CS 395T: Sublinear Algorithms Fall 2014

Lecture 5 — Sept. 11, 2014

Prof. Eric Price *Scribe: Alex Knaust*

In today's lecture, we will cover the following topics:

1. Complete our analysis of count-sketch point query [MP2014]
2. An algorithm with $O(k \log^2 n)$ recovery *time* due to [GLPS2012]

1 Count-sketch Analysis Continued

Recall we had the following definitions

$$h_u : [n] \to [B] \quad \text{A family of independent hash functions}$$

$$s_u : [n] \to +1, -1 \quad \text{A family of random sign functions}$$

A hashtable Y with R rows and B columns, being used as follows:

$$Y_{u,v} = \sum_{i:h(i)=v} s_u(i) x_i$$

$$\hat{x_i}^{(u)} = s_u(i) Y_{u,h(i)}$$

$$\hat{x_i} = \underset{u \in [R]}{\text{median}}\, \hat{x_i}^{(u)}$$

$$\hat{x_i} = s_u(i) Y_{u,h(i)}$$

We would like to estimate the error

$$\Delta_i^u = \hat{x_i}^{(u)} - x_i$$

Which we can rewrite as

$$\Delta_i^u = \sum_{j \neq i} \underbrace{s_u(j) x_j \underset{h_u(i)=h_u(j)}{\mathbb{I}}}_{z_{u,j}}$$

Splitting into the largest coordinates $H = (1...k)$ and the rest $T = (k+1...n)$

$$\Delta_i^u = \underbrace{\sum_{H \setminus \{i\}} z_{u,j}}_{=0 \text{ with prob. } .9} + \sum_{T \setminus \{i\}} z_{u,j}$$

and by the same argument from lecture 4 (cross terms cancelling due to s_u being independent)

$$\mathbb{E}\left[\left(\sum_{T \setminus \{i\}} z_{u,j}\right)^2\right] \leq \frac{||x_T||_2^2}{B}$$

$$\Rightarrow |\Delta_i^u| \leq \frac{||x_T||_2^2}{k} \quad \text{with } \frac{4}{5} \text{ prob.}$$

$$\Rightarrow |\Delta_i| = |x - \hat{x_i}| \leq \frac{||x_T||_2^2}{k} \quad \text{with } 1 - e^{-\Omega(R)} \text{ prob.}$$

1.1 Using the Fourier Transform

Recall for a symmetric random variable we defined the Fourier transform as

$$\mathop{\mathcal{F}}_{X}(t) = \mathop{\mathbb{E}}_{x \sim X}[\cos(2\pi x t)]$$

For $z_{u,i}$ we then have

$$z_{u,i} = s_u(j) x_j \mathop{\mathbb{I}}_{h_u(i) = h_u(j)}$$

Which is 0 (no collision) with prob. $1 - \frac{1}{B}$, and $\pm x_i$ with prob. $\frac{1}{2B}$

$$\mathbb{E}[\cos(2\pi z_{u,i} t)] = \left(1 - \frac{1}{B}\right) \cos 0 + \frac{1}{B} \cos(2\pi t x_i)$$

$$\geq \left(1 - \frac{2}{B}\right) \geq 0$$

Furthermore, since adding PDFs is equivalent to convolving them, we can write

$$\mathop{\mathcal{F}}_{\sum_{T \setminus \{i\}} z_{u,j}}(t) = \prod_{j \in T \setminus \{i\}} \mathop{\mathcal{F}}_{z_{u,j}}(t) \geq 0$$

Since the sum has a non-negative fourier transform, we can apply our previous lemma (Lemma 3.1 in [MP2014])

$$\Rightarrow \mathbb{P}\left[\left|\sum_{T \setminus \{i\}} z_{u,j}\right| \leq \epsilon \frac{||x_T||_2}{\sqrt{B}}\right] \gtrsim \epsilon \tag{1}$$

The sets H and T are independent, thus

$$\mathbb{P}\left[|\Delta_i^u| \leq \epsilon \frac{||x_T||_2}{\sqrt{B}}\right] \geq \underbrace{\mathbb{P}\left[\sum_{H \setminus \{i\}} z_{u,j} = 0\right]}_{.9 \text{ with prev.}} \cdot \underbrace{\mathbb{P}\left[\left|\sum_{T \setminus \{i\}} z_{u,j}\right| \leq \epsilon \frac{||x_T||_2}{\sqrt{B}}\right]}_{\Omega(\epsilon) \text{ due to Equation 1}} \gtrsim \epsilon \tag{2}$$

Question: So what happens to the median of the errors, $\Delta_i = \hat{x}_i - x_i = \text{median}_u \Delta_i^u$?

