2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Singular Control for Discounted Markov Decision Processes in a General State Space

O.L.V. Costa and F. Dufour

Abstract— **This paper studies the asymptotic optimality of discrete-time Markov Decision Processes (MDP's in short) with general state space and action space and having weak and strong interactions. By using a similar approach as developed in [1], the idea in this paper is to consider a MDP with general state and action spaces and to reduce the dimension of the state space by considering an averaged model. This formulation is often described by introducing a small parameter $\epsilon > 0$ in the definition of the transition kernel, leading to a singularly perturbed Markov model with two time scales. First it is shown that the value function of the control problem for the perturbed system converges to the value function of a limit averaged control problem as ϵ goes to zero. In the sequel it is shown that a feedback control policy for the original control problem defined by using an optimal feedback policy for the limit problem is asymptotically optimal.**

I. Introduction

The objective of this work is to study the asymptotic optimality of discrete-time Markov Decision Processes (MDP's in short) with general state space and action space and having weak and strong interactions. We suppose that the state space X of the controlled Markov chain can be written as the union of different ergodic classes X_i for $i \in \mathcal{I}$, where $\mathcal{I}$ a countable (finite or infinite) set, and a transient part X_*. It is assumed that the transitions within each class X_i occur much more frequently than transitions among different classes. This formulation is often described by introducing a small parameter $\epsilon > 0$ in the definition of the transition kernel, leading to a singularly perturbed Markov model with two time scales.

There exists an extensive literature on singularly perturbed discrete-time stochastic control problems. Without attempting to present an exhaustive panorama of this vast field of research, the interested reader may consult the references [2], [3], [4], [5], [1] and the survey [6] and the book [7] to get a rather complete view on this research field. By using a similar approach as developed in [1], the idea in this paper is to consider a MDP with general state space and control space and to reduce the dimension of the state space by considering an averaged model. Indeed, for such MDP's, an inherent problem is the dimension of the state space, which makes the model difficult to be handled. Our objective is twofold. First it is shown that the value function of the control problem for the perturbed system converges to the value function of a limit averaged control problem as ϵ goes to zero. In the second part of the paper, it is proved that a feedback control policy for the original control problem defined by using an optimal feedback policy for the limit problem is asymptotically optimal.

The paper is organized as follows. In section II we present some general definitions and main assumptions, and in section III some auxiliary results regarding the compactness and convergence properties of the relaxed action space. In section IV we present several important results dealing with the convergence of the value function for the original MDP as the small parameter $\epsilon > 0$ goes to zero, and showing that the limit function satisfies an optimality equation. It will be also shown the existence of deterministic δ-optimal solutions for this optimality equation. In section V the limit control problem is formulated and the convergence result is established. Section VI shows that a feedback control policy for the original control problem defined by using an optimal feedback policy for the limit problem is asymptotically optimal.

II. Definitions and assumptions

The main goal of this section is to introduce the notation, definitions and the main assumptions that will be used throughout the paper. In particular we will introduce the class of relaxed controls, which will allow us to get a compactness property for the action space.

In this work we follow closely the notation used in [8]. Let X be a set and v be a mapping from X to $\mathbb{R}$; then for $A \subset X$, v_A denotes the restriction of v to the set A. Moreover, 1_X is the $\mathbb{R}$-valued function defined on X by $1_X(x) = 1$ for all $x \in X$. We recall that X is a Borel space if it is a Borel subset of a complete and separable metric space, and its Borel σ-algebra is denoted by $\mathcal{B}(X)$. For X, Y Borel spaces, the family of all stochastic kernels on X given Y is denoted by $\mathcal{P}(X|Y)$. $\mathcal{M}(X)$ (respectively, $\mathcal{P}(X)$) denotes the set of all finite (respectively probability) measures on $(X, \mathcal{B}(X))$. Moreover, $\mathcal{P}(X)$ is considered as a topological space equipped with the weak topology. For $B \in \mathcal{B}(X)$, I_B denotes the indicator function of the set B, δ_x is the Dirac measure centered on a fixed point $x \in X$ and $\mathbb{I}_B$ denotes the Markov kernel defined by $\mathbb{I}_B(C|x) = I_B(x)\delta_x(C)$ for any $(x, C) \in X \times \mathcal{B}(X)$.

This work was partially supported by INRIA, project CQFD. The first author also received financial support from CNPq (Brazilian National Research Council), grant 301067/09-0, CAPES, and USP project MaCLinC, and the second author from the ARPEGE program of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004.

O.L.V. Costa is with Departamento de Engenharia de Telecomunicações e Controle, Escola Politécnica da Universidade de São Paulo, CEP: 05508 900-São Paulo, Brazil oswaldo@lac.usp.br

F. Dufour is with MAB - Mathématiques Appliqées de Bordeaux, Universite Bordeaux I, 351 Cours de la Liberation, 33405 Talence Cedex, France. dufour@math.u-bordeaux1.fr

Consider $w : X \to [1, \infty)$ a measurable function, that will be referred to as weight function. If u is a real valued function on X we define its w-norm as: $\|u\|_w = \sup_{x\in X} \frac{|u(x)|}{w(x)}$. A function u is said to be w-bounded if $\|u\|_w < \infty$ (bounded if $\|u\| < \infty$ where $\|.\|$ is the sup-norm). The set of w-bounded (bounded) measurable functions defined on X is denoted by $\mathbb{B}_w(X)$ ($\mathbb{B}(X)$ respectively). For a Borel set $A \in \mathcal{B}(X)$, $\mathbb{C}_w(A)$ ($\mathbb{C}_b(A)$ respectively) denotes the set of continuous and w-bounded (bounded) functions from A to $\mathbb{R}$. For a sequence $\{\mu_n\} \in \mathcal{P}(X)$ and $\mu \in \mathcal{P}(X)$, $\mu_n \Rightarrow \mu$ means that the sequence $\{\mu_n\}$ converges to μ in the weak sense, that is, $\int_X v(x)\mu^n(dx) \to \int_X v(x)\mu(dx)$ for every $v \in \mathbb{C}_b(X)$. For Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, $\mathrm{B}(\mathcal{X},\mathcal{Y})$ denotes the space of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ and, for simplicity, we set $\mathrm{B}(\mathcal{X}) = \mathrm{B}(\mathcal{X},\mathcal{X})$. For $T \in \mathrm{B}(\mathcal{X})$, we denote by $r_\sigma(T)$ the spectral radius of the operator T. We recall that $\mathbb{C}_w(X)$ ($\mathbb{C}_b(X)$ respectively) with the w-norm (sup-norm) is a Banach space.

