J. Appl. Prob. **49,** 1072–1090 (2012)
Printed in England

DISCOUNTED CONTINUOUS-TIME CONTROLLED MARKOV CHAINS: CONVERGENCE OF CONTROL MODELS

TOMÁS PRIETO-RUMEAU,* *Universidad Nacional de Educación a Distancia*
ONÉSIMO HERNÁNDEZ-LERMA,** *CINVESTAV-IPN*

Abstract

We are interested in continuous-time, denumerable state controlled Markov chains (CMCs), with compact Borel action sets, and possibly unbounded transition and reward rates, under the discounted reward optimality criterion. For such CMCs, we propose a definition of a sequence of control models $\{\mathcal{M}_n\}$ converging to a given control model $\mathcal{M}$, which ensures that the discount optimal reward and policies of $\mathcal{M}_n$ converge to those of $\mathcal{M}$. As an application, we propose a finite-state and finite-action truncation technique of the original control model $\mathcal{M}$, which is illustrated by approximating numerically the optimal reward and policy of a controlled population system with catastrophes. We study the corresponding convergence rates.

Keywords: Continuous-time controlled Markov chain; approximation of control models; discount optimality

2010 Mathematics Subject Classification: Primary 90C40; 60J27

1. Introduction

When solving a control problem by following the dynamic programming approach, one usually ends up with a so-called optimality equation (also known as the Bellman or the Hamilton–Jacobi–Bellman equation, depending on the nature of the control problem under study). Except for some particular cases (such as, e.g. linear-quadratic control problems), such optimality equations cannot be explicitly solved because they are 'highly' nonlinear.

One usual tool to obtain numerical solutions to the optimality equation is by means of the Markov chain approximating method. The idea is to define, starting from the original control model, a controlled Markov chain (CMC) with finite state space whose optimal reward and policies approximate the optimal reward and policies of the original control model. Such methods have been developed to approximate, e.g. controlled diffusions [7], [13], average reward continuous-time CMCs [11], discrete-time finite-horizon and infinite-horizon discounted CMCs [8], [15], average reward discrete-time CMCs [9], and discrete-time control models involving constraints [1], among others.

In this paper we are concerned with a continuous-time CMC model $\mathcal{M}$ with denumerable state space, under the discounted reward optimality criterion. As for the already mentioned control problems, the corresponding Bellman optimality equation cannot be explicitly solved. There exist, however, algorithms that are shown to converge to the optimal reward and policies of $\mathcal{M}$. These include the value iteration algorithm—developed in [3] and [5] for discounted CMCs—and the policy iteration algorithm—introduced in [4] for average CMCs, and in [11]

Received 5 July 2011; revision received 26 April 2012.
* Postal address: Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Educación a Distancia, Calle Senda del Rey 9, 28040, Madrid, Spain. Email address: tprieto@ccia.uned.es
** Postal address: Departamento de Matemáticas, CINVESTAV-IPN, México D.F. 07000, México.

for discounted CMCs. For our CMC model $\mathcal{M}$ with denumerable state space, however, the value iteration and the policy iteration algorithms are not viable in practice because they require a 'denumerable' number of calculations at each step. Hence, as in the Markov chain approximation scheme [7], this suggests the idea of considering finite-state and finite-action control models $\mathcal{M}_n$ whose optimal reward and policies we can explicitly compute (by using, e.g. the value or the policy iteration algorithm). Then, the optimal reward and policies of $\mathcal{M}_n$ are used as approximations of those of the original control model $\mathcal{M}$. Following this approach, in this paper we introduce a finite-state and finite-action truncation technique to obtain the approximating control models $\mathcal{M}_n$. Similar discretization procedures can be found in, e.g. [1], [6], and [11].

In addition to the convergence of the truncation technique itself, we are interested here in a more general framework. More precisely, our goal in this paper is to propose a definition of the convergence of discounted CMC models. Namely, given continuous-time, denumerable state CMC models $\mathcal{M}$ and $\mathcal{M}_n$ for $n \geq 1$, the idea is to give a suitable definition of the convergence $\mathcal{M}_n \to \mathcal{M}$ ensuring that the optimal reward and policies of $\mathcal{M}_n$ converge to those of $\mathcal{M}$. Such an approach can be found in [8] for finite-horizon and infinite-horizon discounted discrete-time CMCs, in [2] and [14] for constrained discrete-time models, and in [11] for average reward continuous-time CMCs. Then, as a particular case, we will prove the convergence of the finite-state and finite-action truncations $\mathcal{M}_n$ of our original control model $\mathcal{M}$.

The rest of the paper is organized as follows. In Section 2 we introduce the control model we are interested in, and recall some useful results on discount optimality. In Section 3 we give the definition of convergence of CMCs, and prove our main result. The applications are given in Section 4, in which we solve numerically a controlled population system.

2. Model definition

We are concerned with the control model $\mathcal{M} := \{S, A, (A(i)), (q_{ij}(a)), (r(i,a))\}$, which is defined by the following elements.

- The state space of the system is the denumerable set S. We suppose that $S = \{0, 1, 2, \dots\}$.
- The action space A is a complete and separable metric space.
- The action set in state $i \in S$ is $A(i)$, which is a measurable subset of A. (In this paper, measurability is always referred to the corresponding Borel σ-field.) The family of feasible state-action pairs is defined as
$$\mathbb{K} := \{(i,a) \in S \times A : i \in S,\ a \in A(i)\}.$$
- Let $q_{ij}(a)$ be the transition rate from the state $i \in S$ to the state $j \in S$ under the action $a \in A(i)$. We assume that $a \mapsto q_{ij}(a)$ is measurable for each fixed $i, j \in S$. The transition rates verify that $q_{ij}(a) \geq 0$ for every $(i,a) \in \mathbb{K}$ and $j \neq i$. Finally, we suppose that the transition rates are *conservative*, i.e.
$$\sum_{j \in S} q_{ij}(a) = 0 \quad \text{for all } (i,a) \in \mathbb{K},$$
and *stable*, i.e.
$$q(i) := \sup_{a \in A(i)} \{-q_{ii}(a)\} < \infty \quad \text{for all } i \in S.$$
- The reward rate function $r : \mathbb{K} \to \mathbb{R}$ is assumed to be measurable.

This continuous-time CMC model can be found in, e.g. [3], [5], and [11].

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

2.1. Control policies

Our next definition uses the notation $\mathbb{B}(X)$ for the Borel σ-field of X. Let Φ be the family of functions

$$\varphi \equiv \{\varphi_t(B \mid i) : t \geq 0,\ i \in S,\ B \in \mathbb{B}(A(i))\}$$

that verify the following properties.

(i) The mapping $B \mapsto \varphi_t(B \mid i)$ is a probability measure on $(A(i), \mathbb{B}(A(i)))$ for each $t \geq 0$ and $i \in S$.

(ii) The function $t \mapsto \varphi_t(B \mid i)$ is measurable on $[0, \infty)$ for every $i \in S$ and $B \in \mathbb{B}(A(i))$.

We say that $\varphi \in \Phi$ is a *randomized Markov policy*, or a Markov policy for short.

If $\varphi \in \Phi$ is a Markov policy such that $\varphi_t(B \mid i)$ does not depend on $t \geq 0$, and, moreover, the probability measure $\varphi(B \mid i)$ is a Dirac measure, then we say that φ is a *deterministic stationary policy*. The class of deterministic stationary policies can be identified with the family of functions $f : S \to A$ with $f(i) \in A(i)$ for all $i \in S$. The set of such functions will be denoted by $\mathbb{F}$.

2.2. The controlled Markov process

The transition rates of the Markov policy $\varphi \in \Phi$ are defined as

$$q_{ij}(t, \varphi) := \int_{A(i)} q_{ij}(a)\varphi_t(\mathrm{d}a \mid i) \quad \text{for all } i, j \in S,\ t \geq 0. \tag{2.1}$$

The so-defined transition rates are finite because the $q_{ij}(a)$ are conservative and stable (recall the definition of the control model $\mathcal{M}$). If $f \in \mathbb{F}$ is a deterministic stationary policy then the corresponding transition rates will be written $q_{ij}(f) := q_{ij}(f(i))$ for $i, j \in S$.

For each Markov policy $\varphi \in \Phi$, consider the matrix $[q_{ij}(t, \varphi)]_{i,j}$ for $t \geq 0$, which is a nonhomogeneous Q^{φ}-matrix. By Proposition C.4 of [5, Appendix C], there exists a nonhomogeneous transition function $P_{ij}^{\varphi}(s, t)$ for $i, j \in S$ and $t \geq s \geq 0$ whose transition rates are given by (2.1). To ensure that this transition function is unique and regular, we impose Assumption 2.1(a) below.

