Lecture 8 Part 2 — Oct 29th, 2020

Prof. David Woodruff *Scribe: Michael Zhang*

1 Estimation of the p-norm of a Vector

Previously, we've discussed that when estimating the p-norm of a vector y where $p > 2$, we apply two matrices P and D to y, where D is a $n \times n$ diagonal matrix with the property that

$$\|Dy\|_\infty \in [\frac{\|y\|_p}{10^{1/p}}, 10^{1/p}\|y\|_p] \tag{1}$$

with property at least 4/5. Also recall that P is a countsketch matrix. In this section, our goal is to show that with high probability

$$\|PDy\|_\infty \approx \|Dy\|_\infty \tag{2}$$

1.1 Some notations and Ideas Towards Understanding $\|PDy\|_\infty$

First, let's establish some notations we will use to analyze the bound for $\|PDy\|_\infty$.

- Let s be the number of rows of P. We can also think of s as the number of hash buckets for elements in Dy
- Let $h : [n] \to [s]$ and $\sigma : [n] \to \{-1, 1\}$ be the two hash functions that describes P. For our analysis, we assume they are truly random, although they could be de-randomnized in practice.

To achieve our desired property of $\|PDy\|_\infty \approx \|Dy\|_\infty$, we want

- in each bucket not containing the coordinate for which $\|(Dy)_j\| = \|Dy\|_\infty$, we have $\|(PDy)_i\| \leq \frac{\|y\|_p}{100}$
- in the bucket i containing the coordinate j for which $\|(Dy)_j\| = \|Dy\|_\infty$, we have $\|(PDy)_i - (Dy)_\infty\| \leq \frac{\|y\|_p}{100}$

Observe that if the two properties above hold, we know that the bucket that contains $\|(Dy)_j\| = \|Dy\|_\infty$ is close to the estimate of $\|(Dy)_j\|$ and all other buckets are not too large. This will indeed guarantee a good estimate of $\|PDy\|_\infty$.

1.2 Analyzing $\|PDy\|_\infty$

Now we begin by calculating the expectation and variance of $(PDy)_i$. Let $\delta(E) = 1$ if event E occurs and $\delta(E) = 0$ otherwise.

Claim 1. $\mathbb{E}[(PDy)_i] = 0$.

Proof: This is because $(PDy)_i = \sum_j \delta(h(j) = i)\sigma_j (Dy)_j$ and $\mathbb{E}[\sigma_j] = 0$ for all j.

Claim 2. $\mathbb{E}[(PDy)_i^2] = O(\frac{1}{s})\|y\|_2^2$

Proof:

$$E_p[(PDy)_i^2] = \sum_{j,j'} \mathbb{E}[\delta(h(j) = i)\delta(h(j') = i)\sigma_j \sigma_{j'} (Dy)_j (Dy)_{j'}] \tag{3}$$
$$= \sum_{j,j'} \mathbb{E}[\delta(h(j) = i)\delta(h(j') = i)\sigma_j \sigma_{j'}](Dy)_j (Dy)_{j'} \tag{4}$$
$$= \sum_{j} \mathbb{E}[\delta(h(j) = i)](Dy)_j^2 \tag{5}$$
$$= (\frac{1}{s})\|Dy\|_2^2 \tag{6}$$

Note that from (4) to (5), we case on whether $j = j'$. If $j \neq j'$, since the two indices are independent, the expectation is 0. Therefore, we only need to consider the case where $j = j'$, and thus we get (5).

