Symmetries and Model Minimization in Markov Decision Processes

Balaraman Ravindran
Andrew G. Barto

CMPSCI Technical Report 01-43

September 28, 2001

Department of Computer Science
University of Massachusetts
Amherst, Massachusetts 01003

Abstract

Current solution and modelling approaches to Markov Decision Processes (MDPs) scale poorly with the size of the MDP. Model minimization methods address this issue by exploiting redundancy in problem specification to reduce the size of the MDP model. Symmetries in a problem specification can give rise to special forms of redundancy that are not exploited by existing minimization methods. In this work we extend the model minimization framework proposed by Dean and Givan to include symmetries. We base our framework on concepts derived from finite state automata and group theory.

Symmetries and Model Minimization in Markov Decision Processes

Balaraman Ravindran[1] Andrew G. Barto[2]

CMPSCI Technical Report 01-43
Department of Computer Science
University of Massachusetts, Amherst

Abstract

Current solution and modelling approaches to Markov Decision Processes (MDPs) scale poorly with the size of the MDP. Model minimization methods address this issue by exploiting redundancy in problem specification to reduce the size of the MDP model. Symmetries in a problem specification can give rise to special forms of redundancy that are not exploited by existing minimization methods. In this work we extend Dean and Givan's [5] model minimization framework to include symmetries. We base our framework on concepts derived from finite state automata and group theory.

1 Introduction

Markov Decision Processes (MDPs) [21] are a popular way to model stochastic sequential decision problems. But most modelling and solution approaches to MDPs suffer from the fact that they scale poorly with the size of the problem. While modelling real-world scenarios, often there is a lot of redundancy in the MDP model. Model minimization methods introduced by Dean and Givan [5] exploit such redundancy in the problem specification to derive smaller models, i. e., models with fewer states, by aggregating "equivalent" states.

Figure 1 illustrates the model minimization process. The gridworld on the left is the given MDP. This has the usual gridworld dynamics with 4 deterministic actions $\{N, S, E, W\}$. Each cell in the grid corresponds to a state of the MDP. All the states in the top row are goal states with identical rewards for reaching them. Dean and Givan consider two states equivalent if the effect of each of the available action is equivalent in both the states and if no essential information is lost by aggregating them. In this example, the states in each row can be considered equivalent to one another.[3] The resulting reduced model is just a column of states as depicted in the right of Figure 1. It is evident that solving this reduced model will give us a solution to the original problem.

[1]e-mail: `ravi@cs.umass.edu`

[2]e-mail: `barto@cs.umass.edu`

[3]We formalize the notion of equivalence later. Informally, in this special case, the states in a row are considered equivalent since each action changes the distance to goal by the same amount. Action N takes you one step closer to the goal. Action S takes you one step farther in most cases and keeps you at the same distance in the bottom row. Actions E and W keep you at the same distance from the goal.

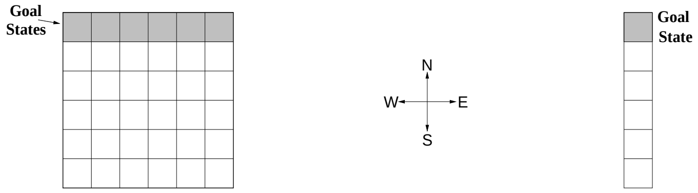

Figure 1: Illustration of Model Minimization in a Simple Gridworld

A special form of redundancy arises in cases where the problem is symmetric and AI researchers have been exploring ways to take advantage of that (e. g., refs. [1, 20]). Symmetries in a problem specification naturally give rise to notions of equivalence. For example consider another simple gridworld with usual dynamics, shown to the left in Figure 2. The goal state is labelled G. Intuitively one can see that there is a symmetry about the NE-SW diagonal. For example taking action E in state A is equivalent to taking action N in state B, in the sense that they go to equivalent states that are one step closer to the goal. One can say that states A and B are symmetrically equivalent. Dean and Givan's model minimization framework cannot accommodate such notions of equivalence and hence considers this gridworld irreducible.[4] In this work we extend the model minimization framework to include such notions of symmetrical equivalence. A reduced model of the gridworld under our framework is shown to the right in Figure 2.

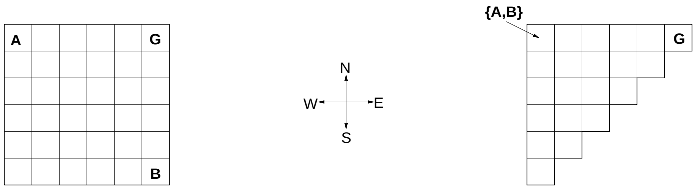

Figure 2: Model Minimization with Symmetric Equivalence

In the next section we present some basic concepts and notation. Then we discuss some related work on minimization of different structures. In Section 4 we present an extension to Dean and Givan's model minimization framework using the notion of homomorphisms derived from classical finite state automata (FSA) [11] theory. Next we develop a formal definition of symmetry in MDPs and show how it relates to our model minimization framework. We conclude with a discussion of certain specializations of our framework, some implications and future directions for research.

[4]States in the same row in the gridworld of Figure 1 are also symmetric to each other. While Dean and Givan's framework does accommodate some cases of symmetries, their theory does not explicitly consider symmetries of MDPs.

2 Basics and Notation

2.1 Markov Decision Processes

A *Markov Decision Process* is a tuple $\langle S, A, \Psi, P, R \rangle$, where S is the set of states, A is the set of actions, $\Psi \subseteq S \times A$ is the set of admissible state-action pairs, $P : \Psi \times S \to [0,1]$ is the transition probability function with $P(s,a,s')$ being the probability of transition from state s to state s' under action a, and $R : \Psi \to \mathbb{R}$ is the expected reward function, with $R(s,a)$ being the expected reward for performing action a in state s. We assume that the rewards are bounded. Let $A_s = \{a | (s,a) \in \Psi\} \subseteq A$ denote the set of actions admissible in state s. We assume that for all $s \in S$, A_s is non-empty. In this work we assume that the set of states and set of actions are finite, but the language of homomorphisms we employ extends to infinite spaces with little work.

A *stochastic policy* π is a mapping from Ψ to the real interval $[0,1]$ with $\sum_{a \in A_s} \pi(s,a) = 1$ for all $s \in S$. For any $(s,a) \in \Psi$, $\pi(s,a)$ gives the probability of picking action a in state s. The *value* of state s under policy π is the expected value of the discounted sum of future rewards starting from state s and following policy π thereafter. The *value function* V^π corresponding to a policy π is the mapping from states to their values under π. It can be shown (e. g., ref. [2]) that V^π satisfies the *Bellman equation*:

$$V^\pi(s) = \sum_{a \in A_s} \pi(s,a) \left[R(s,a) + \gamma \sum_{s' \in S} P(s,a,s') V^\pi(s') \right],$$

where $0 \leq \gamma < 1$ is a discount factor. This formulation is known as the discounted sum of rewards criterion.

Similarly, the value of a state-action pair (s,a) under policy π is the expected value of the discounted sum of future rewards starting from state s, taking action a, and following π thereafter. The *action value function* Q^π corresponding to a policy π is the mapping from state-action pairs to their values and satisfies [2]:

$$Q^\pi(s,a) = R(s,a) + \gamma \sum_{s' \in S} P(s,a,s') V^\pi(s'),$$

where $0 \leq \gamma < 1$ is a discount factor.

The solution of an MDP is an *optimal policy* $\pi^\star$ that uniformly dominates all other possible policies for that MDP. It can be shown [2] that the value functions for all optimal policies is the same. We denote this *optimal value function* by $V^\star$. It satisfies the *Bellman optimality equation*:

$$V^\star(s) = \max_{a \in A_s} \sum_{s' \in S} P(s,a,s') \left[R(s,a) + \gamma V^\star(s') \right].$$

Similarly the *optimal action value function* $Q^\star$ satisfies:

$$Q^\star(s,a) = \sum_{s' \in S} P(s,a,s') \left[R(s,a) + \gamma \max_{a' \in A_{s'}} Q^\star(s',a') \right].$$

These two optimal value functions are related by $V^\star(s) = \max_a Q^\star(s,a)$.

