Equivalence of Deterministic Tree-to-String Transducers Is Decidable

Helmut Seidl, Sebastian Maneth, Gregor Kemper

TU München, U. of Edinburgh

GT-ALGA, April 12, 2016





Overview

- Part 1: The General Setting
- Part 2: Tree-to-Int Transducers
- Part 3: Affine Spaces
- Part 4: Polynomial Invariants





Overview

Part 1: The General Setting

Part 2: Tree-to-Int Transducers

Part 3: Affine Spaces

Part 4: Polynomial Invariants





Input





Output

</frame>

. . .





Output

<frame height=20 width=50> <button>Do not press!</button> ... </frame>

Realized by:

. . .



Output

<frame height=20 width=50> <button>Do not press!</button> ... </frame>

Or realized by:



These two translations are equivalent.





These two translations are equivalent.

Unstructured output, though, can be generated in surprisingly different ways ...

$$egin{array}{rll} q(f(x_1,x_2,x_3)) & o & q(x_3) \ a \ q_1(x_2) \ b \ q(x_2) \ q_1(f(x_1,x_2,x_3)) & o & q_1(x_3) q_1(x_2) q_1(x_2) \ ba \ q_1(e) & o & ba \ q(e) & o & ab \end{array}$$





These two translations are equivalent.

Unstructured output, though, can be generated in surprisingly different ways ...

$$egin{array}{rll} q(f(x_1,x_2,x_3)) & o & q(x_3) \ a \ q_1(x_2) \ b \ q(x_2) \ q_1(f(x_1,x_2,x_3)) & o & q_1(x_3) q_1(x_2) q_1(x_2) \ ba \ q_1(e) & o & ba \ q(e) & o & ab \end{array}$$

versus

$$egin{array}{rll} q'(f(x_1,x_2,x_3)) &
ightarrow & ab \ q'(x_2)q'(x_2)q'(x_3) \ q'(e) &
ightarrow & ab \end{array}$$





Related Work

problem statement

Engelfriet, 1980







Related Work (cont.)

with monadic input

Culik II, Karhumäki, 1986 Ruohonen, 1986 Honkala, 2000





Related Work (cont.)

with monadic input

MSO-definable sequential

Culik II, Karhumäki, 1986 Ruohonen, 1986 Honkala, 2000 Engelfriet, Maneth, 2006

Staworko et al., 2009





Related Work (cont.)

with monadic input

MSO-definable sequential

polynomial program invariants

Culik II, Karhumäki, 1986 Ruohonen, 1986 Honkala, 2000

Engelfriet, Maneth, 2006

Staworko et al., 2009

Letichevsky, Lvov, 1996 Müller-Olm, S., 2004







Obvious:

In-equivalence can be verified by counter example





General Idea

Obvious:

In-equivalence can be verified by counter example

Required:

Complete proof system for equivalence





Overview

Part 1: The General Setting

Part 1: Tree-to-Int Transducers

Part 3: Affine Spaces

Part 2: Polynomial Invariants





Simplification

- A single transducer with states $Q = \{1, \ldots, n\}$.
- The transducer is total.





Simplification

- A single transducer with states $Q = \{1, \ldots, n\}$.
- The transducer is total.
- There is a topdown-deterministic automaton *B* with states *p* ∈ *P*

dom(p) is the set of trees accepted at state p





Simplification

- A single transducer with states $Q = \{1, \ldots, n\}$.
- The transducer is total.
- There is a topdown-deterministic automaton *B* with states *p* ∈ *P*

// generalization of algebraic data type
dom(p) is the set of trees accepted at state p

// trees of type p





Topdown Automaton







Topdown Automaton







Simplified Question

For states q, q' of the transducer, $p_0 \in P$, does it hold that

$$[\![q]\!](t) = [\![q']\!](t) \qquad (t \in \mathsf{dom}(p_0))$$





From Arbitrary Output to Ints Unary Output

 $q(f(x_1, x_2)) \rightarrow dd q_1(x_1) d q_1(x_1) q_2(x_2)$





From Arbitrary Output to Ints Unary Output

 $q(f(x_1, x_2)) \rightarrow dd q_1(x_1) d q_1(x_1) q_2(x_2)$

Succinct representation:

Tree-to-int transducer

$$q(f(x_1, x_2)) \rightarrow 3 + 2 \cdot q_1(x_1) + q_2(x_2)$$





From Arbitrary Output to Ints Unary Output

 $q(f(x_1, x_2)) \rightarrow dd q_1(x_1) d q_1(x_1) q_2(x_2)$

Succinct representation:

Tree-to-int transducer

$$q(f(x_1, x_2)) \rightarrow 3 + 2 \cdot q_1(x_1) + q_2(x_2)$$

Arbitary Output

letters
$$\underline{a}, \underline{b}, \underline{c}, \dots \triangleq \text{ digits } 1, \dots, \underline{h-1}$$

string $\underline{aabc} \triangleq 1 + \underline{h} \cdot (1 + \underline{h} \cdot (2 + \underline{h} \cdot 3))$



Wanted

Transformation of tree-to-string into tree-to-int ...





Wanted

Transformation of tree-to-string into tree-to-int ...

$$q(f(x_1, x_2)) \rightarrow a q_1(x_1) b q_2(x_2)$$

is simulated by:

 $q(f(x_1, x_2)) \rightarrow$





Wanted

Transformation of tree-to-string into tree-to-int ...

$$q(f(x_1, x_2)) \rightarrow a q_1(x_1) b q_2(x_2)$$

is simulated by:

 $q(f(x_1, x_2), \mathbf{y}) \rightarrow$



Wanted

Transformation of tree-to-string into tree-to-int ...

$$q(f(x_1, x_2)) \rightarrow a q_1(x_1) b q_2(x_2)$$

is simulated by:

 $q(f(x_1, x_2), y) \rightarrow 1 + 3 \cdot$



Wanted

Transformation of tree-to-string into tree-to-int ...

$$q(f(x_1, x_2)) \rightarrow a q_1(x_1) b q_2(x_2)$$

is simulated by:

 $q(f(x_1,x_2),\mathbf{y}) \rightarrow \mathbf{1} + \mathbf{3} \cdot q_1(x_1,$



Wanted

Transformation of tree-to-string into tree-to-int ...

$$q(f(x_1, x_2)) \rightarrow a q_1(x_1) b q_2(x_2)$$

is simulated by:

$$q(f(x_1, x_2), y) \rightarrow 1 + 3 \cdot q_1(x_1, 2 + 3 \cdot q_2(x_2, y))$$



Overview

Part 1: The General Setting

Part 2: Tree-to-Int Transducers

Part 3: Affine Spaces

Part 4: Polynomial Ideals





Equivalence of Tree-to-Int Transducers Idea

• The semantics of a tree t can be seen as

$$\llbracket t \rrbracket = (\llbracket 1 \rrbracket(t), \ldots, \llbracket n \rrbracket(t)) \in \mathbb{Q}^n$$





Equivalence of Tree-to-Int Transducers Idea

• The semantics of a tree *t* can be seen as

$$\llbracket t \rrbracket = (\llbracket 1 \rrbracket(t), \ldots, \llbracket n \rrbracket(t)) \in \mathbb{Q}^n$$

• For state p of B, let $V_p = \{ \llbracket t \rrbracket \mid t \in \operatorname{dom}(p) \}$.





Equivalence of Tree-to-Int Transducers Idea

• The semantics of a tree *t* can be seen as

$$\llbracket t \rrbracket = (\llbracket 1 \rrbracket(t), \dots, \llbracket n \rrbracket(t)) \in \mathbb{Q}^n$$

- For state p of B, let $V_p = \{ \llbracket t \rrbracket \mid t \in \operatorname{dom}(p) \}$.
- Consider $H(\mathbf{v}) = \mathbf{v}_q \mathbf{v}_{q'}$.
- The following statements are equivalent:

1. q, q' agree on inputs from $\mathcal{L}(B)$ 2. $H(\mathbf{v}) = \mathbf{0}$ $(\mathbf{v} \in V_{p_0})$




Equivalence of Tree-to-Int Transducers Idea

• The semantics of a tree *t* can be seen as

$$\llbracket t \rrbracket = (\llbracket 1 \rrbracket(t), \dots, \llbracket n \rrbracket(t)) \in \mathbb{Q}^n$$

• For state p of B, let $V_p = \{ \llbracket t \rrbracket \mid t \in \operatorname{dom}(p) \}$.

