4OR-Q J Oper Res (2017) 15:67–84
DOI 10.1007/s10288-016-0321-3

RESEARCH PAPER

Finite approximation for finite-horizon continuous-time Markov decision processes

Qingda Wei[1]

Received: 15 September 2015 / Revised: 26 February 2016 / Published online: 11 June 2016

Abstract In this paper we study the continuous-time Markov decision processes with a denumerable state space, a Borel action space, and unbounded transition and cost rates. The optimality criterion to be considered is the finite-horizon expected total cost criterion. Under the suitable conditions, we propose a finite approximation for the approximate computations of an optimal policy and the value function, and obtain the corresponding error estimations. Furthermore, our main results are illustrated with a controlled birth and death system.

Keywords Continuous-time Markov decision processes · Finite-horizon expected total cost criterion · Unbounded transition rates · Finite approximation

Mathematics Subject Classification 93E20 · 90C40

1 Introduction

Continuous-time Markov decision processes (CTMDPs) have wide applications to many areas, such as queueing systems, epidemiology, and telecommunication; see, for instance, Puterman (1994), Kitaev and Rykov (1995) and Guo and Hernández-Lerma (2009). Since the time interval in the real-world applications is always finite, it is meaningful to study the CTMDPs under the finite-horizon expected criterion. As is well known, a common approach to prove the existence of optimal policies for the finite-horizon expected total cost criterion is via the optimality equation which has been

✉ Qingda Wei
weiqd@hqu.edu.cn

1 School of Economics and Finance, Huaqiao University, Quanzhou 362021, People's Republic of China

established under the different optimality conditions. More precisely, Miller (1968) dealt with the case of finite states and finite actions. Yushkevich (1977) and Bäuerle and Rieder (2011) discussed the case of a denumerable state space and bounded transition rates. Pliska (1975) and Gihman and Skohorod (1979) considered the case of a Borel state space and bounded transition rates. van Dijk (1988) and Guo et al. (2015) investigated the case of a denumerable state space and unbounded transition rates. Wei and Chen (2014) studied the case of a Borel state space and unbounded transition rates. All the aforementioned works discussed the finite-horizon expected total cost criterion in the class of all Markov policies except Yushkevich (1977) and Guo et al. (2015) which deal with the finite-horizon CTMDPs in the more general class of all history-dependent policies. They focus on showing that the value function is a unique solution to the optimality equation and that there exists an optimal Markov policy. Moreover, Guo et al. (2015) gave an example in which the precise forms of an optimal policy and the value function were obtained for some special cases, and solved an open problem proposed in Yushkevich (1977). However, the solution to the optimality equation cannot be solved explicitly in most cases. Hence, it is desirable to study the numerical methods for the approximate computations of an optimal policy and the value function. van Dijk (1988, 1989) used a method of time discretization to develop an approximation for the approximate computations of an optimal policy and the value function for the case of the denumerable state and action spaces and unbounded transition and cost rates, and obtained the orders of the accuracy of the approximation. From the theoretical viewpoint, the approximation in van Dijk (1988, 1989) can be computed by the recursive discrete-time dynamic programming. From the computational viewpoint, how to realize the approximation for the case of the denumerable state and action spaces is not involved in van Dijk (1988, 1989). In view of applications, it is of great importance to investigate the tractable numerical methods for the case of a denumerable state space and an uncountable action space.

In this paper we further study the issue of the approximate computations of an optimal policy and the value function for the finite-horizon expected total cost criterion basing on the existence results in Guo et al. (2015). The state space is a denumerable set and the action space is a Borel space. The transition and cost rates are allowed to be unbounded. Under the suitable conditions, we propose a finite approximation for the approximate computations of the value function and an optimal policy. To be more specific, this approximation can be divided into the following two steps. (I) The first step is to construct a sequence of the control models $\{\mathcal{M}_n, n \geq 1\}$ satisfying that the state space and the set of all admissible actions are finite sets, and that the corresponding value functions converge to the value function of the original control model $\mathcal{M}$. To this end, we define the state space and transition rates of the control model $\mathcal{M}_n$ by employing the finite truncation of the corresponding elements of the original control model $\mathcal{M}$, and choose the set of all admissible actions of the control model $\mathcal{M}_n$ satisfying a certain condition. (II) The second step is to construct a suitable value iteration for approximating the value function of the control model $\mathcal{M}_n$. Since there are uncountable points in the finite time interval, it is infeasible to compute the value function directly from the computational viewpoint. Hence, we divide the time interval into several equal parts, and present a value iteration using a technique of time discretization and the optimality equation [see (4.11)]. Moreover, we can obtain an approximate optimal policy of the control model $\mathcal{M}$ as in the form of (4.12) from executing the iteration procedure. It should be mentioned that the method of time discretization in this paper is different from that in van Dijk (1988, 1989). More specifically, van Dijk (1988, 1989) employed the technique of time discretization to construct a discrete-time MDP with a fixed step size in the time parameter and then gave an approximation for the approximate computations of an optimal policy and the value function by solving the finite-horizon optimality equation of the discrete-time MDP. We use the approach of time discretization to obtain a partition of the time interval and directly provide a finite approximation for the approximate computations of an optimal policy and the value function without introducing an auxiliary discrete-time MDP as in van Dijk (1988, 1989).

Furthermore, we analyze the accuracy of the finite approximation. For the first step, we obtain an error estimation between the value function of the control model $\mathcal{M}_n$ and that of the original control model $\mathcal{M}$ by suitably choosing the set of all admissible actions of the control model $\mathcal{M}_n$ (see Theorem 4.1). For the second step, we give an error estimation between the values obtained from the iteration and the value function of the control model $\mathcal{M}_n$ at the discrete time points, and an error estimation of an approximate optimal policy of the control model $\mathcal{M}$ (see Theorem 4.2). Finally, we use a controlled birth and death system to illustrate the application of the finite approximation.

The rest of this paper is organized as follows. In Sect. 2, we introduce the control model and optimality criterion. In Sect. 3, we give the optimality conditions for the existence of optimal policies and preliminary results. In Sect. 4, we state and prove our main results. In Sect. 5, we illustrate our main results with a controlled birth and death system.

2 The control model

In this paper we consider the following model

$$\mathcal{M} := \{S, A, (A(i), i \in S), q(j|i, a), c(i, a)\},$$

where the state space S is assumed to be the set of all nonnegative integers endowed with the discrete topology and the action space A is assumed to be a Borel space with the Borel σ-algebra $\mathcal{B}(A)$. $A(i) \in \mathcal{B}(A)$ represents the set of all admissible actions when the state of the system is $i \in S$. Let $K := \{(i, a)|i \in S, a \in A(i)\}$ be the set of all the feasible state-action pairs. The transition rate $q(j|i, a)$ is supposed to satisfy the following properties:

- For each fixed $i, j \in S$, $q(j|i, a)$ is measurable in $a \in A(i)$;
- $q(j|i, a) \geq 0$ for all $(i, a) \in K$ and $j \neq i$;
- $\sum_{j\in S} q(j|i, a) = 0$ for all $(i, a) \in K$;
- $q^*(i) := \sup_{a\in A(i)} |q(i|i, a)| < \infty$ for all $i \in S$.

Finally, the real-valued cost rate function $c(i, a)$ is measurable in $a \in A(i)$ for each $i \in S$.

