Dimensionality reduction via sparse matrices

Jelani Nelson
Harvard

September 19, 2013

based on works with Daniel Kane (Stanford) and Huy Nguyễn (Princeton)
Metric Johnson-Lindenstrauss lemma

Metric JL (MJL) Lemma, 1984

Every set of $N$ points in Euclidean space can be embedded into $O(\varepsilon^{-2} \log N)$-dimensional Euclidean space so that all pairwise distances are preserved up to a $1 \pm \varepsilon$ factor.

Uses:
- Speed up geometric algorithms by first reducing dimension of input [Indyk, Motwani '98], [Indyk '01]
- Faster/streaming numerical linear algebra algorithms [Sarl´os '06], [LWMRT '07], [Clarkson, Woodruff '09]
- Essentially equivalent to RIP matrices from compressed sensing [Baraniuk et al. '08], [Krahmer, Ward '11] (used for recovery of sparse signals)
Metric JL (MJL) Lemma, 1984

Every set of \( N \) points in Euclidean space can be embedded into \( O(\varepsilon^{-2} \log N) \)-dimensional Euclidean space so that all pairwise distances are preserved up to a \( 1 \pm \varepsilon \) factor.

Uses:

- Speed up geometric algorithms by first reducing dimension of input [Indyk, Motwani ’98], [Indyk ’01]
- Faster/streaming numerical linear algebra algorithms [Sarlós ’06], [LWMRT ’07], [Clarkson, Woodruff ’09]
- Essentially equivalent to RIP matrices from compressed sensing [Baraniuk et al. ’08], [Krahmer, Ward ’11] (used for recovery of sparse signals)
How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma

For any $0 < \varepsilon, \delta < 1/2$ there exists a distribution $\mathcal{D}_{\varepsilon,\delta}$ on $\mathbb{R}^{m \times n}$ for $m = O(\varepsilon^{-2} \log(1/\delta))$ so that for any $u$ of unit $\ell_2$ norm

$$\mathbb{P}_{\Pi \sim \mathcal{D}_{\varepsilon,\delta}} \left( |\|\Pi u\|_2^2 - 1| > \varepsilon \right) < \delta.$$
How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma

For any $0 < \varepsilon, \delta < 1/2$ there exists a distribution $\mathcal{D}_{\varepsilon, \delta}$ on $\mathbb{R}^{m \times n}$ for $m = O(\varepsilon^{-2} \log(1/\delta))$ so that for any $u$ of unit $\ell_2$ norm

$$P_{\Pi \sim \mathcal{D}_{\varepsilon, \delta}} \left( |\|\Pi u\|_2^2 - 1| > \varepsilon \right) < \delta.$$

Proof of MJL: Set $\delta = 1/N^2$ in DJL and $u$ as the difference vector of some pair of points. Union bound over the $\binom{N}{2}$ pairs.
How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma

For any $0 < \varepsilon, \delta < 1/2$ there exists a distribution $\mathcal{D}_{\varepsilon,\delta}$ on $\mathbb{R}^{m \times n}$ for $m = O(\varepsilon^{-2} \log(1/\delta))$ so that for any $u$ of unit $\ell_2$ norm

$$\mathbb{P}_{\Pi \sim \mathcal{D}_{\varepsilon,\delta}} \left( \left| \| \Pi u \|_2^2 - 1 \right| > \varepsilon \right) < \delta.$$ 

Proof of MJL: Set $\delta = 1/N^2$ in DJL and $u$ as the difference vector of some pair of points. Union bound over the $\binom{N}{2}$ pairs.

Theorem (Alon, 2003)

For every $N$, there exists a set of $N$ points requiring target dimension $m = \Omega(\varepsilon^{-2} / \log(1/\varepsilon)) \log N)$.

Theorem (Jayram-Woodruff, 2011; Kane-Meka-N., 2011)

For DJL, $m = \Theta(\varepsilon^{-2} \log(1/\delta))$ is optimal.
Proving the distributional JL lemma

Older proofs

• [Johnson-Lindenstrauss, 1984], [Frankl-Maehara, 1988]: Random rotation, then projection onto first $m$ coordinates.

• [Indyk-Motwani, 1998], [Dasgupta-Gupta, 2003]: Random matrix with independent Gaussian entries.

• [Achlioptas, 2001]: Independent $\pm 1$ entries.

• [Clarkson-Woodruff, 2009]: $O(\log(1/\delta))$-wise independent $\pm 1$ entries.

• [Arriaga-Vempala, 1999], [Matousek, 2008]: Independent entries having mean 0, variance $1/m$, and subGaussian tails.
Proving the distributional JL lemma

Older proofs

- [Johnson-Lindenstrauss, 1984], [Frankl-Maehara, 1988]: Random rotation, then projection onto first $m$ coordinates.
- [Achlioptas, 2001]: Independent $\pm 1$ entries.
- [Clarkson-Woodruff, 2009]: $O(\log(1/\delta))$-wise independent $\pm 1$ entries.
- [Arriaga-Vempala, 1999], [Matousek, 2008]: Independent entries having mean 0, variance $1/m$, and subGaussian tails

**Downside:** Performing embedding is dense matrix-vector multiplication, $O(m \cdot \|x\|_0)$ time
Fast JL Transforms

- [Ailon-Chazelle, 2006]: $x \mapsto PHDx$, $O(n \log n + m^3)$ time
- $P$ random + sparse, $H$ Fourier, $D$ has random ±1 on diagonal
- Also follow-up works based on similar approach which improve the time while, for some, slightly increasing target dimension
- [Ailon, Liberty ’08], [Ailon, Liberty ’11], [Krahmer, Ward ’11], [N., Price, Wootters ’14], …
Fast JL Transforms

- [Ailon-Chazelle, 2006]: $x \mapsto PHDx$, $O(n \log n + m^3)$ time
  $P$ random+sparse, $H$ Fourier, $D$ has random $\pm 1$ on diagonal
- Also follow-up works based on similar approach which improve the time while, for some, slightly increasing target dimension
  [Ailon, Liberty ’08], [Ailon, Liberty ’11], [Krahmer, Ward ’11],
  [N., Price, Wootters ’14], …

**Downside:** Slow to embed sparse vectors: running time is
$\Omega(\min\{m \cdot \|x\|_0, n \log n\})$. 
Where Do Sparse Vectors Show Up?

- **Document as bag of words:** $u_i =$ number of occurrences of word $i$. Compare documents using cosine similarity.
  
