

# Complexity of Conditional Planning under Partial Observability and Infinite Executions

Jussi Rintanen

Griffith University and the Australian National University  
Australia

**Abstract.** The computational properties of many classes of conditional and contingent planning are well known. The main division in the field is between probabilistic planning (typically infinite or unbounded executions, reward rather than goal-based, and focus on expected costs or rewards) and non-probabilistic planning (ignoring probabilities, focus on plans that reach goal states.) In this work, we address the middle ground between these problems: planning with infinite executions and designated goal states. We address *worst case* rather than *expected* costs measures for the problem we consider. We analyze the structure of the plans for two possible goal-based specifications such plans may have to satisfy, maintaining a goal property indefinitely as well as visiting a goal state infinitely often, and establish their complexity under different observability assumptions.

## 1 INTRODUCTION

Much work on planning focuses on a bounded horizon problem in which a sequence of actions is taken in order to reach a specific goal state. However, this kind of bounded horizon problem is often an unfaithful abstraction of what is going on in most applications to which such planning could be applied. For example, in intelligent robotics, the robot reaching a desired goal state is typically only an intermediate stage in the execution, which will continue further, to reach other goals. In general, infinite (or unbounded) executions would often seem to be the practically most relevant problem. Think of a robot that is repeatedly given the goal of collecting a valuable object, without having knowledge of its future goals. This robot would happily descend into a hole it cannot escape if there was an object there, as it would not care about its future goals. The constraint “don’t descend into a hole you cannot escape” could be manually added as a further requirement, but, in general, stating all such constraints would be difficult. A more robust solution is to let the planner to deal with the whole complexity of the problem.

In this paper, we consider two models of infinite and unbounded horizon problems. The first model is a form of a dual to the traditional goal-reaching plans, *maintaining* a given property by taking an infinite sequence of actions that never take the system out of the desired states. The second model involves repeatedly reaching a goal state, which disallows solutions in which reaching the goals once makes it impossible to reach them again.

The results complement those on planning in the Markov Decision Process model [11] and with goal-oriented planning without probabilities [14, 2]. The works on MDPs focus on *expected rewards* or *costs* as the plan quality measure. With infinite horizons and partial observability this measure leads to undecidability [10], which mo-

tivates the restriction to bounded horizon lengths [11]. The earlier work on conditional planning with partial observability has considered only the bounded horizon problem in which plan executions end after a goal state has been reached.

The model we consider covers *worst-case* plan cost/reward measures, such as “cost not exceeding  $N$  over any  $M$  step segment of the execution”. Our results show these worst-case measures to be computationally easier, avoiding the undecidability in the corresponding POMDP models. Intuitively, the reason for this is that worst-case measures can be handled for every execution separately: if each execution satisfies a cost bound, then the plan as a whole satisfies it. These cost bounds can be encoded in the states as additional state variables. Expected costs involve looking at the set of all executions and requires calculating the probability weight of each, which is an infinitary problem unlike the one arising with worst-case costs.

Our results show that two natural infinite-horizon contingent planning problems, one with maintenance and the other with repeated reaching of goals, are both decidable and no more difficult than the basic goal reachability objective [9, 14]. This is a positive result, possibly even surprising in the light of the undecidability results for the infinite-horizon POMDP models [10], as it means that there is – asymptotically and in the general case – no penalty in solving the infinite horizon problem in comparison to using its (unsound) decomposition into a sequence of separate bounded horizon tasks.

The structure of the paper is as follows. We first briefly recall the basic definitions related to planning and computational complexity in Section 2. In Section 3 we define the new infinite horizon problems. In Section 4 we outline the proof ideas for showing that the problems are hard for EXP and 2-EXP. In Section 5 we give algorithms for solving the problems, establishing membership in EXP and 2-EXP.<sup>1</sup>

## 2 PRELIMINARIES

We use a compact (succinct) representation of planning problems, which does not a priori require the enumeration of all states. Each *state*  $s$  is a valuation  $s : X \rightarrow \{0, 1\}$  of the finite set  $X$  of state variables. *Actions* can be viewed as binary relations on the set of states, associating with each state zero or more possible successor states. Such binary relations can be given a compact representation by using Binary Decision Diagrams or other logic-based representations [5, 8]. We identify an action with the corresponding binary relation compactly represented as a propositional formula.

---

<sup>1</sup> Full proofs will be available in a technical report.

**Definition 1** A succinct transition system is a 5-tuple  $\Pi = \langle X, I, A, G, V \rangle$  where  $X$  is a finite set of state variables,  $I$  is a formula over  $X$  describing the initial states,  $A$  is a finite set of actions over  $X$ ,  $G$  is a formula over  $X$  describing the goal states, and  $V \subseteq X$  is the set of observable state variables.

