Module - 1A

Randomized Linear Algebra (RLA): An Introduction

S. Lakshmivarahan

---

② **Two basic approaches in RLA**

- Random Sampling (RS)
- Random Projection (RP)

---

③ **Random Sampling (RS)**

- Randomly down sample the data
- uses two types of sampling distributions
  - Uniform sampling - data agnostic
  - Importance sampling - data dependent

---

④ **Random Projection (RP)**

- Projects / rotates the data onto a lower dimensional space
- Involves embedding data from a higher to a lower dimensional space
- Data agnostic

---

⑤ **Problems of interest**

- Matrix - Matrix product - Get an approximate product with small

- Obtain a sketch of a matrix, A, a rank k approximation to $A \in \mathbb{R}^{m \times n}$ with small error in Frobenius norm
- Solve approximately large scale linear least squares problem.

---

6 Measures of approximations

- Errors in the Frobenius norm
- Errors in spectral norm
- Quite a variety of approximations are available in both the norms
- How fast an approximation can be generated.

---

7 References

[1] A. Blum, J. Hopcroft and R. Kannan (2020) Foundations of Data Science, Hindustan Book Agency

[2] M. W. Mahoney (2010) "Randomized Algorithms for Matrices and Data", Foundations and Trends in Machine Learning, Vol 3, 2010 pp 123-224.

(1)

Module 1B

PROJECTION MATRIX

S. Lakshmivarahan

---

(2) Properties of matrices associated with $H \in R^{m \times n}$

- $H \in R^{m \times n}$ and $m > n$ – H is a tall matrix

  $m < n$ – H is a broad matrix

- Let H be a full rank matrix: $\text{Rank}(H) = \min\{m, n\}$
- $H^T H \in R^{n \times n}$ and $H H^T \in R^{m \times n}$ – Two Gramians of H
- When $m > n$: $H^T H$ is SPD
- When $m < n$: $H H^T$ is SPD

---

(3) Moore-Penrose inverse H^+ of H

- $H^+ = (H^T H)^{-1} H^T$ when $m > n$. $H^+ \in R^{n \times m}$
- $H^+ = H^T (H H^T)^{-1}$ when $m < n$. $H^+ \in R^{n \times m}$
- Properties of H^+:

$$H H^+ H = H \qquad (H H^+)^T = H H^+$$

$$H^+ H H^+ = H^+ \qquad (H^+ H)^T = H^+ H$$

---

(4) [orthogonal] Projection matrix $P_H \in R^{m \times m}$

- $P_H = H (H^T H)^{-1} H^T$ when $m > n$
- $P_H^T = P_H$ – Symmetric
- $P_H^2 = P_H$ – Idempotent
- $\det(P_H) = 0$ – Singular

(5) Spectral Properties of $P_H \in R^{m \times m}$:

- From $Ax = \lambda x$ it follows that the eigenvector belongs to the range of A
- Let $x = Hy$, that $x \in \text{Range}(H)$
- Consider

$$P_H x = H(H^T H)^{-1} H^T x$$
$$= H(H^T H)^{-1} H^T (Hy)$$
$$= H(H^T H)^{-1} (H^T H) y = Hy = 1 . x$$

(ie) $(1, x)$ is an eigenpair of P_H

- Since $\text{DIM}(\text{Range}(H)) = n$, there are n distinct eigendirections and $\lambda = 1$ is a multiple eigenvalue - n times

- $P_H \in R^{m \times n}$ has $n < m$ non-zero eigenvalues and hence singular

---

(6) Frobenius norm $\|P_H\|_F^2$

- Verify that $\|A\|_F^2 = tr(AA^T) = tr(A^T A)$
- $\|P_H\|_F^2 = tr(P_H P_H^T) = tr(P_H^2) = tr(P_H)$

$$= tr(H(H^T H)^{-1} H^T) = tr((H^T H)(H^T H)^{-1})$$
$$= tr(I_n) = n$$

---

(7) Statistical Leverage.

- The diagonal elements of P_H are called statistical leverage and are used in diagnostic analysis of linear regression.

to identify outliers.

- Define $L = \mathrm{diag}(l_{11}, l_{22}, \dots l_{mm})$

where $l_{ii} = i^{th}$ diagonal element of P_H

- If l_{ii} is large, then the i^{th} variable is said to be "influential"

---

(8) SVD of H ~~and leverage~~

- ~~Let~~ $H = U\Sigma^{1/2}V^T$ be the SVD of H
- $U \in R^{m\times n}$, $\Sigma^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots \lambda_n^{1/2})$, $V \in R^{n\times n}$
- $U^TU = I_n$, $UU^T \in R^{m\times m}$ is a projection matrix
- $VV^T = V^TV = I_n$, $V^{-1} = V^T$

---

(9) Leverage using SVD

- $P_H = H(H^TH)^{-1}H^T$
- $H^TH = (U\Sigma^{1/2}V^T)^T(U\Sigma^{1/2}V^T)$

$= V\Sigma^{1/2}U^TU\Sigma^{1/2}V^T = V\Sigma V^T$

- $(H^TH)^{-1} = (V^T\Sigma V)^{-1} = V\Sigma^{-1}V^T$
- $P_H = U\Sigma^{1/2}V^T[V\Sigma^{-1}V]\,V\Sigma^{1/2}U^T$

---

(9) ~~Leverage using SVD~~ SVD and P_H

- $(H^TH)^{-1} = [(U\Sigma^{1/2}V^T)^T(U\Sigma^{1/2}V^T)]^{-1}$

$= [V\Sigma^{1/2}(U^TU)\Sigma^{1/2}V^T]^{-1}$

$$= (V \Sigma V^T)^{-1} = (V^T)^{-1} \Sigma^{-1} V^{-1} = V \Sigma^{-1} V^T$$

- $P_H = H (H^T H)^{-1} H^T$

$$= (U \Sigma^{1/2} V^T)(V \Sigma^{-1} V^T)(V \Sigma^{1/2} U^T)$$

$$= U \Sigma^{1/2} . \Sigma^{-1} \Sigma^{\frac{1}{2}} U^T = U U^T$$

---

(10) <u>Leverage using SVD</u>

- $P_H = U U^T = \begin{bmatrix} u_{1*} \\ u_{2*} \\ \vdots \\ u_{m*} \end{bmatrix} [u_{1*}^T, u_{2*}^T, \ldots u_{m*}^T]$

$$= [u_{i*} u_{j*}^T] \qquad 1 \le i, j \le n$$

- Hence the i^{th} diagonal element of P_H is

$$l_{ii} = u_{i*} u_{i*}^T = \text{inner product of } i^{th} \text{ row of } U \text{ with itself}$$

$$= \| u_{i*} \|_2^2$$

---

(11) <u>Example</u>

- Let $H = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$

- $H^T H = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$, $\Delta = 42 - 36 = 6$

