Lecture Tue 3/24

Let $x =$ ( ) (( ) ((( ) ( ))))

Diff: 0 1 0 1 2 1 2 3 4 3 4 3 2 1 0

$Y =$ (( ) ( ))) (( ) ( )

0 1 2 1 2 1 0 -1 0 1 0 1 0

↖ no good.

Is ε a balanced paren string? Dictum: yes. Otherwise, "Rule 0" [X must start with (and X must end with).

Rule 1: Every (has a corresponding) to its right.

(( )) ? The string complies with Rule 1 but the matching is wrong.

~~(~~) (~~)~~ — shows "Rule 0" is not enough.

How can we define a criterion that we can program & verify?

$Z =$ (( ) ( ) ( ))

Increment $Diff_x$ for each '('
Decrement $Diff_x$ for each ')'

Numerical Specification. For each i, $0 \leq i \leq n = |X|$, define $Diff_x(i)$ by

$Diff_x(i) = \#\text{'('}[x_1 \cdots x_i] - \#\text{')'}[x_1 \cdots x_i]$

$i = 0$, $X = \varepsilon$: both counts are 0, so $Diff_x(0) = 0$

$\boxed{\textbf{Rule: } X \text{ is } \underline{\text{balanced}} \text{ if } Diff_x(n) = 0 \text{ and for each } i, 0 \leq i < n, Diff_x(i) \geq 0.}$

Define $\underline{Bal} = \{x \in \{(,)\}^* : x \text{ is balanced } \underline{\text{by this rule}}.\}$. Includes ε.

Consider $G = S \to \varepsilon \mid SS \mid \text{"("} S \text{")"}$. $\boxed{\text{Is } L(G) = Bal?}$

and $G' = S \to \varepsilon \mid (S)S$.   How about G'?

I.e. Is $L(G) \subseteq Bal$? and Is $Bal \subseteq L(G)$? (Ditto G'?)

Is G sound (w.r. to the Rule for Bal)? Is G comprehensive? (for that rule)

Intuitively yes, ditto G', but how to prove? We will use "Strong Induction": (aka Course-of-values Ind)

- If you can prove $P(b)$ for some basis value b (often 0 or 1)
- and If for any n you can prove that the truth of $P(m)$ for all $m < n$ implies $P(n)$, then you may conclude: $(\forall n \geq b)\, P(n)$.

$m = n-1$ is enough, if you can use it.

Proof of $L(G) \subseteq Bal$ by strong Induction on the length n of derivations.

$P(n) =_{def}$ " If $S \Rightarrow^{n} x$ then x is balanced" (when $x \in \Sigma^*$)

So "$\forall n\, P(n)$" says "If $S \Rightarrow^* x$ – that is, x is any string in $L(G)$ – then x is balanced". So $L(G) \subseteq Bal$.

$n=0$ doesn't apply.

Basis: $n=1$ The only way $S \Rightarrow^1 x$ in one step is $x=\varepsilon$.
But ε is balanced (by our interpretation of Bal)

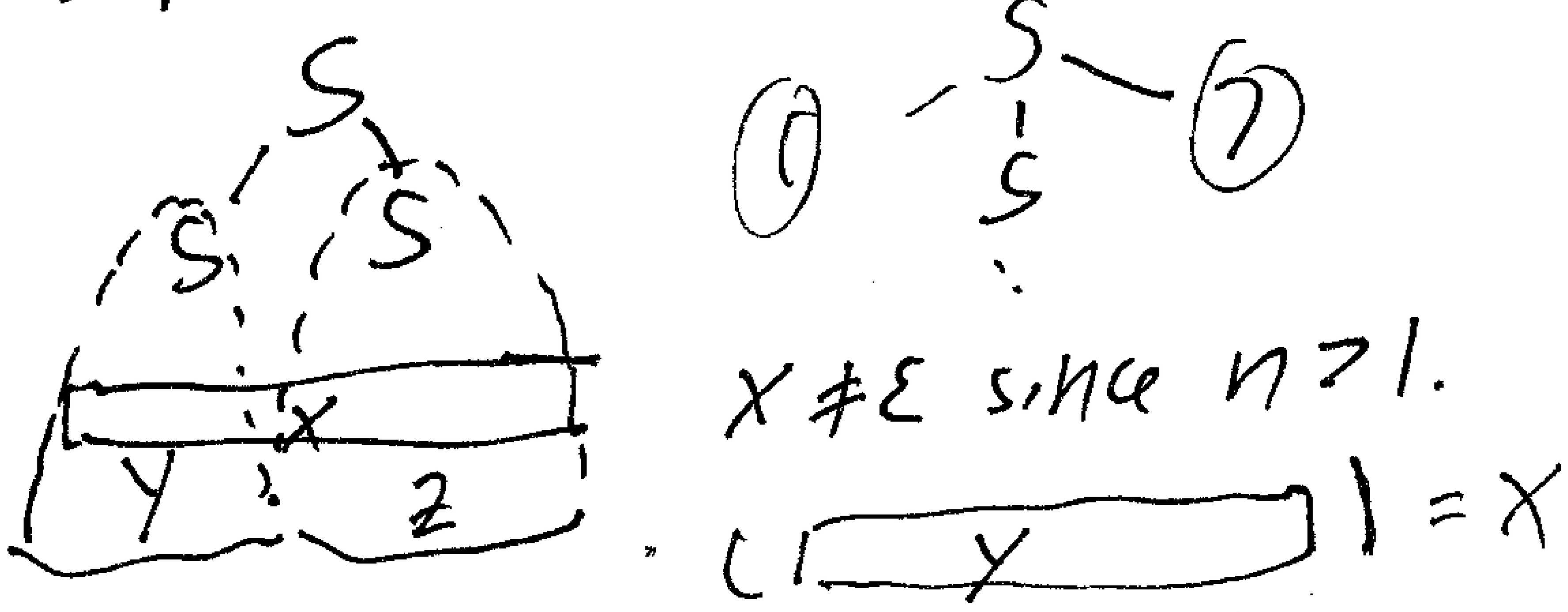
So $P(1)$ holds ✓. General "proof script" for the induction part

