Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures*

Jonatha Anselmi † François Dufour‡ Tomás Prieto-Rumeau§

September 29, 2015

Abstract

In this paper, we propose an approach for approximating the value function and computing an ϵ-optimal policy of a continuous-time Markov decision processes with Borel state and action spaces, with possibly unbounded cost and transition rates, under the total expected discounted cost optimality criterion. Under the assumptions that the controlled process satisfies a Lyapunov type condition and the transition rate has a density function with respect to a reference measure, together with piecewise Lipschitz continuity of the elements of the control model, one can approximate the original controlled process by a sequence of models that are computationally solvable. Convergence of the approximations takes place at an exponential rate in probability.

Keywords: Continuous-time Markov decision processes; Piecewise Lipschitz continuous control models; Approximation of the optimal value function; ϵ-optimal policy.
AMS 2010 Subject Classification: 90C40, 90C39.

1 Introduction

This paper is concerned with numerical methods for solving continuous-time Markov decision processes (CTMDPs). We are interested in approximating numerically the value function and in computing an ϵ-optimal policy of a general Markov decision process (MDP) with Borel state and action spaces, and possibly unbounded cost and transition rates, under the total expected discounted cost optimality criterion.

From a theoretical point of view, CTMDPs have been extensively studied. There exist two main techniques to analyze such optimization problems: the dynamic programming and the linear programming approaches. While these two methods are known to be very efficient for establishing different mathematical properties (such as existence of optimal policies, smoothness of the value function, sufficiency of sub-classes of particular policies, etc.) the problem of solving explicitly a CTMDP remains a critical issue. Indeed, except for very few specific models, the determination of an optimal policy and the value function is an extremely

*This research was supported by grant MTM2012-31393 from the Spanish Ministerio de Economía y Competitividad.

†INRIA Bordeaux Sud Ouest, France. e-mail: jonatha.anselmi@inria.fr

‡Institut Polytechnique de Bordeaux; INRIA Bordeaux Sud Ouest, Team: CQFD and IMB, Institut de Mathématiques de Bordeaux, Université de Bordeaux, France. e-mail: dufour@math.u-bordeaux1.fr

§UNED, Madrid, Spain. e-mail: tprieto@ccia.uned.es (Author for correspondence)

difficult problem to tackle. In this context, the standard approach for solving an MDP is to develop numerical methods to get quasi-optimal solutions. This topic is, therefore, of crucial importance to demonstrate the practical interest of CTMDP as a powerful modeling tool.

In the discrete-time framework, several techniques have been proposed to solve numerically an MDP, which can be classified into two groups. The first class is dedicated to the study of MDPs with discrete (large finite or countable) state and action spaces. These approaches are mainly related to stochastic approximation techniques such as reinforcement learning, neuro-dynamic programming, approximate dynamic programs, and simulation-based methods; see for example the survey [17] and the books [1, 2, 12, 16]. The second category is focused on general MDPs with uncountable Borel state and action spaces. For such models, the traditional approach is to approximate the original control problem by a controlled Markov chain with finite state and action spaces in such a way that the optimal cost and policies approximate those of the primary control model; see [3, 4, 5, 6, 15] and the references therein.

In the continuous-time context, the situation is radically different and there exist very few results on this topic. In [7, 13, 14], an approximation procedure in which a control model with denumerable state space and possibly unbounded transition rate is approximated by a sequence of auxiliary control models is studied. Conditions are provided on the sequence of approximating control models ensuring that the corresponding optimal discounted reward and optimal policies converge to the value function and the optimal policies of the original model. Such an approach can be found in [13] for unconstrained infinite-horizon discounted CTMDPs, in [7] for constrained CTMDPs, and in [14] for average reward continuous-time CTMDPs.

Our objective in this paper is to propose a method for approximating the value function and computing an ϵ-optimal policy of a continuous-time control model $\mathcal{M}$ with Borel state space $\mathbf{S}$ and action space $\mathbf{A}$, with possibly unbounded cost and transition rates, under the total expected discounted cost criterion. Therefore, we do not restrict our attention to the countable state space case. Our approach mainly follows two steps.

(1) The first one is related to the construction of a sequence of control models $\{\mathcal{M}_k\}_{k\geq 1}$ with bounded transition and reward rates. The dynamic of each model $\mathcal{M}_k$ is similar to that of $\mathcal{M}$ as long as the original process remains in a specific subset $\mathbf{S}_k$ of the state space (with the property that $\mathbf{S}_k \uparrow \mathbf{S}$), and it is absorbed at no future cost upon leaving the set $\mathbf{S}_k$. The construction of such models $\mathcal{M}_k$ is based on a Lyapunov type condition, which allows to control the approximation error when approximating $\mathcal{M}$ with $\mathcal{M}_k$.

(2) The second step consists in approximating the dynamic programming optimality equation of $\mathcal{M}_k$ by means of a discretization of the state and action spaces of $\mathcal{M}_k$. To do so, we assume that the positive part $q_k^+(dy|x,a)$ of the transition rate $q_k(dy|x,a)$ governing the dynamics of the control model $\mathcal{M}_k$ has a density function $p_k(y|x,a)$ with respect to a reference probability measure μ_k, that is, $q_k^+(dy|x,a) = p_k(y|x,a)\mu_k(dy)$. The idea is to approximate μ_k with its empirical distribution (for a sample of size n) and, in addition, to replace the action sets $\mathbf{A}(x)$ of the control model $\mathcal{M}_k$ with finite sets $\mathbf{A}_\delta(x)$ satisfying weak technical hypotheses guaranteeing that the Hausdorff distance between $\mathbf{A}(x)$ and $\mathbf{A}_\delta(x)$ is of (small) order $\delta > 0$.

Following these two steps, one can explicitly compute an approximation of the optimal value function of $\mathcal{M}$ and an ϵ-optimal policy, for a given precision $\epsilon > 0$. The accuracy of the approximation is characterized in terms of a concentration inequality, measuring the

non-asymptotic deviation between the value function of the original model $\mathcal{M}$ and its approximations (see Theorems 4.12 and 4.16). It is shown that the approximation errors converge in probability to zero, at an exponential speed in the sample size n.

Our main assumptions consist in supposing the existence of a strictly unbounded function w, which somehow bounds the transition and cost rates of the control model, and which also satisfies suitable Lyapunov type conditions. Moreover, we suppose that the elements of the control model (cost rate, densities, action sets multifunction, etc.) are piecewise Lipschitz continuous. This allows dealing with, e.g., discontinuous transition and cost rates, which is an important departure point from previous works; see, for instance, [4, 5, 6].

Finally, we would like to mention that the approaches developed in [4, 5, 6] cannot be used to approximate the value function of the control model $\mathcal{M}_k$. Indeed, the underlying stochastic kernel Q_k related to the optimality equation is absolutely continuous with respect to a probability measure which is state-dependent, that is $Q_k(dy|x,a) = r_k(y|x,a)\lambda_k(dy|x)$, ruling out one of the main condition of [4, 5, 6].

The rest of the paper is organized as follows. After introducing some notation in Section 1.1, we define the control model $\mathcal{M}$ and state our assumptions in Section 2. We study the approximation of $\mathcal{M}$ with bounded control models $\mathcal{M}_k$ in Section 3, while we show how to approximate the solution of $\mathcal{M}_k$ is Section 4. Finally, we study a numerical application in Section 5.

1.1 Notation

The set of nonnegative integers is $\mathbb{N} = \{0, 1, 2, \ldots\}$, while the real numbers set is $\mathbb{R}$. The subscript $_*$ and the superscript $^+$ will refer to the nonzero an nonnegative elements in the corresponding set, respectively. By $\overline{\mathbb{R}}$ we will denote the set of extended real numbers. Combinations of these indices will yield the corresponding sets. The symbols $\wedge$ and $\vee$ stand for "minimum" and "maximum", respectively.

Given a Borel space $\mathbf{Y}$ with metric $d_{\mathbf{Y}}$, its Borel σ-algebra will be denoted by $\mathcal{B}(\mathbf{Y})$. In this paper, measurability is always referred to the Borel σ-algebra. The indicator function of a set B will be denoted by $\mathbf{I}_B$.

The family of bounded and measurable real-valued functions on $\mathbf{Y}$ is denoted by $\mathbb{B}(\mathbf{Y})$. The supremum norm of $v \in \mathbb{B}(\mathbf{Y})$ is $||v||$. We say that the function $v : \mathbf{Y} \to \mathbf{Z}$, on the Borel space $\mathbf{Z}$ with metric $d_{\mathbf{Z}}$, is Lipschitz continuous if there exists some constant $L^v \geq 0$ such that

$$d_{\mathbf{Z}}(v(y), v(y')) \leq L^v \cdot d_{\mathbf{Y}}(y, y') \quad \text{for all } y, y' \in \mathbf{Y},$$

and we say that v is L^v-Lipschitz continuous. The function $v : \mathbf{Y} \to \mathbf{Z}$ is piecewise Lipschitz continuous if there exist disjoint measurable sets $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ such that $\cup_{j=1}^m \mathbf{Y}_j = \mathbf{Y}$ and such that the restriction of v to each set $\mathbf{Y}_j$ is Lipschitz continuous: for all $1 \leq j \leq m$ there exists some constant $L_j^v \geq 0$ such that

$$d_{\mathbf{Z}}(v(y), v(y')) \leq L_j^v \cdot d_{\mathbf{Y}}(y, y') \quad \text{for all } y, y' \in \mathbf{Y}_j.$$

In general, for a piecewise Lipschitz continuous function v on $\mathbf{Y}$ for the partition $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$, we will denote by L_j^v, for $1 \leq j \leq m$, any constant $L_j^v \geq 0$ satisfying the above inequality. The family of real-valued bounded and piecewise Lipschitz continuous functions on $\mathbf{Y}$, for the partition $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$, is denoted by $\mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$.

On the family of nonempty closed subsets of $\mathbf{Z}$ we will consider the Hausdorff metric defined as

$$d_H(C_1, C_2) = \sup_{z_1 \in C_1} \inf_{z_2 \in C_2} \{d_{\mathbf{Z}}(z_1, z_2)\} \vee \sup_{z_2 \in C_2} \inf_{z_1 \in C_1} \{d_{\mathbf{Z}}(z_1, z_2)\}.$$

It is known that d_H is a metric except that it might not be finite. Lipschitz continuity and piecewise Lipschitz continuity of a closed-valued multifunction ψ from $\mathbf{Y}$ to $\mathbf{Z}$ are defined accordingly. In particular, the constants $L_j^{\psi} \geq 0$ are given a definition similar to the L_j^v for a piecewise Lipschitz continuous multifunction ψ.

The family of probability measures on $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$ is $\mathcal{P}(\mathbf{Y})$. Given $y \in \mathbf{Y}$, the Dirac probability measure at y will be denoted by δ_y. The family of probability measures $\mu \in \mathcal{P}(\mathbf{Y})$ with finite first moment (that is, $\int_{\mathbf{Y}} d_{\mathbf{S}}(y, y')\mu(dy) < \infty$ for some, and then for all, $y' \in \mathbf{S}$) is denoted by $\mathcal{P}_1(\mathbf{Y})$. The 1-Wasserstein distance between $\mu, \nu \in \mathcal{P}_1(\mathbf{Y})$ (also referred to as the Kantorovich-Rubinshtein metric) is defined as

$$\mathcal{W}_1(\mu, \nu) = \sup \Big\{ \int_{\mathbf{Y}} f d\mu - \int_{\mathbf{Y}} f d\nu \Big\}$$

where the supremum ranges over all functions $f : \mathbf{Y} \to \mathbb{R}$ which are 1-Lipschitz continuous. We say that the probability measure $\mu \in \mathcal{P}(\mathbf{Y})$ has a finite exponential moment if there exists $\gamma > 0$ such that

$$\int_{\mathbf{Y}} \exp\{\gamma d_Y(y, y_0)\}\mu(dy) < \infty$$

for some, and then for all, $y_0 \in \mathbf{Y}$. Let $\mathcal{P}_{\exp}(\mathbf{Y})$ be the family of such probability measures with, clearly, $\mathcal{P}_{\exp}(\mathbf{Y}) \subseteq \mathcal{P}_1(\mathbf{Y})$.

Let us now recall some facts on the convergence of the empirical probability measures. Suppose that $\mu \in \mathcal{P}_{\exp}(\mathbf{Y})$. Let $\zeta_1, \zeta_2, \ldots$ be random variables defined on some probability space $(\Theta, \mathcal{F}_\Theta, \mathbb{P})$ taking values in $\mathbf{Y}$ and such that the ζ_i are i.i.d. with distribution μ. The empirical probability measure μ_n is a random probability measure in $\mathcal{P}_1(\mathbf{Y})$ defined as

$$\mu_n(dy; \theta) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\zeta_i(\theta)}(dy) \quad \text{for } \theta \in \Theta.$$

In the sequel we will use the following convergence result, which yields the speed of convergence (in probability) of μ_n to μ in the 1-Wasserstein metric.

Proposition 1.1 *Let $\mathbf{Y}$ be a Borel space and μ a probability measure in $\mathcal{P}_{\exp}(\mathbf{Y})$. Let μ_n, for each $n \geq 1$, be the empirical probability measure for a sample of size n. For every $\epsilon > 0$ there exist positive constants C and D (depending on ϵ but not on $n \geq 1$) such that*

$$\mathbb{P}\{\mathcal{W}_1(\mu, \mu_n) \geq \epsilon\} \leq Ce^{-Dn} \quad \textit{for all } n \geq 1.$$

We say that $Q : \mathcal{B}(\mathbf{Y}) \times \mathbf{Z} \to \overline{\mathbb{R}}^+$ is a transition measure on the Borel space $\mathbf{Y}$ given the Borel space $\mathbf{Z}$ if $B \mapsto Q(B|z)$ is a (nonnegative) measure on $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$ for all $z \in \mathbf{Z}$ and $z \mapsto Q(B|z)$ is measurable for every $B \in \mathcal{B}(\mathbf{Y})$. For measurable $v : \mathbf{Y} \to \mathbb{R}$, we will denote by Qv the function on $\mathbf{Z}$ defined as

$$Qv(z) = \int_{\mathbf{Y}} v(y)Q(dy|z). \tag{1.1}$$

2 The control model $\mathcal{M}$: definition and assumptions

The main goal of this section is to introduce the notation, the parameters defining the model, and to present the construction of the controlled process. In particular, we construct a canonical measure space $(\Omega, \mathcal{F})$ consisting of the sample paths of the multivariate point process (Θ_n, X_n). Having defined the class of admissible strategies, we show the existence of a probability measure with respect to which the controlled process (Θ_n, X_n) has the required conditional distributions.

2.1 Elements of the control model $\mathcal{M}$

We deal with a control model $\mathcal{M} = \{\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x)\}_{x\in\mathbf{X}}, q, c\}$ with the following elements:

- The state space $\mathbf{X}$ is a Borel space with metric $d_{\mathbf{X}}$.
- The action space $\mathbf{A}$ is a Borel space with metric $d_{\mathbf{A}}$. The set of feasible actions in state $x \in \mathbf{X}$ is $\mathbf{A}(x)$, which is a nonempty measurable subset of $\mathbf{A}$. The set of admissible state-action pairs is

  $$\mathbf{K} = \{(x, a) \in \mathbf{X} \times \mathbf{A} : a \in \mathbf{A}(x)\} \in \mathcal{B}(\mathbf{X} \times \mathbf{A}).$$

  It is assumed that $\mathbf{K}$ contains the graph of a measurable function from $\mathbf{X}$ to $\mathbf{A}$. The multifunction from $\mathbf{X}$ to $\mathbf{A}$ given by $x \mapsto \mathbf{A}(x)$ will be denoted by Ψ.
- The transition rate q is a signed kernel on $\mathbf{X}$ given $\mathbf{K}$. This means that $B \mapsto q(B|x, a)$ is a signed measure on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ for all $(x, a) \in \mathbf{K}$, and that $(x, a) \mapsto q(B|x, a)$ is measurable for all $B \in \mathcal{B}(\mathbf{X})$. It satisfies $q(B|x, a) \geq 0$ for all $B \in \mathcal{B}(\mathbf{X})$ such that $x \notin B$. It is conservative, i.e.,

  $$q(\mathbf{X}|x, a) = 0 \quad \text{for all } (x, a) \in \mathbf{K},$$

  and stable, that is,

  $$\overline{q}(x) = \sup_{a\in\mathbf{A}(x)} \{-q(\{x\}|x, a)\} < \infty \quad \text{for any } x \in \mathbf{X}.$$
- The cost rate is the measurable function $c : \mathbf{K} \to \mathbb{R}$.

We will use the following notations. Denote by q^+ the transition measure on $\mathbf{X}$ given $\mathbf{K}$ defined, for $B \in \mathcal{B}(\mathbf{X})$ and $(x, a) \in \mathbf{K}$, as $q^+(B|x, a) = q(B - \{x\}|x, a)$. We can write, equivalently,

$$q(dy|x, a) = q^+(dy|x, a) + q(\{x\}|x, a)\delta_x(dy). \tag{2.1}$$

Also, given a set $\mathbf{Y} \in \mathcal{B}(\mathbf{X})$ we define the set $\mathrm{Gr}(\mathbf{Y}) \in \mathcal{B}(\mathbf{X} \times \mathbf{A})$ as

$$\mathrm{Gr}(\mathbf{Y}) = \{(x, a) \in \mathbf{X} \times \mathbf{A} : x \in \mathbf{Y}, a \in \mathbf{A}(x)\}.$$

In particular, we have that $\mathrm{Gr}(\mathbf{X}) = \mathbf{K}$. Finally, in the next definition we introduce some terminology.

Definition 2.1 *If* $\sup_{x\in\mathbf{X}} \overline{q}(x)$ *and* $\sup_{(x,a)\in\mathbf{K}} |c(x, a)|$ *are both finite then we say that the control model* $\mathcal{M}$ *is bounded.*

2.2 Construction of the process

Let $\mathbf{X}_\infty = \mathbf{X} \cup \{x_\infty\}$, where x_∞ is an isolated point. We put

$$\Omega_n = \mathbf{X} \times (\mathbb{R}_+^* \times \mathbf{X})^n \times \big(\{\infty\} \times \{x_\infty\}\big)^\infty.$$

The canonical space, denoted by Ω, is defined as

$$\Omega = \big(\mathbf{X} \times (\mathbb{R}_+^* \times \mathbf{X})^\infty\big) \ \cup \ \bigcup_{n=1}^{\infty} \Omega_n$$

and it is endowed with its Borel σ-algebra, denoted by $\mathcal{F}$. For notational convenience, $\omega \in \Omega$ will be written

$$\omega = (x_0, \theta_1, x_1, \theta_2, x_2, \ldots).$$

The canonical space Ω is given the following interpretation. Let $x_0 \in \mathbf{X}$ be the initial state of the dynamic system. Given $n \geq 0$, if $x_n \in \mathbf{X}$ then

- either $0 < \theta_{n+1} < \infty$, and we interpret θ_{n+1} as the sojourn time in state $x_n \in \mathbf{X}$, while $x_{n+1} \in \mathbf{X}$ is the post-jump location of the process;
- or $\theta_{n+1} = \infty$; this means that the dynamic system has been absorbed by x_n. In this case we let $x_m = x_\infty$ and $\theta_m = \infty$ for all $m > n$. Such sample paths belong to Ω_n.

For every $n \in \mathbb{N}$ and $\omega \in \Omega$, let

$$h_n = (x_0, \theta_1, x_1, \theta_2, x_2, \ldots \theta_n, x_n)$$

be the path up to n (we do not make ω explicit in the notation), and denote the collection of all such paths by $\mathbf{H}_n$.

