Next

- Non-CF languages
- CFL pumping lemma

Non-CF Languages

The language $L = \{ a^n b^n c^n \mid n \geq 0 \}$ does not appear to be context-free.

Informal: The problem is that every variable can (only) act 'by itself' (*context-free*).

The problem of $A \Rightarrow^* vAy$:

If $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz \in L$,

then $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* \ldots \Rightarrow^* uv^iAy^iz$

$\Rightarrow^* uv^ixy^iz \in L$ as well, for all $i=0,1,2,\ldots$

"Pumping Lemma for CFLs"

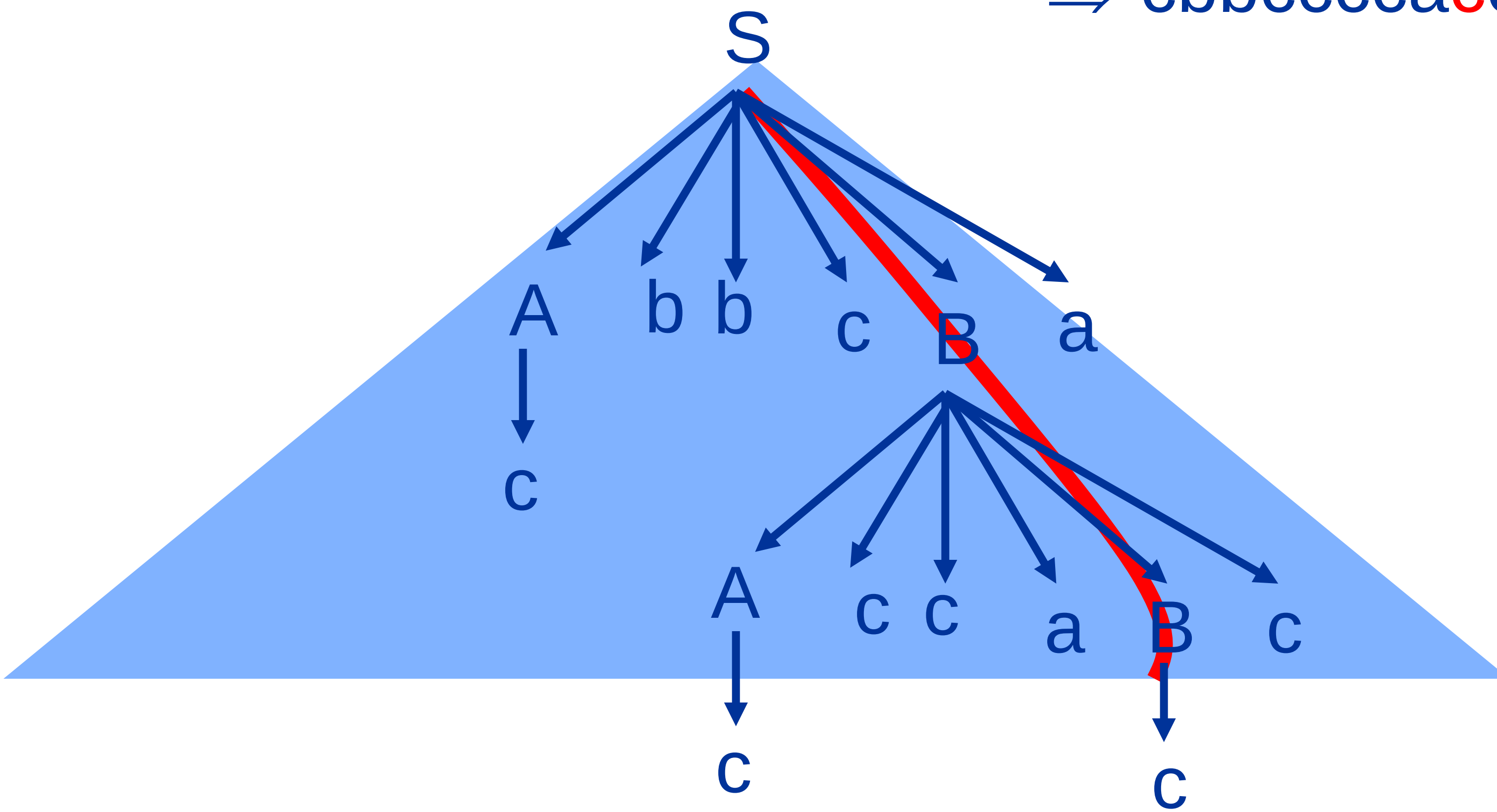
Idea: If we can prove the existence of derivations for elements of the CFL L that use the step $A \Rightarrow^* vAy$, then a new form of 'v-y pumping' holds: $A \Rightarrow^* vAy \Rightarrow^* v^2Ay^2 \Rightarrow^* v^3Ay^3 \Rightarrow^* \ldots$)