Lemma 1.1. *(Lemma 3.3 from [MP2014]) Let Δ_i^u for $u \in [R]$ be symmetric independent random variables. And let equation 2 apply, then*

$$\mathbb{P}\left[\left|\mathop{\text{median}}_{u \in [R]} \Delta_i^u\right| \geq \epsilon \frac{||x_T||_2}{\sqrt{B}}\right] < 2e^{-\Omega(R\epsilon^2)} \tag{3}$$

Proof. Let $\mathbb{I}_u$ denote the indicator of the event $\Delta_i^u \geq \epsilon \frac{||x_T||_2}{\sqrt{B}}$. These $\mathbb{I}_u$ are bounded and therefore subgaussian, so their sum $\sum_u^R \mathbb{I}_u$ is also subgaussian with parameter $\sigma = \frac{\sqrt{R}}{2}$ and mean $\mu = \frac{R}{2}(1 - \Omega(\epsilon))$. Hence using the Chernoff bound (see lecture 1) we have

$$\mathbb{P}\left[\sum_u^R \mathbb{I}_u \geq \mu + \Omega(\epsilon R)\right] \leq e^{\frac{-\Omega(\epsilon R)^2}{R/4}} = e^{-\Omega(2\epsilon^2 R)} \tag{4}$$

Equation 4 also applies if $\mathbb{I}_u = \Delta_i^u \leq -\epsilon \frac{||x_T||_2}{\sqrt{B}}$. If neither of the events occurs then the median must lie in $(-\epsilon \frac{||x_T||_2}{\sqrt{B}}, \epsilon \frac{||x_T||_2}{\sqrt{B}})$. □

If we let $\epsilon = \sqrt{\frac{t}{R}}$ and use 1.1, we arrive at

$$\mathbb{P}\left[|\Delta_i| \geq \sqrt{\frac{t}{R}} \frac{||x_T||_2}{\sqrt{B}}\right] < 2e^{-\Omega(t)} \tag{5}$$

2 Gilbert-Li-Porat-Strauss Fast Recovery

The methods we have looked at so far optimize for space only. Gilbert-Li-Porat-Strauss proposed an alternate method in 2009 [GLPS2012] that only take $O(k \log^2 n)$ time as well as space.

For their paper, they use the the constraint $||\hat{x} - x||_2 \leq (1 + \epsilon) \underbrace{||x_T||_2}_{= \quad \mathrm{Err}_2(x,k)}$. (Reminder that x_T is the vector x with the k largest elements zeroed out)

In class we prove it for a weaker L_1 constraint instead, $||\hat{x} - x||_1 \leq (1 + \epsilon)\mathrm{Err}_1(x, k)$.

Definition 2.1. *Let H be the set of "heavy-hitters"*

$$H = \left\{ i \quad | \quad |x_i| \geq \frac{Err_1(x, k)}{k} \right\}$$

There can be at most $2k$ heavy hitters $|H| \leq 2k$

It suffices to find a superset S such that $|S| \leq o(k), S \supset H$. If we had such a set, the we could estimate x_S using count-min and get

$$||\hat{x}_S - x_S||_1 \leq |S| \frac{\mathrm{Err}_1(x, k)}{k} = O\left(\frac{\mathrm{Err}_1(x, k)}{k}\right) \tag{6}$$

If we split the error $||\hat{x_S} - x||_1$ into the heavy-hitters and non-heavy hitters that are not in S

$$||\hat{x_S} - x||_1 = \underbrace{||\hat{x_S} - x_S||_1}_{=O(\mathrm{Err}_1(x,k))} + \underbrace{||x_{(\text{top } k) \cap \bar{S}}||_1}_{=k \cdot \frac{\mathrm{Err}_1(x,k)}{k}} + \underbrace{||x_{(\text{not top } k) \cap \bar{S}}||_1}_{=\mathrm{Err}_1(x,k)} = O(\mathrm{Err}_1(x, k)) \tag{7}$$

Unfortunately, it is still a bit difficult to find such an $S \supset H$. We can at least find a set S that has 'most' of the heavy hitters.

