

Supermartingales and Termination Analysis of Statistical PCF

August 30, 2020

[Tentative title] ◀ L0

Abstract

[TODO] ◀ L0

Contents

1 Statistical PCF	2
1.1 Syntax of SPCF	2
1.2 Sampling semantics	3
2 Alternative Semantics	4
3 Ranking functions	17
4 Supermartingales	18
5 Constructing Ranking Functions	23
5.1 Partial ranking functions	24
5.2 Examples	27
6 Antitone ranking functions	28
7 Y stopping times are finite	32
8 Relativised completeness	33

Various theorems about probabilistic programs rely on the assumption that the program terminates almost surely. Proof rules based on relating the program state to supermartingales already exist for first-order imperative programs [McIver et al. 2018]. This paper’s contribution is to extend this method to a higher-order setting.

1 Statistical PCF

1.1 Syntax of SPCF

The language SPCF is a simply-typed lambda calculus with sampling of real numbers from $[0, 1]$ and unbounded scoring, following Mak et al. [2020a]. Types and terms are defined as follows, where r is a real number, x is a variable, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is any measurable function, and Γ is an environment:

types $A, B ::= \mathbb{R} \mid A \rightarrow B$

values $v ::= \lambda x.s \mid \underline{r}$

terms $s, t ::= v \mid x \mid t_1 t_2 \mid \underline{f}(s_1, \dots, s_n) \mid \mathbf{Y}s \mid \text{if}(s < 0, t_1, t_2) \mid \text{sample} \mid \text{score}(s)$

$$\frac{}{\Gamma; x : A \vdash x : A} \quad \frac{\Gamma; x : A \vdash s : B}{\Gamma \vdash \lambda x.s : A \rightarrow B} \quad \frac{}{\underline{r} : \mathbb{R}} \quad \frac{\Gamma \vdash s : A \rightarrow B \quad \Gamma \vdash t : A}{\Gamma \vdash s t : B}$$

$$\frac{\Gamma \vdash s_1 : \mathbb{R} \dots \Gamma \vdash s_n : \mathbb{R} \quad f : \mathbb{R}^n \rightarrow \mathbb{R}}{\Gamma \vdash \underline{f}(s_1, \dots, s_n) : \mathbb{R}} \quad \frac{\Gamma \vdash s : (A \rightarrow B) \rightarrow (A \rightarrow B)}{\Gamma \vdash \mathbf{Y}s : (A \rightarrow B)}$$

$$\frac{\Gamma \vdash c : \mathbb{R} \quad \Gamma \vdash s_1 : A \quad \Gamma \vdash s_2 : A}{\Gamma \vdash \text{if}(c < 0, s_1, s_2) : A} \quad \frac{}{\Gamma \vdash \text{sample} : \mathbb{R}} \quad \frac{\Gamma \vdash s : \mathbb{R}}{\Gamma \vdash \text{score}(s) : \mathbb{R}}$$

The set of all terms is denoted Λ , and the set of closed terms is denoted Λ^0 .

To define the reduction relation, let *evaluation contexts* be of the form:

$$E ::= [] \mid E t \mid v E \mid \underline{f}(r_1, \dots, r_{k-1}, E, s_{k+1}, \dots, s_n) \mid \mathbf{Y}E \mid \text{if}(E < 0, s_1, s_2) \mid \text{score}(E)$$

then a term reduces if it is formed by substituting a redex in a context i.e.

$$\begin{aligned} E[(\lambda x.s)v] &\rightarrow E[s[v/x]] \\ E[\underline{f}(\underline{r}_1, \dots, \underline{r}_n)] &\rightarrow E[f(r_1, \dots, r_n)] \\ E[\mathbf{Y}\lambda x.s] &\rightarrow E[\lambda z.s[(\mathbf{Y}\lambda x.s)/x]z] \text{ where } z \text{ is not free in } s \\ E[\text{if}(\underline{r} < 0, s_1, s_2)] &\rightarrow E[s_1] \text{ where } r < 0 \\ E[\text{if}(\underline{r} < 0, s_1, s_2)] &\rightarrow E[s_2] \text{ where } r \geq 0 \\ E[\text{sample}] &\rightarrow E[\underline{r}] \text{ where } r \in [0, 1] \\ E[\text{score}(\underline{r})] &\rightarrow E[r]. \end{aligned}$$

We write \rightarrow^* for the reflexive, transitive closure of \rightarrow . [The **score** rule should have a side condition: if $r > 0$.] ↔ L0 ↵ L0

Every closed well-typed term either is a value or reduces to another term.

[NOTATION. A minor point: metavariables for terms are s, t , etc. up to this point, but M, N , etc. here after all.] ↵ L0

1.2 Sampling semantics

This version of the reduction relation allows **sample** to reduce to any number in $[0, 1]$. To more precisely specify the probabilities, an additional argument is needed to determine the outcome of random samples. Let $I = [0, 1] \subset \mathbb{R}$, and let $S = I^{\mathbb{N}}$, with the **Borel σ -algebra \mathcal{F}** and the probability measure, denoted μ , given by the limit of $1 \leftarrow I \leftarrow I^2 \leftarrow \dots$, where the maps are the projections that ignore the last element. Equivalently, a basis of measurable sets is $\prod_{i=0}^{\infty} X_i$ where X_i are all Borel and all but finitely many are I , and $\mu(\prod_{i=0}^{\infty} X_i) = \prod_{i=0}^{\infty} \text{Leb}(X_i)$. The maps $\pi_h : S \rightarrow I$, $\pi_t : S \rightarrow S$ popping the first element are then measurable. **Following Culpepper and Cobb [2017], we call the probability space (S, \mathcal{F}, μ) the *entropy space*.**

The σ -algebra and measure on Λ are defined by considering it as a disjoint union of equivalence classes under replacing all the real constants by a placeholder, where the measure on each class is that of \mathbb{R}^n , where n is the number of real constants. **Precisely, following Borgström et al. [2016], we view Λ as $\bigcup_{m \in \omega} (\text{Sk}_m \times \mathbb{R}^m)$, where Sk_m is the set of SPCF terms with exactly m place-holders (a.k.a. *skeleton terms*) for numerals. Thus identified, we give Λ the countable disjoint union topology of the product topology of the discrete topology on Sk_m and the standard topology on \mathbb{R}^m . Note that the connected components of Λ have the form $\{M\} \times \mathbb{R}^m$, with M ranging over Sk_m , and m over ω . We fix the Borel algebra of this topology to be the σ -algebra on Λ .**

The one-step reduction is given by the function $\text{red} : \Lambda^0 \times S \rightarrow \Lambda^0 \times S$ where

$$\text{red}(M, s) = \begin{cases} (E[N], s) & \text{if } M = E[R], R \rightarrow N \text{ and } R \neq \text{sample} \\ (E[\pi_h(s)], \pi_t(s)) & \text{if } M = E[\text{sample}] \\ (M, s) & \text{if } M \text{ is a value} \end{cases} \quad (1)$$

[Both \rightarrow and red are called reduction relation.]

The result after n steps is then simply $\text{red}^n(M, s) = \text{red}(\dots \text{red}(M, s) \dots)$, and the limit red^∞ can then be defined as a partial function as $\lim_{n \rightarrow \infty} \text{red}^n(M, s)$ whenever that sequence becomes constant by reaching a value. A term M terminates for a sample sequence s if the limit $\text{red}^\infty(M, s)$ is defined.

The reduction function is measurable, and the set of values is measurable, therefore the set of s such that M terminates at s within n steps is measurable for any n , therefore $\{s \mid M \text{ terminates at } s\}$ is measurable. A term M is said to terminate almost surely if $\mu(\{s \mid M \text{ terminates at } s\}) = 1$.

[Alternatively, define the *runtime of M* to be the random variable

$$T_M(s) := \begin{cases} \min\{n \mid \pi_0(\text{red}^n(M, s)) \text{ is a value}\} & \text{if } \text{red}^\infty(M, s) \text{ is defined} \\ \infty & \text{otherwise} \end{cases}$$

Equivalently, we say that M is *almost-surely terminating* (AST) if $T_M < \infty$ a.s.; and M is *positively almost-surely terminating* (PAST) if $\mathbb{E}[T_M] < \infty$.

For example, the term

$$(\mathbf{Y}\lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1))) \underline{0},$$

which generates a geometric distribution, terminates on the set $S \setminus [0.5, 1]^{\mathbb{N}}$, which has measure 1, therefore it terminates almost surely, whereas

$$\text{if}(\text{sample} - 0.5 < 0, \underline{0}, (\mathbf{Y}\lambda x.x)\underline{0}),$$

which terminates on the set $\pi_h^{-1}[[0, 0.5]]$, has probability 0.5 of failing to terminate.

Note that the requirement that M be closed is not important for the proofs of the theorems, but is simply required to ensure that M reaches a value, rather than a general normal form like xy .

This definition of almost sure termination is equivalent to that given in [Mak et al. 2020a], although the program semantics is stated in a slightly different way. In particular, the argument to `score` is not relevant to termination (except for the possibility that its argument's evaluation wouldn't terminate).

[Your operational semantics does not maintain a record of the current weight of the reduction. A.s. termination does depend on `score`: see [Mak et al. 2020b, §4.3]¹. I think it important to take the behaviour of `score` into account; you should do it as a future task.] ◀ L0

2 Alternative Semantics

[This version of the semantics is much more convenient for allowing multiple different reduction orders. I don't currently have a proof of the equivalence of this with the more usual semantics. I expect that that would be somewhat complicated, but not terribly difficult. I don't know whether a semantics like this has already been defined elsewhere. I just thought I'd write it out so you can see it and in case I use it later. If I do end up actually using it I'll write it up nicer, but hopefully this is at least enough for you to understand what I mean.] ◀ AK-R

Having thought a bit more about this, I realise it may not be quite right. I'll see if I can fix it later.] [I'll read this after you fix it ;-)] ◀ L0

When proving almost sure termination in this way [I guess I should move this section somewhere later.], it is necessary to consider which terms a given term may reduce to. Sometimes however, the reduction that the programmer has in mind may not be strictly the call-by-value order defined so far, or considering an alternative reduction order may be simpler or more intuitive. ◀ AK-R

Non-probabilistic lambda calculi generally have the Church-Rosser property, that if a term A reduces to both B_1 and B_2 , there is some C to which both B_1

¹<https://arxiv.org/abs/2004.03924>

and B_2 reduce, so the reduction order mostly doesn't matter. In the probabilistic case, this may not be true, because β -reduction can duplicate **samples**, so the outputs of the copies of the sample may be identical or independent, depending on whether the sample is taken before or after β -reduction. There are, however, some restricted variations on the reduction order that do not have this problem.

Even with this restriction, a sampling semantics in the style of the one already defined would not be entirely Church-Rosser, as, for example, $\text{red}^3(\text{sample} - \text{sample}, (1, 0, \dots))$ would be either 1 or -1 depending on the order of evaluation of the **samples**, as that determines which sample from the pre-selected sequence is used for each one. To fix this, rather than pre-selecting samples according to the order they'll be drawn in, select them according to the position in the term where they'll be used instead.

A *position* is a finite sequence of steps into a term, defined inductively as

$$P ::= \cdot \mid \lambda; P \mid @_1; P \mid @_2; P \mid \underline{f}_i; P \mid Y; P \mid \text{if}_1; P \mid \text{if}_2; P \mid \text{if}_3; P \mid \text{score}; P.$$

The *subterm of M at a position P* , denoted $M \mid P$, is defined as

$$\begin{aligned} M \mid \cdot &= M \\ \lambda x.M \mid \lambda; P &= M \mid P \\ M_1 M_2 \mid @_i; P &= M_i \mid P \quad \text{for } i = 1, 2 \\ \underline{f}(M_1, \dots, M_n) \mid \underline{f}_i; P &= M_i \mid P \quad \text{for } i \leq n \\ YM \mid Y; P &= M \mid P \\ \text{if}(M_1 < 0, M_2, M_3) \mid \text{if}_i; P &= M_i \mid P \quad \text{for } i = 1, 2, 3 \\ \text{score}(M) \mid \text{score}; P &= M \mid P \end{aligned}$$

so that every subterm is located at a unique position, but not every position corresponds to a subterm (e.g. $xy \mid \lambda; \cdot$ is undefined). A position such that $M \mid P$ does exist is said to *occur* in M . *Substitution* of N at position P in M , written $M[N/P]$, is defined similarly. For example, let

$$M = \lambda x y.y(\text{if}(x < 0, y(\underline{f}(x)), 3)) \quad P = \lambda; \lambda; @_2; \text{if}_2; @_2$$

then $M[\text{sample}/P] = \lambda x y.y(\text{if}(x < 0, y \text{sample}, 3))$.

Two subterms N_1 and N_2 of a term M , corresponding to positions P_1 and P_2 , can overlap in a few different ways. If P_1 is an initial segment (i.e. prefix) of P_2 (written as $P_1 \leq P_2$), then N_2 is also a subterm of N_1 . If neither $P_1 \leq P_2$ nor $P_2 \leq P_1$, the positions are said to be disjoint. This is mostly relevant in that in any substitution at some position, the subterms at all other positions are unaffected.

With this notation, a more general reduction relation \rightarrow can be defined. [\[It would almost work to have one of the neater versions of \$Y\$ -reduction here.\]](#) \triangleleft AK-R

Definition 1. The binary relation \rightarrow is defined by the following rules, each is conditional on a redex occurring at position P in the term M :

- if $M \mid P = (\lambda x.s)v$, $M \rightarrow M[s[v/x]/P]$
- if $M \mid P = \underline{f}(r_1, \dots, r_n)$, $M \rightarrow M[\underline{f}(r_1, \dots, r_n)/P]$
- if $M \mid P = \Upsilon \lambda x.s$, $M \rightarrow M[\lambda z.s[(\Upsilon \lambda x.s)/x]z/P]$ where z is not free in s
- if $M \mid P = \text{if}(\underline{r} < 0, s_1, s_2)$, $M \rightarrow M[s_1/P]$ where $r < 0$
- if $M \mid P = \text{if}(\underline{r} < 0, s_1, s_2)$, $M \rightarrow M[s_2/P]$ where $r \geq 0$
- if $M \mid P = \text{sample}$ and λ does not occur after $@_2$ or Υ in P ,
 $M \rightarrow M[\underline{r}/P]$ where $r \in [0, 1]$
- if $M \mid P = \text{score}(r)$, $M \rightarrow M[r/P]$.

In each of these cases, $M \mid P$ is the *redex*, and reduction takes place at P .

Labelling the pre-chosen samples by the positions in the term would also not work because in some cases, a **sample** will be duplicated before being reduced, for example, in $(\lambda x.x \ 0 \pm x \ 0)(\lambda y.\text{sample})$, both of the **sample** redexes that eventually occur originate at $@_2; \lambda$. It is therefore necessary to consider possible positions that may occur in other terms reachable from the original term. Even this is itself inadequate because some of the positions in different reachable terms need to be considered the same, and the number of reachable terms is in general uncountable, which leads to measure-theoretic issues.

