

Interpretation of Stream Programs

Romain Péchoux

Université de Nancy 2, LORIA

24 october 2010

Outline

Goal

There is a need of general studies in order to develop new formal methods and tools for the analysis of stream programs.

Several studies [Sijtsma 89, Klop et al. 89,97,07] have focused on the notion of productivity, i.e. the ability to produce a normal form of the n -th output element.

However formal methods on the complexity analysis are lacking.

The methodology

This talk presents a combination of works with :

- ▶ Marco Gaboardi (Universita di Bologna)
- ▶ Mathieu Hoyrup (INRIA), Emmanuel Hainry (Nancy 1) and Hugo Férée (ENS Lyon)

using [quasi-intrepretation-based techniques](#) that are useful to prove quantitative properties of stream programs.

Setting and Motivations

Motivations

- ▶ Program [resource control](#) and resource usage certification.
- ▶ Properties for programs working on infinite data structures, notably [streams](#).

Settings

- ▶ Lazy Functional Programming Languages, like Haskell, enable a treatment of stream programs in a [finitary way](#).
- ▶ Use of static analysis interpretation methods inspired by :
 - ▶ [Quasi-interpretation](#) [Bonfante et al. 01] and
 - ▶ [Sup-interpretation](#) [Marion et al. 06]
- ▶ The presented criteria permit to build [upper bounds on](#) the number and size of computed stream elements.

Streams

- ▶ Streams can be used to formalize network communication flows, audio and video signals flows, etc.
- ▶ For simplicity, I will consider streams as **infinite sequences** of natural numbers :

$0 : 1 : 2 : 3 : 4 : 5 : 6 : \dots : 65512 : 65513 \dots$

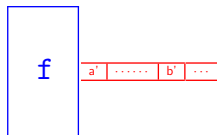
$1 : 1 : 2 : 3 : 5 : 8 : 13 : \dots : 46368 : 75025 : \dots$

$0 : 1 : 1 : 0 : 1 : 0 : 0 : 1 : 1 : 0 : 0 : 1 : 0 : 1 : 1 : 0 \dots$

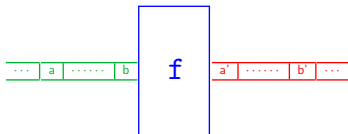
- ▶ In the sequel we will consider programs built over numerals and addition and extended by particular function definitions.

Representation of Stream programs

- ▶ It is convenient to distinguish between two different kinds of programs dealing with stream :
 - ▶ **Stream constructor** programs



- ▶ **Stream function** programs



Clearly the former are a particular case of the latter.

Streams in lazy functional languages

- ▶ Lazy functional programming languages like Haskell allow us to define streams by means of (typed) expressions :

```
repeat : Nat → [Nat]
repeat x = x : (repeat x)
fibonacci : [Nat]
fibonacci = 1 : 1 : (sadd fibonacci (tail fibonacci))
```

- ▶ We restrict the considered programs to a first order fragment of Haskell.
- ▶ We use a **lazy evaluation semantics** \Downarrow without sharing :

```
fibonacci  $\Downarrow$  1 : 1 : (sadd fibonacci (tail fibonacci))
repeat 3  $\Downarrow$  3 : (repeat 3)
```

Preliminaries on Programs

- ▶ We will extensively use the standard function definition :

$\text{take} :: \text{Nat} \rightarrow [\alpha] \rightarrow [\alpha]$

$\text{take } 0 \text{ s} = \text{nil}$

$\text{take } (x + 1) \text{ nil} = \text{nil}$

$\text{take } (x + 1) (y : \text{ys}) = y : (\text{take } x \text{ ys})$

$!! :: [\alpha] \rightarrow \text{Nat} \rightarrow \alpha$

$(x : \text{xs})!!0 = x$

$(x : \text{xs})!!(y + 1) = \text{xs} !! y$

- ▶ "Approximation lemma" : provides finitary methods

$$s = r \iff \forall n ((\text{take } n \text{ s}) = (\text{take } n \text{ r}))$$

