An Algebraic Generalization of the Chomsky-Schützenberger-Theorem

Hans Leiß
leiss@cis.uni-muenchen.de
extending work by/joint with Mark Hopkins

2017 retired from:
Universität München
Centrum für Informations- und Sprachverarbeitung

Oberseminar Theoretische Informatik, LMU, 10.5. 2019
Oberseminar Mathematische Logik, LMU, 22.5. 2019

Plan

- Algebraic generalization of formal language theory: itempotent semirings D with sums $\sum U$ of suitable subsets $U \subseteq D$:

$$\mathcal{R}D = \text{regular subsets}, \quad \mathcal{C}D = \text{context-free subsets}, \quad \text{etc.}$$

- The Chomsky-Schützenberger-Theorem: how to obtain $\mathcal{C}X^*$ from $\mathcal{R}(X \cup \Delta)^*$ and a single language $Dyck \in \mathcal{C}(X \cup \Delta)^*$
- Quotients and tensor products: $\mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}\Delta^*/\rho$
- A generalization of the CST for arbitrary monoids

$$\mathcal{C}M = Q(\mathcal{R}M) = Z_2(\mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}\Delta^*/\rho)$$

  i.e. $\mathcal{C}M$ is an algebraic function of $\mathcal{R}M$.
- Useful to name $L \in \mathcal{C}X^*$ by regular expressions over $X \cup \Delta$.

Reminder: Formal Language Theory for free monoid X^*

language family	set X finite	Construction of $L \subseteq X^*$ by	Automaton recognizing $\{w \in X^* \mid w \in L\}$
Finite	$\mathcal{F}X^*$	$\emptyset, \{x\}, \cup, \cdot$	finite dir.acyclic graph
Regular	$\mathcal{R}X^*$	$\emptyset, \{x\}, \cup, \cdot, ^*$	finite automaton
Context-free	$\mathcal{C}X^*$	$\emptyset, \{x\}, \cup, \cdot, \mu$	push-down automaton
Context-sensitive	$\mathcal{S}X^*$	...	linearly bounded TM
Rec. enumerable	$\mathcal{T}X^*$	...	Turing machine (TM)
Arbitrary languages	$\mathcal{P}X^*$	...	—

Construction more precisely:

- elementwise product: $A \cdot B = \{a \cdot b \mid a \in A, b \in B\}$
- iteration: $A^* = \bigcup\{A^n \mid n \in \mathbb{N}\}$, $A^0 = \{1\}$, $A^{n+1} = A \cdot A^n$
- recursion: for polynomial $p(y, \bar{z}) \in X^*[y, \bar{z}]$ and $\bar{A} \in (\mathcal{C}X^*)^m$, the least solution $\mu y p(\bar{A})$ of $y \supseteq p(y, \bar{A})$ in $\mathcal{P}X^*$ is in $\mathcal{C}X^*$

Chomsky-hierarchy: $\mathcal{F}X^* \subset \mathcal{R}X^* \subset \mathcal{C}X^* \subset \mathcal{S}X^* \subset \mathcal{T}X^* \subset \mathcal{P}X^*$

Algebraization of Formal Language Theory

We work with

$\mathbb{M}$	the category of monoids $(M, \cdot, 1)$ and homomorphisms,
$\mathbb{D}$	the category of dioids $(D, +, \cdot, 0, 1)$ (= idempotent semirings) and semiring homomorphisms.
$\mathbb{M}_{\leq}$	the category of partially ordered monoids $(M, \cdot, 1, \leq)$ and monotone homomorphisms.

Each dioid D implicitly has a partial order $\leq$ defined by

$$d \leq d' \iff d + d' = d'.$$

The power-set functor $\mathcal{P} : \mathbb{M} \to \mathbb{D}$ assigns to a monoid M a dioid

$\mathcal{P}M = (|\mathcal{P}M|, \cup, \cdot, \emptyset, \{1\}),$ where $A \cdot B := \{a \cdot b \mid a \in A, b \in B\}$,

and to each homomorphism $f : M \to N$ a dioid-homomorphism

$$\mathcal{P}f \ = \ \lambda A \, \{f(a) \mid a \in A\} : \mathcal{P}M \to \mathcal{P}N.$$

A monadic opertor $\mathcal{A}$ (Hopkins 2008) is a subfunctor of the power-set functor $\mathcal{P} : \mathbb{M} \to \mathbb{D}$ that satisfies, for each monoid M,

- A_0 $\mathcal{A}M$ is a set of subsets of M: $\mathcal{A}M \subseteq \mathcal{P}M$,
- A_1 $\mathcal{A}M$ contains each finite subset of M: $\mathcal{F}M \subseteq \mathcal{A}M$,
- A_2 $\mathcal{A}M$ is closed under product (hence a monoid),
- A_3 $\mathcal{A}M$ is closed under union of sets from $\mathcal{A}\mathcal{A}M$ (hence a dioid),
- A_4 $\mathcal{A}$ preserves homomorphisms: if $f : M \to N$ is a homomorphism, so is $\mathcal{A}f := \lambda U\ \{f(u) \mid u \in U\} : \mathcal{A}M \to \mathcal{A}N$.

We say $\mathcal{A} \leq \mathcal{A}'$ iff $\mathcal{A}M \subseteq \mathcal{A}'M$ for each monoid M.

Theorem (Hopkins 2008)

$\mathcal{F} \leq \mathcal{R} \leq \mathcal{C} \leq \mathcal{T} \leq \mathcal{P}$ are monadic operators. ($\mathcal{S}$ is not: A_4)

Remark

A monadic operator $\mathcal{A}$ is the left adjoint of an adjunction $(\mathcal{A}, \widehat{\mathcal{A}}, \eta, \epsilon) : \mathbb{M} \to \mathbb{D}\mathcal{A}$ between $\mathbb{M}$ and a category $\mathbb{D}\mathcal{A} \subseteq \mathbb{D}$, where

- $\widehat{\mathcal{A}} : \mathbb{D}\mathcal{A} \to \mathbb{M}$ is the forgetful functor, and if $M \in \mathbb{M}, D \in \mathbb{D}\mathcal{A}$,
- $\eta_M : M \to \mathcal{A}M$ is $m \mapsto \{m\}$, $\epsilon_D : \mathcal{A}D \to D$ is $U \mapsto \sum U$.

This adjunction gives rise to a monad $T_{\mathcal{A}} = (\widehat{\mathcal{A}} \circ \mathcal{A}, \eta, \mu)$ in $\mathbb{M}$, an endofunctor $T = \widehat{\mathcal{A}} \circ \mathcal{A} : \mathbb{M} \to \mathbb{M}$, where

- the unit $\eta : I \to T$ maps $m \in M$ to $\{m\} \in \mathcal{A}M$,
- the product $\mu : TT \to T$ maps $\mathcal{U} \in \mathcal{A}\mathcal{A}M$ to $\bigcup \mathcal{U} \in \mathcal{A}M$.

$\widehat{\mathcal{A}}$ is called a monadic functor in category theory.

$\mathbb{D}\mathcal{A} \simeq \mathbb{M}^T$, the Eilenberg-Moore category of T-algebras $(D, \sum_D)$.

The category $\mathbb{D}\mathcal{A} \subseteq \mathbb{D}$ of $\mathcal{A}$-dioids

For a partial order M, by $x > X$ means x is an upper bound of X.

A map $f : M \to N$ between partially ordered monoids M, N is $\mathcal{A}$-continuous, if for all $U \in \mathcal{A}M$ and $n > (\mathcal{A}f)(U)$ there is some $m > U$ with $n \geq f(m)$.

An $\mathcal{A}$-morphism is an $\mathcal{A}$-continuous monotone homomorphism.

An $\mathcal{A}$-dioid is a partially ordered monoid $M = (M, \cdot, 1, \leq)$ which is

- $\mathcal{A}$-complete: every $U \in \mathcal{A}M$ has a supremum $\sum U \in M$, and
- $\mathcal{A}$-distributive: for all $U, V \in \mathcal{A}M$, $\sum(UV) = (\sum U)(\sum V)$.

Let $\mathbb{D}\mathcal{A}$ be the category of $\mathcal{A}$-dioids with $\mathcal{A}$-morphisms.

- $\mathbb{D}\mathcal{F}$ is the category $\mathbb{D}$ of dioids ($a + b := \sum\{a, b\}, 0 := \sum \emptyset$).
- $\mathbb{D}\mathcal{P}$ is the category of (unital) quantales.

Since every $\mathcal{A}$-dioid $(M, \cdot, 1, \leq)$ is a ($\mathcal{F}$-) dioid via

$$a + b := \sum \{a, b\}, \qquad 0 := \sum \emptyset,$$

we often write $\mathcal{A}$-dioids in the dioid-signature: $D = (D, +, \cdot, 0, 1)$.