Assumption 2.1 uses the notion of a *Lyapunov function*. We say that $w : S \to [1, \infty)$ is a Lyapunov function on S if w is monotone nondecreasing and, in addition, $\lim_{i \to \infty} w(i) = \infty$. Next, we state all our hypotheses on the control model $\mathcal{M}$.

Assumption 2.1. *There exists a Lyapunov function w on S such that the following hypotheses hold.*

(a) *There exist constants $c \neq 0$ and $b \geq 0$, with $b \geq c$, such that*

$$\sum_{j \in S} q_{ij}(a)w(j) \leq -cw(i) + b \quad \textit{for all } (i, a) \in \mathbb{K}.$$

(b) *For each $i \in S$, $q(i) \leq w(i)$.*

(c) *There exist constants $c' \in \mathbb{R}$ and $b' \geq 0$, with $b' \geq c'$, such that*

$$\sum_{j \in S} q_{ij}(a)w^2(j) \leq -c'w^2(i) + b' \quad \textit{for all } (i, a) \in \mathbb{K}.$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

(d) *There exists an* $M > 0$ *such that* $|r(i,a)| \le Mw(i)$ *for all* $(i,a) \in \mathbb{K}$.

(e) *For each fixed* $i, j \in S$, *the functions* $a \mapsto r(i,a)$ *and* $a \mapsto q_{ij}(a)$ *are continuous on* $A(i)$.

(f) *The action set* $A(i)$ *is compact for every* $i \in S$.

Let us make some comments on Assumption 2.1. As already mentioned, Assumption 2.1(a) ensures that the transition function $\{P_{ij}^{\varphi}(s,t)\}$ is regular for each $\varphi \in \Phi$; see [5, Theorem 2.3]. Hence, for each Markov policy $\varphi \in \Phi$ and every initial state $i \in S$, there exists a regular S-valued Markov process, denoted by $\{x^{\varphi}(t)\}_{t\ge 0}$, or $\{x(t)\}_{t\ge 0}$ when there is no risk of confusion, with transition rates (2.1). The corresponding expectation operator will be denoted by E_i^{φ}. As a consequence of Assumption 2.1(a) and [5, Lemma 6.3], we have

$$\mathrm{E}_i^{\varphi}[w(x(t))] \le \mathrm{e}^{-ct} w(i) + \frac{b}{c}(1 - \mathrm{e}^{-ct}) \quad \text{for all } \varphi \in \Phi,\ i \in S,\ t \ge 0. \tag{2.2}$$

Assumption 2.1(b) and (c) are needed to ensure that the Dynkin formula holds; see [5, Appendix C.3].

Assumption 2.1(d) ensures that the growth of the reward rate function $r(i,a)$ is bounded by $w(i)$ uniformly in a. We will also need the following notation. For a Markov policy $\varphi \in \Phi$, let

$$r(t,i,\varphi) := \int_{A(i)} r(i,a)\varphi_t(\mathrm{d}a \mid i) \quad \text{for all } t \ge 0,\ i \in S.$$

For a deterministic stationary policy $f \in \mathbb{F}$, we will simply write $r(i,f) := r(i,f(i))$ for $i \in S$.

Finally, Assumption 2.1(e)–(f) are the standard continuity and compactness conditions.

We now introduce some notation and terminology. Let $\mathcal{B}_w(S)$ be the family of functions $u\colon S \to \mathbb{R}$ such that

$$\|u\|_w := \sup_{i\in S}\left\{\frac{|u(i)|}{w(i)}\right\}$$

is finite. We observe that $\|\cdot\|_w$ is a norm, with respect to which $\mathcal{B}_w(S)$ is a Banach space. We will call $\|\cdot\|_w$ the w-norm.

2.3. The total expected discounted reward optimality criterion

We suppose that the rewards depreciate at a constant *discount rate* $\alpha > 0$, which satisfies the next condition.

Assumption 2.2. *The discount rate* $\alpha > 0$ *satisfies* $\alpha + c > 0$, *where* c *is the constant in Assumption 2.1(a).*

The *total expected discounted reward* (or discounted reward) of the Markov policy $\varphi \in \Phi$ when $i \in S$ is the initial state is defined as

$$v(i,\varphi) := \mathrm{E}_i^{\varphi}\left[\int_0^{\infty} \mathrm{e}^{-\alpha t} r(t, x(t), \varphi)\,\mathrm{d}t\right].$$

It follows from Assumptions 2.1(d) and 2.2, together with inequality (2.2), that the discounted reward verifies

$$|v(i,\varphi)| \le \frac{Mw(i)}{\alpha + c} + \frac{bM}{\alpha(\alpha + c)} \quad \text{for all } i \in S,\ \varphi \in \Phi. \tag{2.3}$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

Therefore, the *optimal discounted reward*, defined as

$$v^*(i) := \sup_{\varphi\in\Phi} v(i,\varphi) \quad \text{for all } i \in S,$$

is finite, and we say that a Markov policy $\varphi \in \Phi$ is *discount optimal* if $v(i,\varphi) = v^*(i)$ for all $i \in S$. We note that, as a consequence of (2.3), and letting $C := M(b+\alpha)/\alpha(c+\alpha)$, we have

$$\|v(\cdot,\varphi)\|_w \le C \quad \text{for all } \varphi \in \Phi, \quad \text{and} \quad \|v^*\|_w \le C. \tag{2.4}$$

Our next result summarizes the main results on the dynamic programming optimality equation for $\mathcal{M}$ and the existence of discount optimal policies. Regarding Theorem 2.1(a) below and as a consequence of Assumption 2.1, for every $u \in \mathcal{B}_w(S)$, the function $r(i,a) + \sum_{j\in S} q_{ij}(a)u(j)$ is continuous in $a \in A(i)$ for each fixed $i \in S$.

Theorem 2.1. *Let the control model $\mathcal{M}$ satisfy Assumptions 2.1 and 2.2.*

(a) *The optimal discounted reward v^* is the unique solution in $\mathcal{B}_w(S)$ of the* discounted reward optimality equation *(DROE)*

$$\alpha u(i) = \max_{a\in A(i)} \left\{ r(i,a) + \sum_{j\in S} q_{ij}(a)u(j) \right\} \quad \textit{for all } i \in S.$$

(b) *A policy $f \in \mathbb{F}$ is discount optimal if and only if it attains the maximum in the DROE, i.e.*

$$\alpha v^*(i) = \max_{a\in A(i)} \left\{ r(i,a) + \sum_{j\in S} q_{ij}(a)v^*(j) \right\} = r(i,f) + \sum_{j\in S} q_{ij}(f)v^*(j) \quad \textit{for all } i \in S.$$

The proof of Theorem 2.1 given in [3, Theorem 3.2] and [5, Chapter 6] uses the value iteration algorithm. In [10, Theorem 1], however, Theorem 2.1 is established by showing the convergence of the policy iteration algorithm.

3. Convergence of control models

Suppose that $\{\mathcal{M}_n\}_{n\ge1}$ is a sequence of control models that converges (in a suitably defined sense) to the control model $\mathcal{M}$ in Section 2. We want to find conditions on the sequence $\{\mathcal{M}_n\}_{n\ge1}$ of 'approximating' control models ensuring that the optimal discounted reward of $\mathcal{M}_n$, say v_n^*, converges as $n \to \infty$ to the optimal discounted reward v^* of $\mathcal{M}$, and that, in addition, discount optimal policies for $\mathcal{M}_n$ converge to discount optimal policies for $\mathcal{M}$. Such approximation procedures have been studied for continuous-time CMCs under the average reward optimality criterion in [11], and for discrete-time constrained control models in [1], [2], and [14].

For each $n \ge 1$, the control model $\mathcal{M}_n$ is

$$\mathcal{M}_n := \{S_n, A, (A_n(i)), (q_{ij}^n(a)), (r_n(i,a))\},$$

with the following elements.

- The state space S_n, which is assumed to be a subset of S. (Note that S_n may be finite or infinite.)
- The action space A, which is the same as the action space of the control model $\mathcal{M}$ defined in Section 2.

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

- The action sets $A_n(i)$ for $i \in S_n$, which are measurable subsets of $A(i)$. Also, let $\mathbb{K}_n$ be the family of state-action pairs for $\mathcal{M}_n$, i.e.

$$\mathbb{K}_n := \{(i, a) \in S \times A : i \in S_n,\ a \in A_n(i)\}.$$

- The transition rates $q_{ij}^n(a)$, which are assumed to be measurable on $A_n(i)$ for each fixed $i, j \in S_n$. The transition rates are conservative and stable, that is, $\sum_{j \in S_n} q_{ij}^n(a) = 0$ for all $(i, a) \in \mathbb{K}_n$, and

$$q_n(i) := \sup_{a \in A_n(i)} \{-q_{ii}^n(a)\} < \infty \quad \text{for all } i \in S_n.$$

- The reward rates $r_n(i, a)$, which are measurable in $a \in A_n(i)$ for each fixed $i \in S_n$.

