CS 15-851: Algorithms for Big Data	Spring 2025

Lecture 3 — 1/30/2025

Prof. David Woodruff *Scribe: Faisal Baqai*

1 Preliminaries; CountSketch

Continuing from last time, $\exists$ a sketching matrix (called CountSketch) with faster subspace embedding than the subsampled randomized Hadamard transform (SRHT). Returning to the regression framework

$$\min_{X \in R^d} \left|\left| \begin{bmatrix} A_{1,1} & \cdots & A_{1,d} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \cdots & A_{n,d} \end{bmatrix} X - \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \right|\right|_2$$

SRHT's time complexity (from previous lecture): $O(nd \log n + poly(d \log(n)/\epsilon))$

Definition. The CountSketch matrix is a sparse k by n matrix such that each column has one uniformly chosen nonzero entry drawn from a Rademacher random variable, where k is $O(\frac{d^2}{\epsilon^2})$

An example of a CountSketch matrix is

$$\begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 1 & -1 & 0 \end{bmatrix}$$

The time complexity of sketching with CountSketch is $O(nnz(A)) + poly(\frac{d}{\epsilon})$ which is still dominated by matrix multiplication, but that only takes $O(nnz(A))$ time now, where nnz(A) is the number of nonzero entries of A. (This is because each column of S only has one non-zero entry, which implies that each nonzero entry of A is counted only once and with constant operations per nonzero entry of A during the multiplication.)

2 Subspace Embedding from Matrix Product

We now need to show correctness: that CountSketch is a subspace embedding. Symbolically: $|SAx|_2 = 1 \pm \epsilon$ for all unit x. The following proof is due to Nguyen.

1. Assume the columns of A are orthonormal wlog (Change of Basis)
2. Choose $k = \frac{6d^2}{\delta\epsilon^2}$ so that SA is a $(6d^2)/(\delta\epsilon^2) \times d$ matrix.

Claim 1. In order to show $|SAx|_2 \leq 1 \pm \epsilon$ for all unit x, it suffices to show $|A^T S^T SA - I|_F \leq \epsilon$

Proof. Assume $|A^TS^TSA-I|_F \leq \epsilon$. Then $|A^TS^TSA-I|_2 \leq \epsilon$ because the frobenius norm dominates the operator norm.
Then by definition of the operator norm: $\sup_{unitx} |(SAx)^T(SAx) - x^Tx|_2 \leq \epsilon$
$\implies ||SAx|_2^2 - 1| \leq \epsilon$ by the triangle inequality Rearranging,
$\implies |SAx|_2^2 \leq 1 \pm \epsilon$
$\implies |SAx|_2 \leq 1 \pm \epsilon$

■

Therefore the goal has transformed, in order to show subspace embedding, now we need to bound $|A^TS^TSA - I|_F \leq \epsilon$. We will do this via the approximate matrix product result.

Claim 2. Matrix Product Property:

$$\mathbf{Pr}[|CS^TSD - CD|_F^2 \leq \frac{6}{\delta k} * |C|_F^2|D|_F^2] \geq 1 - \delta$$

The proof of the matrix product result is deferred to the next section, but for now, assuming the matrix product result is true and using pattern matching, we have $C = A^T$ and $D = A$ yielding $|C|_F^2 = |D|_F^2 = d$ because of the orthonormal assumption on A. Finally, solving for k gives the result with the number of rows of S $\geq \frac{6d^2}{\delta\epsilon^2}$

3 The JL Property yields Matrix Product

The goal is now to show that CountSketch satisfies the matrix product result. The proof of this is due to Kane and Nelson. First, we need a definition.

Definition. JL Property: A distribution on matrices S has the (ϵ, δ, l)-JL moment property if for all unit x,

$$\mathbb{E}[||Sx|_2^2 - 1]|^l \leq \epsilon^l\delta$$

Claim 3. If the JL property holds with $\epsilon, \delta \in (0, \frac{1}{2})$ and $l \geq 2$ then
$\mathbf{Pr}[|A^TS^TSB - A^TB|_F \geq 3\epsilon|A|_F|B|_F] \leq \delta$

Before proving Claim 3, we need some notation and a lemma on Minkowski's Inequality

Definition. Let the p-norm of a random scalar variable= $|X|_p = \mathbb{E}[X^p]^{\frac{1}{p}}$ with $p \geq 1$

Lemma 1. *Minkowksi's Inequality :* $|X + Y|_p \leq |X|_p + |Y|_p$

Proof. By convexity, $|\frac{1}{2}(X + Y)|_p \leq \frac{1}{2}|X|_p + \frac{1}{2}|Y|_p$
$\implies |X + Y|^p \leq 2^{p-1}(|X|^p + |Y|^p)$
Taking expectations of both sides yields $\mathbb{E}[|X + Y|_p^p] \leq \mathbb{E}[2^{p-1}|X|_p^p|Y|_p^p]$, showing we don't need to worry about infinities.

