Problem Set 3 Solutions

November 9, 2019

1 Problem 1

Set $R \in \mathbb{R}^{r \times n}$, where $r = 2^{2^{O(d)}}$ and each $R_{i,j}$ is an i.i.d. Cauchy scaled. Now fix any $x \in \mathbb{R}^n$ and let $z = DRx \in \mathbb{R}^r$, where $D = I_r(r\log(r))^{-1}$ is the scaling matrix. Define the level sets $S_j = \{i \in [r] \mid 2^{j-1}\|x\|_1 \leq |(Rx)_i| \leq 2^j\|x\|_1\}$ for $j = 1, 2, \ldots$. Note that $\mathbb{E}[|S_j|] = (1 \pm 1/2)r2^{-j}$, so $\mathbb{E}[\sum_{i \in S_j} |z_i|] = c_0\|x\|_1/\log(r)$. First, for any $j < \log(\frac{r}{10\log r})$, we have $|S|_j = (1 \pm 1/2)r2^{-j}$ with probability $1 - r^{-10}$ (because the expected number of elements is $\Omega(\log(r))$, and we can then apply a Chernoff bound). Now note that the probability that a coordinate $(Rx)_i$ is larger than $10\log(r)r$ is at most $\frac{1}{5\log(r)r}$. Thus with probability $1 - \frac{1}{5\log(r)}$, we have $|S_j| = 0$ for all j with $2^j \geq 10r\log(r)$, so we can condition on this now. Given this, we have

$$\sum_{j < \log(\frac{r}{10\log r})} \sum_{i \in S_j} z_i = \sum_{j < \log(\frac{r}{10\log r})} c_0\|x\|_1/\log(r) = c_1\|x\|_1$$

for another constant c_1. We must now bound the contribution of the z_i's for $i \in S_j$, and $\log(\frac{r}{10\log r}) \leq j \leq \log(r)$ Now let S^- be the set of all coordinates $(Rx)_i$ with $\|x\|_1\frac{r}{20\log r} \leq (Rx)_i \leq r\|x\|_1$. Note that $\mathbb{E}[|S^-|] \leq 40\log(r)$, and $|S^-| \leq c_3\log(r)$ with probability $1 - 1/r^10$ by a Chernoff bound. Conditioned on this, we have:

$$\sum_{\log(\frac{r}{10\log r}) \leq j \leq \log(r)} \sum_{i \in S_j} z_i = \sum_{i \in S^-} z_i \leq c_2\|x\|_1$$

Finally, we bound the coordinates $(Rx)_i$ for $i \in S_j$, and $\log(r) \leq j \leq \log(10r\log(r))$. Note that

$$\sum_{r < 2^j < 10r\log r} \mathbb{E}[|S_j|\|x\|_1 2^j] = O(\log\log(r)\|x\|r)$$

So by Markov's, with probability $1 - \frac{\log\log r}{10\log r}$ we have

$$\sum_{r < 2^j < 10r\log r} \sum_{i \in S_j} z_i \leq c_3\frac{r\log r\|x\|_1}{r\log r} \leq c_4\|x\|_1$$

Altogether, by a union bound, we have with probability $1 - \frac{\log\log r}{\log r} \geq 1 - d2^{-\Omega(d)} > 1 - 2^{-\Omega(d^2)}$ that

$$c_1\|x\|_1 \leq \|DRx\|_1 \leq (2c_1 + c_2 + c_3)\|x\|_1$$

Where c_1, c_2, c_3 are constants. By a constant factor rescaling of D, and replacing x with Ay we obtain:

$$\|Ay\|_1 \leq \|DRAy\|_1 \leq C\|Ay\|_1$$

For some constant C. Now set $\gamma = \frac{1}{100C} = \Omega(1)$. We now construct a γ-net for the column span of A. We do so greedily. For $x \in \mathbb{R}^d$, let $\|x\|_A = \|Ax\|_1$, which is a norm assuming A has linearly independent columns (otherwise we can replace A with a maximally independent basis of columns). Then we greedily add points to M: while there exists a $x \in \mathbb{R}^d$ with $\|x\|_A = 1$ and $\|x - y\|_A > \gamma$ for all $y \in M$, we add x to M. At the end, we have that $\|y - y'\|_A \geq \gamma/2$ for all $y, y' \in M$. Moreover, the A-norm ball $\mathcal{B}_{\mathcal{A}}(x, \gamma/2) = \{y \in \mathbb{R}^d \mid \|x - y\|_A \leq \gamma/2\}$ is just a polytope which is similar to the polytope $\mathcal{B}_{\mathcal{A}}(0, 1 + \gamma/2)$, so the ratio of the volumes is $(1 + \gamma/2)^2/(\gamma/2)^d \leq 2^{O(d)}$, which upper bounds the size of M. It follows that the set $N = \{Ax \mid x \in M\}$ satisfies that for every $x \in \mathbb{R}^d$ with $\|Ax\|_1 = 1$, there is a $y \in N$ with $\|y - Ax\|_1 \leq \gamma$.

Given this, we can then union bound over all points in N. Since $|N| \leq 2^{O(d)}$ and our failure probability is $2^{-\beta d}$ for some constant β which we can make arbitrarily large by increasing the hidden constant in the second exponent of $r = 2^{2^{O(d)}}$, after a union bound the success probability is still at least $1 - 2^{-\Omega(d)}$. We then obtain

$$\|y'\|_1 \leq \|DRy'\|_1 \leq C\|y'\|_1$$

for all $y' \in N$. Now fix some $y = Ax$ for any $x \in R^d$, and let $y_1 \in M$ be such that $\|y - y_1\|_1 \leq \gamma$, and let $\alpha \geq 1/\gamma$ be such that $\|\alpha(y - y_1)\|_1 = 1$, and let y_2' be such that $\|\alpha(y - y_1) - y_2'\|_1 \leq \gamma$ Setting $y_2 = y_2'/\alpha$, we have $\|y - y_1 - y_2\|_1 \leq \gamma^{-2}$. We can repeat this, obtaining $y_1, y_2, y_3, \ldots$, such that for any $i \geq 1$ we have

$$\|y - y_1 - y_2 - \cdots - y_i\|_1 \leq \gamma^{-i}$$

By the reverse triangle inequality, we have $\|y_i\|_1 \leq \gamma^{-i-1} - \gamma^{-i} \leq 2\gamma^{-i-1}$, and $\|DRy\|_1 = \|DR\sum_i y_i\|_1$. We have

$$\begin{aligned}
\|DR\sum_i y_i\|_1 &\leq \sum_i \|DRy_i\|_1 \\
&\leq \|DRy_1\|_1 + \sum_{i\geq 2} \|DRy_i'\|_1 2\gamma^{i-1} \\
&\leq C\|y_1\|_1 + \sum_{i\geq 2} C2\gamma^{i-1} \\
&\leq C + 2C \\
&\leq 3C
\end{aligned} \tag{1}$$

Similarly,

$$\begin{aligned}
\|DR\sum_i y_i\|_1 &\geq \|DRy_1\|_1 - \|\sum_{i\geq 2} DRy_i\|_1 \\
&\geq \|DRy_1\|_1 - \sum_{i\geq 2} \|DRy_i\|_1 \\
&\geq \|y_1\|_1 - \sum_{i\geq 2} \|DRy_i'\|_1 2\gamma^{i-1} \\
&\geq 1 - \sum_{i\geq 1} C2(\frac{1}{100C})^i \\
&\geq 1/2
\end{aligned} \tag{2}$$

Thus, for all $x \in \mathbb{R}^d$, we have that

$$\frac{1}{2}\|Ax\|_1 \leq \|DRAx\|_1 \leq 3C\|Ax\|_1$$

which completes the proof.

