Stability and probabilistic programs

Thomas Ehrhard Michele Pagani Christine Tasson

Institut de Recherche en Informatique Fondamentale
Université Paris Diderot — Paris 7 (FR)
{ehrhard, pagani, tasson}@irif.fr

Workshop PPS – Paris 2017

PCF with discrete probabilistic distributions

$$\frac{}{\Delta, x : A \vdash x : A} \qquad \frac{\Delta, x : A \vdash M : B}{\Delta \vdash \lambda x^A.M : A \Rightarrow B} \qquad \frac{\Delta \vdash M : A \Rightarrow B \quad \Delta \vdash N : A}{\Delta \vdash (MN) : B} \qquad \frac{\Delta \vdash M : A \Rightarrow A}{\Delta \vdash (\mathtt{Y} M) : A}$$

$$\frac{r \in \mathbb{N}}{\Delta \vdash \underline{n} : \mathtt{Nat}} \qquad \frac{\Delta \vdash M : \mathtt{Nat}}{\Delta \vdash \mathtt{succ}(M) : \mathtt{Nat}} \qquad \frac{\Delta \vdash M : \mathtt{Nat}}{\Delta \vdash \mathtt{pred}(M) : \mathtt{Nat}} \qquad \frac{\Delta \vdash P : \mathtt{Nat} \quad \Delta \vdash M : \mathtt{Nat} \quad \Delta \vdash N : \mathtt{Nat}}{\Delta \vdash \mathtt{ifz}(P, M, N) : \mathtt{Nat}}$$

$$\frac{}{\Delta \vdash \mathtt{Coin} : \mathtt{Nat}} \qquad \frac{\Delta \vdash M : \mathtt{Nat} \quad \Delta, x : \mathtt{Nat} \vdash N : \mathtt{Nat}}{\Delta \vdash \mathtt{let}(x, M, N) : \mathtt{Nat}}$$

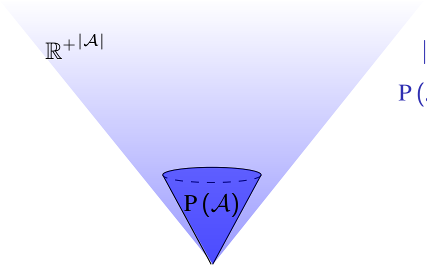
Operational semantics **Red** : $\Lambda \times \Lambda \to [0, 1]$

$$\mathbf{Red}(M, N) = \begin{cases} \delta_N(\{E[T]\}) & \text{if } M = E[R],\ R \to T \text{ and } R \neq \mathtt{Coin}, \\ \frac{1}{2}\delta_N(\{E[\underline{0}], E[\underline{1}]\}) & \text{if } M = E[\mathtt{Coin}], \\ \delta_N(\{M\}) & \text{if } M \text{ normal form.} \end{cases}$$

$$\mathbf{Prob}(M, V) = \sup_{n=0}^{\infty}\left(\mathbf{Red}^n(M, V)\right)$$

How do we model types,
e.g. $[\![\mathtt{Nat}]\!]$, $[\![\mathtt{Nat} \Rightarrow \mathtt{Nat}]\!]$?

A probabilistic coherence space $\mathcal{A} = (|\mathcal{A}|, \mathrm{P}(\mathcal{A}))$



$|\mathcal{A}|$ a countable set, called *web*

$\mathrm{P}(\mathcal{A}) \subseteq (\mathbb{R}^+)^{|\mathcal{A}|}$ s.t. $\mathrm{P}(\mathcal{A})^{\perp\perp} = \mathrm{P}(\mathcal{A})$, with:

$$P^{\perp} = \left\{ v \in \mathbb{R}^{+|\mathcal{A}|} \ ; \ \forall u \in P, \sum_{a \in |\mathcal{A}|} v_a u_a \leq 1 \right\}$$

(+completeness, boundedness)

Example

$$|\mathtt{Unit}| = \{\mathtt{skip}\} \qquad \mathrm{P}(\mathtt{Unit}) = [0,1]$$

$$|\mathtt{Bool}| = \{\mathtt{t}, \mathtt{f}\} \qquad \mathrm{P}(\mathtt{Bool}) = \{(p,q) \ ; \ p+q \leq 1\}$$

