Proceedings of the 30th Conference
on Decision and Control
Brighton, England • December 1991

Aggregation-Disaggregation Algorithm for ϵ^2-Singularly Perturbed Limiting Average Markov Control Problems

Mohammed Abbad and Jerzy A. Filar
Department of Mathematics and Statistics
University of Maryland at Baltimore County
Baltimore, Maryland

1. Introduction

Finite state and action Markov Decision Processes (MDPs, for short) are dynamic, stochastic, systems controlled by a controller, sometimes referred to as "decision-maker". These models have been extensively studied since the 1950's by applied probabilists, operations researchers, and engineers. Engineers typically refer to these models as "Markov control problems", and in this paper we shall use these labels interchangibly. The early MDP models were studied by Howard [13] and Blackwell [5] and, following the latter, are sometimes referred to as "Discrete Dynamic Programming".

During the 1960's and 1970's the theory of classical MDP's evolved to the extent that there is now a complete existence theory, and a number of good algorithms for computing optimal policies, with respect to criteria such as maximization of limiting average expected output, or the discounted expected output. These models were applied in a variety of contexts, ranging from water-resource models, through communication networks, to inventory and maintenance models.

One class of problems that began to be addressed in recent years focussed around the following question:

- How is the analysis of an MDP model affected by perturbations (typically small) of the problem data?

From the practical point of view the above question is of obvious importance; however, it leads to challenging mathematical problems arising from the following natural phenomenon:

- If the perturbation of a Markov Chain **alters the ergodic structure** of that chain, then the stationary distribution of the perturbed process has a **discontinuity at the zero value of the disturbance parameter.** This phenomenon was illustrated by Schweitzer [20] with the following example:

  Let
  $$P_\epsilon = \begin{pmatrix} 1-\epsilon/2 & \epsilon/2 \\ \epsilon/2 & 1-\epsilon/2 \end{pmatrix}$$
  be the perturbed Markov Chain whose stationary distribution matrix is
  $$P_\epsilon^* = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$
  for all $\epsilon \in (0,2]$. Thus we have
  $$\lim_{\epsilon \downarrow 0} P_\epsilon^* = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \neq P_0^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
  where P_0^* is the stationary distribution matrix of the unperturbed Markov chain P_0.

The above difficulty has led researchers to differentiate between the case that avoids the above mentioned discontinuity, and the cases that permit it. Somewhat imprecisely, perhaps, the former is often referred to as a **regular perturbation**, and the latter as a **singular perturbation**. Of course, it is possible to study the properties of perturbed MDPs without performing the asymptotic analysis (as the perturbation tends to zero), and in such a case the distinction between the regular and singular perturbations is not essential (see, for instance, Van Dijk and Puterman [12], and Van Dijk [11]).

In Abbad, Bielecki and Filar [1] we considered a singularly perturbed Markov Decision Process with the limiting average reward criterion. The singular perturbation arises from the assumption that the underlying process is composed of n separate irreducible processes, and that the small ϵ-perturbation is such that it "unites" these processes into a single irreducible process: that is a singular perturbation of order 1. This structure corresponds to the Markov chains admitting "strong and weak interactions" that arises in applications such as management of hydrodams, and control of queueing network models of computer systems. Intuitively, the n irreducible processes correspond to nearly independent components of a larger system, that are united only by an infrequent "interference" from some central controller. Recent studies that address problems with this structure include the contributions of Delebecque and Quadrat [8], Phillips and Kokotovic [18], Coderch et al. [6], Kokotovic [16], Schweitzer [21], Rohlicek and Willsky [19], and Aldhaheri and Khalil [3].

In this paper we consider a singular perturbation of order 2 for a Markov decision process with the limiting average reward criterion.

We define a singular perturbation of order 2 in the following sense: we assume that the underlying process is composed of n separate irreducible processes, and that a small ϵ-perturbation is such that it "unites" these processes into m separate irreducible processes. Then another small ϵ^2-perturbation is such that it "unites" these latter processes into a single irreducible process.

The present paper is organized as follows:

In Section 2, we formulate the singular perturbation of order 2.