As in Definition 2.2.1 of [8] we consider for $\epsilon > 0$, a five-tuple for a Markov control model

$$\Big(X, A, \{A(x)|x \in X\}, P^\epsilon, c\Big) \tag{1}$$

consisting of

(a) a Borel space X, representing the state space.
(b) a Borel space A, representing the control or action set.
(c) a family $\{A(x)|x \in X\}$ of non-empty measurable subsets $A(x)$ of A, where $A(x)$ denotes the set of feasible controls or actions when the system is in state $x \in X$, and with the property that

$$\mathbb{K} := \{(x,a)|x \in X, a \in A(x)\} \tag{2}$$

is a measurable subset of $X \times A$.
(d) stochastic kernels P and $\overline{P}$ on X given $\mathbb{K}$, and the perturbed stochastic kernel P^ϵ defined as follows:

$$P^\epsilon = P + \epsilon G \text{ with } G = \overline{P} - I. \tag{3}$$

It is assumed that for some $\epsilon_0 > 0$ and every $0 < \epsilon \leq \epsilon_0$, P^ϵ defined in (3) is a stochastic kernel on X given $\mathbb{K}$.
(e) a measurable function $c : \mathbb{K} \to \mathbb{R}$.
(f) a constant $\beta > 0$ such that $\beta\epsilon_0 < 1$.

Definition 2.1: The set of all stochastic kernels φ in $\mathcal{P}(A|X)$ such that $\varphi(A(x)|x) = 1$ for all $x \in X$ is denoted by Φ, and $\mathbb{F}$ stands for the set of all measurable functions $f : X \to A$, satisfying that $f(x) \in A(x)$ for all $x \in X$. We use the notation P^φ to denote the stochastic kernel $P^\varphi(C|x) = \int_{A(x)} P(C|x,a)\varphi(da|x)$, similarly for $\overline{P}^\varphi$ and $P^{\varphi,\epsilon}$. It is assumed that the set $\mathbb{F}$ is nonempty.

To introduce the optimal control problem we are concerned with, it is necessary to introduce different classes of control policy.

Definition 2.2: Define $H_0 = X$ and $H_n = \mathbb{K} \times H_{n-1}$ for $n \geq 1$. A control policy is a sequence $\pi = \{\pi_n\}$ of stochastic kernels π_n on A given H_n satisfying the following constraint: for all $h_n \in H_n$ and $n \geq 1$, $\pi_n(A(x_n)|h_n) = 1$, where $h_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n)$. Let Π be the class of all policies. A policy $\pi = \{\pi_n\}$ is said to be a relaxed policy if there exists $\phi \in \Phi$ such that $\pi_n(.|h_n) = \phi(.|x_n)$. A policy $\pi = \{\pi_n\}$ is said to be a stationary policy if there exists $f \in \mathbb{F}$ such that $\pi_n(.|h_n) = \delta_{f(x_n)}(.)$.
According to a standard convention, we identify $\mathbb{F}$ (respectively, Φ) with the class of all stationary (respectively, relaxed) policies. Therefore, we have $\mathbb{F} \subset \Phi \subset \Pi$. Let $(\Omega, \mathcal{F})$ be the canonical space consisting of the sample path $\Omega = (X \times A)^\infty$ and the associated σ-algebra $\mathcal{F}$. For any policy $\pi \in \Pi$ and any initial distribution ν on X, it can be defined a probability, labeled P^π_ν, and a stochastic processes $\{(x_t, a_t)\}_{t\in\mathbb{N}}$ where $\{x_t\}_{t\in\mathbb{N}}$ is the state process and $\{a_t\}_{t\in\mathbb{N}}$ is the control process satisfying for any $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$ and $h_t \in H_t$ with $t \in \mathbb{N}$, $P^\pi_\nu(x_0 \in B) = \nu(B)$, $P^\pi_\nu(a_t \in C|h_t) = \pi_t(C|h_t)$, and $P^\pi_\nu(x_{t+1} \in B|h_t, a_t) = P^\epsilon(B|x_t, a_t)$, see for example [8, Chapter 2] for such a construction. The expectation with respect to P^π_ν is denoted by E^π_ν. If $\nu = \delta_x$ for $x \in X$, we write P^π_x for P^π_ν and E^π_x for E^π_ν. We consider the following expected discounted Markov control problem:

$$V^\epsilon(x) = \inf_{\pi\in\Pi} V^\epsilon(x,\pi), \tag{4}$$

$$V^\epsilon(x,\pi) = \epsilon E^\pi_x \Big(\sum_{t=0}^{\infty} (1-\beta\epsilon)^t c(x_t, a_t)\Big) \tag{5}$$

We make the following assumptions on the parameters of the MDP. These assumptions are similar to those in Assumption 2.7 of [9]. In section 5 of [10] it is presented an example of a MDP satisfying these assumptions. Let d_1 and d_2 be the metrics on X and A respectively, and let d be the metric on $\mathbb{K}$ defined as $d((x,a),(y,a')) := max\{d_1(x,y), d_2(a,a')\}$, for all (x,a) and (y,a') in $\mathbb{K}$. We shall suppose the following:

H1) For each $x \in X$, $A(x)$ is compact.
H2) The compact-valued multifunction $\Psi : X \to A$ defined by $\Psi(x) = A(x)$ is continuous with respect to the Hausdorff metric (see [11]).
H3) There exists a $[1,+\infty)$-valued continuous function w defined on X satisfying
  a) for all $x \in X$

$$\sup_{a\in A(x)} |c(x,a)| \leq c_0 w(x), \tag{6}$$

  where c_0 is a constant.
  b) for all $x \in X$ and for any $\epsilon \leq \epsilon_0$,

$$\sup_{a\in A(x)} P^\epsilon w(x,a) \leq w(x). \tag{7}$$

H4) For each $x \in X$ there exists a positive nondecreasing function ψ^c_x with $\lim_{t\downarrow 0} \psi^c_x(t) = 0$ such that for all $a \in A(x)$ and $(y,a') \in \mathbb{K}$,

$$|c(x,a) - c(y,a')| \leq \psi^c_x(d((x,a),(y,a'))). \tag{8}$$

H5) For each $x \in X$ there exists a positive nondecreasing function ψ^G_x with $\lim_{t\downarrow 0} \psi^G_x(t) = 0$ such that for all $a \in$

$A(x)$ and $(y,a') \in \mathbb{K}$,

$$\left\|G(.,(x,a)) - G(.,(y,a'))\right\|_w \leq \psi_x^G(d((x,a),(y,a'))), \qquad (9)$$

where $\|.\|_w$ is the w-norm.

H6) For each $x \in X$ there exists a positive nondecreasing function ψ_x^P with $\lim_{t\downarrow 0} \psi_x^P(t) = 0$ such that for every $v \in \mathbb{C}_w(X)$, and all $a \in A(x)$, $(y,a') \in \mathbb{K}$

$$|Pv(x,a) - Pv(y,a')| \leq \max\{|v(x) - v(y)|, \|v\|_w \psi_x^P(d((x,a),(y,a')))\}. \qquad (10)$$

H7) There exist a countable (finite or infinite) set $\mathcal{I}$ and Borel disjoint sets $\{X_i\}_{i\in\mathcal{I}}$ and X_* such that $X = \bigcup_{i\in\mathcal{I}} X_i \cup X_*$ and

H7.i) for any $\varphi \in \Phi$, P^φ restricted to X_i is a w-ergodic kernel with unique invariant probability measure on X_i denoted by π_i^φ, $i \in \mathcal{I}$. Therefore, $\pi_i^\varphi(X_i) = 1$ and X_i is an absorbing set for P^φ. Moreover, assume that there exist moments $M_i \geq 1$ defined on X_i for $i \in \mathcal{I}$ such that for all $i \in \mathcal{I}$

$$\sup_{\varphi\in\Phi} \int_{X_i} M_i(x) \pi_i^\varphi(dx) < \infty. \qquad (11)$$

H7.ii) there exists operators $S_* \in \mathrm{B}(\mathbb{C}_w(X_*))$ and $S_i \in \mathrm{B}(\mathbb{C}_w(X_i), \mathbb{C}_w(X_*))$, $i \in \mathcal{I}$, with S_* invertible, and for each $f \in \mathbb{F}$ there exists an operator $R^f \in \mathrm{B}(\mathbb{C}_w(X_*))$ with R^f invertible, such that

$$r_\sigma(R^f S_* + I) < 1 \qquad (12)$$

and for any function $v \in \mathbb{B}_w(X)$, we have for every $x \in X_*$ that

$$P^f \mathbb{I}_{X_i} v(x) = R^f S_i v_{X_i}(x), \qquad (13)$$

$$P^f \mathbb{I}_{X_*} v(x) = (R^f S_* + I) v_{X_*}(x). \qquad (14)$$

H8) For each $i \in \mathcal{I}$, $\mathbb{K}_i := \{(x,a) | x \in X_i, a \in A(x)\}$ is a measurable set of $X \times A$ and if D_i is a compact subset of X_i then

$$\mathbb{D}_i := \{(x,a) | x \in D_i, a \in A(x)\} \qquad (15)$$

is a relatively compact set of $\mathbb{K}_i$.

H9) For each $i \in \mathcal{I}$, introduce the real-valued mapping a_i defined on X_* by

$$a_i(x) = -S_*^{-1} S_i 1_{X_i}(x). \qquad (16)$$

We assume that for each $i \in \mathcal{I}$ the functions $F_i(x,a)$ defined as

$$F_i(x,a) := \overline{P}(X_i|x,a) + \int_{X_*} a_i(y) \overline{P}(dy|x,a) \qquad (17)$$

are continuous on $\mathbb{K}$ and the mapping defined on $\mathbb{K}$ by $(x,a) \to \sum_{j\in\mathcal{I}} F_j(x,a)\overline{w}(j)$ is upper semicontinuous on $\mathbb{K}$ where $\overline{w}(i) = \inf_{x\in X_i} w(x)$. Furthermore we assume that for every $(x,a) \in \mathbb{K}$,

$$\sum_{j\in\mathcal{I}} F_j(x,a)\overline{w}(j) \leq w(x). \qquad (18)$$

Remark 2.3: If the cost function c is bounded by a constant then we can set $w = 1_X$ and (18) is automatically satisfied.

Remark 2.4: From (7) it is easy to see, by taking $\epsilon = 0$, that

$$\sup_{a\in A(x)} Pw(x,a) \leq w(x) \qquad (19)$$

and, from (3), that

$$\sup_{a\in A(x)} \overline{P}w(x,a) \leq \Big(1 + \frac{1}{\epsilon_0}\Big) w(x). \qquad (20)$$

Remark 2.5: From hypothesis H3) it follows that $\left\|V^\epsilon\right\|_w \leq \frac{c_0}{\beta}$ since that

$$|V^\epsilon(x,\pi)| \leq \epsilon \sum_{t=0}^{\infty} (1-\beta\epsilon)^t c_0 w(x) = \frac{c_0}{\beta} w(x).$$

Remark 2.6: From assumption H6) it follows that the transition law P is weakly continuous on $\mathbb{K}$, that is, for every $v \in \mathbb{C}_w(X)$ the function $Pv(x,a)$ is continuous in $\mathbb{K}$. Similarly from assumption H5) we have that $\overline{P}v$ is weakly continuous on $\mathbb{K}$.

