EE270
Large scale matrix computation, optimization and learning

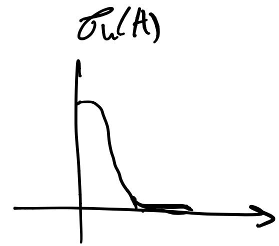
Instructor : Mert Pilanci

Stanford University

Tuesday, March 9 2021

Randomized Linear Algebra and Optimization
Lecture 17: Randomized Matrix Decompositions and Randomized SVD

Randomized Matrix Decompositions



- Suppose that A is an $n \times d$ data matrix of rank r
- Singular Value Decomposition (SVD) provides the best rank k approximation:

  Let $A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ where $\{\sigma_i\}_{i=1}^{r}$ are the singular values sorted in non-increasing order

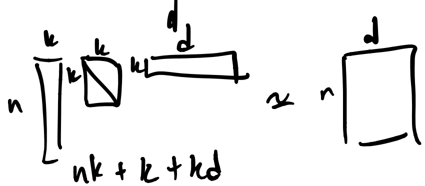
  Define $A_k := \boxed{U_k \Sigma_k V_k} := \sum_{i=1}^{k} \sigma_i u_i v_i^T$. We have

  $$\|A - A_k\|_2 \leq \sigma_{k+1}$$

  spect. norm

  $\boxed{A^T A}$

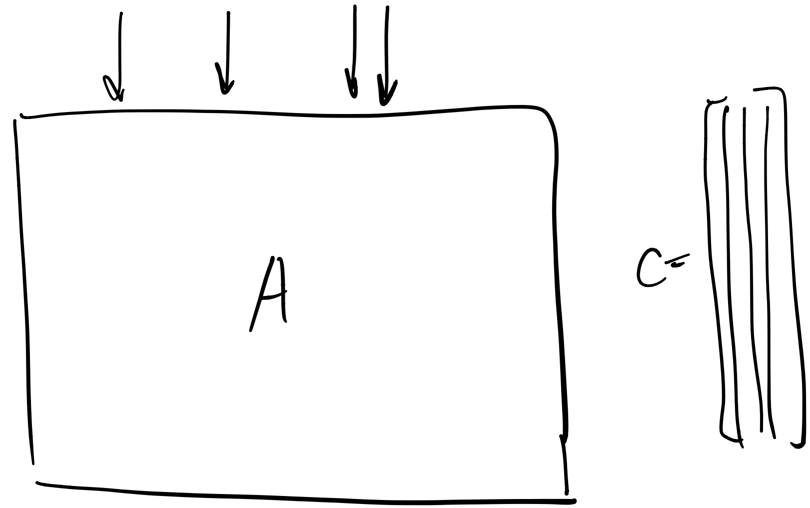
- computational cost of computing the SVD is $O(nd^2)$ for $n \geq d$



$O(dn^2)$ for $d \geq n$

Randomized Matrix Decompositions

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample some columns of A to get $C \in \mathbb{R}^{n \times m}$



$$C = AS$$

Right sketching

sub-sampling

Randomized Matrix Decompositions

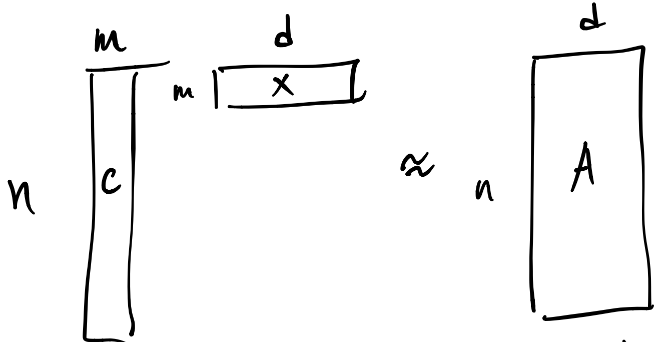
sketch size: m

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample some columns of A to get $C \in \mathbb{R}^{n \times m}$
- form an approximation of A using these sampled columns

$C = AS$

$AA^T \approx CC^T$

$$\min_X \|CX - A\|_F^2 = \min_X \|ASX - A\|_F^2$$



CX decomposition

X^* opt soln $\Rightarrow$ $\underbrace{CX^*}_{\text{rank is at most } m} \approx A$