For $n \in \mathbb{N}$, define the mapping $X_n : \Omega \to \mathbf{X}_\infty$ as $X_n(\omega) = x_n$. For $n \geq 1$, define Θ_n and T_n from Ω to $\overline{\mathbb{R}}_+^*$ as

$$\Theta_n(\omega) = \theta_n \quad \text{and} \quad T_n(\omega) = \theta_1 + \ldots + \theta_n.$$

We make the convention that $\Theta_0(\omega) = T_0(\omega) = 0$ for all $\omega \in \Omega$. Define also $T_\infty(\omega) = \lim_{n\to\infty} T_n(\omega)$. The random variable T_∞ is referred to as the explosion time of the process. We denote by $H_n = (X_0, \Theta_1, X_1, \ldots, \Theta_n, X_n)$ the n-term history process, which takes values in $\mathbf{H}_n$ for $n \in \mathbb{N}$.

The random measure μ associated with $(\Theta_n, X_n)_{n\in\mathbb{N}}$ is a measure defined on $\mathbb{R}_+^* \times \mathbf{X}$ by

$$\mu(\omega; dt, dx) = \sum_{n\geq 1} I_{\{T_n(\omega)<\infty\}} \delta_{(T_n(\omega), X_n(\omega))}(dt, dx).$$

Informally, we can say that $\mu(\omega; dt, dx)$ puts a mass equal to one in each pair $(\theta_1 + \ldots + \theta_n, x_n)$ provided that all $\theta_1, \ldots, \theta_n$ are finite. For notational convenience the dependence on ω will be suppressed and we will simply write $\mu(dt, dx)$. Define

$$\mathcal{F}_t = \sigma\{H_0\} \vee \sigma\big\{\mu\big((0, s] \times B\big) : \ s \leq t, \ B \in \mathcal{B}(\mathbf{X})\big\} \quad \text{for } t \geq 0.$$

Finally, the continuous-time process $\{\xi_t\}_{t\geq 0}$ with values in $\mathbf{X}_\infty$ is given by

$$\xi_t(\omega) = \begin{cases} X_n(\omega), & \text{if } T_n(\omega) \leq t < T_{n+1}(\omega) \text{ for } n \in \mathbb{N}, \\ x_\infty, & \text{if } t \geq T_\infty(\omega). \end{cases}$$

Obviously, the process $\{\xi_t\}_{t\geq 0}$ can be equivalently described by the sequence $(\Theta_n, X_n)_{n\in\mathbb{N}}$.

2.3 Admissible policies and distribution of the controlled process

Define $\mathbf{A}_\infty = \mathbf{A} \cup \{a_\infty\}$, where a_∞ is an isolated action associated to the cemetery state x_∞, and let $\mathbf{A}(x_\infty) = \{a_\infty\}$. We can extend the transition rate q to be a signed kernel on $\mathbf{X}_\infty$ given $\mathbf{K} \cup \{(x_\infty, a_\infty)\}$ by letting $q(\{x_\infty\}|x,a) = 0$ for all $(x,a) \in \mathbf{K}$ and $q(\cdot|x_\infty, a_\infty) \equiv 0$. We define q^+ as in (2.1).

An admissible control policy is a sequence $u = (\pi_n)_{n\in\mathbb{N}}$ where, for any $n \in \mathbb{N}$, π_n is a stochastic kernel (or transition probability measure) on $\mathbf{A}_\infty$ given $\mathbf{H}_n \times \mathbb{R}_+^*$ satisfying

$$\pi_n(\mathbf{A}(x_n)|h_n, t) = 1 \quad \text{for any } h_n = (x_0, \theta_1, x_1, \ldots \theta_n, x_n) \in \mathbf{H}_n \text{ and } t \in \mathbb{R}_+^*.$$

The set of admissible control policies is denoted by $\mathcal{U}$.

Given an admissible control policy $u = (\pi_n)_{n\in\mathbb{N}}$, we denote by π the random process with values in $\mathcal{P}(\mathbf{A}_\infty)$ as

$$\pi(da|t) = \sum_{n\in\mathbb{N}} I_{\{T_n < t \le T_{n+1}\}} \pi_n(da|H_n, t - T_n) + \mathbf{I}_{\{t \ge T_\infty\}} \delta_{a_\infty}(da) \tag{2.2}$$

for $t > 0$. We have that π is an $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$-predictable random process with values in $\mathcal{P}(\mathbf{A}_\infty)$.

Suppose that a control policy $u = (\pi_n)_{n\in\mathbb{N}} \in \mathcal{U}$ is fixed. We introduce the intensity of the jumps

$$\lambda_n(\Gamma, h_n, t) \;=\; \int_{\mathbf{A}_\infty} q^+(\Gamma|x_n, a)\pi_n(da|h_n, t),$$

and the rate of the natural jumps

$$\Lambda_n(\Gamma, h_n, t) \;=\; \int_0^t \lambda_n(\Gamma, h_n, s)ds$$

for any $n \in \mathbb{N}$, $\Gamma \in \mathcal{B}(\mathbf{X}_\infty)$, $h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \mathbf{H}_n$, and $t \in \overline{\mathbb{R}}_+^*$. Now, for any $n \in \mathbb{N}$, the stochastic kernel G_n on $\overline{\mathbb{R}}_+^* \times \mathbf{X}_\infty$ given $\mathbf{H}_n$ is defined by

$$\begin{aligned} G_n(\Gamma|h_n) \;=\; & \delta_{(\infty, x_\infty)}(\Gamma)\Big[\delta_{x_n}(\{x_\infty\}) + \delta_{x_n}(\mathbf{X})e^{-\Lambda_n(\mathbf{X}, h_n, \infty)}\Big] \\ & + \delta_{x_n}(\mathbf{X}) \int_{\Gamma \cap (\mathbb{R}_+^* \times \mathbf{X})} \lambda_n(dx, h_n, t)e^{-\Lambda_n(\mathbf{X}, h_n, t)}dt, \end{aligned}$$

for any $\Gamma \in \mathcal{B}(\overline{\mathbb{R}}_+^* \times \mathbf{X}_\infty)$ and $h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \mathbf{H}_n$.

Consider an admissible policy $u \in \mathcal{U}$ and an initial state $x \in \mathbf{X}$. From Remark 3.43 in [9], there exists a probability $\mathbb{P}_x^u$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}_x^u\{X_0 = x\} = 1$$

and such that, for $\Gamma \in \mathcal{B}(\overline{\mathbb{R}}_+^* \times \mathbf{X}_\infty)$ and $n \ge 0$,

$$\mathbb{P}_x^u\{(\Theta_{n+1}, X_{n+1}) \in \Gamma \mid H_n\} = G_n(\Gamma|H_n)$$

almost surely. Denote by $\mathbb{E}_x^u$ the expectation operator associated to $\mathbb{P}_x^u$.

Remark 2.2 *Observe that $\mathcal{F}_{T_n}$ is the σ-algebra generated by the random variable H_n for $n \in \mathbb{N}$. The conditional distribution of (Θ_{n+1}, X_{n+1}) given $\mathcal{F}_{T_n}$ under $\mathbb{P}_x^u$ is determined by $G_n(\cdot|H_n)$ and the conditional survival function of Θ_{n+1} given $\mathcal{F}_{T_n}$ under $\mathbb{P}_x^u$ is given by $G_n([t, +\infty] \times \mathbf{X}_\infty|H_n)$.*

2.4 Optimality criterion

Now we introduce the infinite-horizon performance criterion we are concerned with. Recall that c is the cost rate function and suppose that a discount rate $\alpha > 0$ is fixed. Given an admissible control policy $u \in \mathcal{U}$ the corresponding total expected discounted cost for the initial state $x \in \mathbf{X}$ up to explosion time is defined as

$$\mathcal{V}(u,x) = \mathbb{E}_x^u\Big[\int_0^{T_\infty} e^{-\alpha s}\int_{\mathbf{A}(\xi_s)} c(\xi_s,a)\pi(da|s)ds\Big], \tag{2.3}$$

with π as in (2.2).

Definition 2.3 *Given an initial state $x \in \mathbf{X}$, the optimization problem consists in minimizing the performance criterion $\mathcal{V}(u,x)$ within the class of admissible strategies $u \in \mathcal{U}$. We define the value function*

$$\mathcal{V}^*(x) = \inf_{u\in\mathcal{U}} \mathcal{V}(u,x) \quad \textit{for } x \in \mathbf{X}.$$

A control strategy $u \in \mathcal{U}$ is called

- *stationary* if $\pi_n(\cdot|h_n,t) = \varphi(\cdot|x_n)$, where φ is a stochastic kernel on $\mathbf{A}_\infty$ given $\mathbf{X}_\infty$ satisfying $\varphi(\mathbf{A}(y)|y) = 1$ for any $y \in \mathbf{X}_\infty$. Let $\mathcal{U}^s$ be the class of stationary policies.
- *deterministic stationary* if $\pi_n(\cdot|h_n,t) = \delta_{\varphi(x_n)}(\cdot)$ where $\varphi : \mathbf{X}_\infty \to \mathbf{A}_\infty$ is a measurable mapping satisfying $\varphi(y) \in \mathbf{A}(y)$ for any $y \in \mathbf{X}_\infty$.

The set of deterministic stationary policies is nonempty by hypothesis and, therefore, so is $\mathcal{U}^s$. For a stationary policy $u \in \mathcal{U}^s$, the expression (2.3) becomes

$$\mathcal{V}(u,x) = \mathbb{E}_x^u\Big[\int_0^{T_\infty} e^{-\alpha s}\int_{\mathbf{A}(\xi_s)} c(\xi_s,a)\varphi(da|\xi_s)ds\Big],$$

with φ the kernel associated to u.

2.5 Assumptions and basic results

In this section we state our main assumptions on the control model $\mathcal{M}$. Under these assumptions, we will have that the controlled process $\{\xi_t\}_{t\geq 0}$ is nonexplosive with probability one under any control policy $u \in \mathcal{U}$ and for any initial state $x \in \mathbf{X}$, and a so-called Dynkin formula will be satisfied. We prove a preliminary result that will be useful in the sequel.

Lemma 2.4 *Let $h : \mathbf{X} \to [1,\infty)$ be a measurable function and suppose that there exist constants $c, d \geq 0$ such that*

$$\int_{\mathbf{X}} h(y)q(dy|x,a) \leq ch(x) + d \quad \textit{for all } (x,a) \in \mathbf{K}. \tag{2.4}$$

Under these conditions, given some power $0 < \gamma < 1$ we have

$$\int_{\mathbf{X}} h^\gamma(y)q(dy|x,a) \leq c\gamma h^\gamma(x) + d\gamma \quad \textit{for all } (x,a) \in \mathbf{K}.$$

Proof. Fix $(x,a) \in \mathbf{K}$ and $\epsilon > 0$. Define the probability measure $p_\epsilon(\cdot|x,a)$ on $\mathbf{X}$ as

$$p_\epsilon(dy|x,a) = \delta_x(dy) + \frac{q(dy|x,a)}{\overline{q}(x)+\epsilon}.$$

The inequality (2.4) can be equivalently written as

$$\int_{\mathbf{X}} h(y) p_\epsilon(dy|x,a) \leq h(x) + \frac{ch(x)+d}{\overline{q}(x)+\epsilon}.$$

Apply now Jensen's inequality to the increasing concave function $z \mapsto z^\gamma$, defined on $[0,\infty)$, to obtain

$$\int_{\mathbf{X}} h^\gamma(y) p_\epsilon(dy|x,a) \leq \Big(\int_{\mathbf{X}} h(y) p_\epsilon(dy|x,a) \Big)^\gamma \leq \Big(h(x) + \frac{ch(x)+d}{\overline{q}(x)+\epsilon} \Big)^\gamma.$$

In particular, we have

$$\frac{1}{\overline{q}(x)+\epsilon} \int_X h^\gamma(y) q(dy|x,a) \leq \Big(h(x) + \frac{ch(x)+d}{\overline{q}(x)+\epsilon} \Big)^\gamma - h^\gamma(x).$$

In the righthand side of this inequality, use $(t+s)^\gamma \leq t^\gamma + \gamma t^{\gamma-1} s$, which holds for every $t > 0$ and $s \geq 0$, to obtain

$$\int_{\mathbf{X}} h^\gamma(y) q(dy|x,a) \leq \gamma h^{\gamma-1}(x) \cdot \big(ch(x) + d \big) \leq c\gamma h^\gamma(x) + d\gamma,$$

where we use the fact that $h(x) \geq 1$. This completes the proof. □

Next, we introduce some assumptions on the control model. In Assumption (A2) below, recall that $\alpha > 0$ is the discount rate.

Assumption A. There exist a measurable function $w : \mathbf{X} \to [1,\infty)$ and an increasing sequence $\{\mathbf{S}_k\}_{k\in\mathbb{N}}$ of measurable subsets of $\mathbf{X}$, with $\mathbf{X} = \cup_{k\in\mathbb{N}} \mathbf{S}_k$, such that:

(A1) For every $k \in \mathbb{N}$ we have $\mathbf{w}_k = \sup_{x\in\mathbf{S}_k} w(x) < \infty$, and

$$\lim_{k\to\infty} \inf\{w(x) : x \in \mathbf{X} - \mathbf{S}_k\} = \infty.$$

(A2) There exist constants $0 \leq \rho < \alpha$, $b \geq 0$, and $\beta > 0$ such that

$$\int_{\mathbf{X}} w^{1+\beta}(y) q(dy|x,a) \leq \rho w^{1+\beta}(x) + b \quad \text{for any } (x,a) \in \mathbf{K}.$$

(A3) There exists a constant $L > 0$ with $\overline{q}(x) \leq L w(x)$ for all $x \in \mathbf{X}$.

(A4) There exists a constant $M > 0$ such that $|c(x,a)| \leq M w(x)$ for each $(x,a) \in \mathbf{K}$.

(A5) There exists a nonnegative measurable function w' on $\mathbf{X}$, and nonnegative constants L', ρ', b' such that

$$\overline{q}(x) w^{1+\beta}(x) \leq L' w'(x) \quad \text{and} \quad \int_{\mathbf{X}} w'(y) q(dy|x,a) \leq \rho' w'(x) + b' \quad \text{for all } (x,a) \in \mathbf{K}.$$

In Assumption (A1) note that if $\mathbf{X} = \mathbf{S}_k$ then the infimum is computed over an empty set, and we adopt the usual convention that it equals ∞. As a consequence of Lemma 2.4 (just let $\gamma = (1+\beta)^{-1}$), Assumption (A2) implies the following inequality:

$$\int_{\mathbf{X}} w(y)q(dy|x,a) \leq \rho w(x) + b \quad \text{for every } (x,a) \in \mathbf{K}. \tag{2.5}$$

We note that, in many references dealing with continuous-time MDPs, the usual requirement is that the function w satisfies precisely the inequality (2.5) above. In this paper, as we shall see, the condition (2.5) does not suffice for our purposes and we need to impose a stronger condition, namely, Assumption (A2).

Definition 2.5 *Given a measurable function $h : \mathbf{X} \to [1,\infty)$, the family of measurable functions $v : \mathbf{X} \to \mathbb{R}$ such that*

$$||v||_h = \sup_{x\in\mathbf{X}} \big\{|v(x)|/h(x)\big\} < \infty$$

will be denoted by $\mathbb{B}_h(\mathbf{X})$.

In this paper, we shall deal with the families $\mathbb{B}_w(\mathbf{X})$ and $\mathbb{B}_{w^{1+\beta}}(\mathbf{X})$, with the function w and the constant $\beta > 0$ as in Assumption A.

Theorem 2.6 *Suppose that Assumptions (A1)–(A3) are satisfied.*

(i) For every initial state $x \in \mathbf{X}$ and every control policy $u \in \mathcal{U}$ we have $\mathbb{P}_x^u\{T_\infty = \infty\} = 1$. Therefore, explosion does not occur and $\mathbb{P}_x^u\{\xi_t \in \mathbf{X} \text{ for all } t \geq 0\} = 1$.

(ii) Given $u \in \mathcal{U}$, $x \in \mathbf{X}$, and $t \geq 0$

$$\mathbb{E}_x^u[w(\xi_t)] \leq e^{\rho t}w(x) + \frac{b}{\rho}\big(e^{\rho t} - 1\big) \quad \text{and} \quad \mathbb{E}_x^u[w^{1+\beta}(\xi_t)] \leq e^{\rho t}w^{1+\beta}(x) + \frac{b}{\rho}\big(e^{\rho t} - 1\big),$$

or $\mathbb{E}_x^u[w(\xi_t)] \leq w(x) + bt$ and $\mathbb{E}_x^u[w^{1+\beta}(\xi_t)] \leq w^{1+\beta}(x) + bt$ when $\rho = 0$.

(iii) If in addition Assumption (A4) holds then, for any $u \in \mathcal{U}$ and $x \in \mathbf{X}$, the total expected discounted cost $\mathcal{V}(u,x)$ is

$$\mathcal{V}(u,x) = \mathbb{E}_x^u\Big[\int_0^\infty e^{-\alpha s}\int_{\mathbf{A}(\xi_s)} c(\xi_s,a)\pi(da|s)ds\Big]$$

and it satisfies

$$|\mathcal{V}(u,x)| \leq \mathbf{m}Mw(x) \quad \text{with } \mathbf{m} = \tfrac{b+\alpha}{\alpha(\alpha-\rho)}.$$

Moreover, for every initial state $x \in \mathbf{X}$,

$$\mathcal{V}^*(x) = \inf_{u\in\mathcal{U}} \mathcal{V}(u,x) = \inf_{u\in\mathcal{U}^s} \mathcal{V}(u,x),$$

and so stationary policies are a sufficient class.

(iv) Suppose that Assumptions (A1)–(A3) and (A5) are satisfied. Given $v \in \mathbb{B}_{w^{1+\beta}}(\mathbf{X})$, a stationary policy $u \in \mathcal{U}^s$ defined by the stochastic kernel $\varphi(da|x)$ on $\mathbf{A}$ given $\mathbf{X}$, and an initial state $x \in \mathbf{X}$, the process

$$M_t^u = e^{-\alpha t} v(\xi_t) - \int_0^t e^{-\alpha s} \Big[-\alpha v(\xi_s) + \int_{\mathbf{A}(\xi_s)} \int_{\mathbf{X}} v(y) q(dy|\xi_s, a) \varphi(da|\xi_s) \Big] ds$$

is a $\{\mathcal{F}_t\}$-martingale. Hence, we have

$$\mathbb{E}_x^u[e^{-\alpha\tau} v(\xi_\tau)] - v(x) = \mathbb{E}_x^u \left[\int_0^\tau e^{-\alpha s} \Big[-\alpha v(\xi_s) + \int_{\mathbf{A}(\xi_s)} \int_{\mathbf{X}} v(y) q(dy|\xi_s, a) \varphi(da|\xi_s) \Big] ds \right]$$

for any bounded stopping time τ for $\{\mathcal{F}_t\}_{t \geq 0}$.

Proof. Statements (i) and (ii) follow from Proposition 2.1 in [10] or Theorem 1 in [11]. The martingale property for $\{M_t^u\}$ follows from Dynkin's formula in Theorem 3 in [11] and the fact that u is stationary. The optional sampling theorem can be used because $\{M_t^u\}$ has càdlàg paths.

