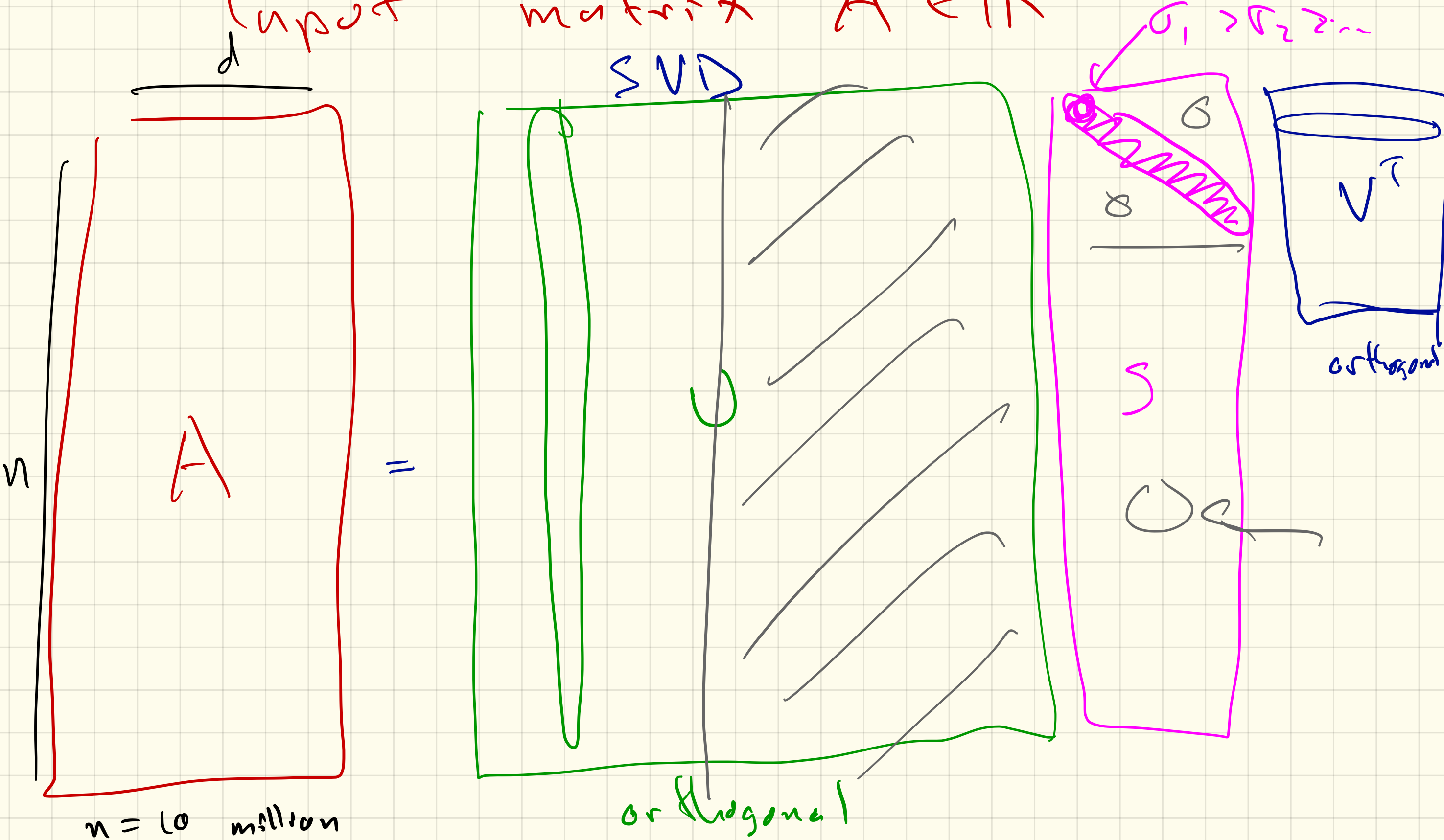
L17: Matrix Sketching

March 25, 2020

Input matrix $A \in \mathbb{R}^{n \times d}$



$n = 10$ million

$d = 100{,}000$

Eigen value Decomposition

input : square matrix $M \in \mathbb{R}^{d \times d}$

$$M v = v \lambda$$ ← eigenvalue

eigenvector $v \in \mathbb{R}^d$

$\|v\| = 1$

$$M = V L V^{-1}$$

V orthogonal $V^{-1} = V^T$

$$L = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_d \end{bmatrix}$$

$\lambda \geq 0$ real if M positive semidefinite

$$M_R = A^T A \in \mathbb{R}^{d \times d}$$

$$M_L = A A^T \in \mathbb{R}^{n \times n}$$

→ positive semidefinite

→ U, S^2

$$M_R = A^T A V \qquad A = USV^T$$

$$= (VSU^T)(USV^T)V = (VS\cancel{U^T})(\cancel{U}SV^T)(\cancel{V}) \qquad \cancel{U^T}\cancel{U} = I, \quad \cancel{V^T}\cancel{V} = I$$

$$= VS^2$$

$$S^2 = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_d^2 \end{bmatrix}$$

right sing. vectors v_j of A

are eigenvectors of M_R

sing values squared $\sigma_i^2 = \lambda_i$

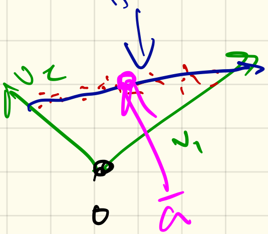
eigenvalues of M_R

Find subspace B (k-dim) $V_B = \{v_1, \dots, v_k\}$

PCA

minimize $SSE(A, B) = \sum_{i=1}^{n} \| a_i - \pi_B(a_i) \|^2$

(if say B contains 0, then SVD opt)

