EE 270: Large Scale Matrix Computation and Optimization **Winter 2020**

Lecture 9 — February 4

Lecturer: Mert Pilanci *Scribe: Ethan Liang*

In the last lecture, we used Johnson-Lindenstrauss (JL) embeddings to approximately solve the least squares problem in the case where A is a tall matrix with full column rank. We used a left-sketching matrix S to reduce the number of rows in matrix A solved a reduced-dimension least squares problem. One solution of the reduced-dimension problem is the left pseudo-inverse of the matrix SA and we analyzed the expectation, variance, and distribution of our estimate for the solution of the original least squares problem under the assumption that $S \overset{\text{iid}}{\sim} \mathcal{N}$.

In this lecture, we summarize some of the main results of last lecture and discuss ways to reduce the variance of our estimate. We then consider the closely related problem of finding the minimum-norm (min-norm) solution to the least squares problem in the case of an over-parameterized solution space. First, we introduce and prove the correctness of the right-inverse of matrix A as a classical method for solving the minimum-norm least squares problem. Next, we apply a right sketch matrix S to reduce the dimension to the matrix A and solve the new optimization problem. Under the assumption that $S \overset{\text{iid}}{\sim} \mathcal{N}$, we compute the find the mean, variance, and distribution of the estimate for the min-norm. We then introduce a number JL embeddings that can be used in lieu of an i.i.d. Gaussian matrix. The lectured is concluded by introducing the basic inequality.

9.1 Faster Least Squares Optimization: Random Projection

Recall from last lecture the least squares optimization problem

$$\min_{x\in\mathbb{R}^d} \|SAx - Sb\|_2^2,$$

where $A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$, $x \in \mathbb{R}^d$, and $S \in \mathbb{R}^{m\times n}$ is a random projection matrix. By forming the matrices SA and Sb, we can formulate the smaller optimization problem above and solve in $O(md^2)$ using any classical method.

9.1.1 Gaussian Sketch

Let S be $\frac{1}{\sqrt{m}}\times$ i.i.d. Gaussian. We have shown in previous lectures that $\mathbb{E}[S^TS] = I$. Also, define the solution to the left-sketch least squares problem as

$$\tilde{x} = \arg\min_{x\in\mathbb{R}^d} \|SAx - Sb\|_2^2.$$

Last lecture, we showed that the solution to this optimization problem is unbiased, i.e.,

$$\mathbb{E}[\tilde{x}] = x_{\text{LS}}.$$

The unbiased property of $\tilde{x}$ followed from the fact that, conditioned on a fixed AS, $\tilde{x} = x_{\text{LS}} + (A^TS^TSA)^{-1}A^TS^TSb^\perp$, where $(A^TS^TSA)^{-1}A^TS^TSb^\perp$ is a zero-mean random variable. We also showed that

$$\mathbb{E}\,\|A(\tilde{x} - x_{\text{LS}})\|_2^2 = f(x_{\text{LS}})\frac{d}{m-d-1},$$

which is valid for $m > d+1$ and $f(x) = \|Ax-b\|_2^2$.

Now let's consider $f(\tilde{x})$. Using the fact that $f(x_{\text{LS}}) = \left\|b^\perp\right\|_2^2 = \|Ax_{\text{LS}} - b\|_2^2$ and $b^\perp \perp \text{range}(A)$, we can expand $f(\tilde{x})$ as follows:

$$\begin{aligned}
f(\tilde{x}) &= \|A\tilde{x} - b\|_2^2 \\
&= \|A\tilde{x} - Ax_{\text{LS}} + Ax_{\text{LS}} - b\|_2^2 \\
&= \|A(\tilde{x} - x_{\text{LS}}) + Ax_{\text{LS}} - b\|_2^2 \\
&= \|A(\tilde{x} - x_{\text{LS}})\|_2^2 + \|Ax_{\text{LS}} - b\|_2^2 && (b^\perp \perp \text{range}(A)) \\
&= \|A(\tilde{x} - x_{\text{LS}})\|_2^2 + f(x_{\text{LS}}) && (f(x_{\text{LS}}) = \|Ax_{\text{LS}} - b\|_2^2)
\end{aligned}$$

Moving $f(x_{\text{LS}})$ to the left side and taking the expectation, we get that

$$\mathbb{E}[f(\tilde{x}) - f(x_{\text{LS}})] = \mathbb{E}\,\|A(\tilde{x} - x_{\text{LS}})\|_2^2 = f(x_{\text{LS}})\frac{d}{m-d-1}.$$

9.1.2 Variance Reduction by Averaging

In the final part of the subsection above, we showed that the average deviation of the left-sketched least squares solution from the original is proportional to $f(x_{\text{LS}})$. Using the fact that this deviation is equal to the variance of the random variable $A(\tilde{x} - x_{\text{LS}})$, we can compute the average of i.i.d. instances of $\tilde{x}$ to reduce the variance and improve our estimate $f(\tilde{x})$.

