CSCE 658: Randomized Algorithms – Spring 2024
Problem Set 3

Due: Tuesday, February 20, 2024, 5:00 pm CT

Problem 1. (30 points total) CountSketch tail bounds.

For any vector $x \in \mathbb{R}^n$ and any integer $k \geq 0$, we define $\textsc{Tail}_k(x)$ to be the vector x, but with the k entries of largest magnitude to be set to 0, breaking ties arbitrarily. For example if $x = (-100, 40, 40, 1)$, then $\textsc{Tail}_2(x)$ can be either $(0, 0, 40, 1)$ or $(0, 40, 0, 1)$.

1. (5 points) Show that for any parameter $\alpha \geq 1$ and $k \leq n-1$, there exists $x \in \mathbb{R}^n$ such that
$$\alpha \cdot \|\textsc{Tail}_k(x)\|_2 < \|x\|_2.$$
That is, the length of a tail vector of x can be arbitrarily smaller than the length of the vector x.

2. (20 points) Show that CountSketch actually provides an L_2 tail guarantee. More specifically, for $\varepsilon \in (0,1)$, suppose we use CountSketch with $\mathcal{O}\left(\frac{1}{\varepsilon^2} \cdot \log n\right)$ buckets to extract estimates $\widehat{x_i}$ for the value of each coordinate x_i. Show that with probability $1 - \frac{1}{n^2}$, we simultaneously have that for all $i \in [n]$,
$$|\widehat{x_i} - x_i| \leq \varepsilon \cdot \|\textsc{Tail}_k(x)\|_2,$$
where $k = \frac{1}{\varepsilon^2}$.

3. (5 points) Conclude that at the end of an insertion-deletion stream, CountSketch with $\mathcal{O}(k \log n)$ buckets can with high probability, recover the exact coordinates of a vector that is k-sparse, even if at intermediate times in the stream, the underlying frequency is not k-sparse.

Problem 2. (30 points total) AMS Sketch for F_p

Let $p \geq 1$. Suppose $f \in \mathbb{R}^n$ is defined by an insertion-only stream of length m, where each update increments a coordinate of f. Suppose we sample an update $t \in [m]$ in the stream, uniformly at random, and set a counter c to be the number of times the item appears in the stream after time t (including time t). After the stream ends, we set $Z = c^p - (c-1)^p$.

For example, suppose the stream consists of the updates $1, 2, 2, 1, 4, 1, 2, 1$, which induces the frequency vector $f = (4, 3, 0, 1)$ and suppose we sample the fourth update of the stream, corresponding to a 1. Then we see a total of three instances of 1, after that time (inclusive), so that $c = 3$ and $Z = 3^p - 2^p$. For $p = 3$ then, we would have $Z = 27 - 8 = 19$.

1. (5 points) Show that $\mathbb{E}[Z] = f_j^p$, *conditioned* on sampling $j \in [n]$.

2. (5 points) Let $F = m \cdot Z$. Show that $\mathbb{E}[F] = \|f\|_p^p$.

3. (10 points) Show that $\mathrm{Var}[F] \leq p \cdot \|f\|_1 \cdot \|f\|_{2p-1}^{2p-1}$.

   HINT: You may use the fact that for all $x \geq 1$ and $p \geq 1$, we have $x^p - (x-1)^p \leq p x^{p-1}$.

4. (10 points) Given an algorithm that uses $O\left(\frac{1}{\varepsilon^2}n^{1-1/p}\right) \cdot \log(nm)$ bits of space and with probability at least $\frac{2}{3}$, outputs an estimate $\widehat{F}$ such that

$$(1-\varepsilon)\|f\|_p^p \leq \widehat{F} \leq (1+\varepsilon)\|f\|_p^p.$$

Justify both its correctness-of-approximation and space complexity.

HINT: You may use the fact that for all $\|f\|_1 \cdot \|f\|_{2p-1}^{2p-1} \leq n^{1-1/p}\|f\|_p^{2p}$.

Problem 3. (30 points total) Easy as 123.

1. (2 points) Suppose we want to count the number of updates, i.e., the length of a data stream. Describe a naïve streaming algorithm that uses $\mathcal{O}(\log m)$ bits of space if the stream has length m, where m is not known in advance.

Consider the following algorithm:

Algorithm 1 Approximate counting

```
C ← 0
for each stream update do
    Flip a coin that is HEADS with probability 1/2^Z
    if the coin is HEADS then
        C ← C + 1
return Z = 2^X − 1
```

2. (10 points) Compute, with proof, $\mathbb{E}[Z]$.
3. (10 points) Compute, with proof, $\mathrm{Var}[Z]$.
4. (5 points) Given an algorithm that uses $\mathcal{O}(\log\log m)$ bits of space and with probability at least $\frac{2}{3}$, outputs an estimate $\widehat{M}$ such that

$$\frac{m}{2} \leq \widehat{M} \leq 2m,$$

where m is the length of the stream, but is not known in advance. Justify both its correctness-of-approximation and space complexity.

Problem 4. (30 points total) Approximate matrix multiplication

Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times d}$ be matrices. We will show there exists an algorithm that samples $s = \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$ columns of A to form a matrix $X \in \mathbb{R}^{n\times s}$ and s rows of B to form a matrix $Y \in \mathbb{R}^{s\times d}$ such that with probability at least $\frac{2}{3}$,

$$\|AB - XY\|_F \leq \varepsilon\|A\|_F \cdot \|B\|_F.$$