Define a *skeleton* to be a term but, instead of having real constants \underline{r} , it has placeholders X , so that each term has a skeleton and each skeleton can be converted to a term given the appropriate number of real numbers to substitute in. [Skeleton is defined in Sec. 1.2.] The positions in a skeleton and the reduction relation on skeletons can be extended from the definitions on terms in the obvious way, with $\text{if}(X < 0, A, B)$ reducing nondeterministically to both A and B , and $M \rightarrow M[X'/P]$ for a freshly generated placeholder X' , in case $M \mid P = \text{sample}$ and λ does not occur after $@_2$ or Υ in P . (By abuse of notation, we use the same metavariables M, N , etc. for skeletons as for terms.) For example, we have $(\lambda x.\text{if}(x < 0, x, X_1)) \text{sample} \rightarrow (\lambda x.\text{if}(x < 0, x, X_1)) X_2 \rightarrow \text{if}(X_2 < 0, X_2, X_1) \rightarrow X_i$ where i is 1 or 2.

Given a closed term M , let $L_0(M)$ be the set of pairs, the first element of which is a \rightarrow -reduction sequence of skeletons starting at the skeleton of M , and the second of which is a position in the final skeleton of the reduction sequence. [I've yet to go through and replace references to terms with references to skeletons, but it should be reasonably obvious where the substitution should take place. This is potentially even more confusing though because I'm now implicitly identifying terms, skeletons, reduction sequences and skeletal reduction sequences.] Reduction sequences are used rather than reachable skeletons because if the same skeleton is reached twice, different samples may be needed, [Give a simple example illustrating this.] but they will be discussed as though they were terms, by abuse of notation.

[Readers would benefit greatly from some examples of an $L_0(M)$ -pair, and the relations \sim_p and \sim_c , here, before the formal definitions. Explain the underlying intuition.] < L0

Definition 2. The relation \sim is defined as the union of the symmetric relations \sim_p (p for parent-child) and \sim_c (c for cousin) where [As defined, \sim_p is not a symmetric relation.] < L0
= L0

1. If N reduces to O with the redex at position P (i.e. $M \rightarrow^* N \rightarrow O = N[\Delta/P]$ with $O \mid P = \Delta$ for some redex contraction $\Delta \rightarrow \Delta$), and Q is a position in N disjoint from P , then $(N, Q) \sim_p (O, Q)$. < L0
= L0
2. If N β -reduces to O at position P , Q is a position in $N \mid P; @_1; \lambda$ and $N \mid P; @_1; \lambda; Q$ is not the variable involved in the reduction, $(N, P; @_1; \lambda; Q) \sim_p (O, P; Q)$
3. If N if-reduces to O at position P , with the first resp. second branch being taken, and $P; \text{if}_i; Q$ occurs in N (where $i = 2$ resp. 3), $(N, P; \text{if}_i; Q) \sim_p (O, P; Q)$
4. If N , O_1 and O_2 match any of the following cases:
 - (a) N contains redexes at disjoint positions P_1 and P_2 , O_1 is N reduced first at P_1 then P_2 and O_2 is N reduced first at P_2 then at P_1 .
 - (b) $N \mid P = \text{if}(\underline{r} < 0, N_1, N_2)$, where $r < 0$ (or, respectively, $r \geq 0$), $(N_2 \text{ resp. } N_1) \mid Q$ is a redex, and O_1 is N reduced at P and O_2 is N reduced first at $P; (\text{if}_3 \text{ resp. } \text{if}_2); Q$ then at P
 - (c) $N \mid P = \text{if}(\underline{r} < 0, N_1, N_2)$, where $r < 0$ (or, respectively, $r \geq 0$), $(N_1 \text{ resp. } N_2) \mid Q$ is a redex, and O_1 is N reduced first at P then at $P; Q$ and O_2 is N reduced first at $P; (\text{if}_2 \text{ resp. } \text{if}_3); Q$ then at P
 - (d) $N \mid P = (\lambda x.A)B$, there is a redex in A at position Q , O_1 is N reduced first at P then at $P; Q$, and O_2 is N reduced first at $P; @_1; \lambda; Q$ then at P
 - (e) $N \mid P = (\lambda x.A)B$, $B \mid Q$ is a redex, $(R_i)_i$ is a list of all the positions in A where $A \mid R = x$, ordered from left to right, O_1 is N reduced first at $P; @_2; Q$ then at P , and O_2 is N reduced first at P then at $P; R_i; Q$ for each i in order.
 - (f) $N \mid P = Y(\lambda x.A)$, A reduced at Q is A' , $(R_i)_i$ is a list of all the positions where $A' \mid R = x$, ordered from left to right, O_1 is N reduced first at $P; Y; \lambda; Q$ then at P , and O_2 is N reduced first at P then at $P; \lambda; @_1; Q$ then at $P; \lambda; @_1; R_i; Y; \lambda; Q$ for each i in order.

(in which case O_1 and O_2 are equal as terms, but with different reduction sequences), O'_1 and O'_2 are the results of applying some reduction sequence to each of O_1 and O_2 (the same reductions in each case, which is always possible because they're equal terms), and S is a position in O'_1 (or equivalently O'_2), then $(O'_1, S) \sim_c (O'_2, S)$.

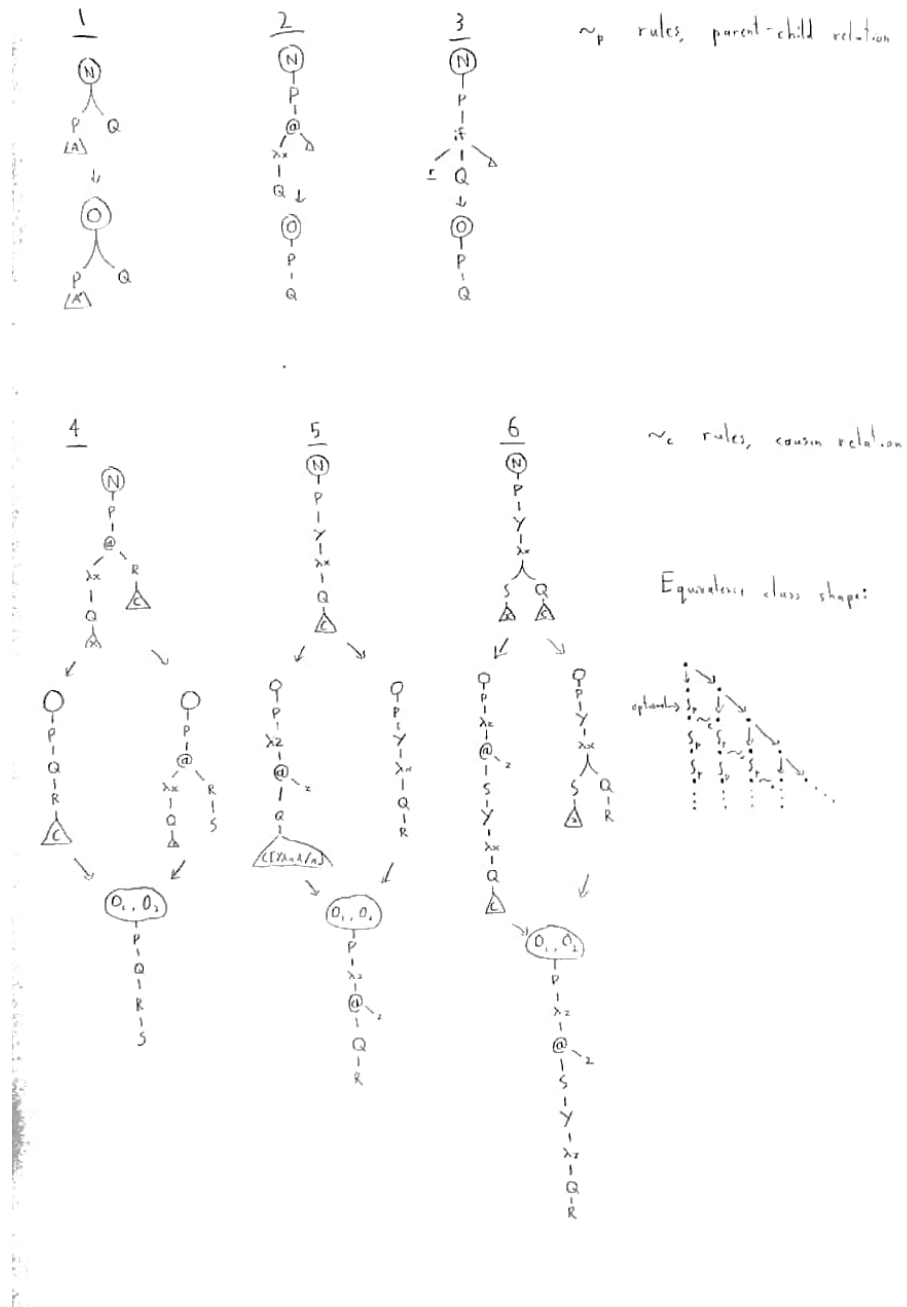


Figure 1: Some examples of the old version of the \sim_p - and \sim_c -rules

See Figure 1 for a graphical depiction of \sim_p - and \sim_c -rules.

[TODO: Typeset Fig. 1 in vector graphics using PGF/TikZ.]

◁ L0

The reflexive transitive closure \sim^* of this relation is used to define the set of *potential positions* $L(M) = L_0(M) / \sim^*$, and each equivalence class can be considered as the same position as it may occur across multiple reachable *skeletons*.

↔ L0

If $(N, P) \sim^* (O, Q)$, then $N \mid P$ and $O \mid Q$ both have the same shape (i.e. they're either *both placeholders*, both variables, both applications, both *samples* etc.), therefore it's well-defined to talk of the set of potential positions where there is a *sample*, $L_s(M)$. Formally $L_s(M) := \{[(X, P)]_{\sim_M^*} : X \mid P = \text{sample}\}$. The new

↔ L0

↔ L0

sample space is then defined as $I^{L_s(M)}$, with the Borel σ -algebra and product measure. Since $I^{L_s(M)}$ is a countable product, the measure space is well-defined [Ash and Doléans-Dade 2000, Cor. 2.7.3].

↔ L0

[For now, here's a summary of the lemmas and theorems I intend to prove about this alternative semantics:

◁ AK-R

- If $M \rightarrow N$, the relation \sim_N^* defined on $Rch(N)$ is equal to the restriction of \sim_M^* to $Rch(N)$
- Given the definition below of \Rightarrow , which is still applicable to this version of \sim , and which is the equivalent of red in this version of the semantics, \Rightarrow is Church-Rosser.
- If $M \rightarrow N$ at P , no position in any term in $Rch(N)$ is related to (M, P) by \sim_M^* , and therefore each sample is used at most once in any reduction sequence.
- The subprobability distribution of values obtained by reducing a term in the usual call-by-value semantics is equal to that obtained by any reduction strategy on \Rightarrow (i.e. a sub-relation of \Rightarrow that is a function) (with some allowance made for the fact that some values are not normal forms of \Rightarrow , so the reduction strategy might overshoot the value that red^∞ reaches. I'm not quite sure what form this will take.)
- Similarly, the probability of termination is equal in red^∞ and any reduction strategy on \Rightarrow , so almost sure termination can be equivalently defined based on \Rightarrow .
- For any closed term N , if there is any reduction strategy for N and ranking function corresponding to that reduction strategy, N is AST.
- Corresponding sparse and antitone-strict ranking function theorems.

]

Before defining the new version of the reduction relation red , a few lemmas about the properties of \sim are necessary for it to be well-defined.

Lemma 1. *If $(N_1, P_1) \sim^* (N_2, P_2)$, there is some descendants N'_1, N'_2 of N_1 resp. N_2 , and a position P' in both of them such that $(N_1, P_1) \sim_p^* (N'_1, P') \sim_c^* (N'_2, P') \sim_p^* (N_2, P_2)$.*

Proof. The \sim relation can be split into $\sim_\downarrow \cup \sim_\uparrow$, where $(A, P_A) \sim_\downarrow (B, P_B)$ if either $A \rightarrow B$ and $(A, P_A) \sim_p (B, P_B)$ or $(A, P_A) \sim_c (B, P_B)$, and similarly $(A, P_A) \sim_\uparrow (B, P_B)$ if either $B \rightarrow A$ and $(A, P_A) \sim_p (B, P_B)$ or $(A, P_A) \sim_c (B, P_B)$. The sequence of \sim relations linking N_1 to N_2 can then be expressed as an alternating sequence of \sim_\downarrow^* and \sim_\uparrow^* s, starting with \sim_\downarrow^* and ending with \sim_\uparrow^* , and the proof proceeds by induction on the number of these alternations.

In the base case, that $(N_1, P_1) \sim_\downarrow^* X \sim_\uparrow^* (N_2, P_2), \dots$

[Actually proving that this process terminates will be a pain although I'm confident it will work, but the basic idea is that the sequence of \sim steps can be altered by repeatedly replacing the subsequence $\sim_c, \sim_{p,\downarrow}$ with $\sim_{p,\downarrow}, \sim_c$, the subsequence $\sim_{p,\uparrow}, \sim_c$ with $\sim_c, \sim_{p,\uparrow}$, and the subsequence $\sim_{p,\uparrow}, \sim_{p,\downarrow}$ with $\sim_{p,\downarrow}^*, \sim_c, \sim_{p,\uparrow}^*$, until everything is in the correct order.] ◀ AK-R

If $(X, P) \sim_c (Y, P)$ then the final skeletons in X and Y are equal, and $(X, Q) \sim_c (Y, Q)$, therefore \sim_c can be considered to apply to reduction sequences even without positions, $X \sim_c Y$ iff $(X, \cdot) \sim_c (Y, \cdot)$.

Lemma 2. *For any reduction sequence X , there is a unique reduction sequence X_{cbv} that is related to X by \sim_c (for each position in the final term in X) and is in call-by-value order, and if $X \sim_c^* Y$, then $X_{cbv} = Y_{cbv}$.*

[As defined, \sim_c is not symmetric.] ◀ L0

Proof. Note that although the reductions that occur in X_{cbv} are in call-by-value order, it may not be the actual call-by-value reduction sequence because some redexes may remain un-reduced even when other reductions that should happen afterwards occur. It is only required that the reductions that do occur occur in the correct order. [This paragraph (especially the last sentence) is ambiguous. What does it mean to say that that a reductions that occur in a reduction sequences are in CBV order?] ◀ L0

Proof sketch: Starting at X , take the last pair of adjacent reductions which aren't in CBV order, and switch them (with \sim_c). Repeat this process until CBV order is reached.

If $Y \sim_c X \sim_c^* X_{cbv}$, and the reductions involved in $Y \sim_c X$ are not the last pair of reductions in Y that aren't in CBV order, the order of the \sim_c s can be rearranged so that the sequence $Y \sim_c^* X_{cbv}$ is the canonical sequence given above, therefore $Y_{cbv} = X_{cbv}$.

