An Algebra of Sequential Decision Problems

Technical Report

Robert Krook
Computer Science and Engineering
University of Gothenburg
Sweden
guskrooro@student.gu.se

Patrik Jansson
Computer Science and Engineering
Chalmers University of Technology
Sweden
patrik.jansson@chalmers.se

ACM Reference Format:

1 Introduction

Sequential decision processes and problems are a well established concept in decision theory, with the Bellman equation [1] as a popular choice for describing them. Botta et al [4] have formalised the notion of such problems in Idris. Using dependent types to bridge the gap between description and implementation of complex systems, for purposes of simulation, has been shown to be a good choice [5]. They have illustrated how to use their formulation to model e.g. climate impact research [3], a very relevant problem today.

Evidence based policy making (when dealing with climate change or other global systems challenges), requires computing policies which are verified to be correct. There are several possible notions of “correctness” for a policy: computing feasible system trajectories through a state space, avoiding “bad” states, or even computing optimal policies. The concepts of feasibility and avoidability have been formalised and presented in Botta et al. [2].

Although motivated by the complexity of modelling in climate impact research, we focus on simpler examples of sequential decision processes and how to combine them.

Examples: Assume that we have a process \( p : SDP \) that models something moving through a 1-D coordinate system with a natural number as the state and \(+1\), \(0\), and \(-1\) as actions. If the circumstances change and we need to model how something moves in a 2-D coordinate system, it would be convenient if we could reuse the one dimensional system and get the desired system for free. We seek a combinator \( SDP \times SDP \) such that

\[
p^2 = p \times SDP \ p
\]

Both \( p \) and \( p^2 \) use a fixed state space, but we can also handle time dependent processes. Assume \( p' : SDProcT \) is similar to \( p \) but time dependent: not all states are available at all times, meaning \( p' \) is more restricted in the moves it can make. If we want to turn this into a process that can also move around in a second dimension, we want to be able to reuse both \( p' \) and \( p \). We can use a combinator \( SDProcT \rightarrow SDProcT \) such that

\[
p^{2'} = p' \times SDP \ (\text{embed} \ p)
\]

As a last example consider the case where we want a process that moves either in a 3-D coordinate system \( p^3 = p^2 \times SDP \ p \) or in \( p'' \). You could think of this as choosing a map in a game. Then we would want a combinator \( SDProcT \rightarrow SDProcT \rightarrow SDProcT \) such that

\[
\text{game} = p^{2'} \cup SDP \ (\text{embed} \ p^3)
\]

These combinators, and more, make up an Algebra of SDPs.

2 Sequential Decision Problems

First, we formalise the notion of a Sequential Decision Process in Agda. A process always has a state, and depending on what that state is there are different controls that describe what actions are possible in that state. The last component of a sequential decision process is a function \( \text{step} \) that when applied to a state and a control for that state returns the next state. To better see the type structure we introduce a type synonym for the family of controls depending on a state:

\[
\begin{align*}
\text{Con} : \ & \text{Set} \rightarrow \text{Set}, \\
\text{Con} \ S & \ = \ S \rightarrow \text{Set}
\end{align*}
\]

and for the the type of step functions defined in terms of a state and a family of controls on that state:

\[
\begin{align*}
\text{Step} : \ & (S : \text{Set}) \rightarrow \text{Con} \ S \rightarrow \text{Set} \\
\text{Step} \ S \ C & \ = \ (s : S) \rightarrow C \ s \rightarrow S
\end{align*}
\]

With these in place we define a record type for SDPs:

\[
\text{record} \ SDProc : \ Set1\ where
\begin{align*}
\text{constructor} \ SDP \\
\text{field} \ State & \ : \ \text{Set} \\
Control & \ : \ \text{Con} \ State \\
\text{step} & \ : \ \text{Step} \ State \ Control
\end{align*}
\]

We can extend this idea of a sequential decision process to that of a problem by adding an additional field \( \text{reward} \).
where Val is often \( \mathbb{R} \). From the type we conclude that the reward puts a value on the steps taken by the step function, based on the state transition and the control used. The problem becomes that of finding the sequence of controls that produces the highest sum of rewards. Or, in more realistic settings with uncertainty (which can be modelled by a monadic step function), finding a sequence of policies which maximises the expected reward. The system presented here aims at describing finite horizon problems, meaning that the sum of rewards is over a finite list. Furthermore, rewards are usually discounted as the time passes. One action now is worth more than the same action a few steps later. Rewards, and problems, are not the focus of this abstract but are mentioned for completeness.

A policy is a function from states to controls:

\[
\text{Policy} : (S : \text{Set}) \rightarrow \text{Con} S \rightarrow \text{Set}
\]

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

Given a list of policies to apply, one for each time step, we can compute the trajectory of a process from a starting state. Here the \( \#_{st} \) and \( \#_{df} \) functions extract the state and step component from the SDProc respectively.

\[
\text{trajectory} : \{n : \mathbb{N}\} \rightarrow (p : \text{SDProc}) \rightarrow \text{Vec} (\text{Policy} (\#_{st} p) (\#_{c} p)) n \rightarrow \#_{st} p \rightarrow \text{Vec} (\#_{st} p) n
\]

\[
\text{trajectory} \ (p :: ps) \ x_0 = x_1 :: \text{trajectory} \ ps \ x_1
\]

where \( x_1 :: \#_{st} \) sys

\[
\text{sys} x_0 = (\#_{df} \) sys \ x_0 \ (p \ x_0)
\]

As an example of a trajectory computation we return to the one dimensional process \( \#l\)-sys (called just \( p \) in the intro) and an example policy sequence \( \text{pseq} \). Ideally \( \text{pseq} \) is the result of an optimization computed using Bellmans backwards induction. Here we just illustrate one trajectory:

\[
\text{pseq} = \text{tryleft} :: \text{tryleft} :: \text{right} :: \text{stay} :: \text{right} :: []
\]

\[
\text{test1} : \text{trajectory} \ \#l\)-sys \ pseq \ 0 \equiv 0 :: 0 :: 1 :: 0 :: 0 :: 0 :: 0 :: 1 :: 2 :: []
\]

\[
\text{test1} = \text{refl}
\]

In an applied setting many trajectories would be computed to explore the system behaviour. This brief example is fully presented in an accompanying technical report [6].

