Theory of Computation 7
Deterministic Membership Testing

Frank Stephan

Department of Computer Science

Department of Mathematics

National University of Singapore

fstephan@comp.nus.edu.sg

Repetition 1: Normal Forms

The normal forms are named after Noam Chomsky (born 1928) and Sheila Greibach (born 1939).

Assume that the language does not contain ε.

Chomsky Normal Form

All rules are of the form $\mathbf{A \to BC}$ or $\mathbf{A \to d}$ where $\mathbf{A, B, C}$ are non-terminals and $\mathbf{d}$ is a terminal.

Greibach Normal Form

All rules are of the form $\mathbf{A \to bw}$ where $\mathbf{b}$ is a terminal and $\mathbf{A}$ a non-terminal and $\mathbf{w}$ a (possibly empty) word of non-terminals.

If the language contains ε, one allows in both normal forms $\mathbf{S} \to \varepsilon$ for the start symbol $\mathbf{S}$ which then is not allowed to appear on the right side of a rule.

Repetition 2: Algorithms

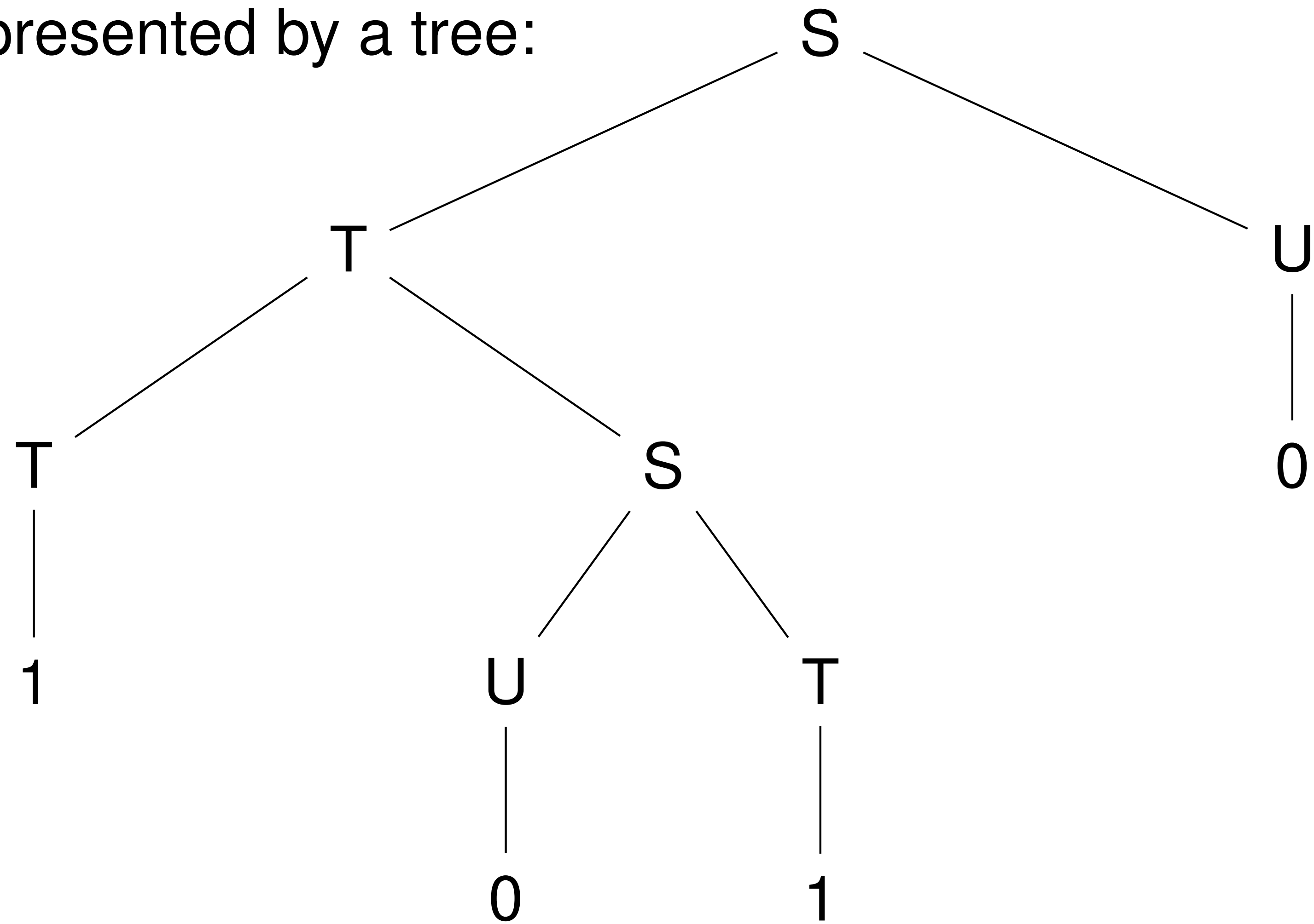
Given a context-free grammar as input, algorithms for the following tasks where presented:

- Converting the grammar into Chomsky Normal Form;
- Testing whether the grammar generates a word;
- Testing whether the grammar generates infinitely many words.

