CS 15-851: Algorithms for Big Data Spring 2024

Lecture 3 — 02/01/2024

Prof. David Woodruff *Scribe: Harrison Wiseblatt*

1 Approximate matrix product guarantees

1.1 From vectors to matrices

In this section, we prove the following theorem:

Theorem 1. *For $\varepsilon, \delta \in (0, \frac{1}{2})$, if D is a distribution on matrices $S \in \mathbb{R}^{k \times n}$ that satisfies the (ε, δ, l)-JL moment property for some $l \geq 2$, then we have for any matrices A, B with n rows*

$$\mathbb{P}\left[|A^T S^T SB - A^T B|_F \geq 3\varepsilon |A|_F |B|_F\right] \leq \delta$$

As a reminder,

Definition. A distribution on matrices $S \in \mathbb{R}^{k \times n}$ has the (ϵ, δ, l)-JL moment property if $\forall x \in \mathbb{R}^n$ with $|x|_2 = 1$, we have $E_S||Sx|_2^2 - 1|^l \leq \epsilon^l \cdot \delta$

Also, recall last time we used, for a random scalar X, the p-norm. This is defined as $|X|_p = (E|X|^p)^{1/p}$, and for $p \geq 1$ we have Minkowski's inequality which says that

$$|X + Y|_p \leq |X|_p + |Y|_p$$

Proof of theorem 1. Last lecture, we proved that for arbitrary vectors x, y with constant probability,

$$\frac{|\langle Sx, Sy\rangle - \langle x, y\rangle|_l}{|x|_2 |y|_2} \leq 3\epsilon * \delta^{\frac{1}{l}}$$

assuming S satisfies the (ϵ, δ, l)-JL moment property

Now, if we define $X_{i,j} = \frac{1}{|A_i|_2 |B_j|_2} \cdot (\langle SA_i, SB_j\rangle - \langle A_i, B_j\rangle)$, we can rearrange terms to get

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

Want to show $\mathbb{P}[|CS^T SD - CD|_F^2 \leq [\frac{6}{\delta * \text{num rows of S}}) * |C|_F^2 |D|_F^2] \geq 1 - \delta$ (ie., with constant probability S gives an approximate matrix product).

$$
\begin{aligned}
&\left||A^TS^TSB - A^TB|_F^2\right|_{l/2} \\
&= |\sum_i \sum_j |A_i|_2^2 |B_j|_2^2 X_{i,j}^2|_{l/2} && \text{(Plug in } A_i, B_j \text{ as above)} \\
&\leq \sum_i \sum_j |A_i|_2^2 \cdot |B_j|_2^2 |X_{i,j}^2|_{l/2} && \text{(Triangle inequality for 1/2 norm)} \\
&= \sum_i \sum_j |A_i|_2^2 \cdot |B_j|_2^2 |X_{i,j}|_l^2 && (|X|_{l/2} = |X^2|_l) \\
&\leq (3\epsilon\delta^{\frac{1}{l}})^2 \sum_i \sum_j |A_i|_2^2 |B_j|_2^2 && \text{(JL-moment property (property (*) above))} \\
&= (3\epsilon\delta^{\frac{1}{l}})^2 |A|_F^2 |B|_F^2 && \text{(Definition of Frobenius norm)}
\end{aligned}
$$

Note that $E[|A^TS^TSB - A^TB|_F^l] = ||A^TS^TSB - A^TB|_F^2|_{\frac{l}{2}}^{l/2}$ (by definition of l and $l/2$ norms).

Now, we can apply Markov's inequality (using that $\mathbb{P}\left[|(A^TS^TSB - A^TB|_F)^{1/l} > (3\epsilon|A|_F|B|_F)^{1/l}\right] a = \mathbb{P}[|A^TS^TSB - A^TB]_F > 3\epsilon|A|_F|B|_F])$ to get

$$\mathbb{P}[|A^TS^TSB - A^TB]_F > 3\epsilon|A|_F|B|_F] \leq \left(\frac{1}{3\epsilon|A|_F|B|_F}\right)^l E[|A^TS^TSB - A^TB|_F^l \leq \delta$$

.

■

1.2 Proof that CountSketch satisfies the JL property

So, we have shown that if CountSketch satisfies the JL-moment property, then it is an approximate matrix product. So, now we just need to show that it satisfies the JL-moment property. Luckily, we have the following theorem:

Theorem 2. *The distribution D over CountSketch matrices satisfies the (ε, δ, l)-JL moment property for $l = 2$*

This will just involve doing an elementary second moment argument with no super deep math facts. We'll show it for $l = 2$ since that is the smallest that worked above (because we needed triangle inequality of the $\frac{l}{2}$ norm above so we needed $\frac{l}{2} \geq 1$).

We'll require some basic hashing definitions for this proof

Definition. A hash function $h : [n] \to [m]$ is k-wise independent if $\forall i_1 \neq i_2 \neq \cdots \neq i_k, \forall j_1, j_2, \cdots j_k \in [m]$ we have $\mathbb{P}[h(i_1) = j_1 \wedge h(i_2) = j_2 \wedge \cdots h(i_k) = j_k] = \frac{1}{m^k}$ (ie., the elements independent and uniform over the output)

Also, k-wise independence gives us the following neat (but irrelevant for the sake of our proof) fact:

Fact 1. 2 and 4-wise independent hash function can be stored with $O(\lg n)$ bits.

Proof of theorem 2. Consider $E_S[|Sx|_2^2]$. For a CountSketch matrix S, let $h : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, k\}$ define the location of the non-zero entry on the column, and let $\sigma : \{1, 2, \cdots, n\} \to \{1, -1\}$ give the sign of the non-zero entry in column i (so S is parameterized exactly by h and σ).

