CHAPTER 8

Properties Of Context-Free Languages

By R.Ameri

Today's Lecture

- The Pumping Lemma For
  - Context-free languages
  - Linear context-free languages
- Closure properties for context-free languages
- Decidable problems for context-free languages

Linear Context-Free Grammar

A linear grammar is a context-free grammar that has at most one non-terminal / variable in the right hand side of each of its productions.

Linear languages are a strict subset of the context-free languages.

Linear Context-Free Language

A **linear context-free language** is a language generated by some linear grammar.

$$L = \{a^n b^n : n \geq 1\}$$

$$S \to aSb \mid ab$$

Non-Linear Context-Free Language

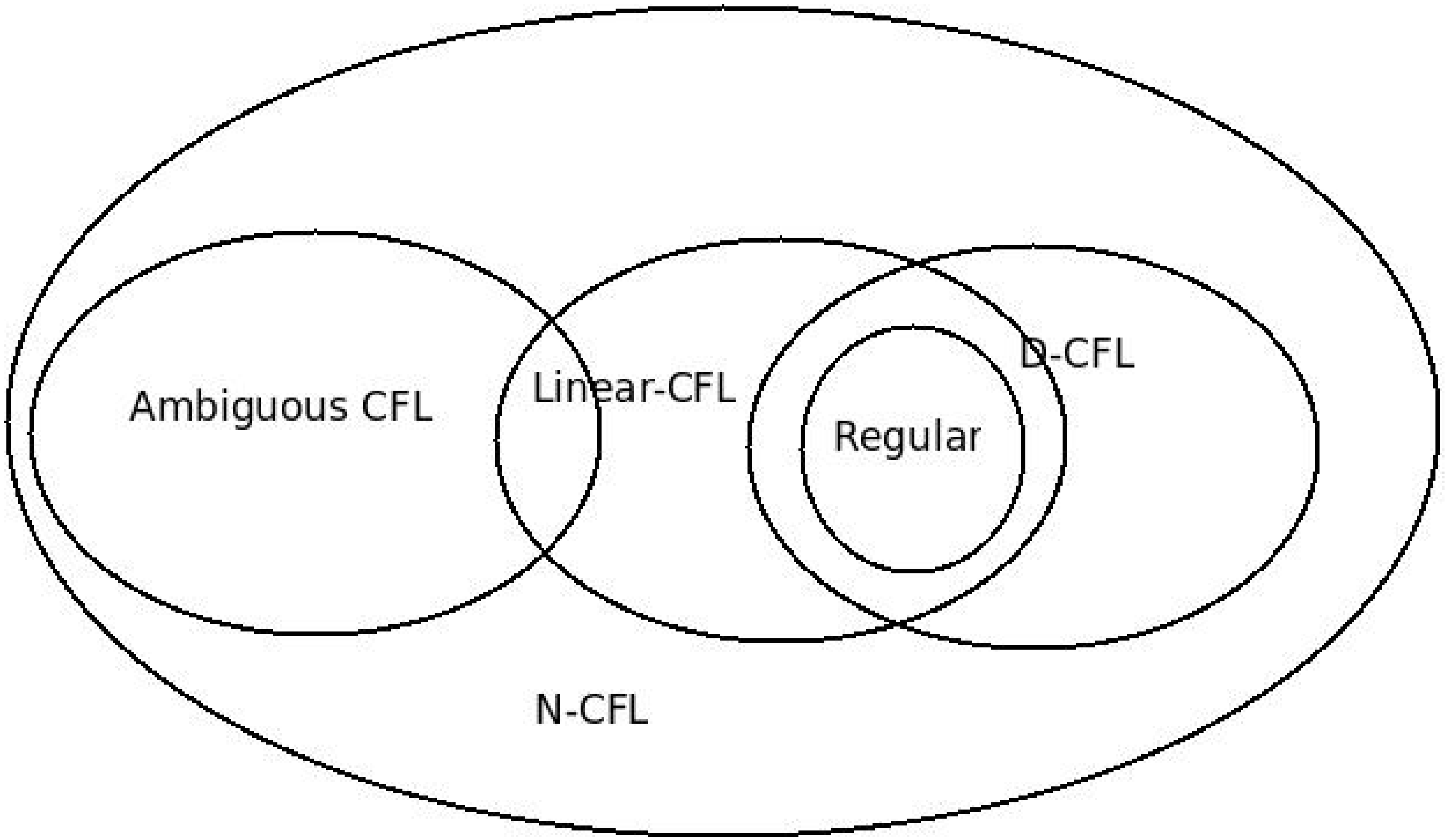
A non-**linear context-free language** is a language that can't generated by any linear grammar.

$$L = \{w \mid w \in \{a,b\}^*,\ n_a = n_b\}$$

$$G1: S \rightarrow aSbS \mid bSaS \mid \lambda$$

$$G2: S \rightarrow SS \mid asb \mid bSa \mid \lambda$$

Venn-diagram for Chomsky classification of formal languages



The Pumping Lemma for Context-Free Languages

The Pumping Lemma:

For infinite context-free language L

there exists an integer m such that

for any string $w \in L, \quad |w| \geq m$

we can write $w = uvxyz$

with lengths $|vxy| \leq m$ and $|vy| \geq 1$

and it must be:

$$uv^i xy^i z \in L, \quad \text{for all } i \geq 0$$

Non-context free languages

$$\{a^n b^n c^n : n \geq 0\}$$

Context-free languages

$$\{a^n b^n : n \geq 0\}$$

Theorem: The language

$$L = \{a^n b^n c^n : n \geq 0\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages

$$L = \{a^n b^n c^n : n \geq 0\}$$

Assume for contradiction that L is context-free

Since L is context-free and infinite we can apply the pumping lemma

$$L = \{a^n b^n c^n : n \geq 0\}$$

Pumping Lemma gives a magic number m such that:

Pick any string $w \in L$ with length $|w| \geq m$

We pick: $w = a^m b^m c^m$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

We can write: $w = uvxyz$

with lengths $|vxy| \leq m$ and $|vy| \geq 1$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Pumping Lemma says:

$uv^i xy^i z \in L$ for all $i \geq 0$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 1: vxy is within a^m

$$\overbrace{\underbrace{aa}_{u}\underbrace{a\ldots aa}_{vxy}a}^{m}\underbrace{\overbrace{bbb\ldots bbb}^{m}\;\overbrace{ccc\ldots ccc}^{m}}_{z}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 1: v and y consist from only a

$$\underbrace{\overbrace{\underbrace{aa}_{u}\underbrace{a...aa}_{vxy}\underbrace{a}}^{m}\underbrace{\overbrace{bbb...bbb}^{m}\,\overbrace{ccc...ccc}^{m}}_{z}}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 1: Repeating v and y

$$k \geq 1$$

$$\underbrace{a}_{u}\underbrace{\overbrace{aaaaa...aaaaaa}^{m+k}}_{v^2xy^2}\underbrace{\overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}}_{z}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 1: From Pumping Lemma: $uv^2xy^2z \in L$

$$k \geq 1$$

$$\underbrace{\overbrace{aaaaaa...aaaaaa}^{m+k}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}}_{}$$

$$\underbrace{aa}_{u}\underbrace{aaaa...aaaaa}_{v^2xy^2}\underbrace{a\ bbb...bbb\ ccc...ccc}_{z}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 1: From Pumping Lemma: $uv^2xy^2z \in L$

$$k \geq 1$$

However: $uv^2xy^2z = a^{m+k}b^mc^m \notin L$

Contradiction!!!

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 2: vxy is within b^m

$$\overbrace{aaa...aaa}^{m}\,\overbrace{bbb...bbb}^{m}\,\overbrace{ccc...ccc}^{m}$$

$$\underbrace{aaa...aaa\,bb}_{u}\underbrace{b...bb}_{vxy}\underbrace{b\,ccc...ccc}_{z}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 2: Similar analysis with case 1

$$\underbrace{\overbrace{aaa...aaa}^{m}\ \overbrace{bbb}^{}}_{u}\!\!\!\!\!\!\!\!\!\!\!\!\underbrace{...bbb}_{vxy}\ \underbrace{\overbrace{ccc...ccc}^{m}}_{z}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 3: vxy is within c^m

$$\underbrace{\overbrace{aaa\ldots aaa}^{m}\ \overbrace{bbb\ldots bbb}^{m}\ cc}_{u}\overbrace{\underbrace{c\ldots c}_{vxy}\ \underbrace{cc}_{z}}^{m}$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 3: Similar analysis with case 1

$$\overbrace{aaa...aaa}^{m}\;\overbrace{bbb...bbb}^{m}\;\overbrace{ccc...ccc}^{m}$$

$$\underbrace{aaa...aaa\;bbb...bbb\;c}_{u}\underbrace{cc...c}_{vxy}\underbrace{cc}_{z}$$

In all cases we obtained a contradiction

Therefore: The original assumption that

$$L = \{a^n b^n c^n : n \geq 0\}$$

is context-free must be wrong

Conclusion: L is not context-free

Theorem: The language

$$L = \{a^n b^n\} \cup \{a^n b^{2n}\} \cup \{a^n b^n c^n\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages ???

Theorem: The language

$$L = \{ww : w \in \{a,b\}^*\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages

$$L = \{ww : w \in \{a,b\}^*\}$$

Assume for contradiction that L is context-free

Since L is context-free and infinite we can apply the pumping lemma

$$L = \{ww : w \in \{a,b\}^*\}$$

$$w = a^m b^m a^m b^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Pumping Lemma says:

$uv^i xy^i z \in L$ for all $i \geq 0$

$$L = \{ww : w \in \{a,b\}^*\}$$

$$w = a^m b^m a^m b^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case:

$$\overbrace{aaa\ldots aaa}^{m}\;\overbrace{bbb\ldots bbb}^{m}\;\overbrace{aaa\ldots aaa}^{m}\;\overbrace{bbb\ldots bbb}^{m}$$

$$u \quad v \quad x \quad y \quad z$$

$$L = \{ww : w \in \{a,b\}^*\}$$

Case:

From Pumping Lemma: $uv^0xy^0z \in L$

$$\overbrace{aaa...aaa}^{m}\;\overbrace{bbb...bbb}^{m}\;\overbrace{aaa...aaa}^{m}\;\overbrace{bbb...bbb}^{m}$$

$$\underbrace{aaa}_{u}\underbrace{...aa}_{v^0}\underbrace{a\;bbb}_{x}\underbrace{...bbb}_{y^0}\;\underbrace{aaa...aaa\;bbb...bbb}_{z}$$

However: $uv^0xy^0z = a^k b^j a^m b^n \notin L, k < m; j < m$

Contradiction!!!

Theorem: The language

$$L = \{a^{n!} : n \geq 0\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages

$$L = \{a^{n!} : n \geq 0\}$$

Assume for contradiction that L is context-free

Since L is context-free and infinite we can apply the pumping lemma

$$L = \{a^{n!} : n \geq 0\}$$

$$w = a^{m1} a^{k} a^{m2} a^{j} a^{m3}; k + j \geq 1$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

$$m1 + k + m2 + j + m3 = m!$$

Pumping Lemma says:

$$uv^i xy^i z \in L$$ for all $$i \geq 0$$

$$L = \{a^{n!} : n \geq 0\}$$

$\boxed{\text{Case 1:}}$ From Pumping Lemma: $uv^0xy^0z \in L$

However:

$$uv^0xy^0z = a^{m1}a^{m2}a^{m3} = a^{m!-(k+j)} \notin L,$$

$$(m-1)! \leq m! - (k+j) \leq m!$$

Contradiction!!!