All these algorithms run in time polynomial in the size of the grammar; for grammars in Chomsky Normal Form, the number of non-terminals can serve as the size.

Repetition 3: Derivation Tree

For grammar $(\{\mathbf{S}, \mathbf{T}, \mathbf{U}\}, \{\mathbf{0}, \mathbf{1}\}, \{\mathbf{S} \rightarrow \mathbf{SS}|\mathbf{TU}|\mathbf{UT}, \mathbf{U} \rightarrow \mathbf{0}|\mathbf{US}|\mathbf{SU}, \mathbf{T} \rightarrow \mathbf{1}|\mathbf{TS}|\mathbf{ST}\}, \mathbf{S})$, a derivation $\mathbf{S} \Rightarrow \mathbf{TU} \Rightarrow \mathbf{TSU} \Rightarrow \mathbf{TUTU} \Rightarrow \mathbf{1UTU} \Rightarrow \mathbf{10TU} \Rightarrow \mathbf{101U} \Rightarrow \mathbf{1010}$ can be represented by a tree:



Repetition 4

Marked Symbols

Mark the first four $\mathbf{1}$ in word $\mathbf{0000011111}$ as $\mathbf{0000011111}$. The word can be pumped such that at least one but at most four marked symbols are pumped or between the pumped parts: $\mathbf{000\,0^\ell\,01\,1^\ell\,111}$.

Ogden's Lemma

Let $\mathbf{L} \subseteq \mathbf{\Sigma}^*$ be an infinite context-free language generated by a grammar $(\mathbf{N}, \mathbf{\Sigma}, \mathbf{P}, \mathbf{S})$ in Chomsky Normal Form with $\mathbf{h}$ non-terminals. Then the constant $\mathbf{k} = \mathbf{2^{h+1}}$ satisfies that for every $\mathbf{u} \in \mathbf{L}$ with at least $\mathbf{k}$ marked symbols, there is a representation $\mathbf{vwxyz} = \mathbf{u}$ such that $\mathbf{wxy}$ contains at most $\mathbf{k}$ marked symbols, $\mathbf{wy}$ contains at least $\mathbf{1}$ marked symbol and $\mathbf{vw^\ell xy^\ell z} \in \mathbf{L}$ for all $\ell \in \mathbb{N}$.

Membership Testing

For a language $\mathbf{L}$ and a word $\mathbf{w}$ of length $\mathbf{n}$, one wants to decide whether $\mathbf{w} \in \mathbf{L}$. The following will be shown:

Regular language: Done by a finite automaton, time $\mathbf{O(n)}$.

Linear language: Special case of Cocke, Kasami and Younger's algorithm, time $\mathbf{O(n^2)}$.

Context-free language: Cocke, Kasami and Younger's algorithm, time $\mathbf{O(n^3)}$.

Context-sensitive language: Savitch's algorithm, space $\mathbf{O(n^2)}$, time $\mathbf{O(c^{n^2})}$ for some $\mathbf{c}$.

Recursively enumerable language: No algorithm (undecidable).

Cocke, Kasami and Younger

Let (N, Σ, P, S) be in Chomsky Normal Form and $a_1 a_2 \ldots a_n$ be the input word.

1. Initialisation: For all k,

$$E_{k,k} = \{A \in N : A \to a_k \text{ is a rule}\}.$$

2. Loop: Go through all pairs (i, j) such that they are processed in increasing order of $j - i$ and let

$$\begin{aligned} E_{i,j} = \{ & A : \exists \text{ rule } A \to BC \, \exists k \\ & [i \leq k < j \text{ and } B \in E_{i,k} \text{ and } C \in E_{k+1,j}]\}. \end{aligned}$$

3. Decision: Word is generated by the grammar iff $S \in E_{1,n}$.

Set $E_{i,j}$ contains all non-terminals generating $a_i \ldots a_j$.
Time $O(n^3)$: $O(n^2)$ values $E_{i,j}$ with $O(n)$ choices of k.

Example 7.2, Word 0011

Grammar $(\{\mathbf{S},\mathbf{T},\mathbf{U}\},\{\mathbf{0},\mathbf{1}\},\{\mathbf{S}\rightarrow\mathbf{SS}|\mathbf{TU}|\mathbf{UT},$
$\mathbf{U}\rightarrow\mathbf{0}|\mathbf{US}|\mathbf{SU},\mathbf{T}\rightarrow\mathbf{1}|\mathbf{TS}|\mathbf{ST}\},\mathbf{S})$. Entries $\mathbf{E_{i,j}}$:

$$E_{1,4}=\{S\}$$

$$E_{1,3}=\{U\} \qquad E_{2,4}=\{T\}$$

$$E_{1,2}=\emptyset \qquad E_{2,3}=\{S\} \qquad E_{3,4}=\emptyset$$

$$E_{1,1}=\{U\} \quad E_{2,2}=\{U\} \quad E_{3,3}=\{T\} \quad E_{4,4}=\{T\}$$

$$0 \qquad\qquad 0 \qquad\qquad 1 \qquad\qquad 1$$

As $\mathbf{S}\in\mathbf{E_{1,4}}$, the word $\mathbf{0011}$ is in the language.

