_i| \leq \|A\|_F \leq n$ for all i.
- $\sum \lambda_i^2 = \|A\|_F^2 \leq n^2$. So there are at most $1/\epsilon^2$ outlying eigenvalues with $|\lambda_i| \geq \epsilon \cdot n$.
- These are the only eigenvalues for which we give a non-trivial approximation.
- It is easy to see that additive error scaling linearly in n is necessary.

Some Remarks

How good are $\pm\epsilon n$ additive error approximations to each of A's eigenvalues?

- $|\lambda_i| \leq \|A\|_F \leq n$ for all i.
- $\sum \lambda_i^2 = \|A\|_F^2 \leq n^2$. So there are at most $1/\epsilon^2$ outlying eigenvalues with $|\lambda_i| \geq \epsilon \cdot n$.
- These are the only eigenvalues for which we give a non-trivial approximation.
- It is easy to see that additive error scaling linearly in n is necessary.
- Could equivalently remove the bounded entry assumption, and obtain additive error $\epsilon \cdot n \cdot \|A\|_\infty$.

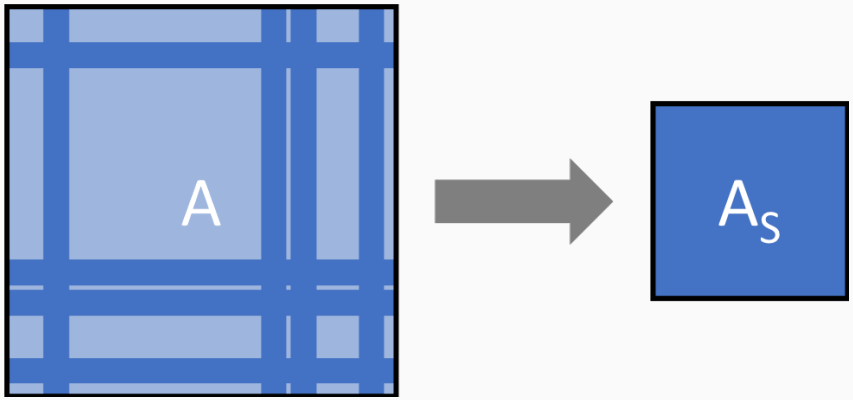
Lower Bound Instance



Only $\approx \epsilon^2 n^2$ entries differ across these matrices. Need to read at least $\Omega(1/\epsilon^2)$ entries before you can distinguish them with good probability.

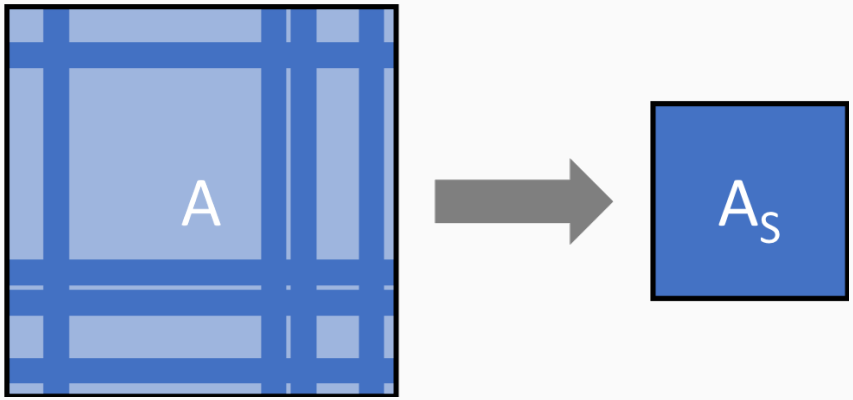
The Algorithm

The algorithm just computes the eigenvalues of a small random principal submatrix of A.



The Algorithm

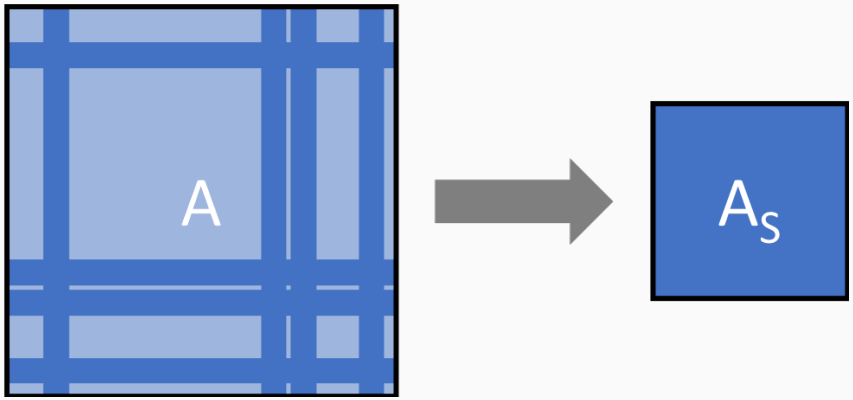
The algorithm just computes the eigenvalues of a small random principal submatrix of A.



1. Let $s = \frac{c \log^3 n}{\epsilon^3}$, and let A_S be the random principal submatrix of A where each row/column is included independently with probability $\frac{s}{n}$.
2. Compute all eigenvalues of $\frac{n}{s} \cdot A_S$.
3. Use these eigenvalues to approximate all eigenvalues of A.

The Algorithm

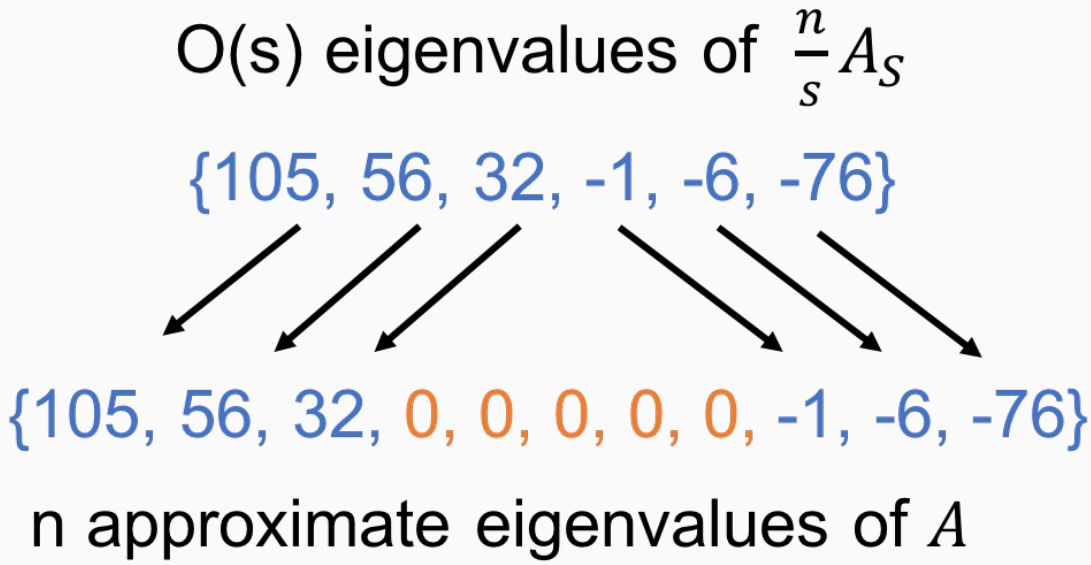
The algorithm just computes the eigenvalues of a small random principal submatrix of A.



1. Let $s = \frac{c \log^3 n}{\epsilon^3}$, and let A_S be the random principal submatrix of A where each row/column is included independently with probability $\frac{s}{n}$.
2. Compute all eigenvalues of $\frac{n}{s} \cdot A_S$.
3. Use these eigenvalues to approximate all eigenvalues of A. Observe that A_S has $O(s)$ eigenvalues while A has n.

Eigenvalue Alignment

Approximate the large positive eigenvalues using the positive eigenvalues of A_S, the large negative ones using the negative eigenvalues of A_S, and the rest by 0.



Improved Bounds for Sparse Matrices

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with entries bounded in magnitude by 1, $\text{nnz}(A)$ non-zero entries, and $\text{nnz}(A_i)$ entries in row i.

