
HPC for numerical methods and data analysis

Fall Semester 2023

Prof. Laura Grigori

Assistant: Mariana Martinez

Session 7 – October 31, 2023

Randomized SVD

Exercise 1: SRHT

In the context of overdetermined least-squares problems, we need to find $x \in \mathbb{R}^n$ such that it minimizes:

$$\|Ax - b\|_2^2,$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, m > n$. There is a class of randomized algorithms for solving this problem based on sketching method. Sketching methods involve using a random matrix $\Omega \in \mathbb{R}^{r \times m}$ to project the data A (and maybe also b) to a lower dimensional space with $r \ll m$. Then they approximately solve the least-squares problem using the sketch ΩA (and/or Ωb). One relaxes the problem to finding a vector x so that

$$\|Ax - b\| \leq (1 + \varepsilon)\|Ax^* - b\|,$$

where x^* is the optimal solution. The overview of sketching applied to solve linear least squares is:

a) Sample/build a random matrix Ω

b) Compute ΩA and Ωb

c) Output the exact solution to the problem $\min_x \|(\Omega A)x - (\Omega)b\|_2$.

Given a data matrix, $X \in \mathbb{R}^{m \times n}$, we want to reduce the dimensionality of X by defining a random orthonormal matrix $\Omega \in \mathbb{R}^{r \times m}$ with $r \ll m$. For $m = 2^q, q \in \mathbb{N}$, the Subsampled Randomized Hadamard Transform (SRHT) algorithm defined a $r \times m$ matrix as:

$$\Omega = \sqrt{\frac{m}{r}} P H_m D,$$

where:

- $D \in \mathbb{R}^{m \times m}$ is a diagonal matrix whose elements are independent random signs, i.e. it's diagonal entries are just -1 or 1.

- $H \in \mathbb{R}^{m\times m}$ is a **normalized** Walsh-Hadamard matrix. If you're going to use a library that implements this transform then check that it implements the normalized Walsh-Hadamard matrix. This matrix is defined recursively as:

$$H_m = \begin{bmatrix} H_{m/2} & H_{m/2} \\ H_{m/2} & -H_{m/2} \end{bmatrix} \qquad H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$
$$H = \frac{1}{\sqrt{m}} H_m \in \mathbb{R}^{m\times m}.$$

- $P \in \mathbb{R}^{r\times m}$ is a subset of randomly sampled r columns from the $m \times m$ identity matrix. The purpose of using P is to uniformly sample r columns from the rotated data matrix $X_{\text{rot}} = H_m DX$.

The following theorem help us get an idea for the size of r.

Theorem 1 (Subsampled Randomized Hadamard Transform) *Let* $\Omega = \sqrt{\frac{m}{r}} P H_m D$ *as previously defined. Then if*

$$r \geq \mathcal{O}((\varepsilon^{-2}\log(n))(\sqrt{n} + \sqrt{\log m})^2)$$

with probability $0,99$ *for any fixed* $U \in \mathbb{R}^{m\times n}$ *with orthonormal columns:*

$$\|I - U^\top \Omega \Omega^\top U\|_2 \leq \varepsilon.$$

Further, for any vector $x \in \mathbb{R}^m, \Omega x$ *can be computed in* $\mathcal{O}(n \log r)$ *time.*

Choose a data set from [https://www.kaggle.com/datasets?tags=13405-Linear+Regression]. Compare the randomized least squares fit using SRHT vs the deterministic least squares fit. Use the previous theorem to estimate r. *Hint: you can use the fast Hadamard transform from scipy or pytorch*

Exercise 2: Randomized SVD

Rokhlin, Szlam, and Tygert introduced an algorithm called *Blanczos* such that it computes the whole approximation $U\Sigma V^\top$ to an SVD of a matrix $A \in \mathbb{R}^{m\times n}$.

Test this algorithm by constructing a rank$-k$ approximation with $k = 10$ to a matrix $A \in \mathbb{R}^{m\times 2m}$ via its SVD:

$$A = U^{(A)}\Sigma^{(A)}V^{(A)\top},$$

where:

- $U \in \mathbb{R}^{m\times m}$ is a Hadamard matrix
- $V \in \mathbb{R}^{2m\times 2m}$ is a Hadamard matrix
- $\Sigma \in \mathbb{R}^{m\times 2m}$ is a diagonal matrix whose diagonal entries are defined as:

$$\Sigma_{jj} = \sigma_j = (\sigma_{k+1})^{\lfloor j/2 \rfloor/5},$$

  for $j = 1, 2, ..., 9, 10$ and

$$\Sigma_{jj} = \sigma_j = \sigma_{k+1}\frac{m-j}{m-11},$$

  for $j = 11, 12, ..., m-1, m$. Thus $\sigma_1 = 1$ and $\sigma_k = \sigma_{k+1}$.

Algorithm 1 Blanczos

Input: $A \in \mathbb{R}^{m \times n}$, i, l such that $k < l$ and $(i+1)l \leq m - k$

Output: U, Σ, V

Form a real $l \times n$ matrix G such that its entries are i.i.d. Gaussian random variables with mean zero and unit variance. Compute:

$$\begin{aligned} R^{(0)} &= GA \\ R^{(1)} &= R^{(0)} A^\top A \\ &\vdots \\ R^{(i)} &= R^{(i-1)} A^\top A. \end{aligned}$$

Form the $(i+1)l \times n$ matrix:

$$R^\top = \begin{bmatrix} (R^{(0)})^\top & (R^{(1)})^\top & \dots & (R^{(i)})^\top \end{bmatrix}$$

Form a real $n \times (i+1)l$ matrix Q whose columns are orthonormal and such that there is a real $(i+1)l \times (i+1)l$ matrix S in such way that $R^\top = QS$

$T \leftarrow AQ$

Form the SVD of T, $T = U \Sigma W^\top$

$V \leftarrow QW$

Set $l = k+12, i = 1$ test this algorithm for $m = 2^{11}$, $\sigma_{k+1} = 0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001$. Plot the decay of the singular values of A and compare such decay with the accuracy of the approximation, $\|A - U\Sigma V^\top\|_{\mathrm{F}}$ and the relative error, $\frac{\|A - U\Sigma V^\top\|_{\mathrm{F}}}{\|A\|_{\mathrm{F}}}$.