Let $S_1, \ldots, S_r$ be $\frac{1}{m}\times$ i.i.d. Gaussian. It is easy to show that $\mathbb{E}[S_i^TS_i] = I \;\forall\; i$. We can define $\tilde{x}_i$ as follows:

$$\tilde{x}_i = \underset{x\in\mathbb{R}^d}{\arg\min}\,\|S_iAx - S_ib\|_2^2\,.$$

Further, let's define $\tilde{x}$ as

$$\tilde{x} = \frac{1}{r}\sum_{i=1}^{r}\tilde{x}_i.$$

$\tilde{x}$ is unbiased because

$$\mathbb{E}[\tilde{x}] \;= \mathbb{E}\left[\frac{1}{r}\sum_{i=1}^{r}\tilde{x}_i\right] = \frac{1}{r}\sum_{i=1}^{r}\mathbb{E}[\tilde{x}_i] = \frac{1}{r}\sum_{i=1}^{r}x_{\text{LS}} = x_{\text{LS}}.$$

Additionally, using the fact that the $\tilde{x}_i$'s are independent, the variance is reduced by $\frac{1}{r}$, i.e.,

$$\mathbb{E}[f(\tilde{x}) - f(x_{\text{LS}})] = \mathbb{E}\,\|A(\tilde{x} - x_{\text{LS}})\|_2^2 = \frac{1}{r}f(x_{\text{LS}})\frac{d}{m-d-1}.$$

9.2 High-dimensional Least Squares Problems

Now we consider the problem of finding the solution to the problem $Ax = b$, where $A \in \mathbb{R}^{n\times d}$ and $d > n$. Although there is no unique solution in general, the minimum (ℓ_2) norm solution is unique. The min-norm solution is defined as follows:

$$x_{\text{min-norm}} = \arg\min_{Ax=b} \|x\|_2^2 .$$

9.2.1 Minimum Norm Solution and SVD

One classical solution to the minimum norm problem is

$$x_{\text{min-norm}} = A^T(AA^T)^{-1}b.$$

The above matrix $A^T(AA^T)^{-1}$ exists as long as A has full row rank. To show that $x_{\text{min-norm}}$ is a valid solution, we can substitute $x_{\text{min-norm}}$ into the constraint equation:

$$Ax_{\text{min-norm}} = AA^T(AA^T)^{-1}b = b.$$

To prove that $x_{\text{min-norm}}$ has the smallest (ℓ_2) norm among all solutions of $Ax = b$, we first show that the difference vector $(x - x_{\text{min-norm}})$ is orthogonal to $x_{\text{min-norm}}$. For any $x' \in \mathbb{R}^d$ such that $Ax' = b$, we have

$$\begin{aligned}(x_{\text{min-norm}} - x')^T x_{\text{min-norm}} &= (x_{\text{min-norm}} - x')^T A^T(AA^T)^{-1}b \\ &= (A(x_{\text{min-norm}} - x'))^T(AA^T)^{-1}b \\ &= (b-b)^T(AA^T)^{-1}b \\ &= 0.\end{aligned}$$

Now that we have shown that $(x_{\text{min-norm}} - x') \perp x_{\text{min-norm}}$, we can rewrite x' as follows:

$$\begin{aligned}\|x'\|_2^2 &= \|x' - x_{\text{min-norm}} + x_{\text{min-norm}}\|_2^2 \\ &= \|x' - x_{\text{min-norm}}\|_2^2 + \|x_{\text{min-norm}}\|_2^2 \\ &\geq \|x_{\text{min-norm}}\|_2^2 .\end{aligned}$$

Now that we have proved that $x_{\text{min-norm}} = A^T(AA^T)^{-1}b$ is a solution to the minimum-norm problem, let's rewrite A using its singular value decomposition. Recall that any matrix A can be written as

