Categorical semantics of compositional reinforcement learning

Georgios Bakirtzis, Michail Savvas, and Ufuk Topcu

Abstract

Reinforcement learning (RL) often requires decomposing a problem into subtasks and composing learned behaviors on these tasks. Compositionality in RL has the potential to create modular subtask units that interface with other system capabilities. However, generating compositional models requires the characterization of minimal assumptions for the robustness of the compositional feature. We develop a framework for a *compositional theory* of RL using a categorical point of view. Given the categorical representation of compositionality, we investigate sufficient conditions under which learning-by-parts results in the same optimal policy as learning on the whole. In particular, our approach introduces a category MDP, whose objects are Markov decision processes (MDPs) acting as models of tasks. We show that MDP admits natural compositional operations, such as certain fiber products and pushouts. These operations make explicit compositional phenomena in RL and unify existing constructions, such as puncturing hazardous states in composite MDPs and incorporating state-action symmetry. We also model sequential task completion by introducing the language of zig-zag diagrams that is an immediate application of the pushout operation in MDP.

1 Introduction

Compositionality is necessary to learn increasingly complex tasks, manage and produce trustworthy system behavior, and reduce hazardous control actions. Exploiting the compositionality feature requires that we have a precise way of comprehending the behavior of the whole by examining the behavior of the parts and vice versa [Szabó, 2012]. As reinforcement learning (RL) raises the flexibility of the behaviors systems can adapt to, it is essential to trust and control systems that include RL such that they will not cause us harm.

One approach to addressing this concern is to exploit the compositionality feature directly by making the interconnection of the parts as explicit as possible [Gur et al., 2021, Li et al., 2021]. By making exact how different parts of the system interconnect and in the precise way they do so, it is possible to examine the total system behavior by its components, thereby reducing the possibility of unwanted emergent behaviors. Exploiting the compositionality features in this context suggests that regardless of the data or flexibility of the RL algorithms we deploy, it is beneficial to equip a symbolic and semantic framework for reasoning in (un)wanted behaviors the system should produce.

Compositionality is one part of the framework in engendering increasingly adaptable systems. In particular, we take the control and RL algorithms as a given and aim to provide a unifying framework for studying compositional phenomena in systems incorporating RL. When compositionality occurs, the behavior of a system is *predictable* by the combination of the behaviors *of* its parts and the combination *preserves* properties emerging *from* the parts [Genovese, 2018].

To achieve this explicit definition of compositionality in RL, we give a symbolic and semantic interpretation of compositional phenomena by working within the notion of a category. Category theory is well-suited for posing and solving *structural* problems. Categorical semantics, in particular, is a language of relationships between mathematical objects with additional structure, such as algebras. As such, categorical constructions, like fiber products and pushouts, describe general operations that model (de)compositions of objects, thereby giving meaning to clusters of objects *as a composite* and how they interface with each other.

The Markov decision process (MDP) is the structural instantiation of RL models. We study compositionality from the perspective of creating subprocess MDPs and prove the properties of MDPs used in RL that create general interfaces with each other. Such interfaces can take the form of bottom-to-top horizontal composition, the classical understanding of composition due to Frege [Szabó, 2022], and top-to-bottom vertical composition with other models, which address new types of composition as systems become increasingly complex [Coecke, 2021]. Categorical semantics make explicit the horizontal and vertical compositions of MDPs in different configurations. In this setting, categories define a precise *syntax* for MDP instantiations and assign *semantics* to particular compositional behaviors we want to examine (section 2).

To show the expressiveness of this compositional RL framework, we apply categorical semantics to three concrete cases. The first shows how the operation of fiber products can model a superposition of obstacles in the grid world problem (section 3). The second is a general treatment of state-action symmetry, where an MDP admits an action by a group to more economically derive policies for symmetrically equivalent actions (section 4). The third is a design of sequential task completion, where we show how zig-zag diagrams can totally model a fetch-and-place robot (section 5). Zig-zag diagrams denote composite MDPs produced by gluing together subprocess MDPs, preserving the relationship between types of actions and states. While this paper aims to present a compositional *theory* of RL, our constructions have a computational interpretation that can partition problem spaces and manage the increasing complexity of RL systems.

Our contributions:

1. We define the category MDP of MDPs—the structural unit of our theory—which is a general language that unifies compositional constructions (subprocess MDPs, safe grid worlds, state-action symmetry, sequential task completion, to name a few).
2. We prove interesting properties of MDP that exploit the compositionality feature, i.e., the existence of certain fiber products and pushouts, which have physical implications for robust RL.
3. We develop categorical semantics that gives precise meaning to a composite RL system derived from its parts and, additionally, make explicit how they ought to interface with the other systems *as a composite*. These rules give properties that have the potential of reducing the hand-written nature of heuristic rules since they apply generally.

Conventions Subscripts for MDP elements refer to a particular instantiation of the definition. For example, $S_{\mathcal{N}}$ refers to the state space of the MDP $\mathcal{N}$. We summarize the fundamental mathematical constructs we use in appendix A. We include the constructions and write all proofs in the appendices.

2 Categorical semantics of compositional reinforcement learning

We start by developing a definition for MDPs containing some subtle generalizations, although the definition is congruent with the traditional MDP instantiation.

Definition 1 (MDP). *An MDP $\mathcal{M} = (S, A, \psi, T)$ is composed of the data:*

- *The state space S, a measurable space with a fixed σ-algebra.*
- *The state-action space A (which also includes the data of the current state $s \in S$) in the action.*
- *A function $\psi : A \to S$ that maps $a \in A$ to its associated state $s \in S$.*
- *The information of the transition probabilities, given as a function $T : A \to \mathcal{P}_S$, where $\mathcal{P}_S$ denotes the space of probability measures on S.*

In the above definition, the actions that an agent can take at a particular state s are given by the set $\psi^{-1}(s) \subseteq A$. We denote this set of available actions for each state s by

A_s. Knowing the action spaces A_s for all $s \in S$, one may recover A and ψ by

$$A = \coprod_{s \in S} A_s,\ \psi : A_s \longmapsto s \in S. \tag{1}$$

Remark 1. If the actions in the MDP are not state-specific, then we can take $A = S \times A_0$ for a fixed set of actions A_0 and $\psi : A \to S$ is just the projection onto the first factor.

The following definition constructs a category whose objects are MDPs (definition 1).

Definition 2 (Category of MDPs)**.** *MDPs form a category* MDP *whose morphisms are as follows. Let $\mathcal{M}_i = (S_i, A_i, \psi_i, T_i)$, with $i = 1, 2$, be two MDPs.*

A morphism $m = (f, g) : \mathcal{M}_1 \to \mathcal{M}_2$ is the data of a measurable function $f : S_1 \to S_2$ and a function $g : A_1 \to A_2$ satisfying the following compatibility conditions:

1. *The diagram*

$$\begin{tikzcd} A_1 \arrow[r, "g"] \arrow[d, "\psi_1"'] & A_2 \arrow[d, "\psi_2"] \\ S_1 \arrow[r, "f"'] & S_2 \end{tikzcd} \tag{2}$$

is commutative.

2. *The diagram*

$$\begin{tikzcd} A_1 \arrow[r, "g"] \arrow[d, "T_1"'] & A_2 \arrow[d, "T_2"] \\ \mathcal{P}_{S_1} \arrow[r, "f_*"'] & \mathcal{P}_{S_2} \end{tikzcd} \tag{3}$$

is commutative, where f_* maps a probability measure $\mu_1 \in \mathcal{P}_{S_1}$ to its pushforward $\mu_2 = f_*\mu_1 \in \mathcal{P}_{S_2}$ under f.

The constant MDP pt is the MDP pt whose state space and action spaces are the one-point set. Every MDP $\mathcal{M}$ admits a unique, natural morphism $\mathcal{M} \to$ pt and pt is the terminal object in MDP.

To account for learning, we augment MDPs with the immediate reward function to obtain a category MDP$_+$: Its objects are pairs $(\mathcal{M}, R)$ where $\mathcal{M}$ is a MDP and $R : A \to \mathbb{R}$ is the reward function. Morphisms between pairs $(\mathcal{M}_1, R_1)$ and $(\mathcal{M}_2, R_2)$ are morphisms $(f, g) : \mathcal{M}_1 \to \mathcal{M}_2$ which in addition satisfy $R_1 = R_2 \circ g : A_1 \to \mathbb{R}$. The category MDP$_+$ recovers, e.g., reward machines, which are a specific notion of operationalizing tasks compositionally in RL [Neary et al., 2021].

This category spans a large number of structures, e.g., beyond MDPs we can see automata used in verification [Jothimurugan et al., 2021] as residing in a subcategory of MDP. The category itself is agnostic to a particular learning algorithm, as long as it appropriately produces policies based on the compositional structure. We can think of the category MDP as a generalization over any type of MDP that is compositional in nature.

In particular, the two commutative diagrams above show us when two MDPs are *compatible* in the sense that their interfaces agree. Namely, diagram (2) guarantees that if an action a_1 in MDP $\mathcal{M}_1$ is associated to a state $s_1 \in S_1$, then its image action $a_2 = g(a_1)$ under m is associated to the image state $s_2 = f(s_1)$. Similarly, diagram (3) ensures (in a discrete setting) that the transition probability from any state s_1 to another state s_1' under taking action a_1 in $\mathcal{M}_1$ is equal to the transition probability from the state $s_2 = f(s_1)$ to $s_2' = f(s_1')$ under action $a_2 = g(a_1)$ in $\mathcal{M}_2$.

In the rest of the section, we explore more of the properties of the category MDP, namely the definition of subobjects, fiber products, and pushouts.

An important definition in relation to the category MDP is the notion of subprocess, which allows us to speak more concretely about compositionality.

Subprocesses The definition of morphism correctly captures the notion of a subprocess of an MDP.

Definition 3 (Subprocess of MDP)**.** *We say that $\mathcal{M}_1$ is a subprocess of the MDP $\mathcal{M}_2$ if there exists a morphism $(f, g) : \mathcal{M}_1 \to \mathcal{M}_2$ such that f and g are injective.*

We say that $\mathcal{M}_1$ is a full subprocess if diagram 2 is moreover cartesian.

Since f is injective, we may consider the state space S_1 as a subset of S_2. Moreover, the condition that diagram 2 is cartesian means that the only available actions on S_1 come from MDP $\mathcal{M}_1$. Thus, $\mathcal{M}_1$ being a full subprocess of $\mathcal{M}_2$ implies that an agent following the MDP $\mathcal{M}_2$ who finds themself at a state $s_1 \in S_1$ will remain within S_1 no matter which action $a_1 \in A_1$ they elect to apply.

Conversely, for a MDP $\mathcal{M}_2$ and any subset $S_1 \subseteq S_2$ there is a canonical subprocess $\mathcal{M}_1$ with state space S_1, whose action space A_1 is defined by

$$A_1 := \psi_2^{-1}(S_1) \cap T_2^{-1}(f_\star(\mathcal{P}_{S_1})). \qquad (4)$$

In fact, $\mathcal{M}_1$ is uniquely characterized as the maximal such subprocess.