Sparse Matrix Result: Given the ability to sample $i \in [n]$ with probability $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$, there is an algorithm that reads $O\left(\frac{\log^{16} n}{\epsilon^{16}}\right)$ entries of A and outputs $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that, for all $i \in [n]$,

$$|\lambda_i - \tilde{\lambda}_i| \leq \epsilon \cdot \sqrt{\text{nnz}(A)}.$$

Improved Bounds for Sparse Matrices

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with entries bounded in magnitude by 1, $\text{nnz}(A)$ non-zero entries, and $\text{nnz}(A_i)$ entries in row i.

Sparse Matrix Result: Given the ability to sample $i \in [n]$ with probability $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$, there is an algorithm that reads $O\left(\frac{\log^{16} n}{\epsilon^{16}}\right)$ entries of A and outputs $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that, for all $i \in [n]$,

$$|\lambda_i - \tilde{\lambda}_i| \leq \epsilon \cdot \sqrt{\text{nnz}(A)}.$$

- Observe that $|\lambda_i| \leq \|A\|_F \leq \sqrt{\text{nnz}(A)} \leq n$ for all i.

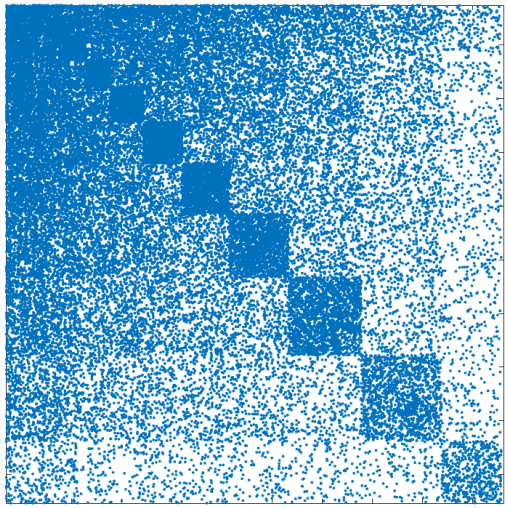
Improved Bounds for Sparse Matrices

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with entries bounded in magnitude by 1, $\text{nnz}(A)$ non-zero entries, and $\text{nnz}(A_i)$ entries in row i.

Sparse Matrix Result: Given the ability to sample $i \in [n]$ with probability $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$, there is an algorithm that reads $O\left(\frac{\log^{16} n}{\epsilon^{16}}\right)$ entries of A and outputs $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that, for all $i \in [n]$,

$$|\lambda_i - \tilde{\lambda}_i| \leq \epsilon \cdot \sqrt{\text{nnz}(A)}.$$

- Observe that $|\lambda_i| \leq \|A\|_F \leq \sqrt{\text{nnz}(A)} \leq n$ for all i.



Improved Bounds for Sparse Matrices

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with entries bounded in magnitude by 1, $\text{nnz}(A)$ non-zero entries, and $\text{nnz}(A_i)$ entries in row i.

Sparse Matrix Result: Given the ability to sample $i \in [n]$ with probability $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$, there is an algorithm that reads $O\left(\frac{\log^{16} n}{\epsilon^{16}}\right)$ entries of A and outputs $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that, for all $i \in [n]$,

$$|\lambda_i - \tilde{\lambda}_i| \leq \epsilon \cdot \sqrt{\text{nnz}(A)}.$$

- Observe that $|\lambda_i| \leq \|A\|_F \leq \sqrt{\text{nnz}(A)} \leq n$ for all i.
- Sparsity sampling requires sublinear queries per sample in the standard graph query model, where A is the adjacency matrix.
- Also possible via sampling a random non-zero entry when A is stored in sparse matrix format.

Improved Bounds for Sparse Matrices

Consider a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with entries bounded in magnitude by 1, $\text{nnz}(A)$ non-zero entries, and $\text{nnz}(A_i)$ entries in row i.

Sparse Matrix Result: Given the ability to sample $i \in [n]$ with probability $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$, there is an algorithm that reads $O\left(\frac{\log^{16} n}{\epsilon^{16}}\right)$ entries of A and outputs $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that, for all $i \in [n]$,

$$|\lambda_i - \tilde{\lambda}_i| \leq \epsilon \cdot \sqrt{\text{nnz}(A)}.$$

- Observe that $|\lambda_i| \leq \|A\|_F \leq \sqrt{\text{nnz}(A)} \leq n$ for all i.
- Sparsity sampling requires sublinear queries per sample in the standard graph query model, where A is the adjacency matrix.
- Also possible via sampling a random non-zero entry when A is stored in sparse matrix format.
- Surprisingly, simply computing the eigenvalues of a random submatrix does not suffice here. Need to carefully zero out some entries of the sampled matrix.

Related Work

Recently, several works have looked at sublinear time property testing algorithms for matrices in the bounded entry model.

Related Work

Recently, several works have looked at sublinear time property testing algorithms for matrices in the bounded entry model.

- [Balcan, Li, Woodruff, Zhang '18] give algorithms for testing rank, stable rank, and matrix norms. E.g., $\tilde{O}(1/\epsilon^c)$ queries for testing if A's Schatten-p norm is $\geq cn^p$ or at least an ϵ fraction of A's entries must be changed for it to have this property.
- [Bakshi, Chepurko, Jayaram '20] give $\tilde{O}(1/\epsilon^c)$ query algorithms for testing if A is either positive semidefinite or has at least one negative eigenvalue $< -\epsilon n$.

Related Work

Recently, several works have looked at sublinear time property testing algorithms for matrices in the bounded entry model.

- [Balcan, Li, Woodruff, Zhang '18] give algorithms for testing rank, stable rank, and matrix norms. E.g., $\tilde{O}(1/\epsilon^c)$ queries for testing if A's Schatten-p norm is $\geq cn^p$ or at least an ϵ fraction of A's entries must be changed for it to have this property.
- [Bakshi, Chepurko, Jayaram '20] give $\tilde{O}(1/\epsilon^c)$ query algorithms for testing if A is either positive semidefinite or has at least one negative eigenvalue $< -\epsilon n$.
- Our point-wise approximation guarantee immediately implies such a testing result, but can be stronger. However, our ϵ dependence is worse.
- Our techniques are related to those of Bakshi, Chepurko, Jayaram.

Related Work

Several other works look at sublinear time spectral density estimation for normalized graph adjacency matrices, which are a special class of bounded entry matrices.

- The goal is to approximate the spectral density: the distribution placing mass $1/n$ at each eigenvalue.
- [Cohen-Steiner, Kong, Sohler, and Valiant '18] give a $2^{O(1/\epsilon)}$ time algorithm for ϵ error approximation in the Wasserstein-1 distance.
- [Braverman, Krishnan, and Musco '22] give a $\tilde{O}(n/\epsilon^c)$ time algorithm for the same task.

Related Work

Several other works look at sublinear time spectral density estimation for normalized graph adjacency matrices, which are a special class of bounded entry matrices.

- The goal is to approximate the spectral density: the distribution placing mass $1/n$ at each eigenvalue.
- [Cohen-Steiner, Kong, Sohler, and Valiant '18] give a $2^{O(1/\epsilon)}$ time algorithm for ϵ error approximation in the Wasserstein-1 distance.
- [Braverman, Krishnan, and Musco '22] give a $\tilde{O}(n/\epsilon^c)$ time algorithm for the same task.
- Our result gives ϵn error approximation in the Wasserstein-1 distance.
- Note that the eigenvalues of a general bounded entry matrix lie in $[-n, n]$. Those of a normalized adjacency matrix lie in $[-1, 1]$.

Broader Context

Broader Research Goal: Use randomness to give sublinear time algorithms for natural linear algebraic problems. Typically under some assumption on the input matrix structure.

Broader Context

Broader Research Goal: Use randomness to give sublinear time algorithms for natural linear algebraic problems. Typically under some assumption on the input matrix structure.

