Adv. Appl. Prob. **42,** 953–985 (2010)
Printed in Northern Ireland
© *Applied Probability Trust* 2010

NEW DISCOUNT AND AVERAGE OPTIMALITY CONDITIONS FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES

XIANPING GUO * ** AND
LIUER YE,* *** *Sun Yat-Sen University*

Abstract

This paper deals with continuous-time Markov decision processes in Polish spaces, under the discounted and average cost criteria. All underlying Markov processes are determined by given transition rates which are allowed to be *unbounded*, and the costs are assumed to be *bounded below*. By introducing an occupation measure of a randomized Markov policy and analyzing properties of occupation measures, we first show that the family of all randomized stationary policies is 'sufficient' within the class of all randomized Markov policies. Then, under the semicontinuity and compactness conditions, we prove the existence of a discounted cost optimal stationary policy by providing a value iteration technique. Moreover, by developing a *new* average cost, minimum nonnegative solution method, we prove the existence of an average cost optimal stationary policy under some reasonably mild conditions. Finally, we use some examples to illustrate applications of our results. Except that the costs are assumed to be bounded below, the conditions for the existence of discounted cost (or average cost) optimal policies are much *weaker* than those in the previous literature, and the minimum nonnegative solution approach is *new*.

Keywords: Occupation measure; optimality inequality/equation; discounted cost optimal policy; average cost optimal policy; minimum nonnegative solution method

2010 Mathematics Subject Classification: Primary 90C40; 60J27

1. Introduction

Continuous-time Markov decision processes (MDPs) have received considerable attention because many optimization models, such as in communication engineering, queueing systems, and control of epidemics, are based on the processes involving continuous time. As is well known, the *expected discount* and *average criteria* are most commonly used in continuous-time MDPs; see, for instance, [3], [4], [6], [7], [8], [9], [11], [12], [15, Chapter 4], [16], [18], [19, Section 5, Chapter 11], [20, Chapter 10], [24], [25], and their extensive references. The main focus of the aforementioned works is the so-called optimality conditions that ensure the existence/calculation of optimal policies. For this reason, the state spaces in [4], [8], [9], [11], [12], [15], [16], [18], [19], and [20] were assumed to be *denumerable*, and taken to be Polish spaces in [3], [7], [6], [24], and [25]. In this paper we will further study the case of continuous-time MDPs in Polish spaces, and, thus, only state results from some of the aforementioned

Received 14 January 2009; revision received 28 June 2010.

* Postal address: School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, 510275, P. R. China.

Research supported by the NSFC and GDUPS (2010).

** Email address: mcsgxp@mail.sysu.edu.cn
*** Email address: yeliuer@hotmail.com

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

works. For the discount criterion of continuous-time MDPs in Polish spaces, rewards in [3] were assumed to be bounded, and, for the case of unbounded rewards in [7], *additional* assumptions, such as the expected growth and absolute integrability conditions, were made. For the average criterion of continuous-time MDPs in Polish spaces, the main optimality conditions are based on the corresponding approaches. Roughly speaking, the optimality equation approach in [25] requires the uniformly *exponentially ergodic* condition; the *optimality two-inequality approach* in [10] requires that a *drift (Lyapunov type)* function can dominate the relative difference of the discount optimal value function; the *convex analysis method* in [6] needs another set of *convergence conditions* for sequences of measures; and the *optimality inequality approach* in [24] requires some additional *absolute integrability* conditions and the assumption that the relative difference is bounded below. Furthermore, the absolute integrability conditions, as well as the *continuity assumption* imposed on the class of policies, were required in [3], [7], [6], [10], [24], and [25], owing to the use of Dynkin's formula, and the existing examples verifying the average optimality conditions in [10] and [11] need the *monotonicity assumption* to be imposed on the rewards. In this paper we further study both the expected discount and average criteria, and aim to drop both the integrability conditions and the continuity assumption required in [3], [6], [7], [10], [24], and [25] for Polish state spaces. To this end, we develop some new techniques, and also give some *new* and *verifiable* conditions as well as examples for both discount and average optimalities.

More precisely, the state and action spaces in our model are allowed to be Polish spaces, the costs are bounded below, the transition rates may be unbounded, and each randomized Markov policy may *not* satisfy the continuity condition in [3], [6], [7], [10], [24], and [25]. Since we established the existence of a transition function without the continuity condition in [23], the discount and average optimality problems can be well defined when the costs are bounded below. To prove the existence of discounted cost optimal policies, we introduce an occupation measure of a randomized Markov policy and analyze its properties. These properties are used to prove that the family of all randomized stationary policies is 'sufficient' within the class of all randomized Markov policies (see Theorem 3.1). Then, using a value iteration technique, we not only establish the discount cost optimality equation and show the existence of a discounted cost optimal stationary policy, but also prove that such value iteration techniques can be used to calculate (or at least approximate) the discounted cost optimal value (see Theorem 3.2). For the average criterion, we first give reasonably mild conditions and present some *new* sufficient conditions for the verification of these average optimality conditions (see Theorem 3.3). Then, in order to prove the existence of average cost optimal policies, we obtain a *key* fact by developing a *new average cost minimum nonnegative solution technique* (see Theorem 3.4), which together with the optimality inequality approach is used to prove the existence of an average cost optimal policy (see Theorem 3.5). Furthermore, we illustrate the applications of our results with examples, which satisfy all of our conditions, but in which the monotonicity assumption commonly used in the examples in the literature fails to hold.

It is worth noting that, except for the fact that the costs are bounded below, the conditions for the existence of discounted cost (or average cost) optimal policies are *much weaker* than those in the literature [3], [6], [7], [10], [24], [25]. More precisely, the improvement in the corresponding optimality conditions is twofold: (i) the expected growth condition in [3] and [7] for the discount criterion has been dropped (see Remark 3.4 for details), and (ii) the exponentially ergodic condition in [25] and the drift conditions imposed on the relative difference of the discounted cost optimal value in [10] and [25] have been removed (see Remark 3.8). Moreover, the absolute integrability condition and the continuity assumption in [3], [6], [7], [10], [24], and

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

[25] are no longer needed in this paper (see Remarks 3.4 and 3.8). Furthermore, some *new* sufficient conditions and examples verifying our assumptions are given (see Remark 3.6 and Examples 4.1–4.3). The minimum nonnegative solution technique and the results of a transition function constructed from transition rates under the measurability condition are *new* and play key roles in our arguments (see Remark 3.7).

The rest of this paper is organized as follows. In Section 2 we introduce the model that we are concerned with. The main optimality results for the two optimality criteria are stated in Section 3, and are illustrated with examples in Section 4. The proofs of these results are postponed to Section 5.

2. The nonnegative cost model

Notation. If X is a Polish space (that is, a complete and separable metric space), we denote by $\mathscr{B}(X)$ the Borel σ-algebra, and by $\mathrm{P}(X)$ the set of all probability measures on $\mathscr{B}(X)$ endowed with the topology of weak convergence.

The model of continuous-time MDPs is defined by

$$\{S, (A(x) \subseteq A,\ x \in S), q(\cdot \mid x, a), c(x, a)\}, \tag{2.1}$$

where $q(\cdot \mid x, a)$ denotes the transition rates, $c(x, a)$ is the cost function, S is a *state space*, A is an *action space*, and $A(x) \in \mathscr{B}(A)$ denotes the set of *admissible actions* at state $x \in S$. We suppose that S and A are Polish spaces. The set

$$K := \{(x, a) \mid x \in S,\ a \in A(x)\} \tag{2.2}$$

is a Borel subset of $S \times A$.

The transition rates, $q(\cdot \mid x, a)$, satisfy the following properties.

(T1) For each fixed $(x, a) \in K$, $q(\cdot \mid x, a)$ is a signed measure on $\mathscr{B}(S)$, whereas, for each fixed $D \in \mathscr{B}(S)$, $q(D \mid \cdot)$ is a real-valued Borel-measurable function on K.

(T2) $0 \leq q(D \mid x, a) < \infty$ for all $(x, a) \in K$ and $x \notin D \in \mathscr{B}(S)$.

(T3) $q(S \mid x, a) = 0$ for all $(x, a) \in K$.

The model is assumed to be *stable*, which means that

$$q^*(x) := \sup_{a \in A(x)} |q(\{x\} \mid x, a)| < \infty \quad \text{for all } x \in S. \tag{2.3}$$

The cost function, $c(x, a)$, is assumed to be bounded below and measurable on K.

A continuous-time MDP evolves as follows. The decision maker *continuously* observes the current state of a system. Whenever the system is at state $x(t) \in S$ at time $t \geq 0$, he/she chooses an action $a(t) \in A(x(t))$ according to some rule. Consequently, he/she incurs an immediate cost $c(x(t), a(t))$ and the system moves to a new state set governed by a possibly nonhomogeneous transition probability function, which is determined by the transition rates $q(\cdot \mid x(t), a(t))$. Thus, the goal of the decision maker is to minimize his/her costs with respect to some performance criterion, such as $V_\alpha(\cdot, \cdot, \cdot)$ or $J(\cdot, \cdot, \cdot)$, respectively defined in (2.9) and (2.10) below.

To define a rule precisely, we introduce some notation.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Definition 2.1. A *(randomized Markov) policy* is a family $\pi := (\pi_t,\ t \geq 0)$ of stochastic kernels π_t that satisfy the following conditions.

(a) For each $t \geq 0$, π_t is a stochastic kernel on A given S such that $\pi_t(A(x) \mid x) = 1$ for all $x \in S$.

(b) For each $B \in \mathcal{B}(A)$, $\pi_t(B \mid x)$ is Borel measurable in $(t, x) \in [0, \infty) \times S$.

We denote by Π the family of all randomized Markov policies.

By Definition 2.1(a), without loss of generality, we regard $\pi_t(\cdot \mid x)$ as a probability measure on $A(x)$.

A policy $\pi = (\pi_t,\ t \geq 0) \in \Pi$ is called *randomized stationary* if there exists a stochastic kernel ϕ on A given S such that

$$\pi_t(\cdot \mid x) = \phi(\cdot \mid x) \quad \text{for all } t \geq 0 \text{ and } x \in S.$$

The set of all randomized stationary policies is denoted by Π_s.

A randomized stationary policy $\phi \in \Pi_s$ is called *deterministic stationary* or simply *stationary* if there exists a Borel-measurable function f on S with $f(x) \in A(x)$ for all $x \in S$ such that

$$\phi(\{f(x)\} \mid x) = 1 \quad \text{for all } x \in S.$$

For simplicity, we denote such a policy ϕ by f. The set of all stationary policies is denoted by F, which means that F is the set of all Borel-measurable functions f on S such that $f(x) \in A(x)$ for all $x \in S$. Obviously, $\Pi \supset \Pi_s \supset F$.

For each $\pi = (\pi_t,\ t \geq 0) \in \Pi$, we define the associated transition rates $q_\pi(\cdot \mid x, \pi_t)$ by

$$q_\pi(D \mid x, \pi_t) := \int_{A(x)} q(D \mid x, a)\pi_t(\mathrm{d}a \mid x)$$

for all $x \in S$, $D \in \mathcal{B}(S)$, and $t \geq 0$.

The function $q_\pi(\cdot \mid x, \pi_t)$ is also called an infinitesimal generator (for any fixed policy $\pi \in \Pi$); see, e.g. [3]. As is well known, any (possibly substochastic and nonhomogeneous) *transition function* $\tilde{p}_\pi(s, x, t, D)$ depending on π such that

$$\lim_{\varepsilon \to 0^+} \frac{\tilde{p}_\pi(t, x, t + \varepsilon, D) - \delta_x(D)}{\varepsilon} = q_\pi(D \mid x, \pi_t)$$

for all $x \in S$, $t \geq 0$, and $D \in \mathcal{B}(S)$ is called a $Q(t, \pi)$-*transition function* with transition rates $q_\pi(\cdot \mid x, \pi_t)$, where $\delta_x(D)$ denotes the Dirac measure at point $x \in S$.

By Theorem 1 of [23] (see, e.g. [5] and [22]), we have the following fact.

Lemma 2.1. *For each policy $\pi = (\pi_t,\ t \geq 0)$ in Π, there exists a $Q(t, \pi)$-transition function with transition rates $q_\pi(\cdot \mid x, \pi_t)$.*

Lemma 2.1 guarantees the existence of a $Q(t, \pi)$-transition function, such as the *minimum* $Q(t, \pi)$-transition function denoted by $p_\pi^{\min}(s, x, t, D)$, which can be constructed from $q_\pi(\cdot \mid x, \pi_t)$ (see [5] and [23]). But, as is well known (see, e.g. [1, Theorem 2.2.2]), such a $Q(t, \pi)$-transition function might not be regular, that is, we might have $p_\pi^{\min}(s, x, t, S) < 1$ for some $x \in S$ and $t \geq s \geq 0$.

To ensure the regularity of a $Q(t, \pi)$-transition function, we propose the following 'drift condition'.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Assumption A. There exist a measurable function $w \geq 1$ on S, and constants $c_0 \in (-\infty, \infty)$, $b_0 \geq 0$, and $M_0 > 0$ such that

(i) $\int_S w(y)q(\mathrm{d}y \mid x, a) \leq c_0 w(x) + b_0$ for all $(x, a) \in K$; and

(ii) $q^*(x) \leq M_0 w(x)$ for all $x \in S$, with $q^*(x)$ as in (2.3).

Remark 2.1. (a) Assumption A is satisfied when the transition rates are bounded (that is, $\sup_{x \in S} q^*(x) < \infty$).

(b) Assumption A(i) is an extension of the 'drift condition' (2.4) of [17] for a *homogeneous* Q-transition function.

(c) Under Assumption A, it follows from Theorem 2 of [23] that $p_\pi^{\min}(s, x, t, S) \equiv 1$. Hence, the $Q(t, \pi)$-transition function with transition rates $q_\pi(\cdot \mid x, \pi_t)$ is *regular* and *unique*. Thus, we write $p_\pi^{\min}(s, x, t, D)$ simply as $p_\pi(s, x, t, D)$.

For each initial distribution $\nu \in P(S)$, initial time $s \geq 0$, and $\pi = (\pi_t,\ t \geq 0) \in \Pi$, as is well known, there exists a unique probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{P}_{s,\nu}^\pi)$, in which the probability measure $\mathrm{P}_{s,\nu}^\pi$ is completely determined by ν and $p_\pi(s, x, t, D)$. Then, Lemma 2.1 of [6] ensures the existence of a 'state and action' process $\{x(t), a(t),\ t \geq s\}$ such that

$$\mathrm{P}_{s,\nu}^\pi((x(t), a(t)) \in K) = 1, \tag{2.4}$$

$$\mathrm{P}_{s,\nu}^\pi(x(s) \in D,\ a(s) \in C) = \int_D \pi_s(C \mid y)\nu(\mathrm{d}y), \tag{2.5}$$

$$\mathrm{P}_{s,\nu}^\pi(x(t) \in D,\ a(t) \in C) = \int_S \int_D \pi_t(C \mid y) p_\pi(s, x, t,\ \mathrm{d}y)\nu(\mathrm{d}x), \tag{2.6}$$

for all $t \geq 0$, $D \in \mathcal{B}(S)$, and $C \in \mathcal{B}(A)$.

