# Approximate Matrix Multiplication

Lecture 21

April 11, 2019

# Matrix data

Lot of data can be viewed as defining a matrix. We have already seen vectors modeling data/signals. More generally we can use tensors too.

$n$ data items and each data item $a_i$ is a vector over some features (say $m$ features)

$A$ is the matrix defined by the $n$ data items.

Assuming $a_1, \ldots, a_n$ are columns then $A$ is a $m \times n$ matrix

Combinatorial objects such as graphs can also be modeled via graphs

# Numerical Linear Algebra

Basic problems in linear algebra:

- Matrix vector product: compute $Ax$
- Matrix multiplication: compute $AB$
- Linear equations: solve $Ax = b$
- Matrix inversion: compute $A^{-1}$
- Least squares: solve $\min_x \|Ax - b\|$
- Singular value decomposition, eigen values, principal component analysis, low-rank approximations
- $\ldots$

Fundamental in all areas of applied mathematics and engineering.
Many applications to statistics and data analysis.

# Numerical Linear Algebra

NLA has a vast literature

In practice iterative methods are used that converge to an optimum solution. They can take advantage of sparsity in the input data better than exact methods

Some TCS contributions in the recent past:

- randomized NLA for faster algorithms with provable approximation guarantees - sampling and JL based techniques and others
- revisit preconditioning methods for Laplacians and beyond
- Many powerful applications in theory and practice

# Norms and matrix norms

## Definition

A norm $\|\|$ in a real vector space $V$ is a real valued function that has three properties: (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ implies $x = 0$, (ii) $\|ax\| = |a|\|x\|$ for all scalars $a$ (iii) $\|x + y\| \leq \|x\| + \|y\|$

Familiar vector norms: $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$

If $A$ is a injective linear transformation $\|Ax\|$ is also a norm in the original space.

Norms and metrics: $d(x, y) = \|x - y\|$ is a metric

# Matrix norms

Consider vector space of all matrices $A \in \mathbb{R}^{m \times n}$

What are useful norms over matrices?

- Treat matrix like a vector of dimension $m \times n$ and apply vector norm. For instance $\|A\|_F$ (Frobenius norm) is $(\sum_{i,j} |A_{i,j}|^2)^{1/2}$.
- Treat matrix as linear operator and see what it does to norms of vectors it operates on. Spectral norm is $\sup_{\|x\|_2=1} \|Ax\|_2$.
- Schatten $p$-norms based on singular values of $A$
- Trace norm, nuclear norm, ...
- Norms are related in some cases (different perspective on the same norm)

# Frobenus and Spectral norms

Submultiplicative property:

$$\|AB\|_F \leq \|A\|_F \|B\|_F$$

$$\|AB\|_2 \leq \|A\|_2 \|B\|_2$$

# Matrix Multiplication

**Problem:** Given matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ compute the matrix $AB$

- Standard algorithm based on definition: $O(mnp)$ time
- Faster algorithms via non-trivial Strassen-like divide and conquer.

# Matrix Multiplication

**Problem:** Given matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ compute the matrix $AB$

- Standard algorithm based on definition: $O(mnp)$ time
- Faster algorithms via non-trivial Strassen-like divide and conquer.

**Approximation:** Compute $D \in \mathbb{R}^{m \times p}$ such that $\|D - AB\|$ is small in some appropriate matrix norm.

Two methods

- random sampling
- random projections (fast JL)

# Matrix Multiplication

**Problem:** Given matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ compute the matrix $AB$

**Notation:** $M^{(j)}$ for $j$'th column of $M$ and $M_{(i)}$ for $i$'th row of $M$ both interpreted as vectors

From textbook definition: $D_{i,h} = \langle A_{(i)}, B^{(h)} \rangle = \sum_{k=1}^{n} A_{i,k} B_{k,h}$

Consider $A^T$ consisting of $m$ column vectors from $\mathbb{R}^n$ and $B$ as $p$ column vectors from $\mathbb{R}^n$

We want to compute all $mp$ inner products of these vectors.

# Approximate Matrix Multiplication

Want to approximate $AB$ in the Frobenius norm.
Want $D$ such that $\|D - AB\|_F \leq \epsilon \|AB\|_F$ but $\|AB\|_F$ can be $\mathbf{0}$.