- ▶ We will write $e_{\underline{n}}$ and $e \upharpoonright_{\underline{n}}$ as a short for $e !! \underline{n}$ and $\text{take } \underline{n} \text{ e}$.

eval

- ▶ We also need a program **forcing the evaluation** of programs :

$\text{eval} :: A \rightarrow A$

$\text{eval } (\mathbf{c} \ e_1 \ \cdots \ e_n) = \hat{\mathbf{C}} (\text{eval } e_1) \ \cdots \ (\text{eval } e_n)$

where $\hat{\mathbf{C}}$ is a program representing the **strict version** of the primitive constructor \mathbf{c} .

- ▶ in particular when it is applied to stream expressions it diverges :

$\text{eval fibo} \uparrow \quad \text{eval (repeat 3)} \uparrow$
 $\text{eval fibo} \downarrow^\infty 1 : 1 : 2 : 3 : \cdots 75025 \cdots$
 $\text{eval (repeat 3)} \downarrow^\infty 3 : 3 : 3 \cdots 3 \cdots$
 $\text{eval (take 5 fibo)} \downarrow 1 : 1 : 2 : 3 : 5$

- ▶ for this reason in what follows we write $e \downarrow_v v$ for $\text{eval } e \downarrow v$.

Interpretation

A program P admits an **interpretation** if there is an assignment $\llbracket - \rrbracket$:

▶ **total** : $\forall \mathbf{t}, \llbracket \mathbf{t} \rrbracket : (\mathbb{R}^+)^n \rightarrow \mathbb{R}^+$

▶ **monotonic** : $X_i \geq Y_i \Rightarrow \llbracket \mathbf{t} \rrbracket(\dots, X_i, \dots) \geq \llbracket \mathbf{t} \rrbracket(\dots, Y_i, \dots)$,

such that for each definition in P of the shape $\mathbf{f} \ p_1 \cdots p_n = \mathbf{e}$:

$$\llbracket \mathbf{f} \ p_1 \cdots p_n \rrbracket \geq \llbracket \mathbf{e} \rrbracket$$

Moreover, a program P admits an **additive interpretation** if $\llbracket - \rrbracket$ is such that for every symbol \mathbf{c} of arity n :

▶ $\llbracket \mathbf{c} \rrbracket = 0$ if $n = 0$

▶ $\llbracket \mathbf{c} \rrbracket(X_1, \dots, X_n) = \sum_{i=1}^n X_i + \alpha_{\mathbf{c}}$, with $\alpha_{\mathbf{c}} \geq 1$ otherwise

An Example of Interpretation

Consider again the example :

```
repeat : Nat → [Nat]
repeat x = x : (repeat x)
```

the assignment $\llbracket - \rrbracket$ defined as

$$\llbracket x \rrbracket = X \quad \llbracket : \rrbracket (X, Y) = \max(X, Y) \quad \llbracket \text{repeat} \rrbracket (X) = X$$

It is total and monotonic and we check that

$$\llbracket \text{repeat } x \rrbracket \geq \llbracket x : (\text{repeat } x) \rrbracket$$

We have :

$$\begin{aligned} \llbracket \text{repeat } x \rrbracket &= \llbracket \text{repeat} \rrbracket (X) = X \\ &= \max(X, X) = \llbracket : \rrbracket (X, X) \\ &= \llbracket : \rrbracket (X, \llbracket \text{repeat} \rrbracket X) \\ &= \llbracket x : (\text{repeat } x) \rrbracket \end{aligned}$$

Properties of Interpretations

- ▶ Interpretation respects evaluation

Given a program P admitting the interpretation $\langle \! \langle - \! \rangle \! \rangle$, for every closed expression e we have

- ▶ if $e \Downarrow v$ then $\langle \! \langle e \! \rangle \! \rangle \geq \langle \! \langle v \! \rangle \! \rangle$
- ▶ if $e \Downarrow_v v$ then $\langle \! \langle e \! \rangle \! \rangle \geq \langle \! \langle v \! \rangle \! \rangle$

