1

HW III

Part 1

Randomized Matrix Multiplication

very large

$B \rightarrow m \times N \qquad C \rightarrow N \times n$

$A \rightarrow BC \rightarrow m \times n$

$$A = BC = \begin{bmatrix} b_1 & b_2 & \rightarrow & b_N \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \downarrow \\ c_N \end{bmatrix}$$

$$= \underbrace{\underset{m \times 1}{b_1} \, \underset{1 \times n}{c_1}}_{m \times n} + b_2 c_2 + \ldots + b_N c_N$$

$$\therefore A = \sum_{i=1}^{N} b_i c_i$$

b_i is i^{th} column of B

c_i is i^{th} corresponding row of C

2.

Consider

$B = [b_1 \longrightarrow b_N]$.

$C^T \rightarrow n \times N$

$C^T \rightarrow \begin{bmatrix} c_1^T \\ \downarrow \\ c_n^T \end{bmatrix} \longrightarrow 1 \times N$

And $b_i c_i^T \rightarrow (m \times 1)(1 \times N) \rightarrow m \times N$

$\downarrow$

Not the dimension of $A \rightarrow m \times n$

$\rightarrow$ C^T is a typo, it should just be C.

3

$$\hat{A} = \frac{1}{r} \sum_{t=1}^{r} \frac{1}{p_{J(t)}} b_{J(t)} c_{J(t)}$$

As discussed c^T is a typo,

c_k is a row of C

b_k is a column of B

$B = [b_1 \ b_2 \rightarrow b_N]$ $\quad C = \begin{bmatrix} c_1 \\ c_2 \\ \downarrow \\ c_N \end{bmatrix}$.

Also $A = \sum_{i=1}^{N} b_i c_i$

Thus we have $\hat{A}$ over r randomly selected columns of B and corresponding rows of C rather than full N columns & rows since N is very large

4

Understanding $J(t)$

→ $J(1)$

→ $J(2)$

↓

→ $J(n)$

all are independent RV's having identical distribution to J

$$J \rightarrow \{\overset{p_1}{\overset{\uparrow}{1}}, \overset{p_2}{\overset{\uparrow}{2}}, \longrightarrow, \overset{p_N}{N}\}$$

Discrete RV

with pmf →

$J = k$ has $p(J = k) = p_k$.

→ In each of the n summands, a column of b and row of c is correspondingly drawn at random from J and multiplied, each such selection is independent & identical

5

(i) Provide an U.B for

$$\frac{E\left[\|\hat{A} - A\|\right]}{\|B\| \; \|C\|} \text{ ?}$$

Consider

$$E[\hat{A}] = \frac{1}{\eta} E\left[\sum_{t=1}^{r} \frac{1}{P_{J(t)}} b_{J(t)} c_{J(t)}\right]$$

b, c, p are deterministic, but J(t) is random variable

Thus $\left[\frac{bc}{P}\right]_{J(t)} \rightarrow f(J(t))$

$$E[\hat{A}] = \frac{1}{\eta} E\left[\sum_{t=1}^{r} f(J(t))\right].$$

$$= \frac{1}{\eta} E\left[f(J(1)) + f(J(2)) + \rightarrow + f(J(\eta))\cdot\right]$$

6

7

→ Now

$$f(J(K)) = \underbrace{\left[\frac{bc}{P}\right]_{J(K)}}_{\text{random vector defined using } J(K)}.$$

This $f(J(K))$ is a Random vector (discrete)

say $\bar{Z}_K$

$$\bar{Z}_K \rightarrow \left\{\left[\frac{bc}{P}\right]_1, \longrightarrow, \left[\frac{bc}{P}\right]_N\right\}.$$

$$\underbrace{P_1 \longrightarrow P_N}_{\text{probability}}$$

So $E(\hat{A}) = \frac{1}{n} E\left[\bar{Z}_1 + \longrightarrow \bar{Z}_K\right].$

$\bar{Z}_1$ to $\bar{Z}_K$ are all iid (as seen).

8

So

$$E(\hat{A}) = \frac{1}{n}\left\{E\left[f(J(t))\right]\right\}_{\forall t} \times n$$

$$= E\left[\frac{1}{p_J} b_J c_J\right].$$

$$= \frac{b_1 c_1}{p_1} p_1 + \longrightarrow + \frac{b_N c_N}{p_N} p_N$$

$$= b_1 c_1 + \longrightarrow + b_N c_N$$

$$= A$$

So $E[\hat{A}] = A$

$\hat{A} \rightarrow$ Unbiased Estimator of A (deterministic A, $\hat{A}$ random).