A continuous-time Markov decision process evolves as follows. A decision-maker observes continuously the state of the system. If the system occupies state i, he/she chooses an action $a \in A(i)$ according to some decision rule. As a consequence of this, the following happens: (i) a cost takes place at the rate $c(i, a)$; (ii) the system remains in the state i for a random time following the exponential distribution with the tail function given by $e^{q(i|i,a)t}$, and then jumps to a new state j with the probability $-\frac{q(j|i,a)}{q(i|i,a)}$ (we make a convention that $\frac{0}{0} := 0$).

To define the optimality criterion, we need to introduce the definition of a policy. From Theorem 4.1 in Guo et al. (2015), we have that there exists an optimal Markov policy over the class of all randomized history-dependent policies for the finite-horizon expected total cost criterion. Thus, without loss of generality, we restrict the discussions to the class of all randomized Markov policies throughout the paper.

Definition 2.1 A randomized Markov policy is a family $\pi := \{\pi_t, t \geq 0\}$ of stochastic kernels that satisfy

(i) for each $t \geq 0$, π_t is a stochastic kernel on A given S such that $\pi_t(A(i)|i) = 1$ for all $i \in S$;
(ii) for each $i \in S$ and $D \in \mathcal{B}(A)$, $\pi_t(D|i)$ is Borel measurable in $t \in [0, \infty)$.

A policy π is called deterministic Markov if there exists a measurable function f on $S \times [0, \infty)$ with $f(i, t) \in A(i)$, such that $\pi_t(\cdot|i)$ is the Dirac measure concentrated at $f(i, t) \in A(i)$ for all $(i, t) \in S \times [0, \infty)$.

We denote by Π the set of all randomized Markov policies and by Π^D the set of all deterministic Markov policies.

For any $\pi \in \Pi$ and any initial state $i \in S$, by Theorem 4.27 in Kitaev and Rykov (1995), there exist a unique probability measure P_i^π on some measurable space $(\Omega, \mathcal{B}(\Omega))$ and a state process $\{\xi_t, t \geq 0\}$. Let E_i^π be the corresponding expectation operator with respect to P_i^π.

Fix an arbitrary constant $T > 0$ denoting the horizon of the CTMDPs. For any $i, j \in S$ and $\pi \in \Pi$, we define the finite-horizon expected total cost from time t to the terminal time T as follows:

$$V^\pi(t, i) := E_j^\pi \left[\int_t^T \int_A c(\xi_s, a)\pi_s(da|\xi_s)ds \middle| \xi_t = i \right]. \tag{2.1}$$

Since for each $\pi \in \Pi$, $\{\xi_t, t \geq 0\}$ is a Markov jump process, the definition of $V^\pi(t, i)$ is independent of the state $j \in S$. The corresponding value function is defined by

$$V^*(t, i) := \inf_{\pi \in \Pi} V^\pi(t, i) \text{ for all } (t, i) \in [0, T] \times S. \tag{2.2}$$

Definition 2.2 A policy $\pi^* \in \Pi$ is said to be optimal if $V^{\pi^*}(0, i) = V^*(0, i)$ for all $i \in S$.

3 Preliminaries

In this section, we give some basic assumptions and preliminary results.

The following conditions are used to avoid the explosion of the process $\{\xi_t, t \geq 0\}$ and guarantee the finiteness of the finite-horizon expected total cost criterion $V^\pi(\cdot)$ and the value function $V^*(\cdot)$; see, for instance, Guo and Hernández-Lerma (2009) and Guo and Zhang (2014).

Assumption 3.1 There exist a nondecreasing function $w \geq 1$ on S with $\lim_{i\to\infty} w(i) = \infty$, and constants $\rho_1 \in \mathbb{R} := (-\infty, +\infty)$, $d_1 \geq 0$, $Q > 0$ and $M > 0$ such that

(i) $\sum_{j\in S} w(j)q(j|i,a) \leq \rho_1 w(i) + d_1$ for all $(i,a) \in K$;
(ii) $q^*(i) \leq Qw(i)$ for all $i \in S$;
(iii) $|c(i,a)| \leq Mw(i)$ for all $(i,a) \in K$.

In addition to Assumption 3.1, we also need the following conditions.

Assumption 3.2 (i) There exist constants $\rho_2 \in \mathbb{R}$ and $d_2 \geq 0$ such that

$$\sum_{j\in S} w^2(j)q(j|i,a) \leq \rho_2 w^2(i) + d_2 \text{ for all } (i,a) \in K,$$

where w is as in Assumption 3.1.
(ii) For each $i \in S$, the set $A(i)$ is compact.
(iii) For any $i, j \in S$ there exist constants $L_i > 0$ and $L_{ij} > 0$ such that

$$|c(i,a) - c(i,b)| \leq L_i d_A(a,b) \text{ and } |q(j|i,a) - q(j|i,b)| \leq L_{ij} d_A(a,b)$$

for all $a, b \in A(i)$, where d_A denotes the metric of the space A.
(iv) For each $i \in S$, the function $\sum_{j\in S} w(j)q(j|i,a)$ is continuous in $a \in A(i)$.

Remark 3.1 Assumption 3.2(i) is used to obtain the Ito–Dynkin formula; see Theorem 3.1 in Guo et al. (2015). Assumption 3.2(ii) and (iv), the standard continuity and compactness conditions, together with Assumption 3.2(iii), are used to ensure the existence of optimal policies; see, for instance, Guo and Hernández-Lerma (2009), Guo and Zhang (2014), Wei and Chen (2014) and Guo et al. (2015). Assumption 3.2(iii), the so-called Lipschitz continuity condition, is also used to give the error estimations of the finite approximation.

Then we have the following results.

Lemma 3.1 *Suppose that Assumptions 3.1 and 3.2 are satisfied. Then the following statements hold.*

(i) $E_j^\pi[w^l(\xi_t)|\xi_s = i] \leq e^{\rho_l(t-s)}w^l(i) + \frac{d_l}{\rho_l}(e^{\rho_l(t-s)} - 1)$ *for all* $i, j \in S$, $\pi \in \Pi$, $t \geq s \geq 0$ *and* $l = 1, 2$ *(if* $\rho_l = 0$*, the righthand term is* $w^l(i) + d_l(t-s)$*).*
(ii) $|V^\pi(t,i)| \leq G_1 w(i)$ *for all* $(t,i) \in [0,T] \times S$ *and* $\pi \in \Pi$*, where* $G_1 = M[(\frac{1}{\rho_1} + \frac{d_1}{\rho_1^2})(e^{\rho_1 T} - 1) - \frac{d_1}{\rho_1}T]$ *(if* $\rho_1 = 0$*,* $G_1 = M(T + \frac{1}{2}d_1 T^2)$*).*

(iii) *The function* V^* *on* $[0, T] \times S$ *satisfies the following equation:*

$$-\frac{\partial V^*}{\partial t}(t, i) = \inf_{a \in A(i)} \left\{ c(i, a) + \sum_{j \in S} V^*(t, j) q(j|i, a) \right\} \tag{3.1}$$

for all $(t, i) \in [0, T] \times S$*, where* $\frac{\partial V^*}{\partial t}$ *denotes the derivative of* V^* *with respect to the variable* t*. Moreover, there exists* $f^* \in \Pi^D$ *with* $f^*(i, t) \in A(i)$ *attaining the infimum in (3.1) and the policy* f^* *is optimal.*

(iv) $|V^*(t_1, i) - V^*(t_2, i)| \le \left[M + G_1(\rho_1 I_{\{\rho_1 > 0\}} + d_1 + 2Q)\right] w^2(i)|t_1 - t_2|$ *for all* $i \in S$ *and* $t_1, t_2 \in [0, T]$*, where* I_D *denotes the indicator function of a set* D*.*

Proof (i) Part (i) follows from Lemma 6.3 in Guo and Hernández-Lerma (2009).

