$AB = \sum A_{*t} B_{t*}$: $\quad A = \begin{bmatrix} | & | & | \end{bmatrix} \quad B = \begin{bmatrix} — \\ — \\ — \end{bmatrix}$

$= \sum_{t=1}^{n} a_{*t} b_{t*}$

$$AB = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji} b_{ij} = \sum_{i=1}^{n} a_{*i} b_{i*}$$

Use independent & identically distributed (iid) trials to randomly sample rank-one mats $(A_{*i} B_{i*})$

Gives idea for algorithm:

Alg 1

Input: $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $c \in [1:n]$, $\{p_k\}$

Function:

For $t = 1:c$

- $i_t = \text{RandChoose}([1:n], \{p_k\}, \text{Replace})$ ← Why replace?
- $C_{*t} = \dfrac{1}{\sqrt{c\, p_{i_t}}} A_{*i_t}$
- $R_{t*} = \dfrac{1}{\sqrt{c\, p_{i_t}}} B_{i_t *}$

Output: $\sum_{t=1}^{c} C_{*t} R_{t*} = \sum_{t=1}^{c} \frac{1}{c\, p_{i_t}} A_{*i_t} B_{i_t *} = CR$

Before addressing the effectiveness, consider the following:

Define $S \in \mathbb{R}^{n \times c}$ by $S_{k,j} = \begin{cases} \frac{1}{\sqrt{c \cdot p_{i_t}}}, & k = i_t, j = t \\ 0, & \text{otherwise} \end{cases}$

Then, $C = AS$ & $R = S^T B$

$\boxed{\text{Sampling \& Rescaling Formalism!}}$

Note: 1) Approach can use any algorithm for producing matrix multiplication

2) Approach unaffected by sparsity/density of matrix

3) Algorithm can be implemented in one pass over the input matrices (Given $\{p_k\}$)

Huh?

Lemma 21 $E[(CR)_{ij}] = (AB)_{i,j}$

$$\operatorname{Var}[(CR)_{ij}] \le \frac{1}{c}\sum_{k=1}^{n} \frac{A_{ik}^2 B_{kj}^2}{p_k}$$

Pf

Fix i,j. For $t=[1:c]$, def $X_t := \left(\frac{A_{*i_t} B_{i_t *}}{c\, p_{i_t}}\right)_{i,j}$

Then $X_t = \frac{A_{i i_t} B_{i_t j}}{c\, p_{i_t}}$. Hence,

$$E[X_t] = E\left[\frac{A_{i i_t} B_{i_t j}}{c\, p_{i_t}}\right] = \sum_{k=1} p_k \frac{A_{ik} B_{kj}}{c\, p_k}$$

$E[X] = \sum_{x\text{'s possible}} x \Pr[X=x]$ pg 13

$$= \frac{1}{c}\sum_{k=1}^{n} A_{ik} B_{kj} = \frac{1}{c}(AB)_{ij}$$

$$(CR)_{i,j} = \sum_{t=1}^{c} X_t \Rightarrow E[(CR)_{ij}] = \sum_{t=1}^{c} E[X_t] = (AB)_{ij} \checkmark$$

$\Rightarrow CR$ unbiased estimator, regardless of $\{p_k\}$...

$$\operatorname{Var}[(CR)_{ij}] = \operatorname{Var}\left[\sum_{t=1}^{c} X_t\right] = \sum_{t=1}^{c} \operatorname{Var}[X_t]$$

Pg 13: $\operatorname{Var}[X] = E[X^2] - E[X]^2 \Rightarrow \operatorname{Var}[X] \le E[X^2]$.

$$\Rightarrow \operatorname{Var}[X_t] \le E[X_t^2] = \sum_{k=1}^{n} \frac{A_{ik}^2 B_{kj}^2}{c^2\, p_k}$$

$$\therefore \operatorname{Var}[(CR)_{ij}] \le c\sum_{k=1}^{n} \frac{A_{ik}^2 B_{kj}^2}{c^2\, p_k} = \frac{1}{c}\sum_{k=1}^{n} \frac{A_{ik}^2 B_{kj}^2}{p_k}$$

$\square$

Analysis Cont

Thm 22 $$E\left[\|AB - CR\|_F^2\right] \le \sum_{k=1}^{n} \frac{\|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2}{c\, p_k}$$

$$E\left[\|AB - CR\|_F^2\right] = \sum_{i=1}^{m} \sum_{j=1}^{p} E\left[(AB - CR)_{ij}^2\right]$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 \quad \& \quad E\left[(AB - CR)_{ij}\right] = 0$$

$$\Rightarrow E\left[(AB - CR)_{ij}^2\right] = \mathrm{Var}\left[(CR - AB)_{ij}\right] = \mathrm{Var}\left[(CR)_{ij}\right]$$

