Monday April 25

To fully specify a PDA we could give its 6-tuple formal definition or we could give its input alphabet, stack alphabet, and state diagram. An informal description of a PDA is a step-by-step description of how its computations would process input strings; the reader should be able to reconstruct the state diagram or formal definition precisely from such a descripton. The informal description of a PDA can refer to some common modules or subroutines that are computable by PDAs:

- PDAs can "test for emptyness of stack" without providing details. *How?* We can always push a special end-of-stack symbol, $, at the start, before processing any input, and then use this symbol as a flag.

- PDAs can "test for end of input" without providing details. *How?* We can transform a PDA to one where accepting states are only those reachable when there are no more input symbols.

Big picture: PDAs were motivated by wanting to add some memory of unbounded size to NFA. How do we accomplish a similar enhancement of regular expressions to get a syntactic model that is more expressive?

DFA, NFA, PDA: Machines process one input string at a time; the computation of a machine on its input string reads the input from left to right.

Regular expressions: Syntactic descriptions of all strings that match a particular pattern; the language described by a regular expression is built up recursively according to the expression's syntax

Context-free grammars: Rules to produce one string at a time, adding characters from the middle, beginning, or end of the final string as the derivation proceeds.

Term	Typical symbol	Definition
Context-free grammar (CFG)	G	$G = (V, \Sigma, R, S)$
Variables	V	Finite set of symbols that represent phases in production pattern
Terminals	Σ	Alphabet of symbols of strings generated by CFG $V \cap \Sigma = \emptyset$
Rules	R	Each rule is $A \to u$ with $A \in V$ and $u \in (V \cup \Sigma)^*$
Start variable	S	Usually on LHS of first / topmost rule
Derivation		Sequence of substitutions in a CFG
	$S \implies \cdots \implies w$	Start with start variable, apply one rule to one occurrence of a variable at a time
Language generated by the CFG G	$L(G)$	$\{w \in \Sigma^* \mid \text{there is derivation in } G \text{ that ends in } w\} = \{w \in \Sigma^* \mid S \implies^* w\}$
Context-free language		A language that is the language generated by some CFG
Sipser pages 102-103		

Examples of context-free grammars, derivations in those grammars, and the languages generated by those grammars

$\Sigma =$

$S \Rightarrow$ · · · · $\Rightarrow$ ___ string over {0}.

one variable   start variable

$S \to 0S$
$S \to 0$

In $L(G_1)$ ...

$\bigcirc$ witnessed by $S \Rightarrow 0$ *second rule*

$\bigcirc\bigcirc$ witnessed by $S \Rightarrow 0S \Rightarrow 00$ *first rule* *second rule*

$\bigcirc\bigcirc\bigcirc$ witnessed by $S \Rightarrow 0S \Rightarrow 00S \Rightarrow 000$ *first rule* *first rule*

Not in $L(G_1)$ ...

(string over {0})

ε because any string of terminals that is at the end of derivation will have pos length because both rules add 1 to num of terminals in string.

$G_2 = (\{S\}, \{0,1\}, R, S)$

$G_2 = (\{S\}, \{0,1\}, \{S \to 0S, S \to 1S, S \to \varepsilon\}, S)$ $S \to 0S \mid 1S \mid \varepsilon$

In $L(G_2)$ ...

NOTE: shorthand for
$S \to 0S$
$S \to 1S$
$S \to \varepsilon$

$S \Rightarrow 0S \Rightarrow 0$

1101 $S \Rightarrow 1S \Rightarrow 11S \Rightarrow 110S \Rightarrow 1101S \Rightarrow 1101$

~~Not in $L(G_2)$ ...~~

$L(G_2) = \{0,1\}^* =$ the set of all strings over $\{0,1\}$.

$(\{S, T\}, \{0, 1\}, R, S)$ with rules

string 1 *string* 1 *string* 1 *string.*

$$S \to T1T1T1T$$
$$T \to 0T \mid 1T \mid \varepsilon$$

how to start

how to replace a T.
looks like G_2.

In $L(G_3)$ ...

1101   as witnessed by

$S \Rightarrow T1T1T1T \Rightarrow 1T1T1T \Rightarrow 11T1T \Rightarrow 110T1T \Rightarrow 1101T \Rightarrow 1101$

Not in $L(G_3)$ ...