From definitions we have $\int |X + Y||X + Y|^{p-1} \leq \int |X||X + Y|^{p-1} + \int |Y||X + Y|^{p-1}$ by double application of Holder's inequality,

$\implies \leq ((\int |X|^p)^{\frac{1}{p}} + (\int |Y|^p))^{\frac{1}{p}} (\int |x+y|^p)^{\frac{p-1}{p}}$
Using the p norm notation $\implies |X+Y|_p |X+Y|_p^{p-1} \leq (|X|_p + |Y|_p)|X+Y|_p^{p-1}$
Finally $|X+Y|_p \leq |X|_p + ||Y|_p$

■

3.1 From vectors to matrices:

We now have the ingredients to prove Claim 3.

Proof. Starting with arbitrary unit vectors, we need a bound on $| < Sx, Sy > - < x, y > |_l$ where $||_l$ is the same p-norm from before.
Expanding, $\frac{1}{2}||Sx|_2^2 - 1 + |Sy|_2^2 - 1 - (|S(x-y)|_2^2 - |x-y|_2^2)|_l$
Applying lemma 3 (Minkowski) $\implies \leq \frac{1}{2}||Sx|_2^2 - 1|_l + ||Sy|_2^2 - 1|_l + |(|S(x-y)|_2^2 - |x-y|_2^2)|_l$
Using the JL-Property $\implies \leq \frac{1}{2}(\epsilon\delta^{1/l} + \epsilon\delta^{1/l} + |x-y|_2^2 \epsilon\delta^{1/l})$
$\leq 3\epsilon\delta^{1/l}$

From unit vectors to arbitrary vectors x,y, linearity gives

$$\frac{| < Sx, Sy > - < x, y > |_l}{|x|_2 |y|_2} \leq 3\epsilon\delta^{1/l}$$

Definition. $X_{i,j} = \frac{|<SA_i, SB_j> - <A_i, B_j>|_l}{|A_i|_2 |B_j|_2}$

From arbitrary vectors to the columns A_i and B_j of matrices A and B, we now write the left hand side of the matrix product property

$$|A^T S^T SB - A^T B|_F^2 = \sum_i \sum_j |A_i|_2^2 |B_j|_2^2 X_{ij}^2$$

$$||A^T S^T SB - A^T B|_F^2|_{l/2} = |\sum_i \sum_j |A_i|_2^2 |B_j|_2^2 X_{ij}^2|_{l/2}$$

Using triangle inequality to pull out the sums,

$$\implies \leq \sum_i \sum_j ||A_i|_2^2 |B_j|_2^2 X_{ij}^2|_{l/2}$$

$$\implies \sum_i \sum_j |A_i|_2^2 |B_j|_2^2 |X_{ij}|_l^2$$

Using the vector result and definition of Frobenius norm

$$\implies \leq (3\epsilon\delta^{1/l})^2 |A|_F^2 |B|_F^2$$

Since the p-norm is an expectation, we can use the Markov inequality to write (raised to the l)

$$\mathbf{Pr}[|A^T S^T SB - A^T B|_F^l \geq (3\epsilon |A|_F |B|_F)^l] \leq (\frac{1}{3\epsilon |A|_F |B|_F})^l \mathbb{E}[|A^T S^T SB - A^T B|_F^l]$$

By the definition of p-norm,

$$\mathbb{E}[|A^TS^TSB - A^TB|_F^l] = ||A^TS^TSB - A^TB|_F^l|_{l/2}^{l/2}$$

and so substituting in the earlier results

$$\mathbf{Pr}[|A^TS^TSB - A^TB|_F^l \geq (3\epsilon|A|_F|B|_F)^l] \leq (\frac{1}{3\epsilon|A|_F|B|_F})^l((3\epsilon\delta^{1/l})^2|A|_F^2|B_F^2)^{l/2} = \delta$$

■

Now the final step to close the loop is showing that CountSketch with the correct number of rows has the JL property. That will be in the next section.