Typically MDPs are solved by approximating the solution to the Bellman optimality equations (e. g., refs. [2, 23]). Given the optimal action value function, an optimal policy is given by

$$\begin{aligned} \pi^\star(s,a) &\geq 0 \quad \text{if } Q^\star(s,a) = \max_{a' \in A_s} Q^\star(s,a') \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

2.2 Partitions, maps and equivalence relations

A *partition* B of a set X is a collection of disjoint subsets, or *blocks*, $b_i \subseteq X$ such that $\bigcup_i b_i = X$. For any $x \in X$, $[x]_B$ denotes the block of B to which x belongs. Let B_1 and B_2 be partitions of X. We say that B_1 is *coarser than* B_2 (or B_2 is a *refinement of* B_1), denoted $B_1 \geq B_2$, if for all $x, x' \in X$, $[x]_{B_2} = [x']_{B_2}$ implies $[x]_{B_1} = [x']_{B_1}$. The relation $\geq$ is a partial order on the set of partitions of X.

To any partition B of X there corresponds an equivalence relation, $\equiv_B$, on X with $x \equiv_B x'$ if and only if $[x]_B = [x']_B$ for all $x, x' \in X$. Any function f from a set X into a set Y defines an equivalence relation on X with $x \equiv_f x'$ if and only if $f(x) = f(x')$. We say that x and x' are *f-equivalent* when $x \equiv_f x'$, and we denote the partition of X corresponding to this equivalence relation by B_f.

Let B be a partition of $Z \subseteq X \times Y$, where X and Y are arbitrary sets. For any $x \in X$, let $B(x)$ denote the set of distinct blocks of B containing pairs of which x is a component, that is, $B(x) = \{[(w,y)]_B \mid (w,y) \in Z, w = x\}$. The *projection of B onto X* is the partition $B|X$ of X such that for any $x, x' \in X$, $[x]_{B|X} = [x']_{B|X}$ if and only if $B(x) = B(x')$. In other words, $x \equiv_{B|X} x'$ if and only if every block of B containing a pair in which x (x') is a component also contains a pair in which x' (x) is a component.[5] Note that if B_1 and B_2 are partitions of Z, then $B_1 \geq B_2$ implies that $B_1|X \geq B_2|X$.

A *partition of an MDP* $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ is a partition of Ψ. Given a partition B of $\mathcal{M}$, the *block transition probability of* $\mathcal{M}$ is the function $T : \Psi \times B|S \to [0,1]$ defined by $T(s,a,[s']_{B|S}) = \sum_{s'' \in [s']_{B|S}} P(s,a,s'')$. In other words, when applying action a in state s, $T(s,a,[s']_{B|S})$ is the probability that the resulting state is in the block $[s']_{B|S}$. It is clear that since $B|S$ is a partition of S, each of these block transition probabilities is in the interval $[0,1]$.

Example 1

Let $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ be an MDP with $S = \{s_1, s_2, s_3\}$, $A = \{a_1, a_2\}$ and $\Psi = \{(s_1,a_1), (s_1,a_2), (s_2,a_1), (s_2,a_2), (s_3,a_1)\}$. We give the projections under both our definition and the traditional one (see footnote).

[5] The more traditional definition of a projection is: $x \equiv_{B|X} x'$ if and only if $(x,y) \equiv_B (x',y)$ for all $y \in Y$. This projection is always a refinement of our projection. We need the modified definition to facilitate the development of some concepts below.

i) If $B_1 = \Big\{\{(s_1, a_1), (s_2, a_2)\}, \{(s_1, a_2), (s_2, a_1), (s_3, a_1)\}\Big\}$,
then $B_1|S = \Big\{\{s_1, s_2\},\ \{s_3\}\Big\}$ (ours); $\Big\{\{s_1\},\ \{s_2\}, \{s_3\}\Big\}$ (traditional).

ii) If $B_2 = \Big\{\{(s_2, a_1)\}, \{(s_1, a_1), (s_1, a_2), (s_2, a_2), (s_3, a_1)\}\Big\}$,
then $B_2|S = \Big\{\{s_1, s_3\},\ \{s_2\}\Big\}$; $\quad \Big\{\{s_1\},\ \{s_2\}, \{s_3\}\Big\}$.

iii) If $B_3 = \Big\{\{(s_1, a_1), (s_2, a_2)\}, \{(s_1, a_2), (s_3, a_1)\}, \{(s_2, a_1)\}\Big\}$,
then $B_3|S = \Big\{\{s_1\},\ \{s_2\}, \{s_3\}\Big\}$; $\quad \Big\{\{s_1\},\ \{s_2\}, \{s_3\}\Big\}$.

3 Related Work

There has been extensive work on minimization of FSAs [11]. Minimization techniques derive the "smallest" model that is equivalent to the given model. This simplifies the search for an efficient implementation. See Hartmanis and Stearns [11] for more details. Similar techniques exist for Probabilistic Automata [19], Probabilistic Transition Systems [17], Concurrent Processes [18, 7], Finite Markov Chains [15] and Markov Processes [22].

Dean, Givan and colleagues have explored minimization of MDPs in detail. Dean and Givan [5] introduce a framework for model minimization and explore its relation to some existing algorithms. They also give algorithms for finding reduced models of MDPs with special representations. They base their definition of equivalence on the notion of *homogeneous* partitions of the state set. This concept of equivalence is related to the *substitution property* of finite state machines [11] and the notion of *lumpability* of markov chains [15]. Givan et al. [9] explore minimization based on certain relaxed equivalence criteria, and Dean et al. [6] extend the framework to facilitate elimination of redundant actions. Givan et al. [8] formulate the model minimization problem in terms of *stochastic bisimulations* derived from the notion of bisimulations of concurrent processes [12, 17, 18] and establish all their previous results in this framework.

Minimization techniques frequently exploit symmetries of the underlying structure (e. g., see ref. [14] for FSAs, ref. [10] for Markov Chains and refs. [13, 7] from model checking for concurrent processes). But there has not been much work on exploiting symmetries of MDPs for minimization. Recently Zinkevich and Balch [24] defined symmetries of MDPs and derived algorithms that take advantage of such symmetries. But their work did not relate to the existing research on model minimization.

In this article we extend the model minimization framework of Dean and Givan to include symmetrical equivalence. This gives us additional power and sometimes enables greater reduction as outlined in the introduction. We base our framework on the notion of MDP homomorphisms derived from the concept of homomorphisms of FSAs. Traditionally symmetries are defined via groups of morphisms (e. g. ref. [16]) and hence employing homomorphisms makes it easier to include symmetries in our framework.

In the next section we present extensions of some of the key results in Givan et al. [8] using our framework. In Section 5 we define symmetries of MDPs using group theoretic concepts and show that our extended minimization framework can exploit symmetrical equivalence.

4 Homomorphisms and model minimization

In this section we extend the concept of *machine homomorphism* from the FSA literature to MDPs and develop a notion of equivalence of states and state-action pairs based on this extended homomorphism. Informally, a homomorphism of a system with transition dynamics is a transformation that preserves some aspects of the dynamics.