Prop. $\mathcal{A}X^*$ is the free $\mathcal{A}$-dioid generated by the set X.
$\mathcal{A}M$ is the free $\mathcal{A}$-dioid extension of the monoid M.

Prop. (i) For $f : M \to N$ between $\mathcal{A}$-dioids M, N:

f is $\mathcal{A}$-continuous iff for all $U \in \mathcal{A}M$: $f(\sum U) = \sum (\mathcal{A}f)(U)$

(ii) An $\mathcal{A}$-complete po-monoid $(M, \cdot, 1, \leq)$ is $\mathcal{A}$-distributive iff

$$\text{forall } a, b \in M, U \in \mathcal{A}M : a(\sum U)b = \sum aUb.$$

Theorem

Hopkins 2008: $\mathbb{D}\mathcal{R}$ *is the category of* **-continuous Kleene algebras.*
HL 2018: $\mathbb{D}\mathcal{C}$ *is the category of* μ*-continuous Chomsky algebras.*

Kleene-Algebra (Kozen 1990): right/left-linearly closed dioid

$x \geq ax + b$ and $x \geq xa + b$ have least solutions a^*b resp. ba^*, for all values a, b.

*-continuity: $a \cdot c^* \cdot b = \sum\{a \cdot c^n \cdot b \mid n \in \mathbb{N}\}$, for all $a, b, c \in M$.

Chomsky-Algebra (Grathwohl e.a. 2015): algebraically closed dioid

every polynomial system $x_1 \geq p_1(\bar{x}, \bar{y}), \ldots, x_n \geq p_n(\bar{x}, \bar{y})$ has a least solution in $\bar{x} = x_1, \ldots x_n$, for each value of $\bar{y}$.

μ-continuity: $a \cdot \mu x p \cdot b = \sum\{a \cdot p^n(0) \cdot b \mid n \in \mathbb{N}\}$, all $p \in M[x]$.

Theorem (Hopkins 2008)

For all monadic operators $\mathcal{A} \leq \mathcal{B}$ there is an adjunction

$$Q^{\mathcal{B}}_{\mathcal{A}} : \mathbb{D}\mathcal{A} \rightleftarrows \mathbb{D}\mathcal{B} : Q^{\mathcal{A}}_{\mathcal{B}}$$

where $Q^{\mathcal{B}}_{\mathcal{A}}(K)$ is a completion of K by order-ideals of $\mathcal{B}$-subsets of K, and $Q^{\mathcal{A}}_{\mathcal{B}}$ is the forgetful functor (resp. restriction of $\sum$).

For monoids M, $\mathcal{C}M = Q^{\mathcal{C}}_{\mathcal{R}}(\mathcal{R}M)$ is the algebraic closure of $\mathcal{R}M$.

Problem

Can we provide an algebraic construction of $Q^{\mathcal{C}}_{\mathcal{R}}$?

Intended advantage:
algebraic expressions for context-free languages, instead of μ-terms.

The classical CST for free monoids

$$\begin{array}{rcl} X^* & = & (X^*, \cdot, 1) \text{ the free monoid generated by the fin.set } X \\ M[\Delta] & = & \text{the free extension of the monoid } M \text{ by the set } \Delta \\ & = & \text{all interleaved sequences of elements of } M \text{ and } \Delta^* \end{array}$$

Theorem (Chomsky/Schützenberger 1963)

Let

- $\Delta = \{b, d, p, q\}$ *consist of two pairs* b, d *and* p, q *of brackets,*
- $h : X^*[\Delta] \to X^*$ *be the "bracket-erasing" homomorphism,*
- $D \in \mathcal{C}(X^*[\Delta])$ *be the Dyck-language, the least* $S \subseteq X^*[\Delta]$ *s.t.*

$$S \geq 1 + X + bSd + pSq + SS$$

Then: $\mathcal{C}X^* = \{h(R \cap D) \mid R \in \mathcal{R}(X^*[\Delta])\}$.

This is not yet $\mathcal{C}X^* = Q^{\mathcal{C}}_{\mathcal{R}}(\mathcal{R}X^*)$, but a first step towards our goal.

The polycyclic monoid P'_n and $\mathcal{R}$-dioid C'_n

We are looking for an algebra in which $h(R \cap D)$ can be performed.

An inverse semigroup is a semigroup $(M, \cdot)$ where each element p has a "generalized inverse" p^{-1}, i.e. a q such that

$$p = pqp \quad \text{and} \quad q = qpq.$$

Example: partial bijections $p \subseteq X \times X$ of X under composition.

Let $\Delta_n = P_n \dot\cup Q_n$, for $P_n = \{p_0, \ldots, p_{n-1}\}$, $Q_n = \{q_0, \ldots, q_{n-1}\}$, and $(\Delta_n^*)_0$ the extension of Δ_n^* by an annihilating element 0.

The polycyclic monoid P'_n is the inverse monoid $(\Delta_n^*)_0/\rho_n$ where

$$\rho_n = \{p_i q_i = 1 \mid i < n\} \cup \{p_i q_j = 0 \mid i, j < n, i \neq j\}.$$

Here, p_i and q_i are generalized inverses of each other, as $p_i q_i = 1$.

Note: the Dyck-language over Δ_n is $D_n = \{w \in \Delta_n^* \mid w/\rho_n = 1\}$.

The Cayley-graph of P'_n

The Cayley-graph of P'_n is the graph $(P'_n, \xrightarrow{p_0}, \ldots, \xrightarrow{q_{n-1}})$ with

$$u/\rho_n \xrightarrow{p_i} up_i/\rho_n, \qquad u/\rho_n \xrightarrow{q_i} uq_i/\rho_n.$$

We have

$$\Delta_n^* \;=\; \Delta_n^* P_n Q_n \Delta_n^* \;\dot\cup\; Q_n^* P_n^*.$$

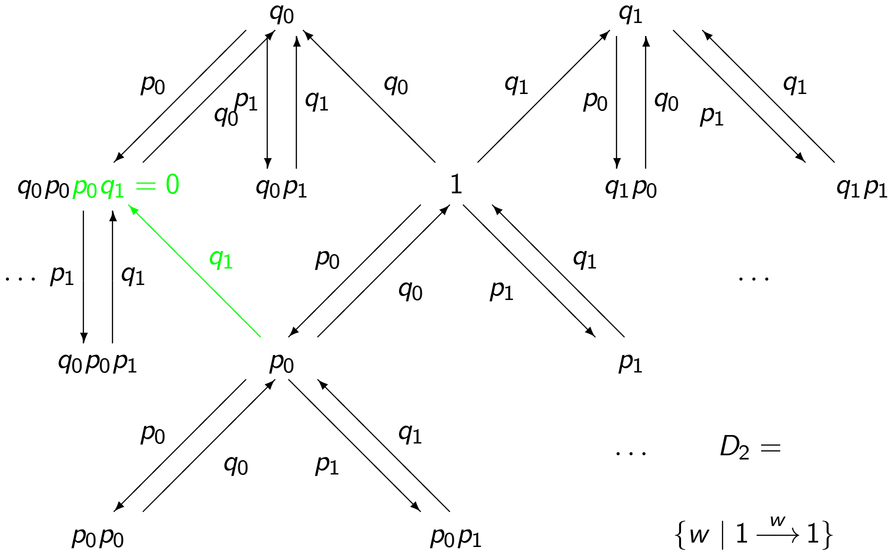
Hence every $w \in \Delta_n^*$ has a *normal form* $nf_{\rho_n}(w)$ in

$$\{0\} \cup Q_n^* P_n^*.$$

The normal forms represent the elements of P'_n, and $1 \xrightarrow{w} nf_{\rho_n}(w)$.

The monoid $P'_n \simeq (Q_n^* P_n^* \cup \{0\}, \cdot', 1)$ with $u \cdot' v = nf_{\rho_n}(uv)$ is the *partial* monoid $Q_n^* P_n^*$ if $u \cdot' v = 0$ is read as "$u \cdot' v$ is undefined".

The Cayley-graph of P_2' (without 0 and edges related to 0):



If we could restrict $\xrightarrow{p_i}$, $\xrightarrow{q_j}$ to $P_2^* \subseteq P_2'$, we had a stack!