Theorem: The language

$$L = \{ww^R w \mid w \in \{a,b\}^*\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages

$$L = \{ww^R w \mid w \in \{a,b\}^*\}$$

Assume for contradiction that L is context-free

Since L is context-free and infinite we can apply the pumping lemma

$$L = \{ww^R w \mid w \in \{a,b\}^*\}$$

$$w = a^m b^m b^m a^m a^m b^m$$

$$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

$$v = a^k; y = a^{k1}; k + k1 \geq 1$$

Pumping Lemma says:

$$uv^i xy^i z \in L$$ for all $i \geq 0$

$$L = \{ww^R w \mid w \in \{a,b\}^*\}$$

Case 1: From Pumping Lemma: $uv^0xy^0z \in L$

However:

$$uv^0xy^0z = a^{m-k-k1}b^m b^m a^m a^m b^m \notin L$$

Contradiction!!!

The Pumping Lemma for Linear Context-Free Languages(LCFL)

The Pumping Lemma for LCFL:

For infinite linear context-free language L

there exists an integer m such that

for any string $w \in L, \quad |w| \geq m$

we can write $w = uvxyz$

with lengths $|uvyz| \leq m$ and $|vy| \geq 1$

and it must be:

$$uv^i xy^i z \in L, \quad \text{for all } i \geq 0$$

Theorem: The language

$$L(M) = \{w \mid w \in \{a,b\}^*: \quad n_a(w) = n_b(w)\}$$

is **not** linear context free

Proof: Use the Pumping Lemma
for linear context-free languages

$$L(M) = \{w \mid w \in \{a,b\}^*: \quad n_a(w) = n_b(w)\}$$

Assume for contradiction that L is linear context-free

Since L is linear context-free and infinite we can apply the pumping lemma

$$L(M) = \{w \mid w \in \{a,b\}^* : \quad n_a(w) = n_b(w)\}$$

$$w = a^m b^m b^m a^m$$

$$w = uvxyz \qquad |uvyz| \le m \quad \text{and} \quad |vy| \ge 1$$

$uv^i xy^i z \in L$ for all $i \ge 0$

$$\overbrace{aaa\ldots aaa}^{m}\;\overbrace{bbb\ldots bbb}^{m}\;\overbrace{bbb\ldots bbb}^{m}\overbrace{aaa\ldots aaa}^{m}$$

$$u \qquad v \qquad\qquad\qquad x \qquad\qquad\qquad y \qquad z$$

$$L(M)=\{w \mid w \in \{a,b\}^*: \quad n_a(w)=n_b(w)\}$$

Case: From Pumping Lemma: $uv^0xy^0z \in L$

$$\overbrace{aaa...aaa}^{m}\,\overbrace{bbb...bbb}^{m}\,\overbrace{bbb...bbb}^{m}\overbrace{aaa...aaa}^{m}$$

$$\underbrace{aa}_{u}\underbrace{a...aaa}_{v}\,\underbrace{bbb...bbb\,bbb...bbba}_{x}\underbrace{aa...}_{y}\underbrace{aaa}_{z}$$

However:

$$uv^0xy^0z = a^{k1}(a^k)a^{k2}b^{2m}a^{k3}(a^{k4})a^{k5}$$

$$= a^{k1}a^{k2}b^{2m}a^{k3}a^{k5} \notin L, k+k4 \geq 1$$

Contradiction!!!

Closure properties for context-free languages

Union

Context-free languages
are closed under: **Union**

L_1 is context free

L_2 is context free

$\Rightarrow$ $L_1 \cup L_2$ is context-free

Example

Language | **Grammar**

$$L_1 = \{a^n b^n\}$$

$$S_1 \to aS_1b \mid \lambda$$

$$L_2 = \{ww^R\}$$

$$S_2 \to aS_2a \mid bS_2b \mid \lambda$$

Union

$$L = \{a^n b^n\} \cup \{ww^R\}$$

$$S \to S_1 \mid S_2$$

In general:

For context-free languages $L_1, \; L_2$

with context-free grammars $G_1, \; G_2$

and start variables $S_1, \; S_2$

The grammar of the **union** $L_1 \cup L_2$

has new start variable S

and additional production $S \rightarrow S_1 \mid S_2$

Concatenation

Context-free languages are closed under: **Concatenation**

L_1 is context free

L_2 is context free

$\Rightarrow$ $L_1 L_2$ is context-free

Example

Language	**Grammar**
$L_1 = \{a^n b^n\}$	$S_1 \to aS_1 b \mid \lambda$
$L_2 = \{ww^R\}$	$S_2 \to aS_2 a \mid bS_2 b \mid \lambda$

<u>Concatenation</u>

$L = \{a^n b^n\}\{ww^R\}$ | $S \to S_1 S_2$

In general:

For context-free languages $L_1, \ L_2$

with context-free grammars $G_1, \ G_2$

and start variables $S_1, \ S_2$

The grammar of the **concatenation** $L_1 L_2$

has new start variable S

and additional production $S \rightarrow S_1 S_2$

Star Operation

Context-free languages
are closed under: **Star-operation**

L is context free ⟹ L^* is context-free

Example

Language

$$L = \{a^n b^n\}$$

Grammar

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

Star Operation

$$L = \{a^n b^n\}^*$$

$$S_1 \rightarrow SS_1 \mid \lambda$$

In general:

For context-free language L
with context-free grammar G
and start variable S

The grammar of the **star operation** L^*
has new start variable S_1
and additional production $S_1 \rightarrow SS_1 \mid \lambda$

Intersection

Context-free languages

are <u>**not**</u> closed under: **intersection**

L_1 is context free

L_2 is context free

$\Longrightarrow$ $L_1 \cap L_2$

<u>**not**</u> necessarily context-free

Example

$$L_1 = \{a^n b^n c^m\}$$

Context-free:

$$S \to AC$$

$$A \to aAb \mid \lambda$$

$$C \to cC \mid \lambda$$

$$L_2 = \{a^n b^m c^m\}$$

Context-free:

$$S \to AB$$

$$A \to aA \mid \lambda$$

$$B \to bBc \mid \lambda$$

Intersection

$$L_1 \cap L_2 = \{a^n b^n c^n\}$$ **NOT** context-free

Complement

Context-free languages
are **not** closed under: **complement**

L is context free $\Rightarrow$ $\overline{L}$ **not** necessarily context-free

Example

$$L_1 = \{a^n b^n c^m\}$$

Context-free:

$$S \rightarrow AC$$

$$A \rightarrow aAb \mid \lambda$$

$$C \rightarrow cC \mid \lambda$$

$$L_2 = \{a^n b^m c^m\}$$

Context-free:

$$S \rightarrow AB$$

$$A \rightarrow aA \mid \lambda$$

$$B \rightarrow bBc \mid \lambda$$

Complement

$$\overline{\overline{L_1} \cup \overline{L_2}} = L_1 \cap L_2 = \{a^n b^n c^n\}$$

