CS 15-859: Algorithms for Big Data Fall 2020

Lecture lecture 3.1 — 9/24/2020

Prof. David Woodruff *Scribe: Divyansh Jhunjhunwala*

1 Recap

We begin by recalling the JL Property definition and the approximate matrix product result from last lecture.

Definition. (JL Property): A distribution on matrices $S \in \mathbb{R}^{k\times n}$ has the (ϵ, δ, ℓ)-JL moment Property if for all $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$,

$$\mathbb{E}_S\left[\|Sx\|_2^2 - 1\right]^\ell \leq \epsilon^\ell \cdot \delta$$

Approximate matrix product result: For fixed matrices C, D with the appropriate dimensions, $S \in \mathbb{R}^{k\times d}$ a distribution over matrices satisfying JL Property and a fixed δ we have,

$$\mathbf{Pr}\left[\left\|CS^\top SD - CD\right\|_F^2 \leq \frac{6}{\delta \cdot k}\|C\|_F^2\|D\|_F^2\right] \geq 1 - \delta$$

In the last lecture, we had assumed that CountSketch satisfies the JL Property which allowed us to use the approximate matrix product result to show that CountSketch is a valid subspace embedding. In this lecture we will prove the following:

- JL Property implies the approximate matrix product result.
- CountSketch satisfies the JL Property

In order to show the approximate matrix product result we first prove the following theorem.

Theorem: (From vectors to matrices) *For $\epsilon, \delta \in (0, \frac{1}{2})$, let D be a distribution on matrices $S \in \mathbb{R}^{k\times n}$ that satisfies the $(\epsilon, \delta, \ell)-$JL Property for some $\ell \geq 2$. Then for matrices A, B with n rows,*

$$\mathbf{Pr}\left[\left\|A^\top S^\top SB - A^\top B\right\|_F \geq 3\epsilon\|A\|_F\|B\|_F\right] \leq \delta$$

Proof:

We had completed the proof of Minkowski's inequality in the last lecture which allows us to apply the triangle inequality on the $p-$norm of random variables which we had defined. We now show that JL Property implies that S almost preserves inner-product between unit vectors x, y

$$\begin{aligned}
\|\langle Sx, Sy\rangle - \langle x, y\rangle\|_\ell &= \frac{1}{2}\left\|(\|Sx\|_2^2 - 1) + (\|Sy\|_2^2 - 1) - (\|S(x-y)\|_2^2 - \|x-y\|_2^2)\right\|_\ell && (1)\\
&\le \frac{1}{2}\left[\left\|(\|Sx\|_2^2 - 1)\right\|_\ell + \left\|(\|Sy\|_2^2 - 1)\right\|_\ell + \left\|(\|S(x-y)\|_2^2 - \|x-y\|_2^2)\right\|_\ell\right] && (2)\\
&\le \frac{1}{2}\left(\epsilon \cdot \delta^{1/l} + \epsilon \cdot \delta^{1/l} + \|x-y\|_2^2 \epsilon \cdot \delta^{1/l}\right) && (3)\\
&\le 3\epsilon \cdot \delta^{1/l} && (4)
\end{aligned}$$

where (2) follows from the triangle inequality, (3) follows from the JL Property definition and (4) follows from bounding the squared norm of the difference of two unit vectors.

For arbitrary pair of vectors x, y we can now normalize by the product of their norms so that (4) still holds. This gives us,

$$\|\langle Sx, Sy\rangle - \langle x, y\rangle\|_\ell \le 3\epsilon \cdot \delta^{1/l}\|x\|_2\|y\|_2 \tag{5}$$

We now attempt to use our last inequality to bound the product of sketched matrices. Let A be any matrix with columns $A_1, \cdots, A_d$ and B be any matrix with columns $B_1, \cdots B_e$. We observe that the (i,j)-th entry of the matrix $(SA)^\top(SB) - A^\top B$ will be $\langle SA_i, SB_j\rangle - \langle A_i, B_j\rangle$

We define $X_{i,j} = \frac{1}{\|A_i\|_2\|B_j\|_2}\left(\langle SA_i, SB_j\rangle - \langle A_i, B_j\rangle\right)$

Using our earlier observation we can now write,

$$\|(SA)^\top(SB) - A^\top B\|_F^2 = \sum_i \sum_j \|A_i\|_2^2\|B_j\|_2^2 X_{i,j}^2$$

Using the previous equation we can now write for any $\ell \ge 2$

$$\begin{aligned}
\left\|\|(SA)^\top(SB) - A^\top B\|_F^2\right\|_{\ell/2} &= \left\|\sum_i \sum_j \|A_i\|_2^2\|B_j\|_2^2 X_{i,j}^2\right\|_{\ell/2} && (6)\\
&\le \sum_i \sum_j \left\|\|A_i\|_2^2\|B_j\|_2^2 X_{i,j}^2\right\|_{\ell/2} \qquad \text{(using triangle inequality)} && (7)\\
&= \sum_i \sum_j \|A_i\|_2^2\|B_j\|_2^2 \|X_{i,j}^2\|_{\ell/2} && (8)\\
&= \sum_i \sum_j \|A_i\|_2^2\|B_j\|_2^2 \, \|X_{i,j}\|_\ell^2 && (9)\\
&\le (3\epsilon \cdot \delta^{1/l})^2 \sum_i \sum_j \|A_i\|_2^2\|B_j\|_2^2 \qquad \text{(from (5))} && (10)\\
&= (3\epsilon \cdot \delta^{1/l})^2 \|A\|_F^2\|B\|_F^2 && (11)
\end{aligned}$$