Randomized Matrix Decompositions

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample some columns of A to get $C \in \mathbb{R}^{n \times m}$
- form an approximation of A using these sampled columns $C = AS$

$$\min_X \|CX - A\|_F^2 = \min_X \|ASX - A\|_F^2$$

- column-wise decomposable problem

$$\arg\min_{X^{(j)}} \sum_{k=1}^{d} \|ASX^{(k)} - A^{(k)}\|_2^2 = (AS)^\dagger A^{(k)}$$

Randomized Matrix Decompositions

$AS = U\Sigma V^T$

$AS(AS)^+ = UU^T$

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample some columns of A to get $C \in \mathbb{R}^{n \times m}$
- form an approximation of A using these sampled columns $C = AS$

$$\min_X \|CX - A\|_F^2 = \min_X \|ASX - A\|_F^2$$

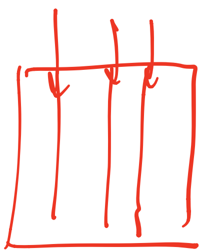
Approx of A

$A \approx \boxed{AS(AS)^+A}$

different than

SA

- column-wise decomposable problem



$$\arg\min_{X^{(j)}} \sum_{k=1}^{d} \|ASX^{(k)} - A^{(k)}\|_2^2 = (AS)^\dagger A^{(k)}$$

$$\arg\min_X \|ASX - A\|_F^2 = (AS)^\dagger A$$

- matrix A is approximated by $(AS)(AS)^\dagger A = CC^\dagger A$

Randomized Matrix Decompositions: Spectral Norm Error

(1) Apply sketch to A $\downarrow$ $C=AS$ (2) Apply S.V. extraction to C

- Given a large matrix $A \in \mathbb{R}^{n\times d}$
- idea: sample some ~~rows~~ column of A to get $C \in \mathbb{R}^{n\times m}$ $C=AS$
- approximation $CC^{\dagger}A \approx A$ $CC^{\dagger}=UU^T$
- calculate top left singular values of $C \approx \underbrace{U_k\Sigma_k V_k^T}_{\sum_{i=1}^{k}\sigma_i u_i v_i^T}$ where $C=U\Sigma V^T$
- then we have $CC^{\dagger} \approx U_kU_k^T$

optional additional approximation

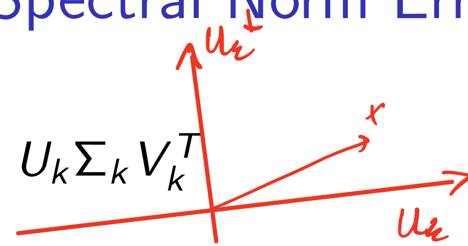
Lemma 1 (Drineas et al. 2006)

$$\|A - \underbrace{\overbrace{U_kU_k^TA}^{\text{at most rank } k}}_{U_kU_k^TA}\|_2^2 \le \underbrace{\|A - \overbrace{U_{A_k}U_{A_k}^TA}^{\text{best rank-}k\text{ approx. of }A}\|_2^2}_{\sigma^2_{k+1,A}} + 2\overbrace{\|AA^T - CC^T\|_2}^{\text{additional error of Randomized Sampling}}$$

- first term is the approximation error of the exact SVD
- second term is the spectral norm approximate matrix multiplication error
- approximate matrix multiplication results can be used!

Randomized Matrix Decompositions: Spectral Norm Error

- $C = AS$, approximation $CC^{\dagger}A \approx A$
- calculate top k left singular values of $C \approx U_k \Sigma_k V_k^T$
- approximate $A \approx U_k U_k^T A$



Proof of Lemma 1

defn of spectral norm

if $x \in U_k$ (Range of U_k)

$x^T A - \underbrace{x^T U_k U_k^T}_{x} A = 0$

$$
\begin{aligned}
&\|A - U_k U_k^T A\|_2^2 \\
&= \max_{\|x\|_2=1} \|x^T(A - U_k U_k^T A)\|_2 \\
&= \max_{\substack{\|y\|_2=\|z\|_2=1, y\in U_k, z\in U_k^\perp \\ \alpha^2+\beta^2=1}} \|(\alpha y + \beta z)^T (A - U_k U_k^T A)\|_2 \\
&\le \max_{\|z\|_2=1, z\in U_k^\perp} \|z^T(A - U_k U_k^T A)\|_2 + \\
&\quad \max_{\|y\|_2=1, y\in U_k} \|y^T \overset{0}{\cancel{(A - U_k U_k^T A)}}\|_2 \\
&= \max_{\|z\|_2=1, z\in U_k^\perp} \|z^T(A - \cancel{U_k U_k^T} A)\|_2
\end{aligned}
$$