In this abstract we focus on non-monadic, time-independent, sequential decision processes, but the algebra extends nicely to the more general case.

### 3 The Product Combinator

To compute \( p^2 \) we need to define a product combinator for SDPs. We illustrate what this combinator does in Figure 1. The state of the product of two processes is the cartesian product of the two separate states.

**Figure 1.** The product process holds components of both states and applies the step function to both components simultaneously. Each component of the next state has two incoming arrows as the policy that computes the control that is used has access to both components of the previous state.

Given two control families, we can compute the control family for pairs of states. The inhabitants (the controls) of each family member are pairs of controls for the two state components.

\[
\begin{align*}
\text{SDP } S_1 C_1 &\rightarrow \text{SDP } S_2 C_2 \\
\text{SDP } S_1 C_1 \times \text{SDP } S_2 C_2 &\rightarrow \text{SDP } (S_1 \times S_2) (C_1 \times C_2)
\end{align*}
\]

Given two step functions we can define a new step function for the product process by returning the pair computed by applying the individual step functions to the corresponding components of the input.

\[
\begin{align*}
\text{SDP } S_1 C_1 &\rightarrow \text{SDP } S_2 C_2 \\
\text{SDP } (S_1 \times S_2) (C_1 \times C_2) &\rightarrow \text{SDP } (S_1 \times S_2) (C_1 \times C_2)
\end{align*}
\]

Finally, we can compute the product of two sequential decision processes by applying the combinators componentwise.

\[
\begin{align*}
\text{SDP } S_1 C_1 &\rightarrow \text{SDP } S_2 C_2 \\
\text{SDP } (S_1 \times S_2) (C_1 \times C_2) &\rightarrow \text{SDP } (S_1 \times S_2) (C_1 \times C_2)
\end{align*}
\]

To illustrate how the combinator works we apply it to the system (\( \#l\)-sys) mentioned previously.

\[
\begin{align*}
\text{2d-system} &= \text{2d-system} \times \text{SDP} \ \#l\)-sys
\end{align*}
\]

Now 2d-system is a process of two dimensions rather than one, as illustrated by the type of test2.

\[
\begin{align*}
\text{2d-pseq} &= \text{zipWith} \ \text{pseq} \ \text{pseq} \ \text{pseq} \\
\text{test2} &= \text{trajectory} \ 2d\text{-system} \ 2d\text{-pseq} \ (0, 5) \\
&\equiv (0, 4 :: (0, 3) :: (1, 4) :: (1, 4) :: (2, 5) :: []
\end{align*}
\]

### 4 Wrapping up

In the technical report [6] we present more combinators for time dependent and time independent processes and policies. We implement the example of a coordinate system described above, and make it even more precise as a time dependent process. Future work includes generalising to monadic SDPs and applying our combinators to the green house gas emission problem [3].
We thank the anonymous reviewers for their helpful comments and the Agda developers for a great tool!

References


5 Technical Report

The rest of this report makes up the technical report. The technical report describes further combinators, discusses time dependent processes and provide more thorough examples.

The work described in this text is the result of a project carried out at Gothenburg University by Robert Krook, under the supervision of Patrik Jansson. Robert has worked independently and met once a week with Patrik over the course of 8 weeks to discuss progress, potential bottlenecks and what to do next. Patrik has been an invaluable source of information, both regarding sequential decision problems and how to write a scientific text.

6 Example

Let us consider a sequential decision process where the state space is a one dimensional coordinate system represented by natural numbers.

\[ \text{kl-state} : \text{Set} \]
\[ \text{kl-state} = \mathbb{N} \]

Seeing how the state space are the natural numbers we emphasize that the coordinate system has only positive coordinates. In any given state, generally, we can choose to either increment, decrement or do nothing to the state. In the edge case where the state is 0 we can not decrement the state. We can encode this control as a type family in Agda.

```agda
data 1d-control : kl-state → Set where  
  Right : {n : kl-state} → 1d-control n  
  Stay : {n : kl-state} → 1d-control n  
  Left : {n : kl-state} → 1d-control (suc n)
```

We implement the step function by pattern matching on the control. In the case of the Left control Agda recognises that the state must be a successor. We increment or decrement the state accordingly and leave it unchanged for the Stay control.

```agda
1d-step : (x : kl-state) → 1d-control x → kl-state
1d-step x Right = suc x
1d-step x Stay = x
1d-step (suc x) Left = x
```

We define a policy to be a function that given a state select a control. The policies right, stay and tryleft are all policies of this kind. tryleft is special in the sense that if the state is zero it will do nothing, as it can not go left.

```agda
kl-Policy = (x : kl-state) → 1d-control x
right stay tryleft : kl-Policy
right _ = Right
stay _ = Stay
tryleft zero = Stay
tryleft (suc s) = Left
```

We can parameterise a policy over a coordinate and define a policy that will select controls that moves the system towards this coordinate.

```agda
towards : \mathbb{N} → 1d-Policy
towards goal n with compare n goal
  ... | less _ _ = Right
  ... | equal _ _ = Stay
  ... | greater _ _ = Left
```

With the three components state, control and step, we can instantiate a sequential decision process.

```agda
1d-sys = SDP kl-state 1d-control 1d-step
```

A policy sequence is now just a vector of policies.

```agda
pseq = tryleft :: tryleft :: right :: stay :: right :: []
```

