Question 11 **(10 marks)**

This question uses the same grammar as the previous question.

Let L be the language generated by this grammar.

(a) Prove by induction that every string of the form $\mathtt{a}^m\mathtt{b}^n$, where $m \geq n-1$ and $n \geq 1$, belongs to L.

We prove this by induction on the length of the string, i.e., on $m+n$.

Inductive Basis:

If the string has length 1, then it just consists of the string $\mathtt{b}$, which is in L by the derivation $S \Rightarrow T \Rightarrow \mathtt{b}$.

Inductive step:

Assume that any string of the form $\mathtt{a}^m\mathtt{b}^n$ with length $< k$, and $m \geq n-1$ and $n \geq 1$, belongs to L, where $k \geq 2$. This is the **Inductive Hypothesis**.

Now suppose we have a string $\mathtt{a}^m\mathtt{b}^n$ of length $k \geq 2$. FIrstly, observe that $m \geq 1$, since if $m = 0$ then $n = 1$ (by $m \geq n-1$ and $n \geq 1$) which implies the string length, k, is only 1 (whereas here we have $k \geq 2$). So we know that the string starts with $\mathtt{a}$ and ends with $\mathtt{b}$.

If the string has only one $\mathtt{b}$ (i.e., $n = 1$), then it has the form $\mathtt{a}^m\mathtt{b}$ (where $m \geq 1$, since the length k is ≥ 2). Consider the shorter string $\mathtt{a}^{m-1}\mathtt{b}$. Since its length is $k-1$, the Inductive Hypothesis can be used. This tells us that this shorter string belongs to L. So there is a derivation

$$S \Rightarrow T \Rightarrow \cdots \Rightarrow \mathtt{a}^{m-1}\mathtt{b}.$$

(Observe that the first step of any derivation of a nonempty string in this grammar must be $S \Rightarrow T$.) Removing the first step, we have a derviation of $\mathtt{a}^{m-1}\mathtt{b}$ from T, i.e.,

$$T \Rightarrow \cdots \Rightarrow \mathtt{a}^{m-1}\mathtt{b}.$$

Prefixing each string in this derivation gives a derivation of $\mathtt{a}^m\mathtt{b}$ from $\mathtt{a}T$:

$$\mathtt{a}T \Rightarrow \cdots \Rightarrow \mathtt{a}^m\mathtt{b}.$$

But we also have a derivation of $\mathtt{a}T$ from S, namely $S \Rightarrow T \Rightarrow \mathtt{a}T$ (using rule (2) then (4)). Putting these together gives a derivation of $\mathtt{a}^m\mathtt{b}$ from S:

$$S \Rightarrow T \Rightarrow \mathtt{a}T \Rightarrow \cdots \Rightarrow \mathtt{a}^m\mathtt{b}.$$

So our original string of length k does indeed belong to L.

It remains to deal with the case where our string has more than one $\mathtt{b}$, i.e., $n \geq 2$.

Since $m \geq n-1$ and $n \geq 2$, we have $m \geq 1$. The string $\mathtt{a}^{m-1}\mathtt{b}^{n-1}$ has length $k-2$, which is $< k$. Also, $m-1 \geq (n-1)-1$ and $n-1 \geq 1$. So the Inductive Hypothesis applies. We deduce that this string belongs to L, so there must be some derivation

$$S \Rightarrow T \Rightarrow \cdots \Rightarrow \mathtt{a}^{m-1}\mathtt{b}^{n-1}.$$

This includes a derivation from T:

$$T \Rightarrow \cdots \Rightarrow \mathtt{a}^{m-1}\mathtt{b}^{n-1}.$$

Adding a at the start, and b at the end, of each string in this derivation gives a new derivation,

$$\mathtt{a}\,T\,\mathtt{b} \Rightarrow \cdots \Rightarrow \mathtt{a}^{m}\mathtt{b}^{n}.$$

We also have the derviation $S \Rightarrow T \Rightarrow \mathtt{a}\,T\,\mathtt{b}$ (using rule (2) then (3)). Putting these derviations together gives a derivation of $\mathtt{a}^{m}\mathtt{b}^{n}$ from S:

$$S \Rightarrow T \Rightarrow \mathtt{a}\,T\,\mathtt{b} \Rightarrow \cdots \Rightarrow \mathtt{a}^{m}\mathtt{b}^{n}.$$

So the string $\mathtt{a}^{m}\mathtt{b}^{n}$ belongs to L.