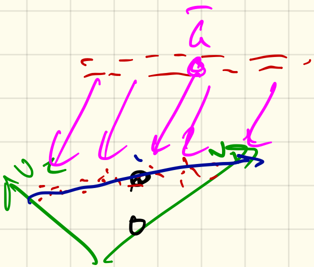


Sol: first center the data

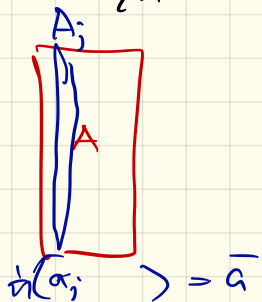
For each dimension $j \in [1 \dots d]$

find average value $\bar{a}_j = \frac{1}{n} \sum_{i=1}^{n} A_{i,j}$

$\bar{a} = (\bar{a}_1, \bar{a}_2, \dots, \bar{a}_d)$



$\tilde{A} = [A_{ij} - \bar{a}_j]_{ij}$



Centering Matrix

$$C_n = I_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T \qquad \mathbb{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$\underset{\text{identity}}{I_n}$

$$\tilde{A} = C_n A = A - \frac{1}{n} \mathbb{1}\mathbb{1}^T A$$

$$\text{svd}(\tilde{A}) = \bar{U} \hat{S} \tilde{V}^T$$

store $\bar{a} \in \mathbb{R}^d$

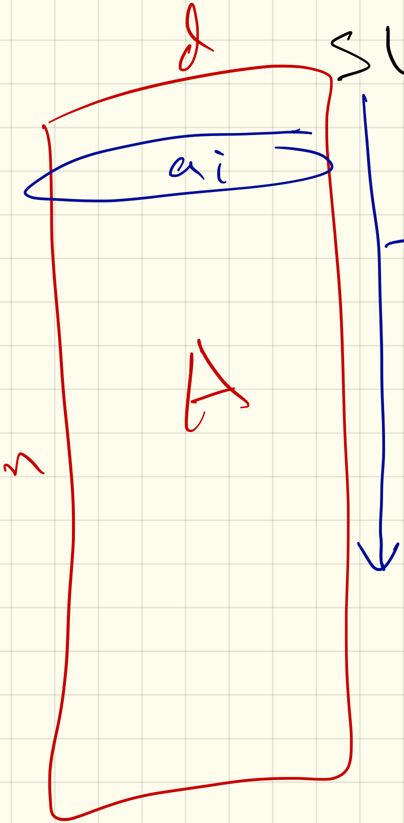
PCA

$\hat{V}^T$ = principal components

$\hat{S}$ = principal value

Very large scale

SVD take $O(nd^2)$ time $\quad n \gg d$



small space → $\hat{S}\,\hat{V}^T$

if d^2 fits in memory

```
B = zeros(d×d)
for i = 1 to n
    B += a_i a_i^T ∈ R^{d×d}
```

$B \mathrel{+}= a_i\, a_i^T \in \mathbb{R}^{d \times d}$

Return $B = M_R$

$l = k/\varepsilon$

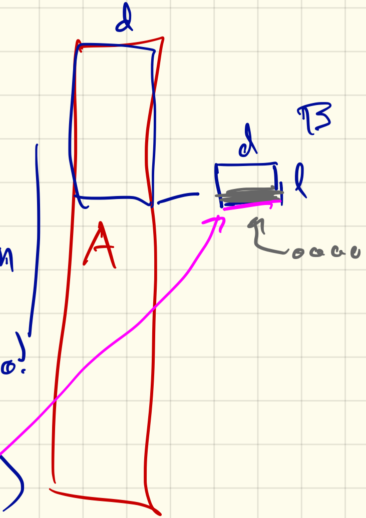
If d^2 too big but $\underline{(10 \cdot dk)}$ ok to fit in memory

Frequent Direction (Misra-Gries but for Matrix)

0. B zeros$(2\ell \times d)$

runtime $O(nd\ell^2)$ → $O(nd\ell)$

1. for $a_i \in A$
2. Insert a_i into all zero row of B
3. if (no more zero rows) ←
4. $[U, S, V^T] = \text{svd}(B)$
5. set $\delta_i = \sigma_\ell^2$ ← still ℓ
6. set $S' = \text{diag}\left(\sqrt{\sigma_1^2 - \delta_i}, \sqrt{\sigma_2^2 - \delta_i}, \ldots, \underbrace{0}_{\ell+1}\right)$ zero! 000
7. $B = S'V^T$
8. Return B



