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)
We studied Cauchy random variables in class, which are 1-stable and have the property that if $a_1, \ldots, a_n$ are fixed scalars, and $C_1, \ldots, C_n$ are i.i.d. Cauchy random variables, then $\sum_{i=1}^n a_i C_i$ is distributed as $\|a\|_1 \cdot C$, where C is a Cauchy random variable. In fact, p-stable random variables exist for any $0 < p < 2$, and they have the property that if $a_1, \ldots, a_n$ are fixed scalars and $X_1, \ldots, X_n$ are i.i.d. p-stable random variables, then $\sum_{i=1}^n a_i X_i$ is distributed as $\|a\|_p \cdot X$, where X is a p-stable random variable.

Let $1 < r < p$, and let $T \in \mathbb{R}^{m \times d}$ be a matrix of i.i.d. p-stable random variables. Consider a fixed $y \in \mathbb{R}^n$. We are going to look at $\|Ty\|_r^r$.

1. It turns out that since $r < p$, that $\|Ty\|_r^r$ exists. Although a closed-form expression for the probability density function of p-stable random variables is not known for general p, it is known that the density function $f(x)$ of a standard p-stable random variable is $\Theta\left(\frac{1}{1+|x|^{p+1}}\right)$. Prove that for any fixed y that $\mathbf{E}[\|Ty\|_r^r] = \Theta(m \cdot \|y\|_p^r)$.

2. It turns out that for any fixed y that $\mathbf{E}[|Ty\|_r^r] = \alpha_{p,r}^r \cdot m\|y\|_p^r$, where $\alpha_{p,r} > 0$ is a constant that does not depend on y. In this part of the problem we let $T \in \mathbb{R}^{m \times d}$ be a matrix of i.i.d. p-stable random variables, each scaled by $\frac{1}{m^{1/r} \cdot \alpha_{p,r}}$. In this way, we have $\mathbf{E}[\|Ty\|]_r^r = \|y\|_p^r$. Our next goal will be to show that $\|Ty\|_r^r$ is concentrated around its expectation. We will state the following two facts that you can use without proof:

   Fact 1 *(Fact 1) Suppose that $r, s \geq 1$ and X is a random variable with $\mathbf{E}[|X|^{rs} < \infty$. Then*
   $$\mathbf{E}\left[\left||X|^r - \mathbf{E}[|X|^r]\right|^s\right] \leq 2^s \mathbf{E}[|X|^{rs}].$$

   and

   Fact 2 *(Fact 2) Suppose that $1 \leq r \leq 2$. Let $X_1, \ldots, X_n$ be independent zero mean random variables with $\mathbf{E}[|X_i|^r] < \infty$. Then we have that*
   $$\mathbf{E}\left[\left(\sum_{i=1}^n |X_i|\right)^r\right] \leq 2 \sum_{i=1}^n \mathbf{E}[|X_i|^r].$$

Assume that $p - r = \Omega(1)$ and $1 < r < p$ as before. Show that if we choose m to be a large enough $\mathrm{poly}(1/\epsilon)$, that $|\|Ty\|_r^r - \|y\|_p^r| \le \epsilon \|y\|_p^r$ with probability at least $99/100$.

HINT: For a random variable S and $\beta > 1$, by Markov's inequality we have that

$$\mathbf{Pr}[|S - \mathbf{E}[S]| \ge \varepsilon \mathbf{E}[S]] = \mathbf{Pr}[|S - \mathbf{E}[S]|^\beta \ge \varepsilon^\beta \mathbf{E}[S]^\beta] \le \frac{\mathbf{E}[|S - \mathbf{E}[S]|^\beta]}{\varepsilon^\beta \mathbf{E}[S]^\beta}$$

It will be useful to consider the r'/r-th moment of $|\|Ty\|_r^r - \|y\|_p^r|$ for an r' with $r < r' < p$. You may also find it useful to apply the previous two facts in your analysis, though other proof strategies are also possible.

