Properties of Context Free Languages

Pallab Dasgupta,
Professor, Dept. of Computer Sc & Engg

Theorem: CFLs are closed under concatenation

If L_1 and L_2 are CFLs, then L_1L_2 is a CFL.

Proof:

1. Let L_1 and L_2 be generated by the CFG, $G_1 = (V_1, T_1, P_1, S_1)$ and $G_2 = (V_2, T_2, P_2, S_2)$, respectively.
2. Without loss of generality, subscript each nonterminal of G_1 with a 1, and each nonterminal of G_2 with a 2 (so that $V1 \cap V2 = \Phi$).
3. Define the CFG, G, that generates L_1L_2 as follows:
   $G = (V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, P_1 \cup P_2 \cup \{S \rightarrow S_1S_2\}, S)$.
4. Each word generated thus is a word in L_1 followed by a word in L_2.

Example

- **Let L_1 be PALINDROME, defined by:**

  $$S \rightarrow aSa \mid bSb \mid a \mid b \mid \lambda$$

- **Let L_2 be $\{ a^n b^n \mid n \geq 0\}$ defined by:**

  $$S \rightarrow aSb \mid \lambda$$

- **Then the concatenation language is defined by:**

  $$S \rightarrow S_1 S_2$$
  $$S_1 \rightarrow aS_1a \mid bS_1b \mid a \mid b \mid \lambda$$
  $$S_2 \rightarrow aS_2b \mid \lambda$$

Theorem: CFLs are closed under Kleene star

If L_1 is a CFL, then L_1^* is a CFL.

Proof:

1. Let L_1 be generated by the CFG, $G_1 = (V_1, T_1, P_1, S_1)$.
2. Without loss of generality, subscript each nonterminal of G_1 with a 1.
3. Define the CFG, G, that generates L_1^* as follows:
   $G = (V_1 \cup \{S\}, T_1, P_1 \cup \{S \rightarrow S_1 S \mid \lambda\}, S)$.
4. Each word generated is either λ or some sequence of words in L_1.
5. Every word in L_1^* (i.e., some sequence of 0 or more words in L_1) can be generated by G.

Example

- Let L_1 be $\{ a^n b^n \mid n \geq 0\}$ defined by:
  $S \rightarrow aSb \mid \lambda$
- Then L_1^* is generated by:
  $S \rightarrow S_1 S \mid \lambda$
  $S_1 \rightarrow aS_1 b \mid \lambda$

None of these example grammars is necessarily the most *compact* CFG for the language it generates.

Intersection and Complement

Theorem: CFLs are not closed under intersection

If L_1 and L_2 are CFLs, then $L_1 \cap L_2$ may not be a CFL.

Proof:

1. $L_1 = \{ a^n b^n a^m \mid n,m \geq 0 \}$ is generated by the following CFG:
   $S \rightarrow XA$
   $X \rightarrow aXb \mid \lambda$
   $A \rightarrow Aa \mid \lambda$

2. $L_2 = \{ a^n b^m a^m \mid n,m \geq 0 \}$ is generated by the following CFG:
   $S \rightarrow AX$
   $X \rightarrow aXb \mid \lambda$
   $A \rightarrow Aa \mid \lambda$

3. $L_1 \cap L_2 = \{ a^n b^n a^n \mid n \geq 0 \}$, which is known not to be a CFL (pumping lemma).

Theorem: CFLs are not closed under complement

If L_1 is a CFL, then $\overline{L_1}$ may not be a CFL.

Proof:

They are closed under union. If they are closed under complement, then they are closed under intersection, which is false. More formally,

1. Assume the complement of every CFL is a CFL.
2. Let L_1 and L_2 be 2 CFLs.
3. Since CFLs are close under union, and we are assuming they are closed under complement,
   $\overline{\overline{L_1} \cup \overline{L_2}} = L_1 \cap L_2$ is a CFL.
4. However, we know there are CFLs whose intersection is not a CFL.
5. Therefore, our assumption that CFLs are closed under complement is false.