For example, consider two MDPs $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ and $\mathcal{M}' = \langle S', A, \Psi', P', R'\rangle$ that have deterministic actions. By abusing notation, we employ the shorthand $P(s, a)$ to denote s_1 in S, such that $P(s, a, s_1) = 1$. A map $f : S \to S'$ is a homomorphism if $P'(f(s), a) = f(P(s, a))$ and $R(s, a) = R'(f(s), a)$ for all $(s, a) \in \Psi$. The homomorphism f is said to *commute* with the dynamics of the MDPs. We can depict this using commutative diagrams as follows:

$$
\begin{tikzcd}
s_1 \arrow[r, mapsto, "P(\cdot, a)"] \arrow[d, mapsto, "f"'] & s_2 \arrow[d, mapsto, "f"] \\
s'_1 \arrow[r, mapsto, "P'(\cdot, a)"] & s'_2
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
s_1 \arrow[r, mapsto, "R(\cdot, a)"] \arrow[d, mapsto, "f"'] & \overline{r} \\
s'_1 \arrow[ur, mapsto, "R'(\cdot, a)"'] &
\end{tikzcd}
$$

Figure 3: Homomorphisms Represented by Commutative Diagrams

More generally a homomorphism from an MDP $\mathcal{M}$ to an MDP $\mathcal{M}'$ is a map from Ψ to Ψ' that commutes with the transition dynamics and preserves the reward function:

Definition: An *MDP homomorphism* h from an MDP $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ to an MDP $\mathcal{M}' = \langle S', A', \Psi', P', R'\rangle$ is a surjection from Ψ to Ψ', defined by a tuple of surjections $\langle f, \{g_s | s \in S\}\rangle$, with $h((s, a)) = (f(s), g_s(a))$, where $f : S \to S'$ and $g_s : A_s \to A'_{f(s)}$ for $s \in S$, such that:

$$
\begin{aligned}
P'(f(s), g_s(a), f(s')) &= T(s, a, [s']_{B_h|S}),\ \forall s, s' \in S, a \in A_s &\quad (1)\\
R'(f(s), g_s(a)) &= R(s, a),\ \forall s \in S, a \in A_s &\quad (2)
\end{aligned}
$$

We call $\mathcal{M}'$ the *homomorphic image* of $\mathcal{M}$ under h. We use the shorthand $h(s, a)$ to denote $h((s, a))$.

Let $P_{sa} : S \to [0, 1]$ be the distribution over states resulting from taking action a in state s, i. e., $P_{sa}(s_1) = P(s, a, s_1)$ for any s_1 in S. The *aggregation* hP_{sa}, of P_{sa} over h, is the distribution over S' such that $hP_{sa}(s') = \sum_{s_1 \in f^{-1}(s')} P_{sa}(s_1)$ for each $s' \in S'$. Here

$f^{-1}(s') = \{s \in S | f(s) = s'\}$ is the pre-image of s' in S. A homomorphism commutes with the one step dynamics of the MDP in the sense that the aggregation hP_{sa} is the same distribution as $P'_{f(s)g_s(a)}$ for all $(s,a) \in \Psi$. We can depict this using commutative diagrams as follows:

$$
\begin{tikzcd}
(s,a) \arrow[r, mapsto, "P"] \arrow[d, mapsto, "h"'] & P_{sa} \arrow[d, mapsto, "agg."] \\
(s',a') \arrow[r, mapsto, "P'"] & P'_{s'a'}
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
(s,a) \arrow[r, mapsto, "R"] \arrow[d, mapsto, "h"'] & \overline{r} \\
(s',a') \arrow[ur, mapsto, "R'"'] &
\end{tikzcd}
$$

Figure 4: An MDP Homomorphism as Commutative Diagrams

Apart from the preservation of block transition behaviour, the usefulness of homomorphisms lie in the fact that they help establish the following equivalences.

Definition: State action pairs (s_1, a_1) and $(s_2, a_2) \in \Psi$ are *equivalent* if $hP_{s_1a_1} = hP_{s_2a_2}$, i. e., the aggregation of their next state distributions is the same. Note that any h-equivalent state-action pairs are also equivalent in this sense.

Definition: States s_1 and $s_2 \in S$ are *equivalent* if for every action $a_1 \in A_{s_1}$, there is an action $a_2 \in A_{s_2}$ such that (s_1, a_1) and (s_2, a_2) are equivalent and for every action $a_2 \in A_{s_2}$, there is an action $a_1 \in A_{s_1}$, such that (s_1, a_1) and (s_2, a_2) are equivalent.

These notions of equivalence lead us to the following theorem on optimal value equivalence. This theorem is an extension of the optimal value equivalence theorem developed in Givan et al [8] for stochastic bisimulations.

Theorem 1: (*Optimal value equivalence*) Let $\mathcal{M}' = \langle S', A', \Psi', P', R' \rangle$ be the homomorphic image of the MDP $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ under the MDP homomorphism $h = \langle f, \{g_s | s \in S\} \rangle$. For any $(s,a) \in \Psi$, $Q^\star(s,a) = Q^\star(f(s), g_s(a))$.

Proof: (Along the lines of [8]) Let us define the m-step optimal discounted action value function recursively for all $(s,a) \in \Psi$ and for all non-negative integers m as

$$Q_m(s,a) = R(s,a) + \gamma \sum_{s_1 \in S} \left[P(s,a,s_1) \max_{a_1 \in A_{s_1}} Q_{m-1}(s_1, a_1) \right]$$

and set $Q_{-1}(s_1, a_1) = 0$. Letting $V_m(s_1) = \max_{a_1 \in A_{s_1}} Q_m(s_1, a_1)$, we can rewrite this as:

$$Q_m(s,a) = R(s,a) + \gamma \sum_{s_1 \in S} \left[P(s,a,s_1) V_{m-1}(s_1) \right].$$

Now we prove by induction on m that the theorem is true. For the base case of $m = 0$, we have that $Q_0(s,a) = R(s,a) = R'(f(s), g_s(a)) = Q_0(f(s), g_s(a))$. Now let us assume that $Q_j(s,a) = Q_j(f(s), g_s(a))$ for all values of j less than m and all state-action pairs in Ψ. Now we have,

$$\begin{aligned}
Q_m(s,a) &= R(s,a) + \gamma \sum_{s' \in S} P(s,a,s') V_{m-1}(s') \\
&= R(s,a) + \gamma \sum_{[s']_{B_h|S} \in B_h|S} T(s,a,[s']_{B_h|S}) V_{m-1}(s') \quad \text{(since h is a homomorphism)} \\
&= R'(f(s), g_s(a)) + \gamma \sum_{s' \in S'} P'(f(s), g_s(a), s') V_{m-1}(s') \qquad (\text{''}) \\
&= Q_m(f(s), g_s(a))
\end{aligned}$$

Since R is bounded it follows by induction that $Q^\star(s,a) = Q^\star(f(s), g_s(a))$ for all $(s,a) \in \Psi$. □

Corollary:

1. For any h-equivalent $(s_1, a_1), (s_2, a_2) \in \Psi$, $Q^\star(s_1, a_1) = Q^\star(s_2, a_2)$.
2. For all equivalent $s_1, s_2 \in S$, $V^\star(s_1) = V^\star(s_2)$.
3. For all $s \in S$, $V^\star(s) = V^\star(f(s))$.

Proof: Corollary 1 follows from Theorem 1. Corollaries 2 and 3 follow from Theorem 1 and the fact that $V^\star(s) = \max_{a \in A_s} Q^\star(s,a)$. □

The above theorem establishes optimal value equivalence. As shown by Givan et al. [8], this is not a sufficient notion of equivalence. In many cases even when the optimal values are equal, the policies might not be related and hence we cannot easily transform solutions of the image MDP to the original MDP. The optimal policies of an MDP and its homomorphic images are closely related and the following establishes the correspondence.

