Available online at www.sciencedirect.com

ScienceDirect

Artificial Intelligence 172 (2008) 945–954

Artificial Intelligence

www.elsevier.com/locate/artint

Reachability analysis of uncertain systems using bounded-parameter Markov decision processes

Di Wu, Xenofon Koutsoukos *

EECS Department, Vanderbilt University, Nashville, TN 37235, USA

Received 8 September 2006; received in revised form 8 December 2007; accepted 17 December 2007

Available online 23 December 2007

Abstract

Verification of reachability properties for probabilistic systems is usually based on variants of Markov processes. Current methods assume an exact model of the dynamic behavior and are not suitable for realistic systems that operate in the presence of uncertainty and variability. This research note extends existing methods for Bounded-parameter Markov Decision Processes (BMDPs) to solve the reachability problem. BMDPs are a generalization of MDPs that allows modeling uncertainty. Our results show that interval value iteration converges in the case of an undiscounted reward criterion that is required to formulate the problems of maximizing the probability of reaching a set of desirable states or minimizing the probability of reaching an unsafe set. Analysis of the computational complexity is also presented.

Keywords: Reachability analysis; Uncertain systems; Markov decision processes

1. Introduction

Verification of reachability properties for probabilistic systems is usually based on variants of Markov processes. 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. Such problems are important in many application domains such as planning for autonomous systems [1], system biology [2], and finance [3].

Algorithms for verification of MDPs have been presented in [4,5]. Several other probabilistic models based on variants of MDPs also have been considered [6,7]. These methods assume exact values of the transition probabilities which typically are computed either based on detailed models using discrete approximation techniques [8] or have to be estimated from data [9]. However, realistic systems often operate in the presence of uncertainty and variability, and modeling and estimation errors can affect the transition probabilities and impact the solution. Such systems can be best described using uncertain transition probabilities. An uncertain MDP which describes the routing of an aircraft based on past weather data is presented in [10]. Computing uncertain transition probabilities for a robot path-finding

[*] Corresponding author.
E-mail addresses: wudipku@gmail.com (D. Wu), xenofon.koutsoukos@vanderbilt.edu (X. Koutsoukos).

0004-3702/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.artint.2007.12.002

example based on a model of the continuous dynamics is described in [11]. Existing reachability analysis methods are insufficient for dealing with such uncertainty.

This research note extends existing methods for Bounded-parameter Markov Decision Processes (BMDPs) to solve the reachability problem. Proposed by Givan, Leach and Dean [12], 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 and policies are compared on the basis of interval value functions. An overview of BMDPs is presented in Section 2.

The paper focuses on the problem of maximizing the probability of reaching a set of desirable states. The results presented in [12] are for dynamic programming methods assuming a discounted reward criterion. A discount factor ensures the convergence of the iterative methods for the interval value functions. Probabilistic verification can be formulated based on the Expected Total Reward Criterion (ETRC) [13]. Under ETRC, the discount factor is set to 1, and the convergence of the iterative algorithms for BMDPs is more involved because the contraction property of the iteration operators does not hold globally and the interval value function may not be well defined unless proper restrictions on the intervals of transition probabilities and rewards are applied. The interval expected total reward for BMDPs is analyzed in Section 3. Further, proving the polynomial computational complexity of the algorithm requires a different method using an appropriate weighted norm. Based on the ETRC, this paper presents a detailed analysis of the convergence and the computational complexity for the maximum probability reachability problem in Sections 4 and 5 respectively. Minimum probability reachability and other problems based on the ETRC [13] can be addressed in a similar fashion. A simplified robot path-finding example and numerical results that illustrate the approach can be found in [11].

Optimal solutions to several variants of uncertain MDP problems have been studied previously. MDPs with uncertain transition probabilities and a discounted reward criterion have been considered in [14,15]. Related methods that consider a discounted reward include the work in [16] which computes the optimal policy in models with compact convex uncertain sets, the approach in [17] which computes the Pareto optimal policy which maximizes the average expected reward over all stationary policies under a specific partial order, and the work in [10] which solves a robust control problem. The average reward problem for BMDPs has been studied in [18] and a similar average performance criterion has been considered in [19]. An algorithm based on real-time dynamic programming for uncertain stochastic shortest path problems is presented in [20]. The algorithm requires that a goal state is reachable from any visited state and proposes a reachability analysis pre-processing step which is based on graph analysis. Probabilistic reachability analysis of uncertain MDPs is a significant problem which requires an undiscounted reward criterion and cannot be treated with these algorithms.

Probabilistic verification of uncertain systems has been addressed also using model checking methods. A variant of uncertain MDPs has been presented in [21,22]. The main characteristic of the model is that uncertainty is resolved through nondeterminism, i.e. at every step an adversary picks a probability distribution that satisfies the uncertain transition probabilities. This differs from BMDPs where the transition probabilities are uncertain for a given action selected by an external agent. The approach presented in [21] computes the probability distribution over the states for finite number of steps while the algorithms in [22] reduce the uncertain system to an MDP of a larger size for verifying a subset of probabilistic computation tree logic specifications without steady state operators.