Example 7.2, Word 0111

Grammar $(\{\mathbf{S}, \mathbf{T}, \mathbf{U}\}, \{\mathbf{0}, \mathbf{1}\}, \{\mathbf{S} \rightarrow \mathbf{SS}|\mathbf{TU}|\mathbf{UT}, \mathbf{U} \rightarrow \mathbf{0}|\mathbf{US}|\mathbf{SU}, \mathbf{T} \rightarrow \mathbf{1}|\mathbf{TS}|\mathbf{ST}\}, \mathbf{S})$. Entries $\mathbf{E_{i,j}}$:

$$
\begin{array}{ccccccc}
 & & & E_{1,4} = \emptyset & & & \\
 & & E_{1,3} = \{T\} & & E_{2,4} = \emptyset & & \\
 & E_{1,2} = \{S\} & & E_{2,3} = \emptyset & & E_{3,4} = \emptyset & \\
E_{1,1} = \{U\} & & E_{2,2} = \{T\} & & E_{3,3} = \{T\} & & E_{4,4} = \{T\} \\
0 & & 1 & & 1 & & 1
\end{array}
$$

As $\mathbf{S} \notin \mathbf{E_{1,4}}$, the word $\mathbf{0111}$ is not in the language.

Quiz 7.3

Make the table for word $\mathbf{1001}$.

Cocke, Kasami and Younger

Exercise 7.4

Consider the grammar $(\{\mathbf{S}, \mathbf{T}, \mathbf{U}, \mathbf{V}, \mathbf{W}\}, \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}, \mathbf{P}, \mathbf{S})$ with $\mathbf{P}$ consisting of the rules $\mathbf{S} \rightarrow \mathbf{TT}$, $\mathbf{T} \rightarrow \mathbf{UU}|\mathbf{VV}|\mathbf{WW}$, $\mathbf{U} \rightarrow \mathbf{VW}|\mathbf{WV}|\mathbf{VV}|\mathbf{WW}$, $\mathbf{V} \rightarrow \mathbf{0}$, $\mathbf{W} \rightarrow \mathbf{1}$. Make the entries of the Algorithm of Cocke, Kasami and Younger for the words $\mathbf{0011}$, $\mathbf{1100}$ and $\mathbf{0101}$.

Exercise 7.5

Consider the grammar $(\{\mathbf{S}, \mathbf{T}, \mathbf{U}, \mathbf{V}, \mathbf{W}\}, \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}, \mathbf{P}, \mathbf{S})$ with $\mathbf{P}$ consisting of the rules $\mathbf{S} \rightarrow \mathbf{ST}|\mathbf{0}|\mathbf{1}$, $\mathbf{T} \rightarrow \mathbf{TU}|\mathbf{1}$, $\mathbf{U} \rightarrow \mathbf{UV}|\mathbf{0}$, $\mathbf{V} \rightarrow \mathbf{VW}|\mathbf{1}$, $\mathbf{W} \rightarrow \mathbf{0}$. Make the entries of the Algorithm of Cocke, Kasami and Younger for the word $\mathbf{001101}$.

Linear Grammars

Linear languages sit between regular and context-free languages.

Definition

A grammar is linear iff on every left side of a rule exactly one non-terminal and on every right side of a rule at most one non-terminal and arbitrary many terminals.

Normal Form

A linear grammar for a language not containing ε is in normal form iff all rules are of one of the following forms: $\mathbf{A} \rightarrow \mathbf{Bc}$, $\mathbf{A} \rightarrow \mathbf{cB}$, $\mathbf{A} \rightarrow \mathbf{c}$ where $\mathbf{A}, \mathbf{B}$ are non-terminals and $\mathbf{c}$ is a terminal.
If ε is in the language then one has the rule $\mathbf{S} \rightarrow \varepsilon$ for the start symbol $\mathbf{S}$ and $\mathbf{S}$ does not appear on any right side of a rule.

Parsing Linear Grammars

Let (N, Σ, P, S) be a linear grammar in normal form and $a_1 a_2 \ldots a_n$ be an input word.