- $\tilde{O}(nk/\epsilon^c)$ time algorithms for near optimal rank-k approximation of positive semidefinite and distance matrices [Musco Musco '17, Musco Woodruff '17, Bakshi Woodruff '18, Indyk et al. '19]
- $\tilde{O}(d \cdot n^{1.173})$ time algorithm for estimating the top eigenvalue of a Gaussian kernel matrix [Backurs Indyk Musco Wagner '21]
- Sublinear time algorithms for structured matrices via sublinear time matrix vector multiplication [Shi Woodruff '19]
- 'Quantum-inspired' algorithms for linear algebra [Tang '18, Chepurko Clarkson Horesh Lin Woodruff '21]
- Classic additive error randomized SVD [Frieze Kannan Vempala '04, Drineas Kannan Mahoney '06].

Proof Approach

Proof Approach

Recall: For a uniformly random principal submatrix A_S, need to show that the eigenvalues of $\frac{n}{s} \cdot A_S$, appropriately padded with zeros, approximate all eigenvalues of A to error $\pm\epsilon n$.

- A_S will be $O(s) \times O(s)$ for $s = \text{poly}(\log n, 1/\epsilon)$.

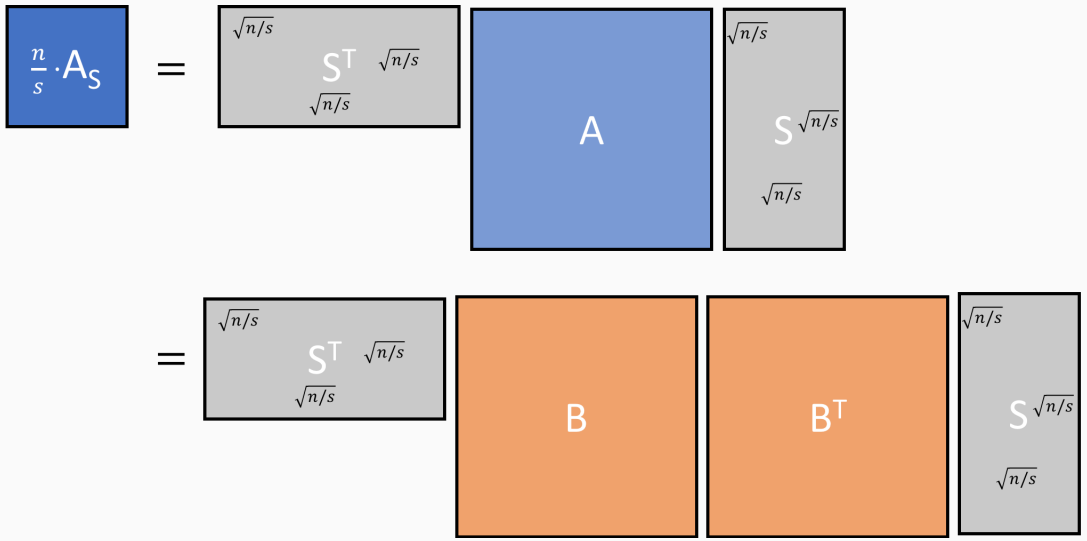
Positive Semidefinite Case

The proof is simple when A is positive semidefinite (PSD) — i.e., has all non-negative eigenvalues.

Positive Semidefinite Case

The proof is simple when A is positive semidefinite (PSD) — i.e., has all non-negative eigenvalues.

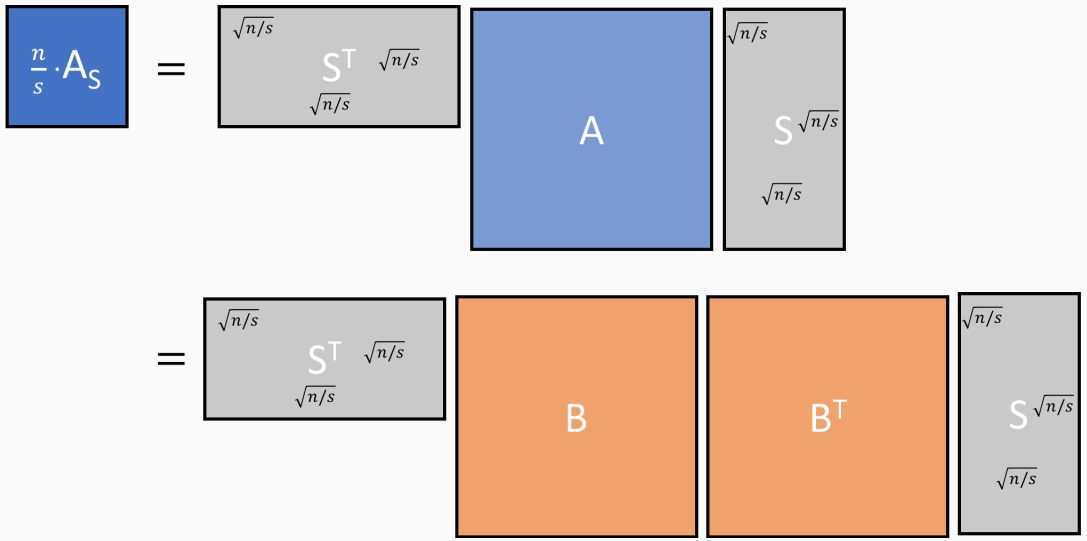
- Let $B \in \mathbb{R}^{n \times n}$ be such that $BB^T = A$.
- Let $\frac{n}{s} \cdot A_S = S^T A S$ be our random principal submatrix, where $S \in \mathbb{R}^{n \times s}$ is an appropriately scaled sampling matrix.



Positive Semidefinite Case

The proof is simple when A is positive semidefinite (PSD) — i.e., has all non-negative eigenvalues.

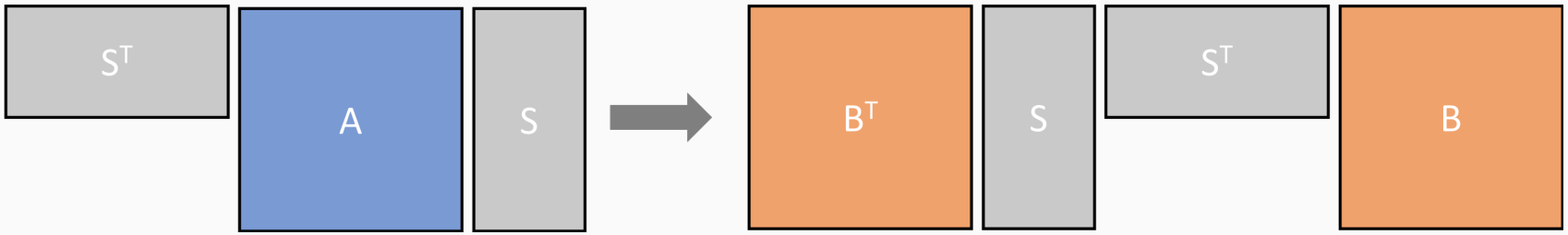
- Let $B \in \mathbb{R}^{n \times n}$ be such that $BB^T = A$.
- Let $\frac{n}{s} \cdot A_S = S^T A S$ be our random principal submatrix, where $S \in \mathbb{R}^{n \times s}$ is an appropriately scaled sampling matrix.
- The non-zero eigenvalues of $\frac{n}{s} \cdot A_S = S^T A S = S^T B B^T S$ are identical to those of $B^T S S^T B$. And those of $A = BB^T$ are identical to those of $B^T B$.



Positive Semidefinite Case

The proof is simple when A is positive semidefinite (PSD) — i.e., has all non-negative eigenvalues.