2. Bounded-parameter Markov decision processes

We first review some basic notions of BMDPs from [12] and establish the notation. A BMDP is defined as $\mathcal{M} = \langle \mathcal{Q}, \mathcal{A}, \hat{F}, \hat{R} \rangle$ 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 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 $M = \langle \mathcal{Q}^M, \mathcal{A}^M, F^M, R^M \rangle$ be an MDP. If $\mathcal{Q}^M = \mathcal{Q}$, $\mathcal{A}^M = \mathcal{A}$, $R^M(p) \in \hat{R}(p)$, and $F^M_{p,q}(\alpha) \in \hat{F}_{p,q}(\alpha)$ for any $\alpha \in \mathcal{A}$ and $p, q \in \mathcal{Q}$, then we say $M \in \mathcal{M}$. To simplify the presentation, the rewards are assumed to be tight, however, the results can be easily generalized to the case of interval rewards.

Policies are defined as $\pi : \mathcal{Q} \to \mathcal{A}$ and are restricted into the set of stationary Markov policies Π. Let $\mathcal{V}$ denote the set of value functions on $\mathcal{Q}$. For an exact MDP M, policy π, and discount factor $\gamma \in (0, 1)$, the value function is the solution of the equation

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

and can be computed by iteratively applying the policy evaluation operator denoted as $VI_{M,\pi} : \mathcal{V} \to \mathcal{V}$. For any policy π and state p, the interval value function of the BMDP $\mathcal{M}$ for π at p is the closed interval

$$\hat{V}_\pi(p) = \Big[\inf_{M \in \mathcal{M}} V_{M,\pi}(p), \sup_{M \in \mathcal{M}} V_{M,\pi}(p)\Big]. \tag{1}$$

An MDP $M \in \mathcal{M}$ is π-maximizing if for any $M' \in \mathcal{M}$, $V_{M,\pi} \geqslant_{\text{dom}} V_{M',\pi}$[1] and likewise, M is π-minimizing if for any $M' \in \mathcal{M}$, $V_{M,\pi} \leqslant_{\text{dom}} V_{M',\pi}$. It is proved in [12] (Theorem 7 and Corollary 1) that for any policy $\pi \in \Pi$ and any ordering of the states Q, there exist a π-maximizing MDP $\overline{M}(\pi)$ and a π-minimizing MDP $\underline{M}(\pi)$. This implies that for input $\overline{V}$ (or $\underline{V}$) there exists a single MDP independent of $\overline{V}$ (or $\underline{V}$) which simultaneously maximizes (or minimizes) $V_{M,\pi}(p)$ for all states $p \in Q$. Therefore, we can define the interval policy evaluation operator $\widehat{IVI}_\pi$ as

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

where

$$\underline{IVI}_\pi(\underline{V}) = \min_{M \in \mathcal{M}} VI_{M,\pi}(\underline{V}) = VI_{\underline{M}(\pi),\pi}(\underline{V}) \quad \text{and} \quad \overline{IVI}_\pi(\overline{V}) = \max_{M \in \mathcal{M}} VI_{M,\pi}(\overline{V}) = VI_{\overline{M}(\pi),\pi}(\overline{V}).$$

In order to define the optimal value function for a BMDP, two different orderings on closed real intervals are introduced: $[l_1, u_1] \leqslant_{\text{opt}} [l_2, u_2] \iff (u_1 < u_2 \vee (u_1 = u_2 \wedge l_1 \leqslant l_2))$ and $[l_1, u_1] \leqslant_{\text{pes}} [l_2, u_2] \iff (l_1 < l_2 \vee (l_1 = l_2 \wedge u_1 \leqslant u_2))$. In addition, $\hat{U} \leqslant_{\text{opt}} \hat{V}$ ($\hat{U} \leqslant_{\text{pes}} \hat{V}$) if and only if $\hat{U}(q) \leqslant_{\text{opt}} \hat{V}(q)$ ($\hat{U}(q) \leqslant_{\text{pes}} \hat{V}(q)$) for each $q \in \mathcal{Q}$. Then the optimistic optimal value function $\hat{V}_{\text{opt}}$ and the pessimistic optimal value function $\hat{V}_{\text{pes}}$ are defined as the upper bounds over all stationary policies using $\leqslant_{\text{opt}}$ and $\leqslant_{\text{pes}}$ respectively to order interval value functions, i.e.

$$\hat{V}_{\text{opt}} = \max_{\pi \in \Pi, \leqslant_{\text{opt}}} \hat{V}_\pi \text{ and } \hat{V}_{\text{pes}} = \min_{\pi \in \Pi, \leqslant_{\text{pes}}} \hat{V}_\pi,$$

respectively.

The value iteration for $\hat{V}_{\text{opt}}$ is used when the objective is to maximize the upper bound $\overline{V}$ while $\hat{V}_{\text{pes}}$ is used to maximize the lower bound $\underline{V}$. In the subsequent sections, we focus on the optimistic case for the optimal interval value functions. Results for the pessimistic case can be inferred analogously.

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

$$\widehat{IVI}_{\text{opt}}(\hat{V})(p) = \max_{\alpha \in \mathcal{A}, \leqslant_{\text{opt}}} \Big[\min_{M \in \mathcal{M}} VI_{M,\alpha}(\underline{V})(p), \max_{M \in \mathcal{M}} VI_{M,\alpha}(\overline{V})(p)\Big]. \tag{2}$$

Due to the nature of $\leqslant_{\text{opt}}$, $\widehat{IVI}_{\text{opt}}$ 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. To capture this behavior, we define the action selection function

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

and

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

Then (2) can be rewritten as

$$\widehat{IVI}_{\text{opt}}(\hat{V}) = \big[\underline{IVI}_{\text{opt},\overline{V}}(\underline{V}), \overline{IVI}_{\text{opt}}(\overline{V})\big]. \tag{4}$$

[1] $V \geqslant_{\text{dom}} U$ if and only if for all $q \in \mathcal{Q}$, $V(q) \geqslant U(q)$; $\leqslant_{\text{dom}}$ is defined similarly.

