ON THE COMPLEXITY OF UNIVERSALITY OF UNAMBIGUOUS CONTEXT-FREE GRAMMARS

LORENZO CLEMENTE
UNIVERSITY OF WARSAW

$VPT'20 \subseteq VPT'21$

27/03/2021 A.E

OUTLINE

- Context-free grammars, universality, inclusion, unambiguity.
- NFA $\overset{?}{\subseteq}$ unambiguous CFG ↘ universality unambiguous CFG → zeroness of convolution-recursive sequences $\in$ PSPACE.
  CFG $\overset{?}{\subseteq}$ unambiguous FA ↗
- Generating functions and classes of recursive sequences $\mathbb{N} \to \mathbb{Q}$.
- SQRT-SUM-hardness of computing the coin-flip measure of UCFG.
- Connection with stochastic CFG.
- Conclusions & open problems.

CONTEXT-FREE GRAMMARS over alphabet $\Sigma = \{a, b\}$

$X \to aXb \mid XX \mid \varepsilon$ $\quad L(X) = \{$balanced words over $\Sigma\}$

- Nondeterministic CFG.
- Ambiguous CFG: $ababab \in L(X)$ has two distinct parse trees.

$Y \to aYbY \mid \varepsilon$: $L(Y) = L(X)$, deterministic CFG!

$Z \to aZa \mid bZb \mid \varepsilon$: unambiguous CFG: every $w \in \Sigma^*$ has ≤ 1 parse tree.
<u>not</u> a deterministic CFL!

$\{a^m b^m a^n\} \cup \{a^m b^n a^n\}$: Ambiguous CFG, inherently ambiguous CFL.

→ Establishing whether a given CFG recognises an unambiguous CFL is undecidable.
[Ginsburg & Ullian JACM'66]

→ Establishing whether a given CFG is unambiguous is undecidable.

UNAMBIGUITY in THEORETICAL COMPUTER SCIENCE

[Colcombet DCFS'15]

- PTIME ⊆ UPTIME ⊆ NPTIME

  Parity games ∈ UPTIME ∩ coUPTIME [Jurdziński IPL'98].

- unambiguous Büchi automata over ω-words are expressively complete.

  [Carton & Michel TCS'03]

- universality PSPACE-hard for nondeterministic finite automata,

  BUT in PTIME for unambiguous finite automata.

  [Stearns & Hunt SFCS'81]

- universality is not primitive recursive for VASS/Petri nets,

  BUT in EXPSPACE for unambiguous " " ".

  [Czerwiński & Figueira & Hofman CONCUR'20]

- universality undecidable for nondeterministic register automata,

  BUT decidable for unambiguous register automata.

  [Mottet & Quaas STACS'19]

- universality undecidable for nondeterministic CFGs,

  BUT decidable for unambiguous CFGs. ...

UNIVERSALITY & INCLUSION PROBLEMS

Fix a finite alphabet Σ.

Given a presentation (automaton, grammar...) of $L, M \subseteq \Sigma^*$:

- Inclusion problem: $L \subseteq M$?
- Universality problem: $\Sigma^* \subseteq M$? < undecidable for CFG. / decidable for DCFG in PTIME!

INCLUSION → EMPTINESS

$$\underbrace{L}_{\text{NFA}} \subseteq \underbrace{M}_{\text{DCFG}} \iff \underbrace{L \cap \underbrace{(\Sigma^* \setminus M)}_{\text{DCFG}}}_{\text{CFG}} = \emptyset$$

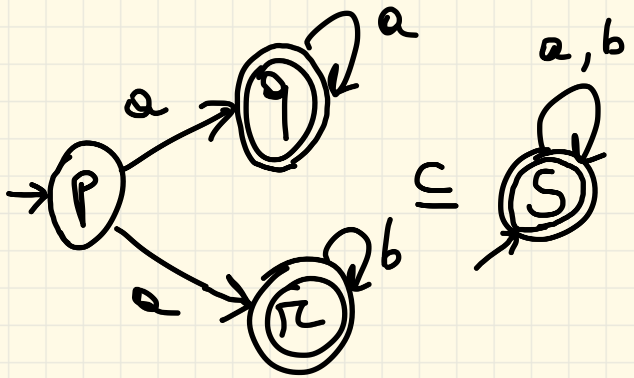
Unambiguous CFG (UCFG) not closed under complement!