Definition 2.7: For each $i \in \mathcal{I}$, Φ_i denotes the set of all stochastic kernels φ in $\mathcal{P}(A|X_i)$ such that $\varphi(A(x)|x) = 1$ for all $x \in X_i$. Similarly $\mathbb{F}_i$ stands for the set of all measurable functions $f : X_i \to A$, satisfying that $f(x) \in A(x)$ for all $x \in X_i$.

Definition 2.8: For $i \in \mathcal{I}$, $\mathcal{P}_i(\mathbb{K})$ denotes the set of probability measures $\mu^\varphi \in \mathcal{P}(\mathbb{K})$ such that for $B \in \mathcal{B}(\mathbb{K})$

$$\mu^\varphi(B) = \int_{\mathbb{K}_i \cap B} \varphi(da|x) \pi_i^\varphi(dx) \qquad (21)$$

for some $\varphi \in \Phi_i$ (notice that according to assumption H7), π_i^φ is uniquely defined by φ).

For $i \in \mathcal{I}$, the set of the restrictions of probability measures of $\mathcal{P}_i(\mathbb{K})$ to $\mathcal{B}(\mathbb{K}_i)$ is denoted by $\mathcal{P}_i$.

Definition 2.9: For each $i \in \mathcal{I}$, consider a sequence $\{\varphi_n\}$ in Φ_i. Then, $\{\varphi_n\}$ converges to φ if the sequence of probability measures $\{\mu^{\varphi_n}\}$ in $\mathcal{P}(\mathbb{K}_i)$ converges weakly to μ^φ.

III. Auxiliary results

Related to the definitions in the previous section, we have the following proposition showing the compact property of the relaxed control set.

Proposition 3.1: For each $i \in \mathcal{I}$, $\mathcal{P}_i$ (respectively, $\mathcal{P}_i(\mathbb{K})$) is a compact set of $\mathcal{P}(\mathbb{K}_i)$ (respectively, $\mathcal{P}(\mathbb{K})$) in the topology of the weak convergence of measures.

Proof: See [12]. □

An important corollary from the previous proposition is as follows.

Corollary 3.2: Consider a sequence $\{\varphi_n\}$ in Φ_i for $i \in \mathcal{I}$. Then there exists a subsequence of $\{\varphi_n\}$ that converges in Φ_i.

Proof: It follows from Proposition 3.1. □

Throughout the paper whenever we consider a sequence $\{\varphi_n\}$ in Φ_i we set the sequence $\mu_n(B \times C) = \int_B \varphi_n(C|x)\pi^{\varphi_n}(dx)$, so that $\mu_n \in \mathcal{P}_i$. From Corollary 3.2 there exists a subsequence of $\{\mu_n\}$, still denoted by $\{\mu_n\}$, and some $\mu \in \mathcal{P}_i$, such that $\mu_n \Rightarrow \mu$. We write $\mu(B \times C) = \int_B \varphi(C|x)\pi^{\varphi}(dx)$ for some $\varphi \in \Phi_i$.

Proposition 3.3: For any $i \in \mathcal{I}$ and $\{\varphi\}$ in Φ_i, $\pi_i^{\varphi} w = \overline{w}(i)$ and $w = \overline{w}(i)$, π_i^{φ} – a.s. on X_i. Consider a sequence $\{\varphi_n\}$ in Φ_i. Then there exists a subsequence still denoted by $\{\varphi_n\}$ and $\varphi \in \Phi_i$ such that $\{\varphi_n\}$ converges to φ and $\lim_{n\to\infty} \pi_i^{\varphi_n} c^{\varphi_n} = \pi_i^{\varphi} c^{\varphi}$.

Proof: See [12]. □

The next technical result characterizes some properties of the mappings a_i for $i \in \mathcal{I}$ defined in equation (16) that will be used to prove the main results of the paper.

Proposition 3.4: The mappings a_i for $i \in \mathcal{I}$ defined in (16) are positive and belong to $\mathbb{C}_w(X_*)$, and for any $x \in X_*$, $\sum_{i\in\mathcal{I}} a_i(x) = 1$.

Proof: See [12]. □

IV. Properties of the value function for the perturbed MDP

From Theorem 8.3.6 in [13] V^{ϵ} is the unique solution $v^{\epsilon} \in \mathbb{B}_w(X)$ satisfying

$$v^{\epsilon}(x) = \min_{a\in A(x)} \Big(\epsilon c(x,a) + (1-\beta\epsilon)P^{\epsilon}v^{\epsilon}(x,a)\Big) \qquad (22)$$

and moreover there is a measurable selector $f^{\epsilon} \in \mathbb{F}$ such that $f^{\epsilon}(x) \in A(x)$ attains the minimum in (22) for every $x \in X$. Thus for any sequence $\{\epsilon_n\}$, there is for each n a measurable selector $f_n \in \mathbb{F}$ such that

$$\begin{aligned} v^{\epsilon_n}(x) = \epsilon_n\Big(c^{f_n}(x) + (1-\beta\epsilon_n)G^{f_n}v^{\epsilon_n}(x)\Big) \\ + (1-\beta\epsilon_n)P^{f_n}v^{\epsilon_n}(x). \end{aligned} \qquad (23)$$

This section is devoted to show some crucial convergence results of $\{V^{\epsilon}\}$ as $\epsilon \downarrow 0$ and that the limit function will satisfy an optimality equation. We start with Proposition 4.1, showing the equicontinuity of the family $\{V^{\epsilon}\}$ in $\mathbb{C}_w(X)$, and Proposition 4.2, which states that the infimum and supremum of $\{V^{\epsilon}\}$ are also in $\mathbb{C}_w(X)$. In Proposition 4.3 we have an important result showing that, for a sequence $\{\epsilon_n\} \downarrow 0$, we have the convergence of $V^{\epsilon_n}(x)$ to a constant value for $x \in X_i$ and a linear combination of $a_i(x)$ for $x \in X_*$. The section is concluded with the crucial Propositions 4.4 and 4.5 which show that the limit of the value function satisfy an optimality equation that will be used in the next section for the limit control problem. It is also shown the existence of deterministic δ-optimal solutions. We begin with the equicontinuity result.

