Matrix Sketching

STREAMING ALGORITHMS FOR LOW-RANK MATRIX APPROXIMATIONS

Review …

- Last time
  - Singular Value Decomposition
  - Randomized SVD
- Today
  - Matrix Sketching

Upcoming Lectures

- Next Lecture
  - Security aspects of Big-Data
  - Guest lecture by Prof. Denning
- Next Week
  - Spark tutorial by TA
- Next Assignment
  - Choice of programming model
    - MPI, Hadoop, Spark
  - Low-rank approximation
    - Stochastic Gradient Descent, Randomized SVD, Sketching

Matrix Sketching

- Data is usually represented as a matrix
- For most Big Data applications, this matrix is too large for one machine
- In many cases, the matrix is too large to even fit in distributed memory
- Need to optimize for data access
  - Similar to our arguments for SGD for UV decomposition
- Streaming algorithm
  - Generate approximation by accessing (streaming) data once

Matrix Sketching

- Efficiently compute a concisely representable matrix B such that

$$\mathrm{B} \approx A \quad \text{or} \quad BB^T \approx AA^T$$

- Working with B is often “good enough”
  - Dimension reduction
  - Classification
  - Regression
  - Matrix Multiplication (approximate)
  - Recommendation systems

Frequent Directions (Edo Liberty '13)

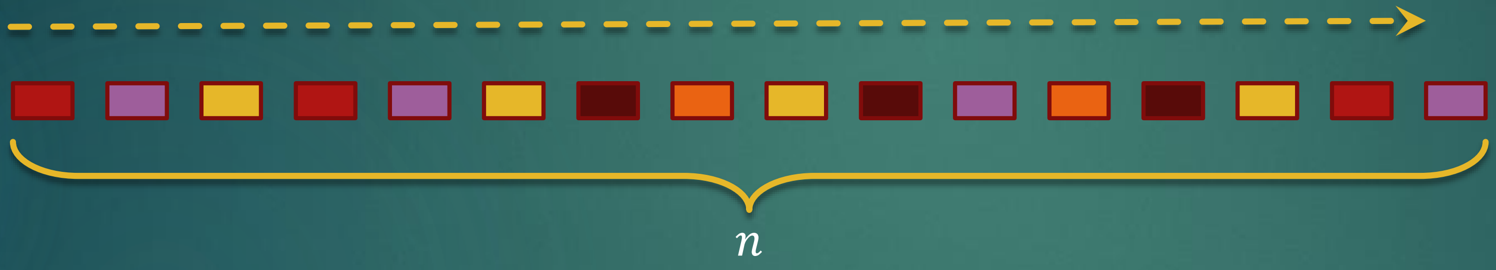
- Efficiently maintain a matrix B with only $l = 2/\epsilon$ columns such that,

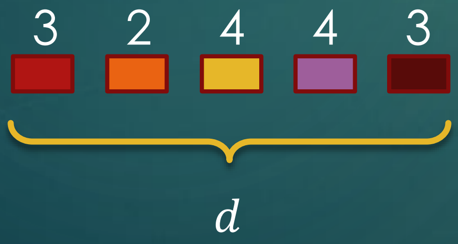
$$\|AA^T - BB^T\|_2 \leq \epsilon \|A\|_f^2$$

- Intuitive approach → extend frequent items
  - How to estimate the frequency of (frequent) items in a streaming fashion?

Frequent Items

Obtain the frequency $f(i)$ of each item in the stream





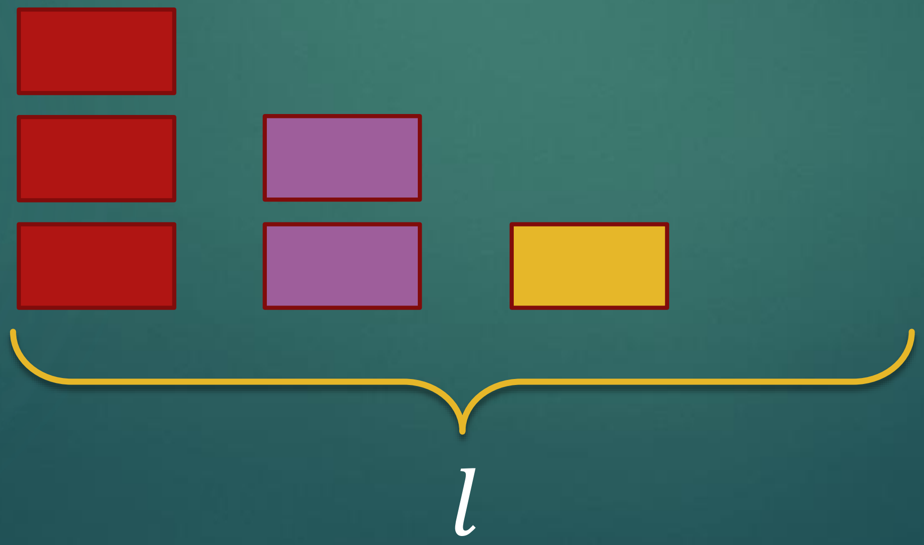
$f(\square) = 4$

What if d is very large?

