Problem Set 1

Due May 8th, 2021

Solve at least three of the following four problems.

Problem 1. Suppose we are receiving updates to integer variables $X_1, \cdots, X_n$ in a turnstile stream, where each X_i is initially set to 0, and for any $S \subseteq [n]$, $\sum_{i \in S} X_i$ always remains in $[N]$. Our goal is to estimate $Z = \sum_i Z_i$ at the end of the stream where $Z_i = f(X_i)$. f is a mapping from $[N]$ to $[m]$, with two properties: i) $f(0) = 0$, and ii) for any a, b, we have $|f(a) - f(b)| \leq f(a+b) \leq f(a) + f(b)$. Give a streaming algorithm that uses $\text{poly}(m, \log n, \log N)$ space for estimating Z at the end of the stream upto a constant factor and with high probability, i.e., $(1 - 1/n)$.

Problem 2.

- i) Alice has a set of numbers $x_1, \cdots, x_n$ where each x_i is in $[-1, 1]$. She wants to send a message to Bob. Bob has a query $q \in [-1, 1]$, and he wants to compute $X = \sum_{i: x_i \leq q} (q - x_i)$. Give a sketching algorithm that Alice can use and send a message of size $\tilde{O}(\sqrt{1/\epsilon})$[1] to Bob so that he can approximate X upto an additive factor of $n\epsilon$ for any query.
- ii) Show that Alice cannot send a message of size better than $\tilde{\Omega}(\sqrt{1/\epsilon})$.
- iii) Turn the above algorithm into a streaming algorithm: Assume that a set of numbers $x_1, \cdots, x_n$ are coming in a stream where each number x_i is in $[-1, 1]$. At the end of the stream a query $q \in [-1, 1]$ also comes. The goal is to compute $X = \sum_{i: x_i \leq q} (q - x_i)$. Give and analyze a streaming algorithm that uses space $\tilde{O}(\sqrt{1/\epsilon})$, and can approximate X upto an additive factor of ϵn for any query.

Problem 3. Assume that the edges of an n vertex graph are coming in a stream. Prove that any streaming algorithm for finding a perfect matching in a single pass requires $\tilde{\Omega}(n^2)$ space.

Problem 4. Give an algorithm for computing a t-robust c-coreset for diversity maximization under the Min-Dist notion of diversity. That is, we want an algorithm that given a point set P_i, computes a subset $S_i \subseteq P_i$ such that for any set of outliers O of size at most $|O| \leq t$, we have that

$$\text{Div}_k(\bigcup_i S_i \setminus O) \geq (1/c) \cdot \text{Div}_k(\bigcup_i P_i \setminus O).$$

Here, for a set of points X, we define $\text{Div}_k(X) = \max_{Y \subseteq X, |Y| = k} \min_{a, b \in Y} \text{dist}(a, b)$, where dist is any metric distance. Finally, we want the size of the core-set to be at most $|S_i| \leq O(t \cdot k)$.

[1] here $\tilde{O}$ is used to hide polylog factors in $1/\epsilon$ and n