Theorem: The intersection of a CFL and an RL is a CFL.

If L_1 is a CFL and L_2 is regular, then $L_1 \cap L_2$ is a CFL.

Proof:

1. We do this by constructing a PDA I to accept the intersection that is based on a PDA A for L_1 and a FA F for L_2.
2. Convert A, if necessary, so that all input is read before accepting.
3. Construct a set Y of all A's states y_1, y_2, ... and a set X of all F's states x_1, x_2, ……
4. Construct { (y,x) | for all y ϵ Y, for all x ϵ X }.
5. The start state of I is (y_0, x_0), where y_0 is the label of A's start state, and x_0 is F's initial state.

Theorem (contd.)

6. Regarding the next state function, the x component changes only when the PDA is in a READ state:

- If in (y_i , x_j) and y_i is not a READ state, its successor is (y_k , x_j), where y_k is the appropriate successor of y_i .

- If in (y_i , x_j) and y_i is a READ state, reading a, its successor is (y_k , x_l), where
  - -- y_k is the appropriate successor of y_i on an a.
  - -- $\delta(x_j , a) = x_l$.

7. I's ACCEPT states are those where the y component is ACCEPT and the x component is final.

If the y component is ACCEPT and the x component is not final, the state in I is REJECT (or omitted, implying a crash).

Example

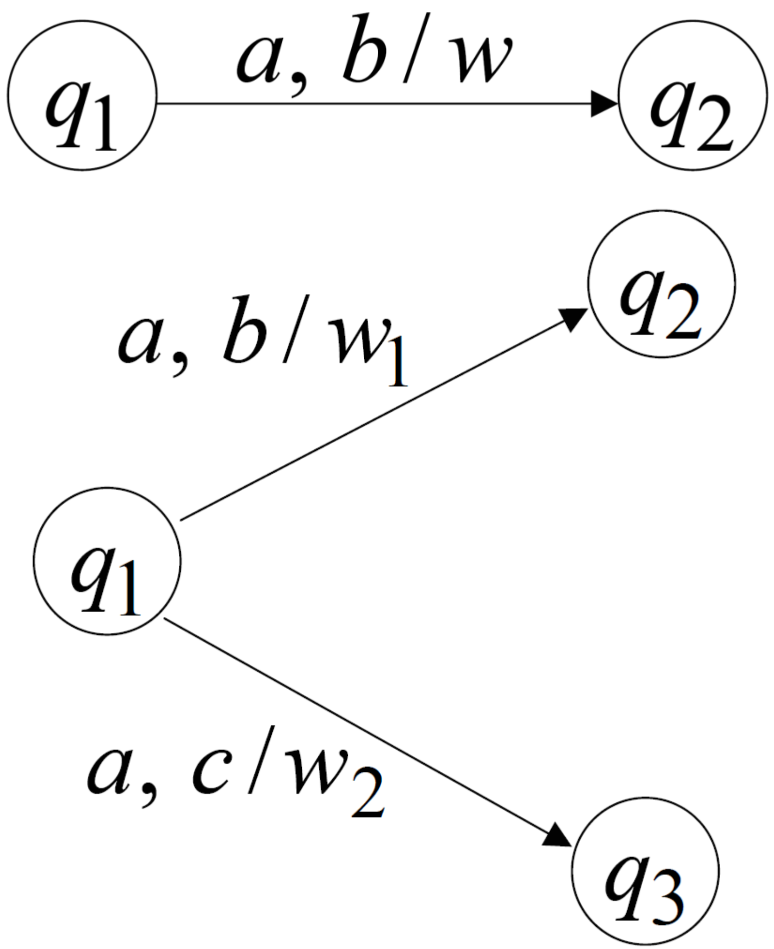
- Let L_1 be the CFL EQUAL of words with an equal number of a's and b's.
  Draw its PDA.

- Let L2 = $(a + b)^*a$.
  Draw its FA.