Instead will settle for $\|D - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$

# Part I

# Random Sampling for Approx Matrix Mult

# Matrix Multiplication and Outer Products

Alternate definition of matrix multiplication based on outer product:

$$AB = \sum_{j=1}^{n} A^{(j)} B_{(j)}$$

$A^{(j)} B_{(j)}$ is a $m \times h$ matrix of rank 1

# Importance Sampling

$$AB = \sum_{j=1}^{n} A^{(j)} B_{(j)}$$

- Pick a probability distribution over $[n]$, $p_1 + p_2 + \ldots + p_n = 1$
- For $\ell = 1$ to $t$ do pick an index $j_\ell \in [n]$ according to distribution $p$ (independent with replacement)
- Output $C = \frac{1}{t} \sum_{\ell=1}^{t} \frac{1}{p_{i_\ell}} A^{(j_\ell)} B_{(j_\ell)}$

# Importance Sampling

$$AB = \sum_{j=1}^{n} A^{(j)} B_{(j)}$$

- Pick a probability distribution over $[n]$, $p_1 + p_2 + \ldots + p_n = 1$
- For $\ell = 1$ to $t$ do pick an index $j_\ell \in [n]$ according to distribution $p$ (independent with replacement)
- Output $C = \frac{1}{t} \sum_{\ell=1}^{t} \frac{1}{p_{i_\ell}} A^{(j_\ell)} B_{(j_\ell)}$

$C = \frac{1}{t} \sum_\ell C_\ell$ where $\mathbf{E}[C_\ell] = AB$.
By linearity of expectation: $\mathbf{E}[C] = AB$

# Importance Sampling

**Question:** How should we choose $p_1, p_2, \ldots, p_n$?

# Importance Sampling

**Question:** How should we choose $p_1, p_2, \ldots, p_n$? $p_j$ should correspond to contribution of $A^{(j)} B_{(j)}$ to $\|AB\|_F$

# Importance Sampling

**Question:** How should we choose $p_1, p_2, \ldots, p_n$? $p_j$ should correspond to contribution of $A^{(j)}B_{(j)}$ to $\|AB\|_F$

Use spectral norm of $A^{(j)}B_{(j)}$ which is $\|A^{(j)}B_{(j)}\|_2$

# Importance Sampling

**Question:** How should we choose $p_1, p_2, \ldots, p_n$? $p_j$ should correspond to contribution of $A^{(j)} B_{(j)}$ to $\|AB\|_F$

Use spectral norm of $A^{(j)} B_{(j)}$ which is $\|A^{(j)} B_{(j)}\|_2$

**Claim:** $\|A^{(j)} B_{(j)}\|_2 = \|A^{(j)}\|_2 \|B_{(j)}\|_2$.

Choose $p_j = \frac{\|A^{(j)}\|_2 \|B_{(j)}\|_2}{\sum_\ell \|A^{(\ell)}\|_2 \|B_{(\ell)}\|_2}$

Due to [Drineas-Kannan-Mahoney]

# Running time

- For all $j$ compute $\|A^{(j)}\|_2$ and $\|B_{(j)}\|_2$. Takes one pass over $A$ and $B$
- Allows one to compute $p_1, p_2, \ldots, p_n$
- $C = \frac{1}{t}\sum_{\ell=1}^{t} \frac{1}{p_{i_\ell}} A^{(j_\ell)} B_{(j_\ell)}$
- At most $O(tmh + N_A + N_B)$ time where $N_A$ and $N_B$ is number of non-zeroes in $A$ and $B$.
- Full computation takes $O(nmh)$ time.

# Analysis of approximation

Want to analyse $\mathbf{Pr}[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F]$.

# Analysis of approximation

Want to analyse $\mathbf{Pr}[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F]$.