3. Interval expected total reward for BMDPs

In this paper, we are primarily interested in the problem of maximizing the probability that the system 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. The ERTC can be viewed as the expected total discounted reward with a discount factor $\gamma = 1$. For $\gamma = 1$ the convergence results in [12] no longer hold, because the iteration operators $\widehat{IVI}_\pi$, $\widehat{IVI}_{\text{opt}}$ and $\widehat{IVI}_{\text{pes}}$ are not global 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.

To simplify the notation, we use vector notation where R and V are column vectors, whose ith element is the scalar reward and value function of the ith state respectively. F_M is the transition probability function of MDP M, and $F_{M,\pi}$ is the transition probability matrix given a policy π. For an exact MDP M and a policy π, the value function for the ETRC is the solution of the equation

$$V_{M,\pi} = R + F_{M,\pi} V$$

and can be computed using the policy evaluation operator $VI_{M,\pi}$ [13]. The interval value function is defined by Eq. (1) similarly to the discounted case. Further, because the existence of a π-minimizing and a π-maximizing MDP does not depend on the discount factor [12], we can define a π-maximizing MDP $\overline{M}(\pi)$ and a π-minimizing MDP $\underline{M}(\pi)$ in $\mathcal{M}$ for the ETRC.

For an MDP M and a policy π, we denote $E_q^{M,\pi}$ the expectation of functionals given the initial state q. 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 $q \in \mathcal{Q}$

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

Let $R^+(q) = \max\{R(q), 0\}$ and $R^-(q) = \max\{-R(q), 0\}$. We define the expected total rewards for R^+ and R^- by

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

that is $\overline{V}^+$ ignores negative rewards and $\overline{V}^+$ ignores positive rewards. Since the summands are non-negative, both of the above limits exist.[2] 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.

Let $\hat{V}_{\text{opt}}$ denote the optimal interval value function for the ETRC. 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.[3]

Theorem 1. *Suppose Assumption* 1 *holds. Then*

(a) *The upper bound of the optimal interval value function* $\overline{V}_{\text{opt}}$ *satisfies the equation*

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

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

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

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

[2] This includes the case when the limit is ±∞.

[3] All proofs are presented in detail in Appendix A for readability.

Based on Theorem 1, the value iteration operator $\widehat{IVI}_{\mathrm{opt}}$ can be defined as in Eq. (2) and the following lemma establishes the monotonicity of the operator.

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

Clearly, Assumption 1 is necessary for any computational approach. In the general case of 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.

4. Maximum probability reachability problem

The maximum probability reachability problem is based on a special case of a class of BMDP models known as the non-negative models (named similarly to MDP models [13]). A BMDP model is called non-negative if it satisfies Assumption 1 and its rewards are all non-negative.

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 $\pi \in \Pi$, $\overline{V}_{\pi}^{+}(q) < \infty$.

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

Lemma 3. *Suppose Assumptions* 2 *and* 3 *hold. Then*
(a) $\overline{V}_{\mathrm{opt}}$ *is the minimal solution of $V = \overline{IVI}_{\mathrm{opt}}(V)$ in $\mathcal{V}^+$, where $\mathcal{V}^+ = \{V \in \mathcal{V} : 0 \leqslant V(p) < \infty$ for each $p \in Q\}$,*
(b) $\underline{V}_{\mathrm{opt},W}$ *is the minimal solution of $V = \underline{IVI}_{\mathrm{opt},W}(V)$ in $\mathcal{V}^+$ for any value function W and the associated action selection function* (3).

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

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.

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

where

$$\underline{U}_{\mathcal{M},\pi}(p) = \min_{M \in \mathcal{M}} Pr_{M,\pi}\big(\exists t.X_t(p) \in \mathcal{T}\big) \quad \text{and} \quad \overline{U}_{\mathcal{M},\pi}(p) = \max_{M \in \mathcal{M}} Pr_{M,\pi}\big(\exists t.X_t(p) \in \mathcal{T}\big).$$

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

$$\underline{U}_{\mathcal{M},\pi}(p) = \begin{cases} \min_{M \in \mathcal{M}} \sum_{q \in \mathcal{Q}} F^M_{p,q}(\pi(p))\underline{U}_{\mathcal{M},\pi}(q) & \text{if } p \in \mathcal{Q} - \mathcal{T} \\ 1 & \text{if } p \in \mathcal{T}. \end{cases} \tag{5}$$