Proposition 1. *Any subprocess $\mathcal{M}_1' \to \mathcal{M}_2$ with state space S_1 factors uniquely through the subprocess $\mathcal{M}_1 \to \mathcal{M}_2$.*

Fiber products

Interesting categorical properties usually are *universal*. Universal properties represent specific ideals of behavior within a defined category [Spivak, 2014]. A fiber product is a categorical generalization of the notion of cartesian product whose universal property is determined by it being maximal in a certain sense. Besides the cartesian product, another common example is the intersection $S_1 \cap S_2$ of two subsets $S_i \subseteq S_3$, obtained as the fiber product of the two inclusions of sets $S_1 \to S_3$ and $S_2 \to S_3$.

We prove that well-formed MDPs have partial fiber products in the category MDP, meaning that there exists a way to compose the data of one MDP with another to create an interface between MDPs (horizontal composition).

The intuition behind a fiber product of MDPs is the idea of “intersecting” two MDPs $\mathcal{M}_1$ and $\mathcal{M}_2$ viewed as components of a third MDP $\mathcal{M}_3$ through morphisms $m_1 : \mathcal{M}_1 \to \mathcal{M}_3$ and $m_2 : \mathcal{M}_2 \to \mathcal{M}_3$. In general, the morphisms m_i do not need to defined subprocesses and this allows for valuable flexibility. For example, the cartesian product $\mathcal{M}_1 \times \mathcal{M}_2$ of two MDPs will be obtained when $\mathcal{M}_3 = \mathrm{pt}$ is the constant MDP (remark 4). In the other extreme, we can obtain intersections of subprocesses, as for the grid world problem.

Write $\mathcal{M}_i = (S_i, A_i, \psi_i, T_i)$ for $i = 1, 2, 3$. The structure of the state space as a measure space requires some care in the construction and is the reason for the failure of existence of a fully universal fiber product. We examine the simpler subprocess case and the case of S_3 finite and develop a general construction that support the following theorem, such as fiber products. When the state spaces are finite or discrete (which is the case in practical applications), we can summarize our results as follows.

Theorem 1. *There exists an MDP $\mathcal{M} = \mathcal{M}_1 \times_{\mathcal{M}_3} \mathcal{M}_2$ with state space $S = S_1 \times_{S_3} S_2$ and action space $A = A_1 \times_{A_3} A_2$ which fits in a commutative diagram in* MDP :

$$\begin{tikzcd}
\mathcal{M} \arrow[r] \arrow[d] & \mathcal{M}_1 \arrow[d, "m_1"'] \\
\mathcal{M}_2 \arrow[r, "m_2"] & \mathcal{M}_3
\end{tikzcd}$$

$\mathcal{M}$ is universal with respect to diagrams whose morphisms are conditionally independent.

Pushouts

The categorical notion of pushout is the contravariant analogue of a fiber product. It thus models gluing two objects along a third object with morphism to each. Its universal property is determined by it being minimal in an appropriate sense. A standard example of a pushout are coproducts. In the category of sets, this reduces to the disjoint union $S_1 \coprod S_2$, which can be viewed as the pushout of the two morphisms $\emptyset \to S_1$ and $\emptyset \to S_2$.

Intuitively, the pushout is the result of gluing two MDPs $\mathcal{M}_1$ and $\mathcal{M}_2$ along a third MDP $\mathcal{M}_3$ which is expressed as a component of both through morphisms $m_1 : \mathcal{M}_1 \to \mathcal{M}_3$ and $m_2 : \mathcal{M}_2 \to \mathcal{M}_3$.

Theorem 2. *There exists an MDP $\mathcal{M} = \mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ which is the pushout of the diagram in* MDP :

$$\begin{tikzcd}
\mathcal{M}_3 \arrow[r, "m_1"] \arrow[d, "m_2"'] & \mathcal{M}_1 \\
\mathcal{M}_2 &
\end{tikzcd}$$

Gluing behaves well with respect to subprocesses.

Proposition 2. *Suppose that $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$. Then $\mathcal{M}_1$ and $\mathcal{M}_2$ are subprocesses of $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$.*

Incorporating rewards

In order to *do* RL we have to assign rewards in the enlarged category MDP_+. We can define fiber products and pushouts MDP_+ to extend our previously proven propositions and theorems for MDPs with reward, showing the generality of the category MDP and its associated properties.

For fiber products, let $m_i = (f_i, g_i)\colon (\mathcal{M}_i, R_i) \to (\mathcal{M}_3, R_3)$ be two morphisms in MDP_+. Their fiber product is the pair $(\mathcal{M}, R)$ where $\mathcal{M} = \mathcal{M}_1 \times_{\mathcal{M}_3} \mathcal{M}_2$ and $R = R_3 \circ f_1 \circ pr_1 = R_3 \circ f_2 \circ pr_2$.

For pushouts, let $m_i = (f_i, g_i)\colon (\mathcal{M}_3, R_3) \to (\mathcal{M}_i, R_i)$ be two morphisms in MDP_+. Their pushout is the pair $(\mathcal{M}, R)$ where $\mathcal{M} = \mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ and R is defined to be R_1 on the image $i_1(S_3) \subseteq S$ and R_2 on the image $i_2(S_3)$. This is well-defined since by definition $R_1 \circ f_1 = R_3 = R_2 \circ f_2$.

3 Safe grid worlds

We consider the case of a grid world [Leike et al., 2017] constructed as a 4×4 grid, where an agent attempts to navigate from a starting position to a destination position given some obstacles (yellow, green and red respectively in figure 1).

Definition 3 also well-suited to removing a subset $\mathbb{O} \subseteq S$ from the state space of a MDP to obtain a new MDP.

Definition 4 (Puncturing undesired states and actions)**.** *Let $\mathcal{M} = (S, A, \psi, T)$ be a MDP and $\mathbb{O} \subseteq S$. The MDP $\mathcal{M}^\circ$ obtained by puncturing $\mathcal{M}$ along $\mathbb{O}$ is the MDP $(S^\circ, A^\circ, \psi^\circ, T^\circ)$ where:*

1. $S^\circ = S \setminus \mathbb{O}$.
2. *Let $B = \{a \in A \mid T(a)(\mathbb{O}) > 0\}$. Then $A^\circ = A \setminus (\psi^{-1}(\mathbb{O}) \cup B)$.*
3. $\psi^\circ = \psi|_{A^\circ}$.
4. $T^\circ = T|_{A^\circ}$.

There is a canonical morphism $\mathcal{M}^\circ \to \mathcal{M}$, which exhibits $\mathcal{M}^\circ$ as a subprocess of $\mathcal{M}$. In fact, $\mathcal{M}^\circ$ is the canonical maximal subprocess associated to the subset $S^\circ \subseteq S$, constructed above.

For a morphism of MDPs $m = (f, g)\colon \mathcal{M}_1 \to \mathcal{M}_2$, the punctured MDP $\mathcal{M}_2^\circ$ obtained by puncturing $\mathcal{M}_2$ along $\mathcal{M}_1$ is defined as the puncture of $\mathcal{M}_2$ along the subset $f(S_1) \subseteq S_2$.

Gluing (section 2) also behaves well with respect to puncturing (proposition 2 and 3).

Proposition 3. *Suppose $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$ and any action $a_2 \in A_2 \setminus g_2(A_3)$ is not supported on S_3, meaning that there is some measurable subset $U_2 \subseteq S_2$ disjoint from $f_2(S_3)$ such that $T(a_2)(U_2) > 0$. Let $\mathcal{M}_2^\circ$ be the MDP obtained by puncturing $\mathcal{M}_2$ along $\mathcal{M}_3$. Then the MDP $(\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2)^\circ$ obtained by puncturing $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ along the subprocess $\mathcal{M}_2^\circ$ is the MDP $\mathcal{M}_1$.*

This construction is well-defined, since for any action in A° the probability of ending in $\mathbb{O}$ is zero, as we removed exactly these actions, which formed the set B. This is where allowing the action spaces to vary along the state space S gives us flexibility to modify the actions locally to avoid the set $\mathbb{O}$.

We give an example of the naturality of the above constructions in the context of puncturing MDPs to enforce a safety condition.

Example 1 (Static obstacles)**.** *Fix an MDP of represented by the form $\mathcal{M} = (S, A, \psi, T, R)$ and consider two disjoint subsets $\mathbb{O}_1, \mathbb{O}_2 \subseteq S$. We can then construct the MDPs:*

1. *The punctured MDPs $\mathcal{M}_i^\circ$ along $\mathbb{O}_i$ for $i = 1, 2$.*
2. *The punctured MDP $\mathcal{M}_{12}^\circ$ along $\mathbb{O}_1 \cup \mathbb{O}_2$.*

Proposition 4. *There exists a commutative diagram*

$$\begin{tikzcd}
\mathcal{M}_{12} \arrow[r] \arrow[d] & \mathcal{M}_1 \arrow[d] \\
\mathcal{M}_2 \arrow[r] & \mathcal{M}
\end{tikzcd}$$

which is simultaneously a fiber and pushout diagram in MDP*, meaning that $\mathcal{M}_{12} = \mathcal{M}_1 \times_{\mathcal{M}} \mathcal{M}_2$ and $\mathcal{M} = \mathcal{M}_1 \cup_{\mathcal{M}_{12}} \mathcal{M}_2$.*

Example 2 (Collision avoidance)**.** *Fix an MDP $\mathcal{M}_1 = (S, A, \psi, T, R)$, which we consider as a model for an agent moving through the state space S with possible actions A.*

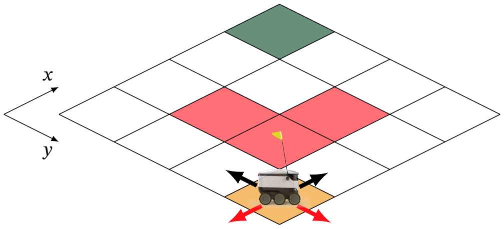

Figure 1: A grid world with a starting state (yellow, s), a destination state (green, $\mathbb{D}$), and obstacles (red, $\mathbb{O}$). We can have arbitrary complex worlds of this form; we use the simplest example to illustrate some of the properties we can model within the categorical theory of RL.

To model the movement of two independent agents, we may work with the product $\mathcal{M}_2 = \mathcal{M} \times \mathcal{M}$.

If, in addition, we would like to make sure that the two agents never collide, we may puncture $\mathcal{M} \times \mathcal{M}$ along the diagonal

$$\mathbb{O}_2 = \Delta = \{(s, s) \mid s \in S\} \subseteq S \times S$$

to get the MDP $\mathcal{M}_2^{\circ}$.

We can use, e.g., the above construction to model the movement of N independent agents, ensuring that no two of them collide, by puncturing the product MDP with N factors $\mathcal{M}_N = \mathcal{M} \times \ldots \times \mathcal{M}$ along the big diagonal

$$\begin{aligned} \mathbb{O}_N &= \bigcup_{1 \le i < j \le N} \Delta_{ij} \\ &= \bigcup_{1 \le i < j \le N} \{(s_1, \ldots, s_N) \mid s_i = s_j\} \subseteq S^N \end{aligned}$$

to define the MDP $\mathcal{M}_N^{\circ}$.

