

AlgSDP: Algebra of Sequential Decision Problems

formalised in Agda

Robert Krook **Patrik Jansson**

2020-01-06, WG2.1 #79 in Otterlo, NL

Abstract

Sequential decision problems are a well established concept in decision theory, with the Bellman equation as a popular choice for describing them. Botta, Jansson, Ionescu have formalised the notion of such problems in Idris (presented by Jansson at #75 Uruguay, by Botta at #77 Brandenburg). Here we focus on an Algebra of SDPs (in Agda): combinators for building more complex SDPs from simpler ones.

AlgSDP by example in one slide

A 1D-coord. syst. with \mathbb{N} as state and $+1$, 0 , and -1 as actions.

$p : SDProc$

AlgSDP by example in one slide

A 1D-coord. syst. with \mathbb{N} as state and $+1$, 0 , and -1 as actions.

$$p : SDProc$$

We define a product to enable reusing p in a 2D setting:

$$\begin{aligned} _ \times_{SDP} _ & : SDProc \rightarrow SDProc \rightarrow SDProc \\ p^2 & = p \times_{SDP} p \end{aligned}$$

AlgSDP by example in one slide

A 1D-coord. syst. with \mathbb{N} as state and $+1$, 0 , and -1 as actions.

$$p : SDProc$$

We define a product to enable reusing p in a 2D setting:

$$\begin{aligned} _ \times_{SDP} _ & : SDProc \rightarrow SDProc \rightarrow SDProc \\ p^2 & = p \times_{SDP} p \end{aligned}$$

Both p and p^2 use a fixed state space, but we can also handle time dependent processes (for example p' of type $SDProcT$).

$$\begin{aligned} _ \times_{SDP}^T _ & : SDProcT \rightarrow SDProcT \rightarrow SDProcT \\ embed & : SDProc \rightarrow SDProcT \\ p^{2'} & = p' \times_{SDP}^T (embed\ p) \\ p^3 & = p^2 \times_{SDP} p \end{aligned}$$

AlgSDP by example in one slide

A 1D-coord. syst. with \mathbb{N} as state and $+1$, 0 , and -1 as actions.

$$p : SDProc$$

We define a product to enable reusing p in a 2D setting:

$$\begin{aligned} _ \times_{SDP} _ & : SDProc \rightarrow SDProc \rightarrow SDProc \\ p^2 & = p \times_{SDP} p \end{aligned}$$

Both p and p^2 use a fixed state space, but we can also handle time dependent processes (for example p' of type $SDProcT$).

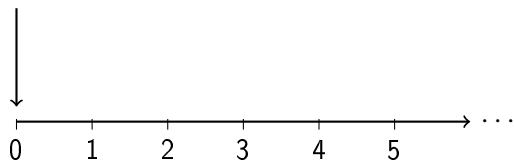
$$\begin{aligned} _ \times_{SDP}^T _ & : SDProcT \rightarrow SDProcT \rightarrow SDProcT \\ embed & : SDProc \rightarrow SDProcT \\ p^{2'} & = p' \times_{SDP}^T (embed\ p) \\ p^3 & = p^2 \times_{SDP} p \end{aligned}$$

Final example: a process that moves either in 3D or in 2D.

$$\begin{aligned} _ \uplus_{SDP}^T _ & : SDProcT \rightarrow SDProcT \rightarrow SDProcT \\ game & = p^{2'} \uplus_{SDP}^T (embed\ p^3) \end{aligned}$$

You could think of this as choosing a map in a game.

