15-851 Algorithms for Big Data — Spring 2024

Problem Set 3

Due: Thursday, March 21, before class

Please see the following link for collaboration and other homework policies:
`http://www.cs.cmu.edu/afs/cs/user/dwoodruf/www/teaching/15851-spring24/grading.pdf`

Problem 1: Embedding ℓ_p into ℓ_r (25 points)

1. We first consider each coordinate of Ty. For $(Ty)_i$ from the property of p-stable random variable we have $(Ty)_i = \sum_j T_{ij} y_i \sim \|y\|_p \cdot C$, where C is a p-stable random variable. Hence we have that

$$\mathbf{E}[|(Ty)_i|^r] = \mathbf{E}[\|y\|_p^r \cdot |C|^r] = \|y\|_p^r \cdot \mathbf{E}[|C|^r] = \|y\|_p^r \int_{x=0}^{\infty} \Theta\left(\frac{|x|^r}{1+|x|^{p+1}}\right) dx = \Theta(1) \cdot \|y\|_p^r \ .$$

To get this, notice that

$$\begin{aligned}\int_{x=0}^{\infty} \frac{|x|^r}{1+|x|^{p+1}} dx &= \int_{x=0}^{1} \frac{|x|^r}{1+|x|^{p+1}} dx + \int_{x=1}^{\infty} \frac{|x|^r}{1+|x|^{p+1}} dx \\ &\leq \int_{x=0}^{1} \frac{|x|^r}{1} dx + \int_{x=1}^{\infty} \frac{|x|^r}{|x|^{p+1}} dx = \Theta(1) \ .\end{aligned}$$

From the linearity of the expectation we have that

$$\mathbf{E}[\|Ty\|_r^r] = \mathbf{E}\left[\sum_i |(Ty)_i|^r\right] = \sum_i \mathbf{E}\left[|(Ty)_i|^r\right] = \Theta(m \cdot \|y\|_p^r) \ .$$

2. From part 1, we have that for every i, $\mathbf{E}|(Ty)_i|^r = \frac{\alpha_{p,r}^r}{\alpha_{p,r}^r \cdot m}\|y\|_p^r = \frac{1}{m}\|y\|_p^r$. To get concentration, we pick an $r' \in (r,p)$ and consider the r'/r-moment of $(Ty)_i^r$. Similarly, we have that $\mathbb{E}[|(Ty)|_i^{r'}] = \frac{\beta_{p,r,r'}}{m^{r'/r}}\|y\|_p^{r'}$ is bounded, where $\beta_{p,r,r'}$ is a constant depending on p, r, r' only. Let $S = \sum_i |(Ty)_i|^r$ and we have that $\mathbf{E}[S] = \|y\|_p^r$. Consider the (r/r')-th moment of

S. We then have

$$\begin{aligned}
\mathbf{E}\left[(S-\mathbf{E}[S])^{r'/r}\right] &= \mathbf{E}\left[\left(\sum_i \left(|(Ty)_i|^r - \frac{1}{m}\|y\|_p^r\right)\right)^{r'/r}\right] \\
&\le 2\left(\sum_i \mathbf{E}\left[|(Ty)_i|^r - \frac{1}{m}\|y\|_p^r|\right]^{r'/r}\right) && \text{(Fact 2)} \\
&\le 2^{r'/r+1}\left(\sum_i \mathbf{E}|(Ty)_i|^{r'}\right) && \text{(Fact 1)} \\
&\le C\left(\sum_i \frac{1}{m^{r'/r}}\|y\|_p^{r'}\right) \\
&= C\|y\|_p^{r'}/m^{r'/r-1} \ ,
\end{aligned}$$

where C is a constant that depends only on r, r', and p. By Markov's inequality, we have that

$$\begin{aligned}
\mathbf{Pr}\left[|S-\mathbf{E}[S]| \ge \varepsilon \mathbf{E}[S]\right] &\le \mathbf{Pr}\left[|S-\mathbf{E}[S]|^{r'/r} \ge (\varepsilon\mathbf{E}[S])^{r'/r}\right] \\
&\le \frac{\mathbf{E}\left[(S-\mathbf{E}[S])^{r'/r}\right]}{\varepsilon^{r'/r}\|y\|_p^{r'}} \\
&\le \frac{C_{r'/r}}{\varepsilon^{r'/r}m^{r'/r-1}} \ .
\end{aligned}$$

Hence, we can see that when $m = \Omega(1/\varepsilon^{\frac{r'}{r'-r}}) = 1/\varepsilon^{\Omega(1)}$, $|\|Ty\|_r - \|y\|_p| \le \varepsilon\|y\|_p$ holds with large constant probability.