Remark 2.7 *If the control model $\mathcal{M}$ is bounded then Assumption A is satisfied. To see this, put $w \equiv w' \equiv 1$ and $S_0 = \mathbf{X}$, and let $\rho = 0$. Therefore, Theorem 2.6 holds for a bounded control model.*

We impose another assumption on the control model $\mathcal{M}$. In Assumption B below, the sets $\mathbf{S}_k$ are taken from Assumption A. We use the notion of piecewise Lipschitz continuity which was introduced in Section 1.1.

Assumption B.

(B1) For all $x \in \mathbf{X}$ the action set $\mathbf{A}(x)$ is compact.

(B2) There exists a measure μ on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ and a nonnegative measurable function p on $\mathbf{X} \times \mathbf{K}$ such that $q^+(dy|x,a) = p(y|x,a)\mu(dy)$ for all $(x,a) \in \mathbf{K}$ (recall (2.1)).

(B3) The measure μ satisfies that for every $k \in \mathbb{N}$ there exist $\gamma > 0$ and $x_0 \in \mathbf{X}$ such that $\int_{\mathbf{S}_k} \exp\{\gamma d_{\mathbf{X}}(x, x_0)\} \mu(dx)$ is finite.

(B4) For each $k \in \mathbb{N}$ there are nonempty measurable disjoint sets $\mathbf{Z}_1, \ldots, \mathbf{Z}_r$ with $\cup_{i=1}^r \mathbf{Z}_i = \mathbf{S}_k$ such that

- *the restriction of p to $\mathbf{S}_k \times \mathrm{Gr}(\mathbf{S}_k)$ is in $\mathbb{L}(\mathbf{Z}_i \times \mathrm{Gr}(\mathbf{Z}_j)); i, j = 1, \ldots, r)$;*
- *the function $(x,a) \mapsto q(\{x\}|x,a)$ is in $\mathbb{L}(\mathrm{Gr}(\mathbf{Z}_1), \ldots, \mathrm{Gr}(\mathbf{Z}_r))$;*
- *the restriction of the cost rate function c to $\mathrm{Gr}(\mathbf{S}_k)$ is in $\mathbb{L}(\mathrm{Gr}(\mathbf{Z}_1), \ldots, \mathrm{Gr}(\mathbf{Z}_r))$;*
- *the multifunction $x \mapsto \mathbf{A}(x)$ is piecewise Lipschitz continuous on $\mathbf{S}_k$ for the partition $\mathbf{Z}_1, \ldots, \mathbf{Z}_r$, with Lipschitz constants denoted by L_j^k for $1 \leq j \leq r$.*

(B5) For every $\delta > 0$ and $k \in \mathbb{N}$, there exist nonempty finite sets $\mathbf{A}_\delta(x) \subseteq \mathbf{A}(x)$, for $x \in \mathbf{S}_k$, such that

(1) $d_H(\mathbf{A}_\delta(x), \mathbf{A}(x)) \leq \delta$ for all $x \in \mathbf{S}_k$, and

(2) the multifunction $x \mapsto \mathbf{A}_\delta(x)$ is piecewise Lipschitz continuous on $\mathbf{S}_k$ for the partition $\mathbf{Z}_1, \ldots, \mathbf{Z}_r$, with the Lipschitz constants L_j^k for $1 \leq j \leq r$ as in Assumption (B4).

Observe that, under Assumption (B1), it is always possible to find finite sets $\mathbf{A}_\delta(x)$ with the property given in part (1) in Assumption (B5). Therefore, the important fact is that we can choose finite sets as in (1) which satisfy, in addition, (2). Note also that we must necessarily have $p(x|x,a)\cdot\mu\{x\}=0$ for all $(x,a)\in\mathbf{K}$ because $q^+(\{x\}|x,a)=0$.

3 Approximation of $\mathcal{M}$ by bounded control models

Under our hypotheses, we will show that the control model $\mathcal{M}$ can be approximated by a sequence of *bounded* control models $\{\mathcal{M}_k\}_{k\geq 1}$. As we shall explain, such approximation should be understood as approximation of the value function and ϵ-optimal policies.

3.1 Definition of the control models $\mathcal{M}_k$

The control models $\mathcal{M}_k$ defined next will be constructed by "killing" the process $\{\xi_t\}_{t\geq 0}$ when it leaves the set $\mathbf{S}_k$, which has the property that the cost incurred outside $\mathbf{S}_k$ is small.

Definition 3.1 *Suppose that the control model $\mathcal{M}$ satisfies Assumption A and fix $k\in\mathbb{N}$. The elements of the control model $\mathcal{M}_k$ are $\{\mathbf{X}_k,\mathbf{A},\{\mathbf{A}(x)\}_{x\in\mathbf{X}_k},q_k,c_k\}$, and we distinguish two cases:*

(i) *If $\mathbf{X}-\mathbf{S}_k$ is nonempty then choose $x_\Delta\in\mathbf{X}-\mathbf{S}_k$ and let $\mathbf{X}_k=\mathbf{S}_k\cup\{x_\Delta\}$. The transition rates are defined as follows: given $B\in\mathcal{B}(\mathbf{X}_k)$, let*

$$q_k(B|x,a)=q(B\cap\mathbf{S}_k|x,a)+q(\mathbf{X}-\mathbf{S}_k|x,a)\mathbf{I}_B(x_\Delta)\quad \text{for } (x,a)\in \mathrm{Gr}(\mathbf{S}_k),$$

and let $q_k(B|x_\Delta,a)=0$ for $a\in\mathbf{A}(x_\Delta)$. Let $c_k(x,a)=c(x,a)\mathbf{I}_{\mathbf{S}_k}(x)$ for $(x,a)\in\mathrm{Gr}(\mathbf{X}_k)$.

(ii) *If $\mathbf{S}_k=\mathbf{X}$, then $\mathcal{M}_k$ is defined as $\mathcal{M}$ itself.*

Observe that, in the case described in Definition 3.1(i), the controlled system $\mathcal{M}_k$ "behaves" as the original control model $\mathcal{M}$ as long as it remains in $\mathbf{S}_k$. Upon leaving $\mathbf{S}_k$, the process is absorbed, at no future cost, by the state $x_\Delta\notin\mathbf{S}_k$. The notation $q_k^+(B|x,a)$ is defined as q^+ in (2.1) but now for the control model $\mathcal{M}_k$, and so we add the index k. Also note that the multifunction Ψ_k from $\mathbf{X}_k$ to $\mathbf{A}$ that defines the actions sets of $\mathcal{M}_k$ is just the restriction of Ψ to $\mathbf{X}_k$.

Proposition 3.2 *If the control model $\mathcal{M}$ satisfies Assumption A then, for each $k\in\mathbb{N}$, the control model $\mathcal{M}_k$ is bounded and, therefore, it satisfies Assumption A.*

Proof. Observe that for all $(x,a)\in\mathrm{Gr}(\mathbf{X}_k)$ we have

$$-q_k(\{x\}|x,a)\leq L\mathbf{w}_k\quad\text{and}\quad |c_k(x,a)|\leq M\mathbf{w}_k,$$

thus proving that $\mathcal{M}_k$ is bounded. So, Remark 2.7 is in order and $\mathcal{M}_k$ satisfies Assumption A (though not necessarily for the same functions and constants as $\mathcal{M}$). □

Given $k\in\mathbb{N}$, let $\mathcal{U}_k$ and $\mathcal{U}_k^s$ be the classes of admissible control policies and stationary policies, respectively, for the control model $\mathcal{M}_k$. For the control model $\mathcal{M}_k$, we will also consider the α-discounted control problem. Given an initial state $x\in\mathbf{X}_k$ and a control

policy $u \in \mathcal{U}_k$, let $\mathcal{V}_k(u,x)$ denote the corresponding total expected discounted cost, under the control model $\mathcal{M}_k$. We deduce from Theorem 2.6(iii) that the optimal discounted cost $\mathcal{V}_k^*(x)$ of the control model $\mathcal{M}_k$ verifies

$$\mathcal{V}_k^*(x) = \inf_{u \in \mathcal{U}_k} \mathcal{V}_k(u,x) = \inf_{u \in \mathcal{U}_k^s} \mathcal{V}_k(u,x) \quad \text{for all } x \in \mathbf{X}_k. \tag{3.1}$$

Regarding the control policies of the control models $\mathcal{M}$ and $\mathcal{M}_k$, notice that there exists an onto function $\mathbf{p}_k : \mathcal{U} \to \mathcal{U}_k$, where $\mathbf{p}_k(u)$ is the restriction of u to the family of feasible state-sojourn times for $\mathcal{M}_k$. Note also that $\mathbf{p}_k(\mathcal{U}^s) = \mathcal{U}_k^s$, the family of stationary policies for $\mathcal{M}_k$.

3.2 First approximation results

Suppose that Assumption A holds. Consider the control model $\mathcal{M}$ and, for some initial state $x \in \mathbf{X}$ and some control policy $u \in \mathcal{U}$, consider the $\mathbf{X}$-valued process $\{\xi_t\}_{t \geq 0}$. Given any $k \in \mathbb{N}$ let

$$\tau_k = \min\{t \geq 0 : \xi_t \notin \mathbf{S}_k\},$$

with the measurable set $\mathbf{S}_k$ as in Assumption A, and $\tau_k = \infty$ when $\xi_t \in \mathbf{S}_k$ for all $t \geq 0$. We have that τ_k is a $\{\mathcal{F}_t\}$-stopping time. Note also that τ_k can be defined as a "min", rather than an "inf", because the paths of $\{\xi_t\}$ are piecewise constant functions which are right-continuous. In particular, $\xi_{\tau_k} \notin \mathbf{S}_k$ if $\tau_k < \infty$.

We make the following important remark. Given an initial state $x \in \mathbf{X}_k$ and a control policy $u \in \mathcal{U}$, the total expected discounted cost until τ_k for the control model $\mathcal{M}$ equals the total expected discounted cost of the policy $\mathbf{p}_k(u)$ for the control model $\mathcal{M}_k$. Thus we can write

$$\mathcal{V}_k(\mathbf{p}_k(u),x) = \mathbb{E}_x^u\Big[\int_0^{\tau_k} e^{-\alpha t} \int_{\mathbf{A}(\xi_t)} c(\xi_t,a)\pi(da|t)dt\Big] \quad \text{for all } x \in \mathbf{X}_k. \tag{3.2}$$

Lemma 3.3 *Let Assumption A hold and fix $k \in \mathbb{N}$. For every initial state $x \in \mathbf{S}_k$ and any stationary policy $u \in \mathcal{U}^s$ we have*

$$\lim_{T \to \infty} \mathbb{E}_x^u[e^{-\alpha(T \wedge \tau_k)} w(\xi_{T \wedge \tau_k})] = \mathbb{E}_x^u[e^{-\alpha \tau_k} w(\xi_{\tau_k}) \mathbf{I}_{\{\tau_k < \infty\}}] \leq w(x) + b/\alpha.$$

and

$$\lim_{T \to \infty} \mathbb{E}_x^u[e^{-\alpha(T \wedge \tau_k)} w^{1+\beta}(\xi_{T \wedge \tau_k})] = \mathbb{E}_x^u[e^{-\alpha \tau_k} w^{1+\beta}(\xi_{\tau_k}) \mathbf{I}_{\{\tau_k < \infty\}}] \leq w^{1+\beta}(x) + b/\alpha.$$

Proof. We prove the first statement. We have the following (almost sure) convergence

$$e^{-\alpha(T \wedge \tau_k)} w(\xi_{T \wedge \tau_k}) \to e^{-\alpha \tau_k} w(\xi_{\tau_k}) \mathbf{I}_{\{\tau_k < \infty\}}.$$

Indeed, this is obvious if $\tau_k < \infty$, while if $\tau_k = \infty$ then $w(\xi_T) \leq \mathbf{w}_k$ for all $T \geq 0$ (this is because the function w is bounded above by $\mathbf{w}_k$ on $\mathbf{S}_k$), and the limit also holds. Now, by

Theorem 2.6(iv),

$$\begin{aligned}
&\mathbb{E}_x^u[e^{-\alpha(T\wedge\tau_k)}w(\xi_{T\wedge\tau_k})] - w(x)\\
&= \mathbb{E}_x^u\Big[\int_0^{T\wedge\tau_k} e^{-\alpha t}[-\alpha w(\xi_t) + \int_{\mathbf{A}(\xi_t)}\int_{\mathbf{X}} w(y)q(dy|\xi_t,a)\varphi(da|\xi_t)\Big]dt\Big]\\
&\le \mathbb{E}_x^u\Big[\int_0^{T\wedge\tau_k} e^{-\alpha t}[-\alpha w(\xi_t) + \rho w(\xi_t) + b]dt\Big] \quad \text{(by (2.5))}\\
&\le \mathbb{E}_x^u\Big[\int_0^{T\wedge\tau_k} be^{-\alpha t}dt\Big] \le b/\alpha,
\end{aligned}$$

and so for all $T \ge 0$ we have $\mathbb{E}_x^u[e^{-\alpha(T\wedge\tau_k)}w(\xi_{T\wedge\tau_k})] \le w(x) + b/\alpha$. By Fatou's lemma we obtain that

$$\mathbb{E}_x^u[e^{-\alpha\tau_k}w(\xi_{\tau_k})\mathbf{I}_{\{\tau_k<\infty\}}] \le w(x) + b/\alpha.$$

Observe now that for all $T \ge 0$

$$e^{-\alpha(T\wedge\tau_k)}w(\xi_{T\wedge\tau_k}) \le \mathbf{w}_k + e^{-\alpha\tau_k}w(\xi_{\tau_k})\mathbf{I}_{\{\tau_k<\infty\}}$$

(to see this, just consider the cases $\tau_k = \infty$ and, when $\tau_k < \infty$, distinguish between $T < \tau_k$ and $T \ge \tau_k$). Now we can use dominated convergence (indeed, now we know that the expectation of the righthand term is finite) to establish the limit. The second statement follows by using similar arguments. □

Our next result uses the following notation. Define the function $\mathbf{H}$ as

$$\mathbf{H}(u,x) = \mathbb{E}_x^u\Big[\int_0^\infty e^{-\alpha t}w(\xi_t)dt\Big]$$

for $u \in \mathcal{U}$ and $x \in \mathbf{X}$. It satisfies $\mathbf{H}(u,x) \le \mathbf{m}w(x)$ for all $u \in \mathcal{U}$ and $x \in \mathbf{X}$ (recall Theorem 2.6(ii)–(iii)).

Theorem 3.4 *Suppose that Assumption A holds and fix $k \in \mathbb{N}$. For every $x \in \mathbf{S}_k$ we have*

$$\sup_{u\in\mathcal{U}^s}\big|\mathcal{V}(u,x) - \mathcal{V}_k(\mathbf{p}_k(u),x)\big| \le \frac{\mathbf{m}M(w^{1+\beta}(x) + b/\alpha)}{\inf\{w^\beta(y) : y \notin \mathbf{S}_k\}}$$

and

$$|\mathcal{V}^*(x) - \mathcal{V}_k^*(x)| \le \frac{\mathbf{m}M(w^{1+\beta}(x) + b/\alpha)}{\inf\{w^\beta(y) : y \notin \mathbf{S}_k\}}.$$

Proof. Fix $k \in \mathbb{N}$, $x \in \mathbf{S}_k$, and $u \in \mathcal{U}^s$. Recalling (3.2), we have

$$\mathcal{V}(u,x) - \mathcal{V}_k(\mathbf{p}_k(u),x) = \mathbb{E}_x^u\Big[\int_{\tau_k}^\infty e^{-\alpha t}\int_{\mathbf{A}(\xi_t)} c(\xi_t,a)\varphi(da|\xi_t)dt\Big],$$

where the integral is defined as 0 when $\tau_k = \infty$. Therefore,

$$|\mathcal{V}(u,x) - \mathcal{V}_k(\mathbf{p}_k(u),x)| \le M\cdot\mathbb{E}_x^u\Big[\int_{\tau_k}^\infty e^{-\alpha t}w(\xi_t)dt\Big] = M\cdot\lim_{T\to\infty}\mathbb{E}_x^u\Big[\int_{T\wedge\tau_k}^\infty e^{-\alpha t}w(\xi_t)dt\Big] \quad (3.3)$$

by dominated convergence. On the other hand,

$$\begin{aligned}\mathbb{E}_x^u\Big[\int_{T\wedge\tau_k}^{\infty} e^{-\alpha t}w(\xi_t)dt\Big] &= \mathbb{E}_x^u\Big[\mathbb{E}_x^u\big[\int_{T\wedge\tau_k}^{\infty} e^{-\alpha t}w(\xi_t)dt \mid T\wedge\tau_k, \xi_{T\wedge\tau_k}\big]\Big]\\ &= \mathbb{E}_x^u\Big[e^{-\alpha(T\wedge\tau_k)}\mathbb{E}_x^u\big[\int_{T\wedge\tau_k}^{\infty} e^{-\alpha(t-T\wedge\tau_k)}w(\xi_t)dt \mid T\wedge\tau_k, \xi_{T\wedge\tau_k}\big]\Big]\\ &= \mathbb{E}_x^u\Big[e^{-\alpha(T\wedge\tau_k)}\mathbf{H}(u,\xi_{T\wedge\tau_k})\Big].\end{aligned}$$

Therefore,

$$\mathbb{E}_x^u\Big[\int_{T\wedge\tau_k}^{\infty} e^{-\alpha t}w(\xi_t)dt\Big] \le \mathbf{m}\mathbb{E}_x^u\Big[e^{-\alpha(T\wedge\tau_k)}w(\xi_{T\wedge\tau_k})\Big].$$

Recalling (3.3) and using Lemma 3.3, we conclude that

$$|\mathcal{V}(u,x) - \mathcal{V}_k(\mathbf{p}_k(u),x)| \le \mathbf{m}M\mathbb{E}_x^u[e^{-\alpha\tau_k}w(\xi_{\tau_k})\mathbf{I}_{\{\tau_k<\infty\}}].$$

Consequently,

$$\begin{aligned}&\inf\{w^\beta(y) : y\notin\mathbf{S}_k\}\cdot|\mathcal{V}(u,x) - \mathcal{V}_k(\mathbf{p}_k(u),x)|\\ &\le \mathbf{m}M\cdot\mathbb{E}_x^u\Big[e^{-\alpha\tau_k}w(\xi_{\tau_k})\mathbf{I}_{\{\tau_k<\infty\}}\inf\{w^\beta(y) : y\notin\mathbf{S}_k\}\Big]\\ &\le \mathbf{m}M\cdot\mathbb{E}_x^u\Big[e^{-\alpha\tau_k}w^{1+\beta}(\xi_{\tau_k})\mathbf{I}_{\{\tau_k<\infty\}}\Big]\end{aligned}$$

because, as already mentioned, when $\tau_k<\infty$ we have $\xi_{\tau_k}\notin\mathbf{S}_k$. By Lemma 3.3 again,

$$|\mathcal{V}(u,x) - \mathcal{V}_k(\mathbf{p}_k(u),x)| \le \frac{\mathbf{m}M}{\inf\{w^\beta(y) : y\notin\mathbf{S}_k\}}\cdot\big(w^{1+\beta}(x)+b/\alpha\big).$$

But $u\in\mathcal{U}^s$ being arbitrary, by Theorem 2.6(iii) and (3.1),

$$|\mathcal{V}^*(x) - \mathcal{V}_k^*(x)| \le \frac{\mathbf{m}M}{\inf\{w^\beta(y) : y\notin\mathbf{S}_k\}}\cdot(w^{1+\beta}(x)+b/\alpha).$$

The proof is now complete. □

As a consequence of Theorem 3.4 above, given an initial state $x\in\mathbf{X}$ for the control model $\mathcal{M}$ and any fixed precision $\epsilon>0$, for $k\in\mathbb{N}$ large enough the value functions of $\mathcal{M}$ and $\mathcal{M}_k$ at x verify $|\mathcal{V}^*(x)-\mathcal{V}_k^*(x)|\le\epsilon$. Moreover, from the initial data of the problem (namely, the function w and the constants involved in Assumption A) we can *explicitly* determine the value of k needed to reach such precision. Indeed, it suffices to choose $k\in\mathbb{N}$ such that $x\in\mathbf{S}_k$ and

$$w^\beta(y) \ge \frac{1}{\epsilon}\cdot\frac{M(b+\alpha)(w^{1+\beta}(x)+b/\alpha)}{\alpha(\alpha-\rho)} \quad \text{for all } y\in\mathbf{X}-\mathbf{S}_k.$$

Moreover, suppose that the stationary policy $u\in\mathcal{U}^s$ is such that $\mathbf{p}_k(u)\in\mathcal{U}_k^s$ is η-optimal for the control model $\mathcal{M}_k$ and the initial state $x\in\mathbf{S}_k$, meaning that $\mathcal{V}_k(\mathbf{p}_k(u),x)\le\mathcal{V}_k^*(x)+\eta$. By Theorem 3.4 we have that $\mathcal{V}(u,x)\le\mathcal{V}_k(\mathbf{p}_k(u),x)+\epsilon$, and so

$$\mathcal{V}^*(x)\ge\mathcal{V}_k^*(x)-\epsilon\ge\mathcal{V}_k(\mathbf{p}_k(u),x)-\eta-\epsilon\ge\mathcal{V}(u,x)-\eta-2\epsilon.$$

So, the policy $u\in\mathcal{U}^s$ is $(\eta+2\epsilon)$-optimal for the control model $\mathcal{M}$ and the initial state x. Loosely speaking we can say: if a stationary policy in $\mathcal{U}_k^s$ is η-optimal for $\mathcal{M}_k$, then extend it arbitrarily to a policy in $\mathcal{U}^s$ and this yields an $(\eta+2\epsilon)$-optimal policy for $\mathcal{M}$.