In Section 3, we give explicitely the **limit Markov Control Problem** (limit MCP), that is entirely different from the original unperturbed MDP, which forms an appropriate asymptotic approximation to a whole **family of perturbed problems.** Thus only the single limit MCP needs to be solved.

In Section 4, we construct an **aggregation-disaggregation algorithm** for solving the limit MCP, which is the main contribution of this paper.

2. Definitions and Preliminaries

In this Section we shall formulate precisely the notion of a **singular perturbation of order 2**. We consider a Markov decision process Γ defined by:
$\Gamma = \langle\ S,\ \{A(s) : s \in S\},\ \{r(s,a) : (s,a) \in S \times A(s)\},\ \{p(s' \mid s,a) : s,s' \in S;\ a \in A(s)\}\ \rangle$.

Since from Abbad and Filar [2] the limit Markov control problem has an optimal deterministic strategy, in this paper we shall concern ourselves only with the class $\Pi := C(S)$ of all stationary strategies.

We shall assume that:
(A1) $S = \cup_{i=1}^{n} S_i$ where $S_i \cap S_j = \emptyset$ if $i \neq j$, $n > 1$, card$S_i = n_i$, $n_1 + \ldots + n_n = N$,
and
(A2) $p(s' \mid s,a) = 0$ whenever $s \in S_i$ and $s' \in S_j$, $i \neq j$.

Consequently we can think of Γ as being the "union" of n smaller MDP's Γ_i, defined on the state space S_i, for each $i = 1,2,\ldots,n$, respectively.
Note that if Π_i is the class of stationary strategies in Γ_i, then a strategy $\pi \in \Pi$ in Γ can be written in the natural way as $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$, where $\pi_i \in \Pi_i$.
The probability transition matrix in Γ_i corresponding to π_i is, of course, defined by: $P_i(\pi_i) := \left(p_{ss'}(\pi_i)\right)_{s,s' \in S_i}$, and the generator $G_i(\pi_i)$ and the Cesaro-limit $P_i^*(\pi_i)$ matrices can be defined in a manner analogous to that in the original process Γ.
In addition, we assume that:
(A3) For every $i = 1,2,\ldots,n$ and for all $\pi_i \in \Pi_i$ the transition matrix $P_i(\pi_i)$ is irreducible, and
(A4) $\{1,2,\ldots,n\} = \cup_{k=1}^{m} I_k$ where $I_k \cap I_l = \emptyset$ if $k \neq l$, card $I_k = m_k$.
In view of (A3), the Cesaro-limit matrix $P_i^*(\pi_i)$ is a matrix with identical rows. We shall denote any row of $P_i^*(\pi_i)$ by $\mathbf{p}_i^*(\pi_i)$.
We define:

$$S^k := \cup_{i \in I_k} S_i\ ,\quad k = 1,2,\ldots,m.$$

We shall now consider the situation where the transition probabilities of Γ are perturbed slightly by a perturbation of order 2. Towards this goal we shall define the **first disturbance law** as the set:

$$D_1 = \{d_1(s' \mid s,a) \mid (s,a,s') \in S \times A(s) \times S\}$$

where the elements of D_1 satisfy:
(i) $\sum_{s' \in S} d_1(s' \mid s,a) = 0$ for all $(s,a) \in S \times A(s)$,
(ii) $-\frac{1}{2} \leq d_1(s \mid s,a) \leq 0$ for all $(s,a) \in S \times A(s)$,
(iii) $d_1(s' \mid s,a) \geq 0$ whenever $s \neq s'$,
(iv) $d_1(s' \mid s,a) = 0$ whenever $s' \in S^k$, $s \in S^l$, and $k \neq l$.
We can think of Π as $\Pi = \Pi^1 \times \Pi^2 \times \ldots \times \Pi^m$ where Π^k, $k = 1,2,\ldots,m$ is the set of stationary strategies of the MDP defined by the restriction of the MDP Γ to the state space S^k.
Now, with every $\pi^k \in \Pi^k$ we can associate a **perturbation generator matrix** $D_{1,k}(\pi^k) = \left(d^1_{ss'}(\pi^k)\right)_{s,s' \in S^k}$, where $d^1_{ss'}(\pi^k) := \sum_{a \in A(s)} d_1(s' \mid s,a)\pi^k(s,a)$; and a transition matrix $P_k(\pi^k) := \left(p^k_{ss'}(\pi^k)\right)_{s,s' \in S^k}$, where $p^k_{ss'}(\pi^k) := \sum_{a \in A(s)} p(s' \mid s,a)\pi^k(s,a)$.