1. **Initialisation:** For all k,

$$E_{k,k} = \{A \in N : A \to a_k \text{ is in } P\}.$$

2. **Loop:** Process all pairs (i, j) with $i < j$ in increasing order of $j - i$ and let

$$E_{i,j} = \{A : \begin{array}{ll} \exists \text{ rule } A \to Bc & [B \in E_{i,j-1} \text{ and } c = a_j] \text{ or} \\ \exists \text{ rule } A \to cB & [c = a_i \text{ and } B \in E_{i+1,j}]\}. \end{array}$$

3. **Decision:** The word is generated by the grammar iff $S \in E_{1,n}$.

Example 7.8, Word 0110

The grammar

$$(\{\mathbf{S},\mathbf{T},\mathbf{U}\},\{\mathbf{0},\mathbf{1}\},\{\mathbf{S}\rightarrow\mathbf{0}|\mathbf{1}|\mathbf{0T}|\mathbf{1U},\mathbf{T}\rightarrow\mathbf{S0}|\mathbf{0},\mathbf{U}\rightarrow\mathbf{S1}|\mathbf{1}\},\mathbf{S})$$

is a linear grammar for palindromes. For the word $\mathbf{0110}$, the entries are

$$
\begin{array}{cccc}
 & & E_{1,4}=\{S\} & \\
 & E_{1,3}=\emptyset & & E_{2,4}=\{T\} \\
E_{1,2}=\{U\} & & E_{2,3}=\{S,U\} & E_{3,4}=\{T\} \\
E_{1,1}=\{S,T\} & E_{2,2}=\{S,U\} & E_{3,3}=\{S,U\} & E_{4,4}=\{S,T\} \\
0 & 1 & 1 & 0
\end{array}
$$

and as $\mathbf{S}\in\mathbf{E}_{\mathbf{1},\mathbf{4}}$, the word is accepted.

Example 7.8, Word 1110

For processing the word 1110, one gets the following table:

$$
\begin{array}{cccc}
 & & & E_{1,4} = \{T\} \\
 & & E_{1,3} = \{S,U\} & E_{2,4} = \{T\} \\
 & E_{1,2} = \{S,U\} & E_{2,3} = \{S,U\} & E_{3,4} = \{T\} \\
E_{1,1} = \{S,U\} & E_{2,2} = \{S,U\} & E_{3,3} = \{S,U\} & E_{4,4} = \{S,T\} \\
1 & 1 & 1 & 0
\end{array}
$$

Exercise 7.9

Make the entries for the word 0110110 for above grammar.

Exercises 7.10 and 7.11

Exercise 7.10
Consider the following linear grammar:

$$(\{S, T, U\}, \{0, 1\}, \{S \to 0T|T0|0U|U0,\ T \to 0T00|1,\ U \to 00U0|1\}, S).$$

Convert the grammar into the linear grammar normal form and determine the $E_{i,j}$ for input 00100.

Exercise 7.11
Which two of the following languages are linear? Provide linear grammars for these two languages:

- $L = \{0^n 1^m 2^k : n + k = m\}$;
- $H = \{0^n 1^m 2^k : n + m = k\}$;
- $K = \{0^n 1^m 2^k : n \neq m \text{ or } m \neq k\}$.

Kleene Star

Algorithm 7.12

Let $\mathbf{L}$ be generated by a linear grammar and $\mathbf{a_1 a_2 \ldots a_n}$ be a word. To check whether $\mathbf{a_1 a_2 \ldots a_n \in L^*}$, do the following:

First Part: Compute for each $\mathbf{i, j}$ with $\mathbf{1 \leq i \leq j \leq n}$ the set $\mathbf{E_{i,j}}$ of all non-terminals which generate $\mathbf{a_i a_{i+1} \ldots a_j}$.

Initialise Loop for Kleene Star: Let $\mathbf{F_0 = 1}$.

Loop for Kleene Star: For $\mathbf{m = 1, 2, \ldots, n}$ Do
Begin If there is a $\mathbf{k < m}$ with $\mathbf{S \in E_{k+1,m}}$ and $\mathbf{F_k = 1}$
Then let $\mathbf{F_m = 1}$ Else let $\mathbf{F_m = 0}$ End.

Decision: $\mathbf{w \in L^*}$ iff $\mathbf{F_n = 1}$.

First Part is $\mathbf{O(n^2)}$, Loop for Kleene Star is $\mathbf{O(n^2)}$.

Other Combinations

Assume that $\mathbf{H}, \mathbf{K}, \mathbf{L}$ are linear languages and that one has computed for $\mathbf{a_1 \ldots a_n}$ the entries $\mathbf{E^L_{i,j}}, \mathbf{E^H_{i,j}}, \mathbf{E^K_{i,j}}$ say whether the word $\mathbf{a_i \ldots a_j}$ is in $\mathbf{L}, \mathbf{H}, \mathbf{K}$, respectively. This information can be computed in $\mathbf{O(n^2)}$. Complete the algorithm to do the following check in $\mathbf{O(n^2)}$ time:

Exercise 7.13
Is $\mathbf{a_1 a_2 \ldots a_n \in L \cdot H \cdot K}$?