Observation: We can prove this existence if the parse-tree is tall enough.

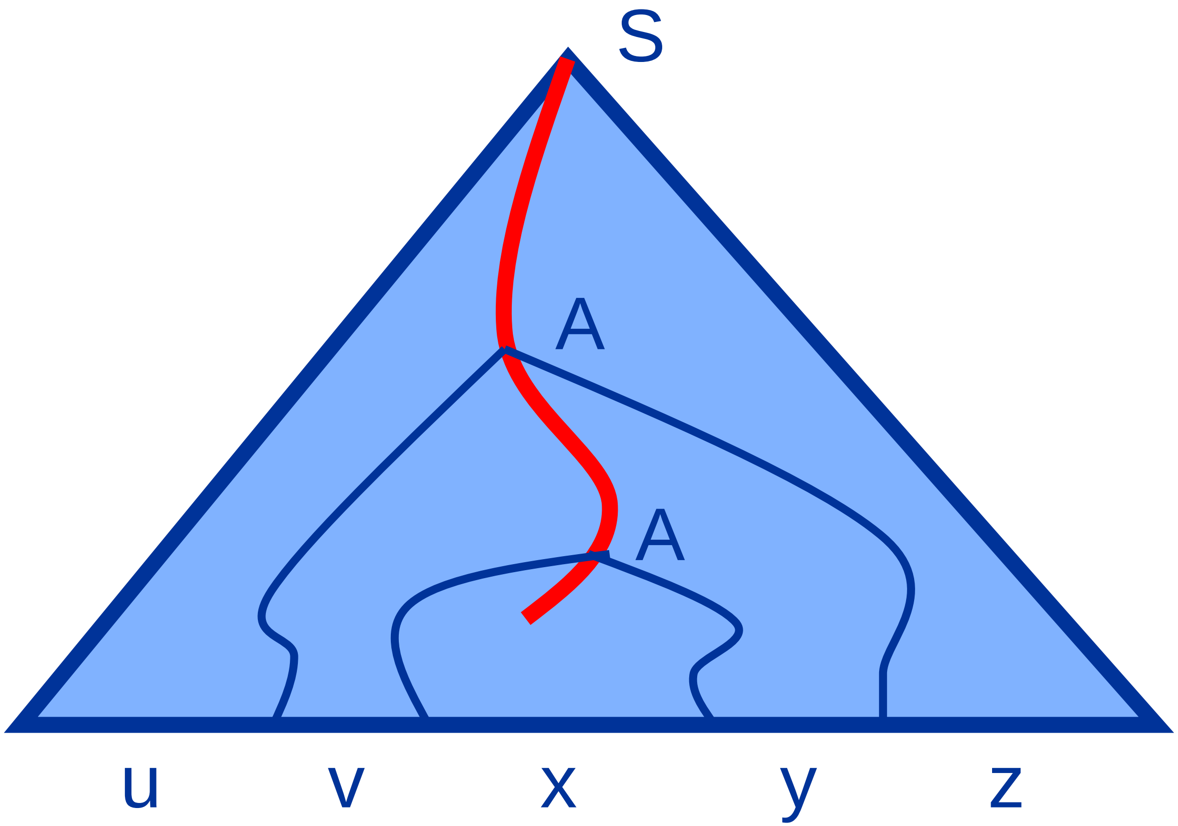
Remember Parse Trees

Parse tree for S $\Rightarrow$ AbbcBa $\Rightarrow^*$ cbbccccaBca

$\Rightarrow$ cbbccccacca



Pumping a Parse Tree



If s = uvxyz $\in$ L is long, then its parse-tree is tall. Hence, there is a path on which a variable A repeats itself. We can pump this A–A part.