The polycyclic $\mathcal{R}$-dioid C'_n is $\mathcal{R}\Delta_n^*/\rho_n$, viewing ρ_n as semiring equations

$$\{p_i\}\{q_i\} = \{1\}, \qquad \{p_i\}\{q_j\} = \emptyset, \quad (i \neq j).$$

This is essentially the $\mathcal{R}$-dioid $\mathcal{R}P_n$ obtained from the polycyclic monoid $P'_n \simeq (Q_n^* P_n^* \cup \{0\}, \cdot', 1)$ with $u \cdot' v = nf_{\rho_n}(uv)$; we only have to remove the "non-existing" monoid-element 0:

Prop. $C'_n \simeq \mathcal{R}P'_n/(\{0\} = \emptyset)$. [Quotients in $\mathbb{D}\mathcal{R}$: later]

We can code $n \geq 2$ bracket pairs p_i, q_i by two, i.e. in $\Delta^* = \Delta_2^*$ by

$$p_i := bp^i, \qquad q_i := q^i d \qquad (i < m).$$

Question

Can we perform $h(R \cap D_n(M))$ algebraically, for $R \in \mathcal{R}(M[\Delta_n])$?

Assume we can move the monoid elements aside, i.e. take $M = 1$.

To go from $R \in \mathcal{R}\Delta_n^*$ to $R \cap D_n$, let $p = p_{n+1}, q = q_{n+1}$. Then

- $w \in D_n \iff w/\rho_n = 1 \iff pwq/\rho_{n+1} = 1$,
- $w \notin D_n \iff w/\rho_n \neq 1 \iff pwq/\rho_{n+1} = 0$. $(Q_n^* P_n^* \setminus \{1\})$

Since $w \in \Delta_n^* \mapsto (w, w/\rho_{n+1}) \in \Delta_n^* \times P'_{n+1}$ is a homomorphism,

$$R' = \{(w, w/\rho_{n+1}) \mid w \in R\} \quad \in \quad \mathcal{R}(\Delta_n^* \times P'_{n+1}).$$

Therefore, multiplying R' by $\{(1, p/\rho_{n+1})\}$ and $\{(1, q/\rho_{n+1})\}$,

- $S' := \{(w, pwq/\rho_{n+1}) \mid w \in R\} \in \mathcal{R}(\Delta_n^* \times P'_{n+1})$,
- $S' \subseteq \Delta_n^* \times \{0, 1\}$.

Then, with a suitable product $\cdot$ and $\sum$ over the regular set S',

$$\sum\{w \cdot c \mid (w, c) \in S'\} = \sum\{w \cdot 1 \mid (w, 1) \in S'\} = R \cap D_n.$$

To make the final parts precise, we will have to do three things:

I. show that $\mathbb{D}\mathcal{R}$ has quotients and tensor products, such that

- there is an $\mathcal{R}$-dioid $\mathcal{R}\Delta_n^*/\rho_n = C_n'$
- for $\mathcal{R}$-dioids D_1 and D_2, there is an $\mathcal{R}$-dioid $D_1 \otimes_{\mathcal{R}} D_2$
- there is an infinite sum $\sum : \mathcal{R}(D_1 \otimes_{\mathcal{R}} D_2) \to D_1 \otimes_{\mathcal{R}} D_2$
- elements of D_1 and D_2 commute in $D_1 \otimes D_2$
- $\mathcal{R}(M_1 \times M_2) \simeq \mathcal{R}M_1 \otimes_{\mathcal{R}} \mathcal{R}M_2$, for monoids M_1, M_2

II. use $\mathcal{R}(M \times P'_{n+1}) \simeq \mathcal{R}M \otimes \mathcal{R}P'_{n+1}$ to get

- $w \cdot c$ by $\{w\} \otimes \{c\}$,
- $S := \{\{w\} \otimes \{c\} \mid (w, c) \in S'\} \in \mathcal{R}(\mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}P'_{n+1})$
- $\sum\{w \cdot c \mid (w, c) \in S'\}$ by $\sum S \in \mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}P'_{n+1}$

III. replace $\mathcal{R}P'_{n+1}$ by $C'_{n+1} = \mathcal{R}P'_{n+1}/(\{0\} = \emptyset)$ in II. to get

- $w \cdot 0 = \{w\} \otimes \{0\} = \{w\} \otimes \emptyset = 0$ (and so eliminate $R \cap \overline{D_n}$).

For the general case of monoid M instead of Δ_n^*, we also have to

IV. get (finitely generated) M "out of the way" by

- going from $M[\Delta_n]$ to $M \times \Delta_n^*$ and intersect $R \in \mathcal{R}(M \times \Delta_n^*)$ with $M \times D_n \in \mathcal{C}(M \times \Delta_n^*)$ to get

$$h(R \cap (M \times D_n)) = \sum \{m \cdot pwq/\rho_{n+1} \mid (m, w) \in R\}$$

V. for $L \in \mathcal{C}M$, find n and $R \in \mathcal{R}(M[\Delta_n])$ with $L = h(R \cap D_n(M))$

- by a proof of (CST) $\mathcal{C}X^* \subseteq \{h(R \cap D_n(X)) \mid R \in \mathcal{R}(X^*[\Delta_n])\}$

Therefore, we should try to prove

$$\mathcal{C}M \subseteq \mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}\Delta_n^*/\rho_n = \mathcal{R}M \otimes_{\mathcal{R}} C_n'$$

and read off *which* elements of $\mathcal{R}M \otimes_{\mathcal{R}} C_n'$ belong to $\mathcal{C}M$.

Application: algebraic terms for $L \in \mathcal{C}X^*$, not μ-terms

Context-free language over X are named by regular expressions evaluated in $\mathcal{R}X^* \otimes_{\mathcal{R}} C_2'$:

$$r, s \quad := \quad x \mid 0 \mid 1 \mid (r + s) \mid (r \cdot s) \mid r^* \mid \langle 0| \mid |0\rangle \mid \langle 1| \mid |1\rangle$$

Example: $X = \{a, b\}$, $L = \{a^n b^n \mid n \in \mathbb{N}\}$.
Write a for its value $\{a\} \otimes 1$ in $\mathcal{R}X^* \otimes_{\mathcal{R}} C_2'$, $\langle 0|$ for $\{1\} \otimes \{\langle 0|\}$ etc.

$$\begin{aligned}
\langle 0|(a\langle 1|)^*(|1\rangle b)^*|0\rangle &= \langle 0|(\sum_n (a\langle 1|)^n)(\sum_m (|1\rangle b)^m)|0\rangle \\
&= \sum_{n,m} \langle 0|(a\langle 1|)^n(|1\rangle b)^m|0\rangle \quad (\mathcal{R}\text{-distrib.}) \\
&= \sum_{n,m} a^n b^m \langle 0|\langle 1|^n|1\rangle^m|0\rangle \quad (\text{rel.comm.}) \\
&= \sum_n a^n b^n \qquad (\langle 0|\langle 1|^n|1\rangle^m|0\rangle = \delta_{n,m})
\end{aligned}$$

Plan for the remaining parts

We now proceed as follows:

1. give a proof of the CST for free monoids X^* and monoids M
2. sketch why $\mathbb{D}\mathcal{A}$ has quotients D/ρ
3. sketch why $\mathbb{D}\mathcal{A}$ has tensor products $D_1 \otimes_{\mathcal{A}} D_2$
4. prove $\mathcal{C}M \lesssim \mathcal{R}M \otimes_{\mathcal{R}} C_2'$ and notice $\mathcal{C}M \lesssim Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$
5. prove $Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2') \lesssim \mathcal{C}M$ for monoids M

Outlook for algebraic closure of $\mathcal{R}$-dioids:

6. Sketch why $Q^{\mathcal{C}}_{\mathcal{R}} D \simeq Z_{C_2'}(D \otimes_{\mathcal{R}} C_2')$ might follow, for $D \in \mathbb{D}\mathcal{R}$.

The CST for free monoids X^*

Of the intended

$$\mathcal{C}X^* = \{h(R \cap D_2(X)) \mid R \in \mathcal{R}(X^*[\Delta_2])\},$$

the inclusion $\supseteq$ is by well-known facts:

- $R \in \mathcal{R}Y^*, D \in \mathcal{C}Y^* \Longrightarrow R \cap D \in \mathcal{C}Y^*$,
- $h : Y^* \to X^*$ a homomorphism, $L \in \mathcal{C}Y^* \Longrightarrow h(L) \in \mathcal{C}X^*$.

For $\subseteq$, via coding n brackets by two it is sufficient to show:

Theorem (Chomsky/Schützenberger 1963)

For every $L \in \mathcal{C}X^*$ there is $n \in \mathbb{N}$ and $R \in \mathcal{R}(X^*[\Delta_n])$ such that

$$L = h(R \cap D_n(X)), \qquad \textit{for } \Delta_n\textit{-erasure } h : X^*[\Delta_n] \to X^*.$$

Proof of the CST for X^* and M

a) The original proof (similar: Kozen 1997, Kanazawa 2010)

$L \in \mathcal{C}X^*$ has a context-free grammar $G = (Y, X, P, S)$ with finite set $P \subseteq Y \times (X^*[Y])$ of "grammar rules" and $S \in Y$.