4 Approximations of the control models $\mathcal{M}_k$

In the previous section, we have shown how we can approximate the control model $\mathcal{M}$ by means of the control models $\mathcal{M}_k$ with bounded transition and cost rates. It remains to study how to approximate, in turn, the value function and the optimal policies of the bounded control models $\mathcal{M}_k$. This is precisely the purpose of this section.

Our next result gives some useful properties of $\mathcal{M}_k$ when the control model $\mathcal{M}$ satisfies our previous Assumptions A and B. We recall that the elements $\{\mathbf{X}_k, \mathbf{A}, \{\mathbf{A}(x)\}_{x\in\mathbf{X}_k}, q_k, c_k\}$ of the control model $\mathcal{M}_k$ have been introduced in Definition 3.1.

Theorem 4.1 *Suppose that the control model $\mathcal{M}$ verifies Assumptions A and (B1)–(B4). Fix $k \in \mathbb{N}$ and consider the control model $\mathcal{M}_k$. There exist $\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k}$ nonempty disjoint measurable subsets of $\mathbf{X}_k$ with $\cup_{i=1}^{m_k} \mathbf{Y}_i = \mathbf{X}_k$, and a probability measure $\mu_k \in \mathcal{P}_{\exp}(\mathbf{X}_k)$ such that:*

(i) *For all $B \in \mathcal{B}(\mathbf{X}_k)$ and $(x,a) \in \mathrm{Gr}(\mathbf{X}_k)$ we have*

$$q_k^+(B|x,a) = \int_B p_k(y|x,a)\mu_k(dy),$$

where the nonnegative function p_k is in $\mathbb{L}(\mathbf{Y}_i \times \mathrm{Gr}(\mathbf{Y}_j) : i,j = 1,\ldots,m_k)$;

(ii) *The cost rate function c_k is in $\mathbb{L}(\mathrm{Gr}(\mathbf{Y}_1), \ldots, \mathrm{Gr}(\mathbf{Y}_{m_k}))$;*

(iii) *The multifunction Ψ_k given by $x \mapsto \mathbf{A}(x)$ is piecewise Lipschitz continuous on $\mathbf{X}_k$ for the partition $\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k}$, with Lipschitz constants $L_j^{\Psi_k}$ for $j = 1, \ldots, m_k$.*

If in addition the control model $\mathcal{M}$ satisfies Assumption (B5) then for every $\delta > 0$ there exist finite sets $\mathbf{A}_\delta(x) \subseteq \mathbf{A}(x)$, for $x \in \mathbf{X}_k$, such that

(iv) *$d_H(\mathbf{A}_\delta(x), \mathbf{A}(x)) \leq \delta$ and, moreover, $x \mapsto \mathbf{A}_\delta(x)$ is piecewise Lipschitz continuous on $\mathbf{X}_k$ for the partition $\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k}$, with Lipschitz constants $L_j^{\Psi_k}$ for $j = 1, \ldots, m_k$ as in part (iii) of this theorem.*

Proof. Regarding the definition of the control model $\mathcal{M}_k$ in Definition 3.1, suppose that we are in case (i). Recalling that the control model $\mathcal{M}$ satisfies Assumption B, and using the notation in Assumption (B4), consider the following partition of the state space $\mathbf{X}_k = \mathbf{S}_k \cup \{x_\Delta\}$:

$$\mathbf{Y}_1 = \mathbf{Z}_1, \ldots, \mathbf{Y}_r = \mathbf{Z}_r, \mathbf{Y}_{r+1} = \{x_\Delta\}.$$

Fix arbitrary $\epsilon > 0$ and define the probability measure $\mu_k \in \mathcal{P}(\mathbf{X}_k)$ as

$$\mu_k(B) = \frac{1}{\mu(\mathbf{S}_k) + \epsilon}\mu(B \cap \mathbf{S}_k) + \frac{\epsilon}{\mu(\mathbf{S}_k) + \epsilon}\mathbf{I}_B(x_\Delta) \quad \text{for } B \in \mathcal{B}(\mathbf{X}_k)$$

(indeed, $\mu(\mathbf{S}_k)$ is finite as a consequence of Assumption (B3)). A direct consequence of Assumption (B3) is that $\mu_k \in \mathcal{P}_{\exp}(\mathbf{X}_k)$.

Let us check statement (i) of the theorem. With p be the function in Assumption (B2), define the nonnegative measurable function p_k on $\mathbf{X}_k \times \mathrm{Gr}(\mathbf{X}_k)$ as

$$p_k(y|x,a) = (\mu(\mathbf{S}_k) + \epsilon)p(y|x,a) \quad \text{for } y \in \mathbf{S}_k \text{ and } (x,a) \in \mathrm{Gr}(\mathbf{S}_k),$$

$$p_k(x_\Delta|x,a) = \frac{\mu(\mathbf{S}_k)+\epsilon}{\epsilon} q^+(\mathbf{X}-\mathbf{S}_k|x,a) \quad \text{for } (x,a)\in \mathrm{Gr}(\mathbf{S}_k),$$

and $p_k(y|x_\Delta,a)=0$ for $y\in\mathbf{X}_k$ and $a\in\mathbf{A}(x_\Delta)$. Given $B\in\mathcal{B}(\mathbf{X}_k)$ and any $(x,a)\in\mathrm{Gr}(\mathbf{S}_k)$

$$\begin{aligned} q_k^+(B|x,a) &= q_k(B|x,a) - q_k(\{x\}|x,a)\mathbf{I}_B(x) \\ &= q_k(B|x,a) - q(\{x\}|x,a)\mathbf{I}_B(x) \\ &= q(B\cap\mathbf{S}_k|x,a) + q(\mathbf{X}-\mathbf{S}_k|x,a)\mathbf{I}_B(x_\Delta) - q(\{x\}|x,a)\mathbf{I}_B(x) \\ &= q^+(B\cap\mathbf{S}_k|x,a) + q^+(\mathbf{X}-\mathbf{S}_k|x,a)\mathbf{I}_B(x_\Delta) \\ &= \int_{B\cap\mathbf{S}_k} p(y|x,a)\mu(dy) + q^+(\mathbf{X}-\mathbf{S}_k|x,a)\mathbf{I}_B(x_\Delta), \end{aligned}$$

so that we can indeed write

$$q_k^+(B|x,a) = \int_B p_k(dy|x,a)\mu_k(dy) \quad \text{for } B\in\mathcal{B}(\mathbf{X}_k) \text{ and } (x,a)\in\mathrm{Gr}(\mathbf{S}_k), \tag{4.1}$$

while, obviously, $0 = q_k^+(B|x_\Delta,a) = \int_B p_k(y|x_\Delta,a)\mu_k(dy)$ for all $B\in\mathcal{B}(\mathbf{X}_k)$ and $a\in\mathbf{A}_\Delta$. Summarizing, we have proved that (4.1) holds in fact for every $B\in\mathcal{B}(\mathbf{X}_k)$ and $(x,a)\in\mathrm{Gr}(\mathbf{X}_k)$. Regarding piecewise Lipschitz continuity of p_k of $\mathbf{X}_k\times\mathrm{Gr}(\mathbf{X}_k)$, we deduce from Assumption (B4) and the definition of p_k that p_k is indeed bounded and piecewise Lipschitz continuous on $\mathbf{S}_k\times\mathrm{Gr}(\mathbf{S}_k)$ for the partition $\mathbf{Y}_i\times\mathrm{Gr}(\mathbf{Y}_j)$, for $1\le i,j\le r$. By definition, p_k is also bounded and Lipschitz continuous on the sets $\mathbf{Y}_i\times\mathrm{Gr}(\mathbf{Y}_{r+1})$ for $1\le i\le r+1$. Therefore, it remains to prove that p_k is bounded and Lipschitz continuous on each set $\mathbf{Y}_{r+1}\times\mathrm{Gr}(\mathbf{Y}_j)$ for $1\le j\le r$ or, equivalently, that $(x,a)\mapsto q^+(\mathbf{X}-\mathbf{S}_k|x,a)$ is bounded and Lipschitz continuous on each set $\mathrm{Gr}(\mathbf{Y}_j)$, for $1\le j\le r$. To see this, write

$$q^+(\mathbf{X}-\mathbf{S}_k|x,a) = -q(\{x\}|x,a) - q^+(\mathbf{S}_k|x,a) \quad \text{for } (x,a)\in\mathrm{Gr}(\mathbf{Y}_j).$$

By Assumption (B4), $(x,a)\mapsto q(\{x\}|x,a)$ is Lipschitz continuous on $\mathrm{Gr}(\mathbf{Y}_j)$. Given (x,a) and (x',a') in $\mathrm{Gr}(\mathbf{Y}_j)$ we have by Assumption (B2)

$$q^+(\mathbf{S}_k|x,a) - q^+(\mathbf{S}_k|x',a') = \int_{\mathbf{S}_k} \big(p(y|x,a) - p(y|x',a')\big)\mu(dy)$$

and so for the Lipschitz constants L_{ij}^p of p on $\mathbf{Y}_i\times\mathrm{Gr}(\mathbf{Y}_j)$

$$\begin{aligned} |q^+(\mathbf{S}_k|x,a) - q^+(\mathbf{S}_k|x',a')| &\le \sum_{i=1}^r \int_{\mathbf{Y}_i} |p(y|x,a) - p(y|x',a')|\mu(dy) \\ &\le \Big(\sum_{i=1}^r L_{ij}^p \mu(\mathbf{Y}_i)\Big)\cdot\big(d_{\mathbf{X}}(x,x') + d_{\mathbf{A}}(a,a')\big). \end{aligned}$$

Recalling that $\mu(\mathbf{S}_k)$ is finite, we can conclude that $(x,a)\mapsto q^+(\mathbf{X}-\mathbf{S}_k|x,a)$ is indeed Lipschitz continuous on $\mathrm{Gr}(\mathbf{Y}_j)$. The fact that it is bounded follows easily from the inequality $q^+(\mathbf{X}-\mathbf{S}_k|x,a)\le -q(\{x\}|x,a)$. This completes the proof of part (i).

Notice that statements (ii) and (iii) hold as a direct consequence of Assumption (B4) and the definition of the control model $\mathcal{M}_k$. Finally, concerning statement (iv), choose $\mathbf{A}_\delta(x)\subseteq\mathbf{A}(x)$ as in Assumption (B5) for $x\in\mathbf{S}_k$, and choose any finite $\mathbf{A}_\delta(x_\Delta)\subseteq\mathbf{A}(x_\Delta)$ with the property $d_H(\mathbf{A}_\delta(x_\Delta),\mathbf{A}(x_\Delta))\le\delta$ (this is indeed posible because the action sets are compact).

The so-defined $\mathbf{A}_\delta(x)$, for $x \in \mathbf{X}_k$, satisfy part (iv) of the theorem.

When part (ii) of Definition 3.1 holds, that is, when $\mathbf{S}_k = \mathbf{X}$ let

$$\mu_k = \frac{1}{\mu(\mathbf{X})}\,\mu, \quad p_k = \mu(\mathbf{X}) \cdot p, \quad \text{and} \quad \mathbf{Y}_1 = \mathbf{Z}_1, \ldots, \mathbf{Y}_r = \mathbf{Z}_r,$$

which is a measurable partition of the state space $\mathbf{X}_k = \mathbf{S}_k = \mathbf{X}$. With these definitions, the statements of this theorem directly follow from Assumption B. □

Hence, Theorem 4.1 states that if the control model $\mathcal{M}$ satisfies Assumption B, then the elements of the bounded control model $\mathcal{M}_k$ are (loosely) piecewise Lipschitz continuous.

4.1 Piecewise Lipschitz continuity of the value function of $\mathcal{M}_k$

In what follows we will suppose that the control model $\mathcal{M}$ satisfies Assumptions A and (B1)–(B4), and so the control models $\mathcal{M}_k$ verify Theorem 4.1(i)–(iii). In the sequel, we will suppose that $k \in \mathbb{N}$ is fixed. This section is devoted to prove that the value function $\mathcal{V}_k^*$ of the control model $\mathcal{M}_k$ is in $\mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$.

We introduce some notation. In Theorem 4.1(i), letting $L_{ij}^{p_k}$ be a Lipschitz constant for p_k on $\mathbf{Y}_i \times \mathrm{Gr}(\mathbf{Y}_j)$, define

$$\mathcal{L}_j^{p_k} = \sum_{i=1}^{m_k} L_{ij}^{p_k}\,\mu_k(\mathbf{Y}_i) \quad \text{for each } 1 \le j \le m_k.$$

By Proposition 3.2, the control model $\mathcal{M}_k$ is bounded and, in particular, $|\mathcal{V}^*(x)| \le ||c_k||/\alpha$ for all $x \in \mathbf{X}_k$. (The cost function c_k of $\mathcal{M}_k$ being bounded, recall that $||c_k||$ is its supremum norm.) Also, there exists a constant $\overline{q}_k > 0$ and a (small) constant $0 < \eta_k < 1$ with

$$-q_k(\{x\}|x,a) \le (1-\eta_k)\overline{q}_k < \overline{q}_k \quad \text{for all } (x,a) \in \mathrm{Gr}(\mathbf{X}_k). \tag{4.2}$$

Define the transition probability measure Q_k on $\mathbf{X}_k$ given $\mathrm{Gr}(\mathbf{X}_k)$ as

$$Q_k(dy|x,a) = \frac{q_k(dy|x,a)}{\overline{q}_k} + \delta_x(dy).$$

By Theorem 4.1 and recalling the notation in (2.1) for the control model $\mathcal{M}_k$, for every $(x,a) \in \mathrm{Gr}(\mathbf{X}_k)$ we have

$$\begin{aligned} Q_k(dy|x,a) &= \frac{q_k^+(dy|x,a)}{\overline{q}_k} + \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k} \cdot \delta_x(dy) \\ &= \frac{p_k(y|x,a)}{\overline{q}_k}\mu_k(dy) + \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k}\delta_x(dy). \end{aligned} \tag{4.3}$$

Given a measurable function $v : \mathbf{X}_k \to \mathbb{R}$ and recalling the notation $Q_k v$ in (1.1), we have

$$Q_k v(x,a) = \frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} v(y)p_k(y|x,a)\mu_k(dy) + \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k}\,v(x). \tag{4.4}$$

Lemma 4.2 *The control model $\mathcal{M}_k$ satisfies the following properties.*

(i) *The function* $(x,a) \mapsto q_k(\{x\}|x,a)$ *is in* $\mathbb{L}(\mathrm{Gr}(\mathbf{Y}_1),\ldots,\mathrm{Gr}(\mathbf{Y}_{m_k}))$ *and it is* $\mathcal{L}_j^{p_k}$*-Lipschitz continuous on* $\mathrm{Gr}(\mathbf{Y}_j)$ *for each* $j=1,\ldots,m_k$.

(ii) *If* $v \in \mathbb{L}(\mathbf{Y}_1,\ldots,\mathbf{Y}_{m_k})$ *then* $Q_k v \in \mathbb{L}(\mathrm{Gr}(\mathbf{Y}_1),\ldots,\mathrm{Gr}(\mathbf{Y}_{m_k}))$. *For each* $1 \le j \le m_k$ *and all* (x,a) *and* (y,b) *in* $\mathrm{Gr}(\mathbf{Y}_j)$, *we have*

$$\big|Q_k v(x,a) - Q_k v(y,b)\big| \le \frac{2\mathcal{L}_j^{p_k}\|v\|}{\overline{q}_k}\big(d_{\mathbf{X}}(x,y) + d_{\mathbf{A}}(a,b)\big) + L_j^v d_{\mathbf{X}}(x,y).$$

Proof. (i). By (4.2), we have that the function $(x,a) \mapsto q_k(\{x\}|x,a)$ is in $\mathbb{B}(\mathrm{Gr}(\mathbf{X}_k))$. Now, given $(x,a) \in \mathrm{Gr}(\mathbf{X}_k)$, note that

$$-q_k(\{x\}|x,a) = q_k^+(\mathbf{X}_k|x,a) = \int_{\mathbf{X}_k} p_k(z|x,a)\mu_k(dz) = \sum_{i=1}^{m_k}\int_{\mathbf{Y}_i} p_k(z|x,a)\mu_k(dz).$$

Therefore, for fixed $1 \le j \le m_k$, when $(x,a) \in \mathrm{Gr}(\mathbf{Y}_j)$ and $(y,b) \in \mathrm{Gr}(\mathbf{Y}_j)$,

$$\begin{aligned}
|q^+(\mathbf{X}_k|x,a) - q^+(\mathbf{X}_k|y,b)| &\le \sum_{i=1}^{m_k}\int_{\mathbf{Y}_i}\big|p_k(z|x,a) - p_k(z|y,b)\big|\mu_k(dz)\\
&\le \sum_{i=1}^{m_k} L_{ij}^{p_k}\big(d_{\mathbf{X}}(x,y) + d_{\mathbf{A}}(a,b)\big)\mu_k(\mathbf{Y}_i)\\
&= \mathcal{L}_j^{p_k}\big(d_{\mathbf{X}}(x,y) + d_{\mathbf{A}}(a,b)\big).
\end{aligned}$$

Notice that we could also write the above Lipschitz constants as functions of the Lipschitz constants of $(x,a) \mapsto q(\{x\}|x,a)$ in Assumption (B4). We prefer the above expressions, however, to have a uniform notation (see (ii) below).

