Randomised Strategies in the λ-calculus

Gabriele Vanoni

Politecnico di Milano (on leave)

October 8, 2018

Joint work with Ugo Dal Lago

The usual untyped λ-calculus

Given a denumerable set of variables $\mathcal{V}$,

$$(\texttt{terms})\, M, N ::= x \in \mathcal{V} \;\Big|\; MN \;\Big|\; \lambda x.M$$

$$(\texttt{contexts})\, C, D ::= \Box \;\Big|\; MC \;\Big|\; CM \;\Big|\; \lambda x.C$$

$$(\beta\text{-}\texttt{rule})\, (\lambda x.M)N \longrightarrow_\beta M\{N/x\}$$

β-$\texttt{reduction}$ relation is the contextual closure of the β-$\texttt{rule}$:

$$\frac{M \longrightarrow_\beta N}{C[M] \longrightarrow_\beta C[N]}$$

Strategies in the λ-calculus

Example

How do we reduce $M = (\lambda x.x)N$ where $N = (\lambda x.x)y$?In M there are two contexts in which we can apply the β-rule.

In fact β-`reduction` is a relation, not a function!

How do we pick the redex to reduce in a λ-term?

We need a **reduction strategy**, i.e. a computable (partial) function that given a reducible λ-term selects the redex to reduce.

Example

Leftmost-outermost: $M \longrightarrow_{\mathsf{LO}} N \longrightarrow_{\mathsf{LO}} y$.
Rightmost-innermost: $M \longrightarrow_{\mathsf{RI}} (\lambda x.x)y \longrightarrow_{\mathsf{LO}} y$.

Problems?

A few.

- Not all strategies are **normalizing**.

Let $\omega = \lambda x.xx$ and $\boldsymbol{\Omega} = \omega\omega$. Note that $\boldsymbol{\Omega} \longrightarrow_{\beta} \boldsymbol{\Omega}$.

$$(\lambda x.y)\boldsymbol{\Omega} \longrightarrow_{\mathsf{LO}} y$$
$$(\lambda x.y)\boldsymbol{\Omega} \longrightarrow_{\mathsf{RI}} (\lambda x.y)\boldsymbol{\Omega} \longrightarrow_{\mathsf{RI}} (\lambda x.y)\boldsymbol{\Omega} \longrightarrow_{\mathsf{RI}} \cdots$$

- Reductions of the same term to normal form can be of **different length** under different strategies.

Let $\mathbf{I} = \lambda x.x$.

$$(\lambda x.xx)(\mathbf{II}) \longrightarrow_{\mathsf{LO}} (\mathbf{II})(\mathbf{II}) \longrightarrow_{\mathsf{LO}} \mathbf{I}(\mathbf{II}) \longrightarrow_{\mathsf{LO}} \mathbf{II} \longrightarrow_{\mathsf{LO}} \mathbf{I}$$
$$(\lambda x.xx)(\mathbf{II}) \longrightarrow_{\mathsf{RI}} (\lambda x.xx)\mathbf{I} \longrightarrow_{\mathsf{RI}} \mathbf{II} \longrightarrow_{\mathsf{RI}} \mathbf{I}$$

Minimal Strategies

Natural question

There exists a strategy that is **minimal** in the number of steps to **normal form**?

Predictable answer [Barendregt '84]

NO! (Of course such a function exists, but it is not computable in the general case)

So what?

Restricting ourselves to some **sub-λ-calculi** the problem of minimality becomes decidable.

Theorem

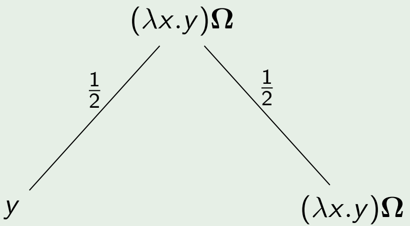
In λA, where copy is forbidden, LO *is minimal.*
In λI, where erasing is forbidden, RI *is minimal.*

Any insight about the general case?

What if we choose the redex to reduce according to some probability distribution?

Example

We can devise a **randomised strategy** U such that for each term M, every redex in M is reduced with the same probability i.e. $\frac{1}{|\mathcal{R}_M|}$.



In order to study properties of such a strategy we need to develop a suitable mathematical framework.

Definition

$\rho \in \mathrm{PDist}\,(S)$ if $\rho : S \to [0,1]$ and $|\rho| = \sum_{s \in S} \rho(s) \leq 1$.

$\rho \in \mathrm{Dist}\,(S)$ if $\rho \in \mathrm{PDist}\,(S)$ and $|\rho| = \sum_{s \in S} \rho(s) = 1$.

Definition (Randomised Strategies)

Given an ARS $(S, \to)$, a randomised reduction strategy P for $(S, \to)$ is a partial function such that if $s \in S$ is in normal form, then $\mathrm{P}(s) = \perp$, otherwise $\mathrm{P}(s) = \mu$, and $\mathrm{Supp}(\mu) \subseteq \{t \mid s \to t\}$.