NOT context-free

Reverse

Context-free languages
are closed under: **Reverse**

L is context free $\Rightarrow$ L^R context-free

Subtraction

Context-free languages
are **<u>not</u>** closed under: **Subtraction**

L_1 is context free

L_2 is context free

$\Longrightarrow$

$L_1 - L_2 =$

$L_1 \cap \overline{L_2}$

<u>not</u> necessarily
context-free

Homomorphism

Let h be a homomorphism. If L is a Context-free language, then its homomorphic image $h(L)$ is also Context-free.

The family of Context-free languages is therefore closed under arbitrary homomorphisms.

Intersection of Context-free languages and Regular Languages

The intersection of

a context-free language and

a regular language

is a context-free language

L_1 context free

L_2 regular

$\Rightarrow$ $L_1 \cap L_2$ context-free

Machine M_1

NPDA for L_1

context-free

Machine M_2

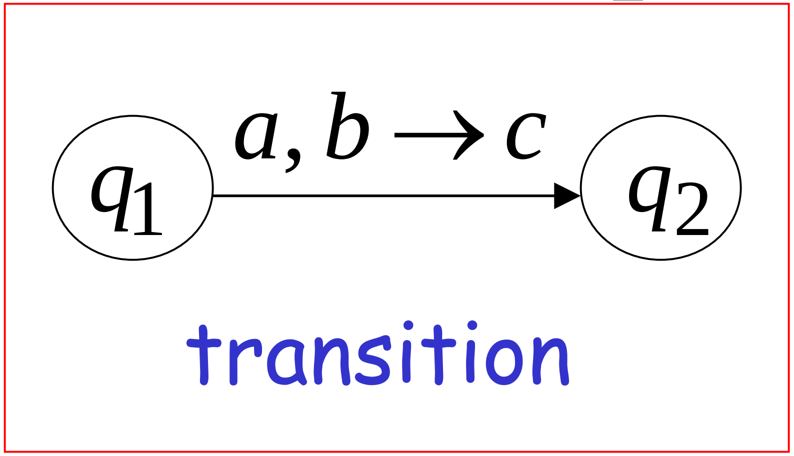
DFA for L_2

regular

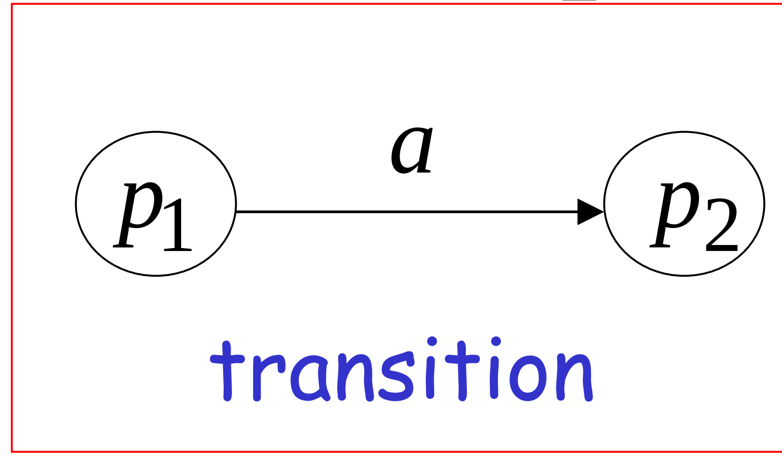
Construct a new NPDA machine M
that accepts $L_1 \cap L_2$