The $w \in X^*$ that belong to $L = L(G)$ are those which have a "parse tree" with respect to G. A parse tree $t = [A, t_1, \ldots, t_k]$ with root category $A \in Y$ is projected to a string over $X^*[\Delta_m]$ via

$$\tau([A, t_1, \ldots, t_k]) = [_A\, \tau(t_1) \ldots \tau(t_k)\,]_A, \quad \tau([x]) = x \text{ for } x \in X.$$

From the rules in G one reads off a finite relation $I \subseteq (X \cup \Delta_m)^2$ such that the well-bracketed projections of parse trees belong to $D_2(X) \cap R$, where the regular set R is

$$R = [_S X^*[\Delta_n]]_S \setminus (X^*[\Delta_n]((X \cup \Delta_n)^2 \setminus I)X^*[\Delta_n]).$$

b) The proof in Harrison 1979: uses push-down automata

c) New proof. Let G' be CFG $\bar{Y} \geq \bar{p}'(\bar{Y}, \bar{x})$ over $X = \bar{x}$ be given, with n variable occurrences in $\bar{p}'$.

Let G or $\bar{Y} \geq \bar{p}(\bar{Y}, \bar{x})$ be G' with the i-th variable occurrence surrounded by brackets $\langle i|$ and $|i\rangle$ of Δ_n, $i < n$.

There is a bijection between the parse trees of G' and G, so it is sufficient to find $R \in \mathcal{R}(X^*[\Delta_n])$ with $L(G) = R \cap D_n(X)$.

Let G_F be the right-linear approximation of G obtained as follows:

- for each recursion variable Y_i, add a "continuation variable" $Y_{i,F}$ and distribute these to the summands in $\bar{Y} \geq \bar{p}(\bar{Y})\bar{Y}_F$:

$$\begin{array}{rcl} Y_1 & \geq & p_1(\bar{Y})Y_{1,F}, \\ Y_2 & \geq & p_2(\bar{Y})Y_{2,F}, \\ & \vdots & \\ Y_m & \geq & p_m(\bar{Y})Y_{m,F}, \end{array}$$

- state $S_F \to 1$ for the continuation of the main variable S and break the above rules (summands of $Y \geq p_Y(\bar{Y})Y_F$)

$$Y \to v_1\langle i_1|Y_1|i_1\rangle \dots v_k\langle i_k|Y_k|i_k\rangle v_{k+1}Y_F \quad (v_j \in X^*)$$

  into "right-linear pieces"

$$Y \to v_1\langle i_1|Y_1, \;\; Y_{1,F} \to |i_1\rangle v_2\langle i_2|Y_2, \;\; \dots \;\; Y_{k,F} \to |i_k\rangle v_{k+1}Y_F$$

Then $L(G_F) \in \mathcal{R}(X^*[\Delta_n])$, since G_F is right-linear. $L(G_F) \supseteq L(G)$, since the $Y_{i,F}$ collect right contexts of *all* occurrences of Y_i in G.

To show $L(G) = L(G_F) \cap D_n(X)$, notice $L(G) \subseteq D_n(X)$ and prove by induction on m: for variable Y and $w \in X^*[\Delta_n]$,

(i) if $Y \Rightarrow_G^m w$, then $Y \Rightarrow_{G_F}^* wY_F$,

(ii) if $Y \Rightarrow_{G_F}^m wY_F$ and $w \in D_n(X)$, then $Y \Rightarrow_G^* w$.

In (ii), we use $w \in D_n(X)$ to show that for $m > 1$ there are $i_1, \ldots, i_k$, $v_1, \ldots, v_{k+1} \in X^*$ and $w_1, \ldots, w_k \in D(X)$ such that $w = v_1\langle i_1|w_1|i_1\rangle \ldots v_k\langle i_k|w_k|i_k\rangle v_{k+1}$ and there are G_F-derivations

$$\begin{aligned} &A \Rightarrow_{G_F} v_1\langle i_1|A_1, \\ &A_j \Rightarrow^{m_j}_{G_F} w_j A_{j,F} \Rightarrow_{G_F} w_j|i_j\rangle v_{j+1}\langle i_{j+1}|A_{j+1} \quad \text{for } j < k, \\ &A_k \Rightarrow^{m_k}_{G_F} w_k A_{k,F} \Rightarrow_{G_F} w_k|i_k\rangle v_{k+1} A_F. \end{aligned}$$

By induction, $A_j \Rightarrow^*_G w_j$ for $j < k$, and by construction of G_F, there is $A \to v_1\langle i_1|A_1|i_1\rangle \ldots v_k\langle i_k|A_k|i_k\rangle v_{k+1}$ in G, so $A \Rightarrow^*_G w$.

The proof works for any monoid M instead of X^*, since the monoid product is never used (because of the brackets).

Example

Source grammar G'

$$\begin{aligned} S &\geq x + yTz, \\ T &\geq SS \end{aligned}$$

brackets added: G

$$\begin{aligned} S &\geq x + y\langle 1| T |1\rangle z, \\ T &\geq \langle 2|S|2\rangle\langle 3|S|3\rangle \end{aligned}$$

with continuation vars

$$\begin{aligned} S &\geq xS_F + y\langle 1| T |1\rangle zS_F, \\ T &\geq \langle 2|S|2\rangle\langle 3|S|3\rangle T_F \end{aligned}$$

right-linearized: G_F

$$\begin{aligned} S &\geq xS_F + y\langle 1| T, \\ T &\geq \langle 2|S \\ S_F &\geq |2\rangle\langle 3|S + |3\rangle T_F + 1 \\ T_F &\geq |1\rangle zS_F \end{aligned}$$

Minimal solution in $\mathcal{R}(X^*[\Delta_3])$:

$$\begin{aligned} S_F &= |3\rangle|1\rangle zS_F + |2\rangle\langle 3|S + 1 = (|3\rangle|1\rangle z)^*(|2\rangle\langle 3|S + 1) \\ S &= xS_F + y\langle 1|\langle 2|S = (y\langle 1|\langle 2| + x(|3\rangle|1\rangle z)^*|2\rangle\langle 3|)^* x(|3\rangle|1\rangle z)^* \end{aligned}$$

Corollary

Let M be a monoid. For each $L \in \mathcal{C}M$ there are $n \in \mathbb{N}$ and $R \in \mathcal{R}(M \times \Delta_n^*)$ such that

$$L = \pi_1(R \cap (M \times D_n)),$$

where $\pi_1 : M \times \Delta_n^* \to M$ is the first projection and $D_n \in \mathcal{C}\Delta_n$ the (pure) Dyck-language over Δ_n.

Proof: Take n, Δ_n and $R' \in \mathcal{R}(M[\Delta_n])$ and $h : M[\Delta_n] \to M$ as in the CST, such that $L = h(R' \cap D_n(M))$.

Let $e = (h, h_\Delta) : M[\Delta_n] \to M \times \Delta_n^*$ be the homomorphism where $h_\Delta : M[\Delta_n] \to \Delta_n^*$ erases elements of M. The claim follows for

$$R := e(R') \in \mathcal{R}(M \times \Delta_n^*).$$

Think of $(m, t) \in R \cap (M \times D_n)$ as m with a $\underbrace{\text{parse tree } t \text{ of } m}_{\text{proof of } m \in L}$.

The category $\mathbb{D}\mathcal{A}$ has quotients and tensor products

For a dioid D and $U, V \subseteq D$, put $U \simeq V :\iff U^{\leq} = V^{\leq}$, where

$$U^{\leq} := \{d \in D \mid d \leq u \text{ for some } u \in U\}.$$

For a dioid-congruence ρ on D, the set D/ρ of congruence classes is a dioid under the operations defined as expected.

An $\mathcal{A}$-congruence on an $\mathcal{A}$-dioid D is a dioid-congruence ρ s.th. for all $U, V \in \mathcal{A}D$, if $U/\rho \simeq V/\rho$, then $(\sum U)/\rho = (\sum V)/\rho$.

For any $E \subseteq D \times D$, there is a least $\mathcal{A}$-congruence on D above E.

Lemma

Let $q : D \to Q$ be an $\mathcal{A}$-morphism between $\mathcal{A}$-dioids D, Q. Then

$$ker(q) := \{(a, b) \mid q(a) = q(b),\ a, b \in D\}$$

is an $\mathcal{A}$-congruence on D.

Prop. If D is an $\mathcal{A}$-dioid and ρ an $\mathcal{A}$-congruence on D, then D/ρ is an $\mathcal{A}$-dioid and the canonical map $d \mapsto d/\rho$ is an $\mathcal{A}$-morphism.

Proof.

