Randomized Linear Algebra and Optimization
Lecture 18: Generalized Least Squares Problems, Randomized Low Rank Approximations and Power Iteration
Recap: Low-rank matrix approximations

- **Singular Value Decomposition (SVD)**

  \[ A = U \Sigma V^T \]

  takes \( O(nd^2) \) time for \( A \in \mathbb{R}^{n \times d} \)

- **best rank-\( k \) approximation** \( A_k := U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T \)

  \[ \| A - A_k \|_2 \leq \sigma_{k+1} \]
Recap: Randomized low-rank matrix approximations

idea: sample some rows/sketch $A \in \mathbb{R}^{n \times d}$ to get $C \in \mathbb{R}^{n \times m}$

$C = AS$ where $S \in \mathbb{R}^{d \times m}$ is a sampling/sketching matrix

▶ we have an approximate matrix multiplication $AA^T \approx CC^T$.

Next, consider the best approximation of $A$ in the range of $C = AS$

$$\min_{X} \|CX - A\|_F$$

also called CX decomposition

▶ $\tilde{A}_m := CX^* = CC^\dagger A$ is a randomized rank-m approximation

$$(AS)(AS)^\dagger A \approx A$$
Recap: Randomized Singular Value Decomposition

- CX decomposition provides the approximation
  \[(AS)(AS)^\dagger A \approx A\]

- calculate QR decomposition of \(AS = QR\)
  \(QQ^T A \approx A\), i.e., \(Q\) approximates the range space of \(A\)

- calculate the SVD \(Q^T A = U\Sigma V^T\)

- approximate SVD of \(A\) is \(A \approx (QU) \underbrace{\Sigma V^T}_{U'}\)

  note that \(U' := QU\) and \(V\) are orthogonal and \(\Sigma\) is diagonal
  we have an approximation of the left and right singular vectors and singular values of \(A\)
Analysis of Randomized Low Rank Approximations

- **CX decomposition**: form sketch $AS$, and find the best approximation of $A$ in the range of $AS$

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

- Approximation $ASX^* = (AS)(AS)^\dagger A \approx A$
yields randomized SVD: $AS = QR$ and $Q^TA = U\Sigma V^T$

- Let $A = U\Sigma V^T$ and $A_k = \sum_{i=1}^k \sigma_k u_k v_k^T$, i.e.,
best rank-$k$ approximation of $A$

Note that

$$\|AS (AS)^\dagger A - A\|_F^2 \leq \|AS(A_kS)^\dagger A_k - A\|_F^2$$

$$= \|A_k^T(S^TA_k^T)^\dagger S^TA^T - A^T\|_F^2$$
Analysis of Randomized Low Rank Approximations

**approximation error**

\[
\|A S (A S)^\dagger A - A\|_F^2 \leq \|A S (A_k S)^\dagger A_k - A\|_F^2
\]

\[
\leq \|A_k^T (S^T A_k^T)^\dagger S^T A^T - A^T\|_F^2
\]

\[
= \|A_k^T X - A^T\|_F^2
\]

where

\[
\tilde{X} := \arg \min_X \|S^T A_k^T X - S^T A^T\|_F^2
\]
Analysis of Randomized Low Rank Approximations

- approximation error

\[ \|AS(AS)^\dagger A - A\|_F^2 \leq \|AS(A_k S)^\dagger A_k - A\|_F^2 \]

\[ = \|A_k^T (S^T A_k^T)^\dagger S^T A^T - A^T\|_F^2 \]

\[ = \|A_k^T \tilde{X} - A^T\|_F^2 \]

where

\[ \tilde{X} := \arg\min_X \|S^T A_k^T X - S^T A^T\|_F^2 \]

- identical to left sketching the Generalized Least Squares problem

\[ \min_X \|A_k^T X - A^T\|_F^2 \]
Generalized Least Squares Problems

\[
\min_X \|AX - B\|_F^2
\]

▶ Least Squares problem with multiple right-hand-sides
\[
B = [b_1, ..., b_r]
X = [x_1, ..., x_r]
\]

\[
\min_{x_1, \ldots, x_r} \sum_{i=1}^r \|Ax_i - b_i\|_2^2
\]