11

since first production in each derivation
adds three 1s and no characters can be
erased.

$L(G_3) = \{ w \in \{0,1\}^* \mid w \text{ has at least three 1s} \} = L(\Sigma^* 1 \Sigma^* 1 \Sigma^* 1 \Sigma^*)$

$G_4 = (\{A, B\}, \{0, 1\}, R, A)$ with rules

start
var.

in the middle

$$A \to 0A0 \mid 0A1 \mid 1A0 \mid 1A1 \mid 1$$
$$ 1 \quad\;\; 2 \quad\;\; 3 \quad\;\; 4 \quad\; 5$$

In $L(G_4)$ ...

$A \overset{①}{\Rightarrow} 0A0 \overset{③}{\Rightarrow} 01A00 \overset{④}{\Rightarrow} 011A100 \overset{⑤}{\Rightarrow} 0111100$

$A \overset{⑤}{\Rightarrow} 1$

Not in $L(G_4)$ ...

11 is not in $L(G_4)$ because even length and ...

000 is not in $L(G_4)$

$L(G_4) = \{ w \in \{0,1\}^* \mid w \text{ has odd length and middle char is } 1 \}$.

Extra practice: Is there a CFG G with $L(G) = \emptyset$?

Design a CFG to generate the language $\{abba\}$ — *very simple language*

a lot of options

$$(\{S, T, V, W\}, \{a, b\}, \{S \to aT, T \to bV, V \to bW, W \to a\}, S)$$

$$(\{Q\}, \{a, b\}, \{Q \to abba\}, Q)$$

$$(\{X, Y\}, \{a, b\}, \{X \to aYa, Y \to bb\}, X)$$

Design a CFG to generate the language $\{a^n b^n \mid n \geq 0\}$ — *more complicated language.*

$\varepsilon \qquad ab \qquad aabb \qquad aaabbb \qquad aaaabbbb$

$G = (\{S\}, \{a, b\}, R, S)$ where R is given by

$$S \to \varepsilon \mid aSb$$

Sample derivation:

Design a CFG to generate the language $\{a^i b^j \mid j \geq i \geq 0\}$ *Extra practice*

Sample derivation:

Review: Week 5 Monday

Recall: Review quizzes based on class material are assigned each day. These quizzes will help you track and confirm your understanding of the concepts and examples we work in class. Quizzes can be submitted on Gradescope as many times (with no penalty) as you like until the quiz deadline: the three quizzes each week are all due on Friday (with no penalty late submission open until Sunday).

Please complete the review quiz questions on Gradescope about context-free grammars

Pre class reading for next time: Example 2.3 and Example 2.4

Wednesday April 27

$\mathcal{P}(\Sigma^*)$

many more languages

Set of Regular languages

Set of Context-free languages

Theorem 2.20: A language is generated by some context-free grammar if and only if it is recognized by some push-down automaton.

Definition: a language is called **context-free** if it is the language generated by a context-free grammar. The class of all context-free language over a given alphabet Σ is called **CFL**.

Consequences:

- Quick proof that every regular language is context free

- To prove closure of the class of context-free languages under a given operation, we can choose either of two modes of proof (via CFGs or PDAs) depending on which is easier

Over $\Sigma = \{a, b\}$, let $L = \{a^n b^m \mid n \neq m\}$. **Goal**: Prove L is context-free.

Building PDA that recognizes L.

$L = \underbrace{\{a^n b^m \mid n > m\}}_{L_1} \cup \underbrace{\{a^n b^m \mid n < m\}}_{L_2}$

Informal description of PDA recognizing L_2: read a's and push them onto stack, when see first b, pop a off the stack for each b we read and when stack is empty, if rest of string has positive number of b (and nothing else), accept (otherwise reject).