Zig-zag diagrams

For designing compositional tasks, we desire to operationalize using the categorical semantics of RL, that involve accomplishing tasks sequentially. In a general setting, we consider the setup given by, what we term, a *zig-zag* diagram of MDPs

$$\begin{tikzcd}
& \mathcal{N}_0 \arrow[dl] \arrow[dr] & & \mathcal{N}_1 \arrow[dl] \arrow[dr] & & \ldots & & \mathcal{N}_{n-1} \arrow[dl] \arrow[dr] & \\
\mathcal{M}_0 & & \mathcal{M}_1 & & \mathcal{M}_2 \ \ldots & & \mathcal{M}_{n-1} & & \mathcal{M}_n
\end{tikzcd} \tag{5}$$

where for each $i = 0, \ldots, n-1$, $\mathcal{N}_i$ is a subprocess of $\mathcal{M}_i$ (cf. definition 3).

The composite MDP associated to the above diagram is the MDP $\mathcal{C}_n$ defined by the inductive rule

$$\mathcal{C}_0 := \mathcal{M}_0, \mathcal{C}_1 := \mathcal{C}_0 \cup_{\mathcal{N}_0} \mathcal{M}_1, \cdots, \mathcal{C}_n := \mathcal{C}_{n-1} \cup_{\mathcal{N}_{n-1}} \mathcal{M}_n.$$

The intuitive interpretation of the above zig-zag diagram is that it black boxes compositional task completion. In particular, each subprocess $\mathcal{N}_i \to \mathcal{M}_i$ models the completion of a task in the sense that the goal of an agent is to eventually find themself at a state of $\mathcal{N}_i$. Once the i-th goal is accomplished inside the environment given by $\mathcal{M}_i$, we allow for the possibility of a changing environment and more options for states and actions in order to achieve the next goal modeled by the subprocess $\mathcal{N}_{i+1} \to \mathcal{M}_{i+1}$.

The composite MDP $\mathcal{C}_n$ is a single environment capturing all the tasks at the same time.

⟨?⟩ Suppose that an agent has learned an optimal policy for each MDP $\mathcal{M}_i$ given the reward function R_i for achieving the i-th goal for each $i = 0, \ldots, n$. Under what conditions do these optimal policies determine the optimal policy for the joint reward on the composite MDP $\mathcal{C}_n$?

A scenario in which this is true is when the zig-zag diagram is forward-moving meaning that $\mathcal{N}_i$ is a full subprocess of $\mathcal{M}_i$ and, moreover, the optimal value function $v_\star(s)$ for any state s in the state space of a component $\mathcal{M}_i$, considered as a state of $\mathcal{C}_n$, is *monotonic* with respect to subsequent subprocesses $\mathcal{M}_{i+1}, \ldots, \mathcal{M}_n$. Monotonicity here means that the expressions

$$\sum_{s' \in S_i} T(a)(s')(R_i(a) + \gamma \cdot v_\star^{\mathcal{C}_n}(s'))$$

$$\sum_{s' \in S_i} T(a)(s')(R_i(a) + \gamma \cdot v_\star^{\mathcal{C}_{[i,n]}}(s'))$$

are maximized by the same action $a \in (A_i)_s$. Here $\mathcal{C}_{[i,n]}$ denotes the composite MDP associated to the truncated zig-zag diagram

$$\begin{tikzcd}
& \mathcal{N}_i \arrow[dl] \arrow[dr] & & \mathcal{N}_{i+1} \arrow[dl] \arrow[dr] & & \ldots & & \mathcal{N}_{n-1} \arrow[dl] \arrow[dr] & \\
\mathcal{M}_i & & \mathcal{M}_{i+1} & & \mathcal{M}_{i+2} \ \ldots & & \mathcal{M}_{n-1} & & \mathcal{M}_n.
\end{tikzcd}$$

A zig-zag diagram can always be made forward-moving by removing the actions of $\mathcal{N}_i$ that can potentially move the agent off $\mathcal{N}_i$ back into $\mathcal{M}_i$. This is the operation of puncturing $\mathcal{M}_i$ along the complement of $\mathcal{N}_i$ and intersecting the result with $\mathcal{N}_i$.

This scenario occurs in practice if the way of achieving a task is independent of the subsequent behaviour of the agent, e.g., for a fetch-and-place robot in the example below whether the object to be fetched is stationary (section 5). Once the robot fetches the object, the environment is independent of how this was achieved.

Theorem 3. *Suppose that the zig-zag diagram is forward-moving and the optimal value function of $\mathcal{C}_n$ is monotonic. Then, following the policy π_i on each component $\mathcal{M}_i$ gives an optimal policy on the composite MDP $\mathcal{C}_n$.*

Example 3 (Visiting regions sequentially). *Consider a point mass robot sequentially visiting three regions, R_1, R_2, R_3, in a grid world, while always avoiding obstacles.*

We can model this problem compositionally by the following zig-zag diagram. We consider the MDPs punctured (proposition 4)

$$\begin{tikzcd}
& R_1 & & R_2 & & R_3 & \\
& \mathsf{pt} \arrow[dl] \arrow[dr] & & \mathsf{pt} \arrow[dl] \arrow[dr] & & \mathsf{pt} \arrow[dl] \arrow[dr] & \\
\mathcal{M}_1^\circ & & \mathcal{M}_2^\circ & & \mathcal{M}_3^\circ & & \mathcal{M}_4^\circ .
\end{tikzcd}$$

At any given position in the grid the five options for the point mass robot are forward, backwards, left, right, and stay in the same position. Here each $\mathcal{M}_i^\circ$ denotes the MDP in which all the obstacles have been punctured and the actions forward, backwards, left, right have been removed at the region R_i. Each intermediate subprocess $\mathsf{pt} \to \mathcal{M}_i$ maps the stationary point to the (point) region R_i. This requires us moving from MDP to MDP to puncture the actions that lead to moving away from the next subsequent sequence we want.

The problem is, therefore, totally defined by the composite MDP:

$$\mathcal{C}_{\text{robot}} = \mathcal{M}_1^\circ \cup_{\mathsf{pt}} \mathcal{M}_2^\circ \cup_{\mathsf{pt}} \mathcal{M}_3^\circ \cup_{\mathsf{pt}} \mathcal{M}_4^\circ .$$

Observe that this zig-zag diagram is forward-moving and the optimal value function of $\mathcal{C}_{\text{robot}}$ is monotonic. Thus, by theorem 3, the optimal policy on $\mathcal{C}_{\text{robot}}$ is given by the optimal policy of each component $\mathcal{M}_i^\circ$.

4 State-action symmetry

One of the benefits of working with algebras in some category is that we can use gadgets from a particular algebraic domain to speak about the geometry of the problem. Using algebraic gadgets becomes apparently helpful when we consider efficient approaches to *symmetric* RL problems, such as homomorphic networks [van der Pol et al., 2020]. We will investigate how homomorphic networks port smoothly within our categorical theory.

MDP homomorphisms in our theory can be viewed as morphisms between MDPs, which preserve composition and *forget* unrelated structure. In the concept of symmetry, we will rely on the construction of a *quotient* MDP, which we detail below. The symmetric structure of the RL problems we consider in this section alters the state-action space such that it is more economical to use the geometry of the problem. The construction removes the state-action pairs that can be represented as an inverse relationship. These symmetries show up in several RL problems and are common to mechanical systems, where we can think of symmetry as moving up compared to down or left compared to the right. In the larger context of designing RL systems we can think of the below operations as engineering within an *abstraction*—related to the "forget" operation. In some cases, we would instead need to add structure, which within our framework would be the same as applying some sort of *refinement* operation.

The goal of discussing symmetries is, therefore, twofold. First, we study symmetries to show the generality of our categorical formalism. Second, we study symmetries to show how design operations, as they occur in the engineering of RL systems, such as abstraction, reflect precisely within the categorical semantics of RL.

Denote by $Aut(\mathcal{M})$ the set of isomorphisms of an object of MDP; a group under composition.

A group action of a group G on an MDP $\mathcal{M}$ is a group homomorphism $\rho : G \to Aut(\mathcal{M})$. Concretely this means that for every group element $g \in G$ there is an isomorphism $\rho_g = (\alpha_g, \beta_g) : \mathcal{M} \to \mathcal{M}$ satisfying the composition identity $\rho_{gh} = \rho_g \circ \rho_h$. Intuitively a group action gives a set of symmetries for an MPD $\mathcal{M}$. We would like to perform RL keeping this mind and obtain policies that are invariant under the given domain symmetries. In order to do this efficiently, we need to have

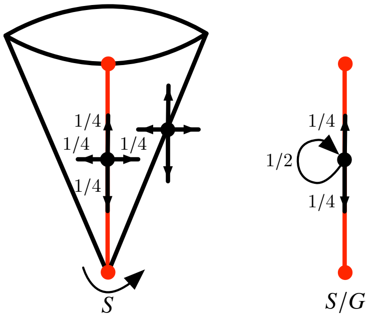

Figure 2: A conical state space collapses to a line in the quotient by axial rotation. The actions of going up and down stay the same, while left and right merge into one.

access to a quotient MDP $\mathcal{M}/G$.

For a group G, we do have a natural quotient MDP $\mathcal{M}/G$ constructed as follows (e.g., figure 2):

1. Let $\widehat{\mathcal{M}} := \mathcal{M} \times G$ be the MDP with state space $\widehat{S} = S \times G$ and action space $\widehat{A} = A \times G$. We have $\widehat{\psi} = \psi \times \mathrm{id}_G$ and the transition probabilities are given by the map

$$\widehat{T}(a, g) = (\mathrm{id}_S \times \iota_g)_\star T(a)$$

   where $\iota_g : S \to G$ denotes the constant map with value $g \in G$.

2. There are natural group action and projection morphisms $\rho, pr_1 : \widehat{\mathcal{M}} \to \mathcal{M}$ defined as:

   - For ρ, the map on state and action spaces is given by the group actions $\alpha := S \times G \to S$ and $\beta := A \times G \to A$. This gives a morphism of MDPs since for any $(a, g) \in A \times G$ mapping to $(s, g) \in S \times G$ under $\widehat{\psi}$ we have

$$\begin{aligned} \alpha_\star \widehat{T}(a, g) &= \alpha_\star (\mathrm{id}_S \times \iota_g)_\star T(a) \\ &= (\alpha_g)_\star T(a) \\ &= T(\beta_g(a)) \\ &= T(\beta(a, g)) \end{aligned}$$

     where we used that by definition $\alpha \circ (\mathrm{id}_S \times \iota_g) = \alpha_g$.

   - For pr_1, the same argument works for the first projection maps $S \times G \to S$ and $A \times G \to A$.

3. Define $\mathcal{M}/G$ as fitting in the pushout diagram

$$\begin{tikzcd} \mathcal{M} \times G \arrow[r, "pr_1"] \arrow[d, "\rho"'] & \mathcal{M} \arrow[d, "q"] \\ \mathcal{M} \arrow[r, "q"'] & \mathcal{M}/G. \end{tikzcd}$$

   Thus $\mathcal{M}/G := \mathcal{M} \cup_{\mathcal{M} \times G} \mathcal{M}$ and $q : \mathcal{M} \to \mathcal{M}/G$ is the canonical quotient morphism.