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

where $U \in \mathbb{R}^{n\times n}$, $\Sigma \in \mathbb{R}^{n\times d}$, and $V \in \mathbb{R}^{d\times d}$. Also recall that the columns U and V are orthogonal and unit norm, so that $U^TU = I$ and $V^TV = I$ and Σ is a diagonal matrix. Without loss of generality, assume that A is full row rank. Substituting in for A, we get that

$$\begin{aligned}x_{\text{min-norm}} &= A^T(AA^T)^{-1}b \\ &= V\Sigma U^T(U\Sigma V^TV\Sigma U^T)^{-1}b \\ &= V\Sigma U^T(U\Sigma^2U^T)^{-1}b \\ &= V\Sigma U^TU\Sigma^{-2}U^Tb \\ &= V\Sigma^{-1}U^Tb.\end{aligned}$$

9.2.2 Random Projection to Reduce Dimension: Right Sketch

Consider the original optimization problem:

$$x_{\text{min-norm}} = \arg\min_{Ax=b} \|x\|_2^2 .$$

Similar to the least squares case, we can multiply A to form AS where $S \in \mathbb{R}^{d\times m}$ and solve

$$\arg\min_{ASx=z} \|z\|_2^2 .$$

Because we right multiplied A by S, we changed the dimension of our optimization parameter, i.e., the vector whose norm we minimized is no longer the same dimension as x. We deal with this problem by using the fact that $x = Sz$ and hope that Sz is a good approximation for x (we will see later that this is true).

Let's define

$$\tilde{z} := \arg\min_{ASx=z} \|z\|_2^2 ,$$

and use the approximation that $\tilde{x} = S\tilde{z}$. A solution for $\tilde{z}$ is $(AS)^\dagger b$. Plugging into the constraint formula, we see that

$$\begin{aligned} A\tilde{x} &= AS\tilde{z} \\ &= AS(AS)^\dagger b \\ &= b. \end{aligned}$$

This shows that $\tilde{x} = S\tilde{z}$ is a valid solution the original constraint $Ax = b$ and exists as long as AS has full row rank.

9.2.3 Right Sketch

Suppose that $S \overset{\text{iid}}{\sim} \mathcal{N}(0, \frac{1}{\sqrt{m}})$ is the right sketch matrix and $\tilde{z}$ is the solution to the sketched least-norm problem. Our estimate for $x_{\text{min-norm}}$, $\tilde{x} = S\tilde{z}$, has a number of important properties.

Lemma 1. For a fixed AS, $\tilde{x}$ is unbiased estimator of $x_{\text{min-norm}}$, i.e., $\mathbb{E}[\tilde{x}] = x_{\text{min-norm}}$.

Proof. Consider the non-compact right singular vector matrix $[V \ V^\perp]$, where $V^T \in \mathbb{R}^{d\times m}$ and $V^{\perp T} \in \mathbb{R}^{(m-d)\times m}$. Left multiplying $\tilde{x}$ by V^T, we get that

$$\begin{aligned} V^T\tilde{x} &= V^T S\tilde{z} \\ &= V^T S(AS)^\dagger b \\ &= V^T SS^T A^T (ASS^T A^T)^{-1} b \\ &= V^T SS^T V\Sigma U^T (U\Sigma V^T SS^T V\Sigma U^T)^{-1} b \\ &= V^T SS^T V\Sigma U^T U\Sigma^{-1} (V^T SS^T V)^{-1}\Sigma^{-1} U^T b \qquad (A \text{ full row rank}) \\ &= V^T SS^T V (V^T SS^T V)^{-1}\Sigma^{-1} U^T b \\ &= \Sigma^{-1} U^T b \\ &= V^T V\Sigma^{-1} U^T b \\ &= V^T x_{\text{min-norm}}. \end{aligned}$$

Similarly, left multiplying $\tilde{x}$ by $V^{\perp^T}$, we get that

$$\begin{aligned} V^{\perp^T}\tilde{x} &= V^{\perp^T}S\tilde{z} \\ &= V^{\perp^T}SS^TV(V^TSS^TV)^{-1}\Sigma^{-1}U^Tb \qquad \text{(steps identical to above)} \\ &= V^{\perp^T}SS^TV(V^TSS^TV)^{-1}V^TV\Sigma^{-1}U^Tb \\ &= V^{\perp^T}SS^TV(V^TSS^TV)^{-1}V^Tx_{\text{min-norm}} \end{aligned}$$