- $(H^T H)^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$

- verify: $\frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = I_2$

---

12 Example

$$P_H = H(H^T H)^{-1} H^T$$

$$= \frac{1}{6}\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

$$= \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

---

13 Leverage Scores

$$L = \frac{1}{6}\begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

---

14 Problems

1) Compute the SVD of H in slide 11 and compute $l_{ii} = \|U_{i*}\|^2$ and verify.

2) Compute the following:

a) $\|H_{*j}\|^2 \quad 1 \le j \le 2$ – Square of the norm of the j^{th} of H

b) $\|H_{i*}\|^2 \quad 1 \le i \le 3$ – Square of the norm of i^{th} row of H

3) Compute the following probability distributions:

a) $p_i = \dfrac{l_{ii}}{\sum_{i=1}^{3} l_{ii}}$, $1 \le i \le 3$, $\quad q_j = \dfrac{\|H_{*j}\|^2}{\|H_{*1}\|^2 + \|H_{*2}\|^2}$, $1 \le j \le 2$

$$\Lambda_k = \frac{\| H_{k*} \|^2}{\sum_{k=1}^{3} \| H_{k*} \|^2} \quad k = 1, 2, 3$$

(1)

Module 1-C

Hadamard Transform

S. Lakshmi Narayan

---

(2) **Hadamard transform (HT)**

- known by other names: Walsh-Hadamard, Walsh, Rademacher-Walsh, Walsh-Fourier transform
- It is an example of a generalized Fourier transform
- HT performs an orthogonal, symmetric, involutive linear operation on 2^m real real numbers

---

~~3 Matrix of HT = H_m~~

~~H_m is a $2^m \times 2^m$~~

---

(3) **Matrix representation of Hadamard transform: H_m**

- H_m is a $2^m \times 2^m$ real matrix that is defined recursively
- $H_0 = 1$, $H_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} H_0 & H_0 \\ H_0 & -H_0 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
- $H_m = \frac{1}{\sqrt{2}}\begin{bmatrix} H_{m-1} & H_{m-1} \\ H_{m-1} & -H_{m-1} \end{bmatrix}$ – a symmetric matrix
- $\frac{1}{\sqrt{2}}$ is a normalizing constant

- Rows of H_m are known as Walsh functions

(4) Examples: Real H_m matrices

- $H_0 = 1, \quad H_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

$$H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & -\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

(5) Hadamard product and HT

- Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad B = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$
- Hadamard product $A \otimes B = \begin{bmatrix} aB & bB \\ cB & dB \end{bmatrix}$
- Verify $H_m = H_1 \otimes H_{m-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} H_{m-1} & H_{m-1} \\ H_{m-1} & -H_{m-1} \end{bmatrix}$

(6) Equivalent definition of HT

- Let $H_m = [h_{k,n}], \quad 0 \le k, n, \le 2^m - 1$
- Express k and n as binary strings of length m
- $k = \sum_{i=0}^{m-1} k_i 2^i = k_{m-1} 2^{m-1} + k_{m-2} 2^{m-2} + \dots + k_1 2 + k_0$
- $n = \sum_{i=0}^{m-1} n_i 2^i = n_{m-1} 2^{m-1} + n_{m-2} 2^{m-2} + \dots + n_1 2 + n_0$
- $k_i, n_i \in \{0, 1\}$
- Verify $h_{kn} = \frac{1}{2^{m/2}} (-1)^{\sum_{i=1}^{m} n_i k_i}$

(7) Linear transformation

- H_m transforms $X \in R^{2^m}$ to $Y \in R^{2^m}$ where $Y = H_m X$

- Y is called Walsh spectrum of x
- Let $x = (1, 0, 1, 0, 0, 1, 1, 0) \in \mathbb{R}^8$ and $m = 3$
- Then $Y = H_3 x = (4, 2, 0, 2, 2, 0, 2) \in \mathbb{R}^8$
- Matrix-vector product : $O(n^2)$ where $n = 8$

---

(8) <u>Patterns in $H_m x$</u> : Let $m = 3$

- $H_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix}$ and $x = \begin{pmatrix} x_t \\ x_b \end{pmatrix}$ $x_t \in \mathbb{R}^4, x_b \in \mathbb{R}^4$
- x_t is top half and x_b is bottom half of x
- $H_3 x = \frac{1}{\sqrt{2}} \begin{bmatrix} H_2 x_t & H_2 x_b \\ H_2 x_t & -H_2 x_b \end{bmatrix}$

---

(9) <u>Patterns (continued)</u>

- $H_2 x_t = \frac{1}{\sqrt{2}} \begin{bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{bmatrix} \begin{bmatrix} x_{tt} \\ x_{tb} \end{bmatrix}$ $x_t = \begin{bmatrix} x_{tt} \\ x_{tb} \end{bmatrix}$
- $H_2 x_b = \frac{1}{\sqrt{2}} \begin{bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{bmatrix} \begin{bmatrix} x_{bt} \\ x_{bb} \end{bmatrix}$ $x_b = \begin{bmatrix} x_{bt} \\ x_{bb} \end{bmatrix}$
- $H_2 x_t = \frac{1}{\sqrt{2}} \begin{bmatrix} H_1 x_{tt} + H_1 x_{tb} \\ H_1 x_{tt} - H_1 x_{tb} \end{bmatrix}$
- $H_2 x_b = \frac{1}{\sqrt{2}} \begin{bmatrix} H_1 x_{bt} + H_1 x_{bb} \\ H_1 x_{bt} - H_1 x_{bb} \end{bmatrix}$