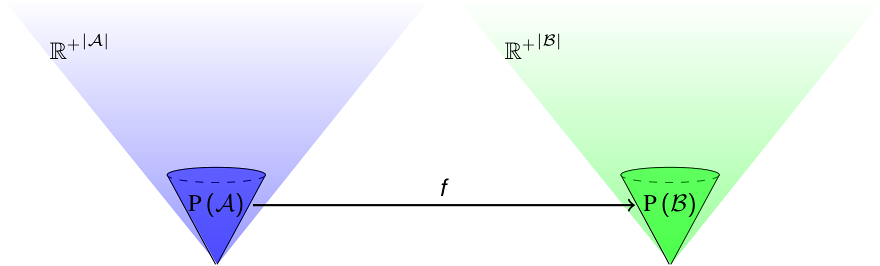
$$|\mathtt{Nat}| = \{0,1,2,3,\dots\} \qquad \mathrm{P}(\mathtt{Nat}) = \{v \in [0,1]^{\mathbb{N}} \ ; \ \textstyle\sum_n v_n \leq 1\}$$

$$|\mathtt{Bool} \Rightarrow \mathtt{Unit}| = \mathcal{M}_{\mathrm{f}}(\{\mathtt{t}, \mathtt{f}\})$$

$$\mathrm{P}(\mathtt{Bool} \Rightarrow \mathtt{Unit}) = \left\{ v \ ; \ \forall p \in [0,1], \sum_{n,m=0}^{\infty} v_{[\mathtt{t}^n, \mathtt{f}^m]} p^n (1-p)^m \leq 1 \right\}$$

How do we model programs ?

e.g. $[\![\Gamma \vdash M : A]\!] : [\![\Gamma]\!] \mapsto [\![A]\!]$



The map f is given by a matrix in $\mathbb{R}^{+\mathcal{M}_{\mathrm{f}}(|\mathcal{A}|)\times|\mathcal{B}|}$, i.e.:

$$f(x)_b = \sum_{[a_1^{n_1},\dots,a_k^{n_k}]\in\mathcal{M}_{\mathrm{f}}(|\mathcal{A}|)} f_{[a_1^{n_1},\dots,a_k^{n_k}],b}\, x_{a_1}^{n_1}\dots x_{a_k}^{n_k}$$

We require that: $\forall x \in \mathrm{P}(\mathcal{A}), f(x) \in \mathrm{P}(\mathcal{B})$.

Example

Let $T = \mathtt{Y}(\lambda fx.\mathtt{ifz}(x, \mathtt{ifz}(\mathtt{pred}(x), \underline{0}, fx), \mathtt{ifz}(x, \underline{0}, fx)))$ then:

$$[\![T]\!]_{\underline{0}} = \sum_{n,m=0}^{\infty} \frac{2(n+m)!}{n!m!} x_{\underline{0}}^{2n+1} x_{\underline{1}}^{2m+1}, \qquad [\![T]\!]_{\underline{n+1}} = 0$$

What do we gain with Probabilistic Coherence Spaces?

The benefits of having a (fully-abstract!) model

- Compositional definition of contextual equivalence:

Theorem (Ehrhard,P.,Tasson 2014)

For every type A and terms $P, Q : A$,

$$\forall C \textit{ context}, \ \textbf{Prob}(C[P], \underline{0}) = \textbf{Prob}(C[Q], \underline{0}) \quad \textit{iff} \quad [\![P]\!] = [\![Q]\!]$$

  - A variant for call-by-push-value in (Ehrhard-Tasson 2017)

- More tools for program analysis:
  - derivation, Taylor expansion, norm, distance…

How to extend Probabilistic Coherence Spaces to continuous data types?