Using Markov:

$$\mathbf{Pr}[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F] \leq \frac{\mathbf{E}\left[\|C - AB\|_F^2\right]}{\epsilon^2 \|A\|_F^2 \|B\|_F^2}$$

**Lemma**

$$\mathbf{E}\left[\|C - AB\|_F^2\right] \leq \frac{1}{t}\left(\sum_{j=1}^{n} \|A^{(j)}\|_2 \|B_{(j)}\|_2\right)^2 - \frac{1}{t}\|AB\|_F^2$$

# Analysis continued

$$\mathbf{Pr}[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F] \leq \frac{\mathbf{E}\left[\|C - AB\|_F^2\right]}{\epsilon^2 \|A\|_F^2 \|B\|_F^2}$$

$$\begin{aligned}\mathbf{E}\left[\|C - AB\|_F^2\right] &\leq \frac{1}{t}\left(\sum_{j=1}^{n} \|A^{(j)}\|_2 \|B_{(j)}\|\right)^2 - \frac{1}{t}\|AB\|_F^2 \\ &\leq \frac{1}{t}\|A\|_F^2 \|B\|_F^2.\end{aligned}$$

Thus, if $t = \frac{1}{\epsilon^2 \delta}$ then

$$\mathbf{Pr}[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F] \leq \delta.$$

# Median trick

Recall that we used median trick to improve dependence on $\delta$ from $1/\delta$ to $\log(1/\delta)$.

If $t = \frac{3}{\epsilon^2}$ then

$$\Pr[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F] \leq 1/3.$$

Repeat independently to obtain $C_1, C_2, \ldots, C_r$ where $r = \Theta(\log(1/\delta))$

By Chernoff bounds majority of estimators are good. How do we pick the “median” matrix?

# Median trick

If $t = \frac{3}{\epsilon^2}$ then

$$\Pr[\|C - AB\|_F \geq \epsilon \|A\|_F \|B\|_F] \leq 1/3.$$

Repeat independently to obtain $C_1, C_2, \ldots, C_r$ where $r = \Theta(\log(1/\delta))$

For each $1 \leq i \leq r$ compute

$$\rho_i = |\{j \mid j \neq i, \|C_i - C_j\| \leq 2\epsilon \|A\|_F \|B\|_F\}|$$

Output $C_s$ such that $\rho_s \geq r/2$

[Clarkson-Woodruff]

# Median trick

For each $1 \le i \le r$ compute

$$\rho_i = |\{j \mid j \neq i, \|C_i - C_j\| \le 2\epsilon\|A\|_F\|B\|_F\}|$$

Output $C_s$ such that $\rho_s \ge r/2$

Correctness follows from triangle inequality.

$$\|C_i - C_j\|_F \le \|C_i - AB\|_F + \|C_j - AB\|_F$$

and

$$\|C_i - C_j\|_F \ge \|C_i - AB\|_F - \|C_j - AB\|_F.$$

# Median trick

For each $1 \leq i \leq r$ compute

$$\rho_i = |\{j \mid j \neq i, \|C_i - C_j\| \leq 2\epsilon \|A\|_F \|B\|_F\}|$$

Output $C_s$ such that $\rho_s \geq r/2$

Correctness follows from triangle inequality.

$$\|C_i - C_j\|_F \leq \|C_i - AB\|_F + \|C_j - AB\|_F$$

More than half of $C_i$'s have $\|C_i - AB\|_F < \epsilon \|A\|_F \|B\|_F$. If $C_s$ is good then $\rho_s \geq r/2$.

# Median trick

For each $1 \le i \le r$ compute

$$\rho_i = |\{j \mid j \neq i, \|C_i - C_j\| \le 2\epsilon\|A\|_F\|B\|_F\}|$$

Output $C_s$ such that $\rho_s \ge r/2$

Correctness follows from triangle inequality.

$$\|C_i - C_j\|_F \ge \|C_i - AB\|_F - \|C_j - AB\|_F.$$

More than half of $C_i$'s have $\|C_i - AB\|_F < \epsilon\|A\|_F\|B\|_F$. If $C_s$ is bad ($\|C_s - AB\|_F > 3\|A\|_F\|B\|_F$) then
$\|C_s - C_j\| > 2\epsilon\|A\|_F\|B\|_F$ which means $\rho_s < r/2$.

# Running time again

- For all $j$ compute $\|A^{(j)}\|_2$ and $\|B_{(j)}\|_2$. Takes one pass over $A$ and $B$
- Allows one to compute $p_1, p_2, \ldots, p_n$
- $C = \frac{1}{t}\sum_{\ell=1}^{t} \frac{1}{p_{i_\ell}} A^{(j_\ell)} B_{(j_\ell)}$
- At most $O(tmh + N_A + N_B)$ time where $N_A$ and $N_B$ is number of non-zeroes in $A$ and $B$.
- Full computation takes $O(nmh)$ time.