Proposition 4.1: For $0 < \epsilon \le \epsilon_0$, the family $\{V^{\epsilon}\}$ in $\mathbb{C}_w(X)$ is equicontinuous and $\|V^{\epsilon}\|_w \le \frac{c_0}{\beta}$.

Proof: See [12]. □

The following auxiliary result states that the infimum and supremum of the value functions of the perturbed MDP are in $\mathbb{C}_w(X)$.

Proposition 4.2: Consider a sequence $\{\epsilon_n\} \downarrow 0$ and set $U_k(x) = \inf_{n\ge k} v^{\epsilon_n}(x)$, $V_k(x) = \sup_{n\ge k} v^{\epsilon_n}(x)$. Then $U_k \in \mathbb{C}_w(X)$ and $V_k \in \mathbb{C}_w(X)$.

Proof: See [12]. □

From the previous propositions we have the following important result, showing the convergence of $V^{\epsilon_n}(x)$ to constant values for $x \in X_i$ and a linear combination of $a_i(x)$ for $x \in X_*$.

Proposition 4.3: For any sequence $\{\gamma_n\} \downarrow 0$, there exists a subsequence $\{\epsilon_n\} \downarrow 0$ and constants v_i^0 such that $\lim_{n\to\infty} V^{\epsilon_n}(x) = v_i^0$ for each $x \in X_i$, and $\lim_{n\to\infty} V^{\epsilon_n}(x) = \sum_{i\in\mathcal{I}} v_i^0 a_i(x)$ for each $x \in X_*$, where the mappings $a_i(x)$ are defined in equation (16).

Proof: See [12]. □

From the previous propositions we have the following inequality.

Proposition 4.4: For any sequence $\{\gamma_n\} \downarrow 0$, the constants $\{v_i^0\}_{i\in\mathcal{I}}$ associated to a subsequence $\{\epsilon_n\}$ as defined in Proposition 4.3 satisfy

$$\begin{aligned} &(1+\beta)v_i^0 \le \\ &\inf_{\varphi\in\Phi_i} \Big(\pi_i^{\varphi}c^{\varphi} + \int_{X_i}\int_{A(x)} \overline{P}v^0(x,a)\varphi(da|x)\pi_i^{\varphi}(dx)\Big) \end{aligned} \qquad (24)$$

Proof: See [12]. □

The reverse inequality and the existence of δ-optimal deterministic controls is established in the next proposition.

Proposition 4.5: For any sequence $\{\gamma_n\} \downarrow 0$, consider the subsequence $\{\epsilon_n\} \downarrow 0$ as defined in Proposition 4.3 and the associated constants $\{v_i^0\}_{i\in\mathcal{I}}$. For each n let $f_n \in \mathbb{F}$ be a measurable selector such that (23) is satisfied and set $\varphi_n \in \Phi_i$ such that $\varphi_n(f_n(x)|x) = 1$ and $\mu_n(B \times C) = \int_B \varphi_n(C|x)\pi^{\varphi_n}(dx)$. Then for some subsequence of $\{\mu_n\}$, still denoted by $\{\mu_n\}$, and some $\hat{\mu} \in \mathcal{P}_i$, we have that $\mu_n \Rightarrow \hat{\mu}$, where $\hat{\mu}(B \times C) = \int_B \hat{\varphi}(C|x)\pi^{\hat{\varphi}}(dx)$ for some $\hat{\varphi} \in \Phi_i$ and the following results hold:

a) $$\begin{aligned} &\lim_{n\to\infty} \int_{X_i}\int_{A(x)} \overline{P}V^{\epsilon_n}(x,a)\varphi_n(da|x)\pi_i^{\varphi_n}(dx) = \\ &\int_{X_i}\int_{A(x)} \overline{P}v^0(x,a)\hat{\varphi}(da|x)\pi_i^{\hat{\varphi}}(dx), \end{aligned} \qquad (25)$$

b) $$\begin{aligned} &\lim_{n\to\infty} \int_{X_i}\int_{A(x)} \overline{P}v^0(x,a)\varphi_n(da|x)\pi_i^{\varphi_n}(dx) = \\ &\int_{X_i}\int_{A(x)} \overline{P}v^0(x,a)\hat{\varphi}(da|x)\pi_i^{\hat{\varphi}}(dx), \end{aligned} \qquad (26)$$

c) $$\begin{aligned} &(1+\beta)v_i^0 = \pi_i^{\hat{\varphi}}c^{\hat{\varphi}} \\ &+ \int_{X_i}\int_{A(x)} \overline{P}v^0(x,a)\hat{\varphi}(da|x)\pi_i^{\hat{\varphi}}(dx), \end{aligned} \qquad (27)$$

d) $$\begin{aligned} &(1+\beta)v_i^0 = \\ &\min_{\varphi\in\Phi_i} \Big(\pi_i^{\varphi}c^{\varphi} + \int_{X_i}\int_{A(x)} \overline{P}v^0(x,a)\varphi(da|x)\pi_i^{\varphi}(dx)\Big) = \\ &\inf_{f\in\mathbb{F}_i} \Big(\pi_i^{f}c^{f} + \int_{X_i}\int_{A(x)} \overline{P}v^0(x,f(x))\pi_i^{f}(dx)\Big) \end{aligned} \qquad (28)$$

where the real mapping v^0 defined on X is given by $v^0(x) = v_i^0$ if $x \in X_i$, and $v^0(x) = \sum_{j\in\mathcal{I}} a_j(x)v_j^0$ if $x \in X_*$.