We can evaluate the system using this sequence, starting from different points. We can use \(\equiv\) and \(\text{refl}\) to assert that the system behaves as intended.

```agda
test1 : trajectory 1d-sys pseq 0 ≡ 0 :: 0 :: 1 :: 1 :: 2 :: []
test2 = refl
test3 = refl
test3 = refl
```

We can use the clever policy to steer the process towards a goal.

```agda
test3 : trajectory kl-sys (replicate (towards 5)) 2 ≡
  3 :: 4 :: 5 :: 5 :: []
test3 = refl
```

To turn a process into a problem we need to introduce a notion of a goal, described by a \text{reward} function. For our example we define the reward function to be parameterised over a target coordinate. The reward function could then reward a proposed step based on how close to the target it lands.

```agda
1d-reward : kl-state → (x : kl-state) → 1d-control x → kl-state → \mathbb{N}
1d-reward target x₀ y x₁ = if distance target x₁ < distance target x₀ then 2
                           else (if distance target x₀ < distance target x₁ then 0
                                  else 1)
```

We can redefine the sequential decision process above to be a sequential decision problem simply by instantiating the SDProb record.
problem : 1d-state → SDProblem
problem target = SDProb 1d-state 1d-control
1d-step (1d-reward target)

7 Combinators for sequential decision processes

Now that we’ve seen an example of a sequential decision process and are getting comfortable with the concept, it is suitable to move forward and see what we can do with processes. This section will explore different ways sequential decision processes can be combined in order to produce more sophisticated processes.

We already defined the product combinator, and before we move on to additional combinators we’d like to make a few notes on the product combinator. An observation to be made is that in order for the new system to exist in any state, it has to hold components of both prior states. This has the consequence that if the state space of one of the prior processes is empty, the new problems state space is also empty. Similarly, if one of the components reaches a point where there are no available controls, and thus can not progress, the other component will not be able to progress either.

Looking back at the example of the one dimensional coordinate system we find ourselves wondering if we would now indeed get a process of a two dimensional coordinate system seemingly for free. The answer, unsurprisingly, is yes.

2d-system = 1d-sys ×_sys 1d-sys

In section 11 we introduce combinators for policy sequences. Here we use the product combinator to produce a policy sequence that is compatible with the new process.

2d-pseq : PolicySeq (#_u 2d-system) (#_c 2d-system) 5
2d-pseq = zipWith _xP_ pseq pseq

And now we can evaluate this new process like we did with the one dimensional system.

run2d = trajectory 2d-system 2d-pseq

2d-test1 : run2d (0, 5)
≡
(0, 4) :: (0, 3) :: (1, 4) ::
(1, 4) :: (2, 5) :: (2, 5) :: []
2d-test1 = refl

2d-test2 : run2d (5, 5)
≡
(4, 4) :: (3, 3) :: (4, 4) ::
(4, 4) :: (5, 5) :: (5, 5) :: []
2d-test2 = refl

Functional programmers will often find they are in need of a unit, e.g when using reduce or other frequently appearing constructs from the functional paradigm. Before we begin implementing a unit for the product case we want to clarify what we mean by a unit. A unit to a process is one that when combined with another process, produces a process where the change at each step is exactly that of the other process.

What we are after is a process that will not carry any extra information, or rather one that can not alter the information it carries. Recall that in order for the state space of the product process to not be empty, both state spaces of the separate processes has to be non-empty. In order to call the step function the control space also has to be inhabited. In an effort to minimise the information the unit carries we declare its state space and control space to be singletons. The step function becomes a constant function that given the only state and the only control, will return the only state.

singleton : SDProc
singleton = record {
State = T;
Control = λ state → T;
step = λ state → λ control → T
}

An example of evaluating the singleton process is illustrated in Figure 2

Taking the product of any process and the singleton process would produce a process where the only change of information during each step is that of the process which is not the singleton. Of course, the other process could itself be the singleton process also, in which case the only change in each step is exactly that of the singleton process, which is no change at all.

7.1 Coproduct

Seeing how we defined a product combinator of two processes, we are interested in also defining a sum combinator for processes. The approach is similar to that of the product case.

The inhabitants of the sum control is the sum of the inhabitants of the prior controls.

₀ ⊎₀ C₀ : {S₁, S₂ : Set}
→ Con S₁ → Con S₂ → Con (S₁ ⊔ S₂)
(C₁ ⊔₀ C₂) (inj₁ s₁) = C₁ s₁
(C₁ ⊔₀ C₂) (inj₂ s₂) = C₂ s₂

Calculating a new step function from two prior step functions is relatively straightforward. The first input is the sum of the two states. Depending on which state the first argument
The process has no states, no controls and the step function keeps the two problems truly independent. In fact, the coproduct combinator will produce a process that acts exactly as that of the given process. There is no way to begin advancing the empty process, and so the only available option is to select an initial state from the other process and start progressing that.

7.2 Yielding Coproduct

Computing the coproduct of two processes and getting a process that behaves like either of the two, without actually considering the other process, leaves us wondering what this is useful for. It would be more useful if we could jump between the two processes. To do this, we first need to define a relation between states. We define a relation on two state and define it to be a mapping from an inhabitant of one state to an inhabitant of the other.

\[ \_ \rightsquigarrow \_ : (S_1, S_2 : \text{Set}) \rightarrow \text{Set} \]

\[ s_1 \rightsquigarrow s_2 = (s_1 \rightarrow s_2) \times (s_2 \rightarrow s_1) \]

Combining the two controls on the states is similar to that of the coproduct case, when looking at the type. However, instead of the new control being defined as either of the two prior ones, it is now \(\text{Maybe} \) either of the two previous ones. The idea is that we extend the control space to have one more inhabitant, the value \(\text{nothing}\). If we select this control the process should yield in favour of the other process.

