A short proof that O_2 is an MCFL

Mark-Jan Nederhof
University of St Andrews

NBSAN 2016-03-17

Context-free grammars

All strings over $\{a, b\}$ consisting of two consecutive palindromes of even length

$$\begin{array}{lcl} S & \rightarrow & P\ P \\ P & \rightarrow & a\ P\ a \\ P & \rightarrow & b\ P\ b \\ P & \rightarrow & \varepsilon \end{array}$$

$$S \Rightarrow PP \Rightarrow aPaP \Rightarrow aaP \Rightarrow aabPb \Rightarrow aabaPab \Rightarrow aabaab$$

Context-free grammars (alternative view)

Nonterminals become predicates of one argument

A variable occurs once in LHS and once in RHS

Change direction of arrows (logical 'implies')

$$
\begin{array}{rcll}
S(x\ y) & \leftarrow & P(x)\ P(y) & (1) \\
P(a\ x\ a) & \leftarrow & P(x) & (2) \\
P(b\ x\ b) & \leftarrow & P(x) & (3) \\
P(\varepsilon) & \leftarrow & & (4)
\end{array}
$$

(3) $P(aa) \Rightarrow P(baab)$

(1) $P(aa)\ ,\ P(baab) \Rightarrow S(aabaab)$

Multiple context-free grammars (MCFGs)

Predicates can now have several arguments

I.e. **fan-out** can be more than 1

MCFL(n): languages generated by MCFGs with fan-out n

Exponent of parsing complexity increases with fan-out

Example with fan-out 2:

$$\begin{array}{rcl} S(x\ y) & \leftarrow & E(x, y) \\ E(xp, yq) & \leftarrow & E(x, y)\ E(p, q) \\ E(a, a) & \leftarrow & \\ E(b, b) & \leftarrow & \end{array}$$

Generates copy language $\{ww \mid w \in \{a, b\}^+\}$

Linguistic motivations

MCFG is **mildly context-sensitive** formalism

(Further generalizes 'linear indexed grammars')

Believed to be powerful enough for natural language

And unable to generate anything that is unlike natural language

MIX language

$$\text{MIX} = \{w \in \{a, b, c\}^* \mid |w|_a = |w|_b = |w|_c\}$$

One combination of *a*, *b*, *c* represents phrase with:

a is main verb
b is its subject
c is its object

Any number of such triples scrambled in any order

Models extreme free word order

Doesn't seem to happen in natural language

So people didn't expect this to be MCFL

MIX is MCFL !

Sylvain Salvati:

- MIX is rationally equivalent to O_2 (to be discussed)
- So MIX is MCFL **iff** O_2 is MCFL
- Proof that O_2 is generated by MCFG
- Geometric arguments (two-dimensional)
- Uses $z \mapsto e^{2i\pi z}$, for $z \in \mathbb{C} \setminus \{0\}$

Journal of Computer and System Sciences 81, pp. 1252-1277, 2015

O_n languages

For $n \geq 1$, let $\Sigma_n = \{a_1, \ldots, a_n, \overline{a_1}, \ldots, \overline{a_n}\}$

$$O_n = \{w \in \Sigma_n^* \mid \forall_i |w|_{a_i} = |w|_{\overline{a_i}}\}$$

State of the art:

- O_1 is an MCFL(1) = CFL **Easy**
- O_2 is an MCFL(2) **Salvati's proof**
- O_3 is an MCFL(3) **???**

There seems to be no way to generalize Salvati's proof

Proof for O_1

$$
\begin{array}{rcll}
S(x) & \leftarrow & R(x) & (1) \\
R(x\ y) & \leftarrow & R(x)\ R(y) & (2) \\
R(a\ x\ \overline{a}) & \leftarrow & R(x) & (3) \\
R(\overline{a}\ x\ a) & \leftarrow & R(x) & (4) \\
R(\varepsilon) & \leftarrow & & (5)
\end{array}
$$

$R(\overline{a}\,\overline{a}aaa\overline{a})$? Use (2) , $R(\overline{a}\,\overline{a}aa)$, $R(a\overline{a})$

$R(\overline{a}\,\overline{a}aa)$? Use (4) , $R(\overline{a}a)$

Etc.