The following proposition confirms that the quotient $\mathcal{M}/G$ satisfies the desired universal property (vertical composition). Identical reasoning produces quotients in the category of MDPs with rewards within the enlarged category MDP_+.

Proposition 5. *$\mathcal{M}/G$ is the quotient of the MDP $\mathcal{M}$ by the action of G in the sense that for any MDP $\mathcal{N}$ and a G-invariant morphism $m : \mathcal{M} \to \mathcal{N}$, there is a unique factorization $\mathcal{M} \overset{q}{\to} \mathcal{M}/G \to \mathcal{N}$.*

5 A design for compositional task completion

Compositional RL problems are *sequential* in nature. An action follows another, given some rules of engagement.[1] Those rules might include how actions modify given the presence of another agent or how the actions of agents ought to intertwine in one sequential task description. We will study such compositional designs and derive a diagrammatic language for formulating them—in category theory, diagrams are explicit math operations and, therefore, constitute a formalism on their own. They also have a computational interpratation meaning that as long as we correctly define the type of operation, the diagrams themselves can abstract and eventually compute reasonable policies based on the instantiation

[1] The assumption that tasks are sequential in nature also relates to the concept of time. A categorical formulation of time can relax some of these assumptions such that we can model other types of problems that might not be sequential, but admit *some* compositional structure otherwise.

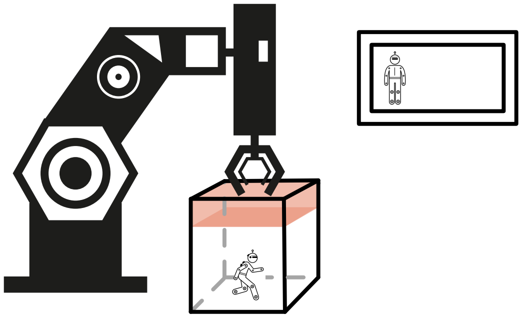

Figure 3: Fetch-and-place robot. The object in the box is moving, we denote that we a small humanoid but it could equally be represented by a simple sphere. Overlap in pink.

of the diagram. These diagrams are used, e.g., for manipulating quantum processes [Abramsky and Coecke, 2009, Coecke, 2010] and databases [Schultz et al., 2016]. The user of these diagrams does not have to be an expert in category theory because most of the categorical machinery sits in the background, providing helpful compositional checks for correctness. We see a fruitful direction for developing diagrammatic reasoning in the context of modeling RL problems via zig-zag diagrams.

Specifically, we detail how our categorical theory of RL works beyond grid world design problems (section 3).

Besides potentially simplifying learning for an agent, as we will see in Example 4, forming a composite MDP can also result in more efficient problem representations. We adapt a fetch-and-place robot problem (figure 3), where the learning algorithm controls actuation and rewards are assigned at completion of a task [Sutton and Barto, 2018, chapter 3].

Example 4 (Fetch-and-place robot)**.** *Suppose that a robotic arm wants to fetch a moving object from inside a box and then place it on a shelf outside the box.*

This is captured by the diagram of MDPs

$$\begin{tikzcd}
& \mathsf{Fetch} \arrow[dl] \arrow[dr] & & \mathsf{Place} \arrow[dl] \arrow[dr] & \\
\mathsf{Box} & & \mathsf{Move} & & \mathsf{Place}
\end{tikzcd}$$

We give a simple intuitive description of each:

Box *The state space is $B\times B$ where each factor records the position of the robot arm tip and the moving object and the action space is $A \times A$.*

Fetch *The object has been fetched and, thus, the position and actions of the arm and the object coincide. The state and action spaces are given by the diagonals in $B \times B$ and $A \times A$. More generally,* Fetch *is the maximal subprocess associated with the diagonal as a subset of $B \times B$. Observe that at this point we have made our state and action spaces smaller by recording data of half the dimension.*

Move *Since the arm needs to move the object outside of the box, we need to enlarge the state space. Thus,* $\mathsf{Move} = \mathsf{Fetch} \cup_{\mathsf{Overlap}} \mathsf{Outside}$ *where* Overlap *is a common region of the box and the outside environment. The actions are defined to allow the arm and object to move within the whole environment.*

Place *This is a full subprocess of* $\mathsf{Outside} \to \mathsf{Move}$. *If the ending position is a point in* Outside, *then we may take* $\mathsf{Place} = \mathsf{pt}$ *and a subprocess* $\mathsf{pt} \to \mathsf{Outside}$.

The composite MDP in this setup can be expressed as

$$\begin{aligned}
\mathcal{C} &= (\mathsf{Box} \cup_{\mathsf{Fetch}} \mathsf{Move}) \cup_{\mathsf{Place}} \mathsf{Place} \\
&= \mathsf{Box} \cup_{\mathsf{Fetch}} \mathsf{Move} \\
&= \mathsf{Box} \cup_{\mathsf{Fetch}} (\mathsf{Fetch} \cup_{\mathsf{Overlap}} \mathsf{Outside}).
\end{aligned}$$

If the object is stationary, we can make the diagram forward-moving by deleting the actions of Fetch *inside* Box *which separate the arm and the object. Then,* Fetch *is full in* Box, *but it is no longer full in composite $\mathcal{C}$, allowing for continuation of movement in order to complete the final task* Place *which remains full in $\mathcal{C}$. In that case, we can apply theorem 3 to compute the policies componentwise, meaning that we prove that learning-by-parts is the same as learning on the whole given some conditions for the composition of MDPs.*

6 Related work

Dealing with (de)composition of tasks requires to *operationalize* [Todorov, 2009] a given overall mission. This is often encoded in hierarchical MDPs [Parr and Russell, 1997, Dietterich, 2000, Nachum et al., 2018]. Theoretical problems with hierarchical MDPs center around the relaxation of *equivalence* relations between MDPs that may define "similar" tasks [Wen et al., 2020]. Our categorical framework gives different levels of *similarity* and different notions of isomorphic structures. Compositionality is often enforced using logical specifications [Jothimurugan et al., 2021, Araki et al., 2021, León et al., 2021, Vaezipoor et al., 2021], which adds more structure to RL. These logical specification can straightforwardly port into our theory via the representation of the product MDP, therefore no change is needed to incorporate logic into these semantics. In general, expressing MDPs as residing within a category gives us the flexibility to relate to different gadgets from other areas of mathematics, thereby giving us tools to address the intersection of formal verification and RL. Work in algebra [Ravindran and Barto, 2003, Perny et al., 2005, Feys et al., 2018, Tasse et al., 2020] develops abstractions over MDPs; our theory subsumes these algebraic structures within the notion of a category using, e.g., results from Patterson [2020]. We are not the first to capture MDPs as categories [Baez et al., 2016], but we significantly develop these structures and apply them to specific problems within RL. Compositionality can also be learned, rather than defined explicitely as we do with categorical semantics [Lake, 2019]. Combinations of parallel and sequential composition of tasks can be represented using the algebra of wiring diagrams [Schultz et al., 2016, Bakirtzis et al., 2021], which another diagrammatic syntax for computing with categories.

7 Conclusion

We develop a compositional theory for RL problems that involve MDPs based on the language of category semantics. Our theoretical results prove that, for the class of RL problems, the composition of MDPs is a fiber product and has a well-defined measure, there exists a gluing operation of MDPs, and that as long as MDP composites are forward-moving, then the learning-by-parts corresponds to learning the optimal policy on the whole. Practically, we show how our compositional framework deals with problems that involve more structure than just an MDP, i.e., group actions on an MDP, to model state-action symmetry (vertical composition) and introduce zig-zag diagrams to model compositional task completion RL problems (horizontal composition). These theoretical and practical mathematical tools have the potential to manage the complexity of RL in the context of autonomous system design, where a relationship between formalisms is an open problem [Luckcuck et al., 2019]. In addition, the generality of the derived properties can give meaning to robustness in the context of safe autonomous systems and potentially reduce the number of handwritten heuristic rules for this particular class of RL problems.

References

S. Abramsky and B. Coecke. Categorical quantum mechanics. In *Handbook of Quantum Logic and Quantum Structures*. Elsevier, 2009. doi: 10.1016/B978-0-444-52869-8.50010-4.

P. Aluffi. *Algebra: Chapter 0*. American Mathematical Society, 2021.

B. Araki, X. Li, K. Vodrahalli, J. A. DeCastro, M. J. Fry, and D. Rus. The logical options framework. In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, Proceedings of Machine Learning Research. PMLR, 2021. URL http://proceedings.mlr.press/v139/araki21a.html.

J. C. Baez, B. Fong, and B. S. Pollard. A compositional framework for Markov processes. *Journal of Mathematical Physics*, 2016. doi: 10.1063/1.4941578.

G. Bakirtzis, C. H. Fleming, and C. Vasilakopoulou. Categorical semantics of cyber-physical systems theory. *ACM Trans. Cyber Phys. Syst.*, 2021. doi: 10.1145/3461669.

P. Billingsley. *Probability and Measure*. Wiley, 1986.

A. Brandenburger and H. J. Keisler. Fiber products of measures and quantum foundation. *Logic and algebraic structures in quantum computing*, 2016.

B. Coecke. Quantum picturalism. *Contemporary physics*, 2010. doi: 10.1080/00107510903257624.

Bob Coecke. Compositionality as we see it, everywhere around us, 2021.

Thomas G. Dietterich. Hierarchical reinforcement learning with the MAXQ value function decomposition. *J. Artif. Intell. Res.*, 2000. doi: 10.1613/jair.639.

F. M. V. Feys, H. H. Hansen, and L. S. Moss. Long-term values in Markov decision processes, (co)algebraically. In *Proceedings of the 14th IFIP WG 1.3 International Workshop on Coalgebraic Methods in Computer Science (CMCS 2018)*, Lecture Notes in Computer Science. Springer, 2018. doi: 10.1007/978-3-030-00389-0_6.

F. Genovese. Modularity vs compositionality: A history of misunderstandings. URL https://perma.cc/B5SZ-G9JW, 2018. Statebox.

I. Gur, N. Jaques, Y. Miao, J. Choi, M. Tiwari, H. Lee, and A. Faust. Environment generation for zero-shot compositional reinforcement learning. In *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/file/218344619d8fb95d504ccfa11804073f-Paper.pdf.

K. Jothimurugan, S. Bansal, O. Bastani, and R. Alur. Compositional reinforcement learning from logical specifications. *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/file/531db99cb00833bcd414459069dc7387-Paper.pdf.

B. M. Lake. Compositional generalization through meta sequence-to-sequence learning. In *Proceedings of the 2019 Annual Conference on Neural Information Processing Systems on 32nd Advances in Neural Information Processing Systems (NeurIPS 2019)*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/f4d0e2e7fc057a58f7ca4a391f01940a-Abstract.html.

F. W. Lawvere and S. H. Schanuel. *Conceptual mathematics: a first introduction to categories*. Cambridge University Press, 2009.