We have now established this conclusion for any string of the form $\mathtt{a}^{m}\mathtt{b}^{n}$ of length k, with $m \geq n-1$ and $n \geq 1$, belongs to L.

By the Principle of Mathematical Induction, it follows that any string of this form, of any length, belongs to L.

(b) Prove or disprove: some string in L has a derivation, using this grammar, that is neither a leftmost derivation nor a rightmost derivation.

This is not true. At any stage of any derivation in this grammar, there is only one non-terminal. So there is no choice of which nonterminal in the string is to be used for the next production rule. In particular, whichever production rule we use, we will always be applying it to the leftmost nonterminal (since it is the only nonterminal) — and, also to the rightmost nonterminal. So, in fact, every derivation is a leftmost derivation, and also a rightmost derivation.

(There might sometimes be a choice of which production rule to use for L, since there are several different production rules for L. But this is not relevant to the issue at hand.)

Official use only
10

Question 9 **(12 marks)**

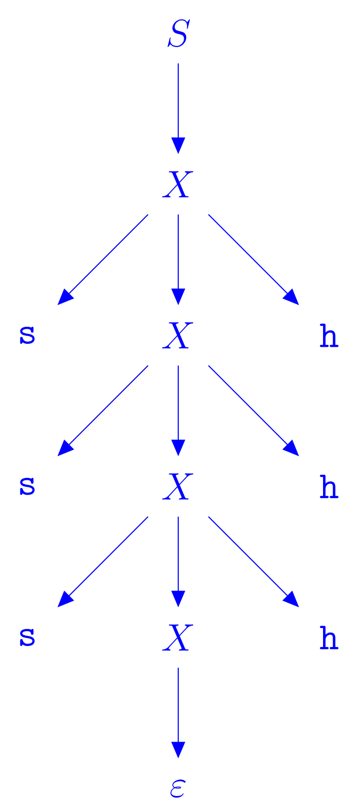
Consider the following Context-Free Grammar:

$$S \rightarrow X \qquad (1)$$
$$X \rightarrow \mathtt{s}X\mathtt{h} \qquad (2)$$
$$X \rightarrow \varepsilon \qquad (3)$$

(a) Give a derivation for the string `ssshhh`.
Each step in your derivation must be labelled, on its right, by the number of the rule used.

$$\begin{aligned} S &\Rightarrow X && (1) \\ &\Rightarrow \mathtt{s}X\mathtt{h} && (2) \\ &\Rightarrow \mathtt{ss}X\mathtt{hh} && (2) \\ &\Rightarrow \mathtt{sss}X\mathtt{hhh} && (2) \\ &\Rightarrow \mathtt{ssshhh} && (3) \end{aligned}$$

(b) Give a parse tree for the same string, `ssshhh`.



(c) Prove by induction on n, that for all $n \geq 0$, the string $\mathtt{s}^n\mathtt{h}^n$ has a derivation in this grammar of $n+2$ steps.

Inductive basis: when $n = 0$, the string is empty. The empty string can be derived from this grammar in two steps: $S \Rightarrow X \Rightarrow \varepsilon$, which is $n+2$ steps with $n = 0$. So the statement is true for $n = 0$.

Now suppose $n \geq 1$, and assume that the statement is true for $n-1$, i.e., any string $\mathtt{s}^{n-1}\mathtt{h}^{n-1}$ has a derivation of length $(n-1)+2 = n+1$ steps. (This is our Inductive Hypothesis.)

This derivation looks like

$$\underbrace{S \Rightarrow X \Rightarrow \cdots \Rightarrow \mathtt{s}^{n-1}\mathtt{h}^{n-1}}_{n+1 \text{ steps}}.$$

Observe here that the first rule *must* be rule (1), since that is the only one that uses S.