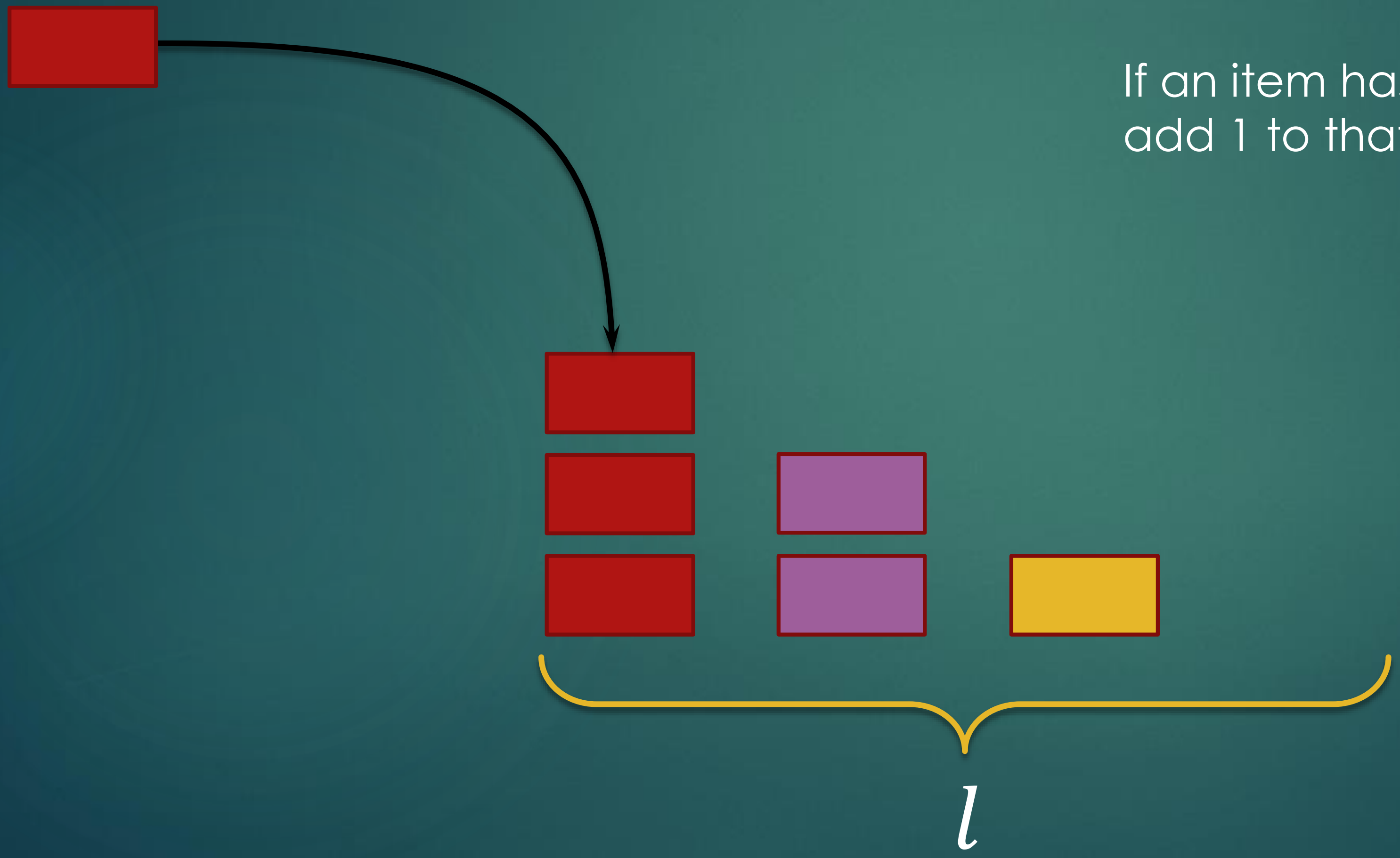
Frequent Items

Lets maintain less than $l < d$ counters (Misra-Gries)