(ii). Obviously, $Q_k v$ is bounded because $||Q_k v|| \le ||v||$ and so $Q_k v \in \mathbb{B}(\mathrm{Gr}(\mathbf{X}_k))$. Consider now the lefthand term of (4.4). For fixed $1 \le j \le m_k$ and for any $(x,a) \in \mathrm{Gr}(\mathbf{Y}_j)$ and $(y,b) \in \mathrm{Gr}(\mathbf{Y}_j)$,

$$\begin{aligned}
&\left|\frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} v(z)p_k(z|x,a)\mu_k(dz) - \frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} v(z)p_k(z|y,b)\mu_k(dz)\right|\\
&\le \frac{1}{\overline{q}_k}\sum_{i=1}^{m_k}\Big|\int_{\mathbf{Y}_i} v(z)\big(p_k(z|x,a) - p_k(z|y,b)\big)\mu_k(dz)\Big|\\
&\le \frac{||v||}{\overline{q}_k}\sum_{i=1}^{m_k} L_{ij}^{p_k}\mu_k(\mathbf{Y}_i)\cdot\big(d_{\mathbf{X}}(x,y) + d_{\mathbf{A}}(a,b)\big)\\
&= \frac{||v||}{\overline{q}_k}\mathcal{L}_j^{p_k}\cdot\big(d_{\mathbf{X}}(x,y) + d_{\mathbf{A}}(a,b)\big).
\end{aligned}$$

Concerning the rightmost term of (4.4), as a function of $(x,a) \in \mathrm{Gr}(\mathbf{Y}_j)$ for some $1 \le j \le m_k$ it is the product of the function

$$(x,a) \mapsto \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k},$$

which is $\mathcal{L}_j^{p_k}/\overline{q}_k$-Lipschitz continuous (part (i) of this lemma) and which takes values in $[0,1]$, and the function $x \mapsto v(x)$, which is L_j^v-Lipschitz continuous and bounded by $||v||$. The stated result now follows. □

In our next result we introduce the dynamic programming equation for $\mathcal{M}_k$ and we show that the value function $\mathcal{V}_k^*$ can be characterized as a solution to this optimality equation. We also define the operator T_k for functions v in $\mathbb{B}(\mathbf{X}_k)$ as

$$T_k v(x) = \inf_{a \in \mathbf{A}(x)} \Big\{ \frac{c_k(x,a)}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} \int_{\mathbf{X}_k} v(y) Q_k(dy|x,a) \Big\} \quad \text{for } x \in \mathbf{X}_k. \tag{4.5}$$

Proposition 4.3 *Suppose that the control model $\mathcal{M}$ satisfies Assumptions A and (B1)–(B4), and consider the control model $\mathcal{M}_k$ for $k \in \mathbb{N}$.*

(i) *The unique solution v in $\mathbb{B}(\mathbf{X}_k)$ of the discounted cost optimality equation equation for the control model $\mathcal{M}_k$*

$$\alpha v(x) = \inf_{a \in \mathbf{A}(x)} \big\{ c_k(x,a) + \int_{\mathbf{X}_k} v(y) q_k(dy|x,a) \big\} \quad \textit{for } x \in \mathbf{X}_k$$

is $\mathcal{V}_k^*(x)$, for $x \in \mathbf{X}_k$.

(ii) *The unique solution v in $\mathbb{B}(\mathbf{X}_k)$ of the fixed point equation $v = T_k v$, that is,*

$$v(x) = \inf_{a \in \mathbf{A}(x)} \Big\{ \frac{c_k(x,a)}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} \int_{\mathbf{X}_k} v(y) Q_k(dy|x,a) \Big\} \quad \textit{for } x \in \mathbf{X}_k$$

is $\mathcal{V}_k^*(x)$, for $x \in \mathbf{X}_k$.

(iii) *If a deterministic stationary policy $\varphi \in \mathcal{U}_k^s$ given by $\varphi : \mathbf{X}_k \to \mathbf{A}$ attains the infimum in the discounted cost optimality equation (see (i)) or in the fixed point equation $v = T_k v$ (see (ii)), and such a φ indeed exists, then φ is an optimal policy for $\mathcal{M}_k$ for any initial state $x \in \mathbf{X}_k$.*

Proof. (i). This result easily follows from Theorem 5 in [11]. Note that Condition 6(a) in that reference, namely, the continuity of

$$a \mapsto \int_{\mathbf{X}_k} v(y) q_k(dy|x,a) \quad \text{for all } x \in \mathbf{X}_k \text{ and } v \in \mathbb{B}(\mathbf{X}_k), \tag{4.6}$$

follows from the proof of Lemma 4.2(ii). Indeed, proceeding as in the proof of Lemma 4.2(ii), we can prove that, for fixed $x \in \mathbf{X}_k$, the function $a \mapsto Q_k v(x,a)$ is Lipschitz continuous on $\mathbf{A}(x)$ because, to show this, measurability of v suffices.

(ii). It should be clear that the discounted cost optimality equation is equivalent to the fixed point equation for the operator T_k. Note that T_k indeed maps $\mathbb{B}(\mathbf{X}_k)$ into itself because, for $v \in \mathbb{B}(\mathbf{X}_k)$, we have that $T_k v$ is measurable (Proposition D5(a) in [8] and

$$||Tv|| \leq \frac{\overline{q}_k ||v|| + ||c_k||}{\alpha + \overline{q}_k}.$$

Observe also that T_k is a contraction operator on $\mathbb{B}(\mathbf{X}_k)$ with modulus $\overline{q}_k/(\alpha + \overline{q}_k)$, and thus the result also follows from standard (discrete-time) dynamic programming arguments.

(iii). The fact that there exists measurable $\varphi : \mathbf{X}_k \to \mathbf{A}$ attaining the minimum in the discounted cost optimality equation in (i) or in the fixed point equation in (ii) follows from the continuity of $a \mapsto c_k(x,a) + \int \mathcal{V}_k^*(y) q_k(dy|x,a)$ for all $x \in \mathbf{X}_k$ (recall Theorem 4.1(ii) and (4.6)) and from Proposition D5 in [8]. □

Now we are ready to prove our main result in this section.

Proposition 4.4 *Suppose that the control model $\mathcal{M}$ satisfies Assumptions A and (B1)–(B4), and consider the control model $\mathcal{M}_k$ for $k \in \mathbb{N}$.*

(i) *If $v \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ then $T_k v \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$, with*

$$L_j^{T_k v} = \frac{1}{\alpha + \overline{q}_k} \Big(\big(L_j^{c_k} + 2\mathcal{L}_j^{p_k} ||v|| \big) \cdot \big(1 + L_j^{\Psi_k} \big) + \overline{q}_k L_j^v \Big)$$

on each set $\mathbf{Y}_j$, for $j = 1, \ldots, m_k$.

(ii) *The optimal discounted cost function $\mathcal{V}_k^*$ is in $\mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ with*

$$L_j^{\mathcal{V}_k^*} = \frac{1}{\alpha} (L_j^{c_k} + 2\mathcal{L}_j^{p_k} ||c_k|| / \alpha) \cdot (1 + L_j^{\Psi_k}) \quad \textit{for each } 1 \le j \le m_k.$$

Proof. (i). Fix $x, y \in \mathbf{Y}_j$ for some $1 \le j \le m_k$. Given $a \in \mathbf{A}(x)$ and $b \in \mathbf{A}(y)$ define

$$\mathbf{T}(x,a,y,b) = \frac{|c_k(x,a) - c_k(y,b)|}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} \big| Q_k v(x,a) - Q_k v(y,b) \big|.$$

By the definition of $T_k v(x)$ and $T_k v(y)$ we have

$$|T_k v(x) - T_k v(y)| \le \sup_{a \in \mathbf{A}(x)} \inf_{b \in \mathbf{A}(y)} \{\mathbf{T}(x,a,y,b)\} \vee \sup_{b \in \mathbf{A}(y)} \inf_{a \in \mathbf{A}(x)} \{\mathbf{T}(x,a,y,b)\}.$$

The function c being $L_j^{c_k}$-Lipschitz continuous on $\mathrm{Gr}(\mathbf{Y}_j)$ (recall Theorem 4.1(ii)) and by Lemma 4.2(ii),

$$|T_k v(x) - T_k v(y)| \le \frac{1}{\alpha + \overline{q}_k} \big(L_j^{c_k} + 2\mathcal{L}_j^{p_k} \|v\| \big) \cdot \Big(d_{\mathbf{X}}(x,y) + d_H(\mathbf{A}(x), \mathbf{A}(y)) \Big) + \frac{\overline{q}_k}{\alpha + \overline{q}_k} L_j^v d_{\mathbf{X}}(x,y).$$

Since $d_H(\mathbf{A}(x), \mathbf{A}(y)) \le L_j^{\Psi_k} d_{\mathbf{X}}(x,y)$ (recall Theorem 4.1(iii)), the result follows.

(ii). The operator T_k being a contraction operator on $\mathbb{B}(\mathbf{X}_k)$, given an arbitrary $V_0 \in \mathbb{B}(\mathbf{X}_k)$ it follows that $\{V_r\}$, defined as $V_{r+1} = T_k V_r$ for $r \ge 0$, converges in the supremum norm to $\mathcal{V}_k^*$. Let $V_0 \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ be such that $||V_0|| \le ||c_k||/\alpha$. It follows from part (i) of this proposition that $V_r \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ and that $||V_r|| \le ||c_k||/\alpha$ for all $r \ge 0$. From part (i) of this proposition also, the Lipschitz constants L_j^r for V_r on $\mathbf{Y}_j$ satisfy

$$L_j^{r+1} = \frac{1}{\alpha + \overline{q}_k} \big(L_j^{c_k} + 2\mathcal{L}_j^{p_k} ||c_k|| / \alpha \big) \big(1 + L_j^{\Psi_k} \big) + \frac{\overline{q}_k}{\alpha + \overline{q}_k} L_j^r.$$

The sequence $\{L_j^r\}_{r \ge 0}$ converges because the factor of L_j^r above is less than one. It turns out that $\mathcal{V}_k^* = \lim_r V_r$ is Lipschitz continuous on $\mathbf{Y}_j$, with a Lipschitz constant

$$L_j^{\mathcal{V}_k^*} = \lim_r L_j^r = \frac{1}{\alpha} \big(L_j^{c_k} + 2\mathcal{L}_j^{p_k} ||c_k|| / \alpha \big) \big(1 + L_j^{\Psi_k} \big).$$

This completes the proof. □

4.2 Approximation of the value function $\mathcal{V}_k^*$ of $\mathcal{M}_k$

Now we study the problem of approximating the value function $\mathcal{V}_k^*$ of the control model $\mathcal{M}_k$ taking advantage of the fact that it is piecewise Lipschitz continuous (Proposition 4.4 above). The technique will consist in discretizing both the state space $\mathbf{X}_k$ and the action sets $\mathbf{A}(x)$ for $x \in \mathbf{X}_k$. *Recall that we are supposing that the control model $\mathcal{M}$ satisfies Assumptions A and (B1)–(B4), and so $\mathcal{M}_k$ satisfies Theorem 4.1(i)–(iii) for each $k \in \mathbb{N}$. In what follows, $k \in \mathbb{N}$ remains fixed.*

Given $1 \leq i \leq m_k$, define the probability measure $\nu_{k,i} \in \mathcal{P}_{\exp}(\mathbf{Y}_i)$ as

$$\nu_{k,i}(B) = \mu_k(B)/\mu_k(\mathbf{Y}_i) \quad \text{for } B \in \mathcal{B}(\mathbf{Y}_i)$$

provided that $\mu_k(\mathbf{Y}_i) > 0$, and as a Dirac probability measure on some point in $\mathbf{Y}_i$ when $\mu_k(\mathbf{Y}_i) = 0$. We therefore have

$$\mu_k = \sum_{i=1}^{m_k} \mu_k(\mathbf{Y}_i)\nu_{k,i}. \tag{4.7}$$

For each $1 \leq i \leq m_k$ let $\{\zeta_{k,i}^n\}_{n\geq 1}$ be i.i.d random variables with distribution $\nu_{k,i}$, defined on the same probability space $(\Theta, \mathcal{F}_\Theta, \mathbb{P})$. For $1 \leq i \leq m_k$ and $n \geq 1$, let $\nu_{k,i}^n$ be the empirical probability measure of size n of the probability measure $\nu_{k,i}$. Define the (random) probability measure $\mu_k^n \in \mathcal{P}_{\exp}(\mathbf{X}_k)$ as the convex linear combination

$$\mu_k^n = \sum_{i=1}^{m_k} \mu_k(\mathbf{Y}_i)\nu_{k,i}^n, \tag{4.8}$$

and let $\Gamma_k^n \subseteq \mathbf{X}_k$ be its finite (random) support, which consists of at most $n \cdot m_k$ points. Note that for each $n \geq 1$ and $1 \leq i \leq m_k$ we have $\mu_k^n(\mathbf{Y}_i) = \mu_k(\mathbf{Y}_i)$. Define now for each $n \geq 1$

$$\mathcal{W}^*(\mu_k, \mu_k^n) = \max_{1\leq i\leq m_k} \mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n),$$

which does not necessarily coincide with the 1-Wasserstein distance between μ_k and μ_k^n.

Remark 4.5 *We are now dealing with two indices: k which represents the "truncation" of the control model $\mathcal{M}$ to $\mathcal{M}_k$, and n which is related to the sampling used to discretize the state space of $\mathcal{M}_k$.*

To make the reading easier we propose the following rule of thumb: the indices related to the "truncation" of $\mathcal{M}$, such as k, will be written as subscripts. The indices related to the discretization of $\mathcal{M}_k$, such as n or δ below, will be written as superscripts.

For instance, μ_k is related to the restriction of μ to $\mathbf{X}_k$, and μ_k^n is obtained from μ_k by taking a sample.

Let R_k^n be transition measure on $\mathbf{X}_k$ given $\mathrm{Gr}(\mathbf{X}_k)$ defined as

$$R_k^n(dy|x,a) = \frac{p_k(y|x,a)}{\overline{q}_k}\mu_k^n(dy) + \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k}\delta_x(dy) \tag{4.9}$$

for $(x,a) \in \mathrm{Gr}(\mathbf{X}_k)$, and observe that R_k^n is given the same definition as Q_k in (4.3) except that μ_k in (4.3) is replaced with μ_k^n. By the definition of the constant η_k in (4.2),

$$R_k^n(\mathbf{X}_k|x,a) \geq \eta_k > 0 \quad \text{for all } (x,a) \in \mathrm{Gr}(\mathbf{X}_k) \text{ and } n \geq 1.$$

We can therefore normalize R_k^n so as to define the following transition probability measure on $\mathbf{X}_k$ given $\mathrm{Gr}(\mathbf{X}_k)$:

$$Q_k^n(dy|x,a) = \frac{R_k^n(dy|x,a)}{R_k^n(\mathbf{X}_k|x,a)} \quad \text{for } (x,a) \in \mathrm{Gr}(\mathbf{X}_k).$$

Our next result compares Q_k^n with Q_k, when applied to a piecewise Lipschitz continuous function. It uses the following notation. Given a function $v \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ with Lipschitz constants L_j^v for $1 \le j \le m_k$, define

$$\mu_k(L^v) = \sum_{i=1}^{m_k} \mu_k(\mathbf{Y}_i) L_i^v.$$

Lemma 4.6 *Fix any $n \ge 1$. Given $1 \le j \le m_k$ and $(x,a) \in \mathrm{Gr}(\mathbf{Y}_j)$, the following inequalities hold.*

(i) $|1 - R_k^n(\mathbf{X}_k|x,a)| \le \left(\mathcal{L}_j^{p_k}/\overline{q}_k\right) \cdot \mathcal{W}^*(\mu_k, \mu_k^n)$.

(ii) *For any function $v \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$*

$$|Q_k v(x,a) - Q_k^n v(x,a)| \le \frac{2\mathcal{L}_j^{p_k}||v|| + \mu_k(L^v)||p_k||}{\overline{q}_k} \cdot \mathcal{W}^*(\mu_k, \mu_k^n).$$

Proof. (i). Fix $(x,a) \in \mathrm{Gr}(\mathbf{Y}_j)$ for some $1 \le j \le m_k$. Observe that by the definition of Q_k and R_k^n in (4.3) and (4.9), and by (4.7)–(4.8), we have

$$\begin{aligned} 1 - R_k^n(\mathbf{X}_k|x,a) &= \frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} p_k(y|x,a)\mu_k(dy) - \frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} p_k(y|x,a)\mu_k^n(dy) \\ &= \frac{1}{\overline{q}_k}\sum_{i=1}^{m_k} \mu_k(\mathbf{Y}_i)\int_{\mathbf{Y}_i} p_k(y|x,a)\,(\nu_{k,i} - \nu_{k,i}^n)(dy) \end{aligned}$$

and so

$$|1 - R_k^n(\mathbf{X}_k|x,a)| \le \frac{1}{\overline{q}_k}\sum_{i=1}^{m_k} \mu_k(\mathbf{Y}_i) L_{ij}^{p_k} \mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n) \le \frac{\mathcal{L}_j^{p_k}}{\overline{q}_k}\,\mathcal{W}^*(\mu_k, \mu_k^n).$$

(ii). Given $(x,a) \in \mathrm{Gr}(\mathbf{Y}_j)$ for some $1 \le j \le m_k$, we have

$$\begin{aligned} Q_k v(x,a) &= \frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} v(y)p_k(y|x,a)\mu_k(dy) + \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k} v(x) \\ Q_k^n v(x,a) &= \frac{1}{R_k^n(\mathbf{X}_k|x,a)}\left(\frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} v(y)p_k(y|x,a)\mu_k^n(dy) + \frac{\overline{q}_k + q_k(\{x\}|x,a)}{\overline{q}_k} v(x)\right). \end{aligned}$$

Consequently,

$$\begin{aligned} Q_k v(x,a) - Q_k^n v(x,a) &= \frac{1}{\overline{q}_k}\int_{\mathbf{X}_k} v(y)p_k(y|x,a)(\mu_k - \mu_k^n)(dy) + (1 - R_k^n(\mathbf{X}_k|x,a))Q_k^n v(x,a) \\ &= \frac{1}{\overline{q}_k}\sum_{i=1}^{m_k} \mu_k(\mathbf{Y}_i)\int_{\mathbf{Y}_i} v(y)p_k(y|x,a)(\nu_{k,i} - \nu_{k,i}^n)(dy) \\ &\quad + (1 - R_k^n(\mathbf{X}_k|x,a))Q_k^n v(x,a). \end{aligned}$$

Recalling that the function $y \mapsto p_k(y|x,a)$ is $L_{ij}^{p_k}$-Lipschitz continuous on each $\mathbf{Y}_i$, it follows that $y \mapsto v(y)p_k(y|x,a)$ is $\big(L_{ij}^{p_k}||v|| + L_i^v||p_k||\big)$-Lipschitz continuous on each $\mathbf{Y}_i$, and so

$$\left|\frac{1}{\overline{q}_k}\sum_{i=1}^{m_k}\mu_k(\mathbf{Y}_i)\int_{\mathbf{Y}_i} v_k(y)p_k(y|x,a)(\nu_{k,i}-\nu_{k,i}^n)(dy)\right| \leq \frac{1}{\overline{q}_k}\big(\mathcal{L}_j^{p_k}||v|| + \mu_k(L^v)||p_k||\big)\cdot\mathcal{W}^*(\mu_k,\mu_k^n).$$

On the other hand, by statement (i),

$$\big|(1-R_k^n(\mathbf{X}_k|x,a))Q_k^n v(x,a)\big| \leq \frac{\mathcal{L}_j^{p_k}||v||}{\overline{q}_k}\cdot\mathcal{W}^*(\mu_k,\mu_k^n).$$

The stated result now follows. □

We propose the following definition of the sets $\mathbf{A}_\delta(x)$ for $x \in \mathbf{X}_k$.