9.

$$\frac{E\left[\|\hat{A} - A\|\right]}{\|B\|\|C\|} = ?$$

$$\hat{A} = \frac{1}{r}\sum_{t=1}^{r}\left[\frac{b_j c_j}{p_j}\right]_{j=J(t)}$$

Define

$$\left[\frac{b_j \ c_j}{p_j}\right]_{j=J} = R.$$

Then $\hat{A} = \frac{1}{r}\sum_{t=1}^{r} R_t$

where $\hat{A}$ is averaged over independently obtained R distributed identically

10.

$$R = \frac{b_J c_J}{p_J}$$

$$\|R\| \leq \max_J \left\| \frac{b_J c_J}{p_J} \right\|.$$

→ for $J \in \{1, 2, \rightarrow, N\}$

→ Noting : Matrix Bernstein

(Continued).

→ A result : Matrix Bernstein

$S_1, S_2, \ldots, S_r \to d_1 \times d_2$

$E(S_i) = 0 \ |$ matrix obviously. $\forall i \in \{1, 2, \ldots, n\}$

$S_i \to$ all independent & identical

$Z = \sum_{k=1}^{n} S_k$. $\qquad \|S_k\| \leq L$.

Define $\quad V(Z) = \max\left[\|E(ZZ^*)\|, \|E(Z^*Z)\|\right]$

$= \max\left[\|\sum_{k=1}^{n} E(S_k S_k^*)\|, \|\sum_{k=1} E(S_k^* S_k\right.$

Then

$$E(\|Z\|) \leq \sqrt{2V(Z)\log(d_1+d_2)} + \frac{1}{3}L\log(d_1+d_2)$$

12.

Lets try to apply Bernstein to obtain our bound.

$$E(\|\hat{A} - A\|).$$

$$\hat{A} - A \rightsquigarrow Z.$$

$S = ?$

Now $\hat{A} - A = \frac{1}{n}\sum_{t=1}^{r} R_t - A$

(R_t defined before)

Now,

$$\hat{A} - A = \frac{1}{n}\left\{\sum_{t=1}^{r} R_t - n A\right\}.$$

$$= \frac{1}{n}\left[\sum_{t=1}^{r} (R_t - A)\right].$$

$$= \sum_{t=1}^{r}\left(\frac{R_t - A}{n}\right)$$

13.

So $z \longrightarrow \hat{A} - A$

$$S_t = \frac{R_t - A}{n}$$

$$Z = \sum_{t=1}^{n} S_t$$

Assume total draws 'n' is fixed & not random (as such it is mostly fixed).

$$E\left(\frac{R_t - A}{n}\right) = ?$$

$$E(S_t) = \frac{1}{n}\left[E(R_t) - E(A)\right].$$

$$= \frac{1}{n}\left[E(R_t) - A\right].$$

14.

$$E(R_t) = E\left(\frac{b_{J(t)}\, c_{J(t)}}{p_{J(t)}}\right)$$

$$= \frac{b_1 c_1}{p_1} p_1 + \longrightarrow + \frac{b_N c_N}{p_N} p_N$$

$$= \sum_{i=1}^{N} b_i c_i = A.$$

So $E(S_t) = E\left(\frac{R_t - A}{\eta}\right) = 0$

Now to find $\mathcal{L}$,

$$\|S_t\| = \left\|\frac{R_t - A}{\eta}\right\| = \left\| \frac{\frac{b_{J(t)}\, c_{J(t)}}{p_{J(t)}} - A}{\eta} \right\|$$

$$\|S_t\| \leq \max_{J(t)} \left\| \frac{\frac{b_{J(t)}\, c_{J(t)}}{p_{J(t)}} - A}{\eta} \right\|$$

Thus $\mathcal{L}$ is obtained as

$$\mathcal{L} = \max_{J(t)} \left\| \underbrace{\frac{\frac{b_{J(t)}\, c_{J(t)}}{P_{J(t)}} - A}{\pi}} \right\|.$$

say π is constant.

$$\left\| \frac{\varphi}{\pi} \right\|, \exists \varphi$$

$$\longrightarrow \left\| \frac{\varphi}{\pi} \right\| = \max_{\|y\|=1} \left\| \frac{\varphi}{\pi} y \right\|$$

$$= \max_{\|y\|=1} \frac{1}{\pi} \| \varphi y \|$$

$$= \frac{1}{\pi} \max_{\|y\|=1} \| \varphi y \| = \frac{1}{\pi} \| \varphi \|.$$

$$\mathcal{L} = \max_{J(t)} \frac{1}{\pi} \left\| \frac{b_{J(t)}\, c_{J(t)}}{P_{J(t)}} - A \right\|.$$

$$= \frac{1}{\pi} \max_{J(t)} \left\| \frac{b_{J(t)}\, c_{J(t)}}{P_{J(t)}} - A \right\|.$$

→ Consider additionally

$$\|M - N\| \quad \forall\, M, N \text{ matrices of same dimns.}$$

Now,

$$\|M - N\| = \max_{\|y\|=1} \|(M - N)y\|$$

$$= \max_{\|y\|=1} \|My - Ny\|$$

Now

$$\left.\begin{aligned} \|a+b\| &\leq \|a\| + \|b\| \\ \|a-b\| &\geq \big|\, \|a\| - \|b\| \,\big| \end{aligned}\right\} \text{Triangle Inequality}$$

$$\|My - Ny\| \geq \big|\, \|My\| - \|Ny\| \,\big|$$

$$\max_{\|y\|=1} \|My - Ny\| \geq \max_{\|y\|=1} \big|\, \|My\| - \|Ny\| \,\big|$$

(obviously).