Definition: Let $\mathcal{M}'$ be the image of $\mathcal{M}$ under homomorphism h. For any $s \in S$, $g_s^{-1}(a')$ denotes the set of actions that have the same image $a' \in A'_{f(s)}$ under g_s. Let π be a stochastic policy in $\mathcal{M}'$. Then π *lifted to* $\mathcal{M}$ is the policy $\pi_{\mathcal{M}}$ such that for any $a \in g_s^{-1}(a')$, $\pi_{\mathcal{M}}(s,a) = \pi(f(s), a') \Big/ |g_s^{-1}(a')|$.
Note: It is sufficient if $\sum_{a \in g_s^{-1}(a')} \pi_{\mathcal{M}}(s,a) = \pi(f(s), a')$, but we use the above definition to make the lifted policy unique.

Example 2

Consider MDP $\mathcal{M}$ from example 1 and $\mathcal{M}' = \langle S', A', \Psi', P', R' \rangle$ with $S' = \{s'_1, s'_2\}$, $A' = \{a'_1, a'_2\}$ and $\Psi' = \{(s'_1, a'_1), (s'_1, a'_2), (s'_2, a'_1)\}$. Let $h = \langle f, \{g_s | s \in S\} \rangle$ be a surjection from $\mathcal{M}$ to $\mathcal{M}'$ defined by

$$\begin{array}{rrr}
f(s_1) = s'_1 & f(s_2) = s'_2 & f(s_3) = s'_2 \\
g_{s_1}(a_1) = a'_2 & g_{s_2}(a_1) = a'_1 & g_{s_3}(a_1) = a'_1 \\
g_{s_1}(a_2) = a'_1 & g_{s_2}(a_2) = a'_1 &
\end{array}$$

Let π be a policy in $\mathcal{M}'$ with

$$\pi(s_1', a_1') = 0.6 \qquad \pi(s_1', a_2') = 0.4 \qquad \pi(s_2', a_1') = 1.0$$

Now π lifted to $\mathcal{M}$, the policy $\pi_{\mathcal{M}}$, is derived as follows:

$$\begin{array}{ll} \pi_{\mathcal{M}}(s_1, a_1) = \pi(s_1', a_2') = 0.4 & \pi_{\mathcal{M}}(s_1, a_2) = \pi(s_1', a_1') = 0.6 \\ \pi_{\mathcal{M}}(s_2, a_1) = \pi(s_2', a_1')/2 = 0.5 & \pi_{\mathcal{M}}(s_2, a_2) = \pi(s_2', a_1')/2 = 0.5 \\ \pi_{\mathcal{M}}(s_3, a_1) = \pi(s_2', a_1') = 1.0 & \end{array}$$

Theorem 2: Let $\mathcal{M}' = \langle S', A', \Psi', P', R' \rangle$ be the image of $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ under the homomorphism $h = \langle f, \{g_s | s \in S\} \rangle$. If $\pi^\star$ is an optimal policy for $\mathcal{M}'$, then $\pi^\star_{\mathcal{M}}$ is an optimal policy for $\mathcal{M}$.

Proof: Let $\pi^\star$ be an optimal policy in $\mathcal{M}'$. Consider some $(s, a) \in \Psi$ such that $\pi^\star(f(s), g_{s_1}(a_1))$ is greater than zero. Then $Q^\star(f(s_1), g_{s_1}(a_1))$ is the maximum value of the $Q^\star$ function in state $f(s_1)$. From Theorem 1, we know that $Q^\star(s, a) = Q^\star(f(s), g_s(a))$ for all $(s, a) \in \Psi$. Therefore $Q^\star(s_1, a_1)$ is the maximum value of the $Q^\star$ function in state s_1. Thus a_1 is an optimal action in state s_1 and hence $\pi^\star_{\mathcal{M}}$ is an optimal policy for $\mathcal{M}$. $\square$

Theorem 2 establishes that an MDP can be solved by solving one of its homomorphic images. To achieve the most impact, we need to derive the smallest possible homomorphic image of the MDP, i. e., an image with the least number of admissible state-action pairs. The following definitions help formalize this notion.

Definition: An MDP $\mathcal{M}$ is a *minimal MDP* if for every MDP $\mathcal{M}'$ that is a homomorphic image of $\mathcal{M}$, there exists a homomorphism from $\mathcal{M}'$ to $\mathcal{M}$.

Definition: A *minimal image* of an MDP $\mathcal{M}$ is a homomorphic image of $\mathcal{M}$ that is also a minimal MDP.

A minimal image of an MDP $\mathcal{M}$ is the smallest MDP whose solution can be lifted to yield a solution to $\mathcal{M}$. Finding a minimal image is the goal of model minimization. Since this can be computationally prohibitive, we frequently settle for a reasonably reduced model, even if it is not a minimal MDP.

4.1 Homomorphisms and Partitions

As mentioned earlier any map on a set induces a partition of the set. Thus a homomorphism from $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ to $\mathcal{M}' = \langle S', A', \Psi', P', R' \rangle$ induces a partition on Ψ. Classical FSA literature employs such partitions of the state set in minimization of machines. There are various algorithms for identifying a suitable partition that gives rise to a reduced image of a machine. Dean and Givan [5] propose several such algorithms for MDP model minimization and demonstrate that they are effective in finding minimal images. The basic idea behind all these algorithms is to start with a very coarse partition satisfying some conditions and successively refine it until one obtains a suitable partition that can be induced by a homomorphism. In this section, we explore the relationship between partitions of Ψ and homomorphisms, and we establish conditions under which a partition corresponds to a homomorphism. We can then extend algorithms that identify suitable partitions of S to identify suitable partitions of Ψ.

Definition: A partition B of an MDP $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ is said to be *reward respecting* if $B_R \geq B$.[6] In other words B is reward respecting if $(s_1, a_1) \equiv_B (s_2, a_2)$ implies $R(s_1, a_1) = R(s_2, a_2)$ for all $(s_1, a_1), (s_2, a_2) \in \Psi$.

Definition: A partition B of an MDP $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ has the *stochastic substitution property* if for all $(s_1, a_1), (s_2, a_2) \in \Psi$, $(s_1, a_1) \equiv_B (s_2, a_2)$ implies $T(s_1, a_1, [s]_{B|S}) = T(s_2, a_2, [s]_{B|S})$ for all $[s]_{B|S} \in B|S$.

In other words, the block transition probability is the same for all state-action pairs in a given block. A partition that satisfies the stochastic substitution property is an *SSP partition.* This is an extension of the substitution property for finite state machines [11]. The *SSP block transition probability* is the function $T_b : B \times B|S \rightarrow [0, 1]$, defined by $T_b([(s_1, a_1)]_B, [s]_{B|S}) = T(s_1, a_1, [s]_{B|S})$. This quantity is well-defined only for SSP partitions.

Theorem 3: Let h be an MDP homomorphism from an MDP $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ to an MDP $\mathcal{M}' = \langle S', A', \Psi', P', R' \rangle$. Then B_h, the partition of Ψ induced by h, is a reward respecting SSP partition.

Proof: Let $h = \langle f, \{g_s | s \in S\} \rangle$ be the homomorphism from $\mathcal{M}$ to $\mathcal{M}'$. We need to show that the partition B_h is a reward respecting SSP partition.

First let us tackle the stochastic substitution property. Let $(s_1, a_1), (s_2, a_2) \in \Psi$, be h-equivalent. From the definition of a homomorphism we have that $f(s_1) = f(s_2) = s' \in S'$ and $g_{s_1}(a_1) = g_{s_2}(a_2) = a' \in A'_{s'}$. Thus, for any $s \in S$, $T(s_1, a_1, [s]_{B_h|S}) = P'(s', a', f(s)) = T(s_2, a_2, [s]_{B_h|S})$. Hence B_h is an SSP partition.

