Probabilistic Verification of Uncertain Systems Using Bounded-Parameter Markov Decision Processes

Di Wu and Xenofon Koutsoukos

Institute for Software Integrated Systems
Department of Electrical Engineering & Computer Science
Vanderbilt University, Nashville, TN 37235, USA
{Di.Wu, Xenofon.Koutsoukos}@Vanderbilt.Edu

Abstract. Verification of probabilistic systems is usually based on variants of Markov processes. For systems with continuous dynamics, Markov processes are generated using discrete approximation methods. These methods assume an exact model of the dynamic behavior. However, realistic systems operate in the presence of uncertainty and variability and they are described by uncertain models. In this paper, we address the problem of probabilistic verification of uncertain systems using Bounded-parameter Markov Decision Processes (BMDPs). Proposed by Givan, Leach and Dean [1], BMDPs are a generalization of MDPs that allow modeling uncertainty. In this paper, we first show how discrete approximation methods can be extended for modeling uncertain systems using BMDPs. Then, we focus on the problem of maximizing the probability of reaching a set of desirable states, we develop a iterative algorithm for probabilistic verification, and we present a detailed mathematical analysis of the convergence results. Finally, we use a robot path-finding application to demonstrate the approach.

1 Introduction

Verification of probabilistic systems is usually based on variants of Markov processes. For systems with continuous dynamics, Markov processes are generated using discrete approximation methods. Probabilistic verification aims at establishing bounds on the probabilities of certain events. Typical problems include the maximum and the minimum probability reachability problems, where the objective is to compute the control policy that maximizes the probability of reaching a set of desirable states, or minimize the probability of reaching an unsafe set. Algorithms for verification of MDPs have been presented in [2, 3]. These methods assume an exact model of the dynamic behavior for defining the transition probabilities. However, realistic systems operate in the presence of uncertainty and variability and they are described by uncertain models. Existing verification methods are insufficient for dealing with such uncertainty.

In this paper, we address the problem of probabilistic verification of uncertain systems using Bounded-parameter Markov Decision Processes (BMDPs). Proposed by Givan, Leach and Dean [1], BMDPs are a generalization of MDPs that allows modeling uncertainty. A BMDP can be viewed as a set of exact MDPs (sharing the same state and action space) specified by intervals of transition probabilities and rewards. Policies are compared on the basis of interval value functions. Optimistic and pessimistic criteria for optimality are used to define partial order relations between pairs of interval value functions.

Our approach is motivated by a robot path-finding application. Under uncertainty, the motion of the robot can be described by a set of stochastic differential equations with uncertain parameters. We show how discrete approximation methods can be extended for modeling such uncertain systems using BMDPs. Although we focus on a robotic example, the approach can be extended for probabilistic verification of stochastic hybrid (discrete-continuous) systems [4].

The paper focuses on the problem of maximizing the probability of reaching a set of desirable states. We develop a iterative algorithm for probabilistic verification, and we present a detailed mathematical analysis of the convergence results. The results presented in [1] are for dynamic programming methods assuming a discounted reward criterion. A discount factor of less than 1 ensures the convergence of the iterative methods for the interval value functions. Probabilistic verification for BMDPs can be formulated based on the Expected Total Reward Criterion (ETRC) for MDPs [5]. Under ETRC, the discount factor is set to 1, and the convergence of the iterative algorithms for BMDPs is more involved because the iteration operators are no longer contraction mappings. Furthermore, the interval value function may be not well defined unless proper restrictions on the intervals of transition probabilities and rewards are applied. Based on the ETRC, we solve the maximum probability reachability problems for BMDPs. Finally, we demonstrate our approach using a robot path-finding application.

Variants of uncertain MDPs have been also studied in [6–9]. These approaches consider a discounted reward. In addition, the work [10] considers an average performance criterion. Probabilistic verification of uncertain systems is a significant problem which requires an undiscounted criterion and cannot be treated with these algorithms.

In the next section, we review the basic notions of BMDPs. In Section 3, we explain how we can use BMDPs to model uncertain systems. In Section 4, we formulate and solve the maximum probability reachability problem for BMDPs. In Section 5, we present a robot path-finding application to demonstrate our approach. We close with conclusions and some future directions in Section 6.

2 Bounded-parameter Markov Decision Processes

We first review some basic notions of BMDPs, establish the notation that we use in this paper, and briefly summarize the main results in [1].

A BMDP is a four tuple $\mathcal{M} = \langle \mathcal{Q}, \mathcal{A}, \hat{F}, \hat{R} \rangle$ [1] where $\mathcal{Q}$ is a set of states, $\mathcal{A}$ is a set of actions, $\hat{R}$ is an interval reward function that maps each $q \in \mathcal{Q}$ to a

[1] In this paper, we use $\hat{X}$ to denote an interval, i.e. $\hat{X} = [\underline{X}, \overline{X}] \subseteq \mathbb{R}$.

closed interval of real values $[\underline{R}(q), \overline{R}(q)]$, and $\hat{F}$ is an interval state-transition distribution so that for $p, q \in \mathcal{Q}$ and $\alpha \in \mathcal{A}$,

$$\underline{F}_{p,q}(\alpha) \leqslant Pr(X_{t+1} = q | X_t = p, U_t = \alpha) \leqslant \overline{F}_{p,q}(\alpha).$$

For any action α and state p, the sum of the lower bounds of $\hat{F}_{p,q}(\alpha)$ over all states q is required to be less than or equal to 1, while the sum of the upper bounds is required to be greater than or equal to 1.