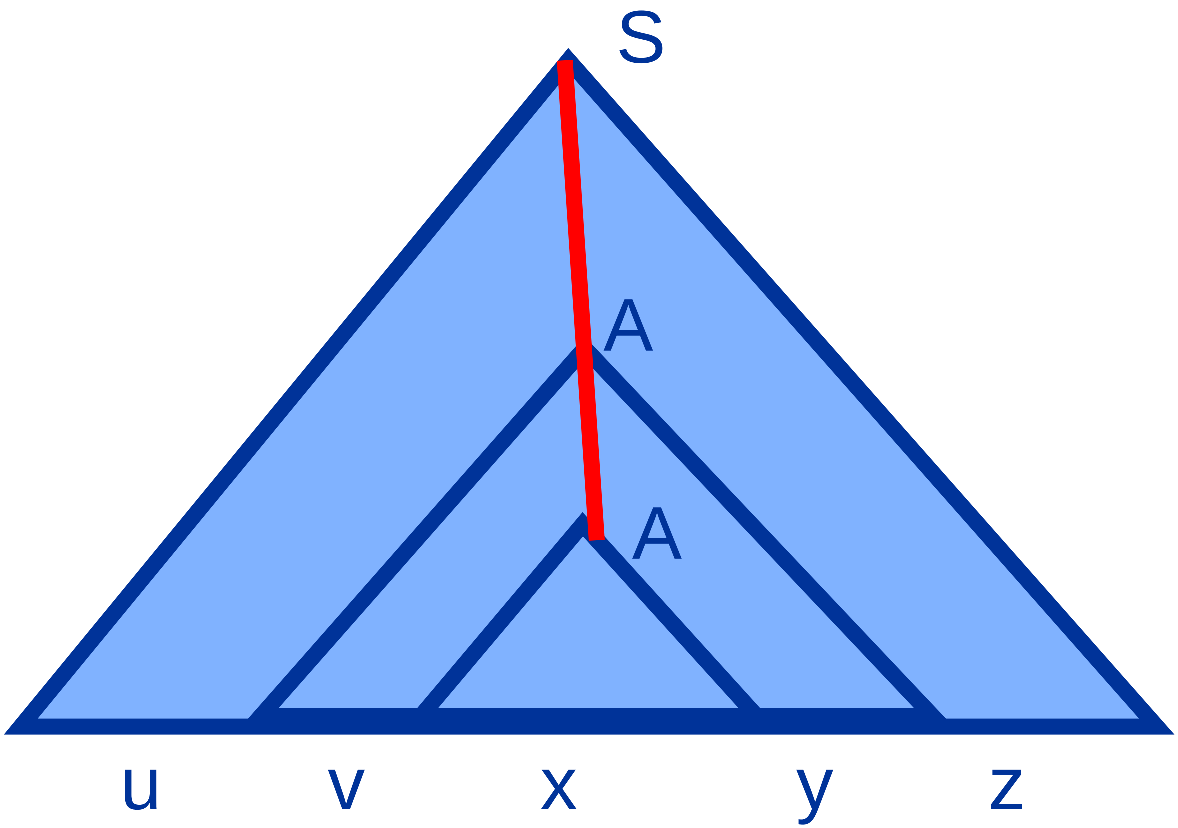
A Tree Tall Enough

Let L be a context-free language, and let G be its grammar with maximal b symbols on the right side of the rules: $A \rightarrow X_1 \ldots X_b$

A parse tree of depth h produces a string with maximum length of b^h.
Long strings implies tall trees.