3. In this part of the problem we will fix a given $n \times d$ matrix A and now let $T \in \mathbb{R}^{m \times d}$ be a matrix of i.i.d. p-stable random variables, each scaled by $\frac{1}{m^{1/r} \cdot \alpha_{p,r}}$, where $m = d \cdot \log(nd) \cdot \mathrm{poly}(1/\epsilon)$. Show that with probability at least $99/100$, we have that $\|TAx\|_r \ge (1 - \epsilon)\|Ax\|_p$ simultaneously for all $x \in \mathbb{R}^d$.

   HINT: you may first show that for a single vector y, $\|Ty\|_r^r \ge (1 - \varepsilon)\|y\|_p^r$ holds with probability at least $1 - \exp(-d \log(nd))$. You may want to borrow some analysis from the previous part and apply it to "chunks" of $\mathrm{poly}(1/\epsilon)$ rows of T at a time. Then you may want to set up indicator random variables for a Chernoff bound to achieve exponentially small failure probability, and finally apply a net argument. You can freely use the fact that a $1/\mathrm{poly}(nd)$-net of a d-dimensional subspace of ℓ_p has size $(nd)^{O(d)}$ and with high probability we have simultaneously for all x that:

$$\|TAx\|_p \le \mathrm{poly}(nd)\|Ax\|_p.$$

   This can be shown by showing that with high probability, every entry of T is at most $\mathrm{poly}(nd)$.

   The net argument for ℓ_1 in class might also be helpful for this part of the problem.

4. Although we do not have $\|TAx\|_r = (1 \pm \epsilon)\|Ax\|_p$ simultaneously for all x, explain why the guarantees in parts 2 and 3 are sufficient to solve the ℓ_p-regression problem $\min_x \|Ax - b\|_p$ up to a $(1 + \epsilon)$ factor by letting x' be the minimizer to $\|TAx - Tb\|_r$ and outputting x'. In particular, which fixed y do you apply part 2 to? It turns out that r-norm regression for any $1 < r$ can be solved efficiently, where the running time depends polynomially on the dimensions of the problem.

Problem 2: Communication Complexity and Streaming (25 points)
We consider the following two-player communication game. Assume n is a power of 2. Alice is given a permutation on $\{1, 2, \ldots, n\}$, which we represent as a list of n numbers $\sigma(1), \sigma(2), \ldots, \sigma(n)$, each of which is $\log_2 n$ bits long. Let σ be the length-$n \log_2 n$ bitstring, which is the concatenation of these numbers. Note that this is not the smallest possible

encoding of the permutation since not every list of n numbers forms a permutation. Nevertheless, we shall use this encoding. Bob is given an index $i \in \{1, 2, \ldots, n \log_2 n\}$. Alice sends a single message M to Bob and Bob would like to output the i-th bit in the encoding σ. Suppose Bob is required to succeed with probability at least 99/100, over the random coins tossed by Alice and Bob.

1. Show that for at least one choice of random coins, Alice's message must be $\Omega(n \log n)$ bits long.

   HINT: Try to adapt the lower bound on the randomized 1-way communication complexity of the INDEX problem from class. Stirling's approximation may also help, namely, that $n! = \Theta(\sqrt{n}(n/e)^n)$.

2. Use the communication problem in the previous part to prove an $\Omega(n \log n)$ bit lower bound on the memory required of any data stream algorithm which succeeds with probability at least 99/100, in solving the following problem: the algorithm sees a stream of edges, one after the other, in an undirected graph, and the algorithm needs to decide if the graph is connected.

   HINT: Start with an instance of the communication problem on permutations on $n/2$ numbers and interpret Alice's input as a perfect matching on two sets of $n/2$ numbers, 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 $n/2$. Bob will add certain edges to this graph so that the output of the streaming algorithm can be used to solve the communication problem. In particular, suppose that the input to Bob is i, which corresponds to the ℓ-th bit in $\sigma(j)$. Let $S \subseteq R$ denote the subset of vertices whose ℓ-th bit is equal to 0. If Bob connects the vertices on $(L \setminus \{j\}) \cup S$, what will happen for the two different cases?