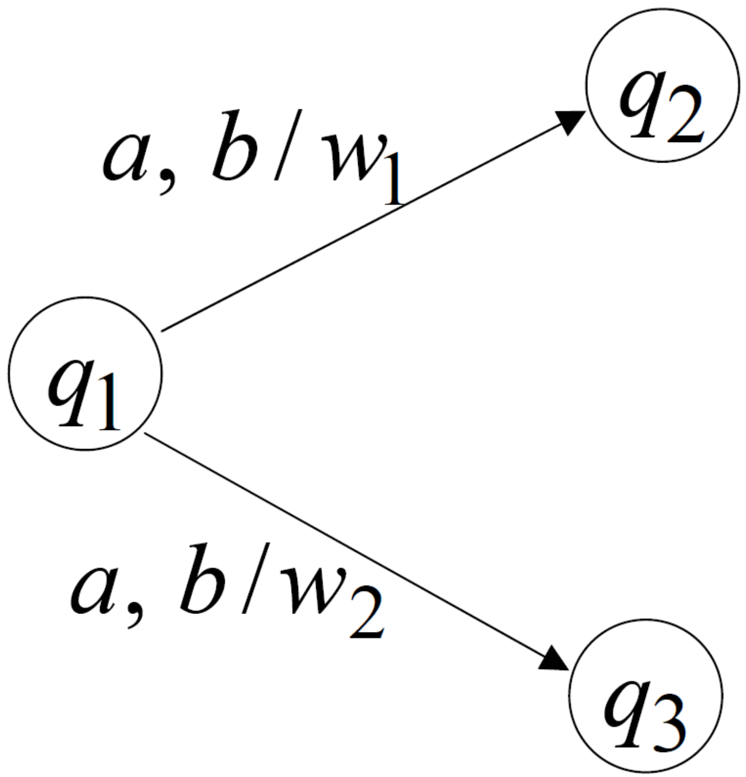
- Perform the construction of the intersection PDA.