Let $\mathrm{E}_{s,\nu}^\pi$ denote the expectation operator associated with $\mathrm{P}_{s,\nu}^\pi$. In particular, if ν is concentrated on the 'initial state' x at time s (i.e. $\nu(\{x\}) = 1$), we write $\mathrm{P}_{s,\nu}^\pi$ and $\mathrm{E}_{s,\nu}^\pi$ as $\mathrm{P}_{s,x}^\pi$ and $\mathrm{E}_{s,x}^\pi$, respectively. Furthermore, if $s = 0$, we write $\mathrm{P}_{s,x}^\pi$ and $\mathrm{E}_{s,x}^\pi$ as P_x^π and E_x^π, respectively.

For each $\pi = (\pi_t,\ t \geq 0) \in \Pi$, let

$$c(x, \pi_t) := \int_{A(x)} c(x, a)\pi_t(\mathrm{d}a \mid x) \quad \text{for each } x \in S \text{ and } t \geq 0. \tag{2.7}$$

Then, by (2.6) and (2.7), we have

$$\mathrm{E}_{s,x}^\pi\, c(x(t), a(t)) = \mathrm{E}_{s,x}^\pi\, c(x(t), \pi_t) := \int_S c(y, \pi_t) p_\pi(s, x, t,\ \mathrm{d}y) \quad \text{for each } t \geq s. \tag{2.8}$$

Since $c(x, a)$ is measurable in $(x, a) \in K$, and $\pi_t(C \mid x)$ is measurable in $(t, x) \in \bar{S} := [0, \infty) \times S$ (for each fixed $C \in \mathcal{B}(A)$), it follows from (2.7) that $c(x, \pi_t)$ is measurable in $(t, x) \in \bar{S}$, and so is the expected cost $\mathrm{E}_{s,x}^\pi\, c(x(t), a(t))$ because $p_\pi(s, x, t,\ \mathrm{d}y)$ is continuous in $t \geq s$; see, e.g. [23].

Let $\alpha (> 0)$ be a fixed discount factor. For each policy $\pi \in \Pi$, $s \geq 0$, and $x \in S$, the *expected discounted cost* and *average cost criteria* are defined as

$$V_\alpha(s, x, \pi) := \int_s^\infty \mathrm{e}^{-\alpha(t-s)}\, \mathrm{E}_{s,x}^\pi\, c(x(t), a(t))\, \mathrm{d}t \tag{2.9}$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

and

$$J(s,x,\pi) := \limsup_{T\to\infty} \frac{1}{T-s} \int_s^T \mathrm{E}^{\pi}_{s,x}\, c(x(t),a(t))\,\mathrm{d}t, \tag{2.10}$$

respectively. The corresponding *discounted cost* and *average cost optimal value* functions are given by

$$V_\alpha{}^*(s,x) := \inf_{\pi\in\Pi} V_\alpha(s,x,\pi) \quad \text{for all } s \geq 0 \text{ and } x \in S$$

and

$$J^*(s,x) := \inf_{\pi\in\Pi} J(s,x,\pi) \quad \text{for all } s \geq 0 \text{ and } x \in S,$$

respectively.

Definition 2.2. A policy $\pi^* \in \Pi$ is said to be discounted cost optimal if

$$V_\alpha(s,x,\pi^*) \leq V_\alpha{}^*(s,x) \quad \text{for all } s \geq 0 \text{ and } x \in S.$$

Similarly, a policy $\pi^* \in \Pi$ is said to be average cost optimal if

$$J(s,x,\pi^*) \leq J^*(s,x) \quad \text{for all } s \geq 0 \text{ and } x \in S.$$

Remark 2.2. Note that, since $p_\pi(s,x,t,D)$ is regular under Assumption A (see Remark 2.1(c)), without loss of generality, we may replace the costs $c(x,a)$ in (2.9)–(2.10) with $c(x,a)+L$ for any constant L. Therefore, in the following arguments, we will assume that '$c(x,a) \geq 0$' since $c(x,a)$ in model (2.1) is bounded below.

3. Main results

In this section we state our main results. Their applications are illustrated with examples in Section 4, and their proofs are postponed to Section 5.

3.1. On discount optimality

In this subsection we present the main results of discounted cost optimality. To this end, we introduce the concept of an occupation measure of a policy.

Definition 3.1. (a) The occupation measure of a policy $\pi \in \Pi$ is a measure μ^π (depending on π) on $S \times A$, which is defined by

$$\mu^\pi(\Gamma) := \int_s^\infty \mathrm{e}^{-\alpha(t-s)}\, \mathrm{P}^{\pi}_{s,\nu}((x(t),a(t)) \in \Gamma)\,\mathrm{d}t \quad \text{for } \Gamma \in \mathcal{B}(S\times A).$$

(Obviously, μ^π also depends on the initial distribution ν at $s \geq 0$, but it is still denoted as μ^π for simplicity.)

(b) Two policies π^1 and π^2 in Π are called equivalent if $\mu^{\pi^1} = \mu^{\pi^2}$.

Remark 3.1. (a) By Definition 2.1(a) and (2.4)–(2.6), we have $\mu^\pi(S\times A) = 1/\alpha$, and μ^π is concentrated on K in (2.2), i.e.

$$\mu^\pi(K^{\mathrm{c}}) = 0,$$

where K^{c} denotes the complement of K.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

(b) The *marginal* (or *projection*) $\hat{\mu}_S^\pi$ of μ^π on S is given by

$$\hat{\mu}_S^\pi(D) := \mu^\pi(D \times A) = \int_s^\infty \mathrm{e}^{-\alpha(t-s)}\, \mathrm{P}_{s,\nu}^\pi(x(t) \in D)\, \mathrm{d}t \quad \text{for each } D \in \mathcal{B}(S). \tag{3.1}$$

Theorem 3.1. *Let $\nu \in P(S)$ be any initial distribution. Suppose that Assumption A holds, that $\int_S w(y)\nu(\mathrm{d}y) < \infty$, and that $\alpha > c_0$, with α the discount factor, and c_0 and w as in Assumption A. Then the following assertions hold.*

(a) *For each fixed policy $\pi \in \Pi$, the occupation measure μ^π is a solution to the equation*

$$\alpha\hat{\mu}_S^\pi(D) = \nu(D) + \int_{S\times A} q(D \mid x, a)\mu^\pi(\mathrm{d}x,\, \mathrm{d}a) \quad \textit{for all } D \in \mathcal{B}(S).$$

(b) *Conversely, if a measure μ on $\mathcal{B}(S \times A)$ concentrated on K satisfies $\mu(K) = 1/\alpha$, $\int_S w(y)\hat{\mu}_S(\mathrm{d}y) < \infty$, and*

$$\alpha\hat{\mu}_S(D) = \nu(D) + \int_{S\times A} q(D \mid x, a)\mu(\mathrm{d}x,\, \mathrm{d}a) \quad \textit{for all } D \in \mathcal{B}(S), \tag{3.2}$$

then there exists a randomized stationary policy $\phi^\mu \in \Pi_s$ (depending on μ) such that $\mu^{\phi^\mu} = \mu$, and ϕ^μ can be given by

$$\mu(D \times C) = \int_D \phi^\mu(C \mid x)\hat{\mu}_S(\mathrm{d}x) \quad \textit{for all } D \in \mathcal{B}(S) \textit{ and } C \in \mathcal{B}(A). \tag{3.3}$$

(c) $V_\alpha{}^*(x) := \inf_{\phi\in\Pi_s} V_\alpha(0, x, \phi) = \inf_{\pi\in\Pi} V_\alpha(s, x, \pi)$ *for all $x \in S$ and $s \geq 0$.*

Proof. See Section 5.

Remark 3.2. Theorem 3.1(a) and (b) state some properties of an occupation measure of a policy. Moreover, Theorem 3.1(c) shows that the family Π_s of all randomized stationary policies is 'sufficient' within the class Π of all randomized Markov policies for the discount optimality when the costs are *bounded below*.

To guarantee the existence of a discounted cost optimal stationary policy, we also need the following assumption.

Assumption B. Suppose that the following conditions hold:

(i) $\alpha > c_0$, with c_0 as in Assumption A;

(ii) $A(x)$ is compact for each $x \in S$;

(iii) for each $x \in S$ and $D \in \mathcal{B}(S)$, the function $q(D \mid x, \cdot)$ is continuous on $A(x)$;

(iv) for each $x \in S$, the function $c(x, \cdot)$ is lower semicontinuous (l.s.c.) on $A(x)$.

Remark 3.3. Assumption B(i) follows from the condition in Theorem 3.1. It is required for the finiteness of $\int_S w(x)\mu^\pi(\mathrm{d}x)$, but not needed when the transition rates are bounded. We call Assumption B(ii) and (iii) the so-called 'semi-continuity and compactness conditions', which are imposed on the *primitive data* of model (2.1) and are an extension of the standard *continuity-compactness* conditions in [13, Assumptions 4.2.1 and 4.2.2], [14, Assumptions 8.3.1 and 8.3.3], and [19, Theorem 6.2.10] for discrete-time MDPs.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

To state our second main result for discounted cost optimality, we define a sequence $\{u_n\}$ as follows. For each $n \geq 0$, let

$$u_{n+1}(x) := \inf_{a \in A(x)} \left\{ \frac{c(x,a)}{\alpha + q(x,a)} + \frac{1}{\alpha + q(x,a)} \int_{S-\{x\}} u_n(y) q(\mathrm{d}y \mid x, a) \right\} \tag{3.4}$$

for $x \in S$, where $u_0 := 0$ and $q(x,a) := -q(\{x\} \mid x, a) (\geq 0)$.

Theorem 3.2. *Suppose that Assumptions A and B hold. Then*

(a) $\lim_{n\to\infty} u_n = V_\alpha^*$*, with* u_n *as in (3.4) and* V_α^* *as in Theorem 3.1(c);*

(b) V_α^* *satisfies the following discounted cost optimality equation:*

$$V_\alpha^*(x) = \inf_{a \in A(x)} \left\{ \frac{c(x,a)}{\alpha + q(x,a)} + \frac{1}{\alpha + q(x,a)} \int_{S-\{x\}} V_\alpha^*(y) q(\mathrm{d}y \mid x, a) \right\} \tag{3.5}$$

for all $x \in S$*;*

(c) *any policy* $f \in F$ *realizing the minimum in the right-hand side of (3.5) is discounted cost optimal;*

(d) *there exists a discounted cost optimal stationary policy.*

Proof. See Section 5.

Remark 3.4. (a) Theorem 3.2 not only shows the existence of a discounted cost optimal stationary policy, but also provides a value iteration algorithm to approximate the discounted cost optimal value function V_α^*.

(b) Except for the fact that the costs are bounded below, the other conditions are much weaker than those in [7]. For example, we have *dropped* the following three assumptions required in [7]: (i) the *continuity* condition imposed on the class of all randomized Markov policies (see Definition 2.1 of [7]), (ii) Assumption B(1) of [7] (i.e. the so-called *expected growth* condition) for the finiteness of the expected discounted cost function, and (iii) Assumption C(4) of [7] (i.e. the *absolute integrability* condition) for the interchange of integrals and sums.

3.2. On average optimality

In this subsection we focus on the main results of the existence of an average cost optimal stationary policy.

In addition to Assumptions A and B, to ensure the existence of an average cost optimal policy, we need the following hypothesis.

Assumption C. For some decreasing sequence $\{\alpha_n\}$ tending to 0 and some state $x_0 \in S$, there exist a constant L^* and a nonnegative real-valued function H on S such that

(i) $\alpha_n V_{\alpha_n}^*(x_0)$ is bounded in $n \geq 1$ (this implies that $V_{\alpha_n}^*(x_0) < \infty$, and so we may define the relative difference of the discount optimal value function $h_{\alpha_n}(x) := V_{\alpha_n}^*(x) - V_{\alpha_n}^*(x_0)$ on S for each $n \geq 1$);

(ii) $L^* \leq h_{\alpha_n}(x) \leq H(x)$ for all $n \geq 1$ and $x \in S$.

Remark 3.5. Assumption C is a continuous-time version of Assumption 5.4.1 of [13] for discrete-time MDPs. Such a hypothesis is commonly used in discrete-time MDPs, and examples

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

satisfying this hypothesis are given in [19, pp. 421–425], but for the case of discrete-time MDPs with denumerable states.

For the verification of Assumption C, since Theorem 3.3(a) of [7] can be used to verify Assumption C(i), we need to verify only Assumption C(ii). To this end, we introduce some notation.

Suppose that there is a set $B \in \mathcal{B}(S)$ such that, for each $f \in F$ and $x \notin B$, either $q(B \mid x, f(x)) > 0$ or there are some distinct sets $B_1, B_2, \dots, B_n \in \mathcal{B}(S)$ (depending on f and x) satisfying

$$q(B_1 \mid x, f(x)) > 0, \qquad q(B_{k+1} \mid x_k, f(x_k)) > 0, \quad k = 1, \dots, n-1,$$
$$\text{and} \quad q(B \mid x_n, f(x_n)) > 0 \quad \text{for all } x_k \in B_k,\ k = 1, 2, \dots, n\,.$$

Then we know that such a set B can be reached from state $x \notin B$ under any $f \in F$, which is denoted by '$x \hookrightarrow B$'. For the aforementioned $B \in \mathcal{B}(S)$ and $f \in F$, we denote by

$$\tau_B^f := \begin{cases} \inf\{t > 0 \colon x(t) \in B\} & \text{if } \{t > 0 \colon x(t) \in B\} \neq \varnothing, \\ +\infty & \text{otherwise,} \end{cases}$$

the first entrance time to B. We can see that $\tau_B^f < \infty$, P_x^f-almost surely (P_x^f-a.s.). When B is a singleton set $\{x_0\}$, we denote τ_B^f simply by $\tau_{x_0}^f$. Furthermore, for any $\delta \geq 0$ and any *nonnegative* Borel-measurable function g on K, define

$$U_\delta^B(x, f) := \mathrm{E}_x^f\left[\int_0^{\tau_B^f} \mathrm{e}^{-\delta t} g(x(t), f)\, \mathrm{d}t\right] \quad \text{for all } x \notin B \text{ and } f \in F, \tag{3.6}$$

where $g(x, f) := g(x, f(x))$.

Then we have the following fact for the verification of Assumption C(ii).

