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 Singular Value Decomposition and CX Decomposition

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 $A_k := U_k \Sigma_k V_k := \sum_{i=1}^k \sigma_i u_i v_i^T$. We have
  $$\|A - A_k\|_2 \leq \sigma_{k+1}$$
- computational cost of computing the SVD is $O(nd^2)$ for $n \geq d$

Randomized Matrix Decompositions

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

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 $C = AS$

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

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

- 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

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

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

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

Randomized Matrix Decompositions: Spectral Norm Error

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

  Lemma 1 (Drineas et al. 2006)

$$\|A - U_k U_k^T A\|_2^2 \leq \|A - U_A U_A^T A\|_2^2 + 2\|AA^T - CC^T\|_2$$

- 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 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

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

Randomized Matrix Decompositions: Spectral Norm Error

Proof of Lemma 1 cont'd

taking squares

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

Randomized Matrix Decompositions: Spectral Norm Error

Proof of Lemma 1 cont'd

taking squares

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

- 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 columns 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

- Given a large matrix $A \in \mathbb{R}^{n \times d}$
- idea: sample some columns of A to get $C \in \mathbb{R}^{n \times m}$
- $C = AS$
- approximation $CC^{\dagger} = (AS)(AS)^{\dagger}A \approx A$
- calculate QR decomposition of $AS = QR$
- 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 $A \approx (QU)\Sigma V^{T}$

Randomized Singular Value Decomposition

- Given a large matrix $A \in \mathbb{R}^{n\times d}$
- idea: sample/sketch some columns 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)

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

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

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

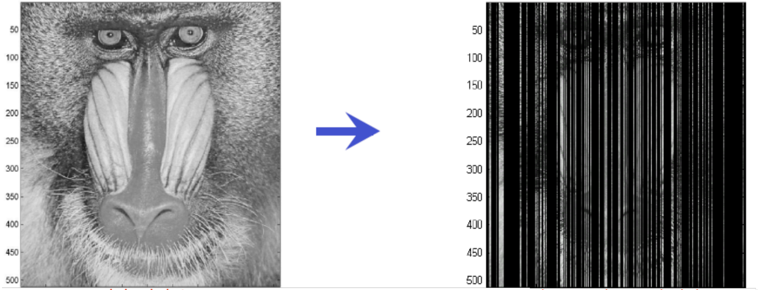
Low-rank matrix approximations

- 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 $A_k := U_k\Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T$
- $\|A - A_k\|_2 \leq \sigma_{k+1}$

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
- $\|A - \tilde{A}_k\|_2^2 \leq \sigma_{k+1}^2 + \epsilon\|A\|_2^2$

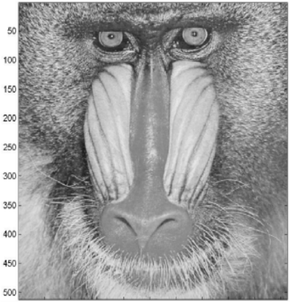
Randomized low-rank approximation example



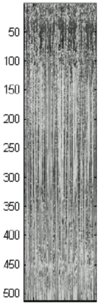
original data

column subsampled data

Randomized low-rank approximation example

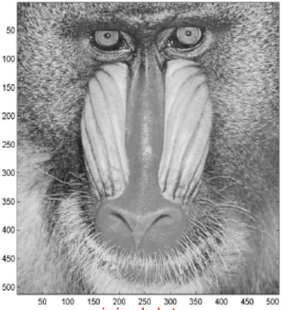


original data

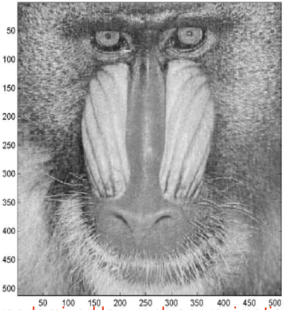


sketch

Randomized low-rank approximation example



original data



randomized low rank approximation