We only need h to be a 2-wise independent hash function and $\sigma : [n] \to \{-1, 1\}$ to be 4-wise independent (and h and σ independent of each other). This means that we can store S with only $O(\lg n)$ bits.

Notation: Let $\delta(E)$ be the indicator for event E.

Note that $E[|Sx|_2^2] = \sum_{j\in[k]} E[(\sum_{i\in[n]} \delta(h(i) = j)\sigma_i x_i)^2]$, by applying linearity of expectation and noting that the (i, j) element in the matrix only contributes to the ith entry in the vector when $\delta(h(i) = j) = 1$, in which case it has value $\sigma_i x_i$. Then,

$$\begin{aligned}
&E[|Sx|_2^2]\\
&= \sum_{j\in[k]} E[(\sum_{i\in[n]} \delta(h(i) = j)\sigma_i x_i)^2]\\
&= \sum_{j\in[k]} \sum_{i_1,i_2\in[n]} E[\delta(h(i_1) = j)\delta(h(i_2) = j)\sigma_{i_1}\sigma_{i_2}]x_{i_1}x_{i_2}
\end{aligned}$$

(Linearity of expectation again by expanding the square)

If $i_1 \neq i_2$, then

$$\begin{aligned}
&\sum_{j\in[k]} \sum_{i_1,i_2\in[n]} E[\sigma_{i_1}]E[\sigma_{i_2}]E[\delta(h(i_1) = j)\delta(h(i_2) = j)]x_{i_1}x_{i_2} && \text{(2-wise independence of } \sigma)\\
&= 0
\end{aligned}$$

So, we have

$$\begin{aligned}
&\sum_{j\in[k]} \sum_{i_1,i_2\in[n]} E[\delta(h(i_1) = j)\delta(h(i_2) = j)\sigma_{i_1}\sigma_{i_2}]x_{i_1}x_{i_2}\\
&= \sum_{j\in[k]} \sum_{i\in[n]} E[\delta(h(i) = j)^2]x_i^2 && \text{(Clear terms where } i_1 \text{ and } i_2 \text{ are not equal)}\\
&= \frac{1}{k} \sum_{j\in[k]} \sum_{i\in[n]} x_i^2 && \text{(Square of an indicator is itself an indicator)}\\
&= |x|_2^2
\end{aligned}$$

Now, to prove that S satisfies the JL property for $l = 2$, we also need to calculate $E[|Sx|_2^4]$, because that term appears in the definition when we set $l = 2$.

$$\begin{aligned}
&E[|Sx|_2^4]\\
&= E[\sum_{j\in[k]} \sum_{j'\in[k]} \left(\sum_{i\in[n]} \delta(h(i) = j)\sigma_i x_i \right)^2 (\delta(h(i') = j')\sigma_{i'} x_{i'})^2]\\
&= \sum_{j_1,j_2,i_1,i_2,j_3,i_4} E[\sigma_{i_1}\sigma i 2\sigma_{i_3}\sigma_{i_4}\delta(h(i_1) = j_1)\delta(i_2) = j_1)\delta(h(i_3) = j_2)\delta(h(i_4) = j_2)]x_{i_1}x_{i_2}x_{i_3}
\end{aligned}$$

(i_1, i_2 are from expanding the first squared norm, i_3, i_4 from the second)

By 4-wise independence of σ, the only non-zero terms is if $i_1 = i_2 = i_3 = i_4$ or there are 2 pairs of equal values.

Case 1 If $i_1 = i_2 = i_3 = i_4$ then necessarily $j_1 = j_2$ (since each column has 1 non-zero entry), so it's $\sum_j \frac{1}{k} \sum_i x_i^4 = |x|_4^4$.

Case 2 If $i_1 = i_2$ and $i_3 = i_4$ but $i_1 \neq i_3$, then we can apply 2-wise independence of h and get $\sum_{j_1,j_2,i_1,i_3,i_1 \neq i_3} \frac{1}{k^2} x_{i_1}^2 x_{i_3}^2 = |x|_2^4 - |x|_4^4$ (we subtract $|x|_4^4$ because $i_1 \neq i_3$ so we don't get any terms that look like x_i^4).

Case 3 If $i_1 = i_3$ and $i_2 = i_4$ but $i_1 \neq i_2$, then we need $j_1 = j_2$ for it to not be 0. Then, we get

$$\sum_j \frac{1}{k^2} \sum_{i_1,i_2,i_1 \neq i_2} x_{i_1}^2 x_{i_2}^2 \leq \sum_j \frac{1}{k^2} \sum_{i_1,i_2} x_{i_1}^2 x_{i_2}^2 \frac{1}{k} |x|_2^4$$

Note this case is lower bounded by 0.

Case 4 If $i_1 = i_4$ and $i_2 = i_3$, then it's the same as case 3.

Putting this all together, we get $E[|Sx|_2^4] \in [|x|_2^4, |x|_2^(1 + \frac{2}{k})] = [1, 1 + \frac{2}{k}]$. The only inexactness came from cases 3 and 4 where we needed the upper bound.

So, setting $k = \frac{2}{\varepsilon^2 \delta}$

$$E_S||Sx|_2^2 - 1|^2 = E_S[|Sx|_2^4] - 2E[Sx]_2^2 + 1 \leq (1 + \frac{2}{k}) - 2 + 1 = \frac{2}{k} = \varepsilon^2 \delta$$

which is exactly the JL property for $l = 2$.

■

Recall that we needed $\mathbb{P}[|CS^T SD - CD|_F^2 \leq (6/\delta k) * |C|_F^2 |D|_F^2] \geq -1\delta$, just pattern match and recall $|C|_F^2 = |D|_F^2 = d$ since A is orthonormal, set $C = A^T, D = A$ and we have the desired result.

To get better failure probability bounds, you can look at higher moments.