- ▶ Interpretation does not respect infinitary evaluation

If $\text{fibonacci} \Downarrow^\infty u = 1 : 1 : 2 : 3 : 5 : 8 : \dots$, we do not expect :

$$\langle \! \langle \text{fibonacci} \! \rangle \! \rangle \geq \langle \! \langle u \! \rangle \! \rangle$$

- ▶ Interpretation vs. expression size

Given a program P admitting the interpretation $\langle \! \langle - \! \rangle \! \rangle$, there is a function $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that for each expression e :

$$\langle \! \langle e \! \rangle \! \rangle \leq G(|e|)$$

First property - Global Upper Bound

- ▶ Each element is **bounded** by a **function in the maximal size** of the input elements.
- ▶ A program generated using :

```
repeat :: Nat → [Nat]
repeat x = x : (repeat x)
zip :: [α] → [α] → [α]
zip (x : xs) ys = x : (zip ys xs)
square :: [Nat] → [Nat]
square (x : xs) = (mul x x) : (square xs)
```

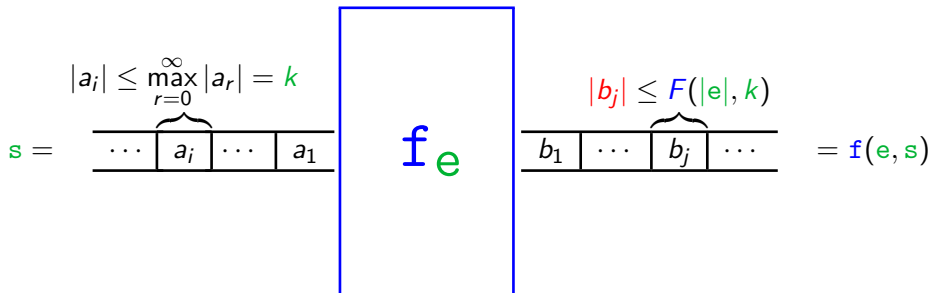
is **bounded by some k** .

- ▶ For example :

```
square (zip (repeat 5) (square (zip (repeat 7) (repeat 4))))
```

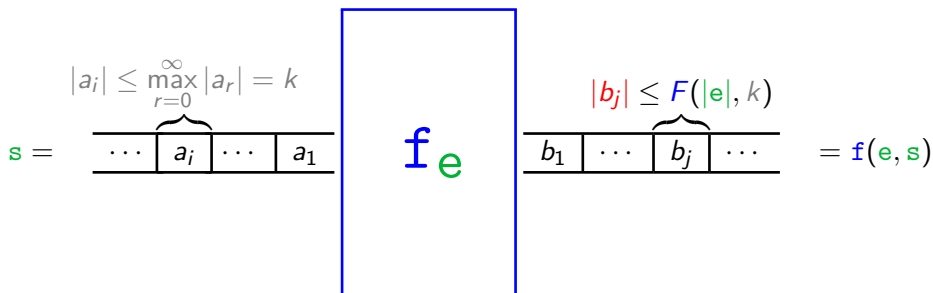
is bounded by **$k = 2401 = 7^4$** .

Global Upper Bound - Intuition



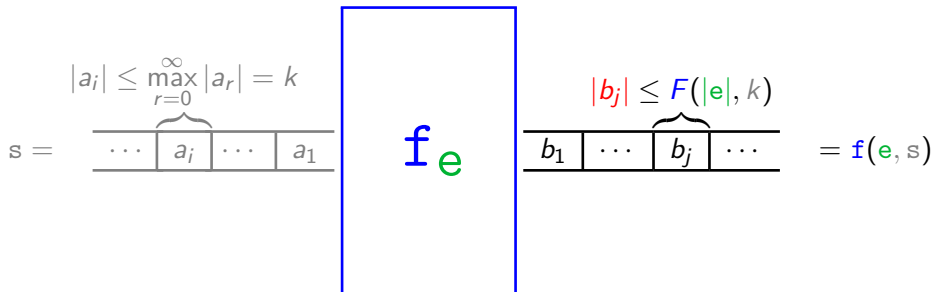
Global Upper Bound - Intuition

Clearly the input stream may have no maximal element :

