# Complexity of the Diagonal Problem for Recursion Schemes

**Paweł Parys** 

University of Warsaw

# <u>Higher-order recursion schemes – what is this?</u>

### **Definition**

Recursion schemes = simply-typed lambda-calculus + recursion or: a context-free grammar in which nonterminals can take arguments

### In other words:

- programs with recursion
- higher-order functions (i.e., functions taking other functions as parameters)
- every function/parameter has a fixed type
- no data values, only functions

We consider here nondeterministic schemes.

# <u>Higher-order recursion schemes – what is this?</u>

### **Definition**

Recursion schemes = simply-typed lambda-calculus + recursion or: a context-free grammar in which nonterminals can take arguments

### In other words:

- programs with recursion
- higher-order functions (i.e., functions taking other functions as parameters)
- every function/parameter has a fixed type
- no data values, only functions

We consider here nondeterministic schemes.

<u>actual objects</u> <u>abstraction</u>

sequential programs ——> finite-state systems

programs with H-O recursion =>H-O recursion schemes

### Rules:

```
S \rightarrow A a

A f \rightarrow A (D f)

A f \rightarrow f b

D f x \rightarrow f (f x)
```

Here: S,A,D are nonterminals (where S – starting nonterminal), a,b are terminals, f,x are variables (parameters)

### Rules:

```
S \rightarrow A a

A f \rightarrow A (D f)

A f \rightarrow f b

D f x \rightarrow f (f x)
```

Here: S,A,D are nonterminals (where S – starting nonterminal), a,b are terminals, f,x are variables (parameters)

The same using another syntax:

```
fun S() = \{ A(a); \}
fun A(f) = \{ if ? then A(\lambda x.D(f,x)) else f(b); \}
fun D(f,x) = \{ f(f(x)); \}
```

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

Here: S,A,D are nonterminals (where S – starting nonterminal), a,b are terminals, f,x are variables (parameters)

In our world, everything is required to have a fixed type:

```
S: o, A: (o \rightarrow o) \rightarrow o, D: (o \rightarrow o) \rightarrow o \rightarrow o

a: o \rightarrow o, b: o

f: o \rightarrow o, x: o
```

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

$$S \rightarrow A a$$

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

$$S \rightarrow A a \rightarrow A (D a)$$

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

$$S \rightarrow A a \rightarrow A (D a) \rightarrow A (D (D a))$$

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

$$S \rightarrow A a \rightarrow A (D a) \rightarrow A (D (D a)) \rightarrow D (D a) b$$

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

$$S \rightarrow A a \rightarrow A (D a) \rightarrow A (D (D a)) \rightarrow D (D a) b \rightarrow D a (D a b)$$

### Rules:

$$S \rightarrow A a$$

$$A f \rightarrow A (D f)$$

$$A f \rightarrow f b$$

$$D f x \rightarrow f (f x)$$

$$S \rightarrow A \ a \rightarrow A \ (D \ a) \rightarrow A \ (D \ (D \ a)) \rightarrow D \ (D \ a) \ b \rightarrow D \ a \ (D \ a \ b)$$
  
  $\rightarrow a \ (a \ (D \ a \ b))$ 

### Rules:

$$S \rightarrow A a$$

$$A f \rightarrow A (D f)$$

$$A f \rightarrow f b$$

$$D f x \rightarrow f (f x)$$

$$S \rightarrow A \ a \rightarrow A \ (D \ a) \rightarrow A \ (D \ (D \ a)) \rightarrow D \ (D \ a) \ b \rightarrow D \ a \ (D \ a \ b)$$
  
 $\rightarrow a \ (a \ (a \ (a \ b)))$ 

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

### Dynamics:

$$S \rightarrow A \ a \rightarrow A \ (D \ a) \rightarrow A \ (D \ (D \ a)) \rightarrow D \ (D \ a) \ b \rightarrow D \ a \ (D \ a \ b)$$
  
 $\rightarrow a \ (a \ (a \ (a \ b)))$ 

# Basic setting:

• Terminals are either of type  $o \rightarrow o$  (one argument of ground type), or of type o (no arguments)

Then a scheme generates a word (e.g., a a a a b)

### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

### Dynamics:

```
S \rightarrow A \ a \rightarrow A \ (D \ a) \rightarrow A \ (D \ (D \ a)) \rightarrow D \ (D \ a) \ b \rightarrow D \ a \ (D \ a \ b)
\rightarrow a \ (a \ (a \ (a \ (a \ b)))
```

### Basic setting:

• Terminals are either of type  $o \rightarrow o$  (one argument of ground type), or of type o (no arguments)