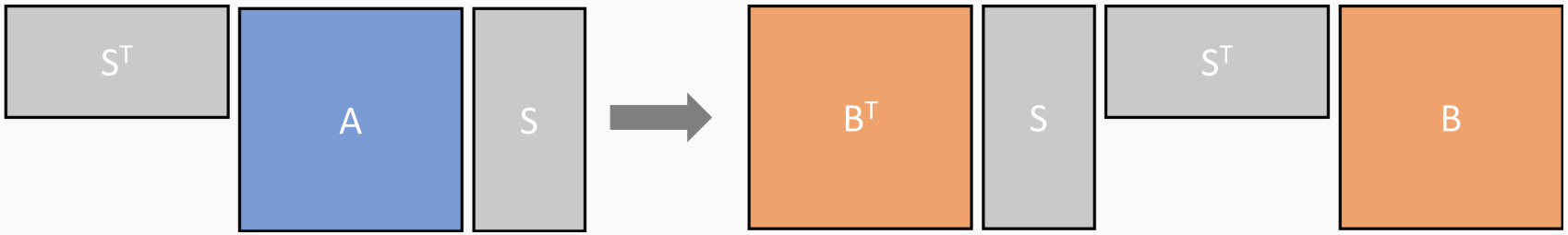
- Let $B \in \mathbb{R}^{n \times n}$ be such that $BB^T = A$.
- Let $\frac{n}{s} \cdot A_S = S^T AS$ be our random principal submatrix, where $S \in \mathbb{R}^{n \times s}$ is an appropriately scaled sampling matrix.
- The non-zero eigenvalues of $\frac{n}{s} \cdot A_S = S^T AS = S^T BB^T S$ are identical to those of $B^T SS^T B$. And those of $A = BB^T$ are identical to those of $B^T B$.



Positive Semidefinite Case

The proof is simple when A is positive semidefinite (PSD) — i.e., has all non-negative eigenvalues.

- Let $B \in \mathbb{R}^{n \times n}$ be such that $BB^T = A$.
- Let $\frac{n}{s} \cdot A_S = S^T AS$ be our random principal submatrix, where $S \in \mathbb{R}^{n \times s}$ is an appropriately scaled sampling matrix.
- The non-zero eigenvalues of $\frac{n}{s} \cdot A_S = S^T AS = S^T BB^T S$ are identical to those of $B^T SS^T B$. And those of $A = BB^T$ are identical to those of $B^T B$.



- So it suffices to analyze how well the eigenvalues of $B^T SS^T B$ approximate those of $B^T B$.

Positive Semidefinite Case

So Far: To show that the eigenvalues of $\frac{n}{s} \cdot A_S$ approximate those of A, it suffices to show that those of $B^T S S^T B$ approximate those of $B^T B$.

Positive Semidefinite Case

So Far: To show that the eigenvalues of $\frac{n}{s} \cdot A_S$ approximate those of A, it suffices to show that those of $B^T S S^T B$ approximate those of $B^T B$.

- Via a standard approximate matrix multiplication analysis (c.f. [Drineas Kannan '01]), with high probability when $s = O(1/\epsilon^2)$,

$$\|B^T B - B^T S S^T B\|_F \leq \epsilon n.$$

- By an eigenvalue version of the Hoffman–Wielandt perturbation bound (c.f. [Bhatia '13]), letting $\Lambda(\cdot)$ denote the eigenvalue vector of a matrix,

$$\|\Lambda(B^T B) - \Lambda(B^T S S^T B)\|_\infty \leq \|\Lambda(B^T B) - \Lambda(B^T S S^T B)\|_2 \leq \epsilon n.$$

- This gives that $|\lambda_i - \tilde{\lambda}_i| \leq \epsilon n$ for all i (padding the eigenvalues of $\frac{n}{s} \cdot A_S$ with zeros accounts for the $n - O(s)$ zero eigenvalues of $B^T S S^T B$ that are not present in $\frac{n}{s} \cdot A_S$.).

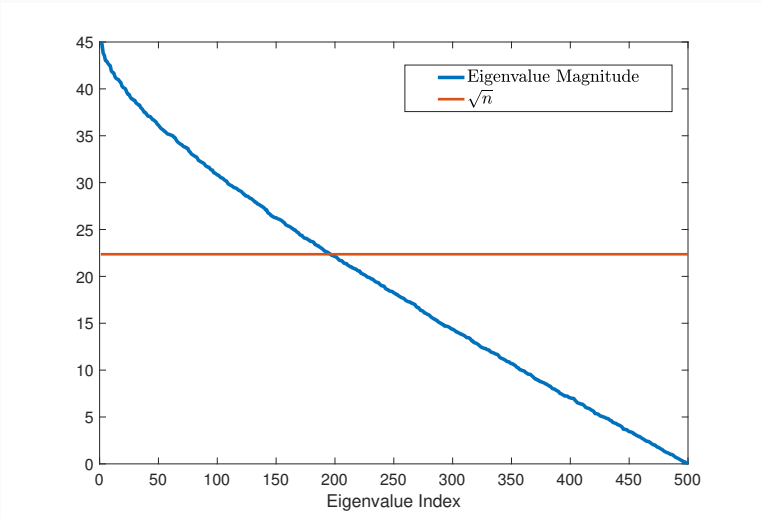
General Case

In the general bounded entry setting, the previous proof breaks down for a simple reason: when A is not PSD, it does not admit a real square root $B \in \mathbb{R}^{n \times n}$ with $BB^T = A$.

General Case

In the general bounded entry setting, the previous proof breaks down for a simple reason: when A is not PSD, it does not admit a real square root $B \in \mathbb{R}^{n \times n}$ with $BB^T = A$.

- Feels like a technicality, but observe that when A is PSD, $\|\Lambda(A)\|_1 = \sum_{i=1}^{n} \lambda_i = \text{tr}(A) \leq n$.
- When A is not PSD, we can have $\|\Lambda(A)\|_1 = O(n^{3/2})$. I.e., there can be significantly more eigenvalue mass overall.
- E.g., have $\Theta(n)$ eigenvalues with $\lambda_i = \Theta(\sqrt{n})$, when A is random ± 1.



General Case

In the general bounded entry setting, the previous proof breaks down for a simple reason: when A is not PSD, it does not admit a real square root $B \in \mathbb{R}^{n \times n}$ with $BB^T = A$.

- Feels like a technicality, but observe that when A is PSD, $\|\Lambda(A)\|_1 = \sum_{i=1}^{n} \lambda_i = \mathrm{tr}(A) \leq n$.
- When A is not PSD, we can have $\|\Lambda(A)\|_1 = O(n^{3/2})$. I.e., there can be significantly more eigenvalue mass overall.
- E.g., have $\Theta(n)$ eigenvalues with $\lambda_i = \Theta(\sqrt{n})$, when A is random ± 1.
- Relatedly, we can never hope to prove an ℓ_2 error bound as we did in the PSD case, where $\|\Lambda - \tilde{\Lambda}\|_\infty \leq \|\Lambda - \tilde{\Lambda}\|_2 \leq \epsilon n$.
- We approximate almost all eigenvalues by 0, so in the random matrix case will have $\|\Lambda - \tilde{\Lambda}\|_2 \approx \|\Lambda\|_2 = n$.

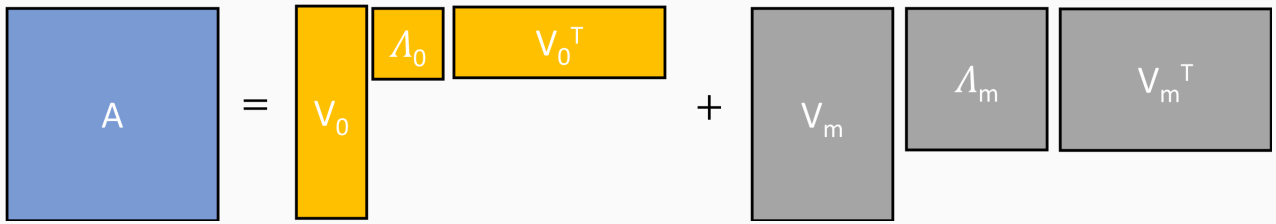
Eigenvalue Split

Key Idea: Split A into its outlying eigenvalues, for which we give non-trivial approximations, and its middle eigenvalues, and analyze these components separately.

Eigenvalue Split

Key Idea: Split A into its outlying eigenvalues, for which we give non-trivial approximations, and its middle eigenvalues, and analyze these components separately.