If Y is in CBV order and $Y \sim_c^* X$, then $Y_{cbv} = X_{cbv}$ but also $Y_{cbv} = Y$ therefore X_{cbv} is unique. ◻

Lemma 3. *The relation \sim is defined on $L_0(M)$ with reference to a particular starting term M , so different versions, \sim_M and \sim_N , can be defined starting at different terms. If $M \rightarrow N$, then \sim_N^* is equal to the restriction of \sim_M^* to $L_0(N)$.*

Proof. \sim_N^* is trivially a subset of \sim_M^* because \sim_N is a subset of \sim_M .

In the other direction, suppose $(N_1, P_1) \sim_M^* (N_2, P_2)$ where both N_1 and N_2 are descendants of N . By Lemma 1, take $(N_1, P_1) \sim_{p,M}^* (N'_1, P') \sim_{c,M}^* (N'_2, P') \sim_{p,M}^* (N_2, P_2)$. The \sim_p^* steps remain within $Rich(N)$, and \sim_p does not depend on the history of the reduction sequences, therefore $(N_1, P_1) \sim_{p,N}^* (N'_1, P') \sim_{c,M}^* (N'_2, P') \sim_{p,N}^* (N_2, P_2)$.

Let X be the call-by-value reduction sequence related to both N'_1 and N'_2 by $\sim_{c,M}$ given by Lemma 2. As $M \rightarrow N$ is the first reduction in the sequence N'_1 , it is the last to be affected by the $\sim_{c,M}$ sequence $N'_1 \sim_{c,M}^* X$ given by Lemma 2 therefore it can be split into $N'_1 \sim_{c,M}^* Y_1 \sim_{c,M}^* X$ where Y_1 is in CBV order except possibly for its first reduction, which is still $M \rightarrow N$. Let the position of the reduction $M \rightarrow N$ be Q . [TODO: fill in the details that are missing here. < AK-R
It shouldn't be too hard, it's just a bit sketchy at present.] The rearrangement of the reduction sequence $Y_1 \sim_{c,M}^* X$ consists of moving the reduction at Q down past the other reductions in Y_1 , possibly duplicating or deleting it in the process, but not affecting the positions or the correct order for any of the other reductions. The reductions derived from $M \rightarrow N$ can be identified as follows:

A position in some (reduction sequence of) term(s) is *derived from* another if it is related by the reflexive transitive closure of \rightsquigarrow , where $(V, R) \rightsquigarrow (W, T)$ just if $V \rightarrow W$ and one of the following cases holds:

- $(V, R) \sim_p (W, T)$
- $R = T$, $V \rightarrow W$ at A , and $A > R$
- $V \mid A = (\lambda x.U)Z, U \mid B = x, R = A; @_2; C$ and $T = A; B; C$
- $V \mid A = Y(\lambda x.U), R = A; Y; \lambda; B$ and $T = A; \lambda; @_1; B$
- $V \mid A = Y(\lambda x.U), U \mid B = x, R = A; C$ and $T = A; \lambda; @_1; B; C$

Crucially, the reductions in X can be partitioned into 2 sets: those at positions derived from (M, Q) , and those at positions equal to the reductions in Y_1 (in the same order as they occur in Y_1). Using the same construction for N'_2 shows that the positions of the reductions in Y_2 are also the positions of the reductions in X other than those derived from (M, Q) , therefore $Y_1 = Y_2$ therefore $N'_1 \sim_{c,M} Y_1 \sim_{c,M} N'_2$ and every reduction sequence in this sequence starts with $M \rightarrow N$ therefore $N'_1 \sim_{c,N} N'_2$.

Combining this with the $\sim_{p,N}^*$ at the beginning and end then yields the desired result that $(N_1, P_1) \sim_N^* (N_2, P_2)$, therefore the restriction of \sim_M^* to \sim_N 's domain ($Rich(N)$) is a subset of \sim_N^* therefore the two versions of \sim match on this domain, as desired. \square

At each reduction step $M \rightarrow N$, the sample space must be restricted from $I^{L_s(M)}$ to $I^{L_s(N)}$. The injection $L_0(N) \rightarrow L_0(M)$ is trivial to define by prepend- ↔ L0

ing $M \rightarrow N$ to each path, and using Lemma 3, this induces a corresponding injection on the quotient, $L(N) \rightarrow L(M)$. The corresponding map $L_s(N) \rightarrow L_s(M)$ is then denoted $i(M \rightarrow N)$.

Unlike in the purely call-by-value case, the version of the reduction relation that takes into account samples is still a general relation rather than a function, so it is denoted “ \Rightarrow ” instead of “red”, and it relates $\biguplus_{M \in \Lambda_0} I^{L_s(M)}$ to itself.

$(M, s) \Rightarrow (N, s \circ i(M \rightarrow N))$ if either

$M \rightarrow N$ and the redex is not **sample**, or

$M \mid P = \text{sample}$, $N = M[s(M, P)/P]$ and λ does not occur after $@_2$ or Y in P

Lemma 4. *The relation \Rightarrow is Church-Rosser.*

Proof. Suppose that $(M, s) \Rightarrow (M_1, s_1)$ and also $(M, s) \Rightarrow (M_2, s_2)$, then it is required to prove that there is some (M', s') such that both $(M_1, s_1) \Rightarrow^* (M', s')$ and $(M_2, s_2) \Rightarrow^* (M', s')$. Let the positions of the redexes in $M \rightarrow M_1$ and $M \rightarrow M_2$ be P_1 and P_2 respectively.

First consider the case that P_1 and P_2 are disjoint. Let $M_1 = M[X_1/P_1]$ and similarly for X_2 , then let $M' = M[X_1/P_1][X_2/P_2]$. As the positions are disjoint, the substitutions commute and both M_1 and M_2 reduce (with \rightarrow) to M' . Let $s' = s \circ i(M \rightarrow M_1) \circ i(M_1 \rightarrow M')$. The injection $i(M \rightarrow M_1) \circ i(M_1 \rightarrow M')$ consists of **prepending** $M \rightarrow M_1 \rightarrow M'$ to each reduction sequence, but by case 4.a of \sim , this is equivalent to **prepending** $M \rightarrow M_2 \rightarrow M'$, which is $i(M \rightarrow M_2) \circ i(M_2 \rightarrow M')$. In the case that the reduction $M \rightarrow M_2$ isn't a **sample**-reduction, this is enough to establish that $(M_1, s_1) \Rightarrow (M', s')$, and vice-versa. If it is **sample** though, in order for it to be the case that $(M_1, s_1) \Rightarrow (M', s')$, it is additionally necessary that X_2 , the result of the reduction at P_2 , be $s_1(M_1, P_2)$, which follows from the fact that $(M, P_2) \sim (M_1, P_2)$ by case 1. The case that the reduction $M \rightarrow M_1$ is a **sample**-reduction is similar. ↔ LO
↔ LO

The case that $P_1 = P_2$ is trivial, because there is at most one possible \Rightarrow reduction at any given position, therefore $(M_1, s_1) = (M_2, s_2)$ already.

The remaining case is that $P_1 < P_2$ or $P_1 > P_2$. Assume without loss of generality that $P_1 < P_2$. For each possible case of what type of redex $M|P_1$ is, and P_2 's position within it, there is a corresponding case of \sim_c , and similarly to the case where P_1 and P_2 are disjoint, the term $O_1 = O_2$ from the definition of \sim_c is a suitable value of M' . The term $M \mid P_1$ can't be **sample**, because it has strict subterms, but the case that $M \mid P_2 = \text{sample}$ is still somewhat more complicated. P_2 can't be within $P_1; @_2$ or $P_1; Y$ (cases e and f of \sim_c) because those are not valid positions to reduce a **sample**, and the case that P_2 is within the branch of an if statement that is deleted by the reduction at P_1 (case b) doesn't actually present a problem because there is no corresponding reduction to $M \rightarrow M_2$ in the other branch, which leaves cases c and d, that P_2 is in the $\text{if}(\cdot, \cdot, \cdot)$ branch that isn't deleted, and that $P_2 < P_1; @_1; \lambda$. The values that the **samples** take in these cases match because $(M, P) \sim_p (M_1, P; Q)$ by cases 3 and 2 of \sim_p respectively. \square

Lemma 5. *If A is some descendant of M and $(A, D) \sim^* (M, E)$, then $(A, D) \sim_p^* (M, E)$, with the length of the reduction sequences decreasing by one each step from A to M .*

Proof. This is a simple induction on \sim^* . In the base case, $(A, D) = (M, E)$ therefore $(A, D) \sim_p^* (M, E)$ trivially. Otherwise, suppose that $(M, E) \sim_p^* (B, F) \sim (A, D)$. Either $(B, F) \sim_p (A, D)$ or $(B, F) \sim_c (A, D)$. In the first case, either $B \rightarrow A$, in which case the result follows directly, or $A \rightarrow B$, in which case the fact that each (reduction sequence of) term(s) has only one parent implies that a $(M, E) \sim_p^* (A, D)$ directly as a subsequence of the path to (B, F) .

In the \sim_c case, consider the definition of \sim_c . Either $O'_1 = A$ and $O'_2 = B$ or vice-versa. As $M \rightarrow^* N \rightarrow^* O_1 \rightarrow^* B$ and $(M, E) \sim_p^* (B, F)$, and \sim_p only relates positions in a term and its parent, there are some positions G, H such that $(M, E) \sim_p^* (N, G) \sim_p^* (O_1, H) \sim_p^* (B, F)$. It follows that $(O_2, H) \sim_p^* (A, D)$ by following the same path, therefore it suffices to provide the only missing portion of the path from M to A , i.e. to prove that $(N, G) \sim_p^* (O_2, H)$ given $(N, G) \sim_p^* (O_1, H)$ (or vice-versa).

If G is disjoint from all the positions of reduction from N to O_1 and O_2 (and consequently $G = H$), this follows from case 1 of \sim_p . Otherwise, this can be proved by taking cases from the definition of \sim_c . This is rather long, but all of the cases are similar. The general idea is that the reductions from N to O_1 correspond to the reductions from N to O_2 , so that if a position is related by \sim_p across that reduction, it is related in the other branch for the same reason. Case d, where $B = O'_1$ rather than O'_2 , is given here in more detail as an illustrative example:

Let I be N reduced at P , and J be N reduced at $P; @_1; \lambda; Q$, so that $N \rightarrow I \rightarrow O_1$ and $N \rightarrow J \rightarrow O_2$. All of the reduction positions are $\geq P$, and G is not disjoint from all of them, therefore G is not disjoint from P . Let K be the position such that $(N, G) \sim_p (I, K) \sim_p (O_1, H)$. The fact that $(N, G) \sim_p (I, K)$ implies that $G > P; @_1; \lambda$. Let $G = P; @_1; \lambda; L$ and $H = P; L'$. If L is disjoint from Q , then $(N, G) \sim_p (J, G) \sim_p (O_2, P; L') = (O_2, H)$ by cases 1 and 2 of \sim_p . In the other case, that L isn't disjoint from Q , $L > Q$ because none of the positions $\leq P; @_1; \lambda; Q$ in N are related to any position in I by \sim_p . As $(I, P; L) \sim_p (O_1, P; L')$ (with the redex at $P; Q$), for exactly the same reason $(N, P; @_1; \lambda; L) \sim_p (J, P; @_1; \lambda; L')$ (with the redex at $P; @_1; \lambda; Q$). Because $(N, P; @_1; \lambda; L) \sim_p (J, P; @_1; \lambda; L')$, $N|P; @_1; \lambda; L = J|P; @_1; \lambda; L'$, and $N|P; @_1; \lambda; L \neq$ the variable of $N|P; @_1$, therefore $J|P; @_1; \lambda; L'$ is also not the variable therefore $(J, P; @_1; \lambda; L') \sim_p (O_2, P; L') = (O_2, H)$ by case 2 of \sim_p . Combining these results, $(N, G) \sim_p (O_2, H)$ as desired. \square

Lemma 6. *If $M \rightarrow N$, with the redex at position P , then no position in any term reachable from N is related by \sim^* to (M, P) .*

Proof. Suppose on the contrary that $(M, P) \sim^* (A, D)$, where $N \rightarrow^* A$, then by Lemma 5, $(M, P) \sim_p^* (A, D)$. $M \neq A$ therefore $(M, P) \sim_p$ some position in

N , but in all the cases of the definition of \sim_p , no position in the child term is related to the position of the redex. \square

The reduction relation \Rightarrow is nondeterministic, so it admits multiple possible reduction strategies. A *reduction strategy* starting from a closed term M is a partial function f from $Rich(M)$ to positions, such that for any reachable term N where f is defined, $f(N)$ is a position of a redex in N , and if $f(N)$ is not defined, N is a value. Using a reduction strategy f , a subset of \Rightarrow that isn't nondeterministic, \Rightarrow_f , can be defined by $(N, s) \Rightarrow_f (N', s')$ just if $(N, s) \Rightarrow (N', s')$ and N reduces to N' with the redex at $f(N)$.

Essentially what Lemma 6 demonstrates is that the samples taken during any reduction sequence are independent of each other. This is made more precise in the following lemmas.

Lemma 7. *For any skeletons $M \rightarrow N$, with the redex at position P , and measurable set of samples $S \subset I^{L_s(N)}$, $\mu(S) = \mu(\{s \in I^{L_s(M)} \mid s \circ i(M \rightarrow N) \in S\})$, and furthermore, if $M|P = \text{sample}$, for any $S \subset I \times I^{L_s(N)}$, $\mu(S) = \mu(\{s \in I^{L_s(M)} \mid (s(M, P), s \circ i(M \rightarrow N)) \in S\})$.*

Proof. If the result holds for all sets S of the form $\{s \in I^{L_s(N)} \mid \forall j : s(j) \in x_i\}$, where $(x_j)_{j \in J}$ is a family of measurable subsets of I indexed by some finite set $J \subset L_s(N)$ of positions, it also holds for all other S by taking limits and disjoint unions.

Take such a $(x_j)_{j \in J}$, where $x_j = x_k$ and define $K = i(M \rightarrow N)[J]$. Because $i(M \rightarrow N)$ is injective, it defines a bijection between J and K . We can then calculate the measure $\mu(\{s \in I^{L_s(M)} \mid s \circ i(M \rightarrow N) \in S\}) = \mu(\{s \in I^{L_s(M)} \mid \forall k : s(k) \in x_{i(M \rightarrow N)^{-1}(k)}\}) = \prod_{k \in K} \mu_I(x_{i(M \rightarrow N)^{-1}(k)}) = \prod_{j \in J} \mu_I(x_j) = \mu(S)$.