• Consider
$$H(\mathbf{v}) = \mathbf{v}_q - \mathbf{v}_{q'}$$
.

• The following statements are equivalent:

1.
$$q, q'$$
 agree on inputs from $\mathcal{L}(B)$
2. $H(\mathbf{v}) = \mathbf{0}$ $(\mathbf{v} \in V_{p_0})$
3. $H(\mathbf{v}) = \mathbf{0}$ $(\mathbf{v} \in \operatorname{Aff}(V_{p_0}))$
 $\#$ affine closure



$$\llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k) = (\llbracket T_1 \rrbracket (\mathbf{x}), \dots, \llbracket T_n \rrbracket (\mathbf{x})) \quad \text{where} \\ q(f(x_1, \dots, x_k)) \to T_q$$





$$\llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k) = (\llbracket T_1 \rrbracket (\mathbf{x}), \dots, \llbracket T_n \rrbracket (\mathbf{x})) \quad \text{where} \\ q(f(x_1, \dots, x_k)) \to T_q \quad \text{and}$$

$$[[3 \cdot q(x_1) + q'(x_1) + 2 \cdot q'(x_2) + 5]](\mathbf{x}_1, \mathbf{x}_2) =$$





$$\begin{split} \llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k) &= (\llbracket T_1 \rrbracket (\mathbf{x}), \dots, \llbracket T_n \rrbracket (\mathbf{x})) & \text{where} \\ q(f(x_1, \dots, x_k)) &\to T_q & \text{and} \\ \\ \llbracket \mathbf{3} \cdot q(x_1) + q'(x_1) + \mathbf{2} \cdot q'(x_2) + \mathbf{5} \rrbracket (\mathbf{x}_1, \mathbf{x}_2) &= \\ \mathbf{3} \cdot \mathbf{x}_{1q} + \mathbf{x}_{1q'} + \mathbf{2} \cdot \mathbf{x}_{2q'} + \mathbf{5} \end{split}$$





$$\begin{split} \llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k) &= (\llbracket T_1 \rrbracket (\mathbf{x}), \dots, \llbracket T_n \rrbracket (\mathbf{x})) & \text{where} \\ q(f(x_1, \dots, x_k)) &\to T_q & \text{and} \\ \\ \llbracket \mathbf{3} \cdot q(x_1) + q'(x_1) + \mathbf{2} \cdot q'(x_2) + \mathbf{5} \rrbracket (\mathbf{x}_1, \mathbf{x}_2) &= \\ \mathbf{3} \cdot \mathbf{x}_{1q} + \mathbf{x}_{1q'} + \mathbf{2} \cdot \mathbf{x}_{2q'} + \mathbf{5} \end{split}$$

$$\Rightarrow \qquad \llbracket f \rrbracket : \mathbb{Q}^n \times \ldots \times \mathbb{Q}^n \to \mathbb{Q}^n \text{ is affine.}$$





Computing Affine Closures (cont.) Consequence

 $V'_{\rho} = \operatorname{Aff}(V_{\rho})$ is the least solution of:

$$V'_p \supseteq \llbracket f \rrbracket (V'_{p_1}, \ldots, V'_{p_k})$$

 $((p, f) \mapsto p_1 \dots p_k$ transition of *B*) over the complete lattice of affine sub-spaces of \mathbb{Q}^n !





Computing Affine Closures (cont.) Consequence

 $V'_{\rho} = \operatorname{Aff}(V_{\rho})$ is the least solution of:

$$V'_p \supseteq \llbracket f \rrbracket (V'_{p_1}, \ldots, V'_{p_k})$$

 $((p, f) \mapsto p_1 \dots p_k$ transition of *B*) over the complete lattice of affine sub-spaces of \mathbb{Q}^n !

Theorem

• Equivalence of total tree-to-int transducers relative to some *B* is decidable in polynomial time.