Exercise 7.14
Is $\mathbf{a_1 a_2 \ldots a_n \in (L \cap H)^* \cdot K}$?

Regular Closure of CTF

A language $\mathbf{H}$ is said to be in the regular closure of the context-free languages iff it is obtained by combining finitely many context-free languages with intersection, union, concatenation, set-difference, Kleene Star and Kleene Plus. An example is

$$\mathbf{H} = (\mathbf{L_1} \cap \mathbf{L_2})^* \cdot \mathbf{L_3} \cdot (\mathbf{L_1} \cap \mathbf{L_2})^* - \mathbf{L_4}$$

and here $\mathbf{L_1}, \mathbf{L_2}, \mathbf{L_3}, \mathbf{L_4}$ are context-free.

Exercise 7.15

Prove by structural induction that every language $\mathbf{H}$ which is in the regular closure of the context-free languages has an $\mathbf{O(n^3)}$ decision algorithm.

Counting Derivation Trees

One can modify the algorithm of Cocke, Kasami and Younger to count derivation trees.

Given grammar in CNF (N, Σ, P, S), a word $w = a_1 a_2 \ldots a_n$ and all $A \in N$, let $E_{i,j}$ denote the set of all nonterminals which generate the characters $a_i \ldots a_j$ and $D_{i,j,A}$ denote the number of derivation trees which can, with root A, derive a word $a_i \ldots a_j$. Let P be the set of rules. For each $A \in N$, if $A \in E_{i,i}$, there is exactly one tree with $A \to a_i$ and so $D_{i,i,A} = 1$ else $D_{i,i,A} = 0$. If $i < j$ then

$$D_{i,j,A} = \sum_{(B,C):\ A \to BC \in P} \ \sum_{k:\ i \leq k < j} D_{i,k,B} \cdot D_{k+1,j,C}$$

and the overall number of derivation trees is $D_{1,n,S}$.

Exercises 7.17 - 7.20

Let $\mathbf{P}$ contain the rules $\mathbf{V} \to \mathbf{VV}|\mathbf{WW}|\mathbf{0}$ and $\mathbf{W} \to \mathbf{VW}|\mathbf{WV}|\mathbf{1}$. Consider the grammars $\mathbf{G} = (\{\mathbf{V},\mathbf{W}\},\{\mathbf{0},\mathbf{1}\},\mathbf{P},\mathbf{W})$ and $\mathbf{H} = (\{\mathbf{U},\mathbf{V},\mathbf{W}\},\{\mathbf{0},\mathbf{1},\mathbf{2}\},\mathbf{P} \cup \{\mathbf{U} \to \mathbf{VU}|\mathbf{UV}|\mathbf{2}\},\mathbf{U})$.

Exercise 7.17
How many derivation trees has $\mathbf{0011100}$ in $\mathbf{G}$?

Exercise 7.18
How many derivation trees has $\mathbf{0000111}$ in $\mathbf{G}$?

Exercise 7.19
How many derivation trees has $\mathbf{021111}$ in $\mathbf{H}$?

Exercise 7.20
How many derivation trees has $\mathbf{010012}$ in $\mathbf{H}$?

Polynomial Space

An algorithm can be measured by

- The time needed to do the computation;
- The space (size of variables and arrays and ...) needed to do the computation.

Let $\mathbf{n}$ be a parameter to measure the size of the input.

If one can be computed in time $\mathbf{F(n)}$ then, under certain assumptions to the machine model, it can also be done in space $\mathbf{F(n)}$.

However, for the converse, only a rough estimate is known: If something can be computed in space $\mathbf{F(n)}$ then one can compute it in time $\mathbf{2^{F(n)}}$.

Theorem of Savitch

Algorithm 7.21
Context-senstive grammar (N, Σ, P, S), input word w.

Recursive Call: Function $Check(u, v, t)$
Begin If $u = v$ or $u \Rightarrow v$ Then Return(1);
If $t \leq 1$ and $u \neq v$ and $u \not\Rightarrow v$ Then Return(0);
Let $t' = t/2$; Let $r' = 0$;
For all $u' \in (N \cup \Sigma)^*$ with $|u| \leq |u'| \leq |v|$ Do
Begin If $Check(u, u', t') = 1$ and $Check(u', v, t') = 1$
Then $r' = 1$ End; Return(r') End.

Decision: If $Check(S, w, k^n) = 1$ Then $w \in L$ Else $w \notin L$.

Space Complexity, per call $O(n)$, in total $O(n^2)$;
Value of t: $k^n/2^h$ in depth h of recursion $(k = |\Sigma| + |N| + 1)$;
Number of nested calls: $O(\log(k^n)) = O(\log(k) \cdot n)$.
Runtime: $O(c^{n^2})$ for any $c > (2k)^{\log(k)}$.