[Hibbard & Ullian '66]

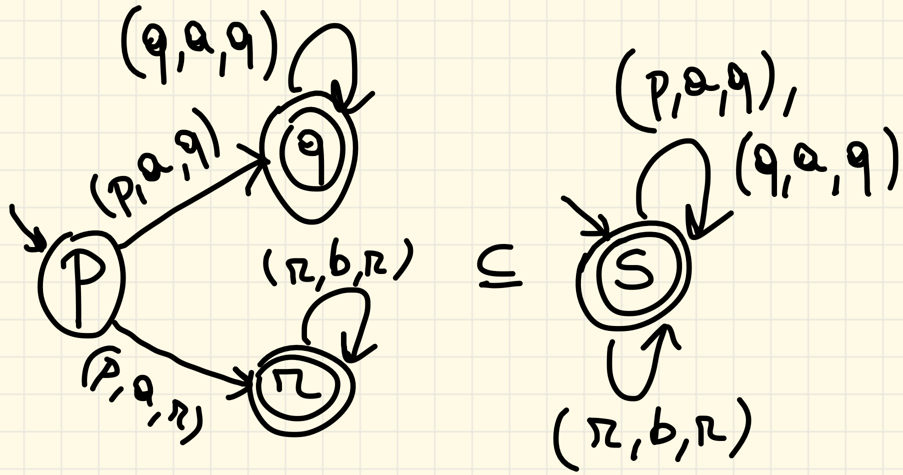
INCLUSION → UNIVERSALITY

STEP 1 : Inclusion → Inclusion with deterministic l.h.s. :

$$\underbrace{L}_{\text{NFA}} \subseteq \underbrace{M}_{\text{UCFG}} \iff \underbrace{L'}_{\text{DFA}} \subseteq \underbrace{M'}_{\text{UCFG}}$$



iff



INCLUSION → UNIVERSALITY

STEP 1 : Inclusion → Inclusion with deterministic l.h.s. :

$$\underbrace{L}_{NFA} \subseteq \underbrace{M}_{UCFG} \iff \underbrace{L'}_{DFA} \subseteq \underbrace{M'}_{UCFG}$$

STEP 2 : Inclusion with deterministic l.h.s. → Universality

$$\underbrace{L}_{DFA} \subseteq \underbrace{M}_{UCFG} \iff \underbrace{\underbrace{(M \cap L)}_{UCFG} \cup \underbrace{(\Sigma^* \setminus L)}_{\text{DFA, thus UCFG}}}_{\text{UCFG since the union is disjoint}} = \Sigma^*$$

Applications :

NFA ⊆ UCFG → universality of UCFG

CFG ⊆ UFA → universality of UCFG

register automata NRA ⊆ URA → universality of URA

Parikh automata NPA ⊆ UPA → universality of UPA

UNIVERSALITY PROBLEM for UNAMBIGUOUS CFG

Decidable :

(1 quantifier alternation → $\exists\forall^*$)

- By reduction to the $\exists\forall^*$-fragment of Tarski's algebra $\mathbb{R}(+,\cdot,0,1)$ [Salomaa & Soittola '78], even in EXPTIME by [Grigoriev '88].
- By elimination theory [Kuich & Salomaa '86].

Better complexity bound :

- In PSPACE by reduction to the $\exists^*$ fragment of Tarski's algebra, which is in PSPACE by [Canny '88] (c.f. master thesis of S. Purgal, 2018).
- In PSPACE by reduction to the zeroness problem of convolution-recursive sequences of rationals $\mathbb{N} \to \mathbb{Q}$ + PSPACE algorithm for the latter problem.

UCFG UNIVERSALITY → CONV-REC ZERONESS

(a.k.a. Chomsky-Schützenberger's approach)

$X \to a X b X \mid \varepsilon$

$f_X(n) :=$ "# words of length n accepted by X" $: \mathbb{N} \to \mathbb{Q}$

- $L(X) = \{a, b\}^*$ iff $\forall n \cdot f_X(n) = 2^n$ iff $\underbrace{\forall n \cdot f_X(n) - 2^n = 0}_{\text{ZERONESS PROBLEM}}$.
- Issue: f_X is complicated in general.

$g_X(n) :=$ "# derivations of words of length n accepted by X".