Theorem 3.3. *Suppose that Assumption A holds, and let $f \in F$, $x \hookrightarrow B \in \mathcal{B}(S)$, and $\delta \geq 0$. Then the following statements hold.*

(a) *$U_\delta^B(x, f)$ is the minimum nonnegative solution to the equation*

$$\delta u(x) \geq g(x, f(x)) + \int_{S-B} u(y) q(\mathrm{d}y \mid x, f(x)) \quad \textit{for all } x \notin B. \tag{3.7}$$

(b) *If, in addition, Assumptions B and C(i) hold, there exists a nonnegative measurable function u on S that satisfies*

$$c(x, a) + \int_{S-\{x_0\}} u(y) q(\mathrm{d}y \mid x, a) \leq 0 \quad \textit{for all } x \neq x_0 \textit{ and } a \in A(x),$$

with x_0 as in Assumption C(i), and $h_{\alpha_n}(x) \leq u(x)$ for all $n \geq 1$ and $x \in S$.

(c) *If, in addition, there exist some constant $\beta > 0$ and $B \in \mathcal{B}(S)$ such that $q(B \mid x, a) \geq \beta$ for all $a \in A(x)$ and $x \notin B$, then $\mathrm{E}_x^f[\tau_B^f] \leq 1/\beta$ for all $x \notin B$ and $f \in F$.*

(d) *If the conditions in (b) and (c) (with $B = \{x_0\}$) hold, then Assumption C is satisfied.*

(e) *If the conditions in (b) hold, and $\mathrm{E}_x^f[\tau_{x_0}^f]$ is bounded in $f \in F$ and $x \neq x_0$ (with x_0 as in Assumption C(i)), then Assumption C holds.*

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Proof. See Section 5.

Remark 3.6. Theorem 3.3 is *new* and can be applied to the case of Polish spaces; see Examples 4.1–4.3 below. In particular, the condition in Theorem 3.3(d) does not require any monotonicity assumption, and, thus, it is different from the monotonicity condition imposed on the transition rates in [9], [10], and [19, pp. 426–427], which further require the *additional* monotonicity assumption imposed on the rewards; see, e.g. Lemma 3.3 and Example 5.1 of [10], Assumption C of [9], and Theorems 8.11.3 and 8.11.4 of [19].

To prove the existence of average cost optimal policies, we need some facts and concepts. For each $f \in F$, since the transition function $p_f(s, x, t, D)$ is homogeneous, $p_f(s, x, s+t, D)$ is independent of $s \geq 0$. Thus, we may write $p_f(x, t, D) := p_f(0, x, 0+t, D)$ for all $x \in S$, $D \in \mathcal{B}(S)$, and $t \geq 0$. Define the *t-horizon expectation total cost* under policy f by

$$J_f(x, t) := \int_0^t \mathrm{E}_x^f\, c(x(s), a(s))\,\mathrm{d}s = \int_0^t \int_S c(y, f(y)) p_f(x, s, \mathrm{d}y)\,\mathrm{d}s \tag{3.8}$$

for all $x \in S$ and $t \geq 0$. We then have the following key result.

Theorem 3.4. *For any fixed $f \in F$, let $q(x, f) := -q(\{x\} \mid x, f(x)) \geq 0$, $q(\cdot \mid x, f) := q(\cdot \mid x, f(x))$, and $c(x, f) := c(x, f(x))$ for all $x \in S$. Suppose that Assumption A holds, then*

(a) *$J_f(x, t)$ is the minimum nonnegative solution to the inequality*

$$u(x, t) \geq c(x, f)t\mathrm{e}^{-q(x,f)t} + \int_0^t \mathrm{e}^{-q(x,f)s}\biggl[q(x, f)c(x, f)s + \int_{S-\{x\}} u(y, t-s)q(\mathrm{d}y \mid x, f)\biggr]\mathrm{d}s \tag{3.9}$$

for all $x \in S$ and $t \geq 0$, satisfying (3.9) with equality;

(b) *if there exist a constant $\rho \geq 0$ and a real-valued measurable function u on S bounded below, such that*

$$\rho + u(x)q(x, f) \geq c(x, f) + \int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x, f) \qquad \textit{for all } x \in S, \tag{3.10}$$

then $\rho \geq J(0, x, f)$ for all $x \in S$.

Proof. See Section 5.

Remark 3.7. (a) Theorem 3.4 needs only Assumption A and allows unbounded costs and transition rates, whereas similar results in [10] and [24] require some *additional* conditions.

(b) We call the method used to prove Theorem 3.4 a *minimum nonnegative solution approach*, which is *new* and rather *different* from those in [10] and [24] for continuous-time MDPs and those in [13] and [19] for the discrete-time case. Moreover, results of a transition function constructed from transition rates under the measurability condition play key roles in our arguments.

The following theorem establishes the existence of average cost optimal policies.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Theorem 3.5. *Under Assumptions A, B, and C, the following assertions hold.*

(a) *There exist a nonnegative constant ρ^*, a stationary policy $f^* \in F$, and a real-valued measurable function h^* on S satisfying the average cost optimality inequality*

$$\begin{aligned}\rho^* &\geq c(x, f^*(x)) + \int_S h^*(y) q(\mathrm{d}y \mid x, f^*(x)) \\ &\geq \inf_{a \in A(x)} \biggl\{ c(x, a) + \int_S h^*(y) q(\mathrm{d}y \mid x, a) \biggr\} \quad \textit{for all } x \in S. \qquad (3.11)\end{aligned}$$

(b) *Any policy $f \in F$ realizing the minimum in (3.11) is average cost optimal. Therefore, f^* in (a) is an average cost optimal stationary policy, and, moreover, ρ^* is the average cost optimal value.*

Proof. See Section 5.

Remark 3.8. The conditions and results of Theorem 3.5 are the *continuous-time versions* of those for discrete-time MDPs; see, e.g. [13]. Note that the exponentially ergodic condition in [25] and the drift condition imposed on the relative difference in [10] have been dropped. Moreover, both the absolute integrability condition and the continuity assumption (see Remark 3.4(b) above) in [3], [6], [7], [10], [24], and [25] have been removed. Since the proof of Theorem 3.5 is based on Theorem 3.4, we also call it an *average cost minimum nonnegative solution approach*, which is rather different from those in [10], [13, Theorem 5.4.3], [19, Theorem 8.10.7], and [24].

4. Examples

In this section we illustrate our conditions and show applications of our results with Examples 4.1–4.3.

Example 4.1. Consider a management problem of a water reservoir with finite capacity $C(> 0)$. The state variable $x(t)$ denotes the amount of water in the reservoir at time $t \geq 0$. Water in the reservoir can be replenished and used at positive constant rates λ and μ, respectively. The quantity of water available to replenish the reservoir depends on the amount of rainfall, and the largest amount of 'replenishment' water is assumed to be $M(< C)$. Moreover, the quantity of water in the reservoir is divided into three zones: the inactive zone $S_1 := [0, \theta]$, the active zone $S_2 := (\theta, C - M]$, and the flood control zone $S_3 := (C - M, C]$, where the constant θ $(0 < \theta < C - M)$ denotes the lowest quantity of water that is not normally used. Suppose that the amount of water in the reservoir decreases to 0 at a constant rate $\beta > 0$ due to some risk, and that the quantity of water to be used can be controlled by a decision maker. Assume that the amount of water in the reservoir is x and that the decision maker plans for a quantity, a, of water to be used. Then the decrease and increase in the quantity of water in the reservoir are measured by Lebesgue's measure on $[x - a, x]$ and $[x, \min\{x + M, C\}]$, respectively, and the cost incurred per unit time is denoted by $c(x, a)$. We then obtain a model for continuous-time MDPs as follows. The state space herein is $S := S_1 \cup S_2 \cup S_3 = [0, C]$, the sets of feasible actions $A(x)$ at $x \in S$ are $A(x) := \{0\}$ for $x \in [0, \theta]$, $A(x) := [0, x - \theta]$ for $x \in (\theta, C - M]$, and $A(x) := \{x - C + M\}$ for $x \in (C - M, C]$. The transition rates $q(\cdot \mid x, a)$ are given as follows. For each $D \in \mathcal{B}(S)$,

$$q(D \mid x, a) := \beta\delta_0(D) + \lambda m_L(D \cap [x, x + M]) - (\beta + \lambda M)\delta_x(D) \quad \text{for } x \in S_1 \text{ and } a \in A(x);$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

when $x \in S_2$ and $a \in A(x)$, we have

$$q(D \mid x, a) := \beta\delta_0(D) + \mu m_L(D \cap [x-a, x]) + \lambda m_L(D \cap [x, x+M]) - (\beta + \mu a + \lambda M)\delta_x(D);$$

moreover, for $x \in S_3$ and $a \in A(x)$,

$$\begin{aligned} q(D \mid x, a) := {}& \beta\delta_0(D) + \mu m_L(D \cap [C-M, x]) + \lambda m_L(D \cap [x, C]) \\ & - [\beta + \mu(x - C + M) + \lambda(C - x)]\delta_x(D). \end{aligned}$$

(Indeed, the $q(\cdot \mid x, a)$ defined above are transition rates because they satisfy properties (T1)–(T3).)

For the management problem of the water reservoir, we aim to find conditions that ensure the existence of discounted cost and average cost optimal stationary policies. To this end, we consider the following hypotheses.

(H1) Assume that $c(x, \cdot)$ is bounded below and l.s.c. in $a \in A(x)$ for each fixed $x \in S$.

(H2) $\beta > \frac{1}{2}\lambda M^2$, and $\sup_{a \in A(x)} |c(x, a)| \le L_1(x + 1)$ for all $x \in S$, with some constant $L_1 > 0$.

Then we obtain the following result.

Proposition 4.1. *Under (H1), the following statements hold.*

(a) *There exists a discounted cost optimal stationary policy for Example 4.1.*

(b) *If, in addition, (H2) holds, then Example 4.1 satisfies Assumptions A, B, and C. Hence, (by Theorem 3.5) there exists an average cost optimal stationary policy.*

Proof. (a) The proof follows from Theorem 3.2; however, in order to appeal to this theorem, we need to verify Assumptions A and B. Since (ii)–(iv) of Assumption B follow from the definition of $q(\cdot \mid x, a)$ above, (H1), and the description of Example 4.1, we need only verify Assumptions A and B(i). Let $w(x) := x + 1$ for all $x \in S$. Then, by the definition of $q(\cdot \mid x, a)$ above and a straightforward calculation, we have

$$\begin{aligned} q^*(x) &\le (\mu + \lambda M + \beta)(x + 1) \quad \text{for all } x \in S, \\ \int_S w(y) q(\mathrm{d}y \mid x, 0) &= \beta w(0) + \lambda \int_x^{x+M} (y + 1)\,\mathrm{d}y - (\beta + \lambda M) w(x) \\ &\le -\beta(x + 1) + \beta + \tfrac{1}{2}\lambda M^2 \quad \text{for } x \in S_1, \\ \int_S w(y) q(\mathrm{d}y \mid x, a) &= \beta w(0) + \mu \int_{x-a}^{x} (y + 1)\,\mathrm{d}y + \lambda \int_x^{x+M} (y + 1)\,\mathrm{d}y \\ &\quad - (\beta + \mu a + \lambda M) w(x) \\ &= \beta - \tfrac{1}{2}\mu a^2 + \tfrac{1}{2}\lambda M^2 - \beta(x + 1) \\ &\le -\beta(x + 1) + \beta + \tfrac{1}{2}\lambda M^2 \quad \text{for } x \in S_2,\ a \in [0, x - \theta], \\ \int_S w(y) q(\mathrm{d}y \mid x, a) &= \beta w(0) + \mu \int_{C-M}^{x} (y + 1)\,\mathrm{d}y + \lambda \int_x^{C} (y + 1)\,\mathrm{d}y \\ &\quad - [\beta + \mu(x - C + M) + \lambda(C - x)] w(x) \end{aligned}$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

$$\begin{aligned}&= -\beta x - \tfrac{1}{2}\mu(x - C + M)^2 + \tfrac{1}{2}\lambda(C+1)(C-x)\\ &\quad - \tfrac{1}{2}\lambda(C-x)(x+1)\\ &\le -\beta(x+1) + \beta + \tfrac{1}{2}\lambda M^2 \quad \text{for } x \in S_3,\ a = x - C + M.\end{aligned}$$

Then Assumptions A and B(i) follow immediately from these inequalities with $c_0 := -\beta$, $b_0 := \beta + \frac{1}{2}\lambda M^2$, and $M_0 := \mu + \lambda M + \beta$.

(b) Since Assumptions A and B are verified, it remains to verify Assumption C. By Theorem 3.2, for each $0 < \alpha < 1$, there exists an α-discounted cost optimal stationary policy $f_\alpha \in F$, for which $V_\alpha(x, f_\alpha) = V_\alpha^*(x)$ for all $x \in S$. Let $x_0 = 0$. Then, by (a), (H2), and Theorem 3.3(a) of [7], we have

$$\hat{L}_1 \le \alpha V_\alpha{}^*(0) = \alpha V_\alpha(0, f_\alpha) \le \frac{L_1 b_0}{\alpha + \beta} + \frac{\alpha L_1}{\alpha + \beta} < \infty, \tag{4.1}$$

where $\hat{L}_1$ is a constant. This implies Assumption C(i).

To verify Assumption C(ii), let $w_1(x) := L_1'(x+1)$ with $L_1' := L_1/(\beta - \lambda M^2/2) > 0$. For $x \neq 0$ and $a \in A(x)$, by (H2), we have, for $x \in S_1$ and $a = 0$,

$$\begin{aligned}L_1(x+1) + \int_{S-\{0\}} w_1(y) q(\mathrm{d}y \mid x, 0) &= L_1(x+1) + L_1'\big(\tfrac{1}{2}\lambda M^2 - \beta(x+1)\big)\\ &\le L_1(x+1) - L_1'\big(\beta - \tfrac{1}{2}\lambda M^2\big)(x+1)\\ &\le 0;\end{aligned} \tag{4.2}$$

for $x \in S_2$ and $a \in [0, x - \theta]$,

$$\begin{aligned}L_1(x+1) + \int_{S-\{0\}} w_1(y) q(\mathrm{d}y \mid x, a) &= L_1(x+1) + L_1'\big(-\tfrac{1}{2}\mu a^2 + \tfrac{1}{2}\lambda M^2 - \beta(x+1)\big)\\ &\le L_1(x+1) - L_1'\big(\beta - \tfrac{1}{2}\lambda M^2\big)(x+1)\\ &\le 0;\end{aligned} \tag{4.3}$$

and, for $x \in S_3$ and $a = c - x + M$,

$$\begin{aligned}&L_1(x+1) + \int_{S-\{0\}} w_1(y) q(\mathrm{d}y \mid x, a)\\ &\quad = L_1(x+1) - L_1'\big(\tfrac{1}{2}\mu(x - C + M)^2 + \beta(x+1) - \tfrac{1}{2}\lambda(C+1)(C-x)\\ &\qquad\qquad + \tfrac{1}{2}\lambda(C-x)(x+1)\big)\\ &\quad \le L_1(x+1) - L_1'\big(\tfrac{1}{2}\mu(x - C + M)^2 + \beta(x+1) - \tfrac{1}{2}\lambda M^2\big)\\ &\quad \le 0.\end{aligned} \tag{4.4}$$

Moreover, since $q(\{0\} \mid x, a) \ge \beta$ for all $x \neq 0$ and $a \in A(x)$, by (4.1)–(4.4) and Theorem 3.3(d), we can see that Assumption C(ii) is satisfied.