  $n =$ lexicon size; most documents aren’t dictionaries

- **Network traffic:** $u_{i,j} =$ #bytes sent from $i$ to $j$
  
  $n = 2^{64} \ (2^{256} \text{ in I Pv6});$ most servers don’t talk to each other

- **User ratings:** $u_{i,j}$ is user $i$’s score for movie $j$ on Netflix
  
  $n =$ #movies; most people haven’t rated all movies

- **Streaming:** $u$ receives a stream of updates of the form: “add $v$ to $u_i$”. Maintaining $\Pi u$ requires calculating $v \cdot \Pi e_i$.

- ...
Sparse JL transforms
One way to embed sparse vectors faster: use sparse matrices.
Sparse JL transforms

One way to embed sparse vectors faster: use sparse matrices.

\[ s = \# \text{non-zero entries per column in } \Pi \]

(so embedding time is \( s \cdot \|x\|_0 \))

<table>
<thead>
<tr>
<th>reference</th>
<th>value of ( s )</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>[JL84], [FM88], [IM98], ...</td>
<td>( m \approx 4\varepsilon^{-2} \ln(1/\delta) )</td>
<td>dense</td>
</tr>
<tr>
<td>[Achlioptas01]</td>
<td>( m/3 )</td>
<td>sparse Bernoulli</td>
</tr>
<tr>
<td>[WDALS09]</td>
<td>no proof</td>
<td>hashing</td>
</tr>
<tr>
<td>[DKS10]</td>
<td>( \tilde{O}(\varepsilon^{-1} \log^3(1/\delta)) )</td>
<td>hashing</td>
</tr>
<tr>
<td>[KN10a], [BOR10]</td>
<td>( \tilde{O}(\varepsilon^{-1} \log^2(1/\delta)) )</td>
<td>&quot;</td>
</tr>
<tr>
<td>[KN12]</td>
<td>( O(\varepsilon^{-1} \log(1/\delta)) )</td>
<td>modified hashing</td>
</tr>
</tbody>
</table>
Sparse JL transforms

One way to embed sparse vectors faster: use sparse matrices.

\( s = \#\text{non-zero entries per column in } \Pi \)  
(so embedding time is \( s \cdot \|x\|_0 \))

<table>
<thead>
<tr>
<th>reference</th>
<th>value of ( s )</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>[JL84], [FM88], [IM98], ...</td>
<td>( m \approx 4\varepsilon^{-2} \ln(1/\delta) )</td>
<td>dense</td>
</tr>
<tr>
<td>[Achlioptas01]</td>
<td>( m/3 )</td>
<td>sparse Bernoulli</td>
</tr>
<tr>
<td>[WDALS09]</td>
<td>no proof</td>
<td>hashing</td>
</tr>
<tr>
<td>[DKS10]</td>
<td>( \tilde{O}(\varepsilon^{-1} \log^3(1/\delta)) )</td>
<td>hashing</td>
</tr>
<tr>
<td>[KN10a], [BOR10]</td>
<td>( \tilde{O}(\varepsilon^{-1} \log^2(1/\delta)) )</td>
<td>&quot;</td>
</tr>
<tr>
<td>[KN12]</td>
<td>( O(\varepsilon^{-1} \log(1/\delta)) )</td>
<td>modified hashing</td>
</tr>
</tbody>
</table>

[N., Nguyễn '13]: for any \( m \leq \text{poly}(1/\varepsilon) \cdot \log N \), \( s = \Omega(\varepsilon^{-1} \log N/ \log(1/\varepsilon)) \) is required, even for metric JL, so [KN12] is optimal up to \( O(\log(1/\varepsilon)) \).

*[Thorup, Zhang '04] gives \( m = O(\varepsilon^{-2} \delta^{-1}) \), \( s = 1 \).
Sparse JL Constructions

\[ s = \tilde{\Theta}(\varepsilon^{-1} \log^2(1/\delta)) \]
## Sparse JL Constructions

<table>
<thead>
<tr>
<th>[DKS, 2010]</th>
<th>$s = \tilde{\Theta}(e^{-1} \log^2(1/\delta))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[KN12]</td>
<td>$s = \Theta(e^{-1} \log(1/\delta))$</td>
</tr>
</tbody>
</table>
Sparse JL Constructions

<table>
<thead>
<tr>
<th>[DKS, 2010]</th>
<th>[s = \tilde{\Theta}(\varepsilon^{-1} \log^2(1/\delta))]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[KN12]</td>
<td>[s = \Theta(\varepsilon^{-1} \log(1/\delta))]</td>
</tr>
<tr>
<td>[KN12]</td>
<td>[s = \Theta(\varepsilon^{-1} \log(1/\delta))]</td>
</tr>
</tbody>
</table>
Sparse JL Constructions (in matrix form)

Each black cell is $\pm 1/\sqrt{s}$ at random
Analysis

- In both constructions, can write $\Pi_{i,j} = \delta_{i,j} \sigma_{i,j} / \sqrt{s}$

\[
\|\Pi u\|_2^2 - 1 = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u_i u_j = \sigma^T B \sigma
\]
Analysis

• In both constructions, can write $\Pi_{i,j} = \delta_{i,j} \sigma_{i,j} / \sqrt{s}$

$$\|\Pi u\|^2 - 1 = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u_i u_j = \sigma^T B \sigma$$

$$B = \frac{1}{s} \cdot \begin{bmatrix}
B_1 & 0 & \ldots & 0 \\
0 & B_2 & \ldots & 0 \\
0 & 0 & \ddots & 0 \\
0 & \ldots & 0 & B_m
\end{bmatrix}$$

• $(B_r)_{i,j} = \delta_{r,i} \delta_{r,j} x_i x_j$
Analysis

- In both constructions, can write $\Pi_{i,j} = \delta_{i,j} \sigma_{i,j} / \sqrt{s}$

$$\|\Pi u\|_2^2 - 1 = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u_i u_j = \sigma^T B \sigma$$

$$B = \frac{1}{s} \cdot \begin{bmatrix}
B_1 & 0 & \ldots & 0 \\
0 & B_2 & \ldots & 0 \\
0 & 0 & \ddots & 0 \\
0 & \ldots & 0 & B_m
\end{bmatrix}$$