When  $V = X$  the system is *fully observable* and without restrictions on  $V$  it is *partially observable*. A succinct transition system can be expanded to an enumerative representation in which states are atomic objects and state sets and actions are represented as binary relations. The size of this representation may be exponential in the size of the succinct representation. Many of our later results assume the reduction from the succinct to the enumerative representation has been performed.

**Definition 2** A transition system is a 5-tuple  $\Pi = \langle S, I, A, G, (C_1, \dots, C_w) \rangle$  where  $S$  is a finite set of states,  $I \subseteq S$  is the initial states,  $A$  is a set of actions  $a \in 2^{S \times S}$  over  $S$ ,  $G \subseteq S$  is the goal states, and  $(C_1, \dots, C_w)$  is a partition of  $S$  to sets of states that are observationally indistinguishable.

In the reduction from succinct to enumerative transition systems, the partition  $(C_1, \dots, C_w)$  of  $S$  is obtained from  $V$ : each  $C_i$  consists of states that assign the same value to all variables in  $V$ . A transition system with  $w = |S|$  is fully observable.

The complexity classes we need in this work are EXP and 2-EXP, which represent all decision problems solvable by a deterministic Turing machine in  $\mathcal{O}(2^n)$  time and in  $\mathcal{O}(2^{2^n})$  time, respectively. EXP contains the complexity classes PSPACE and NP, and is known to properly contain P, therefore representing provably intractable computation unlike for example PSPACE and NP [12].

### 3 PROBLEM DEFINITION

In probabilistic planning [7], the focus is on maximizing discounted or average rewards over an infinite horizon. Works that ignore probabilities usually focus on plans that are guaranteed to reach given goal states [3, 1]. Also, the satisfaction of temporal logic specifications over an infinite horizon have been considered [13]. In this work, we focus on plan objectives defined in terms of goal states. Two objectives are obvious and are motivated by several practical applications. First, *maintenance*, requires that a plan keeps the system in the goal states indefinitely. The goal states specify the acceptable states of the system. For example, a robot may be required to keep an area clean or a system functioning. Second, *repeated reachability* requires not only that the goals are reached once, but that they will be reached over and over indefinitely. For example, a robot’s task may be to fetch some objects (e.g. mail) from area A and bring them to area B, and repeat this task.

These objectives can be combined with measures for the quality of plans. Instead of considering *expected* costs as in most of probabilistic planning with MDPs and POMDPs, we focus on *worst-case* cost measures. Our decision is motivated, as already mentioned in the introduction, by the high worst-case complexity of expected costs.

The results of this paper also cover a range of worst-case cost measures which can be easily reduced to the basic framework. The reduction of e.g. cost bound on fixed length segments of the sequences of executed actions can be easily encoded in the state variables. Consider segments of length 10. We need 10 multi-valued variables to store the costs of the past ten actions; action preconditions require that the sum of these costs does not exceed a bound; every action shifts the cost vector to the left by one and adds the cost of the current

action as the new 10th component. Executability of plans requires obeying the cost bound.

The key property of worst-case cost criteria of this nature is that they only involve looking at the action sequence leading to the current state, unlike expected cost criteria, which require looking at all possible executions starting from any state that could be reached.

We give three formalization of the work the plans do in terms of visiting goal states next.

1. RG reachability goals (finite executions) (Cimatti et al. [5]): It is required that a goal state is reachable from every non-goal state that is reachable from an initial state. This allows iterative trial-and-error strategies which do not have a finite upper bound on the length of executions.
2. MG maintenance goals (infinite executions): All states reached from the initial states must be goal states.
3. RRG Repeated reachability goals (infinite executions): It is required that under a plan, a goal state is reachable by one or more steps from every state that is reachable from an initial state. The difference to RG is that execution continues and goal states must be reachable also after reaching a goal state.

### 4 HARDNESS FOR EXP AND 2-EXP

The fully observable problems with maintenance and repeated reachability goals are EXP-hard, and the partially observable problems are 2-EXP-hard. We only sketch the proofs, which are relatively straightforward. For RRG, these hardness results can easily be established by reductions from RG, with a known complexity [9, 14]. The reductions add a new action that allows staying in a goal state indefinitely once it has been reached. The RRG problem is solvable with this new action if and only if a goal state can be reached with the RG problem.

Hardness proofs for maintenance are obtained by modification from the hardness proofs for the standard reachability goal problem, yielding simulations of EXP and 2-EXP Turing machines [14] as planning with maintenance goals. The key modification is to add a counter for the number of transitions so far, and to consider as goal states all accepting states as well as states with counter  $< 2^n$  for fully observable problems or  $< 2^{2^n}$  for partially observable problems, where  $n$  is the number of state variables. In the partially observable case the counter has an exponential number of bits, which can be encoded in the belief states [14]. Additionally, we add a dummy action that allows staying in the goal state indefinitely.<sup>2</sup> Hence the goal can be maintained indefinitely if and only if the Turing machine accepts.