17.

Going through text reference, a usual approach used is thus:

Triangle inequality for matrix norms

$$\|M - N\| \leq \|M\| + \|N\|$$

(triangle inequality for matrix norms).

$$\max_{\|y\|=1} \|(M-N)y\| = ?$$

$$\|(M-N)y\| \leq \|My\| + \|Ny\|$$

$$\max_{\|y\|=1} \|(M-N)y\| \leq \max_{\|y\|=1} \|My\| + \max_{\|y\|=1} \|Ny\|$$

→ This in turn gives us:

$$L = \frac{1}{n} \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} - A \right\|$$

(for any t, over all values $J(t)$ takes):

$\|S_t\| \leq L$ (Conjectured bound).

$$\underbrace{\left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} - A \right\|}_{1} \leq \underbrace{\left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} \right\|}_{2} + \|A\|$$

$$\left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} - A \right\|_{J(t)=m} \leq \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} - A \right\|.$$

$$\leq \max_{J(t)=m} \left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} \right\| + \|A\|$$

Interpretation

Term

2 $\geq$ 1 } for same $J(t)$.

$2_{max=m} \geq 1_{at\ m}$ (obviously)

But $2_{k} \geq 1_{max=k}$. (obviously)

So $2_{max=m} \geq 2_{k}$

Combining, we have

$$2_{abs.\ max} \geq 2_{\text{when 1 is max}} \geq 1_{\text{abs max}} \geq 1_{\text{where 2 is max}}$$

$$2_{max} \geq 2_{corr} \geq 1_{max} \geq 1_{corr}.$$

$[2 \ \& \ 1]_{corr.}$ are points of 2 & 1 evaluation where the other maximizes

→ Thus we can conclude this :

$$\underbrace{① _{max} \le ② _{max}}_{\text{absolutes}}$$

So

$$\underbrace{\max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} - A \right\| \le \max_{J(t)} \left\{ \left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} \right\| + \|A\| \right\}}_{\text{individual absolute maximization of both}}.$$

$$\|S_k\| \le \left\{ \lambda = \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} - A \right\| \right\}$$

$$\le \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{P_{J(t)}} \right\| + \|A\|.$$

$\mathcal{L}$ is a stronger bound than

$$\mathcal{L}' = \left\{ \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| \right\} + \|A\|.$$

Now

$$R_t = \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \quad \left.\begin{array}{l} \forall t = \{1, 2, \ldots, q\} \\ J(t) \sim J \\ \text{iid note again} \end{array}\right\}$$

Now

$$\hat{A} = \frac{1}{n} \sum_{t=1}^{n} R_t.$$

Assumption : It is typical to assume this :

$$\|R_t\| \leq l \qquad \text{(small } l\text{)}.$$

$$\forall t.$$

Thus

$$\left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| \leq l \qquad \forall t.$$

$$\left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| \leq \underbrace{\max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\|} \leq l \qquad \forall t.$$

$\longrightarrow$ this maximum $\forall t$ is same and $\exists t$

we could have $J(t) \rightarrow$ as that value that maximizes $\left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\|$.

Specifically:

$$\left. \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| \leq \max_J \left\| \frac{b_J c_J}{p_J} \right\| \leq l \right\} \forall t$$

Thus also we have,

$\|R_t\| \leq l$ $\forall t$, over each t taking all values in $\{J / J(t)\}$.

$E(R_t) = A$ (proven before).

By Jensens Inequality (pg 82 of given ref.) we have,

$$\|E(R_t)\| \leq E(\|R_t\|).$$

$$\|A\| \leq E(\|R_t\|).$$

$$\|R_t\| \leq l$$

23

$$R_t = \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \rightarrow J(t) \sim J \quad \forall t, \text{ independent}$$

Also $R_t \rightarrow$ Random Matrix

$$\left.\begin{array}{l} \left\{ \dfrac{b_1 c_1}{p_1}, \dfrac{b_2 c_2}{p_2} \longrightarrow, \dfrac{b_N c_N}{p_N} \right\} \\ \quad \downarrow \qquad\quad \downarrow \qquad\qquad\qquad \downarrow \\ \quad p_1 \qquad\quad p_2 \qquad\qquad\qquad p_N \end{array}\right\} \forall t$$

$$\|R_t\| \rightarrow \left\{ \left\|\frac{b_1 c_1}{p_1}\right\|, \longrightarrow \left\|\frac{b_N c_N}{p_N}\right\| \right\}$$

$$p_1 \longrightarrow p_N.$$

each of these $\leq l$

Thus $E(\|R_t\|) = \sum_{i=1}^{N} p_i \left\|\frac{b_i c_i}{p_i}\right\|$.