DPDAs

Allowed

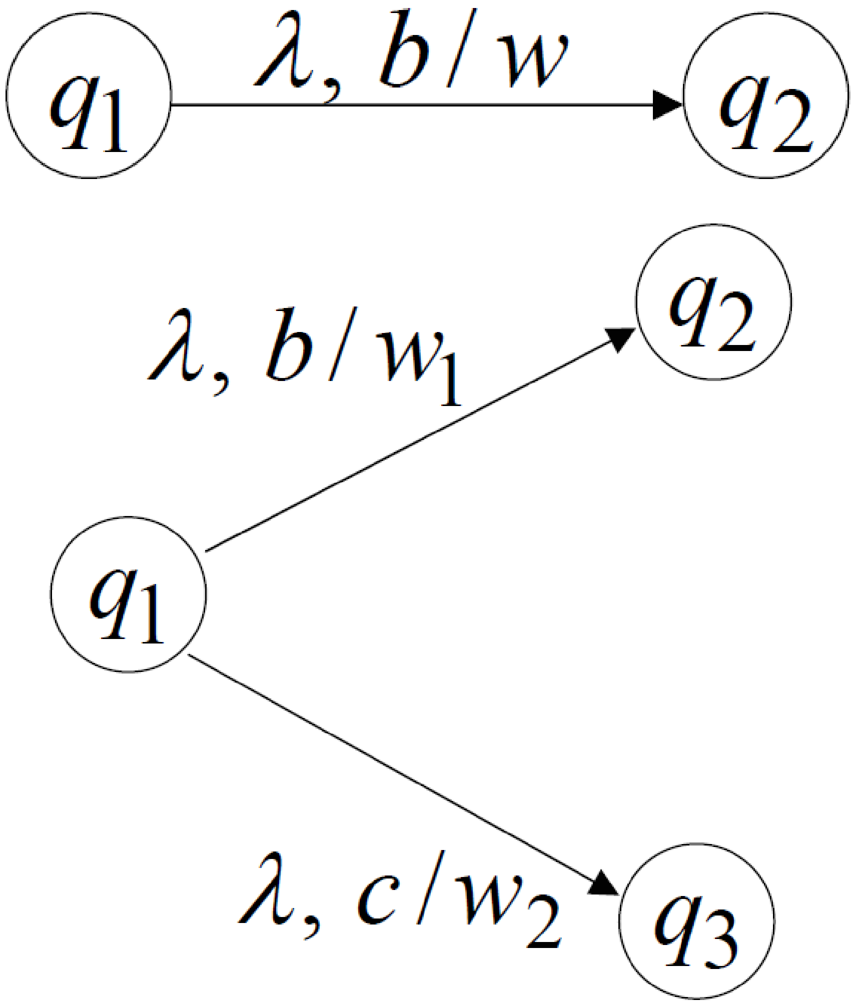


Not Allowed

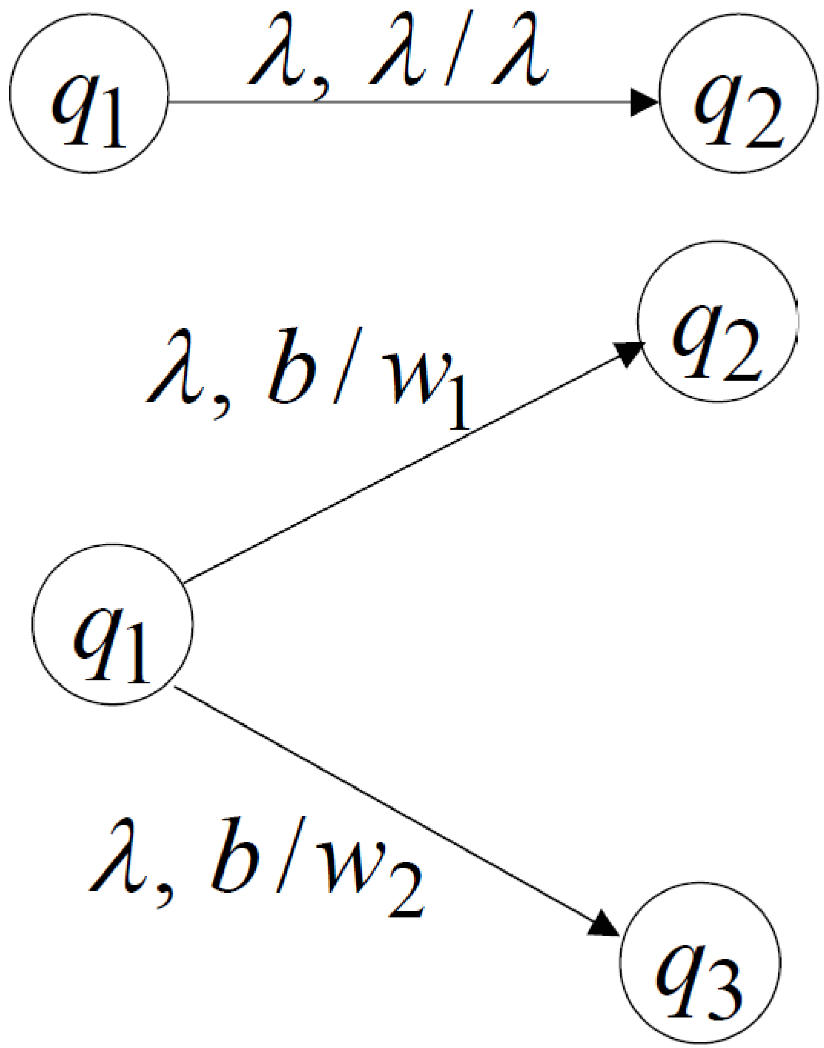


DPDAs

Allowed:
Something must be matched from the stack



Not Allowed



NPDAs are more powerful than DPDAs

We will show that:

- There is a context-free language L (accepted by an NPDA)
- L is not deterministic context-free (not accepted by a DPDA)

The language is: $L = \{a^n b^n\} \cup \{a^n b^{2n}\}\ n \geq 0$

The language L is context-free. The Context-free grammar for L

- $S \rightarrow S1 \mid S2$
- $S1 \rightarrow aS1b \mid \lambda$
- $S2 \rightarrow aS2bb \mid \lambda$

There is an NPDA that accepts L

Theorem: $L = \{a^n b^n\} \cup \{a^n b^{2n}\}$ is not deterministic context-free

The language : $L = \{a^n b^n\} \cup \{a^n b^{2n}\}$ CFL (i.e. there is no DPDA that accepts L).