Distribute... Pg 14

By Lemma 21,

$$\mathrm{Var}\left[(CR)_{ij}\right] \le \frac{1}{c} \sum_{k=1}^{n} \frac{A_{ik}^2 \, B_{kj}^2}{p_k}.$$

Hence,

$$\sum_{i=1}^{m} \sum_{j=1}^{p} E\left[(AB - CR)_{ij}^2\right] \le \frac{1}{c} \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} \frac{A_{ik}^2 \, B_{kj}^2}{p_k}$$

$$= \frac{1}{c} \sum_{k=1}^{n} \frac{1}{p_k} \left(\sum_{i=1}^{m} A_{ik}^2\right) \left(\sum_{j=1}^{p} B_{kj}^2\right)$$

$$\equiv \frac{1}{c} \sum_{k=1}^{n} \frac{1}{p_k} \|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2$$

$$\therefore E\left[\|AB - CR\|_F^2\right] \le \sum_{k=1}^{n} \frac{\|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2}{c\, p_k}$$

Analysis Cont

Let $$p_k = \frac{\|A_{*k}\|_2 \, \|B_{k*}\|_2}{\sum_{\ell=1}^{n} \|A_{*\ell}\|_2 \, \|B_{\ell *}\|_2}.$$

Then $$\frac{1}{c}\sum_{k=1}^{n} \frac{1}{p_k} \|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2 =$$

$$= \frac{1}{c} \sum_{k=1}^{n} \|A_{*k}\|_2 \, \|B_{k*}\|_2 \sum_{\ell=1}^{n} \|A_{*\ell}\|_2 \|B_{\ell *}\|_2 = \frac{1}{c}\left(\sum_{k=1}^{n} \|A_{*k}\|_2 \|B_{k*}\|_2\right)^2$$

We show this choice of $\{p_k\}$ minimizes $E\left[\|AB - CR\|_F^2\right]$:

Def $$f(p_1, \ldots, p_n) = \sum_{k=1}^{n} \frac{1}{p_k} \|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2.$$

Minimize f subject to $\sum p_k = 1$. Use Lag multiplier!

$$g(p_1, \ldots, p_n, \lambda) := f(p_1, \ldots, p_n) + \lambda\left(\sum_{k=1}^{n} p_k - 1\right)$$

Diff w.r.t p_k & λ, then set $= 0$.

$$\frac{\partial g}{\partial p_k} = -\frac{1}{p_k^2} \|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2 + \lambda = 0$$

$$\Rightarrow p_k = \frac{\|A_{*k}\|_2^2 \, \|B_{k*}\|_2^2}{\sqrt{\lambda}}$$

$$\frac{\partial g}{\partial \lambda} = \sum p_k - 1 = 0 \Rightarrow \sqrt{\lambda} = \sum_{\ell=1}^{n} \|A_{*\ell}\|_2 \, \|B_{\ell *}\|_2$$

(Minimizers $\because \frac{\partial^2 g}{\partial p_k^2} > 0 \quad \forall k$)

Other choices for $\{p_k\}$ $\boxed{H.2}$

1. Suppose $\sum p_k = 1$ & $p_k \geq \dfrac{\beta \|A_{*k}\|_2 \|B_{k*}\|_2}{\sum_{\ell}^{n} \|A_{*\ell}\|_2 \|B_{\ell *}\|_2}$ for $0 < \beta \leq 1$. Then,

$$E\left[\|AB - CR\|_F^2\right] \leq \frac{1}{\beta c}\left(\sum_{k=1}^{n} \|A_{*k}\|_2 \|B_{k*}\|_2\right)^2$$

<u>Pf</u>

$$E\left[\|AB - CR\|_F^2\right] \leq \frac{1}{c}\sum_{k=1}^{n} \frac{1}{p_k}\|A_{*k}\|_2^2 \|B_{k*}\|_2^2$$

$$\leq \frac{1}{\beta c}\left(\sum_{k=1}^{n} \|A_{*k}\|_2 \|B_{k*}\|_2\right)^2$$

2. $\sum p_k = 1$ & $p_k \geq \dfrac{\beta_k \|A_{*k}\|_2^2}{\|A\|_F^2}$, $0 < \beta \leq 1$.