In the case that $M|P = \text{sample}$, we can similarly consider only those sets S of the form $x_{\text{sample}} \times \{s \in I^{L_s(N)} \mid \forall j : s(j) \in x_i\}$. The position P in M is not related to any position in $L_s(N)$, therefore $(M, P) \notin K$. Again, we can calculate the measure $\mu(\{s \in I^{L_s(M)} \mid (s(M, P), s \circ i(M \rightarrow N)) \in S\}) = \mu(\{s \in I^{L_s(M)} \mid s(M, P) \in x_{\text{sample}}, \forall k : s(k) \in x_{i(M \rightarrow N)^{-1}(k)}\}) = \mu_I(x_{\text{sample}}) \prod_{k \in K} \mu_I(x_{i(M \rightarrow N)^{-1}(k)}) = \mu_I(x_{\text{sample}}) \prod_{j \in J} \mu_I(x_j) = \mu(S)$. \square

Lemma 8. *For any initial term M , reduction strategy f on M , natural number n , skeleton N with k holes, measurable set $T \subset \mathbb{R}^k$ and measurable set S of samples in $I^{L_s(N)}$, $\mu(\{s \in I^{L_s(M)} \mid \exists r \in T, s' \in S : (M, s) \Rightarrow_f^n (N[r], s')\}) = \mu(\{s \in I^{L_s(M)} \mid \exists r \in T, s' \in I^{L_s(N)} : (M, s) \Rightarrow_f^n (N[r], s')\})\mu(S)$*

Proof. Suppose, to begin with, that $n = 0$. Either $\exists r \in T : N[r] = M$, in which case both sides of the equation are $\mu(S)$, or there is no such r , in which case both sides of the equation are 0.

For $n > 0$, suppose for induction that the lemma is true for $n - 1$, for all N, T and S . If $\exists r \in T, s' \in S : (M, s) \Rightarrow_f^n (N[r], s')$, the r and s' are necessarily unique, therefore we may assume that T is the product of k measurable setsets of \mathbb{R} , $(T_j)_{0 \leq j < k}$, as all of the other cases follow by taking unions and limits of

$$\begin{aligned} R_s &= \{\text{the skeleton of } O \mid O \in \text{Rch}(M), r \in \mathbb{R}^k, O|f(O) = \text{sample}, O \rightarrow N[r]\} \\ R_d &= \{\text{the skeleton of } O \mid O \in \text{Rch}(M), r \in \mathbb{R}^k, O|f(O) \neq \text{sample}, O \rightarrow N[r]\}. \end{aligned}$$
$$\begin{aligned}
& \mu(\{s \in I^{L_s(M)} \mid \exists r \in T, s' \in S : (M, s) \Rightarrow_f^n (N[r], s')\}) \\
&= \mu(\{s \in I^{L_s(M)} \mid \exists O \in R_s, r \in \mathbb{R}^{k-1}, s' \in I^{L_s(O)}, r' \in T, s'' \in S : (M, s) \Rightarrow_f^{n-1} (O[r], s') \Rightarrow_f (N[r'], s'')\}) + \mu(\{s \in I^{L_s(M)} \mid \exists O \in R_s, r \in \mathbb{R}^{k-1}, s' \in I^{L_s(O)} : g_O(r, s'(O, f(O))) \in T, s' \circ i(O[r] \rightarrow N[g_O(r, s'(O, f(O))))\}) \\
&= \sum_{O \in R_s} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \mathbb{R}^{k-1}, s' \in I^{L_s(O)} : g_O(r, s'(O, f(O))) \in T, s' \circ i(O[r] \rightarrow N[g_O(r, s'(O, f(O))))\}) \\
&= \sum_{O \in R_s} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \prod_{j \neq h(O)} T_j, s' \in I^{L_s(O)} : s'(O, f(O)) \in T_{h(O)}, s' \circ i(O[r] \rightarrow N[g_O(r, s'(O, f(O))))\}) \\
&= \sum_{O \in R_s} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \prod_{j \neq h(O)} T_j, s' \in I^{L_s(O)} : (M, s) \Rightarrow_f^{n-1} (O[r], s')\}) \mu(\{s' \in I^{L_s(O)} \mid s'(O, f(O)) \in T_{h(O)}\}) \\
&= \sum_{O \in R_s} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \prod_{j \neq h(O)} T_j, s' \in I^{L_s(O)} : (M, s) \Rightarrow_f^{n-1} (O[r], s')\}) \mu(T_{h(O)}) \mu(S) + \sum_{O \in R_d} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \prod_{j \neq h(O)} T_j, s' \in I^{L_s(O)} : s'(O, f(O)) \in T_{h(O)}, (M, s) \Rightarrow_f^{n-1} (O[r], s')\}) \mu(S) + \sum_{O \in R_d} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \prod_{j \neq h(O)} T_j, s' \in I^{L_s(O)} : s'(O, f(O)) \in T_{h(O)}, (M, s) \Rightarrow_f^{n-1} (O[r], s')\}) \mu(S) + \sum_{O \in R_d} \mu(\{s \in I^{L_s(M)} \mid \exists r \in \mathbb{R}^{k-1}, s' \in I^{L_s(O)} : g_O(r, s'(O, f(O))) \in T, (M, s) \Rightarrow_f^{n-1} (O[r], s')\}) \mu(S) + \sum_{O \in R_d} \mu(\{s \in I^{L_s(M)} \mid \exists O \in R_s, r \in \mathbb{R}^{k-1}, s' \in I^{L_s(O)}, r' \in T, s'' \in I^{L_s(N)} : (M, s) \Rightarrow_f^{n-1} (O[r], s') \Rightarrow_f (N[r'], s'')\}) \\
&= \mu(\{s \in I^{L_s(M)} \mid \exists r \in T, s' \in I^{L_s(N)} : (M, s) \Rightarrow_f^n (N[r], s')\}) \mu(S)
\end{aligned}$$

The usual call-by-value semantics can be implemented as one of these reduction strategies, given by (with V a value and T a term that isn't a value and M

a general term)

$$\begin{aligned}
\text{cbv}(TM) &= @_1; \text{cbv}(T) \\
\text{cbv}(VT) &= @_2; \text{cbv}(T) \\
\text{cbv}(\underline{f}(V_1, \dots, V_{k-1}, T, M_{k+1}, \dots, M_n)) &= \underline{f}_k; \text{cbv}(T) \\
\text{cbv}(\mathbf{Y}T) &= \mathbf{Y}; \text{cbv}(T) \\
\text{cbv}(\text{if}(T < 0, M_1, M_2)) &= \text{if}_1; \text{cbv}(T) \\
\text{cbv}(\text{score}(T)) &= \text{score}; \text{cbv}(T) \\
\text{cbv}(V) &\text{ is undefined} \\
\text{cbv}(T) &= \cdot \text{ otherwise}
\end{aligned}$$

(this last case covers redexes at the root position).

A closed term M terminates with a given reduction strategy f and samples s if there is some natural number n such that $(M, s) \Rightarrow_f^n (N, s')$ where f gives no reduction at N . The term terminates almost surely with respect to f if it terminates with f for almost all s .

Theorem 1. *A closed term M is AST with respect to cbv iff it is AST.*

Proof. For both red and \Rightarrow_{cbv} , the probability measure on the sample space can be used to define a measure on the reduction sequences of each finite length. The distributions are equal for each length, by induction, as follows.

The base case, length 0, is trivial, because in both cases the distribution has probability 1 on the sequence containing only M .

For the inductive case, let the distribution of terms after n steps be D_n . This can be split into a random and a deterministic part, $D_n^r + D_n^d$, where D_n^r has weight only on those terms of the form $E[\text{sample}]$, and D_n^d only has weight on the terms not of that form. \square

Theorem 2. *If M terminates with some reduction strategy f and samples s , it terminates with cbv and s .*

Proof. \square

Corollary 1. *If M is AST with respect to any reduction strategy, it is AST.*

Proof. Suppose M is AST with respect to f . Let the set of samples with which it terminates with this reduction strategy be X . By Theorem 2, M also terminates with cbv and every element of X , and X has measure 1, by assumption, therefore M is AST with respect to cbv therefore by Theorem 1 it is AST. \square

[Give an example or two of a reduction strategy (for \Rightarrow) that is not cbv .]

\triangleleft L0

[NOTE. For the standard CBV semantics, this example is indeed not a problem.

\triangleleft AK-R

In the alternative semantics, defining

$$\begin{aligned}
A[x] &= \text{if}(\text{if}(x > 0, I, I) \text{ sample} - 0.5 > 0, 0, \Omega) \\
B &= \text{if}(\text{sample} - 0.5 > 0, 0, \Omega)
\end{aligned}$$

The set of samples where it terminates is

$$\bigcup_{r \in [0,1]} \{s \in I^{L_s(A[\text{sample}])} \mid s([A[\text{sample}]], \text{if}_1; \sqsubseteq_1; @_1; \text{if}_1) = r, \\ s([A[\text{sample}], A[r], \dots, B, \text{if}_1] > 0.5)\}.$$

This is a rather unwieldy expression, but the crucial part is that r occurs twice in the conditions on s : once as the value a sample must take, and once in the location of a sample. It is this set that's unmeasurable.]

3 Ranking functions

Given a probabilistic program (i.e. a closed term M), in order to construct a supermartingale to prove its a.s. termination, a function to assign values to each reachable program state is necessary.

Let $Rch(M) := \{N \in \Lambda \mid M \rightarrow^* N\}$, with the σ -algebra induced as a subset of Λ . ↔ L0

Definition 3. A *ranking function* on M is a measurable function $f : Rch(M) \rightarrow \mathbb{R}$ such that $f(N) \geq 0$ for all N , and

1. $f(E[Y\lambda x.N]) \geq 1 + f(E[\lambda z.N[(Y\lambda x.N)/x]z])$ where z is not free in N
2. $f(E[\text{sample}]) \geq \int_I f(E[x]) \text{Leb}(dx)$ [I'm not sure whether $Rch(M)$ is a measurable subset of Λ , so it might not have a natural measure, but I'm pretty sure this doesn't actually matter, and the σ -algebra at least is well-defined. $Rch(M)$ could fail to be measurable if there are measurable functions $\mathbb{R}^n \rightarrow \mathbb{R}$ with non-measurable ranges. I don't know whether these exist. They do for the Lebesgue σ -algebra, but I'm using the Borel ones.] [It is worth inserting the preceding (including the example of a measurable function with non-measurable range) as a remark (or as a "commenting.sty" note).] [I'd like to include such a remark, but the only reference I've been able to find for the fact that these functions exist is a Stack Exchange question² which implies that the construction is difficult but doesn't actually say what it is (so I probably won't be able to find it on my own).] < AK-R
3. $f(E[R]) \geq f(E[R'])$ for any other redex R , where $R \rightarrow R'$. [Metavariables: M and N for terms and programs; R / R' for redex / contractum.] ↔ L0 < L0

We say that the ranking function f is *strict* if there exists $\epsilon > 0$ such that $f(N) = 0$ iff N is a value, and for all E and $R \rightarrow R'$, $f(E[R']) \leq f(E[R]) - \epsilon$.

Any closed term for which a ranking (respectively, strict ranking) function exists is called *rankable* (respectively, *strictly rankable*).

²<https://math.stackexchange.com/questions/1717282/construction-of-a-borel-measurable-function-mapping-borel-samples-to-values> rq=1

4 Supermartingales

[A brief explanatory note on supermartingales by way of motivation is needed here.] ◁ L0

[Supermartingales is a fitting title of this section. However the details of the supermartingale construction are suppressed in the proof. I propose that you reformulate the theorem as follows.] ◁ L0

Theorem 3. *If a closed SPCF term is rankable by f , then $(f(M_n))_{n \in \omega}$ is a supermartingale adapted to the filtration $(\mathcal{F}_n)_{n \in \omega}$ where $\mathcal{F}_n = \sigma(M_1, \dots, M_n)$.* ↔ L0

[We also need to show $f(M_n)$ is integrable.] ◁ L0

Given a closed term M and a ranking function f for it, define random variables on the probability space S (where s is a random variable) by

$$\begin{aligned} (M_n, s_n) &= \text{red}^n(M, s) \\ y_0 &= 0 \\ y_{n+1} &= \min\{k \mid k > y_n, M_k \text{ a value or of the form } E[\text{YN}]\} \\ M'_n &= M_{y_n} \\ X_n &= f(M'_n) \end{aligned}$$

and define a filtration $\mathcal{F}_n = \sigma(M_k, k \leq n)$ (i.e. all the samples used up to step n).

[This is quite a pedantic point but technically the proof wouldn't be correct without it.] In the following proof, conditional expectations (conditioning on \mathcal{F}_n) will be used. Conditional expectations are in general only defined up to a null set, but in this case there is a natural measure on each element of \mathcal{F}_n (as they're always isomorphic to S), and this is the version of the conditional expectation that will be used. [I find this a little cryptic. Perhaps I am missing something, but the definition of conditional expectation $\mathbb{E}[f(M_{n+1}) \mid \mathcal{F}_n]$ is standard. The stochastic process $(M_n)_{n \in \omega}$ is trivially adapted to the natural filtration $(\mathcal{F}_n)_{n \in \omega}$.] ◁ AK-R ◁ L0

The expectation of $f(M_{n+1})$ given \mathcal{F}_n is trivially less than or equal to $f(M_n)$ in the cases that $M_n \neq E[\text{sample}]$, and in the case of sample , [This case analysis is problematic. The r.v. M_n is a measurable function of type $S \rightarrow \Lambda^0$ defined by $M_n(s) := N$ where $\text{red}^n(M, s) = (N, s_n)$. In general, it does not make sense to say " $M_n = E[\text{sample}]$ " without referencing a trace (i.e. an element of S). Take $M = \text{if}(\text{sample} - 0.5 < 0, \text{sample}, 1)$. Then $M_3([0.1, \dots]) = \text{sample}$ but $M_3([0.6, \dots]) = 1$. The case analysis should take place "inside the conditional

expectation”.]

$$\begin{aligned}
& \mathbb{E}[f(M_{n+1}) \mid \mathcal{F}_n] \\
&= \mathbb{E}[f(M_{n+1}) \mid M_n = E[\text{sample}], \mathcal{F}_n] \\
&= \mathbb{E}[f(E[\pi_h(s_n)]) \mid \mathcal{F}_n] \\
&= \int_0^1 f(E[x]) \, dx && \text{as } s_n \text{ is independent of } \mathcal{F}_n \\
&\leq f(E[\text{sample}]) && \text{by assumption on } f \\
&= f(M_n),
\end{aligned}$$

[Problem: Some of the above terms (conditional expectation and $f(M_n)$) are random variables (i.e. functions) but the integral is a real number.] therefore the values of the ranking function $f(M_n)$ are a supermartingale with respect to \mathcal{F}_n . ◁ L0

Given M'_n , the reduction relation \rightarrow , *excluding* Y-reduction steps, is strongly normalising because of the type system, therefore there is some finite bound on the number of reduction steps that can take place from M'_n without a Y-reduction step, therefore y_{n+1} is (conditional on $\mathcal{F}_{y_{n+1}}$) a bounded stopping time, therefore $\mathbb{E}[f(M_{y_{n+1}}) \mid \mathcal{F}_{y_{n+1}}] \leq f(M_{y_{n+1}})$. For $n > 0$, if M_{y_n} isn't already a value, then $M_{y_n} = E[Y\lambda x.N]$ for some E, N , therefore $M_{y_{n+1}} = E[\lambda z.N[(Y\lambda x.N)/x]z]$ and $f(M_{y_{n+1}}) \leq f(M_{y_n}) - 1$, therefore if M_{y_n} isn't a value, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_{y_{n+1}}] \leq X_n - 1$.