(ii) Fix any $i, j \in S$, $t \in [0, T]$ and $\pi \in \Pi$. Then using (2.1), Assumption 3.1(iii) and part (i), we have

$$\begin{aligned}
|V^\pi(t, i)| &\le M \int_t^T E_j^\pi[w(\xi_s)|\xi_t = i] ds \\
&\le M w(i) \int_t^T \left[e^{\rho_1(s-t)} + \frac{d_1}{\rho_1}(e^{\rho_1(s-t)} - 1) \right] ds \\
&= M \left[\frac{1}{\rho_1}(e^{\rho_1(T-t)} - 1) + \frac{d_1}{\rho_1^2}(e^{\rho_1(T-t)} - 1) - \frac{d_1}{\rho_1}(T - t) \right] w(i) \\
&\le M \left[\frac{1}{\rho_1}(e^{\rho_1 T} - 1) + \frac{d_1}{\rho_1^2}(e^{\rho_1 T} - 1) - \frac{d_1}{\rho_1} T \right] w(i)
\end{aligned}$$

for $\rho_1 \neq 0$, which implies the desired assertion.

(iii) Part (iii) follows from the same techniques of Theorem 4.1 in Guo et al. (2015) and Theorem 4.1 in Wei and Chen (2014).

(iv) By (2.2) and (3.1), we have

$$V^*(t, i) = \int_t^T \inf_{a \in A(i)} \left\{ c(i, a) + \sum_{j \in S} V^*(s, j) q(j|i, a) \right\} ds$$

for all $(t, i) \in [0, T] \times S$. Thus, direct calculations give

$$\begin{aligned}
|V^*(t_1, i) - V^*(t_2, i)| &= \left| \int_{t_1}^{t_2} \inf_{a \in A(i)} \left\{ c(i, a) + \sum_{j \in S} V^*(s, j) q(j|i, a) \right\} ds \right| \\
&\le \left[M + G_1(\rho_1 I_{\{\rho_1 > 0\}} + d_1 + 2Q)\right] w^2(i)|t_1 - t_2|
\end{aligned}$$

for all $i \in S$ and $t_1, t_2 \in [0, T]$, where the inequality follows from part (ii) and Assumption 3.1. This completes the proof of the lemma. □

4 Finite approximation

In this section, we present a finite approximation for the approximate computations of an optimal policy and the value function. To do so, we need to introduce the following notation.

For each integer $n \geq 1$, we define the control model

$$\mathcal{M}_n := \{S_n, A, (A_n(i), i \in S_n), q_n(j|i,a), c(i,a)\}$$

with the following elements.

- The state space is $S_n := \{0, 1, \ldots, j_n\}$, where the sequence $\{j_n, n \geq 1\}$ is increasing and $\lim_{n\to\infty} j_n = \infty$.
- The action space A is the same as in the model $\mathcal{M}$.
- $A_n(i)$, the set of all admissible actions in the state $i \in S_n$, is an arbitrary finite set. Let $K_n := \{(i,a)|i \in S_n, a \in A_n(i)\}$.
- For each $(i,a) \in K_n$ and $j \in S_n$, the transition rate $q_n(j|i,a)$ is given by

$$q_n(j|i,a) := \begin{cases} q(j|i,a), & \text{if } j \neq j_n, \\ \sum_{k \geq j_n} q(k|i,a), & \text{if } j = j_n. \end{cases}$$

- We still denote by c the restriction of the cost rate function c in the model $\mathcal{M}$ to K_n.

We denote by Π_n the set of all randomized Markov policies and by Π_n^D the set of all deterministic Markov policies for the control model $\mathcal{M}_n$. Moreover, for any $\pi \in \Pi_n$ and any initial state $i \in S_n$, employing Theorem 4.27 in Kitaev and Rykov (1995), there exists a probability measure $P_n^{i,\pi}$ on some measurable space $(\Omega_n, \mathcal{B}(\Omega_n))$. Denote by $E_n^{i,\pi}$ the corresponding expectation operator with respect to $P_n^{i,\pi}$. As in (2.1) and (2.2), we can also define the functions V_n^π and V_n^* on $[0,T] \times S_n$ with $E_n^{i,\pi}$ and Π_n in lieu of E_i^π and Π, respectively.

Let $\mathcal{C}$ be the set of all closed subsets of A. The Hausdorff metric on $\mathcal{C}$ is defined as

$$d_H(B_1, B_2) := \max\left\{\sup_{a\in B_1}\inf_{b\in B_2} d_A(a,b), \sup_{b\in B_2}\inf_{a\in B_1} d_A(a,b)\right\}$$

for all $B_1, B_2 \in \mathcal{C}$.

In order to obtain the error estimations of the finite approximation, we also need to impose the following condition.

Assumption 4.1 There exist constants $\delta > 2$, $\rho_\delta \in \mathbb{R}$ and $d_\delta \geq 0$ such that

$$\sum_{j\in S} w^\delta(j) q(j|i,a) \leq \rho_\delta w^\delta(i) + d_\delta \quad \text{for all } (i,a) \in K,$$

where w comes from Assumption 3.1.

Below we state our first main result on the error estimation between the value function of the control model $\mathcal{M}_n$ and that of the original control model $\mathcal{M}$.

Theorem 4.1 *Suppose that Assumptions 3.1, 3.2 and 4.1 hold. Let G_1 be as in Lemma 3.1. If there exists a constant $\widetilde{M} > 0$ such that for each $n \geq 1$ and $i \in S_n$, the set $A_n(i)$ satisfies*

$$d_H(A(i), A_n(i)) \leq \frac{\widetilde{M} w^\delta(i)}{w^{\delta-2}(j_n)\left[L_i + 2G_1 w(j_n) \sum_{j=0}^{j_n-1} L_{ij}\right]}, \tag{4.1}$$

then we have

$$|V^*(t,i) - V_n^*(t,i)| \leq \frac{Q_1 w^\delta(i)}{w^{\delta-2}(j_n)}$$

for all $t \in [0,T]$, where $Q_1 = \left[\widetilde{M} + 2G_1(|\rho_\delta| + d_\delta + Q)\right] \times [(\frac{1}{\rho_\delta} + \frac{d_\delta}{\rho_\delta^2})(e^{\rho_\delta T} - 1) - \frac{d_\delta}{\rho_\delta} T]$ (if $\rho_\delta = 0$, $Q_1 = \left[\widetilde{M} + 2G_1(d_\delta + Q)\right](T + \frac{1}{2} d_\delta T^2)$).