Lemma 2.2. *In $O(k \log n)$ time and space we can recover*

$$S, \quad |S| \leq o(k) \quad \forall i \in H, i \in S \text{ with 4/5 probability}$$

Proof. **Idea:** Use a hashtable with some clever signing.

Let $h : [n] \to [B]$ be a hash function, and let $c_i = \{j \quad | h(j) = h(i), \quad j \neq i\}$

We know from previous arguments about the number of heavy hitter collisions in a hashtable that

$$||x_{c_i}|| \leq \frac{\text{Err}_1(x,k)}{k} \quad \text{with 4/5 probability} \tag{8}$$

Also, by definition

$$\frac{\text{Err}_1(x,k)}{k} \leq ||x_i||_1 \quad \forall i \in H \tag{9}$$

For each bucket, make $O(\log n)$ measurements that sum the contents with different signs

$$\begin{array}{ll} i = 0 & ++++++++++ \\ i = 1 & +++++++++- \\ i = 2 & ++++++++-+ \\ \vdots & \qquad \vdots \end{array}$$

The signs are the bit representation of the index. i.e.

$$Y_{1,v} = \sum_{h(j)=v} x_j \cdot (-1)^{j\&1}$$

$$Y_{t,v} = \sum_{h(j)=v} x_j \cdot (-1)^{(j>>t-1)\&1}$$

If i dominates a bucket:

$$\text{sign}(Y_{t,h(i)}) = (-1)^{(i>>t-1)\&1} \quad \forall t \in [R]$$

So we can recover i with good probability. □

For an L_2 approximation we can instead use an error correcting code [GLPS2012]

Idea: If S really contains H, we previously showed

$$S \supset H \Rightarrow ||\hat{x}_S - x||_1 \leq (1+\epsilon)\text{Err}_1(x,k)$$

But S only mostly contains H. So instead, the first time we get at least $k/2$ of the top k, if we subtract these from x we can try again to get half of the remaining top $k/2$

$$\text{Err}_1\left(x - \hat{x}_S, \frac{k}{2}\right) \leq (1+\epsilon)\text{Err}_1(x,k)$$

So in $O(k \log n)$ time and space we get a linear sketch $Ax \to \hat{x}_S$. The trick is to repeatedly perform this algorithm on $\hat{x}_S$. I.e. we compute a new $A'x$ with $k/2$ and also $A'\hat{x}_S$ to get $A'(x - \hat{x}_S) \to \hat{x}_{S'}$.

$$\mathrm{Err}_1(x - \hat{x}_S - \hat{x}_{S'}, k/4) \leq (1+\epsilon)\mathrm{Err}_1(x - \hat{x}_S, k/2) \leq (1+\epsilon)^2 \mathrm{Err}_1(x, k) \tag{10}$$

Now we can repeatedly perform the algorithm for $i = 1, ..., \log n$

- Let $k_i = \frac{k}{2^i}$
- Let $\epsilon_i = \frac{1}{10}\left(\frac{2}{3}\right)^i$ decay exponentially
- Compute $A^{(i)}$ with k_i and ϵ_i

Since ϵ_i are decaying, their sum forms a geometric series, and

$$\prod_{i=1}^{\log n}(1+\epsilon_i) \leq e^{\sum \epsilon_i} = e^{O(1)} = O(1) \tag{11}$$

Then using the same argument as in Equation 10, and Equation 11

$$\left\| x - \sum_{i=1}^{\log n} \hat{x}_{S_i}^{(i)} \right\|_1 \leq \prod_{i=1}^{\log n}(1+\epsilon_i)\mathrm{Err}_1(x,k) \lesssim \mathrm{Err}_1(x,k) \tag{12}$$

Question: So what are the total costs of this algorithm?

The space needed to perform this algorithm is $\sum \frac{k_i}{\epsilon_i} \log n = O(k \log n)$

The analysis of the running time has two different parts

- The time to do the recovery algorithm : $O(k \log n)$
- The time to perform the subtractions $A'(x - \hat{x}_S)$: $O(k \log^2 n)$

References

[GLPS2012] Anna C. Gilbert, Yi Li, Ely Porat, Martin J. Strauss. Approximate Sparse Recovery: Optimizing Time and Measurements *SIAM Journal on Computing* 41(2):436–453, 2012

[MP2014] Gregory T. Minton, Eric Price Improved Bounds for Count-Sketch *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 51:669–686, 2014