If $\|R_t\|$ cannot, $\forall t$, be larger than l then its expectation, thus also must be smaller than or at the most equal to l

$$E(\|R_t\|) \leq (E(l) = l) \quad — *$$

$\star$ can be more rigorously proved, but is skipped here

$\star$ is nothing but Jensen's Inequality again

$\|A\| \leq E(\|R_t\|) \leq l$ $\quad$ ($\because \|R_t\| \leq l$)
(Jensen)
(see $\star$).

Thus

$\mathcal{L}' \leq 2l$

Also $\mathcal{L} \leq \mathcal{L}' \leq 2l$

where we assume

$\|R\| \leq l$.

25

For L.

$$\left.\begin{array}{l} \|S_t\| \leq \mathcal{L} \leq \mathcal{L}' \leq 2l \\ \|R\| \leq l \end{array}\right\}.$$

Tighter the bound, better the final bound in Bernstein.

→ For

$$\gamma(Z) = \max\left\{ \left\|\sum_k E(S_k S_k^*)\right\|, \left\|\sum_k E(S_k^* S_k)\right\| \right\}.$$

↳ Using reference & class notes, assuming this form.

$$S_k = \frac{1}{r}(R_k - A)$$

Using t instead of r

$$V(z) = \max \left\{ \left\| \sum_{t=1}^{n} E(S_t S_t^*) \right\| , \left\| \sum_{t=1}^{n} E(S_t^* S_t) \right\| \right\}.$$

$$S_t = \frac{1}{n}(R_t - A).$$

$*$ is conjugate transpose

Everything is real so $* \Rightarrow T$.

$$S_t S_t^T = \frac{1}{n^2}(R_t - A)(R_t - A)^T$$

$$= \frac{1}{n^2}\left((R_t - A)(R_t^T - A^T)\right)$$

$$= \frac{1}{n^2}\left(R_t R_t^T - R_t A^T - A R_t^T + A A^T\right).$$

$$= \frac{1}{n^2}\left(R_t R_t^T + A A^T - (R_t A^T + A R_t^T)\right).$$

$$S_t^T S_t = \frac{1}{n^2}(R_t - A)^T (R_t - A)$$

$$= \frac{1}{n^2}\left[(R_t^T - A^T)(R_t - A)\right]$$

$$= \frac{1}{n^2}\left[R_t^T R_t - R_t^T A - A^T R_t + A^T A\right]$$

$$= \frac{1}{n^2}\left[R_t^T R_t + A^T A - (R_t^T A + A^T R_t)\right].$$

$$\sum_{t=1}^{n} E(S_t S_t^T) = n\, E(S_t S_t^T) \quad \forall\, t$$

(obvious
can be proved
more elaborately
using theory
of Independence
& functions of
independent
RV's)

Thus

$$n\, E\left[\frac{1}{n^2}\left(R_t R_t^T + AA^T - (R_t A^T + A R_t^T)\right)\right]$$

$$= \frac{1}{n}\left\{E(R_t R_t^T) + E(AA^T) - E(R_t A^T) - E(A R_t^T)\right\}$$

$$R_t R_t^T = \frac{b_j c_j}{p_j}\left(\frac{b_j c_j}{p_j}\right)^T$$

$$= \frac{1}{p_j^2} b_j c_j c_j^T b_j^T$$

$$E(R_t R_t^T) = \sum_{i=1}^{N} \frac{1}{p_i} b_i c_i c_i^T b_i^T$$

28

$$E(AA^T) = AA^T$$
$$= \left(\sum_{i=1}^{N} b_i c_i\right)\left(\sum_{i=1}^{N} b_i c_i\right)^T$$
$$= \left(\sum_{i=1}^{N} b_i c_i\right)\left(\sum_{i=1}^{N} c_i^T b_i^T\right).$$

$$R_t A^T = \left(\frac{b_j c_j}{p_j}\right)\left(\sum_{i=1}^{N} c_i^T b_i^T\right)$$

$$E(R_t A^T) = \sum_{j=1}^{N} \frac{b_j c_j}{p_j}\left(\sum_{i=1}^{N} c_i^T b_i^T\right) p_j$$
$$= \sum_{j=1}^{N} b_j c_j \left\{\sum_{i=1}^{N} c_i^T b_i^T\right\}.$$
$$= \sum_{j=1}^{N} \left[b_j c_j \left[\sum_{i=1}^{N} c_i^T b_i^T\right]\right]$$

$$E(A R_t^T) = E\left(\left(\sum_{i=1}^{N} b_i c_i\right) \frac{c_j^T b_j^T}{p_j}\right).$$
$$= \sum_{j=1}^{N} \left[\left(\sum_{i=1}^{N} b_i c_i\right) c_j^T b_j^T\right]$$