---

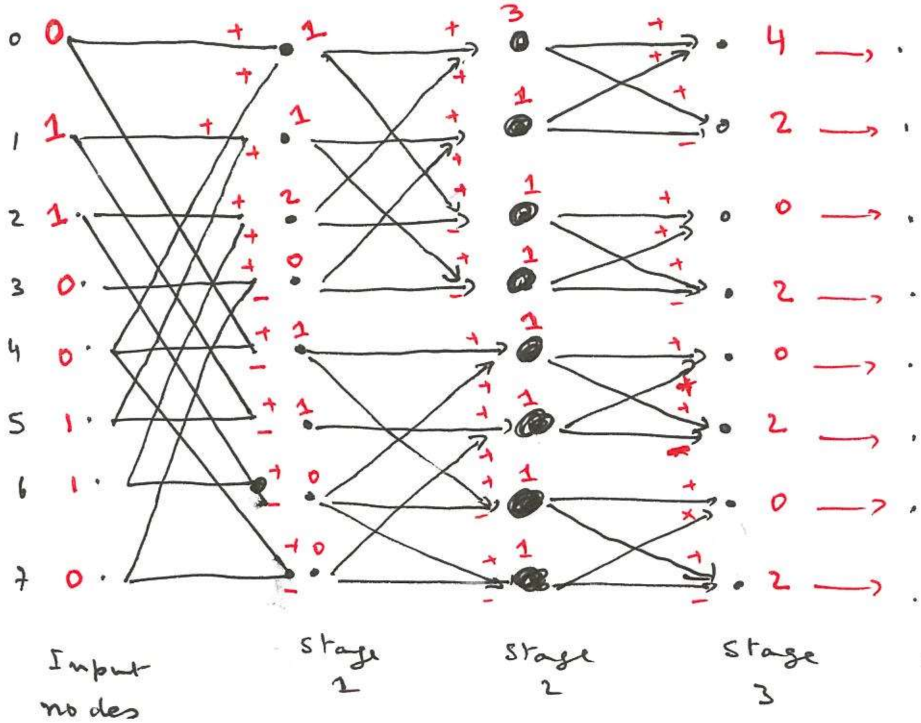
(10) <u>Patterns Continued</u>

- $H_1 x_{tt} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{ttt} \\ x_{ttb} \end{bmatrix}$ $x_{tt} = \begin{bmatrix} x_{ttt} \\ x_{ttb} \end{bmatrix}$
- $H_1 x_{tb} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{tbt} \\ x_{tbb} \end{bmatrix}$ $x_{tb} = \begin{bmatrix} x_{tbt} \\ x_{tbb} \end{bmatrix}$
- $H_1 x_{bt} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{btt} \\ x_{btb} \end{bmatrix}$ $x_{bt} = \begin{bmatrix} x_{btt} \\ x_{btb} \end{bmatrix}$

- $H_1 x_{bb} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{bbt} \\ x_{bbb} \end{bmatrix} \qquad x_{bb} = \begin{bmatrix} x_{bbt} \\ x_{bbb} \end{bmatrix}$

---

(11) A graphical flow diagram:



Time complexity $= O(n \log n)$ where $n = 2^m$

---

(12) Properties of H_m: H_m is an orthogonal matrix

- Consider $H_1 H_1^T = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = I_2$

(ie) H_1 is an orthogonal matrix. (ie) $H_1^{-1} = H_1^T$

and $H_1^T H_1 = H_1 H_1^T = I_2$

- $H_2 H_2^T = \frac{1}{\sqrt{2}} \begin{bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} H_1^T & H_1^T \\ H_1^T & -H_1^T \end{bmatrix}$

$= \frac{1}{2} \begin{bmatrix} H_1 H_1^T + H_1 H_1^T & H_1 H_1^T - H_1 H_1^T \\ H_1 H_1^T - H_1 H_1^T & H_1 H_1^T + H_1 H_1^T \end{bmatrix} = I_4$

- Recursively verify that $H_m H_m^T = H_m^T H_m = I_n$ with $n = 2^m$ (ie) H_m is orthogonal for all m.

(13) ~~H_m~~ ~~Involution~~ Involutive operator - Definition

- Let T be matrix denoting a linear operator, Then T is said to be involutive if $T^2 =$ Identity operator
- That is, $T^{-1} = T$ (ie) T is self-inverse

(14) Examples of involution operation

- $T_1(x) = -x$ $\quad T_4 = \dfrac{b-x}{1+cx}$, $b, c \in R$ and $bc \neq -1$
- $T_2(x) = \dfrac{1}{x}$ $\quad T_5(x) = \log\left(\dfrac{e^x+1}{e^x-1}\right)$
- $T_3(x) = \dfrac{x}{x-1}$ $\quad T_6(x) =$ Reflection through a plane in Euclidean Geometry
  $T_7(x) =$ Reflection w.r. to origin
- Plot of involution is symmetric w.r. to line $y = x$

(15) H_m is an involution

- Since H_m is symmetric and orthogonal, we get
  $H_m^2 = H_m H_m^T = I_n$ with $n = 2^m$
- That is, H_m is an involution

(16) Fast computation of $Y = H_m X$ and $X = H_m$

- Since H_m has only entries from the set $\{1, -1\}$

there is no multiplication but only additions and subtractions.

- Thus, $Y = H_m X$ and $X = H_m Y$ can be computed requiring only $O(n \log n)$ additions/subtractions where $n = 2^m$

<u>Problem</u>

1) Plot the functions $T_1(x)$ to $T_5(x)$ in Slide 14 and their inverse. Comment on what you observe.

2) Let $X = (a, b)^T$ and $Y = H_1 X$. For what values of a and b, we get $Y = X$? Can you generalize it X, Y being 2^m vectors using H_m?

4) Find the eigenvalues and eigenvectors of H_1, H_2, H_3. Can you generalize it?