M simulates in parallel M_1 and M_2

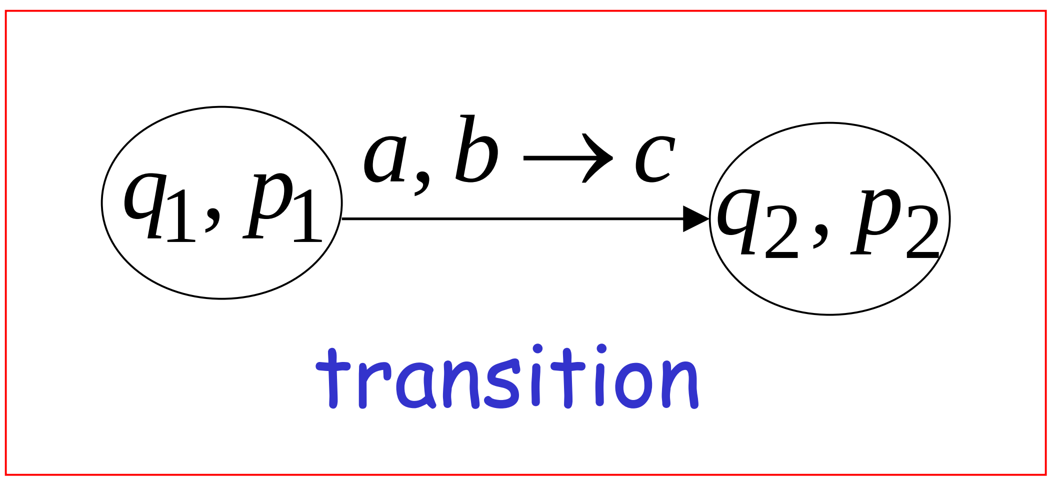
NPDA M_1



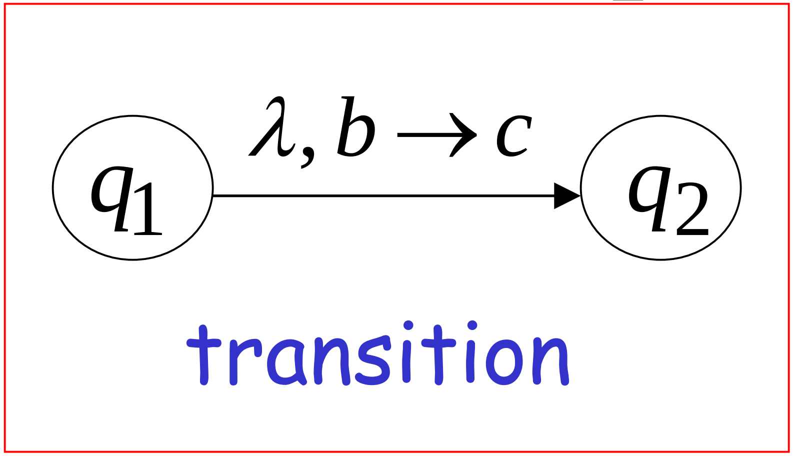
DFA M_2



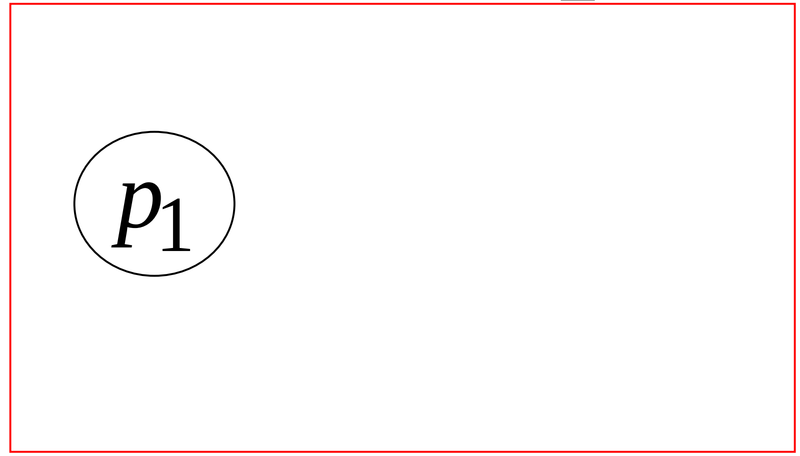
NPDA M



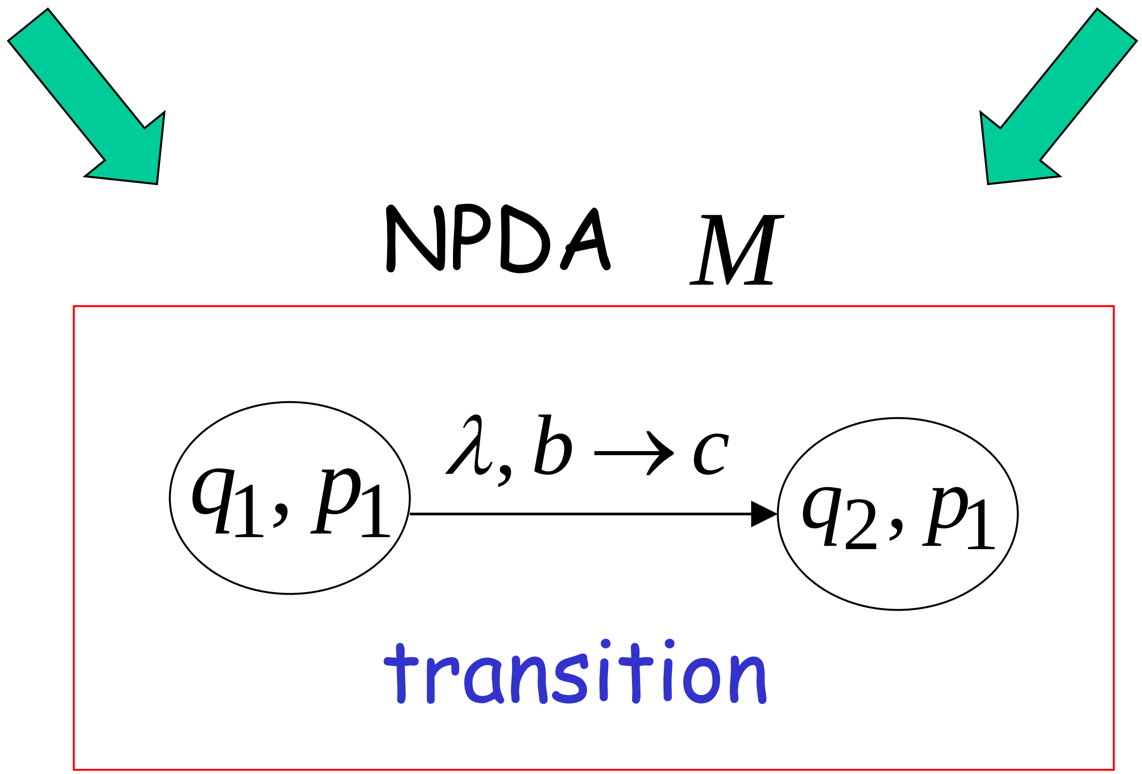