A BMDP $\mathcal{M}$ defines a set of exact MDPs. Let $M = \langle \mathcal{Q}', \mathcal{A}', F', R' \rangle$ be an MDP. If $\mathcal{Q} = \mathcal{Q}'$, $\mathcal{A} = \mathcal{A}'$, $R'(p) \in \hat{R}(p)$ and $F'_{p,q}(\alpha) \in \hat{F}_{p,q}(\alpha)$ for any $\alpha \in \mathcal{A}$ and $p, q \in \mathcal{Q}$, then we say $M \in \mathcal{M}$. To simplify discussions, in the following paragraphs the rewards are assumed to be tight, i.e. a single value. The approaches in this paper can be easily generalized to the case of interval rewards.

A policy is a mapping from states to actions, $\pi : \mathcal{Q} \rightarrow \mathcal{A}$. We use $\mathit{\Pi}$ to denote the set of stationary Markov policies. The policy in this paper is restricted to be in $\mathit{\Pi}$. For any policy π and state p, the interval value function of π at p is a closed interval defined as

$$\hat{V}_\pi(p) = [\min_{M \in \mathcal{M}} V_{M,\pi}(p), \max_{M \in \mathcal{M}} V_{M,\pi}(p)]$$

where

$$V_{M,\pi}(p) = R(p) + \gamma \sum_{q \in \mathcal{Q}} F^M_{p,q}(\pi(p)) V_{M,\pi}(q)$$

where $0 < \gamma < 1$ is called the discount factor.

An MDP $M \in \mathcal{M}$ is π-maximizing if for any $M' \in \mathcal{M}$, $V_{M,\pi} \geq_{dom} V_{M',\pi}$[2]. Likewise, M is π-minimizing if for any $M' \in \mathcal{M}$, $V_{M,\pi} \leq_{dom} V_{M',\pi}$. For any policy $\pi \in \mathit{\Pi}$, there exist a π-maximizing MDP $\overline{M}(\pi)$ and a π-minimizing MDP $\underline{M}(\pi)$ in $\mathcal{M}$.

The interval policy evaluation operator $\widehat{IVI}_\pi$ for each state p is defined as

$$\widehat{IVI}_\pi(\hat{V})(p) = [\underline{IVI}_\pi(\underline{V})(p), \overline{IVI}_\pi(\overline{V})(p)]$$

where

$$\underline{IVI}_\pi(\underline{V}) = \min_{M \in \mathcal{M}} VI_{M,\pi}(\underline{V}) = VI_{\underline{M}(\pi),\pi}(\underline{V}),$$

$$\overline{IVI}_\pi(\overline{V}) = \max_{M \in \mathcal{M}} VI_{M,\pi}(\overline{V}) = VI_{\overline{M}(\pi),\pi}(\overline{V})$$

and $VI_{M,\pi} : \mathcal{V} \rightarrow \mathcal{V}$ is the policy evaluation operator for the exact MDP M and policy π

$$VI_{M,\pi}(V)(p) = R(p) + \gamma \sum_{q \in \mathcal{Q}} F^M_{p,q}(\pi(p)) V(q).$$

To define the π-minimizing MDP $\underline{M}(\pi)$, we only need to compute its transition function $F_{\underline{M}(\pi)}$. Let q_1, q_2, ..., q_k ($k = |\mathcal{Q}|$) be an ordering of $\mathcal{Q}$ so that $\underline{V}(q_i) \leqslant \underline{V}(q_j)$ for any $1 \leqslant i < j \leqslant k$. Let r be the index $1 \leqslant r \leqslant k$ which

[2] $V \geq_{dom} U$ if and only if for all $q \in \mathcal{Q}$, $V(q) \geqslant U(q)$.

maximizes $\sum_{i=1}^{r-1}\overline{F}_{p,q_i}(\alpha)+\sum_{i=r}^{k}\underline{F}_{p,q_i}(\alpha)$ without letting it exceed 1. Then the transition function of the π-minimizing MDP $\underline{M}(\pi)$ is given by

$$F_{p,q_j}^{\underline{M}(\pi)}(\alpha)=\begin{cases}\overline{F}_{p,q_j}(\alpha) & \text{if } j<r\\ \underline{F}_{p,q_j}(\alpha) & \text{if } j>r\end{cases} \quad\text{and}\quad F_{p,q_r}^{\underline{M}(\pi)}(\alpha)=1-\sum_{i=1,i\neq r}^{i=k}F_{pq_i}(\alpha).$$

The definition of the π-maximizing MDP is similar.