2)

$$\sum_{t=1}^{r} E(S_t S_t^T) =$$

$$\frac{1}{r}\left[\left[\sum_{i=1}^{N} \frac{1}{p_i} b_i c_i c_i^T b_i^T\right] + \left[\left(\sum_{i=1}^{N} b_i c_i\right)\left(\sum_{i=1}^{N} c_i^T b_i^T\right)\right] - \left[\sum_{j=1}^{N} b_j c_j \left(\sum_{i=1}^{N} c_i^T b_i^T\right)\right] - \left[\sum_{j=1}^{N} \left(\sum_{i=1}^{N} b_i c_i\right) c_j^T b_j^T\right]\right]$$

Imp. Note, it is clear that

$$AA^T = E(R_t A^T) = E(A R_t^T).$$

$$\sum_{t=1}^{r} E(S_t S_t^T) = \frac{1}{r}\left[E(R_t R_t^T) - E(A R^T)\right]$$

$$= \frac{1}{r}\left[E\left\{R_t R_t^T - A R_t^T\right\}\right].$$

$$= \frac{1}{r} E((R_t - A) R_t^T).$$ ↳ Expectations always distributive absolutely.

30

$$\sum_{t=1}^{n} E(S_t S_t^T) =$$

$$\frac{1}{n}\left[\sum_{i=1}^{N}\frac{1}{p_i} b_i c_i c_i^T b_i^T - \sum_{i=1}^{N}\sum_{j=1}^{N} b_i c_i c_j^T b_j^T\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{N}\frac{1}{p_i} b_i c_i c_i^T b_i^T - \sum_{i=1}^{N}\left\{b_i c_i\left[\sum_{j=1}^{N} c_j^T b_j^T\right]\right\}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N}\left\{\frac{b_i c_i c_i^T b_i^T}{p_i} - b_i c_i\left(\sum_{j=1}^{N} c_j^T b_j^T\right)\right\}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{N}\left[b_i c_i\left(\frac{c_i^T b_i^T}{p_i} - \left(\sum_{j=1}^{N} c_j^T b_j^T\right)\right)\right]\right]$$

$$\left\|\sum_{t=1}^{n} E(S_t S_t^T)\right\| = \left\|\frac{1}{n}\sum_{i=1}^{N}\left[b_i c_i\left(\frac{c_i^T b_i^T}{p_i} - \sum_{j=1}^{N} c_j^T b_j^T\right)\right]\right\|$$

$$= \frac{1}{n}\left\|\sum_{i=1}^{N} b_i c_i\left(\frac{c_i^T b_i^T}{p_i} - \sum_{j=1}^{N} c_j^T b_j^T\right)\right\|$$

31.

$$\sum_{t=1}^{r} E(S_t^T S_t) = \frac{1}{r}\left[E(R_t^T R_t) + E(A^T A) - E(R_t^T A) - E(A^T R_t)\right].$$

$E(R_t^T R_t) =$

$$E\left[\frac{c_j^T b_j^T}{p_j}\frac{b_j c_j}{p_j}\right] = \sum_{i=1}^{N} \frac{1}{p_i} c_i^T b_i^T b_i c_i$$

$$E(A^T A) = A^T A = \left(\sum_{i=1}^{N} c_i^T b_i^T\right)\left(\sum_{i=1}^{N} b_i c_i\right)$$

$$E(R_t^T A) = E\left(\frac{c_j^T b_j^T}{p_j}\sum_{i=1}^{N} b_i c_i\right)$$

$$= \sum_{j=1}^{N} c_j^T b_j^T \sum_{i=1}^{N} b_i c_i$$

$$E(A^T R_t) = E(R_t^T A) = E(A^T A)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i^T b_i^T b_j c_j \quad \text{(as before)}.$$

$$\sum_{t=1}^{r} E(S_t^T S_t) = \frac{1}{r}\left[E(R_t^T R_t) - E(A^T R_t)\right]$$

$$= \frac{1}{r}\left[\sum_{i=1}^{N} \frac{1}{p_i} c_i^T b_i^T b_i c_i - \sum_{i=1}^{N} c_i^T b_i^T \sum_{j=1}^{N} b_j c_j\right].$$

32.

$$\sum_{t=1}^{r} E(s_t^T s_t) =$$

$$\frac{1}{r}\sum_{i=1}^{N}\left[c_i^T b_i^T\left(\frac{b_i c_i}{p_i} - \sum_{j=1}^{N} b_j c_j\right)\right]$$

$$\left\|\sum_{t=1}^{r} E(s_t^T s_t)\right\| =$$

$$\frac{1}{r}\left\|\sum_{i=1}^{N}\left[c_i^T b_i^T\left(\frac{b_i c_i}{p_i} - \sum_{j=1}^{N} b_j c_j\right)\right]\right\|$$