In the other case that X_n is a value, $X_{n+1} = X_n$, and $X_n = \mathbb{E}[X_n \mid \mathcal{F}_{y_{n+1}}]$ therefore in either case $\mathbb{E}[X_{n+1} \mid \mathcal{F}_{y_{n+1}}] - \mathbb{E}[X_n \mid \mathcal{F}_{y_{n+1}}] \leq -P[M_{y_n} \text{ not a value} \mid \mathcal{F}_{y_{n+1}}]$ (as the probability is 0 or 1). Marginalising over $\mathcal{F}_{y_{n+1}}$, this implies that $\mathbb{E}[X_{n+1}] - \mathbb{E}[X_n] \leq -\mathbb{P}[M_{y_n} \text{ not a value}]$ therefore as $\mathbb{E}[X_n]$ is bounded below, $\mathbb{P}[M_{y_n} \text{ not a value}]$ must tend to 0 as $n \rightarrow \infty$. M_{y_n} a value $\Rightarrow M_{y_{n+1}}$ a value therefore

$$\begin{aligned}
\mathbb{P}[M_{y_n} \text{ not a value for infinitely many values of } n] &= \mathbb{P}[\forall n : M_{y_n} \text{ not a value}] \\
&\leq \inf_n \mathbb{P}[M_{y_n} \text{ not a value}] \\
&= 0
\end{aligned}$$

therefore M terminates almost surely.

Reorganised proof

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We have in mind $\Omega = S$, and $\mathbb{P} = \mu$.

Definition 4. A sequence of random variables $(Y_n)_{n \in \omega}$ adapted to a filtration $(\mathcal{F}_n)_{n \in \omega}$ is a *supermartingale* if for all $n \in \omega$, $\mathbb{E}[|Y_n|] < \infty$, and $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \leq Y_n$. [The preceding means: $\forall A \in \mathcal{F}_n, \int_A d\mu Y_{n+1} \leq \int_A d\mu Y_n$.] For $\epsilon > 0$, it is a ϵ -*ranking supermartingale* if, in addition, for all n , $Y_n \geq 0$ and $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \leq Y_n - \epsilon \cdot \mathbf{1}_{\{Y_n > 0\}}$. [Fioriti and Hermanns 2015; Chatterjee et al. 2016] ◁ L0

Intuitively Y_n gives the rank of the program after n steps of computation. (Assume that the rank of a term is 0 iff it is a value.) In a ϵ -ranking supermartingale, each computation step causes a strict decrease in rank of at least ϵ , provided the term in question is not a value.

Let T and T' be stopping times adapted to $(\mathcal{F}_n)_{n \in \omega}$. Recall the σ -algebra (consisting of measurable subsets “priori to T ”)

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid \forall i \in \omega. A \cap \{T \leq i\} \in \mathcal{F}_i\}$$

and if $T \leq T'$, then $\mathcal{F}_T \subseteq \mathcal{F}_{T'}$.

The following is an iterated version of Doob’s well-known Optional Sampling Theorem (see, e.g., [Ash and Doléans-Dade 2000, §6.7]).

Theorem 4 (Optional Sampling). *Let $(X_n)_{n \in \omega}$ be a supermartingale, and $(T_n)_{n \in \omega}$ a sequence of increasing stopping times, then $(X_{T_n})_{n \in \omega}$ is a supermartingale adapted to $(\mathcal{F}_{T_n})_{n \in \omega}$ if one of the following conditions holds:*

1. *each T_n is bounded i.e. $T_n < c_n$ where c_n is a constant*
2. *$(X_n)_{n \in \omega}$ is uniformly integrable.*

Theorem 5 ([Fioriti and Hermanns 2015, Lemma 5.5], [Chatterjee et al. 2016, Prop 1]). *Let $(Y_n)_{n \in \omega}$ be a ϵ -ranking supermartingale, and define the r.v. T by $T(s) := \min\{n \mid Y_n(s) = 0\}$.*

Then $T < \infty$ almost surely, and $\mathbb{E}[T] \leq \frac{\mathbb{E}[Y_0]}{\epsilon}$.

Proof Sketch. First establish by induction on n :

$$\forall n \in \omega. \mathbb{E}[Y_n] \leq \mathbb{E}[Y_0] - \epsilon \cdot \left(\sum_{i=0}^{n-1} \mathbb{P}[Y_i > 0] \right).$$

It follows that $\sum_{i=0}^{\infty} \mathbb{P}[Y_i > 0]$ converges; hence $\lim_{n \rightarrow \infty} \mathbb{P}[Y_n > 0] = 0$ and so $\mathbb{P}[T < \infty] = 1$. Then it suffices to prove: if $T < \infty$ a.s., then $\mathbb{E}[T] = \sum_{i=1}^{\infty} \mathbb{P}[T \geq i]$. \square

A main result is the following theorem.

Theorem 6. *If a closed SPCF term is rankable (respectively, strictly rankable) by f , then $(f(M_n))_{n \in \omega}$ is a supermartingale (respectively, ranking supermartingale) adapted to the filtration $(\mathcal{F}_n)_{n \in \omega}$ where $\mathcal{F}_n = \sigma(M_1, \dots, M_n)$.*

Proof of Theorem 6. *Notation.* Given an ω -sequence (e.g. $s \in S$) and $m \in \omega$, we write $s_{\leq m} \in I^m$ to mean the m -long prefix of s .

Fix a closed SPCF term $M \in \Lambda^0$, and a ranking function f for it, satisfying $f(N) = 0$ iff N is a value. Let $n \in \omega$, define the following random variables on the probability space (S, \mathcal{F}, μ) :

$$\begin{aligned} M_n(s) &:= \pi_0(\text{red}^n(M, s)) \\ \# \text{draw}_n(s) &:= l \text{ where } \pi_1(\text{red}^n(M, s)) = \underbrace{\pi_t(\dots(\pi_t(s)))}_l \end{aligned}$$

and the filtration $\mathcal{F}_n = \sigma(M_1, \dots, M_n)$. The \mathcal{F}_n -measurability of M_n (and hence of $\#draw_n$) follows from [Borgström et al. 2016]; $\#draw_n$ is a stopping time (bounded by n). Take $s \in A \in \mathcal{F}_n$ with $\#draw_n(s) = l$. For any $s' \in S$, if $s_{\leq l} = s'_{\leq l}$ then $s' \in A$. It follows that $\{s_{\leq l}\} \cdot I^\omega \subseteq A$. [NOTE. To fix notation: \triangleleft L0
 $\sigma(M_n) = \sigma(\{M_n^{-1}(\alpha_j, U_{\alpha_j}) \mid \alpha_j \in \mathbf{Sk}_j, U_{\alpha_j} \in \mathcal{B}(\mathbb{R}^j), j \geq 0\})$]

We say that a given SPCF term is *type-1* (respectively 2, 3 and 4) if it has the shape $E[\mathbf{Y} \lambda x. N]$ (respectively $E[\mathbf{sample}]$, $E[R]$ where R is any other redex, and of a value). Define $\mathbf{T}_i := \{s \mid M_n(s) \text{ is type-}i\}$. It is straightforward to see that each $\mathbf{T}_i \in \mathcal{F}_n$, and $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \mathbf{T}_4\}$ is a partition of S . Hence it suffices to prove the following lemma.

Lemma 9. *For all $i \in \{1, 2, 3, 4\}$ and $A \in \mathcal{F}_n$*

$$\int_A \mu(ds) f_{n+1}(M)[s \in \mathbf{T}_i] \leq \int_A \mu(ds) f(M_n)[s \in \mathbf{T}_i]$$

where $[s \in \mathbf{T}_i]$ is the Iverson bracket.

Proof. We show the non-trivial case of $i = 2$. First we express

$$f_{n+1}(M)[s \in \mathbf{T}_2] = \sum_{i \in \mathcal{I}} f(E_i[\underline{\sigma_i(s)}][\underline{\rho(s)}])[s \in U_i] \quad (2)$$

where

- \mathcal{I} is a countable indexing set
- $E_i[\cdot][\mathbf{sample}] \in \mathbf{Sk}_{j_i}$, and $\sigma_i : S \rightarrow \mathbb{R}^{j_i}$, and $\rho(s) = \pi_h(\pi_1(\text{red}^n(M, s))) \in \mathbb{R}$
- $\{U_i\}_{i \in \mathcal{I}}$ is a partition of \mathbf{T}_2 , where each U_i is determined by a skeletal term $E_i[\cdot][\mathbf{sample}]$ and a number (of draws) $l_i \leq n$; precisely

$$U_i := \#draw_n^{-1}(l_i) \cap M_n^{-1}(E_i[\cdot][\mathbf{sample}], \mathbb{R}^{j_i}) \in \mathcal{F}_n.$$

$$[\text{I.e. } (E_i[\underline{\sigma_i(s)}][\mathbf{sample}], \pi_i^{l_i}(s)) = \text{red}^n(M, s) \text{ iff } s \in U_i.]$$

\triangleleft L0

Observe that if $s \in U_i$ then $\{s_{\leq l_i}\} \cdot I^\omega \subseteq U_i$; in fact $(U_i)_{\leq l_i} \cdot I^\omega = U_i$. This means that for any measurable $g : S \rightarrow \mathbb{R}_{\geq 0}$, if $g(s)$ only depends on the $(l_i + 1)$ -long prefix of s , then, writing $g' : I^{l_i+1} \rightarrow \mathbb{R}_{\geq 0}$ where $g(s) = g'(s_{\leq l_i+1})$

$$\int_{A \cap U_i} \mu(ds) g(s) = \int_{(A \cap U_i)_{\leq l_i+1}} \text{Leb}_{l_i+1}(dt) g'(t) \quad (3)$$

(For simplicity, we will write $g' = g$ in the following.)

Take $s \in U_i$, and set $l = l_i$. Now $\sigma_i(s)$ depends on $s_{\leq l}$, and $\rho(s)$ depends on $s_{\leq l+1}$. Let u range over $(U_i)_{\leq l}$. Then, it follows from the definition of f that

$$\int_I \text{Leb}(dr) f(E_i[\underline{\sigma_i(u)}][r]) \leq f(E_i[\underline{\sigma_i(u)}][\mathbf{sample}]).$$

Take $A \in \mathcal{F}_n$, and integrating both sides, we get

$$\begin{aligned} & \int_{(A \cap U_i) \leq l} \text{Leb}_l(du) \int_I \text{Leb}(dr) f(E_i[\underline{\sigma_i}(u)][\underline{r}])) \\ & \leq \int_{(A \cap U_i) \leq l} \text{Leb}_l(du) f(E_i[\underline{\sigma_i}(u)][\text{sample}])) \end{aligned}$$

Since Leb_{l+1} is the (unique) product measure satisfying $\text{Leb}_{l+1}(V \times B) = \text{Leb}_l(V) \cdot \text{Leb}(B)$, and $(U_i)_{\leq l_i} \cdot I^\omega = U_i$, we have

$$\begin{aligned} & \int_{(A \cap U_i) \leq l+1} \text{Leb}_{l+1}(du') f(E_i[\underline{\sigma_i}(u')][\underline{\rho}(u')])) \\ & \leq \int_{(A \cap U_i) \leq l+1} \text{Leb}_{l+1}(du') f(E_i[\underline{\sigma_i}(u')][\text{sample}])) \end{aligned}$$

and so

$$\int_{A \cap U_i} \mu(ds) f(E_i[\underline{\sigma_i}(s)][\underline{\rho}(s)])) \leq \int_{A \cap U_i} \mu(ds) f(E_i[\underline{\sigma_i}(s)][\text{sample}])). \quad (4)$$

Finally, integrating both sides of (2), we have

$$\begin{aligned} \int_A \mu(ds) f_{n+1}(M)[s \in \mathbf{T}_i] &= \int_A \mu(ds) \sum_{i \in \mathcal{I}} f(E_i[\underline{\sigma_i}(s)][\underline{\rho}(s)])(s \in U_i) \\ &= \sum_{i \in \mathcal{I}} \int_{A \cap U_i} \mu(ds) f(E_i[\underline{\sigma_i}(s)][\underline{\rho}(s)])) \\ &\leq \sum_{i \in \mathcal{I}} \int_{A \cap U_i} \mu(ds) f(E_i[\underline{\sigma_i}(s)][\text{sample}])) \quad \because (4) \\ &= \int_A \mu(ds) \sum_{i \in \mathcal{I}} f(E_i[\underline{\sigma_i}(s)][\text{sample}]))[s \in U_i] \\ &= \int_A \mu(ds) f_n(M)[s \in \mathbf{T}_2] \end{aligned}$$

□

As an immediate corollary of Lemma 9, each $f(M_n)$ is integrable:

$$\int_S \mu(ds) |f(M_n)(s)| = \int_S \mu(ds) f(M_n)(s) \leq \int_S \mu(ds) f(M_0)(s) = f(M).$$

This concludes the proof of Theorem 3. □

As before, fix a $M \in \Lambda^0$. Define random variables on (S, \mathcal{F}, μ) :

$$\begin{aligned} T_0(s) &:= 0 \\ T_{n+1}(s) &:= \min\{k \mid k > T_n(s), M_k(s) \text{ a value or of form } E[\mathbf{Y}\lambda x.N]\} \\ X_n(s) &:= f(M_n(s)) \\ Y_n(s) &:= X_{T_n}(s) \\ T(s) &:= \min\{n \mid Y_n(s) = 0\} \end{aligned} \quad (5)$$

[Define the Y -runtime of M to be the random variable T_M^Y :

◁ L0

- if $T_M(s) = n < \infty$ then $T_M^Y(s) :=$ number of Y -reduction steps in the n steps of reduction to $\text{red}^n(M, s)$
- if $T_M(s) = \infty$ then $T_M^Y(s) := \infty$.

Observe that $T = T_M^Y$.] Note that

1. $T_M < \infty$ a.s. iff $T_M^Y < \infty$ a.s. (because $T_M(s) < \infty$ iff $T_M^Y(s) < \infty$)
2. $\mathbb{E}[T_M] < \infty$ implies $\mathbb{E}[T_M^Y] < \infty$ (because $T_M^Y \leq T_M$), but the converse is false (Example 1).