Taking the expectation of $V^{\perp^T}\tilde{x}$:

$$\begin{aligned} \mathbb{E}[V^{\perp^T}\tilde{x}] &= \mathbb{E}[V^{\perp^T}SS^TV(V^TSS^TV)^{-1}V^Tx_{\text{min-norm}}] \\ &= \mathbb{E}[S^TV^{\perp}]S^TV(V^TSS^TV)^{-1}V^Tx_{\text{min-norm}} \\ &= 0. \end{aligned}$$

We can move $S^TV(V^TSS^TV)^{-1}V^Tx_{\text{min-norm}}$ out of the expectation because $S^TV^{\perp}$ and S^TV are uncorrelated, and thus independent, Gaussian vector random variables. Now considering the expectation of the orthogonal matrix-vector product $[V\ V^{\perp}]^T\tilde{x}$, we get that

$$E\begin{bmatrix} V^T\tilde{x} \\ V^{\perp^T}\tilde{x} \end{bmatrix} = \begin{bmatrix} \mathbb{E}[V^T\tilde{x}] \\ \mathbb{E}[V^{\perp^T}\tilde{x}] \end{bmatrix} = \begin{bmatrix} V^Tx_{\text{min-norm}} \\ 0 \end{bmatrix}$$

Because the expectation of $\tilde{x}$ times an orthogonal matrix is equal to $x_{\text{min-norm}}$ times an orthogonal matrix, we can conclude that $\mathbb{E}[\tilde{x}] = x_{\text{min-norm}}$. □

Lemma 2. For a fixed AS, $\tilde{x} \sim \mathcal{N}(x_{\text{min-norm}}, \frac{1}{m}(I - VV^T)b^T(ASS^TA^T)^{-1}b)$.

Proof. Using the fact above that $\tilde{x} = S\tilde{z} = SS^TV(V^TSS^TV)^{-1}V^Tx_{\text{min-norm}}$, we have that

$$\begin{aligned} &\mathbb{E}[(\tilde{x} - x_{\text{min-norm}})(\tilde{x} - x_{\text{min-norm}})^T] \\ &\qquad = \mathbb{E}[\tilde{x}\tilde{x}^T] - \mathbb{E}[x_{\text{min-norm}}x_{\text{min-norm}}^T] \\ &\qquad = \mathbb{E}[SS^TV(V^TSS^TV)^{-1}V^Tx_{\text{min-norm}}x_{\text{min-norm}}^TV(V^TSS^TV)^{-T}V^TSS^T] \\ &\qquad\quad - \mathbb{E}[x_{\text{min-norm}}x_{\text{min-norm}}^T] \\ &\qquad = \mathbb{E}[SS^TV(V^TSS^TV)^{-1}V^TV\Sigma^{-1}U^Tb(V\Sigma^{-1}U^Tb)^TV(V^TSS^TV)^{-T}V^TSS^T] \\ &\qquad\quad - \mathbb{E}[V\Sigma^{-1}U^Tb(V\Sigma^{-1}U^Tb)^T] \\ &\qquad = \mathbb{E}[SS^TV(V^TSS^TV)^{-1}\Sigma^{-1}U^Tbb^TU\Sigma^{-1}(V^TSS^TV)^{-T}V^TSS^T] \\ &\qquad\quad - \mathbb{E}[V\Sigma^{-1}U^Tb(V\Sigma^{-1}U^Tb)^T] \\ &\qquad = \mathbb{E}[SS^TV(U\Sigma V^TSS^TV)^{-1}bb^T(V^TSS^TV\Sigma U^T)^{-T}V^TSS^T] \\ &\qquad\quad - \mathbb{E}[V\Sigma^{-1}U^Tb(V\Sigma^{-1}U^Tb)^T] \end{aligned}$$

□

Lemma 3. The error vector $\tilde{x} - x_{\text{min-norm}} \in \text{NULL}(A)$.