Since, $\mathbb{E}\left[\left\|A^\top S^\top SB - A^\top B\right\|_F^\ell\right] = \left\|\left\|A^\top S^\top SB - A^\top B\right\|_F^\ell\right\|_{\ell/2}^{\ell/2}$, we can write,

$$\mathbf{Pr}\left[\left\|A^\top S^\top SB - A^\top B\right\|_F \geq 3\epsilon\|A\|_F\|B\|_F\right] = \mathbf{Pr}\left[\left\|A^\top S^\top SB - A^\top B\right\|_F^\ell \geq (3\epsilon\|A\|_F\|B\|_F)^\ell\right] \quad (12)$$

$$\leq \left(\frac{1}{3\epsilon\|A\|_F\|B\|_F}\right)^\ell \mathbb{E}\left[\left\|A^\top S^\top SB - A^\top B\right\|_F^\ell\right] \quad (13)$$

$$\leq \left(\frac{1}{3\epsilon\|A\|_F\|B\|_F}\right)^\ell \left((3\epsilon \cdot \delta^{1/l})^2\|A\|_F^2\|B\|_F^2\right)^{\ell/2} \quad (14)$$

$$\leq \delta \quad (15)$$

where (13) follows from applying Markov's inequality and (14) follows from (11). This completes the proof of our theorem. In the next section we will see a bound on the number of rows k in S in terms of ϵ^2 and δ which allows us to recover the approximate matrix product result.

2 CountSketch satisfies the JL Property

We note that for $\ell = 1$ the JL Property checks whether the matrix S preserves squared norm. We try to prove the JL Property for CountSketch for $\ell = 2$.

Proof:

We see that a CountSketch matrix $S \in \mathbb{R}^{k\times n}$ is designed in a way that can be described by two hash functions defined as follows,

- $h : [n] \to [k]$ is a 2-wise independent hash function. $h(i)$ is the row index of the non-zero entry in the $i-$th column of S.
- $\sigma : [n] \to \{-1, 1\}$ is a 4-wise independent hash function. $\sigma(i)$ is the sign of the non-zero entry in the $i-$th column of S.

We define the notation $\delta(E) = 1$ if events E holds and $\delta(E) = 0$ otherwise. Please also note that we use the shorthand notation $\sigma(i) = \sigma_i$ in our proofs for clarity. We first show that CountSketch preserves the norm of a vector in expectation.

$$\mathbb{E}\left[\|Sx\|_2^2\right] = \sum_{j\in[k]} \mathbb{E}\left[\left(\sum_{i\in[n]} \delta(h(i) = j)\sigma_i x_i\right)^2\right] \quad (16)$$

$$= \sum_{j\in[k]} \left(\sum_{i_1,i_2\in[n]} \mathbb{E}\left[\delta(h(i_1) = j)\delta(h(i_2) = j)\sigma_{i_1}\sigma_{i_2}\right] x_{i_1}x_{i_2}\right) \quad (17)$$

$$= \sum_{j\in[k]} \left(\sum_{i\in[n]} \mathbb{E}\left[\delta(h(i) = j)^2\right] x_i^2\right) \quad (18)$$

$$= \sum_{j\in[k]} \frac{1}{k} \sum_{i\in[n]} x_i^2 \quad (19)$$

$$= \|x\|_2^2 \quad (20)$$

(17) follows from expanding the square of sums in (16) and linearity of expectation. (18) follows from (17) using the following observations,

- h and σ are independent of each other
- When $\sigma_{i_1} \neq \sigma_{i_2}, \mathbb{E}[\sigma_{i_1}\sigma_{i_2}] = \mathbb{E}[\sigma_{i_1}]\mathbb{E}[\sigma_{i_2}]$
- $\mathbb{E}[\sigma_i] = 0$, $\sigma_i^2 = 1$ always.

Using these observations we see that all cross terms in (17) become zero leaving us only with the expression in (18).

Thus we see that CountSketch preserves norm in expectation.