3) Compute $\det(H_m)$ for $m = 1, 2, 3$ and generalize

(5) Let
$$T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Verify that T is orthogonal and not involutive

(1)

Module 1 - D

Uncertainty Principle in Fourier Transform

S. Lakshmi Narahari

---

(2) **Fourier Transform: Continuous domain**

- $x(t)$ - time signal
- $f(\omega)$ - frequency representation
- $f(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi \omega t}\, dt$; $\quad x(t) = \int_{-\infty}^{\infty} f(\omega)\, e^{2\pi \omega}\, d\omega$

---

(3) **Discrete FT: F and F^{-1}**

- Let $F = [F_{ij}] \in R^{n \times n}$, $\quad F_{ij} = \frac{1}{\sqrt{n}}\, \omega^{ij}$, $\quad \underline{0 \le i, j \le n-1}$

where $\omega^n = 1$. F is a symmetric matrix

- ω is the principal n^{th} root of unity

- $F^{-1} = [F^{-1}_{ij}] \in R^{n \times n}$, $\quad F^{-1}_{ij} = \frac{1}{\sqrt{n}}\, \omega^{-ij}$, $\quad \underline{0 \le i, j \le n-1}$

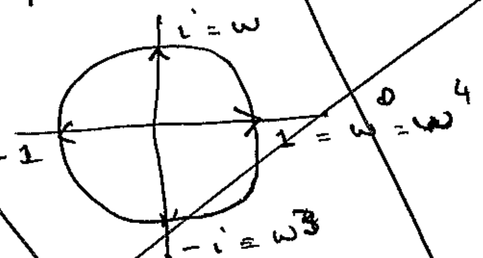
and F^{-1} is also a symmetric matrix.

---

(4) Examples

Let $n = 1$ and $\omega^4 = 1$ has 4 solutions: $\{i, -1, -i, 1\}$

- $F = \frac{1}{2} \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & \omega & \omega^2 & \omega^3 \\ 2 & 1 & \omega^2 & \omega^4 & \omega^6 \\ 3 & 1 & \omega^3 & \omega^6 & \omega^9 \end{array}$



- Verify: $\omega^4 = 1$, $\quad \omega^6 = \omega^{4+2} = \omega^4 \cdot \omega^2 = \omega^2$

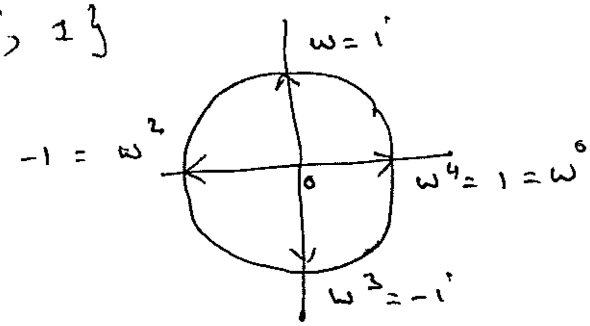
$\omega^9 = \omega^{2*4+1} = (\omega^4)^2 \cdot \omega = \omega$

- $\omega^i = \omega^{i \pmod n}$

---

(4) Example: $n = 4$

- $w^4 = 1 \Rightarrow w = \{ i, -1, -i, 1 \}$



- $$F = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$$

- $$F^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}.$$ verify $FF^{-1} = I_4$

- F is orthogonal and symmetric

(5) A basic Property of F

- Let $x \in R^n$ and $f \in R^n$ be such that
$$f = Fx$$
- x is a representation in the std. basis and f is a representation in the orthogonal basis defined by F
- These two representations are inherently incoherent. That is, a time signal and its Fourier transforms are incoherent (i) if x is sparse only a few randomly chosen coordinates of its Fourier transform are needed to reconstruct x

(6) Main result

- A signal $x(t)$ and its Fourier transform $f(w)$ <u>cannot</u> be sparse in both representation.
- Let n_x and n_f be the number of non-zero

components of x and f respectively. Then

$n_x n_y \geq n$

- This is called the <u>uncertainty principle</u> in fourier analysis

---

(7) Proof of Main result

- Let $x = (x_0, x_1, \dots \cancel{x_{n-1}})^T$ and $f = (f_0, f_1, \dots f_{n-1})^T$
- From $f = Fx$ we get (here $i = \sqrt{-1}$, unit imaginary)

$$f_j = \sum_{k=0}^{n-1} F_{jk} x_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} x_k e^{-\left(\frac{2\pi i}{n}\right) jk}$$

for $0 \leq j \leq n-1$

- If some of the x_k's are zero, then remove those corresponding columns of F

---

(8) Proof Continued

- Since n_x is the number of non-zero elements in $x(t)$, select n_x consecutive ~~columns~~ rows of F to create an $n_x * n_x$ matrix say **E**
- Now normalize the colums of E by dividing each element in the column of E by the column element in the first row

---

(9) Vandermonde matrix

- Then E takes the form of a Vandermode matrix:

$$E = \begin{bmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \\ a^3 & b^3 & c^3 & d^3 \end{bmatrix}$$

which is known to be non-singular

Note: Verify that the basic Fourier transform matrix $F = [w^{ij}]$, $0 \le i, j \le n-1$ also inherits the Vandermonde structure

---

(10) A Lemma

. If $x = (x_0, x_1, \cdots x_{n-1})^*$ has n_x non-zero elements, then its Fourier trans-form $f = (f_0, f_1, \cdots f_{n-1})^*$ cannot have n_x consecutive zeros.

---

(11) Proof of Lemma

. Let $i_1, i_2, \ldots i_{n_x}$ be the locations of non-zero elements in x

. ~~Then~~ $\cancel{f_k = \frac{1}{\sqrt{n}} \sum_{j=1}^{n_x} x_{i_j} e^{-\left(\frac{2\pi i}{n}\right) k i_j}}$

. Define $Z = [z_{kj}]$ a matrix of order $n_x \times n_x$

where

$$z_{kj} = \frac{1}{\sqrt{n}} e^{-\left(\frac{2\pi i}{n}\right) k i_j}$$

. Then $f_k = \sum_{j=1}^{n_x} z_{kj} x_{i_j}$ for $k = m+1, m+2, \cdots m+n_x$

a consecutive set of n_x components of f.