In order to define the optimal value function for a BMDP, two different orderings on closed real intervals are introduced: $[l_1,u_1]\leq_{opt}[l_2,u_2]\iff(u_1<u_2\vee(u_1=u_2\wedge l_1\leqslant l_2))$ and $[l_1,u_1]\leq_{pes}[l_2,u_2]\iff(l_1<l_2\vee(l_1=l_2\wedge u_1\leqslant u_2))$. In addition, $\hat{U}\leq_{opt}\hat{V}$ ($\hat{U}\leq_{pes}\hat{V}$) if and only if $\hat{U}(q)\leq_{opt}\hat{V}(q)$ ($\hat{U}(q)\leq_{pes}\hat{V}(q)$) for each $q\in\mathcal{Q}$. Then the optimistic optimal value function $\hat{V}_{opt}$ and the pessimistic optimal value function $\hat{V}_{pes}$ are given by

$$\hat{V}_{opt}=\sup_{\pi\in\mathit{\Pi},\leq_{opt}}\hat{V}_{\pi}\text{ and }\hat{V}_{pes}=\sup_{\pi\in\mathit{\Pi},\leq_{pes}}\hat{V}_{\pi},$$

respectively. The value interation for $\hat{V}_{opt}$ is used when the agent aims at maximizing the upper bound $\overline{V}$ while $\hat{V}_{pes}$ is used when the agent aims at maximizing the lower bound $\underline{V}$. In the subsequent sections, we focus on the optimistic case for the optimal interval value functions. Unless noted, results for the pessimistic case can be inferred analogously.

The interval value iteration operator $\widehat{IVI}_{opt}$ for each state p is defined as

$$\widehat{IVI}_{opt}(\hat{V})(p)=\max_{\alpha\in\mathcal{A},\leq_{opt}}\Big[\min_{M\in\mathcal{M}}VI_{M,\alpha}(\underline{V})(p),\ \max_{M\in\mathcal{M}}VI_{M,\alpha}(\overline{V})(p)\Big]. \tag{1}$$

Due to the nature of $\leq_{opt}$, $\widehat{IVI}_{opt}$ evaluates actions primarily based on the interval upper bounds, breaking ties on the lower bounds. For each state, the action that maximizes the lower bound is chosen from the subset of actions that equally maximize the upper bound. Hence (1) can be rewritten as

$$\widehat{IVI}_{opt}(\hat{V})=[\underline{IVI}_{opt}(\hat{V}),\overline{IVI}_{opt}(\overline{V})] \tag{2}$$

where

$$\underline{IVI}_{opt}(\hat{V})=\underline{IVI}_{opt,\overline{V}}(\underline{V})$$

and for any $q\in\mathcal{Q}$,

$$\overline{IVI}_{opt}(\overline{V})(q)=\max_{\alpha\in\mathcal{A}}\max_{M\in\mathcal{M}}VI_{M,\alpha}(\overline{V})(q),$$

$$\underline{IVI}_{opt,\overline{V}}(\underline{V})(q)=\max_{\alpha\in\rho_{\overline{V}}(q)}\min_{M\in\mathcal{M}}VI_{M,\alpha}(\underline{V})(q)$$

where

$$\rho_W(p)=\arg\max_{\alpha\in\mathcal{A}}\max_{M\in\mathcal{M}}VI_{M,\alpha}(W)(p). \tag{3}$$

Methods similar to those used in proving the convergence of total discounted reward optimality for exact MDPs can be used to prove that iterating $\widehat{IVI}_{opt}$ converges to $\hat{V}_{opt}$. Detailed proofs of convergence results can be found in [1].

3 Modeling Uncertain Systems by BMDPs

In this section, we describe how BMDPs can be generated for uncertain systems and we illustrate the approach using a robot-path finding application. Consider a continuous system with dynamics described by a stochastic differential equation (SDE) $dx = f(x,u)dt + \sigma(x)dw$ where $x \in X$ is the state of the system, $u \in U$ is the control action, $\sigma(x)$ is a diffusion term of appropriate dimensions, and $w(t)$ is a Wiener process. The SDE is approximated by a controlled Markov process that evolves in a state space that is a discretization of the state space X. The criterion which must be satisfied by the approximating MDP is *local consistency* [11]. Local consistency means that the conditional mean and covariance of the MDP are proportional to the local mean and covariance of the original process. An approximation parameter h analogous to a "finite element size" parameterizes the approximating Markov process. As h goes to zero, the local properties of the MDP resemble the local properties of the original stochastic process.

The transition probabilities of the MDP can be computed systematically from the parameters of the SDE (details can be found in [11]). If the diffusion matrix $a(x) = \sigma(x)\sigma^T(x)$ is diagonal and we consider a uniform grid with e_i denoting the unit vector in the i^{th} direction, the transition probabilities are

$$F_{x,x\pm he_i}(u) = \frac{a_{ii}(x)/2 + hf_i^{\pm}(x,u)}{Q(x,u)}, \tag{4}$$

where $\Delta t(x,u) = h^2/Q(x,u)$, $Q(x,u) = \sum_i [a_{ii}(x) + h|f_i(x,u)|]$ and $a^+ = \max\{a,0\}$ and $a^- = \max\{-a,0\}$.