\[ \_ \rightsquigarrow \_ : (S_1, S_2 : \text{Set}) \rightarrow (S_1 \times S_2) \rightarrow \text{Maybe} \]

\[ (C_1 \rightsquigarrow \_ ) \times (C_2 \rightsquigarrow \_ ) = \text{Maybe} \]

The new step function needs to accommodate for this scenario where the process should yield in favour of the other. To implement this the new step function needs to know how to yield. We describe how to yield by supplying an element of type \(S_1 \rightleftarrows S_2\). If the selected control is \(\text{nothing}\) the step function will apply the appropriate component of this element to the current state.

\[ \_ \rightsquigarrow \_ : (S_1, S_2 : \text{Set}) \times (C_1 : \text{Set}) \times (C_2 : \text{State}) \rightarrow (S_1 \rightleftarrows S_2) \times (C_1 \rightleftarrows C_2) \]

\[ \rightsquigarrow (\_ , \_ ) = \begin{cases} \text{nothing} & \text{if } \rightleftarrows (\_ , \_ ) = \text{nothing} \\ \text{inj}_1 (v_1, s_1) \text{sf}_1 (\_ ) \text{sf}_2 (\_ ) & \text{if } \rightleftarrows (\_ , \_ ) = \text{inj}_1 (v_1, s_1) \text{sf}_1 (\_ ) \text{sf}_2 (\_ ) \\ \text{inj}_2 (v_2, s_2) \text{sf}_1 (\_ ) \text{sf}_2 (\_ ) & \text{if } \rightleftarrows (\_ , \_ ) = \text{inj}_2 (v_2, s_2) \text{sf}_1 (\_ ) \text{sf}_2 (\_ ) \end{cases} \]

Since the other operators were infix, we give a syntax declaration that mimics the same style.

\[ \text{syntax } \rightsquigarrow (r) \text{sf}_1 \text{sf}_2 = \text{sf}_1 (r) \text{sf}_2 \]

Now we can compute the yielding coproduct of two processes by applying the new operations componentwise.
7.3 Interleaving processes

The next combinator we introduce is one that interleaves processes. The state of such a process holds components of both prior states, but takes turns applying the step function to each of them. This behaviour could be similar to that of a game, where two players take turns making their next move. However, the users do not know what moves the other player has made, and can therefore not make particularly smart moves. In section 11 it is reasoned that writing new policies for a process like this will be a policy that does know what move the other ‘player’ has made.

Similar to the product combinator the new state needs to hold components of both prior states. It should apply the step function to them one at a time, alternating between the two. In order to know which components turn it is to advance we extend the product to also hold an index.

\[ \neg \neg S - : \text{Set} \rightarrow \text{Set} \rightarrow \text{Set} \]
\[ S_1 \leftarrow S \rightarrow S_2 = \text{Fin} \times S_1 \times S_2 \]

The control space for the interleaved process is the sum of the two prior control spaces. If the value of the first component is zero, we select the first control. If the value is one, we select the second control.

one : Fin 2

one = suc zero

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. In section 11 we discuss how a policy for such a process is something of a game leader.

A unit to the yielding coproduct combinator is the same one as that for the regular coproduct combinator. If the state space is not inhabited, the process could never progress as we will not be able to call the step function. Further more, we would not be able to give a definition for a function \( S_1 \rightarrow S_2 \).

Defining a unit for the interleaved process is not possible. With a combinator such as this one could you model e.g a game leader.

An Algebra of Sequential Decision Problems
TyDe’19, August 18, 2019, Berlin, Germany

Figure 4. Illustration of the yielding coproduct process. It is capable of switching between the two processes, as illustrated by the calls to v1 and v2.

\[
\neg \neg \neg \wedge_{\text{SDP}} : (p_1 : \text{SDProc}) \rightarrow (p_2 : \text{SDProc}) \\
\rightarrow (\#_a p_1 \rightarrow \#_a p_2) \\
\rightarrow \text{SDProc} \\
((\text{SDP } S_1 \text{ C}_1 \text{ sf}_1) \wedge_{\text{SDP}} (\text{SDP } S_2 \text{ C}_2 \text{ sf}_2)) \text{ rel} \\
= \text{SDP } (S_1 \odot S_2) (C_1 \wedge_{\text{C}} C_2) (\text{sf}_1 \text{ rel } \text{sf}_2)
\]

Defining a step function in terms of the two previous ones is done by pattern matching on the state. Specifically we are interested in the first component, the index. If the index is zero we apply the first step function to the second component of the state, leave the last component unchanged and increment the index by one. Similarly if the index is zero we apply the second step function to the last component, leave the second one unchanged and decrement the index by one.

\[
\neg \neg \neg \text{C} - : \{ S_1 \text{ S}_2 := \text{Set} \} \\
\rightarrow \{ C_1 : \text{Con } S_1 \} \rightarrow \{ C_2 : \text{Con } S_2 \} \\
\rightarrow \text{Step } S_1 \text{ C}_1 \rightarrow \text{Step } S_2 \text{ C}_2 \\
\rightarrow \text{Step } (S_1 \leftarrow_S S_2) (C_1 \leftarrow_C C_2) \\
(sf_1 \leftarrow_{\text{sf}} sf_2)(\text{zero } , s_1 , s_2) c = (\text{one } , sf_1 s_1 c , s_2) \\
(sf_1 \leftarrow_{\text{sf}} sf_2)(\text{suc zero } , s_1 , s_2) c = (\text{zero } , s_1 , sf_2 s_2 c) \\
(sf_1 \leftarrow_{\text{sf}} sf_2)(\text{suc } (), \ldots , \ldots)
\]

Combining two processes to capture this interleaved behaviour is once again simply done by combining the components componentwise.

\[
\neg \neg \neg \text{SDP} - : \text{SDProc} \rightarrow \text{SDProc} \rightarrow \text{SDProc} \\
\text{SDP } S_1 \text{ C}_1 \text{ sf}_1 \leftarrow_{\text{SDP}} \text{SDP } S_2 \text{ C}_2 \text{ sf}_2 \\
= \text{SDP } (S_1 \leftarrow_S S_2) (C_1 \leftarrow_C C_2) (sf_1 \leftarrow_{\text{sf}} sf_2)
\]

The final process behaves as illustrated in figure 5.