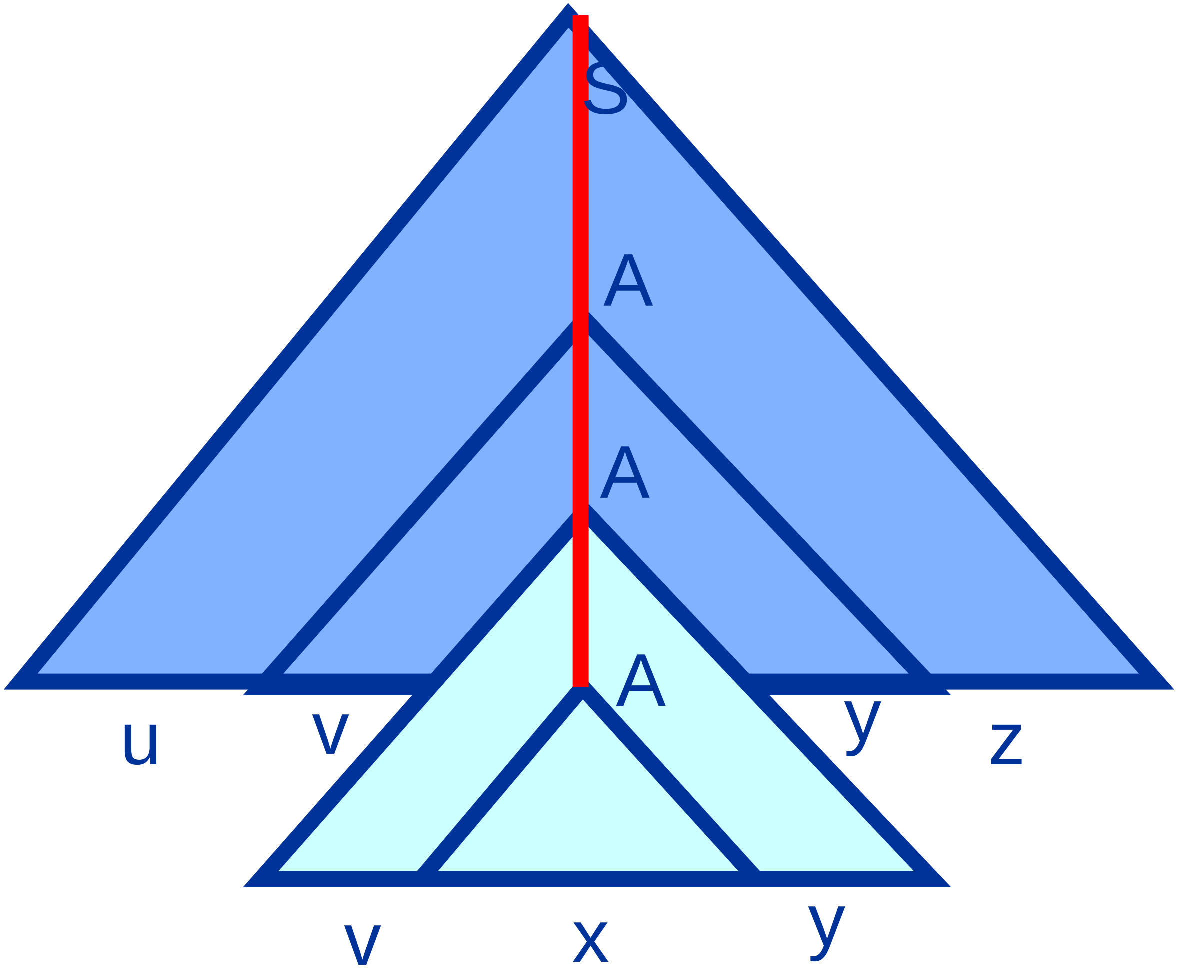
Let |V| be the number of variables of G.
If h = |V|+2 or bigger, then there is a variable on a 'top-down path' that occurs more than once.