Next, we state our hypotheses on the sequence $\{\mathcal{M}_n\}_{n \geq 1}$ of approximating control models. Roughly speaking, we will suppose that the conditions in Assumption 2.1 hold 'uniformly' in $n \geq 1$.

Assumption 3.1. *Let w be the Lyapunov function in Assumption 2.1.*

(a) *For every $n \geq 1$,*

$$\sum_{j \in S_n} q_{ij}^n(a) w(j) \leq -cw(i) + b \quad \textit{for all } (i, a) \in \mathbb{K}_n,$$

where the constants $c \neq 0$ and $b \geq 0$ are as in Assumption 2.1(a).

(b) *For each $n \geq 1$ and $i \in S_n$, $q_n(i) \leq w(i)$.*

(c) *For all $n \geq 1$,*

$$\sum_{j \in S_n} q_{ij}^n(a) w^2(j) \leq -c' w^2(i) + b' \quad \textit{for all } (i, a) \in \mathbb{K}_n,$$

with $c' \in \mathbb{R}$ and $b' \geq 0$ as in Assumption 2.1(c).

(d) *For every $n \geq 1$ and $(i, a) \in \mathbb{K}_n$, we have $|r_n(i, a)| \leq M w(i)$, with the constant $M > 0$ as in Assumption 2.1(d).*

(e) *For each $n \geq 1$ and each fixed $i, j \in S_n$, $a \mapsto r_n(i, a)$ and $a \mapsto q_{ij}^n(a)$ are continuous on $A_n(i)$.*

(f) *For every $n \geq 1$ and $i \in S_n$, the action set $A_n(i)$ is compact.*

Our definitions in Section 2 can be easily modified to account for the control models $\mathcal{M}_n$. For instance, given $u : S_n \to \mathbb{R}$, its w-norm is defined as

$$\|u\|_w := \sup_{i \in S_n} \left\{ \frac{|u(i)|}{w(i)} \right\},$$

and $\mathcal{B}_w(S_n)$ is the family of functions on S_n with finite w-norm. The class of Markov policies for the control model $\mathcal{M}_n$ is denoted by Φ_n. Also, we denote by $\mathbb{F}_n$ the set of deterministic stationary policies for the control model $\mathcal{M}_n$; that is, $f : S_n \to A$ is in $\mathbb{F}_n$ if $f(i) \in A_n(i)$ for all $i \in S_n$. The expectation operator under the control model $\mathcal{M}_n$ for the control policy $\varphi \in \Phi_n$,

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

given the initial state $i \in S_n$, is denoted by $\mathrm{E}^{\varphi}_{i,n}$. Notation such as, e.g. $\{x^{\varphi}(t)\}_{t\geq 0}$, or $r_n(t, i, \varphi)$ for $\varphi \in \Phi_n$, $i \in S_n$, and $t \geq 0$, is given the obvious definition.

We suppose that the discount rate $\alpha > 0$ satisfies Assumption 2.2 (that is, $\alpha + c > 0$, with c the constant in Assumptions 2.1(a) and 3.1(a)). The total expected discounted reward of the policy $\varphi \in \Phi_n$ for the control model $\mathcal{M}_n$, $n \geq 1$, is defined as

$$v_n(i, \varphi) := \mathrm{E}^{\varphi}_{i,n}\left[\int_0^{\infty} \mathrm{e}^{-\alpha t} r_n(t, x(t), \varphi)\, \mathrm{d}t\right] \quad \text{for all } i \in S_n.$$

Obviously, each control model $\mathcal{M}_n$ satisfies Assumption 2.1 and, therefore, the results in Section 2 hold for every $\mathcal{M}_n$. In particular, the optimal discounted reward of the control model $\mathcal{M}_n$, which we will denote by v_n^*, is the unique solution in $\mathcal{B}_w(S_n)$ of the corresponding DROE. Moreover, a policy in $\mathbb{F}_n$ is discount optimal for $\mathcal{M}_n$ if and only if it attains the maximum in the DROE of the control model $\mathcal{M}_n$. Furthermore, as in (2.4), we obtain the bounds

$$\|v_n(\cdot, \varphi)\|_w \leq C \quad \text{for all } \varphi \in \Phi_n \quad \text{and} \quad \|v_n^*\|_w \leq C, \tag{3.1}$$

where the bounding constant $C := M(b + \alpha)/\alpha(c + \alpha)$ is uniform in $n \geq 1$ because the constants b, c, and M are the same for every control model $\mathcal{M}_n$ (recall Assumption 3.1).

Lemma 3.1. *Suppose that the control models $\mathcal{M}_n$ for $n \geq 1$ verify Assumption 3.1(b)–(c). Given $i \in S$ and $\varepsilon > 0$, there exists $K > i$ such that, for every $n \geq 1$ with $i \in S_n$ and all $a \in A_n(i)$,*

$$\sum_{j \in S_n,\, j \geq K} q_{ij}^n(a) w(j) < \varepsilon.$$

Proof. We suppose that $i \in S$ and $\varepsilon > 0$ are fixed. If $K > i$ and n is such that $i \in S_n$, then, for all $a \in A_n(i)$,

$$\begin{aligned} \sum_{j \in S_n,\, j \geq K} q_{ij}^n(a) w(j) &\leq \frac{1}{w(K)} \sum_{j \in S_n,\, j \geq K} q_{ij}^n(a) w^2(j) \qquad (3.2)\\ &\leq \frac{1}{w(K)} \left(\sum_{j \in S_n} q_{ij}^n(a) w^2(j) - q_{ii}^n(a) w^2(i) \right) \\ &\leq \frac{-c' w^2(i) + b' + w^3(i)}{w(K)}, \qquad (3.3) \end{aligned}$$

where (3.2) is derived from the fact that $q_{ij}^n(a) \geq 0$ because $j \geq K > i$, while (3.3) follows from Assumption 3.1(b)–(c). Finally, it is clear from (3.3) that we can choose K large enough (note that K depends only on i and ε, and not on n or a) such that

$$\sum_{j \in S_n,\, j \geq K} q_{ij}^n(a) w(j) < \varepsilon$$

for all $(i, a) \in \mathbb{K}_n$.

Remark 3.1. We note that, as a consequence of Assumption 2.1(b)–(c), the result in Lemma 3.1 is also valid for the control model $\mathcal{M}$. More precisely, for arbitrary $i \in S$ and $\varepsilon > 0$, there exists $K > i$ such that $\sum_{j \geq K} q_{ij}(a) w(j) < \varepsilon$ for all $a \in A(i)$.

Next, we give the definition of the convergence of control models.

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

Definition 3.1. Suppose that $\mathcal{M}$ is the control model defined in Section 2, and let $\{\mathcal{M}_n\}_{n\geq 1}$ be the sequence of control models defined above. We say that $\{\mathcal{M}_n\}_{n\geq 1}$ converges to $\mathcal{M}$ as $n \to \infty$, which will be denoted by $\mathcal{M}_n \to \mathcal{M}$, if the following conditions are satisfied.

(a) The sequence $\{S_n\}_{n\geq 1}$ of state spaces is monotone nondecreasing, and $S_n \uparrow S$.

As a consequence, if, for each $i \in S$, we define $n(i) := \min\{n \geq 1 : i \in S_n\}$, we have $i \in S_n$ if and only if $n \geq n(i)$.

(b) For each $i \in S$, the action sets $A_n(i)$ converge to $A(i)$ as $n \to \infty$ in the Kuratowski sense. Equivalently, for all $i \in S$,

$$\lim_{n\to\infty} \inf\{d_A(a, a') : a' \in A_n(i)\} = 0 \quad \text{for all } a \in A(i),$$

where d_A is the metric in the action space A.

(c) For each fixed $i, j \in S$, if the sequence $a_n \in A_n(i)$ for $n \geq n(i) \vee n(j)$ converges to $a \in A(i)$ as $n \to \infty$, then $q_{ij}^n(a_n) \to q_{ij}(a)$.

(d) For each fixed $i \in S$, if the sequence $a_n \in A_n(i)$ for $n \geq n(i)$ converges to $a \in A(i)$ as $n \to \infty$, then $r_n(i, a_n) \to r(i, a)$.

The condition in Definition 3.1(b) is equivalent to the following statement. For each fixed $(i, a) \in \mathbb{K}$ and every subsequence $\{n'\}$, there exists a further subsequence $\{n''\}$ and actions $a_{n''} \in A_{n''}(i)$ for all $n'' \geq n(i)$ such that $a_{n''} \to a$ as $n'' \to \infty$.

The conditions in Definition 3.1(c) and (d) state, roughly speaking, that $r_n(i, a)$ and $q_{ij}^n(a)$ converge to $r(i, a)$ and $q_{ij}(a)$, respectively, uniformly in a for each fixed states i and j.