- $(B_r)_{i,j} = \delta_{r,i} \delta_{r,j} x_i x_j$

- $P(\|\|\Pi u\|_2^2 - 1\| > \varepsilon) < \varepsilon^{-\ell} \cdot E \|\|\Pi u\|_2^2 - 1\|^{\ell}$. Use moment bound for quadratic forms, which depends on $\|B\|$, $\|B\|_F$ (Hanson-Wright inequality).
What next?
Natural “matrix extension” of sparse JL

[Kane, N. ’12]

**Theorem**

Let \( u \in \mathbb{R}^n \) be arbitrary, unit \( \ell_2 \) norm, \( \Pi \) sparse sign matrix. Then

\[
\mathbb{P} \left( \| \Pi u \|^2 - 1 \geq \varepsilon \right) < \delta
\]

as long as

\[
m \gtrsim \frac{\log(1/\delta)}{\varepsilon^2}, \quad s \gtrsim \frac{\log(1/\delta)}{\varepsilon}, \quad \ell = \log(1/\delta)
\]

or

\[
m \gtrsim \frac{1}{\varepsilon^2 \delta}, \quad s = 1, \quad \ell = 2 \quad ([\text{Thorup, Zhang’04}])
\]
Natural “matrix extension” of sparse JL

[Kane, N. ’12]

**Theorem**

Let $u \in \mathbb{R}^{n \times 1}$ be arbitrary, o.n. cols, $\Pi$ sparse sign matrix. Then

$$
\mathbb{P}(\|\Pi(\Pi u)^T(\Pi u) - l_1\| > \varepsilon) < \delta
$$

as long as

$$
m \gtrsim \frac{1 + \log(1/\delta)}{\varepsilon^2}, \ s \gtrsim \frac{\log(1/\delta)}{\varepsilon}, \ l = \log(1/\delta)
$$

or

$$
m \gtrsim \frac{1^2}{\varepsilon^2 \delta}, \ s = 1, \ l = 2
$$
Conjecture

Theorem

Let $u \in \mathbb{R}^{n \times d}$ be arbitrary, o.n. cols, $\Pi$ sparse sign matrix. Then

$$
\mathbb{P}(\|(\Pi u)^T (\Pi u) - I_d\| > \varepsilon) < \delta
$$

as long as

$$
m \gtrsim \frac{d + \log(1/\delta)}{\varepsilon^2}, \ s \gtrsim \frac{\log(d/\delta)}{\varepsilon}, \ \ell = \log(d/\delta)
$$

or

$$
m \gtrsim \frac{d^2}{\varepsilon^2 \delta}, \ s = 1, \ \ell = 2
$$
Natural “matrix extension” of sparse JL

What we prove [N., Nguyễn ’13]

Theorem

Let $u \in \mathbb{R}^{n \times d}$ be arbitrary, o.n. cols, $\Pi$ sparse sign matrix. Then

$$\mathbb{P}(\|\Pi (\Pi u)^T (\Pi u) - I_d\| > \varepsilon) < \delta$$

as long as

$$m \gtrsim \frac{d \cdot \log^c(d/\delta)}{\varepsilon^2}, s \gtrsim \frac{\log^c(d/\delta)}{\varepsilon} \quad \text{or} \quad m \gtrsim \frac{d^{1.01}}{\varepsilon^2}, s \gtrsim \frac{1}{\varepsilon}$$

or

$$m \gtrsim \frac{d^2}{\varepsilon^2 \delta}, s = 1$$
Remarks

• [Clarkson, Woodruff ’13] was first to show
  \( m = d^2 \cdot \text{polylog}(d/\varepsilon), s = 1 \) bound via other methods

• \( m = O(d^2/\varepsilon^2), s = 1 \) also obtained by [Mahoney, Meng ’13].

• \( m = O(d^2/\varepsilon^2), s = 1 \) also follows from [Thorup, Zhang ’04] + [Kane, N. ’12] (observed by Nguyễn)
Remarks

- [Clarkson, Woodruff ’13] was first to show
  $m = d^2 \cdot \text{polylog}(d/\varepsilon), s = 1$ bound via other methods
- $m = O(d^2/\varepsilon^2), s = 1$ also obtained by [Mahoney, Meng ’13].
- $m = O(d^2/\varepsilon^2), s = 1$ also follows from [Thorup, Zhang ’04] + [Kane, N. ’12] (observed by Nguyên)
- What does the “moment method” mean for matrices?

$$\mathbb{P}(\|\Pi \mathbf{u}^T \mathbf{u} - I_d\| > \varepsilon) < \varepsilon^{-\ell} \cdot \mathbb{E} \|\Pi \mathbf{u}^T \mathbf{u} - I_d\|^\ell$$

$$\leq \varepsilon^{-\ell} \cdot \mathbb{E} \text{tr}(((\Pi \mathbf{u})^T (\Pi \mathbf{u}) - I_d)^\ell)$$

- Classical “moment method” in random matrix theory; e.g. [Wigner, 1955], [Füredi, Komlós, 1981], [Bai, Yin, 1993]
Who cares about this matrix extension?
Motivation for matrix extension of sparse JL

- $\|(\Pi U)^T (\Pi U) - I\| \leq \varepsilon$ equivalent to $\|\Pi x\| = (1 \pm \varepsilon)\|x\|$ for all $x \in V$, where $V$ is the subspace spanned by the columns of $U$ (up to changing $\varepsilon$ by a factor of 2). “subspace embedding”.

- Subspace embeddings can be used to speed up algorithms for many numerical linear algebra problems on big matrices [Sarlos, 2006], [Dasgupta, Drineas, Harb, Kumar, Mahoney, 2008], [Clarkson, Woodruff, 2009], [Drineas, Magdon-Ismail, Mahoney, Woodruff, 2012], [Clarkson, Woodruff, 2013], [Clarkson, Drineas, Magdon-Ismail, Mahoney, Meng, Woodruff, 2013], [Woodruff, Zhang, 2013], ...

- Sparse $\Pi$: can multiply $\Pi A$ in $s \cdot \text{nnz}(A)$ time for big matrix $A$. 
Motivation for matrix extension of sparse JL

- $\|(\Pi U)^T (\Pi U) - I\| \leq \varepsilon$ equivalent to $\|\Pi x\| = (1 \pm \varepsilon)\|x\|$ for all $x \in V$, where $V$ is the subspace spanned by the columns of $U$ (up to changing $\varepsilon$ by a factor of 2). “subspace embedding”.