Freq. Dir $\quad B = (2\ell \times d)$

for all unit vectors $x \in \mathbb{R}^d$

$$0 \le \|Ax\|^2 - \|Bx\|^2 \le \frac{\|A - A_k\|_F^2}{(\ell - k)}$$

$\ell = k + 1/\varepsilon \qquad \varepsilon \cdot \|A - A_k\|_F^2$

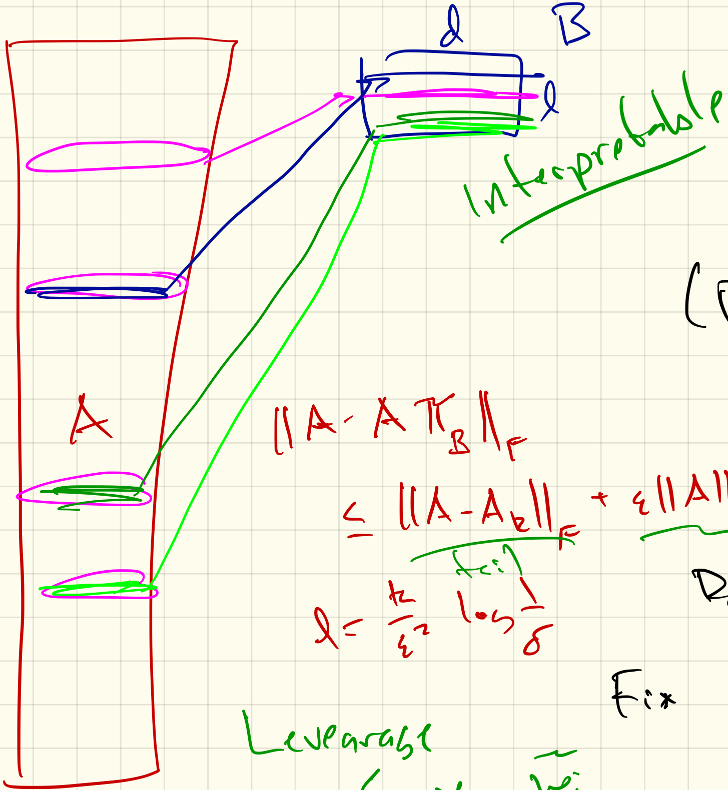
$$\|A - A\underbrace{\Pi_{B_k}}_{\tilde{A}_k}\|_F^2 \le \frac{\ell}{\ell - k}\|A - A_k\|_F^2$$

$\ell = k + k/\varepsilon$

$$\le \varepsilon\|A - A_k\|_F^2$$

Row Sampling

(Column Sampling)
$d > m$



Sample a_i proportional to $\|a_i\|^2 = w_i$

(Reservoir Sampling – sampling in stream)

$$W_i = \sum_{j=1}^{i} w_i$$

replace w/ prob $\frac{w_i}{W_i}$

Do ℓ times independently

Fix duplicate problem w/ Priority Sampling

$$\|A - A\pi_B\|_F \le \underbrace{\|A - A_k\|_F}_{\text{tail}} + \underline{\varepsilon \|A\|_F^2}$$

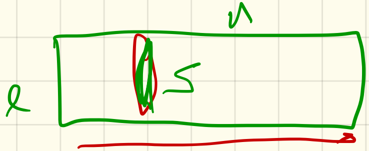
$$\ell = \frac{k}{\varepsilon^2} \log \frac{1}{\delta}$$

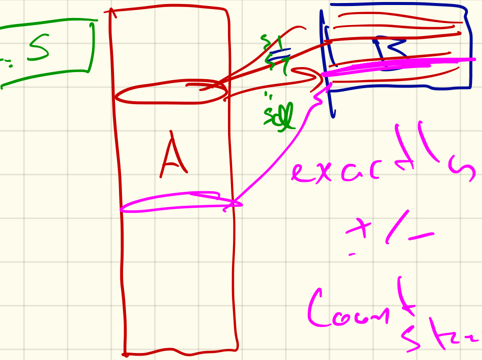
Leverage Score $\tilde{w}_i$

Random Projection / Count Sketch

$S \in \mathbb{R}^{\ell \times n}$ $\qquad$ $S_{ij} \sim N(0,1)\sqrt{\frac{n}{\ell}}$

sketch $B = SA \in \mathbb{R}^{\ell \times d}$





exactly 1 row

+/-

Count sketch

Let $[AV]_k$ = best rank k U RSV B

$$\| A - \underbrace{[AV]_k V^T}_{\tilde{A}_k} \|_F \le (1+\varepsilon) \| A - A_k \|_F^2$$

$\ell \approx k/\varepsilon$

$$(1-\varepsilon) \le \frac{\|Ax\|}{\|Bx\|} \le (1+\varepsilon) \qquad \ell \approx d/\varepsilon^2$$

$\forall x \in \mathbb{R}^d$

$\ell = d^2/\varepsilon^2$