Remark 3.2. Let us comment on Definition 3.1. Note that we allow all the elements of the control models $\mathcal{M}_n$ (namely, the state space, the action sets, and the transition and reward rates) to depend on $n \geq 1$.

When dealing with similar definitions of convergence of control models, the state space is usually allowed to depend on n; see [1], [6], and [11]. The transition and reward rates may as well depend on n. In this case, the uniform convergence property in Definition 3.1(c)–(d) is a usual requirement; see, e.g. Condition (2) of [1, Theorem 6.1], and Assumptions 3.1(c) and 3.3(c) of [2].

The notion of the Kuratowski convergence for the approximation of control models was used in [8]. Let us also mention that the Kuratowski convergence of the action sets $A_n(i)$ is related to the discretization of the state space made in [6, Section 6.3] for a discrete-time Markov control process. We note however that in [1], [2], [6], and [11] the action sets of $\mathcal{M}_n$ are the same as the action sets of the original control model $\mathcal{M}$.

Before stating our main result, we prove the following preliminary fact.

Lemma 3.2. *Suppose that the control models $\mathcal{M}$ and $\mathcal{M}_n$ for $n \geq 1$ satisfy Assumptions 2.1 and 3.1, respectively, as well as Definition 3.1(a). Also, let the discount rate α verify Assumption 2.2, and let $v_n^* \in \mathcal{B}_w(S_n)$ and $f_n^* \in \mathbb{F}_n$ be the optimal discounted reward and a discount optimal policy for $\mathcal{M}_n$, $n \geq 1$, respectively. Then the following statements hold.*

(a) *There exists a subsequence $\{n'\}$ and some $u \in \mathcal{B}_w(S)$ such that*

$$\lim_{n'\to\infty} v_{n'}^*(i) = u(i) \quad \textit{for all } i \in S.$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

(b) *There exists a subsequence* $\{n'\}$ *and a policy* $f \in \mathbb{F}$ *with*

$$\lim_{n'\to\infty} f_{n'}^*(i) = f(i) \quad \textit{for all } i \in S. \tag{3.4}$$

In this case, we say that f *is a limit policy of* $\{f_n^*\}_{n\geq 1}$.

Proof. (i) Note that, for each fixed $i \in S$, and as a consequence of (3.1), the sequence of optimal discounted rewards $v_n^*(i)$ for $n \geq n(i)$ is bounded. Therefore, it has a convergent subsequence. Moreover, by using a diagonal argument we deduce that, for some further subsequence, denoted by $\{n'\}$, the sequence $\{v_{n'}^*\}$ converges pointwise to some $u \in \mathcal{B}_w(S)$. More explicitly, there exists $u \in \mathcal{B}_w(S)$ such that, for every $i \in S$, the sequence $\{v_{n'}^*(i)\}_{n'\geq n(i)}$ converges to $u(i)$. Moreover, (3.1) yields $\|u\|_w \leq C$.

(ii) Fix $i \in S$, and suppose that $n \geq n(i)$, that is, $i \in S_n$. Then we have $f_n^*(i) \in A_n(i) \subseteq A(i)$. Thus, the sequence $\{f_n^*(i)\}_{n\geq n(i)}$ is a sequence in the compact space $A(i)$. Therefore, it has a convergent subsequence. Hence, as in the proof of (i), it follows that there exists a subsequence $\{n'\}$ and $f \in \mathbb{F}$ such that the sequence $\{f_{n'}^*(i)\}_{n'\geq n(i)}$ converges to $f(i)$ for all $i \in S$.

Finally, we state our main result.

Theorem 3.1. *Suppose that the control models* $\mathcal{M}$ *and* $\{\mathcal{M}_n\}_{n\geq 1}$ *satisfy Assumptions 2.1 and 3.1, respectively, and let the discount rate* $\alpha > 0$ *satisfy Assumption 2.2. If* $\mathcal{M}_n \to \mathcal{M}$ *then the following statements hold.*

(a) *For every* $i \in S$, $\lim_{n\to\infty} v_n^*(i) = v^*(i)$.

(b) *If* $f_n^* \in \mathbb{F}_n$ *for* $n \geq 1$ *is a discount optimal policy for* $\mathcal{M}_n$ *then any limit policy* $f^* \in \mathbb{F}$ *of* $\{f_n^*\}_{n\geq 1}$ *is discount optimal for* $\mathcal{M}$.

Proof. Suppose that $\{n'\}$ is a subsequence such that $\{v_{n'}^*\}$ converges pointwise to some $u \in \mathcal{B}_w(S)$ (with, necessarily, $\|u\|_w \leq C$), and such that $\{f_{n'}^*\}$ converges to some $f^* \in \mathbb{F}$; recall (3.4). The existence of such $\{n'\}$ is given by Lemma 3.2.

Fix an arbitrary state $i \in S$, an action $a \in A(i)$, and $\varepsilon > 0$. By Definition 3.1(b) (recall the comment after Definition 3.1), there exists a subsequence $n'' \geq n(i)$ of $\{n'\}$ and actions $a_{n''} \in A_{n''}(i)$ such that $a_{n''} \to a$ as $n'' \to \infty$. To ease the notation, and without loss of generality, this subsequence will still be denoted by $\{n'\}$. For such n', from the DROE for the control model $\mathcal{M}_{n'}$ we obtain

$$\alpha v_{n'}^*(i) \geq r_{n'}(i, a_{n'}) + \sum_{j\in S_{n'}} q_{ij}^{n'}(a_{n'}) v_{n'}^*(j). \tag{3.5}$$

Now, by Lemma 3.1 and Remark 3.1, there exists $K > i$ (which depends only on i and ε) such that

$$\left| \sum_{j\geq K} q_{ij}(a)u(j) \right| \leq C \sum_{j\geq K} q_{ij}(a)w(j) \leq C\varepsilon$$

and, by (3.1), for all $n' \geq n(i)$,

$$\left| \sum_{j\in S_{n'},\, j\geq K} q_{ij}^{n'}(a_{n'}) v_{n'}^*(j) \right| \leq C \sum_{j\in S_{n'},\, j\geq K} q_{ij}^{n'}(a_{n'}) w(j) \leq C\varepsilon.$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

Hence, from (3.5), we deduce that, for all $n' \geq n(i)$,

$$\alpha v^*_{n'}(i) \geq r_{n'}(i, a_{n'}) + \sum_{j \in S_{n'},\, j<K} q^{n'}_{ij}(a_{n'}) v^*_{n'}(j) + \sum_{j \geq K} q_{ij}(a) u(j) - 2C\varepsilon. \tag{3.6}$$

Observe now that, as a consequence of Definition 3.1(c) and (d),

$$r_{n'}(i, a_{n'}) \to r(i, a) \quad \text{and} \quad q^{n'}_{ij}(a_{n'}) \to q_{ij}(a)$$

as $n' \to \infty$. On the other hand, for large n', we have $\{0, 1, \ldots, K-1\} \subseteq S_{n'}$. Finally, for every state $0 \leq j < K$, the limit $v^*_{n'}(j) \to u(j)$ as $n' \to \infty$ holds. Hence, taking the limit as $n' \to \infty$ in (3.6) yields

$$\alpha u(i) \geq r(i, a) + \sum_{j<K} q_{ij}(a) u(j) + \sum_{j \geq K} q_{ij}(a) u(j) - 2C\varepsilon = r(i, a) + \sum_{j \in S} q_{ij}(a) u(j) - 2C\varepsilon.$$

Since $\varepsilon > 0$ and $(i, a) \in \mathbb{K}$ are arbitrary, it follows that

$$\alpha u(i) \geq r(i, a) + \sum_{j \in S} q_{ij}(a) u(j) \quad \text{for all } (i, a) \in \mathbb{K}.$$

Hence, from [5, Theorem 6.9] we conclude that $v^* \leq u$.