Example

Check whether $\mathbf{S} \Rightarrow^* \mathbf{w}$ within $\mathbf{8}$ steps. One call $\mathbf{Check}(\mathbf{S}, \mathbf{w}, \mathbf{8})$.

Is there a word $\mathbf{v}$ such that $\mathbf{S} \Rightarrow^* \mathbf{v}$ and $\mathbf{v} \Rightarrow^* \mathbf{w}$ both within $\mathbf{4}$ steps? For each $\mathbf{v}$, two calls $\mathbf{Check}(\alpha, \beta, \mathbf{4})$ one after each other with $(\alpha, \beta) = (\mathbf{S}, \mathbf{v}), (\mathbf{v}, \mathbf{w})$.

Is there a word $\mathbf{v}'$ such that $\alpha \Rightarrow^* \mathbf{v}'$ and $\mathbf{v}' \Rightarrow^* \beta$ both within $\mathbf{2}$ steps? For each word $\mathbf{v}'$, two calls $\mathbf{Check}(\alpha', \beta', \mathbf{2})$ one after each other with $(\alpha', \beta') = (\alpha, \mathbf{v}'), (\mathbf{v}', \beta')$.

Is there a word $\mathbf{v}''$ such that $\alpha' \Rightarrow^* \mathbf{v}''$ and $\mathbf{v}'' \Rightarrow^* \beta'$ both within $\mathbf{1}$ steps? For each word $\mathbf{v}''$, two calls $\mathbf{Check}(\alpha'', \beta'', \mathbf{1})$ one after each other with $(\alpha'', \beta'') = (\alpha', \mathbf{v}''), (\mathbf{v}'', \beta')$.

Is $\alpha'' = \beta''$ or $\alpha'' \Rightarrow \beta''$ true? No further call needed, bottom of recursion reached.

Example 7.22

Consider grammar $(\{S,T,U,V,W\},\{0,1\},P,S)$ with rules P being the following:

$S \to 0S|U$, $U \to V|0$, $0V \to 1U$, $V \to 1$, $1V \to WU$, $1W \to W0$, $0W \to 10$.

This gives a binary counter with $S \to 0S$ generating enough digits before doing $S \to U$. U stands for last digit 0, V stands for last digit 1, W stands for a digit 0 still having a carry bit to pass on.

Deriving a binary number k needs at least k steps; as the length n of k is logarithmic in k, the derivation length can be exponential in n. In particular deriving 1^n needs more than 2^n steps.

Exercises 7.23 and 7.24

Exercise 7.23

Give a proof that there are $\mathbf{k^n}$ or less words of length up to $\mathbf{n}$ over the alphabet $\mathbf{\Sigma \cup N}$ with $\mathbf{k-1}$ symbols.

Exercise 7.24

Modify Savitch's Algorithm such that it computes the length of the shortest derivation of a word $\mathbf{w}$ in the context-sensitive grammar, provided that such derivation exists. If it does not exist, the algorithm should return the special value ∞.

Naive Algorithm

Exercise 7.25
What is the time and space complexity of this naive algorithm?

Recursive Call: Function $\mathbf{Check(u, w, t)}$
Begin If $\mathbf{u = w}$ or $\mathbf{u \Rightarrow w}$ Then Return$\mathbf{(1)}$;
If $\mathbf{t \leq 1}$ and $\mathbf{u \neq v}$ and $\mathbf{u \not\Rightarrow w}$ Then Return$\mathbf{(0)}$;
Let $\mathbf{r' = 0}$;
For all $\mathbf{v \in (N \cup \Sigma)^*}$ with $\mathbf{u \Rightarrow v}$ and $\mathbf{|v| \leq |w|}$ Do
Begin If $\mathbf{Check(v, w, t-1) = 1}$ Then $\mathbf{r' = 1}$ End;
Return$\mathbf{(r')}$ End;

Decision: If $\mathbf{Check(S, w, k^n) = 1}$ Then $\mathbf{w \in L}$ Else $\mathbf{w \notin L}$.

Growing Grammars

Definition [Dahlhaus and Warmuth 1986]
A grammar $(\mathbf{N}, \mathbf{\Sigma}, \mathbf{P}, \mathbf{S})$ is growing context-sensitive iff $|\mathbf{l}| < |\mathbf{r}|$ for all rules $\mathbf{l} \to \mathbf{r}$ in the grammar.

Theorem [Dahlhaus and Warmuth 1986]
Given a growing context-senstive grammar there is a polynomial time algorithm which decides membership of the language generated by this growing grammar.

In this result, polynomial time means here only with respect to the words in the language, the dependence on the size of the grammar is not polynomial time. So if one asks the uniform decision problem for an input consisting of a pair of a grammar and a word, no polynomial time algorithm is known for this problem. As the problem is NP-complete, the algorithm is unlikely to exist.