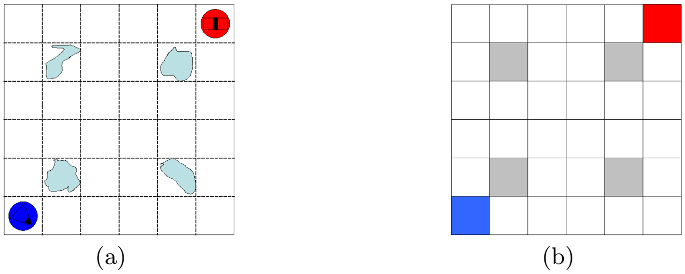
The approximation described above assumes that the system model is known exactly. For many practical systems, however, model parameters are not known exactly. Uncertain continuous systems are usually modeled assuming that some parameters take values in a pre-defined (usually convex) set. In this case, the approximation outlined above will result in BMDPs where the transition probabilities are replaced by interval transition probabilities.

In the following, we illustrate the approximation approach with a robot-path finding example. For simplicity, we assume that mobile robots operate in planar environments and we do not model the orientation or any nonholonomic constraints. The behavior of the robot is described by

$$\begin{aligned} dx &= u_1 dt + \sigma_1 dw \\ dy &= u_2 dt + \sigma_2 dw \end{aligned}$$

where $(x,y)^{\mathrm{T}}$ is the coordinate of the robot, $(u_1,u_2)^{\mathrm{T}}$ is the control input representing the command velocity, and $w(t)$ is a Wiener process modeling noise.

Figure 1(a) shows the original model of the operating environment of the robot. The robot is initially at the lower left corner and the destination is at the upper right corner. We discretize the robot's operating environment using a uniform grid and we assume that there are only 4 control actions, {Up, Down, Left, Right}. As shown in Figure 1(b), we also approximate the position of the robot, the destination, and the obstacles as MDP states. Consider a fixed control



(a) (b)

Fig. 1. Robot path-finding problem. (a) Original model of the path-finding problem. (b) The approximated MDP model.

action denoted by u. Because of uncertainty in the system such as motor friction or an unknown workload, it is reasonable to assume that the control action corresponds to the command velocity $u = (u_1, u_2)^{\mathrm{T}}$ where u_i $(i = 1, 2)$ is not exact but takes values in the interval $[\underline{u}_i, \overline{u}_i]$. Define the set $\tilde{\mathcal{U}} = \{(u_1, u_2)^{\mathrm{T}} : u_i \in [\underline{u}_i, \overline{u}_i]\}$, then the interval transition probabilities can be computed by

$$\hat{F}_{x,x\pm he_i}(u) = \left[\min_{u\in\tilde{\mathcal{U}}} F_{x,x\pm he_i}(u), \max_{u\in\tilde{\mathcal{U}}} F_{x,x\pm he_i}(u)\right]. \tag{5}$$

Assuming that $\mathcal{U}$ is compact, the function F has well-defined extrema. The intervals can be computed either analytically if the functions F are monotone with respect to the uncertain parameters or using numerical optimization methods. Thus, for each state-action pair we obtain an interval for the transition probabilities and by repeating for all state-action pairs we obtain a BMDP model.

Note that in our approach, (4) and (5) are applied to one dimension of the uncertain system at a time, so the approach is actually applicable to more general systems in higher dimensions than the above robotic example.

4 Maximum Probability Reachability Problem

In this section, we formulate the maximum probability reachability problem, we present a value iteration algorithm, and we analyze its convergence.

4.1 Interval Expected Total Reward for BMDPs

In this paper, we are primarily interested in the problem of maximizing the probability that the agent will reach a desirable set of states. By solving this problem, we can establish bounds on the probabilities of reaching desirable configurations used in probabilistic verification of discrete systems. This problem can be formulated using the Expected Total Reward Criterion (ETRC) for BMDPs (see

Section 4.3). Under the ETRC, we compare policies on the basis of the interval expected total reward $\hat{V} = [\underline{V}_\pi, \overline{V}_\pi]$ where for any $q \in \mathcal{Q}$

$$\overline{V}_\pi(q) = E_{\overline{M}(\pi),\pi}\left\{\sum_{t=1}^{\infty} R(X_t(q))\right\} \text{ and } \underline{V}_\pi(q) = E_{\underline{M}(\pi),\pi}\left\{\sum_{t=1}^{\infty} R(X_t(q))\right\}.$$

We may regard these as the expected total discounted reward with a discount factor $\gamma = 1$. However, for $\gamma = 1$ the convergence results in [1] no longer hold, because the iteration operators $\widehat{IVI}_\pi$, $\widehat{IVI}_{opt}$ and $\widehat{IVI}_{pes}$ are not contraction mappings. Furthermore, the interval value function may not be well defined unless proper restrictions on the intervals of the transition probabilities and rewards are applied.

