5 Context Free Grammars

[Linz § 5.1]

Definition: A **grammar** $G = (N, \Sigma, S, P)$ is **context-free** if all producitons in P have the form

$$A \longrightarrow x$$

where $A \in N$ and $x \in (N \cup \Sigma)^*$.

A **language** is **context-free (CF)** if $L = L(G)$ for some **context-free grammar (CFG)**, G.

Note:
Clearly, all linear grammars are CF,
$\therefore$ all regular languages are CF,
but the converse is **not** true.

Note:
CFG's are important in the definition of **programming languages**.
(See, e.g., Linz chapter 5 Intro and Section 5.3.)

Examples of CFL's:

Example 1: Recall (pp. 1-26/30) the language **Bal(Σ)** of **balanced words** over $\Sigma = \{a, b\}$:

$$\mathbf{Bal}(\Sigma) = \{u \in \Sigma^* \mid n_a(u) = n_b(u)\}$$

We write $\mathbf{Bal}(a, b) = \mathbf{Bal}(\{a, b\})$

We saw that $\mathbf{Bal}(a, b)$ is generated by the CFG $\boldsymbol{G}$, with productions:

$$S \longrightarrow aSb \mid bSa \mid SS \mid \lambda$$

We proved:

(a) $L(G) \subseteq \mathbf{Bal}(a, b)$

(b) $\mathbf{Bal}(a, b) \subseteq L(G)$

(a) is clear. (Every sentential form generated by $\boldsymbol{G}$ is **balanced**.)

(b) was proved by showing:

$\boldsymbol{u} \in \mathbf{Bal}(a, b) \implies \boldsymbol{u}$ is generated by $\boldsymbol{G}$,

– by induction (CVI) on $|\boldsymbol{u}|$.

Example 2: Take $\Sigma = \{\, [\,,\ ]\, \}$ and let $WN_{[\,]} \subseteq \Sigma^*$ be the set of all **well-nested bracket strings**. (See Kozen, Lecture 20.)

E.g. [[ ]] [ ] is **well-nested**,

but [ ]] [[ ] is **not**.

For $u \in \Sigma^*$, **define** $E(u) = n_{[}(u) - n_{]}(u)$.

What is the difference between $WN_{[\,]}$ and **Bal**([,])?

For $u \in \{\, [\,,\,]\, \}^*$ to be **balanced**, it is **sufficient** that:

(1) $E(u) = 0$

but for u to be **well-nested**, we also need:

(2) for **every prefix** v of u:

$$n_{[}(v) \;\geq\; n_{]}(u)$$

i.e.

$$E(v) \geq 0$$

i.e. $\forall k = 0, \ldots, |u|$:

$$\begin{aligned} E(u \restriction k) &= n_{[}(u \restriction k) - n_{]}(u \restriction k) \\ &\geq 0. \end{aligned}$$

(See p. 1-26 for notation.)

This suggests, as a CFG for $WN_{[\,]}$

$$S \longrightarrow \underline{[\, S\,] \mid SS \mid \lambda}$$

(Recall the grammar G for $\mathbf{Bal}(a,b)$ on p. 5-2.)

Call this $G_{[\,]}$.

Show $L(G_{[\,]}) = WN_{[\,]}$

We must show:

(a)$L(G_{[\,]}) \subseteq WN_{[\,]}$

(b)$WN_{[\,]} \subseteq L(G_{[\,]})$

(a) is clear. (Why?)

Every sentential form derived from $G_{[\,]}$ is well-nested (ignoring non-terminal 'S').