Example 7.28

Consider the grammar

$$(\{S, T, U\}, \{0, 1\}, \{S \rightarrow 011|T11, T \rightarrow T0U|00U, U0 \rightarrow 0UU, U1 \rightarrow 111\}, S)$$

which is growing. This grammar has derivations like $S \Rightarrow T11 \Rightarrow 00U11 \Rightarrow 001111$ and $S \Rightarrow T11 \Rightarrow T0U11 \Rightarrow 00U0U11 \Rightarrow 00U01111 \Rightarrow 000UU1111 \Rightarrow 000U111111 \Rightarrow 00011111111$. The language of the grammar is

$$\{0^n 1^{2^n} : n > 0\} = \{011, 001111, 0^3 1^8, 0^4 1^{16}, 0^5 1^{32}, \ldots\}$$

and not context-free, as infinite languages satisfying the context-free pumping lemma can only have constant gaps (sequence of lengths without a word). This grammar has growing gaps.

Exercises 7.29 - 7.31

Exercise 7.29

Show that every context-free language is the union of a language generated by a growing grammar and a language containing only words up to length 1.

Exercise 7.30

Modify the proof of Theorem 5.38 to prove that every recursively enumerable language, that is, every language generated by some grammar is the homomorphic image of a language generated by a growing context-sensitive grammar.

Exercise 7.31

Construct a growing grammar for the language $\{1^{2^n}0^{2n}1^{2^n} : n > 0\}$ which is the “palindromisation” of the language from Example 7.28.

Midterm Test Friday 8 Oct 2021

Midterm Test is Closed Book Exam with one A4 Helpsheet (can be used two-sided) allowed.

Zoom Starts 10:00 hrs; Identity Checking.
File with Question Paper on Luminus Files Opens at 10:30 hrs. Folder name “Midterm Test Question Paper”. Write student number onto solutions, not name. Number pages by question number and order them ascendingly. Either handwrite on paper or type (with pdf-viewer) on question pdf-file.
Stop Writing at 11:30 hrs.
Complete Uploading into Luminus by 11:45 hrs. Folder name “Midterm Test Submission”. The file should be NUMBER-MT-CS3231.pdf, so if the student number is “A01234567X” then the file name should be “A01234567X-MT-CS3231.pdf”.

Screen Recording

Upload Screen Recording AFTER uploading the solution. Folder name is “Midterm Screen Recordings”. There are several days time for it.

Name files as above with SR (for screen recording) in it, for example “A01234567-MT-SR-CS3231.pdf”.

Use Panopto for it as recommended by SoC. Claimed to be the easiest system to use for screen recording.

Download Panopto at least one week before the exam. Test it by creating a test submission (for example recording how you do a one homework) and submit it to the Media Web folder “Screen-Recording Test” in the Mediaweb. Practise the screen recording so that it works during the midterm test.

Please consult “https://mysoc.nus.edu.sg/academic/e-exam-sop-for-students/” before the midterm test.

Midterm and Final Questions A–B

Some old additional exercises for Preparation of the Midterm and Final Exam. Starred exercises are NOT PART of midterm syllabus.

Exercise A.
Provide a regular expression for all words which have either the subword $\mathbf{110}$ or the subword $\mathbf{120}$ over the ternary alphabet $\{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$.

Exercise B.
Use structural induction to define $\mathbf{L^{mi}} = \{\mathbf{w^{mi}} : \mathbf{w} \in \mathbf{L}\}$ for all regular sets $\mathbf{L}$. So first define this for all sets containing words up to length $\mathbf{1}$ and then explain, how one goes on with concatenation, union, Kleene star and Kleene plus. Sets consisting of words of length at least $\mathbf{2}$ can be obtained from the base cases this way.

Midterm and Final Questions C–E

Exercise C.
Provide a DFA accepting those binary numbers which are multiples of 3: $11, 110, 1001, 1100, 1111, 10010, 10101$, $11000, 11011$ and so on.

Exercise D.
Construct over the ternary alphabet $\{0, 1, 2\}$ a DFA which accepts all words which contain each of the digits an odd number of times.

Exercise E.
Provide a non-deterministic finite automaton of thirteen states which accepts all decimal numbers which are not a multiple of 210; for this note that 0 is also a multiple of 210 and should be rejected. The numbers accepted should not having leading zeroes.

Midterm and Final Questions F–H

Exercise F.
Construct a context-free grammar in Chomsky Normal Form for the language $\{0^n 1^m 2^n : n, m \in \mathbb{N}\}$.

Exercise G.
Construct a context-free grammar in Greibach Normal Form for the language $\{0^n 1^m 2^n : n, m \in \mathbb{N}\}$.

Exercise H*.
Use the grammar from F and the Cocke Kasami Younger algorithm to check whether F generates the following words: 00122 and 00112.

Midterm and Final Questions I–K

Exercise I.