### 5 MEMBERSHIP IN EXP AND 2-EXP

Our membership proofs for EXP and 2-EXP are constructive: we give algorithms for solving the problems, respectively with exponential and doubly exponential worst-case runtimes.

For MG, planning with partial observability easily reduces to the fully observable case, by the obvious exponential time reduction in which each *belief state* (set of possible current states) is viewed as a state. This is similar to RG, for which the membership in EXP in the fully observable case trivially yields membership in 2-EXP for the partially observable case [14].

So in the next section we show that MG with full observability is in EXP, and this trivially shows that MG with partial observability is in 2-EXP.

<sup>2</sup> Notice that a reduction from RG to MG does not work because the the RG objective allows unbounded trial-and-error, invalidating the use of bounded precision counters.

We will use the image, weak preimage and strong preimage operations of relations  $R$  (actions), respectively defined as follows [5].

$$\begin{aligned} \text{img}_R(S) &= \{s' \mid s \in S, \langle s, s' \rangle \in R\} \\ \text{preimg}_R(S) &= \{s \mid s' \in S, \langle s, s' \rangle \in R\} \\ \text{spreimg}_R(S) &= \{s \mid s' \in S, \langle s, s' \rangle \in R, \text{img}_R(s) \subseteq S\} \end{aligned}$$

The weak preimage includes all of the possible predecessor states of  $S$ , whereas the strong preimage limits to those from which reaching  $S$  by  $R$  is guaranteed.

We introduce some terminology. Let  $S$  be a set of states,  $A$  a set of actions, and  $\pi : S \rightarrow A$  a mapping from states to actions. A sequence  $s_0, \dots, s_n$  of states is an *execution* if for every  $i \in \{1, \dots, n\}$  there is  $a \in A$  such that  $s_i \in \text{img}_a(s_{i-1})$ . It is an *execution of  $\pi$*  if  $s_i \in \text{img}_{\pi(s_{i-1})}(s_{i-1})$  for every  $i \in \{1, \dots, n\}$ .

## 5.1 Maintenance

Figure 1 gives an algorithm for finding plans for maintenance goals. The algorithm starts with the set  $G$  of all states that satisfy the property to be maintained. Then iteratively such states are removed from  $G$  for which the satisfaction of the property cannot be guaranteed in longer executions, first with one step, then with two and more. The iteration ends when maintaining the property is guaranteed indefinitely, corresponding to the limit/fixpoint of the sequence of state sets in the loop. Similarly to RG [5], plans  $\pi$  with MG can always be represented as mappings  $\pi : S \rightarrow A$  from states to actions.

```

1: procedure MAINTENANCE( $I, A, G$ )
2: repeat
3:    $G_0 := G$ ;
4:    $G := \bigcup_{a \in A} (\text{spreimg}_a(G_0) \cap G_0)$ ;
5: until  $G = G_0$ ;
6: if  $I \subseteq G$  then return true else return false;

```

**Figure 1.** Testing existence of plans for maintenance goals

**Theorem 3** *Let  $I$  be a set of initial states,  $A$  a set of actions and  $G$  a set of goal states. Then MAINTENANCE( $I, A, G$ ) returns true if and only if there is a plan for  $(S, I, A, G, P)$  under MG.*

*Proof:* We sketch the main idea of the proof.

Let  $G'$  be final value of the variable  $G$  in the procedure MAINTENANCE in Figure 1. The induction proof shows that  $G' \subseteq G$  and there is a plan  $\pi$  such that  $\text{img}_{\pi(s)}(s) \subseteq G'$  for every  $s \in G'$ , and for every  $s \in G \setminus G'$  and every plan  $\pi'$  there is  $n \geq 1$  and an execution  $s_0, \dots, s_n$  of  $\pi'$  with  $s_0 = s$  such that  $s_n \notin G$ .

Now, the procedure returns *true* iff all initial states are in  $G'$  if and only if all states reachable from the initial states under some plan  $\pi$  are in  $G'$ .  $\square$

This algorithm can be trivially lifted to the partially observable case with belief states replacing the role of states. A goal belief state is a belief state that consists of goal states only. Membership in 2-EXP trivially follows.

## 5.2 Repeated reachability

We give a new algorithm for solving the planning problem with full observability and repeated reachability goals. This problem generalizes the problem solved by an algorithm given by Cimatti et al.

[5] (the global strong cyclic algorithm) for our simplest objective of reachability goals. Our algorithm runs in polynomial time in the size of the state space and hence yields an exponential time upper bound for the plan existence problem of succinct transition systems. Our algorithm uses the subprocedure PRUNE given in Figure 2, which is arguably simpler than a similar algorithm by Cimatti et al. because it does not explicitly construct a state-action table.