Proof Fix any $n \geq 1$ and $i \in S_n$. Let $f^* \in \Pi^D$ be as in Lemma 3.1(iii). Then we have

$$\begin{aligned} -\frac{\partial V^*}{\partial s}(s,i) &= c(i, f^*(i,s)) + \sum_{j \in S} V^*(s,j) q(j|i, f^*(i,s)) \\ &= c(i, f^*(i,s)) + \sum_{j=0}^{j_n-1} (V^*(s,j) - V^*(s,j_n)) q(j|i, f^*(i,s)) \\ &\quad + \sum_{j > j_n} (V^*(s,j) - V^*(s,j_n)) q(j|i, f^*(i,s)) \end{aligned} \tag{4.2}$$

for all $s \in [0,T]$. Moreover, direct calculations give

$$\begin{aligned} &\sum_{j > j_n} (V^*(s,j) - V^*(s,j_n)) q(j|i, f^*(i,s)) \\ &\geq -2G_1 \sum_{j > j_n} w(j) q(j|i, f^*(i,s)) \\ &\geq -\frac{2G_1}{w^{\delta-2}(j_n)} \sum_{j > j_n} w^{\delta-1}(j) q(j|i, f^*(i,s)) \\ &\geq -\frac{2G_1}{w^{\delta-2}(j_n)} \left[\sum_{j \in S} w^{\delta-1}(j) q(j|i, f^*(i,s)) - w^{\delta-1}(i) q(i|i, f^*(i,s))\right] \\ &\geq -2G_1(|\rho_\delta| + d_\delta + Q) \frac{w^\delta(i)}{w^{\delta-2}(j_n)} \end{aligned} \tag{4.3}$$

for all $s \in [0, T]$, where the first inequality follows from Lemma 3.1(ii), the second one is due to the monotonicity of w, and the last one follows from Assumptions 3.1(ii) and 4.1. For each $s \in [0, T]$, there exists $\widetilde{f}(i, s) \in A_n(i)$ satisfying

$$d_A(f^*(i,s), \widetilde{f}(i,s)) = \min_{a \in A_n(i)} d_A(f^*(i,s), a) \le d_H(A(i), A_n(i)),$$

which together with Lemma 3.1(ii) and Assumption 3.2(iii) yields

$$\begin{aligned} c(i, f^*(i,s)) - c(i, \widetilde{f}(i,s)) &\ge -L_i d_A(f^*(i,s), \widetilde{f}(i,s)) \\ &\ge -L_i d_H(A(i), A_n(i)) \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} &\sum_{j=0}^{j_n-1} (V^*(s,j) - V^*(s,j_n)) \left[q(j|i, f^*(i,s)) - q(j|i, \widetilde{f}(i,s)) \right] \\ &\ge -2G_1 w(j_n) \sum_{j=0}^{j_n-1} \left| q(j|i, f^*(i,s)) - q(j|i, \widetilde{f}(i,s)) \right| \\ &\ge -2G_1 w(j_n) \sum_{j=0}^{j_n-1} L_{ij} d_A(f^*(i,s), \widetilde{f}(i,s)) \\ &\ge -2G_1 w(j_n) d_H(A(i), A_n(i)) \sum_{j=0}^{j_n-1} L_{ij}. \end{aligned} \tag{4.5}$$

Hence, by (4.2)–(4.5) we obtain

$$\begin{aligned} -\frac{\partial V^*}{\partial s}(s,i) &\ge c(i, \widetilde{f}(i,s)) + \sum_{j=0}^{j_n-1} (V^*(s,j) - V^*(s,j_n)) q(j|i, \widetilde{f}(i,s)) \\ &\quad - \left[L_i + 2G_1 w(j_n) \sum_{j=0}^{j_n-1} L_{ij} \right] d_H(A(i), A_n(i)) \\ &\quad -2G_1(|\rho_\delta| + d_\delta + Q) \frac{w^\delta(i)}{w^{\delta-2}(j_n)} \\ &\ge c(i, \widetilde{f}(i,s)) + \sum_{j \in S_n} V^*(s,j) q_n(j|i, \widetilde{f}(i,s)) - \frac{\widetilde{M} w^\delta(i)}{w^{\delta-2}(j_n)} \\ &\quad -2G_1(|\rho_\delta| + d_\delta + Q) \frac{w^\delta(i)}{w^{\delta-2}(j_n)} \\ &= c(i, \widetilde{f}(i,s)) + \sum_{j \in S_n} V^*(s,j) q_n(j|i, \widetilde{f}(i,s)) - \frac{\widetilde{Q} w^\delta(i)}{w^{\delta-2}(j_n)} \end{aligned} \tag{4.6}$$

for all $s \in [0, T]$, where $\widetilde{Q} := \widetilde{M} + 2G_1(|\rho_\delta| + d_\delta + Q)$. Note that under Assumptions 3.1(i), (ii), 3.2(i) and 4.1, we have

$$q_n^*(i) := \sup_{a \in A_n(i)} |q_n(i|i,a)| \le q^*(i) \le Qw(i) \text{ and } \tag{4.7}$$

$$\begin{aligned}\sum_{j \in S_n} w^l(j) q_n(j|i,a) &= \sum_{j \in S_n} w^l(j) q(j|i,a) + w^l(j_n) \sum_{j \notin S_n} q(j|i,a) \\ &\le \sum_{j \in S} w^l(j) q(j|i,a) \le \rho_l w^l(i) + d_l \end{aligned}\tag{4.8}$$

for all $a \in A_n(i)$, which together with Lemma 6.3 in Guo and Hernández-Lerma (2009) imply

$$E_n^{j,\pi}[w^l(\xi_t)|\xi_s = i] \le e^{\rho_l(t-s)} w^l(i) + \frac{d_l}{\rho_l}(e^{\rho_l(t-s)} - 1) \tag{4.9}$$

for all $j \in S_n, \pi \in \Pi_n, t \ge s \ge 0$ and $l = 1, 2, \delta$ (if $\rho_l = 0$, the righthand term of (4.9) is $w^l(i) + d_l(t-s)$). Direct calculations give

$$\begin{aligned}&E_n^{j,\pi}\left[\int_t^T \left|\frac{\partial V^*}{\partial s}(s,\xi_s)\right| ds \middle| \xi_t = i\right] \le \left[M + G_1(\rho_1 I_{\{\rho_1>0\}} + d_1 + 2Q)\right] \\ &\times E_n^{j,\pi}\left[\int_t^T w^2(\xi_s) ds \middle| \xi_t = i\right] \le \left[M + G_1(\rho_1 I_{\{\rho_1>0\}} + d_1 + 2Q)\right] \\ &\times \left[\left(\frac{1}{\rho_2} + \frac{d_2}{\rho_2^2}\right)(e^{\rho_2 T} - 1) - \frac{d_2}{\rho_2} T\right] w^2(i) < \infty\end{aligned}$$

(if $\rho_2 = 0$, the last term is $\left[M + G_1(\rho_1 I_{\{\rho_1>0\}} + d_1 + 2Q)\right](T + \frac{1}{2}d_2T^2)w^2(i)$) for all $j \in S_n, \pi \in \Pi_n$ and $t \in [0, T]$, where the first inequality follows from Lemma 3.1 and Assumption 3.1, and the second one is due to (4.9). Moreover, by (4.6) we obtain

$$\begin{aligned}&-E_n^{j,\widetilde{f}}\left[\int_t^T \frac{\partial V^*}{\partial s}(s,\xi_s) ds \middle| \xi_t = i\right] \ge E_n^{j,\widetilde{f}}\left[\int_t^T c(\xi_s, \widetilde{f}(\xi_s, s)) ds \middle| \xi_t = i\right] \\ &+E_n^{j,\widetilde{f}}\left[\int_t^T \sum_{j \in S_n} V^*(s,j) q_n(j|\xi_s, \widetilde{f}(\xi_s, s)) ds \middle| \xi_t = i\right] \\ &-\frac{\widetilde{Q}}{w^{\delta-2}(j_n)} E_n^{j,\widetilde{f}}\left[\int_t^T w^\delta(\xi_s) ds \middle| \xi_t = i\right]\end{aligned}\tag{4.10}$$