J. Leike, M. Martic, V. Krakovna, P. A. Ortega, T. Everitt, A. Lefrancq, L. Orseau, and S. Legg. AI safety gridworlds. arXiv:1711.09883 [cs.LG], 2017.

T. Leinster. *Basic Category Theory*. Cambridge University Press, 2014.

B. G. León, M. Shanahan, and F. Belardinelli. In a nutshell, the human asked for this: Latent goals for following temporal specifications. *arXiv:2110.09461 [cs.AI]*, 2021. doi: 10.48550/ARXIV.2110.09461.

Yunfei Li, Yilin Wu, Huazhe Xu, Xiaolong Wang, and Yi Wu. Solving compositional reinforcement learning problems via task reduction. In *Proceedings of the 9th International Conference on Learning Representations (ICLR 2021)*, 2021. URL https://openreview.net/forum?id=9SS69KwomAM.

M. Luckcuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher. Formal specification and verification of autonomous robotic systems: A survey. *ACM Comput. Surv.*, 2019. doi: 10.1145/3342355.

S. Mac Lane. *Categories for the working mathematician*. Springer, 1998.

O. Nachum, S. S. Gu, H. Lee, and S. Levine. Data-efficient hierarchical reinforcement learning. *Advances in neural information processing systems*, 2018. URL https://proceedings.neurips.cc/paper/2018/file/e6384711491713d29bc63fc5eeb5ba4f-Paper.pdf.

Cyrus Neary, Zhe Xu, Bo Wu, and Ufuk Topcu. Reward machines for cooperative multi-agent reinforcement learning. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '21)*. ACM, 2021. doi: 10.5555/3463952.3464063.

R. Parr and S. Russell. Reinforcement learning with hierarchies of machines. *Advances in neural information processing systems*, 1997. URL https://proceedings.neurips.cc/paper/1997/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Evan Patterson. *The algebra and machine representation of statistical models*. PhD thesis, Stanford University, 2020.

P. Perny, O. Spanjaard, and P. Weng. Algebraic Markov decision processes. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI 2005)*, 2005. URL http://ijcai.org/Proceedings/05/Papers/1677.pdf.

B. Ravindran and A. G. Barto. SMDP homomorphisms: An algebraic approach to abstraction in semi-Markov decision processes. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI 2003)*. Morgan Kaufmann, 2003. URL http://ijcai.org/Proceedings/03/Papers/145.pdf.

P. Schultz, D. I. Spivak, C. Vasilakopoulou, and R. Wisnesky. Algebraic databases. *Theory & Applications of Categories*, 2016.

David I Spivak. *Category theory for the sciences*. MIT Press, 2014.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

JM Swart. A conditional product measure theorem. *Statistics & probability letters*, 1996. doi: 10.1016/0167-7152(95)00107-7.

Z. Szabó. The case for compositionality. *The Oxford Handbook of Compositionality*, 2012.

Z. G. Szabó. Compositionality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2022 edition, 2022.

G. N. Tasse, S. D. James, and B. Rosman. A boolean task algebra for reinforcement learning. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems on Advances in Neural Information Processing Systems (NeurIPS 2020)*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6ba3af5d7b2790e73f0de32e5c8c1798-Abstract.html.

Emanuel Todorov. Compositionality of optimal control laws. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems on Advances in Neural Information Processing Systems (NeurIPS 2009)*. Curran Associates, Inc., 2009. URL https://proceedings.neurips.cc/paper/2009/hash/3eb71f6293a2a31f3569e10af6552658-Abstract.html.

P. Vaezipoor, A. C. Li, R. T. Icarte, and S. A. McIlraith. LTL2Action: Generalizing LTL instructions for multi-task RL. In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, Proceedings of Machine Learning Research. PMLR, 2021. URL http://proceedings.mlr.press/v139/vaezipoor21a.html.

E. van der Pol, D. E. Worrall, H. van Hoof, F. A. Oliehoek, and M. Welling. MDP homomorphic networks: Group symmetries in reinforcement learning. In *Advances in Neural Information Processing Systems*, 2020. URL https://proceedings.neurips.cc/paper/2020/file/2be5f9c2e3620eb73c2972d7552b6cb5-Paper.pdf.

Z. Wen, D. Precup, M. Ibrahimi, A. Barreto, B. Van Roy, and S. Singh. On efficiency in hierarchical reinforcement learning. In *Advances in Neural Information Processing Systems*, 2020. URL https://proceedings.neurips.cc/paper/2020/file/4a5cfa9281924139db466a8a19291aff-Paper.pdf.

A Fundamental definitions

Here we give brief definitions on some of the mathematical structures we use. Consult Lawvere and Schanuel [2009], Leinster [2014], or Mac Lane [1998] for an in-depth treatment of category theory.

Definition 5 (Category)**.** *A* category C *consists of a collection of set of objects* Ob *and for any two* $X, Y \in \mathrm{Ob}$ *a set of arrows* $f : X \to Y$*, along with a composition rule*

$$(f : X \to Y,\ g : Y \to Z) \longmapsto g \circ f : X \to Z$$

and an identity arrow $\mathrm{id}_X : X \to X$ *for all objects, subject to associativity and unity conditions:* $(f \circ g) \circ h = f \circ (g \circ h)$ *and* $f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f$*.*

This definition encompasses a vast variety of structures in mathematics and other sciences: to name a couple, Set is the category of sets and functions, whereas Lin is the category of k-linear real vector spaces and k-linear maps between them.

Definition 6 (Commutative diagrams)**.** *A standard diagrammatic way to express composites is* $X \xrightarrow{f} Y \xrightarrow{g} Z$ *and equations via commutative diagrams of the following form*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "h"'] & X \arrow[d, "g"] \\ & Y \end{tikzcd} \qquad \textit{which stands for } g \circ f = h.$$

A commutative square is instead,

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "g"'] & Z \arrow[d, "g'"] \\ Y \arrow[r, "f'"] & W \end{tikzcd} \qquad \textit{which stands for } g' \circ f = f' \circ g.$$

Definition 7 (Isomorphism)**.** *A morphism* $f : X \to Y$ *is called* invertible *or an* isomorphism *when there exists another* $g : Y \to X$ *such that* $f \circ g = \mathrm{id}_Y$ *and* $g \circ f = \mathrm{id}_X$*.*

Definition 8 (Functor)**.** *A* functor $F : \mathsf{C} \to \mathsf{D}$ *between two categories consists of a function between objects and a function between morphisms, where we denote* $Ff : FX \to FY$*, such that it preserves composition and identities:* $F(f \circ g) = Ff \circ Fg$ *and* $F(\mathrm{id}_X) = \mathrm{id}_{FX}$*.*

A functor can informally be thought of as a structure-preserving map between domains of discourse. Interestingly, categories and functors form a category on their own, denoted Cat, in the sense that functors compose and the rest of the axioms hold.

Definition 9 (Cartesian category)**.** *A category is called cartesian closed if it has all finite products and exponentials.*

The category of sets Set is cartesian closed where the product is the common product

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

We can think of this operation as, e.g., containing data on a table and the projections are then giving us the particular column and row respectively.

Fiber products are a generalization of this notion.

Definition 10 (Fiber product). *A category is said to have a fiber product for morphisms $f : A \to C$ and $g : B \to C$ when there exists an object W together with*

morphisms $a : W \to A$ and $b : W \to B$ such that the square

$$\begin{tikzcd}
W \arrow[r, "b"] \arrow[d, "a"'] & B \arrow[d, "g"] \\
A \arrow[r, "f"] & C
\end{tikzcd}$$

commutes, where the morphisms a, b are thought of as projections such that $f \circ a = g \circ b$, and which is universal in the following way: for any object W_o with morphisms $a_o : W_o \to A$ and $b_o : W_o \to B$ such that $f \circ a_o = g \circ b_o$, there exist a unique morphism $w : W_o \to W$ such that $a \circ w = a_o$ and $b \circ w = b_o$.

W is then the limit of the diagram

$$\begin{tikzcd}
 & B \arrow[d, "g"] \\
A \arrow[r, "f"] & C
\end{tikzcd}$$

and we write $W = A \times_C B$ for the fiber product. We also say that W is the pullback of A along the map $g : B \to C$ and also the pullback of B along the map $f : A \to C$.

Definition 11 (Pushout). *A pushout for morphisms $f : C \to A$ and $g : C \to B$ is an object W together with morphisms $a : A \to W$ and $b : B \to W$ such that the square*

$$\begin{tikzcd}
C \arrow[r, "g"] \arrow[d, "f"'] & B \arrow[d, "b"] \\
A \arrow[r, "a"] & W
\end{tikzcd}$$

commutes, where the morphisms a, b are thought of as inclusions such that $a \circ f = b \circ g$, and which is universal in the following way: for any object W_o with morphisms $a_o : A \to W_o$ and $b_o : B \to W_o$ such that $a_o \circ f = b_o \circ g$, there exist a unique morphism $w : W \to W_o$ such that $w \circ a = a_o$ and $w \circ b = b_o$.

W is then the colimit of the diagram

$$\begin{tikzcd}
C \arrow[r, "g"] \arrow[d, "f"'] & B \\
A &
\end{tikzcd}$$

and we write $W = A \cup_C B$ for the pushout.

We see that a pushout is a contravariant version of a fiber product. The pushout in the category Set of two morphisms $\emptyset \to A$ and $\emptyset \to B$ is the standard disjoint union $A \coprod B$, labelling the composed set with which elements come from set A and which elements come from set B. This is also the coproduct in the category of sets. For another example, when we have $A \cap B \subseteq A$ and $A \cap B \subseteq B$ and $a : A \cap B \to A$ and $b : A \cap B \to B$, the union $A \cup B$ is naturally isomorphic with the pushout of a and b.

For a treatment of measure theory consult Billingsley [1986].

Definition 12 (Pushforward measure). *Given measurable spaces (Ω, Σ) and (X, T), a probability measure μ on (Ω, Σ) and a measurable function $\psi : (\Omega, \Sigma) \to (X, T)$, we write $\psi_\star \mu$ for the pushforward measure obtained from μ by applying ψ:*

$$\psi_\star \mu(A) := \mu\left(\psi^{-1}(A)\right) \text{ for all } A \in T.$$

For, e.g., a deterministic function with random inputs, the pushforward measure gives us an explicit description of the possible distribution of outcomes.

For an in-depth treatment of algebra consult Aluffi [2021].

Definition 13 (Group). *A set G endowed with a binary operation $\bullet$, $(G, \bullet)$ is a group if the following conditions hold.*

1. *The operation $\bullet$ is associative; that is,* $(\text{for all } g, h, k \in G)\colon\ (g \bullet h) \bullet k = g \bullet (h \bullet k)$.
2. *There exists an identity element e_G for $\bullet$; that is,*
$$(\text{there exists } e_G \in G)(\text{for all } g \in G)\colon\ g \bullet e_G = e_G \bullet g.$$
3. *Every element in G is invertible with respect to $\bullet$; that is,*
$$(\text{for all } g \in G)(\text{there exists } h \in G)\colon\ g \bullet h = h \bullet g = e_G.$$

Consider as an example the set of integers $\mathbb{Z}$:

1. Addition in $\mathbb{Z}$ is an associative operation.
2. The identity element is the integer 0.
3. The inverse map sends an integer $n \in \mathbb{Z}$ to another integer $-n$.