Now consider the string $\mathtt{s}^n\mathtt{h}^n$. Since $n \geq 1$, we can write this as:

$$\mathtt{s}^n\mathtt{h}^n = \mathtt{s}\mathtt{s}^{n-1}\mathtt{h}^{n-1}\mathtt{h}$$

Take the derivation $S \Rightarrow X \Rightarrow \cdots \Rightarrow \mathtt{s}^{n-1}\mathtt{h}^{n-1}$ given above, and do the following. First, omit S and the first production, at the very start. Then we have a derivation from X to $\mathtt{s}^{n-1}\mathtt{h}^{n-1}$ in n steps. Then, for each string in this derivation, put an extra $\mathtt{s}$ at the start and an extra $\mathtt{h}$ at the end. This gives us the partial derivation:

$$\underbrace{\mathtt{s}X\mathtt{h} \Rightarrow \cdots \Rightarrow \mathtt{s}\mathtt{s}^{n-1}\mathtt{h}^{n-1}\mathtt{h}}_{n \text{ steps}}$$

(The fact that it's still a valid derivation follows from the context-free property.)

We now prepend this derivation with the two-step derivation of $\mathtt{s}X\mathtt{h}$ from S, which is $S \Rightarrow X \Rightarrow \mathtt{s}X\mathtt{h}$, to give the $n+2$-step derivation of $\mathtt{s}^n\mathtt{h}^n$:

$$\underbrace{S \Rightarrow X \Rightarrow \mathtt{s}X\mathtt{h} \Rightarrow \cdots \Rightarrow \mathtt{s}\mathtt{s}^{n-1}\mathtt{h}^{n-1}\mathtt{h}}_{n+2 \text{ steps}} = \mathtt{s}^n\mathtt{h}^n.$$

This completes the inductive step.

The statement is therefore true for all $n \geq 0$, by the Principle of Mathematical Induction.

Question 3 **(6 marks)**

Let E_n be the following Boolean expression in variables $x_1, x_2, \ldots, x_n$:

$$((\cdots((((\neg x_1 \vee x_2) \wedge \neg x_2) \vee x_3) \wedge \neg x_3) \cdots) \vee x_n) \wedge \neg x_n$$

For example, $E_1 = \neg x_1$, and $E_2 = (\neg x_1 \vee x_2) \wedge \neg x_2$. In general, $E_n = (E_{n-1} \vee x_n) \wedge \neg x_n$.

Prove by induction on n that, for all $n \geq 1$, the expression E_n is satisfiable.

Inductive basis:

If $n = 1$, we have $E_1 = \neg x_1$, which is satisfiable because $x_1 =$ False satisfies it.

Inductive step:

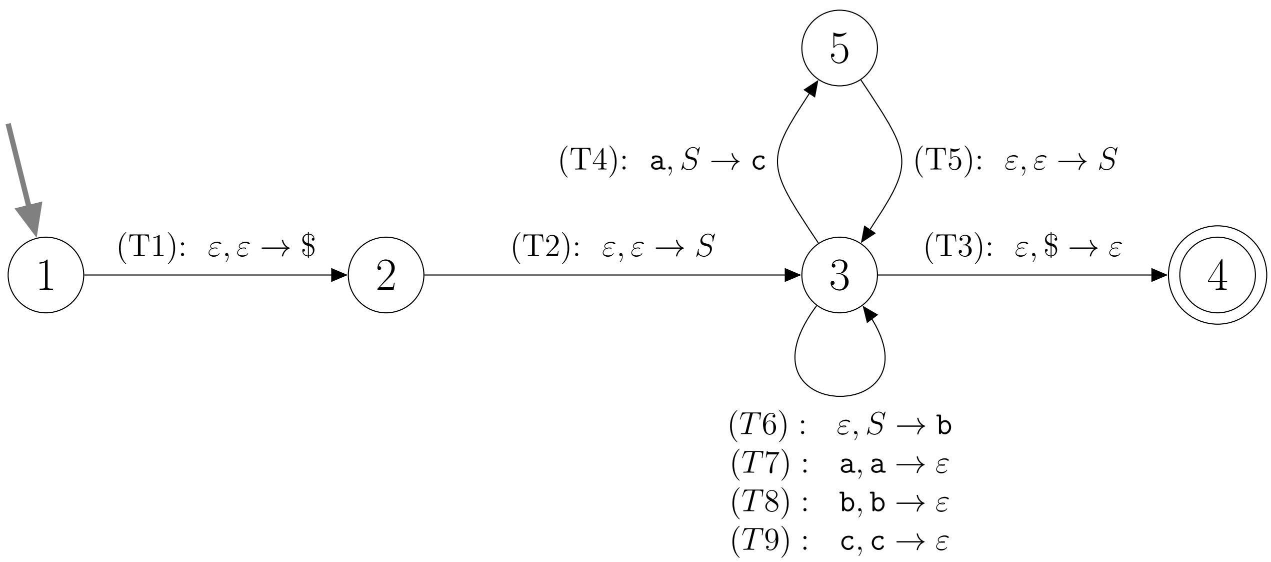
Suppose $n \geq 2$. Assume E_{n-1} is satisfiable (the Inductive Hypothesis). So it has a satisfying truth assignment to its variables, $x_1, x_2, \ldots, x_{n-1}$. Consider $E_n = (E_{n-1} \vee x_n) \wedge \neg x_n$. (Note, this uses $n \geq 2$.) We have a satisfying truth assignment that makes E_{n-1} true. Under this assignment, $E_{n-1} \vee x_n$ is also true, so in this case we have $E_n = (E_{n-1} \vee x_n) \wedge \neg x_n = \text{True} \wedge \neg x_n = \neg x_n$. This can be satisfied by putting $x_n =$ False. So E_n is satisfiable.

Therefore E_n is satisfiable for all n, by Mathematical Induction.

Official use only
6

Question 8 **(12 marks)**

Consider the following Pushdown Automaton (PDA), with input alphabet {a,b,c} and extra stack symbols S and \$.



(a) Show that the single-letter string `b` is accepted by this PDA, by giving the sequence of transitions that leads to acceptance of `b`.

Use the names of the transitions, i.e., $T1, T2, \ldots$, etc.

T1, T2, T6, T8, T3

(b) Prove the following statement by induction on n:

For all $n \geq 0$:
If the PDA is in State 3, the top symbol on the stack is S, and the remaining input begins with $\mathtt{a}^n\mathtt{b}\mathtt{c}^n$, **then** after reading $\mathtt{a}^n\mathtt{b}\mathtt{c}^n$ the PDA is again in State 3 and the stack is the same except that the S on the top has been removed.

Inductive basis: $n = 0$: the remaining input is $\mathtt{b}$, and if S is on top of the stack, the transitions T6, T8 leave us still in state 3, having read the input, and S has been removed from the top of the stack. No other change has been made to the stack.

Inductive step:

Suppose the statement is true for n (the *Inductive Hypothesis*). Suppose the PDA is in state 3, the remaining input starts with $\mathtt{a}^{n+1}\mathtt{b}\mathtt{c}^{n+1}$, and that S is on top of the stack. Then the combined effect of transitions T4, T5 is to read the first $\mathtt{a}$ of input and replace the S on top of the stack by $\mathtt{c}$ and then S, i.e., the only change to the stack is that $\mathtt{c}$ is underneath S; the portion of the stack below this $\mathtt{c}$ is unchanged, and S is still on top of the stack.

So the situation is now that the input begins with $\mathtt{a}^n\mathtt{b}\mathtt{c}^{n+1}$, which is $\mathtt{a}^n\mathtt{b}\mathtt{c}^n\mathtt{c}$, which begins with $\mathtt{a}^n\mathtt{b}\mathtt{c}^n$; also, we are again in state 3, and S is on top of the stack. This is the "if"-part of the given statement, for n. By the Inductive Hypothesis, after reading $\mathtt{a}^n\mathtt{b}\mathtt{c}^n$, the PDA is again in State 3 and the stack is the same except that the S on the top has been removed. The next thing on the stack is the $\mathtt{c}$ we put there before (using T4 and T5). The next unread letter is $\mathtt{c}$. So we can apply T9, which reads that letter and pops the $\mathtt{c}$ off the stack. We are back at state 3 again, and the stack is the same as it was when we began the inductive step except that the S that was then on the top of it is not there any more. This is exactly our desired conclusion, so we have established the given statement for $n + 1$.