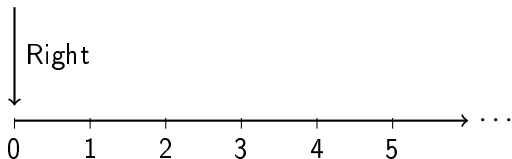
Example: 1-dimensional coordinate system



1d-state : Set

1d-state = \mathbb{N}

Example: 1-dimensional coordinate system



1d-state : Set

1d-state = \mathbb{N}

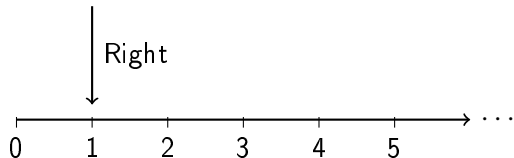
data *1d-control* : *1d-state* \rightarrow Set **where**

Right : $\{n : 1d\text{-state}\} \rightarrow 1d\text{-control } n$

Stay : $\{n : 1d\text{-state}\} \rightarrow 1d\text{-control } n$

Left : $\{n : 1d\text{-state}\} \rightarrow 1d\text{-control } (\text{suc } n)$

Example: 1-dimensional coordinate system



1d-state : Set

1d-state = \mathbb{N}

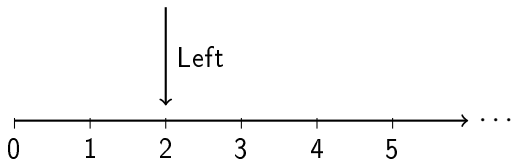
data *1d-control* : *1d-state* \rightarrow Set **where**

Right : { *n* : *1d-state* } \rightarrow *1d-control* *n*

Stay : { *n* : *1d-state* } \rightarrow *1d-control* *n*

Left : { *n* : *1d-state* } \rightarrow *1d-control* (*suc n*)

Example: 1-dimensional coordinate system



1d-state : Set

1d-state = \mathbb{N}

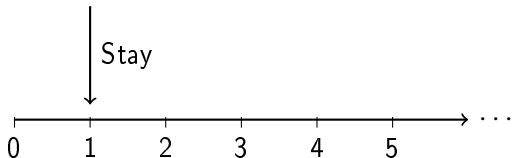
data *1d-control* : *1d-state* \rightarrow Set **where**

Right : $\{n : 1d\text{-state}\} \rightarrow 1d\text{-control } n$

Stay : $\{n : 1d\text{-state}\} \rightarrow 1d\text{-control } n$

Left : $\{n : 1d\text{-state}\} \rightarrow 1d\text{-control } (\text{succ } n)$

Example: 1-dimensional coordinate system



1d-state : Set

1d-state = \mathbb{N}

data *1d-control* : *1d-state* \rightarrow Set **where**

Right : { *n* : *1d-state* } \rightarrow *1d-control* *n*

Stay : { *n* : *1d-state* } \rightarrow *1d-control* *n*

Left : { *n* : *1d-state* } \rightarrow *1d-control* (*suc n*)

Example: 1-dimensional coordinate system

1d-state : Set

1d-state = \mathbb{N}

data *1d-control* : *1d-state* \rightarrow Set **where**

Right : { *n* : *1d-state* } \rightarrow *1d-control* *n*

Stay : { *n* : *1d-state* } \rightarrow *1d-control* *n*

Left : { *n* : *1d-state* } \rightarrow *1d-control* (*suc* *n*)

1d-step : (*x* : *1d-state*) \rightarrow *1d-control* *x* \rightarrow *1d-state*

1d-step *x* *Right* = *suc* *x*

1d-step *x* *Stay* = *x*

1d-step (*suc* *x*) *Left* = *x*

Sequential Decision Process

```
record SDProc : Set1 where  
  constructor SDP  
  field  
    State : Set
```

Sequential Decision Process

```
record SDProc : Set1 where  
  constructor SDP  
field  
  State : Set  
  Control : State → Set
```

Sequential Decision Process

record *SDProc* : *Set1* **where**

constructor *SDP*

field

State : *Set*

Control : *State* \rightarrow *Set*

step : (*x* : *State*) \rightarrow *Control* *x* \rightarrow *State*

Sequential Decision Process

```
record SDProc : Set1 where  
  constructor SDP  
  field  
    State    : Set  
    Control  : State → Set  
    step     : (x : State) → Control x → State
```