B Constructions

Fiber products

Suppose that we have three MDPs $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ together with two morphisms $m_1 = (f_1, g_1)\colon \mathcal{M}_1 \to \mathcal{M}_3$ and $m_2 = (f_2, g_2)\colon \mathcal{M}_2 \to \mathcal{M}_3$.

We would like to define a fiber product $\mathcal{M} = \mathcal{M}_1 \times_{\mathcal{M}_3} \mathcal{M}_2 = (S, A, \psi, T)$ in the category MDP. While this is generally not possible, as one would then be able to deduce a construction of fiber products for the category of measurable spaces, which is known to not exist, we will succeed allowing for a weakening of the universal properties we consider.

For the state and action spaces, we set $S = S_1 \times_{S_3} S_2,\ A = A_1 \times_{A_3} A_2$. By standard properties of fiber products (of sets), the maps $\psi_i\colon A_i \to S_i$ for $i = 1, 2, 3$ induce a canonical morphism $\psi\colon A \to S$. Write $pr_i\colon S \to S_i$ and $\rho_i\colon A \to A_i$ for the projection maps, where $i = 1, 2$.

Since we need S to be a measurable space, we endow it with the σ-algebra generated by all subsets
$$pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2),$$
where U_1, U_2 are any two measurable subsets of S_1 and S_2 respectively. The projections pr_i are measurable functions.

Remark 2. This σ-algebra on S is potentially rather small. However, in the case where S_1 and S_2 are finite measure spaces, whose σ-algebras are their power sets, then the σ-algebra is the power set of S.

Let $a \in A$ with projections $a_1 \in A_1$ and $a_2 \in A_2$ mapping in turn to an action $a_3 \in A_3$. For brevity, write $\mu_i = T_i(a_i) \in \mathcal{P}_{S_i}$ for $i = 1, 2$ and $\mu_3 = (f_i)_\star \mu_i \in \mathcal{P}_{S_3}$.

Our goal is to construct $\mu = T(a) \in \mathcal{P}_S$ and obtain for each index $i = 1, 2$ a commutative diagram

$$\begin{tikzcd}
A \arrow[r, "\rho_i"] \arrow[d, "T"'] & A_i \arrow[d, "T_i"] \\
\mathcal{P}_S \arrow[r, "(pr_i)_\star"'] & \mathcal{P}_{S_i}.
\end{tikzcd}$$

Having that, there is an MDP $\mathcal{M}$ fitting in the commutative diagram

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "(pr_1{,}\rho_1)"] \arrow[d, "(pr_2{,}\rho_2)"'] & \mathcal{M}_1 \arrow[d, "m_1"] \\
\mathcal{M}_2 \arrow[r, "m_2"'] & \mathcal{M}_3,
\end{tikzcd} \tag{6}$$

which we would like to be universal among commutative diagrams of the form

$$\begin{tikzcd}
\mathcal{N} \arrow[r, "(\alpha_1{,}\beta_1)"] \arrow[d, "(\alpha_2{,}\beta_2)"'] & \mathcal{M}_1 \arrow[d, "m_1"] \\
\mathcal{M}_2 \arrow[r, "m_2"'] & \mathcal{M}_3
\end{tikzcd} \tag{7}$$

with certain properties, in the sense that any commutative diagram (7) should be induced by a canonical morphism $\mathcal{N} \to \mathcal{M}$ and composing with the projection maps $\mathcal{M} \to \mathcal{M}_1$ and $\mathcal{M} \to \mathcal{M}_2$.

The subprocess case Suppose that one of the two $\mathcal{M}_1 \to \mathcal{M}_3$ and $\mathcal{M}_2 \to \mathcal{M}_3$ is a subprocess. Without loss of generality, we assume that $\mathcal{M}_2 \to \mathcal{M}_3$ is a subprocess.

It follows that the maps $pr_1 : S \to S_1$ and $\rho_1 : A \to A_1$ are injective.

Moreover, since $(f_1)_\star \mu_1 = (f_2)_\star \mu_2 = \mu_3$, it follows that μ_1 is supported on S. We define $T(a) = \mu_1$ considered as a measure on S.

Proposition 6. *The diagram* (6) *is universal among diagrams of the form* (7) *for which the morphism $\mathcal{N} \to \mathcal{M}_1$ defines a subprocess.*

Proof. Suppose that $(\alpha_1, \beta_1) : \mathcal{N} \to \mathcal{M}_1$ defines a subprocess. We have induced morphisms $\alpha = (\alpha_1, \alpha_2) : S_\mathcal{N} \to S = S_1 \times_{S_3} S_2$ and $\beta = (\beta_1, \beta_2) : A_\mathcal{N} \to A = A_1 \times_{A_3} A_2$.

Observe that both maps are injective, since the compositions $pr_1 \circ \alpha = \alpha_1$ and $\rho_1 \circ \beta = \beta_1$ are injective.

For any $a \in A_\mathcal{N}$ we need to check that $\alpha_\star T_\mathcal{N}(a) = T(\beta(a))$.

But $T_1(\beta_1(a))$ is supported on $\mathcal{N}$ and we have $T_1(\beta_1(a)) = T(\beta(a))$.

On the other hand, $(pr_1)_\star \alpha_\star T_\mathcal{N}(a) = (pr_1 \circ \alpha)_\star T_\mathcal{N}(a) = (\alpha_1)_\star T_\mathcal{N}(a) = T_1(\beta_1(a))$, as wanted. □

The finite case Assume that the spaces S_1, S_2, S_3 are discrete or finite and their σ-algebras are their power sets.

Consider the function $\nu : S_1 \times_{S_3} S_2 \to \mathbb{R}$ defined by the formula

$$\nu(s_1, s_2) = \begin{cases} \frac{\mu_1(s_1)\mu_2(s_2)}{\mu_3(f_1(s_1))} = \frac{\mu_1(s_1)\mu_2(s_2)}{\mu_3(f_2(s_2))} & \text{if } \mu_3(f_1(s_1)) = \mu_3(f_2(s_2)) > 0 \\ \\ 0 & \text{if } \mu_3(f_1(s_1)) = 0. \end{cases} \tag{8}$$

We define the measure $\mu = T(a) \in \mathcal{P}_S$ to be the one with probability density function ν. Thus,

$$\mu\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) := \sum_{s \in pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)} \nu(s). \tag{9}$$

This satisfies the desired properties to define MDP morphisms $\mathcal{M} \to \mathcal{M}_i$, $i = 1, 2$ (proposition 8).

Consider now two morphisms $(\alpha_i, \beta_i) : \mathcal{N} \to \mathcal{M}_i$, $i = 1, 2$, in MDP. We say that they form a *conditionally independent pair* if for any $a \in A_{\mathcal{N}}$ with images $a_i = \beta_i(a) \in A_i$ and any measurable subsets $U_i \subseteq S_i$, we have

$$T_{\mathcal{N}}(a)\left(\alpha_1^{-1}(U_1) \cap \alpha_2^{-1}(U_2)\right) = \sum_{(s_1, s_2) \in pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)} \nu(s_1, s_2). \tag{10}$$

This assignment determines the measure $T_{\mathcal{N}}$ only on the σ-algebra generated by sets $\alpha_1^{-1}(U_1) \cap \alpha_2^{-1}(U_2)$.

Proposition 7. *The morphisms $\mathcal{M} \to \mathcal{M}_1$ and $\mathcal{M} \to \mathcal{M}_2$ in diagram* (6) *form a conditionally independent pair. Moreover, diagram* (6) *is universal among diagrams* (7) *where the morphisms $\mathcal{N} \to \mathcal{M}_i$ form a conditionally independent pair.*

Proof. The independence of $\mathcal{M} \to \mathcal{M}_1$ and $\mathcal{M} \to \mathcal{M}_2$ follows immediately by observing that equations (11) and (10) coincide.

For the universality statement, suppose as above that we have two independent morphisms $(\alpha_i, \beta_i) : \mathcal{N} \to \mathcal{M}_i$, $i = 1, 2$, in MDP, that fit in a commutative square (7).

Since $\mathcal{M}$ has state and action spaces given by $S = S_1 \times_{S_3} S_2$ and $A = A_1 \times_{A_3} A_2$ respectively and the diagram gives that $f_1 \circ \alpha_1 = f_2 \circ \alpha_2$ and $g_1 \circ \beta_1 = g_2 \circ \beta_2$, we get canonical morphisms $\gamma : S_{\mathcal{N}} \to S$ and $\delta : A_{\mathcal{N}} \to A$. These satisfy

$$pr_i \circ \gamma = \alpha_i, \quad \rho_i \circ \delta = \beta_i, \quad i = 1, 2,$$

and moreover by construction $\gamma \circ \psi_{\mathcal{N}} = \psi \circ \delta$.

To obtain a morphism $\mathcal{N} \to \mathcal{M}$, it remains to check that diagram (3) commutes. For $a \in A_{\mathcal{N}}$, write $\beta_i(a) = \rho_i(\delta(a)) = a_i \in A_i$. We then have, using formula (10) for the third equality and formula (9) for the last equality,

$$\begin{aligned} \gamma_\star T_{\mathcal{N}}(a)\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) &= T_{\mathcal{N}}(a)\left(\gamma^{-1}(pr_1^{-1}(U_1)) \cap \gamma^{-1}(pr_2^{-1}(U_2))\right) \\ &= T_{\mathcal{N}}(a)\left(\alpha_1^{-1}(U_1) \cap \alpha_2^{-1}(U_2)\right) \\ &= \sum_{(s_1, s_2) \in pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)} \nu(s_1, s_2) \\ &= T(\delta(a))\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) \end{aligned}$$

which implies that $\gamma_\star \circ T_{\mathcal{N}} = T \circ \delta$, completing the proof. □

The case of S_3 finite Suppose that S_3 is finite or discrete with weights given by a function ν_3.

Suppose in addition that S_1, S_2 are measurable subsets of ambient measure spaces $(K_1, \tau_1), (K_2, \tau_2)$ with density functions ν_1, ν_2 so that for any measurable subset $U_i \subseteq S_i$ we have $\mu_i(U_i) = \int_{U_i} \nu_i \, \mathrm{d}\tau_i$. This is not restrictive and we allow for it for purposes of exhibition and consistency with the preceding case. One can always take ν_i to be identically 1 and then $\mu_i = \tau_i$.

As above, we consider the function

$$\nu : S_1 \times_{S_3} S_2 \to \mathbb{R},$$
$$\nu(s_1, s_2) = \begin{cases} \frac{\nu_1(s_1)\nu_2(s_2)}{\nu_3(f_1(s_1))} = \frac{\nu_1(s_1)\nu_2(s_2)}{\nu_3(f_2(s_2))} & \text{if } \nu_3(f_1(s_1)) = \nu_3(f_2(s_2)) > 0 \\ \\ 0 & \text{if } \nu_3(f_1(s_1)) = 0, \end{cases}$$

and define the measure $\mu = T(a) \in \mathcal{P}_S$ to be the one with probability density ν with respect to the product measure $\tau_1 \otimes \tau_2$ on $K_1 \times K_2$ so that

$$\mu\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) := \int_{pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)} \nu \, \mathrm{d}(\tau_1 \otimes \tau_2). \tag{11}$$

Remark 3. The case of S_3 finite encompasses the finite case above, since for S_1, S_2 we may take the ambient space K to be $S_1 \coprod S_2$ with measure given by set cardinality. The reason we treat the finite case separately is that we can obtain a clean and simpler to write down universality statement.