Informal description of PDA recognizing L_1

$\varepsilon, \varepsilon, \varepsilon$ ... $\varepsilon, \varepsilon, \varepsilon$

M_1 ... M_2
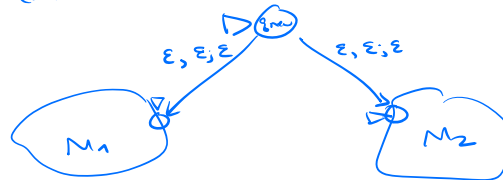
Build CFG G that generates $L = L_1 \cup L_2$

$L_1 = \{a^n b^m \mid n > m\}$

$\quad = \{a^k a^m b^m \mid m \in \mathbb{Z}^{\geq 0}, k > 0\}$

$\quad \underset{aaagbbbb}{}$

$G_1 = (\{S_1, T_1\}, \{a, b\}, R_1, S_1)$

$\quad \underset{\cdots aga}{} S_1 \to a S_1 b \mid T_1$

$R_i \quad T_1 \to a T_1 \mid a$

$S \Rightarrow a S_1 b \Rightarrow aa S_1 bb \Rightarrow aa T_1 bb \Rightarrow aaabb$

$L_2 = \{a^n b^m \mid n < m\}$

$\quad = \{a^n b^n b^k \mid n \in \mathbb{Z}^{\geq 0}, k > 0\}$

$G_2 = (\{S_2, T_2\}, \{a, b\}, R_2, S_2)$

$\quad S_2 \to a S_2 b \mid T_2$

$\quad T_2 \to T_2 b \mid b$

$S_2 \Rightarrow a S_2 b \Rightarrow a T_2 b \Rightarrow a T_2 b b \Rightarrow a T_2 b b b \Rightarrow abbbb$

$G = (\{S, S_1, T_1, S_2, T_2\}, \{a, b\}, R, S)$

$R: \quad S \to S_1 \mid S_2 \qquad S_1 \to a S_1 b \mid T_1 \qquad S_2 \to a S_2 b \mid T_2$

$\qquad\qquad\qquad\qquad\quad T_1 \to a T_1 \mid a \qquad T_2 \to T_2 b \mid b$

$S \Rightarrow S_1 \Rightarrow a S_1 b \Rightarrow aa S_1 bb \Rightarrow aa T_1 bb \Rightarrow aaabb$

The class of context-free languages is closed under union.

Suppose L_1 and L_2 are context-free languages over Σ. **Goal**: $L_1 \cup L_2$ is also context-free.
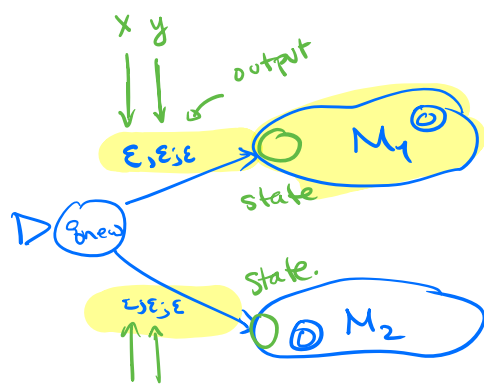
Approach 1: with PDAs

Let $M_1 = (Q_1, \Sigma, \Gamma_1, \delta_1, q_1, F_1)$ and $M_2 = (Q_2, \Sigma, \Gamma_2, \delta_2, q_2, F_2)$ be PDAs with $L(M_1) = L_1$ and $L(M_2) = L_2$.

Define $M = \left(Q_1 \cup Q_2 \cup \{q_{new}\}, \; \Sigma, \; \Gamma_1 \cup \Gamma_2, \; \delta, \; q_{new}, \; F_1 \cup F_2 \right)$

$T =$

$Q_1 \cap Q_2 = \varnothing$

$q_{new} \notin Q_1 \cup Q_2$

$\delta: Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \to \mathcal{P}\left(Q \times \Gamma_\varepsilon \right)$

$$\delta((q, x, y)) = \begin{cases} \{(q_1, \varepsilon), (q_2, \varepsilon)\} & q = q_{new} \\ & x = \varepsilon \\ & y = \varepsilon \\[6pt] \delta_1((q, x, y)) & q \in Q_1 \\ & x \in \Sigma_\varepsilon \\ & y \in \Gamma_{1\varepsilon} \\[6pt] \delta_2((q, x, y)) & q \in Q_2 \\ & x \in \Sigma_\varepsilon \\ & y \in \Gamma_{2\varepsilon} \\[6pt] \varnothing & \text{otherwi} \end{cases}$$
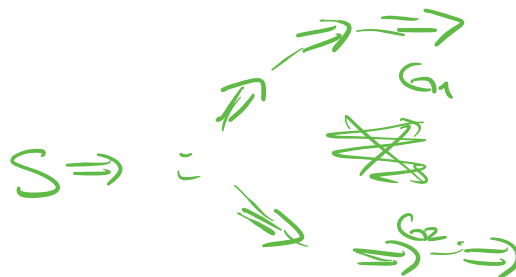
Approach 2: with CFGs