NPDA M_1

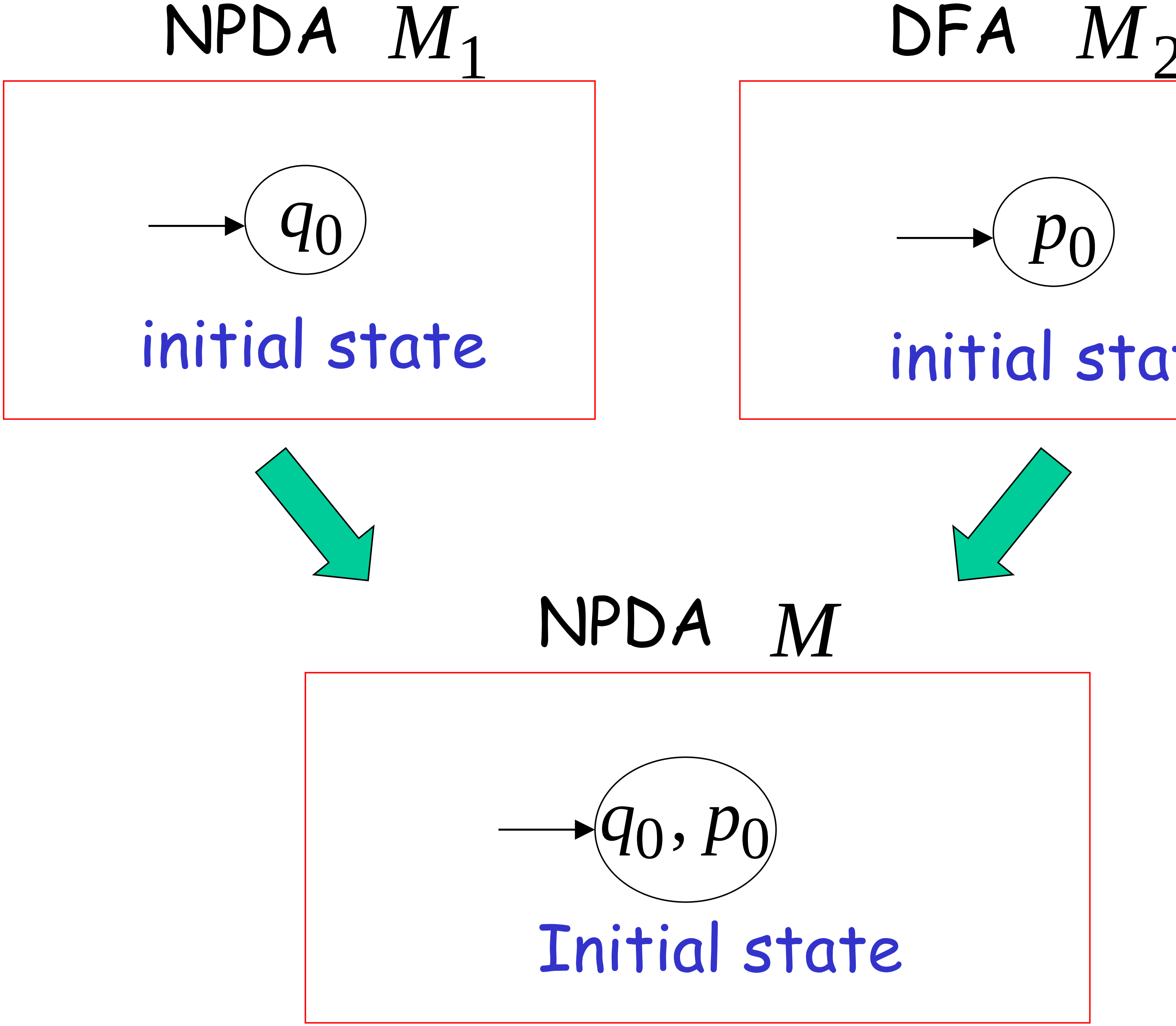


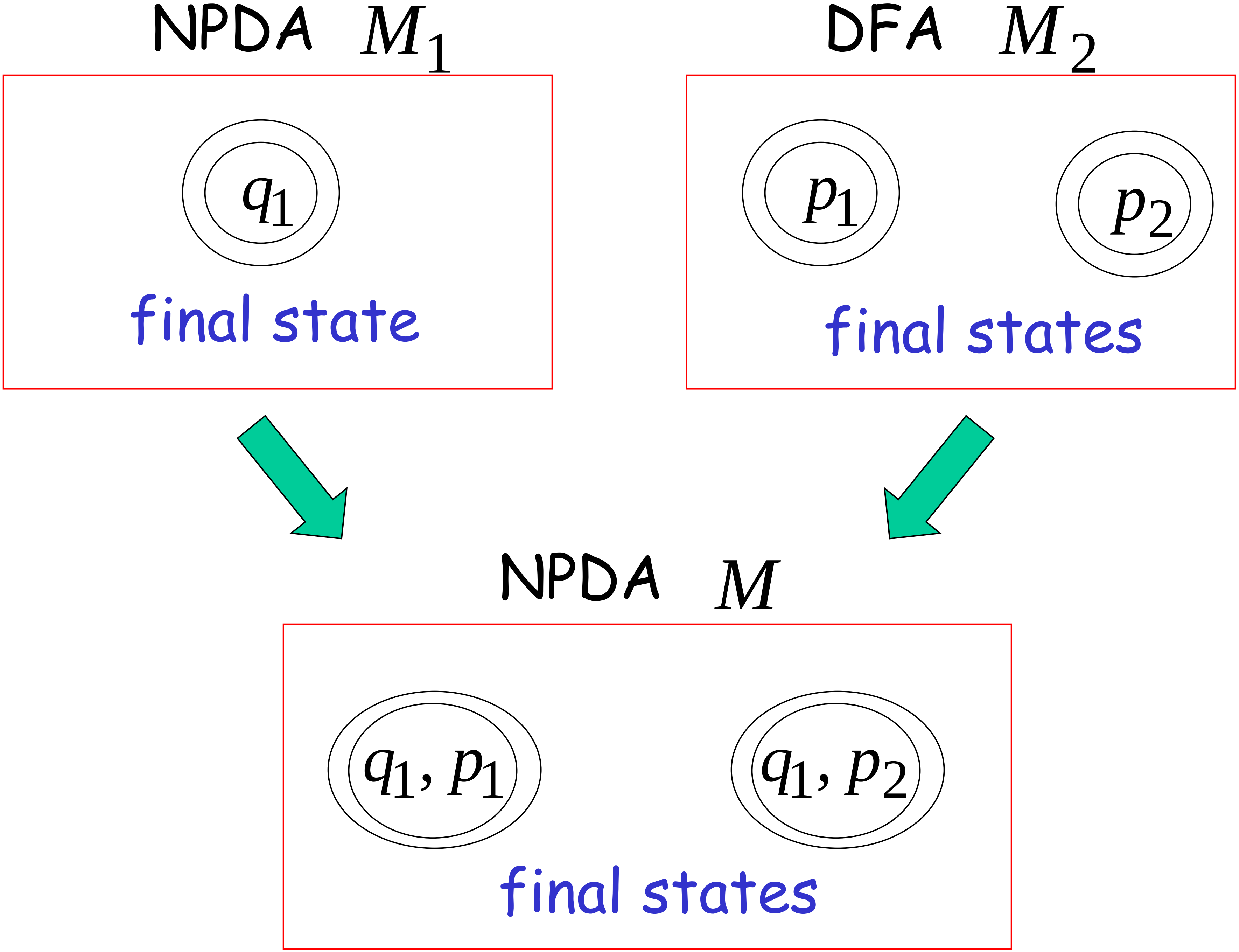
DFA M_2



NPDA M




NPDA M_1
q_0
initial state
DFA M_2
p_0
initial state
NPDA M
q_0, p_0
Initial state