Then a scheme generates a word (e.g., a a a a b)

Recognized language:  $\{a^{2^n}b: n \in \mathbb{N}\}$ 

### Rules:

$$S \rightarrow A a$$

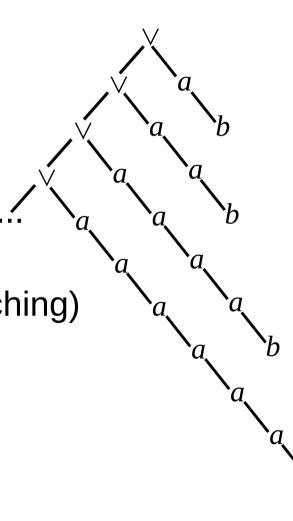
$$A f \rightarrow A (D f)$$

$$A f \rightarrow f b$$

$$D f x \rightarrow f (f x)$$

Recognized language:  $\{a^{2^n}b: n \in \mathbb{N}\}$ 

Another view: generating a single infinite tree, containing all possible derivations of the scheme (" $\vee$ " = branching)



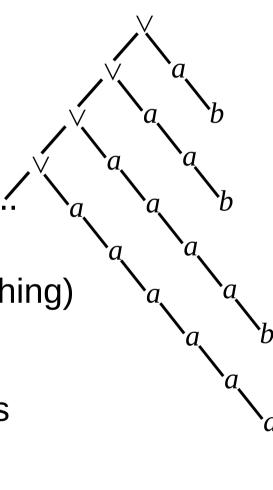
### Rules:

$$S \rightarrow A a$$
  
 $A f \rightarrow A (D f)$   
 $A f \rightarrow f b$   
 $D f x \rightarrow f (f x)$ 

Recognized language:  $\{a^{2^n}b:n\in\mathbb{N}\}$ 

Another view: generating a single infinite tree, containing all possible derivations of the scheme (" $\vee$ " = branching)

General goal: verifying properties of languages/trees generated by schemes



# Model checking for schemes

General goal: verifying properties of languages/trees generated by schemes

# Theorem [Ong 2006]

There is an algorithm, which given a recursion scheme G and an MSO formula  $\phi$  checks whether  $\phi$  holds in the tree generated by G.

# Model checking for schemes

General goal: verifying properties of languages/trees generated by schemes

# Theorem [Ong 2006]

There is an algorithm, which given a recursion scheme G and an MSO formula  $\phi$  checks whether  $\phi$  holds in the tree generated by G.

Moreover: there exist (experimental) tools that take as input a program in a functional language, and a property, and check whether the program satisfies this property. They approximate the program by a recursion scheme, and then they check the property in the tree generated by the scheme.

# Model checking for schemes

General goal: verifying properties of languages/trees generated by schemes

# Theorem [Ong 2006]

There is an algorithm, which given a recursion scheme G and an MSO formula  $\phi$  checks whether  $\phi$  holds in the tree generated by G.

Moreover: there exist (experimental) tools that take as input a program in a functional language, and a property, and check whether the program satisfies this property. They approximate the program by a recursion scheme, and then they check the property in the tree generated by the scheme.

Aim of this work: verify properties not expressible in MSO.

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$ 

containing every letter from A at least n times?

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### Example:

```
A = \{a, b\}
L_1 = \{a^nb \mid n \in \mathbb{N}\} \cup \{ab^n \mid n \in \mathbb{N}\} - \text{answer} = \mathbb{NO}
L_2 = a^*b^* - \text{answer} = \text{YES}
L_3 = \{a^nb^{n+3} \mid n \in \mathbb{N}\} - \text{answer} = \text{YES}
L_4 = \{bbb(ab)^n \mid n \in \mathbb{N}\} - \text{answer} = \text{YES}
```

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### Remark

Notice that this is not a regular property of the tree generated by S, i.e., the tree containing all words from L on particular branches. We say here that "something is unbounded".

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### **Motivation?**

1. Interesting on its own: sometimes we want to know whether some event may happen arbitrarily many times, or not.

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### **Motivation?**

- 1. Interesting on its own: sometimes we want to know whether some event may happen arbitrarily many times, or not.
- 2. It is easy to reduce the problem of computing the downward closure of the language recognized by a scheme to the diagonal problem The downward closure  $L\downarrow$  of a language L contains all words that can be obtained from words in L by removing some letters [Zetzsche 2015]. The downward closure is always a regular language.

```
examples:

(a^*b^*)\downarrow = a^*b^*

\{a^nb^{n+3} \mid n \in \mathbb{N}\}\downarrow = a^*b^*

\{bbb(ab)^n \mid n \in \mathbb{N}\}\downarrow = \{a,b\}^*
```

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### **Motivation?**

- 1. Interesting on its own: sometimes we want to know whether some event may happen arbitrarily many times, or not.
- 2. It is easy to reduce the problem of computing the downward closure of the language recognized by a scheme to the diagonal problem The downward closure  $L\downarrow$  of a language L contains all words that can be obtained from words in L by removing some letters [Zetzsche 2015].
- 2'. For languages recognized by schemes, it is undecidable whether  $L=A^*$ ,  $L_1=L_2$ , etc. But we can check this approximately, by checking whether  $L\downarrow=A^*$ ,  $L_1\downarrow=L_2\downarrow$ , etc.