---

(12) ~~Proof of Lemma~~

. ~~By definition~~ $\cancel{x_{i_j} \ne 0}$ ~~for~~ $\cancel{j = 1, 2, \ldots n_x}$. ~~Hence~~ $\cancel{X \ne 0}$

. ~~Need to prove that~~ $\cancel{(f_{m+1}, f_{m+2}, \cdots f_{m+n_x}) \ne 0}$

(12) Vandermonde structure in Z

- Rescale each column of Z by dividing the elements of a column by the leading entry in that column.
- The resulting matrix inherits the Vandermonde struc[ture] and hence is non-singular

---

(13) Proof of lemma

- By definition $x_{i_j} \neq 0$ for $j = 1, 2, \dots n_x$. Hence $x \neq 0$.
- Need to prove that $(f_{m+1}, f_{m+2}, \dots f_{m+n_x}) \neq 0$
- Since $f = Zx$, this claim is true and the lemma follows

---

(14) Proof of the main result

- From lemma, it follows that if x has n_x non-zero components, then f cannot have consecutive blocks (of length n_x) zeros.

~~We can divide divide $f \in R^n$ into $\left\lceil \frac{n}{n_x} \right\rceil$ ($n_f =$) consecutive blocks each containing at least one non zero element~~

~~with at least $\left\lfloor \frac{n}{n_x} \right\rfloor$ blocks each containing at least one non-zero element.~~

~~Hence $\left\lfloor \frac{n}{n_x} \right\rfloor \cdot n_x \geq n$~~

- clearly, we can divide f into $n_f = \left\lceil \frac{n}{n_x} \right\rceil = \left\lfloor \frac{n}{n_x} \right\rfloor + 1$ consecutive blocks ~~ea~~ each containing at least one non-zero element.

- Hence $n_f n_x = \left(\left[\frac{n}{n_x}\right]+1\right) n_x \geq n$

---

(15) <u>This bound is tight</u>

- Let $x = (x_0, x_1, \dots x_{n-1})$ with $\sqrt{n}$ number of 1's separated by $(\sqrt{n}-1)$ zeros
- Let $n = 9$. $x = (1, 0, 0, 1, 0, 0, 1, 0, 0)^T$
- Then $f = x$ and $n_x n_f = n$ (ii) n_x divides n.

---

(16) <u>Proof</u>:

- Let $\delta(\sqrt{n}, \sqrt{n})$ be a time signal with $\sqrt{n}$ number of 1's spaced $\sqrt{n}$ apart
- Then $\delta(\sqrt{n}, \sqrt{n})$ is its Fourier transform
- Consider columns: $0, \sqrt{n}, 2\sqrt{n}, \dots (\sqrt{n}-1)\sqrt{n}$ of δ that contain 1's.
- Then the row element $j\sqrt{n}$ of column $k\sqrt{n}$ is given by $\omega^{nkj} = 1$ for $1 \leq k < \sqrt{n}$

---

(17) <u>Proof</u>

- The product of these rows of F with $\delta(\sqrt{n}, \sqrt{n})$ is $\sqrt{n}$ and scaling by $\left(\frac{1}{\sqrt{n}}\right)$ yields $f_{j\sqrt{n}} = 1$
- For rows with index not of this form $j\sqrt{n}$,

~~the elements in columns $\{0, \sqrt{n}, 2\sqrt{n},$~~

the row b, $b \neq j\sqrt{n}$, $j \in \{0, 1, 2, \dots \sqrt{n}-1\}$

the elements in the row b in columns

$0, \sqrt{n}, 2\sqrt{n}, \dots (\sqrt{n}-1)\sqrt{n}$

are $1, z^b, z^{2b}, \dots z^{(\sqrt{n}-1)b}$.

· Then $f_b = \frac{1}{\sqrt{n}}\left(1 + z^b + z^{2b} + \dots + z^{(\sqrt{n}-1)b}\right)$

$$= \frac{1}{\sqrt{n}} \frac{z^{\sqrt{n}b} - 1}{z^b - 1} = 0 \quad \text{since } z^{\sqrt{n}b} = 1 \text{ and } z^b \neq 1$$

· Hence the Fourier transform of $s(\sqrt{n}, \sqrt{n}) = s(\sqrt{n}, \sqrt{n})$

(1)

Module - 2

MATRIX - MATRIX PRODUCT :

Importance Sampling

S. Lakshmivarahan

---

(2) **Three ways of defining matrix-matrix product**

. $A \in R^{m \times n}$, $B \in R^{n \times p}$, $C \in R^{m \times p}$, $C = AB$

1) $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$ - inner product

2) $c_{*j} = \sum_{k=1}^{n} a_{*k} b_{kj}$ - Saxpy

3) $C = \sum_{k=1}^{n} a_{*k} b_{k*}$ - outer product

. Time = $O(mnp)$

. $m = n = p$ $\Rightarrow$ Time = $O(n^3)$

---

(3) ~~Stra~~ **Faster method**

. Strassen's method - originally $O(n^{2.8})$

. Current algorithms - $O(n^{2.5 \pm \epsilon})$

. Recursive procedure with large overhead

. For it to be effective $n = O(10^3)$

---

(4) **A sampling strategy**

. Let $[n] = \{1, 2, \dots n\}$ - $\left\{ \begin{array}{l} \text{column index of } A \\ \text{row index of } B \end{array} \right\}$

. Let Z be a random variable with values in $[n]$

. Define $p_k = \text{Prob}[Z = k]$

. $0 \le p_k \le 1$ and $\sum_{k=1}^{n} p_k = 1$. (1)

- Get uniform sampling if $p_k = \frac{1}{n}$
- Importance sampling: p_k varies with the length of the columns of A and lengths of the rows of B

---

(5) Basic idea

- Pick a subset of columns of A and the corresponding ~~columns~~ rows of B using the distribution $\{p_k\}_{k=1}^{n}$
- Scaled Product of a column ~~and~~ the corresponding row gives an outer product matrix of rank = 1
- ~~Add all the~~ Weighted average of these outer product matrices is an approximation to $C = AB$

---

(6) Matrix-valued random variable X

- Define X as follows:

$X = X_k$

$$P\Big[\underset{m \times n}{X} = \underbrace{\frac{1}{p_k} a_{*k} b_{k*}}_{\text{scaled outer product matrix}}\Big] = p_k \quad 1 \le k \le n \longrightarrow (2)$$

- That is, X takes on n values whose distribution is given above

---

(7) Expected value of X

$$E(X) = \sum_{k=1}^{n} p_k \cdot \left(\frac{1}{p_k} a_{*k} b_{k*}\right) = \sum_{k=1}^{n} p_k X_k$$

$$= \sum_{k=1}^{n} a_{*k} b_{k*} = AB \longrightarrow (3)$$

- Since $E(X) = AB$, the true value, estimating the product this way provides an unbiased estimate