PCF with continuous probabilistic distributions as well

$$\frac{}{\Delta, x : A \vdash x : A} \qquad \frac{\Delta, x : A \vdash M : B}{\Delta \vdash \lambda x^A.M : A \Rightarrow B} \qquad \frac{\Delta \vdash M : A \Rightarrow B \quad \Delta \vdash N : A}{\Delta \vdash (MN) : B} \qquad \frac{\Delta \vdash M : A \Rightarrow A}{\Delta \vdash (\mathrm{Y} M) : A}$$

$$\frac{r \in \mathbb{R}}{\Delta \vdash \underline{r} : \mathtt{Real}} \qquad \frac{f \text{ meas. } \mathbb{R}^n \to \mathbb{R} \quad \Delta \vdash M_i : \mathtt{Real}, \forall i \leq n}{\Delta \vdash \underline{f}(M_1, \dots, M_n) : \mathtt{Real}} \qquad \frac{\Delta \vdash P : \mathtt{Real} \quad \Delta \vdash M : \mathtt{Real} \quad \Delta \vdash N : \mathtt{Re}}{\Delta \vdash \mathtt{ifz}(P, M, N) : \mathtt{Real}}$$

$$\frac{}{\Delta \vdash \mathtt{sample} : \mathtt{Real}} \qquad \frac{\Delta \vdash M : \mathtt{Real} \quad \Delta, x : \mathtt{Real} \vdash N : \mathtt{Real}}{\Delta \vdash \mathtt{let}(x, M, N) : \mathtt{Real}}$$

Operational semantics **Red** : $\Lambda \times \Sigma_\Lambda \to [0, 1]$

$$\mathbf{Red}(M, U) = \begin{cases} \delta_{E[N]}(U) & \text{if } M = E[R],\ R \to N \text{ and } R \neq \mathtt{sample}, \\ \lambda\{r \in [0,1] \text{ s.t. } E[\underline{r}] \in U\} & \text{if } M = E[\mathtt{sample}], \\ \delta_M(U) & \text{if } M \text{ normal form.} \end{cases}$$

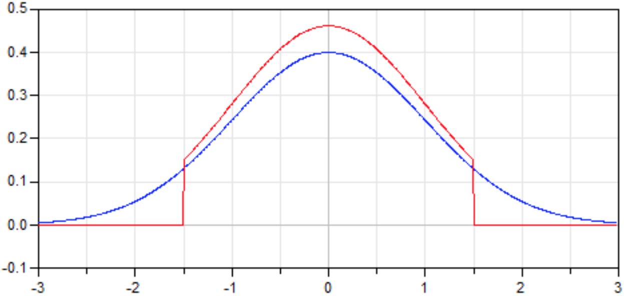
$$\mathbf{Prob}(M, U) = \sup_{n=0}^{\infty} \left(\mathbf{Red}^n(M, U) \right)$$

Examples

$$\texttt{Coin} = \texttt{let}(x, \texttt{sample}, x \leq 0.5)$$

$$\texttt{normal} = \texttt{let}(x, \texttt{sample}, \texttt{let}(y, \texttt{sample}, \underline{(-2\log(x))}^{\frac{1}{2}}\ \underline{\cos(2\pi y)}))$$

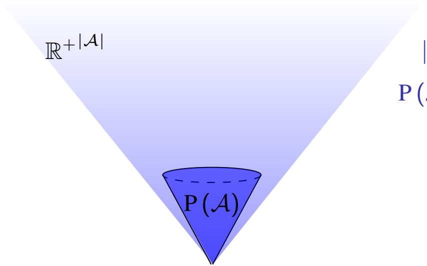
$$\texttt{truncated\_normal} = \texttt{Y}(\lambda y.\texttt{let}(x, \texttt{normal}, \texttt{ifz}(x \in [-1.5, 1.5], x, y)))$$



How do we model types,

e.g. the type `Real` of real numbers?

A probabilistic coherence space $\mathcal{A} = (|\mathcal{A}|, \mathrm{P}(\mathcal{A}))$



$|\mathcal{A}|$ a countable set, called *web*

$\mathrm{P}(\mathcal{A}) \subseteq (\mathbb{R}^+)^{|\mathcal{A}|}$ s.t. $\mathrm{P}(\mathcal{A})^{\perp\perp} = \mathrm{P}(\mathcal{A})$, with:

$$P^\perp = \left\{ v \in \mathbb{R}^{+|\mathcal{A}|} \; ; \; \forall u \in P, \sum_{a \in |\mathcal{A}|} v_a u_a \leq 1 \right\}$$

(+completeness, boundedness)