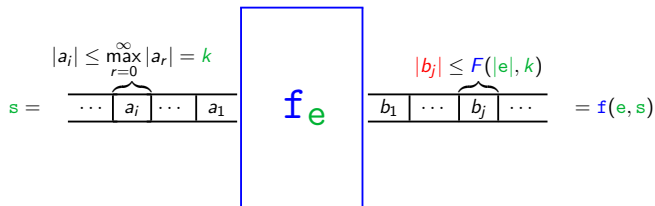


Global Upper Bound - Intuition

Stream constructors are just a particular case :



Global Upper Bound - Formally



- ▶ A function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ of P has a **global upper bound** if there is $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that, for every stream expression $s \in P$ and for every expression $e \in P$ if

$$(f \ s \ e)_{\underline{n}} \Downarrow_v \underline{v}$$

then

$$F(\max(|s|, |e|)) \geq |v|$$

- ▶ A program P has a **global upper bound** if **every function symbol** in it enjoys this property.

GUB Criterion

- ▶ A program P satisfies the **GUB** criterion if it admits an **additive interpretation** (\cdot) such that :

$$(\cdot)(X, Y) = \max(X, Y)$$

- ▶ The interpretations of **GUB programs** well behave with respect to the selection function $!!$. In particular if

$$e_{\underline{n}} \Downarrow_v v$$

then

$$(\underline{e}) \geq (\underline{v})$$

- ▶ **Theorem**

If a program is **GUB** then each stream function f has a **global upper bound** F .

GUB Examples

- ▶ The program including `repeat`, `zip` and `square` is **GUB**, it admits the interpretation :

$$\langle \text{repeat} \rangle (X) = X + 1 \quad \langle \text{zip} \rangle (X, Y) = \max(X, Y)$$

$$\langle \text{square} \rangle (X) = X^2 \quad \langle \text{:} \rangle (X, Y) = \max(X, Y)$$

- ▶ Define `fun x = square (zip (repeat 9) x)`.
Taking $F(X) = \langle \text{square} \rangle (\langle \text{zip} \rangle (\langle \text{repeat} \rangle 9), G(X))$

If $(\text{fun } e) \Downarrow_v \underline{m}$

then $F(|e|) = (\max(10, G(|e|)))^2 \geq \underline{m}$

Second property - Local Upper Bound

- ▶ Each element is **bounded** by a **function in the maximal size** of the input elements and of its **position in the output stream**.
- ▶ Consider the following stream program :

```
nats :: Nat → [Nat]
nats x = x : (nats (x + 1))
sad :: [Nat] → [Nat] → [Nat]
sad (x : xs) (y : ys) = (add x y) : (sad xs ys)
```

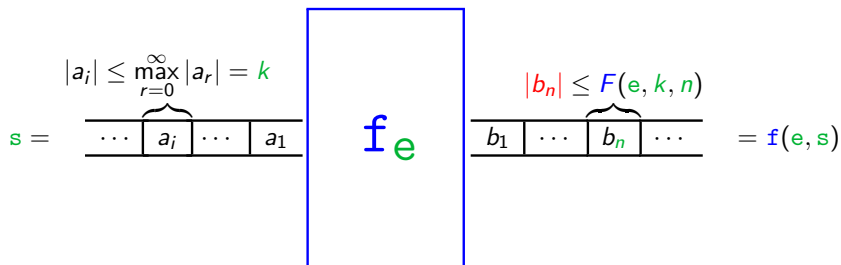
- ▶ the size of each element of a stream s built only using `nats` and `sad` as

$$\text{sad nats } 3 \text{ (sad nats } 5 \text{ nats } 4)$$

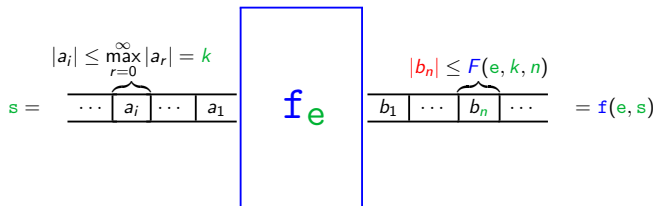
is **not globally bounded** but is bounded by the function $(3 \times 5) \times (n + 1)$ where n is **its position in the stream**.