For simplicity, we use vector notation. For example, R and V are column vectors, whose i-th element is respectively the scalar reward and value function of the i-th state p_i; F_M is the transition probability function of MDP M and $F_{M,\pi}$ is the transition probability matrix of the Markov Chain reduced from M when given a policy π, whose (i,j)-th element is the probability of transitioning from state p_i to state p_j when executing action $\pi(p_i)$.

Let $R^+(q) = \max\{R(q), 0\}$ and $R^-(q) = \max\{-R(q), 0\}$ and define

$$\overline{V}_\pi^{\pm}(q) \equiv \lim_{N\to\infty} E_{\overline{M}(\pi),\pi}\left\{\sum_{t=1}^{N-1} R^{\pm}(X_t(q))\right\}.$$

Since the summands are non-negative, both of the above limits exist[3]. The limit defining $\overline{V}_\pi(q)$ exists whenever at least one of $\overline{V}_\pi^+(q)$ and $\overline{V}_\pi^-(q)$ is finite, in which case $\overline{V}_\pi = \overline{V}_\pi^+(q) - \overline{V}_\pi^-(q)$. $\underline{V}_\pi^+(q)$, $\underline{V}_\pi^-(q)$ and $\underline{V}_\pi(q)$ can be similarly defined. Noting this, we impose the following finiteness assumption which assures that $\hat{V}_\pi$ is well defined.

Assumption 1 *For all $\pi \in \Pi$ and $q \in \mathcal{Q}$, (a) either $\overline{V}_\pi^+(q)$ or $\overline{V}_\pi^-(q)$ is finite, and (b) either $\underline{V}_\pi^+(q)$ or $\underline{V}_\pi^-(q)$ is finite.*

Consider the optimal interval value functions $\hat{V}_{opt}$ defined in Section 2. The following theorem establishes the optimality equation for the ETRC and shows that the optimal interval value function is a solution of the optimality equation.

Theorem 1 *Suppose Assumption 1 holds. Then (a) The upper bound of the optimal interval value function $\overline{V}_{opt}$ satisfies the equation*

$$V = \sup_{\pi\in\Pi} \max_{M\in\mathcal{M}} VI_{M,\pi}(V) = \sup_{\pi\in\Pi}\{R + F_{\overline{M}(\pi),\pi}V\} \equiv \overline{IVI}_{opt}(V),$$

(b) The lower bound of the optimal interval value function $\underline{V}_{opt,W}$ satisfies the equation

$$V = \sup_{\pi\in\rho_W} \min_{M\in\mathcal{M}} VI_{M,\pi}(V) = \sup_{\pi\in\rho_W}\{R + F_{\underline{M}(\pi),\pi}V\} \equiv \underline{IVI}_{opt,W}(V)$$

for any value function W and the associated action selection function (3)[4].

[3] This includes the case when the limit is $\pm\infty$.

[4] Proofs are omitted due to length limitation, and can be found in [12].

Based on Theorem 1, the value iteration operator $\widehat{IVI}_{opt}$ can be defined as in Equation (1). The following lemma establishes the monotonicity of the iteration operators.

Lemma 2 *Suppose U and V are value functions in $\mathcal{V}$ with $U \leq_{dom} V$, then (a) $\overline{IVI}_{opt}(U) \leq_{dom} \overline{IVI}_{opt}(V)$, (b) $\underline{IVI}_{opt,W}(U) \leq_{dom} \underline{IVI}_{opt,W}(V)$ for any value function W and the associated action selection function (3).*

Lemma 2 also suggests that the iteration operator $\widehat{IVI}_{opt}$ has the following property: for any $\hat{U} \leq_{opt} \hat{V}$ in $\hat{\mathcal{V}}$, $\widehat{IVI}_{opt}(\hat{U}) \leq_{opt} \widehat{IVI}_{opt}(\hat{V})$. These properties are essential in the proof of the convergence results of the interval value iteration algorithm.

Clearly, Assumption 1 is necessary for any computational approach. In the general case for the expected total reward criterion (ETRC), we cannot validate that the assumption holds. However, in the maximum probability reachability problem, the (interval) value function is interpreted as (interval) probability and therefore Assumption 1 can be easily validated as shown in Section 4.3.

4.2 Interval Value Iteration for Non-negative BMDP models

In order to prove convergence of the value iteration, we consider the following assumptions in addition to Assumption 1:

Assumption 2 *For all $q \in Q$, $R(q) \geqslant 0$.*

Assumption 3 *For all $q \in Q$ and $\pi \in \Pi$, $\overline{V}^{+}_{\pi}(q) < \infty$ and $\underline{V}^{+}_{\pi}(q) < \infty$.*

If a BMDP is consistent with both Assumption 2 and 3, it is a non-negative BMDP model, and its value function under the ETRC is called non-negative interval expected total reward. Note that Assumption 3 implies Assumption 1, so Theorem 1 and Lemma 2 hold for non-negative BMDP models. Lemma 3 suggests that $\hat{V}_{opt}$ is the minimal solution of the optimality equation, and Theorem 4 establishes the convergence result of interval value iteration for non-negative BMDPs.