Remark 4.1. It should be mentioned that the state space in Example 4.1 is *not* denumerable, any monotonicity assumptions imposed on both the rewards and transition rates in [9], [10], and [11] are *not* required.

Next we illustrate the applications of our results with another two examples with unbounded transition rates.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Example 4.2. Consider a control problem of hypertension, in which we are interested in how to control the average time when the blood pressure in a body is 'stable'. As is well known, by the normalization technique we can describe the blood pressure with the standard normal distribution $N(0, 1)$, and, thus, the normalized quantity of blood pressure may take values in $S := (-\infty, \infty)$. When the current amount of blood pressure is at $x \in S$ and a controlled amount a is given, we suppose that the holding time of the 'stable' blood pressure has an exponential distribution with parameter $(\gamma|x| + a)^{-1}$, where γ is a fixed constant. Thus, the rate of change of blood pressure is given as

$$q(D \mid x, a) := \beta\delta_0(D) + (\gamma|x| + a)\int_{D-\{x\}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2}\,\mathrm{d}y - (\gamma|x| + \beta + a)\delta_x(D) \quad (4.5)$$

for each $D \in \mathcal{B}(S)$, where the constant β represents the rate at which a risk may happen.

We denote by $c(x, a)$ the cost of taking control a when the current amount of blood pressure is at $x \in S$, and regard a as an action, which takes values in $[0, \kappa]$, with some constant $\kappa > 0$. Then, the model of continuous-time MDPs is specified with S, $q(\cdot \mid x, a)$, and $c(x, a)$ as above, and $A = A(x) := [0, \kappa]$ for all $x \in S$.

Our goal is to find conditions that ensure the existence of discounted cost and average cost optimal stationary policies. To this end, we need the following hypotheses.

(H3) Assume that $c(x, \cdot)$ is bounded below and l.s.c. in $a \in A(x)$ for each fixed $x \in S$.

(H4) $\sup_{a\in A(x)} |c(x, a)| \le L_2(x^2 + 1)$ for all $x \in S$, with some constant $L_2 > 0$.

Then we obtain the following result.

Proposition 4.2. *Under (H3), the following statements hold.*

(a) *If $\alpha + \beta > \frac{1}{2}\gamma$ then there exists a discounted cost optimal stationary policy for Example 4.2.*

(b) *If, in addition, (H4) holds and $\beta > \kappa + \frac{1}{2}\gamma$, then Example 4.2 verifies Assumptions A, B, and C. Hence, (by Theorem 3.5) there exists an average cost optimal stationary policy.*

Proof. Since the proof of this proposition is similar as that of Proposition 4.1, we only describe the skeleton of the proof, and omit the details.

(a) The proof follows from Theorem 3.2 and, therefore, it suffices to verify Assumptions A and B. Since (ii)–(iv) of Assumption B follow from (4.5) and the description of Example 4.2, it remains to verify Assumptions A and B(i). Let $w(x) := x^2 + 1$ for all $x \in S$. Then, by (4.5) and a straightforward calculation, we can see that Assumptions A and B(i) are indeed satisfied with $c_0 := -\beta + \frac{1}{2}\gamma$, $b_0 := \kappa + \beta$, and $M_0 := \beta + \kappa + \frac{1}{2}\gamma$.

(b) Since Assumptions A and B are verified above, we now need to verify only Assumption C. Take $x_0 = 0$. Then, by (a), (H3), and Theorem 3.3(a) of [7], we see that Assumption C(i) is also true. To verify Assumption C(ii), let

$$w_2(x) := \frac{L_2}{\beta - (\kappa + \gamma/2)}(x^2 + 1).$$

Then, under (H4) and $\beta > \kappa + \frac{1}{2}\gamma$, a direct calculation shows that

$$L_2(x^2 + 1) + \int_{S-\{0\}} w_2(y)q(\mathrm{d}y \mid x, a) \le 0. \quad (4.6)$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Moreover, it follows from (4.5) that $q(\{0\} \mid x, a) = \beta > 0$ for all $x \neq 0$ and $a \in A(x)$. Therefore, by (4.6) and Theorem 3.3(d), Assumption C(ii) is thus satisfied, and the proof is complete.

Example 4.3. (*A controlled birth-and-death process.*) Consider a birth-and-death process with controlled birth and death parameters, in which the state variable denotes a system's size at any time $t \geq 0$. There are 'natural' birth and death rates represented by positive constants λ and μ, respectively, and additional birth and death parameters (a_1 and a_2), which are controlled by a decision maker. When the state of the process is $x \in S := \{0, 1, \dots\}$, the decision maker takes an action $a := (a_1, a_2)$ from a given set $A(x)$, which may admit ($a_1 \geq 0$) or expel ($a_1 \leq 0$) the birth rate, and also increase ($a_2 \geq 0$) or decrease ($a_2 \leq 0$) the death rate. Moreover, this action a incurs a cost at rate $c(x, a)$.

We now formulate this system as a model of continuous-time MDPs. The corresponding function of transition rates $q(y \mid x, a)$ is given as follows. For $x = 0$ and $a = (a_1, a_2) \in A(0)$,

$$q(1 \mid 0, a) = -q(0 \mid 0, a) := a_1,$$

and, for each $x \geq 1$ and $a = (a_1, a_2) \in A(x)$,

$$q(y \mid x, a) = \begin{cases} \lambda x^2 + a_1 & \text{if } y = x + 1, \\ -(\mu + \lambda)x^2 - a_1 - a_2 & \text{if } y = x, \\ \mu x^2 + a_2 & \text{if } y = x - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.7}$$

The cost function $c(x, a)$ is defined as

$$c(x, a) = px^2 + h(x, a) \quad \text{for all } (x, a) \in K, \tag{4.8}$$

with some fixed constant $p > 0$ and $h(\cdot, \cdot)$ a Borel-measurable function on K.

To ensure the existence of discounted cost and average cost optimal stationary policies, we consider the following hypotheses.

(H5) $A(0) := [0, \frac{1}{4}\lambda]$, and assume that $A(x)$ is a compact subset of $[-\lambda, \frac{1}{4}\lambda] \times [-\frac{1}{4}\mu, \mu]$ for each $x \geq 1$.

(H6) The function $h(x, a)$ is bounded below and l.s.c. in $a \in A(x)$ for each fixed $x \in S$.

(H7) $\mu \geq \frac{3}{2}\lambda$.

(H8) $\sup_{a \in A(x)} |h(x, a)| \leq L_3(x^2 + 1)$ for all $x \geq 1$ with some fixed constant $L_3 > 0$.

Under these hypotheses, we obtain the following result.

Proposition 4.3. *Under (H5) and (H6), the following assertions hold.*

(a) *If $\mu > \lambda$ then there exists a discounted cost optimal stationary policy for Example 4.3.*

(b) *If, in addition, (H7) and (H8) hold, then Example 4.3 satisfies Assumptions A, B, and C. Therefore, (by Theorem 3.5) there exists an average cost optimal stationary policy.*

Proof. (a) It follows from (H6) and (4.8) that the cost function is bounded below. Moreover, since (ii)–(iv) of Assumption B follow from (4.7)–(4.8) and (H5)–(H6), we need to verify only

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Assumptions A and B(i). Let $w(x) := x^2 + 1$ for all $x \in S$. Then, by (4.7) and (H5), we can derive

$$q^*(x) \le \lambda(x^2 + 1) + \mu(x^2 + 1) = M_0 w(x) \quad \text{for all } x \in S, \tag{4.9}$$

$$\sum_{y \in S} w(y) q(y \mid x, a) = 2(\lambda - \mu)x^3 + (\lambda + \mu)x^2 + 2(a_1 - a_2)x + a_2 + a_1. \tag{4.10}$$

Moreover, by $\mu > \lambda$ and (4.10), we have, for each $a \in A(x)$ with $x \ge 1$,

$$\sum_{y \in S} w(y) q(y \mid x, a) \le (\lambda - \mu)(x^2 + 1) + b_0 \tag{4.11}$$

for some constant $b_0 \ge 0$. Moreover, for each $a \in A(0)$, we have

$$\sum_{y \in S} w(y) q(y \mid 0, a) = a_1 \le (\lambda - \mu) w(0) + b_0. \tag{4.12}$$

Therefore, Assumptions A and B(i) follow from (4.9)–(4.12) with $c_0 := \lambda - \mu$, b_0 as above, and $M_0 := \lambda + \mu > 0$. Thus, (a) follows from Theorem 3.2.

(b) Since we have verified Assumptions A and B in part (a), it follows from Theorem 3.2 that, for each $0 < \alpha < 1$, there exists an α-discounted cost optimal stationary policy $f_\alpha \in F$ for which $V_\alpha(x, f_\alpha) = V_\alpha^*(x)$ for all $x \in S$. To verify Assumption C(i), we take $x_0 = 0$. By (4.8), (H8), and Theorem 3.3(a) of [7], we obtain

$$\hat{L} \le \alpha V_\alpha{}^*(0) = \alpha V_\alpha(0, f_\alpha) \le \frac{(p + L_3)b_0}{\alpha - \lambda + \mu} + \frac{\alpha(p + L_3)}{\alpha - \lambda + \mu} < \infty,$$

since $\hat{L} \le c(x, a) \le (p + L_3)(x^2 + 1)$ for all $x \in S$, with some constant $\hat{L}$ (by (H6) and (H8)). Hence, Assumption C(i) is verified.

To verify Assumption C(ii), let $\{\alpha_m\}$ be a decreasing sequence that tends to 0, and let $f_{\alpha_m} \in F$ be the α_m-discounted cost optimal stationary policies. Then

$$h_{\alpha_m}(x) = V_{\alpha_m}(x, f_{\alpha_m}) - V_{\alpha_m}(0, f_{\alpha_m}) \quad \text{for all } x \in S \text{ and } m \ge 1.$$

Let

$$w_3(x) := \frac{2(p + L_3)}{13\mu - 19\lambda}(x^2 + 8x + 9).$$

By (H5), (H7), and (H8), we have, for $x = 1$ and $a = (a_1, a_2) \in A(1)$,

$$\begin{aligned} &2(p + L_3) + \sum_{y \ge 1} w_3(y) q(y \mid 1, a) \\ &\quad = 2(p + L_3) \\ &\qquad + \frac{2(p + L_3)}{13\mu - 19\lambda}[(-\mu - \lambda - a_1 - a_2)(1 + 8 + 9) + (\lambda + a_1)(4 + 16 + 9)] \\ &\quad \le 2(p + L_3) + \frac{2(p + L_3)}{13\mu - 19\lambda}\left(-\frac{27}{2}\mu + \frac{55}{4}\lambda\right) \\ &\quad < 0, \end{aligned}$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

and, for $x \geq 2$ and $a = (a_1, a_2) \in A(x)$,

$$\begin{aligned}
&(p + L_3)(x^2 + 1) + \sum_{y \geq 1} w_3(y) q(y \mid x, a) \\
&\quad = (p + L_3)(x^2 + 1) \\
&\qquad + \frac{2(p + L_3)}{13\mu - 19\lambda}[2(\lambda - \mu)x^3 + (9\lambda - 7\mu)x^2 + 2(a_1 - a_2)x + 9a_1 - 7a_2] \\
&\quad \leq (p + L_3)(x^2 + 1) \\
&\qquad + \frac{2(p + L_3)}{13\mu - 19\lambda}\biggl[16(\lambda - \mu) + (9\lambda - 7\mu)x^2 + \frac{1}{2}(\lambda + \mu)x + \frac{9}{4}\lambda + \frac{7}{4}\mu\biggr] \\
&\quad \leq (p + L_3)(x^2 + 1) + \frac{2(p + L_3)}{13\mu - 19\lambda}\biggl[\biggl(\frac{19}{2}\lambda - \frac{13}{2}\mu\biggr)(x^2 + 1) + \frac{35}{4}\lambda - \frac{31}{4}\mu\biggr] \\
&\quad < 0.
\end{aligned}$$

The two inequalities imply that

$$(p + L_3)(x^2 + 1) + \sum_{y \geq 1} w_2(y) q(y \mid x, a) \leq 0 \quad \text{for all } x \geq 1,\ (x, a) \in K.$$

Hence, by Theorem 3.3(b) (with $x_0 = \{0\}$ and $u = w_3$), we obtain

$$h_{\alpha_m}(x) \leq w_3(x) = \frac{2(p + L_3)}{13\mu - 19\lambda}(x^2 + 8x + 9) \quad \text{for all } x \in S.$$

We now estimate $\mathrm{E}_x^f\bigl[\tau_0^f\bigr]$ (for $f \in F$ and $x \neq 0$). Denote by

$$R_f := \sum_{x=1}^{\infty}\biggl(\frac{1}{\mu_f(x)} + \sum_{k=0}^{\infty}\frac{\lambda_f(x)\lambda_f(x+1)\cdots\lambda_f(x+k)}{\mu_f(x)\mu_f(x+1)\cdots\mu_f(x+k)\mu_f(x+k+1)}\biggr)$$

the mean time of first reaching 0 from state '∞' under policy f (see [21, pp. 146–148] for details), with $\lambda_f(x) := \lambda x^2 + f_1(x)$, $\mu_f(x) := \mu x^2 + f_2(x)$, and $f(x) =: (f_1(x), f_2(x)) \in A(x)$. Then by (H5) we have $\mu_f(x) \geq \frac{3}{4}\mu x^2$ and $\lambda_f(x) \leq \lambda(x+1)^2$ for all $x \geq 1$. Therefore, from (H5) and (H6), we obtain

$$\begin{aligned}
R_f &\leq \sum_{x=1}^{\infty}\biggl(\frac{1}{3\mu x^2/4} + \sum_{k=0}^{\infty}\frac{[\lambda(x+1)^2][\lambda(x+2)^2]\cdots[\lambda(x+k+1)^2]}{[3\mu x^2/4][3\mu(x+1)^2/4]\cdots[3\mu(x+k+1)^2/4]}\biggr) \\
&= \sum_{x=1}^{\infty}\frac{4}{3\mu x^2}\biggl(1 + \sum_{k=0}^{\infty}\biggl(\frac{4\lambda}{3\mu}\biggr)^{k+1}\biggr) \\
&=: M^* \\
&< \infty \quad \text{for all } f \in F,
\end{aligned}$$

and so (by Theorem 2 of [21, Chapter 5, p. 149]), we have

$$\mathrm{E}_x^f[\tau_0^f] \leq R_f \leq M^* \quad \text{for all } x \geq 1.$$

This together with Theorem 3.3(e) verifies Assumption C(ii), and thus completes the proof.