for all $j \in S_n$ and $t \in [0, T]$. Employing Theorem 3.1 in Guo et al. (2015) we have

$$E_n^{j,\widetilde{f}}\left[\int_t^T \sum_{j\in S_n} V^*(s,j)q_n(j|\xi_s,\widetilde{f}(\xi_s,s))ds \middle| \xi_t = i\right]$$
$$= -V^*(t,i) - E_n^{j,\widetilde{f}}\left[\int_t^T \frac{\partial V^*}{\partial s}(s,\xi_s)ds \middle| \xi_t = i\right]$$

for all $j \in S_n$ and $t \in [0, T]$. Hence, the last equality, (4.9) and (4.10) imply

$$V^*(t,i) \geq V_n^{\widetilde{f}}(t,i) - \frac{\widetilde{Q}}{w^{\delta-2}(j_n)} E_n^{j,\widetilde{f}}\left[\int_t^T w^\delta(\xi_s)ds \middle| \xi_t = i\right]$$
$$\geq V_n^*(t,i) - \widetilde{Q}\left[\left(\frac{1}{\rho_\delta} + \frac{d_\delta}{\rho_\delta^2}\right)(e^{\rho_\delta T} - 1) - \frac{d_\delta}{\rho_\delta}T\right]\frac{w^\delta(i)}{w^{\delta-2}(j_n)}$$

(if $\rho_\delta = 0$, the last term is $V_n^*(t,i) - \widetilde{Q}(T + \frac{1}{2}d_\delta T^2)\frac{w^\delta(i)}{w^{\delta-2}(j_n)}$) for all $j \in S_n$ and $t \in [0, T]$. On the other hand, using the similar arguments, we have

$$V^*(t,i) \leq V_n^*(t,i) + \widetilde{Q}\left[\left(\frac{1}{\rho_\delta} + \frac{d_\delta}{\rho_\delta^2}\right)(e^{\rho_\delta T} - 1) - \frac{d_\delta}{\rho_\delta}T\right]\frac{w^\delta(i)}{w^{\delta-2}(j_n)}$$

for all $t \in [0, T]$. Therefore, we obtain the desired assertion. □

For each integer $m \geq 1$, a partition of the interval $[0, T]$ is as follows:

$$T =: t_0 > t_1 > \cdots > t_m := 0,$$

where $t_l := t_0 - \frac{T}{m}l$ for all $l = 0, 1, \ldots, m$. For each $n \geq 1$, we define the following iteration

$$W_m(t_l,i) := W_m(t_{l-1},i) + \frac{T}{m}\min_{a\in A_n(i)}\left\{c(i,a) + \sum_{j\in S_n} W_m(t_{l-1},j)q_n(j|i,a)\right\}$$
$$\text{with } W_m(t_0,i) = 0 \tag{4.11}$$

for all $i \in S_n$ and $l = 1, \ldots, m$. For each $n \geq 1$, $i \in S_n$ and $l \in \{1, \ldots, m\}$, let $\mathcal{D}_{n,l}(i)$ be the set of all the minimizers attaining the minimum in (4.11). For each $n \geq 1$ and $m \geq 1$, denote by $\mathcal{O}_{n,m}$ the set of all the policies with the following form:

$$f_{n,m}(i,t) := \begin{cases} g_{n,m}(i), & \text{if } t \in [0, t_{m-1}], \\ g_{n,l}(i), & \text{if } t \in (t_l, t_{l-1}] (l = 1, \ldots, m-1), \\ a^*, & \text{if } t > T, \end{cases} \tag{4.12}$$

for all $i \in S_n$, where $g_{n,l}(i)$ belongs to $\mathcal{D}_{n,l}(i)$ and $a^* \in A_n(i)$ is arbitrarily fixed.

Next, we give our second main result on the error estimations of the approximate computations of the value function and an optimal policy for the control model $\mathcal{M}$ using the above iteration defined by (4.11).

Theorem 4.2 *Suppose that the conditions in Theorem 4.1 are satisfied. Let Q_1 be as in Theorem 4.1. Then the following statements hold: for each $m \geq 1$ and $n \geq 1$,*

(i) $|V_n^*(t_l, i) - W_m(t_l, i)| \leq \frac{M_1 T w^2(j_n)}{2Qm}\left(e^{2QTw(j_n)} - 1\right)$ *for all $i \in S_n$ and $l = 0, 1, \ldots, m$, where $M_1 = \left[M + G_1(\rho_1 I_{\{\rho_1>0\}} + d_1 + 2Q)\right](\rho_2 I_{\{\rho_2>0\}} + d_2 + 2Q)$.*

(ii) $|V^*(0, i) - W_m(0, i)| \leq \frac{Q_1 w^\delta(i)}{w^{\delta-2}(j_n)} + \frac{M_1 T w^2(j_n)}{2Qm}\left(e^{2QTw(j_n)} - 1\right)$ *for all $i \in S_n$.*

(iii) *Any policy $f_{n,m} \in \mathcal{O}_{n,m}$ satisfies*

$$|V^*(0, i) - V_n^{f_{n,m}}(0, i)| \leq \frac{Q_1 w^\delta(i)}{w^{\delta-2}(j_n)} + \frac{M_1 T w^2(j_n)}{Qm}\left(e^{2QTw(j_n)} - 1\right)$$

for all $i \in S_n$.

Proof (i) Fix any $n \geq 1$, $i \in S_n$, $m \geq 1$ and $l \in \{1, \ldots, m\}$. By (4.7), (4.8) and the description of the control model $\mathcal{M}_n$, we see that Assumptions 3.1, 3.2 and 4.1 also hold for the transition rate $q_n(j|i, a)$. Then it follows from Lemma 3.1 that

$$-\frac{\partial V_n^*}{\partial t}(t, i) = \min_{a \in A_n(i)}\left\{c(i, a) + \sum_{j \in S_n} V_n^*(t, j) q_n(j|i, a)\right\} \tag{4.13}$$

and

$$|V_n^*(s, i) - V_n^*(t, i)| \leq \left[M + G_1(\rho_1 I_{\{\rho_1>0\}} + d_1 + 2Q)\right] w^2(i)|t - s| \tag{4.14}$$

for all $s, t \in [0, T]$. Thus, direct calculations give

$$\begin{aligned}
&|V_n^*(t_l, i) - W_m(t_l, i)| \\
&= \Bigg| V_n^*(t_{l-1}, i) + \int_{t_l}^{t_{l-1}} \min_{a \in A_n(i)}\left\{c(i, a) + \sum_{j \in S_n} V_n^*(t, j) q_n(j|i, a)\right\} dt \\
&\quad - W_m(t_{l-1}, i) - \frac{T}{m}\min_{a \in A_n(i)}\left\{c(i, a) + \sum_{j \in S_n} W_m(t_{l-1}, j) q_n(j|i, a)\right\}\Bigg| \\
&\leq |V_n^*(t_{l-1}, i) - W_m(t_{l-1}, i)| + \int_{t_l}^{t_{l-1}} \max_{a \in A_n(i)}\left|\sum_{j \in S_n}(V_n^*(t, j) - W_m(t_{l-1}, j)) q_n(j|i, a)\right| dt \\
&\leq |V_n^*(t_{l-1}, i) - W_m(t_{l-1}, i)| + \int_{t_l}^{t_{l-1}} \max_{a \in A_n(i)}\left|\sum_{j \in S_n}(V_n^*(t, j) - V_n^*(t_{l-1}, j)) q_n(j|i, a)\right| dt
\end{aligned}$$

