Problem Set 1

- Due Date: **21 Sep (Wed), 2011 (TIFR) / 30 Sep (Fri), 2011 (IMSc)**
- If you submit handwritten solutions, start each problem on a fresh page.
- Collaboration is encouraged, but all writeups must be done individually and must include names of all collaborators.
- Referring sources other than the lectures is strongly discouraged. But if you do use an outside source (eg., other text books, lecture notes, any material available online), ACKNOWLEDGE it in your writeup.
- The points for each problem are indicated on the side.
- If you don't know the answer to a problem, then just don't answer it. Do not try to convince yourself or others into believing a false proof.
- Be clear in your writing.

---

1. **[Median Computation]** (10)

   Given sets $x, y \subseteq [n]$, $\mathsf{MED}_n(x,y)$ is defined to be the median of the multiset $x \cup y$ (if $x \cup y$ contains an even number $t = 2k$ (not necessarily distinct) elements, then the median is defined as the k-th smallest element).

   Show that the deterministic communication complexity of MED_n is $O(\log n)$ (i.e., $\mathrm{D}(\mathsf{MED}_n) = O(\log n)$).

   [Hint: First, show that you can assume that x, y have equal number of elements. Having justified this assumption, show that with constant communication, one can either reduce the size of both the sets by a constant fraction or reduce the potential range where the median lies by a constant factor.]

2. **[Rank vs. communication complexity]** (10+10)

   Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a Boolean function and let M_f be the $2^n \times 2^n$ matrix defined as $M_f(x,y) = f(x,y)$.

   (a) Prove that $\mathrm{D}(f) \leq \mathrm{rank}(M_f) + 1$.

   [Hint: If $\mathrm{rank}(M_f) = r$, then there exists a $r \times r$ sub-matrix of M_f that has full rank.]

   (b) Show that $\mathrm{D}(f) = O(\mathrm{N}(f) \cdot \log \mathrm{rank}(M_f))$ (similarly, $\mathrm{D}(f) = O(\mathrm{N}(\bar{f}) \cdot \log \mathrm{rank}(M_f))$).

   [Hint: Recall the (combinatorial) proof of $D(f) = O(\mathrm{N}(f) \cdot \mathrm{N}(\bar{f}))$.]

3. **[Tribes function]** (15+5+5)

   Let n be a square number. View an n bit string x as a $\sqrt{n} \times \sqrt{n}$ matrix with entries x_{ij}. The $\mathsf{TRIBES}_n : \{0.1\}^n \times \{0,1\}^n \to \{0,1\}$ function is defined as

   $$\mathsf{TRIBES}_n(x,y) = \bigwedge_{i=1}^{\sqrt{n}} \bigvee_{j=1}^{\sqrt{n}} x_{ij} \wedge y_{ij}.$$

   Show that $\mathrm{D}(\mathsf{TRIBES}_n) = \Omega(n)$, yet $\mathrm{N}(\mathsf{TRIBES}) = O(\sqrt{n} \log n)$ and $\mathrm{N}(\overline{\mathsf{TRIBES}_n}) = O(\sqrt{n})$. This is another example that shows the relationship given between deterministic and nondeterministic complexity in class is almost tight.

   [Hint: For the deterministic lower bound, use the rank argument.]

4. **[Clique Independent Set]** (10+5)

Given a graph G on n vertices, the Clique Independent Set (CLIS_G) problem is defined as follows: Alice is given a clique C in the graph G and Bob is given an independent set I in G. They need to compute $|C \cap I|$ (note that this quantity is either 1 or 0 1, depending on whether C and I intersect or not).

(a) Show that $\Omega(\log n) \leq \mathrm{D}(\mathsf{CLIS}_G) \leq O(\log^2 n)$.

(b) Show that if $\mathrm{D}(\mathsf{CLIS}_G) = O(\log^c n)$, then $\mathrm{D}(f) = O((\log C^D(f))^c)$, for any function $f : X \times Y \to \{0,1\}$. Here $C^D(f)$ is the smallest number of monochromatic rectangles in a partition (i.e., disjoint cover) of $X \times Y$.

5. **[Balancing a protocol tree]** (7)

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Prove the following statement: If f has a determinstic communication protocol tree with l leaves, then f has a protocol tree with depth at most $O(\log l)$.

[Hint: The following might come useful. Every binary tree with N leaves has a node whose deletion results in trees with at most $2N/3$ leaves. Provide a proof of this fact, in case you are using it]

6. **[Zero-error randomized communication protocols]** (8)

Recall that $\mathrm{R}_0^{\mathrm{pub}}(f)$ denotes the minimum expected communication of a randomized protocol with public coins that computes$f$ with zero error. Show that if f has a fooling set of size t, then

$$\mathrm{R}_0^{\mathrm{pub}}(f) \geq \log t.$$

Hence, conclude that $\mathrm{R}_0^{\mathrm{pub}}(\mathsf{EQ}_n) = \Omega(n)$.

7. **[Duplicates in a stream]** (7+8)

Consider the following FindDuplicate problem.

Input: A stream $y_1, y_2, \ldots, y_{m+1} \in [m]^{m+1}$; for $j \in [m]$, let $f(j) = |\{i \in [m+1] : y_i = j\}|$.

Output: An element $j \in [m]$ such that $f(j) \geq 2$.

Fix a constant $r \geq 1$.

(a) Show that there is an r-pass deterministic streaming algorithm that uses $\tilde{O}(m^{\frac{1}{r}})$ space.

(b) Show that every r-pass deterministic streaming algorithm requires space $m^{1/O(r)}$.

For part (b), you may cite the following result of Boppana/Hastad.

> Any monotone circuit computing the majority function on N bits where ANDs and ORs alternate at most r times, requires depth $N^{1/O(r)}$.