Let $G_1 = (V_1, \Sigma, R_1, S_1)$ and $G_2 = (V_2, \Sigma, R_2, S_2)$ be CFGs with $L(G_1) = L_1$ and $L(G_2) = L_2$.

Define $G = \left(V_1 \cup V_2 \cup \{S\}, \; \Sigma, \; R, \; S \right)$

$$R = R_1 \cup R_2 \cup \{S \to S_1 \mid S_2\}$$

$S \notin V_1 \cup V_2$

$V_1 \cap V_2 = \varnothing$

$S \Rightarrow \vdots$

Approach 1: with PDAs

Let $M_1 = (Q_1, \Sigma, \Gamma_1, \delta_1, q_1, F_1)$ and $M_2 = (Q_2, \Sigma, \Gamma_2, \delta_2, q_2, F_2)$ be PDAs with $L(M_1) = L_1$ and $L(M_2) = L_2$.

Define $M =$

*extra practice
(see Hw4)*

Approach 2: with CFGs

Let $G_1 = (V_1, \Sigma, R_1, S_1)$ and $G_2 = (V_2, \Sigma, R_2, S_2)$ be CFGs with $L(G_1) = L_1$ and $L(G_2) = L_2$.

Define $G =$

*extra practice
(see Hw4)*

Review Quiz: Closure of CFL under Star

Summary

Over a fixed alphabet Σ, a language L is **regular**

$$\text{iff it is described by some regular expression}$$
$$\text{iff it is recognized by some DFA}$$
$$\text{iff it is recognized by some NFA}$$

Over a fixed alphabet Σ, a language L is **context-free**

$$\text{iff it is generated by some CFG}$$
$$\text{iff it is recognized by some PDA}$$

Fact: Every regular language is a context-free language.

Fact: There are context-free languages that are not nonregular.

Fact: There are countably many regular languages.

Fact: There are countably inifnitely many context-free languages.

Consequence: Most languages are **not** context-free!

Examples of non-context-free languages

$$\{a^n b^n c^n \mid 0 \le n, n \in \mathbb{Z}\}$$
$$\{a^i b^j c^k \mid 0 \le i \le j \le k, i \in \mathbb{Z}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$$
$$\{ww \mid w \in \{0,1\}^*\}$$

(Sipser Ex 2.36, Ex 2.37, 2.38)

There is a Pumping Lemma for CFL that can be used to prove a specific language is non-context-free: If A is a context-free language, there there is a number p where, if s is any string in A of length at least p, then s may be divided into five pieces $s = uvxyz$ where (1) for each $i \ge 0$, $uv^i x y^i z \in A$, (2) $|uv| > 0$, (3) $|vxy| \le p$. *We will not go into the details of the proof or application of Pumping Lemma for CFLs this quarter.*

Review: Week 4 Wednesday

Please complete the review quiz questions on Gradescope about context-free languages

Pre class reading for next time: Pages 165-167

Friday April 29

Regular languages over Σ

A set X is said to be **closed** under an operation OP if, for any elements in X, applying OP to them gives an element in X.

True/False	Closure claim
True	The set of integers is closed under multiplication. $\forall x \forall y\, (\, (x \in \mathbb{Z} \wedge y \in \mathbb{Z}) \to xy \in \mathbb{Z}\,)$
True	For each set A, the power set of A is closed under intersection. $\forall A_1 \forall A_2\, (\, (A_1 \in \mathcal{P}(A) \wedge A_2 \in \mathcal{P}(A) \in \mathbb{Z}) \to A_1 \cap A_2 \in \mathcal{P}(A)\,)$
True	The class of regular languages over Σ is closed under complementation.
True	The class of regular languages over Σ is closed under union.
True	The class of regular languages over Σ is closed under intersection.
True	The class of regular languages over Σ is closed under concatenation.
True	The class of regular languages over Σ is closed under Kleene star.
False Section 2.2.	The class of context-free languages over Σ is closed under complementation.
True	The class of context-free languages over Σ is closed under union. CFG
False Section 2.2.	The class of context-free languages over Σ is closed under intersection.
True HW4.	The class of context-free languages over Σ is closed under concatenation.
True Review Quiz.	The class of context-free languages over Σ is closed under Kleene star. CFG.