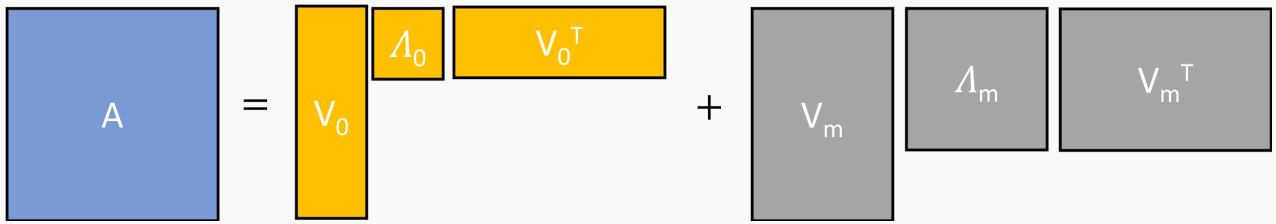
- Let $V_o \in \mathbb{R}^{n \times n_o}$ have columns equal to all eigenvectors with corresponding eigenvalues satisfying $|\lambda_i| \geq \epsilon n$. Let $V_m \in \mathbb{R}^{n \times n_m}$ have columns equal to the remaining eigenvectors.
- Let $\Lambda_o \in \mathbb{R}^{n_o \times n_o}$ and $\Lambda_m \in \mathbb{R}^{n_m \times n_m}$ be the corresponding diagonal eigenvalue matrices.
- Write $A = A_o + A_m$ where $A_o = V_o \Lambda_o V_o^T$ and $A_m = V_m \Lambda_m V_m^T$.



Eigenvalue Split

Key Idea: Split A into its outlying eigenvalues, for which we give non-trivial approximations, and its middle eigenvalues, and analyze these components separately.

- Let $V_o \in \mathbb{R}^{n \times n_o}$ have columns equal to all eigenvectors with corresponding eigenvalues satisfying $|\lambda_i| \geq \epsilon n$. Let $V_m \in \mathbb{R}^{n \times n_m}$ have columns equal to the remaining eigenvectors.
- Let $\Lambda_o \in \mathbb{R}^{n_o \times n_o}$ and $\Lambda_m \in \mathbb{R}^{n_m \times n_m}$ be the corresponding diagonal eigenvalue matrices.
- Write $A = A_o + A_m$ where $A_o = V_o \Lambda_o V_o^T$ and $A_m = V_m \Lambda_m V_m^T$.



- Can similarly write $\frac{n}{s} \cdot A_S = S^T A S = S^T A_o S + S^T A_m S$.

Proof Approach

So Far: Have written $A = A_o + A_m$ and $S^T A S = S^T A_o S + S^T A_m S$.

Proof Approach

So Far: Have written $A = A_o + A_m$ and $S^T A S = S^T A_o S + S^T A_m S$.

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm \epsilon n$ error.

Proof Approach

So Far: Have written $A = A_o + A_m$ and $S^T A S = S^T A_o S + S^T A_m S$.

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm \epsilon n$ error.

Step 2: Show that the eigenvalues of $S^T A_m S$ are all small in magnitude — i.e. $\leq \epsilon n$.

Proof Approach

So Far: Have written $A = A_o + A_m$ and $S^T A S = S^T A_o S + S^T A_m S$.

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm \epsilon n$ error.

Step 2: Show that the eigenvalues of $S^T A_m S$ are all small in magnitude — i.e. $\leq \epsilon n$.

Step 3: By Weyl's inequality and Step 2, the eigenvalues of $S^T A S$ are within $\pm \epsilon n$ of those of $S^T A_0 S$. Thus, by Step 1, they are all either within $\pm 2\epsilon n$ of some eigenvalue of A_o or bounded in magnitude by ϵn.

This is enough to give that the eigenvalues of $\frac{n}{s} \cdot A_S = S^T A S$, appropriately padded with zeros, approximate all eigenvalues of A up to $\pm 2\epsilon n$ error.

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Key Proof Idea: Since A has bounded magnitude entries, its outlying eigenvectors are all incoherent — i.e., their mass is spread across many entries. Thus, they are well approximated via uniform sampling.

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Key Proof Idea: Since A has bounded magnitude entries, its outlying eigenvectors are all incoherent — i.e., their mass is spread across many entries. Thus, they are well approximated via uniform sampling.

For any unit norm eigenvector $v \in \mathbb{R}^n$ with $Av = \lambda \cdot v$ and $|\lambda| \geq \epsilon n$:

$$|v(i)| = \frac{1}{|\lambda|} \cdot |[Av](i)|$$

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Key Proof Idea: Since A has bounded magnitude entries, its outlying eigenvectors are all incoherent — i.e., their mass is spread across many entries. Thus, they are well approximated via uniform sampling.

For any unit norm eigenvector $v \in \mathbb{R}^n$ with $Av = \lambda \cdot v$ and $|\lambda| \geq \epsilon n$:

$$|v(i)| = \frac{1}{|\lambda|} \cdot |[Av](i)| = \frac{1}{|\lambda|} \cdot \langle A_{i,:}, v \rangle \leq \frac{1}{|\lambda|} \cdot \|A_{i,:}\|_2 \cdot \|v\|_2$$

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Key Proof Idea: Since A has bounded magnitude entries, its outlying eigenvectors are all incoherent — i.e., their mass is spread across many entries. Thus, they are well approximated via uniform sampling.

For any unit norm eigenvector $v \in \mathbb{R}^n$ with $Av = \lambda \cdot v$ and $|\lambda| \geq \epsilon n$:

$$|v(i)| = \frac{1}{|\lambda|} \cdot |[Av](i)| = \frac{1}{|\lambda|} \cdot \langle A_{i,:}, v \rangle \leq \frac{1}{|\lambda|} \cdot \|A_{i,:}\|_2 \cdot \|v\|_2 \leq \frac{1}{\epsilon\sqrt{n}}.$$

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Key Proof Idea: Since A has bounded magnitude entries, its outlying eigenvectors are all incoherent — i.e., their mass is spread across many entries. Thus, they are well approximated via uniform sampling.

For any unit norm eigenvector $v \in \mathbb{R}^n$ with $Av = \lambda \cdot v$ and $|\lambda| \geq \epsilon n$:

$$|v(i)| = \frac{1}{|\lambda|} \cdot |[Av](i)| = \frac{1}{|\lambda|} \cdot \langle A_{i,:}, v \rangle \leq \frac{1}{|\lambda|} \cdot \|A_{i,:}\|_2 \cdot \|v\|_2 \leq \frac{1}{\epsilon\sqrt{n}}.$$

I.e., v is within a $1/\epsilon$ factor of being perfectly flat.

Incoherence of Outlying Eigenvalues

Step 1: Show that the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Key Proof Idea: Since A has bounded magnitude entries, its outlying eigenvectors are all incoherent — i.e., their mass is spread across many entries. Thus, they are well approximated via uniform sampling.

For any unit norm eigenvector $v \in \mathbb{R}^n$ with $Av = \lambda \cdot v$ and $|\lambda| \geq \epsilon n$:

$$|v(i)| = \frac{1}{|\lambda|} \cdot |[Av](i)| = \frac{1}{|\lambda|} \cdot \langle A_{i,:}, v \rangle \leq \frac{1}{|\lambda|} \cdot \|A_{i,:}\|_2 \cdot \|v\|_2 \leq \frac{1}{\epsilon\sqrt{n}}.$$

I.e., v is within a $1/\epsilon$ factor of being perfectly flat.

The above bound was an important part of [Bakshi, Chepurko, and Jayaram '20]. We show a related bound, that $\|[V_o]_{i,:}\|_2^2 \leq \frac{1}{\epsilon^2 n}$. I.e., we show that the leverage scores of V_o are uniformly bounded.

Sampling Outlying Eigendirections

So far: Can show that the outlying eigenspace of A is incoherent, with i^{th} leverage score bounded by $\|[V_o]_{i,:}\|_2^2 \leq \frac{1}{\epsilon^2 n}$.

Sampling Outlying Eigendirections

So far: Can show that the outlying eigenspace of A is incoherent, with i^{th} leverage score bounded by $\|[V_o]_{i,:}\|_2^2 \leq \frac{1}{\epsilon^2 n}$.