- Each 'a' is to be matched by either one or two 'b's. An initial choice must be made.

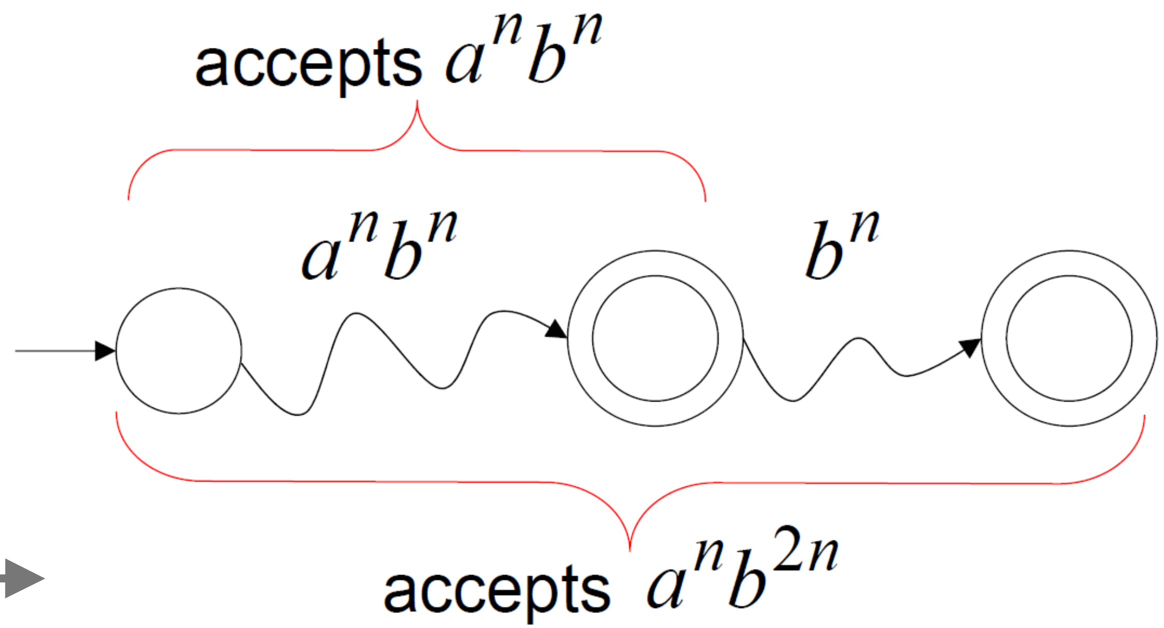
Proof: (by contradiction)

Assume the opposite, i.e. that L is deterministic context free.

Therefore: there is a DPDA M that accepts L.

1. Fact 1: The language $\{a^n b^n c^n\}$ is not context-free. (by Pumping Lemma)
2. Fact 2: The language $L \cup \{a^n b^n c^n\}$ is not context-free. (by Pumping Lemma)

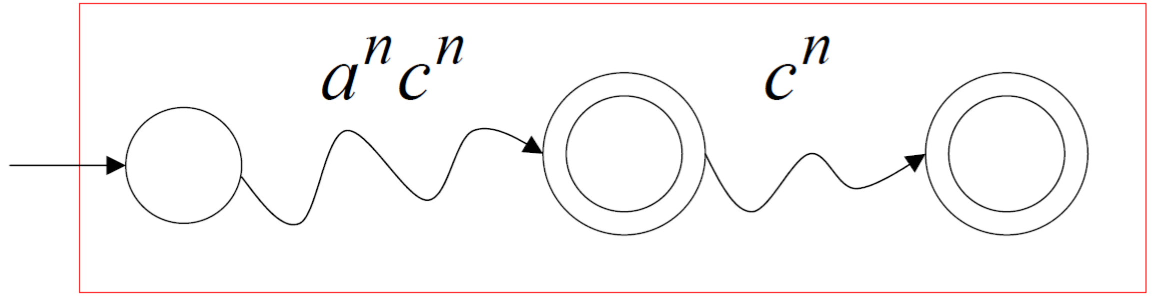
➢ But, if we have a DPDA like this one here