(8) - This explains the role of scaling the outerproduct by $\frac{1}{p_k}$

~~(8) Variance of x~~

$$\cancel{Var(x_{ij}) = E(x_{ij}^2) - [E(x_{ij})]^2}$$

$$\cancel{= p_k \cdot \left(\frac{1}{p_k} a_{*k} b_{k*}\right)^2_{ij} - (a_{*k} b_{k*})^2_{ij}}$$

$$\cancel{= \frac{1}{p_k} (a_{*k} b_{k*})^2_{ij} - (a_{*k} b_{k*})^2_{ij}}$$

(8) Var (x)

- Var(x) = Sum of the variance of all of its entries

$$= \sum_{i=1}^{m} \sum_{j=1}^{p} Var(x_{ij})$$

$$= \sum_{i,j} \underbrace{E(x_{ij}^2)}_{\mathrm{I}} - \underbrace{\sum_{i,j} E^2(x_{ij})}_{\mathrm{II}} \rightarrow (4)$$

(9) Term I

$$\sum_{i,j} E(x_{ij}^2) = \sum_{i,j} \sum_{k=1}^{n} p_k \left(\frac{a_{ik} b_{kj}}{p_k}\right)^2$$

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

$$= \sum_{k=1}^{n} \frac{1}{p_k} \|a_{*k}\|^2 \|b_{k*}\|^2 \text{ - depends on } p_k \rightarrow (5)$$

---

(10) Term II

$$\sum_{i,j} E^2(x_{ij}) = \sum_{i,j} (AB)^2_{ij}$$

$$= \|AB\|^2_F \longrightarrow (6)$$

$$= \text{Independent of } p_k$$

---

(11) Var (x)

- Combining Term I and Term II:

$$\text{Var}(x) = \sum_{k=1}^{n} \frac{1}{p_k} \|a_{*k}\|^2 \|b_{k*}\|^2 - \|AB\|^2_F \longrightarrow (7)$$

---

(12) What is the best choice for p_k?

- It is the one that minimizes the variance
- Seek unbiased, minimum variance estimate of the product AB

---

(13) Optimal choice of p_k

- ~~Choose~~ $\cancel{p_k \propto \|a_{*k}\| \|b_{k*}\|}$
- Choose p_k that minimizes the first term on the RHS of (7)
- Claim: The minimizing p_k is given by

$$p_k = \frac{\|a_{*k}\| \|b_{k*}\|}{\sum_{j=1}^{n} \|a_{*j}\| \|b_{j*}\|} \longrightarrow (8)$$

(14) Expression for minimum Var(x)

. From (7):

$$Var(x) \le \sum_{k=1}^{n} \frac{1}{p_k} \|a_{*k}\|^2 \|b_{k*}\|^2 \rightarrow (9)$$

$$= \sum_{k=1}^{n} \left[\left(\frac{\|a_{*k}\|^2 \|b_{k*}\|^2}{\|a_{*k}\| \|b_{k*}\|} \right) \underbrace{\sum_{j=1}^{n} \|a_{*j}\| \|b_{j*}\|} \right]$$

$$= \sum_{k=1}^{n} \left[\underbrace{(\|a_{*k}\| \|b_{k*}\|)} \underbrace{\left(\sum_{j=1}^{n} \|a_{*j}\| \|b_{j*}\| \right)} \right]$$

$$= \sum_{k=1}^{n} \|a_{*k}\|^2 \cdot \sum_{j=1}^{n} \|b_{j*}\|^2 = \|A\|_F^2 \|B\|_F^2$$
$\rightarrow (10)$

---

(15) FACT: Cauchy-Schwartz Inequality

$$\sum_{k=1}^{3} \left[(a_k b_k) \sum_{j=1}^{3} (a_j b_j) \right] \qquad a = (a_1, a_2, a_3)^T \quad b = (b_1, b_2, b_3)^T$$

$$= a_1 b_1 [a_1 b_1 + a_2 b_2 + a_3 b_3]$$
$$+ a_2 b_2 [a_1 b_1 + a_2 b_2 + a_3 b_3]$$
$$+ a_3 b_3 [a_1 b_1 + a_2 b_2 + a_3 b_3]$$
$$= (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 = (a^T b)^2 \le \|a\|^2 \|b\|^2 \rightarrow (11)$$

. clearly (11) implies (10)

---

statistical

(16) A trick to improve "efficiency" of the estimate

- . Repeat the selection of x using s ~~independent~~ independent trials
- . Estimate AB using the average

$$\frac{1}{s} \sum_{i=1}^{s} x_i$$

(17) Variance of this average

$$Var\left[\frac{1}{s}\sum_{i=1}^{s} x_i\right] = \frac{1}{s^2} Var\left(\sum_{i=1}^{s} x_i\right)$$

$$= \frac{1}{s^2} \cdot \sum_{i=1}^{s} Var(x_i)$$

$$= \frac{1}{s^2} \cdot s \cdot Var(x) = \frac{Var(x)}{s}$$

$\rightarrow (12)$

---

(18) Implementation:

- Let $\{k_1, k_2, \ldots k_s\}$ be the integers chosen from $[n]$ in s-trials using the importance sampling p_k in (8)

---

(19) Estimate of AB

$$\frac{1}{s}\sum_{i=1}^{s} x_i = \frac{1}{s}\sum_{i=1}^{s}\left[\frac{a_{*k_i} b_{k_i *}}{p_{k_i}}\right] \longrightarrow (13)$$

$$= [a_{*k_1}, a_{*k_2} \cdots a_{*k_s}] \begin{bmatrix} \frac{b_{k_1 *}}{s p_1} \\ \frac{b_{k_2 *}}{s p_2} \\ \vdots \\ \frac{b_{k_s *}}{s p_{k_s}} \end{bmatrix}$$

$$= C\tilde{B}$$

---

(20) Structure of the estimate

- C is the $m \times s$ matrix of chosen columns of A
- $\tilde{B}$ is the $s \times n$ matrix of corresponding rows of B scaled by $s p_k$

$$\underset{m\times n}{\begin{bmatrix} \; A \; \end{bmatrix}} \underset{n\times p}{\begin{bmatrix} \; B \; \end{bmatrix}} \approx \underset{m\times s}{\overset{C}{\begin{bmatrix} \text{sampled} \\ \text{columns} \end{bmatrix}}} \underset{\text{of } A}{} \; \underset{s\times p}{\overset{\tilde{B}}{\begin{bmatrix} \text{corresponding} \\ \text{scaled} \\ \text{rows of } B \end{bmatrix}}}$$

