Problem Set 1

Sublinear Algorithms

Due Tuesday, September 22

1. For a data stream received under the non-strict turnstile model, let $x \in \mathbb{Z}^n$ be the final histogram. You may assume that each entry in x is an $O(\log n)$ bit integer.

   (a) Show how to determine whether x is the all-zero vector, using $\mathrm{poly}(\log(n/\delta))$ space. Bonus: use $O(\log(n/\delta))$ space.

   (b) Give an algorithm that detects if x has a single non-zero entry, and if so finds that location. That is, give an algorithm to compute

   $$\begin{cases} i & \text{if } x = \lambda e_i \text{ for some } \lambda \neq 0 \\ \perp & \text{otherwise} \end{cases}$$

   with probability $1 - \delta$ using $\mathrm{poly}(\log(n/\delta))$ space. Hint: first identify whether x has zero, one, or at least two non-zero entries.

2. Recall the AMS sketch from class for $\|\cdot\|_2$ estimation, when applied for constant probability: a random $m \times n$ matrix A with entries $A_{ij} \in \{\pm 1\}$ is drawn for $m = O(1/\epsilon^2)$, and $\|x\|_2^2$ is estimated as $\frac{1}{m}\|Ax\|_2^2$. With at least 3/4 probability, we had

   $$(1-\epsilon)\|x\|_2^2 \leq \frac{1}{m}\|Ax\|_2^2 \leq (1+\epsilon)\|x\|_2^2 \tag{1}$$

   when each row has 4-wise independent entries.

   (a) Consider the following matrix instead: for each $i \in [n]$, let the ith column of A have a single $\pm\sqrt{m}$ in a random row, and 0s elsewhere (with full independence). Because this matrix is sparse, it can be maintained under turnstile updates in *constant* time. Show that this A still satisfies (1) with 3/4 probability for $m = O(1/\epsilon^2)$.

   (b) Show how to generate A using only $O(\log n)$ bits of randomness, while still satisfying the guarantee.