Example

$|\mathtt{Unit}| = \{\mathtt{skip}\}$ $\quad$ $\mathrm{P}(\mathtt{Unit}) = [0,1]$

$|\mathtt{Bool}| = \{\mathtt{t}, \mathtt{f}\}$ $\quad$ $\mathrm{P}(\mathtt{Bool}) = \{(p,q) \; ; \; p + q \leq 1\}$

$|\mathtt{Nat}| = \{0, 1, 2, 3, \dots\}$ $\quad$ $\mathrm{P}(\mathtt{Nat}) = \{v \in [0,1]^{\mathbb{N}} \; ; \; \sum_n v_n \leq 1\}$

$$|\mathtt{Bool} \Rightarrow \mathtt{Unit}| = \mathcal{M}_{\mathrm{f}}(\{\mathtt{t}, \mathtt{f}\})$$

$$\mathrm{P}(\mathtt{Bool} \Rightarrow \mathtt{Unit}) = \left\{ v \; ; \; \forall p \in [0,1], \sum_{n,m=0}^{\infty} v_{[\mathtt{t}^n, \mathtt{f}^m]} p^n (1-p)^m \leq 1 \right\}$$

Complete cones (Selinger 2004)

Normed cone: an $\mathbb{R}^+$-semimodule P with $\mathbb{R}^+$-valued function $\|\_\|_P$ s.t.:

- $x + y = x + y' \Rightarrow y = y'$
- $\|\alpha x\|_P = \alpha \|x\|_P$
- $\|x\|_P = 0 \Rightarrow x = 0$
- $\|x + x'\|_P \leq \|x\|_P + \|x'\|_P$
- $\|x\|_P \leq \|x + x'\|_P$,

where $x \leq_P y$ is defined as $\exists z \in P, x + z = y$.

Complete cone: a normed cone P s.t.:

- the unit ball $\mathcal{B}(P) = \{x \in P \;;\; \|x\|_P \leq 1\}$ is complete wrt. $\leq_P$.

Any $\mathcal{A} = (|\mathcal{A}|, \mathrm{P}(\mathcal{A}))$ gives a complete cone:

- $\bigcup_{\alpha \in \mathbb{R}^+} \alpha \mathrm{P}(\mathcal{A})$,
- $\|x\|_{\mathcal{A}} = \inf\{\alpha > 0 \;;\; \frac{1}{\alpha} x \in \mathrm{P}(\mathcal{A})\}$.

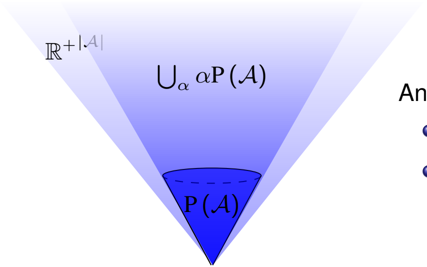
Normed cone: an $\mathbb{R}^+$-semimodule P with $\mathbb{R}^+$-valued function $\|\_\|_P$ s.t.:

- $x + y = x + y' \Rightarrow y = y'$
- $\|\alpha x\|_P = \alpha \|x\|_P$
- $\|x\|_P = 0 \Rightarrow x = 0$
- $\|x + x'\|_P \leq \|x\|_P + \|x'\|_P$
- $\|x\|_P \leq \|x + x'\|_P$,

where $x \leq_P y$ is defined as $\exists z \in P, x + z = y$.

Complete cone: a normed cone P s.t.:

- the unit ball $\mathcal{B}(P) = \{x \in P \,;\ \|x\|_P \leq 1\}$ is complete wrt. $\leq_P$.



Any $\mathcal{A} = (|\mathcal{A}|, \mathrm{P}(\mathcal{A}))$ gives a complete cone:

- $\bigcup_{\alpha \in \mathbb{R}^+} \alpha\mathrm{P}(\mathcal{A})$,
- $\|x\|_\mathcal{A} = \inf\{\alpha > 0 \,;\ \frac{1}{\alpha} x \in \mathrm{P}(\mathcal{A})\}$.