Let us now prove the reverse inequality. Recall that we are assuming that there is a subsequence $\{n'\}$ and a policy $f^* \in \mathbb{F}$ such that $f^*_{n'}(i) \to f^*(i)$ for all $i \in S$. Fix a state $i \in S$, and suppose that $n' \geq n(i)$. With the policy $f^*_{n'} \in \mathbb{F}_{n'}$ being discount optimal for the control model $\mathcal{M}_{n'}$, by Theorem 2.1(b) we have

$$\alpha v^*_{n'}(i) = r_{n'}(i, f^*_{n'}) + \sum_{j \in S_{n'}} q^{n'}_{ij}(f^*_{n'}) v^*_{n'}(j). \tag{3.7}$$

Given $\varepsilon > 0$, by Lemma 3.1 and Remark 3.1 again, there exists $K > i$ (which depends only on i and ε) such that

$$\left| \sum_{j \geq K} q_{ij}(f^*) u(j) \right| \leq C \sum_{j \geq K} q_{ij}(f^*) w(j) \leq C\varepsilon,$$

where $u \in \mathcal{B}_w(S)$ is the pointwise limit of $\{v^*_{n'}\}$, and, for all $n' \geq n(i)$,

$$\left| \sum_{j \in S_{n'},\, j \geq K} q^{n'}_{ij}(f^*_{n'}) v^*_{n'}(j) \right| \leq C \sum_{j \in S_{n'},\, j \geq K} q^{n'}_{ij}(f^*_{n'}) w(j) \leq C\varepsilon.$$

Thus, as a consequence of (3.7),

$$\alpha v^*_{n'}(i) \leq r_{n'}(i, f^*_{n'}) + \sum_{j \in S_{n'},\, j<K} q^{n'}_{ij}(f^*_{n'}) v^*_{n'}(j) + \sum_{j \geq K} q_{ij}(f^*) u(j) + 2C\varepsilon.$$

Letting $n' \to \infty$ in the above inequality, and recalling that (see Definition 3.1(c) and (d))

$$r_{n'}(i, f^*_{n'}) \to r(i, f^*) \quad \text{and} \quad q^{n'}_{ij}(f^*_{n'}) \to q_{ij}(f^*),$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

and also that $v^*_{n'}(j) \to u(j)$ for all $0 \le j < K$, it follows that

$$\alpha u(i) \le r(i, f^*) + \sum_{j \in S} q_{ij}(f^*)u(j) + 2C\varepsilon.$$

With the state $i \in S$ and the constant $\varepsilon > 0$ being arbitrary, we conclude that

$$\alpha u(i) \le r(i, f^*) + \sum_{j \in S} q_{ij}(f^*)u(j) \quad \text{for all } i \in S.$$

From [5, Theorem 6.9] we obtain $u \le v(\cdot, f^*) \le v^*$. Combined with the previously established inequality $v^* \le u$, we find that u equals the optimal discounted reward v^*. In addition, we find that f^* is discount optimal. Hence, we have shown that the pointwise limit of $\{v^*_{n'}\}$ through any convergent subsequence is necessarily v^*. Therefore, the whole sequence $\{v^*_n\}$ converges pointwise to v^*. This proves part (a) of the theorem, while part (b) is a direct consequence of the arguments above.

Remark 3.3. Theorem 3.1 gives the pointwise convergence of the optimal discounted rewards v^*_n of $\mathcal{M}_n$ to v^*, and also the convergence of the optimal policies of $\mathcal{M}_n$ to an optimal policy of $\mathcal{M}$. There still remains one important open issue, which is to study the speed of this convergence, or to give bounds on the approximation errors. For some particular cases, such convergence rates can be explicitly determined—see the example in Section 4. For general control models, however, determining the convergence rates is an open problem.

4. Application to finite approximations

Suppose that $\mathcal{M}$ is a control model whose optimal discounted reward and policies we want to approximate. The simplest way of obtaining a sequence $\{\mathcal{M}_n\}_{n \ge 1}$ of control models that converges to $\mathcal{M}$ is by the finite-state and finite-action truncation procedure defined next.

Let $\mathcal{M} = \{S, A, (A(i)), (q_{ij}(a)), (r(i, a))\}$ be the control model in Section 2, assumed to satisfy the conditions in Assumption 2.1. For each $n \ge 1$, define the control model

$$\mathcal{M}_n := \{S_n, A, (A_n(i)), (q^n_{ij}(a)), (r_n(i, a))\}$$

as follows.

- The state space is $S_n := \{0, 1, \dots, n\}$.
- For each $i \in S_n$, let $A_n(i)$ be a finite subset of $A(i)$. In addition, suppose that the sets $A_n(i)$ verify the condition in Definition 3.1(b).
- Given states $i \in S_n$ and $0 \le j < n$, let

$$q^n_{ij}(a) := q_{ij}(a) \quad \text{and} \quad q^n_{in}(a) := \sum_{j \ge n} q_{ij}(a)$$

  for $a \in A_n(i)$.
- The reward rate is $r_n(i, a) := r(i, a)$ for $i \in S_n$ and $a \in A_n(i)$.

Roughly speaking, the control model $\mathcal{M}_n$ consists in restarting the control process at state n when it reaches the set $\{n + 1, n + 2, \dots\}$.

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

Proposition 4.1. *Suppose that $\mathcal{M}$ is a control model that satisfies Assumption 2.1. If $\{\mathcal{M}_n\}_{n\geq 1}$ is the sequence of finite-state and finite-action control models defined above, then the $\mathcal{M}_n$ for $n \geq 1$ satisfy Assumption 3.1, and, moreover, $\mathcal{M}_n \to \mathcal{M}$.*

Proof. Before proceeding with the proof itself, note that the particular definition of $q_{in}^n(a)$ means that

$$\sum_{j\in S_n} q_{ij}^n(a) = 0 \quad \text{for all } (i,a) \in \mathbb{K}_n,$$

so the transition rates of $\mathcal{M}_n$ are conservative. Their stability will be proved below, together with Assumption 3.1(b).

Let us check that Assumption 3.1(a) is satisfied. Recall that the Lyapunov function w, as well as the constants $c \neq 0$ and $b \geq 0$, are taken from Assumption 2.1. Given $n \geq 1$ and $(i,a) \in \mathbb{K}_n$,

$$\begin{aligned}\sum_{j\in S_n} q_{ij}^n(a)w(j) &= \sum_{0\leq j\leq n} q_{ij}(a)w(j) + \sum_{j>n} q_{ij}(a)w(n)\\ &\leq \sum_{0\leq j\leq n} q_{ij}(a)w(j) + \sum_{j>n} q_{ij}(a)w(j) \qquad (4.1)\\ &= \sum_{j\in S} q_{ij}(a)w(j)\\ &\leq -cw(i) + b, \qquad (4.2)\end{aligned}$$

where (4.1) follows from the monotonicity of w and the fact that $q_{ij}(a) \geq 0$ for $j > n \geq i$, while (4.2) is derived from Assumption 2.1(a). Hence, Assumption 3.1(a) is satisfied, and, obviously, Assumption 3.1(c) is derived similarly.

Concerning Assumption 3.1(b), note that, by Assumption 2.1(b), given $i \in S_n$ and $a \in A_n(i)$,

$$-q_{ii}^n(a) = -q_{ii}(a) \leq w(i) \quad \text{for } 0 \leq i < n,$$

while $-q_{nn}^n(a) \leq -q_{nn}(a) \leq w(n)$. Therefore, for every $n \geq 1$ and $i \in S_n$, we have $q_n(i) \leq w(i)$.

Assumption 3.1(d) is a straightforward consequence of the definition of r_n and Assumption 2.1(d). Finally, Assumption 3.1(e) and (f) hold because the action sets $A_n(i)$ for $n \geq 1$ and $i \in S_n$ are finite.

To conclude the proof of this proposition, let us check that $\mathcal{M}_n \to \mathcal{M}$ as $n \to \infty$. It is clear that the conditions in Definition 3.1(a) and (b) are satisfied. Regarding (c) and (d), suppose that, given $i, j \in S$, the sequence $a_n \in A_n(i)$ for $n \geq i$ converges to $a \in A(i)$. Since, for large n, we have $q_{ij}^n(a_n) = q_{ij}(a_n)$, the convergence $q_{ij}^n(a_n) \to q_{ij}(a)$ follows from the continuity of the transition rates $q_{ij}(a)$ in Assumption 2.1(e). In the same way, we can prove that $r_n(i,a_n)$ converges to $r(i,a)$. Hence, we have shown that $\mathcal{M}_n \to \mathcal{M}$.

By using the policy iteration algorithm—see [10]—we can *explicitly* obtain the optimal discounted reward v_n^* of $\mathcal{M}_n$, as well as the corresponding optimal policies $f_n^* \in \mathbb{F}_n$. This is because, with the state space S_n and the action sets $A_n(i)$ being finite, the set $\mathbb{F}_n$ of deterministic stationary policies is finite, and so the policy iteration algorithm converges in a finite number of steps.

Finally, Theorem 3.1 ensures that $v_n^* \to v^*$, the optimal discounted reward of $\mathcal{M}$, and that any limit policy in $\mathbb{F}$ of $\{f_n^*\}_{n\geq 1}$ is discount optimal for $\mathcal{M}$.

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

4.1. A population system with catastrophes

Our next example is a generalization of the population system proposed in [5, Example 7.2]; see also [11, Section IV].