- With at least 1 empty counter



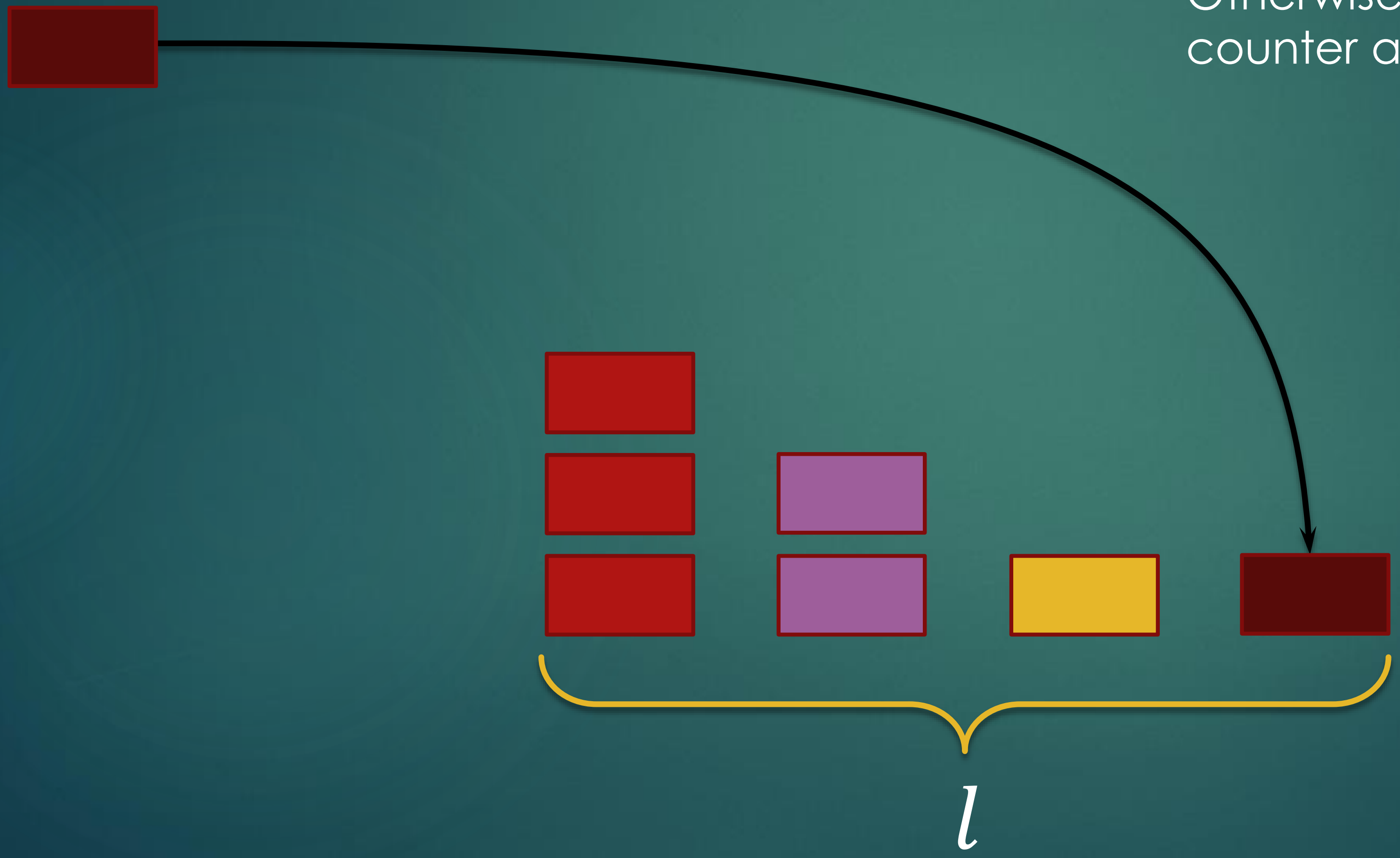
Frequent Items



If an item has a counter, add 1 to that counter

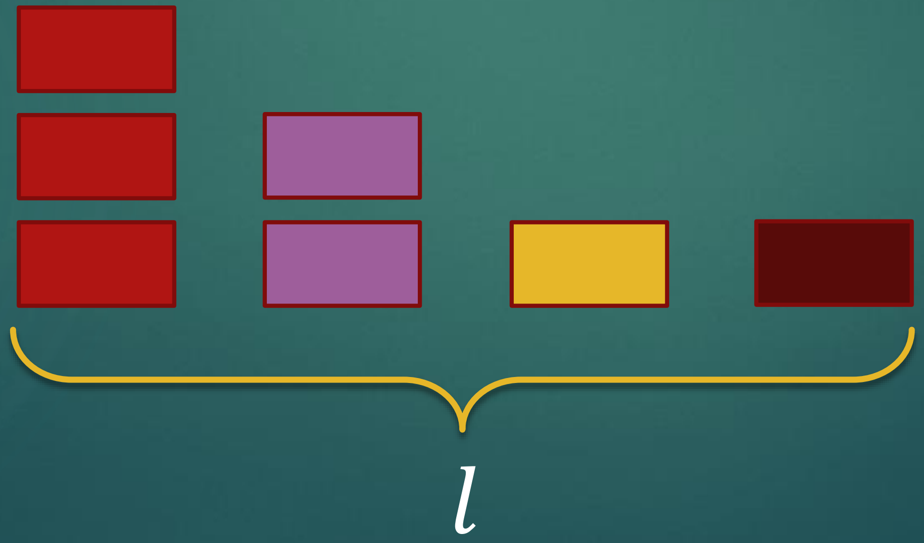
Frequent Items

Otherwise, create a new counter and set it to 1



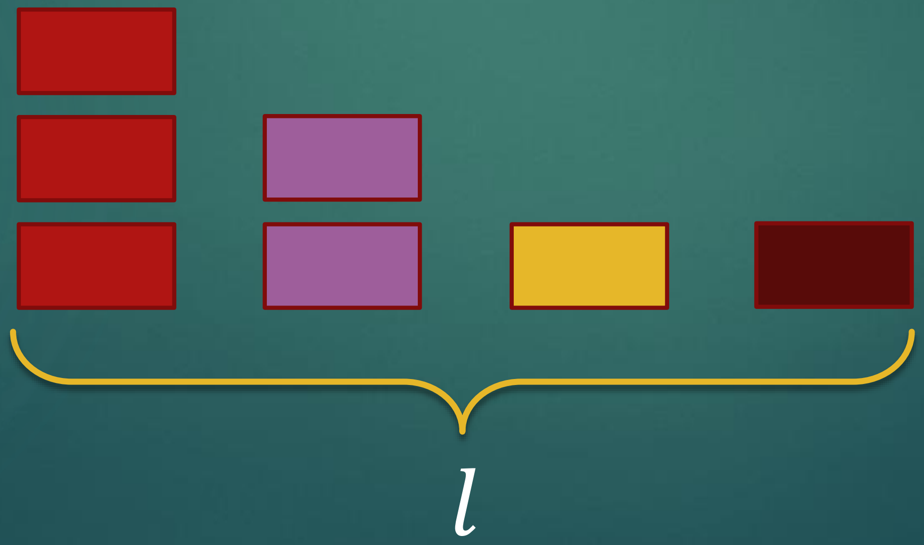
Frequent Items

But now we don't have