Our example:

```
1d-sys : SDProc  
1d-sys = SDP 1d-state 1d-control 1d-step
```

Sequential Decision Problem

In a sequential decision **problem** there is also a fourth field *reward*:

record *SDProb* : *Set1* **where**

constructor *SDP*

field

State : *Set*

Control : *State* \rightarrow *Set*

step : (*x* : *State*) \rightarrow *Control* *x* \rightarrow *State*

reward : (*x* : *State*) \rightarrow *Control* *x* \rightarrow *Val*

(where *Val* is often \mathbb{R}).

- ▶ The Seq. Dec. Problem is: find a sequence of controls that maximises the sum of rewards.
- ▶ Or, in more realistic settings with uncertainty, finding a sequence of *policies* which maximises the *expected* reward.
- ▶ Rewards, and problems, are not the focus of this talk but are mentioned for completeness.

Policy

In general:

$$\text{Policy} : (S : \text{Set}) \rightarrow ((s : S) \rightarrow \text{Set}) \rightarrow \text{Set}$$
$$\text{Policy } S C = (s : S) \rightarrow C s$$

Policy

In general:

$$\text{Policy} : (S : \text{Set}) \rightarrow ((s : S) \rightarrow \text{Set}) \rightarrow \text{Set}$$
$$\text{Policy } S \ C = (s : S) \rightarrow C \ s$$

Specialised:

$$1d\text{-Policy} : \text{Set}$$
$$1d\text{-Policy} = \text{Policy } 1d\text{-state } 1d\text{-control}$$
$$\text{--} \quad = (x : 1d\text{-state}) \rightarrow 1d\text{-control } x$$

Policy

Specialised:

1d-Policy : *Set*

1d-Policy = *Policy 1d-state 1d-control*

-- = $(x : 1d-state) \rightarrow 1d-control\ x$

Example policies:

right stay tryleft : *1d-Policy*

right _ = *Right*

stay _ = *Stay*

tryleft zero = *Stay*

tryleft (suc s) = *Left*

Policy

Example policies:

right stay tryleft : *1d-Policy*

right _ = *Right*

stay _ = *Stay*

tryleft zero = *Stay*

tryleft (suc s) = *Left*

A family of policies (to move *towards* a particular goal coordinate):

towards : $\mathbb{N} \rightarrow$ *1d-Policy*

towards goal n **with** *compare n goal*

... | *less* _ _ = *Right*

... | *equal* _ = *Stay*

... | *greater* _ _ = *Left*

Back to Processes - now with abbreviations

record *SDProc* : *Set1* **where**
 constructor *SDP*
 field *State* : *Set*
 Control : *Con State*
 step : *Step State Control*

Con : *Set* \rightarrow *Set*₁

Con *S* = *S* \rightarrow *Set*

Step : (*S* : *Set*) \rightarrow *Con* *S* \rightarrow *Set*

Step *S* *C* = (*s* : *S*) \rightarrow *C* *s* \rightarrow *S*

Policy : (*S* : *Set*) \rightarrow ((*s* : *S*) \rightarrow *Set*) \rightarrow *Set*

Policy *S* *C* = (*s* : *S*) \rightarrow *C* *s*

Trajectory

Here $\#_{st}$, $\#_c$, $\#_{sf}$ extract the different components of an SDP.

$$\begin{aligned} \text{trajectory} &: (p : SDProc) \rightarrow \{n : \mathbb{N}\} \rightarrow \\ &\rightarrow \text{Vec} (\text{Policy} (\#_{st} p) (\#_c p)) n \\ &\rightarrow \#_{st} p \rightarrow \text{Vec} (\#_{st} p) n \\ \text{trajectory sys } [] &\quad x_0 = [] \\ \text{trajectory sys } (p :: ps) & \quad x_0 = x_1 :: \text{trajectory sys } ps \ x_1 \\ \text{where } x_1 &: \#_{st} \text{ sys} \\ x_1 &= (\#_{sf} \text{ sys}) x_0 (p \ x_0) \end{aligned}$$