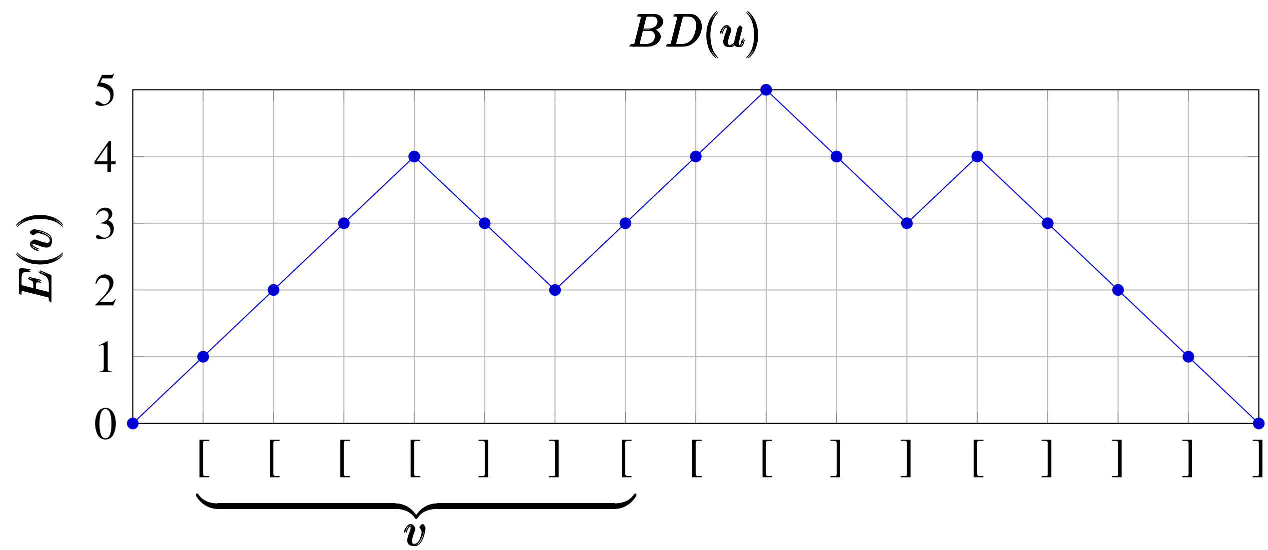
(b) We will show:

$$u \in WN_{[\,]} \Longrightarrow u \textbf{ is generated by } G_{[\,]}$$

by CVI on $|u|$

To prove this, we need the concept of **bracket diagrams**.

The **bracket diagram** of $\boldsymbol{u} = \boldsymbol{BD(u)}$ is the graph of $\boldsymbol{E(v)}$ for all **prefixes** $\boldsymbol{v}$ of $\boldsymbol{u}$:



Note:

(1) For **all prefixes** $\boldsymbol{v}$ of $\boldsymbol{u}$, $\boldsymbol{E(v) \geq 0}$, and

(2) $\boldsymbol{E(u)} = 0$

In other words, if $\boldsymbol{u} \in \{\,[\,,\,]\,\}^*$, then $\boldsymbol{u}$ is **well-nested** iff

(1) its bracket diagram is **always non-negative**, and

(2) its **right end** has value **0**.

We will show: for all $u \in \{ [\, , \,] \}^*$, if $u \in WN_{[\,]}$, then $u \in L(G_{[\,]})$.

By **CVI** on $|u|$

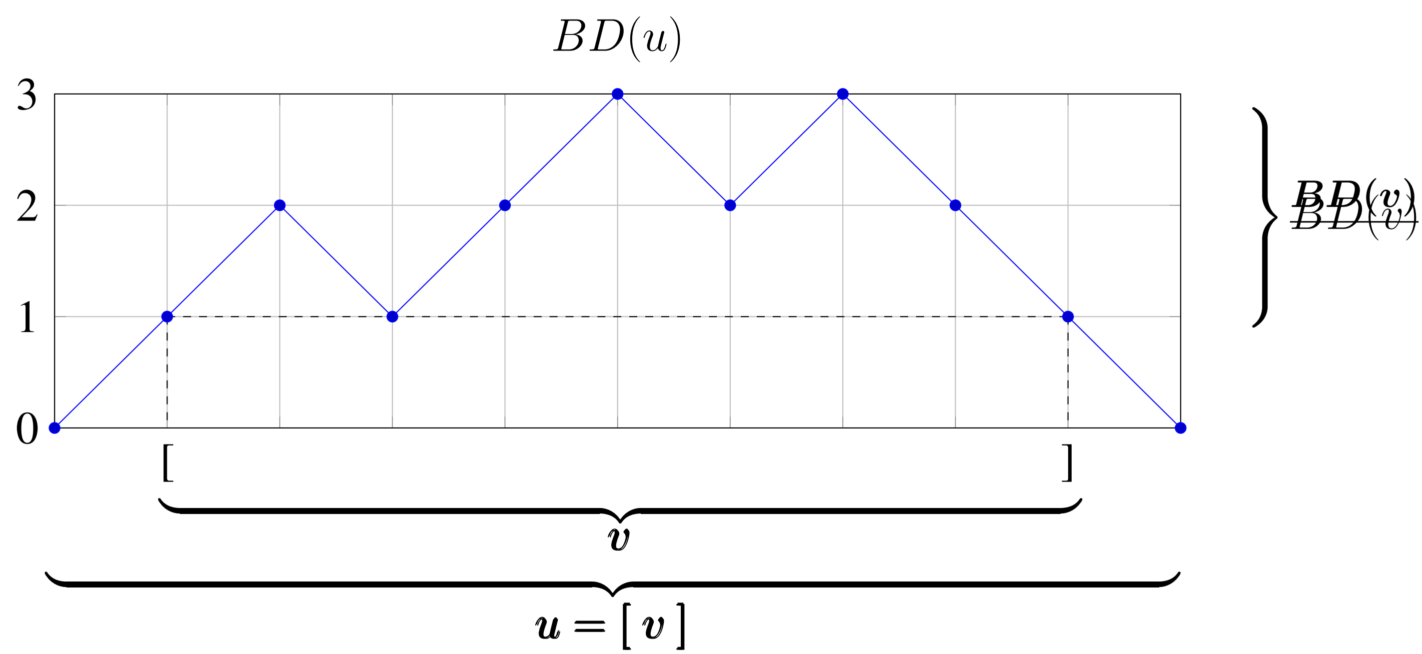
Base: $|u| = 0$. Then $u = \lambda$, and so $S \implies u$.