Local Upper Bound - Intuition

We consider also dependencies on the **local** index n as follows :



Local Upper Bound - Formally



- ▶ A function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ of P has a **local upper bound** if there is $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that, for every stream expression $s \in P$ and for every expression $e \in P$ if

$$(f \ s \ e)_{\underline{n}} \Downarrow_v \underline{v}$$

then

$$F(\max(|s|, |e|, |\underline{n}|)) \geq |\underline{v}|$$

- ▶ A program P has a **local upper bound** if **every stream function symbol** enjoys this property.

Local Upper Bound - Example

The following program computing the Fibonacci sequence :

1 : 1 : 2 : 3 : 5 : 8 : 13 : ... : 46368 : 75025 : ...

has clearly a **local upper bound**. The **size** of each element in position n is bounded by 2^n :

```
fib :: [Nat]                                tail :: [α]
fib = 0 : (1 : (sad fib (tail fib)))         tail x : xs = xs

sad :: [Nat] → [Nat] → [Nat]
sad (x : xs) (y : ys) = (add x y) : (sad xs ys)
```

Parametrized Interpretations

A program P admits a **parametrized interpretation** if there is a total **parametrized assignment** $\langle _ \rangle_L$ over \mathbb{R}^+ , such that :

- ▶ it is extended to expressions by

$$\begin{aligned} \langle \text{hd} : \text{tl} \rangle_L &= \langle _ \rangle_L(\langle \text{hd} \rangle_L, \langle \text{tl} \rangle_{L-1}) && \text{Case :} \\ \langle \text{t } e_1 \cdots e_n \rangle_L &= \langle \text{t} \rangle_L(\langle e_1 \rangle_L, \dots, \langle e_n \rangle_L) && \text{t} \neq _ : \end{aligned}$$

- ▶ for each definition $f \vec{p} = e : \langle f \vec{p} \rangle_L \geq \langle e \rangle_L$

Proposition

- If $e \Downarrow v$ then $\langle e \rangle_r \geq \langle v \rangle_r$
- If $e \Downarrow_v v$ then $\langle e \rangle_r \geq \langle v \rangle_r$
- $\exists G, \forall e, \langle e \rangle_r \leq G(|e|, r)$

LUB Criterion

- ▶ A program P satisfies the **LUB** criterion if it admits an **additive and monotonic (also in L) parametrized additive interpretation** $(-)_L$ such that :

$$(\cdot)_L(X, Y) = \max(X, Y)$$

- ▶ **LUB programs** has suitable properties wrt **!!** :

If $e_{\underline{n}} \Downarrow_v v$

then $(e)_n \geq (v)_0$

- ▶ **Theorem**

If a program is **LUB** then each stream function f has a **local upper bound** F .

LUB Examples

- ▶ The program for the Fibonacci sequence consisting of `tail`, `sad`, and `fib` is **LUB**. It admits the additive interpretation :

$$\begin{aligned} \langle 0 \rangle_L &= 0 & \langle +1 \rangle_L(X) &= X + L + 1 & \langle \cdot \rangle_L(X, Y) &= \max(X, Y) \\ \langle \text{sad} \rangle_L(X, Y) &= \langle \text{add} \rangle_L(X, Y) = X + Y & \langle \text{tail} \rangle_L(X) &= X & \langle \text{fib} \rangle_L &= 2^L \end{aligned}$$

- ▶ The program including `nats` and `sad` is **LUB**, it admits the interp. :

$$\begin{aligned} \langle \text{nats} \rangle_L(X) &= X + L & \langle \text{sad} \rangle(X, Y) &= X + Y \\ \langle \cdot \rangle(X, Y) &= \max(X, Y) \end{aligned}$$