33

$$\left\| \sum_{t=1}^{r} E(S_t S^T) \right\|$$

$$= \frac{1}{r} \left\| \sum_{i=1}^{N} b_i c_i \left[\frac{c_i^T b_i^T}{p_i} - \sum_{j=1}^{N} c_j^T b_j^T \right] \right\|$$

$$= \frac{1}{r} \left\| \sum_{i=1}^{N} b_i c_i \left(\frac{c_i^T b_i^T}{p_i} - A^T \right) \right\|$$

$$= \frac{1}{r} \left\| \sum_{i=1}^{N} \frac{b_i c_i c_i^T b_i^T}{p_i} - \sum_{i=1}^{N} b_i c_i A^T \right\|$$

$$= \frac{1}{r} \left\| \sum_{i=1}^{N} \frac{b_i c_i c_i^T b_i^T}{p_i} - \left(\sum_{i=1}^{N} b_i c_i \right) A^T \right\|$$

$$= \frac{1}{r} \left\| \sum_{i=1}^{N} \frac{b_i c_i c_i^T b_i^T}{p_i} - AA^T \right\|$$

$$\le \frac{1}{r} \left[\left\| \sum_{i=1}^{N} \frac{b_i c_i c_i^T b_i^T}{p_i} \right\| + \left\| AA^T \right\| \right]$$

$\leftarrow$ triangle inequality

34.

$\|AA^T\| \leq \|A\|\|A^T\|$ ($\forall$ matrix norms)

$\|A\| \leq l$

We know,

$\|A^T\| = \|A\|$

So $\|AA^T\| \leq l^2$

($\|AA^T\| = \|A\|^2$;

$\|A\| \leq l \rightarrow \|A\|^2 \leq l^2 \rightarrow$

Also true

$\|AA^T\| \leq l^2$)

So $\|AA^T\| \leq l^2$.

35.

$$\left\| \sum_{t=1}^{r} E(S_t S_t^T) \right\| = \frac{1}{r} \left\| E(R_t R_t^T) - AA^T \right\|$$

$$\leq \frac{1}{r} \| E(R_t R_t^T) \| + \frac{1}{r} \| AA^T \|$$

$$\| E(R_t R_t^T) \| \leq E(\| R_t R_t^T \|)$$

$\hookrightarrow$ Jensen.

$$\left.\begin{array}{l} \| R_t \| \leq l \\ \| R_t^T \| \leq l \end{array}\right\} \text{assumed previously}$$

So $\| R_t R_t^T \| \leq l^2$

$$E(\| R_t R_t^T \|) \leq l^2$$

$$\| E(R_t R_t^T) \| \leq l^2$$

$$\left\| \sum_{t=1}^{r} E(S_t S_t^T) \right\| \leq \left\{ \left[\frac{1}{r} l^2 + \frac{1}{r} l^2 \right] = \frac{2 l^2}{r} \right\}.$$

36.

$$\left\|\sum_{t=1}^{r} E(s_t s_t^T)\right\| \le \frac{2l^2}{n}.$$

$$\left\|\sum_{t=1}^{r} E(s_t^T s_t)\right\| = \frac{1}{n}\left\|E(R_t^T R_t) - A^T A\right\|$$

$$\le \frac{1}{n}\left\|E(R_t^T R_t)\right\| + \frac{1}{n}\left\|A^T A\right\|$$

$$\left\|E(R_t^T R_t)\right\| \le E(\|R_t^T R_t\|) \le \|R_t^T R_t\|$$

$$\le l^2$$

$$\|A^T A\| \le l^2$$

similar to previous case

So $\left\|\sum_{t=1}^{n} E(s_t^T s_t)\right\| \le \frac{2l^2}{n}$

37.

$$V(z) = \max\left[\left\|\sum_{t=1}^{'} E(S_t^T S_t)\right\|, \left\|\sum_{t=1}^{'} E(S_t S_t^T)\right\|\right]$$

$$= \max\left[\leq \frac{2l^2}{n}, \leq \frac{2l^2}{n}\right].$$

$$V(z) \leq \frac{2l^2}{n}$$

Thus

$$V(z) \leq \frac{2l^2}{n}.$$

$$\|R\| \leq l.$$

Finally

$$V(Z) \leq \frac{2\ell^2}{n}$$

$$\|S_k\| \leq 2\ell$$

$$\|R\| \leq \ell$$

Thus

$$\overset{d_1 \times d_2}{\uparrow} \quad (\text{or } S \text{ is } d_1 \times d_2)$$

$$E(\|Z\|) \leq \sqrt{2V(Z)\log(d_1+d_2)} + \frac{L}{3}\log(d_1+d_2)$$

$$Z = \sum S_k$$

$S \rightarrow \dfrac{R - A}{n}$ here

$Z \longrightarrow \hat{A} - A$ here

$d_1 \times d_2 \longrightarrow m \times n$ here.