Then, $E\left[\|AB - CR\|_F^2\right] \leq \dfrac{1}{\beta c}\|A\|_F^2 \|B\|_F^2$

<u>Pf</u>

$$E\left[\|AB - CR\|_F^2\right] \leq \frac{1}{\beta c}\|A\|_F^2 \sum_{k=1}^{n} \|B_{k*}\|_2^2 = \frac{1}{\beta c}\|A\|_F^2 \|B\|_F^2$$

3. $\sum p_k = 1$ & $p_k \geq \frac{\beta \|B_{k*}\|_2^2}{\|B\|_F^2}$, $0 < \beta \leq 1$

Then $E\left[\|AB - CR\|_F^2\right] \leq \frac{1}{\beta c}\|A\|_F^2 \|B\|_F^2$

Pf identical to #2.

$$\left(\sum_{k=1}^{n} \|A_{*k}\|_2 \|B_{k*}\|_2\right)^2 \leq \sum_{k=1}^{n} \|A_{*k}\|_2^2 \sum_{k=1}^{n} \|B_{k*}\|_2^2$$

$$= \|A\|_F^2 \|B\|_F^2$$

So #1 bound generally better than #2 & 3.

4.3

Suppose $col(U) \subseteq col(A)$. Spans "important" part of the $col(A)$.

Interested in calculating $U^T U$.

Let $U \in \mathbb{R}^{n \times d}$ $(n >> d)$ $\quad U = \begin{bmatrix} | & | & | & | \end{bmatrix}$

Use idea of sampling & rescaling: Let $R \in \mathbb{R}^{c \times d}$ be a ~~sample~~ of c rescaled rows of U.

By Thm 22:

$$E\left[\|U^T U - R^T R\|_F^2\right] = E\left[\|I_d - R^T R\|_F^2\right] \leq \frac{d^2}{\beta c}$$

"Leverage Scores"

p_k chosen to sat: $\sum p_k = 1$ & $p_k \geq \frac{\beta \|U_{k*}\|_2}{d}$

Markov Inequality:

$$Pr[X \geq \alpha] \leq \frac{E[X]}{\alpha}$$

Hence,

$$P\left[\|I_d - R^T R\|_F^2 \geq \varepsilon\right] \leq \frac{d^2}{\varepsilon \beta c}$$

Choose $c = \frac{10 d^2}{\beta \varepsilon^2}$

$$P\left[\|I_d - R^T R\|_F^2 \geq \varepsilon\right] \leq \frac{d^2}{\varepsilon \beta}\left(\frac{\beta \varepsilon^2}{10 d^2}\right) = \frac{\varepsilon}{10}$$

$$\Updownarrow$$

$$P\left[\|I_d - R^T R\|_F^2 < \varepsilon\right] > 1 - \frac{\varepsilon}{10}$$

← Why doe notes call this at least 90%?

Since $\|A\|_2 \leq \|A\|_F$, can also conclude

$$\|I_d - R^T R\|_2 \leq \varepsilon$$ w/ 90% confidence

(& using the same c).

Thm 36 $U \in \mathbb{R}^{n \times d}$ $(n >> d)$, $U^T U = I_d$.

Construct R as described at beg. of §4.3.

Let $$c \geq \frac{96\, d}{\beta \varepsilon^2} \ln\left(\frac{96 d}{\beta \varepsilon^2 \sqrt{\delta}}\right).$$

Then w/ prob $\geq 1-\delta$:

$$\|U^T U - R^T R\|_2 = \|I_d - R^T R\|_2 \leq \varepsilon.$$

Need Lemma for pf

Lemma 38 Let $x^1, \ldots, x^c$ be i.i.d copies of a d-dimensional rand vector x, w/

$$\|x\|_2 \leq M \quad \& \quad \|E[x x^T]\|_2 \leq 1.$$

Then $\forall \alpha > 0$:

$$\left\| \frac{1}{c} \sum_{i=1}^{c} x^i x^{i\,T} - E[x x^T] \right\|_2 \leq \alpha$$

holds w/ prob at least

$$1 - (2c^2) \exp\left(- \frac{c\alpha^2}{16M^2 + 8M^2\alpha}\right)$$

Pf of Thm

Define random row vec. $y \in \mathbb{R}^d$ as

$$P\left[y = \frac{1}{\sqrt{p_k}} U_{k*}\right] = p_k \geq \frac{\beta \, \|U_{k*}\|_2^2}{d},$$

for $k = 1:n$, "y is rescaled k^{th} row of U w/ prob p_k"

$$\Rightarrow R = \begin{bmatrix} \text{---} \frac{1}{\sqrt{c}} y^{(1)} \text{---} \\ \text{---} \frac{1}{\sqrt{c}} y^{(2)} \text{---} \\ \vdots \\ \text{---} \frac{1}{\sqrt{c}} y^{(c)} \text{---} \end{bmatrix}$$

i.i.d copies of y

$\|U_{k*}\|_2$ bounded $\Rightarrow \|y\|_2$ bounded.

