
HPC for numerical methods and data analysis

Fall Semester 2024

Prof. Laura Grigori

Assistant: Mariana Martínez Aguilar

Session 6 – October 15, 2024

Sketching techniques

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 methods. 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 $r \ll m$. This can be used to approximately solve the least-squares problem using the sketches ΩA and 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 $\min_x \|(\Omega A)x - (\Omega b)\|_2$.

Exercise 1: General properties of sketching techniques

An ε -subspace embedding for the column space of a $m \times n$ matrix A is a matrix Ω for which for all $x \in \mathbb{R}^n$ the following property is satisfied with high probability:

$$(1 - \varepsilon)\|Ax\|_2 \leq \|\Omega Ax\|_2 \leq (1 + \varepsilon)\|Ax\|_2. \quad (1)$$

Let U be a matrix whose columns form an orthonormal basis for the column space of A . Prove that the requirement of an ε -subspace embedding can be simplified to:

$$\|I - U^\top \Omega^\top \Omega U\|_2 \leq \epsilon. \quad (2)$$

What does 2 mean? How does this compare to 1?

Exercise 2: Gaussian

The most “classical” sketch is a matrix $\Omega \in \mathbb{R}^{r \times m}$ with independent and identically distributed (i.i.d.) Gaussian entries $\mathcal{N}(0, 1/r)$. The following theorem from [1] provides the optimal number of rows of Ω up to a constant factor $\mathcal{O}(r\epsilon^{-2})$:

Theorem 1 *Let $0 < \epsilon, \delta < 1$ and $\Omega = \frac{1}{\sqrt{r}}R \in \mathbb{R}^{r \times m}$ where the entries $R_{i,j}$ of R are independent standard normal random variables. Then if $r = \mathcal{O}((n + \log(1/\delta))\epsilon^{-2})$, then for any fixed $m \times n$ matrix A , with probability $1 - \delta$, Ω is a $(1 \pm \epsilon)$ l_2 -subspace embedding for A , that is, simultaneously for all $x \in \mathbb{R}^n$,*

$$\|\Omega Ax\|_2 = (1 \pm \epsilon)\|Ax\|_2 \quad (3)$$

Choose a data set from [<https://www.kaggle.com/datasets?tags=13405-Linear+Regression>]. Compare the linear regression obtained from solving the deterministic least squares problem vs the one obtained from the randomized least squares problem with $\Omega \in \mathbb{R}^{r \times m}$ a normal random variable. That is, using the previous theorem with $\delta = 0.99$ choose different values of ϵ and compare the difference between the randomized least squares fit vs the deterministic one. Check that (2) holds for every ϵ you choose.

Exercise 3: SRHT

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:

$$\Omega = \sqrt{\frac{m}{r}}PH_mD,$$

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 (if it doesn’t then add the normalizing factor yourself). This matrix is defined recursively as:

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

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

- $P \in \mathbb{R}^{r \times m}$ 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 $X_{\text{rot}} = H_mDX$.

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

Theorem 2 (Subsampled Randomized Hadamard Transform) *Let $\Omega = \sqrt{\frac{m}{r}}PH_mD$ 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 $x \in \mathbb{R}^m$, Ωx can be computed in $\mathcal{O}(n \log r)$ time.

Take the same data set from the previous exercise. Compare the randomized least squares fit using SRHT, Gaussian, and the deterministic least squares fit. Use the previous theorem to estimate r .
Hint: you can use the fast Hadamard transform from scipy or pytorch