CS 15-859: Algorithms for Big Data Fall 2017

Lecture 3a — September 21

Prof. David Woodruff *Scribe: Ellango Jothimurugesan*

1 CountSketch Satisfies the JL Property

Recall that our CountSketch matrix is a sparse $k \times n$ matrix, in which each column contains a single randomly chosen nonzero entry which is ± 1. Every CountSketch matrix can be described by two functions:

1. $h : [n] \to [k]$ is a 2-wise independent hash function. For the ith column, $h(i)$ is the row with a nonzero.
2. $\sigma : [n] \to \{-1, +1\}$ is a 4-wise independent hash function. For the ith column, $\sigma(i)$ is the sign of the entry in that column.

In the previous lecture, we were in the middle of showing that the CountSketch matrix is a subspace embedding. We already saw the approximate matrix product result, under the assumption that CountSketch satisfied the (ε, δ, l)-JL moment property for some $l > 2$. It only remains to show that the CountSketch matrix satisfies this property for $l = 2$ to finish the proof.

Proposition 1. The distribution on CountSketch matrices $S \in \mathbb{R}^{k \times n}$ has the JL property with $l = 2$. That is, for all $x \in \mathbb{R}^n$ with $|x|_2 = 1$,

$$\mathbb{E}_S\left[\left||Sx|_2^2 - 1\right|^2\right] \le \varepsilon^2 \delta.$$

Proof. Let us first compute the term $\mathbb{E}[|Sx|_2^2]$. We will use the notation that $\delta(E) = 1$ if the event E holds and $\delta(E) = 0$ otherwise.

$$
\begin{aligned}
\mathbb{E}[|Sx|_2^2] &= \mathbb{E}\left[\sum_{j=1}^{k}(S_{j*}\cdot x)^2\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{k}\left(\sum_{i=1}^{n}\delta(h(i)=j)\sigma_i x_i\right)^2\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{k}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\delta(h(i_1)=j)\delta(h(i_2)=j)\sigma_{i_1}\sigma_{i_2}x_{i_1}x_{i_2}\right] \\
&= \sum_{j=1}^{k}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\mathbb{E}[\delta(h(i_1)=j)\delta(h(i_2)=j)\sigma_{i_1}\sigma_{i_2}]x_{i_1}x_{i_2} \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n}\mathbb{E}[\delta(h(i)=j)^2]x_i^2 \\
&= \sum_{j=1}^{k}\frac{1}{k}\sum_{i=1}^{n}x_i^2 \\
&= |x|_2^2.
\end{aligned}
$$

In the 5th line, we used pairwise independence of σ, which implies all terms where i_1, i_2 are distinct vanish since $\mathbb{E}[\sigma_i] = 0$, and that $\sigma_i^2 = 1$.

Next let us compute $\mathbb{E}[|Sx|_2^4]$. Following from the 3rd line of the previous calculation,

$$
\begin{aligned}
\mathbb{E}[|Sx|_2^4] &= \mathbb{E}[(|Sx|_2^2)^2] \\
&= \sum_{j_1,j_2}^{k}\sum_{i_1,i_2,i_3,i_4}^{n}\mathbb{E}[\delta(h(i_1)=j_1)\delta(h(i_2)=j_1)\delta(h(i_3)=j_2)\delta(h(i_4)=j_2)\sigma_{i_1}\sigma_{i_2}\sigma_{i_3}\sigma_{i_4}]x_{i_1}x_{i_2}x_{i_3}x_{i_4}
\end{aligned}
$$

By 4-wise independence of σ, if any of the indices $\{i_1, i_2, i_3, i_4\}$ is distinct, that term in the sum vanishes. There are 4 cases:

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 terms. The contribution to the sum is
$$\sum_{j=1}^{k}\sum_{i=1}^{n}\mathbf{Pr}[h(i)=j]x_i^4 = \sum_{j=1}^{k}\frac{1}{k}\sum_{i=1}^{n}x_i^4 = |x|_4^4.$$

2. $i_1 = i_2, i_3 = i_4, i_1 \neq i_3$: Using pairwise independence of h, the contribution to the sum is
$$\sum_{j_1,j_2}^{k}\sum_{i_1\neq i_3}^{n}\mathbf{Pr}[h(i_1)=j_1\wedge h(i_3)=j_2]x_{i_1}^2x_{i_3}^2 = \sum_{j_1,j_2}^{k}\frac{1}{k^2}\sum_{i_1\neq i_3}^{n}x_{i_1}^2x_{i_3}^2 = \sum_{i_1,i_3}^{n}x_{i_1}^2x_{i_3}^2 - \sum_{i_1=i_3}^{n}x_{i_1}^4 = |x|_2^4 - |x|_4^4.$$