Definition 4.7 *For each $\delta > 0$ and every $x \in \mathbf{X}_k$, let $\mathbf{A}_\delta(x) \subseteq \mathbf{A}(x)$ be a finite set such that $d_H(\mathbf{A}_\delta(x), \mathbf{A}(x)) \leq \delta$.*

Note that under Assumption (B1) on the control model $\mathcal{M}$ (namely, the actions sets $\mathbf{A}(x)$ are compact) such a construction is indeed possible. It is important to stress that we are not (yet) imposing Assumption (B5) on $\mathcal{M}$, and so Theorem 4.1(iv) needs hot hold for the control model $\mathcal{M}_k$. Hence, the multifunction $x \mapsto \mathbf{A}_\delta(x)$, for the sets $\mathbf{A}_\delta(x)$ in Definition 4.7, is not necessarily piecewise Lipschitz continuous.

For every $n \geq 1$ and $\delta > 0$, define the operator $T_k^{n,\delta}$ on $\mathbb{B}(\mathbf{X}_k)$ as

$$T_k^{n,\delta}v(x) = \min_{a\in\mathbf{A}_\delta(x)}\Big\{\frac{c_k(x,a)}{\alpha+\overline{q}_k} + \frac{\overline{q}_k}{\alpha+\overline{q}_k}\int_{\mathbf{X}_k} v(y)Q_k^n(dy|x,a)\Big\} \quad \text{for } x \in \mathbf{X}_k. \tag{4.10}$$

The operator $T_k^{n,\delta}$ is defined similarly to T_k in (4.5) except that the sets $\mathbf{A}(x)$ are replaced with $\mathbf{A}_\delta(x)$, and the kernel Q_k is replaced with Q_k^n. We have that $T_k^{n,\delta}$ is a contraction operator with modulus $\frac{\overline{q}_k}{\alpha+\overline{q}_k}$ on $\mathbb{B}(\mathbf{X}_k)$, and let $\mathcal{V}_k^{n,\delta} \in \mathbb{B}(\mathbf{X}_k)$ be its unique fixed point. Note that $||\mathcal{V}_k^{n,\delta}|| \leq ||c_k||/\alpha$.

Remark 4.8 *Given $v \in \mathbb{B}(\mathbf{X}_k)$ we can explicitly compute $T_k^{n,\delta}v(x)$ for every $x \in \mathbf{X}_k$. Indeed, given $x \in \mathbf{X}_k$ and $a \in \mathbf{A}_\delta(x)$, note that $\int_{\mathbf{X}_k} v(y)Q_k^n(dy|x,a)$ depends on v only through its values on the finite set $\Gamma_k^n \cup \{x\}$. The set $\mathbf{A}_\delta(x)$ being finite, the value of $T_k^{n,\delta}v(x)$ can be determined explicitly.*

Now we explain how we can explicitly compute $\mathcal{V}_k^{n,\delta}$. First of all note that $Q_k^n(\cdot|x,a)$ can be seen as a transition probability measure on Γ_k^n given $\mathrm{Gr}(\Gamma_k^n)$. Therefore, the fixed point equation

$$v(x) = \min_{a\in\mathbf{A}_\delta(x)}\Big\{\frac{c_k(x,a)}{\alpha+\overline{q}_k} + \frac{\overline{q}_k}{\alpha+\overline{q}_k}\int_{\mathbf{X}_k} v(y)Q_k^n(dy|x,a)\Big\} \quad \textit{for } x \in \Gamma_k^n$$

corresponds to the discounted cost optimality equation of a discrete-time Markov decision process with state space Γ_k^n, actions sets $\mathbf{A}_\delta(x)$ for $x \in \Gamma_k^n$, cost function $\frac{c_k}{\alpha+\overline{q}_k}$, discount factor $\frac{\overline{q}_k}{\alpha+\overline{q}_k}$, and transition probabilities Q_k^n. This optimality equation can be solved explicitly using, e.g., linear programming or the policy iteration algorithm. Once the value of $\mathcal{V}_k^{n,\delta}$ is known on Γ_k^n, it is straightforward to compute $\mathcal{V}_k^{n,\delta}(x)$ for any $x \in \mathbf{X}_k$.

Our next result compares the operators T_k and $T_k^{n,\delta}$.

Lemma 4.9 *Given $v \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$, for every $n \geq 1$ and $\delta > 0$ we have*

$$\left|T_k v(x) - T_k^{n,\delta} v(x)\right| \leq \frac{L_j^{c_k} + 2\mathcal{L}_j^{p_k}||v||}{\alpha + \overline{q}_k} \cdot \delta + \frac{2\mathcal{L}_j^{p_k}||v|| + \mu_k(L^v)||p_k||}{\alpha + \overline{q}_k} \cdot \mathcal{W}^*(\mu_k, \mu_k^n)$$

for $x \in \mathbf{Y}_j$, for each $1 \leq j \leq m_k$.

Proof. Fix $x \in \mathbf{Y}_j$ for some $1 \leq j \leq m_k$. Note that $T_k v(x) - T_k^{n,\delta} v(x)$ equals

$$\inf_{a \in \mathbf{A}(x)} \left\{ \frac{c_k(x,a)}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} Q_k v(x,a) \right\} - \inf_{a \in \mathbf{A}_\delta(x)} \left\{ \frac{c_k(x,a)}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} Q_k^n v(x,a) \right\},$$

where we use the notation $Q_k^n v$, recall (1.1). Let

$$\mathbf{T}(x,a,a') = \frac{c_k(x,a)}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} Q_k v(x,a) - \frac{c_k(x,a')}{\alpha + \overline{q}_k} + \frac{\overline{q}_k}{\alpha + \overline{q}_k} Q_k^n v(x,a')$$

for $a \in \mathbf{A}(x)$ and $a' \in \mathbf{A}_\delta(x)$. So,

$$|T_k v(x) - T_k^{n,\delta} v(x)| \leq \sup_{a \in \mathbf{A}(x)} \min_{a' \in \mathbf{A}_\delta(x)} |\mathbf{T}(x,a,a')| \ \vee \max_{a' \in \mathbf{A}_\delta(x)} \inf_{a \in \mathbf{A}(x)} |\mathbf{T}(x,a,a')|.$$

On one hand we have $|c_k(x,a) - c_k(x,a')| \leq L_j^{c_k} d_{\mathbf{A}}(a,a')$ and, on the other hand, write

$$|Q_k v(x,a) - Q_k^n v(x,a')| \leq |Q_k v(x,a) - Q_k v(x,a')| + |Q_k v(x,a') - Q_k^n v(x,a')|.$$

By Lemma 4.2(ii) we have

$$|Q_k v(x,a) - Q_k v(x,a')| \leq \frac{2\mathcal{L}_j^{p_k}||v||}{\overline{q}_k} d_{\mathbf{A}}(a,a'),$$

while by Lemma 4.6(ii) we have

$$|Q_k v(x,a') - Q_k^n v(x,a')| \leq \frac{2\mathcal{L}_j^{p_k}||v|| + \mu_k(L^v)||p_k||}{\overline{q}_k} \cdot \mathcal{W}^*(\mu_k, \mu_k^n).$$

Therefore,

$$|\mathbf{T}(x,a,a')| \leq \frac{L_j^{c_k} + 2\mathcal{L}_j^{p_k}||v||}{\alpha + \overline{q}_k} d_{\mathbf{A}}(a,a') + \frac{2\mathcal{L}_j^{p_k}||v|| + \mu_k(L^v)||p_k||}{\alpha + \overline{q}_k} \cdot \mathcal{W}^*(\mu_k, \mu_k^n).$$

Consequently

$$|T_k v(x) - T_k^{n,\delta} v(x)| \leq \frac{L_j^{c_k} + 2\mathcal{L}_j^{p_k}||v||}{\alpha + \overline{q}_k} d_H(\mathbf{A}(x), \mathbf{A}_\delta(x)) + \frac{2\mathcal{L}_j^{p_k}||v|| + \mu_k(L^v)||p_k||}{\alpha + \overline{q}_k} \cdot \mathcal{W}^*(\mu_k, \mu_k^n),$$

and the result now follows. □

We define the following two constants

$$\mathbf{G}_k = \max_{1 \leq j \leq m_k} \left\{ \frac{L_j^{c_k} + 2\mathcal{L}_j^{p_k}||c_k||/\alpha}{\alpha} \right\} \quad \text{and} \quad \mathbf{H}_k = \max_{1 \leq k \leq m_k} \left\{ \frac{\mu_k(L^{\mathcal{V}_k^*})||p_k|| + 2\mathcal{L}_j^{p_k}||c_k||/\alpha}{\alpha} \right\}.$$

Remark 4.10 *It is important to stress that the constants* $\mathbf{G}_k$ *and* $\mathbf{H}_k$ *depend on constants and parameters related to the control model* $\mathcal{M}$*, given in Assumptions A and (B1)–(B4), or related to parameters of the control model* $\mathcal{M}_k$*, given in Theorem 4.1(i)–(iii). In particular note that we indeed have an explicit expression for* $\mu_k(L^{\mathcal{V}_k^*})$*, which can be derived from Proposition 4.4. So, the constants* $\mathbf{G}_k$ *and* $\mathbf{H}_k$ *can be computed explicitly and they neither depend on* $n \geq 1$ *nor on* $\delta > 0$*.*

Now we can provide a bound on the approximation error of $\mathcal{V}_k^*(x)$.

Proposition 4.11 *Suppose that the control model* $\mathcal{M}$ *verifies Assumptions A and (B1)–(B4), and consider the control model* $\mathcal{M}_k$ *for some given* $k \in \mathbb{N}$*. For each* $n \geq 1$ *and* $\delta > 0$ *(recall Definition 4.7), we have*

$$||\mathcal{V}_k^* - \mathcal{V}_k^{n,\delta}|| \leq \mathbf{G}_k \cdot \delta + \mathbf{H}_k \cdot \mathcal{W}^*(\mu_k, \mu_k^n).$$

Proof. Observe that

$$\begin{aligned} ||\mathcal{V}_k^* - \mathcal{V}_k^{n,\delta}|| &= ||T_k\mathcal{V}_k^* - T_k^{n,\delta}\mathcal{V}_k^{n,\delta}|| \leq ||T_k\mathcal{V}_k^* - T_k^{n,\delta}\mathcal{V}_k^*|| + ||T_k^{n,\delta}\mathcal{V}_k^* - T_k^{n,\delta}\mathcal{V}_k^{n,\delta}|| \\ &\leq ||T_k\mathcal{V}_k^* - T_k^{n,\delta}\mathcal{V}_k^*|| + \frac{\overline{q}_k}{\alpha + \overline{q}_k}||\mathcal{V}_k^* - \mathcal{V}_k^{n,\delta}|| \end{aligned}$$

and so

$$||\mathcal{V}_k^* - \mathcal{V}_k^{n,\delta}|| \leq \frac{\alpha + \overline{q}_k}{\alpha}||T_k\mathcal{V}_k^* - T_k^{n,\delta}\mathcal{V}_k^*||.$$

Recall now that $\mathcal{V}_k^* \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ (Proposition 4.4(ii)) with $||\mathcal{V}_k^*|| \leq ||c_k||/\alpha$ and use Lemma 4.9 to obtain the result. □

As already mentioned in Remark 4.8, the value of $\mathcal{V}_k^{n,\delta}(x)$ can be explicitly computed for each fixed $x \in \mathbf{X}_k$, while the constants $\mathbf{G}_k$ and $\mathbf{H}_k$ can be explicitly computed as well (Remark 4.10). In conclusion, Proposition 4.11 provides a *computable approximation* of $\mathcal{V}_k^*(x)$ with an *explicitly known* approximation error.

We can now combine the results in Proposition 1.1 and Theorem 3.4 to obtain our main result on the approximation of the value function $\mathcal{V}^*$ of the control model $\mathcal{M}$.

Theorem 4.12 *Suppose that the control model* $\mathcal{M}$ *satisfies Assumptions A and (B1)–(B4). Fix an initial state* $x \in \mathbf{X}$ *and let* $\epsilon > 0$ *be some given precision. There exist* $k \in \mathbb{N}$*,* $\delta > 0$*, and positive constants* $C = C(\epsilon, k)$ *and* $D = D(\epsilon, k)$ *such that*

$$\mathbb{P}\big\{|\mathcal{V}^*(x) - \mathcal{V}_k^{n,\delta}(x)| > \epsilon\big\} \leq Ce^{-Dn} \quad \textit{for all } n \geq 1.$$

Proof. Choose $k \in \mathbb{N}$ large enough so that $x \in \mathbf{S}_k$ and (cf. Theorem 3.4)

$$\frac{\mathbf{m}M(w^{1+\beta}(x) + b/\alpha)}{\inf\{w^\beta(y) : y \notin \mathbf{S}_k\}} \leq \epsilon/3$$

so that $|\mathcal{V}^*(x) - \mathcal{V}_k^*(x)| \leq \epsilon/3$. For such $k \in \mathbb{N}$, choose $\delta > 0$ with (cf. Proposition 4.11)

$$\mathbf{G}_k \cdot \delta \leq \epsilon/3.$$

Consequently, by Proposition 4.11,

$$\{|\mathcal{V}^*(x) - \mathcal{V}_k^{n,\delta}(x)| > \epsilon\} \subseteq \Big\{\mathcal{W}^*(\mu_k, \mu_k^n) > \frac{\epsilon}{3\mathbf{H}_k}\Big\}.$$

On the other hand,

$$\begin{aligned}\mathbb{P}\Big\{\mathcal{W}^*(\mu_k, \mu_k^n) > \frac{\epsilon}{3\mathbf{H}_k}\Big\} &= \mathbb{P}\Big\{\max_{1\le i\le m_k}\{\mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n)\} > \frac{\epsilon}{3\mathbf{H}_k}\Big\} \\ &\le \sum_{i=1}^{m_k}\mathbb{P}\Big\{\mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n) > \frac{\epsilon}{3\mathbf{H}_k}\Big\}.\end{aligned}$$

By Proposition 1.1, for each $1 \le i \le m_k$ there exist positive constants $\mathfrak{c}(\epsilon, i, k)$ and $\mathfrak{d}(\epsilon, i, k)$ with

$$\mathbb{P}\{\mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n) > \epsilon/3\mathbf{H}_k\} \le \mathfrak{c}(\epsilon, i, k)e^{-\mathfrak{d}(\epsilon,i,k)n} \quad \text{for all } n \ge 1.$$

Letting $C(\epsilon, k) = \sum_{i=1}^{m_k}\mathfrak{c}(\epsilon, i, k)$ and $D(\epsilon, k) = \min_{1\le i\le m_k}\mathfrak{d}(\epsilon, i, k)$ we reach the stated result. □

This theorem states that, given some initial state $x \in \mathbf{X}$ and some precision $\epsilon > 0$, we can find $k \in \mathbb{N}$ and $\delta > 0$ such that, when approximating $\mathcal{V}^*(x)$ by $\mathcal{V}_k^{n,\delta}(x)$, the probability of indeed reaching the precision ϵ goes to zero exponentially in the sample size n. Notice that we do not make any special assumption on the dependence-independence of the families of random variables $\{\zeta_{k,i}^n\}_{n\ge1}$ as i varies because we use the inequality

$$\mathbb{P}\Big\{\max_{1\le i\le m_k}\{\mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n)\} > \frac{\epsilon}{3\mathbf{H}_k}\Big\} \le \sum_{i=1}^{m_k}\mathbb{P}\Big\{\mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n) > \frac{\epsilon}{3\mathbf{H}_k}\Big\},$$

which holds whatever the dependence-independence is.

4.3 Approximation of an optimal policy

Now we are interested in providing computable nearly optimal policies for the control model $\mathcal{M}_k$. *In the sequel, we suppose that the control model $\mathcal{M}$ satisfies Assumptions A and B and so, for each fixed $k \in \mathbb{N}$, the control model $\mathcal{M}_k$ verifies Theorem 4.1.* Our next lemma uses the notation $R_k^n v$ (recall (1.1)).

Lemma 4.13 *The following statements hold for every $n \ge 1$.*

(i) *The function $(x, a) \mapsto R_k^n(\mathbf{X}_k|x, a)$ is in $\mathbb{L}(\mathrm{Gr}(\mathbf{Y}_1), \ldots, \mathrm{Gr}(\mathbf{Y}_{m_k}))$ with Lipschitz constant $2\mathcal{L}_j^{p_k}/\overline{q}_k$ on $\mathrm{Gr}(\mathbf{Y}_j)$, for $1 \le j \le m_k$.*

(ii) *For each $v \in \mathbb{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_{m_k})$ we have $R_k^n v \in \mathbb{L}(\mathrm{Gr}(\mathbf{Y}_1), \ldots, \mathrm{Gr}(\mathbf{Y}_{m_k}))$ with*

$$|R_k^n v(x, a) - R_k^n v(y, b)| \le \frac{2\mathcal{L}_j^{p_k}||v||}{\overline{q}_k}\big(d_{\mathbf{X}}(x, y) + d_{\mathbf{A}}(a, b)\big) + L_j^v d_{\mathbf{X}}(x, y)$$

for (x, a) and (y, b) in $\mathrm{Gr}(\mathbf{Y}_j)$, for $1 \le j \le m_k$.

(iii) *Define* $\mathbf{m}_k^n = \inf_{(x,a)\in \mathrm{Gr}(\mathbf{X}_k)} R_k^n(\mathbf{X}_k|x,a)$. *If* $v \in \mathbb{L}(\mathbf{Y}_1,\ldots,\mathbf{Y}_{m_k})$ *then we have that* $Q_k^n v \in \mathbb{L}(\mathrm{Gr}(\mathbf{Y}_1),\ldots,\mathrm{Gr}(\mathbf{Y}_{m_k}))$ *with*

$$\begin{aligned} |Q_k^n v(x,a) - Q_k^n v(y,b)| \;\leq\; & \frac{2\mathcal{L}_j^{p_k}||v||}{\overline{q}_k^2(\mathbf{m}_k^n)^2}\big(||p_k|| + \overline{q}_k(1+\mathbf{m}_k^n)\big)\cdot\big(d_{\mathbf{X}}(x,y) + d_{\mathbf{A}}(a,b)\big) \\ & + (L_j^v/\mathbf{m}_k^n)\cdot d_{\mathbf{X}}(x,y) \end{aligned}$$

for (x,a) *and* (y,b) *in* $\mathrm{Gr}(\mathbf{Y}_j)$, *for* $1 \leq j \leq m_k$.

Proof. (i). This part is derived from the definition of μ_k^n (in particular it makes use of the fact that $\mu_k(\mathbf{Y}_i) = \mu_k^n(\mathbf{Y}_i)$) and it follows the same technique as Lemma 4.2(ii). Note that $||R_k^n(\mathbf{X}_k|\cdot,\cdot)|| \leq ||p_k||/\overline{q}_k + 1$.

(ii). This statement is derived as Lemma 4.2(ii), with $||R_k^n v|| \leq ||v||\cdot\big(||p_k||/\overline{q}_k + 1\big)$.