uvxyz ∈L



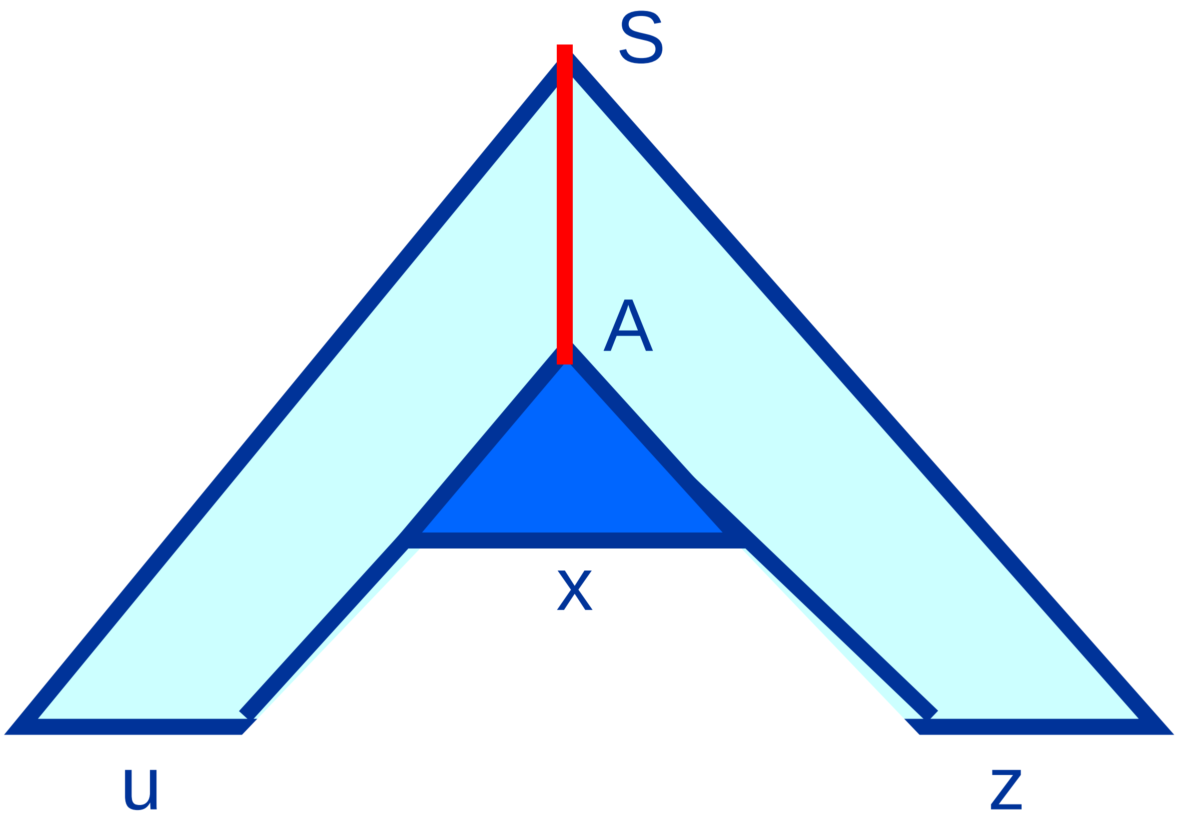
By repeating the A–A part we get…

$uv^2xy^2z \in L$



… while removing the A–-A gives…

Pumping down: uxz $\in$ L



In general $uv^ixy^iz \in L$ for all i=0,1,2,…

Pumping Lemma for CFL

For every context-free language L, there is a pumping length p, such that for every string $s \in L$ and $|s| \geq p$, we can write $s = uvxyz$ with

1) $uv^ixy^iz \in L$ for every $i \in \{0,1,2,\ldots\}$
2) $|vy| \geq 1$
3) $|vxy| \leq p$

Note that 1) implies that $uxz \in L$
2) says that vy cannot be the empty string ε
Condition 3) is not always used

Formal Proof of Pumping Lemma

Let $G=(V,\Sigma,R,S)$ be the grammar of a CFL.
Maximum size of rules is $b \geq 2$: $A \to X_1 \ldots X_b$
A string s requires a <u>minimum</u> tree-depth $\geq \log_b|s|$.
If $|s| \geq p=b^{|V|+2}$, then tree-depth $\geq |V|+2$, hence there is a path and variable A where A repeats itself: $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$
It follows that $uv^ixy^iz \in L$ for all $i=0,1,2,\ldots$
Furthermore:

$|vy| \geq 1$ because tree is minimal

$|vxy| \geq p$ because bottom tree with $\geq p$ leaves has a 'repeating path'

Pumping $a^n b^n c^n$ (Ex. 2.20)

Assume that $B = \{a^n b^n c^n \mid n \geq 0\}$ is CFL

Let p be the pumping length, and $s = a^p b^p c^p \in B$

P.L.: $s = uvxyz = a^p b^p c^p$, with $uv^i xy^i z \in B$ for all $i \geq 0$

Options for |vxy|:

1) The strings v and y are uniform
   (v=a…a and y=c…c, for example).
   Then uv^2xy^2z will not contain the same number of a's, b's and c's, hence $uv^2xy^2z \notin B$
2) v and y are not uniform.
   Then uv^2xy^2z will not be a…ab…bc…c
   Hence $uv^2xy^2z \notin B$

Pumping $a^n b^n c^n$ (cont.)