less than l counters



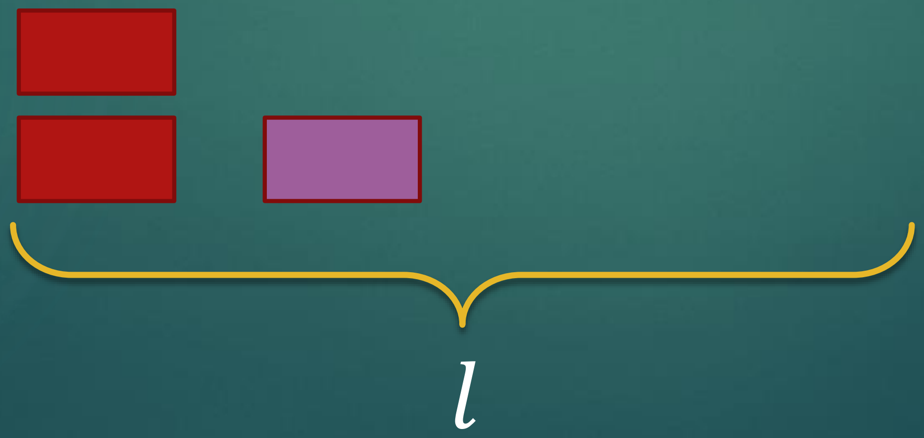
Frequent Items

Let δ be the median counter value at time t

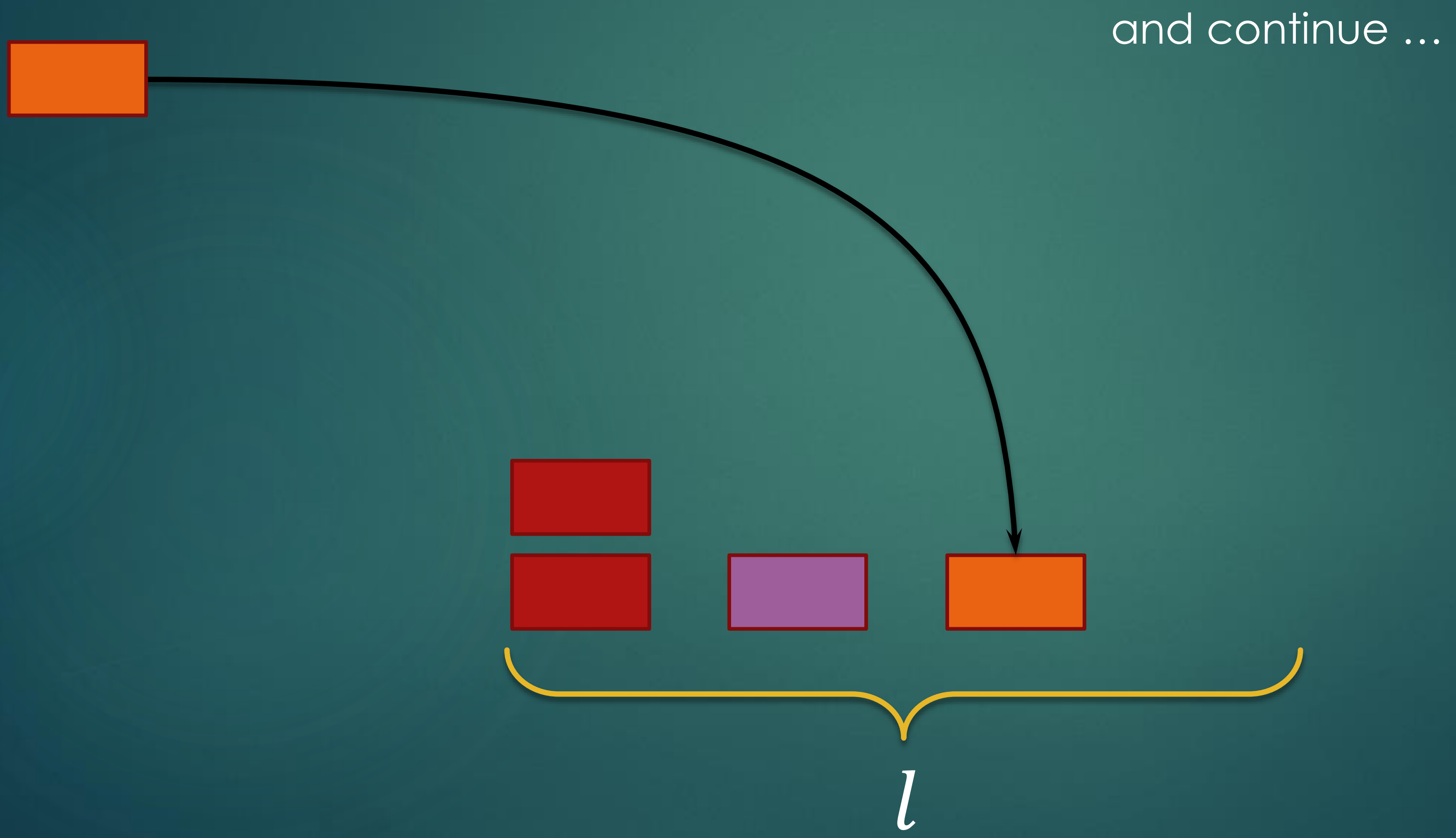


Frequent Items

Decrease all counters by δ

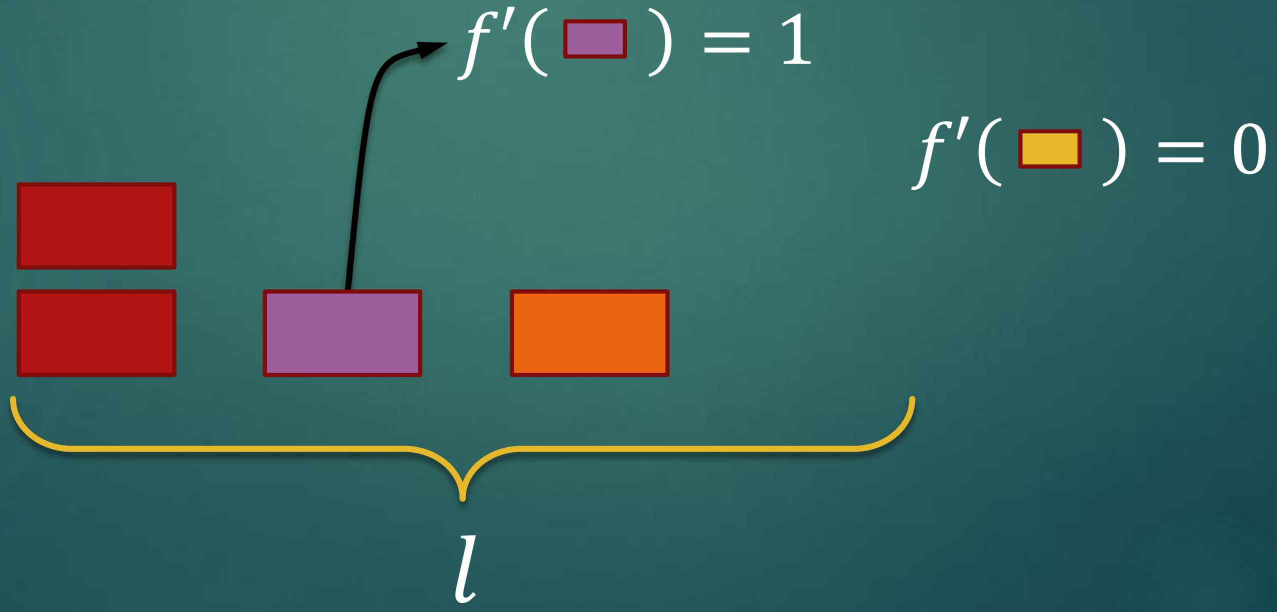


Frequent Items



Frequent Items

The approximated counts are f'