D/ρ is $\mathcal{A}$-complete: Each $U' \in \mathcal{A}(D/\rho)$ is U/ρ for some $U \in \mathcal{A}D$. Since ρ is an $\mathcal{A}$-congruence,

$$\sum U' := (\sum U)/\rho$$

is well-defined, and an upper bound of $U' = U/\rho$. Least: by A_3

Let e/ρ be any upper bound of U'. As $\{U, \{e\}\} \in \mathcal{F}\mathcal{A}D \subseteq \mathcal{A}\mathcal{A}D$, we have $U \cup \{e\} \in \mathcal{A}D$. By choice of e, $(U \cup \{e\})/\rho \simeq \{e\}/\rho$, so

$$(e + \sum U)/\rho = (\sum(U \cup \{e\}))/\rho = (\sum\{e\})/\rho = e/\rho.$$

Hence $(\sum U)/\rho \leq e/\rho$, and $\sum U'$ is a least upper bound of U'. □

In fact, $\mathbb{D}\mathcal{A}$ has coequalizers, and $\mathcal{A}$ lifts coequalizers in $\mathbb{M}$ to coequalizers in $\mathbb{D}\mathcal{A}$. This can be used to show

Theorem

Let E be a congruence on the monoid M, $\mathcal{A}E$ the least $\mathcal{A}$-congruence on $\mathcal{A}M$ above $\{(\{m\},\{m'\}) \mid (m,m') \in E\}$. Then

$$\mathcal{A}(M/E) \simeq \mathcal{A}M/\mathcal{A}E.$$

Writing ρ'_n for the lifting $\mathcal{R}\rho_n$ of the monoid-congruence

$$\rho_n = \{p_i q_j = \delta_{i,j} \mid i,j < n\},$$

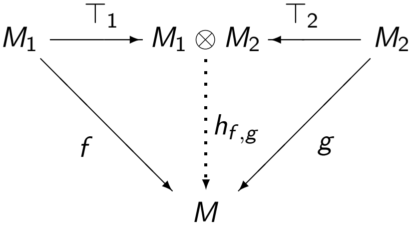
this gives us $\mathcal{R}P'_n := \mathcal{R}((\Delta^*_n)_0/\rho_n) = \mathcal{R}((\Delta^*_n)_0)/\rho'_n$, hence

$$\mathcal{R}P'_n/(\{0\} = \emptyset) = (\mathcal{R}((\Delta^*_n)_0)/\rho'_n)/(\{0\} = \emptyset) = \mathcal{R}\Delta^*_n/\rho'_n =: C'_n.$$

Tensor Product

In $\mathbb{M}$, two morphisms $f : M_1 \to M \leftarrow M_2 : g$ are relatively commuting, if for all $m_1 \in M_1, m_2 \in M_2$, $f(m_1)g(m_2) = g(m_2)f(m_1)$.

In a category whose objects have a monoid structure, a tensor product of two objects M_1 and M_2 is an object $M_1 \otimes M_2$ with two relatively commuting morphisms $\top_1 : M_1 \to M_1 \otimes M_2 \leftarrow M_2 : \top_2$ such that for any pair $f : M_1 \to M \leftarrow M_2 : g$ of relatively commuting morphisms the diagram



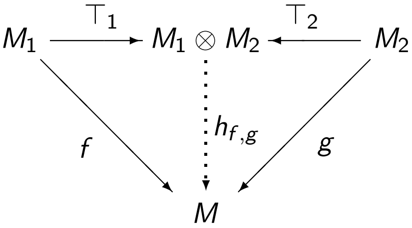
can be uniquely completed as shown.

Intuitively, $M_1 \otimes M_2$ is a free extension of both objects in which elements of one commute with elements of the other.

Example

In $\mathbb{M}$, the tensor product $\top_1 : M_1 \to M_1 \otimes M_2 \leftarrow M_2 : \top_2$ consists of $M_1 \otimes M_2 := M_1 \times M_2$ with $T_1(m) = (m, 1)$, $T_2(m') = (1, m')$.

The induced map $h_{f,g} : M_1 \times M_2 \to M$ is $h_{f,g}(a, b) = f(a)g(b)$.



Theorem

In the category $\mathbb{D}\mathcal{A}$, the tensor product of $\mathcal{A}$-dioids D_1, D_2 is

$$\top_1 : D_1 \to D_1 \otimes_{\mathcal{A}} D_2 \leftarrow D_2 : \top_2,$$

where we use

- *the underlying monoid $M_k := \widehat{\mathcal{A}} D_k$ of D_k, $k = 1, 2$,*
- *their tensor product $\hat{\top}_1 : M_1 \to M_1 \otimes M_2 \leftarrow M_2 : \hat{\top}_2$ in $\mathbb{M}$,*
- *the least $\mathcal{A}$-congruence $\equiv$ on $\mathcal{A}(M_1 \otimes M_2)$ above*

$$\{(\{(\sum A, \sum B)\}, A \times B) \mid A \in \mathcal{A}M_1, B \in \mathcal{A}M_2\},$$

to put $D_1 \otimes_{\mathcal{A}} D_2 := \mathcal{A}(M_1 \otimes M_2)/_{\equiv}$ and $\top_k(d_k) = \{\hat{\top}_k(d_k)\}/_{\equiv}$.

The induced map of $f : D_1 \to D \leftarrow D_2 : g$ is

$$h_{f,g}(U/_{\equiv}) := \sum\{f(a)g(b) \mid (a, b) \in U\}, \quad U \in \mathcal{A}(M_1 \times M_2).$$

The $\mathcal{A}$-congruence $\equiv$ is needed to make T_i be $\mathcal{A}$-morphisms.

Notation: For $a \in D_1, b \in D_2$, we write

$$a \otimes 1 := T_1(a) = \{(a, 1)\}/_{\equiv} \qquad 1 \otimes b := T_2(b) = \{(1, b)\}/_{\equiv}$$

For $U \in \mathcal{A}(D_1 \times D_2)$, $U/_{\equiv} \in D_1 \otimes_{\mathcal{A}} D_2$ can be written as

$$U/_{\equiv} = \sum \{a \otimes b \mid (a, b) \in U\} \; =: [U].$$

Prop. $\mathcal{A}M_1 \otimes_{\mathcal{A}} \mathcal{A}M_2 \simeq \mathcal{A}(M_1 \otimes M_2)$ for monoids M_1, M_2.

Proof (Sketch) Recall $M_1 \otimes M_2 = M_1 \times M_2$ and use

$$\begin{aligned} U/_{\equiv} \quad &\mapsto \quad \{(\sum A, \sum B) \mid (A, B) \in U\}, \quad U \in \mathcal{A}(\mathcal{A}M_1 \times \mathcal{A}M_2) \\ V \quad &\mapsto \quad \{(\{a\}, \{b\}) \mid (a, b) \in V\}/_{\equiv}, \quad V \in \mathcal{A}(M_1 \times M_2). \end{aligned}$$

Algebraic CST for monoids

We claim that the $\mathcal{R}$-dioid $\mathcal{R}M \otimes_{\mathcal{R}} C_2'$ contains a copy of the $\mathcal{C}$-dioid $\mathcal{C}M$ as a subdioid.

In $\mathcal{R}M \otimes_{\mathcal{R}} C_2'$, images of $m \in M$ and $c \in C_2'$ commute:

$$(\{m\} \otimes 1)(\{1\} \otimes c) = \{m\} \otimes c = (\{1\} \otimes c)(\{m\} \otimes 1).$$

Let $Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ be the centralizer of C_2' in $\mathcal{R}M \otimes_{\mathcal{R}} C_2'$, i.e.

$$\{e \in \mathcal{R}M \otimes_{\mathcal{R}} C_2' \mid e(1 \otimes c) = (1 \otimes c)e \text{ for all } c \in C_2'\}.$$

We want to prove that this is the algebraic closure $\mathcal{C}M$ of $\mathcal{F}M$.

Claim: $Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ is an $\mathcal{R}$-dioid.

Proof: Each $\mathcal{V} \in \mathcal{R}(Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')) \subseteq \mathcal{R}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ has a $\sum \mathcal{V} \in \mathcal{R}M \otimes_{\mathcal{R}} C_2'$, which commutes with C_2' by $\mathcal{R}$-distributivity.