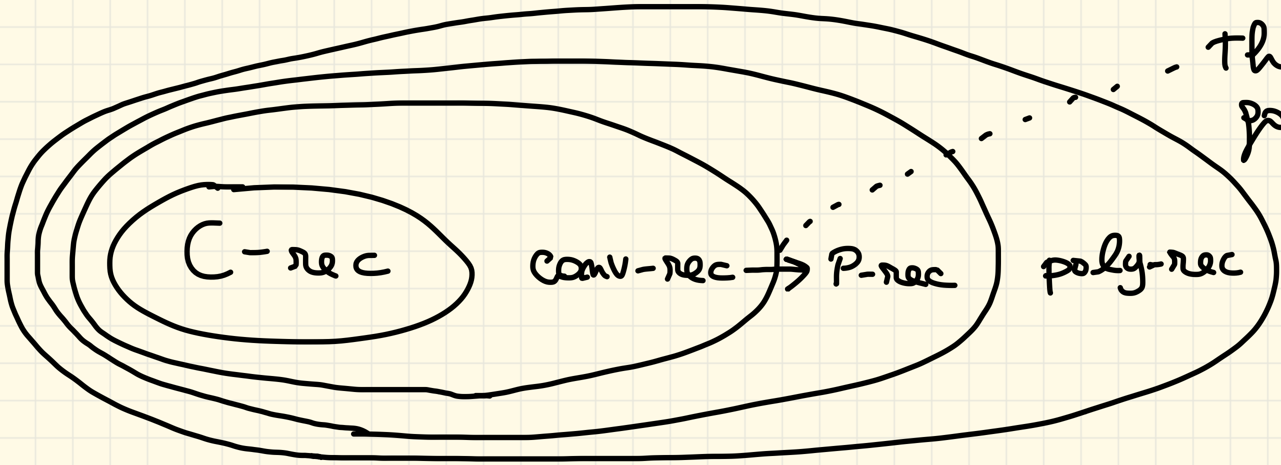
- Unambiguity $\Rightarrow f_X = g_X$.
- Recursive decomposition:

$$g_X(0) = 1,\ g_X(1) = 0,\ g_X(n+2) = \sum_{k=0}^{n} g_X(k) g_X(n-k) = \underbrace{(g_X * g_X)}_{\text{convolution product}}(n)$$

- $g_X : \mathbb{N} \to \mathbb{Q}$ is a convolution-recursive sequence (conv-rec)
- zeroness for conv-rec sequences is in PSPACE.

UNIVERSALITY → ZERONESS PROBLEM

UCFG	conv-rec	$f(n+1) = (f * f)(n)$ Catalan numbers
Unambiguous finite automata	C-recursive	$f(n+2) = 3 \cdot f(n+1) + 7 \cdot f(n)$ Fibonacci
Unambiguous Parikh automata	P-recursive	$f(n+2) = \frac{(17n^2+51n+39)(2n+3)}{(n+2)^3} \cdot f(n+1) + \left(\frac{n+1}{n+2}\right)^2 f(n)$ [Bostan, Carayol, Koechlin, Nicaud | CALP' 20]
Unambiguous register automata	linear-recursive 2 dim. $\mathbb{N}^2 \to \mathbb{Q}$	$f(n+1, k+1) = f(n,k) + (k+1) \cdot f(n, k+1)$ Stirling numbers of the second kind
?	polynomial recursive	$f(n+1) = f^2(n)$



the conversion goes via generating functions is not efficient

GENERATINGFUNCTIONOLOGY [Wilf '89]

$$f : \mathbb{N} \to \mathbb{Q} \quad \longleftrightarrow \quad \hat{f}(x) := \sum_{n=0}^{\infty} f(n) x^n : \mathbb{C} \to \mathbb{C},$$ formal power series

$h = f + g$		$\hat{h}(x) = \hat{f}(x) + \hat{g}(x)$	linearity
$= k \cdot f$		$= k \cdot \hat{f}(x)$	linearity
$= f * g$	convolution	$= \hat{f}(x) \cdot \hat{g}(x)$	product
$= f(n+1)$	shift	$= \frac{1}{x} \cdot \left(\hat{f}(x) - f(0) \right)$	rational transf.
$= n \cdot f(n)$	polynomial mult.	$= x \cdot \frac{\partial}{\partial x} \hat{f}(x)$	differentiation