From condition 2 in the definition of a homomorphism, it is clear that the partition induced is reward respecting. $\square$

Theorem 3 establishes that the partition induced by a homomorphism is a reward respecting SSP partition. But the converse of the theorem, that for every reward respecting SSP partition there exists a homomorphism that induces it, is not true. The following examines how to construct a homomorphic image of an MDP given a reward respecting SSP partition.

Definition: Let B be a reward respecting SSP partition of MDP $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$. Let $\eta(s)$ be the number of distinct blocks of B that contain a state-action pair with s as the state component and let $\{[(s, a_i)]_B \, | i = 1, 2, \cdots, \eta(s)\}$ be the blocks. Note that if $[s_1]_{B|S} = [s_2]_{B|S}$ then $\eta(s_1) = \eta(s_2)$, hence the following is well-defined. The *quotient MDP* $\mathcal{M}/B$ is the MDP $\langle S', A', \Psi', P', R' \rangle$ where, $S' = B|S$; $A' = \bigcup_{[s]_{B|S} \in B|S} A'_{[s]_{B|S}}$ where

[6] Recall, B_R is the partition of Ψ induced by the reward function.

$A'_{[s]_{B|S}} = \{a'_1,\ a'_2,\ \cdots,\ a'_{\eta(s)}\}$ for each $[s]_{B|S} \in B|S$; P' is given by $P'([s]_{B|S}, a'_i, [s']_{B|S}) = T_b([(s, a_i)]_B, [s']_{B|S})$ and R' is given by $R'([s]_{B|S}, a'_i) = R(s, a_i)$.

Theorem 4: Let B be a reward respecting SSP partition of MDP $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$. There exists a homomorphism from $\mathcal{M}$ to the quotient MDP $\mathcal{M}/B$.

Proof: Given a reward respecting SSP partition B of $\mathcal{M}$, we show by construction that there exists a homomorphism h from $\mathcal{M}$ to the quotient MDP $\mathcal{M}/B = \langle S', A', \Psi', P', R'\rangle$.

The homomorphism $h = \langle f, \{g_s | s \in S\}\rangle$ between $\mathcal{M}$ and $\mathcal{M}/B$ is given by $f(s) = [s]_{B|S}$ and $g_s(a) = a'_i$ such that $T(s, a, [s']_{B|S}) = P'([s]_{B|S}, a'_i, [s']_{B|S})$ for all $[s']_{B|S} \in B|S$. In other words, if $[(s, a)]_{B|S}$ is the i-th unique block in the ordering used in the construction of $\mathcal{M}/B$, then $g_s(a) = a'_i$. It is easy to verify that h is indeed a homomorphism. □

The partition induced on $\mathcal{M}$ by h, is only guaranteed to be a refinement of B and is not always the same partition as B. In other words, $B \geq B_h$. In fact B_h is the least coarse partition such that $B_h|S = B|S$, and $\mathcal{M}/B$ is the same MDP as $\mathcal{M}/B_h$ up to a relabelling of states and actions.

Partitions and minimal images

As we said earlier model minimization algorithms work by finding suitable partitions of an MDP. As is evident now, by suitable partitions we mean reward respecting SSP partitions. Here we explore the relationship between reward respecting SSP partitions and minimal images of the MDPs

Definition: A partition B of an MDP $\mathcal{M}$ is the *coarsest* reward respecting SSP partition of $\mathcal{M}$ if and only if for every reward respecting SSP partition B' of $\mathcal{M}$, $B \geq B'$.

It is easy to verify (by contradiction) that there exists an *unique* coarsest reward respecting SSP partition for any MDP $\mathcal{M}$. Intuitively one would expect the quotient MDP corresponding to the coarsest reward respecting SSP partition of an MDP $\mathcal{M}$ to be a minimal image of $\mathcal{M}$. The following theorem states that formally.

Theorem 5: Let B be the coarsest reward respecting SSP partition of MDP $\mathcal{M}$. The quotient MDP $\mathcal{M}/B$ is a minimal image of $\mathcal{M}$.

Proof: We defer the proof of this theorem to the next section, after we define composition of homomorphisms.

Given an MDP $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ the outline of a basic model minimization algorithm is as follows:

1. Start with any reward respecting partition of Ψ. The most obvious choice is to pick the one that is induced by the expected reward function R. This is the coarsest possible reward respecting partition, but any suitable reward respecting partition will do.

2. Repeatedly refine the partition until all violations of the SSP property are resolved. This process might take as much time as solving the original MDP itself. Therefore most modifications of this basic algorithm focus on special representations of $\mathcal{M}$ that make this step simpler. Let B be the resulting partition.

3. Form the quotient MDP $\mathcal{M}/B$ and identify the homomorphism between $\mathcal{M}$ and $\mathcal{M}/B$.

Now one can solve $\mathcal{M}/B$ and lift the optimal policy to get an optimal policy for $\mathcal{M}$. Specific methods for refining the partitions can provide certain guarantees on the quality of the SSP partition derived. For example, see ref. [5] for a method that guarantees finding the coarsest reward respecting SSP partition.

5 Automorphisms and Symmetries

Recall the notion of symmetrical equivalence outlined in Section 3. That notion is a special case of the notion of equivalence we developed in the previous section. In this section we define symmetries using homomorphisms. We also borrow concepts from group theory to define groups of symmetries and show that considering such groups together can lead to a greater reduction in problem size. This is a special case of our earlier framework and unifies the concepts of model minimization and exploiting symmetries.

Definition: An MDP homomorphism $h = \langle f, \{g_s | s \in S\}\rangle$ from MDP $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ to MDP $\mathcal{M}' = \langle S', A', \Psi', P', R'\rangle$ is an *MDP isomorphism* from $\mathcal{M}$ to $\mathcal{M}'$ if and only if f and g_s, $s \in S$, are bijective. $\mathcal{M}$ is said to be *isomorphic* to $\mathcal{M}'$ and vice versa.

Note that property (1) of a homomorphism reduces to a simpler form in this case: $P(s, a, s') = P'(f(s), g_s(a), f(s'))$ for all $s, s' \in S$ and $a \in A_s$. Therefore, when two MDPs are isomorphic, it means that the MDPs are the same except for a relabelling of the states and the actions. Thus we can transfer policies learned for one MDP to the other by simple transformations. Also note that an MDP $\mathcal{M}$ is a minimal MDP if it is isomorphic to all of its homomorphic images.

Definition: An MDP isomorphism from an MDP $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ to itself is an *automorphism* of $\mathcal{M}$.

Intuitively one can see that automorphisms can be used to describe symmetries in a problem specification. In the gridworld example of Figure 2 a reflection of the states along the NE-SW diagonal and a swapping of actions N and E and of actions S and W is an automorphism. It is easy to see that this remapping captures the symmetry that we discussed earlier. When we consider all such symmetries together we achieve greater reduction in the size of an MDP.

Let the set of all automorphisms of an MDP $\mathcal{M}$ be denoted by Aut$\mathcal{M}$. This set forms a group under composition of homomorphisms. This group is the symmetry group of $\mathcal{M}$. Let $\mathcal{G}$ be a subgroup of Aut$\mathcal{M}$ denoted by $\mathcal{G} \leq$ Aut$\mathcal{M}$.

The subgroup $\mathcal{G}$ defines an equivalence relation $\equiv_{\mathcal{G}}$ on Ψ: $(s_1, a_1) \equiv_{\mathcal{G}} (s_2, a_2)$ if and only if there exists $h \in \mathcal{G}$ such that $h(s_1, a_1) = (s_2, a_2)$. Note that since $\mathcal{G}$ is a subgroup, this implies that there exists an $h^{-1} \in \mathcal{G}$ such that $h^{-1}(s_2, a_2) = (s_1, a_1)$. Let $B_{\mathcal{G}}$ be the partition of Ψ induced by $\equiv_{\mathcal{G}}$.