3. We first show that for every $x \in \mathbb{R}^d$, $\|Ty\|_r^r \ge (1-\varepsilon)\|y\|_p^r$ holds with probability at least $1-\exp(-d\log(nd))$. Recall that we have that $\mathbf{E}|(Ty)_i|^r = \frac{1}{m}\|y\|_p^r$ for every i. Fix $k = 1/\varepsilon^{O(1)}$. Let

$$s_i = |(Ty)_{(i-1)k+1}|^r + |(Ty)_{(i-1)k+2}|^r + \cdots + |(Ty)_{ik}|^r \quad (1 \le i \le m/k) \ .$$

We then have $\|Ty\|_r^r = \sum_i s_i$. Similar to (1), one can show that for each i, with probability at least $1-O(\varepsilon)$,

$$\left|s_i - \frac{k}{m}\|y\|_p^r\right| \le \varepsilon\frac{k}{m}\|y\|_p^r \tag{1}$$

By a Chernoff bound, with probability at least $1-\exp(-d\log(nd))$, at least a $(1-\varepsilon)$-fraction of the s_i satisfy (1). Conditioned on this event, it holds that

$$\|Ty\|_r^r = \sum_i s_i \ge \frac{m}{k}(1-\varepsilon)\frac{k}{m}\|y\|_p^r = (1-\varepsilon)\|y\|_p^r \ ,$$

which is what we need. The next is a standard net-argument. Let $\mathcal{S} = \{Ax : x \in \mathbb{R}^d, \|Ax\|_p = 1\}$ be the unit ℓ_p-ball and $\mathcal{N}$ be a γ-net with $\gamma = \text{poly}(1/nd)$ under the ℓ_p distance. It is a standard fact that the size of $\mathcal{N}$ can be $(nd)^{O(d)}$. By a union bound, we have that $\|TAx\|_r \geq (1-\varepsilon)\|Ax\|_p = (1-\varepsilon)$ for all $Ax \in \mathcal{N}$ simultaneously with probability at least 9/10. From the fact the problem gives, we have that with probability at least 9/10, $\|TAx\|_p \leq \text{poly}(nd)\|Ax\|_p$ for all $x \in \mathbb{R}^d$. Conditioned on these events, we then have for all $x \in \mathbb{R}^d$,

$$\|TAx\|_r \leq m^{1/r-1/p}\|TAx\|_p \leq \text{poly}(nd)\|Ax\|_p \ .$$

Then, for each $y = Ax \in \mathcal{S}$, we can choose a y_1 such that $\|y - y_1\|_p \leq \gamma$. Suppose that $y_1 = Ax_1$, then we have

$$\|TAx\|_r \geq \|TAx_1\|_r - \|TA(x - x_1)\|_r \geq (1-\varepsilon) - \gamma \cdot (\text{poly}(nd)) = 1 - O(\varepsilon)$$

given the condition $\frac{1}{\varepsilon} < n$. Rescaling ε, we obtain that $\|TAx\|_r^r \geq (1-\varepsilon)\|Ax\|_p^r$ for all $x \in \mathbb{R}^d$ simultaneously.