$\Rightarrow |\alpha| \le 1, |\beta| \le 1$

$\|\alpha y + \beta z\|_2^2 = \alpha^2 \cdot \underbrace{\|y\|_2^2}_{1} + \beta^2 \cdot \underbrace{\|z\|_2^2}_{1} = 1$

$z^T U_k U_k^T = 0$ since $z \in U_k^\perp$

Randomized Matrix Decompositions: Spectral Norm Error

Proof of Lemma 1 cont'd

$$
\begin{aligned}
&= \max_{\|z\|_2=1, z\in U_k^{\perp}} \|z^T A\|_2^2 = z^T A A^T z \\
&= \max_{\|z\|_2=1, z\in U_k^{\perp}} z^T CC^T z + z^T(AA^T - CC^T)z \\
&\leq \sigma_{k+1}^2(C) + \|AA^T - CC^T\|_2 \\
&\leq \sigma_{k+1}^2(A) + \|AA^T - CC^T\|_2
\end{aligned}
$$

We need a matrix perturbation result

$$\left|\sigma_k(A) - \sigma_k(A+E)\right| \leq \|E\|_2 \quad \forall k$$

$$\left|\, \|A\|_2 - \|A+E\|_2 \right| \leq \|E\|_2$$

Randomized Matrix Decompositions: Spectral Norm Error

Proof of Lemma 1 cont'd

$$
\begin{aligned}
&= \max_{\|z\|_2=1, z\in U_k^\perp} \|z^T A\|_2 \\
&= \max_{\|z\|_2=1, z\in U_k^\perp} z^T C C^T z + z^T (AA^T - CC^T) z \\
&\le \max_{\|z\|_2=1, z\in U_k^\perp} \sigma_{k+1}^2(C) + \|AA^T - CC^T\|_2 \\
&\le \max_{\|z\|_2=1, z\in U_k^\perp} \sigma_{k+1}^2(A) + 2\|AA^T - CC^T\|_2
\end{aligned}
$$

$\sigma_{k+1}^2(C) \le \sigma_{k+1}^2(A) + \|AA^T - CC^T\|_2$

$\sigma_{k+1}(A)^2$

- where we used a matrix perturbation result $\sigma_{k+1}(CC^T) - \sigma_{k+1}(AA^T) \le \|AA^T - CC^T\|_2$
- Hoffman–Wielandt inequality:
- $\max_k |\sigma_k(Q+E) - \sigma_k(Q)| \le \|E\|_2$

‘ ’

Randomized Matrix Decompositions: Frobenius Norm Error

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample some rows of A to get $C \in \mathbb{R}^{n \times m}$
- approximation $CC^{\dagger}A \approx A$
- calculate top left singular values of $C = U_k \Sigma_k V_k^T$

  Lemma 2 (Drineas et al. 2006)

$$\|A - U_k U_k^T A\|_F^2 \leq \|A - U_A U_A^T A\|_F^2 + 2\sqrt{k}\|AA^T - CC^T\|_F$$

- approximate matrix multiplication results can be used

Randomized Singular Value Decomposition $\equiv$ CX decomposition

$A \leftarrow A^T$

$A^T S = (S^T A)^T$

$(A^T S)(A^T S)^\dagger A = (S^T A)^T \cdot (S^T A)^{T\dagger} \cdot A$

- Given a large matrix $A \in \mathbb{R}^{n\times d}$
- idea: sample some ~~rows~~ columns of A to get $C \in \mathbb{R}^{n\times m}$
- $C = AS$
- approximation $CC^\dagger A =$
- calculate QR decomposition of : $QR = AS$
- then $\underbrace{QQ^T A} \approx A$, i.e., Q approximate the range space of A
- calculate the SVD $Q^T A = \underline{U\Sigma V^T}$
- approximate SVD of A is $A \approx (QU)\Sigma V^T = \underbrace{QQ^T A}_{U\Sigma V^T} = CC^\dagger A = \underline{(AS)}(AS)^\dagger A$