$$
\begin{aligned}
&+\int_{t_l}^{t_{l-1}} \max_{a\in A_n(i)} \left| \sum_{j\in S_n} (V_n^*(t_{l-1},j)-W_m(t_{l-1},j))q_n(j|i,a) \right| dt \\
&\le |V_n^*(t_{l-1},i)-W_m(t_{l-1},i)| + \left(\frac{T}{m}\right)^2 M_1 w^3(i) \\
&+\frac{T}{m}\max_{a\in A_n(i)} \sum_{j\in S_n} |(V_n^*(t_{l-1},j)-W_m(t_{l-1},j))q_n(j|i,a)| \\
&\le |V_n^*(t_{l-1},i)-W_m(t_{l-1},i)| + \left(\frac{T}{m}\right)^2 M_1 w^3(i) + \frac{2QT}{m}\max_{j\in S_n} |V_n^*(t_{l-1},j) \\
&- W_m(t_{l-1},j)|w(i)
\end{aligned}
$$

where M_1 equals $\left[M+G_1(\rho_1 I_{\{\rho_1>0\}}+d_1+2Q)\right](\rho_2 I_{\{\rho_2>0\}}+d_2+2Q)$, the first equality is due to (4.11) and (4.13), and the third inequality follows from (4.7), (4.8) and (4.14). Hence, employing the last inequality, we have

$$
\begin{aligned}
&\max_{j\in S_n} |V_n^*(t_l,j)-W_m(t_l,j)| \\
&\quad\le \left(1+\frac{2QT}{m}w(j_n)\right)\max_{j\in S_n} |V_n^*(t_{l-1},j)-W_m(t_{l-1},j)| + \left(\frac{T}{m}\right)^2 M_1 w^3(j_n) \\
&\quad= \left(1+\frac{2QT}{m}w(j_n)\right)\left[\max_{j\in S_n} |V_n^*(t_{l-1},j)-W_m(t_{l-1},j)| + \frac{M_1 T w^2(j_n)}{2Qm}\right] \\
&\qquad -\frac{M_1 T w^2(j_n)}{2Qm},
\end{aligned}
$$

which implies

$$
\begin{aligned}
&\max_{j\in S_n} |V_n^*(t_l,j)-W_m(t_l,j)| + \frac{M_1 T w^2(j_n)}{2Qm} \\
&\quad\le e^{\frac{2QT}{m}w(j_n)}\left[\max_{j\in S_n} |V_n^*(t_{l-1},j)-W_m(t_{l-1},j)| + \frac{M_1 T w^2(j_n)}{2Qm}\right].
\end{aligned}
$$

Thus, it follows from the induction and the last inequality that

$$
\begin{aligned}
\max_{j\in S_n} |V_n^*(t_l,j)-W_m(t_l,j)| &\le \frac{M_1 T w^2(j_n)}{2Qm}\left(e^{\frac{2lQT}{m}w(j_n)}-1\right) \\
&\le \frac{M_1 T w^2(j_n)}{2Qm}\left(e^{2QTw(j_n)}-1\right).
\end{aligned}
$$

Therefore, part (i) follows from the last inequality.

(ii) Using the inequality $|V^*(0,i)-W_m(0,i)| \le |V^*(0,i)-V_n^*(0,i)|+|V_n^*(0,i)-W_m(0,i)|$, we see that part (ii) follows directly from Theorem 4.1 and part (i).

(iii) Fix any $n \geq 1$, $i, i' \in S_n$, $m \geq 1$, $l \in \{1, \ldots, m\}$ and $f_{n,m} \in \mathcal{O}_{n,m}$. Then we have

$$\begin{aligned}
&\int_t^T \sum_{j \in S_n} V_n^{f_{n,m}}(s, j) q_n(j|i, f_{n,m}(i, s)) ds \\
&= \int_t^T \sum_{j \in S_n} \int_s^T \sum_{k \in S_n} c(k, f_{n,m}(k, v)) P_n^{i', f_{n,m}}(\xi_v = k|\xi_s = j) dv \\
&\quad q_n(j|i, f_{n,m}(i, s)) ds \\
&= \int_t^T \sum_{k \in S_n} c(k, f_{n,m}(k, v)) \int_t^v \sum_{j \in S_n} P_n^{i', f_{n,m}}(\xi_v = k|\xi_s = j) \\
&\quad q_n(j|i, f_{n,m}(i, s)) ds dv \\
&= \int_t^T \sum_{k \in S_n} c(k, f_{n,m}(k, v)) P_n^{i', f_{n,m}}(\xi_v = k|\xi_t = i) dv - \int_t^T c(i, f_{n,m}(i, v)) dv \\
&= V_n^{f_{n,m}}(t, i) - \int_t^T c(i, f_{n,m}(i, v)) dv
\end{aligned} \tag{4.15}$$

for all $t \in [0, T]$, where the first and fourth equalities are due to the definition of $V_n^{f_{n,m}}$, and the second and third ones follow from the Fubini theorem and the Kolmogorov backward equation in Guo and Hernández-Lerma (2009, p. 211), respectively. Thus, using (4.15) and following the similar arguments of part (i), we have

$$|V_n^{f_{n,m}}(0, i) - W_m(0, i)| \leq \frac{M_1 T w^2(j_n)}{2Qm} \left(e^{2QTw(j_n)} - 1 \right). \tag{4.16}$$

Hence, the desired assertion follows from (4.16) and part (ii). □

Remark 4.1 (a) For the original control model $\mathcal{M}$, Theorem 4.2 indicates that we can compute numerically the value function by the iteration constructed as in (4.11) and obtain an approximate optimal policy as in the form of (4.12).

(b) The method of time discretization is used to construct a discrete-time MDP with a fixed step size in the time parameter, from which an approximation for the approximate computations of an optimal policy and the value function is given in van Dijk (1988, 1989) for the case of the denumerable state and action spaces and unbounded transition and cost rates. To solve the finite-horizon optimality equation of the constructed discrete-time MDP is a crucial step in the approximation in van Dijk (1988, 1989). We employ the technique of time discretization to obtain a partition of the time interval and directly provide a finite approximation for the case of a denumerable state space, a Borel action space and unbounded transition and cost rates without introducing an auxiliary discrete-time MDP as in van Dijk (1988, 1989).

(c) If the state and action spaces of the model $\mathcal{M}$ are both finite sets, there exist positive constants $\overline{Q}$ and $\overline{M}$ such that $q^*(i) \leq \overline{Q}$ and $|c(i, a)| \leq \overline{M}$ for all $i \in S$ and $a \in A(i)$. For this particular class of the CTMDPs, there is no need to construct a sequence of the control models $\{\mathcal{M}_n, n \geq 1\}$. Hence, using the similar arguments of Theorem 4.2, we can obtain

$$|V^*(t_l, i) - W_m(t_l, i)| \leq \frac{\overline{M}T(1 + 2\overline{Q}T)(e^{2\overline{Q}T} - 1)}{m}$$

and

$$|V^*(0, i) - V^{f_m}(0, i)| \leq \frac{2\overline{M}T(1 + 2\overline{Q}T)(e^{2\overline{Q}T} - 1)}{m}$$

for all $i \in S$, $m \geq 1$ and $l = 0, 1, \ldots, m$, where the policy f^m is as in (4.12). Therefore, for this particular case, the last two inequalities imply that the accuracy of the approximation given by (4.11) and (4.12) is of order m^{-1}.