Proof.

$$A\tilde{x} - Ax_{\text{min-norm}} = A(\tilde{x} - x_{\text{min-norm}}) = 0.$$

□

Lemma 4. $\mathbb{E}\left\|\tilde{x} - x_{\text{min-norm}}\right\|_2^2 = \frac{d-n}{m-n-1}\left\|x_{\text{min-norm}}\right\|_2^2$

Proof. We use the fact that $\mathbb{E}[(ASS^TA^T)^{-1}] = (AA^T)^{-1}\frac{m}{m-n-1}$ and that for $z \sim \mathcal{N}(0, K)$, we have $\mathbb{E}[\|z\|_2^2] = \mathbb{E}[\text{tr}zz^T] = \mathbb{E}[\text{tr}K]$.

$$\begin{aligned}
\mathbb{E}\left\|\tilde{x} - x_{\text{min-norm}}\right\|_2^2 &= \mathbb{E}[(\tilde{x} - x_{\text{min-norm}})^T(\tilde{x} - x_{\text{min-norm}})] \\
&= \mathbb{E}[\text{tr}(\tilde{x} - x_{\text{min-norm}})(\tilde{x} - x_{\text{min-norm}})^T] \\
&= \mathbb{E}[\text{tr}\frac{1}{m}(I - VV^T)b^T(ASS^TA^T)^{-1}b)] \\
&= \frac{1}{m}\text{tr}(I - VV^T)b^T\mathbb{E}[(ASS^TA^T)^{-1}]b \\
&= \frac{1}{m}\frac{m}{m-n-1}\text{tr}(I - VV^T)b^T(AA^T)^{-1}b \\
&= \frac{d-n}{m-n-1}b^T(AA^T)^{-1}b \\
&= \frac{d-n}{m-n-1}b^T(U\Sigma^2U^T)^{-1}b \\
&= \frac{d-n}{m-n-1}b^TU\Sigma^{-2}U^Tb \\
&= \frac{d-n}{m-n-1}b^TU\Sigma^{-1}V^TV\Sigma^{-1}U^Tb \\
&= \frac{d-n}{m-n-1}(V\Sigma^{-1}U^Tb)^TV\Sigma^{-1}U^Tb \\
&= \frac{d-n}{m-n-1}\left\|x_{\text{min-norm}}\right\|_2^2
\end{aligned}$$

□

9.3 Left Sketch vs Right Sketch Summary

In this section, we compare and contrast the main results for approximate least square and minimum norm optimization. Suppose for this section that $A \in \mathbb{R}^{n\times d}$.

- In both cases, using an i.i.d. Gaussian sketching matrix S, the resulting estimate $\tilde{x}$ is unbiased.
- In the case where $n \geq d$, we solved the following least squares optimization problem:

  $$\tilde{x} = \min_{x\in\mathbb{R}^d}\left\|SAx - Sb\right\|_2^2.$$

  Variance: $\mathbb{E}\left\|A(\tilde{x} - x_{\text{LS}})\right\|_2^2 = f(x_{\text{LS}})\frac{d}{m-d-1}$

- In the case where $n < d$, we solved the following minimum norm optimization problem:

$$\tilde{x} = S\tilde{z} \text{ where } \tilde{z} := \underset{ASx=z}{\arg\min} \|z\|_2^2 .$$

Variance: $\mathbb{E}\,\|\tilde{x} - x_{\text{min-norm}}\|_2^2 = \frac{d-n}{m-n-1}\,\|x_{\text{min-norm}}\|_2^2$

9.4 Back to Left Sketch: Which Sketching Matrices are Good?

Up until this point, we have assumed that S is an i.i.d. Gaussian sketching matrix to help simplify our proofs. However, many other sketching matrices are possible. Recall that A can be written as $A = U\Sigma V^T$ in SVD compact form and that the minimum norm solution to $Ax = b$ is given by $x_{\text{min-norm}} = A^T(AA^T)^{-1}b$. Some deterministic options for S are as follows:

- $S = U^T$
  The minimum norm solution for this sketching matrix is the left pseudo-inverse of SA, i.e.,

$$\begin{aligned}
\tilde{x} &= (SA)^\dagger Sb \\
&= (A^T S^T SA)^{-1} A^T S^T Sb \\
&= (V\Sigma U^T UU^T U\Sigma V^T)^{-1} V\Sigma U^T UU^T b \\
&= (V\Sigma U^T U\Sigma V^T)^{-1} V\Sigma U^T b \\
&= (A^T A)^{-1} A^T b \\
&= x_{\text{LS}}
\end{aligned}$$

