Set membership question.

Question. Given a regular set $A \subseteq \Sigma^*$ and $x \in \Sigma^*$ is $x \in A$?

DFA M: Simulate M on the input x

$O(k)$ where $|x| = k$

NFA N: Is $x \in L(N)$?

$O(k n^2)$ $|x| = k$, $|Q| = n$.

Question. Given a CFL $A \subseteq \Sigma^*$ and $x \in \Sigma^*$ is $x \in A$?

Algorithm - Due to Cocke, Kasami and Younger

CKY - algorithm.

Runs in cubic-time

Determines for each substring y of x the set of all nonterminals that generate y.

Defined inductively on the length of y.

Assume G is in Chomsky normal form.

Example. $S \to AB \mid BA \mid SS \mid AC \mid BD$

$A \to a, \quad B \to b, \quad C \to SB, \quad D \to SA$

$L(G)$ - all strings with equal number of a's and b's.

Let $x = aabbab$. Let $n = |x|$ (So $n=6$ here)

$$\begin{array}{ccccccccccccc} & a & & a & & b & & b & & a & & b & \\ 0 & & 1 & & 2 & & 3 & & 4 & & 5 & & 6 \end{array}$$

For $0 \le i < j \le n$, let x_{ij} denote the substring of x between i & j. Ex. $x_{0,3} = aab$, $x_{2,6} = bbab$
$x = x_{0,n}$.

0						
_	1					
_	_	2				
_	_	_	3			
_	_	_	_	4		
_	_	_	_	_	5	
_	_	_	_	_	_	6

Entry $T_{i,j}$ of T - Set of non-terminals of G that generates the substring x_{ij} of x.

Define by induction on the length of the substrings.

Example. $S \to AB \mid BA \mid SS \mid AC \mid BD$

$A \to a, \quad B \to b, \quad C \to SB, \quad D \to SA$

$x \to$ a a b b a b

0 1 2 3 4 5 6

Substring of length 1.

0						
A	1					
	A	2				
		B	3			
			B	4		
				A	5	
					B	6

Substring of length 2

0						
A	1					
ϕ	A	2				
	S	B	3			
		ϕ	B	4		
			S	A	5	
				S	B	6

For each $x_{i,i+2}$: split the string into two substrings $x_{i,i+1}$ and $x_{i+1,i+2}$. Select a non-terminal X from $T_{i,i+1}$ and Y from $T_{i+1,i+2}$.
Look for a production $Z \to XY$ in G.
Label $T_{i,i+2}$ with Z for each such production.

Example. $S \to AB \mid BA \mid SS \mid AC \mid BD$

$A \to a, \quad B \to b, \quad C \to SB, \quad D \to SA$

$x \to$ a a b b a b

0 1 2 3 4 5 6

Substring of length 1.

0						
A	1					
_	A	2				
_	_	B	3			
_	_	_	B	4		
_	_	_	_	A	5	
_	_	_	_	_	B	6

Substring of length ≥ 3

0						
A	1					
ϕ	A	2				
ϕ	S	B	3			
S	C	ϕ	B	4		
D	S	ϕ	S	A	5	
S	C	ϕ	C	S	B	6

For each $x_{i,i+2}$, break into 2 substring $x_{i,i+1}$ & $x_{i+1,i+2}$
Select a nonterminal X from $T_{i,i+1}$ and Y from $T_{i+1,i+2}$
Look for production $Z \to XY$ in G.
Label $T_{i,i+2}$ with Z for each such production.

$S \in T_{0,6}$ indicates $S \xrightarrow[G]{*} x_{0,6} = x$: the input string.
Therefore $x \in L(G)$.

Algorithm.

For $i := 0$ to $n-1$ do [length 1 strings].

begin

$T_{i,i+1} = \phi$ [Initialise to ϕ]

For $A \to a$, a production in G do

if $a = x_{i,i+1}$ then $T_{i,i+1} = T_{i,i+1} \cup \{A\}$

end

For $m := 2$ to n do [length $m \geq 2$]

for $i := 0$ to $n-m$ do [for each substring of length m]

begin

$T_{i,i+m} := \phi$ [Initialise]

For $j = i+1$ to $i+m-1$ do [All possible splits]

For $A \to BC$, a production in G do

if $B \in T_{i,j}$ and $C \in T_{j,i+m}$

then

$T_{i,i+m} = T_{i,i+m} \cup \{A\}$

end

Running Time - $O(n^3 p)$ where $p = |P|$ and $n = |x|$.

Ambiguous Grammar.

Consider the grammar $S \to S+S \mid S*S \mid (S) \mid A$
$A \to a \mid b$

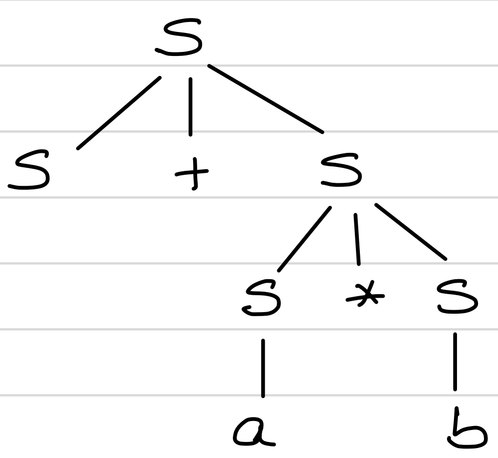
Consider the string $a+a*b$

Derivation 1. $S \to S+S \to a+S \to a+S*S \to a+a*b$

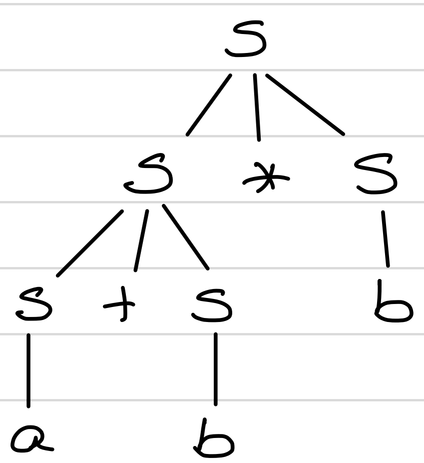
Derivation 2. $S \to S*S \to S+S*S \to a+S*S \to a+a*b$

Parse Tree

Derivation 1



Derivation 2.



A CFG G is ambiguous if $\exists x \in L(G)$ for which there are <u>two different parse trees</u>.

→ Not two different derivations.

Definition

A string x is derived ambiguously in a CFG G if it has two different leftmost derivation.

Grammar G is ambiguous if it generates some string ambiguously.

G is unambiguous if G is not ambiguous.

A CFL $A \subseteq \Sigma^*$ is inherently ambiguous if $\forall$ CFG G s.t $L(G) = A$, G is ambiguous.

Note. There are inherently ambiguous CFLs.

DCFLs – CFLs that can be accepted by a DPDA.

DCFLs always admit an unambiguous grammar.

DCFLs $\subsetneq$ unambiguous CFLs.