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### **Motivation?**

- 1. Interesting on its own: sometimes we want to know whether some event may happen arbitrarily many times, or not.
- 2. It is easy to reduce the problem of computing the downward closure of the language recognized by a scheme to the diagonal problem The downward closure  $L\downarrow$  of a language L contains all words that can be obtained from words in L by removing some letters [Zetzsche 2015].
- 2'. For languages recognized by schemes, it is undecidable whether  $L=A^*$ ,  $L_1=L_2$ , etc. But we can check this approximately, by checking whether  $L\downarrow=A^*$ ,  $L_1\downarrow=L_2\downarrow$ , etc.
- 3. The problem "is there a piecewise testable language (i.e., boolean combination of downward closed languages) containing  $L_1$  and not intersecting with  $L_2$ " reduces to the diagonal problem [Czerwiński, Martens, van Rooijen, Zeitoun 2015]. This gives a more refined approximation for disjointness of  $L_1$  and  $L_2$  than the test  $L_1 \!\!\! \downarrow \cap L_2 \!\!\! \downarrow = \emptyset$ .

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$ 

containing every letter from *A* at least *n* times?

<u>Theorem</u> [Clemente, P., Walukiewicz, Salvati 2016] The diagonal problem for recursion schemes is decidable.

This paper: further improvements

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$ 

containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch this word written in leaves

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch this word written in leaves

# Order of a type:

```
ord(o) = 0

ord(\alpha_1 \rightarrow ... \rightarrow \alpha_k \rightarrow o) = 1+max(ord(\alpha_1), ..., ord(\alpha_k))
```

### For example:

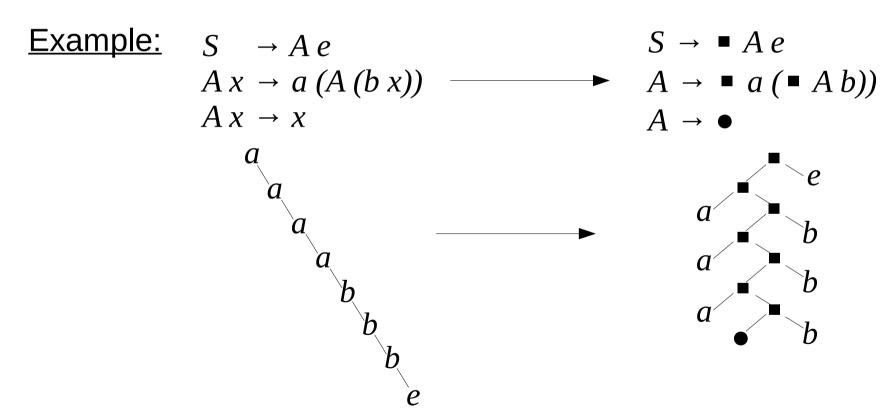
- ord(o) = 0,
- ord $(o \rightarrow o)$  = ord $(o \rightarrow o \rightarrow o)$  = 1,
- ord $(o \rightarrow (o \rightarrow o) \rightarrow o) = 2$

Order of a HORS = maximal order of (a type of) its nonterminal

Input: a scheme S recognizing L, set of letters A Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch this word written in leaves



We have generalized the previous setting: our new scheme recognizes a language of trees, not a language of words

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch this word written in leaves

Example: 
$$S \rightarrow Ae$$

$$Ax \rightarrow a(A(bx))$$

$$Ax \rightarrow x$$

$$S \rightarrow Ae$$

$$A \rightarrow a(Ab)$$

$$A \rightarrow A \rightarrow A$$

<u>Idea:</u> 1) Observe that an argument of type o can be used at most once.