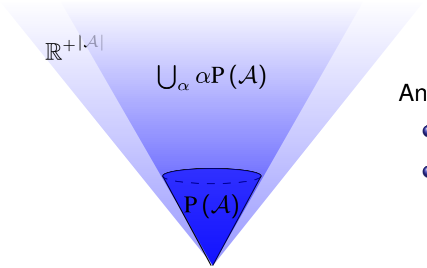
Normed cone: an $\mathbb{R}^+$-semimodule P with $\mathbb{R}^+$-valued function $\|\_\|_P$ s.t.:

- $x + y = x + y' \Rightarrow y = y'$
- $\|\alpha x\|_P = \alpha \|x\|_P$
- $\|x\|_P = 0 \Rightarrow x = 0$
- $\|x + x'\|_P \leq \|x\|_P + \|x'\|_P$
- $\|x\|_P \leq \|x + x'\|_P$,

where $x \leq_P y$ is defined as $\exists z \in P, x + z = y$.

Complete cone: a normed cone P s.t.:

- the unit ball $\mathcal{B}(P) = \{x \in P \,;\, \|x\|_P \leq 1\}$ is complete wrt. $\leq_P$.



Any $\mathcal{A} = (|\mathcal{A}|, \mathrm{P}(\mathcal{A}))$ gives a complete cone:

- $\bigcup_{\alpha \in \mathbb{R}^+} \alpha \mathrm{P}(\mathcal{A})$,
- $\|x\|_\mathcal{A} = \inf\{\alpha > 0 \,;\, \frac{1}{\alpha} x \in \mathrm{P}(\mathcal{A})\}$.

The complete cone $\mathbf{Meas}(\mathbb{R})$ of the bounded measures over $\mathbb{R}$

Given a measurable space (X, Σ_X), we define:

$$\mathbf{Meas}(X, \Sigma_X) = \{\mu : \Sigma_X \mapsto \mathbb{R}^+ \; ; \; \mu \text{ is a (bounded) measure}\}$$

- $\mathbf{Meas}(X, \Sigma_X)$ is endowed with a structure of complete cone:

$$(\mu + \mu')(U) = \mu(U) + \mu'(U), \qquad (\alpha\mu)(U) = \alpha\mu(U), \qquad \|\mu\| = \mu(X)$$

- In particular,

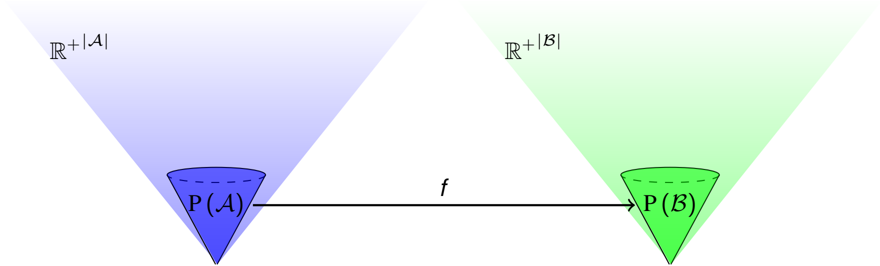
$$\mathcal{B}(\mathbf{Meas}(X, \Sigma_X)) = \text{ the set of sub-probability distributions over } (X, \Sigma_X).$$

We denote by $\mathbf{Meas}(\mathbb{R})$ the complete cone given by the Lebesgue σ-algebra over $\mathbb{R}$.

$$[\![\texttt{Real}]\!] = \mathbf{Meas}(\mathbb{R})$$

How do we model programs, e.g $[\![x : \texttt{Real} \vdash M : \texttt{Real}]\!] : \mathbf{Meas}(\mathbb{R}) \mapsto \mathbf{Meas}(\mathbb{R})$?

A morphism f: $\mathcal{A} \mapsto \mathcal{B}$



The map f is given by a matrix in $\mathbb{R}^{+\mathcal{M}_f(|\mathcal{A}|)\times|\mathcal{B}|}$, i.e.:

$$f(x)_b = \sum_{[a_1^{n_1},\ldots,a_k^{n_k}]\in\mathcal{M}_f(|\mathcal{A}|)} f_{[a_1^{n_1},\ldots,a_k^{n_k}],b}\, x_{a_1}^{n_1}\ldots x_{a_k}^{n_k}$$

We require that: $\forall x \in \mathrm{P}(\mathcal{A}), f(x) \in \mathrm{P}(\mathcal{B})$.