- Subspace embeddings can be used to speed up algorithms for many numerical linear algebra problems on big matrices [Sarlós, 2006], [Dasgupta, Drineas, Harb, Kumar, Mahoney, 2008], [Clarkson, Woodruff, 2009], [Drineas, Magdon-Ismail, Mahoney, Woodruff, 2012], [Clarkson, Woodruff, 2013], [Clarkson, Drineas, Magdon-Ismail, Mahoney, Meng, Woodruff, 2013], [Woodruff, Zhang, 2013], ...
Motivation for matrix extension of sparse JL

- $\|(\Pi U)^T(\Pi U) - I\| \leq \varepsilon$ equivalent to $\|\Pi x\| = (1 \pm \varepsilon)\|x\|$ for all $x \in V$, where $V$ is the subspace spanned by the columns of $U$ (up to changing $\varepsilon$ by a factor of 2). “subspace embedding”.

- Subspace embeddings can be used to speed up algorithms for many numerical linear algebra problems on big matrices [Sarlós, 2006], [Dasgupta, Drineas, Harb, Kumar, Mahoney, 2008], [Clarkson, Woodruff, 2009], [Drineas, Magdon-Ismail, Mahoney, Woodruff, 2012], [Clarkson, Woodruff, 2013], [Clarkson, Drineas, Magdon-Ismail, Mahoney, Meng, Woodruff, 2013], [Woodruff, Zhang, 2013], ... 

- Sparse $\Pi$: can multiply $\Pi A$ in $s \cdot \text{nnz}(A)$ time for big matrix $A$. 

Numerical linear algebra

- $A \in \mathbb{R}^{n \times d}$, $n \gg d$, $\text{rank}(A) = r$
Numerical linear algebra

- $A \in \mathbb{R}^{n \times d}$, $n \gg d$, $\text{rank}(A) = r$

Classical numerical linear algebra problems

- Compute the leverage scores of $A$, i.e. the $\ell_2$ norms of the $n$ standard basis vectors when projected onto the subspace spanned by the columns of $A$. 
Numerical linear algebra

\( A \in \mathbb{R}^{n \times d}, \ n \gg d, \ \text{rank}(A) = r \)

Classical numerical linear algebra problems

- Compute the **leverage scores** of \( A \), i.e. the \( \ell_2 \) norms of the \( n \) standard basis vectors when projected onto the subspace spanned by the columns of \( A \).
- **Least squares regression**: Given also \( b \in \mathbb{R}^n \).
  
  Compute \( x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_2 \)

- **Low-rank approximation**: Given also an integer \( 1 \leq k \leq d \).
  
  Compute \( A_k = \arg\min_{\text{rank}(B) \leq k} \|A - B\|_F \)

- **Preconditioning**: Compute \( R \in \mathbb{R}^{d \times d} \) (for \( d = r \)) so that \( \forall x \|ARx\|_2 \approx \|x\|_2 \)
Numerical linear algebra

- $A \in \mathbb{R}^{n \times d}$, $n \gg d$, $\text{rank}(A) = r$

Classical numerical linear algebra problems

- Compute the leverage scores of $A$, i.e. the $\ell_2$ norms of the $n$ standard basis vectors when projected onto the subspace spanned by the columns of $A$.

- **Least squares regression**: Given also $b \in \mathbb{R}^n$.
  
  Compute $x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$

- **$\ell_p$ regression** ($p \in [1, \infty)$):
  
  Compute $x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_p$
Numerical linear algebra

• $A \in \mathbb{R}^{n \times d}$, $n \gg d$, $\text{rank}(A) = r$

Classical numerical linear algebra problems

• Compute the **leverage scores** of $A$, i.e. the $\ell_2$ norms of the $n$ standard basis vectors when projected onto the subspace spanned by the columns of $A$.

• **Least squares regression**: Given also $b \in \mathbb{R}^n$.

  Compute $x^* = \text{argmin}_{x \in \mathbb{R}^d} \|Ax - b\|_2$

• **$\ell_p$ regression ($p \in [1, \infty)$)**:

  Compute $x^* = \text{argmin}_{x \in \mathbb{R}^d} \|Ax - b\|_p$

• **Low-rank approximation**: Given also an integer $1 \leq k \leq d$.

  Compute $A_k = \text{argmin}_{\text{rank}(B) \leq k} \|A - B\|_F$
Numerical linear algebra

- $A \in \mathbb{R}^{n \times d}$, $n \gg d$, $\text{rank}(A) = r$

Classical numerical linear algebra problems

- Compute the leverage scores of $A$, i.e. the $\ell_2$ norms of the $n$ standard basis vectors when projected onto the subspace spanned by the columns of $A$.

- **Least squares regression**: Given also $b \in \mathbb{R}^n$.
  \[
  \text{Compute } x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_2
  \]

- **$\ell_p$ regression ($p \in [1, \infty)$)**:
  \[
  \text{Compute } x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_p
  \]

- **Low-rank approximation**: Given also an integer $1 \leq k \leq d$.
  \[
  \text{Compute } A_k = \arg\min_{\text{rank}(B) \leq k} \|A - B\|_F
  \]

- **Preconditioning**: Compute $R \in \mathbb{R}^{d \times d}$ (for $d = r$) so that
  \[
  \forall x \quad \|ARx\|_2 \approx \|x\|_2
  \]
Singular Value Decomposition

Theorem

Every matrix $A \in \mathbb{R}^{n \times d}$ of rank $r$ can be written as

$$A = U \Sigma V^T$$

where $U$ has orthonormal columns in $\mathbb{R}^{n \times r}$, $\Sigma$ is a diagonal positive definite matrix in $\mathbb{R}^{r \times r}$, and $V^T$ has orthonormal columns in $\mathbb{R}^{d \times r}$.

Can compute SVD in $\tilde{O}(nd^{\omega-1})$ [Demmel, Dumitriu, Holtz, 2007].

$\omega < 2.373\ldots$ is the exponent of square matrix multiplication [Coppersmith, Winograd, 1987], [Stothers, 2010], [Vassilevska-Williams, 2012]
Computationally efficient solutions

\[ A = \begin{pmatrix} U \\ \Sigma \\ V^T \end{pmatrix} \]

- **Leverage scores**: Output row norms of \( U \).
- **Least squares regression**: Output \( V \Sigma^{-1} U^T b \).
- **Low-rank approximation**: Output \( U \Sigma_k V^T \).
- **Preconditioning**: Output \( R = V \Sigma^{-1} \).
Computationally efficient solutions

\[ A = \begin{bmatrix} U & \Sigma & V^T \end{bmatrix} \]

- **Leverage scores**: Output row norms of \( U \).
- **Least squares regression**: Output \( V\Sigma^{-1}U^Tb \).
- **Low-rank approximation**: Output \( U\Sigma_kV^T \).
- **Preconditioning**: Output \( R = V\Sigma^{-1} \).