Consider the definition `dubnat x = sad (nats 5) (nats x)`.
Take $F(X) = \langle \text{sad} \rangle_X(\langle \text{nats } 5 \rangle_X, \langle \text{nats} \rangle_X(X))$ and
 $k = \max(|e|, |\underline{n}|)$, if

$$\langle \text{dubnat } e \rangle_{\underline{n}} \Downarrow \underline{m}$$

then $F(k) = \langle \text{sad} \rangle_k(\langle \text{nats } 5 \rangle_k, \langle \text{nats} \rangle_k(k)) = 3k + 5 \geq \underline{m}$

The third and fourth properties - Length and size-based I/O Upper Bounds

- ▶ The **number of output elements** is **bounded** by a **function** in the **number (resp. size) of read input elements**.
- ▶ Consider the following stream program :

`zip :: [α] \rightarrow [α] \rightarrow [$\alpha \times \alpha$]`

`zip (x : xs) ys = x : (zip ys xs)`

`double :: [Nat] \rightarrow [Nat]`

`double (x : xs) = x : (x : (double xs))`

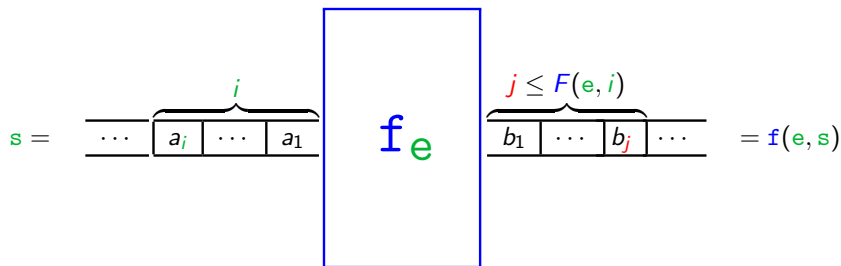
- ▶ the number of output written elements of a stream function `f` defined by :

`f x = double(zip (double x) (double x))`

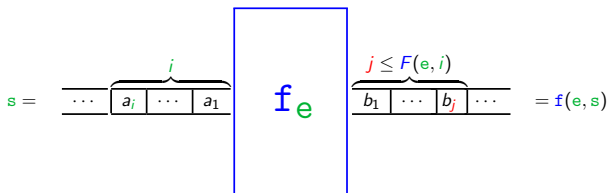
is bounded by a constant `k(=8 in this case)`.

Length-based I/O Upper Bound - Intuition

We consider dependencies between the number of **input readings** and the number of **output writings** :



Length-based I/O Upper Bound - Formally



- ▶ A function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ of P has a **length based I/O upper bound** if there is $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that, for every stream expression $s \in P$ and for every expression $e \in P$ if

$$\text{length}(f \ s \ \underline{n} \ e) \downarrow_v \underline{m}$$

then

$$F(\max(|s|, |e|, |\underline{n}|)) \geq |\underline{m}|$$

- ▶ A program P has a **length based I/O upper bound** if **every stream function** in it enjoys this property.

LBUB Criterion

- ▶ A program is **LBUB** if it admits an additive **interpretation** $(-)$ such that

$$(\cdot)(X, Y) = Y + 1$$

$$(+1)(X) = X + 1$$

- ▶ **Theorem**

If a program is **LBUB** then each **stream function** in it has a **length based I/O upper bound**.