2 Problem 2

2.1 2.1

First pretend we can work with infinite precision Cauchy's. Recall from class, if we take $S \in \mathbb{R}^{k\times n}$ to be a count-sketch with $k = O(1/\epsilon^2)$, we will obtain an estimate $R_2 = \|Sx\|_2^2 = (1 \pm \epsilon)\|x\|_2^2$ with probability 99/100. Now let $C \in \mathbb{R}^{k\times n}$ be a dense matrix of i.i.d. Cauchy random variables. Note by 1-stability that $(Cx)_i \sim Z\|x\|_1$, where Z is again a Cauchy random variable. In class we saw:

$$R_1 = \text{median}_i|(Cx)_i| = (1 \pm \epsilon)\|x\|_1$$

with probability 99/100, which we can union bound over to obtain both results for R_1, R_2 with probability 9/10. Conditioned on this, we have $2R_2 - R_1 = 2\|x\|_2^2 - \|x\|_1 \pm \epsilon(2\|x\|_2^2 + \|x\|_1) = 2\|x\|_2^2 - \|x\|_1 \pm \epsilon(2\|x\|_2^2 - \|x\|_1) = (1 \pm O(\epsilon))2\|x\|_2^2 - \|x\|_1$ as needed. here we used that $\|x\|_1 \leq \|x\|_2^2$ for an integer vector x, so $2\|x\|_2^2 + \|x\|_1 \leq \|x\|_2^2 = O(\|x\|_2^2 + \|x\|_1)$.

Now $\|C\|_\infty \leq n^{2c}$ with probability $1 - 1/n^c$ for large enough constant c, so by truncating each Cauchy C_i into a cauchy C_i' with additive error $\delta = 1/n^{2c}$, we have that $\|C - C'\|_\infty \leq 1/n^{2c}$, and moreover that the resulting bit complexity is such that each $C_{i,j}$ can be represented in $O(\log(n))$ bits. Thus $\|(C - C')x\|_\infty \leq \|x\|_\infty/n^{2c}$, so setting c larger enough such that $n^c > M$, we have that $\|(C - C')x\|_\infty \leq n^{-c}$. Thus the value of our sketch changes by a additive n^{-c} coordinate wise, therefore the median can change by at most this much, and we get an additive n^{-c} error in our estimate of $\|x\|_1$. If $\epsilon < 1/n$, then we can just store the whole vector, otherwise this additive error is relative $(1 \pm \epsilon)$ error unless $x = 0$ (which we know how to test).

2.2 2.2

We perform a reduction from disjointness to F_p estimation. Label the players $P^1, \ldots, P^t$, where $t = n^{1/p}$, and player P^i holds $S_i \subset [n]$. To solve disjointness, the players create a stream, and for $i = 1, 2, \ldots, t$, player i inserts the indicator vector $x \leftarrow x + \vec{1}_{S_i}$ to the stream, where $\vec{1}_{S_i}$ is the vector that is 1 on the entries in S_i and 0 otherwise. They can do this by the first player starting a stream and adding their vector, and then sending it to the coordinator, who then relays the message to the second player, who adds their vector to the stream, and so on. Note that at each point, the message send is the state of the F_p streaming algorithm run on the input vector x. In addition, the players forward $\log(n)$ bits to keep track of the sum $N = \sum_i |S_i|$. First note that if the sets are disjoint $\|x\|_p^p = N$, since $\|x\|_0 = N$ and all non-zero entries are 1. Now if the sets are not disjoint, then $\|x\|_p^p = (N - 1) + t^p = (N - 1) + n$. Now note that $N \leq n$, since x only has n coordinates, thus $\|x\|_p^p$ differs by a factor of 3/2 between the two cases (e.g. $(N - 1) + n > (3/2)(N - 1)$), and so a constant factor F_p estimation algorithm will distinguish these cases.

Now observe that the total communication lower bound is $\Omega(n/t) = \Omega(n^{1-1/p})$. But since there were at most $O(t)$ messages, two for each player to relay to the coordinator and back to the next player, it follows that at least one player must have sent a message of size $\Omega(n/t^2) = \Omega(n^{1-2/p})$. Thus the size of the streaming algorithm must be $\Omega(n^{1-2/p})$, which completes the proof. Note that the additive $\log(n)$ bits the players use to send N is negligible with respects to this polynomial lower bound.