Proof: See [12]. □

V. The limit control problem

The goal of this section is to formulate the limit control problem and to show that its value function coincides with v^0. We first present the limit control problem. Let us define

$$p_{ij}^{\mu} = \int_{\mathbb{K}_i} \Big[\overline{P}(X_j|x,a) + \int_{X_*} a_j(y)\overline{P}(dy|x,a)\Big]\mu(dx,da), \quad (29)$$

for all $(i,j) \in \mathcal{I}^2$, $\mu \in \mathcal{P}_i(\mathbb{K})$. From Proposition 3.4, $p_{ij}^{\mu} \geq 0$ and $\sum_{j\in\mathcal{I}} p_{ij}^{\mu} = 1$. Moreover, by using Assumption H9), the function $\mu \to p_{ij}^{\mu}$ defined on $\mathcal{P}_i(\mathbb{K})$ is continuous for all $(i,j) \in \mathcal{I}^2$. Therefore, the mapping $p : \big(\bigcup_{j\in\mathcal{I}} \{j\} \times \mathcal{P}_j(\mathbb{K})\big) \times 2^{\mathcal{I}} \to [0,1]$ defined by

$$p(i,\mu,B) \quad = \quad \sum_{j\in B} p_{ij}^{\mu} \quad (30)$$

for $i \in \mathcal{I}$, $\mu \in \mathcal{P}_i(\mathbb{K})$ and $B \in 2^{\mathcal{I}}$ is a stochastic kernel on $\mathcal{I}$ given $\bigcup_{j\in\mathcal{I}} \{j\} \times \mathcal{P}_j(\mathbb{K})$.

Introduce now the following parameters of the limit MDP:

- the state space is defined by $\mathcal{I}$ equipped with the discrete topology,
- the action set is given by $\mathcal{P}(\mathbb{K})$ equipped with the topology of weak convergence,
- the set of feasible actions in the state $i \in \mathcal{I}$ is $\mathcal{P}_i(\mathbb{K}) \subset \mathcal{P}(\mathbb{K})$,
- the transition law is given by the stochastic kernel p defined in (30),
- the cost is defined by

$$g(i,\mu) \quad = \quad \frac{1}{1+\beta}\int_{\mathbb{K}_i} c(x,a)\mu(dx,da), \quad (31)$$

  for $i \in \mathcal{I}$ and $\mu \in \mathcal{P}_i(\mathbb{K})$.

Note that for the limit control problem the set of feasible actions in the state $i \in \mathcal{I}$ is defined by $\mathcal{P}_i(\mathbb{K})$ while for the original control problem it is defined by Φ_i. The reason for such definition is mainly technical, in particular for ensuring the measurability of the transition kernel and the cost function with respect to the control. We have the following remarks, collecting some properties on the limit MDP.

Remark 5.1: i) From Proposition 3.1, $\mathcal{P}_i(\mathbb{K})$ is a compact set of $\mathcal{P}(\mathbb{K})$. ii) Since c is continuous on $\mathbb{K}$, the cost g is continuous on $\bigcup_{j\in\mathcal{I}} \{j\} \times \mathcal{P}_j(\mathbb{K})$. iii) For every $u \in \mathbb{B}(\mathcal{I})$ and $i \in \mathcal{I}$, the mapping $\mu \to \sum_{j\in\mathcal{I}} u_j p_{ij}^{\mu}$ defined on $\mathcal{P}_i(\mathbb{K})$ is continuous (see the remark on page 44 in [13]).

Remark 5.2: From (6), (18) and Assumption H9) we have that the mapping $\mu \to \sum_{j\in\mathcal{I}} p_{ij}^{\mu}\overline{w}(i)$ defined on $\mathcal{P}_i(\mathbb{K})$ is continuous.

Definition 5.3: The set of all measurable functions $\lambda : \mathcal{I} \to \mathcal{P}(\mathbb{K})$, satisfying that $\lambda_i \in \mathcal{P}_i(\mathbb{K})$ for every $i \in \mathcal{I}$ is denoted by Λ. Define $E_0 = X$ and $E_n = \big(\bigcup_{j\in\mathcal{I}} \{j\} \times \mathcal{P}_j(\mathbb{K})\big) \times E_{n-1}$ for $n \geq 1$. Let $(\Omega_0, \mathcal{F}_0)$ be the canonical space consisting of the sample path $\Omega = (\mathcal{I} \times \mathcal{P}(\mathbb{K}))^{\infty}$ and the associated σ-algebra $\mathcal{F}_0$. Consequently, for any control $\lambda \in \Lambda$ and any $i \in \mathcal{I}$, it can be defined a probability, labeled P_i^{λ}, and a stochastic processes $\{(y_t,\gamma_t)\}_{t\in\mathbb{N}}$ where $\{y_t\}_{t\in\mathbb{N}}$ is the state process and $\{\gamma_t\}_{t\in\mathbb{N}}$ is the control process satisfying for any $B \in 2^{\mathcal{I}}$, $C \in \mathcal{B}(\mathcal{P}(\mathbb{K}))$, $e_t = (y_0,\gamma_0,\ldots,y_{t-1},\gamma_{t-1},y_t) \in E_t$ with $t \in \mathbb{N}$, $P_i^{\lambda}(y_0 \in B) = \delta_i(B)$, $P_i^{\lambda}(\gamma_t \in C|e_t) = \delta_{\lambda_{y_t}}(C)$, and $P_i^{\lambda}(y_{t+1} \in B|e_t,\gamma_t) = p_{y_t y_{t+1}}^{\lambda_{y_t}}$. The expectation with respect to P_i^{λ} is denoted by E_i^{λ}.

Consider the following countable MDP problem:

$$V_i(\lambda) = \sum_{k=0}^{\infty} \alpha^k E_i^{\lambda}\big[g(y_k,\lambda_{y_k})\big], \quad (32)$$

$$V_i = \inf_{\lambda\in\Lambda} V_i(\lambda). \quad (33)$$

where $\alpha = \frac{1}{1+\beta} < 1$.

We have the following theorem.

Theorem 5.4: For any $y \in X$, $\lim_{\epsilon\to 0} V^{\epsilon}(y) = V^0(y)$ where $V^0(x) = V_i^0$ for $x \in X_i$ and $V^0(x) = \sum_{i\in\mathcal{I}} V_i^0 a_i(x)$ for $x \in X_*$ and the constants $\{V_i^0\}_{i\in\mathcal{I}}$ are the unique solution of the optimality equation

$$V_i^0 = \min_{\mu\in\mathcal{P}_i} \Big(g(i,\mu) + \alpha\sum_{j\in\mathcal{I}} p_{ij}^{\mu} V_j^0\Big) \quad (34)$$

associated to the MDP problem defined in equations (32)-(33).