Defining a unit for the interleaved process is not possible. Where the initial process would advance e.g five steps, the interleaved process would need ten steps to take that component to the same state. We can not give a generic process that when interleaved with another process acts as a unit.

The way we define the interleaved combinator might not be optimal. Combining more than two processes will produce potentially unexpected behaviour. If we combine three processes using this combinator the resulting system would be one where one of the processes advance half the time, and the other two only a quarter of the time each.
If we interleave two processes and then interleave the resulting process with a third we get a situation like this. They are not properly interleaved.

This is the interleaved behaviour we might expect for three processes. A round robin behaviour that gives the processes equally many turns.

Figure 6. Illustrations of why the interleaved combinator might not behave as one would expect. Again the two incoming arrows illustrate that the policy that selects the control has access to all components and can base the choice of control on them.

This does not necessarily mean that the combinator described in Figure 6 is wrong, but rather that there is another combinator we could implement that would have this other behaviour.

8 Time dependent processes
Imagine a process where the state space can vary over time. If we consider the example with the one dimensional coordinate system, if the process is time dependent it could disallow some states at certain points in time. Below we illustrate how this is defined in Agda.

ConT : (N → Set) → Set
ConT S = (t : N) → S t → Set
StepT : (S : N → Set) → ConT S → Set
StepT S C = (t : N) → (s : S t) → C t s → S (suc t)
record SDProcT : Set₁ where
  constructor SDPT
  field
    State : N → Set
    Control : ConT State
    step : StepT State Control

The state is now dependent on a parameter t : N, which allows the state to take on alternate forms. In section 9.1 we illustrate what this means.

A time independent process can be embedded as a time dependent process. The embedding is a process that disregards the time parameter.

\[
\text{embed} : \text{SDProc} → \text{SDProcT}
\]
\[
\text{embed} (\text{SDP} S C sf) = \text{SDPT} (\lambda _ → S) (\lambda _ → C) (\lambda _ → sf)
\]

9 A discussion on the \textit{Fin} type
Before we move on to an example of a time dependent process, we need to briefly present the \textit{Fin} type and its properties. The \textit{Fin} \textit{n} type (for any natural number \textit{n}) has exactly \textit{n} elements.

\[
\text{data Fin} : \mathbb{N} → \text{Set where}
\]
\[
\begin{align*}
\text{zero} & : \{ n : \mathbb{N} \} → \text{Fin} (\text{suc} n) \\
\text{suc} & : \{ n : \mathbb{N} \} (i : \text{Fin} n) → \text{Fin} (\text{suc} n)
\end{align*}
\]

From this definition we see that \text{zero} is an element of \textit{Fin} \textit{n} for any \textit{n} > \text{zero}. The constructor \text{suc} takes an element of type \textit{Fin} \textit{n} and returns an element of type \textit{Fin} (\text{suc} \textit{n}). We illustrate the type for a couple of different \textit{n}'s in figure 7.

We illustrate what the \text{suc} constructor does in figure 8a. It takes an element of type \textit{Fin} \textit{n}, and returns the successor element of type \textit{Fin} (\text{suc} \textit{n}).

What if we want an element of the successor type without using the \text{suc} constructor? We might wish to simply 'promote' the type of an element. In figure 8d it becomes clear that all elements of type \textit{Fin} \textit{n} are also elements of \textit{Fin} (\text{suc} \textit{n}).
we initially can not decrement the state.

To do this promoting we use the function \(\text{inject}_1\), which is illustrated in figure 8b.

\[
\begin{align*}
\text{inject}_1 & : \forall \{m\} \to \text{Fin } m \to \text{Fin } (\text{suc } m) \\
\text{inject}_1 \text{ zero} & = \text{ zero} \\
\text{inject}_1 (\text{suc } i) & = \text{ suc } (\text{inject}_1 i)
\end{align*}
\]

Now, what if we find ourselves in a situation where we have an element of type \(\text{Fin } (\text{suc } n)\), and we want to return its predecessor, but of the successor type \(\text{Fin } (\text{suc } (\text{suc } n))\)? What we want to do is given an element \(\text{suc } x\), return \(x\). We can’t do this as is since the element \(\text{suc } x\) is of type \(\text{Fin } (\text{suc } n)\), the element \(x\) is of type \(\text{Fin } n\). To get the proper type we need to invoke \(\text{inject}_1\) twice, which is illustrated in figure 8c.

### 9.1 Time dependent example

Looking back at the time independent example, we reflect on the choice of state. The natural numbers seemed, and were, a reasonable choice. With the time dependent process at our disposal however we notice a source of ineffectiveness.

We consider the case where the process is always evaluated with zero as the initial state. After 1 step we could either have stayed or we went right, meaning the state is now zero or \(\text{suc } \text{zero}\). After 2 steps we could have gone left, stayed or gone right. In figure 9 this edge case is illustrated, and we note that the number of possible states after \(n\) steps is \(n + 1\).

\[
\begin{align*}
\text{Fin } n \\
1 & \Rightarrow \text{ } 0 \\
2 & \Rightarrow \{0,1\} \\
3 & \Rightarrow \{0,1,2\} \\
4 & \Rightarrow \{0,1,2,3\}
\end{align*}
\]

**Figure 7.** An illustration of the \(\text{Fin } n\) type. We emphasise that there are exactly \(n\) inhabitants of \(\text{Fin } n\).

