Problem Set 2

Sublinear Algorithms

Due Tuesday, September 15

In the *gap Hamming* problem $\texttt{GapHamming}_n$, Alice and Bob are respectively given vectors $x, y \in \{0,1\}^n$. It is guaranteed that the Hamming distance between x and y is either at least $n/2 + \sqrt{n}$ or at most $n/2 - \sqrt{n}$. Alice and Bob must communicate to figure out which situation they are in, i.e., at the end of the communication Bob should output YES if $\|x-y\|_1 > n/2 + \sqrt{n}$ and NO if $\|x-y\|_1 < n/2 - \sqrt{n}$, with 99% probability.

1. Prove that the communication complexity of gap Hamming is $\Omega(n)$, as follows.
   (a) Let $x \in \{-1,1\}^n$ be any fixed vector, and let $r \in \{-1,1\}^n$ be uniformly distributed. For every $i \in [n]$, show that

   $$\Pr_r[\operatorname{sign}(\langle x, r\rangle) \neq x_i r_i] \leq \frac{1}{2} - \Omega(\frac{1}{\sqrt{n}})$$

   when n is odd.

   (b) Let $x \in \{-1,1\}^n$ be any fixed vector, and $i \in [n]$ be fixed. Let $r^{(1)}, \ldots, r^{(m)} \in \{-1,1\}^n$ be iid uniformly random. Define

   $$u_j := \operatorname{sign}(\langle x, r^{(j)}\rangle)$$
   $$v_j := \operatorname{sign}(\langle e_i, r^{(j)}\rangle)$$

   Using your bound from part (a), analyze the mean and variance of $\|u-v\|_1$.

   In particular, you should be able to show that 90% of the time when $x_i = 1$, $\|u-v\|_1$ is "small"; and 90% of the time when $x_i = -1$, $\|u-v\|_1$ is "large".

   (c) Describe how this means a randomized protocol for `GapHamming` leads to a randomized protocol for `Index`. Conclude with an $\Omega(n)$ lower bound for one-way randomized protocols for `GapHamming` on n bits.[1]

2. Using reductions from the gap Hamming problem:
   (a) Show that any insertion-only streaming algorithm for distinct elements over $[n]$ must use $\Omega(\min(n, 1/\epsilon^2))$ space.

   (b) Show that any insertion-only streaming algorithm for estimating $\|x\|_2$ must use $\Omega(\min(n, 1/\epsilon^2))$ bits.

[1] The $\Omega(n)$ bound even holds with general (multiway) communication, but the proof is more involved.