---

(21) <u>Main Result</u>

- Let $A \in R^{m\times n}$ and $B \in R^{n\times s}$. The product AB is estimated by $C\tilde{B}$ where $C \in R^{m\times s}$ and $\tilde{B} \in R^{s\times n}$ with s columns of A and the corresponding scaled rows of B chosen from a non-uniform (importance) ~~distrib~~ distribution $\{p_k\}$ in (8). Then

$$E\left[\|AB - C\tilde{B}\|_F^2\right] \le \frac{\|A\|_F^2 \, \|B\|_F^2}{s} \rightarrow (14)$$

---

(22) <u>Proof of claim in (8)</u>: <u>Constrained minimum</u>

- Consider the analog of the first-term ~~on~~ the RHS of (7).
- Let $p = (p_1, p_2, \cdots p_n)^T$ be such that $0 \le p_i \le 1$ and $\sum_{i=1}^{n} p_i = 1$
- Consider a function $f(p) = \sum_{i=1}^{n} \frac{a_i^2}{p_i}$ $\rightarrow (15)$
- Find the minimizer of (15) subject to the constraint $\sum_{i=1}^{n} p_i = 1$.

---

(23) <u>Lagrangian multiplier</u>

- $L(p, \lambda) = f(p) + \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$ $\longrightarrow (16)$
- $\frac{\partial L}{\partial p_i} = -\frac{a_i^2}{p_i^2} + \lambda = 0, \quad 1 \le i \le n$
- $\lambda = \frac{a_i^2}{p_i^2}$ $\longrightarrow (17)$

- $p_i = \dfrac{a_i}{\sqrt{\lambda}}$ and $\sum_{i=1}^{n} p_i = 1$
- $\sum_{i=1}^{n} \dfrac{a_i}{\sqrt{\lambda}} = \dfrac{1}{\sqrt{\lambda}} \sum_{i=1}^{n} a_i = 1 \Rightarrow \sqrt{\lambda} = \sum_{i=1}^{n} a_i$
- optimal $p_i = \dfrac{a_i}{\sum_{i=1}^{n} a_i} \longrightarrow (18)$

---

(24) optimal p_i

- Substituting $a_i = \| a_{*j} \| \, \| b_{j*} \|$ in (15) We obtain the optimal expression in (8)

---

(25) Special case: $B = A^T$ and compute the Gramian AA^T

- Set $b_{k*} = a_{*k}$ (since $B = A^T$)
- $p_k = \dfrac{\| a_{*k} \|^2}{\| A \|_F^2}$
- Perform s repeated trials to get an estimate of AA^T

---

(26) Exercises

1) Write a pseudo code to implement the three ways of multiplying matrices in Slide 2
2) Generate $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ randomly where $a_{ij} \sim$ i.i.d $N(0,1)$ and $b_{ij} \sim$ i.i.d $N(0,1)$

for $m = n = p = 100$

4. Compute the product using all the three ways and compare your result.

3) Compute p_k (as in (8)) and implement the above algorithm

for $s = 5, 10, 20, 30, 40, 500$.

. compute the error $\|AB - C\tilde{B}\|_F^2$ and plot it as a function of s. Comment on the resulting graph

4) use the uniform sampling $p_k = \frac{1}{n}$ and repeat the experiment for the same set of values of s as in problem (3). Compare the graph of the error $\|AB - C\tilde{B}\|_F^2$ in this case with that of using the importance sampling in problem (3) above.

(1)

Module 3
Sketch of a Matrix

S. Lakshmivarahan

---

(2) Definition of sketch of a matrix

- A sample (or subset) of columns and rows each picked in independent trials based on the squared length of columns/rows is a good sketch of the matrix $A \in R^{m \times n}$

---

(3) ~~An algorithms~~ Importance sampling probabilities

- Let $p_k = \dfrac{\|a_{*k}\|^2}{\|A\|_F^2}$ used for column $k = 1$ to n
- Let $q_j = \dfrac{\|a_{j*}\|^2}{\|A\|_F^2}$ used for row $j = 1$ to m

---

(4) ~~Algorithm~~ Interpolative approximation

- Pick s columns of A using $\{p_k\}$ to form $C \in R^{m \times s}$
- Pick r rows of A using $\{q_k\}$ to form $R \in R^{s \times n}$
- Using C and R, find $U \in R^{s \times s}$ such that

$$A \simeq C\,U\,R \longrightarrow (1)$$

---

(5) Comparison with SVD (A is full rank)

- SVD of $A = U \Sigma V^T \longrightarrow (2)$

- ~~where~~ $U \in R^{m \times n}$ with orthogonal columns, (ie) $U^T U = I_n$ and $U U^T$ is a projection onto R^m
- $V \in R^{n \times n}$ ~~with~~ orthogonal matrix: $V^T V = V V^T = I_n$
- $\Sigma = diag(\lambda_1, \lambda_2, \ldots \lambda_n) : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

~~(5) Best approximation using SVD~~

- ~~For any $1 \leq r \leq n$, by using the first r columns of U, r rows of V^T and~~
- ~~Let $1 \leq r \leq n$~~

---

(6) Property of SVD based approximation

- Let $1 \leq r \leq n$
- $\bar{U} = U(*, 1:r)$ — matrix of first r columns of U
- $\bar{V}^T = V(1:r, *)$ — matrix of first r rows of V
- $\bar{\Sigma} = diag(\lambda_1, \lambda_2, \ldots \lambda_r)$
- Then $A \approx \bar{U} \bar{\Sigma} \bar{V}^T$ is an optimal approximation:

$$\| A - \bar{U} \bar{\Sigma} \bar{V}^T \|_F^2 \leq O(\lambda_{r+1}^2)$$

---

(7) SVD based sketch

- Provably optimal
- Computing SVD is demanding

- The columns of U and rows of $\bar{V}^T$ do not have a direct interpretable relation to the columns and rows of A

---

(8) Interpolative sketch

- In $A \approx CUR$, the rows of R and columns of C are directly obtained from those of A
- Easily interpretable
- Coupling matrix U is dependent on C and R to make the approximation error $\|A - CUR\|_F^2$ small.