$$\left(R^T R = \frac{1}{c} \sum_{t=1}^{c} y^{t^T} y^t\right)$$

$$\Rightarrow \mathbb{E}[y^T y] \equiv \sum_{k=1}^{n} p_k \left(\frac{1}{\sqrt{p_k}} U_{k*}^T\right)\left(\frac{1}{\sqrt{p_k}} U_{k*}\right) = U^T U = I_d.$$

(y is called "isotropic")

So $\|E[y^T y]\|_2 = 1$

Can apply Lemma!

By prev Lemma

$$\left\|\frac{1}{c} \sum y^{t^T} y^t - U^T U\right\|_2 = \|R^T R - I_d\|_2 < \varepsilon$$

With prob $\geq 1 - (2c^2) \exp\left(-\frac{c\,\varepsilon^2}{16M^2 + 8M^2\varepsilon}\right)$.

Let δ be failure prob. Then we seek c s.t.

$(2c)^2 \exp\left(\frac{-c\,\varepsilon^2}{16M^2 + 8M^2\varepsilon}\right) \leq \delta$. ~~Also require~~ $\delta \leq 1$. So,

$$\exp\left(\frac{1}{2} \frac{c\,\varepsilon^2}{16M^2 + 8M^2\varepsilon}\right) \leq \frac{(2c)}{\sqrt{\delta}} \Rightarrow \frac{\frac{1}{2} c\,\varepsilon^2}{16M^2 + 8M^2\varepsilon} \leq \ln\left(\frac{2c}{\sqrt{\delta}}\right)$$

$$\Rightarrow \frac{c}{\ln(2c/\sqrt{\delta})} \geq \frac{2}{\varepsilon^2}\left(16M^2 + 8M^2\varepsilon^2\right)$$

Pf cont

$$\|y\|_2 = \frac{1}{\sqrt{p_k}} \|U_{k*}\|_2 \quad , \quad p_k \geq \frac{\beta \|U_{k*}\|_2^2}{d}$$

$\Rightarrow \|y\|_2 \leq M \leq \sqrt{\frac{d}{\beta}}$ Suppose $\varepsilon < 1$. Then

$$\frac{2}{\varepsilon^2}\left(16M^2 + 8M^2\varepsilon\right) \leq \frac{2}{\varepsilon^2}\left(16\frac{d}{\beta} + 8\frac{d}{\beta}\right) = \frac{48d}{\beta\varepsilon^2}$$

So now find c satisfying: $\frac{c}{\ln\left(\frac{2c}{\sqrt{\delta}}\right)} \geq \frac{48d}{\beta\varepsilon^2}$

$$\Rightarrow \frac{2c/\sqrt{\delta}}{\ln(2c/\sqrt{\delta})} \geq \frac{96d}{\beta\varepsilon^2\sqrt{\delta}} .$$

Consider $\eta \geq 4$. Claim: If $x \geq 2\eta \ln \eta$, then $x/\ln x \geq \eta$

Why?

$$\frac{x}{\ln(x)} \geq \frac{2\eta\ln\eta}{\ln(x)} \overset{?}{\geq} \frac{2\eta\ln\eta}{2\ln\eta} \Rightarrow \ln(x) \leq 2\ln\eta$$

I'll come back to this. Am I missing something obv?

Set $x = \frac{2c}{\sqrt{\delta}}$ & $\eta = \frac{96d}{\beta\varepsilon^2\sqrt{\delta}}$.

Then it suffices for c to satisfy

$$\frac{2c}{\sqrt{\delta}} \geq 2\frac{96d}{\beta\varepsilon^2\sqrt{\delta}} \ln\left(\frac{96d}{\beta\varepsilon^2\sqrt{\delta}}\right)$$ to guarantee

$$\frac{2c/\sqrt{\delta}}{\ln(2c/\sqrt{\delta})} \geq \frac{96d}{\beta\varepsilon^2\sqrt{\delta}} .$$

$\square$