Then we have

$$E_D[\|Dy\|_2^2] = \sum_i y_i^2 \mathbb{E}[D_{i,i}^2] \tag{7}$$

Now we bound $\mathbb{E}[D_{i,i}^2]$. Recall that $D_{i,i} = \frac{1}{E_i^{1/p}}$ and $E_i^{2/p} = \int_{t\geq 0}^{\infty} e^{-t} t^{-2/p} \mathrm{dt}$, we get

$$\mathbb{E}[D_{i,i}^2] = \int_{t\geq 0} t^{-2/p} e^{-t} \mathbf{dt} \tag{8}$$
$$= \int_{0\leq t\leq 1} t^{-2/p} e^{-t} \mathbf{dt} + \int_{t>1} t^{-2/p} e^{-t} \mathbf{dt} \tag{9}$$
$$\leq \int_{0\leq t\leq 1} t^{-2/p} \mathbf{dt} + \int_{t>1} e^{-t} \mathbf{dt} \tag{10}$$
$$= \frac{1}{1-\frac{2}{p}} t^{1-2/p}|_0^1 - e^{-t}|_1^{\infty} \tag{11}$$
$$= O(1) \tag{12}$$

Therefore, combining (6), (7), and (12), we get

$$\mathbb{E}[(PDy)_i^2] = O(\frac{1}{s})\|y\|_2^2 \tag{13}$$

Claim 3. $O(\frac{1}{s})\|y\|_2^2 \leq O(\frac{1}{s})(n^{1-2/p}\|y\|_p^2$

Proof: First, recall Holder's inequality: whenever $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\langle x, y\rangle \leq \|x\|_p \|y\|_q \tag{14}$$

Using this inequality, we get

$$\|y\|_2^2 = \sum y_i^2 \cdot 1 \tag{15}$$

$$\leq (\sum (y_i^2)^{p/2})^{2/p} \cdot (\sum 1^q)^{1/q} \tag{16}$$

$$= (\sum y_i^p)^{2/p} \cdot n^{1-2/p} \tag{17}$$

$$= \|y\|_p^2 n^{1-2/p} \tag{18}$$

Therefore, plugging in claim 3, we get

$$\mathbb{E}[(PDy)_i^2] = O(\frac{1}{s})(n^{1-2/p}\|y\|_p^2 \tag{19}$$

Now that we have a bound for the variance of $(PDy)_i^2$, we introduce Bernstein's bound which will give us a bound on $(PDy)_i^2$.

Bernstein's Bound: Suppose $R_1, R_2, \ldots, R_n$ are independent, and for all j, $\|R_j\| \leq K$, and $\mathbf{Var}[\sum_j R_j] = \sigma^2$, there are constants C, c so that for all $t > 0$:

$$\mathbf{Pr}[\|\sum_j R_j - \mathbb{E}[\sum_j R_j]\| > t] \leq C(e^{\frac{-ct^2}{\sigma^2}} + e^{\frac{-ct}{K}}) \tag{20}$$

While it's tempting to directly apply this bound to $(PDy)_i$ where each R_j is $\delta(h(j) = i)\sigma_j(Dy)_j$, we must first find a good value for K. If we simply define K as the maximum entry in Dy, Bernstein's bound will only give us a constant probability that each $(PDy)_i$ falls within our desired bound, which is not good enough since we cannot get a meaningful union bound over all entries in PDy. And indeed, simply bounding K as the maximum entry defeats our purpose which we discussed in the previous section - if the bucket contains a large element in Dy, we want the absolute value of that bucket to be large instead of being closer to 0. Therefore, we will proceed with a case analysis and only apply Bernstein's bound to the "small" entries in PDy.

1.3 Understanding the Large Elements

As previously suggested, let $R_j = \delta(h(j) = i)\sigma_j(Dy)_j$. We will separately handle those R_j's where $\|R_j\| > \frac{\alpha\|y\|_p}{\log n}$ for a sufficiently small constant $\alpha > 0$. Observe that if $\|R_j\| > \frac{\alpha\|y\|_p}{\log n}$, then necessarily $\|(Dy)_j\| \geq \frac{\alpha\|y\|_p}{\log n}$. We call such a j *large* if $\|(Dy)_j\| \geq \frac{\alpha\|y\|_p}{\log n}$, otherwise we say j is *small*.

Claim 4. The expected number of *large* j is $O(\log^p n)$ with constant probability.