Lemma: For any $h = \langle f, \{g_s | s \in S\}\rangle \in \mathcal{G}$, $f(s) \in [s]_{B_{\mathcal{G}}|S}$.
Proof: The lemma follows from the properties of groups, namely closure and existence of an inverse. □

Theorem 6: Let $\mathcal{G} \leq \mathrm{Aut}\mathcal{M}$ be a group of automorphisms on $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$. The partition $B_{\mathcal{G}}$ is a reward respecting SSP partition of $\mathcal{M}$.

Proof: Consider $(s_1, a_1), (s_2, a_2) \in \Psi$ such that $(s_1, a_1) \equiv_{\mathcal{G}} (s_2, a_2)$. This implies that there exists an $h = \langle f, \{g_s | s \in S\}\rangle$ in $\mathcal{G}$ such that $f(s_1) = s_2$ and $g_{s_1}(a_1) = a_2$.

From the definition of an automorphism we have that for any $s \in S$, $P(s_1, a_1, s) = P(s_2, a_2, f(s))$. Using the lemma, we have $\sum_{s' \in [s]_{B_{\mathcal{G}}|S}} P(s_1, a_1, s') = \sum_{s' \in [s]_{B_{\mathcal{G}}|S}} P(s_2, a_2, s')$. Since we chose s arbitrarily, this holds for all s in S. Hence $B_{\mathcal{G}}$ is an SSP partition.

Again from the definition of an automorphism we have that $R(s_1, a_1) = R(s_2, a_2)$. Hence $B_{\mathcal{G}}$ is reward respecting too. □

Corollary: There exists a homomorphism $h_{\mathcal{G}}$ from $\mathcal{M}$ to $\mathcal{M}/B_{\mathcal{G}}$. We call $\mathcal{M}/B_{\mathcal{G}}$ the $\mathcal{G}$-reduced image of $\mathcal{M}$.
This follows from Theorems 4 and 6. □

Corollary: An optimal policy for $\mathcal{M}/B_{\mathcal{G}}$ lifted to $\mathcal{M}$ is an optimal policy for $\mathcal{M}$.
This follows from the above corollary and Theorem 2. □

Note that the converse of Theorem 6 is not true. It is possible to define SSP partitions that are not generated by groups of automorphisms. We give an example in the next section. Frequently the $\mathrm{Aut}\mathcal{M}$-reduced image of an MDP $\mathcal{M}$ is a minimal image of $\mathcal{M}$, as in the example in the next section. Even when we employ some $G < \mathrm{Aut}\mathcal{M}$ we get useful reductions. Thus model reduction can also be accomplished by finding the symmetry group of an MDP.

Proof of Theorem 5

Definition: Let $h = \langle f, \{g_s | s \in S\}\rangle : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ and $h' = \langle f', \{g'_s | s \in S\}\rangle : \mathcal{M}_2 \rightarrow \mathcal{M}_3$ be two MDP homomorphisms. The *composition* of h and h' denoted by $h \circ h'$ is a map from $\mathcal{M}_1$ to $\mathcal{M}_3$, with $(h \circ h')(s, a) = h'(h(s, a)) = \left(f'(f(s)), g'_{f(s)}(g_s(a))\right)$ for all $(s, a) \in \Psi$. It can be shown that $h \circ h'$ is a homomorphism from $\mathcal{M}_1$ to $\mathcal{M}_3$.

Theorem 5: Let B be the coarsest reward respecting SSP partition of MDP $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$. The quotient MDP $\mathcal{M}/B$ is a minimal image of $\mathcal{M}$.

Proof: We will prove this by proving the contrapositive: if $\mathcal{M}/B$ is not a minimal image of

$\mathcal{M}$, then B cannot be the coarsest reward respecting SSP partition of $\mathcal{M}$.

Let h be the homomorphism from $\mathcal{M}$ to $\mathcal{M}/B$. If $\mathcal{M}/B$ is not a minimal MDP, then there exists a homomorphism h' (that is not an isomorphism) from $\mathcal{M}/B$ to some MDP $\mathcal{M}'$. Therefore there exists a homomorphism $(h \circ h')$ from $\mathcal{M}$ to $\mathcal{M}'$. From the definition of composition, it is evident that $B_h < B_{(h \circ h')}$.

We need to show that B is not coarser than $B_{(h \circ h')}$. In other words we need to show that either $B < B_{(h \circ h')}$ or they are not comparable. From the construction of a quotient MDP it is clear that $B_h|S = B|S$ since we use $B|S$ as the states of $\mathcal{M}/B$. Since $\mathcal{M}'$ is a homomorphic image of $\mathcal{M}/B$ but is not isomorphic to it, either (i) $\mathcal{M}'$ has fewer states than $\mathcal{M}/B$ or (ii) some states in $\mathcal{M}'$ have fewer actions than $\mathcal{M}/B$. In case (i) we have that $B|S < B_{(h \circ h')}|S$. We know that this implies that B is not coarser than $B_{(h \circ h')}$. In case (ii) we have that $B|S = B_{(h \circ h')}|S$. Let $[s]_B \, (= [s]_{B_{(h \circ h')}})$ be a state with fewer admissible actions in $\mathcal{M}'$. This implies that s appears in fewer unique blocks in $B_{(h \circ h')}$ than in B. Thus $B < B_{(h \circ h')}$. Therefore B is not the coarsest reward respecting SSP partition. Hence $\mathcal{M}/B$ is a minimal image if B is the coarsest reward respecting partition of $\mathcal{M}$. □

6 An Example

In this section we work out a slightly detailed example.

Consider the MDP $\mathcal{M} = \langle S, A, \Psi, P, R \rangle$ with $S = \{s_1, s_2, s_3, s_4\}$, $A = \{a_1, a_2\}$, $\Psi = S \times A$, P and R defined as follows:
$P(s_i, a_1, s_j)$ is given by the entry in the i-th row and j-th column of:

	s_1	s_2	s_3	s_4
s_1	0	0.8	0.2	0
s_2	0.2	0	0	0.8
s_3	0.8	0	0	0.2
s_4	0	0	0	1.0

and $P(s_i, a_2, s_j)$ is given by:

	s_1	s_2	s_3	s_4
s_1	0	0.2	0.8	0
s_2	0.8	0	0	0.2
s_3	0.2	0	0	0.8
s_4	0	0	0	1.0

$R(s_2, a_1) = R(s_3, a_2) = 0.8$ and $R(s_2, a_2) = R(s_3, a_1) = 0.2$. For all other values of i and j, $R(s_i, a_j)$ equals zero. Figure 5 gives the transition graph of $\mathcal{M}$.

Consider the partition B of $\mathcal{M}$ given by $B = \Big\{\{(s_1, a_1), (s_1, a_2)\},\ \{(s_2, a_1),\ (s_3, a_2)\},$ $\{(s_2, a_2), (s_3, a_1)\},\ \{(s_4, a_1), (s_4, a_2)\}\Big\}$. B is a reward respecting SSP partition. We can derive the quotient MDP $\mathcal{M}/B = \langle S', A', \Psi', P', R' \rangle$ as follows:
$S' = B|S = \Big\{\{s_1\}, \{s_2, s_3\}, \{s_4\}\Big\}$ are the states of $\mathcal{M}/B$.