Induction step: $|u| = n > 0$

Suppose: $\forall v$ **with** $|v| < n : \big(v \in WN_{[\,]} \implies v \in L(G_{[\,]})\big)$.

Case 1

$BD(u)$ is **0** at **beginning** and **end**, and **positive** in between.



Then $\boldsymbol{u}$ has the form $[\,\boldsymbol{v}\,]$, where $BD(\boldsymbol{v})$ is also:

- 0 at the beginning and end,
- non-negative in between.

So $\boldsymbol{v}$ is **also well-nested**.

Also, $|\boldsymbol{v}| = \underline{|\boldsymbol{u}| - \mathbf{2}} < |\boldsymbol{u}|$.

Hence, by **induction hypothesis**,
in $G_{[\,]}$: $\quad S \overset{*}{\Longrightarrow} \boldsymbol{v}$

Hence

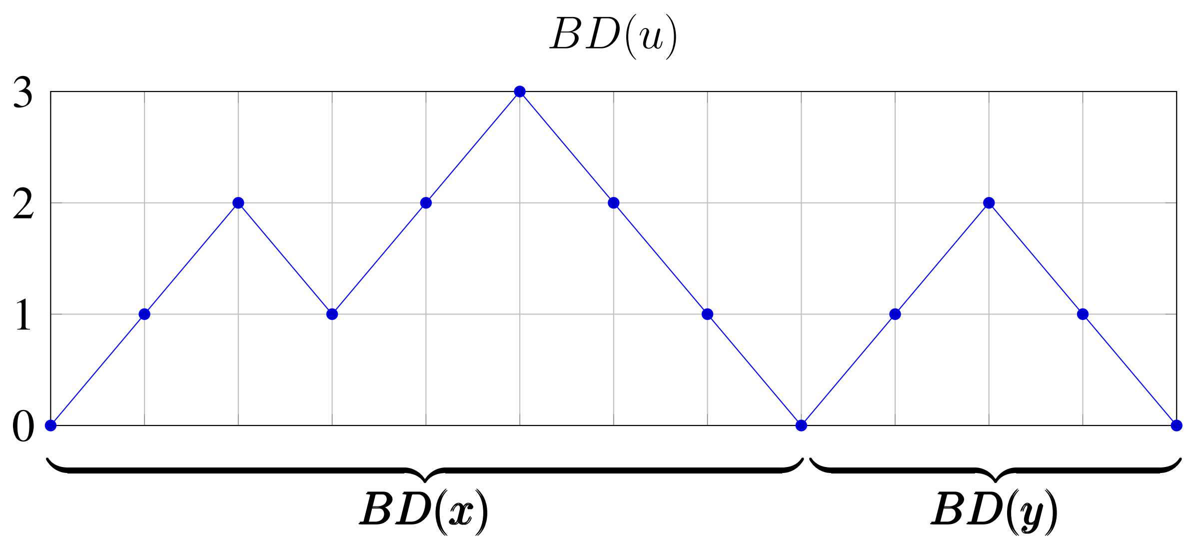
$$S \Longrightarrow [\,S\,] \overset{*}{\Longrightarrow} \boldsymbol{v} = \boldsymbol{u},$$

i.e.

$$S \overset{*}{\Longrightarrow} \boldsymbol{u}.$$

Case 2

$BD(u)$ is **0** at **some point between** the beginning and the end:



So $u = xy$, where $x, y \in WN_{[\,]}$.