Either we choose $t = \frac{1}{\epsilon^2 \delta}$ or we choose $t = \frac{1}{\epsilon^2} \ln(1/\delta)$ but then we need to do pairwise matrix computations so effectively $t = \frac{1}{\epsilon^4} \log^2(1/\delta)$.

# Proof of Lemma

## Lemma

$$\mathbf{E}\left[\|C - AB\|_F^2\right] \leq \frac{1}{t}\left(\sum_{j=1}^{n} \|A^{(j)}\|_2 \|B_{(j)}\|_2\right)^2 - \frac{1}{t}\|AB\|_F^2$$

Recall $C$ is sum of $t$ independent estimators so the lemma is basically about $t = 1$.

$$\mathbf{E}\left[\|C - AB\|_F^2\right] = \sum_{x,y} \mathbf{E}\left[(C_{x,y} - (AB)_{x,y})^2\right]$$

Fix $x, y$. Let $Z = C_{x,y}$. We have $\mathbf{E}[Z] = (AB)_{x,y}$. Hence

$$Var[Z] = \mathbf{E}\left[Z^2\right] - (AB)_{x,y}^2 = \mathbf{E}\left[(C_{x,y} - (AB)_{x,y})^2\right]$$

# Proof of Lemma

$\mathbf{E}\left[Z^2\right] = \sum_{j=1}^{n} p_j (A_{x,j} B_{j,y})^2 / p_j^2 = \sum_{j=1}^{n} (A_{x,j} B_{j,y})^2 / p_j$

Thus $\mathbf{E}\left[\|C - AB\|_F^2\right] = \sum_{x,y} \sum_{j=1}^{n} (A_{x,j} B_{j,y})^2 / p_j - \|AB\|_F^2$.

Simplifying the first term:
$\sum_{x,y} \sum_{j=1}^{n} (A_{x,j)})^2 (B_{j,y})^2 / p_j = \sum_{j=1}^{n} \frac{1}{p_j} \|A^{(j)}\|^2 \|B_{(j)}\|^2$

# Proof of Lemma

$\mathbf{E}\left[Z^2\right] = \sum_{j=1}^{n} p_j (A_{x,j} B_{j,y})^2 / p_j^2 = \sum_{j=1}^{n} (A_{x,j} B_{j,y})^2 / p_j$

Thus $\mathbf{E}\left[\|C - AB\|_F^2\right] = \sum_{x,y} \sum_{j=1}^{n} (A_{x,j} B_{j,y})^2 / p_j - \|AB\|_F^2$.

Simplifying the first term:
$\sum_{x,y} \sum_{j=1}^{n} (A_{x,j)})^2 (B_{j,y})^2 / p_j = \sum_{j=1}^{n} \frac{1}{p_j} \|A^{(j)}\|^2 \|B_{(j)}\|^2$

Recall: $p_j = \frac{\|A^{(j)}\| \|B_{(j)}\|}{\sum_\ell \|A^{(\ell)}\| \|B_{(\ell)}\|}$

Thus $\mathbf{E}\left[\|C - AB\|_F^2\right] = \left(\sum_{j=1}^{n} \|A^{(j)}\| \|B_{(j)}\|\right)^2 - \|AB\|_F^2$.

# Proof of Lemma

$$\mathbf{E}\left[Z^2\right] = \sum_{j=1}^{n} p_j (A_{x,j} B_{j,y})^2 / p_j^2 = \sum_{j=1}^{n} (A_{x,j} B_{j,y})^2 / p_j$$

Thus $\mathbf{E}\left[\|C - AB\|_F^2\right] = \sum_{x,y} \sum_{j=1}^{n} (A_{x,j} B_{j,y})^2 / p_j - \|AB\|_F^2$.