By the Principle of Mathematical Induction, the statement holds for all $n \geq 0$.

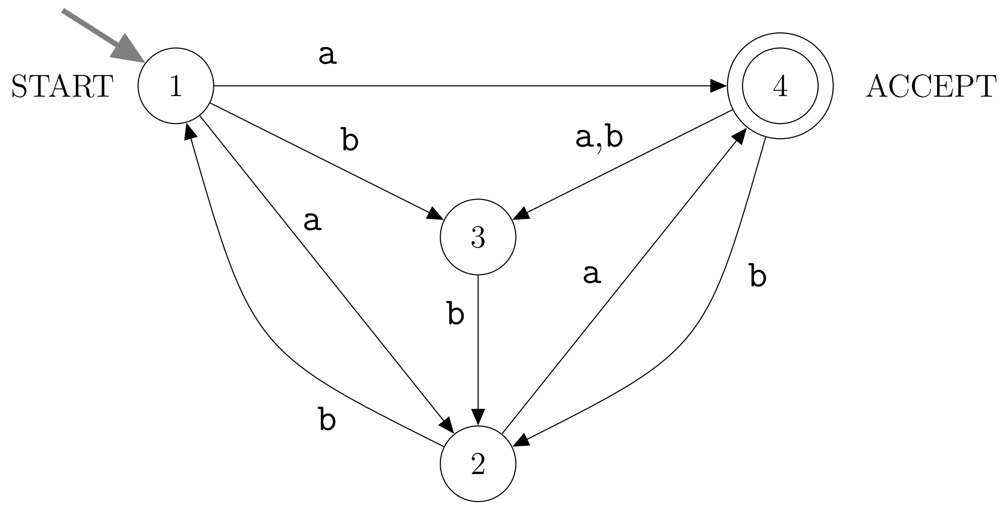
(c) Hence prove that, for all $n \geq 0$, the string $\mathtt{a}^n\mathtt{b}\mathtt{c}^n$ is accepted by this PDA.

When this string is given as input to this PDA, before reading anything it pushes \$ and then S onto the stack. We are then in state 3, with S on top of the stack, and the string to be read begins with (and in fact equals) $\mathtt{a}^n\mathtt{b}\mathtt{c}^n$. By part (b), the PDA reads this string and then is in state 3 with S removed from the top of the stack, and otherwise the stack is unchanged. So all that's left on the stack is the single symbol \$. The PDA now does transition T3. At this stage, the entire input has been read, and the PDA is in its Final state, so the input string is accepted.

Official use only
12

Question 5 **(7 marks)**

Consider the following Nondeterministic Finite Automaton (NFA).



Let L be the language of strings accepted by this NFA.

(a) What are the possible states that this NFA could be in, after reading the input string `abba`?

2, 4

(b) Prove, by induction on n, that for all positive integers n, the string $(\mathtt{abba})^n$ is accepted by this NFA. (This string is obtained by n repetitions of `abba`.)

Base case ($n = 1$): We saw in (a) that `abba` can end up in State 4, which is the Final state. Therefore `abba` is accepted.

Inductive step: suppose $n \geq 2$, and that $(\mathtt{abba})^{n-1}$ is accepted.
Since $(\mathtt{abba})^{n-1}$ is accepted (by Inductive Hypothesis), there is some path it can take through the NFA that ends at the Final State. We also see that, if we are in the Final State, and we then read `abba`, then we can reach the Final State again, by following the path $4 \to 3 \to 2 \to 1 \to 4$. We can put these two paths together, to give a path for the string $(\mathtt{abba})^{n-1}\mathtt{abba}$ that goes from the Initial State to the Final State. Therefore this string is accepted by our NFA. But this string is just $(\mathtt{abba})^n$. So $(\mathtt{abba})^n$ is accepted.