Also, $|x|, |y| < |u|$

Hence by **induction hypothesis**,

$$S \underset{(1)}{\overset{*}{\Longrightarrow}} x \quad \text{and} \quad S \underset{(2)}{\overset{*}{\Longrightarrow}} y$$

Hence in $G_{[\,]}$

$$S \Longrightarrow SS \underset{(1)}{\overset{*}{\Longrightarrow}} xS \underset{(2)}{\overset{*}{\Longrightarrow}} xy = u$$

i.e. $$S \overset{*}{\Longrightarrow} u.$$

Problem: (2 types of brackets)

Let $\Sigma = \{ [,] , (,) \}$

(1) ***Define***, for $\boldsymbol{u} \in \Sigma^*$: "$\boldsymbol{u}$ is **well-nested**".

(2) Let $WN_{[\,]\,(\,)} = \{u \in \Sigma^* \mid u \text{ is well-nested}\}$.

Find a CFG for $WN_{[\,]\,(\,)}$!

Note:

The string '$[\,(\,]\,)$' ***is not well-nested***.

More on CFG's and CFL's

Given Σ,

$Q.$ Is the class of **CFL's** closed under **union**?

I.e. is the **union** of 2 CFL's always a CFL?

$A.$ Let $\boldsymbol{L_1 = L(G_1), \quad G_1 = (N_1, \Sigma, S_1, P_1)}$

$$\boldsymbol{L_2 = L(G_2), \quad G_2 = (N_2, \Sigma, S_2, P_2)}$$

Can assume w.l.o.g.: $\boldsymbol{N_1 \cap N_2 = \varnothing}$ (***Why?***)

Then let $\boldsymbol{G = (V, \Sigma, S, P)}$ where

$$\boldsymbol{N = N_1 \cup N_2 \cup \{S\}},$$

$$P = P_1 \cup P_2 \cup \underline{\boldsymbol{\{S \longrightarrow S_1 \mid S_2\}}}$$

Then $L(G) = \underline{\boldsymbol{L_1 \cup L_2}}$

$Q.$ What about **intersection**?

*** Stay tuned! ***

Examples: Find CFG's for the following languages: $(\Sigma = \{a, b\})$

1. $\{a^n b^{2n} \mid n \geq 0\}$

   $\underline{S \longrightarrow aSbb \mid \lambda}$

2. $\{a^n b^m \mid n \leq m\}$

   $\underline{S \longrightarrow aSb \mid Sb \mid \lambda}$

3. $\{a^n b^m \mid n < m\}$

   $\underline{S \longrightarrow aSb \mid Sbb \mid \lambda}$

4. $\{a^n b^m \mid n \leq m \leq 2n\}$

   $\underline{S \longrightarrow aSb \mid aSbb \mid \lambda}$

5. $\{a^n b^n c^k \mid n, k \geq 0\}$ $(\Sigma = \{a, b, c\})$

   $\underline{S \longrightarrow Sc \mid T}$

   $\underline{S \longrightarrow aTb \mid \lambda}$

6. $\{a^n b^m \mid m \leq n+2\}$

$\underline{S \longrightarrow aSb \mid aS \mid \lambda \mid b \mid bb}$

7. $\{a^n b^m \mid n \neq m\}$

$\underline{S \longrightarrow S_1 \mid S_2}$

$\underline{S_1 \longrightarrow aS_1b \mid S_1b \mid b}$

$\underline{S_2 \longrightarrow aS_2b \mid aS_2 \mid a}$

8. $\{a^n b^m \mid 0 \leq m \leq 2n\}$

$\underline{S \longrightarrow aS \mid aSb \mid aSbb \mid \lambda}$

9. $\{a^n b^m \mid m \geq 2n\}$

$\underline{S \longrightarrow aSbb \mid Sb \mid \lambda}$

10. $\{a^n b^k c^n \mid n, k \geq 0\}$ $(\Sigma = \{a, b, c\})$

$\underline{S \longrightarrow aSc \mid T}$

$\underline{T \longrightarrow Tb \mid \lambda}$