The algorithm first identifies all states  $W_0$  from which a state in  $G$  is reachable (loop at line 5). Then, the loop with  $i$  on line 11 eliminates those states in  $W_i$  for which, under any plan, there is an execution that leads outside  $W_{i-1}$  and hence makes the goal unreachable. The inner loop on line 15 identifies those states in  $S_k$  that can reach  $G$  in  $k$  steps without risking getting outside of  $W_{i-1}$ . The termination conditions of the first and the inner loop correspond to the limit in which the number of steps for reaching the goals can be arbitrarily high. The termination condition of the outer loop 11 corresponds to the requirement that for any execution starting in  $W_i$  we are guaranteed to stay inside  $W_i$  until a goal state is reached.

```

1: procedure PRUNE( $A, G$ );
2:  $W_{-1} :=$  all states;
3:  $k := 0$ ;
4:  $W_{0,0} := \emptyset$ ;
5: repeat
6:    $k := k + 1$ ;
7:    $W_{0,k} := (W_{0,k-1} \cup \bigcup_{a \in A} (\text{preimg}_a(W_{0,k-1} \cup G)))$ ;
8: until  $W_{0,k} = W_{0,k-1}$ ;
9:  $W_0 := W_{0,k}$ ; (*  $G$  is reachable from every  $s \in W_0$ . *)
10:  $i := 0$ ;
11: repeat
12:    $i := i + 1$ ;
13:    $k := 0$ ;
14:    $S_0 := \emptyset$ ;
15:   repeat
16:      $k := k + 1$ ;
17:      $S_k := S_{k-1} \cup \bigcup_{a \in A} \left( \text{preimg}_a(S_{k-1} \cup G) \right. \\ \left. \cap \text{spreimg}_a(W_{i-1} \cup G) \right)$ ;
18:   until  $S_k = S_{k-1}$ ;
19:    $W_i := S_k$ ; (* Reach  $G$  from  $s \in W_i$  while staying in  $W_{i-1}$ . *)
20: until  $W_i = W_{i-1}$ ;
21: return  $W_i$ ;

```

**Figure 2.** Identifying all states from which goals are eventually reached

**Lemma 4 (Procedure PRUNE)** *Let  $S$  be the set of all states,  $G \subseteq S$  a set of states and  $A$  a set of actions. Then the procedure call PRUNE( $A, G$ ) will terminate after a finite number of steps returning  $W \subseteq S$  so that there is function  $\pi : W \rightarrow A$  such that*

1. *for every  $s \in W$  there is an execution  $s_0, s_1, \dots, s_n$  of  $\pi$  with  $n \geq 1$  such that  $s = s_0$  and  $s_n \in G$ ,*
2.  *$\text{img}_{\pi(s)}(s) \subseteq W \cup G$  for every  $s \in W$ , and*
3. *for every  $s \in S \setminus W$  and function  $\pi' : S \rightarrow A$  there is an execution  $s_0, \dots, s_n$  of  $\pi'$  such that  $s = s_0$  and there is no  $m \geq n$  and execution  $s_n, s_{n+1}, \dots, s_m$  such that  $s_m \in G$ .*

*Proof:* By straightforward, but quite involved, nested inductions exactly matching the repeat-until loops in the algorithm.  $\square$

The main procedure of the decision procedure for repeated reachability under full observability is given in Figure 3. The procedure first

assigns  $G_{ne} := G$ , and then repeatedly eliminates – with *PRUNE* – those states from  $G_{ne}$  for which there is no plan that is guaranteed to eventually reach a state in  $G_{ne}$ .

After the last iteration of the loop,  $G_{ne}$  will be the maximal subset of  $G$  from which reaching a state in  $G_{ne}$ , again, is guaranteed with an implicitly represented plan  $\pi$ . The number of iterations is bounded by the number of states, and each iteration has a runtime that is polynomial in the number of states. Hence the total runtime of this stage is exponential in the size of the succinct transition system.

The last line of the procedure tests whether the initial states are included in  $W$ . If they are, then any execution with  $\pi$  from an initial state eventually reaches a state in  $G_{ne}$ , and hence the RRG objective is satisfied.

**Theorem 5** *Testing plan existence for succinct transition systems with full observability under the repeated reachability objective is in EXP.*

*Proof:* Given a succinct transition system  $\Pi$ , we can produce the corresponding transition system  $F(\Pi) = (S, I, A, G, P)$  in exponential time. Then we call the procedure *DECIDE-FO-RRG*( $I, A, G$ ) which is given in Figure 3. The first call to *PRUNE* yields states from which

```

1: procedure DECIDE-FO-RRG( $I, A, G$ )
2:  $G_{ne} := G$ ;
3: repeat
4:    $W := \text{PRUNE}(A, G_{ne})$ ;
5:    $G'_{ne} := G_{ne}$ ;
6:    $G_{ne} := G_{ne} \cap W$ ;
7: until  $G_{ne} = G'_{ne}$ ;
8: if  $I \subseteq W$  then return true else return false;

```

**Figure 3.** Testing existence of RRG plans with full observability

a goal state can be reached once, but there is no guarantee that the goals can be reached again from all of those goal states. Reaching the fixpoint on line 7 guarantees that goal states in  $G_{ne}$  can be reached from states in  $W$  an infinite number of times.  $\square$

The generalization of the EXP membership for RRG to 2-EXP membership for RRG and partial observability is more complicated than with MG. Simple reductions to the fully observable case as for MG don't exist. The issue is that with full observability, it is always known what the current state is, whereas with partial observability it might never be known whether the current state is a goal state, and the problem can still be solvable under RRG. This is radical difference to RG and MG. We illustrate this by an example.