Substituting we get

$$E(\|\hat{A}-A\|) \leq \sqrt{2 \cdot \frac{2l^2}{r} \log(m+n)} + \frac{2l}{3} \log(m+n)$$

$$E(\|\hat{A}-A\|) \leq \sqrt{\frac{4l^2}{r} \log(m+n)} + \frac{2l}{3} \log(m+n)$$

$$E(\|\hat{A}-A\|) \leq 2l \sqrt{\frac{\log(m+n)}{r}} + \frac{2l}{3} \log(m+n)$$

$$\frac{E\|\hat{A}-A\|}{\|B\|\,\|C\|} \leq \frac{2l \sqrt{\frac{\log(m+n)}{r}} + \frac{2l}{3} \log(m+n)}{\|B\|\,\|C\|}$$

$$\frac{E\|\hat{A}-A\|}{\|B\|\|C\|} \leq \frac{2l}{\|B\|\|C\|}\left(\sqrt{\frac{\log(m+n)}{s}} + \frac{\log(m+n)}{3}\right).$$

Few points :

$\|R\| \longrightarrow \left\|\frac{b_j c_j}{p_j}\right\| \leq l$

$p \longrightarrow$ Not Known / Not obtained

→ Not the tightest bound for multiple reasons :

→ Tightest Bound of $\|S\|$, $V(Z)$ <u>not</u> taken, and <u>true</u> but loose tight bounds taken

41

This bound can surely be tightened "more"

(ii) $p_i = ?$

$$\|R_t\| \leq l$$

Consider again a few quantities,

$$\|S_t\| \leq \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} - A \right\|.$$

$$V(z) = \max \left\{ \frac{1}{n} \|E(R_t R_t^T) - AA^T\|, \frac{1}{n} \|E(R_t^T R_t) - A^T A\| \right\}$$

$$E\|\hat{A} - A\| \leq \underbrace{\sqrt{2V(z)\log(m+n)} + \frac{L}{3}\log(m+n)}.$$

To minimize this we need :

1) Tightest bounds on $V(z)$ & L

2) With the smallest value of $V(z)$ & L possible

Thus getting $V(z)$ & L as low as possible whilst obtaining their tightest bounds is our goal

We crudely took $\|R_t\| \leq l$, now we work with

$$\|R_t\| = \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\|.$$

$$\|S_t\| \leq \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} - A \right\|.$$

$$\leq \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| + \|A\|$$

assuming this as tight a bound as possible

$$\|S_t\| \leq \max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| + \|A\|$$

44.

$$\max_{J(t)} \left\| \frac{b_{J(t)} c_{J(t)}}{p_{J(t)}} \right\| \longrightarrow \text{minimized}$$

$\downarrow$

$$\max_{J(t)} \frac{1}{p_{J(t)}} \left\| b_{J(t)} c_{J(t)} \right\| .$$

Doing this "tightens $\|R\|$" bound.
optimizing over $V(z)$ too

$\rightarrow$ this is max say for $\exists\, J(t) = \alpha$

$\rightarrow$ $\frac{1}{p_\alpha} \| b_\alpha c_\alpha \|$

$\sum_{\alpha=1}^{N} p_\alpha = 1$ and p_α, $\forall \alpha \in \{1, 2, \rightarrow, N\}$
$\in [0,1]$

(should exclude 0 & 1 as such)

Satisfying conditions of probability we have, minimizing p_i $\exists i$, $\forall i \in \{1, \dots, N\}$

$$p_i = \frac{\|b_i\|^2 + \|c_i\|^2}{\|B\|_F^2 + \|C\|_F^2}$$

ref. pg 89 (6.4.3).

Actually optimizing over, $v(z)$, $\|S\|$, $\Sigma p = 1$, $p \in (0,1)$, will give these results (ref MIT OCW notes).

However intuitively this is clear as the class case for SVD dealt with only one matrix, here we have two B, C this gives rise towards the sum term

Taking this P_i, referring the text reference, we get (all steps similar), correspondingly

$$E\|\hat{A} - A\| \leq \sqrt{\frac{4\, asr \log(m+n)}{r}} + \frac{2\, asr \log(m+n)}{3r}$$

↓

similar to previous obtained bound

↓

assume $\|B\| = \|C\| = 1$

↓

$$asr = \frac{1}{2}\left[\frac{\|B\|_F^2}{\|B\|^2} + \frac{\|C\|_F^2}{\|C\|^2}\right].$$

$(\|B\| = \|C\| = 1$ → simplifies this to

$$\frac{1}{2}\left(\|B_F\|^2 + \|C\|_F^2\right).$$

↳ same bound for $\frac{E\|\hat{A} - A\|}{\|B\|\|C\|}$ ↑ taken

47.