- Via a standard matrix Bernstein bound, can show that if we take $s = \tilde{O}(1/\epsilon^4)$ samples, with high probability $V_o^T S S^T V_o \approx V_o^T V_o \approx I$.
- Can use this to argue that the nonzero eigenvalues of $S^T A_o S = S^T V_o \Lambda_o V_o S$ are close to those of Λ_o — i.e., close to the outlying eigenvalues in A_o.
- This completes Step 1 of the proof.

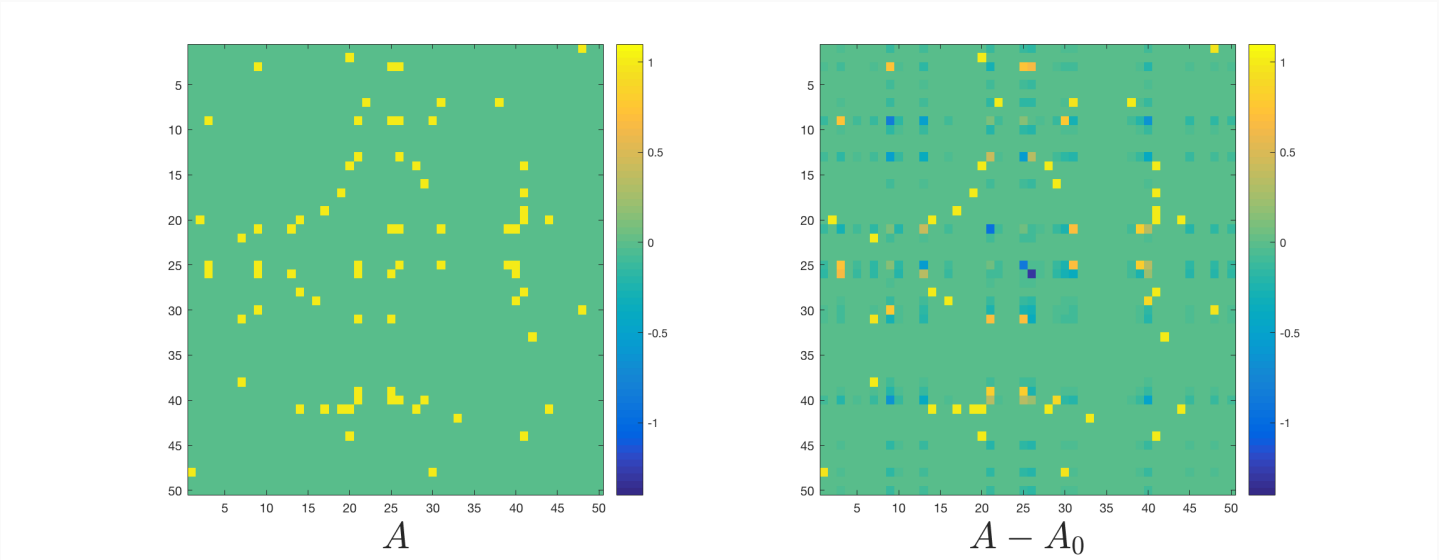
Sampling Middle Eigendirections

It remains to show that the eigenvalues of $S^T A_m S$ are bounded in magnitude by ϵn.

Sampling Middle Eigendirections

It remains to show that the eigenvalues of $S^T A_m S$ are bounded in magnitude by ϵn.

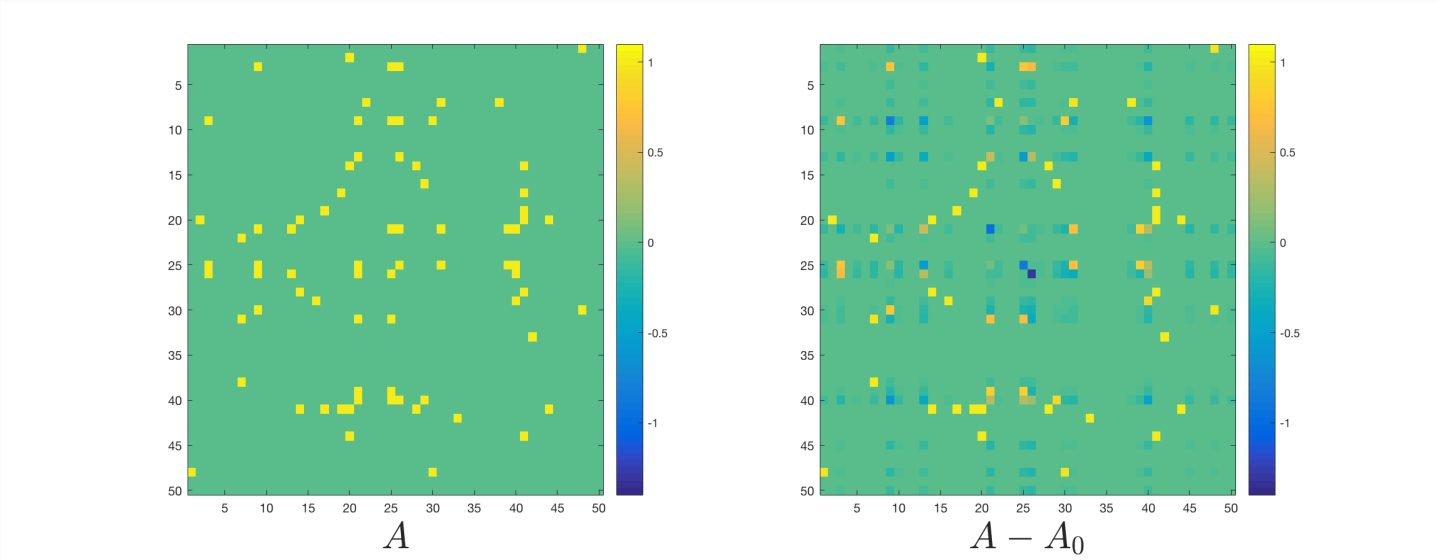
- Via the incoherence of V_o, can show that $\|A_o\|_\infty \leq \frac{1}{\epsilon}$ and so by triangle inequality, $\|A_m\|_\infty \leq \|A\|_\infty + \|A_o\|_\infty \leq 1 + \frac{1}{\epsilon}$.



Sampling Middle Eigendirections

It remains to show that the eigenvalues of $S^T A_m S$ are bounded in magnitude by ϵn.

- Via the incoherence of V_o, can show that $\|A_o\|_\infty \leq \frac{1}{\epsilon}$ and so by triangle inequality, $\|A_m\|_\infty \leq \|A\|_\infty + \|A_o\|_\infty \leq 1 + \frac{1}{\epsilon}$.



- Can then apply spectral norm bounds for random principal submatrices of bounded entry matrices [Rudelson Vershynin '07, Tropp '08], to show that $\|S^T A_m S\|_2 \leq \epsilon n$ when $s = \tilde{O}(1/\epsilon^2)$.

Proof Recap

Step 0: Split $A = A_o + A_m$ into its outlying and middle eigendirections.

Step 1: Prove that the outlying eigendirections of A are incoherent, and thus, uniform sampling approximately preserves the eigenvalues of A_o. I.e., the non-zero eigenvalues of $S^T A_o S$ approximate all the eigenvalues of A_o to $\pm\epsilon n$ error.

Step 2: Use the incoherence of A_o to argue that $A_m = A - A_o$ is entrywise bounded, and thus $\|S^T A_m S\|_2 \leq \epsilon n$.

Step 3: Combine the above to show that, after padding by zeros, the eigenvalues of $\frac{n}{s} \cdot A_S = S^T A S = S^T A_o S + S^T A_m S$ approximate those of A up to $\pm\epsilon n$ error.

Sparse Matrices

Importance Sampling

In many cases in randomized numerical linear algebra, more refined sampling probabilities can give stronger error bounds. It is natural to expect this for eigenvalue estimation.

Importance Sampling

In many cases in randomized numerical linear algebra, more refined sampling probabilities can give stronger error bounds. It is natural to expect this for eigenvalue estimation.

- Uniform Sampling: $\pm\epsilon n$ (just showed)
- Sparsity Sampling $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$: $\pm\epsilon\sqrt{\text{nnz}(A)}$?
- Norm Sampling $\propto \frac{\|A_i\|_2^2}{\|A\|_F^2}$: $\pm\epsilon\|A\|_F$?

