CS Theory (Spring '24)	Assigned: February 19, 2024 (Prob. 2 updated on Feb 20)

Homework 3

Instructor: *Josh Alman* — Due: February 28, 2024 at 11:59 pm

Collaboration on homework is allowed. However, you must write up solutions by yourself and understand everything that you hand in. You may also consult other reference materials, but you may not seek out answers from other sources.

Write the solution to each problem on a separate page. Please do not write your name on any page of the submission; we are using anonymous grading in GradeScope.

Be sure to answer the final question listing your collaborators and other sources used. See the course webpage for more details.

There are 120 total possible points, and also bonus questions worth an additional 20 points.

1 Big-O Notation (10 points; each part 2 points)

Determine whether each of the following statements is true or false. You must justify each your answers with a **short** explanation.

(a) For all functions $f, g : \mathbb{N} \to \mathbb{N}$, if $f(n) = o(g(n))$ then $f(n) + 100 \cdot g(n) = O(g(n))$.

(b) $\log_{10}(n) = O(\log_2(n))$.

(c) $2^n = O(2^{2n})$.

(d) $\log_2 \log_2(n) \cdot \sqrt{n} = \Omega(\log_2(n) \cdot n^{1/4})$.

(e) For all functions $f, g : \mathbb{N} \to \mathbb{N}$, we have $f(n) \cdot g(n) = O(f(n)^2 + g(n)^2)$.

2 Defining Streaming (10 points; 5 points each part)

Recall from homework 2 that a DIA is a DFA $(Q, \Sigma, q_0, F, \delta)$ where the sets Q, F (the set of states and the set of accept states) are allowed to be infinitely large. For every positive integer n, let's furthermore define $Q_n \subseteq Q$ to be the set of states that can be reached by an input string of length $\leq n$.

Recall from class that we defined streaming algorithms and their space usage. In this problem, we'll show that DIAs and Q_n can be used to capture essentially the same notion.

Prove that for every language L over $\Sigma = \{0, 1\}$, the following are true:

(a) If there is a streaming algorithm (using the definition from class) for L that uses $S(n)$ space, then there is a DIA for L such that, for all integers n, we have $|Q_n| = O(2^{S(n)})$.

(b) If there is a DIA for L (with sets Q_n defined above), then there is a streaming algorithm (using the definition from class) which uses space $S(n) = O(\log_2(|Q_n|))$.

3 Streaming and Friends (20 points; each part 10 points)

(a) Prove that if L is a language over $\Sigma = \{0,1\}$, and L has a NFA with k states, then L has a streaming algorithm which uses at most k space.

(b) Prove that if L is a language over $\Sigma = \{0,1\}$, and L has a streaming algorithm which uses $S(n)$ space, then L^* has a streaming algorithm which uses at most $O(2^{S(n)})$ space.

(c) (BONUS, 10 additional points) Prove or disprove: If L is a language over $\Sigma = \{0,1\}$, and L has a streaming algorithm which uses at most $O(\sqrt{\log(n)})$ space, then L is a regular language.

4 Balance (20 points; each part 10 points)

A string $w \in \{0,1\}^*$ is called *balanced* if more than a third of its bits are 1 and more than a third of its bits are 0. Define the language $L = \{w \in \{0,1\}^* \mid w \text{ is balanced}\}$.

(a) Give a streaming algorithm for L which uses at most $O(\log(n))$ space.

(b) Prove that L does not have a streaming algorithm which uses $o(\log(n))$ space.

5 Constant Communication (20 points; each part 10 points)

Let $\Sigma = \{0,1\}$. Recall that for a language L over Σ, there is an associated communication problem $f_L : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$, where $f_L(x,y) = 1$ if $xy \in L$ and $f_L(x,y) = 0$ if $xy \notin L$.

(a) Prove that if L is a regular language, then the communication complexity of f_L is at most a constant (independent of n).

(b) Give (with proof) an example of a language L which is not regular, but for which the communication complexity of f_L is at most a constant.

6 Communication Problems (30 points; each part 10 points)

(a) Define the function $GT : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ by

$$GT(x,y) = \begin{cases} 1, & \text{if } x > y \text{ (when } x \text{ and } y \text{ are interpreted as nonnegative integers written in binary)} \\ 0, & \text{otherwise} \end{cases}$$

Prove that the communication complexity of GT is at least $\Omega(n)$. (Recall from class that n is the length of the inputs to Alice and Bob, i.e., Alice is given $x \in \{0,1\}^n$, Bob is given $y \in \{0,1\}^n$, and their goal is to determine $GT(x,y)$.)

(b) Define the function $ONE : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ by

$$ONE(x,y) = \begin{cases} 1, & \text{if } y \text{ has at most one 1} \\ 0, & \text{otherwise} \end{cases}$$

Prove that the communication complexity of ONE is at most $O(1)$.

(c) Define the function $COMBINE : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ by

$$COMBINE(x,y) = GT(x,y) \cdot ONE(x,y).$$

Prove that the communication complexity of $COMBINE$ is $\Theta(\log n)$.

(d) (BONUS, 10 additional points) Find (with proof) a function $f : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ with the following three properties:

- For all $x, y \in \{0,1\}^*$, if $COMBINE(x,y) = 0$ then $f(x,y) = 0$,
- For all positive integers n,

$$|\{(x,y) \in \Sigma^n \times \Sigma^n \mid f(x,y) = 1\}| \geq \frac{1}{2} \cdot |\{(x,y) \in \Sigma^n \times \Sigma^n \mid COMBINE(x,y) = 1\}|,$$

- The communication complexity of f is $O(1)$.

7 Day-to-day Communication (9 points)

Give an example of a real-life system where a limiting factor is a communication problem that *cannot* be solved with constant communication. You don't need to prove a communication lower bound, but be sure to unambiguously describe what the communication problem is and why it's a limiting factor.

8 Collaborators and References (1 point)

Please list who you collaborated with on each problem, including any TAs or students you discussed the problems with in office hours. Also list any reference materials consulted other than the lectures and textbook for our class.