Proof: Let $Z_j = 1$ if j is *large* and $Z_j = 0$ otherwise. We have $\mathbb{E}[\sum_j Z_j] = \sum_j \mathbf{Pr}[\|(Dy)_j\| \text{is large}]$.

$$\mathbf{Pr}[\|(Dy)_j\| \text{is large}] = \mathbf{Pr}[\frac{\|y_j\|}{E_j^{1/p}} \geq \frac{\alpha\|y\|_p}{\log n}] \tag{21}$$

$$= \mathbf{Pr}[\frac{\|y_j\|^p \log^p n}{\alpha^p \|y\|_p^p} \geq E_j] \tag{22}$$

$$= 1 - e^{-\frac{\|y_j\|^p \log^p n}{\alpha^p \|y\|_p^p}} \tag{23}$$

$$\leq \frac{\|y_j\|^p \log^p n}{\alpha^p \|y\|_p^p} \tag{24}$$

Note that from (23) to (24) we applied the inequality $e^{-x} \geq 1 - x$ for all x.

Therefore, $\mathbb{E}[\sum_j Z_j] = \frac{\log^p n}{\alpha^p} = O(\log^p n)$. By Markov's inequality, with constant probability, the number of *large* j's in Dy is in $O(\log^p n)$.

Now, conditioning on $\|Dy\|_\infty \in [\frac{\|y\|_p}{10^{1/p}}, 10^{1/p}\|y\|_p]$ and that the number of *large* j is in $O(\log^p n)$, we know that with high probability, there is no collision for the large elements in h (i.e. any two large elements are not hashed into the same bucket by h). We can easily verify this since $s = \theta(n^{1-2/p} \log n)$ is way larger than $\log^p n$. Now with large elements dealt with separately, we are ready to apply Bernstein's bound for small elements and wrap up our analysis.

1.4 Understanding Small Elements

We condition on

1. $\|Dy\|_\infty \in [\frac{\|y\|_p}{10^{1/p}}, 10^{1/p}\|y\|_p]$
2. The number of large elements is in $O(\log^p n)$.
3. All large elements are perfectly hashed in h.

Then we can set K to be $\frac{\alpha\|y\|_p}{\log n}$. We set $t = \frac{\|y\|_p}{100}$ and $s = \theta(n^{1-p/2} \log n)$. By Bernstein's bound,

$$\mathbf{Pr}[|\sum_{\text{small j}} \delta(h(j) = i)\sigma_j(Dy)_j| > \frac{\|y\|_p}{100}] \leq C(e^{-\theta(\log n)} + e^{-c\frac{\log n}{100\alpha}} \leq \frac{1}{n^2} \tag{25}$$

Therefore, the "signed sum" of all small elements in each bucket is at most $\frac{\|y\|_p}{100}$ in each bucket.

1.5 Wrapping up

Following our analysis of $(PDy)_i$, we know that for all i,

- $\|(PDy)_i\| \leq \frac{\|y\|_p}{100}$ if there is no large element in the i-th bucket
- $\|(PDy)_i\| = \|Dy\|_j \pm \frac{\|y\|_p}{100}$ if there is a large element $(Dy)_j$ in the i-th bucket.
- No bucket contains more than 1 large element

Since we also conditioned on $\|Dy\|_\infty \in [\frac{\|y\|_p}{10^{1/p}}, 10^{1/p}\|y\|_p]$, we conclude that

$$\|PDy\|_\infty \leq 10^{1/p}\|y\|_p + \frac{\|y\|_p}{100} \tag{26}$$

$$\|PDy\|_\infty \geq \frac{\|y\|_p}{10^{1/p}} - \frac{\|y\|_p}{100} \tag{27}$$

Therefore, $\|PDy\|_\infty$ is a good estimation of $\|y\|_p$, and we output $\|PDy\|_\infty$ as the result of this algorithm. The overall space complexity is therefore $O(n^{1-2/p} \log n)$ words and $O(n^{1-2/p} \log^2 n)$ bits.