Communication Complexity 2022 Midterm

Total Marks Obtainable: 100 Time: 9:30 AM - 12:30 PM

Hint Sheet:

1. The randomized two-sided error public coin communication complexity of the function disjointness DISJ_n is $\Omega(n)$.
2. There exists a randomized public coin protocol for Equality (EQ) with communication complexity $O(1)$ and error at most $1/4$.
3. For vectors $x, y \in \mathbb{R}^n$, we have $||x||_2 \leq ||x||_1$ and $|\langle x, y\rangle| \leq ||x||_2 \cdot ||y||_2$.

1. (a) **(5 marks)** Alice gets an array x and Bob gets an array y as input, where x, y contain elements from $\{1, \dots, n\}$. Moreover, both of them are given a positive integer $\tau \in \{1, \dots, n\}$. The goal for them is to determine whether there is an element that has at least τ occurrences in $x \circ y$, where $\circ$ denotes concatenation. Show an $\Omega(n)$ lower bound on the randomized public coin communication complexity of this problem.

   (b) **(5 marks)** Let $\varepsilon \in (0, 1)$. An element i in an array of length n is called an *ε-heavy hitter* if i occurs at least εn times in the array. In the setting above, Alice and Bob aim to detect all ε-heavy hitters in $x \circ y$. Give a deterministic protocol for this problem that uses $O(\frac{1}{\varepsilon} \log n)$ bits of communication.

2. **(15 marks)** Alice and Bob are given arrays x and y containing elements from $\{0, \dots, n-1\}$, where n is a power of 2. There is an additional promise that the concatenation of x and y contains either all the elements of $\{0, \dots, n-1\}$ exactly once or an element j is replaced by i, thereby resulting in an array where every element in $\{0, \dots, n-1\} \setminus \{i, j\}$ appears exactly once, i appears twice, and j does not appear at all. The goal for Alice and Bob is to determine which of the two cases holds for given x, y. Give a $O(\log n)$-bit protocol for this problem.

3. Consider the greater than function GT: $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined as $\text{GT}(x, y) = 1$ if and only if $x > y$.

   (a) **(5 marks)** Prove that the nondeterministic communication complexity of GT is $\Omega(n)$.

   (b) **(15 marks)** Prove that the randomized public coin communication complexity of GT is $O(\log n \log \log n)$.

4. In the lecture where we proved that $\Omega(k \log(n/k))$ linear measurements are necessary for the sparse recovery problem, we made an assumption that each entry in the measurement matrix $A \in \mathbb{R}^{m \times n}$ can be represented using $O(\log n)$ bits. In this question, we work towards showing that the assumption is without loss of generality.

   Consider a matrix $A \in \mathbb{R}^{m \times n}$ with *orthonormal rows.* Let B be obtained after rounding entries of A to b bits each. Let $C = A - B$.

   (a) **(5 marks)** Prove that each entry of C has absolute value at most 2^{-b}.

   (b) **(5 marks)** For $v \in \mathbb{R}^n$, let $s = A^T C v$. Argue that $Bv = A(v - s)$. That is, taking a linear measurement on v with the rounded-down matrix B is identical to taking a linear measurement on a perturbed vector $v - s$ with A.

   (c) **(10 marks)** Prove that $||s||_1 \leq \frac{n^3}{2^b} \cdot ||v||_1$.

5. Consider the function INTER: $\{0,1\}^n \times \{0,1\}^n \to \{0, \ldots, n-1\}$ defined as $\text{INTER}(x, y) = |\{j : x_j = y_j = 1\}|$.

   (a) **(5 marks)** Prove that $D(\text{INTER})$ is $\Omega(n)$.

   (b) **(10 marks)** Prove that rank of the matrix M_{INTER} is at most n.

   (c) **(5 marks)** Do the above statements violate the log-rank conjecture? Justify.

6. A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is linear if there exists a set $S \subseteq [n]$ such that $f(x) = \sum_{i \in S} x_i \mod 2$ for all $x \in \mathbb{F}_2^n$, where x_i denotes the i-th bit of x. Since there is a bijection between the set $2^{[n]}$ and the set of linear functions, we denote the linear function corresponding to the set $S \subseteq [n]$ as χ_S. The function χ_S is a k-parity function for $k \leq n$ if $|S| = k$. An ε-tester for k-parity gets inputs $k, n \in \mathbb{N}, \varepsilon \in (0, 1)$ and oracle access to a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$. It has to accept, with probability at least $2/3$, if f is k-parity and reject, with probability at least $2/3$, if f is ε-far from every k-parity function. Recall that a function f is ε-far from k-parity if f differs from every k-parity function on at least $\varepsilon \cdot 2^n$ values.

   (a) **(5 marks)** Prove that for $k' \neq k$ a k-parity function f is $\frac{1}{2}$-far from a k'-parity function g.

   (b) **(10 marks)** Consider sets $S, T \subseteq [n]$ such that $|S| = |T| = k$. Let h denote the function $(\chi_S + \chi_T) \mod 2$. Show that:

   - if $S \cap T = \emptyset$, then h is a $2k$-parity function.
   - if $S \cap T \neq \emptyset$, then h has parity at most $2k - 2$.

   (c) **(10 marks)** In the DISJ_k problem, Alice and Bob are given sets $S, T \subseteq [n]$ where $|S| = |T| = k \leq n/2$ and their goal is to determine if $S \cap T = \emptyset$ or not. It is known that the randomized two-sided error public coin communication complexity of DISJ_k is $\Omega(k)$. Using a reduction from this problem, argue that the query complexity of every ε-tester for testing k-parity is $\Omega(k)$ for all $\varepsilon \in (0, 1/2)$. Does your lower bound hold for adaptive testers as well?