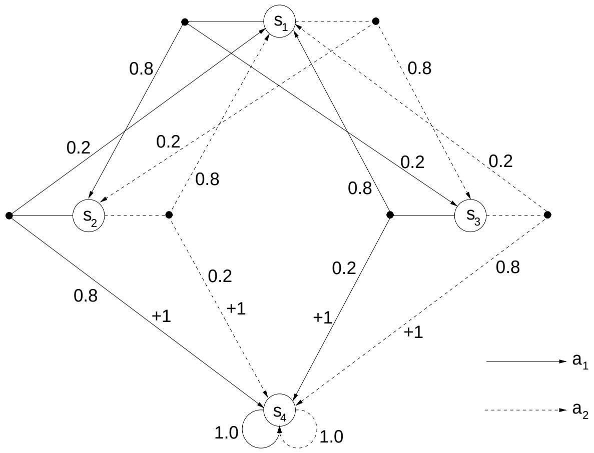

Figure 5: Transition graph of example MDP $\mathcal{M}$

Now, $\eta(s_1) = 1$, $\eta(s_2) = \eta(s_3) = 2$ and $\eta(s_4) = 1$. Hence we set $A'_{\{s_1\}} = \{a'_1\}$, $A'_{\{s_2,s_3\}} = \{a'_1, a'_2\}$ and $A'_{\{s_4\}} = \{a'_1\}$.
Now $P'(\{s_1\}, a'_1, \{s_2, s_3\}) = P(s_1, a_1, s_2) + P(s_1, a_1, s_3) = P(s_1, a_2, s_2) + P(s_1, a_2, s_3) = 1.0$. Proceeding similarly, we have

$$
\begin{array}{ll}
P'(\{s_1\}, a'_1, \{s_2, s_3\}) = 1.0 & P'(\{s_4\}, a'_1, \{s_4\}) = 1.0 \\
P'(\{s_2, s_3\}, a'_1, \{s_1\}) = 0.8 & P'(\{s_2, s_3\}, a'_2, \{s_1\}) = 0.2 \\
P'(\{s_2, s_3\}, a'_1, \{s_4\}) = 0.2 & P'(\{s_2, s_3\}, a'_2, \{s_4\}) = 0.8
\end{array}
$$

The probability of the each of the other transitions is zero. $R'(\{s_2, s_3\}, a'_1) = 0.2$, $R'(\{s_2, s_3\}, a'_2) = 0.8$ and all other rewards are zero. Figure 6 shows the transition graph for $\mathcal{M}/B$.

One can define a homomorphism $\langle f, \{g_s | s \in S\}\rangle$ from $\mathcal{M}$ to $\mathcal{M}/B$ as follows: $f(s_1) = \{s_1\}$, $f(s_2) = \{s_2, s_3\}$, $f(s_3) = \{s_2, s_3\}$ and $f(s_4) = \{s_4\}$. $g_{s_1}(a_i) = g_{s_4}(a_i) = a'_1$, for $i = 1, 2$, $g_{s_2}(a_1) = g_{s_3}(a_2) = a'_2$ and $g_{s_2}(a_2) = g_{s_3}(a_1) = a'_1$.

Let $\mathcal{I}$ be the identity map on Ψ and let h be an automorphism on $\mathcal{M}$ defined by: $h(s_1, a_1) = (s_2, a_2)$, $h(s_2, a_1) = (s_3, a_2)$, $h(s_2, a_2) = (s_3, a_1)$ and $h(s_4, a_1) = (s_4, a_2)$. The set of all automorphisms is given by $\mathrm{Aut}\mathcal{M} = \{\mathcal{I}, h\}$ and with the composition operator is the symmetry group of $\mathcal{M}$. It is easy to see that $B_{\mathcal{G}} = B$. Hence the $\mathcal{M}/B$ is the $\mathcal{G}$-reduced image of $\mathcal{M}$. $\mathcal{M}/B$ is also the minimal image of $\mathcal{M}$.

Consider the partition $B_1 = \Big\{\{(s_1, a_1)\}, \{(s_1, a_2)\}, \{(s_2, a_1), (s_3, a_2)\}, \{(s_2, a_2), (s_3, a_1)\}, \{(s_4, a_1)\}, \{(s_4, a_2)\}\Big\}$. B_1 is also a reward respecting SSP partition, but is not generated by any group of automorphisms on $\mathcal{M}$.

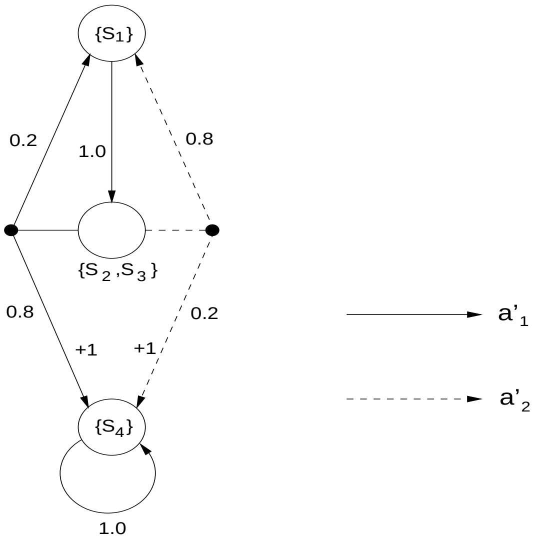

Figure 6: Transition graph of reduced MDP $\mathcal{M}/B$

7 Special forms of Homomorphisms

In some special cases we can study simpler transformations of an MDP that give rise to useful reduced images. In this section, we discuss some special forms of homomorphisms.

If there exists an isomorphism from MDP $\mathcal{M}$ to MDP $\mathcal{M}'$, then they are the same except for a relabelling of states and actions. Frequently the relabelling of actions is independent of the states. In such cases one can consider a simpler definition of a homomorphism as an ordered pair of surjections. Thus a homomorphism h from $\mathcal{M} = \langle S, A, \Psi, P, R\rangle$ to $\mathcal{M}' = \langle S', A', \Psi', P', R'\rangle$ is defined by $\langle f, g\rangle$ where $f : S \to S'$ and $g : A \to A'$. h still needs to satisfy both conditions (1) and (2) of a homomorphism. We assume that in such scenarios each state has the same set of actions admissible in it, i. e., $\Psi = S \times A$.

For example consider the symmetric gridworld example from Section 3. That world is isomorphic to problems with the goal in any of the other corners. If the goal moves from the NE corner to the SE corner, then an isomorphism between the two problems maps the states in the bottom half of the grid to those in the top half and vice versa. Action N goes to S and vice versa. Actions W and E are mapped onto themselves. This certainly is a simpler description than giving action maps for each of the 25 states.

Another interesting specialization is the case of *state homomorphisms*. When the actions admissible in a state and its homomorphic image are the same, i. e., $A_s = A'_{f(s)}$ for all $s \in S$, we can consider homomorphisms with $g_s(a) = a$ for all s. Thus a homomorphism h reduces to just a surjection on states f. This is the case widely studied in model minimization literature. This simplifies the derivation of a reduced image. As Dean and Givan [5] show, it is still a hard problem to derive a minimal image and frequently we have to settle for some reduced image.

This formulation of a homomorphism is sufficient for a large class of problems. But (*full*) homomorphisms as we define them in Section 4 are more powerful and enable greater reduction in MDP size. For example, in the previous section, if we had restricted ourselves to state homomorphisms, the given MDP $\mathcal{M}$ is a minimal MDP. Also certain symmetries such as rotational and reflectional symmetry, which are not captured by state homomorphisms, are captured by (full) homomorphisms.

As mentioned earlier, given a partition, it is a very hard task to identify and refine violations of the SSP property. To make this task easier one can employ different representations of the MDPs. One such method is to use *factored representations* as in refs. [5, 6]. Here the states of the MDP are represented by using various *features*. For example, a gridworld MDP might be represented by the x and y co-ordinates rather than a grid number. With factored representations, one can study partitions that result from projections on to one or more of the features in the cross product. Though this restricts the class of partitions that we examine, it sometimes makes it easier to check for violations of the SSP property. Dean and Givan [5] show that such restrictions lead to useful algorithms.

8 Discussion

In this article, we extended the model minimization framework of Givan and Dean to enable greater reduction in problem size. Givan et al. [8] consider two states equivalent if every action admissible in one state is admissible in the other and is equivalent. We extend the notion of equivalence so that two states are considered equivalent if for every action available in one state there is *some* equivalent action available in the other state.