(iii). Observe that we have $\mathbf{m}_k^n \geq \eta_k > 0$, by (4.2). Using the fact that $Q_k^n v$ is the quotient of two bounded and Lipschitz continuous functions on $\mathrm{Gr}(\mathbf{Y}_j)$, the stated result follows after some elementary calculations. □

For every $\delta > 0$ and $x \in \mathbf{X}_k$, consider the finite sets $\mathbf{A}_\delta(x) \subseteq \mathbf{A}(x)$ as in Definition 4.7 which satisfy, in addition, the piecewise Lipschitz continuity condition in Theorem 4.1(iv). Given $n \geq 1$ and $\delta > 0$, there exists a deterministic stationary policy in $\mathcal{U}_k^s$, which we will identify with some measurable function $\varphi_k^{n,\delta} : \mathbf{X}_k \to \mathbf{A}$, such that (recall the fixed point equation for $T_k^{n,\delta}$ in (4.10))

$$\begin{aligned} \mathcal{V}_k^{n,\delta}(x) &= \min_{a\in\mathbf{A}_\delta(x)}\Big\{\frac{c_k(x,a)}{\alpha+\overline{q}_k} + \frac{\overline{q}_k}{\alpha+\overline{q}_k}\int_{\mathbf{X}_k}\mathcal{V}_k^{n,\delta}(y)Q_k^n(dy|x,a)\Big\} \\ &= \frac{c_k(x,\varphi_k^{n,\delta}(x))}{\alpha+\overline{q}_k} + \frac{\overline{q}_k}{\alpha+\overline{q}_k}\int_{\mathbf{X}_k}\mathcal{V}_k^{n,\delta}(y)Q_k^n(dy|x,\varphi_k^{n,\delta}(x)) \quad \text{for } x\in\mathbf{X}_k. \end{aligned} \tag{4.11}$$

The existence of such measurable selector is a direct consequence of Proposition D.5 in [8].

Our next result is analogous to Proposition 4.4 but now for the operator $T_k^{n,\delta}$.

Proposition 4.14 *Suppose that the control model satisfies Assumptions A and B. Consider the control model* $\mathcal{M}_k$, *which satisfies Theorem 4.1, and let* $n \geq 1$ *and* $\delta > 0$.

(i) *If* $v \in \mathbb{L}(\mathbf{Y}_1,\ldots,\mathbf{Y}_{m_k})$ *then* $T_k^{n,\delta}v \in \mathbb{L}(\mathbf{Y}_1,\ldots,\mathbf{Y}_{m_k})$, *with Lipschitz constant*

$$\frac{1}{\alpha+\overline{q}_k}\left(\Big(L_j^{c_k} + \frac{2\mathcal{L}_j^{p_k}||v||}{\overline{q}_k(\mathbf{m}_k^n)^2}(||p_k|| + \overline{q}_k(1+\mathbf{m}_k^n))\Big)(1+L_j^{\Psi_k}) + \overline{q}_k L_j^v/\mathbf{m}_k^n\right).$$

on each set $\mathbf{Y}_j$, *for* $j = 1,\ldots,m_k$.

(ii) *Suppose that* $n \geq 1$ *is such that*

$$\mathcal{W}^*(\mu_k,\mu_k^n) < \frac{1}{\max\limits_{1\leq j\leq m_k}\mathcal{L}_j^{p_k}}\cdot\frac{\alpha\overline{q}_k}{\alpha+\overline{q}_k}. \tag{4.12}$$

In this case, $\mathcal{V}_k^{n,\delta}$ is in $\mathbb{L}(\mathbf{Y}_1,\dots,\mathbf{Y}_{m_k})$ and there exists $K>0$ such that the Lipschitz constant of $\mathcal{V}_k^{n,\delta}$ is

$$\frac{K}{\alpha}\Big(L_j^{c_k}+\frac{2\mathcal{L}_j^{p_k}||c_k||(\alpha+\overline{q}_k)}{\alpha\overline{q}_k^3}\big(||p_k||(\alpha+\overline{q}_k)+\overline{q}_k(\alpha+2\overline{q}_k)\big)\Big)(1+L_j^{\Psi_k})$$

on $\mathbf{Y}_j$ for each $1\le j\le m_k$.

Proof. (i). This statement is similar to Proposition 4.4(i) and its proof is omitted. It makes use of piecewise Lipschitz continuity of $x\mapsto\mathbf{A}_\delta(x)$ in Theorem 4.1(iv).
(ii). By the hypothesis on $\mathcal{W}^*(\mu_k,\mu_k^n)$ we deduce that there exists $\epsilon>0$ such that

$$\mathcal{W}^*(\mu_k,\mu_k^n)\le\frac{1}{\max\limits_{1\le j\le m_k}\mathcal{L}_j^{p_k}}\cdot\frac{\alpha\overline{q}_k}{\alpha+(1+\epsilon)\overline{q}_k}.$$

Recalling now Lemma 4.6(i), when $(x,a)\in\mathrm{Gr}(\mathbf{Y}_j)$ for some $1\le j\le m_k$ we have

$$R_k^n(\mathbf{X}_k|x,a)\ge 1-(\mathcal{L}_j^{p_k}/\overline{q}_k)\cdot\mathcal{W}^*(\mu_k,\mu_k^n)\ge\frac{(1+\epsilon)\overline{q}_k}{\alpha+(1+\epsilon)\overline{q}_k},$$

and so

$$\mathbf{m}_k^n\ge\frac{(1+\epsilon)\overline{q}_k}{\alpha+(1+\epsilon)\overline{q}_k}\ge\frac{\overline{q}_k}{\alpha+\overline{q}_k}.$$

Using these bounds, it follows from part (i) that, given $v\in\mathbb{L}(\mathbf{Y}_1,\dots,\mathbf{Y}_{m_k})$, the constant

$$\frac{1}{\alpha+\overline{q}_k}\Big(L_j^{c_k}+\frac{2\mathcal{L}_j^{p_k}||v||(\alpha+\overline{q}_k)}{\overline{q}_k^3}(||p_k||(\alpha+\overline{q}_k)+\overline{q}_k(\alpha+2\overline{q}_k))\Big)(1+L_j^{\Psi_k})+\frac{\alpha+(1+\epsilon)\overline{q}_k}{(1+\epsilon)(\alpha+\overline{q}_k)}L_j^v$$

is a Lipschitz constant for $T_k^{n,\delta}v$ on the set $\mathbf{Y}_j$, for any $1\le j\le m_k$. Since the coefficient of L_j^v is strictly less than one, we can proceed as in Proposition 4.4(ii) and, recalling that $||\mathcal{V}_k^{n,\delta}||\le||c_k||/\alpha$, we obtain that $\mathcal{V}_k^{n,\delta}\in\mathbb{L}(\mathbf{Y}_1,\dots,\mathbf{Y}_{m_k})$ and that

$$\frac{1+\epsilon}{\alpha\epsilon}\Big(L_j^{c_k}+\frac{2\mathcal{L}_j^{p_k}||c_k||(\alpha+\overline{q}_k)}{\alpha\overline{q}_k^3}\big(||p_k||(\alpha+\overline{q}_k)+\overline{q}_k(\alpha+2\overline{q}_k)\big)\Big)(1+L_j^{\Psi_k})$$

is a Lipschitz constant for $\mathcal{V}_k^{n,\delta}$ on $\mathbf{Y}_j$ for $1\le j\le m_k$. Letting $K=(1+\epsilon)/\epsilon$, the stated result follows. □

Remark that the condition (4.12) in Proposition 4.14(ii) is satisfied with probability close to one provided that n is large enough. Indeed, we know from Proposition 1.1 that $\mathcal{W}_1(\nu_{k,i},\nu_{k,i}^n)$ converges to zero in probability as $n\to\infty$ for every $1\le i\le m_k$ and, therefore, so does $\mathcal{W}^*(\mu_k,\mu_k^n)=\max_{1\le i\le m_k}\mathcal{W}_1(\nu_{k,i},\nu_{k,i}^n)$.

Proposition 4.15 *Suppose that the control model $\mathcal{M}$ satisfies Assumptions A and B. Given $k\in\mathbb{N}$ consider the control model $\mathcal{M}_k$. Fix $\delta>0$, let $n\ge1$ be such that (cf. (4.12))*

$$\mathcal{W}^*(\mu_k,\mu_k^n)\le\frac{1}{\max\limits_{1\le j\le m_k}\mathcal{L}_j^{p_k}}\cdot\frac{\alpha\overline{q}_k}{\alpha+2\overline{q}_k},\tag{4.13}$$

and let the policy $\varphi_k^{n,\delta}\in\mathcal{U}_k^s$ be as in (4.11). There exist positive constants $\mathbf{G}_k'$ and $\mathbf{H}_k'$ (they depend neither on $n\ge1$ nor on $\delta>0$) such that

$$\mathcal{V}_k^*(x)\le\mathcal{V}_k(\varphi_k^{n,\delta},x)\le\mathcal{V}_k^*(x)+\mathbf{G}_k'\cdot\delta+\mathbf{H}_k'\cdot\mathcal{W}^*(\mu_k,\mu_k^n)\quad\text{for each } x\in\mathbf{X}_k.$$

Proof. Let $n \geq 1$ be such that (4.13) holds and let $\delta > 0$. We have (for $\epsilon = 1$ and $K = 2$ in the proof of Proposition 4.14(ii)) that

$$\frac{2}{\alpha}\Big(L_j^{c_k} + \frac{2\mathcal{L}_j^{p_k}||c_k||(\alpha+\overline{q}_k)}{\alpha\overline{q}_k^3}\big(||p_k||(\alpha+\overline{q}_k) + \overline{q}_k(\alpha+2\overline{q}_k)\big)\Big)(1+L_j^{\Psi_k})$$

is a Lipschitz constant for $\mathcal{V}_k^{n,\delta}$ on $\mathbf{Y}_j$, for $1 \leq j \leq m_k$. So,

$$\mu_k\big(L^{\mathcal{V}_k^{n,\delta}}\big) = \frac{2}{\alpha}\sum_{j=1}^{m_k}\mu_k(\mathbf{Y}_j)\Big(L_j^{c_k} + \frac{2\mathcal{L}_j^{p_k}||c_k||(\alpha+\overline{q}_k)}{\alpha\overline{q}_k^3}\big(||p_k||(\alpha+\overline{q}_k) + \overline{q}_k(\alpha+2\overline{q}_k)\big)\Big)(1+L_j^{\Psi_k})$$

is the same for all $\delta > 0$ and all such n. We will thus write $\mathcal{L}_k = \mu_k\big(L^{\mathcal{V}_k^{n,\delta}}\big)$. We can apply Lemma 4.6(ii) to obtain

$$\begin{aligned}\mathcal{V}_k^{n,\delta}(x) \;\geq\; & \frac{c_k(x,\varphi_k^{n,\delta}(x))}{\alpha+\overline{q}_k} + \frac{\overline{q}_k}{\alpha+\overline{q}_k}Q_k\mathcal{V}_k^{n,\delta}(x,\varphi_k^{n,\delta}(x)) \\ & - \frac{\mathcal{L}_k||p_k|| + 2||c_k||\max_j \mathcal{L}_j^{p_k}/\alpha}{\alpha+\overline{q}_k}\cdot\mathcal{W}^*(\mu_k,\mu_k^n)\end{aligned}$$

for all $x \in \mathbf{X}_k$ or, equivalently,

$$\begin{aligned}\alpha\mathcal{V}_k^{n,\delta}(x) \;\geq\; & c_k(x,\varphi_k^{n,\delta}(x)) + \int_{\mathbf{X}_k}\mathcal{V}_k^{n,\delta}(y)q_k(dy|x,\varphi_k^{n,\delta}(x)) \\ & - \Big(\mathcal{L}_k||p_k|| + 2||c_k||\max_j \mathcal{L}_j^{p_k}/\alpha\Big)\cdot\mathcal{W}^*(\mu_k,\mu_k^n)\end{aligned}$$

for all $x \in \mathbf{X}_k$. Now we are going to use Theorem 2.6(iv) for the control model $\mathcal{M}_k$ (which indeed satisfies Assumption A because it is bounded), the function $\mathcal{V}_k^{n,\delta}$, the control policy $\varphi_k^{n,\delta} \in \mathcal{U}_k^s$, and any initial state $x \in \mathbf{X}_k$. To simplify the notation, we will simply denote by $\mathbb{E}_x$ the expectation operator corresponding to the policy $\varphi_k^{n,\delta}$ and the initial state $x \in \mathbf{X}_k$. We obtain that for all $t \geq 0$

$$\begin{aligned}&\mathbb{E}_x[e^{-\alpha t}\mathcal{V}_k^{n,\delta}(\xi_t)] - \mathcal{V}_k^{n,\delta}(x) \\ &= \mathbb{E}_x\Big[\int_0^t e^{-\alpha s}\big[-\alpha\mathcal{V}_k^{n,\delta}(\xi_s) + \int_{\mathbf{X}_k}\mathcal{V}_k^{n,\delta}(y)q_k(dy|\xi_s,\varphi_k^{n,\delta}(\xi_s))\big]ds\Big] \\ &\leq -\mathbb{E}_x\Big[\int_0^t e^{-\alpha s}c_k(\xi_s,\varphi_k^{n,\delta}(\xi_s))ds\Big] + \Big(\mathcal{L}_k||p_k||/\alpha + 2||c_k||\max_j\mathcal{L}_j^{p_k}/\alpha^2\Big)\cdot\mathcal{W}^*(\mu_k,\mu_k^n).\end{aligned}$$

Letting $t \to \infty$ yields

$$\mathcal{V}_k(\varphi_k^{n,\delta},x) \leq \mathcal{V}_k^{n,\delta}(x) + \Big(\mathcal{L}_k||p_k||/\alpha + 2||c_k||\max_j\mathcal{L}_j^{p_k}/\alpha^2\Big)\cdot\mathcal{W}^*(\mu_k,\mu_k^n).$$

Recalling the bound on $|\mathcal{V}_k^{n,\delta}(x) - \mathcal{V}_k^*(x)|$ given in Proposition 4.11 and letting

$$\mathbf{G}_k' = \mathbf{G}_k \quad\text{and}\quad \mathbf{H}_k' = \mathbf{H}_k + \mathcal{L}_k||p_k||/\alpha + 2||c_k||\max_{1\leq j\leq m_k}\mathcal{L}_j^{p_k}/\alpha^2$$

we obtain the result. □

As a consequence of Remark 4.8, it is possible to determine $\varphi_k^{n,\delta}(x)$ explicitly for each $x \in \mathbf{X}_k$. Therefore, given $\epsilon > 0$ and by choosing $n \geq 1$ with $\mathcal{W}^*(\mu_k, \mu_k^n)$ small enough and small enough $\delta > 0$, we can explicitly compute a deterministic stationary policy in $\mathcal{U}_k^s$ that is ϵ-optimal for the control model $\mathcal{M}_k$ for any initial state.

Extend the deterministic stationary policy $\varphi_k^{n,\delta} : \mathbf{X}_k \to \mathbf{A}$ for the control model $\mathcal{M}_k$ to any deterministic stationary policy $\tilde{\varphi}_k^{n,\delta} : \mathbf{X} \to \mathbf{A}$ for the control model $\mathcal{M}$, that is, such that $\mathbf{p}_k(\tilde{\varphi}_k^{n,\delta}) = \varphi_k^{n,\delta}$.

Theorem 4.16 *Suppose that the control model $\mathcal{M}$ satisfies Assumptions A and B. Let $x \in \mathbf{X}$ be the initial state of the system and let $\epsilon > 0$ be some given precision. There exist $k \in \mathbb{N}$, $\delta > 0$, and positive constants $C' = C'(\epsilon, k)$ and $D' = D'(\epsilon, k)$ such that the deterministic stationary policy $\tilde{\varphi}_k^{n,\delta} \in \mathcal{U}^s$ verifies*

$$\mathbb{P}\big\{\mathcal{V}(\tilde{\varphi}_k^{n,\delta}, x) - \mathcal{V}^*(x) > \epsilon\big\} \leq C' e^{-D' n} \quad \textit{for all } n \geq 1.$$

Proof. Choose $k \in \mathbb{N}$ large enough so that $x \in \mathbf{S}_k$ and

$$\frac{\mathbf{m} M(w^{1+\beta}(x) + b/\alpha)}{\inf\{w^\beta(y) : y \notin \mathbf{S}_k\}} \leq \epsilon/4.$$

For such $k \in \mathbb{N}$, choose $\delta > 0$ with $\mathbf{G}'_k \cdot \delta \leq \epsilon/4$. Suppose for a moment that n is such that (cf. (4.13))

$$\mathcal{W}^*(\mu_k, \mu_k^n) \leq \frac{\epsilon}{4\mathbf{H}'_k} \wedge \Big(\frac{1}{\max\limits_{1 \leq j \leq m_k} \mathcal{L}_j^{p_k}} \cdot \frac{\alpha \overline{q}_k}{\alpha + 2\overline{q}_k}\Big),$$

and consider the deterministic stationary policies $\varphi_k^{n,\delta} \in \mathcal{U}_k^s$ and $\tilde{\varphi}_k^{n,\delta} \in \mathcal{U}^s$. We would have, on one hand, by Theorem 3.4,

$$|\mathcal{V}^*(x) - \mathcal{V}_k^*(x)| \leq \epsilon/4 \quad \text{and} \quad |\mathcal{V}(\tilde{\varphi}_k^{n,\delta}, x) - \mathcal{V}_k(\varphi_k^{n,\delta}, x)| \leq \epsilon/4$$

and, on the other hand, by Proposition 4.15,

$$\mathcal{V}_k(\varphi_k^{n,\delta}, x) \leq \mathcal{V}_k^*(x) + \epsilon/2.$$

Consequently, we would have $\mathcal{V}(\tilde{\varphi}_k^{n,\delta}, x) \leq \mathcal{V}^*(x) + \epsilon$. We have thus proved the following inclusion

$$\big\{\mathcal{V}(\tilde{\varphi}_k^{n,\delta}, x) - \mathcal{V}^*(x) > \epsilon\big\} \subseteq \Big\{\mathcal{W}^*(\mu_k, \mu_k^n) > \frac{\epsilon}{4\mathbf{H}'_k} \wedge \Big(\frac{1}{\max\limits_{1 \leq j \leq m_k} \mathcal{L}_j^{p_k}} \cdot \frac{\alpha \overline{q}_k}{\alpha + 2\overline{q}_k}\Big)\Big\}$$

and so $\mathbb{P}\big\{\mathcal{V}(\tilde{\varphi}_k^{n,\delta}, x) - \mathcal{V}^*(x) > \epsilon\big\}$ is less than or equal to

$$\sum_{i=1}^{m_k} \mathbb{P}\Big\{\mathcal{W}_1(\nu_{k,i}, \nu_{k,i}^n) > \frac{\epsilon}{4\mathbf{H}'_k} \wedge \Big(\frac{1}{\max\limits_{1 \leq j \leq m_k} \mathcal{L}_j^{p_k}} \cdot \frac{\alpha \overline{q}_k}{\alpha + 2\overline{q}_k}\Big)\Big\}.$$

Proceeding as in Theorem 4.12 and by Proposition 1.1, we obtain the result. □

Consequently, given some precision $\epsilon > 0$ and an initial state $x \in \mathbf{X}$, we can find $k \in \mathbb{N}$ and $\delta > 0$ such that the probability that the policy $\tilde{\varphi}_k^{n,\delta}$ is not ϵ-optimal for $\mathcal{M}$ goes to zero

exponentially in the sample size n of the empirical probability measures. Moreover, for each fixed k, n, and δ, the action prescribed by the policy $\tilde{\varphi}_k^{n,\delta}$ can be explicitly computed for any state in $\mathbf{X}$. Summarizing, starting from the control model $\mathcal{M}_k$ and by discretizing its state and actions sets (parameters n and δ) we can construct a policy in $\mathcal{U}^s$ which, when "plugged" intro the dynamics of the original control model $\mathcal{M}$, yields an ϵ-optimal policy with probability close to one.