Lemma 3 *Suppose Assumption 2 and 3 hold. Then (a) $\overline{V}_{opt}$ is the minimal solution of $V = \overline{IVI}_{opt}(V)$ in $\mathcal{V}^{+}$, where $\mathcal{V}^{+} = \mathcal{V} \cap [0, \infty]$, (b) $\underline{V}_{opt,W}$ is the minimal solution of $V = \underline{IVI}_{opt,W}(V)$ in $\mathcal{V}^{+}$ for any value function W and the associated action selection function (3).*

Theorem 4 *Suppose Assumption 2 and 3 hold. Then for $\hat{V}^0 = [0, 0]$, the sequence $\{\hat{V}^n\}$ defined by $\hat{V}^n = \widehat{IVI}^{\,n}_{opt}(\hat{V}^0)$ converges pointwise and monotonically to $\hat{V}_{opt}$.*

It can be shown that the initial value of the interval value function is not restricted to be $[0, 0]$. By choosing a $\hat{V}^0$ with $0 \leqslant \underline{V}^0 \leqslant \underline{V}_{opt}$ and $0 \leqslant \overline{V}^0 \leqslant \overline{V}_{opt}$, interval value iteration converges to $\hat{V}_{opt}$ for non-negative BMDPs. For BMDP models consistent with Assumption 2 and Assumption 3, convergence of the iterative algorithm is guaranteed by Theorem 4 for $\hat{V}^0 = [0, 0]$.

4.3 Verification Based on Non-negative BMDP models

An instance of the maximum probability reachability problem for BMDPs consists of a BMDP $\mathcal{M} = \langle \mathcal{Q}, \mathcal{A}, \hat{F}, R \rangle$ together with a destination set $\mathcal{T} \subseteq \mathcal{Q}$. The objective of maximum probability reachability problem is to determine, for all $p \in \mathcal{Q}$, the maximum interval probability of starting from p and finally reaching any state in $\mathcal{T}$, i.e.

$$\hat{U}^{max}_{\mathcal{M},opt}(p) = \sup_{\pi \in \Pi, \leq_{opt}} [\underline{U}_{\mathcal{M},\pi}(p), \overline{U}_{\mathcal{M},\pi}(p)]$$

where

$$\underline{U}_{\mathcal{M},\pi}(p) = \min_{M \in \mathcal{M}} Pr_{M,\pi}(\exists t. X_t(p) \in \mathcal{T}), \tag{6}$$

$$\overline{U}_{\mathcal{M},\pi}(p) = \max_{M \in \mathcal{M}} Pr_{M,\pi}(\exists t. X_t(p) \in \mathcal{T}). \tag{7}$$

$\underline{U}_{\mathcal{M},\pi}$ and $\overline{U}_{\mathcal{M},\pi}$ are probabilities and therefore by definition take values in $[0,1]$. Thus, the interval value function satisfies Assumption 1. Note that $\underline{U}_{\mathcal{M},\pi}(p)$ can be computed recursively by

$$\underline{U}_{\mathcal{M},\pi}(p) = \begin{cases} \min\limits_{M \in \mathcal{M}} \sum\limits_{q \in \mathcal{Q}} F^M_{p,q}(\pi(p)) \underline{U}_{\mathcal{M},\pi}(q) & \text{if } p \in \mathcal{Q} - \mathcal{T} \\ 1 & \text{if } p \in \mathcal{T} \end{cases} \tag{8}$$

In order to transform the Maximum Probability Reachability Problem to a problem solvable by interval value iteration, we add a terminal state r with transition probability 1 to itself on any action, let all the destination states in $\mathcal{T}$ be absorbed into the terminal state, i.e., transition to r with probability 1 on any action, and set the reward of each destination state to be 1 and of every other state to be 0. Thus we form a new BMDP model $\widetilde{\mathcal{M}} = \langle \tilde{\mathcal{Q}}, \tilde{\mathcal{A}}, \widetilde{F}, \tilde{R} \rangle$, where $\tilde{\mathcal{Q}} = \mathcal{Q} \cup \{r\}$, $\tilde{\mathcal{A}} = \mathcal{A}$ and for any $p, q \in \tilde{\mathcal{Q}}$, and $\alpha \in \mathcal{A}$

$$\tilde{R}(p) = \begin{cases} 1 \text{ if } p \in \mathcal{T} \\ 0 \text{ if } p \notin \mathcal{T} \end{cases},$$

$$\tilde{F}_{p,q}(\alpha) = \begin{cases} \hat{F}_{p,q}(\alpha) & \text{if } p \notin \mathcal{T} \cup \{r\} \\ [0,0] & \text{if } p \in \mathcal{T} \cup \{r\} \text{ and } q \neq r \\ [1,1] & \text{if } p \in \mathcal{T} \cup \{r\} \text{ and } q = r \end{cases}. \tag{9}$$