Needed grammar for O_2

$$
\begin{aligned}
S(xy) &\leftarrow R(x,y) && (1)\\
R(xp,yq) &\leftarrow R(x,y)\ R(p,q) && (2)\\
R(xp,qy) &\leftarrow R(x,y)\ R(p,q) && (3)\\
R(xpy,q) &\leftarrow R(x,y)\ R(p,q) && (4)\\
R(p,xqy) &\leftarrow R(x,y)\ R(p,q) && (5)\\
R(a,\overline{a}) &\leftarrow && (6)\\
R(\overline{a},a) &\leftarrow && (7)\\
R(b,\overline{b}) &\leftarrow && (8)\\
R(\overline{b},b) &\leftarrow && (9)\\
R(\varepsilon,\varepsilon) &\leftarrow && (10)
\end{aligned}
$$

Derivation for O_2

$$S(a\overline{b}\overline{a}\overline{b}\overline{a}bba)$$

(1)

$$R(a\overline{b}\overline{a}\overline{b}\overline{a}b, ba)$$

(4)

$$R(a\overline{b}, \overline{a}b) \qquad R(\overline{a}\overline{b}, ba)$$

(2) (3)

$$R(a, \overline{a}) \quad R(\overline{b}, b) \quad R(\overline{a}, a) \quad R(\overline{b}, b)$$

(6) (9) (7) (9)

Remember:

$$R(xpy, q) \leftarrow R(x, y)\ R(p, q) \tag{4}$$

Does grammar generate O_2 ?

Easy: if there is derivation of $R(x, y)$ then $xy \in O_2$

Difficult: if $xy \in O_2$ then there is derivation of $R(x, y)$

This is all we need !!!

Remember:

$$S(xy) \quad \leftarrow \quad R(x, y) \tag{1}$$

Proof by induction

How to prove $xy \in O_2$ implies $R(x, y)$?

Induction on $|xy|$

Only interesting case requiring inductive hypothesis:

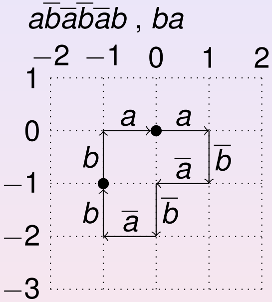
- $|xy| \geq 4$
- no non-empty substring of x or of y is in O_2

To prove:

Some binary rule is always applicable to divide pair (x, y) into four strings, to use inductive hypothesis on two shorter pairs

Geometry for O_2

- a is 'right'
- $\overline{a}$ is 'left'
- b is 'up'
- $\overline{b}$ is 'down'



$P[1] = (-1, -1)$, $P[k] = k * P[1]$ for $k \in \mathbb{Z}$

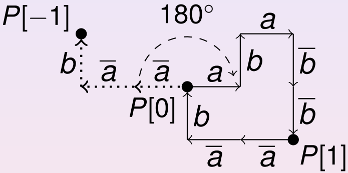
E.g. $P[0] = (0, 0)$, $P[-5] = (5, 5)$

$A[k]$ is path of first string from $P[k]$

$B[k]$ is path of second string from $P[k]$

Three rule applications (1)

$a\,b\,a\,\overline{b}\,\overline{b}\,,\,\overline{a}\,\overline{a}\,b$



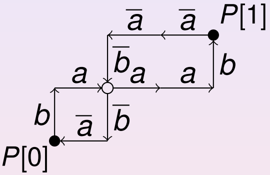
Remember:

$$R(xp, yq) \leftarrow R(x, y)\ R(p, q) \tag{2}$$

Let $x = a$, $y = \overline{a}$

Three rule applications (2)

$baaab\ ,\ \overline{a}\,\overline{a}\overline{b}\,\overline{b}\overline{a}$



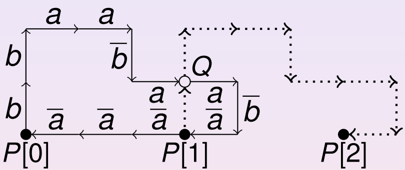
Remember:

$$R(xp, qy) \quad \leftarrow \quad R(x, y)\ R(p, q) \tag{3}$$

Let $x = ba$ and $y = \overline{b}\overline{a}$

Three rule applications (3)