**Conclusion**: In time \( \tilde{O}(nd^{\omega-1}) \) we can compute the SVD then solve all the previously stated problems. Is there a faster way?
Least squares regression: Let $\Pi$ be a subspace embedding for the subspace spanned by $b$ and the columns of $A$. Let $x^* = \arg\min ||Ax - b||$ and $\tilde{x} = \arg\min ||\Pi Ax - \Pi b||$. Then
Least squares regression: Let $\Pi$ be a subspace embedding for the subspace spanned by $b$ and the columns of $A$. Let $x^* = \arg\min ||Ax - b||$ and $\tilde{x} = \arg\min ||\Pi Ax - \Pi b||$. Then

$$||\Pi A\tilde{x} - \Pi b|| \leq ||\Pi Ax^* - \Pi b||$$
How to use subspace embeddings

**Least squares regression**: Let \( \Pi \) be a subspace embedding for the subspace spanned by \( b \) and the columns of \( A \). Let \( x^* = \arg\min \|Ax - b\| \) and \( \tilde{x} = \arg\min \|\Pi Ax - \Pi b\| \). Then

\[
(1 - \varepsilon) \|A\tilde{x} - b\| \leq \|\Pi A\tilde{x} - \Pi b\| \leq \|\Pi Ax^* - \Pi b\| \leq \|\Pi(A\tilde{x} - b)\|
\]
Least squares regression: Let $\Pi$ be a subspace embedding for the subspace spanned by $b$ and the columns of $A$. Let $x^* = \arg\min ||Ax - b||$ and $\tilde{x} = \arg\min ||\Pi Ax - \Pi b||$. Then

$$(1-\varepsilon)||A\tilde{x} - b|| \leq ||\Pi A\tilde{x} - \Pi b|| \leq ||\Pi Ax^* - \Pi b|| \leq (1+\varepsilon)||Ax^* - b||$$

$$\Rightarrow ||A\tilde{x} - b|| \leq \left(\frac{1+\varepsilon}{1-\varepsilon}\right) \cdot ||Ax^* - b||$$
Least squares regression: Let $\Pi$ be a subspace embedding for the subspace spanned by $b$ and the columns of $A$. Let $x^* = \text{argmin} \|Ax - b\|$ and $\tilde{x} = \text{argmin} \|\Pi Ax - \Pi b\|$. Then

$$(1 - \varepsilon)\|A\tilde{x} - b\| \leq \|\Pi A\tilde{x} - \Pi b\| \leq \|\Pi Ax^* - \Pi b\| \leq (1 + \varepsilon)\|Ax^* - b\|$$

$$\Rightarrow \|A\tilde{x} - b\| \leq \left(\frac{1 + \varepsilon}{1 - \varepsilon}\right) \cdot \|Ax^* - b\|$$

Computing SVD of $\Pi A$ takes time $\tilde{O}(md^{\omega-1})$, which is much faster than $\tilde{O}(nd^{\omega-1})$ since $m \ll n$. 

How to use subspace embeddings
Back to the analysis

$$\mathbb{P} \left( \| (\prod U)^T (\prod U) - I_d \| > \varepsilon \right) < \varepsilon^{-\ell} \cdot \mathbb{E} \text{tr}( ((\prod U)^T (\prod U) - I_d)^\ell )$$
Analysis ($\ell = 2$)

$s = 1, \ m = O(d^2/\varepsilon^2)$

Want to understand $S - I, \ S = (\Pi U)^T(\Pi U)$
Analysis ($\ell = 2$)

$s = 1, \ m = O(d^2/\varepsilon^2)$

Want to understand $S - I, \ S = (\Pi U)^T(\Pi U)$

Let the columns of $U$ be $u^1, \ldots, u^d$
Recall $\Pi_{i,j} = \delta_{i,j}\sigma_{i,j}/\sqrt{s}$
Analysis \((\ell = 2)\)
\[s = 1, \; m = O(d^2/\varepsilon^2)\]

Want to understand \(S - I, \; S = (\Pi U)^T(\Pi U)\)

Let the columns of \(U\) be \(u^1, \ldots, u^d\)
Recall \(\Pi_{i,j} = \delta_{i,j}\sigma_{i,j}/\sqrt{s}\)

Some computations yield

\[
(S - I)_{k,k'} = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i}\delta_{r,j}\sigma_{r,i}\sigma_{r,j} u_i^k u_j^{k'}
\]
Analysis ($\ell = 2$)

$s = 1$, $m = O(d^2/\varepsilon^2)$

Want to understand $S - I$, $S = (\prod U)^T(\prod U)$

Let the columns of $U$ be $u^1, \ldots, u^d$

Recall $\Pi_{i,j} = \delta_{i,j}\sigma_{i,j}/\sqrt{s}$

Some computations yield

\[
(S - I)_{k,k'} = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_r,i \delta_r,j \sigma_r,i \sigma_r,j u^k_i u^{k'}_j
\]

Computing $\mathbb{E} \text{tr}((S - I)^2) = \mathbb{E} \|S - I\|_F^2$ is straightforward, and can show $\mathbb{E} \|S - I\|_F^2 \leq (d^2 + d)/m$

\[
\mathbb{P}(\|S - I\| > \varepsilon) < \frac{1}{\varepsilon^2} \frac{d^2 + d}{m}
\]
Analysis ($\ell = 2$)

$s = 1$, $m = O(d^2/\varepsilon^2)$

Want to understand $S - I$, $S = (\Pi U)^T(\Pi U)$

Let the columns of $U$ be $u^1, \ldots, u^d$

Recall $\Pi_{i,j} = \delta_{i,j} \sigma_{i,j}/\sqrt{s}$

Some computations yield

$$(S - I)_{k,k'} = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u^k_i u^{k'}_j$$

Computing $\mathbb{E} \text{tr}((S - I)^2) = \mathbb{E} \|S - I\|_F^2$ is straightforward, and can show $\mathbb{E} \|S - I\|_F^2 \leq (d^2 + d)/m$

$$\mathbb{P}(\|S - I\| > \varepsilon) < \frac{1}{\varepsilon^2} \frac{d^2 + d}{m}$$

Set $m \geq \delta^{-1}(d^2 + d)/\varepsilon^2$ for success probability $1 - \delta$
Analysis (large $\ell$)