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 $\widetilde{\mathcal{M}} = \langle \tilde{\mathcal{Q}}, \tilde{\mathcal{A}}, \tilde{F}, \tilde{R} \rangle$, where $\tilde{\mathcal{Q}} = \mathcal{Q} \cup \{\mathrm{r}\}$, $\tilde{\mathcal{A}} = \mathcal{A}$ and for any $p, q \in \tilde{\mathcal{Q}}$, and $\alpha \in \mathcal{A}$

$$\tilde{R}(p) = \begin{cases} 1 & \text{if } p \in \mathcal{T} \\ 0 & \text{if } p \notin \mathcal{T}, \end{cases} \quad \text{and} \quad \tilde{F}_{p,q}(\alpha) = \begin{cases} \hat{F}_{p,q}(\alpha) & \text{if } p \notin \mathcal{T} \cup \{r\} \\ [0,0] & \text{if } p \in \mathcal{T} \cup \{r\} \text{ and } q \neq r \\ [1,1] & \text{if } p \in \mathcal{T} \cup \{r\} \text{ and } q = r. \end{cases} \tag{6}$$

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 value function of any other states. For any $p \in \mathcal{Q}$, we have

$$\underline{V}_{\widetilde{\mathcal{M}},\pi}(p) = \min_{M \in \widetilde{\mathcal{M}}} \left\{ \tilde{R}(p) + \sum_{q \in \tilde{Q}} F^M_{p,q}\big(\pi(p)\big) \underline{V}_{M,\pi}(q) \right\}. \tag{7}$$

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

$$\underline{V}_{\widetilde{\mathcal{M}},\pi}(p) = \min_{M \in \widetilde{\mathcal{M}}} \left\{ \tilde{R}(p) + \sum_{q \in \tilde{Q}} F^M_{p,q}\big(\pi(p)\big) \underline{V}_{M,\pi}(q) \right\} = \tilde{R}(p) + \underline{V}_{\widetilde{\mathcal{M}},\pi}(r) = 1. \tag{8}$$

From (6), (7) and (8), 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

$$\hat{V}_{\widetilde{\mathcal{M}},\mathrm{opt}} = \sup_{\pi \in \Pi, \leqslant_{\mathrm{opt}}} \big[\underline{V}_{\widetilde{\mathcal{M}},\pi}, \overline{V}_{\widetilde{\mathcal{M}},\pi}\big] = \sup_{\pi \in \Pi, \leqslant_{\mathrm{opt}}} \big[\underline{U}_{\mathcal{M},\pi}, \overline{U}_{\mathcal{M},\pi}\big] = \hat{U}_{\mathcal{M},\mathrm{opt}}. \tag{9}$$

The BMDP $\widetilde{\mathcal{M}}$ constructed above satisfies Assumptions 2 and 3, so the interval value function for each state exists, and further, $\widetilde{\mathcal{M}}$ the maximum probability reachability problem can be solved by interval value iteration and the convergence is assured by Theorem 4.

Note that we do not assume the existence of a proper policy [23]. Convergence is guaranteed without this assumption. The reason for that is twofold: (i) rewards are all 0 except for the destination state, and (ii) the destination state goes with probability 1 to the terminal state that is absorbing and reward-free. Therefore, even if there is a cycle, it does not add to the total reward.

5. Computational complexity

In this section, we show the polynomial time complexity of the interval value iteration for the ETRC. Our discussion is based on BMDP models in which there is a reward-free and absorbing state, i.e. the terminal state. Without loss of generality, we assume that q_1 is the terminal state. The interval value function of the destination state is set to be $[1,1]$. The initial interval value of each of the other states is set to be $[0,0]$.

States from which the terminal state is not accessible do not affect the result of the interval value iteration algorithm, because their interval value functions will remain $[0,0]$ as the algorithm proceeds. This is because if the terminal state is not accessible from a state then the destination states are also not accessible from it.

In order to prove the polynomial complexity of the interval value iteration, we consider the following assumption:

Assumption 4. The terminal state is accessible from all the other states.

Given a BMDP, it is possible that there exists a set of states such that once the set is entered, there exists a policy that will keep the state in the set for ever. In MDPs, such sets of states are called *end components* [5] or *controllably recurrent states* [4] and can be defined similarly for BMDPs. In the verification problem considered in this paper, such an end component will incur zero-reward and it can be replaced by a state so that the transformed model satisfies Assumption 4 [5]. End components can be computed in polynomial time [5], and therefore, if we will show that the interval value iteration algorithms are polynomial under Assumption 4 then it is polynomial for every BMDP.

Lemma 5. *Let* $X = \{x \in \mathbb{R}^n \mid x \geqslant_{\text{dom}} 0, x_1 = 0\}$ *and suppose Assumption* 4 *holds. Then* (a) $\overline{IVI}_{\text{opt}}$ *is a contraction mapping with respect to some weighted norm* $\|\cdot\|_\infty^w$ *over X,* (b) *for any value function W and the associated action selection function* (3), $\underline{IVI}_{W,\text{opt}}$ *is a contraction mapping with respect to some weighted norm* $\|\cdot\|_\infty^w$ *over X.*

The following theorem shows that interval value iteration algorithm is polynomial and is based on a similar argument of [12].

Theorem 6. *Interval value iteration converges to the desired interval value function in a number of steps polynomial in the number of states, the number of actions, and the number of bits used to represent the BMDP parameters.*

6. Conclusions

The results described in this paper show that interval value iteration with proper restrictions on the reward and transition functions can be used to solve BMDPs under the expected total reward criterion. These results allow us to solve a variety of new problems for BMDPs. The paper focuses on the maximum probability reachability problem for uncertain systems. Additional problems and extension to other probabilistic models used for verification are subject of current and future work.

Acknowledgements

This work was partially supported in part by the National Science Foundation under NSF Career award CNS-0347440. We would also like to thank the anonymous reviewer for identifying errors in the presentation of the technical results.