$b\, b\, a\, a\, \overline{b}\, a\, a\, \overline{b}\, \overline{a}\, ,\, \overline{a}\, \overline{a}\, \overline{a}$



Remember:

$$R(xpy, q) \quad \leftarrow \quad R(x, y)\ R(p, q) \qquad (4)$$

Let $x = b$ and $y = a\,\overline{b}\,\overline{a}$

Note $d_{A[0]}(Q) = 6 > d_{A[1]}(Q) = 1$

where $d_C(Q)$ is path distance of Q from start of path C

Suppose no rules are applicable

Then **four constraints** must hold:

(i) angle in $P[0]$ between beginning of $A[0]$ and that of $B[0]$ is *not* 180°

(ii) $A[0] \cap B[1] = \{P[0], P[1]\}$

(iii) $\nexists\ Q \in (A[0] \cap A[1]) \setminus \{P[1]\}$ such that $d_{A[0]}(Q) > d_{A[1]}(Q)$

(iv) $\nexists\ Q \in (B[0] \cap B[1]) \setminus \{P[0]\}$ such that $d_{B[1]}(Q) > d_{B[0]}(Q)$

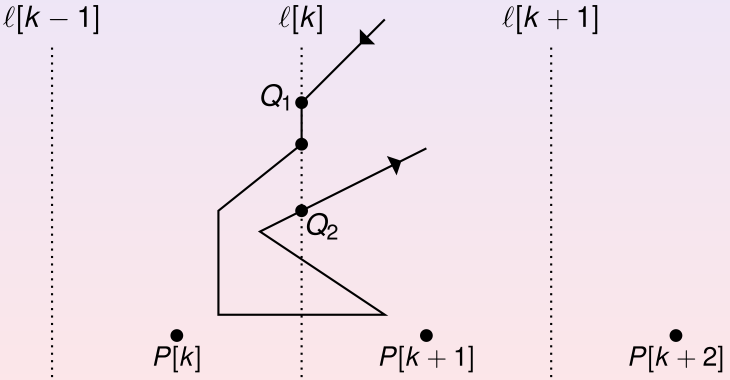
(No self-intersections: no non-empty substring of x or y in O_2)

Can we derive a contradiction from this ?

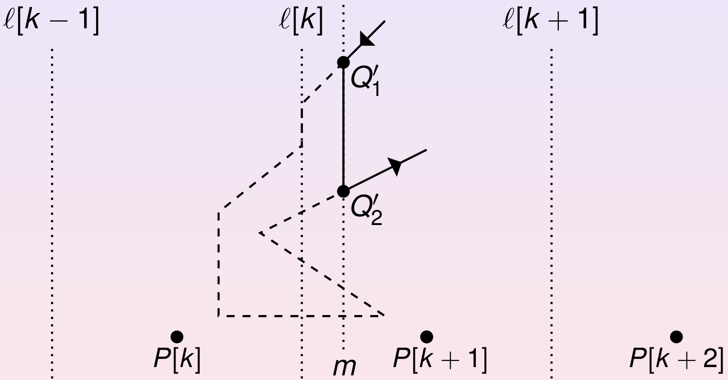
How to 'tame' the myriad possibilities of paths A and B ?

Excursion

Excursion from right between Q_1 and Q_2



Excursion truncated



Truncate excursions without violating **four constraints** !!!

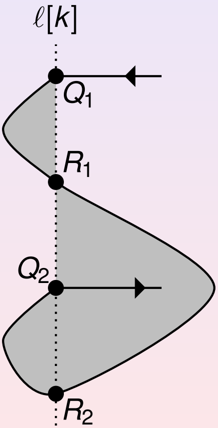
Regions and area of excursion

Regions of excursion:
enclosed by path and line

Area of excursion:
total surface area of regions

Excursion is **filled**:
if its regions contain some $P[k']$

Excursion is **unfilled**:
otherwise



Normal form

A and *B* are in **normal form** if all excursions exhaustively truncated

(without violating **four constraints** or introducing self-intersections)

Suppose some **unfilled** excursions remain

Take one with smallest area and find contradiction

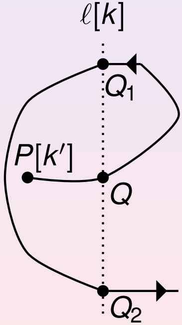
Suppose some **filled** excursions remains and find contradiction

So no excursions remain !!!