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch this word written in leaves

Example: 
$$S \rightarrow Ae$$

$$Ax \rightarrow a(A(bx))$$

$$Ax \rightarrow x$$

$$S \rightarrow Ae$$

$$A \rightarrow a(Ab)$$

$$A \rightarrow A \rightarrow A$$

- Idea: 1) Observe that an argument of type o can be used at most once.
  - 2) All arguments of type o are dropped ( $\Rightarrow$  order decreases).
  - 3) Every subterm MN with N of type o can be replaced
    - a) either by  $\blacksquare MN$  (when the argument is used in M),
    - b) or by M (when the argument is ignored in M).

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch this word written in leaves

Example: 
$$S \rightarrow Ae$$

$$Ax \rightarrow a(A(bx)) \longrightarrow A \rightarrow a(Ab)$$

$$Ax \rightarrow x \longrightarrow Ae$$

$$A \rightarrow a(Ab)$$

- Idea: 1) Observe that an argument of type o can be used at most once.
  - 2) All arguments of type o are dropped ( $\Rightarrow$  order decreases).
  - 3) Every subterm MN with N of type o can be replaced
    - a) either by  $\blacksquare MN$  (when the argument is used in M),
    - b) or by M (when the argument is ignored in M).
  - 4) Additional work is required to choose correctly a) or b).

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$ 

containing every letter from A at least n times?

### How to solve it?

a scheme S of order m with step 1 a scheme S of order m-1 with a word written on a branch step 2 this word written in leaves

a scheme S of order m-1 with a similar word written on a branch

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

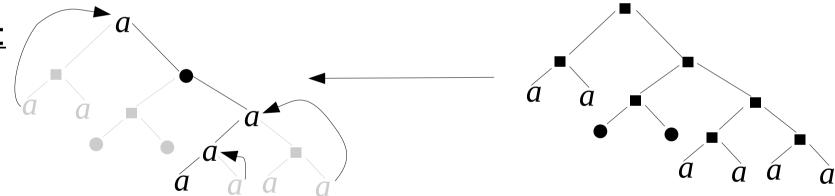
### How to solve it?

a scheme S of order m with a word written on a branch

step 1 a scheme S of order m-1 with this word written in leaves

a scheme S of order m-1 with a similar word written on a branch

# Example:



### <u>Idea:</u>

- 1) Choose (nondeterministically) only one branch.
- 2) For every removed subtree with a, write a new a just above.

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

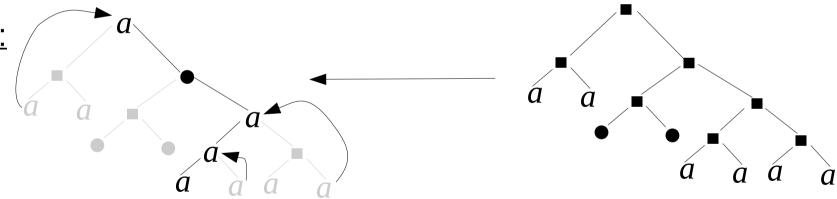
#### How to solve it?

a scheme S of order m with a word written on a branch

step 1 a scheme S of order m-1 with this word written in leaves

a scheme S of order m-1 with a similar word written on a branch

# <u>Example:</u>



#### <u>Idea:</u>

- 1) Choose (nondeterministically) only one branch.
- 2) For every removed subtree with a, write a new a just above.
- 3) The number of a's decreases at most logarithmically, if the branch is chosen correctly (always go to the subtree with more a's).

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

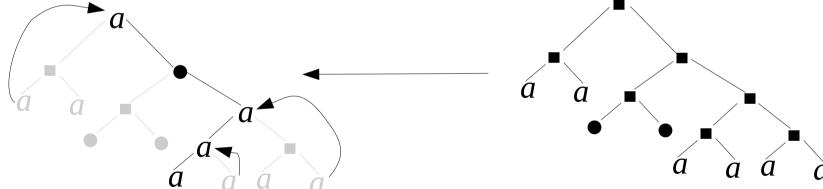
#### How to solve it?

a scheme S of order m with a word written on a branch

step 1 a scheme S of order m-1 with this word written in leaves

a scheme S of order m-1 with a similar "word" written on |A| branches





#### <u>Idea:</u>

- 1) Choose (nondeterministically) only one branch.
- 2) For every removed subtree with a, write a new a just above.
- 3) The number of a's decreases at most logarithmically, if the branch is chosen correctly (always go to the subtree with more a's). We have to this for every letter  $\Rightarrow |A|$  branches

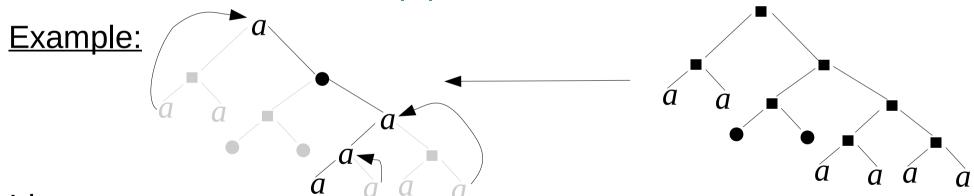
Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