Appendix A. Proofs

Proof of Theorem 1. Denote by e the vector $[1, 1, \ldots, 1]$. For any $\varepsilon > 0$ there exists a $\pi_1 \in \Pi$ which satisfies $\overline{V}_{\pi_1} \geqslant_{\text{dom}} \overline{V}_{\text{opt}} - \varepsilon e$. By the definition of $\overline{V}_{\text{opt}}$, we have $\overline{V}_{\text{opt}} \geqslant_{\text{dom}} VI_{\overline{M}(\pi),\pi}(\overline{V}_{\pi_1}) = R + F_{\overline{M}(\pi),\pi}\overline{V}_{\pi_1}$ for any $\pi \in \Pi$. It follows that

$$\begin{aligned}\overline{V}_{\text{opt}} &\geqslant_{\text{dom}} \sup_{\pi \in \Pi}\{R + F_{\overline{M}(\pi),\pi}\overline{V}_{\pi_1}\} \geqslant_{\text{dom}} \sup_{\pi \in \Pi}\{R + F_{\overline{M}(\pi),\pi}\overline{V}_{\text{opt}}\} - \varepsilon e \\ &= \sup_{\pi \in \Pi} VI_{\overline{M}(\pi),\pi}(\overline{V}_{\text{opt}}) - \varepsilon e = \overline{IVI}_{\text{opt}}(\overline{V}_{\text{opt}}) - \varepsilon e.\end{aligned}$$

Hence $\overline{V}_{\text{opt}} \geqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}_{\text{opt}})$. For any fixed $q \in \mathcal{Q}$ and $\varepsilon > 0$ there exists a $\pi \in \Pi$ which satisfies $\overline{V}_\pi(q) \geqslant \overline{V}_{\text{opt}}(q) - \varepsilon$. It follows that

$$\overline{V}_{\text{opt}}(q) - \varepsilon \leqslant VI_{\overline{M}(\pi),\pi}(\overline{V}_\pi)(q) \leqslant VI_{\overline{M}(\pi),\pi}(\overline{V}_{\text{opt}})(q) \leqslant \sup_{\pi \in \Pi} VI_{\overline{M}(\pi),\pi}(\overline{V}_{\text{opt}})(q) = \overline{IVI}_{\text{opt}}(\overline{V}_{\text{opt}})(q).$$

Hence $\overline{V}_{\text{opt}} \leqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}_{\text{opt}})$. Since both $\overline{V}_{\text{opt}} \leqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}_{\text{opt}})$ and $\overline{V}_{\text{opt}} \geqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}_{\text{opt}})$ hold, (a) follows. The proof of (b) is similar. □

Proof of Lemma 2. For any $\varepsilon > 0$, there exist $M_1 \in \mathcal{M}$ and $\pi_1 \in \Pi$ such that

$$\begin{aligned}\overline{IVI}_{\text{opt}}(U) &= \sup_{\pi \in \Pi} VI_{\overline{M}(\pi),\pi}(U) \leqslant_{\text{dom}} R + F_{M_1,\pi_1}U + \varepsilon e \\ &\leqslant_{\text{dom}} R + F_{M_1,\pi_1}V + \varepsilon e \leqslant_{\text{dom}} \sup_{\pi \in \Pi} VI_{\overline{M}(\pi),\pi}(V) + \varepsilon e\overline{IVI}_{\text{opt}}(V) + \varepsilon e.\end{aligned}$$

Since ε was chosen arbitrarily, (a) holds. The proof of (b) is similar. □

Proof of Lemma 3. We denote $F^n_{M,\pi}$ the n-step transition probability matrix and V^n the sequence $V^n = VI^n_{M,\pi} V^0$. The upper bound of the interval value function for any $\pi \in \Pi$ can be defined as $\overline{V}_\pi = \lim_{N\to\infty} \sum_{n=1}^{N} F^{n-1}_{\overline{M}(\pi),\pi} R$ since by Assumption 2 the limit is well defined. Further, by Assumption 3 $\overline{V}_\pi < \infty$ and we have

$$\overline{V}_{\text{opt}} = \sup_{\pi\in\Pi} \overline{V}_\pi \geqslant_{\text{dom}} 0. \tag{A.1}$$

Suppose that there exists a $V \in \mathcal{V}^+$, for which $V = \overline{IVI}_{\text{opt}}(V)$. By definition of $\overline{IVI}_{\text{opt}}$, $V = \overline{IVI}_{\text{opt}}(V) \geqslant_{\text{dom}} VI_{\overline{M}(\pi),\pi}(V) = R + F_{\overline{M}(\pi),\pi} V$ for all $\pi \in \Pi$. Hence

$$\begin{aligned} V &\geqslant_{\text{dom}} R + F_{\overline{M}(\pi),\pi} V \geqslant_{\text{dom}} R + F_{\overline{M}(\pi),\pi} R + F^2_{\overline{M}(\pi),\pi} V \\ &\geqslant_{\text{dom}} \sum_{n=1}^{N} F^{n-1}_{\overline{M}(\pi),\pi} R + F^N_{\overline{M}(\pi),\pi} V = V^{N+1}_\pi + F^N_{\overline{M}(\pi),\pi} V. \end{aligned}$$