Example

Let $T = \mathtt{Y}(\lambda fx.\mathtt{ifz}(x, \mathtt{ifz}(\mathtt{pred}(x), \underline{0}, fx), \mathtt{ifz}(x, \underline{0}, fx)))$ then:

$$[\![T]\!]_{\underline{0}} = \sum_{n,m=0}^{\infty} \frac{2(n+m)!}{n!m!} x_{\underline{0}}^{2n+1} x_{\underline{1}}^{2m+1}, \qquad [\![T]\!]_{\underline{n+1}} = 0$$

$$\mathcal{B}(P \Rightarrow Q) = \{ f : \mathcal{B}P \to \mathcal{B}Q \; ; \; f \text{ Scott-continuous} \}$$

- ✔ it yields a complete cone $\bigcup_\alpha \alpha\mathcal{B}(P \Rightarrow Q)$ with the operations defined point-wise,
- ✔ it gives a cartesian category:

$$P \times Q = \{(x, y) \; ; \; x \in P, y \in Q\}, \quad \|(x, y)\|_{P\times Q} = \max(\|x\|_P, \|y\|_P)$$

- ✘ it is not cartesian closed:

Example ($\texttt{wpor} : \texttt{Unit} \times \texttt{Unit} \Rightarrow \texttt{Unit}$)

$$\begin{array}{ccc} [0,1] \times [0,1] & & [0,1] \\ (x \quad , \quad y) & \mapsto & x + y - xy \end{array}$$

- $\texttt{wpor}$ is a Scott-continuous function, so in $\texttt{Unit} \times \texttt{Unit} \Rightarrow \texttt{Unit}$
- however, its currying $\lambda x.\lambda y.\texttt{wpor}$ is not Scott-continuous,
- in fact, it is neither non-decreasing in $\texttt{Unit} \Rightarrow \texttt{Unit} \Rightarrow \texttt{Unit}$:
  - $(\lambda x.\lambda y.\texttt{wpor})1 \not\geq_{\texttt{Unit}\Rightarrow\texttt{Unit}} (\lambda x.\lambda y.\texttt{wpor})0$
  - in fact, $(\lambda x.\lambda y.\texttt{wpor})1 - (\lambda x.\lambda y.\texttt{wpor})0$
    which is $y \mapsto 1 - y$
    is not non-decreasing in y, so not in $\texttt{Unit} \Rightarrow \texttt{Unit}$.

An instructive failure: Scott-continuous functions

$$\mathcal{B}(P \Rightarrow Q) = \{ f : \mathcal{B}P \to \mathcal{B}Q \; ; \; f \text{ Scott-continuous} \}$$

- ✔ it yields a complete cone $\bigcup_\alpha \alpha\mathcal{B}(P \Rightarrow Q)$ with the operations defined point-wise,
- ✔ it gives a cartesian category:

$$P \times Q = \{(x, y) \; ; \; x \in P, y \in Q\}, \quad \|(x, y)\|_{P \times Q} = \max(\|x\|_P, \|y\|_P)$$

- ✘ it is not cartesian closed:

Example ($\mathtt{wpor} : \mathtt{Unit} \times \mathtt{Unit} \Rightarrow \mathtt{Unit}$)

$$\begin{array}{ccc} [0,1] \times [0,1] & & [0,1] \\ (x \quad , \quad y) & \mapsto & x + y - xy \end{array}$$

- $\mathtt{wpor}$ is a Scott-continuous function, so in $\mathtt{Unit} \times \mathtt{Unit} \Rightarrow \mathtt{Unit}$
- however, its currying $\lambda x.\lambda y.\mathtt{wpor}$ is not Scott-continuous,
- in fact, it is neither non-decreasing in $\mathtt{Unit} \Rightarrow \mathtt{Unit} \Rightarrow \mathtt{Unit}$:
  - $(\lambda x.\lambda y.\mathtt{wpor})1 \not\geq_{\mathtt{Unit} \Rightarrow \mathtt{Unit}} (\lambda x.\lambda y.\mathtt{wpor})0$
  - in fact, $(\lambda x.\lambda y.\mathtt{wpor})1 - (\lambda x.\lambda y.\mathtt{wpor})0$
    which is $y \mapsto 1 - y$
    is not non-decreasing in y, so not in $\mathtt{Unit} \Rightarrow \mathtt{Unit}$.