Remark 4.2. It is worth noting that the conditions in Examples 4.1–4.3, under which the existence of an average cost optimal stationary policy is ensured, are *different* from those in [6], [9], [10], and [11]. In particular, we do not need the *monotonicity* assumptions imposed on both transition rates and rewards in [9], [10], and [11].

5. Proofs of Theorems 3.1–3.5

Note that the function $V_\alpha(s, x, \phi)$ defined in (2.9) is independent of time s, and, since $p_\phi(s, x, t, D) = p_\phi(0, x, t-s, D)$ for $\phi \in \Pi_s$, we can write $V_\alpha(s, x, \phi) = V_\alpha(0, x, \phi) =: V_\alpha(x, \phi)$ for each $\phi \in \Pi_s$.

5.1. Proof of Theorem 3.1

To prove Theorem 3.1, we need the following result.

Lemma 5.1. *For each $\phi \in \Pi_s$, $x \in S$, and $D \in \mathcal{B}(S)$, let $q(x, \phi) := -q_\phi(\{x\} \mid x, \phi)$ and $q(D \mid x, \phi) := q_\phi(D \mid x, \phi)$. Then, under Assumption A, the following statements hold.*

(a) *$V_\alpha(\phi)$ is the minimum nonnegative solution to the equation*

$$u(x) = \frac{c(x, \phi)}{\alpha + q(x, \phi)} + \frac{1}{\alpha + q(x, \phi)} \int_{S-\{x\}} u(y) q(\mathrm{d}y \mid x, \phi) \quad \textit{for all } x \in S. \tag{5.1}$$

(b) *If a nonnegative measurable function u on S satisfies*

$$u(x) \geq \frac{c(x, \phi)}{\alpha + q(x, \phi)} + \frac{1}{\alpha + q(x, \phi)} \int_{S-\{x\}} u(y) q(\mathrm{d}y \mid x, \phi) \quad \textit{for all } x \in S,$$

then $u \geq V_\alpha(\phi)$.

Proof. (a) Choose an arbitrary $\phi \in \Pi_s$. For each $x \in S$, $D \in \mathcal{B}(S)$, and $n \geq 0$, define

$$\varphi_\phi^{(n)}(x, D) := \begin{cases} \dfrac{\delta_x(D)}{\alpha + q(x, \phi)} & \text{for } n = 0, \\ \dfrac{1}{\alpha + q(x, \phi)} \left[\delta_x(D) + \displaystyle\int_{S-\{x\}} \varphi_\phi^{(n-1)}(y, D) q(\mathrm{d}y \mid x, \phi) \right] & \text{for } n \geq 1. \end{cases} \tag{5.2}$$

Note that $p_\phi(s, x, t, D)$ is homogeneous and $\varphi_\phi^{(n)}(x, D)$ is nondecreasing in $n \geq 0$. In view of the theory of continuous-time Markov processes (see, e.g. Theorem 2.21 of [2]), we obtain

$$\int_0^\infty \mathrm{e}^{-\alpha t} p_\phi(0, x, t, D) \, \mathrm{d}t = \lim_{n \to \infty} \varphi_\phi^{(n)}(x, D).$$

Since $c(x, a) \geq 0$, this equality together with (2.8)–(2.9), the monotone convergence theorem, and Fubini's theorem gives

$$V_\alpha(x, \phi) = \int_S c(y, \phi) \Big[\lim_{n \to \infty} \varphi_\phi^{(n)}(x, \mathrm{d}y) \Big] = \lim_{n \to \infty} \int_S c(y, \phi) \varphi_\phi^{(n)}(x, \mathrm{d}y). \tag{5.3}$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

For any $n \geq 1$, from (5.2) we can derive

$$\begin{aligned}\int_S c(y,\phi)\varphi_\phi^{(n+1)}(x,\,\mathrm{d}y) &\\ &= \int_S \frac{c(y,\phi)}{\alpha+q(x,\phi)}\biggl(\delta_x(\mathrm{d}y)+\int_{S-\{x\}}\varphi_\phi^{(n)}(z,\,\mathrm{d}y)q(\mathrm{d}z\mid x,\phi)\biggr)\\ &= \frac{1}{\alpha+q(x,\phi)}\biggl(c(x,\phi)+\int_{S-\{x\}}\int_S c(y,\phi)\varphi_\phi^{(n)}(z,\,\mathrm{d}y)q(\mathrm{d}z\mid x,\phi)\biggr). \qquad (5.4)\end{aligned}$$

Letting $n\to\infty$ in (5.4), the monotone convergence theorem and (5.3) give

$$V_\alpha(x,\phi)=\frac{c(x,\phi)}{\alpha+q(x,\phi)}+\frac{1}{\alpha+q(x,\phi)}\int_{S-\{x\}}V_\alpha(z,\phi)q(\mathrm{d}z\mid x,\phi),$$

which implies that $V_\alpha(\phi)$ satisfies (5.1).

Let u be a nonnegative solution to (5.1). To prove that $u\geq V_\alpha(\phi)$, in view of (5.3) it suffices to show that

$$\int_S c(y,\phi)\varphi_\phi^{(n)}(x,\,\mathrm{d}y)\leq u(x)\quad\text{for all } x\in S \text{ and } n\geq 0. \qquad (5.5)$$

Obviously, it is valid for $n=0$. In fact, since $u\geq 0$ and $q(D\mid x,\phi)\geq 0$ for $x\notin D$, by (5.1)–(5.2), we have

$$u(x)\geq\frac{c(x,\phi)}{\alpha+q(x,\phi)}=\int_S c(y,\phi)\varphi_\phi^{(0)}(x,\,\mathrm{d}y). \qquad (5.6)$$

Suppose now that (5.5) holds for some $n\geq 0$. Then taking (5.4) and the induction hypothesis into account, we obtain

$$\begin{aligned}\int_S c(y,\phi)\varphi_\phi^{(n+1)}(x,\,\mathrm{d}y) &\\ &= \frac{c(x,\phi)}{\alpha+q(x,\phi)}+\frac{1}{\alpha+q(x,\phi)}\int_{S-\{x\}}\int_S c(y,\phi)\varphi_\phi^{(n)}(z,\,\mathrm{d}y)q(\mathrm{d}z\mid x,\phi)\\ &\leq \frac{c(x,\phi)}{\alpha+q(x,\phi)}+\frac{1}{\alpha+q(x,\phi)}\int_{S-\{x\}}u(z)q(\mathrm{d}z\mid x,\phi)\\ &= u(x).\end{aligned}$$

Hence, (5.5) is valid for all $n\geq 0$, which proves (a).

(b) Under the assumption in assertion (b), there exists a nonnegative measurable function v on S such that

$$u(x)=\frac{c(x,\phi)+v(x)}{\alpha+q(x,\phi)}+\frac{1}{\alpha+q(x,\phi)}\int_{S-\{x\}}u(y)q(\mathrm{d}y\mid x,\phi)\quad\text{for all } x\in S.$$

Thus, in view of (a) and (5.3) (with $c(x,\phi)+v(x)$ in lieu of $c(x,\phi)$), we have

$$u(x)\geq\lim_{n\to\infty}\int_S(c(y,\phi)+v(y))\varphi_\phi^{(n)}(x,\,\mathrm{d}y)\geq\lim_{n\to\infty}\int_S c(y,\phi)\varphi_\phi^{(n)}(x,\,\mathrm{d}y)=V_\alpha(x,\phi)$$

for all $x\in S$, which yields (b). This completes the proof.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Proof of Theorem 3.1. (a) Fix a policy $\pi \in \Pi$. For each $x \in S$ and $D \in \mathcal{B}(S)$, by the Kolmogorov forward equation in [23], we obtain

$$p_\pi(s, x, t, D) = \delta_x(D) + \int_s^t \int_S p_\pi(s, x, \tau, \mathrm{d}y) q(D \mid y, \pi_\tau)\,\mathrm{d}\tau \quad \text{for all } t \geq s.$$

Moreover, since $\alpha > c_0$ and $\int_S w(y)\nu(\mathrm{d}y) < \infty$, by Assumption A and Theorem 3.1 in [7], we have

$$\left| \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \int_S \int_s^t \int_D \left[\int_A q(\{y\} \mid y, a)\pi_\tau(\mathrm{d}a \mid y) \right] p_\pi(s, x, \tau, \mathrm{d}y)\,\mathrm{d}\tau\, \nu(\mathrm{d}x)\,\mathrm{d}t \right| < \infty.$$

Thus, by (2.6) and (3.1), we can derive

$$\begin{aligned}
\hat{\mu}_S^\pi(D) &= \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \int_S p_\pi(s, x, t, D)\nu(\mathrm{d}x)\,\mathrm{d}t \\
&= \frac{\nu(D)}{\alpha} + \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \int_S \int_s^t \int_S q(D \mid y, \pi_\tau) p_\pi(s, x, \tau, \mathrm{d}y)\,\mathrm{d}\tau\, \nu(\mathrm{d}x)\,\mathrm{d}t \\
&= \frac{\nu(D)}{\alpha} + \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \int_S \int_s^t \int_{S-D} \left(\int_A q(D \mid y, a)\pi_\tau(\mathrm{d}a \mid y) \right) \\
&\qquad\qquad \times p_\pi(s, x, \tau, \mathrm{d}y)\,\mathrm{d}\tau\, \nu(\mathrm{d}x)\,\mathrm{d}t \\
&\quad + \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \int_S \int_s^t \int_D \left(\int_A q(D - \{y\} \mid y, a)\pi_\tau(\mathrm{d}a \mid y) \right) \\
&\qquad\qquad \times p_\pi(s, x, \tau, \mathrm{d}y)\,\mathrm{d}\tau\, \nu(\mathrm{d}x)\,\mathrm{d}t \\
&\quad + \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \int_S \int_s^t \int_D \left(\int_A q(\{y\} \mid y, a)\pi_\tau(\mathrm{d}a \mid y) \right) \\
&\qquad\qquad \times p_\pi(s, x, \tau, \mathrm{d}y)\,\mathrm{d}\tau\, \nu(\mathrm{d}x)\,\mathrm{d}t \\
&= \frac{\nu(D)}{\alpha} + \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \\
&\qquad \times \int_s^t \int_{S\times A} q(D \mid y, a) \left(\int_S p_\pi(s, x, \tau, \mathrm{d}y)\pi_\tau(\mathrm{d}a \mid y)\nu(\mathrm{d}x) \right) \mathrm{d}\tau\,\mathrm{d}t \\
&= \frac{\nu(D)}{\alpha} + \int_{S\times A} q(D \mid y, a) \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \left(\int_s^t \mathrm{P}_{s,\nu}^\pi(x(\tau) \in \mathrm{d}y, a(\tau) \in \mathrm{d}a)\,\mathrm{d}\tau \right) \mathrm{d}t \\
&= \frac{\nu(D)}{\alpha} + \int_{S\times A} q(D \mid y, a) \left(\frac{1}{\alpha} \int_s^\infty \mathrm{e}^{-\alpha(\tau-s)}\, \mathrm{P}_{s,\nu}^\pi(x(\tau) \in \mathrm{d}y, a(\tau) \in \mathrm{d}a)\,\mathrm{d}\tau \right) \\
&= \frac{\nu(D)}{\alpha} + \frac{1}{\alpha} \int_{S\times A} q(D \mid y, a)\mu^\pi(\mathrm{d}y, \mathrm{d}a),
\end{aligned}$$

and so (a) follows.

(b) By Lemma 9.4.4 of [14], there exists a stochastic kernel ϕ^μ (depending on μ) on A given S, which is concentrated on $A(x)$ for all $x \in S$, such that

$$\mu(D \times C) = \int_D \phi^\mu(C \mid x)\hat{\mu}_S(\mathrm{d}x) \quad \text{for all } D \in \mathcal{B}(S) \text{ and } C \in \mathcal{B}(A).$$

Obviously, ϕ^μ is in Π_s. To prove (b), it suffices to show that

$$\int_{S\times A} h(x, a)\mu(\mathrm{d}x, \mathrm{d}a) = \int_{S\times A} h(x, a)\mu^{\phi^\mu}(\mathrm{d}x, \mathrm{d}a) \tag{5.7}$$

for each *nonnegative* and *bounded* measurable function h on $S \times A$. To this end, for any nonnegative measurable function h and $x \in S$, define

$$\tilde{V}_h(x, \phi^{\mu}) := \int_0^{\infty} \mathrm{e}^{-\alpha t}\, \mathrm{E}_x^{\phi^{\mu}}[h(x(t), a(t))]\, \mathrm{d}t \quad \text{and} \quad \tilde{V}_h(\nu, \phi^{\mu}) := \int_S \tilde{V}_h(x, \phi^{\mu})\nu(\mathrm{d}x).$$

By Fubini's theorem, Definition 3.1(a), and (2.6), we obtain

$$\tilde{V}_h(\nu, \phi^{\mu}) = \int_{S\times A} h(x, a)\mu^{\phi^{\mu}}(\mathrm{d}x,\ \mathrm{d}a). \tag{5.8}$$

When h is nonnegative and bounded (thus, $\tilde{V}_h(x, \phi^{\mu})$ is finite), by Lemma 5.1(a) (with h in lieu of c) and a straightforward calculation, we obtain

$$\alpha\tilde{V}_h(x, \phi^{\mu}) = \int_A h(x, a)\phi^{\mu}(\mathrm{d}a \mid x) + \int_S \tilde{V}_h(y, \phi^{\mu})q(\mathrm{d}y \mid x, \phi^{\mu}). \tag{5.9}$$

On the other hand, by Assumption A(ii) and $\|h\| := \sup_{(x,a)\in K} |h(x, a)| < \infty$,

$$\begin{aligned}\int_S\biggl(\int_S |\tilde{V}_h(s, y, \phi^{\mu})q(\mathrm{d}y \mid x, \phi^{\mu})|\biggr)\hat{\mu}_S(\mathrm{d}x) &\le \int_S\int_S |q(\mathrm{d}y \mid x, \phi^{\mu})|\frac{\|h\|}{\alpha}\hat{\mu}_S(\mathrm{d}x)\\ &\le \frac{2M_0\|h\|}{\alpha}\int_S w(x)\hat{\mu}_S(\mathrm{d}x)\\ &< \infty.\end{aligned}$$

Thus, from Fubini's theorem we can derive

$$\begin{aligned}&\int_{S\times A} h(x, a)\mu(\mathrm{d}x,\ \mathrm{d}a)\\ &\quad= \int_{S\times A} h(x, a)\phi^{\mu}(\mathrm{d}a \mid x)\hat{\mu}_S(\mathrm{d}x) \quad \text{(by (3.3))}\\ &\quad= \int_S\biggl(\alpha\tilde{V}_h(x, \phi^{\mu}) - \int_S \tilde{V}_h(y, \phi^{\mu})q(\mathrm{d}y \mid x, \phi^{\mu})\biggr)\hat{\mu}_S(\mathrm{d}x) \quad \text{(by (5.9))}\\ &\quad= \int_S \tilde{V}_h(x, \phi^{\mu})\biggl(\nu(\mathrm{d}x) + \int_{S\times A} q(\mathrm{d}x \mid y, a)\mu(\mathrm{d}y,\ \mathrm{d}a)\biggr) \quad \text{(by (3.2))}\\ &\qquad - \int_S\biggl(\int_S \tilde{V}_h(y, \phi^{\mu})q(\mathrm{d}y \mid x, \phi^{\mu})\biggr)\hat{\mu}_S(\mathrm{d}x)\\ &\quad= \int_S \tilde{V}_h(x, \phi^{\mu})\nu(\mathrm{d}x) + \int_S \tilde{V}_h(x, \phi^{\mu})\int_{S\times A} q(\mathrm{d}x \mid y, a)\phi^{\mu}(\mathrm{d}a \mid y)\hat{\mu}_S(\mathrm{d}y)\\ &\qquad - \int_S\biggl(\int_S \tilde{V}_h(y, \phi^{\mu})q(\mathrm{d}y \mid x, \phi^{\mu})\biggr)\hat{\mu}_S(\mathrm{d}x)\\ &\quad= \tilde{V}_h(\nu, \phi^{\mu}) + \int_S\int_S \tilde{V}_h(x, \phi^{\mu})q(\mathrm{d}x \mid y, \phi^{\mu})\hat{\mu}_S(\mathrm{d}y)\\ &\qquad - \int_S\biggl(\int_S \tilde{V}_h(y, \phi^{\mu})q(\mathrm{d}y \mid x, \phi^{\mu})\biggr)\hat{\mu}_S(\mathrm{d}x)\\ &\quad= \tilde{V}_h(\nu, \phi^{\mu}),\end{aligned}$$

which together with (5.8) implies (5.7).