Computing Affine Closures (cont.) Consequence

 $V'_{\rho} = \operatorname{Aff}(V_{\rho})$ is the least solution of:

$$V'_p \supseteq \llbracket f \rrbracket (V'_{p_1}, \ldots, V'_{p_k})$$

 $((p, f) \mapsto p_1 \dots p_k$ transition of *B*) over the complete lattice of affine sub-spaces of \mathbb{Q}^n !

Theorem

- Equivalence of total tree-to-int transducers relative to some *B* is decidable in polynomial time.
- In-Equivalence of linear tree-to-string transducers is decidable in randomized polynomial time.





Overview

- Part 1: The General Setting
- Part 2: Tree-to-Int Transducers
- Part 3: Affine Spaces
- Part 4: Polynomial Invariants





Tree-to-int Transducers with Parameters

$$\begin{array}{rcl} q(f(x_1, x_2), y) & \to & q_1(x_1, q_1(x_2, 1)) \\ q_1(a(x_1), y) & \to & y + q_1(x_1, y) \\ q_1(e, y) & \to & 0 \end{array}$$





Tree-to-int Transducers with Parameters

$$\begin{array}{rcl} q(f(x_1, x_2), y) & \to & q_1(x_1, q_1(x_2, 1)) \\ q_1(a(x_1), y) & \to & y + q_1(x_1, y) \\ q_1(e, y) & \to & \mathbf{0} \\ q'(f(x_1, x_2), y) & \to & q_1(x_2, q_1(x_1, 1)) \end{array}$$





Tree-to-int Transducers with Parameters

$$\begin{array}{rcl} q(f(x_1, x_2), y) & \to & q_1(x_1, q_1(x_2, 1)) \\ q_1(a(x_1), y) & \to & y + q_1(x_1, y) \\ q_1(e, y) & \to & 0 \\ q'(f(x_1, x_2), y) & \to & q_1(x_2, q_1(x_1, 1)) \end{array}$$

The semantics of a tree t is a vector

 $\llbracket t \rrbracket : (\mathbb{Q}' \to \mathbb{Q})^n$

of affine functions in the parameters

can be represented by a matrix $(\llbracket t \rrbracket_{q^i}) \in \mathbb{Q}^{n \times (l+1)}$





The Semantics of Constructors

$$\llbracket f \rrbracket : (\mathbb{Q}^{n \times (l+1)} \times \ldots \times \mathbb{Q}^{n \times (l+1)}) \to \mathbb{Q}^{n \times (l+1)}$$

thus is of the form:

$$(\llbracket f \rrbracket (\mathbf{x}_1, \ldots, \mathbf{x}_k))_{qj} = \text{polynomial in the } \mathbf{x}_{iq'j'}$$





In the Example

$$\begin{array}{rcl} q(f(x_1, x_2), y) & \to & q_1(x_1, q_1(x_2, 1)) \\ q_1(a(x_1), y) & \to & y + q_1(x_1, y) \\ q_1(e, y) & \to & 0 \\ q'(f(x_1, x_2), y) & \to & q_1(x_2, q_1(x_1, 1)) \end{array}$$





In the Example

$$\begin{array}{rcl} q(f(x_1, x_2), y) & \to & q_1(x_1, q_1(x_2, 1)) \\ q_1(a(x_1), y) & \to & y + q_1(x_1, y) \\ q_1(e, y) & \to & 0 \\ q'(f(x_1, x_2), y) & \to & q_1(x_2, q_1(x_1, 1)) \end{array}$$

$$(\llbracket f \rrbracket (\mathbf{x}_1, \mathbf{x}_2))_{q0} = \mathbf{x}_{1q_10} + \mathbf{x}_{1q_11} \cdot (\mathbf{x}_{2q_10} + \mathbf{x}_{2q_11} \cdot \mathbf{1}) \\ (\llbracket f \rrbracket (\mathbf{x}_1, \mathbf{x}_2))_{q'0} = \mathbf{x}_{2q_10} + \mathbf{x}_{2q_11} \cdot (\mathbf{x}_{1q_10} + \mathbf{x}_{1q_11} \cdot \mathbf{1})$$