Since $V \in \mathcal{V}^+$, $F^N_{\overline{M}(\pi),\pi} V \geqslant_{\text{dom}} 0$, so that $V \geqslant_{\text{dom}} V^{N+1}_\pi$ for all N. Thus $V \geqslant_{\text{dom}} V_\pi$ for all $\pi \in \Pi$. Hence

$$\forall V \in \mathcal{V}^+ . V = \overline{IVI}_{\text{opt}}(V) \Longrightarrow V \geqslant_{\text{dom}} \sup_{\pi\in\Pi} V_\pi = \overline{V}_{\text{opt}}. \tag{A.2}$$

Theorem 1 shows that $\overline{V}_{\text{opt}}$ satisfies the optimality equation. Together with (A.1) and (A.2), $\overline{V}_{\text{opt}}$ is the minimal solution of $V = \overline{IVI}_{\text{opt}}(V)$. The proof of (b) is similar. □

Proof of Theorem 4. We first proof that the sequence $\{\overline{V}^n\}$ defined by $\overline{V}^n = \overline{IVI}^n_{\text{opt}}(\overline{V}^0)$ converges point-wise and monotonically to $\overline{V}_{\text{opt}}$. By Assumption 2, $\overline{IVI}_{\text{opt}}(0) \geqslant_{\text{dom}} 0$, so according to Lemma 2(a), $\{\overline{V}^n\}$ increases monotonically. Also, for each $q \in \mathcal{Q}$, $\overline{V}^n(q)$ is finite. Hence $\overline{V}^n(q)$ is a monotonically increasing and bounded series, thus $\lim_{n\to\infty} \overline{V}^n(q) = \sup_n \{\overline{V}^n(q)\} = \overline{V}(q)$ exists. Since $\overline{V} \geqslant_{\text{dom}} \overline{V}^n$, Lemma 2(a) implies that $\overline{IVI}_{\text{opt}}(\overline{V}) \geqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}^n) = \overline{V}^{n+1}$ for all n. Therefore $\overline{IVI}_{\text{opt}}(\overline{V}) \geqslant_{\text{dom}} \overline{V}$. For any $\pi \in \Pi$, and all n,

$$\overline{IVI}_\pi(\overline{V}^n) \leqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}^n) = \overline{V}^{n+1} \leqslant_{\text{dom}} \overline{V} \leqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V}). \tag{A.3}$$

According to the monotone convergence theorem, $\lim_{n\to\infty} \overline{IVI}_\pi(\overline{V}^n) = \overline{IVI}_\pi(\overline{V})$ for each $\pi \in \Pi$. Together with (A.3) we have $\overline{IVI}_{\text{opt}}(\overline{V}) = \sup_{\pi\in\Pi} \overline{IVI}_\pi(\overline{V}) \leqslant_{\text{dom}} \overline{V} \leqslant_{\text{dom}} \overline{IVI}_{\text{opt}}(\overline{V})$, which implies that $\overline{V}$ is a fixed point of $\overline{IVI}_{\text{opt}}$. Choosing $\varepsilon > 0$ and a sequence $\{\varepsilon_n\}$ which satisfies $\sum_{n=1}^{\infty} \varepsilon_n = \varepsilon$, there exists a policy π, which satisfies $\sum_{n=0}^{N} F^n_{\overline{M}(\pi),\pi} R + \sum_{n=1}^{N} \varepsilon_n e \geqslant_{\text{dom}} \overline{IVI}^N_{\text{opt}}(0)$ for all N. Hence $V_\pi + \varepsilon e \geqslant_{\text{dom}} \overline{V}$, implying $\overline{V} \leqslant_{\text{dom}} \overline{V}_{\text{opt}}$. By Lemma 3(a), $\overline{V}_{\text{opt}}$ is the minimal solution of the optimality equation, so $\overline{V} = \overline{V}_{\text{opt}}$. Similarly, we can prove that the sequence $\{\underline{V}^n\}$ defined by $\underline{V}^n = \underline{IVI}^n_{\text{opt},W}(\underline{V}^0)$ converges point-wise and monotonically to $\underline{V}_{\text{opt},W}$ for any value function W and the associated action selection function (3). By the definition of $\widehat{IVI}_{\text{opt}}$ in Section 2, $\hat{V}^n = \widehat{IVI}^n_{\text{opt}}(\hat{V}^0)$ must converge point-wise and monotonically to $\hat{V}_{\text{opt}}$. □

Proof of Lemma 5. We follow the approach of [24]. Suppose $\widehat{\mathcal{M}} = \langle \mathcal{Q}, \mathcal{A}, \hat{F}, R \rangle$ is a BMDP and q_1 is the terminal state for which $\hat{F}_{q_1,q_1}(\alpha) = [1,1]$ for any $\alpha \in \mathcal{A}$ and $R(q_1) = 0$. For any fixed $M \in \mathcal{M}$, the set of remaining states $\mathcal{Q}\backslash\{q_1\} = \{q_2,\ldots,q_n\}$ can be partitioned into nonempty subsets $\mathcal{Q}^M_1,\ldots,\mathcal{Q}^M_{m^M}$ such that for any $1 \leqslant i \leqslant m^M$, $p \in \mathcal{Q}^M_i$ and $\alpha \in \mathcal{A}$, there exists some state $r \in \{q_1\} \cup \mathcal{Q}^M_1 \cup \cdots \cup \mathcal{Q}^M_{i-1}$ such that $\overline{F}^M_{p,r}(\alpha) > 0$ (the choice of r depends on both p and α). Construct a set of weights $\{w^M_2,\ldots,w^M_n\}$ as

$$w^M_j = \sum_{i=1}^{m^M} \left\{1 - (\eta^M)^{2i}\right\}^{I_{\mathcal{Q}^M_i}(q_j)} \tag{A.4}$$

where

$$\eta^M = \min\left\{F^M_{p,q}(\alpha) \mid p,q \in \mathcal{Q}, \alpha \in \mathcal{A} \text{ such that } F^M_{p,q}(\alpha) > 0\right\}.$$