Theorem (algebraic CST for monoids)

For every monoid M, $\mathcal{C}M \precnsim Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ via

$$L \;\mapsto\; \hat{L} := \sum \{\{m\} \otimes 1 \mid m \in L\}.$$

Analogy: for $D := Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$,

$$L = \bigcup \underbrace{\{\{m\} \mid m \in L\}}_{\in \mathcal{C}\mathcal{C}M} \in \mathcal{C}M \quad \mapsto \quad \sum \underbrace{\{\{m\} \otimes 1 \mid m \in L\}}_{\in \mathcal{C}D} \in D.$$

$$\bigcup : \mathcal{C}\mathcal{C}M \to \mathcal{C}M \quad \mapsto \quad \sum : \mathcal{C}D \to D$$

Proof

Suppose $L \in \mathcal{C}M$. By the Corollary, there is $n \geq 2$ and $R \in \mathcal{R}(M \times \Delta_n^*)$ such that

$$L = \{m \mid (m, d) \in R, d \in D_n\}.$$

The isomorphism $\mathcal{R}(M \times \Delta_n^*) \simeq \mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}\Delta_n^*$ maps R to

$$\sum\{\{m\} \otimes \{d\} \mid (m, d) \in R\} \in \mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}\Delta_n^*.$$

By adding new brackets p, q, going from R to

$$\{(1, p)\} \cdot R \cdot \{(1, q)\} \in \mathcal{R}(M \times \Delta_{n+1}^*),$$

and recoding $n + 1$ bracket pairs by n pairs, we can assume that in $(m, d) \in R$, $d = pd'q$ where p, q do not occur in d', so that

$$d \in D_n \iff \{d\}/\rho_n = 1, \qquad d \notin D_n \iff \{d\}/\rho_n = 0.$$

The homomorphism $(1_{\mathcal{R}M}, \cdot/\rho_n) : \mathcal{R}M \times \mathcal{R}\Delta_n \to \mathcal{R}M \times C'_n$ gives

$$S := \{(\{m\}, \{d\}/\rho_n) \mid (m, d) \in R\} \in \mathcal{R}(\mathcal{R}M \times C'_n),$$

and therefore, since $\{m\} \otimes d/\rho_n = \{m\} \otimes 0 = 0$ for $d \notin D_n$,

$$\begin{aligned}\hat{L} &= \sum\{\{m\} \otimes 1 \mid m \in L\} \\ &= \sum\{\{m\} \otimes \{d\}/\rho_n \mid (m, d) \in R, d \in D_n\} \\ &= \sum\{\{m\} \otimes \{d\}/\rho_n \mid (m, d) \in R\} = [S] \in \mathcal{R}M \otimes_{\mathcal{R}} C'_n.\end{aligned}$$

Claim: $\hat{L} \in Z_{C'_2}(\mathcal{R}M \otimes_{\mathcal{R}} C'_2)$

$\hat{L}$ is a $\sum$ of a set $S^{\otimes} \in \mathcal{R}(\mathcal{R}M \otimes_{\mathcal{R}} C'_2)$ of elements $\{m\} \otimes 1$ and $\{m\} \otimes 0$ that commute with C'_2. As $\mathcal{R}M \otimes_{\mathcal{R}} C'_2$ is $\mathcal{R}$-distributive,

$$c'(\sum S^{\otimes}) = \sum(c' S^{\otimes}) = \sum(S^{\otimes} c') = (\sum S^{\otimes})c'$$

for each $c' = \{1\} \otimes c$ with $c \in C'_2$.

Claim: $L \mapsto \hat{L}$ is an embedding of $\mathcal{C}M$ into $Z_{C_n'}(\mathcal{R}M \otimes_{\mathcal{R}} C_n')$.

Proof: We show there are relatively commuting $\mathcal{R}$-morphisms f, g

$$
\begin{array}{ccccc}
\mathcal{R}M & \xrightarrow{\top_1} & Z_{C_n'}(\mathcal{R}M \otimes_{\mathcal{R}} C_n') & \xleftarrow{\top_2} & C_n' \\
 & \searrow^{f} & \downarrow h_{f,g} & \swarrow_{g} & \\
 & & \mathcal{P}M \otimes_{\mathcal{P}} Mat_{P_n^*,P_n^*}(\mathbb{B}) & & \\
 & & \simeq Mat_{P_n^*,P_n^*}(\mathcal{P}M) & &
\end{array}
$$

such that the induced $\mathcal{R}$-morphism $h_{f,g}$ maps $\hat{L}$ to a copy of L.

Let f be defined by $f(A) = A \otimes I$ for $A \in \mathcal{R}M$ and unit matrix I.

For g, let $h : \Delta_n^* \to Mat_{P_n^*,P_n^*}(\mathbb{B})$ map p_i, q_j to the transition relations $\xrightarrow{p_i}$, $\xrightarrow{q_j}$ on P_n', restricted to $P_n^* \times P_n^*$. This hom. h extends to an $\mathcal{R}$-morphism $h^* : \mathcal{R}\Delta_n^* \to Mat_{P_n^*,P_n^*}(\mathbb{B})$ by

$$h^*(U) = \sum \{h(u) \mid u \in U\}.$$

The semiring equations $p_i q_j = \delta_{i,j}$ hold in $Mat_{P_n^*,P_n^*}(\mathbb{B})$ under this interpretation, so h^* is constant on ρ_n-congruence classes, and

$$g(U/\rho_n) = \{1\} \otimes h^*(U), \quad \text{for } U \in \mathcal{R}\Delta_n^*,$$

defines an $\mathcal{R}$-morphism $g : C_n' \to \mathcal{P}M \otimes_{\mathcal{P}} Mat_{P_n^*,P_n^*}(\mathbb{B})$.

Obviously, f and g are relatively commuting.

For $d \in D_n$, $g(\{d\}/\rho_n) = I$, and for $d \notin D_n$, $g(\{d\}/\rho_n) = 0$.

By the choice of R and since $h_{f,g}$ is an $\mathcal{R}$-morphism,

$$\begin{aligned}
h_{f,g}(\hat{L}) &= h_{f,g}(\sum\{\{m\} \otimes 1 \mid m \in L\}) \\
&= h_{f,g}(\sum\{\{m\} \otimes \{d\}/\rho_n \mid (m,d) \in R\}) \\
&= \sum\{h_{f,g}(\{m\} \otimes \{d\}/\rho_n) \mid (m,d) \in R\} \\
&= \sum\{f(\{m\}) \cdot g(\{d\}/\rho_n) \mid (m,d) \in R\} \\
&= \sum\{\{m\} \otimes I \mid m \in L\} = L \otimes_{\mathcal{P}} I.
\end{aligned}$$

Thus, $h_{f,g}$ is essentially an inverse to $L \mapsto \hat{L}$. □

Algebraic reverse CST for monoids

Theorem (algebraic RCST for monoids)

For every monoid M, $Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2') \subsetneq \mathcal{C}M$.

Recall that $C_n' = \mathcal{R}\Delta_n^*/\rho_n$ has a representation by $\mathcal{R}$-sets of ρ_n-reduced strings,

$$C_n' \simeq \mathcal{R}P_n'/(\{0\} = \emptyset).$$

For simplicity, we write

$$C_n' = \mathcal{R}'P_n' = \{B \setminus \{0\} \mid B \in \mathcal{R}P_n'\}.$$

Write $[S]$ for the congruence class $S/_{\equiv} \in \mathcal{R}M \otimes_{\mathcal{R}} C_2'$ of S.

Lemma

$$Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2') =$$

$$\{[S] \mid S \in \mathcal{R}(\mathcal{R}M \times \mathcal{R}'P_2'), S \subseteq \mathcal{R}M \times \{\emptyset, \{1\}\}\}.$$

Proof $\supseteq$: For $c \in C_2'$ and $B \in \{\emptyset, \{1\}\}$, $cB = Bc$, so

$$(1 \otimes c)(A \otimes B) = (A \otimes B)(1 \otimes c).$$

Since $[S] = \sum\{A \otimes B \mid (A, B) \in S\}$, by $\mathcal{R}$-distributivity we get

$$(1 \otimes c)[S] = [S](1 \otimes c).$$

$\subseteq$: Suppose $[R] \in Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ for $R \in \mathcal{R}(\mathcal{R}M \times C_2')$. Add a new pair p, q of brackets and recode C_3' in C_2'. By choice of R,

$$(\{1\} \otimes \{p\})[R](\{1\} \otimes \{q\}) = [R].$$

Since $\mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}'P_2'$ is $\mathcal{R}$-distributive, we have

$$\begin{aligned} [R] &= (\{1\} \otimes \{p\})[R](\{1\} \otimes \{q\}) \\ &= (\{1\} \otimes \{p\})(\sum\{A \otimes B \mid (A, B) \in R\})(\{1\} \otimes \{q\}) \\ &= \sum\{A \otimes \{p\}B\{q\}) \mid (A, B) \in R\}) = [S], \end{aligned}$$

where $S \in \mathcal{R}(\mathcal{R}M \times \mathcal{R}'P_2')$ is

$$\{(A, \{p\}B\{q\}) \mid (A, B) \in R\} = \{(\{1\}, \{p\})\}R\{(\{1\}, \{q\})\}.$$

As $B \subseteq Q_2^* P_2^*$ and p, q do not occur in B, $pBq \subseteq \{\emptyset, \{1\}\}$. $\square$

Proof of the RCST for monoids M

Take $[R] \in Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ with $R \in \mathcal{R}(\mathcal{R}M \times \mathcal{R}' P_2')$. By the lemma, we can assume $R \subseteq \mathcal{R}M \times \{\emptyset, \{1\}\}$. Put

$$L_R := \bigcup \{A \mid (A, B) \in R, 1 \in B\} \subseteq M.$$

Then

$$\begin{aligned} [R] &= \sum \{A \otimes B \mid (A, B) \in R\} \\ &= \sum \{\{w\} \otimes \{v\} \mid (A, B) \in R, w \in A, v \in B\} \\ &= \sum \{\{w\} \otimes \{1\} \mid w \in L_R\}. \end{aligned}$$

It remains to show $L_R \in \mathcal{C}M$. (Notice that then $[R] = \widehat{L_R}$.)