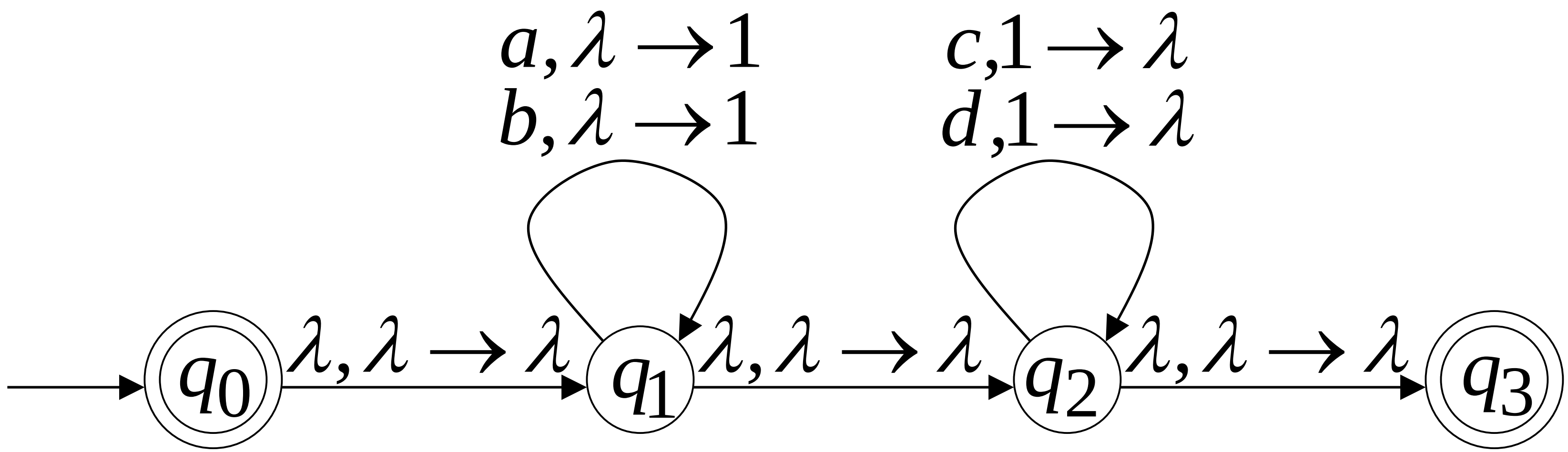

NPDA M_1
DFA M_2
q_1
final state
p_1
p_2
final states
NPDA M
q_1, p_1
q_1, p_2
final states

Example:

context-free

$$L_1 = \{w_1 w_2 : |w_1| = |w_2|,\ w_1 \in \{a,b\}^*,\ w_2 \in \{c,d\}^*\}$$

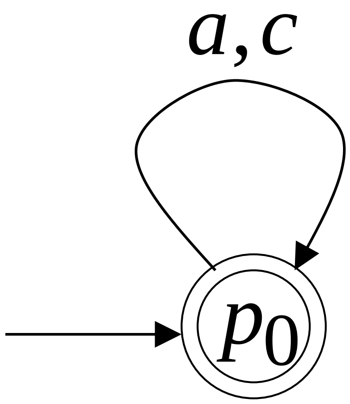
NPDA M_1



regular

$$L_2 = \{a, c\}^*$$

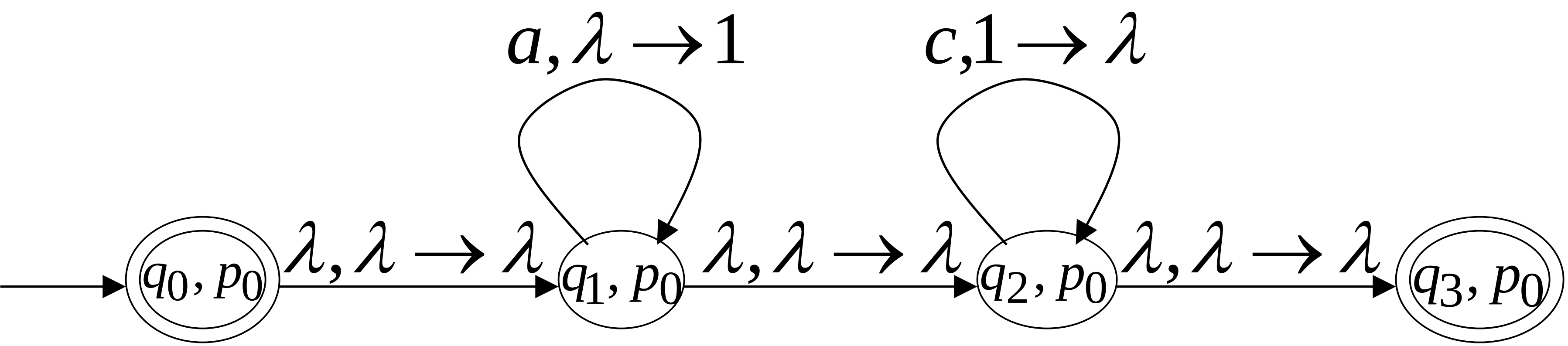
DFA M_2



context-free

Automaton for: $L_1 \cap L_2 = \{a^n c^n : \ n \geq 0\}$

NPDA M



In General:

