CSci 4011, Formal Languages and Automata Theory
Homework 7, Fall 2020
Posted: Nov 30, 2020 Due: Dec 14, 2020

Reminder: Assignments are due by 11:59 p.m. on the indicated date.

Please make sure to consult the guidelines for homework submissions posted at

https://z.umn.edu/csci4011submissions

before you start writing up your homework. In addition, make sure to read the part of the syllabus that discusses the requirements of homework submissions to understand the factors that will impact on the grading. Remember that communicating your approach to the solution of a problem—e.g. the key ideas underlying a proof or a construction—is as important as laying out the details, and sometimes may be more important. For this reason, you should typically begin your answer to each problem with a brief discussion of these key ideas before you go into the specific details. This is not a hard-and-fast rule—sometimes the essential idea may be so apparent as to not need any further explanation—but it is a yardstick for you to consider when writing your solutions. Remember that this is a math course and in accessing math work we must look at precision as well as clarity of thought.

Problem 1 *(1 + 2 points for each part, 9 points in all)*

State whether each of the following is true or false and explain why. Note that the explanation carries two-third of the credit.

1. $3^n = o(2^n)$.
2. $(n * \sqrt{5 * n}) + (31 * (log_2 n)^2) = O(n^2)$
3. $n^2 * log_2(n^5) = O(n^3)$

Problem 2 *(6 points)*

We have seen the definition of a k-clique in an undirected graph in the video-recorded lectures. This definition also appears at the bottom of page 295 in the textbook. Let 5-CLIQUE be the following language

$$\text{5-CLIQUE} = \{\langle G \rangle \mid G \text{ is an undirected graph with a 5-clique in it}\}.$$

Prove that 5-CLIQUE is in P.

Problem 3 *(4 + 4 + 6 + 2 points)*

This problem is essentially Problem 7.13 in the textbook. The version presented here structures your progress through that problem, thereby hopefully making it more accessible.

Assume that it is possible to describe a deterministic single-tape Turing machine to calculate $a \bmod p$, the remainder under integer division of a positive number a with respect to another positive number p, in time that is polynomial in the length of the representation of the larger of a and p. Now suppose that we are given four positive numbers, a, b, c and p and we desire to determine if $a^b \bmod p$ is identical to $c \bmod p$.

1. One way to check this would be to first calculate a^b and to compare the mod of this number with respect to p with that of c. Explain why this calculation will not yield a polynomial time algorithm for a (deterministic) Turing machine. (Hint: Take a to be 2 and think of the size of a^b.)
2. Show that the following is true: $((a * b) \bmod p) = (((a \bmod p) * (b \bmod p)) \bmod p)$.
3. Use the previous part to describe another way to calculate $(a^b \bmod p)$ that yields a polynomial time algorithm for a (deterministic) Turing machine. You may assume that multiplication is an operation that can be carried out by such a machine in time that is polynomial in the size of the two numbers being multiplied. (Hint: To keep the size of the numbers small, you should try to integrate the modulus calculation into the exponentiation using the previous part. You may present pseudo-code to describe the computation and then assess the complexity based on that code.)
4. Conclude from this that the language MODEXP in Problem 7.13 in the book is in P.

Problem 4 *((2 + 4) + 4 points)*

A Hamiltonian path in a directed graph G is a directed path that goes through each node in the graph exactly once. On page 292 of the book, Sipser describes the language HAMPATH that corresponds to the problem of determining if there is a Hamiltonian path in a directed graph G from a given node u to a given node v. We can describe another language HAMGRAPH that corresponds to determining if there is a Hamiltonian path between any pair of nodes in a directed graph G as follows:

$$\text{HAMGRAPH} = \{\langle G\rangle \mid G \text{ is a directed graph in which there is a Hamiltonian path between some pair of nodes}\}$$

Show that HAMGRAPH $\leq_P$ HAMPATH. Recall that what you have to do for this is

1. describe a reduction from HAMGRAPH to HAMPATH and prove that it is a reduction, and
2. show that the reduction you have described can be carried out in a number of steps that is polynomial in the size of the input.

Problem 5 *(3 + 3 + 5 + 6 + 4 points)*

Let SET-SPLITTING be the following language:

$$\{\langle S, C\rangle \mid S \text{ is a finite set and, for } k > 0, C = \{C_1, \ldots, C_k\}\text{ , is a collection of subsets of } S \text{ such that the elements of } S \text{ can be colored } \textit{red} \text{ or } \textit{blue} \text{ so that no } C_i \text{ has all its elements colored with the same color.}\}$$

As concrete examples, the pair $\langle\{1,2,3\},\{\{1,2\},\{1,3\}\}\rangle$ is a member SET-SPLITTING but the pair $\langle\{1,2,3\},\{\{1,2\},\{1,3\},\{2,3\}\}\rangle$ is not and neither is $\langle\{1,2,3\},\{\{1\},\{2,3\}\}\rangle$. Notice also that although we have chosen S to be a set of natural numbers in these examples, S can have any kind of objects as its elements.

This problem requires you to show that SET-SPLITTING is NP-complete in a sequence of steps. *Note that you must answer each part separately to get the credit for it.*

1. As a first step, show that SET-SPLITTING $\in NP$ by describing a certificate that can be used to quickly determine membership in the set.

2. Let S be a set of boolean variables $x_1, \ldots, x_n$ and their complements. In other words, $S = \{x_1, \overline{x_1}, \ldots, x_n, \overline{x_n}\}$. Describe a set C such that $\langle S, C \rangle \in$ SET-SPLITTING but that also forces each variable and its complement to be colored differently; thus such a pair allows you to think of the colours as truth values.

3. Given a formula $F = \phi_1 \wedge \ldots \wedge \phi_n$ where each ϕ_i is a clause with exactly 3 literals, describe a way of constructing a pair $\langle S, C \rangle$ whose size is polynomial in the size of F and that belongs to SET-SPLITTING if and only if F is satisfiable.

   Hint: Think of using the variables appearing in F and their complements to construct S. Then map each clause to a suitable element of C. You will also need to add one additional special object, say a, to S and will need to add this to the elements of C in a way that makes it possible to satisfy the coloring restriction when F is satisfiable.

4. Show that what you have described in the previous step is actually a reduction of 3-SAT to SET-SPLITTING. Remember, there are two directions to show here.

5. Argue that the reduction described in step 3 (and proved correct in step 4) can actually be carried out in time polynomial in the size of F, the original input.

Problem 6 *(4 points)*

Show that if we are able to prove that PATH is not NP-complete, then we would have shown that $P \neq NP$. (This is problem 7.20 in the textbook.)