NUMBER SEQUENCES & THEIR GENERATING FUNCTIONS

$\mathbb{N} \to \mathbb{Q}$

C-recursive

$$\begin{cases} f(n+1) = a f(n) + b g(n) \\ g(n+1) = c f(n) + d g(n) \end{cases}$$

Convolution recursive

$$\begin{cases} f(n+1) = a + b \cdot f(n) + c \cdot (f * g)(n) + d \cdot g(n) \\ g(n+1) = \cdots \end{cases}$$

P-recursive

$$\begin{cases} f(n+1) = (a + bn) f(n) + (c + dn) g(n) \\ g(n+1) = \cdots \end{cases}$$

$\mathbb{C} \to \mathbb{C}$ / formal power series

Rational

$$\begin{cases} \hat{f}(x) = f(0) + x \cdot (a \hat{f}(x) + b \hat{g}(x)) \\ \hat{g}(x) = g(0) + x \cdot (c \hat{f}(x) + d \hat{g}(x)) \end{cases}$$

$$\left(\to \hat{f}(x) = \frac{f(0)(1 - dx) + g(0) \cdot bx}{1 - (a + d + (bc - ad)x)x} \right)$$

Algebraic

$$\begin{cases} \hat{f}(x) = f(0) + x \cdot (a + b \hat{f}(x) + c \hat{f}(x) \hat{g}(x) + d \hat{g}(x)) \\ \hat{g}(x) = \cdots \end{cases}$$

$\cap|$ [Comtet '64]

D-finite / holonomic

$$\begin{cases} \hat{f}(x) = f(0) + x \left(a \hat{f}(x) + bx \frac{\partial}{\partial x} \hat{f}'(x) + c \hat{g}(x) + dx \frac{\partial}{\partial x} \hat{g}'(x) \right) \\ \hat{g}(x) = \cdots \end{cases}$$

CONVOLUTION-RECURSIVE SEQUENCES

$$\begin{cases} f_1(n+1) = P_1(f_1(n), \dots, f_k(n)) \\ \vdots \\ f_k(n+1) = P_k(f_1(n), \dots, f_k(n)) \end{cases}$$

$P_1, \dots, P_k \in \mathbb{Q}[x_1, \dots, x_k]$

multiplication interpreted as convolution

Zeroness problem: $\forall n .\ f_1(n) = 0$? Non-trivial: $f_2, \dots, f_k$ need not be 0

Theorem: Zeroness of conv-rec sequences is in PSPACE.

Proof: Take generating functions $\hat{f}_i(x) = \sum_{n=0}^{\infty} f_i(n) x^n$. R -> R

and study the system of polynomial equations:

$$\begin{cases} \hat{f}_1(x) = f_1(0) + x \cdot p_1(\hat{f}_1(x), \dots, \hat{f}_k(x)) \\ \vdots \\ \hat{f}_k(x) = f_k(0) + x \cdot p_k(\hat{f}_1(x), \dots, \hat{f}_k(x)) \end{cases}$$

1) unique power series solution.

2) $\hat{f}_i(x)$ convergent for $0 \leq x < (\text{total degree})^{-1}$.

$f_1 = 0$ iff $\hat{f}_1 = 0$ iff $\underbrace{\forall (0 < x < (\text{total degree})^{-1}), \bar{y} \cdot \bar{y} = \bar{f}(0) + x \cdot \bar{p}(\bar{y}) \Rightarrow y_1 = 0}_{\text{sentence of } \exists \mathbb{R}(+, \cdot, 0, 1), \text{ in PSPACE [Canny'88]}}$

LOWER BOUNDS for UCFG UNIVERSALITY? NO

Quantitative generalisation of universality : COIN-FLIP measure

$\mu(L)$ = "probability that a random word $w \in \Sigma^*$ is in L" $= \sum_{n=0}^{\infty} |L \cap \Sigma^n| \left(\frac{1}{|\Sigma|+1}\right)^{n+1}$

1) $\mu(L)$ is algebraic for L rec. by UCFG [Chomsky-Schützenberger '63].

2) $\mu(L) = 1$ iff $L = \Sigma^*$.

3) $\mu(L) \overset{?}{\geq} \varepsilon$ is SQRT-SUM hard

- SQRT-SUM ∈ PSPACE
- not known to be in NP
- not known to be NP-hard
- long-standing open problem

Encode SQRT-SUM as $\mu(L(G)) \overset{?}{\geq} \varepsilon$ for G a UCFG

(U: challenging.)