Let \(\text{inject}_0\) be applied to the \(\text{Fin } n\) type. We capture this reasoning in the definition of \(\text{inject}_0\).

\[
\begin{align*}
\text{inject}_0 &: \forall \{m\} \to \text{Fin } m \to \text{Fin } (\text{suc } m) \\
\text{inject}_0 \text{ zero} & = \text{ zero} \\
\text{inject}_0 (\text{suc } i) & = \text{ suc } (\text{inject}_0 i)
\end{align*}
\]

The Agda type \(\text{Fin } n\) is a type with at most \(n\) elements. Using this type gives us a more precise definition of what the possible states are.

The controls are identical to those in the time independent case.

\[
\begin{align*}
1d\text{-}\text{control} &: (n : \mathbb{N}) \to \text{id}\text{-}state \ n \to \text{Set} \\
1d\text{-}\text{control } n \ \text{zero} & = \text{ ZAction} \\
1d\text{-}\text{control } n \ (\text{suc } x) & = \text{ SAction}
\end{align*}
\]

The \(\text{step}\) function says the same thing as in the previous example, but it says it a little differently. If the state is zero there is only two available controls, and we update the state like we did previously. However, if the state is greater than zero we need to change the types as described in section 9. For the \(\text{left}\) control the result has to be injected twice, and for the \(\text{stay}\) control it has to be injected once.

\[
\begin{align*}
1d\text{-}\text{step} &: (n : \mathbb{N}) \to (x : \text{id}\text{-}state \ n) \\
& \to (y : 1d\text{-}\text{control } n \ x) \to \text{id}\text{-}state \ (\text{suc } n) \\
1d\text{-}\text{step } n \ \text{zero} \ \text{ZS} & = \text{ zero} \\
1d\text{-}\text{step } n \ \text{zero} \ \text{ZR} & = \text{ suc } \text{zero} \\
1d\text{-}\text{step } n \ (\text{suc } x) \ \text{SL} & = \text{ inject}_1 (\text{inject}_1 x) \\
1d\text{-}\text{step } n \ (\text{suc } x) \ \text{SS} & = \text{ inject}_1 (\text{succ } x)
\end{align*}
\]

Now the entire system has been defined and we can package it as a \(\text{SDProcT}\).

\[
\begin{align*}
\text{finsystem} &: \text{SDProcT} \\
\text{finsystem} &= \text{SDPT } \text{id}\text{-}state \ 1d\text{-}\text{control} \ 1d\text{-}\text{step}
\end{align*}
\]

### 10 Combinators for the Time Dependent Case

Before we move on we want to highlight that the state now is dependent on the natural numbers. The controls of these time dependent processes are time dependent themselves. We capture this reasoning in the definition of \(\text{Con’}\).

\[
\begin{align*}
\text{Con’ } &: \text{Con } \mathbb{N} \to \text{Set} \\
\text{Con’ } S &= (t : \mathbb{N}) \to \text{Con } (S t)
\end{align*}
\]

Now on to the product combinator for the time dependent case. To combine two states that are time dependent we compute a new time dependent state that is the product of applying the prior states to the time.

\[
\begin{align*}
\text{×}_{\text{Case}} &: (S_1 \ S_2 : \text{Con } \mathbb{N}) \to \text{Con } \mathbb{N} \\
S_1 \ ×_{\text{Case}} S_2 &= \lambda t \to S_1 \ t \times S_2 \ t
\end{align*}
\]

The product combinator for two controls should produce a new \(\text{Con’}\) on \(S_1 \ ×_{\text{Case}} S_2\) defined in terms of two controls \(\text{Con’ } S_1\) and \(\text{Con’ } S_2\). The defining equation is similar to the time independent case, but the extra parameter time is given as the first argument.
The time dependent sum combinator for controls pattern matches on what injection was used, and applies the associated control to the time and the state.

\[ \lambda_{\mathbf{C}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \]
\[ \rightarrow \text{Con'} S_1 \rightarrow \text{Con'} S_2 \rightarrow \text{Con'} (S_1 \times_S S_2) \]
\[ (C_1 \times_C C_2) \text{ time } (s_1, s_2) = C_1 \text{ time } s_1 \times C_2 \text{ time } s_2 \]

Again we capture the type of the step function in a type \( \text{Step} \).
\[ \text{Step} : (S : \text{Con } \mathbb{N}) \rightarrow \text{Con' } S \rightarrow \text{Set} \]
\[ \text{Step } S C = (t : \mathbb{N}) \rightarrow (s : S t) \rightarrow C t s \rightarrow S (\text{suc } t) \]

Combining two such step functions is similar to the time independent case. The only different is that we have an extra parameter \( \text{time} \), and we must apply the step functions to this \( \text{time} \) parameters.

\[ \lambda_{\mathbf{sf}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \]
\[ \rightarrow \{ C_1 : \text{Con' } S_1 \} \rightarrow \{ C_2 : \text{Con' } S_2 \} \]
\[ \rightarrow \text{Step } S_1 C_1 \rightarrow \text{Step } S_2 C_2 \]
\[ \rightarrow \text{Step } (S_1 \times_S S_2) (C_1 \times_C C_2) \]
\[ (s_1 \times_f s_2) \text{ time state control} \]
\[ = s_1 \text{ time proj state} (\text{proj control}), \]
\[ s_2 \text{ time proj state} (\text{proj control}) \]

Finally, combining two time dependent sequential decision processes is done by applying the combinators componentwise.

\[ \lambda_{\mathbf{SDP}} : \text{SDProc}_T \rightarrow \text{SDProc}_T \rightarrow \text{SDProc}_T \]
\[ \text{SDPT } S_1 C_1 s_1 s_2 \xrightarrow{\mathbf{SDP}} \text{SDPT } S_2 C_2 s_2 \]
\[ = \text{SDPT} (S_1 \times_S S_2) (C_1 \times_C C_2) (s_1 \times_f s_2) \]

Just as the product combinator, the defining equation for the coproduct combinator is similar to its time independent counterpart. The difference is again that the parameters are applied to the time.