For $\ell = 2$ we have,

$$\mathbb{E}\left[\|Sx\|_2^2 - 1\right]^2 = \mathbb{E}[\|Sx\|_2^4] - 2\mathbb{E}[\|Sx\|_2^2] + 1 \tag{21}$$

We now attempt to bound $\mathbb{E}[\|Sx\|_2^4]$. We see that,

$$\begin{aligned}
\mathbb{E}[\|Sx\|_2^4] &= \mathbb{E}[(\|Sx\|_2^2)^2] \\
&= \mathbb{E}[\sum_{j\in[k]}\sum_{j'\in[k]}(\sum_{i\in[n]}\delta(h(i)=j)\sigma_i x_i)^2(\sum_{i'\in[n]}\delta(h(i^{'})=j^{'})\sigma_{i'}x_{i'})^2] \\
&= \sum_{j_1,j_2,i_1,i_2,i_3,i_4}\mathbb{E}[\delta(h(i_1)=j_1)\delta(h(i_2)=j_2)\delta(h(i_3)=j_3)\delta(h(i_4)=j_4)\sigma_i\sigma_2\sigma_3\sigma_4]x_{i_1}x_{i_2}x_{i_3}x_{i_4}
\end{aligned}$$

First of all, we make the observation that if any of the indices in $\{i_1, i_2, i_3, i_4\}$ is distinct then the expectation of that term goes to zero by the 4-wise independence of σ.

This observation allows us to break our analysis down to different cases

Case 1:

$i_1 = i_2 = i_3 = i_4$. Since there is only one nonzero in each column, this implies $j_1 = j_2$ for nonzero probability terms. This gives us,

$$\begin{aligned}
&\sum_{j\in[k]}\sum_{i\in[n]}\mathbb{E}[(\delta(h(i)=j)^4]x_i^4 \\
&= \sum_{j\in[k]}\sum_{i\in[n]}\mathbf{Pr}[h(i)=j]x_i^4 \\
&= \sum_{j\in[k]}\frac{1}{k}\sum_{i\in[n]}x_i^4 \\
&= \|x\|_4^4
\end{aligned}$$

Case 2:

$i_1 = i_2, i_3 = i_4, i_1 \neq i_3$. We get,

$$\begin{aligned}
&\sum_{j_1,j_2\in[k]} \sum_{i_1,i_3\in[n],i_1\neq i_3} \mathbb{E}[(\delta(h(i_1)=j_1)^2\delta(h(i_3)=j_2)^2]x_{i_1}^2x_{i_3}^2 \\
&= \sum_{j_1,j_2\in[k]} \sum_{i_1,i_3\in[n],i_1\neq i_3} \mathbf{Pr}[(h(i_1)=j_1)\wedge(h(i_3)=j_2)]x_{i_1}^2x_{i_3}^2 \\
&= \sum_{j_1,j_2\in[k]} \frac{1}{k^2} \sum_{i_1,i_3\in[n],i_1\neq i_3} x_{i_1}^2x_{i_3}^2. && \text{(from 2-wise independence of } h \text{)} \\
&= \sum_{i_1,i_3\in[n]} x_{i_1}^2x_{i_3}^2 - \sum_{i\in[n]} x_i^4 \\
&= \|x\|_2^4 - \|x\|_4^4
\end{aligned}$$

Case 3:

$i_1 = i_3, i_2 = i_4, i_1 \neq i_2$. This again necessitates $j_1 = j_2$ by the same logic in Case 1, which gives us,

$$\begin{aligned}
&\sum_{j} \sum_{i_1,i_2\in[n],i_1\neq i_2} \mathbb{E}[(\delta(h(i_1)=j)^2\delta(h(i_2)=j)^2]x_{i_1}^2x_{i_2}^2 \\
&= \sum_{j} \sum_{i_1,i_2\in[n],i_1\neq i_2} \mathbf{Pr}[(h(i_1)=j)\wedge(h(i_2)=j)]x_{i_1}^2x_{i_2}^2 \\
&= \sum_{j} \frac{1}{k^2} \sum_{i_1,i_2\in[n],i_1\neq i_2} x_{i_1}^2x_{i_2}^2 && \text{(from 2-wise independence of } h \text{)} \\
&= \frac{1}{k} \sum_{i_1,i_2\in[n],i_1\neq i_2} x_{i_1}^2x_{i_2}^2 \\
&\leq \frac{1}{k} \sum_{i_1,i_2\in[n]} x_{i_1}^2x_{i_2}^2 \\
&= \frac{1}{k}\|x\|_2^4
\end{aligned}$$

Case 4:

$i_1 = i_4, i_2 = i_3, i_1 \neq i_2$. We see that this is equivalent to Case 3 and adds a further $\frac{1}{k}\|x\|_2^4$ towards the total sum.

We can now bound (21) for a unit vector x as follows,

$$\begin{aligned}
\mathbb{E}\left[\|Sx\|_2^2 - 1\right]^2 &= \mathbb{E}[\|Sx\|_2^4] - 2\mathbb{E}[\|Sx\|_2^2] + 1 \\
&\leq \left(1 + \frac{2}{k}\right) - 2 + 1 && (\mathbb{E}[\|Sx\|_2^4] \leq \left(1+\frac{2}{k}\right), \mathbb{E}[\|Sx\|_2^2] = 1) \\
&= \frac{2}{k}
\end{aligned}$$

In order for JL Property to hold we set $k = \frac{2}{\epsilon^2 \delta}$. This completes the proof that JL Property holds for CountSketch with $\ell = 2$. Substituting $\epsilon^2 = \frac{2}{k\delta}$ also gives us our approximate matrix product result.