Indian Institute of Technology Kharagpur
Department of Computer Science and Engineering

CS60005: Foundations of Computing Science **Autumn 2023**

Class Test 2 **6th of November 2023, 5:45 PM – 6:45 PM** **Marks = 20**

Answer all questions. State all assumptions you make. Keep your answers concise.

---

1. (a) Construct a context-free grammar generating A^* where A is defined as

$$A = \{x\#x^{\mathbf{R}}\# \mid x \in \{a,b\}^+\}.$$

[$x^{\mathbf{R}}$ denotes the reverse of string x.]

Solution: Let $G = (N = \{S,T\}, \Sigma = \{a,b,\#\}, P, S)$ where P consists of the following production rules.

$$S \to SS \mid \epsilon \mid T\#$$
$$T \to a\#a \mid b\#b \mid aTa \mid bTb$$

It is straightforward to verify that $L(G) = A^*$.

(b) Convert the grammar to Chomsky normal form.

Solution: The only ϵ-production is $S \to \epsilon$ and getting rid of this production results in the same grammar sans the production $S \to \epsilon$. The resulting grammar consists the following productions.

$$S \to SS \mid T\#$$
$$T \to a\#a \mid b\#b \mid aTa \mid bTb$$

Now we introduce three productions $H \to \#$, $A \to a$ and $B \to b$ and replace occurrences of terminals $\#, a, b$ with H, A, B respectively. Resulting set of production rules is given by

$$S \to SS \mid TH$$
$$T \to AHA \mid BHB \mid ATA \mid BTB$$
$$H \to \#$$
$$A \to a$$
$$B \to b$$

Next step is to modify T-productions so that RHS of every production consists of exactly two non-terminals. To this end, we introduce 4 new non-terminals X_A, X_B, U_A, U_B. New set of rules is as follows.

$$S \to SS \mid TH$$
$$T \to AX_A \mid BX_B \mid AU_A \mid BU_B$$
$$X_A \to HA$$
$$X_B \to HB$$
$$U_A \to TA$$
$$U_B \to TB$$
$$H \to \#$$
$$A \to a$$
$$B \to b$$

The above productions define a grammar in Chomsky normal form for the language $A^* \setminus \{\epsilon\}$.

6+2 = 8

2. Consider the language $B = \{a^{n^2} \mid n \geq 1\}$. Is B context-free? If it is, describe a context-free grammar or a pushdown automaton for B. Otherwise, prove that it is not and describe (informally) a Turing machine that decides the language.

Hint: $1 + 3 + 5 + \cdots + (2n-1) = n^2$, for $n \geq 1$. 6

Solution: B is not context-free. We prove this using pumping lemma for CFLs. Suppose that B is a CFL. Let k be the constant guaranteed by pumping lemma. Take $z = a^{k^2}$. Then it is possible to write $z = uvwxy$ such that $vx \neq \epsilon$, $|vwx| \leq k$ and $uv^iwx^iz \in B$ for all $i \geq 0$.

Let $|vx| = \ell$. We know $0 < \ell \leq k$. Consider the string $z' = uv^2wx^2z$. We have $|z'| = |z| + |vx| = k^2 + |vx| = k^2 + \ell$. We have $0 < \ell \leq k < 2k+1$ and so $k^2 < |z'| < k^2 + 2k + 1 = (k+1)^2$. That is, the length of z' cannot be a perfect square and hence $z' \notin B$. This contradicts our assumption that B is a CFL.

We now describe a (total) Turing machine $\mathcal{M}$ that decides B. $\mathcal{M}$ has two tapes. The first tape contains the input string (a string of the form a^k). $\mathcal{M}$'s is to test whether k is a perfect square or not. Let X be a symbol other than $a, \vdash, \sqcup$. The second tape initially contains 1 X followed by blanks. (Both tapes have a left endmarker in the left-most cell.)

$\mathcal{M}$ repeats the following.

- $\mathcal{M}$ advances tape-head 1 (reading one a) to the right and moves tape-head 2 one cell to the right reading an X.
- If both tape-heads have reached blank symbol, then accept and halt.
- If tape-head 1 reaches blank and tape-head 2 is pointing to an X, reject and halt.
- If tape-head 1 is reading a and tape-head 2 is reading blank, then write two more X's on tape-2 at the end of the existing string of X'.
- Move tape-head 2 to the left-end so that it points to the first occurrence of X.

Essentially the machine maintains an odd number of X's on tape 2 – 1 initially, 3 in the second iteration, 5 in the third iteration, and so on. Tape-head 1 never moves backwards; instead it moves one step forward reading a for every occurrence of X on the second tape and this repeats for every odd-length string of X's genrated on tape-2. $\mathcal{M}$ checks if the number of a's in tape 1 is a sum of consecutive odd numbers (starting from 1). If so, the number of a's must be a perfect square, and vice-versa.

3. Identify whether each of the following languages is recursive, *r.e.* but not recursive or not *r.e.* . Justify your answer. In all of the following, $\mathcal{M}$ denotes a Turing machine.

   (a) $\{\mathcal{M} \mid L(\mathcal{M}) \text{ has at least 100 strings}\}$

   Solution: *r.e. but not recursive.*

   Let $T_{\geq 100} = \{\mathcal{M} \mid L(\mathcal{M}) \text{ has at least 100 strings}\}$. The language $T_{\geq 100}$ is *r.e.* since it is possible to design a TM that, given description of a TM $\mathcal{M}$, runs $\mathcal{M}$ on all strings from Σ^* in a round robin fashion; accepts whenever $\mathcal{M}$ accepts 100 strings.

   Let P_1 denote the property on r.e. sets defined as

$$P_1(A) = \begin{cases} \mathsf{T} & \text{if } A \text{ contains atleast 100 strings} \\ \mathsf{F} & \text{otherwise} \end{cases}$$

   Deciding P_1 is equivalent to dedciding membership in $T_{\geq 100}$. We have $P_1(\phi) = \perp$, $P_1(\Sigma^*) = \top$ and ϕ, Σ^* are both *r.e.* sets. So P_1 is a non-trivial property and by Rice's theorem part I, P_1 is not decidable i.e., $T_{\geq 100}$ is not recursive.

   (b) $\{\mathcal{M} \mid \mathcal{M} \text{ has at least 100 states}\}$

   Solution: *Recursive.*

   Check the description of the machine to see if the number of states is ≥ 100.

   (c) $\{\mathcal{M} \mid L(\mathcal{M}) \text{ has at most 100 strings}\}$

   Solution: *Not r.e.*

   Let $T_{\leq 100} = \{\mathcal{M} \mid L(\mathcal{M}) \text{ has at most 100 strings}\}$. Let P_2 denote the property on r.e. sets defined as

$$P_2(A) = \begin{cases} \mathsf{T} & \text{if } A \text{ contains at most 100 strings} \\ \mathsf{F} & \text{otherwise} \end{cases}$$

Deciding P_2 is equivalent to dedciding membership in $T_{\leq 100}$. We have $P_2(\phi) = \top$, $P_2(\Sigma^*) = \bot$ and $\phi \subset \Sigma^*$. So P_2 is a non-monotone property and by Rice's theorem part II, undecidable. Therefore, $T_{\leq 100}$ is not *r.e.*

2+2+2 = 6