$QR(QR)^\dagger A = QQ^T A$

$AS(AS)^\dagger A = QQ^T A$

$(QR)^\dagger = R^\dagger Q^\dagger = R^\dagger Q^T$

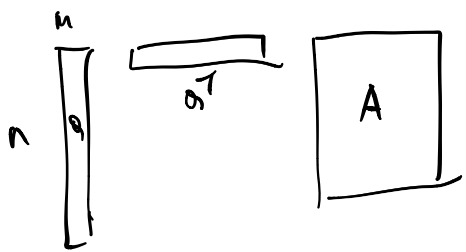
$= Q\underbrace{Q^T A}_{U\Sigma V^T}$

$= \underbrace{QU}_{U'}\Sigma V^T$

$= U'\Sigma V^T$ RSVD

orthogonal

$(QU)^T QU = U^T Q^T Q U = U^T U = I$



Randomized Singular Value Decomposition

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample/sketch some rows of A to get $C \in \mathbb{R}^{n \times m}$
- $C = AS$
- approximation $CC^{\dagger} = (SA)(SA)^{\dagger} A \approx A$
- calculate QR decomposition of $SA = QR$
- then $QQ^T A \approx A$, i.e., Q approximate the range space of A
- calculate the SVD $Q^T A = U \Sigma V^T$
- approximate SVD of A is $A \approx (QU) \Sigma V^T$

Randomized Singular Value Decomposition: Analysis

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- Generate a Gaussian sketching matrix $S \in d \times m$
- $C = AS$
- approximation $CC^{\dagger} = (AS)(AS)^{\dagger} A \approx A$
- calculate QR decomposition of $C = AS = QR$
- then $QQ^T A \approx A$, i.e., Q approximate the range space of A
- calculate the SVD $Q^T A = U \Sigma V^T$
- approximate SVD of A is $A \approx (QU) \Sigma V^T$
- **Lemma** (Halko et al. 2009)

$$\mathbb{E}\|A - QQ^T A\|_2 \leq \left(1 + \frac{4\sqrt{m}}{m-k-1}\sqrt{\min(n,d)}\right)\sigma_{k+1} \approx (1+\epsilon)\cdot\sigma_{k+1}$$

$U_k U_k^T$

$k = m$

not additive

multiplicative

$m = c \cdot k$

Randomized Singular Value Decomposition: Comparison

- Generate a Gaussian sketching matrix $S \in d \times m$
- calculate QR decomposition of $C = AS = QR$
- calculate the SVD $Q^T A = U \Sigma V^T$
- approximate SVD of A is $A \approx (QU) \Sigma V^T$
- **Lemma** (Halko et al. 2009)

$$\mathbb{E}\|A - QQ^T A\|_2 \leq \left(1 + \frac{4\sqrt{m}}{m-k-1} \sqrt{\min(n,d)}\right) \sigma_{k+1}$$

- Exact SVD of $A = U_A \Sigma_A V_A^T$ yields

$$\|A - U_A^k (U_A^k)^T A\|_2 \leq \sigma_{k+1}$$

Low-rank matrix approximations

- Singular Value Decomposition (SVD)
- $A = U\Sigma V^T$
- takes $O(nd^2)$ time for $A \in R^{n\times d}$
- best rank-k approximation is $A_k := U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T$
- $\|A - A_k\|_2 \leq \sigma_{k+1}$