(S, P) can be seen as a *fully* probabilistic abstract reduction system (FPARS), i.e. a purely probabilistic PARS, without any nondeterminism. The configuration of an FPARS is $\rho \in \mathrm{PDist}\,(S)$, its evolution a function $\mathrm{E} : \mathrm{PDist}\,(S) \to \mathrm{PDist}\,(S)$.

Termination

What does it mean that an FPARS (S, P) **terminates**?
In a **probabilistic** setting there are different possible answers.

Definition (AST)

A term is almost-surely terminating if it reaches normal form in a finite number of steps almost-surely i.e. with probability 1.

$\mathrm{Steps}_{\mathrm{P}}(s)$ is the **average** number of steps of s to its normal form.

Definition (PAST)

A term s is positive almost-surely terminating if $\mathrm{Steps}_{\mathrm{P}}(s) < \infty$.

PAST implies AST but the converse is **not** true! Think about the simmetric random walk on $\mathbb{Z}$.

$$\begin{tikzcd} \mathrm{AST} \arrow[rr, bend left, "/" marking] & & \mathrm{PAST} \arrow[ll, bend left] \end{tikzcd}$$

An example

Let us consider an abstract FPARS.

$$\tfrac{1}{2} \circlearrowleft \mathbf{a} \xrightarrow{\quad \frac{1}{2} \quad} \mathbf{b}$$

The computation is a stochastic process:

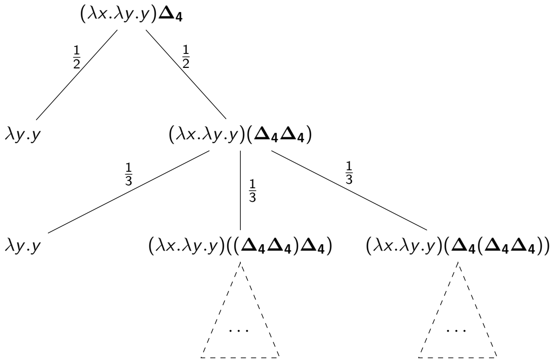
$$\underbrace{\begin{cases} 1 & \mathbf{a} \\ 0 & \mathbf{b} \end{cases}}_{\rho_0} \rightsquigarrow \underbrace{\begin{cases} \frac{1}{2} & \mathbf{a} \\ \frac{1}{2} & \mathbf{b} \end{cases}}_{\rho_1} \rightsquigarrow \underbrace{\begin{cases} \frac{1}{4} & \mathbf{a} \\ \frac{1}{4} & \mathbf{b} \end{cases}}_{\rho_2} \rightsquigarrow \cdots \rightsquigarrow \underbrace{\begin{cases} \frac{1}{2^k} & \mathbf{a} \\ \frac{1}{2^k} & \mathbf{b} \end{cases}}_{\rho_k} \rightsquigarrow \cdots$$

Termination results

- $\lim_{n \to +\infty} |\rho_n| = \lim_{n \to +\infty} \frac{1}{2^{n-1}} = 0 \Rightarrow$ **AST**.
- $\mathsf{Steps}_{\mathsf{P}}(\mathbf{a}) = \sum_{n=1}^{\infty} |\rho_n| = \sum_{n=1}^{\infty} \frac{1}{2^{n-1}} = 2 \Rightarrow$ **PAST**

The uniform strategy is not PAST

Let $\mathbf{\Delta_2} = \lambda x.(xx)(xx)$, $\mathbf{\Delta_4} = \mathbf{\Delta_2}\mathbf{\Delta_2}$ and $M = (\lambda x.\lambda y.y)\mathbf{\Delta_4}$.
Since $\mathbf{\Delta_4} \longrightarrow_\beta \mathbf{\Delta_4}\mathbf{\Delta_4}$, reducing under U leads to the following tree:



$$\text{Steps}_{\text{U}}(M) = \sum_{i=1}^{\infty} |\rho_i| = \sum_{i=1}^{\infty} \frac{1}{i} = +\infty \Rightarrow (\Lambda_{WN}, \text{U}) \text{ is } \textbf{not} \text{ PAST}.$$

Why U is not PAST?

Because there is no **lower bound** on the probability of picking the LO-redex, the one that assures **normalisation**.
Consider again the term $M_0 = (\lambda x.\lambda y.y)\mathbf{\Delta_4}$ and a reduction sequence of length n.

$$
\begin{array}{ccccccc}
M_0 & \longrightarrow_\beta & M_1 & \longrightarrow_\beta & \cdots & \longrightarrow_\beta & M_n \\
\big\downarrow{\scriptstyle \frac{1}{2}} & & \big\downarrow{\scriptstyle \frac{1}{3}} & & & & \big\downarrow{\scriptstyle \frac{1}{n+2}} \\
\lambda y.y & & \lambda y.y & & & & \lambda y.y
\end{array}
$$

As $n \to +\infty$, the probability of picking the LO-redex goes to zero.

Is the uniform strategy AST? Open problem.

A proof method for PAST

For $\varepsilon > 0$ we write $x >_\varepsilon y$ if and only if $x \geq y + \varepsilon$.