Proof: See [12]. □

VI. Asymptotic Control

The goal of this section is to develop a feedback control policy for the original problem using the optimal feedback control policy for the limit problem, and show that this control policy is asymptotically optimal. This result is in the same spirit as Theorem 9.9 of [7], although the tools for deriving it are different. According to Theorem 8.3.6 in [13] we can find an optimal selector for the problem (34), that is a measurable mapping $\widehat{\mu} : \mathcal{I} \to \mathcal{P}(\mathbb{K})$ such that for all $i \in \mathcal{I}$, $\widehat{\mu}_i \in \mathcal{P}_i(\mathbb{K})$ and

$$V_i^0 = g(i,\widehat{\mu}_i) + \alpha\sum_{j\in\mathcal{I}} p_{ij}^{\widehat{\mu}_i} V_j^0. \quad (35)$$

Since, $\widehat{\mu}_i \in \mathcal{P}_i(\mathbb{K})$, there exists $\widehat{\varphi}_i \in \Phi_i$ such that $\widehat{\mu}_i(B) = \int_{\mathbb{K}_i\cap B} \widehat{\varphi}_i(da|x)\pi_i^{\widehat{\varphi}_i}(dx)$. Let Φ_* denote the set of all stochastic kernels φ in $\mathcal{P}(A|X_*)$ such that $\varphi(A(x)|x) = 1$ for all $x \in X_*$. Choose an arbitrary $\widehat{\varphi}_* \in \Phi_*$ and define the stochastic kernel $\widehat{\varphi} \in \Phi$ as follows:

$$\widehat{\varphi}(.|x) = \sum_{i\in\mathcal{I}} \widehat{\varphi}_i(.|x)1_{X_i}(x) + \widehat{\varphi}_*(.|x)1_{X_*}(x).$$

Define also $\widehat{c}(x) = c^{\widehat{\varphi}}(x)$, $\widehat{\overline{P}}v(x) = \overline{P}^{\widehat{\varphi}}v(x)$, $\widehat{G}v(x) = G^{\widehat{\varphi}}v(x)$, $\widehat{P}v(x) = P^{\widehat{\varphi}}v(x)$, $\widehat{\pi}_i = \pi_i^{\widehat{\varphi}}$. We need 2 assumptions. The first one assumes that $\widehat{P}$ restricted to X_i is an w_{X_i}-geometric ergodic kernel, instead of just w-ergodic as it was supposed in H7.i). The second one assumes that H7.ii) holds also for $\widehat{\varphi}$ (if $\widehat{\varphi}$ is deterministic then it is automatically satisfied by assumption H7.ii)). More specifically, the assumptions are:

HC1) For each $i \in \mathcal{I}$ there exists $\chi_i > 0$ and $0 < \varrho_i < 1$ such that

$$\|\widehat{P}^k - \widehat{\pi}_i\|_{w_{X_i}} \leq \chi_i \varrho_i^k, \ \forall k = 0, 1, \ldots. \tag{36}$$

HC2) Consider the operators $S_* \in \mathrm{B}(\mathbb{C}_w(X_*))$ and $S_i \in \mathrm{B}(\mathbb{C}_w(X_i), \mathbb{C}_w(X_*))$, $i \in \mathcal{I}$, as in assumption H7.ii). We suppose that there exists an operator $\widehat{R} \in \mathrm{B}(\mathbb{C}_w(X_*))$ with $\widehat{R}$ invertible, such that

$$r_\sigma(\widehat{R}S_* + I) < 1 \tag{37}$$

and for any function $v \in \mathbb{B}_w(X)$, we have for every $x \in X_*$ that

$$\widehat{P}\mathbb{I}_{X_i} v(x) = \widehat{R}S_i v_{X_i}(x), \tag{38}$$

$$\widehat{P}\mathbb{I}_{X_*} v(x) = (\widehat{R}S_* + I) v_{X_*}(x). \tag{39}$$

For $0 < \epsilon < \epsilon_0$ fixed denote by $\widehat{V}^\epsilon$ the cost (5) associated to the strategy $\widehat{\varphi}$. Define $\widehat{P}^\epsilon = \widehat{P} + \epsilon\widehat{G}$. From Theorem 8.3.6 in [13], $\widehat{V}^\epsilon$ is the unique solution $\widehat{v}^\epsilon \in \mathbb{B}_w(X)$ satisfying

$$\widehat{v}^\epsilon(x) = \epsilon\widehat{c}(x) + (1 - \beta\epsilon)\widehat{P}^\epsilon \widehat{v}^\epsilon(x). \tag{40}$$

Set for each $i \in \mathcal{I}$,

$$\widehat{V}_{\sup} = \limsup_{\epsilon\downarrow 0} \widehat{V}^\epsilon, \ \ \widehat{V}_{\inf} = \varliminf_{\epsilon\downarrow 0} \widehat{V}^\epsilon, \tag{41}$$

$$\widehat{V}_{\sup,i} = \sup_{z\in X_i} \widehat{V}_{\sup}(z). \tag{42}$$

We want to show next that $\widehat{V}_{\sup}(x) = \widehat{V}_{\inf}(x) = V_i^0(x)$ for all $x \in X$, where $V^0(x)$ is as defined in Theorem 5.4. In order to do that we need the following auxiliary results.

Proposition 6.1: For each $i \in \mathcal{I}$,

$$\widehat{V}_{\sup} = \widehat{V}_{\sup,i}, \ \ \widehat{\pi}_i - \text{a.s. on } X_i, \tag{43}$$

$$\widehat{V}_{\sup}(x) \leq \sum_{i\in\mathcal{I}} \widehat{V}_{\sup,i} a_i(x), \ \text{ for } x \in X_*. \tag{44}$$

Proof: See [12]. □

In the next proposition we set $\ell^\epsilon(x) = \widehat{c}(x) + (1 - \beta\epsilon)\widehat{G}\widehat{V}^\epsilon(x)$. From (6), Remarks 2.4 and 2.5, we get that $|\ell^\epsilon(x)| \leq \widehat{c}_0 w(x)$ where we have defined $\widehat{c}_0 = c_0 + \frac{c_0}{\beta}\Big(2 + \frac{1}{\epsilon_0}\Big)$. Thus we can conclude that $\|\ell^\epsilon\|_w \leq \widehat{c}_0$. We have the following result.