Therefore, by the Principle of Mathematical Induction, the string $(\mathtt{abba})^n$ is accepted for all $n \geq 1$.

Official use only
7

(c) Prove by induction on n, that for all $n \geq 0$, the string $\mathtt{h}^n\mathtt{m}^n$ has a derivation in this grammar of $n+1$ steps.

Inductive basis: when $n = 0$, the string is empty. The empty string can be derived from this grammar in one step: $S \Rightarrow \varepsilon$, which is $n+1$ steps with $n = 0$. So the statement is true for $n = 0$.

Now suppose $n \geq 1$, and assume that the statement is true for $n-1$, i.e., any string $\mathtt{h}^{n-1}\mathtt{m}^{n-1}$ has a derivation of length $(n-1)+1 = n$ steps. (This is our Inductive Hypothesis.)

This derivation looks like

$$\underbrace{S \Longrightarrow \cdots \Longrightarrow \mathtt{h}^{n-1}S\mathtt{m}^{n-1} \stackrel{(3)}{\Longrightarrow} \mathtt{h}^{n-1}\mathtt{m}^{n-1}}_{n \text{ steps}}.$$

Observe here that the last rule *must* be rule (3), since that is the only one without any non-terminal symbols on its right-hand side. So the second-last string must have just a single S. Furthermore, it can be seen from the rules that the only terminal symbol that can appear to the *left* of an S is $\mathtt{h}$, and the only terminal that can appear to the *right* of an S is $\mathtt{m}$. From these observations it follows that the solitary S in the second-last string above must lie exactly in the middle, with an $\mathtt{h}$ on its left and an $\mathtt{m}$ on its right.

Take the derivation $S \Longrightarrow \cdots \Longrightarrow \mathtt{h}^{n-1}S\mathtt{m}^{n-1} \stackrel{(3)}{\Longrightarrow} \mathtt{h}^{n-1}\mathtt{m}^{n-1}$ given above, and do the following. Instead of applying the rule (3) to S in the string $\mathtt{h}^{n-1}S\mathtt{m}^{n-1}$, we apply the rule (1). Then our productions take us from S to $\mathtt{h}^{n-1}\mathtt{h}S\mathtt{m}\mathtt{m}^{n-1}$ in n steps. This string is just $\mathtt{h}^nS\mathtt{m}^n$. Then apply rule (1). This gives us the derivation:

$$\underbrace{S \Longrightarrow \cdots \Longrightarrow \mathtt{h}^{n-1}S\mathtt{m}^{n-1} \stackrel{(3)}{\Longrightarrow} \mathtt{h}^nS\mathtt{m}^n \stackrel{(1)}{\Longrightarrow} \mathtt{h}^n\mathtt{m}^n}_{n+1 \text{ steps}}.$$

This completes the inductive step.

The statement is therefore true for all $n \geq 0$, by the Principle of Mathematical Induction.

An alternative approach to the inductive step is to take the n-step derivation

$$S \Longrightarrow \cdots \Longrightarrow \mathtt{h}^{n-1}\mathtt{m}^{n-1},$$

given by the Inductive Hypothesis, and then put $\mathtt{h}$ in front of every string and $\mathtt{m}$ behind every string. This gives the n-step partial derivation

$$\mathtt{h}S\mathtt{m} \Longrightarrow \cdots \Longrightarrow \mathtt{h}\mathtt{h}^{n-1}\mathtt{m}^{n-1}\mathtt{m} = \mathtt{h}^n\mathtt{m}^n.$$

Then put the production $S \stackrel{(2)}{\Longrightarrow} \mathtt{h}S\mathtt{m}$ at the start of this partial derivation, giving the $n+1$-step derivation

$$S \stackrel{(2)}{\Longrightarrow} \mathtt{h}S\mathtt{m} \Longrightarrow \cdots \Longrightarrow \mathtt{h}^n\mathtt{m}^n.$$

Official use only
13