Since $\tilde{R}(r) = 0$, by the structure of $\tilde{F}_{p,q}$, it is clear that $\underline{V}_{\widetilde{\mathcal{M}},\pi}(r)$ will not be affected by the values of any states. For any $p \in \mathcal{Q}$

$$\underline{V}_{\widetilde{\mathcal{M}},\pi}(p) = \min_{M \in \widetilde{M}} \left\{ \tilde{R}(p) + \sum_{q \in \tilde{Q}} F^M_{p,q}(\pi(p)) \underline{V}_{M,\pi}(q) \right\}. \tag{10}$$

Specifically, for $p \in \mathcal{T}$

$$\underline{V}_{\widetilde{\mathcal{M}},\pi}(p) = \min_{M \in \widetilde{M}} \left\{ \tilde{R}(p) + \sum_{q \in \tilde{Q}} F^M_{p,q}(\pi(p)) \underline{V}_{M,\pi}(q) \right\} = \tilde{R}(p) + \underline{V}_{\widetilde{M},\pi}(r) = 1. \tag{11}$$

From (9), (10) and (11), it follows that $\underline{U}_{\mathcal{M},\pi}$ is equivalent to $\underline{V}_{\widetilde{\mathcal{M}},\pi}$. Similarly, $\overline{U}_{\mathcal{M},\pi}$ is equivalent to $\overline{V}_{\widetilde{\mathcal{M}},\pi}$. Therefore

$$\hat{V}_{\widetilde{M},opt} = \sup_{\pi \in \Pi, \leq_{opt}} [\underline{V}_{\widetilde{M},\pi}, \overline{V}_{\widetilde{M},\pi}] = \sup_{\pi \in \Pi, \leq_{opt}} [\underline{U}_{\mathcal{M},\pi}, \overline{U}_{\mathcal{M},\pi}] = \hat{U}_{\mathcal{M},opt}. \tag{12}$$

The BMDP $\widetilde{M}$ constructed as described above is consistent with Assumption 3, so the interval value function for each state exists, which suggests that the MPRP for $\mathcal{M}$ can be solved using the algorithm presented in Section 4.1. Further, $\widetilde{\mathcal{M}}$ satisfies Assumption 2, and therefore the convergence is characterized by Theorem 4.

Note that we don't assume the existence of a proper policy. Convergence is guaranteed without this assumption. In the case of the maximum probability reachability problem, if there is not proper policy (for a particular state) then the algorithm will simply compute the corresponding interval value function (probability) as $[0, 0]$. The approach can be used to validate the existence of a proper policy and actually this is one of the ways that probabilistic verification algorithms can be used in practice.

5 Experimental Results

This section illustrates the approach using a robot path-finding application. In our model, an action succeeds with interval probability $[0.75, 0.9]$ and moves in any other direction with interval probability $[0.05, 0.1]$. For instance, if the robot choose the action "Up", the probability of reaching the adjacent grid to its north is within $[0.75, 0.9]$, the probability of reaching each of the other adjacent grids is within $[0.05, 0.1]$. We also assume the robot will stay where it is with a probability in the same interval probability as if it is not out of bound. Obstacle grids are treated as absorbing states, i.e. transition to itself with interval probability $[1, 1]$ on any action. The goal is to find a policy that maximizes the interval probability that the robot will reach the destination from the initial position.

The layout of the gridworld used in our simulation is shown in Figure 2(a). The (blue) cell in the lower left corner is the initial position of the robot. The (red) grid in the upper right corner is the destination. The (grey) cells represent obstacles. In order to evaluate the computational complexity and scalability of our algorithm, the environment is made up of the same 3 × 3 tiles as shown in Figure 2(b). For instance, the 9 × 9 gridworld shown in Figure 2(c) is made up of 9 such tiles, while the 6 × 6 gridworld in Figure 1(b) in Section 2 is made up of 4 such tiles.

Table 1 shows the interval maximum probabilities for the robot to reach the destination from the initial position, number of iterations and time needed for the iterative algorithm to converge. For example, the optimistic maximum reachability probability for the 9×9 gridworld is $[0.2685, 0.6947]$, the pessimistic maximum reachability probability for the 18×18 gridworld is $[0.1067, 0.4806]$. We can see that the larger the size of the gridworld, the lower the reachability

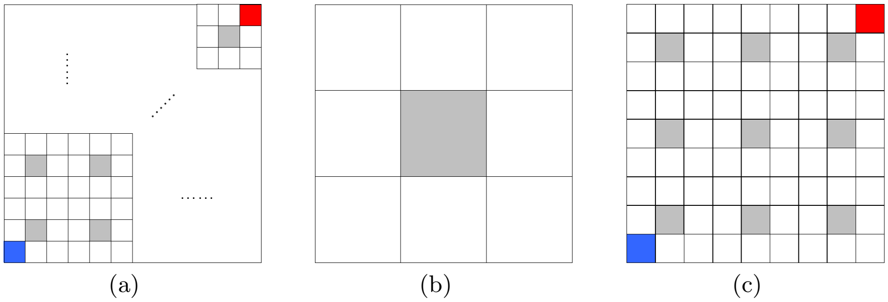


Fig. 2. Robot path-finding problem. (a) The operating environment of the robot. b) 3×3 tile – the basic component of the environment model. c) The 9×9 environment model that is made up of the 3×3 tiles.