Definition

Given an FPARS (S, P), we define a function $V : S \to \mathbb{R}$ as Lyapunov if:

1. There exists $b \in \mathbb{R}$ such that $V(s) \geq b$ for each $s \in S$.
2. There exists $\varepsilon > 0$ such that for every $s \in S$ if $\mathrm{P}(s) = \mu$, then $V(s) >_\varepsilon V(\mu)$, where V is extended to partial distributions as follows:
$$V(\mu) = \sum_{t \in S} V(t) \cdot \mu(t).$$

Theorem ([Foster '53])

If we can define for an FPARS $\mathcal{P} = (S, \mathrm{P})$ a Lyapunov function V, then $\mathcal{P}$ is PAST and $\mathrm{Steps}_{\mathrm{P}}(s)$ is bounded by $\frac{V(s)}{\varepsilon}$.

$(\Lambda_{\mathsf{WN}}, \mathsf{P}_\varepsilon)$ is PAST

Can we exploit Foster criterion to prove that a strategy is PAST?

Theorem

Each FPARS $(\Lambda_{\mathsf{WN}}, \mathsf{R}_\varepsilon)$ *is PAST if* R_ε *reduces the* LO*-redex with probability* $\varepsilon > 0$*.*

Proof.

$\mathsf{Steps}_{\mathsf{LO}}$ is a Lyapunov function for $(\Lambda_{\mathsf{WN}}, \mathsf{R}_\varepsilon)$ □

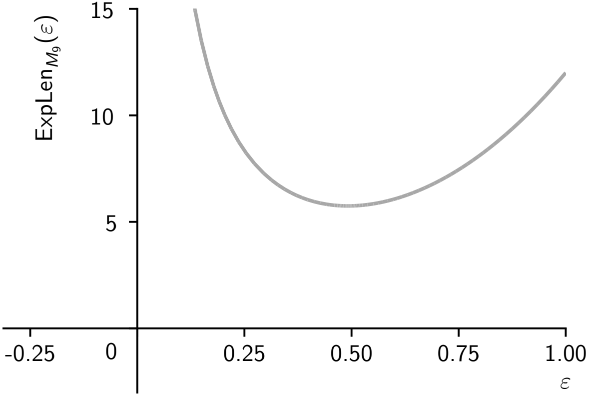
That means that reducing a term with a strategy R_ε brings to normal form in a **finite** number of steps, in **average**. Thus **efficiency** of any strategy R_ε should be compared to the one of LO.

We consider now a very simple strategy which reduces with probability $1 - \varepsilon$ the RI-redex. We call it P_ε.

P_ε is non trivial

Can P_ε be **more efficient**, in average, than LO?
Let us consider a family of terms $M_n = NL_n$ where:

$$N = \lambda x.\underbrace{((\lambda y.z)\boldsymbol{\Omega})}_{P}x \qquad L_n = C_n\underbrace{((\lambda x.x)y)}_{S} \qquad C_n = \lambda x.\underbrace{xx\cdots x}_{n \text{ times}}$$



Experimental work

We have tried to find out which terms benefit from **randomisation**.We adopted an **experimental** approach, developing a **tool** that ouputs $\mathrm{ExpLen}_M(\varepsilon)$ for a given M.

E.g. let $N_{n,m} = ((\lambda y.z)\mathbf{I}^m)(C_n(\mathbf{I}y))$. Where is worth randomising?

$m \backslash n$	**1**	**2**	**3**	**4**	**5**
1	cost	RI	RI	RI	RI
2	LO	P_ε	P_ε	P_ε	P_ε
3	LO	P_ε	P_ε	P_ε	P_ε
4	LO	P_ε	P_ε	P_ε	P_ε
5	LO	P_ε	P_ε	P_ε	P_ε

In all non-trivial cases it is worth randomising!

Analysis of ExpLen(ε)

We would want to analyze the shape of ExpLen(ε). In particular we are interested in its **minima**.If $M \longrightarrow_{\mathrm{LO}} N$ and $M \longrightarrow_{\mathrm{RI}} L$ then

$$\mathrm{ExpLen}_M(\varepsilon) = \varepsilon \cdot \mathrm{ExpLen}_N(\varepsilon) + (1-\varepsilon) \cdot \mathrm{ExpLen}_L(\varepsilon) + 1.$$

Differentiating with respect to ε we have

$$\begin{aligned}\mathrm{ExpLen}'_M(\varepsilon) = & \,\mathrm{ExpLen}_N(\varepsilon) - \mathrm{ExpLen}_L(\varepsilon) + \\ & + \varepsilon \cdot \mathrm{ExpLen}'_N(\varepsilon) + (1-\varepsilon) \cdot \mathrm{ExpLen}'_L(\varepsilon).\end{aligned}$$

By induction we are able to prove **monotonicity** for sub-λ-calculi λA and λI, looking only at the **sign** of $\mathrm{ExpLen}_N(\varepsilon) - \mathrm{ExpLen}_L(\varepsilon)$.

Theorem

In λA ExpLen(ε) is monotonically decreasing.
In λI ExpLen(ε) is monotonically increasing.

Future works

This was the first study on **randomised strategies**. Many questions remain **open**.

- Are there better randomised strategies?
- How to tune ε?
- What about the real world?