Recall that a language $\mathbf{L}$ satisfies the weakest form of the Pumping Lemma iff there is a constant $\mathbf{k}$ such that all words of length at least $\mathbf{k}$ in $\mathbf{L}$ can be split into parts $\mathbf{xyz}$ with $\mathbf{y} \neq \varepsilon$ and $\{\mathbf{x}\} \cdot \{\mathbf{y}\}^* \cdot \{\mathbf{z}\} \subseteq \mathbf{L}$. Which of the following choices for $\mathbf{L}$ satisfy this pumping lemma:

1. $\mathbf{L} = \{\mathbf{0^n 1^m 2^n} : \mathbf{n}, \mathbf{m} \in \mathbb{N}\}$;
2. $\mathbf{L} = \{\mathbf{0^n 1^m 0^n} : \mathbf{n}, \mathbf{m} \in \mathbb{N}\}$;
3. $\mathbf{L} = \{\mathbf{0^n 1^m 2^k} : \mathbf{n} + \mathbf{k} \neq \mathbf{m}\}$?

Exercise J.

Which of the languages in Exercise I have a linear grammar?

Exercise K.

Use Ogden's Lemma to prove that the language $\{\mathbf{0^n 1^m 2^k} : \mathbf{n} \neq \mathbf{m} \wedge \mathbf{n} \neq \mathbf{k} \wedge \mathbf{m} \neq \mathbf{k}\}$ is not context-free.

Midterm and Final Questions L–O

Let $\mathbf{h}$ be a homomorphism mapping $\{0, 1, 2, 3\}^*$ to $\{4, 5\}^*$. Determine the number of homomorphisms for the following tasks:

Exercise L. $\mathbf{h(01) = 444}$ and $\mathbf{h(23) = 55555}$ and no letter is mapped to the empty word.
Exercise M. $\mathbf{h(012) = 4444}$ and $\mathbf{h(030) = 4554}$.
Exercise N. $\mathbf{h(000112) = 4444445445444}$.

Exercise O. Let $\mathbf{L = \{2^m 0^n 1^n 2^n, 3^n 2^m : m, n > 0\}}$.
Is there a homomorphism $\mathbf{h}$ mapping $\mathbf{L}$ to a set which is context-free but not regular and which satisfies the traditional regular pumping lemma? If so, provide this $\mathbf{h}$, if not, explain why $\mathbf{h}$ does not exist.

Midterm and Final Questions P–S

Do for $\mathbf{L} = \{\mathbf{w} \in \{\mathbf{000}, \mathbf{111}, \mathbf{222}\}^+ : \mathbf{w}$ is a palindrome$\}$ the following exercises.

Exercise P. Which is the least constant $\mathbf{k}$ such that every word $\mathbf{w} \in \mathbf{L}$ can be split into three parts $\mathbf{x}, \mathbf{y}, \mathbf{z}$ with $\mathbf{w} = \mathbf{xyz}$ and $\mathbf{1} \leq |\mathbf{y}| \leq \mathbf{k}$ and $\mathbf{xy^*z} \subseteq \mathbf{L}$. Give the answer ∞ if there is no such constant $\mathbf{k}$. Prove the answer.

Exercise Q. Provide a context-free grammar in Chomsky Normal Form for $\mathbf{L}$.

Exercise R. Provide a context-free grammar in Greibach Normal Form for $\mathbf{L}$.

Exercise S*. Provide a PDA accepting by state for $\mathbf{L}$. Can this PDA be made deterministic? Give a short reason for the answer.

Midterm and Final Questions T–W

Prove that the following functions are primitive recursive.

Exercise T*. Function $\mathbf{f(n) = n^2}$.

Exercise U*. Function $\mathbf{g(n) = n^n}$.

Exercise V*. Function $\mathbf{h(m, n) = \binom{m+n}{m}}$.

Exercise W*. Function $\mathbf{k(m, n) = m! + n!}$.

Midterm and Final Questions X-Z

Exercise X*. Is it decidable to check whether a polynomial with integer coefficients and one input variable $\mathbf{x}$ takes on some input the value $\mathbf{1024}$? For example, for input function $\mathbf{f(x) = x^2 + 1}$, one wants to check whether there is an integer $\mathbf{x}$ with $\mathbf{f(x) = 1024}$. Prove the answer.

Exercise Y*. Let $\varphi_0, \varphi_1, \ldots$ be an acceptable numbering of all partial-recursive functions. Is $\mathbf{L} = \{\mathbf{e} : \varphi_\mathbf{e}(\mathbf{x})$ is undefined for some $\mathbf{x}\}$ recursively enumerable? Prove the answer using Rice's Theorem.

Exercise Z*. Let $\varphi_0, \varphi_1, \ldots$ be an acceptable numbering of all partial-recursive functions. Is $\mathbf{H} = \{\mathbf{e} :$ the domain of $\varphi_\mathbf{e}$ is the range of a primitive recursive function$\}$ (a) decidable, (b) recursively enumerable and undecidable, (c) not recursively enumerable? Prove the answer using Rice's Theorem.