Proposition 8. *μ is a probability measure and we have $(pr_i)_\star \mu = \mu_i$ for $i = 1, 2$.*

Proof. We first observe that for any $s_3 \in S_3$

$$\begin{aligned} \nu_3(s_3) &= \mu_3(s_3) \\ &= (f_2)_\star \mu_2(s_3) \\ &= \mu_2(f_2^{-1}(s_3)) \\ &= \int_{f_2^{-1}(s_3)} \nu_2(s_2) \, \mathrm{d}\tau_2. \end{aligned}$$

We then have for any measurable subset $U_1 \subseteq S_1$

$$\begin{aligned} (pr_1)_\star \mu(U_1) = \mu(pr_1^{-1}(U_1)) &= \mu(U_1 \times_{S_3} S_2) \\ &= \int_{U_1 \times_{S_3} S_2} \nu \, \mathrm{d}(\tau_1 \otimes \tau_2) \\ &= \int_{U_1 \times S_2} \mathbb{1}_{U_1 \times_{S_3} S_2} \cdot \nu \, \mathrm{d}(\tau_1 \otimes \tau_2) \end{aligned}$$

so, applying Fubini-Tonelli's theorem, we obtain

$$
\begin{aligned}
\mu(U_1 \times_{S_3} S_2) &= \int_{U_1} \left(\int_{S_2} \mathbb{1}_{U_1 \times_{S_3} S_2} \cdot \nu \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \nu_1(s_1) \left(\int_{f_2^{-1}(f_1(s_1))} \frac{\nu_2(s_2)}{\nu_3(f_2(s_2))} \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \nu_1(s_1) \left(\int_{f_2^{-1}(f_1(s_1))} \frac{\nu_2(s_2)}{\nu_3(f_1(s_1))} \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \frac{\nu_1(s_1)}{\nu_3(f_1(s_1))} \left(\int_{f_2^{-1}(f_1(s_1))} \nu_2(s_2) \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \frac{\nu_1(s_1)}{\nu_3(f_1(s_1))} \nu_3(f_1(s_1)) \, \mathrm{d}\tau_1 \\
&= \int_{U_1} \nu_1(s_1) \, \mathrm{d}\tau_1 = \mu_1(U_1),
\end{aligned}
$$

which is what we want. □

Remark 4. When $\mathcal{M}_3$ is the terminal object pt and we take $S_1 = K_1, S_2 = K_2$ and ν_1, ν_2 identically 1, we obtain the cartesian product $\mathcal{M}_1 \times \mathcal{M}_2$ of two MDPs $\mathcal{M}_1$ and $\mathcal{M}_2$.

The local fibration case We now treat the most general case. We first introduce some terminology.

Definition 14. *A measurable function between two measurable spaces $f : X \to Y$ is a local fibration if there exists a measurable partition $Y = Y_1 \coprod \cdots \coprod Y_N$ such that for every index i, we have that $X_i \simeq Y_i \times F_i$ for some measure space F_i and $f|_{X_i}$ is projection onto Y_i.*

A morphism $(f, g) : \mathcal{M}_1 \to \mathcal{M}_2$ between MDPs is called a local fibration if the map $f : S_1 \to S_2$ is a local fibration.

Theorem 4. *Assume that the state spaces S_1, S_2 and S_3 are Polish spaces and the morphisms $\mathcal{M}_1 \to \mathcal{M}_3$ and $\mathcal{M}_2 \to \mathcal{M}_3$ are local fibrations. Then there is a well-defined measure $\mu = T(a)$, functorial in the morphisms $\mathcal{M}_1 \to \mathcal{M}_3$ and $\mathcal{M}_2 \to \mathcal{M}_3$, giving rise to a fiber product MDP $\mathcal{M} = \mathcal{M}_1 \times_{\mathcal{M}_3} \mathcal{M}_2$.*

Proof. By refining partitions, we can assume that $S_3 = Z_1 \coprod \cdots \coprod Z_N$ is a common partition for the fibration structure of the maps f_i with the measures determined by $a \in A$, so that $S_1 = Z_1 \times X_1 \coprod \cdots \coprod X_N \times F_N$ and $S_2 = Z_1 \times Y_1 \coprod \cdots \coprod Z_N \times Y_N$.

As sets, we then have

$$S_1 \times_{S_3} S_2 = (X_1 \times Y_1 \times Z_1) \coprod \cdots \coprod (X_N \times Y_N \times Z_N)$$

and thus we may reduce to the case $S_3 = Z$ and $S_1 = Z \times X$ and $S_2 = Z \times Y$.

The functoriality statement follows from the existence and functoriality of μ with respect to local fibrations $\mathcal{M}_i \to \mathcal{M}_3$ [Swart, 1996, Brandenburger and Keisler, 2016]. □

We now explain how the previous three situations are applications of theorem 4:

1. For the subprocess case, observe that any injection $f_i : S_i \to S_3$ is a local fibration by taking $S_i \simeq S_i \times \{\bullet\} \coprod (S_3 \setminus S_i) \times \emptyset$.
2. For the finite case, any morphism $f : X \to Z$ between finite spaces is a local fibration, because
$$X = \coprod_{z \in Z} f^{-1}(z) \simeq \coprod_{z \in Z} f^{-1}(z) \times \{z\}$$
and $f^{-1}(z) \times \{z\} \longmapsto z \in Z$ is a fibration with fiber $f^{-1}(z)$.
3. When S_3 is finite, any measurable function $f_i : S_i \to S_3$ is a local fibration
$$S_i = \coprod_{s_3 \in S_3} f_i^{-1}(s_3).$$

Gluing

Suppose that we have three MDPs $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ together with two morphisms $m_1 = (f_1, g_1) : \mathcal{M}_3 \to \mathcal{M}_1$ and $m_2 = (f_2, g_2) : \mathcal{M}_3 \to \mathcal{M}_2$. We wish to glue $\mathcal{M}_1$ and $\mathcal{M}_2$ along their overlap coming from $\mathcal{M}_3$ to obtain a new MDP, denoted by $\mathcal{M} := \mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$, such that there exist natural maps $\mathcal{M}_1 \to \mathcal{M}$ and $\mathcal{M}_2 \to \mathcal{M}$, giving a pushout diagram

$$\begin{tikzcd}
\mathcal{M}_3 \arrow[r, "m_1"] \arrow[d, "m_2"'] & \mathcal{M}_1 \arrow[d] \\
\mathcal{M}_2 \arrow[r] & \mathcal{M}.
\end{tikzcd}$$

We propose the following construction (figure 4). To define the state space S of $\mathcal{M}$, we take

$$S = S_1 \cup_{S_3} S_2 = S_1 \coprod S_2 / \sim \tag{12}$$

where the equivalence relation $\sim$ is generated by identifying $f_1(s_3) \in S_1$ with $f_2(s_3) \in S_2$ for all $s_3 \in S_3$. This gives a pushout diagram in the category of sets:

$$\begin{tikzcd}
S_3 \arrow[r, "f_1"] \arrow[d, "f_2"'] & S_1 \arrow[d, "i_1"] \\
S_2 \arrow[r, "i_2"'] & S.
\end{tikzcd} \tag{13}$$

Observe that the state space is a disjoint union of three components

$$S = (S_1 \setminus f_1(S_3)) \coprod (f_1(S_3) \sim f_2(S_3)) \coprod (S_2 \setminus f_2(S_3)). \tag{14}$$

Set $S_1^\circ = S_1 \setminus f_1(S_3)$, $S_2^\circ = S_2 \setminus f_2(S_3)$. By abuse of notation we write S_3 to denote the middle component, when this is clear from context.

To specify the action space A and the projection map $\psi : A \to S$, by formula (1), it suffices to define the action spaces A_s for each state $s \in S$. We consider each component separately:

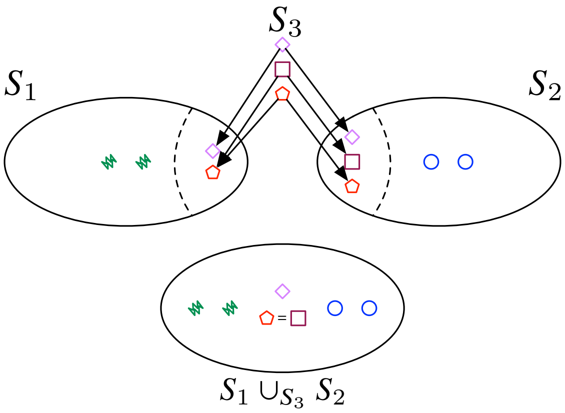

Figure 4: Gluing state spaces S_i. Gluing works similarly for state-action spaces A_i.

1. Over each S_i°, for $s \in S_i^\circ$, since S_i° is naturally a subset of S_i, we define $A_s := (A_i)_s$.

2. Over the overlap S_3, for each state $s = f_1(s_3) = f_2(s_3) \in S$, where $s_3 \in S_3$, we let

$$A_s = (A_1)_s \coprod (A_2)_s / \sim \tag{15}$$

   where $\sim$ is generated by identifying $g_1(a_3)$ with $g_2(a_3)$ for all actions $a_3 \in A_3$.

As for state spaces, we have defined A as the following pushout in the category of sets:

$$\begin{tikzcd} A_3 \arrow[r, "g_1"] \arrow[d, "g_2"'] & A_1 \arrow[d, "j_1"] \\ A_2 \arrow[r, "j_2"'] & A. \end{tikzcd} \tag{16}$$

By the universal property of pushouts, diagrams (14) and (16) imply that we have an induced canonical morphism $\psi : A \to S$.

Finally, we need to specify the information of the transition probabilities; that is, the morphism $T : A \to \mathcal{P}_S$. Again, we proceed componentwise specifying $T_s : A_s \to \mathcal{P}_S$ for s in S_i° and S_3:

1. Over each S_i°, we define $T_s = (T_i)_s$. This makes sense since $A_s = (A_i)_s$ in this case.

2. Over S_3, according to formula (15), the action space consists of components $(A_i)_s$.

   For $a \in (A_i)_s$ which does not lie in the image of g_i, define $T(a)$ to equal $(i_i)_\star T_i(a)$.