Givan et al. [8] examined other notions of equivalence from existing literature before adopting stochastic bisimulations. For example, one such notion from FSA literature is action sequence equivalence. Two machines are considered equivalent if they produce the same sequence of output symbols given the same sequence of input symbols and the same starting state. In an MDP framework, this would translate as MDPs having the same distribution over sequences of rewards received given the same sequence of actions. This is not a sufficient notion of equivalence for MDPs, since we are interested in equivalence of policies and not just sequences of actions. See ref. [8] for an example where MDPs that are action-sequence equivalent have different optimal values.

MDP homomorphisms can be viewed as a form of stochastic bisimulations employed by Givan et al. [8] but they are a more basic concept. Stochastic bisimulation are defined via relations between sets and hence they have a greater expressive power than homomorphisms that are based on surjections. Despite this greater power, one can show that there exists a stochastic bisimulation between two MDPs if and only if they have a common minimal image. Thus, from the view point of model minimization, the same reductions are achievable with both formulations.

Givan et al. [8] also outline several methods for arriving at reward respecting SSP partitions. It should be trivially possible to extend those methods to our extended definitions.

It is also possible to extend their results on structured state spaces. We are working on this presently. Dean and Givan [5] show that model reduction algorithms such as state-space abstraction [3] and structured policy iteration [4] are special cases of model minimization. These results also hold for our extended definition. In fact it is possible to show that a larger class of algorithms fit into our general framework. We outline one such example next.

Zinkevich and Balch [24] define special classes of symmetries of MDPs and develop algorithms for taking advantage of such symmetries by copying values among symmetrically equivalent state-action pairs. Their notion of symmetries is based on equivalence relations on state-action pairs and can be shown to be a special case of our definition. Their algorithm can then be viewed as a special form of model minimization.

The insight that symmetries give rise to reward respecting SSP partitions gives us another way to look for such partitions. One can start from obvious symmetries in a problem and find their closure to generate suitable partitions. In some cases, especially that of spatial problems, it is possible to define the resulting homomorphism $h_{\mathcal{G}}$, and hence the reduced image, without explicitly finding $\mathcal{G}$.

Finding representations that exploit symmetries have always been a challenging problem [1]. Combining model minimization with symmetries gives us some guidance in this direction. By examining the form of the homomorphism one can suitably modify representations so as to make it easier to derive the quotient MDP. This in turn simplifies the solution process. Again consider the symmetrical gridworld in Figure 2. As we discussed earlier, the gridworld is symmetrical around the NE-SW diagonal. If we adopt a scheme that assigns the same representation to states that are symmetrical then we simplify the learning process. One such scheme is to represent each square by the horizontal and vertical projections on the NE-SW diagonal. Actions also should be represented with respect to the diagonal. This representation cuts the state space roughly in half. The resulting MDP can be shown to be isomorphic to that in Figure 2 and is in fact a minimal MDP.

Even when partitions of MDPs do not satisfy the SSP property exactly, sometimes they satisfy some relaxation of it. Givan et al. [9] study model minimization with a weaker criterion. The quotient MDP derived under this weaker condition is a Bounded Parameter MDP where the transition probabilities are given by an interval. Analogously we would like to develop a concept of approximate homomorphisms and approximate symmetries that would let us apply our ideas to a still larger class of problems.

Acknowledgements

We wish to thank Dan Bernstein for many hours of useful discussion; Amy McGovern and Dan Bernstein for commenting on drafts of this report; and Bob Givan and Matt Greig for clarifying certain ideas from their work. This material is based upon work supported by the National Science Foundation under Grant No. ECS-9980062. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Amarel, S. 1968. On representations of problems of reasoning about actions. In *Machine Intelligence 3*, Michie, D. (Ed.), pp. 131-171. Edinburgh Press; reprinted in *Readings in Artificial Intelligence*, Webber, B. L. and Nilsson, N. J. (Eds.), Tioga, 1981.

[2] Bertsekas, D. P. 1987. *Dynamic Programming: Deterministic and Stochastic Models.* Prentice-Hall, Englewood Cliffs, NJ.

[3] Boutilier, C. and Dearden, R. 1994. Using abstractions for decision theoretic planning with time constraints. In *Proceedings of the AAAI-94*, pp. 1016-1022. AAAI

[4] Boutilier, C., Dearden, R. and Goldszmidt, M. 1995. Exploiting structure in policy construction. In *Proceedings of IJCAI 14*, pp. 1104-1111.

[5] Dean, T. and Givan, R. 1997. Model minimization in Markov Decision Processes. In *Proceedings of AAAI-97.*

[6] Dean, T., Givan, R. and Kim, K-E. 1998. Solving planning problems with large state and action spaces. In *The Fourth International Conference on Artificial Intelligence Planning Systems.*

[7] Emerson, E. A. and Sistla, A. P. 1996. Symmetry and model checking. In *Fifth International Conference on Computer Aided Verification*, Crete, Greece.

[8] Givan, R., Dean, T. and Greig, M. 2001. Equivalence notions and model minimization in Markov Decision Processes. Submitted to *Artificial Intelligence.*

[9] Givan, R., Leach, S. and Dean, T. 1997. Bounded parameter Markov Decision Processes. Technical Report CS-97-05, Brown University, Providence, RI.

[10] Glover, J. 1991. Symmetry groups and translation invariant representations of markov processes. In *The Annals of Probability*, Vol 19, N0. 2, pp. 562 - 586.

[11] Hartmanis, J. and Stearns, R. E. 1966. *Algebraic Structure Theory of Sequential Machines.* Prentice-Hall, Englewood Cliffs, NJ.

[12] Hennessy, M. and Milner, R. 1985. Algebraic laws for nondeterminism and concurrency. In *Journal of the Association for Computing Machinery*, Vol. 32, No. 1, pp. 137 - 161.

[13] Ip, C. N. and Dill, D. L. 1993. Better verification through symmetry. In *Proceedings of the 11th International Symposium on Computer Hardware Description Languages.*

[14] Jump, J. R. 1969. A note on the iterative decomposition of finite automata. In *Information and Control*, 15: 424-435.

[15] Kemeny, J. G. and Snell, J. L. 1960. *Finite Markov Chains.* Van Nostrand, Princeton, NJ.

[16] Lang, S. 1967. *Algebraic Structures.* Addison Wesley, Reading, MA.

[17] Larsen, K. G. and Skou, A. 1991. Bisimulation through probabilistic testing. In *Information and Computation*, 94(1), pp. 1 - 28. Academic Press.

[18] Lee, D. and Yannakakis, M. 1992. Online minimization of transition systems. In *Proceedings of* 24^{th} *Annual ACM Symposium on the Theory of Computing.*

[19] Paz, A. 1971. *Introduction to Probabilistic Automata.* Academic Press, New York, NY.

[20] Popplestone, R. and Grupen R. 2000. Symmetries in World Geometry and Adaptive System Behaviour. In *Proceedings of the 2nd International Workshop on Algebraic Frames for the Perception-Action Cycle (AFPAC 2000)*, September 10-11, 2000, Kiel, Germany.

[21] Puterman, M. L. 1994. *Markov Decision Processes.* Wiley, New York, NY.

[22] Sharpe, M. J. 1988. *General Theory of Markov Processes.* Academic, San Diego, CA.

[23] Sutton, R. S. and Barto, A. G. 1998. *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA. RL book

[24] Zinkevich, M. and Balch, T. 2001. Symmetry in markov decision processes and its implications for single agent and multi agent learning. In *Proceedings of the 18th International Conference on Machine Learning*, Williamstown, Massachusetts, pp. 632-640. Morgan Kaufmann, San Francisco, CA.