Proposition 6.2: Fix $i \in \mathcal{I}$ and $z \in X_i$. We have for every $x \in X_i$ and $k = 0, 1, \ldots$ that,

$$|\widehat{P}^k \ell^\epsilon(x) - \widehat{P}^k \ell^\epsilon(z)| \leq \widehat{c}_0 \chi_i \varrho_i^k (w(x) + w(z)). \tag{45}$$

Proof: See [12]. □

For each $i \in \mathcal{I}$ choose $z_i \in X_i$ such that $\widehat{V}_{\sup}(z_i) = \widehat{V}_{\sup,i}$ (this is possible as seen in (43)). Define now for $0 < \epsilon < \epsilon_0$ the functions h^ϵ on X_i as follows: $h^\epsilon(x) = \frac{1}{\epsilon}\Big(\widehat{V}^\epsilon(x) - \widehat{V}^\epsilon(z_i)\Big)$. We have the following result:

Proposition 6.3: For every $x \in X_i$,

$$|h^\epsilon(x)| \leq \frac{\widehat{c}_0 \chi_i}{1 - \varrho_i}\Big(w(x) + w(z_i)\Big). \tag{46}$$

Proof: See [12]. □

The next result shows that in fact the equality in (43) holds for every $x \in X_i$ and, moreover, $\limsup_{\epsilon\downarrow 0} \widehat{\pi}_i \widehat{V}^\epsilon = \widehat{V}_{\sup,i}$.

Proposition 6.4: Consider $i \in \mathcal{I}$ and $x \in X_i$. The following assertions hold: a) $\widehat{V}_{\sup}(x) = \widehat{V}_{\sup,i}$ and b) $\limsup_{\epsilon\downarrow 0} \widehat{\pi}_i \widehat{V}^\epsilon = \widehat{V}_{\sup,i}$.

Proof: See [12]. □

Finally we have the next result which shows that $\widehat{V}_{\sup}(x) = \widehat{V}_{\inf}(x) = V_i^0(x)$ for all $x \in X$, where $V^0(x)$ is as defined in Theorem 5.4.

Proposition 6.5: Consider $V^0(x)$ as defined in Theorem 5.4. We have that for all $x \in X$,

$$\widehat{V}_{\sup}(x) = \widehat{V}_{\inf}(x) = V^0(x). \tag{47}$$

Proof: See [12]. □

Combining Proposition 6.5 and Theorem 5.4 we have the following theorem, showing that $\widehat{\varphi}$ is asymptotically optimal.

Theorem 6.6: For each $x \in X$, $\lim_{\epsilon\downarrow 0} |\widehat{V}^\epsilon(x) - V^\epsilon(x)| = 0$.

Proof: See [12]. □

REFERENCES

[1] R. H. Liu, Q. Zhang, and G. Yin, "Nearly optimal control of singularly perturbed Markov decision processes in discrete time," *Appl. Math. Optim.*, vol. 44, no. 2, pp. 105–129, 2001. [Online]. Available: http://dx.doi.org/10.1007/s00245-001-0016-8

[2] M. Abbad and J. A. Filar, "Perturbation and stability theory for Markov control problems," *IEEE Trans. Automat. Control*, vol. 37, no. 9, pp. 1415–1420, 1992. [Online]. Available: http://dx.doi.org/10.1109/9.159584

[3] M. Abbad, J. A. Filar, and T. R. Bielecki, "Algorithms for singularly perturbed limiting average Markov control problems," *IEEE Trans. Automat. Control*, vol. 37, no. 9, pp. 1421–1425, 1992. [Online]. Available: http://dx.doi.org/10.1109/9.159585

[4] T. R. Bielecki and L. Stettner, "Ergodic control of a singularly perturbed Markov process in discrete time with general state and compact action spaces," *Appl. Math. Optim.*, vol. 38, no. 3, pp. 261–281, 1998. [Online]. Available: http://dx.doi.org/10.1007/s002459900091

[5] T. R. Bielecki and J. A. Filar, "Singularly perturbed Markov control problem: limiting averge cost," *Ann. Oper. Res.*, vol. 28, no. 1-4, pp. 153–168, 1991. [Online]. Available: http://dx.doi.org/10.1007/BF02055579

[6] K. E. Avrachenkov, J. Filar, and M. Haviv, "Singular perturbations of Markov chains and decision processes," in *Handbook of Markov decision processes*, ser. Internat. Ser. Oper. Res. Management Sci. Boston, MA: Kluwer Acad. Publ., 2002, vol. 40, pp. 113–150.

[7] G. G. Yin and Q. Zhang, *Discrete-time Markov chains*, ser. Applications of Mathematics (New York). New York: Springer-Verlag, 2005, vol. 55, two-time-scale methods and applications, Stochastic Modelling and Applied Probability.

[8] O. Hernández-Lerma and J. Lasserre, *Discrete-time Markov control processes: Basic optimality criteria*, ser. Applications of Mathematics. New York: Springer-Verlag, 1996, vol. 30.

[9] E. Gordienko and O. Hernández-Lerma, "Average cost Markov control processes with weighted norms: Existence of canonical policies," *Applicationes Mathematicae*, vol. 23, no. 2, p. 199218, 1995.

[10] ——, "Average cost Markov control processes with weighted norms: Value iteration," *Applicationes Mathematicae*, vol. 23, no. 2, p. 219237, 1995.

[11] K. Hinderer, "Lipschitz continuity of value functions in markovian decision processes," *Math. Meth. Oper. Res.*, vol. 62, pp. 3–22, 2005.

[12] O. Costa and F. Dufour, "Singular discounted Markov control processes in a general state space," *Submitted*, p. 21, 2011.

[13] O. Hernández-Lerma and J. Lasserre, *Further topics on discrete-time Markov control processes*, ser. Applications of Mathematics. New York: Springer-Verlag, 1999, vol. 42.