
HPC for numerical methods and data analysis

Fall Semester 2024

Prof. Laura Grigori

Assistant: Mariana Martínez Aguilar

Session 8 – November 5, 2024

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:

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

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

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

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 $\min_x \|(\Omega W)x - (\Omega)b\|_2$.

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

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

where:

- $D \in \mathbb{R}^{m \times m}$ is a diagonal matrix whose elements are independent random signs, i.e. its 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:

$$H_m = \begin{bmatrix} H_{m/2} & H_{m/2} \\ H_{m/2} & -H_{m/2} \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$H = \frac{1}{\sqrt{m}} H_m \in \mathbb{R}^{m \times m}.$$

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

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

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

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

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 $x \in \mathbb{R}^m$, Ωx can be computed in $\mathcal{O}(n \log l)$ 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 l . *Hint: you can use the fast Hadamard transform from scipy or pytorch*

Exercise 2: Randomized SVD

Consider the following algorithm to compute a randomized SVD factorization:

Algorithm 1 Randomized SVD $q = 1$

Input: $A \in \mathbb{R}^{m \times n}$, desired rank k , $l = p + k$

Output: Approximation $A_k = Q U \Sigma V$

Sample an $n \times l$ test matrix Ω_1 with independent mean-zero, unit-variance Gaussian entries.

Compute $Y = (A A^\top) A \Omega_1$

Construct $Q \in \mathbb{R}^{m \times l}$ with columns forming an orthonormal basis for the range of Y .

Compute $B = Q^\top A$, $B \in \mathbb{R}^{l \times n}$

Compute the rank- k truncated SVD of B as $U \Sigma V^\top$, $U \in \mathbb{R}^{l \times k}$, $V \in \mathbb{R}^{n \times k}$

Remember the following theorem:

Theorem 2 If Ω is chosen to be i.i.d. $\mathcal{N}(0, 1)$, $k, p \geq 2$, then the expectation with respect to the random matrix Ω is:

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

and the probability that the error satisfies

$$\|A - QQ^\top A\|_2 \leq \left(1 + 11\sqrt{k+p}\sqrt{\min(m,n)}\right) \sigma_{k+1}(A)$$

is at least $1 - 6/p^p$. For $p = 6$, the probability becomes 0,99.

Construct a rank- k approximation with $k = 10, p = 6$ to a matrix $A \in \mathbb{R}^{m \times 2m}$ via its SVD, the matrix A is defined as follows

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

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 $j = 1, 2, \dots, 9, 10$ and

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

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

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, $\|A - QQ^\top A\|_2$. Compare it with the theorem presented above.