\[ \lambda_{\mathbf{m}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \rightarrow \text{Con } \mathbb{N} \]
\[ s_1 \xrightarrow{m} s_2 = \lambda t s_1 t \xrightarrow{m} s_2 t \]

The time dependent sum combinator for controls pattern matches on what injection was used, and applies the associated control to the time and the state.

\[ \lambda_{\mathbf{C}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \]
\[ \rightarrow \text{Con' } S_1 \rightarrow \text{Con' } S_2 \rightarrow \text{Con'} (S_1 \times_S S_2) \]
\[ (C_1 \times_C C_2) \text{ time } \{ (\text{inj}_1 s_1) \rightarrow C_1 \text{ time } s_1; \]
\[ (\text{inj}_2 s_2) \rightarrow C_2 \text{ time } s_2 \} \]

Combining the step functions to produce one defined for the new process is, similarly to the time independent case, done by pattern matching on the state. If the state is injected with the first injection, we apply the first step function, and similarly for the second injection.

\[ \lambda_{\mathbf{m}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \]
\[ \rightarrow \{ C_1 : \text{Con' } S_1 \} \rightarrow \{ C_2 : \text{Con' } S_2 \} \]
\[ \rightarrow \text{Step } S_1 C_1 \rightarrow \text{Step } S_2 C_2 \]
\[ \rightarrow \text{Step } (S_1 \times_S S_2) (C_1 \times_C C_2) (s_1 \xrightarrow{m} s_2) \]

To combine two time dependent processes into a yielding coproduct we begin by describing the component that relates the states in one process to states in the other.

\[ \lambda_{\mathbf{m}} : \text{SDProc}_T \rightarrow \text{SDProc}_T \rightarrow \text{SDProc}_T \]
\[ \text{SDPT } S_1 C_1 s_1 s_2 \xrightarrow{\mathbf{m}} \text{SDPT } S_2 C_2 s_2 \]
\[ = \text{SDPT} (S_1 \times_S S_2) (C_1 \times_C C_2) (s_1 \xrightarrow{m} s_2) \]

The first change from the coproduct combinator is again that the control space is extended to contain also the \textit{nothing} constructor.

\[ \lambda_{\mathbf{C}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \]
\[ \rightarrow \text{Con' } S_1 \rightarrow \text{Con' } S_2 \rightarrow \text{Con' } (S_1 \xrightarrow{m} S_2) \]
\[ (C_1 \xrightarrow{m} C_2) \text{ time } (\text{inj}_1 s_1) = \text{Maybe } (C_1 \times_C C_2) \]
\[ (C_1 \xrightarrow{m} C_2) \text{ time } (\text{inj}_2 s_2) = \text{Maybe } (C_2 \times_C C_2) \]

In contrast to the coproduct case, the new step function will switch which process is executing if the control is the \textit{nothing} constructor, and otherwise, depending on which injection was used, apply one of the previous step functions.

\[ \lambda_{\mathbf{m}} : \{ S_1, S_2 : \text{Con } \mathbb{N} \} \]
\[ \rightarrow \{ C_1 : \text{Con' } S_1 \} \rightarrow \{ C_2 : \text{Con' } S_2 \} \rightarrow S_1 \xrightarrow{m} S_2 \]
\[ \rightarrow \text{Step } S_1 C_1 \rightarrow \text{Step } S_2 C_2 \]
\[ \rightarrow \text{Step } (S_1 \xrightarrow{m} S_2) (C_1 \xrightarrow{m} C_2) \]
\[ \xrightarrow{m} \text{ inj}_1 (s_f \times_f s_2) \text{ time } (\text{inj}_1 s_1) \text{ just } c_1 = \text{inj}_2 (s_f \times_f s_2) \text{ time } (\text{inj}_1 s_1) \text{ just } c_1 \]
Again we provide an infix operator to be consistent.

\[
(\text{sf} \, \text{time} \, \text{inj}_1 \, \text{nothing}) = \text{inj}_2 (r \, \text{time} \, s_1)
\]

\[
(\text{sf} \, \text{time} \, \text{inj}_2 \, \text{just} \, c_2) = \text{inj}_2 (s_2 \, \text{time} \, c_2)
\]

\[
(\text{sf} \, \text{time} \, \text{inj}_1 \, \text{nothing}) = \text{inj}_1 (r \, \text{time} \, s_2)
\]

Again we provide an infix operator to be consistent.

\[
\text{syntax } \omega^m_{\text{sf}} \, r \, \text{sf}_1 \, \text{sf}_2 = \text{sf}_1 \, (r \, ) \, \text{sf}_2
\]

To create a yielding coproduct we use the same combinator for the state space, but use the new modified combinators for the control space and step function.

\[
\omega^m_{\text{SDP}} : (p_1 \, p_2 : \text{SDProcT}) \to (\#_{a_1} \, p_1) \leftarrow (\#_{a_2} \, p_2) \to \text{SDProcT}
\]

\[
\omega^m_{\text{SDP}} (\text{SDPT} \, S_1 \, C_1 \, \text{sf}_1) (\text{SDPT} \, S_2 \, C_2 \, \text{sf}_2) \, r = \text{SDPT} (S_1 \, \text{inj}_1 \, S_2) (C_1 \, \text{inj}_2 \, C_2) (\text{sf}_1 \, (r \, ) \, \text{sf}_2)
\]

When we try to implement the interleaved combinator for the time dependent case we run into some problems. The main problem is that since the step function only advances one of the state components, the other one will be of the wrong type. At time \( n \) one of the components get advanced to a state in time \( \text{suc} \, n \), while the other is not changed at all.

### 11 Policy Combinators

Now that we have a way of reusing sequential decision processes to create more sophisticated processes, we want to reuse existing policy sequences also. As the function \( \text{trajectory} \) which observes a process accepts a process and a policy sequence as input, this would simplify combining and observing processes without having to write any new policy sequences. We start by combining single policies.