3. $i_1 = i_3, i_2 = i_4, i_1 \neq i_2$: Necessarily, $j_1 = j_2$, and by pairwise independence of h, the contribution to the sum is at most
$$\sum_{j}^{k}\sum_{i_1\neq i_2}^{n}\mathbf{Pr}[h(i_1)=j\wedge h(i_2)=j]x_{i_1}^2x_{i_2}^2 = \frac{1}{k}\sum_{i_1\neq i_2}^{n}x_{i_1}^2x_{i_2}^2 \leq \frac{1}{k}\sum_{i_1,i_2}^{n}x_{i_1}^2x_{i_2}^2 = \frac{1}{k}|x|_2^4.$$

4. $i_1 = i_4, i_2 = i_3, i_1 \neq i_2$: Identical to case 3, contributing at most $\frac{1}{k}|x|_2^4$.

In total, we have $\mathbb{E}[|Sx|_2^4] \leq (1 + \frac{2}{k})|x|_2^4 = 1 + \frac{2}{k}$.

Finally, we can bound the quantity

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

if we set $k \geq \frac{2}{\varepsilon^2\delta}$.

■

2 Affine Embeddings

Consider the problem

$$\min_X |AX - B|_F^2 \tag{1}$$

where $A : n \times d$, $B : n \times m$, where d is small but m may be large. We can solve the problem by solving the linear regression problem for each column of B. If we want to use sketching however, we cannot directly apply our subspace embeddings; previously we used a matrix S that preserved the column space of A joined with a column vector b, which was at most a $d+1$ dimensional subspace, but here B has many columns.

Let us try to show the desired bound $|SAX - SB|_F = (1 \pm \varepsilon)|AX - B|_F$, and see what properties of S we need for the proof to go through. As usual, we can assume wlog that A has orthonormal columns. Let $B^* = AX^* - B$, where X^* is the optimum in (1). By the normal equations, each column of B^* is orthogonal to the column space of A, so

$$A^T B^* = 0 \tag{2}$$

$$|AX - B|_F^2 = |A(X - X^*)|_F^2 + |B^*|_F^2. \tag{3}$$

Let us show that $|S(AX - B)|_F^2 - |SB^*|_F^2 \in |AX - B|_F^2 - |B^*|_F^2 \pm 2\varepsilon|AX - B|_F^2$:

$$\begin{aligned}
&|S(AX - B)|_F^2 - |SB^*|_F^2 \\
&\quad = |SA(X - X^*) + S(AX^* - B)|_F^2 - |SB^*|_F^2 \\
&\quad = |SA(X - X^*)|_F^2 + 2\,\mathrm{Tr}((X - X^*)^T A^T S^T SB^*) && (|C + D|_F^2 = |C|_F^2 + |D|_F^2 + 2\,\mathrm{Tr}(C^T D)) \\
&\quad \in |SA(X - X^*)|_F^2 \pm 2|X - X^*|_F |A^T S^T SB^*|_F && (|\mathrm{Tr}(CD)| \leq |C|_F |D|_F) \\
&\quad \in |SA(X - X^*)|_F^2 \pm 2\varepsilon|X - X^*|_F |B^*|_F && (\text{if approx matrix product and (2)}) \\
&\quad \in |A(X - X^*)|_F^2 \pm \varepsilon(|A(X - X^*)|_F^2 + 2|X - X^*|_F |B^*|_F) && (\text{if subspace embedding for } A) \\
&\quad \in |A(X - X^*)|_F^2 \pm \varepsilon(|A(X - X^*)|_F + |B^*|_F)^2 && (a^2 + 2ab \leq (a + b)^2) \\
&\quad \in |A(X - X^*)|_F^2 \pm 2\varepsilon(|A(X - X^*)|_F^2 + |B^*|_F^2) && (\text{AM-GM inequality}) \\
&\quad \in |AX - B|_F^2 - |B^*|_F^2 \pm 2\varepsilon|A(X - B)|_F^2 && (\text{using (3) twice})
\end{aligned}$$

Let us say that S preserves the Frobenius norm of a fixed matrix A if $|SA|_F^2 = (1 \pm \varepsilon)|A|_F^2$ with some constant probability. A fact from the homework is that the CountSketch matrix preserves the Frobenius norm of any fixed A. If S preserves the Frobenius norm of B^*, then

$$\begin{aligned}|S(AX - B)|_F^2 &= |AX - B|_F^2 - |B^*|_F^2 + |SB^*|_F^2 \pm 2\varepsilon|AX - B|_F^2 \\ &= (1 + 2\varepsilon)|AX - B|_F^2 + \varepsilon|B^*|_F^2 \\ &= (1 + 3\varepsilon)|AX - B|_F^2\end{aligned}$$

which is what we wanted to show.

In summary, S is an affine embedding if the following three properties hold:

1. S is a subspace embedding for columns of A.
2. S has the approximate matrix product result.
3. S preserves the Frobenius norm.