Simplifying the first term:
$\sum_{x,y} \sum_{j=1}^{n} (A_{x,j)})^2 (B_{j,y})^2 / p_j = \sum_{j=1}^{n} \frac{1}{p_j} \|A^{(j)}\|^2 \|B_{(j)}\|^2$

Recall: $p_j = \frac{\|A^{(j)}\| \|B_{(j)}\|}{\sum_\ell \|A^{(\ell)}\| \|B_{(\ell)}\|}$

Thus $\mathbf{E}\left[\|C - AB\|_F^2\right] = \left(\sum_{j=1}^{n} \|A^{(j)}\| \|B_{(j)}\|\right)^2 - \|AB\|_F^2$.

One can show that the choice of $p_j$ values is optimum for reducing variance in the simple importance sampling scheme.

# Sampling matrix view

$$AB = \sum_{j=1}^{n} A^{(j)} B_{(j)}$$

- Pick a probability distribution over $[n]$, $p_1 + p_2 + \ldots + p_n = 1$
- For $\ell = 1$ to $t$ do pick an index $j_\ell \in [n]$ according to distribution $p$ (independent with replacement)
- Output $C = \frac{1}{t} \sum_{\ell=1}^{t} \frac{1}{p_{i_\ell}} A^{(j_\ell)} B_{(j_\ell)}$

$C = (AS^T)(SB)$ where $S \in \mathbb{R}^{n \times t}$ is a sampling matrix:

$S_{i,j} = \frac{1}{\sqrt{tp_j}}$ if column $j$ is picked in $i$'th sample else $S_{i,j} = 0$

# Part II

# Random Projection for Approx Matrix Mult

# JL Approach for Approx Matrix Multiplication

[Sarlos]

Output $C = (AS^T)(SB)$ where $S$ is a (fast) JL matrix. Works!

Advantage?

# JL Approach for Approx Matrix Multiplication

[Sarlos]

Output $C = (AS^T)(SB)$ where $S$ is a (fast) JL matrix. Works!

Advantage? Oblivious to $A, B$. Can update them etc.

# Recalling JL

## Lemma (Distributional JL Lemma)

*Fix vector* $x \in \mathbb{R}^d$ *and let* $\Pi \in \mathbb{R}^{k \times d}$ *matrix where each entry* $\Pi_{ij}$ *is chosen independently according to standard normal distribution* $\mathcal{N}(0,1)$ *distribution. If* $k = \Omega(\frac{1}{\epsilon^2}\log(1/\delta))$*, then with probability* $(1-\delta)$ *we have* $\|\frac{1}{\sqrt{k}}\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$.

Can choose entries from $\{-1, 1\}$ as well.

## Definition

Let $\mathcal{D}$ be a distribution over $m \times n$ matrices. $D$ is said to have $(\epsilon, \delta)$ JL moment property if for any unit vector $x \in \mathbb{R}^n$,

$$E_{\Pi \sim \mathcal{D}} |\|\Pi x\|_2^2 - 1| \leq \epsilon\delta.$$

# JL Property

## Lemma

*If* $\Pi$ *comes from* $(\epsilon, \delta)$ *JL moment distribution then for all unit vectors* $x, y \in \mathbb{R}^n$, $\mathbf{E}\left[|\langle \Pi x, \Pi y\rangle - \langle x, y\rangle|^2\right] \leq c\epsilon^2\delta$.

# Using JL

## Theorem

*Suppose* $\Pi$ *is chosen from a distribution* $\mathcal{D}$ *that satisfies* $(\epsilon, \delta)$ *JL moment property then*

$$\Pr_{\Pi \sim \mathcal{D}}[\|AB - (A\Pi^T)(B\Pi)\|_F > 3\epsilon\|A\|_F\|B\|_F] \leq \delta.$$

# Using JL

## Theorem

*Suppose* $\Pi$ *is chosen from a distribution* $\mathcal{D}$ *that satisfies* $(\epsilon, \delta)$ *JL moment property then*

$$\Pr_{\Pi \sim \mathcal{D}}[\|AB - (A\Pi^T)(B\Pi)\|_F > 3\epsilon\|A\|_F\|B\|_F] \leq \delta.$$

Let $C = (A\Pi^T)(B\Pi)$.

$$\Pr[\|AB - C\|_F^2 > 3\epsilon\|A\|_F\|B\|_F^2] \leq \frac{\mathbf{E}\left[\|AB - C\|_F^2\right]}{(3\epsilon\|A\|_F\|B\|_F)^2}.$$