**Example 6** *Consider the following transition system, in which state  $b$  is the only goal state.*



*Initially all three states are possible, and no further information about the current state is obtained later. Alternating the two actions, one depicted with a dotted line and the other with a solid line, satisfies the Repeated Reachability objective. Applying either of these two actions exclusively does not.* ■

The underlying issue is that the conventional notion of belief states as a set of possible current states (or a probability distribution over

them) does not carry enough information to decide what to do next. Therefore, the action to be taken is not a function of the set of possible current states.

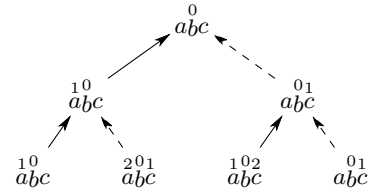
The important insight is that, under a given plan, we need to consider the set of possible current states, as well as the (optimistic) distance (number of actions) to a goal state for all of the possible current states. We will explain this insight in more detail next.

Consider a plan for the problem in Example 6. This plan alternates between the two actions. The set of possible current states at each stage of execution is the same,  $B = \{a, b, c\}$ . At every stage of execution, the distance from  $b$  to the goal state is 0, but the distances from  $a$  and  $c$  depend on which actions will follow. If our plan alternates between the two actions, and the dotted action is taken first, then the distances in the current stage are  $d_0(a) = 2, d_0(b) = 0, d_0(c) = 1$ , and at the stage following it they are  $d_1(a) = 1, d_1(b) = 0, d_1(c) = 2$ .

The extended belief states in this example are  $\langle B_0, d_0 \rangle$  and  $\langle B_1, d_1 \rangle$ , where  $B_0 = B_1 = \{a, b, c\}$ . The plan that only takes the dotted action always stays in the extended belief state  $\langle \{a, b, c\}, d \rangle$  with  $d(a) = \infty, d(b) = 0, d(c) = 1$ , never visiting a goal state when the execution starts from  $a$ .

The algorithm we will give for planning with the RRG objective iteratively generates extended belief states with a finite distance for as many of the constituent states as possible.

The basic intuition is that we start from extended belief states that assign distances to goal states only, and then repeatedly generate the predecessors. With Example 6 we start from  $\langle \{a, b, c\}, \{(b, 0)\} \rangle$  (and all subsets of  $\{a, b, c\}$ ), and as the preimage of the dotted action we obtain  $\langle \{a, b, c\}, \{(b, 0), (c, 1)\} \rangle$ , because the predecessor state of  $b$  w.r.t. the dotted action is  $c$ . Now with the undotted action we get the preimage  $\langle \{a, b, c\}, \{(a, 1), (b, 0), (c, 2)\} \rangle$ . This process, with alternative actions for each extended belief state, is illustrated in Figure 4. The leaves in the tree represent belief states from which



**Figure 4.** Preimages of extended belief states

the root node can be reached by taking the actions on the path to the root. The distances attached to the states indicate the number of steps that are needed to visit the goal state  $b$ . If a distance is not defined, then the distance to a goal state is unknown due to the unknown goal distance for the states in the root node.

We can see that starting from the leaf with distances 201 and taking the dotted and then the undotted action we can get back to the same extended belief state (partially specified in the root node by explicitly stated distance only for state  $b$ .) Similarly for the leaf 102.

This example is very simple, as it does not include nondeterministic actions nor branching of the plans due to observations. Next we will define the weak and strong preimage operations for extended belief states which cover these features.

The state space is partitioned to  $(C_1, \dots, C_w)$  to sets of obser-

vationally indistinguishable states. We consider only belief states  $B$  such that  $B \subseteq C_i$  for some  $i \in \{1, \dots, w\}$ , as any belief state overlapping two classes could be split, corresponding to eliminating those states from consideration that are not compatible with the current observations.

Given plans for belief states in some set  $W$ , we can have a plan for any of the belief states in  $\{B \subseteq S \mid \{B_1, \dots, B_w\} \subseteq W, B_i \subseteq C_i \text{ for all } i \in \{1, \dots, w\}, a \in A, \text{img}_a(B) = \bigcup_{i=1}^w B_i\}$ . These are all the belief states from which a belief state in  $W$  is reached by taking an action in  $A$ , partitioning the set of possible successor states according to the possible observations, and then choosing the belief state that corresponds to the actual observation.

