CS 15-851: Algorithms for Big Data Spring 2025

Lecture 2 — 1/23/2024

Prof. David Woodruff *Scribe: Aksara Bayyapu*

Matrix Chernoff Bound Let $X_1, .., X_n$ be independent copies of a symmetric random matrix $X \in \mathbb{R}^{d \times d}$ with $\mathbb{E}[X] = 0, |X|_2 \leq \gamma$, and $|\mathbb{E}[X^T X]|_2 \leq \sigma^2$. Let $W = \frac{1}{s}\Sigma_{i \in [s]} X_i$. For any $\epsilon > 0$,

$$Pr[|W_2| > \epsilon] \leq 2d \cdot e^{-s\epsilon/(\sigma^2 + \frac{\gamma\epsilon}{3})}$$

We use Matrix Chernoff Bound to show that the Subsampled Randomized Hadamard Transform sketching $S = PHD$ is a subspace embedding. We use the bound to show matrices are concentrated around the mean.

Notice that increase the samples will increase the concentration around the mean. Similarly, increasing the variance will decrease the concentration.

In the sketching, P is a sampling matrix and samples s rows uniformly with replacement. Let $V = HDA$, and recall that V has orthonormal columns. Notice, P samples rows of V.

Let Y_i be the i-th sampled row of $V = HDA$. Note the following property.

$$\begin{aligned} \mathbb{E}[Y_i^T Y_i] &= \Sigma_i Pr[Y_i = v_j] v_j^T v_j \\ &= \frac{1}{n} \Sigma_i v_i^T v_i \\ &= \frac{1}{n} V^T V \end{aligned}$$

Let $X_i = I_d - n \cdot Y_i^T Y_i$.

We verify $\mathbb{E}[X_i] = 0, |X_i|_2 \leq \gamma$, and $|\mathbb{E}[X_i^T X_i]|_2 \leq \sigma^2$ to apply the Matrix Chernoff Bounds.

First, we verify $\mathbb{E}[X_i] = 0$. By definition, X_i and V are orthonormal. Thus, use the above property of $\mathbb{E}[Y_i^T Y_i]$ to show,

$$\begin{aligned} \mathbb{E}[X_i] &= \mathbb{E}[I_d - n \cdot Y_i^T Y_i] \\ &= \mathbb{E}[I_d - n \ \frac{1}{n} V^T V] \\ &= I_d - I_d \\ &= 0^{d \times d} \end{aligned}$$

Next, we verify $|X_i|_2 \leq \gamma$.

We are interested in bounding the norm $|X_i|_2$, which is given by:

$$|X_i|_2 = \sup_{|z|_2 = 1} |X_i z|_2.$$

Using the triangle inequality, we can split the norm as:

$$|X_i|_2 \leq |I_d|_2 + |n \cdot Y_i^\top Y_i|_2.$$

The norm of the identity matrix I_d is the largest singular value, which is:

$$|I_d|_2 = 1.$$

For the second term, consider $|n \cdot Y_i^\top Y_i|_2$. Recall that Y_i is the i-th sampled row of $V = \mathrm{HDA}$. We analyze the squared norm of Y_i, which can be written as:

$$|Y_i|_2^2 = |e_j \cdot \mathrm{HDA}|_2^2,$$

where e_j is the j-th standard basis vector corresponding to the sampled row.

Thus, we can bound the spectral norm contribution as:

$$|n \cdot Y_i^\top Y_i|_2 \leq n \cdot \max |e_j \cdot \mathrm{HDA}|_2^2.$$

Combining these bounds, we have:

$$|X_i|_2 \leq |I_d|_2 + n \cdot \max |e_j \cdot \mathrm{HDA}|_2^2.$$

Substituting $|I_d|_2 = 1$, the result becomes:

$$|X_i|_2 \leq 1 + n \cdot \max |e_j \cdot \mathrm{HDA}|_2^2.$$

Concentration bounds show that for a random sampling of rows from HDA, the maximum contribution $|e_j \cdot \mathrm{HDA}|_2^2$ is bounded by:

$$|e_j \cdot \mathrm{HDA}|_2^2 \leq C^2 \log\left(\frac{nd}{\delta}\right),$$

Thus, the final bound becomes:

$$|X_i|_2 \leq 1 + n \cdot C^2 \log\left(\frac{nd}{\delta}\right),$$

which simplifies further to:

$$|X_i|_2 = \Theta\left(d \log\left(\frac{nd}{\delta}\right)\right),$$

for appropriately chosen values of n, d, and δ.