$s = O(\gamma(1/\varepsilon))$, $m = O(d^{1+\gamma}/\varepsilon^2)$

$$(S - I)_{k,k'} = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u^k_i u^{k'}_j$$
Analysis (large \( \ell \))

\[
s = O_\gamma(1/\varepsilon), \quad m = O(d^{1+\gamma}/\varepsilon^2)
\]

\[(S - I)_{k,k'} = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u_i^k u_j^{k'}
\]

By induction, for any square matrix \( B \) and integer \( \ell \geq 1 \),

\[
(B^\ell)_{i,j} = \sum_{i_1,\ldots,i_{\ell+1}} \prod_{t=1}^{\ell} B_{i_t,i_{t+1}}
\]
Analysis (large $\ell$)

\[ s = O_\gamma(1/\varepsilon), \quad m = O(d^{1+\gamma}/\varepsilon^2) \]

\[(S - I)_{k,k'} = \frac{1}{s} \sum_{r=1}^{m} \sum_{i \neq j} \delta_{r,i} \delta_{r,j} \sigma_{r,i} \sigma_{r,j} u_i^k u_j^{k'} \]

By induction, for any square matrix $B$ and integer $\ell \geq 1$,

\[ (B^\ell)_{i,j} = \sum_{i_1, \ldots, i_{\ell+1}} \prod_{t=1}^{\ell} B_{i_t, i_{t+1}} \]

\[ \Rightarrow \text{tr}(B^\ell) = \sum_{i_1, \ldots, i_{\ell+1}} \prod_{t=1}^{\ell} B_{i_t, i_{t+1}} \]
Analysis (large $\ell$)

\[ s = O_\gamma(1/\varepsilon), \ m = O(d^{1+\gamma}/\varepsilon^2) \]

\[ \mathbb{E} \text{tr}((S-I)^\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell} \left( \mathbb{E} \prod_{t=1}^{\ell} \delta_{r_t,i_t} \delta_{r_t,j_t} \right) \left( \mathbb{E} \prod_{t=1}^{\ell} \sigma_{r_t,i_t} \sigma_{r_t,j_t} \right) \prod_{t=1}^{\ell} u_{i_t}^{k_t} u_{j_t}^{k_{t+1}} \]
Analysis (large $\ell$)

$s = O_\gamma(1/\varepsilon), \ m = O(d^{1+\gamma}/\varepsilon^2)$

\[
\mathbb{E} \text{tr}((S - I)^\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell} \left( \mathbb{E} \prod_{t=1}^{\ell} \delta_{r_t,i_t} \delta_{r_t,j_t} \right) \left( \mathbb{E} \prod_{t=1}^{\ell} \sigma_{r_t,i_t} \sigma_{r_t,j_t} \right) \prod_{t=1}^{\ell} u_{i_t}^{k_t} u_{j_t}^{k_t+1}
\]

**The strategy**: Associate each monomial in summation above with a graph, group monomials that have the same graph, and estimate the contribution of each graph then do some combinatorics

(a common strategy; see [Wigner, 1955], [Füredi, Komlós, 1981], [Bai, Yin, 1993])
Example monomial $\rightarrow$ graph correspondence

$$\text{tr}((S - I)^\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell} \prod_{t=1}^{\ell} \delta_{r_t, i_t} \delta_{r_t, j_t} \cdot \prod_{t=1}^{\ell} \sigma_{r_t, i_t} \sigma_{r_t, j_t} \cdot \prod_{t=1}^{\ell} u_{i_t}^{k_t} u_{j_t}^{k_t+1}$$

$\ell = 4$

$\delta_{r_e, i_a} \delta_{r_e, i_b} \sigma_{r_e, i_a} \sigma_{r_e, i_b} u_{i_a}^{k_1} u_{i_b}^{k_2}$
Example monomial → graph correspondence

\[
\text{tr}((S - I)^\ell) = \sum \prod_{t=1}^{\ell} \delta_{r_t, i_t} \delta_{r_t, j_t} \cdot \prod_{t=1}^{\ell} \sigma_{r_t, i_t} \sigma_{r_t, j_t} \cdot \prod_{t=1}^{\ell} u_{i_t}^{k_t} u_{j_t}^{k_{t+1}}
\]

\[
\ell = 4
\]
Example monomial $\rightarrow$ graph correspondence

$$\text{tr}((S - I)^\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell, r_1, \ldots, r_\ell} \prod_{t=1}^\ell \delta_{r_t, i_t} \delta_{r_t, j_t} \cdot \prod_{t=1}^\ell \sigma_{r_t, i_t} \sigma_{r_t, j_t} \cdot \prod_{t=1}^\ell u_{i_t}^{k_t} u_{j_t}^{k_{t+1}} \cdot \prod_{t=1}^\ell \delta_{r_t, i_t} \delta_{r_t, j_t} \cdot \prod_{t=1}^\ell \sigma_{r_t, i_t} \sigma_{r_t, j_t} \cdot \prod_{t=1}^\ell u_{i_t}^{k_t} u_{j_t}^{k_{t+1}}$$

$$\ell = 4$$

$\delta_{r_t, i_t} \delta_{r_t, j_t}$ and $\sigma_{r_t, i_t} \sigma_{r_t, j_t}$ represent the edge between nodes $i_t$ and $j_t$.

$u_{i_t}^{k_t} u_{j_t}^{k_{t+1}}$ represents the weight of the edge.

The figure on the right shows a graph with nodes labeled $k, i, j, r, a, c, b, e, d, f$. The edges are labeled with numbers: 9, 10, 11, and 12.
Example monomial $\rightarrow$ graph correspondence

\[
\text{tr}((S - I)\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell} \prod_{t=1}^{\ell} \delta_{r_t, i_t} \delta_{r_t, j_t} \cdot \prod_{t=1}^{\ell} \sigma_{r_t, i_t} \sigma_{r_t, j_t} \cdot \prod_{t=1}^{\ell} u_{i_t}^{k_t} u_{j_t}^{k_{t+1}}
\]

$\ell = 4$

\[
\times \delta_{r_f, i_c} \delta_{r_f, i_d} \sigma_{r_f, i_c} \sigma_{r_f, i_d} u_{i_c}^{k_4} u_{i_d}^{k_1}
\]
Example monomial $\rightarrow$ graph correspondence

\[
\text{tr}((S - I)^\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell} \prod_{t=1}^\ell \delta_{r_t, i_t} \delta_{r_t, j_t} \cdot \prod_{t=1}^\ell \sigma_{r_t, i_t} \sigma_{r_t, j_t} \cdot \prod_{t=1}^\ell u_{i_t}^{k_t} u_{j_t}^{k_{t+1}}
\]