Example:

$$\begin{aligned} \text{pseq} &= \text{tryleft} :: \text{tryleft} :: \text{right} :: \text{stay} :: \text{right} :: [] \\ \text{test1} &= \text{trajectory } 1d\text{-sys } \text{pseq } 0 \\ -- &= 0 :: 0 :: 1 :: 1 :: 2 :: [] \end{aligned}$$

In an applied setting many trajectories would be computed to explore the system behaviour.

The Product of SDPs

$$\begin{aligned} _ \times_{SDP} _ &: SDProc \rightarrow SDProc \rightarrow SDProc \\ (SDP \ S_1 \ C_1 \ sf_1) \times_{SDP} (SDP \ S_2 \ C_2 \ sf_2) \\ &= SDP (S_1 \times S_2) (C_1 \times_C C_2) (sf_1 \times_{sf} sf_2) \end{aligned}$$

The Product of SDPs

$$\begin{aligned} _ \times_{SDP} _ &: SDProc \rightarrow SDProc \rightarrow SDProc \\ (SDP \ S_1 \ C_1 \ sf_1) \times_{SDP} (SDP \ S_2 \ C_2 \ sf_2) \\ &= SDP (S_1 \times S_2) (C_1 \times_C C_2) (sf_1 \times_{sf} sf_2) \end{aligned}$$

$$\begin{aligned} Con &: Set \rightarrow Set_1 \\ Con \ S &= S \rightarrow Set \end{aligned}$$

$$\begin{aligned} _ \times_C _ &: \{S_1 \ S_2 : Set\} \rightarrow \\ &Con \ S_1 \rightarrow Con \ S_2 \rightarrow Con (S_1 \times S_2) \\ (C_1 \times_C C_2) (s_1, s_2) &= C_1 \ s_1 \times C_2 \ s_2 \end{aligned}$$

The Product of SDPs

$$\begin{aligned} _ \times_{SDP} _ &: SDProc \rightarrow SDProc \rightarrow SDProc \\ (SDP \ S_1 \ C_1 \ sf_1) \times_{SDP} (SDP \ S_2 \ C_2 \ sf_2) \\ &= SDP (S_1 \times S_2) (C_1 \times_C C_2) (sf_1 \times_{sf} sf_2) \end{aligned}$$

$$Con : Set \rightarrow Set_1$$

$$Con \ S = S \rightarrow Set$$

$$\begin{aligned} _ \times_C _ &: \{S_1 \ S_2 : Set\} \rightarrow \\ &Con \ S_1 \rightarrow Con \ S_2 \rightarrow Con (S_1 \times S_2) \\ (C_1 \times_C C_2) (s_1, s_2) &= C_1 \ s_1 \times C_2 \ s_2 \end{aligned}$$

$$Step : (S : Set) \rightarrow Con \ S \rightarrow Set$$

$$Step \ S \ C = (s : S) \rightarrow C \ s \rightarrow S$$

$$\begin{aligned} _ \times_{sf} _ &: \{S_1 \ S_2 : Set\} \{C_1 : Con \ S_1\} \{C_2 : Con \ S_2\} \\ &\rightarrow Step \ S_1 \ C_1 \rightarrow Step \ S_2 \ C_2 \\ &\rightarrow Step (S_1 \times S_2) (C_1 \times_C C_2) \\ (sf_1 \times_{sf} sf_2) (s_1, s_2) (c_1, c_2) &= (sf_1 \ s_1 \ c_1, sf_2 \ s_2 \ c_2) \end{aligned}$$