4. Let $x^\star \in \mathbb{R}^d$ be the solution to $\min_{x\in\mathbb{R}^d} \|Ax - b\|_r$. Applying part 2 on the vector $Ax^\star - b$ we have that with high constant probability

$$\|T(Ax^\star - b)\|_r \leq (1+\varepsilon)\|Ax^\star - b\|_p \ .$$

From this we can get that

$$\min_{x\in\mathbb{R}^d} \|T(Ax - b)\|_r \leq (1+\varepsilon) \min_{x\in\mathbb{R}^d} \|Ax - b\|_p \ .$$

Suppose that $x' \in \mathbb{R}^d$ is the solution to $\min_{x\in\mathbb{R}^d} \|T(Ax - b)\|_r$, It follows from part 3 that

$$\|Ax' - b\|_p \leq \frac{1}{1-\varepsilon}\|T(Ax' - b)\|_r \leq (1 + O(\varepsilon)) \min_{x\in\mathbb{R}^d} \|Ax - b\|_p \ .$$

Problem 2: Communication Complexity and Streaming (25 points)

1. Let us consider the uniform distribution on σ. Let M be Alice's message to Bob. Similarly to what we did in the lecture, we have for at least one choice of Alice's random coins, the message $|M| \geq H(M) \geq I(\sigma; M)$. Hence, we only need to bound $I(\sigma; M)$. Let σ_i denote the i-th bit of the list $\sigma(1), \sigma(2), ..., \sigma(n)$. By the chain rule, we have that

$$\begin{aligned} I(\sigma; M) &= \sum_{i=1}^{n \log n} I(\sigma_i; M \mid \sigma_{<i}) \\ &= \sum_i (H(\sigma_i \mid \sigma_{<i}) - H(\sigma_i \mid M, \sigma_{<i})) \\ &\geq \sum_i (H(\sigma_i \mid \sigma_{<i}) - H(\sigma_i \mid M)) \\ &= H(\sigma) - \sum_i (H(\sigma_i \mid M)) \ . \end{aligned}$$

Using Stirling's approximation, we have that $H(\sigma) = \log n! \geq n\log(n/e)$. We next consider the $H(\sigma_i \mid M)$. Since M is a randomized protocol that succeeds on every pair of inputs (σ, i) with probability at least 99/100, and M does not depend on i, it follows that from M Bob can predict σ_i for any given i with probability at least 99/100. By Fano's inequality, we have that $H(\sigma_i \mid M) \geq H(1/100)$. Putting these two things we have that $I(\sigma; M) \geq n\log(n/e) - H(1/100) \cdot n\log n = \Omega(n\log n)$, which is what we need.

2. We will perform a reduction from the problem in (1) with size $n/2$. Given a permutation σ, Alice creates a perfect matching from $[n/2]$ to $[n/2]$ where the i-th left vertex connects to the $\sigma(i)$-th right vertex. Let L and R denote the two parts of the vertex set V, each of size $n/2$. Suppose that there is a streaming algorithm $\mathcal{A}$ that solves the graph connectivity problem. Given the input of the permutation σ in the problem in (1), Alice adds the matching edges to $\mathcal{A}$ and sends the memory of $\mathcal{A}$ to Bob.

Suppose the input of the problem in (1) to Bob is i, which corresponds to the ℓ-th bit in $\sigma(j)$. Bob then adds the edges to the graph as follows. Let $S \subseteq R$ denote the subset of vertices whose ℓ-th bit is equal to 0. Bob then creates edges that consist of a spanning tree on $(L \setminus \{j\}) \cup S$ and adds them to $\mathcal{A}$. We can ensure the edges of the spanning tree are disjoint from the matching edges by including a new vertex w and including the edges to w.

Let us consider the underlying graph of $\mathcal{A}$. Observe that since the vertices in $L \setminus \{j\}$ are connected. Since we placed a perfect matching from L to R, any vertex u is connected to any other vertex except possibly to j or $\sigma(j)$. Now, if the $\sigma(j)$-th right vertex has its ℓ-th bit equal to 0, then $\sigma(j)$ is connected to S, and hence to $L \setminus \{j\}$. It follows that the graph is connected. On the other hand, if the $\sigma(j)$-th right vertex has its ℓ-th bit equal to 1, then the edge from the j-th left vertex to the $\sigma(j)$-th right vertex is isolated. In this case the graph is disconnected. From the above discussion, we get that if we can solve the graph connectivity problem, we can solve the permutation problem in (1), which results in a $\Omega(n\log n)$ bits of space lower bound for the graph connectivity problem.