5 Example

We study the model proposed in [11, Section 3.3].

Definition. This example describes a one-channel queuing system where any job that finds the server busy is rejected. Every job is characterized by its volume $x \in (0, 1]$, so that the state space is $\mathbf{X} = [0, 1]$. Hence, $\xi_t = 0$ when the system is idle and $\xi_t \in (0, 1]$ means that the corresponding job is under service. The action space is $\mathbf{A} = \mathbb{R}_+$ and an action represents the service intensity. Let

$$\mathbf{A}(0) = \{0\} \quad \text{and} \quad \mathbf{A}(x) = \big[0, \overline{A}/x\big], \quad \text{for } 0 \leq x < 1,$$

where $\overline{A}$ is a positive constant. The jobs arrive according to a Poisson process with constant intensity $\lambda > 0$, and the volume is distributed according to the density $5x^4$ for $x \in (0, 1]$. The transition rate q can be written as

$$q(dy|x, a) = \begin{cases} 5\lambda y^4 dy - \lambda\delta_0(dy) & \text{when } (x, a) = (0, 0) \\ \frac{a}{x}\big[\delta_0(dy) - \delta_x(dy)\big] & \text{when } 0 < x \leq 1 \text{ and } a \in \mathbf{A}(x), \end{cases}$$

where dy represents the Lebesgue measure on $\mathbf{X}$. In particular, we have

$$\overline{q}(0) = \lambda \quad \text{and} \quad \overline{q}(x) = \frac{\overline{A}}{x^2} \quad \text{for } 0 < x \leq 1.$$

The cost rate function is

$$c(x, a) = C_1 x + C_2 a^2 - \frac{a}{x} \quad \text{for } 0 < x \leq 1 \text{ and } a \in \mathbf{A}(x),$$

and $c(0, 0) = 0$, for some positive constants C_1 and C_2. The goal of the controller is to minimize the total expected α-discount cost, for some discount rate $\alpha > 0$.

Assumptions. Let us verify that this control model satisfies our assumptions. For Assumption A, consider the function w on $\mathbf{X}$ defined as

$$w(0) = 1 \quad \text{and} \quad w(x) = \frac{1}{x^2} \quad \text{for } 0 < x \leq 1,$$

and the sets

$$\mathbf{S}_k = \{0\} \cup \big[1/k, 1\big] \quad \text{for } k \geq 1.$$

Assumptions (A1), (A3), and (A4) are straightforward. Regarding Assumptions (A2) and (A4), observe that if $v : \mathbf{X} \to [1, \infty)$ is any measurable function with $v(0) = 1$ and $\int_0^1 v(y)y^4 dy$ finite, then

$$b_v = \int_{\mathbf{X}} v(y)q(dy|0,0) = \lambda\Big(5\int_0^1 v(y)y^4 dy - 1\Big) \geq 0$$

and

$$\int_{\mathbf{X}} v(y)q(dy|x,a) = \frac{a}{x}\big(v(x) - 1\big) \leq 0 \quad \text{for } 0 < x \leq 1 \text{ and } a \in \mathbf{A}(x).$$

Therefore, the following inequality holds for all $(x, a) \in \mathbf{K}$:

$$\int_{\mathbf{X}} v(y)q(dy|x,a) \leq 0 \cdot v(x) + b_v.$$

By choosing v of the form $v(0) = 1$ and $v(x) = x^{-\gamma}$ for $0 < x \leq 1$, we must have $0 \leq \gamma < 5$ to satisfy the above conditions. Consequently, Assumption (A2) holds for $\beta = 1/4$, $\rho = 0$, and $b = \lambda$, while Assumption (A5) is satisfied by the function w' with

$$w'(0) = 1 \quad \text{and} \quad w'(x) = x^{-9/2} \quad \text{for } 0 < x \leq 1,$$

and the constants $\rho' = 0$ and $b' = 9\lambda$.

Let us now check Assumption B. While Assumption (B1) is trivial, for (B2) consider the measure μ on X given by $\mu(dy) = \delta_0(dy) + dy$, with dy the Lebesgue measure on $\mathbf{X}$, and the function p given by

$$p(y|0,0) = 5\lambda y^4 \quad \text{for } y \in \mathbf{Y}, \quad p(0|x,a) = a/x \quad \text{for } 0 \leq x < 1 \text{ and } a \in \mathbf{A}(x),$$

and p is zero otherwise. With these definitions, Assumptions (B2) and (B3) are satisfied. Given $k \geq 1$, consider the partition $\mathbf{Z}_1 \cup \mathbf{Z}_2$ of the set $\mathbf{S}_k$, with $\mathbf{Z}_1 = \{0\}$ and $\mathbf{Z}_2 = [1/k, 1]$. It is easy to see that Assumption (B4) holds. Note, in particular, that

$$d_H(\mathbf{A}(x), \mathbf{A}(x')) = \overline{A} \cdot \big|1/x - 1/x'\big| \leq \overline{A}k^2 \cdot |x - x'| \quad \text{for } 1/k \leq x, x' \leq 1.$$

Finally, given $k \geq 1$ and $\delta > 0$, define the sets $\mathbf{A}_\delta(0) = \{0\}$ and, for $1/k \leq x \leq 1$,

$$\mathbf{A}_\delta(x) = \Big\{\frac{\overline{A}}{x} \cdot \frac{j}{r_{k,\delta}}\Big\}_{j=0,1,\ldots,r_{k,\delta}}$$

with $r_{k,\delta}$ equal to smallest integer larger than or equal to $\frac{\overline{A}k}{2\delta}$. With this definition, we indeed have $d_H(\mathbf{A}(x), \mathbf{A}_\delta(x)) \leq \delta$ for all $x \in \mathbf{S}_k$. Moreover, for $x, x' \in \mathbf{Z}_2$ we have

$$d_H(\mathbf{A}_\delta(x), \mathbf{A}_\delta(x')) = d_H(\mathbf{A}(x), \mathbf{A}(x')) \leq \overline{A}k^2 \cdot |x - x'|.$$

This shows that Assumption (B5) holds. The proof that the control model verifies Assumption B is now complete.

Approximation of the control model. Given $k \geq 1$, let us now describe the control model $\mathcal{M}_k$. Its state space is

$$\mathbf{X}_k = \{0\} \cup [1/k, 1] \cup \{x_\Delta\},$$

with arbitrary $0 < x_\Delta < 1/k$. Recalling the statement of Theorem 4.1, we consider the partition of the state space $\mathbf{X}_k$ given by

$$\mathbf{Y}_0 = \{0\}, \quad \mathbf{Y}_1 = [1/k, 1], \quad \mathbf{Y}_2 = \{x_\Delta\}.$$

We consider the probability measure μ_k defined as

$$\mu_k(\Gamma) = \gamma \cdot \mathbf{I}_\Gamma(0) + (1 - \gamma - \eta) \cdot \Lambda(\Gamma \cap [1/k, 1]) + \eta \cdot \mathbf{I}_\Gamma(x_\Delta),$$

for measurable $\Gamma \subseteq \mathbf{X}_k$, with Λ is the probability measure on $[1/k, 1]$ with density function $\frac{5y^4}{1-k^{-5}}$, and where $\eta, \gamma > 0$ are arbitrary constants with $\eta + \gamma < 1$. Define the function p_k on $\mathbf{X}_k \times \mathrm{Gr}(\mathbf{X}_k)$ as

- $p_k(y|0,0) = \lambda(1 - \frac{1}{k^5})/(1 - \gamma - \eta)$ when $y \in \mathbf{Y}_1$.
- $p_k(x_\Delta|0,0) = \frac{\lambda}{\eta k^5}$.
- $p_k(0|x,a) = \frac{a}{x\gamma}$ when $(x,a) \in \mathrm{Gr}(\mathbf{Y}_1)$.

Otherwise, $p_k(y|x,a) = 0$. With this definition and recalling the expression of q_k as introduced in item (i) in Definition 3.1, it is easily seen that $q^+(dy|x,a) = p_k(y|x,a)\mu_k(dy)$. Observe that $q_k(\{x\}|x,a) = -a/x$ when $(x,a) \in \mathrm{Gr}(\mathbf{Y}_1)$ and that $q_k(\{0\}|0,0) = -\lambda$. So, we let $\overline{q}_k = 0.1 + \max\{\lambda, \overline{A}k^2\}$.

The probability measures $\nu_{k,i}$ are as follows: $\nu_{k,0}$ and $\nu_{k,2}$ are Dirac measures on 0 and x_Δ, respectively, while $\nu_{k,1}$ is distributed as Λ. When taking samples, $\nu^n_{k,0}$ and $\nu^n_{k,2}$ do not change, and $\nu^n_{k,1}$ is the empirical measure of a sample of size n of the law Λ. Hence, we can write

$$\mu^n_k(dy) = \gamma\delta_0(dy) + \frac{1 - \gamma - \eta}{n} \sum_{i=1}^n \delta_{Z_i}(dy) + \eta \cdot \delta_{x_\Delta}(dy).$$

for $Z_1, \ldots, Z_n$ i.i.d. random variables distributed as Λ. Therefore, the transition probability measure $Q^n_k(dy|x,a)$ is

- Starting from the state 0 and the action 0:

$$Q^n_k(dy|0,0) = \frac{1}{\overline{q}_k} \cdot \Big((\overline{q}_k - \lambda)\delta_0(dy) + \frac{\lambda(1 - \frac{1}{k^5})}{n} \sum_{i=1}^n \delta_{Z_i}(dy) + \frac{\lambda}{k^5}\delta_{x_\Delta}(dy) \Big).$$

- Starting from a state $(x,a) \in \mathrm{Gr}(\mathbf{Y}_1)$,

$$Q^n_k(\{0\}|x,a) = \frac{a}{x\overline{q}_k} \quad \text{and} \quad Q^n_k(\{x\}|x,a) = 1 - \frac{a}{x\overline{q}_k}.$$

- $Q^n_k(dy|x_\Delta, a)$ is concentrated on x_Δ.

Note that these expressions do not depend on the parameters η, γ.

Value of j	1	2	3	4	5
Mean	-1.5210	-1.5489	-1.5511	-1.5536	-1.5493
Std. deviation	0.1339	0.0992	0.0757	0.0660	0.0477
Coeff. var. (%)	8.8028	6.4021	4.8821	4.2457	3.0794

Table 1: Estimation of the optimal value.

Numerical results. We fix the following values of the parameters of the control model

$$\lambda = 3 \quad \overline{A} = 1 \quad C_1 = 0.1 \quad C_2 = 0.05 \quad \alpha = 0.5.$$

We are interested in approximating the optimal discounted value for the initial state 0, that is, the value of $\mathcal{V}^*(0)$.

For the approximations, we will consider parameters of the form

$$k = 3 \cdot j \quad n = 10 \cdot j \quad \delta = 1/j$$

for $j = 1, \ldots, 5$, in order to study the precision of the numerical method as k and n grow, and δ decreases.

For a given $j = 1, \ldots 5$, fix the values of k, n, δ as above. We simulate a sample $Z_1, \ldots, Z_n$ of the distribution Λ and we obtain the value of $\mathcal{V}_k^{n,\delta}$ on the finite set $\Gamma_k^n = \{0, Z_1, \ldots, Z_n, x_\Delta\}$; recall Remark 4.8. We do this by solving the equivalent linear programming formulation of the fixed point equation for the operator $T_k^{n,\delta}$ on Γ_k^n. Such calculations are performed for every particular sample $Z_1, \ldots, Z_n$ of size n. We do this for 100 independent samples, so that we obtain 100 independent realizations of the random variable $\mathcal{V}_k^{n,\delta}(0)$. We make a simple statistical analysis of the results in Table 1.

We observe that the average value stabilizes around -1.55 with standard deviations getting smaller. The coefficient of variation shows a very nice behavior, starting at almost 9% and reaching 3%, thus showing that the distribution of $\mathcal{V}_k^{n,\delta}(0)$ becomes more concentrated as j grows.

For each $j = 1, \ldots, 5$ we make a density estimation of the distribution of $\mathcal{V}_k^{n,\delta}(0)$ for the 100 samples we have obtained. We use the function `ksdensity` of Matlab, which uses normal kernel estimators. In Figure 1 it becomes clear that the distribution of $\mathcal{V}_k^{n,\delta}(0)$ becomes more concentrated as k and n grow and δ decreases, thus providing more accurate estimators.

Regarding the estimation of an optimal policy, for every particular sample $Z_1, \ldots, Z_n$, we obtain the policy $\varphi_k^{n,\delta} \in \mathcal{U}_k^s$ defined on the states of $\mathbf{X}_k$ as in (4.11), which is just the minimizer in the fixed point equation of $T_k^{n,\delta}$, and we let $\varphi_k^{n,\delta}(x_\Delta) = 0$. We extend this policy to a policy in $\mathcal{U}^s$ by letting $\tilde{\varphi}_k^{n,\delta}(x) = 0$ for $0 < x < 1/k$. We simulate a sample path of the process $\{\xi_t\}_{0 \le t \le T}$ with $\xi_0 = 0$ when using this policy, until the "large" time horizon $T = 5000$, and then we compute the total discounted cost until $T = 5000$. We do this 200 times, so as to obtain the estimator of the total expected discounted cost under the policy $\tilde{\varphi}_k^{n,\delta}$, that is, $\mathcal{V}(\tilde{\varphi}_k^{n,\delta}, 0)$, by averaging these 200 estimates. This is repeated for 100 independent samples $Z_1, \ldots, Z_n$. We provide the following statistical analysis of these 100 estimations in Table 2.

We note that the mean of the sample becomes stable around -1.55 and -1.56, with a very small coefficient of variation (around 2%) for all the values of j. The fact that these figures are quite stable and that they do not seem to "depend" on j show that, even for small values

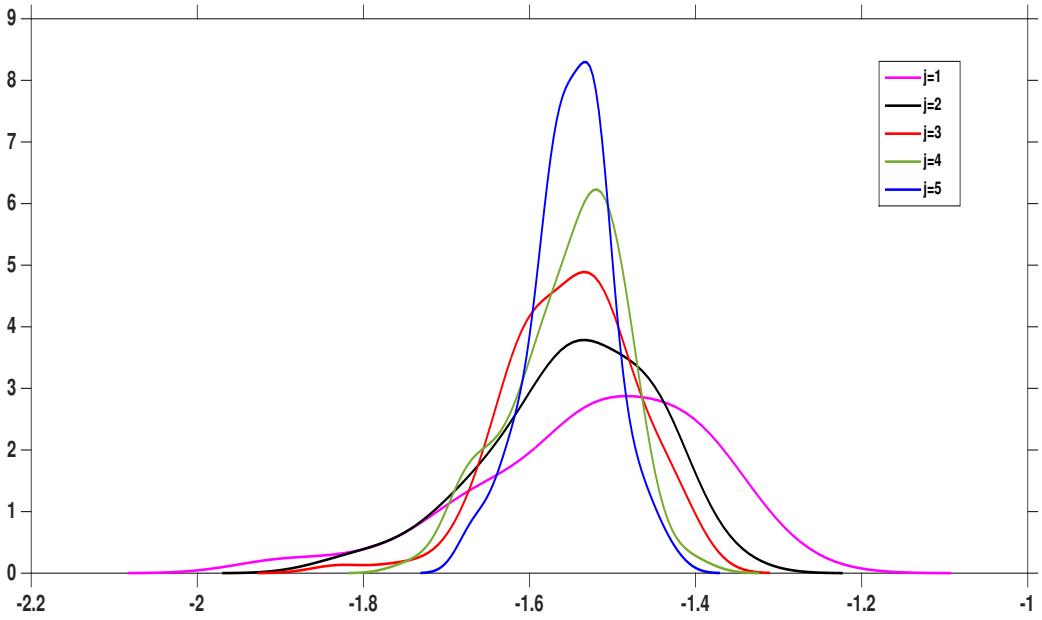

Figure 1: Density estimators.

Value of j	1	2	3	4	5
Mean	-1.5303	-1.5522	-1.5530	-1.5601	-1.5612
Std. deviation	0.0331	0.0320	0.0307	0.0343	0.0356
Coeff. var. (%)	2.1610	2.0603	1.9750	2.1963	2.2833

Table 2: Estimation of the optimal value by using the policy $\varphi_k^{n,\delta}$.

of j, we have a good estimator of the optimal policy. Thus, when evaluating this policy on the real dynamics of the control model, we obtain precise estimators somehow regardless of the goodness of the estimator $\mathcal{V}_k^{n,\delta}(0)$.

To conclude, we observe empirically that our numerical method gives very accurate estimations of the optimal value of the control problem, and that the estimation of an optimal policy is very precise as well, even for the "early" values of the parameters n, k, and δ.

References

[1] D.P. Bertsekas and J.N. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, Belmont, Massachusetts, 1996.

[2] Hyeong Soo Chang, Michael C. Fu, Jiaqiao Hu, and Steven I. Marcus. *Simulation-based algorithms for Markov decision processes.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 2007.

[3] François Dufour and Tomás Prieto-Rumeau. Approximation of Markov decision processes with general state space. *Journal of Mathematical Analysis and Applications*, 388(2):1254–1267, 2012.

[4] François Dufour and Tomás Prieto-Rumeau. Finite linear programming approximations of constrained discounted Markov decision processes. *SIAM Journal on Control and Optimization*, 51(2):1298–1324, 2013.

[5] François Dufour and Tomás Prieto-Rumeau. Stochastic approximations of constrained discounted Markov decision processes. *J. Math. Anal. Appl.*, 413(2):856–879, 2014.

[6] François Dufour and Tomás Prieto-Rumeau. Approximation of average cost Markov decision processes using empirical distributions and concentration inequalities. *Stochastics*, 87(2):273–307, 2015.

[7] Xianping Guo and Wenzhao Zhang. Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints. *European J. Oper. Res.*, 238(2):486–496, 2014.

[8] Onésimo Hernández-Lerma and Jean-Bernard Lasserre. *Discrete-time Markov control processes: Basic optimality criteria*, volume 30 of *Applications of Mathematics*. Springer-Verlag, New York, 1996.

[9] Jean Jacod. *Calcul stochastique et problèmes de martingales*, volume 714 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979.

[10] Alexey Piunovskiy and Yi Zhang. Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.*, 49(5):2032–2061, 2011.

[11] Alexey Piunovskiy and Yi Zhang. Discounted continuous-time Markov decision processes with unbounded rates and randomized history-dependent policies: the dynamic programming approach. *4OR*, 12(1):49–75, 2014.

[12] W.B. Powell. *Approximate dynamic programming*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2007.

[13] Tomás Prieto-Rumeau and Onésimo Hernández-Lerma. Discounted continuous-time controlled Markov chains: convergence of control models. *J. Appl. Probab.*, 49(4):1072–1090, 2012.

[14] Tomás Prieto-Rumeau and José María Lorenzo. Approximating ergodic average reward continuous-time controlled Markov chains. *IEEE Trans. Automat. Control*, 55(1):201–207, 2010.

[15] Naci Saldi, Tamás Linder, and Serdar Yüksel. Asymptotic optimality and rates of convergence of quantized stationary policies in stochastic control. *IEEE Trans. Automat. Control*, 60(2):553–558, 2015.

[16] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, Massachusetts, 1998.

[17] B. Van Roy. Neuro-dynamic programming: overview and recent trends. In *Handbook of Markov decision processes*, volume 40 of *Internat. Ser. Oper. Res. Management Sci.*, pages 431–459. Kluwer Acad. Publ., Boston, MA, 2002.