LBUB Examples

The program including `zip` and `double` is **LBUB**, it admits the interpretation :

$$\langle \text{zip} \rangle (X, Y) = X + Y \quad \langle \text{double} \rangle (X) = 2X \quad \langle \text{:} \rangle (X, Y) = Y + 1$$

Consider the definition

```
fun x = double (zip (double x) (double x))
```

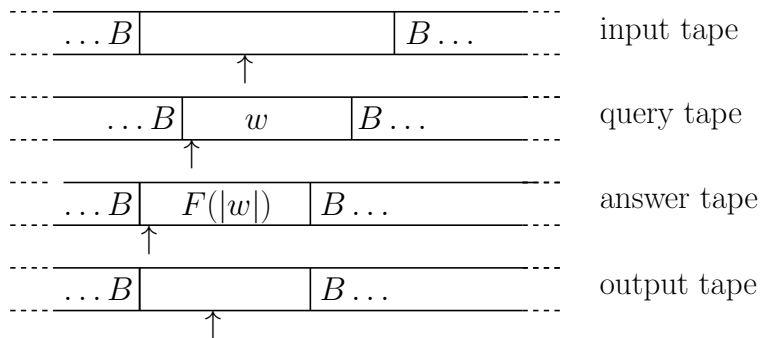
Take $F(X) = \langle \text{double} \rangle (\langle \text{zip} \rangle (\langle \text{double} \rangle (X), \langle \text{double} \rangle (X)))$, and $k = \max(|\underline{n}|, |s|)$, if

$$\text{length}(\text{funone } s \upharpoonright_{\underline{n}}) \Downarrow \underline{m}$$

then

$$F(k) = \langle \text{double} \rangle (\langle \text{zip} \rangle (\langle \text{double} \rangle (k), \langle \text{double} \rangle (X))) = 8 \times k \geq \underline{m}$$

Oracle Turing Machine



Alternative model : OTM' if $F(w)$ is substituted to $F(|w|)$ on the oracle answer tapes

Oracle Turing Machine

- ▶ Costs : Usual step : 1, Oracle step $F(|w|)$
- ▶ An OTM $(\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N} \rightarrow \mathbb{N}$ has time complexity T if it halts before $T(|F|, |x|)$ steps on oracle F and input x , where

$$|F|(n) = \max_{k \leq n} |F(k)|$$

- ▶ Second order poly $P := c \mid X \mid P + P \mid P \times P \mid Y(P)$
- ▶ A function is **OTM-Poly** if it is computed by an OTM of time complexity P , second order polynomial.
- ▶ A function is **BFF** if it is computed by an **OTM'** of time complexity P , second order polynomial.

Second order polynomial interpretations

We slightly modify interpretations by :

- ▶ The interpretation of a function $f : [Nat] \rightarrow Nat \rightarrow [Nat]$ is a function $\llbracket f \rrbracket : (\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N}^2 \rightarrow \mathbb{N}$.
- ▶ The interpretation of a function $f : [Nat] \rightarrow Nat \rightarrow Nat$ is a function $\llbracket f \rrbracket : (\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N} \rightarrow \mathbb{N}$.
- ▶ $\llbracket y \rrbracket(Z) = Y(Z)$
- ▶ $\llbracket !: \rrbracket(X, Y, Z + 1) = 1 + X + Y(Z)$ and $\llbracket !: \rrbracket(X, Y, 0) = 1 + X$.

Second order poly $P := c \mid X \mid P + P \mid P \times P \mid Y(P)$
(polyExp $PE := c \mid X \mid PE + PE \mid PE \times PE \mid Y(2^{PE})$)

Definition A program is WFPoly (WFPolyExp) if it admits a second order polynomial (polyExp) interpretation such that for each definition $f p = e$, $\llbracket f p \rrbracket > \llbracket e \rrbracket$.

Characterizations

Definition A program is WFPoly (WFPolyExp) if it admits a second order polynomial (polyExp) interpretation such that for each definition $f_p = e$, $\langle f_p \rangle > \langle e \rangle$.

Theorem

- ▶ A WFPoly (and WFPolyExp) program is productive
- ▶ The set of functions computed by WFPoly programs is exactly OTM-Poly
- ▶ The set of functions computed by WFPolyExp programs is exactly BFF

Future works

Further developments :

- ▶ Extend the work on OTM
- ▶ Study how to **combine the different criteria** in order to capture a wider set of examples
- ▶ Adapt the techniques presented here to study space properties relevant for **synchronous stream processing languages**, e.g. buffering, memory leak, reachability.