By $\mathcal{R}M \otimes_{\mathcal{R}} \mathcal{R}P_2' \simeq \mathcal{R}(M \times P_2')$, we assume $R \in \mathcal{R}(M \times P_2')$ and

$$L_R = \{m \mid (m, 1) \in R\}$$

to show $L_R \in \mathcal{C}M$. First, "undo" the separation $M[\Delta_2] \to M \times \Delta_2^*$.

For $R \in \mathcal{R}(M \times P_2')$, define $R' \in \mathcal{R}(M[\Delta_2])$ inductively by:

- $\emptyset' = \emptyset = \{(m, 0)\}'$, for $m \in M$,
- $\{(m, \bar{q}\bar{p})\}' = \{\bar{q}m\bar{p}\}$, for $m \in M$, $\bar{q} \in Q_2^*, \bar{p} \in P_2^*$,
- $(R \cup S)' = R' \cup S'$,
- $(R_1 \cdot R_2)' = R_1' \cdot R_2'$
- $(R^*)' = (R')^*$

Claim. $L_R = h_M(R' \cap D_2(X))$ for the erasure $h_M : M[\Delta_2] \to M$ and a finite $X \subseteq M$. (For $M = X^*$, it follows that $L_R \in \mathcal{C}M$.)

Consider $(m, \bar{q}\bar{p}) \in R$ as a relation on P'_2 generating output in M:

from $u \in P'_2$, go to $u \cdot' \bar{q}\bar{p} \in P'_2$ and output m.

For $R \in \mathcal{R}(M \times P'_2)$ and $R' \in \mathcal{R}(M[\Delta_2])$, define ternary relations $\Longrightarrow_R$, $\longrightarrow_{R'} \subseteq P'_2 \times P'_2 \times M$ by

- $u \overset{m'}{\Longrightarrow}_R v :\iff \exists (m, \bar{q}\bar{p}) \in R[u \cdot' \bar{q}\bar{p} = v \wedge m' = m]$
- $u \overset{m'}{\longrightarrow}_{R'} v :\iff \exists \alpha \in R'[nf(h_\Delta(u\alpha)) = v \wedge m' = h_M(\alpha)]$

where $nf : \Delta_2^* \to P'_n = Q_2^* P_2^* \cup \{0\}$ is the ρ_2-normal form and $h_\Delta : M[\Delta_2] \to \Delta_2^*$ is the homomorphism erasing elements of M.

Claim. For $R \in \mathcal{R}(M \times P'_2)$, $\longrightarrow_{R'} = \Longrightarrow_R$.

Proof: by induction on the construction of R.

$R = \emptyset = R'$: there are no m', u, v with $u \xrightarrow{m'}_{R'} v$ or $u \xRightarrow{m'}_{R} v$.

$R = \{(m, \bar{q}\bar{p})\}$: Then $R' = \{\bar{q}m\bar{p}\}$, and $h_M(\bar{q}m\bar{p}) = m$, $nf(h_\Delta(u\bar{q}m\bar{p})) = u \cdot \bar{q}\bar{p}$, so $\xrightarrow{m'}_{R'} = \emptyset = \xRightarrow{m'}_{R}$ for $m' \neq m$ and

$$\xrightarrow{m}_{R'} = \{(u, u \cdot \bar{q}\bar{p}) \mid u \in P_2'\} = \xRightarrow{m}_{R}.$$

$R = R_1 \cup R_2$: Then $R' = R_1' \cup R_2'$, the claim follows by induction.

$R = S^* = \bigcup\{S^n \mid n \in \mathbb{N}\}$: Then $R' = (S')^* = \bigcup\{(S')^n \mid n \in \mathbb{N}\}$. By induction, $\longrightarrow_{S'} = \Longrightarrow_{S}$, and $\longrightarrow_{(S^n)'} = \longrightarrow_{(S')^n} = \Longrightarrow_{S^n}$ by induction from the product case, so

$$\Longrightarrow_{R} = \bigcup\{\Longrightarrow_{S^n} \mid n \in \mathbb{N}\} = \bigcup\{\longrightarrow_{(S^n)'} \mid n \in \mathbb{N}\} = \xrightarrow{R'}.$$

$R = R_1 \cdot R_2$:

$$
\begin{aligned}
& u \xrightarrow{m'}_{R_1' R_2'} v \\
\iff \quad & \exists \alpha_1 \in R_1' \exists \alpha_2 \in R_2' \\
& [nf(h_\Delta(u\alpha_1\alpha_2)) = v \wedge m' = h_M(\alpha_1\alpha_2)] \\
\iff \quad & \exists m_1, m_2 \in M, w \in P_2' \\
& [\exists \alpha_1 \in R_1'(nf(h_\Delta(u\alpha_1)) = w \wedge m_1 = h_M(\alpha_1)) \wedge \\
& \ \exists \alpha_2 \in R_2'(nf(h_\Delta(w\alpha_2)) = v \wedge m_2 = h_M(\alpha_2)) \wedge m' = m_1 m_2] \\
\iff \quad & \exists m_1, m_2 \in M, w \in P_2' \, [u \xrightarrow{m_1}_{R_1'} w \wedge w \xrightarrow{m_2}_{R_2'} v \wedge m' = m_1 m_2] \\
\iff \quad & \exists m_1, m_2 \in M, w \in P_2' \, [u \overset{m_1}{\Longrightarrow}_{R_1} w \wedge w \overset{m_2}{\Longrightarrow}_{R_2} v \wedge m' = m_1 m_2] \\
\iff \quad & \exists (m_1, \bar{q}_1\bar{p}_1) \in R_1, (m_1, \bar{q}_2\bar{p}_2) \in R_2 \\
& [u \cdot' (\bar{q}_1\bar{p}_1 \cdot' \bar{q}_2\bar{p}_2) = v \wedge m' = m_1 m_2] \\
\iff \quad & u \overset{m'}{\Longrightarrow}_{R_1 R_2} v
\end{aligned}
$$

Now let X be the finitely many members of M in the elements of $\mathcal{F}(M \times P_2')$ used in the construction of R (and hence R'). Then

$$\begin{aligned} 1 \xrightarrow{m}_{R'} 1 &\iff \exists \alpha \in R'(nf(h_\Delta(1\alpha)) = 1 \wedge m = h_M(\alpha)) \\ &\iff \exists \alpha \in R'(\alpha \in D_n(X) \wedge m = h_M(\alpha)) \\ &\iff m \in h_M(R' \cap D_n(X)). \end{aligned}$$

By the previous Claim, the claim on L_R follows:

$$\begin{aligned} L_R &= \{m \in M \mid (m, 1) \in R\} \\ &= \{m \mid (m, \bar{q}\bar{p}) \in R, \bar{q}\bar{p} = 1\} \quad (\Leftarrow\text{: Cayley-graph}) \\ &= \{m \mid \exists (m', \bar{q}\bar{p}) \in R\, [1 \cdot' \bar{q}\bar{p} = 1 \wedge m = m']\} \\ &= \{m \mid 1 \overset{m}{\Longrightarrow}_R 1\} = \{m \mid 1 \xrightarrow{m}_{R'} 1\} \\ &= h_M(R' \cap D_n(X)). \end{aligned}$$

This proves the RCST *if* $R' \cap D_n(X) \in \mathcal{C}(M[\Delta_n])$, as for $M = X^*$.

For arbitrary monoid M', one can't use automata to show

$$R' \in \mathcal{R}(M'), L' \in \mathcal{C}(M') \Rightarrow R' \cap L' \in \mathcal{C}(M'),$$

as "inputs" $m \in M'$ need not compose into factors from $R' \cap L'$.

Claim. For finitely generated monoid M,

$$R \in \mathcal{R}(M[\Delta_2]) \Rightarrow R \cap D_2(M) \in \mathcal{C}(M[\Delta_2]).$$

Proof:
R is a component of the least solution $\bar{R}$ of a right-linear system

$$\begin{aligned} y_1 &\geq (w_{1,1} + \ldots)y_1 + \ldots + (w_{1,m} + \ldots)y_m + (w_1 + \ldots) \\ &\vdots \\ y_m &\geq (w_{m,1} + \ldots)y_1 + \ldots + (w_{m,m} + \ldots)y_m + (w_m + \ldots) \end{aligned}$$

over $M[\Delta_2]$, where $w_{i,j}, w_i \in M[\Delta_2]$. Assume $w_{i,j} \in M \cup \Delta_2$ and $w_i \in \{0, 1\}$ (by splitting $y \geq mbz$ into $y \geq my', y' \geq bz$ etc.).