$CA \quad m \times d$

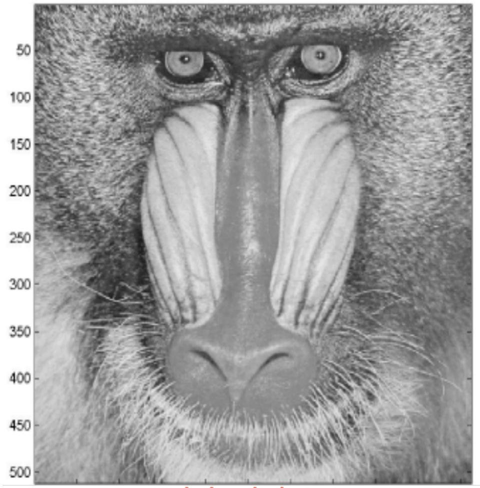
$QR(CA) \quad : \quad O(md^2)$

$SVD(\underbrace{Q^T A}_{n\cdot md}) \quad : \quad O(md^2 + nmd)$

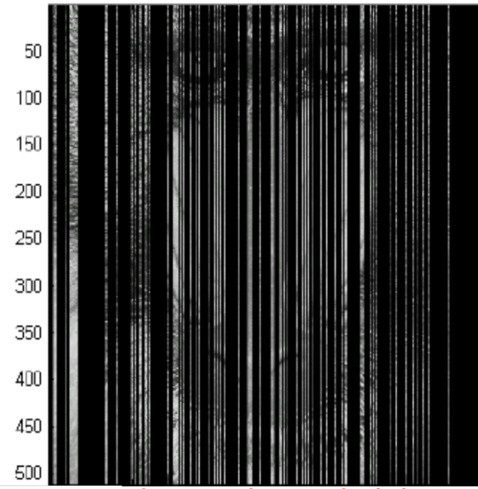
Randomized low-rank matrix approximations

- Randomized (SVD)
- approximation C (e.g. a subset of the columns of A)
- $AA^T \approx CC^T$
- $\tilde{A}_k = CC^\dagger A$ is a randomized rank-k approximation
- $\|A - \tilde{A}_k\|_2^2 \leq \sigma_{k+1}^2 + \epsilon \|A\|_2^2$

Randomized low-rank approximation example



original data



column subsampled data

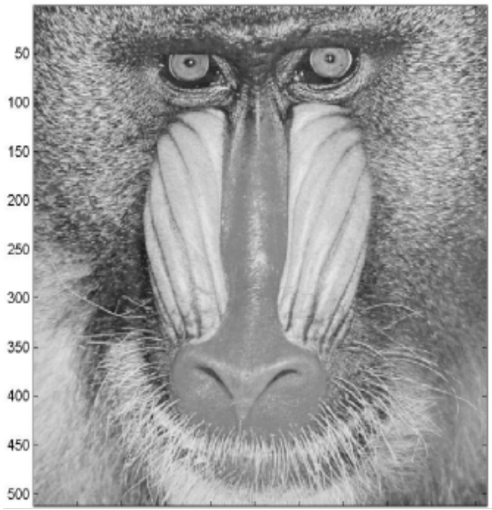
A

AS

$$AS(AS)^{+}A = \underline{QQ^{T}A} \approx A$$

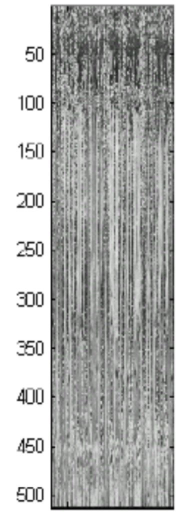
Randomized low-rank approximation example

A



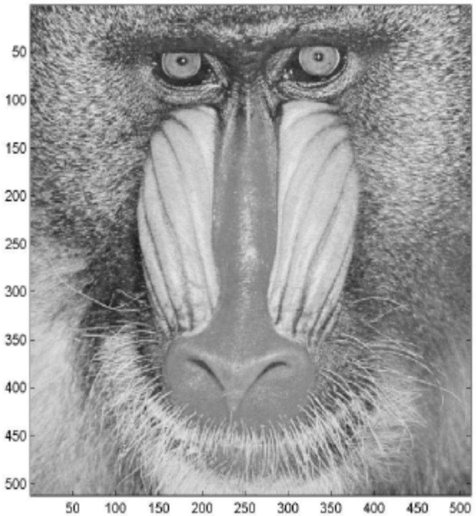
original data

$C = AS$

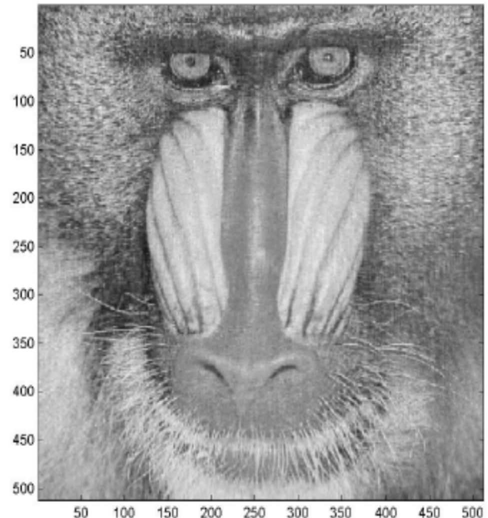


sketch

Randomized low-rank approximation example



original data



randomized low rank approximation

$$CC^{+}A \approx A$$

$$\| A - CC^{+}A \|_2^2 \le \sigma_{k+1}^2 + 2 \cdot \epsilon$$