To obtain a preimage operation for extended belief states we add distances. The distance assigned to each state is optimistic in the sense that a goal state can be reached in the specified number of steps under a given plan, but there is no guarantee that it will. This corresponds to the requirements of the RRG objective. We also define a *weak preimage* operation, which does not require that all successor belief states are in  $W$ . This is analogous to the definition of weak preimages for states. The definitions use the predicate  $\text{Belief}_\Pi(B)$ , which requires that  $B$  is a belief state, i.e. it consists of mutually indistinguishable states in  $\Pi$ , and the predicate  $\text{Distinct}_\Pi(\{B_1, \dots, B_k\})$ , which requires that  $B_1, \dots, B_k$  are belief states corresponding to different observations.

**Definition 7 (Weak Preimages for Extended Belief States)** Let  $\Pi = \langle S, I, A, G, (C_1, \dots, C_n) \rangle$  be a transition system, and  $T$  a set of pairs  $\langle B, d \rangle$  such that  $d : B \rightarrow \mathbb{N}$  is a partial function and  $\text{Belief}_\Pi(B)$  holds. Define  $\text{wpreimg}_a^E(T) =$

$$\begin{aligned} & \{ \langle B, d \rangle \mid \text{Belief}_\Pi(B), \\ & \quad \{ \langle B_1, d_1 \rangle, \dots, \langle B_k, d_k \rangle \} \subseteq T, \\ & \quad \text{Distinct}_\Pi(\{B_1, \dots, B_k\}), \\ & \quad \text{img}_a(B) \cap C_i = B_j \text{ for all } j \in \{1, \dots, k\} \\ & \quad \quad \text{and some } i \in \{1, \dots, n\}, \\ & \quad d(s) = 0 \text{ for all } s \in B \cap G, \\ & \quad d(s) = 1 + \min_{i=1}^k \min_{s' \in \text{img}_a(s) \cap B_i} d_i(s') \text{ for all } s \in B \setminus G \}. \end{aligned}$$

The definition of strong preimages is almost exactly the same, with the difference that *all* successors of  $(B, d)$  have to be in  $T$ .

Given sets  $T$  of extended belief states that occur during the execution of a plan, we can use the preimage operations to find further extended belief states  $\text{spreimg}_a^E(T)$  for which visits to goal states can be guaranteed. This is analogous to the use of the preimage operations for states in the algorithm PRUNE and the decision procedure DECIDE-FO-RRG before.

We can show that mappings from extended belief states to actions can express any plan that is expressible by any finite conditional plan. Plans in general can be defined as program-like structures that map a sequence of observations to the action to be taken next [14], with relevant observation sequences abstractly represented by nodes (program counters) of the plan.

**Lemma 8** *Let there be a plan that satisfies the RRG objective. Then there is a plan that is a mapping from extended belief states to actions.*

*Proof:* We only give a brief sketch. The proof progresses in stages. First, the given plan is expanded so that there are different copies of each original plan node for every possible extended belief state that could be the current one in that node. Then it is shown that all

nodes with the same extended belief states can be combined without affecting the satisfaction of the RRG criterion: for any execution of the plan for which reaching a goal state is always possible, any arc to node  $n_1$  (with a given extended belief state) can be re-directed to any other node  $n_2$  with the same extended belief state, without violating the RRG objective, and the node  $n_1$  can be deleted.  $\square$

Now we are ready to give the algorithm for the RRG objective under partial observability. The structure of the algorithm is the same as in the fully observable case, but the sets of states will be replaced by sets of extended belief states. The subprocedure  $\text{PRUNE}^E$  is obtained from PRUNE by replacing the image operations for states by image operations for extended belief states.

The decision procedure is given in Figure 5. The computation works in the (infinite) space of all extended belief states. Define the predicate  $\text{covered}(B, W)$  iff  $\langle B, d \rangle \in W$  for some  $d$  such that  $(s, n) \in d$  for all  $s \in B$  and some  $n \in \mathbb{N}$ . This means that all states in  $B$  have a finite distance to a goal state in an extended belief state in  $W$ . Initially the set  $G_{ne}$  consists of all extended belief states that

```

1: procedure DECIDE-PO-RRG( $S, I, A, G, (C_1, \dots, C_n)$ )
2:  $G_{ne} := \{ \langle B, (B \cap G) \times \{0\} \rangle \mid 1 \leq i \leq n, B \subseteq C_i \}$ ;
3: repeat
4:  $W := \text{PRUNE}^E(A, G_{ne})$ ;
5:  $G'_{ne} := G_{ne}$ ;
6:  $G_{ne} := \{ \langle B, (B \cap G) \times \{0\} \rangle \in G_{ne} \mid \text{covered}(B, W) \}$ ;
7: until  $G_{ne} = G'_{ne}$ ;
8: if for all  $i \in \{1, \dots, n\}, \langle I \cap C_i, d \rangle \in W$  for some  $d$ 
9: then return true else return false;

```

**Figure 5.** Testing existence of RRG plans under partial observability

assign 0 distance to goal states in the belief state, and don't assign any distance to non-goal states. Subsequent iterations of the main loop on line 3 eliminate (line 6) those members  $(B, d)$  of  $G_{ne}$  with no finite distance found for all  $s \in B$  in the sense that there is no  $(B, d') \in W$  such that  $d'(s)$  is defined for all  $s \in B$ .