We can assume that M is generated by the finite set $X \subseteq M$ of elements of M occurring in this right-linear system, and use X^*.

Variables: $[y, Z, z]$, with $y, z \in \{y_1, \ldots, y_m\}$, $Z \in M \cup \Delta_2 \cup \{D\}$

context-free G for $R \cap D_2(M)$			*definition of R contains*
S	$\geq$	$[y, D, z]$	y as main variable
$[y, D, z]$	$\geq$	m	$y \geq mz$ with $m \in X$
$[y, d, z]$	$\geq$	d	$y \geq dz$ with $d \in \Delta_2$
$[y, D, 1]$	$\geq$	1	$y \geq 1$
$[y, D, y]$	$\geq$	1	
$[y, D, z]$	$\geq$	$[y, D, y'][y', D, z]$	
$[y, D, z]$	$\geq$	$[y, p_i, y'][y', D, z'][z', q_i, z]$	

Claim: if $[y, D, z] \Rightarrow_G^k w \in M[\Delta_2]$, then $y \Rightarrow_R^* wz$, $w \in D_2(X)$

by induction on k. So, if $S \Rightarrow_G^* w \in M[\Delta_2]$, then $w \in R \cap D_2(X)$.

Claim: if $w \in R \cap D_2(X)$, then $w \in L(G)$.

Proof: Suppose $y \Rightarrow_R^k wz$ and $w \in D_2(X)$. There is a sequence

$$y_{i_0} \geq w_1 y_{i_1}, \quad y_{i_1} \geq w_2 y_{i_2}, \quad \ldots, \quad y_{i_{k-1}} \geq w_k y_{i_k}$$

s.th. $y = y_{i_0}$, $w = w_1 \cdots w_k \in D_2(X)$ and $y_{i_k} = z$. We show

$$[y_{i_0}, D, y_{i_k}] \Rightarrow_G^* w_1 \cdots w_k$$

by induction on the construction of $w \in X^*$. If y in $y \Rightarrow_R^k wz$ is the main variable in the definition of R, it follows that

$$S \Rightarrow_G [y, D, z] \Rightarrow_G^* w, \text{ so } w \in L(G).$$

If $k = 0$, then $y_{i_0} \Rightarrow_R^0 wy_{i_0}$, so $w = 1$, and $[y_{i_0}, D, y_{i_0}] \Rightarrow_G 1$.

If $k = 1$, then $y_{i_0} \geq wy_{i_1}$ and $w \in R \cap D_2(X)$, hence $y_{i_1} \geq 1$ and $w \in X$, hence $[y_{i_0}, D, y_{i_1}] \Rightarrow_G w$ (and $[y_{i_1}, D, 1] \Rightarrow_G 1$).

Suppose $k > 1$. By the construction of $D_2(X)$, for $w_1 \cdots w_k \in X^*$ there either is $j < k$ with $u := w_1 \cdots w_j, v := w_{j+1} \cdots w_k \in D_2(X)$, or $u := w_2 \cdots w_{k-1} \in D_2(X)$ and $w = p_0 u q_0$ or $w = p_1 u q_1$.

In the first case, we have $[y_{i_0}, D, y_{i_j}] \Rightarrow_G^* u$ and $[y_{i_{j+1}}, D, y_{i_k}] \Rightarrow_G^* v$ by induction, hence

$$[y_{i_0}, D, y_{i_k}] \Rightarrow_G [y_{i_0}, D, y_{i_j}][y_{i_{j+1}}, D, y_{i_k}] \Rightarrow_G^* uv = w.$$

In the second case, we have $[y_{i_1}, D, y_{i_{k-1}}] \Rightarrow_G^* u$ by induction, hence

$$\begin{aligned} [y_{i_0}, D, y_{i_k}] &\Rightarrow_G [y_{i_0}, p_j, y_{i_1}][y_{i_1}, D, y_{i_{k-1}}][y_{i_{k-1}}, q_j, y_{i_k}] \\ &\Rightarrow_G^2 p_j[y_{i_1}, D, y_{i_{k-1}}]q_j \Rightarrow_G^* p_j u q_j = w. \end{aligned}$$

Conclusion

1. For monoids M, the fixed-point-closure of $\mathcal{R}M$ or $\mathcal{F}M$ is an algebraic function of $\mathcal{R}M$:
$$Q^{\mathcal{C}}_{\mathcal{R}}(\mathcal{R}M) = \mathcal{C}M = Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2').$$
2. All cf-languages $L \in \mathcal{C}X^*$ are values in $Z_{C_2'}(\mathcal{R}X^* \otimes_{\mathcal{R}} C_2')$ of (certain) *regular* expressions over $X \cup \Delta_2$.
3. For $Q^{\mathcal{C}}_{\mathcal{R}} : \mathbb{D}\mathcal{R} \to \mathbb{D}\mathcal{C}$ one can probably show
$$Q^{\mathcal{C}}_{\mathcal{R}} D = Z_{C_2'}(D \otimes_{\mathcal{R}} C_2')$$
by considering quotients $D = \mathcal{R}M/\rho$
$$\begin{aligned} Q^{\mathcal{C}}_{\mathcal{R}}(D) &= Q^{\mathcal{C}}_{\mathcal{R}}(\mathcal{R}M/\rho) =^? Q^{\mathcal{C}}_{\mathcal{R}}(\mathcal{R}M)/\rho' \\ &= (Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2'))/\rho' =^? Z_{C_2'}(\mathcal{R}M/\rho \otimes_{\mathcal{R}} C_2') \\ &= Z_{C_2'}(D \otimes_{\mathcal{R}} C_2') \end{aligned}$$

Remark

Hopkins proves an RCST $Z_{C_2}(\mathcal{R}M \otimes_{\mathcal{R}} C_2) \subseteq \mathcal{C}M$ where

$$C_n := \mathcal{R}\Delta_n^*/(\rho_n \cup \{1 = \sum\{q_i p_i \mid i < n\}\}).$$

But the additional relation $1 = \sum\{q_i p_i \mid i < n\}$ is not valid in a stack: at the empty stack 1, we cannot perform any $q_i p_i$.

To get a stack, we might add an emptyness test e, i.e. use

$$1 = e + \sum\{q_i p_i \mid i < n\}, \quad ee = e, \quad p_i e = 0 = e q_j$$

But these additional relations

- are not inherited from *monoid* equations,
- make understanding C_2 harder than understanding C_2'
- do not fit to the CST $\mathcal{C}M \subseteq Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2')$ with C_2'
- don't(?) give $\mathcal{C}M \subseteq Z_{C_2'}(\mathcal{R}M \otimes_{\mathcal{R}} C_2') \subseteq^? Z_{C_2}(\mathcal{R}M \otimes_{\mathcal{R}} C_2)$.

Questions and References

Questions:

- Closure under $\mathcal{A}$-transductions: if $R \in \mathcal{A}(M_1 \times M_2)$ and $A \in \mathcal{A}M_1$, is $R(A) := \{b \in M_2 \mid (a, b) \in R, a \in A\} \in \mathcal{A}M_2$?
- Closure under matrix ring formation: if $D \in D\mathcal{A}$, is $D^{n \times n} \in \mathbb{D}\mathcal{A}$? Then: $D^{n \times n} \simeq D \otimes_{\mathcal{A}} \mathbb{B}^{n \times n}$. ($\mathcal{F}, \mathcal{R}, \mathcal{C}, \mathcal{P}$: yes)
- How much follows from results in category theory?

References:

Mark Hopkins.
The Algebraic Approach I+II: The Algebraization of the Chomsky Hierarchy + Dioids, Quantales and Monads.
In *Relational Methods in Computer Science/Applications of Kleene Algebra*, Springer LNCS 4988, pages 155–190, 2008.

Mark Hopkins and Hans Leiß.
Coequalizers and tensor products for continuous idempotent semirings.
In *Proc. RAMiCS'17*, pp 37–52, Springer LNCS 11194, 2018.

Niels Grathwohl, Fritz Henglein, and Dexter Kozen.
Infinitary axiomatization of the equational theory of context-free languages.
In *Fixed Points in Computer Science (FICS 2013)*, vol. 126 of *EPTCS*, pp 44–55, 2013.

Hans Leiß and Mark Hopkins.
C-dioids and μ-continuous Chomsky-algebras.
In *Proc. RAMiCS'17*, pp 21–36, Springer LNCS 11194, 2018.