We remind the reader that a policy is defined in terms of a state and a control.

\[
P : (S : \text{Set}) \to (C : S \to \text{Set}) \to \text{Set}
\]

\[
P \, S \, C = (s : S) \to C \, s
\]

A policy for a product process defined in terms of two policies for the individual processes, is created by taking a pair of the two previous policies applied to the components of the state.

\[
\text{\_\_}_\text{p}_\text{-} : \{S_1 \, S_2 : \text{Set}\} \{C_1 : \text{Con} \, S_1\} \{C_2 : \text{Con} \, S_2\}
\]

\[
\to P \, S_1 \, C_1 \to P \, S_2 \, C_2
\]

\[
\to P \, (S_1 \times S_2) \,(C_1 \, \text{\_\_} \, C_2)
\]

\[
(p_1 \, \text{\_\_} \, p_2) \,(s_1 \, s_2) = p_1 \, s_1 \, p_2 \, s_2
\]

A policy for the sum of two processes is defined by pattern matching on the state. If the pattern matches on the left injection, we can reuse the previous policy defined on that state. Similarly, if the pattern matches on the right injection we can reuse the given policy for the other process.

\[
\text{\_\_}_\text{p}_\text{-} : \{S_1 \, S_2 : \text{Set}\} \{C_1 : \text{Con} \, S_1\} \{C_2 : \text{Con} \, S_2\}
\]

\[
\to P \, S_1 \, C_1 \to P \, S_2 \, C_2
\]

\[
\to P \, (S_1 \, \text{\_\_} \, S_2) \,(C_1 \, \text{\_\_} \, C_2)
\]

\[
(p_1 \, \text{\_\_} \, p_2) \,(s_1 \, s_2) = p_1 \, s_1 \, p_2 \, s_2
\]

### 12 A note on Policies

With these policy combinators defined, we make a few observations. These observations intend to illustrate some of the meanings behind policies.

#### 12.1 Product State Policies

Recall a policy for a product state. The type of this policy is \((s : \text{State}) \to \text{Control} \, s\), where State is the type of states and Control is the type family of controls. Since \( s \) is a product, the signature actually is something like \((a, b) : A \times B \to \text{Control} \, (a, b)\). We know from currying that this is the same as \((a : A) \to (b : B) \to \text{Control} \, (a, b)\).

In the case of the product combinator the control space was the product of the separate state components control spaces. Where a policy for the individual process would base the choice of control on the state, the policy for a product process will select a control for the same state component, but base the choice on both state components.

In the case of the interleaved process we related the situation to that of a game, where players take turns making their
moves. In such a scenario a policy could be something of a game leader, making the best choices for each component based both components. The policy can, after all, make a decision for one of the components based on the state of both components.

12.2 Sum state implies Product Policy
Another interesting observation to make is that a policy for a process with a sum state, e.g. a policy for a coproduct, is a pair of policies for the separate processes. We can make this concrete with the following two definitions.

\[ \forall \inj : \{ S_1 S_2 : \text{Set} \} \]
\[ \{ C_1 : \text{Con } S_1 \} \{ C_2 : \text{Con } S_2 \} \]
\[ \rightarrow \text{P } (S_1 \uplus S_2)(C_1 \uplus C \times C_2) \]
\[ \rightarrow \text{P } S_1 C_1 \times P S_2 C_2 \]

\[ \forall \inj \times \text{policy } = \lambda s_1 \rightarrow \text{policy } (\inj_1 s_1), \]
\[ \lambda s_2 \rightarrow \text{policy } (\inj_2 s_2) \]

Then we can further solidify this statement by showing that they are equal by functional extensionality.

\[ \forall \inj \times : \{ S_1 S_2 : \text{Set} \} \]
\[ \{ C_1 : \text{Con } S_1 \} \{ C_2 : \text{Con } S_2 \} \]
\[ (p : \text{P } (S_1 \uplus S_2)(C_1 \uplus C \times C_2)) \]
\[ \rightarrow \text{state } : S_1 \uplus S_2 \]
\[ \rightarrow \times \rightarrow \uplus (\inj \rightarrow x\text{p } \text{state } \equiv p \text{ state} \]

\[ \forall \inj \times - (\inj_1 _) = \text{refl} \]
\[ \forall \inj \times - (\inj_2 _) = \text{refl} \]

13 Conclusions and future work
We have shown that sequential decision processes can quite comfortably be combined. Things start to get interesting when we consider writing new policies rather than combining existing policies. As exemplified in section 11 a policy for a product state is essentially a policy for one process parameterised over another. This becomes particularly interesting when we look at the interleaved process where we only wish to advance one component but we now have information about the other also. Using backwards induction to compute optimal policy sequences could perhaps use this extra information to produce very clever sequences.

There are many avenues left to explore for future work.

- Here we have mainly focused on sequential decision processes, and not so much on problems. It is not entirely clear how problems should be combined. The problems need to offer the same type of reward, normalised to some range and then combined in some way. Even when the type of the reward is the natural numbers, combining them is not unambiguous as there are many ways to combine natural numbers.
- There are definitely many more combinators than those presented by us in this text. Here we focused mainly on generic combinators that required no additional information and then briefly touched the yielding coproduct, which did require additional information in order to switch processes.
- Both processes and problems can be parameterised over a monad. A processes step function can e.g produce a list of possible next states and associate a probability to each of them. Modeling such uncertainties vastly increases the amount of problems that can be modeled using this framework. When combining processes like these combining the step functions becomes less trivial.
- Many claims and ideas in this text has been described but not formalised. As an example the notion of a unit and its properties could be formalised, and then the units presented here could be proven to be correct units. We are also interested in implementing an n-ary combinator for the interleaved combinator, as well as an interleaved combinator for the time dependent case.