A key observation is that DECIDE-PO-RRG will generate all extended belief states that occur in *any* plan that solves the problem instance in question, entailing completeness. The algorithm is also sound, as the preimage operations for extended belief states faithfully represent the relation between extended belief states and their predecessors and successors.

The algorithm, as we have described it so far, has infinite loops because of the arbitrarily high distances that can be found for states. To make the algorithm finitary, we ignore any extended belief state  $\langle B, d \rangle$  in  $\text{PRUNE}^E$  and in DECIDE-PO-RRG as soon as we have some  $\langle B, d' \rangle$  such that  $\{s \mid (s, n) \in d\} = \{s \mid (s, n) \in d'\}$ , that is, it has finite distances for the same constituent states. This makes all sets of relevant extended belief states finite, guarantees the finite termination of all loops, and does not affect completeness or soundness.

**Theorem 9** *Testing plan existence for succinct transition systems with partial observability under the RRG objective is in 2-EXP.*

*Proof:* We sketch the proof, which is analogous to the EXP-membership proof for the fully observable case.

The call to  $\text{PRUNE}^E$  on line 4 identifies those extended belief states from which  $G_{ne}$  can be reached.

Line 6 retains those  $(B, d)$  in  $G_{ne}$  for which there is  $(B, \hat{d}) \in W$  such that  $\hat{d}(s)$  is defined for all  $s \in B$ , and eliminates the rest. Let

$\widehat{G}_{ne}$  be the set of such extended belief states  $(B, \widehat{d})$  for all  $(B, d) \in G_{ne}$ . Now there is a plan so that for every  $(B, \widehat{d}) \in \widehat{G}_{ne}$  we are guaranteed to reach  $\widehat{G}_{ne}$  again by a non-empty execution, and for every  $s \in B$  there is at least one such execution that visits a goal state on the way. We call this one *good cycle* of the plan.

At the end of the  $i$ th iteration of the loop that starts on line 3,  $\widehat{G}_{ne}$  consists of extended belief states such that there is a plan with  $i$  consecutive good cycles.

When the loop terminates,  $G_{ne}$  represents extended belief states for which there are infinitely many consecutive good cycles, satisfying the RRG objective. The set  $W$  at this point consists of those extended belief states from which we are guaranteed to reach an extended belief state in  $G_{ne}$  (but which itself is not necessarily a part of a good cycle.) If the initial belief states are included in  $W$ , then a plan for the problem instance with those initial belief states exists.

The runtime of DECIDE-PO-RRG is polynomial in the size of the set of generated extended belief states. The number of generated extended belief states is  $\mathcal{O}(2^{2^n})$ : each extended belief state is a pair  $\langle B, d \rangle$ , there are  $2^{2^n}$  different sets  $B$  for  $n$  state variables, and the number of functions  $d$  generated by DECIDE-PO-RRG under our pruning criterion is also  $\mathcal{O}(2^{2^n})$  because for every  $B$  we generate at most  $2^{2^n}$  of them, yielding an  $\mathcal{O}(2^{2^n})$  upper bound  $2^{2^n} 2^{2^n} = 2^{2^{n+1}}$  for their number. Their total size is also  $\mathcal{O}(2^{2^n})$  because the size of each is  $\mathcal{O}(2^n)$  and  $\mathcal{O}(2^n 2^{2^n})$  equals  $\mathcal{O}(2^{2^n})$ .  $\square$

## 6 RELATED WORK

In addition to works on MDPs and POMDPs, infinite executions have earlier been considered with temporally extended goals, for example expressed in logics such as Linear Temporal Logic LTL [6] and the Computation Tree Logic [13]. The classical goal reachability, the maintenance and the repeated goal reachability respectively have a meaning intuitively corresponding to the LTL formulas  $F\phi$ ,  $G\phi$ ,  $GF\phi$ , where  $\phi$  is a non-modal formula. For the first two formulas and criteria the correspondence is exact, but for the third one not. Our repeated reachability objective rather corresponds to the Computation Tree Logic formula  $AGEF\phi$ , which says that for all executions, always in the future, there is at least one execution that reaches  $\phi$ . This formula is compatible with the existence of a degenerate execution that never reaches  $\phi$ , as long as reaching  $\phi$  always remains possible. The CTL goal  $AGEF\phi$  agrees with the LTL goal  $GF\phi$  if we assume, for the CTL case, a fairness condition that guarantees that the “wrong” choices, leading to avoiding  $\phi$ , don’t continue forever.