$\ell = 4$

\[
\begin{align*}
\delta_{r_e, i_a} & \delta_{r_e, i_b} \sigma_{r_e, i_a} \sigma_{r_e, i_b} u_{i_a}^{k_1} u_{i_b}^{k_2} \\
\times \delta_{r_e, i_a} & \delta_{r_e, i_b} \sigma_{r_e, i_a} \sigma_{r_e, i_b} u_{i_a}^{k_2} u_{i_b}^{k_3} \\
\times \delta_{r_f, i_c} & \delta_{r_f, i_d} \sigma_{r_f, i_c} \sigma_{r_f, i_d} u_{i_c}^{k_3} u_{i_d}^{k_4} \\
\times \delta_{r_f, i_c} & \delta_{r_f, i_d} \sigma_{r_f, i_c} \sigma_{r_f, i_d} u_{i_c}^{k_4} u_{i_d}^{k_1}
\end{align*}
\]
Example monomial → graph correspondence

\[ \text{tr}((S - I)^\ell) = \sum_{i_1 \neq j_1, \ldots, i_\ell \neq j_\ell} \prod_{t=1}^\ell \delta_{r_t,i_t} \delta_{r_t,j_t} \cdot \prod_{t=1}^\ell \sigma_{r_t,i_t} \sigma_{r_t,j_t} \cdot \prod_{t=1}^\ell \langle u_{i_t}, u_{i_{t+1}} \rangle \]

\[ \ell = 4 \]

\[ \delta_{r_e,i_a} \delta_{r_e,i_b} \sigma_{r_e,i_a} \sigma_{r_e,i_b} u_{i_a}^{k_1} u_{i_b}^{k_2} \]
\[ \times \delta_{r_e,i_a} \delta_{r_e,i_b} \sigma_{r_e,i_a} \sigma_{r_e,i_b} u_{i_a}^{k_2} u_{i_b}^{k_3} \]
\[ \times \delta_{r_f,i_c} \delta_{r_f,i_d} \sigma_{r_f,i_c} \sigma_{r_f,i_d} u_{i_c}^{k_3} u_{i_d}^{k_4} \]
\[ \times \delta_{r_f,i_c} \delta_{r_f,i_d} \sigma_{r_f,i_c} \sigma_{r_f,i_d} u_{i_c}^{k_4} u_{i_d}^{k_1} \]
Grouping monomials by graph

$z$ right vertices, $b$ distinct edges between middle and right

\[
\mathbb{E} \text{tr}((S - I)^\ell) = \sum_{i_1 \neq i_1, \ldots, i_\ell \neq j_\ell} \left( \mathbb{E} \prod_{t=1}^\ell \delta_{r_t, i_t} \delta_{r_t, j_t} \right) \left( \mathbb{E} \prod_{t=1}^\ell \sigma_{r_t, i_t} \sigma_{r_t, j_t} \right) \prod_{t=1}^\ell \langle u_{i_t}, u_{i_{t+1}} \rangle
\]

\[
\leq \sum_{G} m^z \left( \frac{s}{m} \right)^b \left| \sum_{i_1 \neq \ldots \neq i_y} \prod_{e=(\alpha, \beta) \in \hat{G}} \langle u_{i_\alpha}, u_{i_\beta} \rangle \right|
\]

<table>
<thead>
<tr>
<th>$G$</th>
<th>$\hat{G}$</th>
</tr>
</thead>
</table>
| \begin{tikzpicture}
    \node[k] at (0,0) (k) {k};
    \node[i,j] at (1,1) (ij) {i,j};
    \node[r] at (2,2) (r) {r};
    \node[1] at (0,1) (1) {1};
    \node[5] at (0,2) (5) {5};
    \node[6] at (1,2) (6) {6};
    \node[2] at (1,1) (2) {2};
    \node[3] at (1,0) (3) {3};
    \node[4] at (0,0) (4) {4};
    \node[7] at (2,1) (7) {7};
    \node[e] at (2,0) (e) {e};
    \node[8] at (0,-1) (8) {8};
    \node[9] at (-1,-1) (9) {9};
    \node[10] at (-1,-2) (10) {10};
    \node[c] at (-1,-1) (c) {c};
    \node[11] at (-1,-2) (11) {11};
    \node[12] at (-2,-2) (12) {12};
    \node[d] at (1,-2) (d) {d};
    \node[13] at (-1,-3) (13) {13};
    \node[14] at (-2,-3) (14) {14};
    \node[f] at (-2,-2) (f) {f};
    \node[15] at (-2,-3) (15) {15};
    \node[a] at (3,1) (a) {a};
    \node[b] at (3,0) (b) {b};
    \node[c] at (2,-1) (c) {c};
    \node[d] at (2,-2) (d) {d};
\end{tikzpicture} | \begin{tikzpicture}
    \node[k] at (0,0) (k);\node[i,j] at (1,1) (ij);\node[r] at (2,2) (r);\node[a] at (3,1) (a);\node[b] at (3,0) (b);\node[c] at (2,-1) (c);\node[d] at (2,-2) (d);\end{tikzpicture} |
Let $C$ be the number of connected components of $\hat{G}$. It turns out the right upper bound for $F(\hat{G})$ is roughly $d^C$. 

$$F(\hat{G}) = \left| \sum \prod_{i_1 \neq \ldots \neq i_y} \langle u_{i_\alpha}, u_{i_\beta} \rangle \right|$$
Understanding $\hat{G}$

\[ F(\hat{G}) = \left| \sum_{i_1 \neq \ldots \neq i_y} \prod_{e=\langle \alpha, \beta \rangle \in \hat{G}} \langle u_{i_\alpha}, u_{i_\beta} \rangle \right| \]

Let $C$ be the number of connected components of $\hat{G}$. It turns out the right upper bound for $F(\hat{G})$ is roughly $d^C$

- Can get $d^C$ bound if all edges in $\hat{G}$ have even multiplicity
Understanding $\hat{G}$

\[ F(\hat{G}) = \left| \sum_{i_1 \neq \ldots \neq i_y} \prod_{e=(\alpha, \beta) \in \hat{G}} \langle u_{i_\alpha}, u_{i_\beta} \rangle \right| \]