---

(9) Application of Sketch - Document ~~-Term~~ matrix

- Let A be a document matrix for a large collection of documents.
- Here $A \in R^{m \times n}$ with m total number of (documents) ~~words in English dictionary~~ and n number of ~~documents~~ terms in all of documents $(n >> m)$
- $A = [a_{ij}]$ where a_{ij} is an integer that denotes the number of occurrences of the j^{th} term ~~word~~ in the i^{th} document

---

(10) use of document matrix

- Let there be a query ~~with k-dim vector with~~ vector of length n with one term per entry arrives

- Goal is to find a subset of documents that are similar to the query and display pointers to the document as is done in a web search
- Similarity is measured by the dot product of rows of A with the incoming query.

---

(11) Sketch of a document matrix: $(n >> m)$

$$A = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m \end{array} \begin{array}{c} 1 \quad 2 \quad \ldots \text{Terms} \ldots \quad n \\ \left[\begin{array}{ccc} & & \\ & & \\ & & \\ & & \end{array} \right]_{m \times n} \end{array} \text{Documents}$$

$$\approx \left[\begin{array}{c} \text{Sample columns of } A: \\ C \end{array} \right]_{m \times s} \left[\; U \; \right]_{s \times r} \left[\begin{array}{c} \text{sample rows of } A: \\ R \end{array} \right]_{r \times n}$$

- The dot product needs: $(rn + sr + ms)$ operations which is linear in n and m if s, r are $O(1)$

---

(12) Property of sketch

- Pick s and r such that
$$\| A - \text{sketch}(A) \|_2 \text{ is small}$$
where $\| A \|_2 = \max_{\|x\|=1} \| A x \|$ called the spectral norm of A
- sketch(A) contains a subset of terms and a subset of documents that are considered representative terms and documents

(13) Application of Sketch : customer - product matrix

$$A = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m \end{array} \!\! \begin{array}{c} 1 \quad 2 \quad \cdots \quad n \\ \left[\begin{array}{ccc} & & \\ & & \\ & & \\ & & \end{array} \right] \end{array} \text{ customers}$$

Products

- $A = [a_{ij}]$
- a_{ij} = preference of the i^{th} customer for j^{th} product

---

(14) Sketch (A)

- sample rows : Preferences of a subset of customers - r rows - Get R
- sample columns : customers preference for few products - s column - Get C
- Find U such that $A \simeq C\,U\,R$

---

(15) Problems

1) Set $m = 100$, $n = 40$. Generate $H \in \mathbb{R}^{m \times n}$ with

a) $H_{ij} \sim$ i i d $N(0, 1)$

b) $H_{ij} \sim$ i i d Binomial with $H_{ij} = \pm 1$ with probability $1/2$ and $1/2$

c) $H_{ij} \sim$ i. i. d $\{\pm\sqrt{3}, 0\}$ $H_{ij} = 0$ with $p = 2/3$
$H_{ij} = \pm\sqrt{3}$ with $p = 1/6$ each.

- Compute a Sketch for H.

1 MODULE - 4

$A = CUR$ Decomposition

S. Lakshmivarahan

---

2 Given C, and R, what is U?

- $A \approx C\,U\,R$ and $m > n$
- $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{m \times s}$, $R \in \mathbb{R}^{r \times n}$, $U \in \mathbb{R}^{s \times r}$
- If $r < \min(m, n) = n$, and A is non-degenerate, it is likely that r rows of A in R are linearly independent and $RR^T \in \mathbb{R}^{r \times r}$ is non-singular

---

3 Basic idea for computing U

- $A = A I_n$
- Approximate the product $A I_n$ by sampling s-columns of using the importance distribution based on the squared length of columns of A
- $A I_n = C W \longrightarrow (1)$
- W - scaled version of the corresponding s rows of I

(4) <u>Goodness of CW</u>

- From Main Result in Module-2 on product of matrices:

$$E\left[\,\|AI - CW\|^2\,\right] \le \frac{\|A\|_F^2 \, \|I_n\|_F^2}{s} \rightarrow (2)$$

$$= \frac{n}{s}\,\|A\|_F^2 \rightarrow (3)$$

- For this to be useful $s \geq n$ which is of no use
- Need to modify this basic idea

---

(5) <u>Use of projection matrix based on R</u>

- Let $R \in \mathbb{R}^{r \times n}$ be such that RR^T is invertible. Happens when R is of full rank.
- Define $P = R^T (RR^T)^{-1} R \in \mathbb{R}^{n \times n}$ → (4)
- P is symmetric: $P^T = P$
- P is idempotent: $P^2 = P$
- P acts as identity on the space V spanned by rows of R
- Now replace I_n by P in the above development

---

(6) ~~Structure of AA~~ <u>choice of $C \in \mathbb{R}^{m \times s}$</u>

- Sample s columns of A using squared length importance distribution

- Let $C \in \mathbb{R}^{m \times s}$ with the chosen s-columns of A

---

(7) <u>What is U in $A \approx CUR$?</u>

- Consider the product $AP = AR^T(RR^T)^{-1}R \rightarrow (5)$

- Replacing A by C on the R.H.S. of (5):

$$AP \approx C\,R^T(RR^T)^{-1}R = CUR \rightarrow (6)$$

- $U = R^T(RR^T)^{-1} \Rightarrow (7)$

<u>Note</u> $A = AP$ since P is the projection on to the row space of A spanned by r rows of A.

---

(8) <u>Error Bounds</u>:

- $$E\left[\|AP - CUR\|_2^2\right] \le E\left[\|AP - CUR\|_F^2\right]$$

$$\le \frac{\|A\|_F^2\,\|P\|_F^2}{s} \rightarrow (8)$$

$$\le \frac{r}{s}\|A\|_F^2 \rightarrow (9)$$

since $\|P\|_F^2 \le r \rightarrow (10)$

---

(9)