# Analysis

$C_{i,j} = \langle \Pi A_{(i)}, \Pi B^{(j)} \rangle$ while $(AB)_{i,j} = \langle A_{(i)}, B^{(j)} \rangle$

Notation: $a_i$ for $A_{(i)}$ and $b_j$ for $B^{(j)}$

# Analysis

$C_{i,j} = \langle \Pi A_{(i)}, \Pi B^{(j)} \rangle$ while $(AB)_{i,j} = \langle A_{(i)}, B^{(j)} \rangle$

Notation: $a_i$ for $A_{(i)}$ and $b_j$ for $B^{(j)}$

$\|AB - C\|_F^2 = \sum_{i,j} |\langle \Pi a_i, \Pi b_j \rangle - \langle a_i, b_j \rangle|^2$

# Analysis

$C_{i,j} = \langle \Pi A_{(i)}, \Pi B^{(j)} \rangle$ while $(AB)_{i,j} = \langle A_{(i)}, B^{(j)} \rangle$

Notation: $a_i$ for $A_{(i)}$ and $b_j$ for $B^{(j)}$

$\|AB - C\|_F^2 = \sum_{i,j} |\langle \Pi a_i, \Pi b_j \rangle - \langle a_i, b_j \rangle|^2$

Term by term:
$|\langle \Pi a_i, \Pi b_j \rangle - \langle a_i, b_j \rangle|^2 = \alpha |\langle \Pi \frac{a_i}{\|a_i\|_2}, \Pi \frac{b_j}{\|b_j\|_2} \rangle - \langle \frac{a_i}{\|a_i\|_2}, \frac{b_j}{\|b_j\|_2} \rangle|^2$
where $\alpha = \|a_i\|_2^2 \|b_j\|_2^2$

# Analysis

$C_{i,j} = \langle \Pi A_{(i)}, \Pi B^{(j)} \rangle$ while $(AB)_{i,j} = \langle A_{(i)}, B^{(j)} \rangle$

Notation: $a_i$ for $A_{(i)}$ and $b_j$ for $B^{(j)}$

$\|AB - C\|_F^2 = \sum_{i,j} |\langle \Pi a_i, \Pi b_j \rangle - \langle a_i, b_j \rangle|^2$

Term by term:
$|\langle \Pi a_i, \Pi b_j \rangle - \langle a_i, b_j \rangle|^2 = \alpha |\langle \Pi \frac{a_i}{\|a_i\|_2}, \Pi \frac{b_j}{\|b_j\|_2} \rangle - \langle \frac{a_i}{\|a_i\|_2}, \frac{b_j}{\|b_j\|_2} \rangle|^2$
where $\alpha = \|a_i\|_2^2 \|b_j\|_2^2$

Applying JL property and linearity of expectation

$$\mathbf{E}\left[\|AB - C\|_F^2\right] \leq (c\epsilon)^2 \delta \sum_{i,j} \|a_i\|_2^2 \|b_j\|_2^2 \leq (c\epsilon^2)\delta \|A\|_F^2 \|B\|_F^2$$

# Analysis

$$\Pr[\|AB - C\|_F^2 > 3\epsilon\|A\|_F\|B\|_F^2] \leq \frac{\mathbf{E}\left[\|AB - C\|_F^2\right]}{(3\epsilon\|A\|_F\|B\|_F)^2}.$$

and

$$\mathbf{E}\left[\|AB - C\|_F^2\right] \leq (c\epsilon)^2\delta\sum_{i,j}\|a_i\|_2^2\|b_j\|_2^2 \leq (c\epsilon^2)\delta\|A\|_F^2\|B\|_F^2$$

hence

$$\Pr[\|AB - C\|_F^2 > 3\epsilon\|A\|_F\|B\|_F^2] \leq \delta.$$

# Running time

Roughly speaking: $\Pi$ converts vectors of dimension $n$ into vectors of dimension $d = O(\frac{1}{\epsilon^2}\log(1/\delta))$.

Need to compute $\Pi A^T$ and $\Pi B$ and then compute dot products.

$mp$ inner products of vectors of dimension $d$ which is $O(mpd^2)$ time in the worst case

Using Fast JL with very sparse $\Pi$ one can improve running time