5 An example

In this section, we use a controlled birth and death system to illustrate the finite approximation of the finite-horizon expected total cost criterion.

Example 5.1 (*A controlled birth and death system*) The control model is given as follows: $S := \{0, 1, 2, \ldots\}$, $A(0) := [0, \kappa] \times \{0\}$ ($\kappa > 0$), $A(i) := [0, \kappa] \times [\zeta_1, \zeta_2]$ ($\zeta_2 > \zeta_1 > 0$) for all $i \geq 1$, $q(1|0, (a_1, 0)) = -q(0|0, (a_1, 0)) := a_1$ for all $a_1 \in [0, \kappa]$, and for each $i \geq 1$, $a = (a_1, a_2) \in A(i)$,

$$q(j|i, a) := \begin{cases} \lambda i + a_1, & \text{if } j = i + 1, \\ -(\lambda + \mu)i - a_1 - a_2, & \text{if } j = i, \\ \mu i + a_2, & \text{if } j = i - 1, \\ 0, & \text{otherwise,} \end{cases}$$

where the positive constants λ and μ denote the birth and death rates, respectively.

To ensure the existence of optimal policies, we consider the following conditions.

(C1) There exists a constant $M > 0$ such that $|c(i, a)| \leq M(i + 1)$ for all $(i, a) \in K$.

(C2) For each $i \in S$, there exists a constant $L_i > 0$ such that $|c(i, a) - c(i, b)| \leq L_i d_A(a, b)$ for all $a = (a_1, a_2), b = (b_1, b_2) \in A(i)$, where $d_A(a, b) := |a_1 - b_1| + |a_2 - b_2|$.

Proposition 5.1 *Under conditions (C1) and (C2), the controlled birth and death system satisfies Assumptions 3.1, 3.2 and 4.1.*

Proof Let $w(i) := i + 1$ for all $i \in S$. Then direct calculations yield

$$\begin{aligned} q^*(i) &\leq (\lambda + \mu)i + \kappa + \zeta_2 \leq [\lambda + \mu + (\kappa + \zeta_2 - \lambda - \mu)I_{\{\kappa+\zeta_2-\lambda-\mu>0\}}]w(i), \\ \sum_{j\in S} w(j)q(j|i, a) &= (\lambda - \mu)i + a_1 - a_2 \leq (\lambda - \mu)w(i) + \max\{0, \kappa + \mu - \lambda\}, \end{aligned} \tag{5.1}$$

$$\begin{aligned}
\sum_{j\in S} w^2(j)q(j|i,a) &= 2(\lambda-\mu)i^2 + (3\lambda-\mu+2a_1-2a_2)i + 3a_1 - a_2 \\
&= 2(\lambda-\mu)w^2(i) + (2a_1-2a_2-\lambda+3\mu)w(i) + a_1 + a_2 - \lambda - \mu \\
&\le 2(\lambda-\mu)w^2(i) + (2\kappa-2\zeta_1-\lambda+3\mu)w(i) + \kappa + \zeta_2 - \lambda - \mu \\
&\le [2(\lambda-\mu) + (2\kappa-2\zeta_1-\lambda+3\mu)I_{\{2\kappa-2\zeta_1-\lambda+3\mu>0\}}]w^2(i) \\
&\quad + \max\{0, 3\kappa+2\mu+\zeta_2-\lambda\}, \\
\sum_{j\in S} w^3(j)q(j|i,a) &= 3(\lambda-\mu)i^3 + (9\lambda-3\mu+3a_1-3a_2)i^2 + (7\lambda-\mu+9a_1-3a_2)i \\
&\quad + 7a_1 - a_2 \\
&= 3(\lambda-\mu)w^3(i) + (6\mu+3a_1-3a_2)w^2(i) \\
&\quad + (3a_1+3a_2-2\lambda-4\mu)w(i) + \mu - \lambda + a_1 - a_2 \\
&\le [3(\lambda-\mu) + (6\mu+3\kappa-3\zeta_1)I_{\{6\mu+3\kappa-3\zeta_1>0\}} \\
&\quad + (3\kappa+3\zeta_2-2\lambda-4\mu)I_{\{3\kappa+3\zeta_2-2\lambda-4\mu>0\}}]w^3(i) \\
&\quad + \mu - \lambda + \kappa - \zeta_1 \\
&\le [3(\lambda-\mu) + (6\mu+3\kappa-3\zeta_1)I_{\{6\mu+3\kappa-3\zeta_1>0\}} \\
&\quad + (3\kappa+3\zeta_2-2\lambda-4\mu)I_{\{3\kappa+3\zeta_2-2\lambda-4\mu>0\}}]w^3(i) \\
&\quad + \max\{0, 7\kappa+3\mu-\lambda\}
\end{aligned}$$

for all $i \ge 1$ and $a = (a_1, a_2) \in A(i)$, and

$$\begin{aligned}
q^*(0) &\le \kappa w(0) \le [\lambda+\mu+(\kappa+\zeta_2-\lambda-\mu)I_{\{\kappa+\zeta_2-\lambda-\mu>0\}}]w(0), \\
\sum_{j\in S} w(j)q(j|0,a) &= a_1 \le (\lambda-\mu)w(0) + \max\{0, \kappa+\mu-\lambda\}, \qquad (5.2) \\
\sum_{j\in S} w^2(j)q(j|0,a) &= 3a_1 \le [2(\lambda-\mu)+(2\kappa-2\zeta_1-\lambda+3\mu)I_{\{2\kappa-2\zeta_1-\lambda+3\mu>0\}}]w^2(0) \\
&\quad + 3\kappa - 2(\lambda-\mu) \\
&\le [2(\lambda-\mu) + (2\kappa-2\zeta_1-\lambda+3\mu)I_{\{2\kappa-2\zeta_1-\lambda+3\mu>0\}}]w^2(0) \\
&\quad + \max\{0, 3\kappa+2\mu+\zeta_2-\lambda\}, \\
\sum_{j\in S} w^3(j)q(j|0,a) &= 7a_1 \le [3(\lambda-\mu) + (6\mu+3\kappa-3\zeta_1)I_{\{6\mu+3\kappa-3\zeta_1>0\}} \\
&\quad + (3\kappa+3\zeta_2-2\lambda-4\mu)I_{\{3\kappa+3\zeta_2-2\lambda-4\mu>0\}}]w^3(0) \\
&\quad + 7\kappa - 3(\lambda-\mu) \\
&\le [3(\lambda-\mu) + (6\mu+3\kappa-3\zeta_1)I_{\{6\mu+3\kappa-3\zeta_1>0\}} \\
&\quad + (3\kappa+3\zeta_2-2\lambda-4\mu)I_{\{3\kappa+3\zeta_2-2\lambda-4\mu>0\}}]w^3(0) \\
&\quad + \max\{0, 7\kappa+3\mu-\lambda\}
\end{aligned}$$

for all $a = (a_1, a_2) \in A(0)$. Thus, Assumptions 3.1(i), (ii), 3.2(i) and 4.1 hold with $Q := \lambda+\mu+(\kappa+\zeta_2-\lambda-\mu)I_{\{\kappa+\zeta_2-\lambda-\mu>0\}}$, $\rho_1 := \lambda-\mu$, $d_1 := \max\{0, \kappa+\mu-\lambda\}$, $\rho_2 := 2(\lambda-\mu)+(2\kappa-2\zeta_1-\lambda+3\mu)I_{\{2\kappa-2\zeta_1-\lambda+3\mu>0\}}$, $d_2 := \max\{0, 3\kappa+2\mu+\zeta_2-\lambda\}$, $\delta := 3$, $\rho_\delta := 3(\lambda-\mu) + (6\mu+3\kappa-3\zeta_1)I_{\{6\mu+3\kappa-3\zeta_1>0\}} + (3\kappa+3\zeta_2-2\lambda-4\mu)I_{\{3\kappa+3\zeta_2-2\lambda-4\mu>0\}}$ and $d_\delta := \max\{0, 7\kappa+3\mu-\lambda\}$. Moreover, it follows from the description of the model, condition (C1), (5.1) and (5.2) that Assumptions 3.1(iii), 3.2(ii) and (iv) hold. Finally, for each $i, j \in S$, we have

$$|q(j|i,a) - q(j|i,b)| \leq |a_1 - b_1| + |a_2 - b_2|$$

for all $a = (a_1, a_2), b = (b_1, b_2) \in A(i)$. Hence, Assumption 3.2(iii) follows from the last inequality with $L_{ij} = 1$ and condition (C2). This completes the proof of the proposition. □