- $S = A^T$
  The minimum norm solution for this sketching matrix is the pseudo-inverse of AA^T:

$$\begin{aligned}
\tilde{x} &= (SA)^\dagger Sb \\
&= (A^T A)^\dagger A^T b \\
&= (A^T AA^T A)^{-1} A^T AA^T b \\
&= (V\Sigma^2 V^T V\Sigma^2 V^T)^{-1} V\Sigma^2 V^T V\Sigma U^T b \\
&= (V\Sigma^4 V^T)^{-1} V\Sigma^3 U^T b \\
&= V\Sigma^{-4} V^T V\Sigma^3 U^T b \\
&= V\Sigma^{-1} U^T b \\
&= x_{\text{LS}}
\end{aligned}$$

The choices for the S matrix above exactly solve the least squares problem. However, the computations above exactly coincide with some of the classical methods for solving the least squares problem, contradicting the original purpose of using a JL embedding to approximately solve the original optimization problem with fewer computations.

9.5 Basic Inequality Method

Consider the left sketch least squares problem where we compute $\tilde{x} = \min_{x\in\mathbb{R}^d} \|SAx - Sb\|_2^2$. To bound the difference between $\tilde{x}$ and $x_{\text{LS}} = \min_{x\in\mathbb{R}^d} \|Ax - b\|_2^2$, we can use the basic inequality method. For this problem, the basic inequality method proceeds in three steps:

1. Establish two optimality (in)equalities for these variables.
   - $\|Ax_{\text{LS}} - b\|_2^2 \le \|Ax' - b\|_2^2$ for any x', i.e., $A^T(Ax_{\text{LS}} - b) = 0$
   - $\|S(A\tilde{x} - b)\|_2^2 \le \|S(Ax_{\text{LS}} - b)\|_2^2$
2. Define the error $\Delta = \tilde{x} - x_{\text{LS}}$ and re-write these inequalities in terms of Δ.

$$\begin{aligned}
\|S(A\tilde{x} - b)\|_2^2 &= \|S(A(\tilde{x} - x_{\text{LS}} + x_{\text{LS}}) - b)\|_2^2 \\
&= \|S(A(x_{\text{LS}} + \Delta) - b)\|_2^2 \\
&= \|SA(x_{\text{LS}} + \Delta) - Sb\|_2^2 \\
&= \|SA\Delta + SAx_{\text{LS}} - Sb\|_2^2 \\
&= \|SA\Delta\|_2^2 + \|SAx_{\text{LS}} - Sb\|_2^2 + 2\langle SA\Delta, SAx_{\text{LS}} - Sb\rangle
\end{aligned}$$

Recall above because $\|S(A\tilde{x} - b)\|_2^2 \le \|S(Ax_{\text{LS}} - b)\|_2^2$, we can conclude that $\|SA\Delta\|_2^2 + 2\langle SA\Delta, SAx_{\text{LS}} - Sb\rangle \le 0$. Then we can write

$$\begin{aligned}
\|SA\Delta\|_2^2 &\le 2\langle SA\Delta, S(b - Ax_{\text{LS}})\rangle \\
&= 2\left\langle SA\Delta, Sb^{\perp}\right\rangle \\
&= 2(b^{\perp})^T S^T SA\Delta \\
&= 2(b^{\perp})^T S^T SA\Delta - 2(b^{\perp})^T A\Delta \qquad (b^{\perp} \perp \text{RANGE}(A)) \\
&= 2(b^{\perp})^T (S^T S - I)A\Delta
\end{aligned}$$

3. Argue $S^T S \approx I$

$$\begin{aligned}
\max_{\Delta} \left| \frac{\|SA\Delta\|_2^2}{\|A\Delta\|_2^2} - 1 \right| &= \max_z \left| \frac{\|SUz\|_2^2}{\|z\|_2^2} - 1 \right| \\
&= \max_z \left| \frac{z^T}{\|z\|_2} (U^T S^T SU - I) \frac{z}{\|z\|_2} \right| \\
&= \sigma_{\max}^2 (U^T S^T SU - I)
\end{aligned}$$

Using the property of approximate matrix multiplication for JL embeddings (this property actually holds for many different random sampling matrices):

$$\sigma_{\max}(U^T S^T SU - U^T U) \le \left\|U^T S^T SU - U^T U\right\|_F \le \epsilon \left\|U^T U\right\|_F^2$$

Note that we can rescale ϵ with an appropriate m to get $\sigma_{\max}(U^T S^T SU - U^T U) \le m$.