Since $0 < \eta^M < 1$, from (A.4) we have $0 < w^M_j < 1$ for any $2 \leqslant j \leqslant n$. Suppose q_j $(2 \leqslant j \leqslant n)$ is in $\mathcal{Q}^M_i$ $(1 \leqslant i \leqslant m^M)$. For any $\alpha \in \mathcal{A}$, there exists some state $q_l \in \{q_1\} \cup \mathcal{Q}^M_1 \cup \cdots \cup \mathcal{Q}^M_{i-1}$ such that $F^M_{q_j,q_l}(\alpha) > 0$. Let $\gamma^M =$

$(1-\eta^{2m^M-1})/(1-\eta^{2m^M})$. We have

$$\begin{aligned}\left(\sum_{k=1}^{n} F^M_{q_j,q_k}(\alpha)w^M_k\right)/w^M_j &\leqslant \left(\sum_{k=1,k\neq l}^{n} F^M_{q_j,q_k}(\alpha)+F^M_{q_j,q_l}(\alpha)w^M_l\right)/w^M_j\\ &= \big(1+F^M_{q_j,q_l}(\alpha)(w^M_l-1)\big)/w^M_j \leqslant \big(1+\eta^M(w^M_l-1)\big)/w^M_j\\ &\leqslant \big\{1-(\eta^M)^{2i-1}\big\}/w^M_j\big\{1-(\eta^M)^{2i-1}\big\}/\big\{1-(\eta^M)^{2i}\big\}\leqslant \gamma^M,\end{aligned}$$

where the second inequality follows from the fact $F^M_{q_j,q_l}(\alpha)\geqslant \eta^M$ and the third inequality follows from the fact $w^M_l \leqslant 1-(\eta^M)^{2i-2}$.

Let $\hat{U}\in\widehat{\mathcal{V}}$ and $\hat{V}\in\widehat{\mathcal{V}}$ be interval value functions. For fixed $p\neq q_1$, assume $\overline{IVI}_{\text{opt}}(\hat{U})(p)\leqslant\overline{IVI}_{\text{opt}}(\hat{V})(p)$. Select $M\in\mathcal{M}$ and $\alpha\in\mathcal{A}$ to maximize the expression $VI_{M,\alpha}(\overline{V})(p)$. Then

$$\begin{aligned}0&\leqslant \overline{IVI}_{\text{opt}}(\hat{V})(p)-\overline{IVI}_{\text{opt}}(\hat{U})(p)\\ &=\max_{\alpha\in\mathcal{A}}\max_{M\in\mathcal{M}} VI_{M,\alpha}(\overline{V})(p)-\max_{\alpha\in\mathcal{A}}\max_{M\in\mathcal{M}} VI_{M,\alpha}(\overline{U})(p)\\ &\leqslant R(p)+\sum_{q\in\mathcal{Q}}\big(F^M_{p,q}(\alpha)\overline{V}(q)\big)-R(p)-\sum_{q\in\mathcal{Q}}\big(F^M_{p,q}(\alpha)\overline{U}(q)\big)\\ &=\sum_{q\neq q_1}\big(F^M_{p,q}(\alpha)w^M_q\big)\big(\overline{U}(q)-\overline{V}(q)\big)/w^M_q\\ &\leqslant \gamma^M w^M_p\max_q\big\{\big(\overline{U}(q)-\overline{V}(q)\big)/w^M_q\big\}.\end{aligned}$$

It follows that

$$\big\|\overline{IVI}_{\text{opt}}(\hat{U})-\overline{IVI}_{\text{opt}}(\hat{V})\big\|^{\mathbf{w}^M}\leqslant\gamma^M\|\overline{U}-\overline{V}\|^{\mathbf{w}^M},$$

where $\|\cdot\|^{\mathbf{w}^M}$ denotes the maximum norm $\|\cdot\|$ scaled by $\mathbf{w}^M=(1,w^M_2,\dots,w^M_n)$, i.e. $\|x\|^{\mathbf{w}^M}=\|(x_1,x_2/w^M_2,\dots,x_n/w^M_n)\|$. Note that the maximizing MDP M is independent of $\overline{V}$ and γ^M depends only on the transition matrix of M, therefore $\overline{IVI}_{\text{opt}}$ is a contraction mapping.