A language is context-free means there's a CFG that generates it, or there's a PDA that recognizes it.

Assume $\Sigma = \{0, 1, \#\}$

Extra practice

	Σ^*	Regular / ~~nonregular and context-free~~ / ~~not context-free~~
	$\{0^i \# 1^j \mid i \geq 0, j \geq 0\}$	Regular / ~~nonregular and context-free~~ / ~~not context-free~~
	$\{0^i 1^j \# 1^j 0^i \mid i \geq 0, j \geq 0\}$	~~Regular~~ / nonregular and context-free / ~~not context-free~~
	$\{0^i 1^j \# 0^i 1^j \mid i \geq 0, j \geq 0\}$	~~Regular~~ / ~~nonregular and context-free~~ / not context-free

Ch 1 Regular Languages   *Ch 2 CFL*

Turing machines: unlimited read + write memory, unlimited time (computation can proceed without "consuming" input and can re-read symbols of input)

- Division betweeen program (CPU, state diagram) and data

- Unbounded memory gives theoretical limit to what modern computation (including PCs, supercomputers, quantum computers) can achieve

- State diagram formulation is simple enough to reason about (and diagonalize against) while expressive enough to capture modern computation

For Turing machine $M = (Q, \Sigma, \Gamma, \delta, q_0, q_{accept}, q_{reject})$ the **computation** of M on a string w over Σ is:

set of states ↓ *input alphabet* ↓ *tape alphabet* ↓ *transition function* ↓ $w \in \Sigma^*$

- Read/write head starts at leftmost position on tape.

- Input string is written on $|w|$-many leftmost cells of tape, rest of the tape cells have the blank symbol. **Tape alphabet** is Γ with $\sqcup \in \Gamma$ and $\Sigma \subseteq \Gamma$. The blank symbol $\sqcup \notin \Sigma$.

- Given current state of machine and current symbol being read at the tape head, the machine transitions to next state, writes a symbol to the current position of the tape head (overwriting existing symbol), and moves the tape head L or R (if possible). Formally, **transition function** is

$$\delta : Q \times \Gamma \to Q \times \Gamma \times \{L, R\}$$
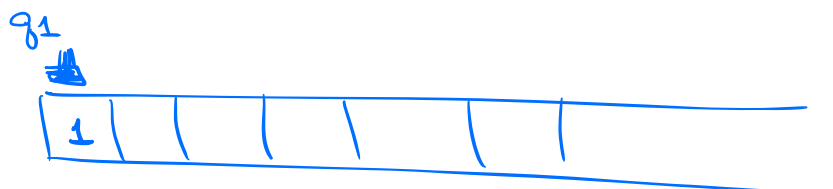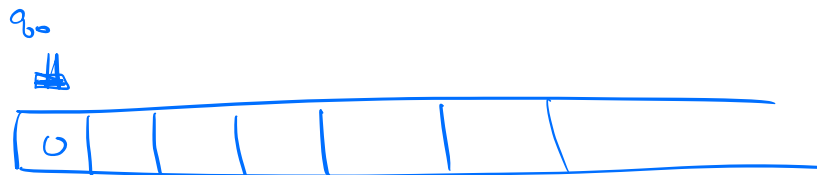
new state · *write symbol to tape* · *move tape head*

- Computation ends if and when machine enters either the accept or the reject state. This is called **halting**. Note: $q_{accept} \neq q_{reject}$.
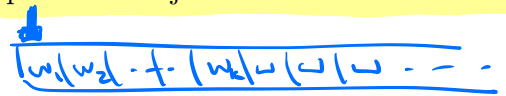
The **language recognized by the Turing machine** M, is