Polynomial Invariant

polynomial equality:

$$\mathbf{z}_{q1} \cdot \mathbf{z}_{q'1} \cdot \mathbf{z}_{q'0} - \mathbf{2} \cdot \mathbf{z}_{q0} + \mathbf{3} \doteq \mathbf{0}$$





Polynomial Invariant

polynomial equality:

$$\mathbf{z}_{q1} \cdot \mathbf{z}_{q'1} \cdot \mathbf{z}_{q'0} - \mathbf{2} \cdot \mathbf{z}_{q0} + \mathbf{3} \doteq \mathbf{0}$$

 $r_1 \doteq 0 \land \ldots \land r_m \doteq 0$ invariant at p iff

$$r_1(\llbracket t \rrbracket) = \ldots = r_m(\llbracket t \rrbracket) = 0$$
 $(t \in \operatorname{dom}(p))$





Polynomial Invariant

polynomial equality:

$$\mathbf{z}_{q1} \cdot \mathbf{z}_{q'1} \cdot \mathbf{z}_{q'0} - \mathbf{2} \cdot \mathbf{z}_{q0} + \mathbf{3} \doteq \mathbf{0}$$

 $r_1 \doteq 0 \land \ldots \land r_m \doteq 0$ invariant at p iff

$$r_1(\llbracket t \rrbracket) = \ldots = r_m(\llbracket t \rrbracket) = \mathbf{0} \qquad (t \in \mathbf{dom}(p))$$

can be described by polynomial ideals ...





Polynomial Ideals: A Primer

R ring. $I \subseteq R$ ideal, if

- $a+b \in I$ whenever $a, b \in I$;
- $r \cdot a \in I$ whenever $a \in I$ and $r \in R$.





Polynomial Ideals: A Primer

R ring. $I \subseteq R$ ideal, if

- $a+b \in I$ whenever $a, b \in I$;
- $r \cdot a \in I$ whenever $a \in I$ and $r \in R$.

I is finitely generated, if

$$I = \langle a_1, \ldots, a_s \rangle_R = \{\sum_{i=1}^s r_i \cdot a_i \mid r_i \in R\}$$





Polynomial Ideals: A Primer

R ring. $I \subseteq R$ ideal, if

- $a+b \in I$ whenever $a, b \in I$;
- $r \cdot a \in I$ whenever $a \in I$ and $r \in R$.

I is finitely generated, if

$$I = \langle a_1, \ldots, a_s \rangle_R = \{\sum_{i=1}^s r_i \cdot a_i \mid r_i \in R\}$$

$\mathbf{R} = \mathbb{Q}[\mathbf{z}]$ polynomial ring





Polynomial Ideals — Basis Theorem

David Hilbert (1890)



Every ideal of $\mathbb{Q}[\mathbf{z}]$ is finitely generated !





Consequence

- Polynomial Invariants can be represented by polynomial ideals!
- Finite conjunctions suffice!





Consequence

- Polynomial Invariants can be represented by polynomial ideals!
- Finite conjunctions suffice!
- There are effective algorithms for
 - membership
 - inclusion
 - equality





Notation:

$$r_{qj}^{(f)} = (\llbracket f \rrbracket (\mathbf{x}_1, \ldots, \mathbf{x}_k))_{qj}$$

z fresh set of variables





Notation:

$$r_{qj}^{(f)} = (\llbracket f \rrbracket (\mathbf{x}_1, \ldots, \mathbf{x}_k))_{qj}$$

z fresh set of variables

 $p\mapsto \mathit{I}_{p}\subseteq \mathbb{Q}[\mathsf{z}]$





Notation:
$$r_{qj}^{(f)} = (\llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k))_{qj}$$

z fresh set of variables

 $\rho \mapsto I_{\rho} \subseteq \mathbb{Q}[\mathbf{z}]$ is inductive if for $\rho \to f(p_1, \dots, p_k)$,

$$egin{aligned} & I_{p} \subseteq \{r \in \mathbb{Q}[\mathbf{Z}] \mid r[r^{(f)}/\mathbf{Z}] \in & & & & & \\ & & & I_{p_{1}} & \oplus \ldots \oplus & I_{p_{k}} & & & & & \end{bmatrix} \end{aligned}$$

holds.