The proof of (b) is similar. The construction of the weights $\mathbf{w}'^M=(1,w'^M_2,\dots,w'^M_n)$ used here is a little different from that of $\mathbf{w}^M$, in the sense that the set of available choices of actions for each state $p\in\mathcal{Q}$ is no longer in $\mathcal{A}$ but in $\rho_W(p)$. □

Proof of Theorem 6. By Eq. (4), the iteration operator $\widehat{IVI}_{\text{opt}}$ works in the following way: the upper bound of the interval value function will first converge; once the upper bound has converged, the iteration will continue until the lower bound converges. The first stage is polynomial because: (a) By Lemma 5, $\overline{IVI}_{\text{opt}}$ is a contraction mapping on the upper bound value function with respect to some weighted norm over a subset of $\mathbb{R}^n$, and thus the successive estimates of $\overline{V}_{\text{opt}}$ produced converge geometrically to the unique fixed-point. (b) By Theorem 4, the unique fixed-point is the desired upper bound value function. (c) The upper bound of the true $\hat{V}_{\text{opt}}$ is the optimal value function in π-maximizing MDP in $\widehat{\mathcal{M}}$. (d) The parameters for the MDPs in $\widehat{\mathcal{M}}$ can be specified with a number of bits polynomial in the number of bits used to specify the BMDP parameters. (e) Since $\overline{IVI}_{\text{opt}}$ is a contraction mapping, by an argument similar to [24], the upper bound will converge in a number of steps that is polynomial in the number of states, the number of actions, and the number of bits used to represent the BMDP parameters. Similarly, the second stage is also polynomial. Hence the value iteration algorithm is polynomial. □

References

[1] H.L.S. Younes, D.J. Musliner, Probabilistic plan verification through acceptance sampling, in: Proceedings of the AIPS-02 Workshop on Planning via Model Checking, 2002, pp. 81–88.

[2] J. Heath, M. Kwiatkowska, G. Norman, D. Parker, O. Tymchyshun, Probabilistic model checking of complex biological pathways, in: Proc. International Conference on Computational Methods in Systems Biology, 2006.

[3] H. Pham, Minimizing shortfall risk and applications to finance and insurance problems, The Annals of Applied Probability 312 (1) (2002) 143–172.

[4] C. Courcoubetis, M. Yannakakis, Markov decision processes and regular events, IEEE Transaction on Automatic Control 43 (10) (1998) 1399–1418.

[5] L. de Alfaro, Formal verification of probabilistic systems, PhD thesis, Tech. Rep. STAN-CS-TR-98-1601, Stanford University, Stanford, CA, 1997.

[6] J. Rutten, M. Kwiatkowska, G. Norman, D. Parker, Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems, CRM Monograph Series, vol. 23, American Mathematical Society, 2004.

[7] M. Jaeger, Probabilistic decision graphs—combining verification and AI techniques for probabilistic inference, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (2004) 19–42.

[8] H.J. Kushner, P. Dupuis, Numerical Methods for Stochastic Control Problems in Continuous Time, second ed., Springer-Verlag, New York, 2001.

[9] E. Feinberg, A. Shwartz, Handbook of Markov Decision Processes: Methods and Applications, Kluwer Academic Publishers, Boston, MA, 2002.

[10] A. Nilim, L.E. Ghaoui, Robust control of Markov decision processes with uncertain transition matrices, Operations Research 53 (5).

[11] D. Wu, X.D. Koutsoukos, Probabilistic verification of bounded-parameter Markov decision processes, in: V. Torra, Y. Narukawa, S. Miyamoto (Eds.), Modeling Decisions for Artificial Intelligence, MDAI 2006, Tarragona, Catalonia, Spain, April 3–5, 2006, in: Lecture Notes in Artificial Intelligence, vol. 3885, Springer, 2006, pp. 283–294.

[12] R. Givan, S. Leach, T. Dean, Bounded-parameter Markov decision process, Artificial Intelligence 122 (1–2) (2000) 71–109.

[13] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, Inc., New York, 1994.

[14] J.K. Satia, R.E. Lave, Markovian decision processes with uncertain transition probabilities, Operations Research 39 (1953) 1095–1100.

[15] C.C. White, H.K. Eldeib, Markov decision processes with imprecise transition probabilities, Operations Research 43 (1994) 739–749.

[16] J.A. Bagnell, A.Y. Ng, J.G. Schneider, Solving uncertain Markov decision problems, Tech. Rep. CMU-RI-TR-01-25, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 2001.

[17] M. Kurano, M. Yasuda, J. Nakagami, Interval methods for uncertain Markov decision processes, Markov Processes and Controlled Markov Chains (2002) 223–232.

[18] A. Tewari, P.L. Bartlett, Bounded parameter Markov decision processes with average reward criterion, in: Proceedings of the 20th Annual Conference on Learning Theory, in: Lecture Notes in Computer Science, vol. 4539, Springer, 2007, pp. 263–277.

[19] S. Kalyanasundaram, E.K.P. Chong, N.B. Shroff, Markovian decision processes with uncertain transition rates: Sensitivity and robust control, in: Proc. 41th IEEE Conference on Decision and Control, 2002.

[20] O. Buffet, Reachability analysis for uncertain ssps, in: ICTAI '05: Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence, 2005, pp. 515–522.

[21] I.O. Kozine, L.V. Utkin, Interval-valued finite Markov chains, Reliable Computing 8 (2).

[22] K. Sen, M. Viswanathan, G. Agha, Model-checking Markov chains in the presence of uncertainties, in: TACAS, in: Lecture Notes in Computer Science, vol. 3920, Springer, 2006, pp. 394–410.

[23] D.P. Bertsekas, J.N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Inc., Upper Saddle River, NJ, 1989.

[24] P. Tseng, Solving H-horizon, stationary Markov decision problems in time proportional to log(H)*, Operations Research Letters 9 (5) (1990) 287–297.