Table 1. Interval maximum reachability probabilities ($\hat{U}^{max}_{opt}$, $\hat{U}^{max}_{pes}$) for the robot path-finding problem, Number of iterations (I_{opt}, I_{pes}) and time (t_{opt}, t_{pes}, in seconds) needed for the iterative algorithms to converge.

Size	**States**	$\hat{U}^{max}_{opt}$	I_{opt}	t_{opt}	$\hat{U}^{max}_{pes}$	I_{pes}	t_{pes}
9×9	81	$[0.2685, 0.6947]$	43	3.98	$[0.4156, 0.6947]$	43	3.98
12×12	144	$[0.1707, 0.6145]$	54	15.04	$[0.2645, 0.6145]$	54	14.94
15×15	225	$[0.1083, 0.5435]$	63	42.10	$[0.1681, 0.5435]$	63	41.84
18×18	324	$[0.0686, 0.4807]$	71	98.34	$[0.1067, 0.4806]$	71	98.54
21×21	441	$[0.0434, 0.4251]$	79	201.49	$[0.0434, 0.4251]$	79	201.55
24×24	576	$[0.0275, 0.3760]$	87	374.92	$[0.0275, 0.3760]$	87	375.95

probability. This is because larger gridworld suggests a longer path for the robot to reach the destination, and greater chance to collide with obstacles. All the simulations are carried out on a Windows XP laptop, 1.60GHz, with 768 MB of RAM, using MATLAB 7.0. Our experimental results suggest that the time complexity of the interval value iteration is polynomial. The exact complexity characterization is a subject of current work.

6 Conclusions

The results described in this paper show that BMDPs can be used for probabilistic verification of uncertain systems. With proper restrictions on the reward and transition functions, the interval value function is well defined and bounded. We also analyze the convergence of iterative methods for computing the interval value function. These results allow us to solve a variety of new problems for BMDPs. The paper focuses on the maximum reachability probability problem. Additional verification problems are subject of current and future work.

Acknowledgments This work was partially supported by the NSF CAREER grant CNS-0448554.

References

1. Givan, R., Leach, S., Dean, T.: Bounded-parameter Markov decision process. Artificial Intelligence **122** (2000) 71–109
2. Courcoubetis, C., Yannakakis, M.: Markov decision processes and regular events. IEEE Transaction on Automatic Control **43** (1998) 1399–1418
3. de Alfaro, L.: Computing minimum and maximum reachability times in probabilistic systems. In: CONCUR '99: Proceedings of the 10th International Conference on Concurrency Theory, London, UK, Springer-Verlag (1999) 66–81
4. Koutsoukos, X.D.: Optimal control of stochastic hybrid systems based on locally consistent Markov Decision Processes. International Journal of Hybrid Systems **4** (2004) 301–318
5. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, NY, USA (1994)
6. Satia, J.K., Lave, R.E.: Markovian decision processes with uncertain transition probabilities. Operations Research **39** (1953) 1095–1100
7. White, C.C., Eldeib, H.K.: Parameter imprecision in finite state, finite action dynamic programs. Operations Research **34** (1986) 120–129
8. White, C.C., Eldeib, H.K.: Markov decision processes with imprecise transition probabilities. Operations Research **43** (1994) 739–749
9. Tang, H., Liang, X., Gao, J., Liu, C.: Robust control policy for semi-markov decision processes with dependent uncertain parameters. In: WCICA 04': Proceedings of the Fifth World Congress on Intelligent Control and Automation. (2004)
10. Kalyanasundaram, S., Chong, E.K.P., Shroff, N.B.: Markovian decision processes with uncertain transition rates: Sensitivity and robust control. In: CDC '02: Proceedings of the 41th IEEE Conference on Decision and Control. (2002)
11. Kushner, H.J., Dupuis, P.: Numerical Methods for Stochastic Control Problems in Continuous Time, 2nd ed. Springer-Verlag, New York, NY, USA (2001)
12. Wu, D., Koutsoukos, X.D.: Probabilistic verification of bounded-parameter markov decision processes. Technical Report ISIS-05-607, Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN (2005)
13. Dean, T., Givan, R., Leach, S.: Model reduction techniques for computing approximately optimal solutions for Markov decision processes. In: Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence (UAI-97), San Francisco, CA, Morgan Kaufmann Publishers (1997) 124–131
14. Rudin, W.: Real and Complex Analysis, 3rd Edition. McGraw-Hill, New York, NY, USA (1994)