#### How to solve it?

a scheme S of order m with a  $\frac{\text{step 1}}{\text{word"}}$  a scheme S of order m-1 with "word" written on |A| branches  $\frac{1}{\text{step 2}}$  this "word" written in leaves

a scheme S of order m-1 with a similar "word" written on |A| branches



#### <u>Idea:</u>

- 1) Choose (nondeterministically) only one branch.
- 2) For every removed subtree with a, write a new a just above.
- 3) The number of a's decreases at most logarithmically, if the branch is chosen correctly (always go to the subtree with more a's). We have to this for every letter  $\Rightarrow |A|$  branches

Input: a scheme S recognizing L, set of letters AQuestion: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$ containing every letter from A at least n times?

How to solve it?

a scheme S of order m with a tous = tous =

Repeat these steps until the order drops down to 0, and solve the diagonal problem for a regular language.

Input: a scheme S recognizing L, set of letters A

Question: is it the case that for every  $n \in \mathbb{N}$  there is a word  $w \in L$  containing every letter from A at least n times?

How to solve it?

a scheme S of order m with a true step 1 true a scheme <math>true step 1 true step

a scheme S of order m-1 with a similar "word" written on |A| branches

Repeat these steps until the order drops down to 0, and solve the diagonal problem for a regular language.

## **Complexity** of this solution?

- Every step applied to a scheme of order m increases its size m-fold exponentially.
- In total the complexity becomes about  $m^2$ -EXPTIME

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

 Basic idea: we give a type system that "simulates" all the steps of the previous algorithm at once. While deriving a type in this type system, we recover the solution to the diagonal problem.

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

 Basic idea: we give a type system that "simulates" all the steps of the previous algorithm at once. While deriving a type in this type system, we recover the solution to the diagonal problem.

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

 Basic idea: we give a type system that "simulates" all the steps of the previous algorithm at once. While deriving a type in this type system, we recover the solution to the diagonal problem.

## **Complexity:**

• For schemes of order m (recognizing languages of words), the number of types is (m-1)-fold exponential.

Thus the problem is in (m-1)-EXPTIME (or in NP for  $m \in \{0,1\}$ ).

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

 Basic idea: we give a type system that "simulates" all the steps of the previous algorithm at once. While deriving a type in this type system, we recover the solution to the diagonal problem.

## **Complexity:**

- For schemes of order m (recognizing languages of words), the number of types is (m-1)-fold exponential.
  - Thus the problem is in (m-1)-EXPTIME (or in NP for  $m \in \{0,1\}$ ).
- We also prove a corresponding lower bound: the problem is (m-1)-EXPTIME-complete.
- For schemes generating languages of trees, the problem is m-EXPTIME-complete (or NP-complete for m=0).

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

## **Complexity:**

- For schemes of order m (recognizing languages of words), the number of types is (m-1)-fold exponential. Thus the problem is in (m-1)-EXPTIME (or in NP for  $m \in \{0,1\}$ ).
- We also prove a corresponding lower bound: the problem is (m-1)-EXPTIME-complete.
- For schemes generating languages of trees, the problem is m-EXPTIME-complete (or NP-complete for m=0).
- This complexity is high. But all known problems for schemes of order m are at least (m-1)-EXPTIME-hard, even the question whether the recognized language is nonempty. Beside of that, for some problems there exist tools that work in practice.
- All algorithms for schemes that were implemented are based on type systems, so also our algorithm have the potential of working in practice (typing is complicated in the worst-case, but usually it is much easier).

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

#### Advantages of our algorithm:

- Faster
- Based on a type system
- It gives also the reflection property:

Recall that we can see a recursion scheme as a generator of an infinite tree, containing all words of the recognized language on branches. Given a scheme G we can construct a scheme G which generates the same tree as G, but enriched with additional labels: the label of every node of the tree says what is the answer to the diagonal problem for the subtree starting in this node (thus we have not only the answer for the whole tree, but also for every subtree).

In this paper we give a new algorithm solving the diagonal problem for recursion schemes.

#### Further work:

- The algorithm for the diagonal can be used for model-checking trees generated by schemes against formulas of the WMSO+U logic [P., submitted to STACS 2018]
- This logic extends the MSO logic (weak variant only quantification over finite sets) with the U quantifier:
  - $UX.\phi$  means that  $\phi$  holds for some arbitrarily large finite sets X