The state space is $S = \{0, 1, 2, \dots\}$, which stands for the size of the population. The birth and death rates of the population are $\lambda > 0$ and $\mu > 0$, respectively. We suppose that the decision maker controls an immigration rate $a \in [0, a_2]$ for some $a_2 > 0$. When the population size is $i \in S$, we assume that a catastrophe occurs at a rate $d(i, b) \geq 0$. This rate is controlled by the action $b \in [b_1, b_2]$ chosen by the controller. We will suppose that the function $b \mapsto d(i, b)$ is continuous on $[b_1, b_2]$ for each fixed $i \in S$, and that there exists $C > 0$ such that

$$\sup_{b \in [b_1, b_2]} d(i, b) \leq C(i + 1) \quad \text{for all } i \in S.$$

Let $\gamma_i(j)$ for $1 \leq j \leq i$ be the probability that j individuals perish if a catastrophe happens when the size of the population is $i > 0$. We have, obviously,

$$\sum_{j=1}^{i} \gamma_i(j) = 1 \quad \text{for each } i > 0.$$

We will denote by E_{γ_i} the expectation operator associated with the distribution $\{\gamma_i(j)\}_{1 \leq j \leq i}$, and by X_i the random number of perished individuals.

The action sets of our control model are $A(0) = [0, a_2]$ and

$$A \equiv A(i) = [0, a_2] \times [b_1, b_2] \quad \text{for all } i > 0.$$

The system's transition rates in state $i = 0$ are given by

$$q_{01}(a) = a = -q_{00}(a) \quad \text{for all } a \in A(0),$$

while, for $i > 0$ and $(a, b) \in A$, they are given by

$$q_{ij}(a, b) = \begin{cases} 0 & \text{for } j > i + 1, \\ \lambda i + a & \text{for } j = i + 1, \\ -(\lambda + \mu)i - a - d(i, b) & \text{for } j = i, \\ \mu i + d(i, b)\gamma_i(1) & \text{for } j = i - 1, \\ d(i, b)\gamma_i(i - j) & \text{for } 0 \leq j < i - 1. \end{cases}$$

When the population size is $i \in S$, the controller receives a reward at a rate pi for some $p > 0$. The cost rate for controlling both the immigration and the catastrophe rates is $c(i, a, b)$. We assume that the function $c(i, \cdot, \cdot)$ is continuous on $A(i)$ for each $i \in S$ and, furthermore, that, for some constant $C' > 0$,

$$\sup_{(a,b) \in A(i)} |c(i, a, b)| \leq C'(i + 1) \quad \text{for all } i \in S.$$

We will consider the net reward rate $r(i, a, b) = pi - c(i, a, b)$ for $i \in S$ and $(a, b) \in A(i)$.

Proposition 4.2. *The controlled population system $\mathcal{M}$ verifies Assumption 2.1.*

Proof. For some constant $R \geq 1$ such that $R > \lambda + \mu + a_2 + C$, we consider the Lyapunov function $w(i) = R(i + 1)$ for $i \in S$.

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

It is easily seen that the conditions in Assumption 2.1(b) and (d)–(f) are satisfied. Hence, it remains to show that Assumption 2.1(a) and (c) hold.

By a direct calculation we obtain, for states $i > 0$ and actions $(a, b) \in A$,

$$\begin{aligned}\sum_{j\in S} q_{ij}(a,b)w(j) &= \left(\lambda - \mu - d(i,b)\,\mathrm{E}_{\gamma_i}\left[\frac{X_i}{i+1}\right]\right)w(i) + R(a-\lambda+\mu)\\ &\le (\lambda-\mu)w(i) + R(a_2-\lambda+\mu) \qquad (4.3)\end{aligned}$$

and

$$\begin{aligned}\sum_{j\in S} q_{ij}(a,b)w^2(j) &= \left(2(\lambda-\mu) + d(i,b)\left(\mathrm{E}_{\gamma_i}\left[\left(\frac{X_i}{i+1}\right)^2\right] - 2E_{\gamma_i}\left[\frac{X_i}{i+1}\right]\right)\right.\\ &\qquad \left. + \frac{2a-\lambda+3\mu}{i+1}\right)w^2(i) + R^2(a-\lambda-\mu)\\ &\le \left(2(\lambda-\mu) + \frac{|2a_2-\lambda+3\mu|}{2}\right)w^2(i) + R^2(a_2-\lambda-\mu).\end{aligned}$$

Assumption 2.1(a) and (c) easily follow from the above inequalities.

Remark 4.1. The proof of Proposition 4.2 shows that the constant c in Assumption 2.1(a) is equal to $\mu - \lambda$. Consequently, for the discounted control problem, we can choose a discount rate $\alpha > 0$ such that $\alpha > \lambda - \mu$. In particular, if $\mu \ge \lambda$ (that is, the death rate is larger than or equal to the birth rate) then we can choose any $\alpha > 0$.

Under some additional hypotheses, however, we can make a sharper choice for c. Suppose, for instance, that there exists $D > 0$ such that

$$\inf_{b\in[b_1,b_2]} d(i,b) \ge Di \quad \text{for all } i > 0. \qquad (4.4)$$

From (4.3), it follows that the constant c in Assumption 2.1(a) is equal to $\mu + D - \lambda$. In this case, the discount rate $\alpha > 0$ must be such that $\alpha > \lambda - \mu - D$. So, if $\mu + D \ge \lambda$ (and, in particular, this allows the birth rate λ to be larger than the death rate μ) then we can choose an arbitrary discount rate $\alpha > 0$.

Now, we give the definition of the truncated control models $\mathcal{M}_n$ for $n \ge 1$. The state space is $S_n = \{0, 1, \dots, n\}$. Regarding the action sets, let

$$A_n(0) = \left\{\frac{\ell_1 a_2}{P_n} : 0 \le \ell_1 \le P_n\right\}$$

and

$$A_n(i) = \left\{\left(\frac{\ell_1 a_2}{P_n}, b_1 + \frac{\ell_2}{P_n}(b_2 - b_1)\right) : 0 \le \ell_1, \ell_2 \le P_n\right\} \quad \text{for } i > 0,$$

where $\{P_n\}_{n\ge1}$ is an arbitrary sequence of positive integers such that $\lim_n P_n = \infty$. Thus, the action sets $A_n(i)$ consist of a finite grid of points in $A(i)$, which are the vertices of rectangles with dimensions $a_2/P_n \times (b_2 - b_1)/P_n$. In particular, the action sets $A_n(i)$ verify the Kuratowski convergence property in Definition 3.1(b). The transition and reward rates of $\mathcal{M}_n$ are defined as at the start of this section. Hence, it is clear that the truncated control models $\mathcal{M}_n$ verify Proposition 4.1.

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

We choose a discount rate $\alpha > 0$ according to Remark 4.1. We can use the policy iteration algorithm to obtain *explicitly* the optimal discounted reward v_n^* and a discount optimal policy f_n^* of $\mathcal{M}_n$. By Theorem 3.1, v_n^* converges pointwise to the optimal discounted reward of $\mathcal{M}$, and the limit policies of $\{f_n^*\}_{n\geq 1}$ are discount optimal for $\mathcal{M}$.

4.2. Convergence rate results

Consider the control model $\mathcal{M}$ and the finite truncations $\mathcal{M}_n$ defined above, and suppose that, in addition, the condition in (4.4) is satisfied. Given an initial state $i \in S$, our goal now is to give lower bounds of $v^*(i)$ based on the approximations $v_n^*(i)$. As we shall see, such lower bounds can be established with a certain level of generality. Upper bounds will be obtained later for a particular choice of the functions $r(i,a,b)$ and $d(i,b)$.

Given a real number $\beta \geq 1$, direct calculations similar to those made in the proof of Proposition 4.2 yield some constant $\tilde{D}_\beta > 0$ such that

$$\sum_{j\in S} q_{ij}(a,b)w^\beta(j) \leq -\beta(\mu + D - \lambda)w^\beta(i) + \tilde{D}_\beta w^{\beta-1}(i) \quad \text{for all } (i,a,b) \in \mathbb{K}. \tag{4.5}$$

If $\mu + D \geq \lambda$ then choose any discount rate $\alpha > 0$ and an arbitrary $\beta_0 > 1$, while if $\mu + D < \lambda$, choose a discount rate $\alpha > \lambda - \mu - D$ and let β_0 satisfy

$$1 < \beta_0 < \frac{\alpha}{\lambda - \mu - D}. \tag{4.6}$$

From (4.5) we derive the existence of $\hat{D} > 0$ such that

$$\sum_{j\in S} q_{ij}(a,b)w^{\beta_0}(j) \leq \alpha w^{\beta_0}(i) + \hat{D} \quad \text{for all } (i,a,b) \in \mathbb{K}. \tag{4.7}$$

In what follows, we fix an arbitrary initial state i in S and $n \geq i$. Given a policy φ_0 in Φ or Φ_n, define the stopping time $\tau_n(\varphi_0)$ as the hitting time of $\{n, n+1, \dots\}$ (or, equivalently, since the population augments at most by one individual at each transition, the hitting time of $\{n\}$):

$$\tau_n(\varphi_0) := \inf\{t \geq 0 \colon x^{\varphi_0}(t) \geq n\} = \inf\{t \geq 0 \colon x^{\varphi_0}(t) = n\}.$$