Notation:
$$r_{qj}^{(f)} = (\llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k))_{qj}$$

z fresh set of variables

 $\rho\mapsto I_{\rho}\subseteq \mathbb{Q}[\mathbf{z}]$ is inductive if for $\rho \to f(p_1,\ldots,p_k)$,

$$\begin{split} \textit{I}_{p} &\subseteq \{\textit{r} \in \mathbb{Q}[\textbf{z}] \mid \textit{r}[\textit{r}^{(f)}/\textbf{z}] \in \\ & \textit{I}_{p_{1}}(\textbf{x}_{1}) \quad \oplus \ldots \oplus \textit{I}_{p_{k}}(\textbf{x}_{k}) \quad \} \end{split}$$

holds.





Notation:
$$r_{qj}^{(f)} = (\llbracket f \rrbracket (\mathbf{x}_1, \dots, \mathbf{x}_k))_{qj}$$

z fresh set of variables

 $\rho \mapsto I_{\rho} \subseteq \mathbb{Q}[\mathbf{z}]$ is inductive if for $\rho \to f(p_1, \dots, p_k)$,

$$\begin{split} I_{\rho} &\subseteq \{ r \in \mathbb{Q}[\mathbf{Z}] \mid r[r^{(f)}/\mathbf{Z}] \in \\ & \langle I_{\rho_1}(\mathbf{X}_1) \rangle_{\mathbb{Q}[\mathbf{X}]} \oplus \ldots \oplus \langle I_{\rho_k}(\mathbf{X}_k) \rangle_{\mathbb{Q}[\mathbf{X}]} \} \end{split}$$

holds.



Main Theorem

• Assume $p \mapsto l_p$ is inductive. Then for every $r \in l_p$, $r(\llbracket t \rrbracket) = 0 \qquad (t \in \operatorname{dom}(p))$





Main Theorem

- Assume $p\mapsto l_p$ is inductive. Then for every $r\in l_p,$ $r(\llbracket t
 rbracket)=0 \qquad (t\in \operatorname{dom}(p))$
- For $p \in P$, let $\overline{l}_p = \{r \in \mathbb{Q}[\mathbf{z}] \mid r(\llbracket t \rrbracket) = 0 \quad (t \in \text{dom}(p))\}$ Then $p \mapsto \overline{l}_p$ is inductive.



Main Theorem

• Assume $p\mapsto l_p$ is inductive. Then for every $r\in l_p,$ $r(\llbracket t
rbracket)=0 \qquad (t\in ext{dom}(p))$

• For
$$p \in P$$
, let
 $\overline{l}_p = \{r \in \mathbb{Q}[\mathbf{z}] \mid r(\llbracket t \rrbracket) = 0 \quad (t \in \text{dom}(p))\}$
Then $p \mapsto \overline{l}_p$ is inductive.

Corollary

Let $H(\mathbf{z}) = \mathbf{z}_{q0} - \mathbf{z}_{q'0}$. Then q, q' are equivalent (relative to p_0) iff

 $H \in I_{p_0}$

for some inductive invariant.





Discussion

• The best inductive invariant $p \mapsto \overline{l}_p$ is a greatest fixpoint.

Greatest fixpoint iteration may not terminate.





Discussion

- The best inductive invariant *p* → *l*_p is a greatest fixpoint.
 Greatest fixpoint iteration may not terminate.
- All inductive invariants, though, can be recursively enumerated!





Discussion

- The best inductive invariant *p* → *l*_p is a greatest fixpoint.
 Greatest fixpoint iteration may not terminate.
- All inductive invariants, though, can be recursively enumerated!
- All potential counter examples can be enumerated ...







Theorem

• Equivalence of deterministic tree-to-int transducers with parameters is decidable.




Wrap-up

Theorem

- Equivalence of deterministic tree-to-int transducers with parameters is decidable.
- Equvalence of general deterministic tree-to-string transducers is decidable.







Parameters allow to encode general output alphabets by means of unaries, i.e., numbers.







Parameters allow to encode general output alphabets by means of unaries, i.e., numbers.

Equivalence for unary transducers can be handled by means of techniques from precise program analysis, i.e., program proving.





Thank you!



