San Skulrattanakulchai
Gustavus Adolphus College
April 8, 2013

Pumping Lemma for CFLs

Sipser Ch 2: p.125–129

Theorem (Pumping Lemma for Context-free Languages)**.** *Every CFL A has a pumping length $p > 0$ (depending on A) such that every string $s \in A$ with $|s| \geq p$ can be written as $uvxyz$ such that $|vy| > 0$, $|vxy| \leq p$, and $uv^ixy^iz \in A$ for all $i = 0, 1, 2, \ldots$.*

Proof. Choose a CFG G in CNF for A. Take any $s \in A$ of length $\geq 2^{|V|}$. Let T be a parse tree for s and let $T' = T - \{\text{leaves of } T\}$. Since T' has $\geq 2^{|V|}$ leaves, the height of T' is $\geq |V|$. Therefore, some variable R occurs at least twice on some longest path in T'. We have $|vy| > 0$ since T is a parse tree for a grammar in CNF and so every internal node of T' has exactly two children, and every child of an internal node of T' yields some terminals. We choose the upper R to have height $\leq |V|$. Thus $|vxy| \leq 2^{|V|}$. Now, choosing $p = 2^{|V|}$ gives the Pumping Lemma. □

Example 1. $A_1 = \{\, a^n b^n c^n : n \geq 0 \,\}$ is not context free.

Proof. Suppose A_1 is context free and let p be its pumping length. Consider $a^p b^p c^p \in A_1$ and write it as $uvxyz$ as in the Pumping Lemma. We see that ≤ 1 distinct symbol occurs in v, for otherwise uv^2xy^2z would have ≥ 4 1-symbol sections, constradicting it being in A_1. Similarly, ≤ 1 distinct symbol occurs in y. Thus, vy is missing at least a symbol, call it σ. (In fact, σ is either a or c.) Hence, uv^2xy^2z has too few σ's to be in A_1. □

Example 2. $A_2 = \{\, a^n b^m a^n b^m : n, m \geq 0 \,\}$ is not context free.

Proof. As in previous example, let p be its pumping constant, choose $s = a^p b^p a^p b^p \in A_2$, and write it as $uvxyz$.

String vxy must straddle the midpoint of s. To see this, suppose vxy is totally contained in the first half of s. Then $ba^p b^p$ would be a suffix of the second half of $t = uv^2xy^2z \in A_2$, contradicting $t \in A_2$. Similarly, vxy being totally contained in the second half of s is impossible.

We have $uv^0xy^0z = uxz \in A_2$ by the Pumping Lemma. Since vxy straddles the midpoint of s and string $vy \neq \varepsilon$, string uxz has the form $a^p b^i a^j b^p$ with $i < p$ or $j < p$. Thus $uxz \notin A_2$, a contradiction. □

Example 3. $A_3 = \{\, ww : w \in \{a, b\}^* \,\}$ is not context free.

Proof. If it were, then $A_3 \cap L(a^* b^* a^* b^*)$ would be a CFL. However, the language of this intersection is A_2, which has just been shown not context-free, a contradiction. □