(c) The desired result follows from (b). Indeed, for any fixed $x \in S$, we take an initial distribution ν such that $\nu(\{x\}) = 1$. Then, by Assumption A and Theorem 3.3(a) of [7], we have

$$\tilde{V}_w(x,\pi) \le \frac{b_0}{\alpha(\alpha - c_0)} + \frac{w(x)}{\alpha - c_0} \quad \text{for all } x \in S \text{ and } \pi \in \Pi,$$

which implies that $\int_S w(y)\hat{\mu}_S^\pi(\mathrm{d}y) = \tilde{V}_w(x,\pi) < \infty$ for all $\pi \in \Pi$. By (2.9) and (b), for any policy $\pi \in \Pi$ and any initial state $x \in S$, there exists a randomized stationary policy $\phi_x^\pi \in \Pi_s$ (depending on π and x) such that

$$V_\alpha(x,\phi_x^\pi) = V_\alpha(0,x,\phi_x^\pi) = V_\alpha(s,x,\pi).$$

Hence,

$$V_\alpha^*(x) := \inf_{\phi\in\Pi_s} V_\alpha(x,\phi) = \inf_{\phi\in\Pi_s} V_\alpha(0,x,\phi) = \inf_{\pi\in\Pi} V_\alpha(s,x,\pi),$$

which is desirable.

5.2. Proof of Theorem 3.2

Before proving Theorem 3.2, we need some general lemmas. Denote by $M_+(S)$ the family of nonnegative measurable functions on S.

Lemma 5.2. *Suppose that Assumption B(iii) holds. Then the function* $\int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x,\cdot)$ *is l.s.c. on* $A(x)$ *for each* $x \in S$ *and* $u \in M_+(S)$.

Proof. Since $u \in M_+(S)$, there exists a nondecreasing sequence of nonnegative simple measurable functions $\{h_n\}$ such that $h_n \uparrow u$. On the other hand, for any fixed $x \in S$ and a sequence $\{a_x^n\}$ in $A(x)$, it follows from Assumption B(i) that there exists a convergent subsequence $\{a_x^k\}$ of $\{a_x^n\}$ such that $a_x^k \to a_x \in A(x)$ as $k \to \infty$. Hence, by Proposition C.4(b) of [13], for any $n \ge 1$,

$$\begin{aligned}\liminf_{k\to\infty} \int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x, a_x^k) &\ge \liminf_{k\to\infty} \int_{S-\{x\}} h_n(y)q(\mathrm{d}y \mid x, a_x^k)\\ &\ge \int_{S-\{x\}} h_n(y)q(\mathrm{d}y \mid x, a_x).\end{aligned}$$

Letting $n \to \infty$ on both sides above, it follows from the monotone convergence theorem that

$$\liminf_{k\to\infty} \int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x, a_x^k) \ge \int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x, a_x),$$

which implies that the function $\int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x,\cdot)$ is l.s.c. on $A(x)$. This completes the proof.

Define operators T_ϕ (for each fixed $\phi \in \Pi_s$) and T on $M_+(S)$ as

$$\begin{aligned}T_\phi u(x) &:= \frac{c(x,\phi)}{\alpha + q(x,\phi)} + \frac{1}{\alpha + q(x,\phi)} \int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x,\phi) \quad \text{for all } x \in S,\\ Tu(x) &:= \inf_{a\in A(x)} \left\{ \frac{c(x,a)}{\alpha + q(x,a)} + \frac{1}{\alpha + q(x,a)} \int_{S-\{x\}} u(y)q(\mathrm{d}y \mid x,a) \right\} \quad \text{for all } x \in S.\end{aligned} \tag{5.10}$$

Note that these operators are *monotone*, that is, $u \ge u'$ implies that $T_\phi u \ge T_\phi u'$, and similarly for T.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Lemma 5.3. *Under Assumption A, the following assertions hold.*

(a) *For each $\phi \in \Pi_s$, define a sequence $\{u_n^\phi\}$ as*

$$u_0^\phi := 0, \qquad u_{n+1}^\phi := T_\phi u_n^\phi \quad \textit{for all } n \geq 0. \tag{5.11}$$

Then $\{u_n^\phi\}$ is increasing in $n \geq 0$, and $\lim_{n\to\infty} u_n^\phi = V_\alpha(\phi)$.

(b) *If, in addition, Assumption B holds, then the sequence $\{u_n\}$ in (3.4) is increasing in $n \geq 0$, and* $\lim_{n\to\infty} u_n \leq V_\alpha(\phi)$ *for all $\phi \in \Pi_s$.*

Proof. (a) In view of the hypotheses on the model (2.1), u_n^ϕ is well defined and $u_n^\phi \in M_+(S)$ for all $n \geq 0$. Then (a) follows from the monotonicity of T_ϕ and the proof of Lemma 5.1(a).

(b) From (a), it suffices to prove that

$$u_n \leq u_n^\phi \quad \text{and} \quad u_n \in M_+(S) \quad \text{for all } n \geq 0 \text{ and } \phi \in \Pi_s. \tag{5.12}$$

By (3.4) and (5.10), we know that

$$u_0 = 0, \qquad u_{n+1} = Tu_n \quad \text{for all } n \geq 0. \tag{5.13}$$

Obviously, (5.12) is true for $n = 0$. Suppose that $u_n \leq u_n^\phi$ and $u_n \in M_+(S)$ for some $n \geq 0$. Since $c(x, a) \geq 0$ and T is monotone, $u_{n+1} \geq u_n$ for all $n \geq 0$. Under Assumptions A and B, from Lemma 5.2, we have $\int_{S-\{x\}} u_n(y)q(\mathrm{d}y \mid x, \cdot)$ is l.s.c. on $A(x)$ for each $x \in S$. Thus, by Proposition A.3(a) of [13] and Lemma 8.3.8(a) of [14], $u_{n+1} \in M_+(S)$. Moreover, (5.11) and (5.13) give

$$u_{n+1} = Tu_n \leq Tu_n^\phi \leq T_\phi u_n^\phi = u_{n+1}^\phi,$$

where the first inequality follows from the inductive hypothesis and the second inequality is shown below. In fact, by (5.10) we have

$$\begin{aligned} Tu_n^\phi(x) \leq{}& \frac{c(x, a)}{\alpha + q(x, a)} \\ &+ \frac{1}{\alpha + q(x, a)} \int_{S-\{x\}} u_n^\phi(y)q(\mathrm{d}y \mid x, a) \quad \text{for all } x \in S \text{ and } a \in A(x). \end{aligned}$$

Multiplying both sides by $\alpha + q(x, a)$ and taking the expectation with respect to $\phi(\cdot \mid x)$, we obtain

$$(\alpha + q(x, \phi))Tu_n^\phi(x) \leq c(x, \phi) + \int_{S-\{x\}} u_n^\phi(y)q_\phi(\mathrm{d}y \mid x, \phi),$$

which divided by $\alpha + q(x, \phi)$ yields $Tu_n^\phi \leq T_\phi u_n^\phi$. Hence, (5.12) is valid for all $n \geq 0$. By (a) and letting $n \to \infty$ in (5.12), we obtain $\lim_{n\to\infty} u_n \leq V_\alpha(\phi)$. This completes the proof.

We are now ready to prove Theorem 3.2.

Proof of Theorem 3.2. (a) Let $\{u_n\}$ be as in (3.4), and let V_α^* be as in Theorem 3.1(c). Since $0 = u_0 \leq u_1 \leq \cdots \leq u_n$, the limit $u^* := \lim_{n\to\infty} u_n$ exists, and $0 \leq u^* \leq V_\alpha(\phi)$ for all $\phi \in \Pi_s$ (by Lemma 5.3(b)), which imply that

$$V_\alpha^* = \inf_{\phi \in \Pi_s} V_\alpha(\phi) \geq u^* \geq u_n \geq 0 \quad \text{for all } n \geq 0. \tag{5.14}$$

Then, to complete the proof of part (a), it suffices to show the converse, that is, $u^* \geq V_\alpha^*$.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

To this end, by the monotonicity of T and u_n, we have $Tu^* \geq Tu_n = u_{n+1}$. Letting $n \to \infty$ on both sides, we obtain

$$Tu^* \geq u^*. \tag{5.15}$$

On the other hand, from the proof of Lemma 5.3(b), we know that $u_n \in M_+(S)$ for any $n \geq 0$. Hence, for any fixed $x \in S$ and any $n \geq 1$, by Lemma 5.2, Proposition A.3(a) of [13], and Lemma 8.3.8(a) of [14], there exists $a_x^n \in A(x)$ (depending on x and n) such that

$$\begin{aligned} u^*(x) &\geq u_{n+1}(x) \\ &= \inf_{a\in A(x)}\left\{\frac{c(x,a)}{\alpha+q(x,a)} + \frac{1}{\alpha+q(x,a)}\int_{S-\{x\}} u_n(y)q(\mathrm{d}y \mid x,a)\right\} \\ &= \frac{c(x,a_x^n)}{\alpha+q(x,a_x^n)} + \frac{1}{\alpha+q(x,a_x^n)}\int_{S-\{x\}} u_n(y)q(\mathrm{d}y \mid x,a_x^n). \end{aligned} \tag{5.16}$$

Since $a_x^n \in A(x)$ for all $n \geq 1$, by Assumption B(ii), there exists a subsequence of $\{a_x^n\}$, denoted by $\{a_x^{nk}\}$, such that $a_x^{nk} \to a_x \in A(x)$ as $k \to \infty$. Then, for any fixed $k_0 > 0$, letting $n = n_k$ in (5.16) and $k \to \infty$, it follows from the monotonicity of u_n that

$$\begin{aligned} u^*(x) &\geq \liminf_{k\to\infty}\left\{\frac{c(x,a_x^{n_k})}{\alpha+q(x,a_x^{nk})} + \frac{1}{\alpha+q(x,a_x^{nk})}\int_{S-\{x\}} u_{n_k}(y)q(\mathrm{d}y \mid x,a_x^{n_k})\right\} \\ &\geq \frac{c(x,a_x)}{\alpha+q(x,a_x)} \\ &\quad + \frac{1}{\alpha+q(x,a_x)}\liminf_{k\to\infty}\int_{S-\{x\}} u_{n_{k_0}}(y)q(\mathrm{d}y \mid x,a_x^{nk}) \quad (\text{for all } k \geq k_0) \\ &\geq \frac{c(x,a_x)}{\alpha+q(x,a_x)} + \frac{1}{\alpha+q(x,a_x)}\int_{S-\{x\}} u_{n_{k_0}}(y)q(\mathrm{d}y \mid x,a_x) \quad (\text{by Lemma 5.2}). \end{aligned}$$

Letting $k_0 \to \infty$, by the monotone convergence theorem we obtain

$$\begin{aligned} u^*(x) &\geq \frac{c(x,a_x)}{\alpha+q(x,a_x)} + \frac{1}{\alpha+q(x,a_x)}\int_{S-\{x\}} u^*(y)q(\mathrm{d}y \mid x,a_x) \\ &\geq \inf_{a\in A(x)}\left\{\frac{c(x,a)}{\alpha+q(x,a)} + \frac{1}{\alpha+q(x,a)}\int_{S-\{x\}} u^*(y)q(\mathrm{d}y \mid x,a)\right\}, \end{aligned}$$

which means that $u^* \geq Tu^*$. This together with (5.15) gives $u^* = Tu^*$.

Moreover, since $u^* \in M_+(S)$, from Lemma 5.2 and Lemma 8.3.8(a) of [14], it follows that there exists a policy $f^* \in F$ for which

$$\begin{aligned} u^*(x) &= \frac{c(x,f^*(x))}{\alpha+q(x,f^*(x))} \\ &\quad + \frac{1}{\alpha+q(x,f^*(x))}\int_{S-\{x\}} u^*(y)q(\mathrm{d}y \mid x,f^*(x)) \quad \text{for all } x \in S. \end{aligned}$$

Thus, by Lemma 5.1(a), we obtain $u^* \geq V_\alpha(f^*) \geq \inf_{\phi\in\Pi_s} V_\alpha(\phi) = V_\alpha^*$, which together with (5.14) gives

$$u^* = V_\alpha(f^*) = V_\alpha^*. \tag{5.17}$$

This completes the proof of part (a).

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

(b) Since $u^* = Tu^*$ (just proved), (b) follows from (5.17) and (5.10).

(c) Suppose that $f \in F$ attains the minimum on the right-hand side of (3.5). By (a) and Lemma 5.1, we have $V_\alpha^* \geq V_\alpha(f)$, which together with Definition 2.2 yields the fact that f is discounted cost optimal.

(d) The existence of a discounted cost optimal policy is ensured by (5.17).

5.3. Proof of Theorem 3.3

In the following, we prove Theorem 3.3 by using some properties of continuous-time Markov chains in [2, Chapter 2].