▶ optimal solution
\[
X^* = [x_1^*, ..., x_r^*]
= [A^\dagger b_1, ..., A^\dagger b_r]
= A^\dagger B
\]
Left Sketching Generalized Least Squares Problems

- original problem

\[ \mathbf{X}^* := \arg \min_{\mathbf{X}} \| \mathbf{AX} - \mathbf{B} \|_F^2 \]

- form sketches of the data \( \mathbf{SA} \) and \( \mathbf{SB} \), e.g.,
  uniform row sampling, weighted sampling, Gaussian, \( \pm 1 \) i.i.d, CountSketch, FJLT...

\[ \hat{\mathbf{X}} := \arg \min_{\mathbf{X}} \| \mathbf{SAX} - \mathbf{SB} \|_F^2 \]

\[ \hat{\mathbf{X}}_i = \arg \min_{\mathbf{x}_i} \| \mathbf{SAx}_i - \mathbf{Sb}_i \|_2^2 \]

\[ = (\mathbf{SA})^\dagger(\mathbf{Sb}_i) \]

- left-sketch applied to simple Least Squares problem

\[ \min_{\mathbf{x}_i} \| \mathbf{Ax}_i - \mathbf{b}_i \|_2^2 \]
Recall Gaussian Sketch Analysis

Let \( A \in \mathbb{R}^{n \times d} \), \( S \in \mathbb{R}^{m \times n} \) be i.i.d. Gaussian

\[
\begin{align*}
x^* := \arg \min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2 & \quad \text{and} \quad \tilde{x} = \arg \min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2^2 \\
\end{align*}
\]

Conditioned on the matrix \( SA \)

\[
A(\tilde{x} - x^*) \sim N\left(0, \frac{f(x^*)}{m} A(A^T S^T SA)^{-1} A\right)
\]
Recall Gaussian Sketch Analysis

Let $A \in \mathbb{R}^{n \times d}$, $S \in \mathbb{R}^{m \times n}$ be i.i.d. Gaussian

$$x^* := \arg \min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2$$

and

$$\tilde{x} = \arg \min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2^2$$

Conditioned on the matrix $SA$

$$A(\tilde{x} - x^*) \sim N\left(0, \frac{f(x^*)}{m} A (A^T S^T SA)^{-1} A\right)$$

Taking expectation over $SA$, and using

$$\mathbb{E}[(A^T S^T SA)^{-1}] = (A^T A)^{-1} \frac{m}{m - d - 1}$$

we get

$$\mathbb{E}\|A(\tilde{x} - x^*)\|_2^2 = \frac{f(x^*)}{m - d - 1} \text{tr} A (A^T A)^{-1} A$$

$$= f(x^*) \frac{\text{rank}(A)}{m - d - 1} = f(x^*) \frac{d}{m - d - 1}$$
Left Sketching Generalized Least Squares Problems

- original problem and left-sketch
  \[
  X^* := \arg \min_X \|AX - B\|_F^2 \quad \text{and} \quad \hat{X} := \arg \min_X \|SAX - SB\|_F^2
  \]

- \(x_i\): \(i\)-th column of \(\hat{X}\) satisfies
  \[
  \hat{x}_i = \arg \min_{x_i} \|SAx_i - Sb_i\|_2^2
  \]

- For a Gaussian sketching matrix \(S\) we have
  \[
  \mathbb{E}\|A(\hat{x}_i - x^*_i)\|_2^2 = \|Ax^*_i - b_i\|_2^2 \frac{d}{m - d - 1}
  \]
  implies
  \[
  \mathbb{E}\|A(\hat{X} - X^*)\|_F^2 = \sum_{i=1}^r \|Ax^*_i - b_i\|_2^2 \frac{d}{m - d - 1}
  \]
  \[
  = \|AX^* - B\|_F^2 \frac{d}{m - d - 1}
  \]
Left Sketching Optimality Gap