Theorem 7 (Soundness). *1. If a closed SPCF term M is rankable, then $T_M^Y < \infty$ a.s. (equivalently M is AST), and $\mathbb{E}[T_M^Y] < \infty$.*
2. If a closed SPCF term M is strictly rankable, then $\mathbb{E}[T_M] < \infty$ i.e. M is PAST.

Proof. For 1, observe that T_0, T_1, T_2, \dots are an increasing sequence of stopping times adapted to $(\mathcal{F}_n)_{n \in \omega}$, and each T_i is bounded. Thanks to Theorem 6, $(X_n)_{n \in \omega}$ is a supermartingale; it then follows from Theorem 4 that $(Y_n)_{n \in \omega}$ is a (1-ranking) supermartingale. Therefore, by Theorem 5, $T_M^Y < \infty$ a.s. and $\mathbb{E}[T_M^Y] < \infty$.

Statement 2 follows immediately from Theorem 5. \square

Thus the method of (strict) ranking function is sound for proving (positive) a.s. termination of SPCF programs. It is in fact also complete in a sense: if $\mathbb{E}[T_N^Y] < \infty$ for all $N \in \text{Rch}(M)$ then M is rankable (Theorem 8).

5 Constructing Ranking Functions

Although rankability implies almost-sure termination, the converse does not hold in general. For example,

$$\text{if}(-\text{sample} < 0, \underline{0}, (\mathbf{Y}\lambda x.x) \underline{0}) \tag{6}$$

terminates in 3 steps with probability 1, but isn't rankable because $(\mathbf{Y}\lambda x.x)\underline{0}$ is reachable, although that has probability 0. Not only is this counterexample AST, it's positively almost surely terminating i.e. the expected time to termination is finite.

A ranking function can be constructed under the stronger assumptions that, for every N reachable from M , the expected number of Y -reduction steps from N to a value is finite. In particular, the expected number of Y -reduction steps from each reachable term is a ranking function. Note that a finite number of expected Y -reduction steps does not necessarily imply a finite number of expected total reduction steps.

Example 1. The term

$$M = (\Upsilon \lambda f n. \text{if}(\text{sample} - 0.5 < 0, n \underline{0}, f(\lambda x. n(nx))))(\lambda x. x + 1)$$

terminates with only 2 Υ -reductions on average i.e. $\mathbb{E}[T_M^\Upsilon] = 2$, but applies the increment function 2^n times with probability 2^{-1-n} for $n \geq 0$, which diverges i.e. $\mathbb{E}[T_M] = \infty$.

Theorem 8. *Given a closed term M , the function $f : Rch(M) \rightarrow \mathbb{R}$ given by $f(N) = \mathbb{E}[\text{the number of } \Upsilon\text{-reduction steps from } N \text{ to a value}]$, if it exists, is the least of all possible ranking functions of M .*

Proof. Let f be the candidate least ranking function defined above, and suppose g is another ranking function such that $f(N) > g(N)$ for some $N \in Rch(M)$. The restrictions of f and g to $Rch(N)$ have the same properties assumed of f and g , so assume w.l.o.g. that $N = M$. The difference $g - f$ is then a supermartingale (with the same setup as in the proof of Theorem 3 therefore $\forall n. \mathbb{E}[g(M_n)] \leq \mathbb{E}[f(M_n)] + g(M) - f(M)$).

$\mathbb{E}[f(M_n)] = \sum_{k=n}^{\infty} \mathbb{P}[M_n = E[\Upsilon N] \text{ for some } E, N] \rightarrow 0$ as $n \rightarrow \infty$ therefore as $g(M) - f(M) < 0$, eventually $\mathbb{E}[g(M_n)] < 0$, which is impossible, therefore $g \geq f$ as required.

In order for f to be the least ranking function of M , it also has to actually be a ranking function itself. Each of the conditions on a ranking function is easily verified from the definition of f . \square

5.1 Partial ranking functions

Even in the case of reasonable simple terms, explicitly constructing a ranking function would be a lot of work, and Theorem 8 makes even stronger assumptions than almost sure termination, so it isn't useful for proving it. Take, for example, the term

$$\begin{aligned} &(\Upsilon \lambda f, n : \\ &\quad \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1)) \\ &\quad) \underline{0} \end{aligned}$$

which generates a geometric distribution. Despite its simplicity, its Rch contains all the terms

1. $(\Upsilon \lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1))) \underline{0}$
2. $(\lambda z. (\lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1))) (\Upsilon \lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1))) z) \underline{i}$
3. $(\lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1))) (\Upsilon \lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n + 1))) \underline{i}$
4. $(\lambda n : \text{if}(\text{sample} - 0.5 < 0, n, (\Upsilon \lambda f, m : \text{if}(\text{sample} - 0.5 < 0, n, f(m + 1)))) (m + 1)) \underline{i}$

5. $\text{if}(\text{sample} - 0.5 < 0, \underline{i}, (\mathbf{Y}\lambda f, m : \text{if}(\text{sample} - 0.5 < 0, m, f(m+1)))(\underline{i}+1))$
6. $\text{if}(\underline{r} - 0.5 < 0, \underline{i}, (\mathbf{Y}\lambda f, m : \text{if}(\text{sample} - 0.5 < 0, m, f(m+1)))(\underline{i}+1))$
7. $\text{if}(\underline{r - 0.5} < 0, \underline{i}, (\mathbf{Y}\lambda f, m : \text{if}(\text{sample} - 0.5 < 0, m, f(m+1)))(\underline{i}+1))$
8. \underline{i}
9. $(\mathbf{Y}\lambda f, m : \text{if}(\text{sample} - 0.5 < 0, m, f(m+1)))(\underline{i}+1)$
10. $(\lambda z.(\lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n+1)))(\mathbf{Y}\lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n+1))))z)(\underline{i}+1).$

[TODO: make this into a diagram with arrows.]

◁ AK-R

[Here is a more readable version.]

◁ L0

Writing $\Theta = \lambda f, n : \text{if}(\text{sample} - 0.5 < 0, n, f(n+1))$, we have

1. $(\mathbf{Y}\Theta) \underline{0}$
2. $(\lambda z. \Theta (\mathbf{Y}\Theta) z) \underline{i}$
3. $\Theta (\mathbf{Y}\Theta) \underline{i}$
4. $(\lambda n : \text{if}(\text{sample} - 0.5 < 0, n, (\mathbf{Y}\lambda f, m : \text{if}(\text{sample} - 0.5 < 0, n, f(m+1)))(m+1)))(\underline{i})$

⇔ AK-R

[I think the preceding should be: $(\lambda n : \text{if}(\text{sample} - 0.5 < 0, n, (\mathbf{Y}\lambda f, m : \text{if}(\text{sample} - 0.5 < 0, m, f(m+1)))(n+1)))(\underline{i})$, and so, more succinctly:

◁ L0

$$(\lambda n : \text{if}(\text{sample} - 0.5 < 0, n, (\mathbf{Y}\Theta) (n+1)))(\underline{i})$$

↓

5. $\text{if}(\text{sample} - 0.5 < 0, \underline{i}, (\mathbf{Y}\Theta) (\underline{i}+1))$
6. $\text{if}(\underline{r} - 0.5 < 0, \underline{i}, (\mathbf{Y}\Theta) (\underline{i}+1))$
7. $\text{if}(\underline{r - 0.5} < 0, \underline{i}, (\mathbf{Y}\Theta) (\underline{i}+1))$
8. \underline{i}
9. $(\mathbf{Y}\Theta) (\underline{i}+1)$
10. $(\lambda z. \Theta (\mathbf{Y}\Theta) z) (\underline{i}+1)$
11. $(\lambda z. (\lambda n : \text{if}(\text{sample} - 0.5 < 0, n, (\mathbf{Y}\Theta) (n+1))))z)(\underline{i})$

[So we have:

◁ L0

$$\begin{aligned} 1 &\rightarrow 11 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7 \rightarrow 8 \\ 7 &\rightarrow 9 \rightarrow 11 \rightarrow \dots \end{aligned}$$

You have 2, 3 and 10 because your Y-redex rule is

$$\mathbf{Y}\lambda x.s \rightarrow \lambda z.(\lambda x.s) (\mathbf{Y}\lambda x.s) z$$

which is not the official definition.]

Even in this simple case, defining a ranking function explicitly is awkward because of the number of cases, although in most cases, because the value need only be greater than or equal to that of the next term in sequence, it suffices to take the ranking function as taking only 3 distinct values. (We will explain why later.)

The definition of rankability is also inconvenient for syntactic sugar. It could be useful, for example, to define $M \oplus_p N = \text{if}(\text{sample} - p < 0, M, N)$, where $M \oplus_p N$ reduces to M or N , depending on the first value of s , with probability p resp. $(1 - p)$. Technically though, it reduces first to $\text{if}(\underline{r} - p < 0, M, N)$ for $r \in I$, so those terms all need values of the ranking function too.

In both of these cases, there are only some values of the ranking function that are semantically important. Define a *partial ranking function* on a closed term M to be a partial function $f : \text{Rch}(M) \rightarrow \mathbb{R}$ such that

- $f(N) \geq 0$ for all N where f is defined.
- f is defined at M .
- For any N in the domain of definition of f , evaluation of N will eventually reach some O which is either a value or in the domain of definition of f , and $f(N) \geq \mathbb{E}[f(O) + \text{the number of } Y\text{-reduction steps from } N \text{ to } O]$ (where $f(O)$ is taken to be 0 if O is a value outside of the domain of f). ◁ AK-R

A partial ranking function that is total is just a ranking function. Providing a partial ranking function is essentially part way between providing a ranking function and directly proving almost sure termination.

Theorem 9. *Every partial ranking function is a restriction of a ranking function.*

Proof. Take a closed term M and partial ranking function f on M . Define $f_1 : \text{Rch}(M) \rightarrow \mathbb{R}$ by

$$\begin{aligned} f_1(N) &= f(N) \text{ whenever } f(N) \text{ is defined,} \\ f_1(V) &= 0 \text{ for values } V \text{ not in the domain of } f. \end{aligned}$$

Define $(\text{next}(N, s), _) = \text{red}^n(N, s)$ for the least $n \geq 0$ such that it's in the domain of f_1 , and $g(N, s) = |\{m < n \mid \text{red}^m(N, s) \text{ is of the form } (E[YN'], s')\}|$. The function next is well-defined (i.e. n is finite) for all $N \in \text{Rch}(M)$ by induction on the path from M to N , by the third condition on partial ranking

functions. Define $f_2(N) = \int_S f_1(\text{next}(N, s)) + g(N, s) \mu(ds)$. The (total) function f_2 agrees with f on f 's domain, and it is a ranking function on M (in fact, the least ranking function of which f is a restriction). [TODO. The preceding sentence deserves a proof.] ◁ L0 □

As a corollary, any term which admits a partial ranking function terminates almost surely.

5.2 Examples

Let $M \oplus_p N = \text{if}(\text{sample} - p < 0, M, N)$, for $p \in (0, 1]$, then there are the pseudo-reduction relations

$$\begin{aligned} E[M \oplus_p N] &\rightarrow^3 E[M] \\ E[M \oplus_p N] &\rightarrow^3 E[N] \\ \text{red}^3(E[M \oplus_p N], s) &= \begin{cases} (E[M], \pi_t(s)) & \text{if } \pi_h(s) < p \\ (E[N], \pi_t(s)) & \text{if } \pi_h(s) \geq p. \end{cases} \end{aligned}$$

A partial ranking function could be defined with respect to this shortcut reduction simply by replacing \rightarrow and red in the definition of a partial ranking function by a version that goes straight from $N \oplus_p O$ to N or O . Such a pseudo-partial ranking function would then be a partial function from a subset of $\text{Rch}(M)$, so it could also be considered as a partial function from all of $\text{Rch}(M)$, and it would in fact also be an actual partial ranking function. It is therefore possible to prove rankability directly using the shortcut reductions.

A similar procedure would work for other forms of syntactic sugar. If a closed term N eventually reduces to one of a set of other terms $\{N_i \mid i \in I\}$ with certain probabilities, a partial ranking function defined with respect to a reduction sequence that skips straight from N to N_i is also a valid partial ranking function for the original reduction function, and therefore its existence implies almost sure termination. There is a caveat, however, that Y -reduction steps skipped over in the shortcut still need to be counted for the expected number of Y -reduction steps.

With this abbreviation, the geometric distribution example from earlier can be written as $(Yf, n.n \oplus_{0.5} f(n+1))\underline{0}$. It is then easy to see that the following is a partial ranking function:

$$\begin{aligned} (Yf, n.n \oplus_{0.5} f(n+1))N &\mapsto 2 \\ \underline{i} \oplus_{0.5} (Yf, n.n \oplus_{0.5} f(n+1))(\underline{i}+1) &\mapsto 1 \\ \underline{i} &\mapsto 0 \end{aligned}$$

[Call the above partial function h . I assume that $h : (Yf, n.n \oplus_{0.5} f(n+1))N \mapsto 2$ for all N (such that $(Yf, n.n \oplus_{0.5} f(n+1))N$ is reachable from M), and the second and third clauses hold for all $i \geq 0$.] ◁ L0

Maybe I have misread / misunderstood the preceding, but I don't see why h is a restriction of a ranking function. Suppose h is a restriction of H . Consider the reduction sequence (writing $\Xi = (\Upsilon f, n.n \oplus_{0.5} f(n+1))$):

$$\Xi 0 \xrightarrow{Y} 0 \oplus_{0.5} \Xi 1 \rightarrow \Xi 1 \xrightarrow{Y} 1 \oplus_{0.5} \Xi 2 \rightarrow \Xi 2 \xrightarrow{Y} 2 \oplus_{0.5} \Xi 3$$

Since $H(\Xi 0) = 2$, we must have $H(\Xi 1) \leq 1$ and $H(\Xi 2) \leq 0$, contradicting the restriction assumption. \rfloor In this case, $H(0 \oplus_{0.5} \Xi 1) \leq 1$, but that term reduces to either 0 or $\Xi 1$ therefore the actual condition is that $0.5H(0) + 0.5H(\Xi 1) \leq 1$, which is possible. \rfloor OK. I see it now. \rfloor \triangleleft AK-R
 \triangleleft LO

6 Antitone ranking functions

Example 2. 1. 1D unbiased random walk

$$M_1 = (\Upsilon f n . \text{if}(n = 0, 0, f(n-1) \oplus_{1/2} f(n+1))) 10$$

2. 1D biased (away from 0) random walk

$$M_2 = (\Upsilon f n . \text{if}(n = 0, 0, f(n-1) \oplus_{1/3} f(n+1))) 10$$

3. 1D biased (towards 0) random walk

$$M_3 = (\Upsilon f n . \text{if}(n = 0, 0, f(n-1) \oplus_{2/3} f(n+2))) 10$$

Problem. Find a generalised notion of ranking function so that M_1 becomes rankable, and then prove it sound i.e. rankable in this generalised sense implies AST.