Next, we take $c(i,a) = (|a_1 - \eta_1| + |a_2 - \eta_2|)i$ ($\eta_1 > 0, \eta_2 > 0$) for all $(i,a) \in K$. Then direct calculations give

$$|c(i,a)| \leq (\max\{\eta_1, |\eta_1 - \kappa|\} + \max\{|\zeta_1 - \eta_2|, |\zeta_2 - \eta_2|\})(i+1) \text{ and}$$
$$|c(i,a) - c(i,b)| \leq (|a_1 - b_1| + |a_2 - b_2|)i$$

for all $i \in S$ and $a = (a_1, a_2), b = (b_1, b_2) \in A(i)$. Hence, conditions (C1) and (C2) are satisfied with $M := \max\{\eta_1, |\eta_1 - \kappa|\} + \max\{|\zeta_1 - \eta_2|, |\zeta_2 - \eta_2|\}$ and $L_i = i$.

For each $n \geq 1$, choose the control model $\mathcal{M}_n$ with $S_n = \{0, 1, \ldots, n\}$, $A_n(0) = \{\frac{\kappa l}{n^3} : l = 0, 1, \ldots, n^3\} \times \{0\}$, $A_n(i) = \{\frac{\kappa l}{n^3} : l = 0, 1, \ldots, n^3\} \times \{\zeta_1 + \frac{(\zeta_2 - \zeta_1)l}{n^3} : l = 0, 1, \ldots, n^3\}$ for all $i = 1, \ldots, n$. Then for each $n \geq 1$ and $i \in S_n$, we have

$$w(n)\left[L_i + 2G_1 w(n) \sum_{j=0}^{n-1} L_{ij}\right] = (n+1)(i + 2G_1 n(n+1)) \leq 2(1+4G_1)n^3,$$

which together with the definition of the Hausdorff metric yields

$$d_H(A(i), A_n(i)) \leq \frac{\kappa + \zeta_2 - \zeta_1}{n^3} \leq \frac{2(1+4G_1)(\kappa + \zeta_2 - \zeta_1)w^3(i)}{w(n)\left[L_i + 2G_1 w(n) \sum_{j=0}^{n-1} L_{ij}\right]}.$$

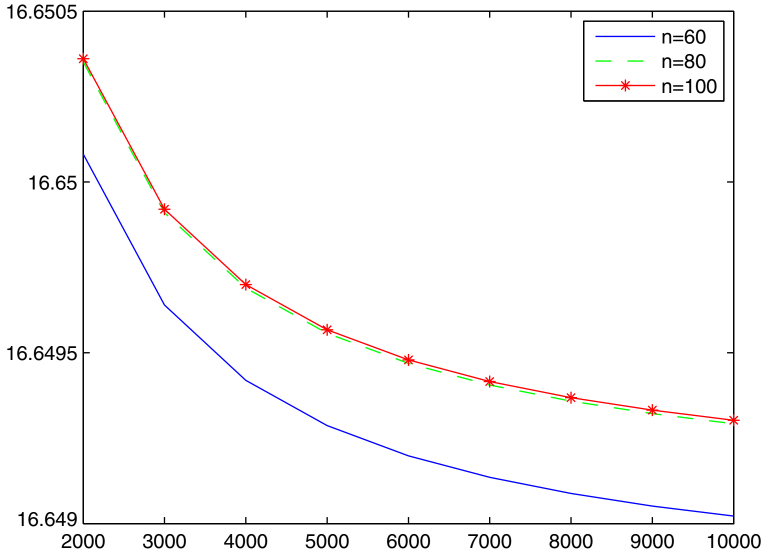

Fig. 1 Value $W_m(0,2)$ of the model $\mathcal{M}_n$ for $n = 60, 80, 100$ and $m = 2000, \ldots, 10{,}000$

Hence, (4.1) holds with $\widetilde{M} := 2(1 + 4G_1)(\kappa + \zeta_2 - \zeta_1)$.

For a numerical experimentation of Example 5.1, we choose the following values of the parameters: $T = 10$, $\lambda = 0.9$, $\mu = 1$, $\kappa = 1$, $\zeta_1 = 0.5$, $\zeta_2 = 1$, $\eta_1 = 2$, $\eta_2 = 3$. Then for $n = 60, 80, 100$ and $m = 2000, \ldots, 10000$, employing the iteration constructed as in (4.11), we obtain the corresponding value $W_m(0, 2)$ as shown in Fig. 1. Empirically, the convergence is faster than that given in Theorem 4.2. This is due to the fact that the bounds used to obtain the error estimations in Theorem 4.2 are very conservative. Moreover, from Fig. 1, we get the approximate value $V^*(0, 2) \simeq 16.6493$.

Acknowledgements I am greatly indebted to the anonymous referees and the associate editor for many valuable comments and suggestions that have greatly improved the presentation. The research was supported by National Natural Science Foundation of China (Grant No. 11526092).

Compliance with ethical standards

Conflict of interest I declare that no conflict of interest exists in this paper.

References

Bäuerle N, Rieder U (2011) Markov decision processes with applications to finance. Springer, Berlin

Gihman II, Skohorod AV (1979) Controlled stochastic processes. Springer, Berlin

Guo XP, Hernández-Lerma O (2009) Continuous-time Markov decision processes: theory and applications. Springer, Berlin

Guo XP, Huang XX, Huang YH (2015) Finite horizon optimality for continuous-time Markov decision processes with unbounded transition rates. Adv Appl Probab 47:1064–1087

Guo XP, Zhang WZ (2014) Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints. Euro J Oper Res 238:486–496

Kitaev MY, Rykov VV (1995) Controlled queueing systems. CRC Press, Boca Raton

Miller BL (1968) Finite state continuous time Markov decision processes with finite planning horizon. SIAM J Control 6:266–280

Pliska SR (1975) Controlled jump processes. Stoch Process Appl 3:259–282

Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, New York

van Dijk NM (1988) On the finite horizon Bellman equation for controlled Markov jump models with unbounded characteristics: existence and approximation. Stoch Process Appl 28:141–157

van Dijk NM (1989) A note on constructing ε-optimal policies for controlled Markov jump models with unbounded characteristics. Stochastics 27:51–58

Wei QD, Chen X (2014) Strong average optimality criterion for continuous-time Markov decision processes. Kybernetika 50:950–977

Yushkevich AA (1977) Controlled Markov models with countable state and continuous time. Theory Probab Appl 22:215–235