- suppose that $\text{rank}(A) = r$

original problem and left-sketich

$$X^* := \arg \min_X \|AX - B\|_F^2 \quad \text{and} \quad \hat{X} := \arg \min_X \|SAX - SB\|_F^2$$

$$\mathbb{E}\|A(\hat{X} - X^*)\|_F^2 = \|AX^* - B\|_F^2 \frac{r}{m - r - 1}$$

$$\mathbb{E}\|A\hat{X} - B\|_F^2 = \mathbb{E}\|AX^* - B + A(\hat{X} - X^*)\|_F^2$$

$$= \|AX^* - B\|_F^2 + \mathbb{E}\|A(\hat{X} - X^*)\|_F^2$$

$$= \|AX^* - B\|_F^2 (1 + \frac{r}{m - r - 1})$$

$$= \|AX^* - B\|_F^2 \frac{m - 1}{m - r - 1}$$
Back to Randomized Low Rank Approximations

**approximation error**

\[ E \| AS (AS)^\dagger X^\star - A \|^2_F \leq E \| AS(A_k S)^\dagger A_k - A \|^2_F \]

\[ = \| A_k^T (S^T A_k^T)^\dagger S^T A^T - A^T \|^2_F \]

\[ = E \| A_k^T \tilde{X} - A^T \|^2_F \]

\[ \leq \frac{m - 1}{m - k - 1} \| A_k^T (A_k^T)^\dagger A^T - A^T \|^2_F \]

\[ \leq \frac{m - 1}{m - k - 1} \| A(A_k A_k^\dagger - I) \|^2_F \]

\[ \leq \frac{m - 1}{m - k - 1} \| A_k - A \|^2_F \]
Randomized Low Rank Approximation and Randomized SVD Error Bound

- CX decomposition and randomized SVD

\[ AS(AS)^\dagger A \approx A \]

- final Frobenious norm error bound

\[ \mathbb{E} \| AS(AS)^\dagger A - A \|_F^2 \leq \frac{m - 1}{m - k - 1} \| A_k - A \|_F^2 \]

valid for any \( k \in \{1, \ldots, \text{rank}(A)\} \)
Randomized Low Rank Approximation and Randomized SVD Error Bound

- **CX decomposition and randomized SVD**
  \[ AS(AS)\dagger A \approx A \]

- **final Frobenious norm error bound**
  \[ \mathbb{E} \|AS(AS)\dagger A - A\|_F^2 \leq \frac{m-1}{m-k-1} \|A_k - A\|_F^2 \]
  valid for any \( k \in \{1, ..., \text{rank}(A)\} \)

- **define the oversampling factor** \( \ell := m - k - 1 \)
  \[ \|AS(AS)\dagger A - A\|_F^2 \leq (1 + \frac{k}{\ell}) \|A_k - A\|_F^2 \]
Reducing the Error: Power Iteration

- error bounds depend on tail singular values

\[ \|A_k - A\|_F^2 = \sum_{j=k+1}^{\text{rank}(A)} \sigma_j^2 \]

- idea: compute the sketch of \((AA^T)^q A\)

\[ C = (AA^T)^q A S \]

where \(q\) is an integer parameter

\[ CC^\dagger A \approx A \]

\(CC^\dagger\) approximates the range of \(A\) better for \(q \geq 1\)

- singular values of \((AA^T)^q A\) are \(\sigma_i(A)^{2q+1}\)

where \(\sigma_i(A)\) are the singular values of \(A\)
Connection to the Classical Power Iteration

- Suppose that $A^T A$ has a unique maximal eigenvector.
- Starting from a nonzero vector $x_0$, the iterations
  \[ x_{t+1} = \frac{A^T A x_t}{\|A^T A x_t\|_2} \]
  converges to a multiple of the maximal eigenvector.
Simultaneous (QR) Iteration

- suppose that the top-\(m\) eigenvalues of \(A^T A\) are distinct
- Starting from a matrix \(X_0\) of rank \(m\), e.g., a random \(d \times m\) matrix, the iterations

\[
Q_t R_t = X_t \\
X_{t+1} = A^T A Q_t
\]

where \(Q_t R_t\) is the QR factorization of \(X_t\), converges to an orthonormal basis for the top-\(m\) eigenspace of \(A^T A\)

- note that the first iteration computes the QR decomposition \(A^T A S = QR\) where \(S\) is random. This is similar to the CR-decomposition and randomized SVD approach, where the QR decomposition of \(A S\) serves as a crude approximation of the top eigenspace basis