Importance Sampling

In many cases in randomized numerical linear algebra, more refined sampling probabilities can give stronger error bounds. It is natural to expect this for eigenvalue estimation.

- Uniform Sampling: $\pm\epsilon n$ (just showed)
- Sparsity Sampling $\propto \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$: $\pm\epsilon\sqrt{\text{nnz}(A)}$?
- Norm Sampling $\propto \frac{\|A_i\|_2^2}{\|A\|_F^2}$: $\pm\epsilon\|A\|_F$?

We show that the improvement for sparsity-based sampling is possible — although not so straightforward. Further improvement e.g., for norm-based sampling is open.

Sparsity Sampling Algorithm

Natural extension of random submatrix algorithm to sparsity-based sampling:

1. Let $s = \text{poly}(\log n, 1/\epsilon)$, and let A_S be the random principal submatrix of A where each row/column is included independently with probability $p_i = s \cdot \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$.
2. Let D be the diagonal matrix with $D_{i,i} = \frac{1}{\sqrt{p_j}}$ if the i^{th} sampled row/column is row j.
3. Compute all eigenvalues of DA_SD.
4. Use these eigenvalues to approximate all eigenvalues of A.

Sparsity Sampling Algorithm

Natural extension of random submatrix algorithm to sparsity-based sampling:

1. Let $s = \text{poly}(\log n, 1/\epsilon)$, and let A_S be the random principal submatrix of A where each row/column is included independently with probability $p_i = s \cdot \frac{\text{nnz}(A_i)}{\text{nnz}(A)}$.
2. Let D be the diagonal matrix with $D_{i,i} = \frac{1}{\sqrt{p_j}}$ if the i^{th} sampled row/column is row j.
3. Compute all eigenvalues of DA_SD.
4. Use these eigenvalues to approximate all eigenvalues of A.

Observe that if the rows have uniform sparsity, $DA_SD = \frac{n}{s} \cdot A_S$, and we have exactly the uniform sampling algorithm.

Challenge 1: The Identity

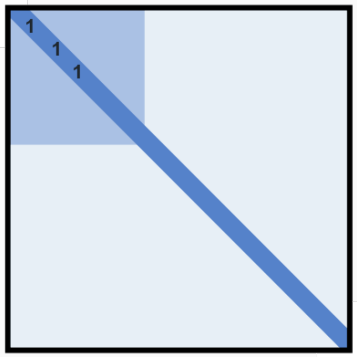
Say that $A = I$, so sparsity-based sampling is just uniform sampling, so $DA_SD = \frac{n}{s} \cdot A_S$. Also $\text{nnz}(A) = n$.

- Want to approximate all eigenvalues up to $\pm\epsilon\sqrt{n}$

Challenge 1: The Identity

Say that $A = I$, so sparsity-based sampling is just uniform sampling, so $DA_SD = \frac{n}{s} \cdot A_S$. Also $\text{nnz}(A) = n$.

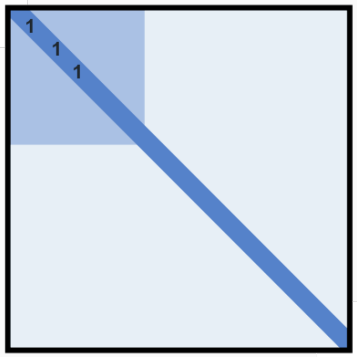
- Want to approximate all eigenvalues up to $\pm\epsilon\sqrt{n}$
- However, A_S is just a smaller identity matrix, so $\frac{n}{s} \cdot A_S$ has all eigenvalues equal to $\frac{n}{s}$.



Challenge 1: The Identity

Say that $A = I$, so sparsity-based sampling is just uniform sampling, so $DA_SD = \frac{n}{s} \cdot A_S$. Also $\text{nnz}(A) = n$.

- Want to approximate all eigenvalues up to $\pm\epsilon\sqrt{n}$
- However, A_S is just a smaller identity matrix, so $\frac{n}{s} \cdot A_S$ has all eigenvalues equal to $\frac{n}{s}$.

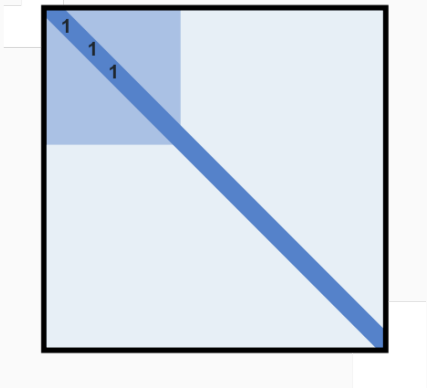


- Would need to set $s \geq \frac{\sqrt{n}}{\epsilon}$ to achieve the desired bound.

Challenge 1: The Identity

Say that $A = I$, so sparsity-based sampling is just uniform sampling, so $DA_SD = \frac{n}{s} \cdot A_S$. Also $\text{nnz}(A) = n$.

- Want to approximate all eigenvalues up to $\pm\epsilon\sqrt{n}$
- However, A_S is just a smaller identity matrix, so $\frac{n}{s} \cdot A_S$ has all eigenvalues equal to $\frac{n}{s}$.

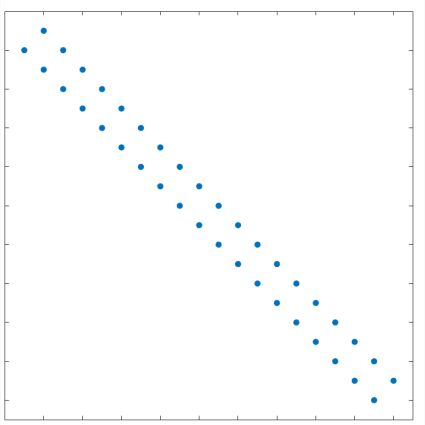


- Would need to set $s \geq \frac{\sqrt{n}}{\epsilon}$ to achieve the desired bound.

Simple Fix: Set the diagonal of A_S to 0. Introduces at most ± 1 error into the eigenvalue estimates and resolves this issue. When $A = I$, $A_S = 0$. So our eigenvalue estimates all have error 1.

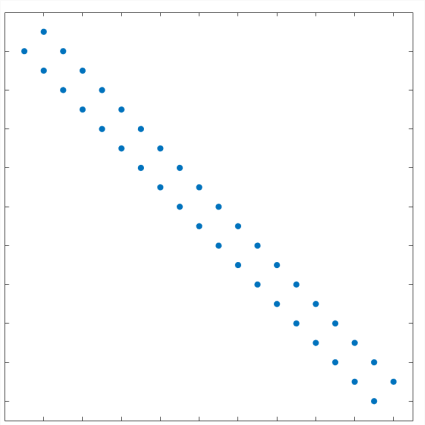
Challenge 2: Sparse Rows and Columns

Consider A with $A_{i,i+1} = A_{i+1,i} = 1$ for all $i = 1, \ldots, n-1$.



Challenge 2: Sparse Rows and Columns

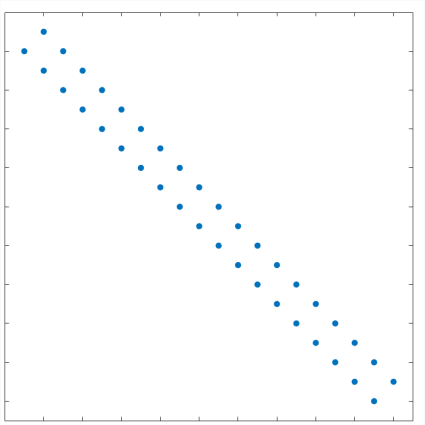
Consider A with $A_{i,i+1} = A_{i+1,i} = 1$ for all $i = 1, \ldots, n-1$.



- $\sqrt{\text{nnz}(A)} \leq \sqrt{2n}$ and $|\lambda_i| \leq 2$ for all i. Roughly, $DA_SD \approx \frac{n}{s} \cdot A_S$.