Proof of Theorem 3.3. (a) Denote by

$$S_0 := 0 \quad \text{and} \quad S_{n+1} := \inf\{t > S_n : x(t) \neq x(S_n)\} \quad \text{for } n = 0, 1, 2, \dots,$$

the nth jumping time. Fix $\delta \geq 0$ and $f \in F$. For any $x \notin B$, by a direct calculation we have

$$\begin{aligned}
U_\delta^B(x, f) &= \mathrm{E}_x^f\left[\int_0^{\tau_B^f} \mathrm{e}^{-\delta t} g(x(t), f)\,\mathrm{d}t\right] \\
&= \mathrm{E}_x^f\left[\int_0^{\infty} \mathbf{1}_{\{\tau_B^f > t\}}\,\mathrm{e}^{-\delta t} g(x(t), f)\,\mathrm{d}t\right] \\
&= \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\,\mathrm{P}_x^f(x(u) \notin B \text{ for all } u \in [0, t],\ x(t) \in \mathrm{d}y)\,\mathrm{d}t \\
&= \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f) \\
&\qquad \times \sum_{m=0}^{\infty} \mathrm{P}_x^f(S_m \leq t < S_{m+1},\ x(S_l) \notin B,\ l = 0, 1, \dots, m,\ x(S_m) \in \mathrm{d}y)\,\mathrm{d}t \\
&= \lim_{n\to\infty} \sum_{m=0}^{n} \int_0^\infty \mathrm{e}^{-\delta t} \\
&\qquad \times \int_S g(y, f)\,\mathrm{P}_x^f(S_m \leq t < S_{m+1},\ x(S_l) \notin B,\ l = 0, 1, \dots, m,\ x(S_m) \in \mathrm{d}y)\,\mathrm{d}t,
\end{aligned}$$

where $\mathbf{1}_D$ denotes the indicator function of D. Define, for any $n \geq 0$,

$$\begin{aligned}
U_\delta^n(x, f) := \sum_{m=0}^{n} \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f) \\
\times \mathrm{P}_x^f(S_m \leq t < S_{m+1},\ x(S_l) \notin B,\ l = 0, 1, \dots, m,\ x(S_m) \in \mathrm{d}y)\,\mathrm{d}t.
\end{aligned} \tag{5.18}$$

Then

$$U_\delta^B(x, f) = \lim_{n\to\infty} U_\delta^n(x, f) \quad \text{for all } x \notin B. \tag{5.19}$$

Now define operator T_B^f as follows. For each nonnegative measurable function u defined on $S - B$,

$$\begin{aligned}
T_B^f u(x) := \frac{g(x, f)}{\delta + q(x, f)} \\
+ \frac{1}{\delta + q(x, f)} \int_{S-B-\{x\}} u(y) q(\mathrm{d}y \mid x, f) \quad \text{for all } x \notin B,
\end{aligned} \tag{5.20}$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

where $g(x, f) := g(x, f(x))$, $q(x, f) := -q(\{x\} \mid x, f(x))$, and $q(\cdot \mid x, f) := q(\cdot \mid x, f(x))$. Then we have

$$U_\delta^{n+1}(x, f) = T_B^f U_\delta^n(x, f) \quad \text{for each } n \geq -1, \tag{5.21}$$

where $U_\delta^{-1}(x, f) := 0$ for any $x \notin B$.

Indeed, by (5.18) and a straightforward calculation, we obtain

$$\begin{aligned}
&U_\delta^{n+1}(x, f)\\
&\quad = \sum_{m=0}^{n+1} \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\\
&\qquad\qquad \times \mathrm{P}_x^f(S_m \leq t < S_{m+1},\ x(S_l) \notin B,\ l = 0, 1, \ldots, m,\ x(S_m) \in \mathrm{d}y)\,\mathrm{d}t\\
&\quad = \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\,\mathrm{P}_x^f(S_0 \leq t < S_1, x(S_0) \notin B, x(S_0) \in \mathrm{d}y)\,\mathrm{d}t\\
&\qquad + \sum_{m=1}^{n+1} \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\\
&\qquad\qquad \times \mathrm{P}_x^f(S_m \leq t < S_{m+1},\ x(S_l) \notin B,\ l = 0, 1, \ldots, m,\ x(S_m) \in \mathrm{d}y)\,\mathrm{d}t\\
&\quad = g(x, f) \int_0^\infty \mathrm{e}^{-\delta t}\,\mathrm{P}_x^f(t < S_1)\,\mathrm{d}t\\
&\qquad + \sum_{m=1}^{n+1} \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\\
&\qquad\quad \times \mathrm{E}_x^f\biggl[\mathrm{P}_x^f\biggl(S_m \leq t < S_{m+1}, \bigcap_{l=0}^{m}\{x(S_l) \notin B\}, x(S_m) \in \mathrm{d}y \biggm| S_0, x(S_0), S_1, x(S_1)\biggr)\biggr]\,\mathrm{d}t\\
&\quad = g(x, f) \int_0^\infty \mathrm{e}^{-\delta t}\mathrm{e}^{-q(x,f)t}\,\mathrm{d}t\\
&\qquad + \sum_{m=1}^{n+1} \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\\
&\qquad\quad \times \mathrm{E}_x^f\biggl[\mathbf{1}_{\{x(S_0)\notin B, x(S_1)\notin B\}}\\
&\qquad\qquad \times \mathrm{P}_x^f\biggl(S_m \leq t < S_{m+1}, \bigcap_{l=2}^{m}\{x(S_l) \notin B\}, x(S_m) \in \mathrm{d}y \biggm| S_0, x(S_0), S_1, x(S_1)\biggr)\biggr]\,\mathrm{d}t\\
&\quad = \frac{g(x, f)}{\delta + q(x, f)}\\
&\qquad + \sum_{m=1}^{n+1} \int_0^\infty \mathrm{e}^{-\delta t} \int_S g(y, f)\\
&\qquad\quad \times \int_{S-B-\{x\}} \int_0^t \mathrm{d}\mathrm{P}_x^f(S_1 \leq v,\ x(S_1) \in \mathrm{d}z)\\
&\qquad\quad \times \mathrm{P}_x^f\biggl(S_m \leq t < S_{m+1}, \bigcap_{l=2}^{m}\{x(S_l) \notin B\}, x(S_m) \in \mathrm{d}y \biggm| 0, x, S_1 = v, x(S_1) = z\biggr)\,\mathrm{d}t
\end{aligned}$$

$$\begin{aligned}
&= \frac{g(x, f)}{\delta + q(x, f)} \\
&\quad + \sum_{m=1}^{n+1} \int_0^{\infty} \mathrm{e}^{-\delta t} \int_S g(y, f) \\
&\qquad \times \int_{S-B-\{x\}} \int_0^t q(\mathrm{d}z \mid x, f) \mathrm{e}^{-q(x,f)v} \,\mathrm{d}v \\
&\qquad \times \mathrm{P}_x^f \Bigg(S_m \le t < S_{m+1}, \bigcap_{l=2}^{m} \{x(S_l) \notin B\}, x(S_m) \in \mathrm{d}y \,\Bigg|\, 0, x, S_1 = v,\ x(S_1) = z \Bigg) \\
&= \frac{g(x, f)}{\delta + q(x, f)} \\
&\quad + \int_{S-B-\{x\}} q(\mathrm{d}z \mid x, f) \int_0^{\infty} \mathrm{e}^{-q(x,f)v} \mathrm{e}^{-\delta v} \\
&\qquad \times \Bigg[\sum_{m=0}^{n} \int_v^{\infty} \mathrm{e}^{-\delta(t-v)} \int_S g(y, f) \\
&\qquad\qquad \times \mathrm{P}_z^f (S_m \le t < S_{m+1},\ x(S_l) \notin B,\ l = 0, 1, \ldots, m, x(S_m) \in \mathrm{d}y) \,\mathrm{d}t \Bigg] \mathrm{d}v \\
&= \frac{g(x, f)}{\delta + q(x, f)} + \int_{S-B-\{x\}} q(\mathrm{d}z \mid x, f) \int_0^{\infty} \mathrm{e}^{-q(x,f)v} \mathrm{e}^{-\delta v} U_\delta^n(z, f) \,\mathrm{d}v \\
&= \frac{g(x, f)}{\delta + q(x, f)} + \frac{1}{\delta + q(x, f)} \int_{S-B-\{x\}} U_\delta^n(z, f) q(\mathrm{d}z \mid x, f).
\end{aligned}$$

Hence, letting $n \to \infty$ in (5.21) and recalling (5.19)–(5.20), we obtain

$$U_\delta^B(x, f) = \frac{g(x, f)}{\delta + q(x, f)} + \frac{1}{\delta + q(x, f)} \int_{S-B-\{x\}} U_\delta^B(y, f) q(\mathrm{d}y \mid x, f) \quad \text{for all } x \notin B,$$

which implies (3.7).

On the other hand, suppose that u is a nonnegative solution of (3.7). Note that (3.7) can be rewritten as

$$u(x) \ge \frac{g(x, f)}{\delta + q(x, f)} + \frac{1}{\delta + q(x, f)} \int_{S-B-\{x\}} u(y) q(\mathrm{d}y \mid x, f) = T_B^f u(x) \quad \text{for all } x \notin B; \tag{5.22}$$

hence, $u(x) \ge U_\delta^{-1}(x, f)$ and also $u(x) \ge U_\delta^0(x, f)$. Suppose that $u(x) \ge U_\delta^n(x, f)$ for some $n \ge -1$. It follows from (5.20)–(5.22) that $u(x) \ge U_\delta^{n+1}(x, f)$ for all $x \notin B$ and $n \ge -1$. Thus, the proof of (a) is complete by (5.19).

(b) Under Assumptions A and B, by Theorem 3.2, for any $\alpha > 0$, there exists an α-discounted cost optimal stationary policy $f_\alpha \in F$ such that $V_\alpha(x, f_\alpha) = V_\alpha^*(x)$ for all $x \in S$. Choose

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

$x_0 \in S$ satisfying Assumption C(i). By (2.9) and the strong Markov property, we derive

$$\begin{aligned} h_\alpha(x) &:= V_\alpha(x, f_\alpha) - V_\alpha(x_0, f_\alpha) \\ &= \mathrm{E}_x^{f_\alpha}\bigg[\int_0^\infty \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg] - V_\alpha(x_0, f_\alpha) \\ &= \mathrm{E}_x^{f_\alpha}\bigg[\int_0^{\tau_{x_0}^{f_\alpha}} \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg] + \mathrm{E}_x^{f_\alpha}\bigg[\int_{\tau_{x_0}^{f_\alpha}}^\infty \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg] - V_\alpha(x_0, f_\alpha) \\ &= \mathrm{E}_x^{f_\alpha}\bigg[\int_0^{\tau_{x_0}^{f_\alpha}} \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg] + \mathrm{E}_x^{f_\alpha}[\exp(-\alpha\tau_{x_0}^{f_\alpha}) V_\alpha(x_0, f_\alpha)] - V_\alpha(x_0, f_\alpha) \\ &= \mathrm{E}_x^{f_\alpha}\bigg[\int_0^{\tau_{x_0}^{f_\alpha}} \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg] + \mathrm{E}_x^{f_\alpha}[\exp(-\alpha\tau_{x_0}^{f_\alpha}) - 1] V_\alpha(x_0, f_\alpha). \end{aligned}$$

Since $c(x, a) \geq 0$ and $\exp(-\alpha\tau_{x_0}^{f_\alpha}) \leq 1$, we have

$$\mathrm{E}_x^{f_\alpha}[\exp(-\alpha\tau_{x_0}^{f_\alpha}) - 1] V_\alpha(x_0, f_\alpha) \leq h_\alpha(x) \leq \mathrm{E}_x^{f_\alpha}\bigg[\int_0^{\tau_{x_0}^{f_\alpha}} \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg]. \tag{5.23}$$

Hence, the desired result follows from (b) with $\delta := \alpha$, $B =: \{x_0\}$, $f := f_\alpha$, and $g := c$ in (3.6)–(3.7).

(c) Let $f \in F$ be any stationary policy. Then, under the current condition in (c), we have $q(B \mid x, f) \geq \beta$ for all $x \notin B$ and $f \in F$. Let $u(x) = 1/\beta$ for all $x \notin B$ in (3.7). Then

$$1 + \int_{S-B} u(y) q(\mathrm{d}y \mid x, f) = 1 - \frac{q(B \mid x, f)}{\beta} \leq 0 \leq \delta u(x) \quad \text{for all } x \notin B \text{ and } \delta > 0,$$

since $q(S \mid x, f) = 0$ for all $x \in S$ and $f \in F$. Hence, it follows from (a) that

$$\mathrm{E}_x^f\bigg[\int_0^{\tau_B^f} \mathrm{e}^{-\delta t}\,\mathrm{d}t\bigg] \leq \frac{1}{\beta} \quad \text{for all } x \notin B \text{ and } f \in F.$$

Therefore, (c) follows by letting $\delta \to 0$ in the above inequality.

(d) Since $\mathrm{e}^{-x} - 1 \geq -x$, it follows from (5.23) and part (b) that

$$-\alpha V_\alpha(x_0, f_\alpha)\,\mathrm{E}_x^{f_\alpha}[\tau_{x_0}^{f_\alpha}] \leq h_\alpha(x) \leq \mathrm{E}_x^{f_\alpha}\bigg[\int_0^{\tau_{x_0}^{f_\alpha}} \mathrm{e}^{-\alpha t} c(x(t), f_\alpha)\,\mathrm{d}t\bigg] \leq u(x), \tag{5.24}$$

with u as in (b). On the other hand, by Assumption C(i), there exists a constant $\tilde{M} > 0$ such that $0 \leq \alpha V_\alpha(x_0, f_\alpha) \leq \tilde{M} < \infty$. This fact together with (5.24), and taking $B = \{x_0\}$ and $f = f_\alpha$ in (c), implies that

$$-\frac{\tilde{M}}{\beta} \leq h_\alpha(x) \leq u(x),$$

which yields Assumption C.

(e) Obviously, (e) follows from (5.24) and the proof of part (d).

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

5.4. Proof of Theorem 3.4

In the proof of Theorem 3.4, we develop a so-called average cost minimum nonnegative solution approach.

Proof of Theorem 3.4. (a) To prove assertion (a), we first show that the function $J_f(x,t)$ satisfies (3.9) with equality, that is, for each $x \in S$ and $t \geq 0$,

$$\begin{aligned} J_f(x,t) &= c(x, f)t\mathrm{e}^{-q(x,f)t} \\ &\quad + \int_0^t \mathrm{e}^{-q(x,f)s}\biggl[q(x, f)c(x, f)s + \int_{S-\{x\}} J_f(y, t-s)q(\mathrm{d}y \mid x, f)\biggr]\,\mathrm{d}s. \end{aligned} \tag{5.25}$$

Then, we further prove that $u(x,t) \geq J_f(x,t)$ for any nonnegative measurable function $u(x,t)$ satisfying (3.9).