Frequent Items

- We increase the count by only 1 for each item appearance

$$f'(i) \leq f(i)$$

- Because we decrease each counter by at most δ_t at time t

$$f'(i) \geq f(i) - \sum_t \delta_t$$

- Calculating the total approximate frequencies:

$$0 \leq \sum_i f'(i) \leq \sum_t 1 - \frac{l}{2}\delta_t = n - \frac{l}{2}\sum_t \delta_t$$

$$\sum_t \delta_t \leq \frac{2n}{l}$$

- Setting $l = 2/\epsilon$ gives us

$$|f(i) - f'(i)| \leq \epsilon n$$

Frequent Directions (Liberty 2013)

maintain a sketch of
at most l columns

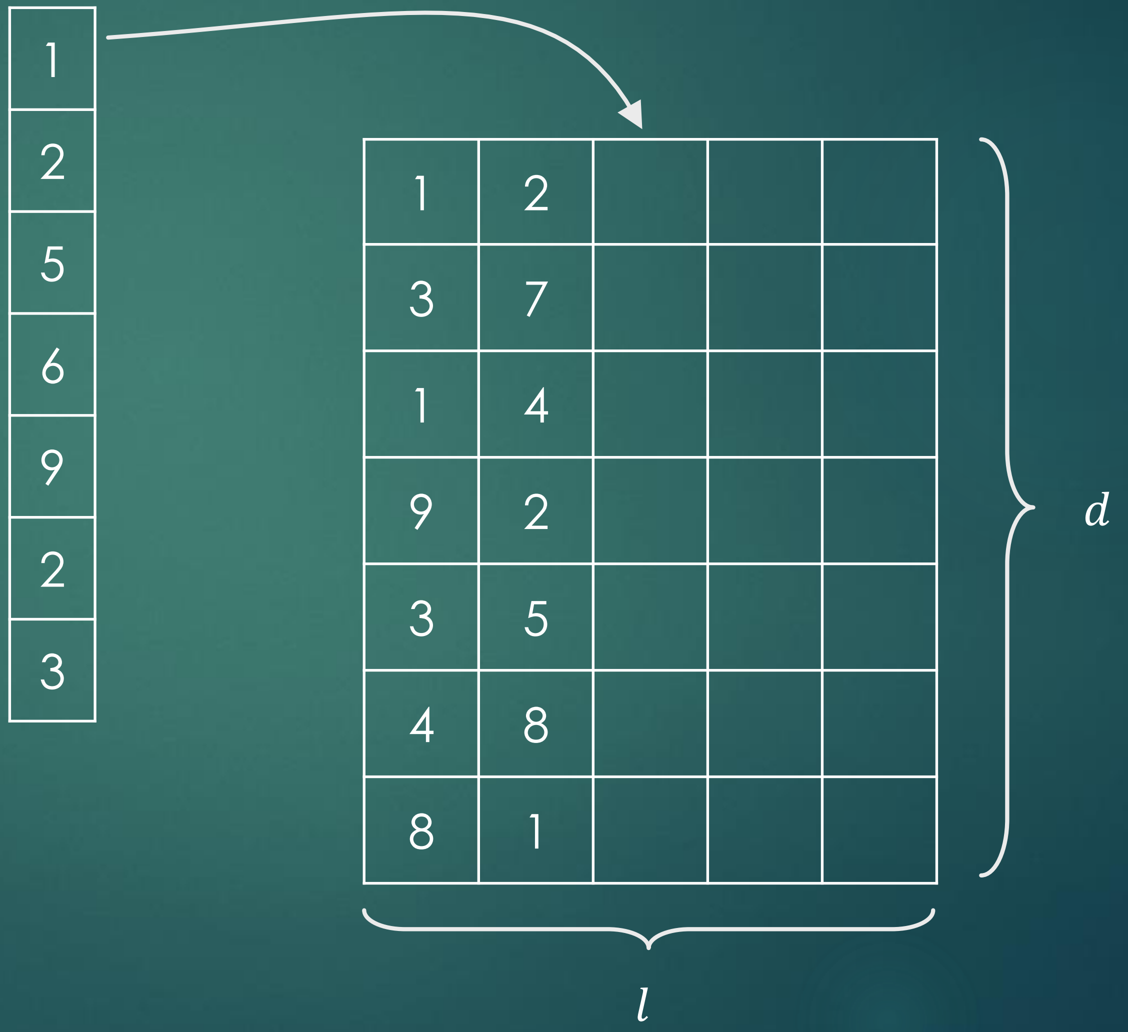
maintain the invariant
that some columns are
empty (zero-valued)