Challenge 2: Sparse Rows and Columns

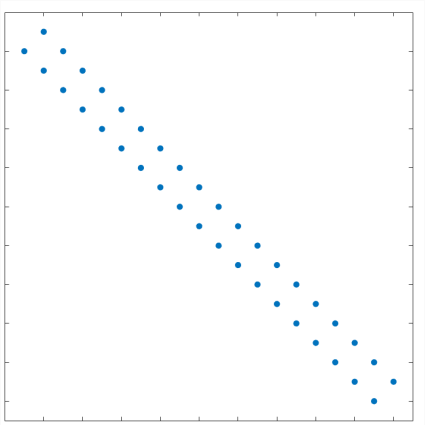
Consider A with $A_{i,i+1} = A_{i+1,i} = 1$ for all $i = 1, \ldots, n-1$.



- $\sqrt{\text{nnz}(A)} \leq \sqrt{2n}$ and $|\lambda_i| \leq 2$ for all i. Roughly, $DA_SD \approx \frac{n}{s} \cdot A_S$.
- If $s = o(\sqrt{n})$, with good probability, $\frac{n}{s} \cdot A_S = 0$, and we get good estimates: ± 2 error.

Challenge 2: Sparse Rows and Columns

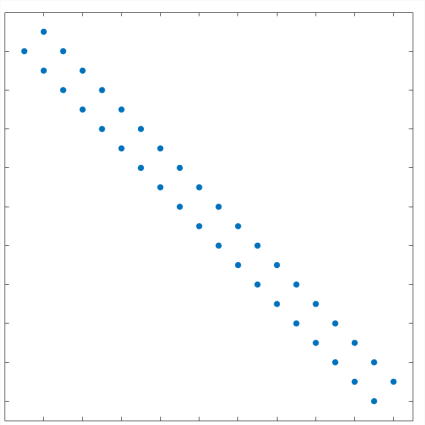
Consider A with $A_{i,i+1} = A_{i+1,i} = 1$ for all $i = 1, \ldots, n-1$.



- $\sqrt{\text{nnz}(A)} \leq \sqrt{2n}$ and $|\lambda_i| \leq 2$ for all i. Roughly, $DA_SD \approx \frac{n}{s} \cdot A_S$.
- If $s = o(\sqrt{n})$, with good probability, $\frac{n}{s} \cdot A_S = 0$, and we get good estimates: ± 2 error.
- When $s = c\sqrt{n}$, with constant probability, $\frac{n}{s} \cdot A_S$ is non-zero (birthday paradox), and we have error $\approx \frac{n}{s} = \frac{\sqrt{n}}{c} > \epsilon\sqrt{n}$.

Challenge 2: Sparse Rows and Columns

Consider A with $A_{i,i+1} = A_{i+1,i} = 1$ for all $i = 1, \ldots, n-1$.



- $\sqrt{\text{nnz}(A)} \leq \sqrt{2n}$ and $|\lambda_i| \leq 2$ for all i. Roughly, $DA_SD \approx \frac{n}{s} \cdot A_S$.
- If $s = o(\sqrt{n})$, with good probability, $\frac{n}{s} \cdot A_S = 0$, and we get good estimates: ± 2 error.
- When $s = c\sqrt{n}$, with constant probability, $\frac{n}{s} \cdot A_S$ is non-zero (birthday paradox), and we have error $\approx \frac{n}{s} = \frac{\sqrt{n}}{c} > \epsilon\sqrt{n}$.
- Can find many related examples: entries at the intersection of sparse rows/columns get scaled up too much in DA_SD, leading to large estimation errors.

Zeroing Out Entries

To handle these cases, we argue that zeroing out the entries of A lying at the intersection of sparse rows and columns does not significantly alter the eigenvalues.

Zeroing Out Entries

To handle these cases, we argue that zeroing out the entries of A lying at the intersection of sparse rows and columns does not significantly alter the eigenvalues.

Key Lemma: Let A' be equal to A but with $A'_{ij} = 0$ if $\sqrt{\text{nnz}(A_i) \cdot \text{nnz}(A_j)} \leq \frac{\epsilon\sqrt{\text{nnz}(A)}}{c \log n}$. Then $\|A - A'\|_2 \leq \epsilon\sqrt{\text{nnz}(A)}$.

Zeroing Out Entries

To handle these cases, we argue that zeroing out the entries of A lying at the intersection of sparse rows and columns does not significantly alter the eigenvalues.

Key Lemma: Let A' be equal to A but with $A'_{ij} = 0$ if $\sqrt{\text{nnz}(A_i) \cdot \text{nnz}(A_j)} \leq \frac{\epsilon\sqrt{\text{nnz}(A)}}{c \log n}$. Then $\|A - A'\|_2 \leq \epsilon\sqrt{\text{nnz}(A)}$.

- Can be thought of as a generalization of Girshgorin's circle theorem, which would show that zeroing out the entries in rows/columns with $\text{nnz}(A_i) \leq \epsilon\sqrt{\text{nnz}(A)}$ does not perturb the eigenvalues by more than $\epsilon\sqrt{\text{nnz}(A)}$.

Zeroing Out Entries

To handle these cases, we argue that zeroing out the entries of A lying at the intersection of sparse rows and columns does not significantly alter the eigenvalues.

Key Lemma: Let A' be equal to A but with $A'_{ij} = 0$ if $\sqrt{\text{nnz}(A_i) \cdot \text{nnz}(A_j)} \leq \frac{\epsilon\sqrt{\text{nnz}(A)}}{c \log n}$. Then $\|A - A'\|_2 \leq \epsilon\sqrt{\text{nnz}(A)}$.

- Can be thought of as a generalization of Girshgorin's circle theorem, which would show that zeroing out the entries in rows/columns with $\text{nnz}(A_i) \leq \epsilon\sqrt{\text{nnz}(A)}$ does not perturb the eigenvalues by more than $\epsilon\sqrt{\text{nnz}(A)}$.
- Ensures that after sampling, no entries are scaled up too much in DA_SD, and lets us extend our uniform sampling proof to give $\pm\epsilon\sqrt{\text{nnz}(A)}$ error with sparsity-based sampling.

Zeroing Out Entries

To handle these cases, we argue that zeroing out the entries of A lying at the intersection of sparse rows and columns does not significantly alter the eigenvalues.

Key Lemma: Let A' be equal to A but with $A'_{ij} = 0$ if $\sqrt{\text{nnz}(A_i) \cdot \text{nnz}(A_j)} \leq \frac{\epsilon\sqrt{\text{nnz}(A)}}{c \log n}$. Then $\|A - A'\|_2 \leq \epsilon\sqrt{\text{nnz}(A)}$.

- Can be thought of as a generalization of Girshgorin's circle theorem, which would show that zeroing out the entries in rows/columns with $\text{nnz}(A_i) \leq \epsilon\sqrt{\text{nnz}(A)}$ does not perturb the eigenvalues by more than $\epsilon\sqrt{\text{nnz}(A)}$.
- Ensures that after sampling, no entries are scaled up too much in DA_SD, and lets us extend our uniform sampling proof to give $\pm\epsilon\sqrt{\text{nnz}(A)}$ error with sparsity-based sampling.
- I'm ignoring many details here — see paper for the full argument. Challenging to obtain bounds on S^TA_mS when S is sampled non-uniformly.

Open Questions

Open Questions

- Can we obtain tight $\tilde{O}(1/\epsilon^2)$ query complexity for computing $\pm\epsilon n$ approximations to all eigenvalues? Requires going beyond principal submatrix sampling, for which a simple $\Omega(1/\epsilon^4)$ lower bound holds. What is even a plausible algorithm here?
- Can we approximate $\|A\|_1 = \sum_{i=1}^{n} |\lambda_i|$ to error $1/2 \cdot n^{3/2}$ using $o(n^2)$ queries to A? [Balcan, Li, Woodruff, Zhang '18] show that $\tilde{\Omega}(n)$ is required. Key challenge problem in understanding how to approximate bulk spectral properties.
- Can we give $\pm\epsilon\|A\|_F$ approximations to all eigenvalues via norm-based sampling?
- For what other classes of structured matrices can we give stronger approximation bounds? E.g., interesting bounds are known for normalized graph adjacency matrices. What else?
- What other spectral properties can we approximate in sublinear time?