Idea. The definition of antitone-convex ranking function f (say) for $M \in \Lambda^0$ is the same as that of ranking function except that in the case of Y -redex, we require the existence of an antitone (meaning: $r < r'$ implies $\epsilon(r) \geq \epsilon(r')$) and convex function $\epsilon : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ such that the ranking function $f : \text{Rch}(M) \rightarrow \mathbb{R}_{\geq 0}$ satisfies

$$f(E[R']) \leq f(E[R]) - \epsilon(f(E[R]))$$

where $R \rightarrow R'$ is the Y -redex rule.

[An aside: This CBV Y -rule seems cleaner: \triangleleft LO

$$(\Upsilon f^{A \rightarrow B} x^A . \theta^B) v \rightarrow \theta[(\Upsilon f x . \theta)/f, v/x].$$

We assume $\Upsilon f^{A \rightarrow B} x^A . \theta^B$ is a value.]

Definition 5. We say that a supermartingale $(Y_n)_{n \in \omega}$ and a stopping time T [Separating the stopping time from $Y_n = 0$ allows for a neater definition of the ranking function of a term] [Nice!] adapted to filtration $(\mathcal{F}_n)_{n \in \omega}$ are *antitone strict* if $Y_n \geq 0$, and there exists an antitone measurable function $\epsilon : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{>0}$ satisfying \triangleleft AK-R
 \triangleleft LO

$$\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \leq Y_n - \epsilon(Y_n) \cdot 1_{\{T > n\}}$$

Theorem 10. Let $(Y_n)_{n \in \omega}$ be an antitone strict supermartingale via function $\epsilon : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{>0}$, and stopping time T . Then $T < \infty$ almost surely. [It's not ◀ AK-R
actually true that $\mathbb{E}[T] \leq \frac{\mathbb{E}[Y_0]}{\epsilon_0}$ where $\epsilon_0 := \epsilon(\mathbb{E}[Y_0])$. This is fortunate because
it means that a rankability theorem based on antitone strict supermartingales
actually has a chance of being stronger than the existing one (which was pretty
much complete for Y-PAST terms already). In fact, for any almost surely finite
stopping time, there is a corresponding antitone strict supermartingale. This
seems promising for the method being conditionally complete (for terms where
all reachable terms are AST).] [See Section 8.] ◀ L0

Proof. First, as (Y_n) is a supermartingale, $\mathbb{E}[Y_n] \leq \mathbb{E}[Y_0]$, therefore

$$\begin{aligned}
&= \frac{\mathbb{E}[Y_n \mid T > n]}{\mathbb{P}[T > n] \mathbb{E}[Y_n \mid T > n] + \mathbb{P}[T \leq n] \mathbb{E}[Y_n \mid T \leq n] - \mathbb{P}[T \leq n] \mathbb{E}[Y_n \mid T \leq n]} \\
&= \frac{\mathbb{E}[Y_n] - \mathbb{P}[T \leq n] \mathbb{E}[Y_n \mid T \leq n]}{\mathbb{P}[T > n]} \\
&\leq \frac{\mathbb{E}[Y_n]}{\mathbb{P}[T > n]} \\
&\leq \frac{\mathbb{E}[Y_0]}{\mathbb{P}[T > n]}
\end{aligned}$$

Claim: for all $0 < x \leq 1$, let $B_x = \left\lceil \frac{\mathbb{E}[Y_0] + 1}{x \epsilon(\mathbb{E}[Y_0] x^{-1})} \right\rceil$, then $\mathbb{P}[T > B_x] \leq x$.

As the convex hull of ϵ (the greatest convex function less than or equal to it) satisfies all the conditions assumed of ϵ , in addition to being convex, assume wlog that ϵ is convex. Suppose for a contradiction that $\mathbb{P}[T > B_x] > x$, then for every $n \leq B_x$ [Change of justification of the second equality below: event, ◀ L0
not discrete r.v. Recall: expectation conditioning on r.v. (resp. event) is a
r.v. (resp. number).]

$$\begin{aligned}
&\mathbb{E}[Y_n - Y_{n+1}] \\
&= \{ \text{definition and linearity of conditional expectation} \} \\
&\mathbb{E}[Y_n - \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n]] \\
&\geq \{ \text{antitone strict assumption} \} \\
&\mathbb{E}[\epsilon(Y_n) \cdot 1_{\{T > n\}}] \\
&= \{ \text{definition of expectation conditioning on an event} \} \\
&\mathbb{P}[T > n] \mathbb{E}[\epsilon(Y_n) \mid T > n] \\
&\geq \{ \text{Jensen's inequality} \}
\end{aligned}$$

$$\begin{aligned}
& \mathbb{P}[T > n] \in (\mathbb{E}[Y_n \mid T > n]) \\
& \geq \quad \{ \text{proved earlier} \} \\
& \mathbb{P}[T > n] \in \left(\frac{\mathbb{E}[Y_0]}{\mathbb{P}[T > n]} \right) \\
& > \quad \{ \text{assumption, } \mathbb{P}[T > n] \geq \mathbb{P}[T > B_x] > x \} \\
& \quad x \in (\mathbb{E}[Y_0] x^{-1}).
\end{aligned}$$

Therefore, by a telescoping sum

$$\mathbb{E}[Y_{B_x}] = \mathbb{E}[Y_{B_x} - Y_0 + Y_0] \leq \mathbb{E}[Y_0] - B_x x \in (\mathbb{E}[Y_0] x^{-1}) \leq -1 < 0$$

which is a contradiction, therefore the claim must be true, therefore $P[T > n]$ tends to 0 as n tends to infinity, therefore $T < \infty$ almost surely. \square

Results to prove:

1. (AST Soundness) If a closed SPCF term M is antitone rankable, then $T_M^Y < \infty$ a.s. (equivalently M is AST). [This should be fine.] < L0
2. A partial ranking function theorem for antitone-convex ranking functions. [TODO] < L0

Example 2 revisited Programs M_1 and M_3 are AST. To construct antitone ranking functions, we appeal [(hopefully)] to a partial ranking function theorem < L0 for antitone ranking functions.

For M_1 , define the function $g_1 : \mathbb{R}_{\geq 0} \rightarrow [0, 2)$ by $g_1(x) = 2 - \frac{1}{2^{x-1}}$. Note that $(g_1(n))_{n \in \omega}$ satisfies the recurrence relation

$$\begin{cases} z_0 = 0 \\ n \geq 1, & z_n > \frac{1}{2} (z_{n-1} + z_{n+1}). \end{cases}$$

Set the antitone function $\epsilon_1(y) := \frac{2-y}{4}$. Then

$$\epsilon_1(g_1(x)) = g_1(x) - \frac{1}{2} (g_1(x-1) + g_1(x+1)) = \frac{1}{2^{x+1}} \quad (7)$$

(Since g_1 is bijective, with inverse $g_1^{-1}(y) = 1 - \log(2-y)$, we can obtain $\epsilon_1(y)$ from Equation (7).)

Using shorthand $\Theta_1 = Y f n . \text{if}(n = 0, 0, f(n-1) \oplus_{1/2} f(n+1))$, we can define a partial antitone ranking function $f_1 : \text{Rich}(M_1) \rightarrow \mathbb{R}_{\geq 0}$ with antitone function ϵ_1 defined as follows:

$$\begin{aligned}
& \Theta_1 n \mapsto g_1(n) \\
& \text{if}(n = 0, 0, \Theta_1(n-1) \oplus_{1/2} \Theta_1(n+1)) \mapsto \frac{1}{2} (g_1(n-1) + g_1(n+1)) \\
& 0 \mapsto 0
\end{aligned}$$

For M_3 , we use the function $g_3(x) := 3 - \frac{1}{3^{x-1}}$ and the antitone function $\epsilon_3(y) := \frac{84-28y}{27}$. Then

$$\epsilon_3(g_3(x)) = g_3(x) - \frac{1}{3}(2 \cdot g_3(x-1) + g_3(x+2)) = \frac{28}{3^{x+2}}.$$

Following M_1 above, we can define a partial antitone ranking function $f_3 : \text{Rich}(M_3) \rightarrow \mathbb{R}_{\geq 0}$ with ϵ_3 as the antitone function.

Example 3 (Continuous random walk). In SPCF we can construct a function whose argument changes by a random amount at each recursive call:

$$\underbrace{(\Upsilon f x. \text{if}(x \leq 0, 0, f(x - \text{sample})))}_{\Theta} 10$$

We can construct a partial antitone ranking function f as follows:

$$\begin{aligned} \Theta l &\mapsto l \\ \text{if}(l \leq 0, 0, \Theta(l - \text{sample})) &\mapsto l - 1/2 \\ 0 &\mapsto 0 \end{aligned}$$

with the constant antitone function $\epsilon(y) = 1/2$.

For a more complex example, consider the following “continuous random walk”

$$\underbrace{(\Upsilon f x. \text{if}(x \leq 0, 0, f(x - \text{sample}) \oplus_{1/2} f(x + \text{sample})))}_{\Xi} 10$$

Let $g(x) := \sqrt{2} - \frac{1}{2^{x-\frac{1}{2}}}$ and let ϵ be the function specified by

$$\epsilon(g(x)) = g(x) - \frac{1}{2} \left(g(x - \frac{1}{2}) + g(x + \frac{1}{2}) \right) = \frac{3 - 2\sqrt{2}}{2^{x+1}}.$$

We define a partial antitone ranking function as follows:

$$\begin{aligned} \Xi l &\mapsto g(l) \\ \text{if}(l \leq 0, 0, \Xi(l - \text{sample}) \oplus_{1/2} \Xi(l + \text{sample})) &\mapsto \frac{1}{2} \left(g(l - \frac{1}{2}) + g(l + \frac{1}{2}) \right) \\ r \leq 0, \quad r &\mapsto 0 \end{aligned}$$

Example 4 (Fair-in-the-limit random walk). [McIver et al. 2018, §5.3]

$$\underbrace{(\Upsilon f x. \text{if}(x \leq 0, 0, f(x-1) \oplus_{\frac{x}{2x+1}} f(x+1)))}_{\Xi} 10$$

To construct an antitone ranking function, we solve the recurrence relation:

$$\begin{cases} z_0 &= 0 \\ n \geq 1, \quad z_n &> \frac{n}{2n+1} z_{n-1} + \frac{n+1}{2n+1} z_{n+1}. \end{cases}$$

It turns out the solution to Example M_1 , $z_n = \frac{2^n-1}{2^{n-1}}$, is nearly a solution (for $n = 1$, we have equality instead).

Let $g(x) = \frac{2^n-1}{2^{n-1}}$. Then

$$\epsilon(g(x)) = g(x) - \frac{x}{2x+1} g(x-1) - \frac{x+1}{2x+1} g(x+1) = \frac{x-1}{2^x(2x+1)} \quad (8)$$

[ongoing]

◁ L0

Example 5 (Escaping spline). [McIver et al. 2018, §5.4]

$$\underbrace{(\Upsilon f x. \text{if}(x \leq 0, 0, 0 \oplus \frac{1}{x+1} f(x+1)))}_{\Xi} 10$$

7 Υ stopping times are finite

As before, fix a $M \in \Lambda^0$. Recall the random variables on (S, \mathcal{F}, μ) :

$$\begin{aligned} T_0(s) &:= 0 \\ T_{n+1}(s) &:= \min\{k \mid k > T_n(s), M_k(s) \text{ a value or of form } E[\Upsilon \lambda x. N]\} \end{aligned}$$

Lemma 10. T_0, T_1, T_2, \dots are an increasing sequence of stopping times adapted to $(\mathcal{F}_n)_{n \in \omega}$, and each T_i is bounded.

We first show that T_1 is bounded i.e. $T_1 \leq n$, for some $n \in \omega$.

A first proof idea is to construct a reduction argument to the strong normalisation of the simply-typed lambda-calculus. There is a classical transform of conditionals into the pure lambda-calculus. However, each evaluation of **sample** (almost surely) returns a different number, which complicates such a transform.

Given a closed SPCF term M , we transform it to a term $\ulcorner M \urcorner$ of the nondeterministic simply-typed lambda-calculus as follows. (We may assume that the nondeterministic calculus is generated from a base type ι with a ι -type constant symbol r , and a function symbol $\perp : A$ for each function type A .)

$$\begin{aligned} \ulcorner \text{sample} \urcorner &:= r \\ \ulcorner y \urcorner &:= y \\ \ulcorner \lambda y. N \urcorner &:= \lambda y. \ulcorner N \urcorner \\ \ulcorner M_1 M_2 \urcorner &:= \ulcorner M_1 \urcorner \ulcorner M_2 \urcorner \\ n \geq 0, \ulcorner f M_1 \dots M_n \urcorner &:= (\lambda z_1 \dots z_n. r) \ulcorner M_1 \urcorner \dots \ulcorner M_n \urcorner \\ \ulcorner \text{if}(B, M_1, M_2) \urcorner &:= (\lambda z. (\ulcorner M_1 \urcorner + \ulcorner M_2 \urcorner)) \ulcorner B \urcorner \\ \ulcorner \Upsilon f y. N \urcorner &:= \lambda y. \ulcorner N \urcorner [\perp / f] \\ \ulcorner \text{score}(N) \urcorner &:= \ulcorner N \urcorner \end{aligned}$$

In the above, variables z, z_1, \dots, z_n are assumed to be fresh; f ranges over primitive functions and numerals. In transforming $\text{score}(N)$, we set aside the requirement that score is undefined on negative numbers.

The idea is that $\ulcorner M \urcorner$ captures the initial, non-recursive operational behaviour of M , so that $\ulcorner M \urcorner$ simulates the reduction of M until the latter reaches a value, or a term of the form $E[\lambda x.N]$.

It is straightforward to see that for every trace $s \in S$, there is a reduction sequence of $\ulcorner M \urcorner$ that simulates (an initial subsequene of) the reduction of M under s ; and the simulation can be made lockstep.

Finally thanks to the following

Theorem 11 (de Groote). *The simply-typed nondeterministic lambda-calculus is strongly normalising.* \square

we have:

1. Every $\ulcorner M \urcorner$ -reduction terminates on reaching r , or a term of the shape $E[\perp N_1 \cdots N_n]$.
2. Further there is a finite bound (say l) on the length of such $\ulcorner M \urcorner$ -reduction sequences; and l bounds the stopping time T_1 .

[Note: de Groote [1994] considered a call-by-name version of a simply-typed nondeterministic lambda-calculus, and proved it strongly normalising, using a reducibility argument à la Tait. One should check that it extends to the call-by-value calculus.] ◀ L0

8 Relativised completeness

If every term reachable from a closed SPCF program is AST then the program is also AST, but not conversely (because a diverging term may be reachable with probability 0 from an AST term).

We aim to prove the following conjecture:

Given a closed SPCF term M , if every term reachable from M is AST then M admits an antitone ranking function.