The complexity of planning with temporal logic goals has been investigated before. The most complex case for conventional action representations, investigated by Giacomo and Vardi [6], is with partial observability and deterministic actions, which is EXPSPACE-complete. Calvanese et al. [4] investigate a very general language in which both actions and the goal specifications are expressed as LTL formulas. In their framework the plan existence problem in the most general case is 2-EXPSPACE-complete, which is far more complex than the 2-EXP-completeness with conventional (non-modal) action representations used earlier [14].

## 7 CONCLUSIONS

We have shown that two natural partially observable infinite horizon conditional planning problems, with maintenance and repeated reachability, are 2-EXP-complete. The results complete the picture

of conditional planning, which has been well understood both in its probabilistic infinite horizon (MDP, POMDP) and non-probabilistic finite horizon variants. The infinite-horizon conditional planning problems we addressed are best viewed as representing worst-case performance criteria. Earlier works on probabilistic *expected cost* criteria have shown the partially observable problems to be computationally very difficult and undecidable in the most general cases [10]. In some applications, there are legal reasons or risk-averseness that make worst-case criteria preferable, because expected cost criteria allow solutions that fail as long as failure probabilities are low enough. Expected cost criteria are preferable in applications that involve a high number of plan executions and in which failures are only assessed in terms of their expected cost.

## REFERENCES

- [1] Piergiorgio Bertoli, Alessandro Cimatti, Marco Roveri, and Paolo Traverso, ‘Planning in nondeterministic domains under partial observability via symbolic model checking’, in *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, ed., Bernhard Nebel, pp. 473–478. Morgan Kaufmann Publishers, (2001).
- [2] Blai Bonet, ‘Conformant plans and beyond: Principles and complexity’, *Artificial Intelligence*, **174**(3-4), 245–269, (2010).
- [3] Blai Bonet and Héctor Geffner, ‘Planning with incomplete information as heuristic search in belief space’, in *Proceedings of the Fifth International Conference on Artificial Intelligence Planning Systems*, eds., Steve Chien, Subbarao Kambhampati, and Craig A. Knoblock, pp. 52–61. AAAI Press, (2000).
- [4] D. Calvanese, G. De Giacomo, and M.Y. Vardi, ‘Reasoning about actions and planning in I1 action theories’, in *Principles of Knowledge Representation and Reasoning: Proceedings of the Eighth International Conference (KR 2002)*, pp. 593–602. Morgan Kaufmann Publishers, (2002).
- [5] Alessandro Cimatti, Marco Pistore, Marco Roveri, and Paolo Traverso, ‘Weak, strong, and strong cyclic planning via symbolic model checking’, *Artificial Intelligence*, **147**(1–2), 35–84, (2003).
- [6] G. De Giacomo and M. Vardi, ‘Automata-theoretic approach to planning for temporally extended goals’, in *Recent Advances in AI Planning. 5th European Conference on Planning, ECP’99, Durham, UK, September 8-10, 1999. Proceedings*, eds., Susanne Biundo and Maria Fox, number 1809 in Lecture Notes in Artificial Intelligence, pp. 226–238. Springer-Verlag, (2000).
- [7] Leslie Pack Kaelbling, M. L. Littman, and Anthony R. Cassandra, ‘Planning and acting in partially observable stochastic domains’, *Artificial Intelligence*, **101**(1-2), 99–134, (1998).
- [8] Henry Kautz and Bart Selman, ‘Planning as satisfiability’, in *Proceedings of the 10th European Conference on Artificial Intelligence*, ed., Bernd Neumann, pp. 359–363. John Wiley & Sons, (1992).
- [9] Michael L. Littman, ‘Probabilistic propositional planning: Representations and complexity’, in *Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97) and 9th Innovative Applications of Artificial Intelligence Conference (IAAI-97)*, pp. 748–754. AAAI Press, (1997).
- [10] Omid Madani, Steve Hanks, and Anne Condon, ‘On the undecidability of probabilistic planning and related stochastic optimization problems’, *Artificial Intelligence*, **147**(1–2), 5–34, (2003).
- [11] Martin Mundhenk, Judy Goldsmith, Christopher Lusena, and Eric Alender, ‘Complexity of finite-horizon Markov decision process problems’, *Journal of the ACM*, **47**(4), 681–720, (2000).
- [12] Christos H. Papadimitriou, *Computational Complexity*, Addison-Wesley Publishing Company, 1994.
- [13] M. Pistore and P. Traverso, ‘Planning as model checking for extended goals in non-deterministic domains’, in *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, pp. 479–486, (2001).
- [14] Jussi Rintanen, ‘Complexity of planning with partial observability’, in *ICAPS 2004. Proceedings of the Fourteenth International Conference on Automated Planning and Scheduling*, eds., Shlomo Zilberstein, Jana Koehler, and Sven Koenig, pp. 345–354. AAAI Press, (2004).