The Product of SDPs

$_ \times_{SDP} _ : SDProc \rightarrow SDProc \rightarrow SDProc$

$(SDP S_1 C_1 sf_1) \times_{SDP} (SDP S_2 C_2 sf_2)$

$= SDP (S_1 \times S_2) (C_1 \times_C C_2) (sf_1 \times_{sf} sf_2)$

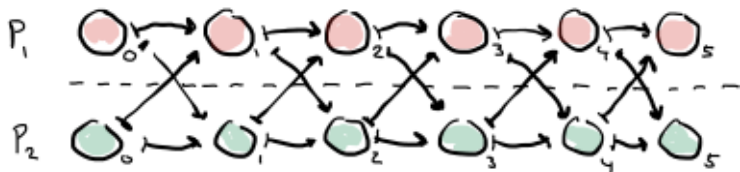
$_ \times_{sf} _ : \{S_1 S_2 : Set\} \{C_1 : Con S_1\} \{C_2 : Con S_2\}$

$\rightarrow Step S_1 C_1 \rightarrow Step S_2 C_2$

$\rightarrow Step (S_1 \times S_2) (C_1 \times_C C_2)$

$(sf_1 \times_{sf} sf_2) (s_1, s_2) (c_1, c_2) = (sf_1 s_1 c_1, sf_2 s_2 c_2)$

Example: $P_1 \times_{SDP} P_2$



Product example

Example:

$$2d\text{-system} = 1d\text{-sys} \times_{SDP} 1d\text{-sys}$$

Now *2d-system* is a process of two dimensions rather than one:

$$\begin{aligned} pseq &= \text{tryleft} :: \text{tryleft} :: \text{right} :: \text{stay} :: \text{right} :: [] \\ 2d\text{-pseq} &= \text{zipWith } _ \times_P _ \text{ pseq pseq} \\ \text{test2} &= \text{trajectory } 2d\text{-system } 2d\text{-pseq } (0, 5) \\ &\text{--} = (0, 4) :: (0, 3) :: (1, 4) :: (1, 4) :: (2, 5) :: [] \end{aligned}$$

where $_ \times_P _$ is a combinator for policies.

Zero and One

zero : *SDProc*

zero = **record** {

State = \perp ;

Control = $\lambda state \rightarrow \perp$;

step = $\lambda state \rightarrow \lambda control \rightarrow state$ }

unit : *SDProc*

unit = **record** {

State = \top ;

Control = $\lambda state \rightarrow \top$;

step = $\lambda state \rightarrow \lambda control \rightarrow tt$ }

Coproduct combinator

$$\begin{aligned} _ \uplus_{SDP} _ & : SDProc \rightarrow SDProc \rightarrow SDProc \\ SDP \ S_1 \ C_1 \ sf_1 \ \uplus_{SDP} \ SDP \ S_2 \ C_2 \ sf_2 \\ & = SDP \ (S_1 \uplus S_2) \ (C_1 \uplus_C C_2) \ (sf_1 \uplus_{sf} sf_2) \end{aligned}$$

$$\begin{aligned} _ \uplus_C _ & : \{S_1 \ S_2 : Set\} \\ & \rightarrow Con \ S_1 \rightarrow Con \ S_2 \rightarrow Con \ (S_1 \uplus S_2) \\ (C_1 \uplus_C C_2) \ (inj_1 \ s_1) & = C_1 \ s_1 \\ (C_1 \uplus_C C_2) \ (inj_2 \ s_2) & = C_2 \ s_2 \end{aligned}$$

$$\begin{aligned} _ \uplus_{sf} _ & : \{S_1 \ S_2 : Set\} \\ & \rightarrow \{C_1 : Con \ S_1\} \rightarrow \{C_2 : Con \ S_2\} \\ & \rightarrow Step \ S_1 \ C_1 \rightarrow Step \ S_2 \ C_2 \\ & \rightarrow Step \ (S_1 \uplus S_2) \ (C_1 \uplus_C C_2) \\ (sf_1 \uplus_{sf} sf_2) \ (inj_1 \ s_1) \ c_1 & = inj_1 \ (sf_1 \ s_1 \ c_1) \\ (sf_1 \uplus_{sf} sf_2) \ (inj_2 \ s_2) \ c_2 & = inj_2 \ (sf_2 \ s_2 \ c_2) \end{aligned}$$