Lastly, we verify $|\mathbb{E}[X_i^T X_i]|_2 \leq \sigma^2$

Start with $\mathbb{E}[X^\top X + I_d]$. We begin with the following decomposition, which is derived by substituting $X_i = I_d - n \cdot Y_i^\top Y_i$ into the definition of $\mathbb{E}[X^\top X]$.

$$\mathbb{E}[X^\top X + I_d] = I_d + I_d - 2n\mathbb{E}[Y_i^\top Y_i] + n^2\mathbb{E}[Y_i^\top Y_i Y_i^\top Y_i].$$
$$= 2I_d - 2I_d + n^2\mathbb{E}[Y_i^\top Y_i Y_i^\top Y_i].$$

$$= n^2\Sigma_i(\frac{1}{n})v_i^T v_i v_i^T v_i$$

$$= n\Sigma_i v_i^T v_i|v_i|_2^2$$

To bound $n^2\mathbb{E}[Y_i^\top Y_i Y_i^\top Y_i]$, we introduce a matrix Z and apply the flattening lemma.

$$Z = n\sum_i v_i v_i^\top C^2 \log\left(\frac{nd}{\delta}\right)\cdot\frac{d}{n},$$

where v_i represents the i-th row vector of V, the matrix with orthonormal columns. Both $\mathbb{E}[X^\top X + I_d]$ and Z are symmetric matrices with non-negative eigenvalues.

Claim:

$$y^\top\mathbb{E}[X^\top X + I_d]y \le y^\top Z y, \quad \forall y.$$

Proof:

Both sides of the inequality can be expressed as,

$$y^\top\mathbb{E}[X^\top X + I_d]y = n\sum_i y^\top v_i v_i^\top y|v_i|_2^2. = n\Sigma_i < v_i, y >^2 |v_i|_2^2$$

$$y^\top Z y = n\sum_i y^\top v_i v_i^\top y\cdot C^2\log\left(\frac{nd}{\delta}\right)\cdot\frac{d}{n} = d\Sigma_i < v_i, y >^2 C^2\log(\frac{nd}{\delta})$$

By construction, Z provides an upper bound for $\mathbb{E}[X^\top X + I_d]$, ensuring:

$$y^\top\mathbb{E}[X^\top X + I_d]y \le y^\top Z y.$$

Since both $\mathbb{E}[X^\top X + I_d]$ and Z are symmetric, we can bound their norms:

$$|\mathbb{E}[X^\top X]|_2 \le |\mathbb{E}[X^\top X + I_d]|_2 + |I_d|_2 = |\mathbb{E}[X^\top X + I_d]|_2 + 1 \le |Z|_2 + 1$$

From the definition of Z, its spectral norm is:

$$|Z|_2 = C^2 d\log\left(\frac{nd}{\delta}\right)$$

Thus, we obtain:

$$|\mathbb{E}[X^\top X]|_2 \le C^2 d\log\left(\frac{nd}{\delta}\right) + 1$$

Finally,

$$|\mathbb{E}[X^\top X]|_2 = \mathcal{O}\left(d\log\left(\frac{nd}{\delta}\right)\right).$$

Now, we are able to use the Matrix Chernoff Bound. Apply the bound as follows.

$$Pr\left[\left|I_d - (\text{PHDA})^T(\text{PHDA})\right|^2 \ge \epsilon\right] \le 2d\cdot e^{\left(-\frac{s\epsilon^2}{\Theta(d\log(nd/\delta))}\right)}$$

Set $s = d\log(\frac{nd}{\delta})\frac{log(\frac{d}{\delta})}{\epsilon^2}$ to make this probability less than $\frac{\epsilon}{2}$

SRHT Wrapup

We have now shown that $\left|I_d - (\text{PHDA})^T(\text{PHDA})\right|^2 \leq \epsilon$ using Matrix Chernoff Bounds. This allows us to construct a subspace embedding. This implied tat for every unit vector x,

$$\begin{aligned}\left|1 - |PHDAx|_2^2\right| &= \left|x^T \left(I_d - (P_{\text{HDA}})^T P_{\text{HDA}}\right) x\right| \\ &= \left|x^T x - x^T (P_{\text{HDA}})^T P_{\text{HDA}} x\right| \\ &= \left|I - |(SAx)|_2^2\right| < \epsilon \\ &\implies |(SAx)|_2^2 \in [1-\epsilon, 1+\epsilon]\end{aligned}$$

Having established the subspace embedding, we apply the technique for Gaussian sketch matrices S to derive a solution to the original regression problem. This approach results in an algorithm with a running time of:

$$O(nd \log n) + \text{poly}\left(\frac{d \log n}{\epsilon}\right)$$

CountSketch Intro

CountSketch matrices are used to obtain even faster subspace embeddings.

Definition (CountSketch Matrix): The matrix S is a $k \times n$ matrix with $k = O\left(\frac{d^2}{\epsilon^2}\right)$. Each column of S contains exactly one non-zero entry, which is either $+1$ or -1 with equal probability.

SA can be computed in $\text{nnz}(A)$ time. We maintain a list of indices for the non-zero entries, and simply index into A to obtain the product. The rest of the entries are zero, which does not affect the result.