$\{w \in \Sigma^* \mid$ the computation of M on w halts after entering the accept state$\} = \{w \in \Sigma^* \mid w$ is accepted by $M\}$

$$\delta\left((q_0, 0)\right) = (q_1, 1, L)$$
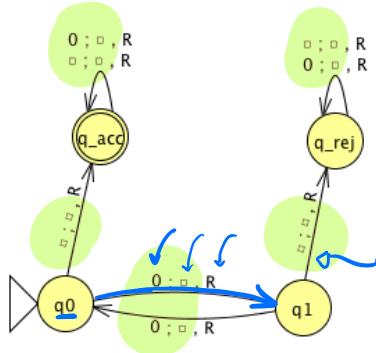
Deterministic (left margin)

An example Turing machine: $\Sigma = \{0\}$, $\Gamma = \{0, \sqcup\}$

$\delta((q0,0)) = (q1, \sqcup, R)$

JFLAP · $\square$ $\sqcup$

Formal definition: $(\{q0, q1, q_{acc}, q_{rej}\}, \{0\}, \{0, \sqcup\}, \delta, q0, q_{acc}, q_{rej})$

Sample computation: input 000

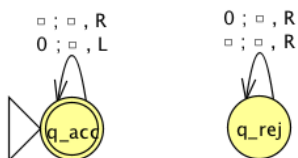| $q0 \downarrow$						
0	0	0	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$
$q1 \downarrow$						
$\sqcup$	0	0	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$
	$q0 \downarrow$					
$\sqcup$	$\sqcup$	0	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$
			$q1 \downarrow$			
$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$
			$q_{rej} \downarrow$			
$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$	$\sqcup$

$\delta((q1,0)) = (q0, \sqcup, R)$

This machine rejects 000.

The language recognized by this machine is ... $\{w \in \{0^*\} \mid w \text{ has even length}\}$
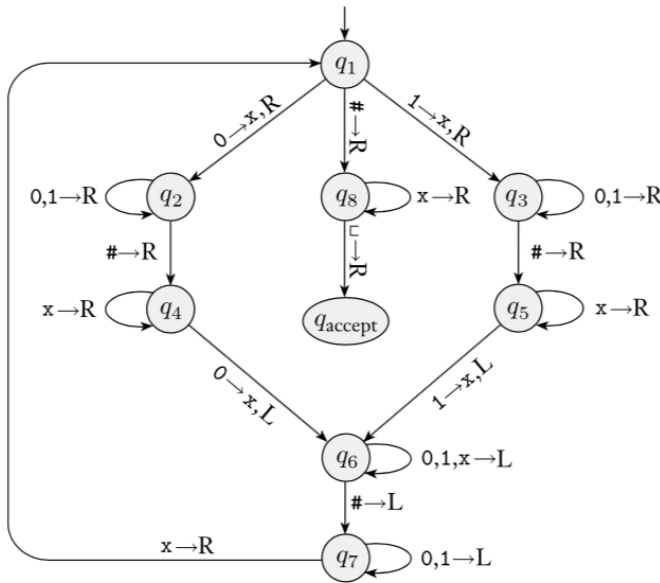
Extra practice:

Formal definition:

Sample computation:

Monday ...

Sipser Figure 3.10

Conventions in state diagram of TM: $b \to R$ label means $b \to b, R$ and all arrows missing from diagram represent transitions with output $(q_{reject}, \llcorner, R)$

Computation on input string $01\#01$

$q_1 \downarrow$						
0	1	#	0	1	⌞	⌞

Implementation level description of this machine:

Zig-zag across tape to corresponding positions on either side of # to check whether the characters in these positions agree. If they do not, or if there is no #, reject. If they do, cross them off.

Once all symbols to the left of the # are crossed off, check for any un-crossed-off symbols to the right of #; if there are any, reject; if there aren't, accept.

The language recognized by this machine is

$$\{w\#w \mid w \in \{0,1\}^*\}$$

Extra practice

Computation on input string 01#1

q₁ ↓						
0	1	#	1	␣	␣	␣

Review: Week 4 Friday

Please complete the review quiz questions on Gradescope about Turing machines.

Pre class reading for next time: Definitions 3.5, 3.6 (page 170)