Coproduct combinator example

Left injection:



P_2

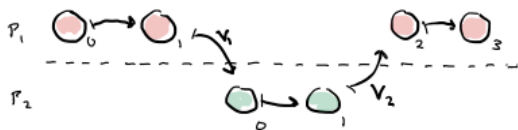
Right injection:

P_1



Yielding coproduct example

Illustration: It is capable of switching between the two processes, as illustrated by the calls to $v1$ and $v2$.



With a combinator such as this one could you model e.g a two player game.

The processes would be the players and the combined process allows each to take turns making their next move.

Yielding coproduct code

$$\begin{aligned} _ \uplus_C^m _ &: \{ S_1 S_2 : \text{Set} \} \\ &\rightarrow \text{Con } S_1 \rightarrow \text{Con } S_2 \rightarrow \text{Con } (S_1 \uplus S_2) \end{aligned}$$

$$(C_1 \uplus_C^m C_2) (\text{inj}_1 s_1) = \text{Maybe } (C_1 s_1)$$

$$(C_1 \uplus_C^m C_2) (\text{inj}_2 s_2) = \text{Maybe } (C_2 s_2)$$

$$_ \rightleftarrows _ : (S_1 S_2 : \text{Set}) \rightarrow \text{Set}$$

$$s_1 \rightleftarrows s_2 = (s_1 \rightarrow s_2) \times (s_2 \rightarrow s_1)$$

$$\begin{aligned} \uplus_{sf}^m &: \{ S_1 S_2 : \text{Set} \} \{ C_1 : \text{Con } S_1 \} \{ C_2 : \text{Con } S_2 \} \\ &\rightarrow (S_1 \rightleftarrows S_2) \end{aligned}$$

$$\rightarrow \text{Step } S_1 C_1 \rightarrow \text{Step } S_2 C_2$$

$$\rightarrow \text{Step } (S_1 \uplus S_2) (C_1 \uplus_C^m C_2)$$

$$\uplus_{sf}^m _ \quad sf_1 \ sf_2 \ (\text{inj}_1 \ s_1) \ (\text{just } c) = \text{inj}_1 \ (sf_1 \ s_1 \ c)$$

$$\uplus_{sf}^m _ \quad sf_1 \ sf_2 \ (\text{inj}_2 \ s_2) \ (\text{just } c) = \text{inj}_2 \ (sf_2 \ s_2 \ c)$$

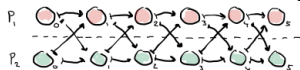
$$\uplus_{sf}^m \ (v_1, _) \ sf_1 \ sf_2 \ (\text{inj}_1 \ s_1) \ \text{nothing} = \text{inj}_2 \ (v_1 \ s_1)$$

$$\uplus_{sf}^m \ (_, v_2) \ sf_1 \ sf_2 \ (\text{inj}_2 \ s_2) \ \text{nothing} = \text{inj}_1 \ (v_2 \ s_2)$$

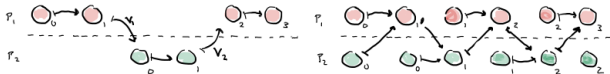
$$\text{syntax } \uplus_{sf}^m \ r \ sf_1 \ sf_2 = sf_1 \langle r \rangle \ sf_2$$

Summary

- ▶ It is possible to implement an algebra of SDPs
- ▶ Products are immediately useful



- ▶ Plain coproducts — not so much
- ▶ Many variants possible: yielding coproducts, interleaving product, etc.



Time-dependent, monadic cases left as exercises for the audience;-)