$$\underbrace{\left.\begin{array}{|c|c|c|c|c|}\hline 1 & 2 & & & \\ \hline 3 & 7 & & & \\ \hline 1 & 4 & & & \\ \hline 9 & 2 & & & \\ \hline 3 & 5 & & & \\ \hline 4 & 8 & & & \\ \hline 8 & 1 & & & \\ \hline \end{array}\right\} d}_{l}$$

Frequent Directions

Stream in matrix A one column at a time

Input vectors are simply stored In empty columns



Frequent Directions

When the sketch is ‘full’
We need to zero out some columns

$$\underbrace{\left.\begin{array}{|c|c|c|c|c|}
\hline
1 & 2 & 3 & 8 & 6 \\ \hline
3 & 7 & 5 & 7 & 0 \\ \hline
1 & 4 & 4 & 3 & 3 \\ \hline
9 & 2 & 2 & 3 & 7 \\ \hline
3 & 5 & 6 & 6 & 1 \\ \hline
4 & 8 & 1 & 5 & 4 \\ \hline
8 & 1 & 2 & 6 & 2 \\ \hline
\end{array}\right\} d}_{l}$$

Frequent Directions

Using SVD, compute

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

And set

$$B_{new} = U\Sigma$$

$$BB^T = B_{new}B_{new}^T$$

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

$$\underbrace{B_{new} = U\Sigma}\quad\underbrace{V^T}$$

-9.2	-0.4	-5.1	1.5	0.9
-10.4	-4.4	2.0	0.5	0.3
-6.5	-1.9	-1.1	-0.9	-1.9
-9.9	6.7	0.1	-1.1	-1.4
-9.6	-3.2	1.1	1.5	-1.4
-10.1	-0.8	0.3	-4.0	1.6
-9.0	3.8	2.1	2.7	1.3

Frequent Directions

The columns of B are now orthogonal and in decreasing magnitude

$$B_{new} = U\Sigma$$

-9.2	-0.4	-5.1	1.5	0.9
-10.4	-4.4	2.0	0.5	0.3
-6.5	-1.9	-1.1	-0.9	-1.9
-9.9	6.7	0.1	-1.1	-1.4
-9.6	-3.2	1.1	1.5	-1.4
-10.1	-0.8	0.3	-4.0	1.6
-9.0	3.8	2.1	2.7	1.3

d

l

Frequent Directions

Let $\delta = \|B_{l/2}\|^2$

$\delta = \|B_{l/2}\|^2$

-9.2	-0.4	-5.1	1.5	0.9
-10.4	-4.4	2.0	0.5	0.3
-6.5	-1.9	-1.1	-0.9	-1.9
-9.9	6.7	0.1	-1.1	-1.4
-9.6	-3.2	1.1	1.5	-1.4
-10.1	-0.8	0.3	-4.0	1.6
-9.0	3.8	2.1	2.7	1.3