- (1) Given an a.s. finite stopping time T (meaning: $T < \infty$ a.s.) adapted to a filtration, there exist supermartingale $(Y_n)_{n \in \omega}$ adapted to the same filtration and antitone function ϵ such that $(Y_n)_{n \in \omega}$ is an antitone strict supermartingale w.r.t. (ϵ, T) .
- (2) Let $(Y_n)_{n \in \omega}$ be an antitone strict supermartingale w.r.t. (ϵ, T) . Then M admits an antitone ranking function.

[Is (1) too general for the conjecture? Claim (1) assumes an arbitrary a.s. finite stopping time T . This seems too general for the conjecture, which assumes a “strongly” AST $M \in \Lambda^0$. Regardless of the validity of (1), in order to prove the conjecture, I think we would need a more specific version of (1) that assumes a stopping time about terminating runs of M — such as the stopping time ◀ L0

of definition (5) which only counts Y -reduction steps — so that the antitone-strict supermartingale that is constructed has the necessary structure to yield a ranking function for M .]

For 1, consider the following:

Proof. If T is bounded by b a.s. (for n constant), the statement is trivially true by taking ϵ to be constantly 1, and $Y_n = b - n$. Otherwise, let $(T_n)_{n \in \omega}$ be defined recursively such that $T_0 = 0$, $T_{n+1} > T_n$ and $\mathbb{P}[T > T_{n+1}] \leq \frac{1}{4} \mathbb{P}[T > T_n]$. [Since T is a stopping time, we can assume that $T_n \in \mathbb{N}$. Problem with T_1 : we require $T_1 > 0$ which may contradict the requirement $\mathbb{P}[T < T_1] \leq \frac{1}{4} \mathbb{P}[T < 0] = 0$ for certain T , such as those that satisfy $\mathbb{P}[T < 1] > 0$. We could define T_1 as the least number satisfying $\mathbb{P}[T < T_1] \geq \frac{1}{4}$.] [Sorry, I just wrote it wrong. I assume this $(\mathbb{P}[T > T_n])$ instead of $\mathbb{P}[T < T_n]$] resolves your concern.] We will then define a supermartingale $(Y_n)_{n \in \omega}$ and nonrandom sequence $(Y'_n)_{n \in \omega}$ such that $Y_n = 0$ iff $T < n$ a.s. [I see no reason that this can't just be true everywhere, as this is part of the definition of Y_n . Likewise for the other “a.s.” you added.], $Y_n = Y'_n$ iff $T \geq n$ a.s., and $Y'_n \geq 2^k$ for $n \geq T_k$.
 $[(T_n)_{n \in \omega}$ and $(Y'_n)_{n \in \omega}$ are ω -sequences of numbers. There is a convention that random variables are denoted by uppercase letters. I suggest we write them as $(t_n)_{n \in \omega}$ and $(y'_n)_{n \in \omega}$ respectively: in this way, we emphasise that they are sequences of numbers, and avoid any possible confusion with r.v.]

The antitone function [It is good practice not to begin a sentence with a maths expression / symbol.] ϵ is defined piecewise and recursively in such a way as to force all the necessary constraints to hold:

$$\begin{aligned} \text{for } x \in [0, 1), \epsilon(x) &= 1 \\ \text{for } x \in [2^k, 2^{k+1}), \epsilon(x) &= \min \left(\epsilon(2^{k-1}), \frac{2^k}{T_{k+1} - T_k} \right). \end{aligned}$$

This ensures that

- ϵ is weakly decreasing.
- ϵ is bounded below by 0.
- For $x \geq 2^k$, $\epsilon(x) \leq \frac{2^k}{T_{k+1} - T_k}$. [This would imply $\frac{2^k}{T_{k+1} - T_k} \geq \epsilon(2^{k-1})$. Why?][Corrected: it should have been min, not max.]

Y'_n is then defined recursively by

$$Y'_0 = 2, \quad Y'_{n+1} = (Y'_n - \epsilon(Y'_n)) \frac{\mathbb{P}[T > n]}{\mathbb{P}[T > n+1]}.$$

The fact $Y'_{T_k} \geq 2^{k+1}$ is proven by induction on k . For base case, $2 \geq 2$, as required. For the inductive case $k+1$, we first do another induction to prove that $Y'_n \geq 2^k$ for all $T_k \leq n \leq T_{k+1}$. The base case of this inner induction

follows from the outer induction hypothesis a fortiori. take the greatest k such that $n > T_k$. By the induction hypothesis, $Y'_{T_k} \geq 2^k$.

$$\begin{aligned}
Y'_n &= Y'_{T_k} \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} - \sum_{m=T_k}^{n-1} \epsilon(Y'_m) \frac{\mathbb{P}[T > m+1]}{\mathbb{P}[T > n]} \\
&\geq \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} (Y'_{T_k} - \sum_{m=T_k}^{n-1} \epsilon(Y'_m)) \\
&\geq \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} (Y'_{T_k} - \sum_{m=T_k}^{n-1} \frac{2^k}{T_{k+1} - T_k}) \\
&= \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} (Y'_{T_k} - \frac{2^k(n - T_k)}{T_{k+1} - T_k}) \\
&\geq \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} (Y'_{T_k} - 2^k) \\
&\geq \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} (2^{k+1} - 2^k) \\
&= \frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > n]} 2^k \\
&\geq 2^k
\end{aligned}$$

as required. Substituting $n = T_{k+1}$ and using the fact that $\frac{\mathbb{P}[T > T_k]}{\mathbb{P}[T > T_{k+1}]} > 4$, the same reasoning gives $Y'_{T_{k+1}} \geq 2^{k+2}$, as required.

Although the stronger induction hypothesis was necessary for the induction to work, the only reason this is needed is to prove that $Y'_n > 0$ for all n , which implies that $Y_n \geq 0$. The other condition for (Y_n) to be an antitone-strict supermartingale is [The inference from the third to the fourth line is actually incorrect if \mathcal{F} is not coarsest filtration to which T is adapted. This error may require a substantial change to the proof to correct, but I thought you'd want to see it anyway rather than waiting for me to fix it.] ◀ AK-R

$$\begin{aligned}
&\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \\
&= \mathbb{E}[Y'_{n+1} 1_{T \geq n+1} \mid \mathcal{F}_n] \\
&= Y'_{n+1} \mathbb{E}[1_{T \geq n+1} \mid \mathcal{F}_n] \\
&= Y'_{n+1} 1_{T \geq n} \frac{\mathbb{P}[T > n+1]}{\mathbb{P}[T > n]} \\
&= (Y'_n - \epsilon(Y'_n)) 1_{T \geq n} \\
&= Y_n - \epsilon(Y'_n) 1_{T \geq n}
\end{aligned}$$

therefore $(Y_n)_n$ is an antitone strict supermartingale with respect to (T, ϵ) . \square

[Ongoing.]

◀ L0

[I have some incomplete thoughts in the direction of proving completeness (both ◁ AK-R in the general case and for terms). For each non-null event measurable with respect to some step in the filtration (or reachable term in the case of proving termination of terms), there is a weakly decreasing function $\mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ with limit 0 representing the conditional probability of having not yet terminated after n more steps. These functions can be linearly combined and have a partial order, and form a sort of generalised supermartingale. It is then necessary to map these to real numbers, which can be done by taking the limit with respect to some ultrafilter on the natural numbers of the ratios between them, and picking some arbitrary starting point (e.g. setting the value for the initial term to 1). The problem with this is that some terms could get assigned the value 0, so maybe it's necessary to take equivalence classes of terms whose values have ratios in the open interval $(0, \infty)$ or something.]

Lemma 11. *If M is strongly AST (i.e. every reachable term from N is AST) then for each $N \in Rch(M)$, there is a non-increasing function $\rho_N : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ such that $\lim_{n \rightarrow \infty} \rho_N(n) = 0$.*

Take $N \in Rch(M)$. Assume N is not a value. Define $\rho_N : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ by setting

$$\rho_N(n) := \mu(\{s \mid \pi_0(\text{red}^n(N, s)) \text{ non-value}\})$$

I.e. $\rho_N(n)$ is the probability that N is not yet terminated after n more reduction steps. [E.g. writing $I^{m+1} := I^m I$ where $I^1 := I$, we have $\rho_{I^m}(n) = 1$ if ◁ L0 $n < m - 1$, and 0 otherwise.] Let $|\rho_N| := \sum_{n \in \omega} \rho_N(n)$.

Conjecture 1. $|\rho_N|$ is finite.

Define $\hat{\rho}_N(n) := \frac{\rho_N(n)}{|\rho_N|}$. Assuming the conjecture, $\hat{\rho}_N$ is a probability mass function (or subprobability mass function if we set $\hat{\rho}_N$ to be 0 in case N is a value).

[Define r.v. $Z_n(s) := \hat{\rho}_N$ where $\text{red}^n(M, s) = (N, -)$. Then $(Z_n)_{n \in \omega}$ is a (gen- ◁ L0 eralised) supermartingale. I am not sure if we need this.]

We want to map functions $\hat{\rho}_N$ where N ranges over $Rch(M)$ to real numbers.

The *total variational distance* between probability measures μ, ν on the same measurable space (Ω, \mathcal{F}) , $\|\mu - \nu\|_{\text{TV}} := \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$. The definition is presented as the supremum over all measurable subsets of the state space Ω , which is not a convenient way to calculate the distance. Fortunately there are two useful alternative characterisations:

1. as a simple sum over the state space: $\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$
2. via *coupling*, a different probabilistic interpretation, i.e., $\|\mu - \nu\|_{\text{TV}}$ measures how close to identical we can force two random variables to be, such that one realises μ and the other realises ν

$$\|\mu - \nu\|_{\text{TV}} = \inf \{\mathbb{P}(X \neq Y) \mid (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

[NOTE. *Coupling* is a powerful method and proof technique in probability theory based on the simple idea of comparing random variables with each other. It has been applied to prove limit theorems, derive inequalities, and obtain approximations, useful in a host of situations, ranging across random walks, Markov chains, percolation, interacting particle systems, and diffusions. A easily accessible reference is [den Hollander 2012]; see also [Levin and Peres 2017, Ch. 4 & 5].] ◁ L0

A candidate for ranking function: for $N \in Rch(M)$, $r(N) := 1 - \|\hat{\rho}_M - \hat{\rho}_N\|_{TV}$. Thus, intuitively, the “further” N is from M , the closer the value $r(N)$ is to 0. [Problem: We want to show that if $\text{red}(N, s) = (N', -)$ by contracting a Y-redex, then $r(N) > r(N')$. However we have $|\rho_{N'}| < |\rho_N|$ and $\rho_{N'}(n) \leq \rho_N(n)$.] ◁ L0
 [When proving that a candidate function is a ranking function, try making use of the sparse ranking function theorem.] ◁ L0

References

- Agrawal, S., Chatterjee, K., and Novotný, P. (2018). Lexicographic ranking supermartingales: an efficient approach to termination of probabilistic programs. *PACMPL*, 2(POPL):34:1–34:32.
- Ash, R. B. and Doléans-Dade, C. (2000). *Probability and measure Theory*. Harcourt Academic Press.
- Borgström, J., Lago, U. D., Gordon, A. D., and Szymczak, M. (2016). A lambda-calculus foundation for universal probabilistic programming. In *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, ICFP 2016, Nara, Japan, September 18-22, 2016*, pages 33–46.
- Breuvar, F. and Lago, U. D. (2018). On intersection types and probabilistic lambda calculi. In *Proceedings of the 20th International Symposium on Principles and Practice of Declarative Programming, PPDP 2018, Frankfurt am Main, Germany, September 03-05, 2018*, pages 8:1–8:13.
- Chatterjee, K., Fu, H., Novotný, P., and Hasheminezhad, R. (2016). Algorithmic analysis of qualitative and quantitative termination problems for affine probabilistic programs. In Bodík, R. and Majumdar, R., editors, *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2016, St. Petersburg, FL, USA, January 20 - 22, 2016*, pages 327–342. ACM.
- Culpepper, R. and Cobb, A. (2017). Contextual equivalence for probabilistic programs with continuous random variables and scoring. In Yang, H., editor, *Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, volume 10201 of *Lecture Notes in Computer Science*, pages 368–392. Springer.

- de Groote, P. (1994). Strong normalization in a non-deterministic typed lambda-calculus. In Nerode, A. and Matiyasevich, Y. V., editors, *Logical Foundations of Computer Science, Third International Symposium, LFCS'94, St. Petersburg, Russia, July 11-14, 1994, Proceedings*, volume 813 of *Lecture Notes in Computer Science*, pages 142–152. Springer.
- den Hollander, F. (2012). Probability theory: The coupling method. unpublished notes.
- Fioriti, L. M. F. and Hermanns, H. (2015). Probabilistic termination: Soundness, completeness, and compositionality. In Rajamani, S. K. and Walker, D., editors, *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2015, Mumbai, India, January 15-17, 2015*, pages 489–501. ACM.
- Hark, M., Kaminski, B. L., Giesl, J., and Katoen, J. (2020). Aiming low is harder: induction for lower bounds in probabilistic program verification. *PACMPL*, 4(POPL):37:1–37:28.
- Kennaway, J., Klop, J. W., Sleep, M., and de Vries, F. (1996). Infinitary lambda calculus.
- Lago, U. D. and Grellois, C. (2017). Probabilistic termination by monadic affine sized typing. In *Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, pages 393–419.
- Levin, D. A. and Peres, Y. (2017). *Markov Chains and Mixing Times*. American Mathematical Society, second edition.
- Mak, C., Ong, C.-H. L., and Paquet, H. (2020a). Universal probabilistic programming and almost everywhere-differentiable densities. Working draft, version of 18 January 2020.
- Mak, C., Ong, C. L., Paquet, H., and Wagner, D. (2020b). Densities of almost-surely terminating probabilistic programs are differentiable almost everywhere. *CoRR*, abs/2004.03924.
- McIver, A., Morgan, C., Kaminski, B. L., and Katoen, J. (2018). A new proof rule for almost-sure termination. *PACMPL*, 2(POPL):33:1–33:28.
- Schreuder, A. and Ong, C.-H. L. (2019). Polynomial probabilistic invariants and the optional stopping theorem. *CoRR*. To appear on arXiv. See also <http://www.cs.ox.ac.uk/people/luke.ong/personal/publications/tacas.pdf>.
- [22 Jan:

◀ L0

Further directions An obvious next step is to extend the result to the **score** construct.

The following are highly topical, and could form the basis of an interesting and novel DPhil thesis.

1. Devise methods for proving (positive) a.s. termination.
 - For example, develop a type system satisfying the property: if a term is typable then it is (positively) a.s. terminating. See [Breuvart and Lago 2018; Lago and Grellois 2017].
 - Another approach is to develop algorithms that synthesise ranking supermartingales, following, for example, [Agrawal et al. 2018].
2. Develop principles (e.g. in the form of “proof rules”) for reasoning about (positively) a.s. termination, in the style of [McIver et al. 2018].
3. Design algorithms that synthesise probabilistic invariants (*qua* martingales), à la Schreuder and Ong [2019]; see also [Hark et al. 2020].

]