$NFA \overset{?}{\subseteq} UCFG \leqslant DFA \overset{?}{\subseteq} UCFG \leqslant$ UCFG universality $\in$ PSPACE

$\lor$|

UCFG universality $\longrightarrow$ $NFA \overset{?}{\subseteq} UCFG$ is "UCFG universality" - Complete.

$CFG \overset{?}{\subseteq} UFA \leqslant DCFG \overset{?}{\subseteq} UFA \leqslant$ UCFG universality $\in$ PSPACE

- Also reduces to

$$\mu_G(A) \overset{?}{=} 1$$

- for a stochastic CFG G and UFA A.
- in PTIME for a DFA A;
  $\mu_G(A)$ can even be approximated in PTIME

[Etessami & Stewart & Yannakakis ICALP'13]

OPEN PROBLEM: approximate $\mu_G(A)$ in PTIME for UFA A?

UCFG vs. SCFG: UNAMBIGUOUS vs. STOCHASTIC

Stochastic CFG: deterministic CFG + probabilities

$$\begin{cases} X \xrightarrow{2/3} aXaX \\ X \xrightarrow{1/3} \varepsilon \end{cases}$$

inducing a probability distribution $\mu_X : \Sigma^* \to \mathbb{R}_{\geq 0}$

Weighted

Unambiguous CFG $\longrightarrow$ Stochastic CFG

$X \to aXbX \mid \varepsilon$

$f(n+2) = (f * f)(n)$

counting function

$\nearrow \; \mu_{cf}(X) = 1 + \frac{1}{9} \cdot [\mu_{cf}(X)]^2 \; \to \begin{cases} X \xrightarrow{1/9} aXbX \\ X \xrightarrow{1} \varepsilon \end{cases}$

Coin-flip measure

$\mu_{cf}(X) = \mu_G(\Sigma^*)$

$$\mu_G(\Sigma^*) \stackrel{?}{=} 1 \text{ in PTIME if } G \text{ is a stochastic CFG}$$

[Etessami & Yannakakis JACM'09, Esparza & Gaiser & Kiefer STACS'10]

SUMMARY: COMPLEXITY & DECIDABILITY of INCLUSION

$\overset{?}{\subseteq}$	DFA	UFA	NFA	DCFG	UCFG	CFG
DFA	PTIME	PTIME	PSPACE-c.	PTIME	= UNIVERSALITY of UCFG	UNDECIDABLE
UFA	PTIME	PTIME	PSPACE-c.	PTIME	= UNIVERSALITY of UCFG	UNDECIDABLE
NFA	PTIME	NEW	PSPACE-c.	PTIME	= UNIVERSALITY of UCFG	UNDECIDABLE
DCFG	PTIME	≤ UNIVERSALITY of UCFG	EXPTIME-c. [Kasai & Iwata '92]	UNDECIDABLE [Hopcroft & Ullman '79]	UNDECIDABLE [Hopcroft & Ullman '79]	UNDECIDABLE [Hopcroft & Ullman '79]
UCFG	PTIME	≤ UNIVERSALITY of UCFG	EXPTIME-c. [Kasai & Iwata '92]	UNDECIDABLE [Hopcroft & Ullman '79]	UNDECIDABLE [Hopcroft & Ullman '79]	UNDECIDABLE [Hopcroft & Ullman '79]
CFG	PTIME	≤ UNIVERSALITY of UCFG	EXPTIME-c. [Kasai & Iwata '92]	UNDECIDABLE [Hopcroft & Ullman '79]	UNDECIDABLE [Hopcroft & Ullman '79]	UNDECIDABLE [Hopcroft & Ullman '79]

DCFG = DCFG decidable [Sénizergues '97]

UCFG = UCFG open, as well as its generalisation:

Multiplicity equivalence of CFG

(same # derivations for every input).

CONCLUSIONS

- Universality of unambiguous CFGs problem:
  - → in PSPACE via reduction to zeroness of conv-rec sequences.
  - → computing the coin-flip measure is SQRT-sum hard.
  - → NFA $\overset{?}{\subseteq}$ UCFG & CFG $\overset{?}{\subseteq}$ UFA reduce to it, so also in PSPACE.

  OPEN PROBLEM: exact complexity? lower bounds?