Every Context-free language can be decided in polynomial time, using the CYK (Cook, Younger, and Kurotowski) dynamic programming algorithm.

Notation: If $A$ is a variable of a context-free grammar with terminal alphabet $\Sigma$, we let $L(A)$ denote the set of strings over $\Sigma$ that can be derived from $A$.

Unless you jump through hoops, a CNF grammar cannot generate the empty string, so we assume that $G$ generates $L(\lambda)$, i.e., all non-empty strings of $L$.

A Chomsky Normal Form grammar is a CF grammar with only two kinds of productions. The left-hand-side of one of these productions is, of course, a variable. The right-hand-side is either a terminal or two variables.

**Example.** The language $L$ of non-empty even length palindromes over $\{a, b\}$ is generated by the grammar $G$ below.

- $S \rightarrow aSa$
- $S \rightarrow bSb$
- $S \rightarrow aa$
- $S \rightarrow bb$

In order to use the CYK algorithm, we need a CNF grammar equivalent to $G$, such as

- $S \rightarrow AB$
- $S \rightarrow AC$
- $C \rightarrow SA$
- $S \rightarrow BD$
- $D \rightarrow SB$
- $A \rightarrow a$
- $B \rightarrow b$

### 0.1 Subproblems of CYK.

Let $L$ be a context-free language, and $G$ a CNF (Chomsky normal form) grammar for $L$ with terminal alphabet $\Sigma$. An instance of the membership problem for $L$ is a string $w \in \Sigma^*$, and the question is, whether $w \in L$.

Let $n = |w|$. We write $w = a_1a_2\ldots a_n$; $w$ has $\binom{n+1}{2}$ substrings. For any $1 \leq \ell \leq i \leq n$ let $w_{i,\ell} = a_i\ldots a_{i+\ell-1}$, the substring of $w$ of length $\ell$ starting at the $i$th symbol of $w$. Note that $w_{i,1} = a_i$.

Let $m$ be the number of variables of $G$, and let $A_p$ be the $p^{th}$ variable. We assume that $A_1 = S$, the start symbol. Let $S[p, i, \ell]$ be 1 if $w_{i,\ell} \in L(A_p)$, 0 otherwise. There are $m\binom{n+1}{2}$ subproblems, namely to compute the values of $\{S[p, i, \ell]\}$.

We write the dynamic program CYK in pseudocode.
for all $1 \leq p \leq m$, $1 \leq \ell \leq i \leq n$

$S[p, i, \ell] = \text{false}$

for all $1 \leq p \leq m$, $1 \leq i \leq n$

if $(A_p \rightarrow a_i)\ S[p, i, 1] = \text{true}$

for all $2 \leq \ell \leq n$

for all $1 \leq i \leq n - \ell$

for all $i + 1 \leq j \leq n - \ell + 1$

for $1 \leq p \leq m$, $1 \leq q \leq m$, $1 \leq r \leq m$

if $(A_p \rightarrow A_q A_r$ and $S[q, i, j - i]$ and $S[r, j, \ell - j + i]$)

$S[p, i, \ell] = \text{true};$

return $S[1, 1, n]$

---

Walking Through CYK by Hand

Recall that $V = \{A_1, \ldots A_m\}$ is the alphabet of variables of $G$. We define $V[i, \ell]$ to be the set of all variables $A_p$ such that $w_{i, \ell} \in L(A_p)$. In terms of our $S$ notation, $V[i, \ell]$ is the set of all variables $A_p$ such that $S[p, i, \ell]$ is true. Hand execution of CYK consists of computing the sets $\{S[p, i, \ell]\}$ in order of increasing $\ell$. In textbook and internet explanations of CYK, each of those sets is shown inside a box which is an entry of a triangular matrix, since $i + \ell \leq n + 2$, and this matrix is oriented in the usual row and column manner. However, I have found it intuitive to rotate the matrix 45 degrees, as in Figure 1 below.

![Figure 1: CYK Matrix](image)

Each box corresponds to one substring of $w$ and holds one of the sets $V[i, \ell]$ The values of those sets are computed from the bottom up: $w \in L$ if and only if $S$ is a member of the top set, $V[1, n]$
Example. Let $G$ be the CNF grammar:

\begin{align*}
S & \rightarrow IS \\
S & \rightarrow WS \\
S & \rightarrow XY \\
X & \rightarrow IS \\
Y & \rightarrow ES \\
S & \rightarrow a \\
E & \rightarrow e \\
I & \rightarrow i \\
W & \rightarrow w
\end{align*}

Here is the CYK matrix with the initial string $iiwae$ written below the first row. The entries of each cell of the matrix are the members of $V[i, \ell]$ Since $S$ is in the top cell, $w \in L$.

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{cyk_matrix.png}
\caption{CYK verifying that $iiwae \in L$.}
\end{figure}

CYK can then be used to show that the string $ieiaea$ is not in $L$, as shown in Figure 3.

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{cyk_matrix2.png}
\caption{CYK verifying that $iiwae \notin L$.}
\end{figure}