(iii) $$\frac{E\|\hat{A}-A\|}{\|B\|\|C\|} \leq \varepsilon$$

how large is r

(with $\|B\| = \|C\| = 1$)

Thus

(assuming)

$$\sqrt{\frac{4asr\log(m+n)}{r}} + \frac{2asr\log(m+n)}{3r} \leq \varepsilon$$

$$2\sqrt{\frac{asr\log(m+n)}{r}} + \frac{2}{3}\frac{asr\log(m+n)}{r} \leq \varepsilon$$

+ve   +ve

$$\sqrt{\frac{asr\log(m+n)}{r}} = q \qquad (q \geq 0)$$

$$q^2 \rightarrow \frac{asr\log(m+n)}{r}$$

$$2q + \frac{2}{3}q^2 \leq \varepsilon \qquad (q \geq 0).$$

So $\frac{2}{3}q^2 + 2q - \varepsilon \leq 0$

$$\frac{2}{3}q^2 + 2q + (-\varepsilon) \leq 0$$

$$D = 4 - 4\left(\frac{2}{3}\right)(-\varepsilon)$$

$$= 4 + \frac{4 \cdot 2}{3}\varepsilon$$

$$= 4 + \frac{8\varepsilon}{3} = \frac{12 + 8\varepsilon}{3} = \frac{4(3 + 2\varepsilon)}{3}$$

$$= \frac{4}{3}(3 + 2\varepsilon).$$

$$= 4\left(1 + \frac{2\varepsilon}{3}\right).$$

$$x_1 = \frac{-2 + 2\sqrt{1 + \frac{2\varepsilon}{3}}}{\frac{4}{3}}$$

$$= \frac{3}{\cancel{4}_2}(\cancel{2})\left(-1 + \sqrt{1 + \frac{2\varepsilon}{3}}\right)$$

$$= \frac{3}{2}\left(\sqrt{1 + \frac{2\varepsilon}{3}} - 1\right)$$

49.

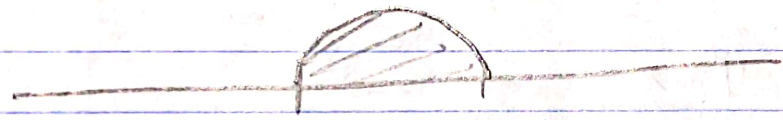
$$r_2 = -\frac{3}{2}\left(\sqrt{1+\frac{2\varepsilon}{3}}+1\right).$$

$r_1 \geq 0$

$r_2 \leq 0$

$(q - r_1)(q - r_2) \leq 0$

$(q - |r_1|)(q + |r_2|) \leq 0$



$-|r_2|$ $\qquad$ $|r_1|$

$q \in [0, |r_1|]$

$q \leq r_1$ $\qquad$ ($q \geq 0$, obvious)

$q^2 \leq r_1^2$

$$\frac{asr \log(m+n)}{n} \leq r_1^2$$

$$n r_1^2 \geq asr \log(m+n)$$

$$n \geq \frac{asr \log(m+n)}{r_1^2}$$

So

$$n \geq \frac{asr \log(m+n)}{\left[\frac{3}{2}\left(\sqrt{1+\frac{2\varepsilon}{3}}-1\right)\right]^2}$$

Thus

$$n \geq \frac{4}{9} \frac{asr \log(m+n)}{\left(\sqrt{1+\frac{2\varepsilon}{3}}-1\right)^2}$$

for error, $e \leq \varepsilon$

asr, m, n, ε all defined before.

as seen if $\varepsilon \downarrow$ n lower bound increases

51.

Computational cost as function of ε, m, n (N removed).

Normal $O(N m n)$

$$\downarrow$$

$$O(r m n)$$

$$\downarrow$$

$$O\left(\frac{4}{9} \frac{a s r \log(m+n)\, m n}{\left(\sqrt{1+\frac{2\varepsilon}{3}}-1\right)^2}\right)$$

Now $\left(1+\frac{2\varepsilon}{3}\right)^{\frac{1}{2}} \simeq 1 + \frac{1}{2}\left(\frac{2\varepsilon}{3}\right) \simeq 1 + \frac{\varepsilon}{3}$

$\left(\varepsilon/3 \lll 1 \text{ generally}\right)$.

$$O\left(\frac{4}{9} \frac{a s r \log(m+n)\, m n}{\left(1+\frac{\varepsilon}{3}-1\right)^2}\right).$$

52

$$O\left(\frac{4}{q}\,\frac{\text{asr}\,\log(m+n)\,mn}{\varepsilon^2}\,q\right)$$

$\downarrow$

$$O\left(4\,\varepsilon^{-2}\,\text{asr}\,\log(m+n)\,mn\right).$$

→ Thus to get

$E\|\hat{A}-A\| \leq \varepsilon$,

$\|B\| = \|C\| = 1$,

Complexity with r instead of N is

$O\left(4\,\varepsilon^{-2}\,\text{asr}\,\log(m+n)\,mn\right)$

as compared to

$O(Nmn)$ → great improvement

PART 1

END.