Given a policy $\varphi \in \Phi_n$ (defined on the states of S_n), consider an arbitrary extension $\widetilde{\varphi}$ of φ to Φ (i.e. to the states of $S - S_n$). It is clear that, for the initial state $i \leq n$, we have $\tau_n := \tau_n(\varphi) = \tau_n(\widetilde{\varphi})$, and that $x^{\varphi}(t)$ and $x^{\widetilde{\varphi}}(t)$ coincide on $0 \leq t \leq \tau_n$. Therefore, for the control model $\mathcal{M}_n$,

$$\begin{aligned} v_n(i,\varphi) &= \mathrm{E}_{n,i}^{\varphi}\left[\int_0^{\tau_n} \mathrm{e}^{-\alpha t} r_n(t, x^{\varphi}(t), \varphi)\,\mathrm{d}t\right] + \mathrm{E}_{n,i}^{\varphi}\left[\int_{\tau_n}^{\infty} \mathrm{e}^{-\alpha t} r_n(t, x^{\varphi}(t), \varphi)\,\mathrm{d}t\right] \\ &= \mathrm{E}_i^{\widetilde{\varphi}}\left[\int_0^{\tau_n} \mathrm{e}^{-\alpha t} r(t, x^{\widetilde{\varphi}}(t), \widetilde{\varphi})\,\mathrm{d}t\right] + \mathrm{E}_{n,i}^{\varphi}[\mathrm{e}^{-\alpha\tau_n}]v_n(n,\varphi) \\ &= \mathrm{E}_i^{\widetilde{\varphi}}\left[\int_0^{\tau_n} \mathrm{e}^{-\alpha t} r(t, x^{\widetilde{\varphi}}(t), \widetilde{\varphi})\,\mathrm{d}t\right] + \mathrm{E}_i^{\widetilde{\varphi}}[\mathrm{e}^{-\alpha\tau_n}]v_n(n,\varphi). \end{aligned} \tag{4.8}$$

On the other hand, for the control model $\mathcal{M}$,

$$\begin{aligned} v(i,\widetilde{\varphi}) &= \mathrm{E}_i^{\widetilde{\varphi}}\left[\int_0^{\tau_n} \mathrm{e}^{-\alpha t} r(t, x^{\widetilde{\varphi}}(t), \widetilde{\varphi})\,\mathrm{d}t\right] + \mathrm{E}_i^{\widetilde{\varphi}}\left[\int_{\tau_n}^{\infty} \mathrm{e}^{-\alpha t} r(t, x^{\widetilde{\varphi}}(t), \widetilde{\varphi})\,\mathrm{d}t\right] \\ &= \mathrm{E}_i^{\widetilde{\varphi}}\left[\int_0^{\tau_n} \mathrm{e}^{-\alpha t} r(t, x^{\widetilde{\varphi}}(t), \widetilde{\varphi})\,\mathrm{d}t\right] + \mathrm{E}_i^{\widetilde{\varphi}}[\mathrm{e}^{-\alpha\tau_n}]v(n,\widetilde{\varphi}). \end{aligned}$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

Now, recalling (2.4) and (3.1), we have $|v(n,\tilde{\varphi})| \le Cw(n)$ and $|v_n(n,\varphi)| \le Cw(n)$. Therefore,

$$|v_n(i,\varphi) - v(i,\widetilde{\varphi})| \le 2Cw(n)\,\mathrm{E}_i^{\widetilde{\varphi}}[\mathrm{e}^{-\alpha\tau_n}].$$

(We note that the above calculations do not exclude the possibility that $\tau_n = \infty$ with positive probability.)

Now we use (4.7) and Dynkin's formula [5, Appendix C.3] for the function $(t,i) \mapsto \mathrm{e}^{-\alpha t} w^{\beta_0}(i)$ (which indeed applies because (4.5) holds for all $\beta \ge 1$) to obtain, for arbitrary $T > 0$,

$$\mathrm{E}_i^{\widetilde{\varphi}}[\mathrm{e}^{-\alpha(\tau_n\wedge T)} w^{\beta_0}(x(\tau_n \wedge T))] \le w^{\beta_0}(i) + \mathrm{E}_i^{\widetilde{\varphi}}\left[\int_0^{\tau_n\wedge T} \mathrm{e}^{-\alpha t}\hat{D}\,\mathrm{d}t\right] \le w^{\beta_0}(i) + \hat{D}\alpha^{-1}.$$

In this inequality we let $T \to \infty$ and, by dominated convergence, we obtain

$$\mathrm{E}_i^{\widetilde{\varphi}}[\mathrm{e}^{-\alpha\tau_n} w^{\beta_0}(n)\,\mathbf{1}\{\tau_n < \infty\}] \le w^{\beta_0}(i) + \hat{D}\alpha^{-1},$$

and so

$$\mathrm{E}_i^{\widetilde{\varphi}}[\mathrm{e}^{-\alpha\tau_n}] \le \frac{w^{\beta_0}(i) + \hat{D}\alpha^{-1}}{w^{\beta_0}(n)}. \tag{4.9}$$

Since $\varphi \in \Phi_n$ is arbitrary, we deduce that

$$\sup_{\varphi\in\Phi_n} |v_n(i,\varphi) - v(i,\tilde{\varphi})| \le \frac{2C(w^{\beta_0}(i) + \hat{D}\alpha^{-1})}{w^{\beta_0-1}(n)}. \tag{4.10}$$

Finally, let $\varphi = f_n^*$ be an optimal policy for $\mathcal{M}_n$ and let $\tilde{f}_n^*$ be an extension of f_n^* to Φ. By (4.10),

$$v_n^*(i) = v(i,\tilde{f}_n^*) + v_n(i,f_n^*) - v(i,\tilde{f}_n^*) \le v^*(i) + \frac{2C(w^{\beta_0}(i) + \hat{D}\alpha^{-1})}{w^{\beta_0-1}(n)}. \tag{4.11}$$

Therefore, we obtain a lower bound of $v^*(i)$ at a rate $1/n^{\beta_0-1}$ as $n \to \infty$. Recall that $\beta_0 > 1$ is arbitrary if $\mu + D \ge \lambda$, while if $\mu + D < \lambda$, the maximal convergence order is given by $\alpha/(\lambda - \mu - D)$ in (4.6). Finally, let us mention that the argument above cannot be used to derive an upper bound of $v^*(i)$. Indeed, the restriction to S_n of a discount optimal policy for $\mathcal{M}$ might not be in Φ_n because the corresponding actions need not belong to the action sets $A_n(i)$ for $i \in S_n$.

4.3. Numerical results

For the numerical experimentation, we fix the values of the parameters as

$$\lambda = 3.05, \qquad \mu = 3, \qquad a_2 = 5, \qquad b_1 = 5, \qquad b_2 = 8.$$

The catastrophe rate is given by $d(i,b) = ib/10$ for $i > 0$ and $b \in [5,8]$. The distribution $\{\gamma_i(j)\}$ of the catastrophe size is a truncated geometric distribution with parameter $\gamma = 0.8$; more precisely, given $i > 0$,

$$\gamma_i(j) = \frac{\gamma^{j-1}(1-\gamma)}{1-\gamma^i} \quad \text{for } 1 \le j \le i.$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

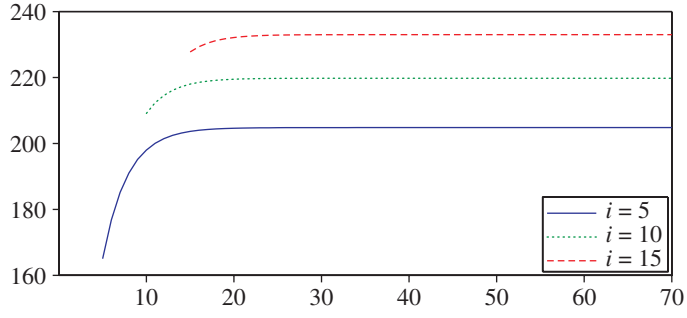
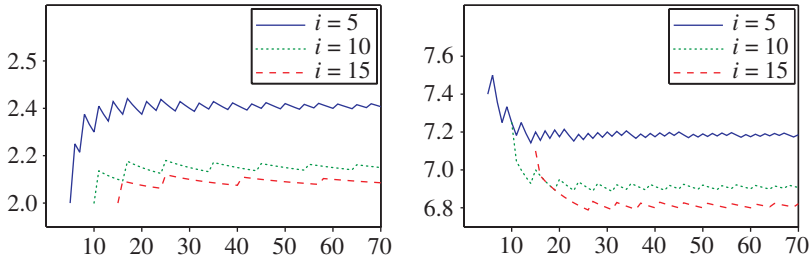


FIGURE 1: The optimal rewards $v_n^*(i)$ for $i = 5, 10, 15$.