Let $C$ be the number of connected components of $\hat{G}$. It turns out the right upper bound for $F(\hat{G})$ is roughly $d^C$

- Can get $d^C$ bound if all edges in $\hat{G}$ have even multiplicity
- How about $\hat{G}$ where this isn’t the case, e.g. as above?
Bounding $F(\hat{G})$ with odd multiplicities

$$= \frac{1}{2} \cdot \left[ \right]$$

Reduces back to case of even edge multiplicities! (AM-GM)
Bounding $F(\hat{G})$ with odd multiplicities

\[
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}

= \quad \begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array} \times \begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}
\begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array}

\leq \frac{1}{2} \cdot \left[ \begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array} + \begin{array}{c}
\text{a} \\
\text{d} \\
\text{b} \\
\text{c}
\end{array} \right]
\]

Reduces back to case of even edge multiplicities! (AM-GM)

Caveat: # connected components increased (unacceptable)
AM-GM trick done right

Theorem (Tutte '61, Nash-Williams '61)

Let $G$ be a multigraph with edge-connectivity at least $2k$. Then $G$ must have at least $k$ edge-disjoint spanning trees.

• If every connected component (CC) of $\hat{G}$ has 2 edge-disjoint spanning trees, we are done
• Otherwise, some CC is not 4 edge-connected. Since each CC is Eulerian, there must be a cut of size 2
Theorem (Tutte ’61, Nash-Williams ’61)

Let $G$ be a multigraph with edge-connectivity at least $2k$. Then $G$ must have at least $k$ edge-disjoint spanning trees.

Using the theorem ($k = 2$)

- If every connected component (CC) of $\hat{G}$ has 2 edge-disjoint spanning trees, we are done
Theorem (Tutte '61, Nash-Williams '61)

Let $G$ be a multigraph with edge-connectivity at least $2k$. Then $G$ must have at least $k$ edge-disjoint spanning trees.

Using the theorem ($k = 2$)

- If every connected component (CC) of $\hat{G}$ has 2 edge-disjoint spanning trees, we are done
- Otherwise, some CC is not 4 edge-connected. Since each CC is Eulerian, there must be a cut of size 2
AM-GM trick done right

\[
\sum_{i_{v} \in T} \left( \prod_{(q, r) \in T} \left\langle u_{iq}, u_{ir} \right\rangle \right) u_{ic}^{T} \left( \sum_{i_{v} \in \bar{T}} u_{ia} \left( \prod_{(q, r) \in \bar{T}} \left\langle u_{iq}, u_{ir} \right\rangle \right) u_{ib}^{T} \right) u_{id}^{T}
\]
AM-GM trick done right

\[
\sum_{i_v \in T} \left( \prod_{(q,r) \in T} \langle u_{i_q}, u_{i_r} \rangle \right) u_{ic} \left( \sum_{i_v \in \bar{T}} \left( \prod_{(q,r) \in \bar{T}} \langle u_{i_q}, u_{i_r} \rangle \right) u_{ib} \right) u_{id}
\]

- Repeatedly eliminate size-2 cuts until every connected component has two edge-disjoint spanning trees
- Show all $M$’s along the way have bounded operator norm
- Show that even edge multiplicities are still possible to handle when all $M$’s have bounded operator norm
Conclusion
Other recent progress

- Can show any oblivious subspace embedding succeeding with probability $\geq 2/3$ must have $\Omega(d/\varepsilon^2)$ rows [N., Nguyên]
Other recent progress

- Can show any oblivious subspace embedding succeeding with probability $\geq 2/3$ must have $\Omega(d/\varepsilon^2)$ rows [N., Nguyễn]
- Can show any oblivious subspace embedding with $O(d^{1+\gamma})$ rows must have sparsity $s = \Omega(1/(\varepsilon\gamma))$* [N., Nguyề́n]
Other recent progress

• Can show any oblivious subspace embedding succeeding with probability $\geq 2/3$ must have $\Omega(d/\varepsilon^2)$ rows [N., Nguyêñ]

• Can show any oblivious subspace embedding with $O(d^{1+\gamma})$ rows must have sparsity $s = \Omega(1/(\varepsilon\gamma))$* [N., Nguyêñ]

• Can provide upper bounds on $m, s$ to preserve an arbitrary bounded set $T \subset \mathbb{R}^n$, in terms of the geometry of $T$, in the style of [Gordon ’88], [Klartag, Mendelson ’05], [Mendelson, Pajor, Tomczak-Jaegermann ’07], [Dirksen ’13] (in the current notation, those works analyzed dense $\Pi$, i.e. $m = s$) [Bourgain, N.]
Other recent progress

- Can show any oblivious subspace embedding succeeding with probability $\geq \frac{2}{3}$ must have $\Omega(d/\varepsilon^2)$ rows [N., Nguyễn]
- Can show any oblivious subspace embedding with $O(d^{1+\gamma})$ rows must have sparsity $s = \Omega(1/(\varepsilon\gamma))$* [N., Nguyễn]
- Can provide upper bounds on $m, s$ to preserve an arbitrary bounded set $T \subset \mathbb{R}^n$, in terms of the geometry of $T$, in the style of [Gordon ’88], [Klartag, Mendelson ’05], [Mendelson, Pajor, Tomczak-Jaegermann ’07], [Dirksen ’13] (in the current notation, those works analyzed dense $\Pi$, i.e. $m = s$) [Bourgain, N.]

* Has restriction that $1/(\varepsilon\gamma) \ll d$. 
Open Problems

- **OPEN:** Improve $\omega$, the exponent of matrix multiplication
- **OPEN:** Find exact algorithm for least squares regression (or any of these problems) in time faster than $\tilde{O}(nd^{\omega-1})$
- **OPEN:** Prove conjecture: to get subsp. embedding with prob. $1 - \delta$, can set $m = O((d + \log(1/\delta))/\varepsilon^2)$, $s = O(\log(d/\delta)/\varepsilon)$. Easier: obtain this $m$ with $s = m$ via moment method.
- **OPEN:** Show that the tradeoff $m = O(d^{1+\gamma}/\varepsilon^2)$, $s = \text{poly}(1/\gamma) \cdot 1/\varepsilon$ is optimal for any distribution over subspace embeddings (the poly is probably linear)
- **OPEN:** Show that $m = \Omega(d^2/\varepsilon^2)$ is optimal for $s = 1$

Partial progress: [N., Nguyên, 2012] shows $m = \Omega(d^2)$