M simulates in parallel M_1 and M_2

M accepts string w if and only if

M_1 accepts string w and

M_2 accepts string w

$$L(M) = L(M_1) \cap L(M_2)$$

Therefore:

M is NPDA

$L(M_1) \cap L(M_2)$ is context-free

$L_1 \cap L_2$ is context-free

Applications
of
Regular Closure

The intersection of

a context-free language and

a regular language

is a context-free language

Regular Closure

L_1 context free

L_2 regular

$\Rightarrow$

$L_1 \cap L_2$

context-free

An Application of Regular Closure

Prove that: $L = \{a^n b^n : \ n \neq 100, n \geq 0\}$

is context-free

We know:

$\{a^n b^n : n \geq 0\}$ is context-free

We also know:

$$L_1 = \{a^{100}b^{100}\}$$ is regular

$$\overline{L_1} = \{(a+b)^*\} - \{a^{100}b^{100}\}$$ is regular

$\{a^n b^n\}$

context-free

$\overline{L_1} = \{(a+b)^*\} - \{a^{100} b^{100}\}$

regular

(regular closure) $\{a^n b^n\} \cap \overline{L_1}$ context-free

$$\{a^n b^n\} \cap \overline{L_1} = \{a^n b^n : \quad n \neq 100, n \geq 0\} = L$$

is context-free

Another Application of Regular Closure

Prove that: $L = \{w: \quad n_a = n_b = n_c\}$

is **not** context-free

If $L = \{w: \ n_a = n_b = n_c\}$ is context-free

(regular closure)

Then $L \cap \{a^*b^*c^*\} = \{a^n b^n c^n\}$

context-free regular context-free

Impossible!!!

Therefore, L is **not** context free

Reverse

linear context-free languages
are closed under: **Reverse**

L Is a linear context free $\Longrightarrow$ L^R linear context-free

Concatenation

linear context-free languages
are **not** closed under: **Concatenation**

L_1 is linear context free

L_2 Is linear context free

$\Rightarrow$ L_1L_2 Is **not** linear context-free

Union

Deterministic context-free language are **not** closed under: **Union**

L_1 is context free

L_2 is context free

$\Rightarrow$ $L_1 \cup L_2$ is **not** necessarily context-free

Assume for contradiction that

$$L = \{a^n b^n\} \cup \{a^n b^{2n}\}$$

is non-deterministic context free

Decidable Properties of Context-Free Languages

Membership Question:

for context-free grammar G

find if string $w \in L(G)$

Membership Algorithms: Parsers

- Exhaustive search parser

Empty Language Question:

for context-free grammar G

find if $L(G) = \emptyset$

Algorithm:

1. Remove useless variables

2. Check if start variable S is useless

Infinite Language Question:

for context-free grammar G

find if $L(G)$ is infinite

Algorithm:

1. Remove useless variables

2. Remove unit and λ productions

3. Create dependency graph for variables

4. If there is a loop in the dependency graph then the language is infinite

Example: $S \rightarrow AB$

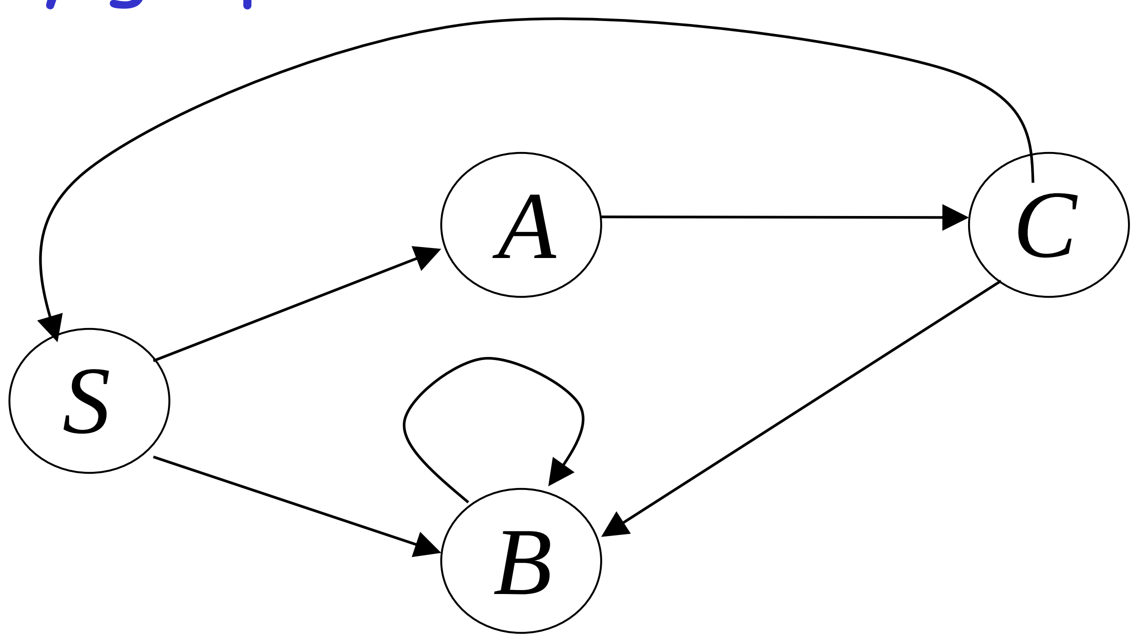
$$A \rightarrow aCb \mid a$$

$$B \rightarrow bB \mid bb$$

$$C \rightarrow cBS$$

Dependency graph

Infinite language



$$S \to AB$$

$$A \to aCb \mid a$$

$$B \to bB \mid bb$$

$$C \to cBS$$

$$S \Rightarrow AB \Rightarrow aCbB \Rightarrow acBSbB \Rightarrow acbbSbbb$$

$$S \overset{*}{\Rightarrow} acbbSbbb \overset{*}{\Rightarrow} (acbb)^2 S (bbb)^2$$

$$\overset{*}{\Rightarrow} (acbb)^i S (bbb)^i$$