   Otherwise, suppose that $a_1 = g_1(a_3) \in A_s$ for some action $a_3 \in A_3$. Write $a_2 = g_2(a_3)$. We observe that

$$(i_1)_\star T_1(a_1) = (i_2)_\star T_2(a_2). \tag{17}$$

This follows from the equality $i_1 \circ f_1 = i_2 \circ f_2$ and diagram (3), since

$$\begin{aligned}
(i_1)_\star T_1(a_1) &= (i_1)_\star T_1(g_1(a_3)) \\
&= (i_1)_\star (f_1)_\star T_3(a_3) \\
&= (i_1 \circ f_1)_\star T_3(a_3), \\
(i_2)_\star T_2(a_2) &= (i_2)_\star T_2(g_2(a_3)) \\
&= (i_2)_\star (f_2)_\star T_3(a_3) \\
&= (i_2 \circ f_2)_\star T_3(a_3).
\end{aligned}$$

We may, thus, define unambiguously

$$T(a) = (i_1)_\star T_1(a_1) = (i_2)_\star T_2(a_2). \tag{18}$$

This expression is independent of the choice of index i, so it respects the equivalence relation $\sim$ on A. If the action is in the image of g_2, we argue in an identical way.

We now verify that this construction indeed gives a pushout.

Theorem 2. *There exists an MDP $\mathcal{M} = \mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ which is the pushout of the diagram in* MDP :

$$\begin{tikzcd}
\mathcal{M}_3 \arrow[r, "m_1"] \arrow[d, "m_2"'] & \mathcal{M}_1 \\
\mathcal{M}_2 &
\end{tikzcd}$$

Proof. By definition, we have commutative diagrams

$$\begin{tikzcd}
A_1 \arrow[r, "j_1"] \arrow[d, "T_1"'] & A \arrow[d, "T"] & A_2 \arrow[r, "j_2"] \arrow[d, "T_2"'] & A \arrow[d, "T"] \\
\mathcal{P}_{S_1} \arrow[r, "(i_1)_\star"'] & \mathcal{P}_S, & \mathcal{P}_{S_2} \arrow[r, "(i_2)_\star"'] & \mathcal{P}_S.
\end{tikzcd} \tag{19}$$

Therefore we obtain natural morphisms $\rho_1 = (i_1, j_1) \colon \mathcal{M}_1 \to \mathcal{M}$ and $\rho_2 = (i_2, j_2) \colon \mathcal{M}_2 \to \mathcal{M}$ fitting in a commutative diagram

$$\begin{tikzcd}
\mathcal{M}_3 \arrow[r, "m_1"] \arrow[d, "m_2"'] & \mathcal{M}_1 \arrow[d, "\rho_1"] \\
\mathcal{M}_2 \arrow[r, "\rho_2"'] & \mathcal{M}.
\end{tikzcd}$$

Now let $\mathcal{N}$ be a MDP fitting in a commutative diagram

$$\begin{tikzcd}
\mathcal{M}_3 \arrow[r, "m_1"] \arrow[d, "m_2"'] & \mathcal{M}_1 \arrow[d, "(\alpha_1{,}\beta_1)"] \\
\mathcal{M}_2 \arrow[r, "(\alpha_2{,}\beta_2)"'] & \mathcal{N}.
\end{tikzcd}$$

We need to check that the diagram is induced by a canonical morphism $m = (f, g): \mathcal{M} \to \mathcal{N}$. By the definition of the state and action spaces of $\mathcal{M}$, they are the pushouts (in the category of sets) of the corresponding spaces of $\mathcal{M}_i$ along those of $\mathcal{M}_3$ so there are natural candidates for f and g. These fit in a commutative diagram (2), so it remains to verify that diagram (3) is commutative.

To avoid confusion, we now use the subscripts $\mathcal{M}$ and $\mathcal{N}$ to indicate the MDP to which state and action spaces correspond below.

Since the maps $j_1 : A_1 \to A$ and $j_2 : A_2 \to A$ are jointly surjective, we only need to check that the outer square in the diagram

$$\begin{tikzcd}
A_1 \arrow[r, "j_1"] \arrow[d, "T_1"'] & A_{\mathcal{M}} \arrow[r, "g"] \arrow[d, "T_{\mathcal{M}}"] & A_{\mathcal{N}} \arrow[d, "T_{\mathcal{N}}"] \\
\mathcal{P}_{S_1} \arrow[r, "(i_1)_\star"'] & \mathcal{P}_{S_{\mathcal{M}}} \arrow[r, "f_\star"'] & \mathcal{P}_{S_{\mathcal{N}}}
\end{tikzcd}$$

commutes and the same is true for the corresponding diagram for j_2. But this is the case by construction of f and g, since $g \circ j_1 = \beta_1$ and $f \circ i_1 = \alpha_1$ and $(\alpha_1, \beta_1): \mathcal{M}_1 \to \mathcal{N}$ is a morphism in MDP. □

Proposition 2. *Suppose that $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$. Then $\mathcal{M}_1$ and $\mathcal{M}_2$ are subprocesses of $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$.*

Proof. We need to show that the functions i_1, i_2 in diagram (14) given that f_1, f_2 are injective and, similarly, that j_1, j_2 in diagram (16) are injective given the injectivity of g_1, g_2.

We check that this is true for i_1. The same argument works for all maps.

Suppose that $i_1(s_1) = i_1(s_1')$. If $i_1(s_1) \in S_1^\circ \subseteq S$, then we must have $s_1 = s_1' \in S_1^\circ$ since $i_1|_{S_1^\circ}$ is the identity map from $S_1^\circ \subseteq S_1$ to $S_1^\circ \subseteq S$.

If $i_1(s_1) \in f_1(S_3) \subseteq S$, then there exists a unique $s_3 \in S_3$ such that $s_1 = f_1(s_3)$, since f_1 is injective. Similarly $s_1' = f_1(s_3')$. Now $i_1(s_1) = i_1(s_1')$ implies that $f_1(s_3) \sim f_1(s_3')$ under the equivalence relation $\sim$ on $S_1 \coprod S_2$, whose quotient is S by definition. Since f_2 is also injective, it follows that the equivalence classes of $\sim$ are pairs $\{f_1(t), f_2(t)\} \subseteq S_1 \coprod S_2$ where t runs through S_3. Hence $f_1(s_3) \sim f_1(s_3')$ is only possible if $f_1(s_3) = f_1(s_3')$, which implies that $s_3 = s_3'$. □

C Proofs

Subprocesses

Proposition 1. *Any subprocess $\mathcal{M}_1' \to \mathcal{M}_2$ with state space S_1 factors uniquely through the subprocess $\mathcal{M}_1 \to \mathcal{M}_2$.*

Proof. This follows from the fact that any action $a_2 \in (A_2)_{s_1}$ such that $T_2(a_2)$ is a measure supported on S_1 is an element of A_1 defined in (4). □

Safe grid worlds

Proposition 3. *Suppose $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$ and any action $a_2 \in A_2 \setminus g_2(A_3)$ is not supported on S_3, meaning that there is some measurable subset $U_2 \subseteq S_2$ disjoint from $f_2(S_3)$ such that $T(a_2)(U_2) > 0$. Let $\mathcal{M}_2^\circ$ be the MDP obtained by puncturing $\mathcal{M}_2$ along $\mathcal{M}_3$. Then the MDP $(\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2)^\circ$ obtained by puncturing $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ along the subprocess $\mathcal{M}_2^\circ$ is the MDP $\mathcal{M}_1$.*

Proof. The state spaces of $(\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2)^\circ$ and $\mathcal{M}_1$ coincide. For the action spaces, by the given condition it follows that puncturing $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ along $\mathcal{M}_2$ will remove precisely the actions $A_2 \setminus \mathrm{g}_2(A_3) \subseteq A$. But, since $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$ and $\mathrm{g}_1(A_3) = \mathrm{g}_2(A_3)$, it follows that $A \setminus (A_2 \setminus \mathrm{g}_2(A_3)) = A_1$. □

Proposition 4. *There exists a commutative diagram*

$$\begin{tikzcd}
\mathcal{M}_{12} \arrow[r] \arrow[d] & \mathcal{M}_1 \arrow[d] \\
\mathcal{M}_2 \arrow[r] & \mathcal{M}
\end{tikzcd}$$

which is simultaneously a fiber and pushout diagram in MDP, *meaning that* $\mathcal{M}_{12} = \mathcal{M}_1 \times_{\mathcal{M}} \mathcal{M}_2$ *and* $\mathcal{M} = \mathcal{M}_1 \cup_{\mathcal{M}_{12}} \mathcal{M}_2$.

Proof. This is a simple check. To see that this is a fiber diagram, observe that the state space of the fiber product is the intersection of state spaces of the two punctured MDPs $\mathcal{M}_i^\circ$, which is the same as the state space of $\mathcal{M}_{12}^\circ$. For the action spaces, the action space of the fiber product consists of the actions in $\mathcal{M}$ that avoid the two obstacles $\mathbb{O}_i$. This coincides with the actions in the MDP $\mathcal{M}_{12}$. The transition probabilities also coincide tautologically.

The reasoning for the pushout is analogous. □

State-action symmetry

Proposition 5. *$\mathcal{M}/G$ is the quotient of the MDP $\mathcal{M}$ by the action of G in the sense that for any MDP $\mathcal{N}$ and a G-invariant morphism $m : \mathcal{M} \to \mathcal{N}$, there is a unique factorization $\mathcal{M} \overset{q}{\to} \mathcal{M}/G \to \mathcal{N}$.*

Proof. This follows immediately from the definition of $\mathcal{M}/G$. A morphism $m : \mathcal{M} \to \mathcal{N}$ is G-invariant if the maps between their state and action spaces are G-invariant and that this is equivalent to the condition that

$$m \circ pr_1 = m \circ \rho : \mathcal{M} \times G \to \mathcal{N}.$$

The conclusion then follows by the universal property of the pushout $\mathcal{M}/G$. □

Compositional learning

Theorem 3. *Suppose that the zig-zag diagram is forward-moving and the optimal value function of $\mathcal{C}_n$ is monotonic. Then, following the policy π_i on each component $\mathcal{M}_i$ gives an optimal policy on the composite MDP $\mathcal{C}_n$.*

Proof. First, consider the Bellman equation for all $s \in S$, $a \in A$, R the reward function, and γ the discount factor we have [Sutton and Barto, 2018, chapter 4]:

$$v_\star(s) = \max_{a \in A_s} \sum_{s' \in S} T(a)(s') \left(R(a) + \gamma v_\star(s') \right).$$

We argue by reverse induction. Suppose that the claim is true for the composite $\mathcal{C}_{[i+1,n]}$.

We then show that it is true for $\mathcal{C}_{[i,n]}$. Since the zig-zag diagram is forward-moving, the optimal values of $\mathcal{C}_{[i,n]}$ and $\mathcal{C}_{[i+1,n]}$ coincide on the state space of the latter, since successor states in $\mathcal{C}_{[i+1,n]}$ are states in $\mathcal{C}_{[i+1,n]}$. For the same reason, the optimal values of states of $\mathcal{M}_i$ are the same for the composite MDPs $\mathcal{C}_n$ and $\mathcal{C}_{[i,n]}$. The monotonicity condition now implies that the following policy π_i on $\mathcal{M}_i$ and the optimal policy on $\mathcal{C}_{[i+1,n]}$ gives an optimal policy on $\mathcal{C}_{[i,n]}$.

For the base case, observe that $\mathcal{M}_n = \mathcal{C}_[n, n]$ which has maximal policy π_n. □