Assume that $B = \{a^n b^n c^n \mid n \geq 0\}$ is CFL

Let p be the pumping length, and $s = a^p b^p c^p \in B$

P.L.: $s = uvxyz = a^p b^p c^p$, with $uv^i xy^i z \in B$ for all $i \geq 0$

We showed: No options for |vxy| such that
$uv^i xy^i z \in B$ for all i. Contradiction.

B is not a context-free language.

Example 2.21 (Pumping down)

Proof that $C = \{a^i b^j c^k \mid 0 \le i \le j \le k\}$ is not context-free.

Let p be the pumping length, and $s = a^p b^p c^p \in C$

P.L.: $s = uvxyz$, such that $uv^ixy^iz \in C$ for every $i \ge 0$

Two options for $1 \le |vxy| \le p$:

1) $vxy = a^*b^*$, then the string uv^2xy^2z has not enough c's, hence $uv^2xy^2z \notin C$

2) $vxy = b^*c^*$, then the string $uv^0xy^0z = uxz$ has too many a's, hence $uv^0xy^0z \notin C$

Contradiction: C is not a context-free language.

$D = \{ ww \mid w \in \{0,1\}^* \}$ (Ex. 2.22)

Carefully take the strings $s \in D$.

Let p be the pumping length, take $s=0^p1^p0^p1^p$.

Three options for $s=uvxyz$ with $1 \leq |vxy| \leq p$:

1) If a part of y is to the left of | in $0^p1^p|0^p1^p$,
   then second half of uv^2xy^2z starts with "1"
2) Same reasoning if a part of v is to the right
   of middle of $0^p1^p|0^p1^p$, hence $uv^2xy^2z \notin D$
3) If x is in the middle of $0^p1^p|0^p1^p$, then uxz
   equals $0^p1^i0^j1^p \notin D$ (because i or $j < p$)

<u>Contradiction</u>: D is not context-free.

Pumping Problems

Using the CFL pumping lemma is more difficult than the pumping lemma for regular languages.

You have to choose the string s carefully, and divide the options efficiently.

Additional CFL properties would be helpful (like we had for regular languages).

What about closure under standard operations?

Next

- Closure properties of CFL

Union Closure Properties

Lemma: Let A_1 and A_2 be two CF languages, then the *union* $A_1 \cup A_2$ is context free as well.

Proof: Assume that the two grammars are $G_1=(V_1,\Sigma,R_1,S_1)$ and $G_2=(V_2,\Sigma,R_2,S_2)$.
Construct a third grammar $G_3=(V_3,\Sigma,R_3,S_3)$ by:
$V_3 = V_1 \cup V_2 \cup \{ S_3 \}$ (new start variable) with
$R_3 = R_1 \cup R_2 \cup \{ S_3 \to S_1 \mid S_2 \}$.

It follows that $L(G_3) = L(G_1) \cup L(G_2)$.

Intersection & Complement?

Let again A_1 and A_2 be two CF languages.

One can prove that, *in general,*
 the intersection $A_1 \cap A_2$,
and
 the complement $\bar{A}_1 = \Sigma^* \setminus A_1$
are not context free languages.

One proves this with specific counter examples of languages.

What do we really know?

Can we always decide if a language L is regular/ context-free or not?

We know:

$\{ 1^x \mid x = 0 \bmod 7 \}$ is regular

$\{ 1^x \mid x \text{ is prime} \}$ is not regular

But what about

$\{ 1^x \mid x \text{ and } x+2 \text{ are prime} \}$?

This is (yet) unknown.

Describing a Language

The problem lies in the informal notion of a description.

Consider:

$\{ n \mid \exists a,b,c: a^n+b^n = c^n \}$

{ x | in year x the first female US president }

{ x | x is “an easy to remember number” }

We have to define what we mean by “description” and “method of deciding”.