d

l

Frequent Directions

Reduce column l_2^2-norms by δ

$$\underbrace{\left.\begin{array}{|c|c|c|c|c|}\hline -8.9 & -0.3 & & & \\ \hline -10.1 & -3.5 & & & \\ \hline -6.3 & -1.5 & & & \\ \hline -9.6 & 5.2 & & & \\ \hline -9.3 & -2.5 & & & \\ \hline -9.8 & -0.7 & & & \\ \hline -8.7 & 2.9 & & & \\ \hline \end{array}\right\} d}_{l}$$

Frequent Directions

1
2
5
6
9
2
3

$$\underbrace{\left.\begin{bmatrix} -8.9 & -0.3 & & & \\ -10.1 & -3.5 & & & \\ -6.3 & -1.5 & & & \\ -9.6 & 5.2 & & & \\ -9.3 & -2.5 & & & \\ -9.8 & -0.7 & & & \\ -8.7 & 2.9 & & & \end{bmatrix}\right\} d}_{l}$$

Start aggregating columns again

Frequent Directions

Input: $l,\ A \in \mathbb{R}^{d\times n}$

$B \leftarrow$ all zeros matrix $\in \mathbb{R}^{d\times l}$

for $i \in [n]$ **do**

Insert A_i into a zero valued column of B

if B has no zero valued columns **then**

$[U, \Sigma, V] \leftarrow SVD(B)$

$\delta \leftarrow \sigma_{l/2}^2$

$\hat{\Sigma} \leftarrow \sqrt{\max(\Sigma^2 - I_l\delta, 0)}$

$B \leftarrow U\hat{\Sigma}$

Return: B

Bounding the error

We first bound $\|AA^T - BB^T\|$

$$\sup_{\|x\|=1} \|xA\|^2 - \|xB\|^2 = \sup_{\|x\|=1} \sum_{t=1}^{n} [\langle x, A_t \rangle^2 + \|xB^{t-1}\|^2 - \|xB^t\|^2]$$

$$= \sup_{\|x\|=1} \sum_{t=1}^{n} [\|xC^t\|^2 - \|xB^t\|^2]$$

$$\leq \sum_{t=1}^{n} \left\| {C^t}^T C^t - {B^t}^T B^t \right\| \cdot \|x\|^2$$

$$= \sum_{t-1}^{n} \delta_t$$

Which gives,

$$\|AA^T - BB^T\| \leq \sum_{t=1}^{n} \delta_t$$

Bounding the error

We compute the Frobenius norm of the final sketch,

$$0 \leq \|B\|_f^2 = \sum_{t=1}^{n} \left[\|B^t\|_f^2 - \|B^{t-1}\|_f^2\right]$$

$$= \sum_{t=1}^{n} \left[\left(\|C^t\|_f^2 - \|B^{t-1}\|_f^2\right) - \left(\|C^t\|_f^2 - \|B^t\|_f^2\right)\right]$$

$$= \sum_{t=1}^{n} \|A_t\|^2 - tr\left({C^t}^T C^t - {B^t}^T B^t\right)$$

$$\leq \|A\|_f^2 - \frac{l}{2}\sum_{t=1}^{n} \delta_t$$

Which gives,

$$\sum_{t=1}^{n} \delta_t \leq \frac{2\|A\|_f^2}{l}$$

Bounding the error

We saw that:

$$\|AA^T - BB^T\| \leq \sum \delta_t$$

and,

$$\sum \delta_t \leq \frac{2\|A\|_f^2}{l}$$

setting $l = \frac{2}{\epsilon}$ gives us,

$$\|AA^T - BB^T\| \leq \epsilon \|A\|_f^2$$

Divide & Conquer

- Sketching can be implemented in a divide & conquer fashion as well
- Let $A = [A_1;\ A_2]$
- Compute the sketches B_1, B_2 of matrices A_1, A_2
- Compute the sketch C of the matrix $[B_1; B_2]$
- It can be shown that

$$\|AA^T - CC^T\| \leq \frac{2\|A\|_f^2}{l}$$

Streaming Matrices

- Can we compute the covariance matrix from a stream?

$$AA^T = \sum_{i=1}^{n} A_i A_i^T$$

- Naïve solution
  - For A of size $d \times n$ → AA^T of size $d \times d$
  - Compute in $\mathcal{O}(nd^2)$ with storage $\mathcal{O}(d^2)$