Class notes for Monday, February 29th

Recall: Randomized ID

(1) Draw gaussian \( G = \text{randn}(n, k + p) \)

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

(3) \( [I_s, X] = ID_{row}(Y, k) \sim mk^2 \)

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 easily convert the ID to an SVD

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

\[
X = QR
\]

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

\[
= Q_{m \times k}\hat{U}_{k \times k}D_{k \times k}V_{k \times n}^* = U_{m \times k}D_{k \times k}V_{k \times n}^*
\]

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

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

(2) \( Y = A\omega \rightarrow \) evaluate \( d\) with subFFT \( (mn\log(k)) \)

(3) \( [I_s, X] = ID_{row}(Y, k) \)

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

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

(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 \( k \) rows 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 columns in \( J_s \)

\[
A(:, J_s) = XA_sZ(:, J_s)
\]

\[ \Rightarrow C = XA_{s(2)} \] (2)

Analogously: \( R = A_s Z \) (3)

(1) \( \iff A = (XA_s)A^{-1}_{s(A_s)}Z \)

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)

\[
I_s = I_r = [1,2] \quad (*)
\]
Then the CUR of $A$ is

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

Suppose $\beta$ 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

$\sigma_j$ vs $j$

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

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

example

$log(\sigma_j)$ vs $j$

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

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

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 $A \approx C^\dagger AR^\dagger R$

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