Class notes for Monday, Februray 29th

Recall: Randomized ID

(1) Draw gaussian $G = randn(n, k+p)$

(2) $Y = AG \quad \sim mnk$

(3) $[I_s, X] = ID\_row(Y, k) \quad \sim mk^2$

Then $A \approx \; XA(I_s, :)$

Observation (1) : It is easy to accelerate to complexity $mn\log(k)$ by replacing the gaussian random matrix by an SRFT

Observation (2): We can eaisly convert the ID to an SVD

$$A_{m\times n} \approx X_{m\times k} A_{k\times n}(I_s, :)$$

$$X = QR$$

$$\Rightarrow \quad A \approx Q_{m\times k}(RA(I_s, :))_{k\times n}$$

$$= Q_{m\times k}\hat{U}_{k\times k}D_{k\times k}V^*_{k\times n} = U_{m\times k}D_{k\times k}V^*_{k\times n}$$

Randomized SVD with complexity $mn\log(k)$

(1) Draw SRFT ω of size $n \times (k+p)$

(2) $Y = A\omega \rightarrow$ evaluate dwith subFFT $(mn\log(k))$

(3) $[I_s, X] = ID_row(Y, k)$

(4) $[Q, R] = qr(Y, 0)$

(5) $[\hat{U}, D, V] = svd(RA(I_s, :),' econ')$

(6) $U = Q\hat{U}$

Then

$$A \approx UDV^*$$

The CUR Decomposition

Let A be an $m \times n$ matrix of exact rank k. Then A admits a factorization

$$A = C_{m\times k}U_{k\times k}R_{k\times n}$$

where

$C = A(:, J_s)$ is a set of k columns of A

$R = A(I_s, :)$ is a set of k ros of A

The CUR and the doublesided ID are "siblings"

We have (1) $A = XA_sZ$ where $A_s = (I_s, J_s)$

Restrict (1) to coluns in J_s

$$A(:, J_s) = XA_sZ(:, J_s)$$

$$\Rightarrow \quad C = XA_S \quad (2)$$

Analoguly: $R = A_sZ$ (3)

(1) $\Leftrightarrow \ A = (XA_s)A_s^{-1}(A_sZ)$

Set $U = A(I_s, J_s)^{-1}$ to obtain

$$A = CUR$$

Algorithm: Computing CUR from ID suppose I_s, J_s have been determined. Then set

$$U = A(I_s, J_s)$$

$$C = A(:, J_s)$$

$$R = A(I_s, :)$$

Fundamental Problem of CUR:

$A(I_s, J_s)$ is typically ill-conditioned:

example

$$A = \begin{bmatrix} I_2 \\ S \end{bmatrix} \begin{bmatrix} 1 & c \\ 0 & \beta \end{bmatrix} \begin{bmatrix} I_2 & T \end{bmatrix}$$

(A is of rank 2)

$$I_s = I_r = [1, 2] \quad (*)$$

$$= \begin{bmatrix} A_s & A_sT \\ SA_s & SA_sT \end{bmatrix}$$

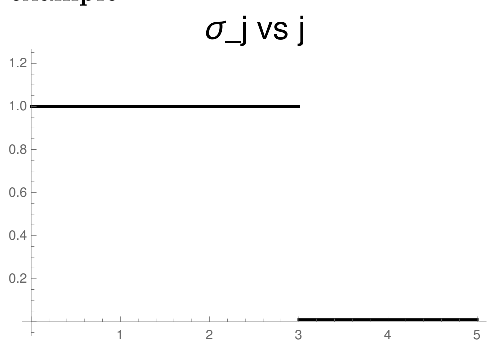
Then the CUR of A is

$$A = A(:,[1,2]) \begin{pmatrix} 1 & 0 \\ 1 & \frac{1}{\beta} \end{pmatrix} A([1,2],:) \quad (**)$$

Suppose β is small, then the factorization in (*) is perfectly benign, but (**) is numerically unstable (large numbers must combine to produce small numbers)

Let us consider the case of a matrix of appropriate rank k

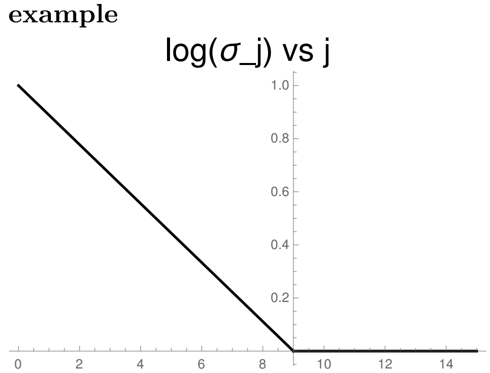
example



The CUR will have $cond(A_s) \approx \frac{\sigma_1}{\sigma_k}$ which is moderate

(something) often has $\sigma_j(A) \approx \sigma_j(A_s), \quad i \leq j \leq k \rightarrow$ unusual!

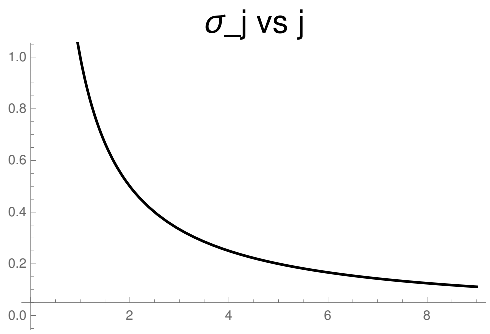
example



$cond(A_s) = \frac{\sigma_1(A_s)}{\sigma_k(A_s)} \approx \frac{1}{10^{-15}} = 10^{15}$

CUR will lead to very bad errors (due to floating point artimetic). You cannot, roughly speaking, get errors better then $\sqrt{\epsilon_{machine}} \approx 10^{-8}$

example



Condition number of A_s will be 10^1 or 10^2 , which is not problematic

---

How do you compute CUR when A has approximate rank k ?

(1) First identify I_s, J_s (use gram-schmidt, use randomized sampling + GS, randomized sampling via "leverage scores")

(2) $C = A(:, J_s) \quad R = A(I_s, :)$

we now have $A \approx C^\dagger A R^\dagger R$

(3) Set $U = C^\dagger A R^\dagger$ (this is probably the optimal U)