Suppose truncation would introduce self-intersection

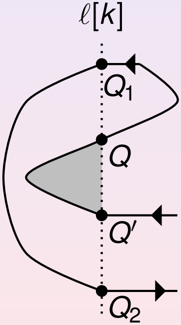
…in unfilled excursion with smallest area

Exactly one crossing



Filled !!!

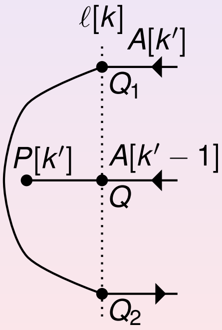
Two or more crossings



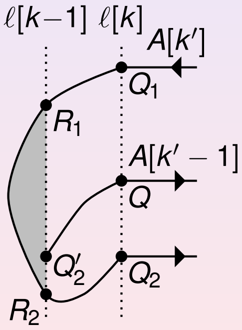
Area not smallest !!!

Suppose truncation would violate constraint (iii)

...in unfilled excursion with smallest area, one crossing, $d_{A[k'-1]}(Q) > d_{A[k']}(Q_2)$



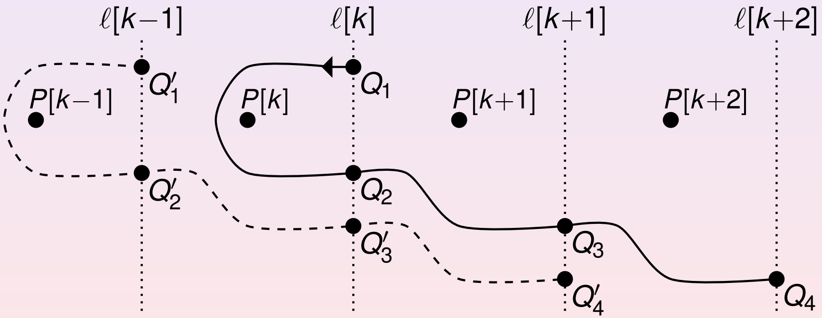
Filled !!!



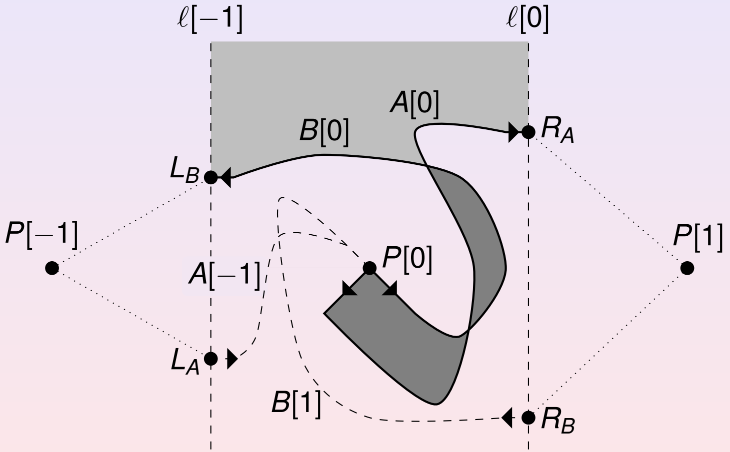
Area not smallest !!!

Filled excursions are impossible

One of two cases:



Final contradiction



Four constraints: L_B above L_A iff R_A above R_B (Impossible !!!)

Outlook

O_3 is likely MCFL too, with fanout 3

Three dimensional arguments required

Partitioning space into ‘top’ and ‘bottom’ applicable to 3D

One more idea needed (related to braid theory)

Unclear yet how to redefine ‘excursion’ for 3D

Are O_4, O_5, . . . also MCFLs ?

Would mean MCFLs are closed under permutation closure

Full paper: `http://arxiv.org/abs/1603.03610`