Non-decreasingness of all iterated differences i.e. absolute monotonicity

Given a function $f : \mathcal{B}P \to \mathcal{B}Q$, we say:

- **f 0-non-decreasing:** whenever f is non-decreasing,
- **f $(n+1)$-non-decreasing:** whenever f is non-decreasing and $\forall x \in P$, the function

$$\Delta_x f : x' \mapsto f(x + x') - f(x')$$

  is n-non-decreasing (of domain $\{x' \in P \ ; \ x' + x \in \mathcal{B}P\}$).
- **f absolutely monotone:** whenever f n-non-decreasing for every $n \in \mathbb{N}$.

Example (wpor)

$\texttt{wpor} : (x, y) \mapsto x + y - xy$ is not absolutely monotone (in fact not 1-non-decreasing).

Theorem (Ehrhard,P.,Tasson, 2017)

The category of complete cones and absolutely monotone and Scott-continuous functions is a cpo-enriched cartesian closed category.

So:

$[\![\texttt{Real} \Rightarrow \texttt{Real}]\!] = \{f : \mathcal{B}(\mathbf{Meas}(\mathbb{R})) \to \mathbf{Meas}(\mathbb{R}) \ ; f \text{ absolutely monote and Scott-contin.}\}$

Non-decreasingness of all iterated differences i.e. absolute monotonicity

Given a function $f : \mathcal{B}P \to \mathcal{B}Q$, we say:

- *f 0-non-decreasing:* whenever f is non-decreasing,
- *f $(n+1)$-non-decreasing:* whenever f is non-decreasing and $\forall x \in P$, the function

$$\Delta_x f : x' \mapsto f(x + x') - f(x')$$

  is n-non-decreasing (of domain $\{x' \in P \ ; \ x' + x \in \mathcal{B}P\}$).
- *f absolutely monotone:* whenever f n-non-decreasing for every $n \in \mathbb{N}$.

Example (wpor)

$\texttt{wpor} : (x, y) \mapsto x + y - xy$ is not absolutely monotone (in fact not 1-non-decreasing).

Theorem (Ehrhard,P.,Tasson, 2017)

The category of complete cones and absolutely monotone and Scott-continuous functions is a cpo-enriched cartesian closed category.

So:

$[\![\texttt{Real} \Rightarrow \texttt{Real}]\!] = \{f : \mathcal{B}(\textbf{Meas}(\mathbb{R})) \to \textbf{Meas}(\mathbb{R}) \ ; f$ absolutely monote and Scott-contin.$\}$

- We call the absolutely monotone and Scott-continuous functions

  the stable functions

- in fact, this notion “corresponds” to Berry’s stability in this quantitative setting, to convince you:
  - take the usual coherence space model
  - replace $+$ with disjoint union, $-$ with set-theoretical difference
  - the algebraic order is then $\subseteq$
  - the absolutely monotone and Scott-continuous functions are exactly the stable functions between cliques

- then, stability has much to do with analyticity and not only with sequentiality

Conclusion

- We call the absolutely monotone and Scott-continuous functions

  the stable functions

- in fact, this notion "corresponds" to Berry's stability in this quantitative setting, to convince you:
  - take the usual coherence space model
  - replace $+$ with disjoint union, $-$ with set-theoretical difference
  - the algebraic order is then $\subseteq$
  - the absolutely monotone and Scott-continuous functions are exactly the stable functions between cliques

- then, stability has much to do with analyticity and not only with sequentiality

Conclusion

- We call the absolutely monotone and Scott-continuous functions

  the stable functions

- in fact, this notion "corresponds" to Berry's stability in this quantitative setting, to convince you:
  - take the usual coherence space model
  - replace $+$ with disjoint union, $-$ with set-theoretical difference
  - the algebraic order is then $\subseteq$
  - the absolutely monotone and Scott-continuous functions are exactly the stable functions between cliques

- then, stability has much to do with analyticity and not only with sequentiality