In order to prove (5.25), we apply the construction of $p_f(x,t,D)$ (for any fixed $f \in F$): for each $x \in S$, $D \in \mathcal{B}(S)$, $t \geq 0$, and $n \geq 1$, let

$$p_0^f(x,t,D) := \mathbf{1}_D(x)\mathrm{e}^{-q(x,f)t},$$

$$p_n^f(x,t,D) := \int_0^t \mathrm{e}^{-q(x,f)s} \int_{S-\{x\}} p_{n-1}^f(y, t-s, D)q(\mathrm{d}y \mid x, f)\,\mathrm{d}s, \tag{5.26}$$

$$S_n^f(x,t,D) := \sum_{k=0}^{n} p_k^f(x,t,D), \tag{5.27}$$

$$\begin{aligned} m_0^f(x,t) &:= \int_0^t \int_S c(y, f)S_0^f(x, s, \mathrm{d}y)\,\mathrm{d}s \\ &= c(x, f)t\mathrm{e}^{-q(x,f)t} + \int_0^t q(x, f)\mathrm{e}^{-q(x,f)s}c(x, f)s\,\mathrm{d}s, \end{aligned} \tag{5.28}$$

$$m_n^f(x,t) := \int_0^t \int_S c(y, f)S_n^f(x, s, \mathrm{d}y)\,\mathrm{d}s. \tag{5.29}$$

Then it follows from (5.27) and Theorem 2.21 of [2] (or Theorem 2 of [5]) that

$$S_n^f(x,t,D) \uparrow p_f(x,t,D) \quad \text{as } n \to \infty.$$

Hence, from (3.8) and (5.29), we have

$$m_n^f(x,t) \uparrow J_f(x,t) \quad \text{for all } x \in S \text{ and } t \geq 0. \tag{5.30}$$

On the other hand, by (5.26) and (5.29), for each $n \geq 1$, we derive

$$\begin{aligned} &m_n^f(x,t) \\ &\quad = \int_0^t \int_S c(y, f)S_n^f(x, s, \mathrm{d}y)\,\mathrm{d}s \\ &\quad = m_0^f(x,t) + \int_0^t \int_S c(y, f)\sum_{k=1}^{n} p_k^f(x, s, \mathrm{d}y)\,\mathrm{d}s \\ &\quad = m_0^f(x,t) \\ &\qquad + \int_0^t \int_S c(y, f)\sum_{k=1}^{n}\biggl(\int_0^s \mathrm{e}^{-q(x,f)r} \int_{S-\{x\}} p_{k-1}^f(z, s-r, \mathrm{d}y)q(\mathrm{d}z \mid x, f)\,\mathrm{d}r\biggr)\,\mathrm{d}s \end{aligned}$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

$$
\begin{aligned}
&= m_0^f(x,t)\\
&\quad + \int_0^t \mathrm{e}^{-q(x,f)r}\int_{S-\{x\}}\biggl(\int_r^t\int_S c(y,f)\sum_{k=1}^{n} p_{k-1}^f(z,s-r,\,\mathrm{d}y)\,\mathrm{d}s\biggr)q(\mathrm{d}z\mid x,f)\,\mathrm{d}r\\
&= m_0^f(x,t)\\
&\quad + \int_0^t \mathrm{e}^{-q(x,f)r}\int_{S-\{x\}}\biggl(\int_0^{t-r}\int_S c(y,f)\sum_{k=0}^{n-1} p_{k}^f(z,s,\,\mathrm{d}y)\,\mathrm{d}s\biggr)q(\mathrm{d}z\mid x,f)\,\mathrm{d}r\\
&= m_0^f(x,t) + \int_0^t \mathrm{e}^{-q(x,f)r}\int_{S-\{x\}} m_{n-1}^f(z,t-r)q(\mathrm{d}z\mid x,f)\,\mathrm{d}r,
\end{aligned}
$$

which together with (5.28) gives

$$
\begin{aligned}
m_n^f(x,t) &= c(x,f)t\mathrm{e}^{-q(x,f)t}\\
&\quad + \int_0^t \mathrm{e}^{-q(x,f)s}\biggl(q(x,f)c(x,f)s + \int_{S-\{x\}} m_{n-1}^f(y,t-s)q(\mathrm{d}y\mid x,f)\biggr)\,\mathrm{d}s.
\end{aligned}
\tag{5.31}
$$

Thus, (5.25) immediately follows from (5.31) and (5.30).

Now suppose that a nonnegative function $u(x,t)$ on $S\times[0,\infty)$ satisfies (3.9). Since $c(x,f)$ and $q(D\mid x,f)$ are nonnegative for all $x\notin D$, it follows from (3.9) and (5.28) that $u(x,t)\ge m_0^f(x,t)$. Then, by induction and (5.31), we know that $u(x,t)\ge m_n^f(x,t)$ for all $n\ge 1$. This fact together with (5.30) completes the proof of part (a).

(b) Since the transition rates in model (2.1) are conservative, (3.10) still holds when the function u is replaced by $u+L$ with any constant L. Therefore, without loss of generality, we may further assume that $u\ge 0$.

Let $\hat u(x,t)=u(x)+\rho t\ge 0$ for all $x\in S$ and $t\ge 0$, with $u\ge 0$ and ρ as in (3.10). Then, by (3.10) and $q(S-\{x\}\mid x,f)=q(x,f)$, we have

$$
\begin{aligned}
&c(x,f)t\mathrm{e}^{-q(x,f)t} + \int_0^t \mathrm{e}^{-q(x,f)s}\biggl(q(x,f)c(x,f)s + \int_{S-\{x\}}\hat u(y,t-s)q(\mathrm{d}y\mid x,f)\biggr)\,\mathrm{d}s\\
&\quad= c(x,f)t\mathrm{e}^{-q(x,f)t}\\
&\qquad + \int_0^t \mathrm{e}^{-q(x,f)s}\biggl(q(x,f)c(x,f)s + \int_{S-\{x\}}[u(y)+\rho(t-s)]q(\mathrm{d}y\mid x,f)\biggr)\,\mathrm{d}s\\
&\quad\le c(x,f)t\mathrm{e}^{-q(x,f)t}\\
&\qquad + \int_0^t \mathrm{e}^{-q(x,f)s}(q(x,f)c(x,f)s + q(x,f)u(x) + \rho - c(x,f)\\
&\qquad\qquad + \rho q(x,f)(t-s))\,\mathrm{d}s\\
&\quad= u(x)+\rho t - u(x)\mathrm{e}^{-q(x,f)t}\quad\text{(by a straightforward calculation)},
\end{aligned}
$$

and so,

$$
\begin{aligned}
\hat u(x,t) &\ge c(x,f)t\mathrm{e}^{-q(x,f)t}\\
&\quad + \int_0^t \mathrm{e}^{-q(x,f)s}\biggl(q(x,f)c(x,f)s + \int_{S-\{x\}}\hat u(y,t-s)q(\mathrm{d}y\mid x,f)\biggr)\,\mathrm{d}s.
\end{aligned}
$$

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

Hence, $\hat{u}(x, t)$ is a nonnegative solution to (3.9). By (a) we obtain

$$u(x) + \rho t = \hat{u}(x, t) \geq J_f(x, t) \quad \text{for all } x \in S \text{ and } t \geq 0.$$

Multiplying both sides by $1/t$ and letting $t \to \infty$, from (3.8) and (2.10), we obtain the desired result.

5.5. Proof of Theorem 3.5

We now prove Theorem 3.5 by using the average cost minimum nonnegative solution approach and the optimality inequality method.

Proof of Theorem 3.5. By Assumption C, there exists a subsequence $\{\alpha_m\}$ of $\{\alpha_n\}$ with $\alpha_m \downarrow 0$, a constant ρ^*, and a real-valued measurable function h^* on S such that

$$\rho^* = \lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(x_0) \geq 0 \quad \text{and} \quad h^*(x) := \liminf_{m\to\infty} h_{\alpha_m}(x) \geq L^*. \tag{5.32}$$

Then, for each $m \geq 1$, Theorem 3.2(b) ensures the existence of a policy $f_m \in F$ (depending on α_m) such that

$$V^*_{\alpha_m}(x) = \frac{c(x, f_m)}{\alpha_m + q(x, f_m)} + \frac{1}{\alpha_m + q(x, f_m)} \int_{S-\{x\}} V^*_{\alpha_m}(y) q(\mathrm{d}y \mid x, f_m),$$

which together with $|V^*_{\alpha_m}(x) q(x, f_m)| < \infty$ (by Assumption C(ii)) implies that

$$\alpha_m V^*_{\alpha_m}(x) = c(x, f_m) + \int_S V^*_{\alpha_m}(y) q(\mathrm{d}y \mid x, f_m) \quad \text{for all } x \in S \text{ and } t \geq 0. \tag{5.33}$$

Moreover, by Assumption C and $q(S \mid x, f_m) \equiv 0$, it follows from (5.33) that

$$\alpha_m V^*_{\alpha_m}(x_0) + \alpha_m h_{\alpha_m}(x) = c(x, f_m) + \int_S [h_{\alpha_m}(y) - L^*] q(\mathrm{d}y \mid x, f_m),$$

and so

$$\begin{aligned} \alpha_m V^*_{\alpha_m}(x_0) + \alpha_m h_{\alpha_m}(x) = c(x, f_m) &+ \int_{S-\{x\}} [h_{\alpha_m}(y) - L^*] q(\mathrm{d}y \mid x, f_m) \\ &+ [h_{\alpha_m}(x) - L^*] q(\{x\} \mid x, f_m(x)). \end{aligned} \tag{5.34}$$

On the other hand, for any fixed $x \in S$, Assumption B gives the existence of a subsequence $\{f_k(x)\}$ of $\{f_m(x)\}$ and $a^*_x \in A(x)$ such that

$$\lim_{k\to\infty} f_k(x) = a^*_x \quad \text{and} \quad \liminf_{k\to\infty} c(x, f_k(x)) \geq c(x, a^*_x).$$

These facts together with Lemma 8.3.7 of [14] (generalized Fatou's lemma) and (5.32)–(5.34) yield

$$\begin{aligned} \rho^* &\geq c(x, a^*_x) + \int_{S-\{x\}} [h^*(y) - L^*] q(\mathrm{d}y \mid x, a^*_x) + [h^*(x) - L^*] q(\{x\} \mid x, a^*_x) \\ &= c(x, a^*_x) + \int_S h^*(y) q(\mathrm{d}y \mid x, a^*_x) \\ &\geq \inf_{a\in A(x)} \{c(x, a) + \int_S h^*(y) q(\mathrm{d}y \mid x, a)\} \quad \text{for all } x \in S, \end{aligned}$$

and so (a) follows.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

(b) Suppose that $f \in F$ realizes the minimum in (3.11), so that

$$\rho^* \geq c(x, f(x)) + \int_S u^*(y) q(\mathrm{d}y \mid x, f(x)) \quad \text{for all } x \in S.$$

Thus, it follows from Theorem 3.4(b) that

$$\rho^* \geq J(0, x, f) \quad \text{for all } x \in S. \tag{5.35}$$

On the other hand, in view of (5.32) and Assumption C(ii), we have

$$\rho^* = \lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(x_0) = \lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(x) \quad \text{for all } x \in S.$$

This implies, by the well-known *Tauberian theorem*, that, for each $\pi \in \Pi$ and $x \in S$,

$$\begin{aligned}
\rho^* &= \lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(x) \\
&\leq \limsup_{m\to\infty} \alpha_m V_{\alpha_m}(0, x, \pi) \\
&= \limsup_{\alpha_m \downarrow 0} \alpha_m \int_0^\infty \mathrm{e}^{-\alpha_m t}\, \mathrm{E}^\pi_x\, c(x(t), a(t))\, \mathrm{d}t \\
&\leq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{E}^\pi_x\, c(x(t), a(t))\, \mathrm{d}t \\
&= J(0, x, \pi).
\end{aligned}$$

This inequality together with (5.35) gives $J(0, x, f) \leq \rho^* \leq J(0, x, \pi)$ for all $\pi \in \Pi$ and $x \in S$, which yields (b).

References

[1] Anderson, W. J. (1991). *Continuous-Time Markov Chains*. Springer, New York.

[2] Chen, M.-F. (2004). *From Markov Chains to Non-Equilibrium Particle Systems*, 2nd edn. World Scientific, River Edge, NJ.

[3] Doshi, B. T. (1976). Continuous time control of Markov processes on an arbitrary state space: discounted rewards. *Ann. Statist.* **4,** 1219–1235.

[4] Feinberg, E. A. (2004). Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Operat. Res.* **29,** 492–524.

[5] Feller, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48,** 488–515.

[6] Guo, X. P. (2007). Constrained optimization for average cost continuous-time Markov decision processes. *IEEE Trans. Automatic Control* **52,** 1139–1143.

[7] Guo, X. P. (2007). Continuous time Markov decision processes with discounted rewards: the case of Polish spaces. *Math. Operat. Res.* **32,** 73–87.

[8] Guo, X. P. and Hernández-Lerma, O. (2003). Continuous-time controlled Markov chains. *Ann. Appl. Prob.* **13,** 363–388.

[9] Guo, X. P. 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.

[10] Guo, X. P. and Rieder, U. (2006). Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Prob.* **16,** 730–756.

[11] Guo, X. P., Hernández-Lerma, O. and Prieto-Rumeau, T. (2006). A survey of resent results on continuous-time Markov decision processes. *TOP* **14,** 177–261.

[12] Hernández-Lerma, O. and Govindan, T. E. (2001). Nonstationary continuous-time Markov control processes with discounted costs on infinite horizon. *Acta Appl. Math.* **67,** 277–293.

[13] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes*. Springer, New York.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press

[14] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.

[15] Kitaev, M. Y. and Rykov, V. V.(1995). *Controlled Queueing Systems*. CRC Press, Boca Raton, FL.

[16] Lewis, M. E. and Puterman, M. L. (2000). A note on bias optimality in controlled queueing systems. *J. Appl. Prob.* **37,** 300–305.

[17] Lund, R. B., Meyn, S. P. and Tweedie, R. L. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. *Ann. Appl. Prob.* **6,** 218–237.

[18] Prieto-Rumeau, T. and Hernández-Lerma, O. (2006). Bias optimality for continuous-time controlled Markov chains. *SIAM J. Control Optimization* **45,** 51–73.

[19] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

[20] Sennott, L. I. (1999). *Stochastic Dynamic Programming and the Control of Queueing Systems*. John Wiley, New York.

[21] Wang, Z. K. and Yang, X. Q. (1992). *Birth and Death Processes and Markov Chains*. Science Press, Beijing.

[22] Ye, L., Guo, X. P. and Hernández-Lerma, O. (2008). Existence and regularity of a nonhomogeneous transition matrix under measurability conditions. *J. Theoret. Prob.* **21,** 604–627.

[23] Ye, L. E. and Guo, X. P. (2010). Construction and regularity of transition functions on Polish spaces under measurablity conditions. Submitted.

[24] Zhu, Q. X. (2007). Average optimality inequality for continuous-time Markov decision processes in Polish spaces. *Math. Meth. Operat. Res.* **66,** 299–313.

[25] Zhu, Q. X. (2008). Average optimality for continuous-time Markov decision processes with a policy iteration approach. *J. Math. Anal. Appl.* **339,** 691–704.

https://doi.org/10.1239/aap/1293113146 Published online by Cambridge University Press