(a) The optimal actions $a_n^*(i)$ (b) The optimal actions $b_n^*(i)$

FIGURE 2: The optimal policies $f_n^*(i)$ for $i = 5, 10, 15$.

Finally, the net reward rate is

$$r(i, a, b) = (10 - (a - 2)^2 - 0.5(b - 8)^2)i.$$

The interpretation of the term $(a - 2)^2$ is that we suppose that there is a natural immigration rate (which equals 2), and that augmenting or diminishing this natural immigration rate implies a cost for the controller. Similarly, the term $(b - 8)^2$ means that there is a natural catastrophe rate (which equals 8), and the controller incurs a cost when decreasing it. The discount rate is $\alpha = 0.1$ and P_n in the definition of the finite action sets $A_n(i)$ is equal to $2n$.

For every $1 \leq n \leq 70$, we solved the discounted control problem for $\mathcal{M}_n$. Given the initial states $i = 5, 10, 15$, the discount optimal rewards $v_n^*(i)$ and actions $a_n^*(i)$ and $b_n^*(i)$ are displayed in Figures 1 and 2, respectively, as functions of n. Empirically, we observe that the optimal reward and actions quickly converge, and become stable for relatively small values of n.

4.4. Convergence rate results (revisited)

Finally, we study the issue of the upper bounds for $v^*(i)$ for a given $i \in S$. We suppose that the functions $r(i, a, b)$ and $d(i, b)$ are as defined above. Let $f^* \in \mathbb{F}$ be a discount optimal policy for the control model $\mathcal{M}$. Given an initial state $i \in S$ and $n \geq i$, let $\tilde{v}_n(i, f^*)$ be the total

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

expected discounted reward of the policy f^* up to time $\tau_n(f^*)$, that is,

$$\tilde{v}_n(i, f^*) = \mathrm{E}_i^{f^*}\left[\int_0^{\tau_n(f^*)} \mathrm{e}^{-\alpha t} r(t, x^{f^*}(t), f^*)\,\mathrm{d}t\right]$$

with, in particular, $\tilde{v}_n(n, f^*) = 0$. By arguments similar to those used to derive (4.10), we have

$$|v^*(i) - \tilde{v}_n(i, f^*)| \le \frac{C(w^{\beta_0}(i) + \hat{D}\alpha^{-1})}{w^{\beta_0 - 1}(n)}. \tag{4.12}$$

Also, since $\{\tilde{v}_n(\cdot, f^*)\}_{i \in S_n}$ is the expected discounted reward of a policy with transition rates equal to $q_{jk}(f^*)$ for $0 \le j < n$ and $0 \le k \le n$, and equal to 0 if $j = n$, and reward rates equal to $r(j, f^*)$ for $0 \le j < n$ and equal to 0 for $j = n$, $\tilde{v}_n(\cdot, f^*)$ verifies (see Theorem 6.9.c of [5])

$$\alpha \tilde{v}_n(j, f^*) = r(j, f^*) + \sum_{0 \le k \le n} q_{jk}(f^*)\tilde{v}_n(k, f^*) \quad \text{for all } 0 \le j < n.$$

For each $0 \le j \le n$, let $f_n(j) \in A_n(j)$ be the closest point to $f^*(j) \in A(j)$. In particular, $\|f^*(j) - f_n(j)\| \le B/P_n$, where $\|\cdot\|$ stands for the Euclidean norm and the constant $B > 0$ does not depend on n. In this way, we define a policy $f_n \in \mathbb{F}_n$. Using the mean value theorem [12, Theorem 5.10], it can be shown after some elementary calculations that, for some constant $\tilde{B} > 0$ that does not depend on n,

$$\alpha \tilde{v}_n(j, f^*) \le r(j, f_n) + \sum_{0 \le k \le n} q_{jk}(f_n)\tilde{v}_n(k, f^*) + \frac{\tilde{B}n^2}{P_n} \quad \text{for all } 0 \le j < n.$$

Equivalently,

$$\alpha\left(\tilde{v}_n(j, f^*) - \frac{\tilde{B}n^2}{\alpha P_n}\right) \le r_n(j, f_n) + \sum_{0 \le k \le n} q_{jk}^n(f_n)\left(\tilde{v}_n(k, f^*) - \frac{\tilde{B}n^2}{\alpha P_n}\right) \quad \text{for all } 0 \le j < n,$$

from which [5, Theorem 6.9.b], for $0 \le i < n$,

$$\tilde{v}_n(i, f^*) - \frac{\tilde{B}n^2}{\alpha P_n} \le \mathrm{E}_{n,i}^{f_n}\left[\int_0^{\tau_n(f_n)} \mathrm{e}^{-\alpha t} r_n(t, x^{f_n}(t), f_n)\,\mathrm{d}t\right].$$

Now, proceeding as in (4.8) and (4.9), we obtain

$$\tilde{v}_n(i, f^*) - \frac{\tilde{B}n^2}{\alpha P_n} \le v_n(i, f_n) + \frac{C(w^{\beta_0}(i) + \hat{D}\alpha^{-1})}{w^{\beta_0 - 1}(n)} \le v_n^*(i) + \frac{C(w^{\beta_0}(i) + \hat{D}\alpha^{-1})}{w^{\beta_0 - 1}(n)}.$$

Together with (4.12), this yields

$$v^*(i) \le v_n^*(i) + \frac{2C(w^{\beta_0}(i) + \hat{D}\alpha^{-1})}{w^{\beta_0 - 1}(n)} + \frac{\tilde{B}n^2}{\alpha P_n}.$$

Therefore, recalling (4.11), it follows that, for sufficiently 'fine' partitions of the state space (in particular, we can choose $P_n = \mathrm{O}(n^{1+\beta_0})$), we obtain

$$|v_n^*(i) - v^*(i)| = \mathrm{O}(n^{-(\beta_0 - 1)}).$$

https://doi.org/10.1239/jap/1354716658 Published online by Cambridge University Press

Hence, the convergence of $v_n^*(i)$ to $v^*(i)$ is of order $\beta_0 - 1$, where

- $\beta_0 > 1$ is arbitrary if $\mu + D \geq \lambda$; and
- $\beta_0 < \alpha/(\lambda - \mu - D)$ if $\mu + D < \lambda$; recall (4.6).

Remark 4.2. We note that the main feature of the functions $r(i, a, b)$, $d(i, b)$, and $q_{ij}(a, b)$ used to derive the above upper bound of $v^*(i)$ is that they are Lipschitz continuous on $A(i)$, with a Lipschitz constant that is $O(i)$.

References

[1] Altman, E. (1994). Denumerable constrained Markov decision processes and finite approximations. *Math. Operat. Res.* **19,** 169–191.

[2] Álvarez-Mena, J. and Hernández-Lerma, O. (2002). Convergence of the optimal values of constrained Markov control processes. *Math. Meth. Operat. Res.* **55,** 461–484.

[3] Guo, X. and Hernández-Lerma, O. (2003). Continuous-time controlled Markov chains with discounted rewards. *Acta Appl. Math.* **79,** 195–216.

[4] Guo, X. and Hernández-Lerma, O. (2003). Drift and monotonicity conditions for continuous-time controlled Markov chains with an average criterion. *IEEE Trans. Automatic Control* **48,** 236–245.

[5] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes*. Springer, Berlin.

[6] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer, New York.

[7] Kushner, H. J. and Dupuis, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time,* 2nd edn. Springer, New York.

[8] Langen, H.-J. (1981). Convergence of dynamic programming models. *Math. Operat. Res.* **6,** 493–512.

[9] Leizarowitz, A. and Shwartz, A. (2008). Exact finite approximations of average-cost countable Markov decision processes. *Automatica J. IFAC* **44,** 1480–1487.

[10] Prieto-Rumeau, T. and Hernández-Lerma, O. (2010). Policy iteration and finite approximations to discounted continuous-time controlled Markov chains. In *Modern Trends in Controlled Stochastic Processes*, ed. A. B. Piunovskiy, Luniver Press, pp. 84–101.

[11] Prieto-Rumeau, T. and Lorenzo, J. M. (2010). Approximating ergodic average reward continuous-time controlled Markov chains. *IEEE Trans. Automatic Control* **55,** 201–207.

[12] Rudin, W. (1976). *Principles of Mathematical Analysis,* 3rd edn. McGraw-Hill, New York.

[13] Song, Q. S. (2008). Convergence of Markov chain approximation on generalized HJB equation and its applications. *Automatica J. IFAC* **44,** 761–766.

[14] Tidball, M. M., Lombardi, A., Pourtallier, O. and Altman, E. (2000). Continuity of optimal values and solutions for control of Markov chains with constraints. *SIAM J. Control Optimization* **38,** 1204–1222.

[15] Whitt, W. (1978). Approximation of dynamic programs. I. *Math. Operat. Res.* **3,** 231–243.