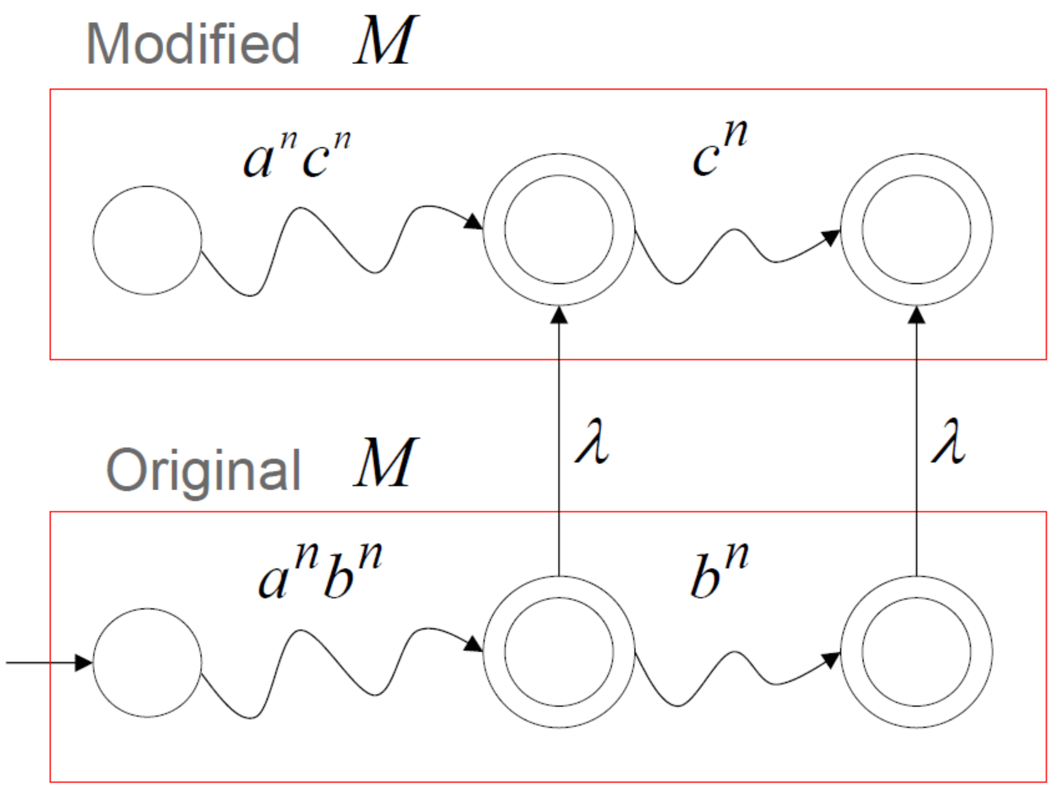


then we can construct an NPDA to accept $L \cup \{a^n b^n c^n\}$. HOW?

Theorem: $L = \{a^n b^n\} \cup \{a^n b^{2n}\}$ is not deterministic context-free

We modify *M*. Replace *b* with *c*





The DPDA for $\{a^n b^n\} \cup \{a^n b^{2n}\}$ and $\{a^n c^n\} \cup \{a^n c^{2n}\}$ are combined to create an NPDA for $L \cup \{a^n b^n c^n\}$

- *Therefore* $L \cup \{a^n b^n c^n\}$ is a CFL.

- *But by the Pumping Lemma we know that* $L \cup \{a^n b^n c^n\}$ is not a CFL.
- Therefore our assumption that we can have a DPFA for L is wrong.
- Therefore L is not deterministic context free