Let any $n \geq b$ be given. Assume (IH) the statements $P(m)$ for all $m < n$ (and $m \geq b$). Goal: Show $P(n)$. ∴ $\forall n\, P(n)$ by induction

To show $P(n)$, suppose $S \Rightarrow^n x$. Goal: Show x is balanced.

The Key "breakdown" is to ask: which rule/production in G was used first?
I.e. which production is used at the top of our parse tree for x?

$G: S \to \varepsilon \mid SS \mid (S)$



$x \neq \varepsilon$ since $n > 1$.

It could be either $S \to SS$ or $S \to (S)$
Suppose first the first step is $S \Rightarrow SS$.

Thus $x = yz$ where $S \Rightarrow^* y$ and $S \Rightarrow^* z$. Moreover, the numbers m_1 giving $S \Rightarrow^{m_1} y$ and m_2 giving $S \Rightarrow^{m_2} z$ are both $< n$ (since they sum to $n-1$). Hence we may apply the induction hypotheses $P(m_1)$ and $P(m_2)$ which conclude that y and z are balanced. Since $x = yz$, and we agreed that the concatenation of two balanced strings is balanced, x is balanced.

⚠ We need to cover the rule $S \to (S)$ first. Suppose $S \Rightarrow^* x$ using this production first (yet)
Then $x = (y)$ where $S \Rightarrow^* y$, in fact $S \Rightarrow^{n-1} y$. Since $n-1 < n$, we may apply $P(n-1)$ to conclude that y is balanced. Hence x is balanced. Since we covered all productions, ∴ $P(n)$. ∴ $\forall n\, P(n)$ ☒

Abbreviation & Top Down Ind Pf.

$G = S \to \varepsilon \mid SS \mid (S)$ Bal = {x : x is balanced}

Prove $L(G) \subseteq$ Bal by "S.I." directly on rules of G.

- $S \to \varepsilon$. Terminal production. $\varepsilon \in$ Bal so this checks out.
- $S \to SS$. Suppose $S \Rightarrow^* x$ utpf. Then $x = yz$ where $S \Rightarrow^* y$ and $S \Rightarrow^* z$. "By IH P_S on LHS of these rules" y is balanced and z is balanced.

"The Property of S"

$\boxed{P_S \equiv \text{"Every string x that I derive is balanced."}}$

With regard to the rule $S \to SS$, whe have IH P_S on RHS twice. to get that y and z are balanced.

New goal: Show P_S on LHS of this rule, ie that x is balanced.

Since $x = yz$ and both y & z are balanced, x is balanced. $\therefore$ "P_S on LHS".

- $\boxed{S \to (S)}$: Suppose $S \Rightarrow^* x$ utpf. Then $x = (y)$ where $S \Rightarrow^* y$. By IH P_S on RHS (of the boxed rule), y is balanced. $\therefore$ x is balanced. $\therefore$ This upholds P_S on LHS of the boxed rule.

Since all three productions in G are upheld, $L(G) \subseteq$ Bal by "S.I."

$S \to (S)S$ idea is y & z balanced $\Rightarrow$ (y)z is balanced.

EXTRA:

Preview of Thu lecture:

Language $E = \{x : \#a(x) = \#b(x)\}$

$L(G) \subseteq E$

G: $S \to \varepsilon \mid SS \mid AB \mid BA$

$A \to a \mid aS \mid$ ~~ABB~~ BAA

$B \to b \mid bS \mid ABB$

Extra wrinkle: now we need properties P_A for A and P_B for B as well as P_S for S.

$P_S \equiv$ "Every x I derive has $\#a(x) = \#b(x)$, ie $x \in E$."

$P_A \equiv$ "Every x I derive has $\#a(x) = \#b(x) + 1$."

$P_B \equiv$ "Every x I derive has $\#b(x) = \#a(x) + 1$."

Thursday's lecture will have this example and then cover "$E \subseteq L(G)$" going back to the parenthesis grammars too.