We shall also require that there exists $\epsilon_0 > 0$ such that:
(A5) For any $\epsilon \in [0, \epsilon_0)$ and $k = 1,2,\ldots,m$; the quantities

$$p_{k,\epsilon}(s' \mid s,a) := p(s' \mid s,a) + \epsilon d_1(s' \mid s,a),\ s \in S^k,\ a \in A(s)$$

define a transition law in the MDP restricted to the state space S^k, and
(A6) for every $\pi^k \in \Pi^k$, $k = 1,2,\ldots,m$, the transition matrix:

$$P_{k,\epsilon}(\pi^k) := P_k(\pi^k) + \epsilon D_{1,k}(\pi^k) \tag{2.1}$$

is irreducible for any $\epsilon \in (0, \epsilon_0)$.
Assumption (A6) is called the **first singular perturbation assumption.**
The **second disturbance law** is defined by the set:

$$D_2 = \{d_2(s' \mid s,a) \mid (s,a,s') \in S \times A(s) \times S\}$$

where the elements of D_2 satisfy:
(i) $\sum_{s' \in S} d_2(s' \mid s,a) = 0$ for all $(s,a) \in S \times A(s)$,
(ii) $-\frac{1}{2} \leq d_2(s \mid s,a) \leq 0$ for all $(s,a) \in S \times A(s)$,
(iii) $d_2(s' \mid s,a) \geq 0$ whenever $s \neq s'$.
We shall assume that for any $\epsilon \in (0, \epsilon_0)$:
(A7) the quantities
$p_\epsilon(s' \mid s,a) := p(s' \mid s,a) + \epsilon d_1(s' \mid s,a) + \epsilon^2 d_2(s' \mid s,a),\ s,s' \in S,\ a \in A(s)$
define a transition law, and
(A8) for every $\pi \in \Pi$, the transition matrix

$$P_\epsilon(\pi) := \Big(\sum_{a \in A(s)} p_\epsilon(s' \mid s,a)\pi(s,a)\Big)_{s,s'=1}^{N}$$

is an irreducible matrix.

Remark 2.1 *Note that as a result of (A8) the rank of $P_\epsilon^*(\pi)$ is 1, which is strictly less than n, the rank of $P^*(\pi)$. Consequently our perturbation is indeed a singular perturbation in the sense of Delebecque [7]. If $m = 1$, we find the case of the singularly problem studied in Abbad, Bielecki and Filar [1].*

Remark 2.2 *Note that for all $\pi = (\pi_1, \ldots, \pi_n) \in \Pi$ we have the following representation of $P^*(\pi)$:*

$$P^*(\pi) = E_1 M_1(\pi)$$

where E_1 is an $N \times n$ matrix with entries:

$$e^1_{sj} = \begin{cases} 1 & \text{if } \sum_{k=1}^{j-1} n_k < s \leq \sum_{k=1}^{j} n_k \\ 0 & \text{otherwise} \end{cases}$$

for $s = 1,2,\ldots,N$ and $j = 1,2,\ldots,n$, and $M_1(\pi)$ is an $n \times N$ matrix with entries:

$$m^1_{js}(\pi) = \begin{cases} [\mathbf{p}_j^*(\pi_j)]_s & \text{if } \sum_{k=1}^{j-1} n_k < s \leq \sum_{k=1}^{j} n_k \\ 0 & \text{otherwise} \end{cases}$$

for $j = 1,2,\ldots,n$ and $s = 1,2,\ldots,N$. Of course we set $\sum_{k=1}^{0} n_k := 0$. Note also that from the above definitions we conclude that:

$$M_1(\pi)E_1 = I_{n \times n}. \tag{2.2}$$

3. The Reduction Process

For each $\pi \in \Pi$, let us define the $n \times n$ matrices $\bar{B}_1(\pi)$ and $\bar{B}_2(\pi)$ by:

$$\bar{B}_1(\pi) := M_1(\pi)D_1(\pi)E_1, \quad \textit{and} \tag{3.1}$$
$$\bar{B}_2(\pi) := M_1(\pi)\Big\{D_2(\pi) + D_1(\pi)H(\pi)D_1(\pi)\Big\}E_1, \tag{3.2}$$

where, $D_1(\pi) := \Big(\sum_{a\in A(s)} d_1(s'|s,a)\pi(s,a)\Big)_{s,s'=1}^{N}$, $D_2(\pi) := \Big(\sum_{a\in A(s)} d_2(s'|s,a)\pi(s,a)\Big)_{s,s'=1}^{N}$, and

$$H(\pi) := \Big\{I_N - P(\pi) + P^*(\pi)\Big\}^{-1} - P^*(\pi). \tag{3.3}$$

The matrix $H(\pi)$ is called the **deviation matrix** associated to the stationary strategy π.

Now, we define:

$$\bar{B}_\epsilon(\pi) := \bar{B}_1(\pi) + \epsilon\bar{B}_2(\pi) \tag{3.4}$$

Remark 3.1 *Note that the above procedure of construction of the matrices (3.1), (3.2) and (3.4) is the "reduction process" considered in Delebecque [7]. The matrix $\bar{B}_2(\pi)$ follows from the reduction process because $M_1(\pi)P^*(\pi) = M_1(\pi)$ and $P^*(\pi)E_1 = E_1$. In general $\bar{B}_\epsilon(\pi)$ is a series, but by assumption (A8) the reduction process stops at step 2, and hence the terms of order $\epsilon^2, \epsilon^3, \ldots$ are not important.*

Remark 3.2 *It can be verified by inspection using the assumptions made on the disturbance law $D_1(\pi)$ and (2.2) that $\bar{B}_1(\pi)$ defines a generator of a Markov chain. Also, from Theorem 1 in Delebecque [7], it follows that the matrices $\bar{B}_1(\pi)$ and $\bar{C}(\pi)$ are generators of Markov chains.*

Let $\bar{P}^*(\pi)$ be the Cesaro limit corresponding to the generator $\bar{B}_1(\pi)$, and define from $\bar{P}^*(\pi)$ the matrices $M_2(\pi)$ and E_2 in the same way as $M_1(\pi)$ and E_1 were defined from $P^*(\pi)$.

Consider the $m \times m$ matrix defined by:

$$\bar{C}(\pi) := M_2(\pi)\bar{B}_2(\pi)E_2. \tag{3.5}$$

Then let $\bar{C}^*(\pi)$ be the Cesaro-limit corresponding to the generator $\bar{C}(\pi)$.

Now, we define the following $N \times N$ matrix by:

$$\hat{P}^*(\pi) := E_1E_2\bar{C}^*(\pi)M_2(\pi)M_1(\pi). \tag{3.6}$$

From Delebecque [7], we derived the following result:

Theorem 3.1 *For any stationary strategy $\pi \in \Pi$,*

$$\lim_{\epsilon\to 0} P_\epsilon^*(\pi) = \hat{P}^*(\pi).$$

Recall from [1] that the optimization problem (L) defined by:

$$\max_{\pi\in\Pi}[\hat{P}^*(\pi)r(\pi)]_s, \quad s \in S$$

is the **the limit Markov Control Problem.**

Example 3.1 *Let $S = \{1,2,3,4\}$, $A(1) = A(2) = \{1\}$, $A(3) = A(4) = \{1,2\}$. The rewards and transition probabilities can be represented in the following format:*

$s=1$	$s=2$	$s=3$	$s=4$
$(0) \to (1,0,0,0)$	$(0) \to (0,1,0,0)$	$\begin{pmatrix}0\\0\end{pmatrix} \to \begin{matrix}(0,0,0,1)\\(0,0,\frac{1}{2},\frac{1}{2})\end{matrix}$	$\begin{pmatrix}0\\0\end{pmatrix} \to \begin{matrix}(0,0,1,0)\\(0,0,\frac{1}{2},\frac{1}{2})\end{matrix}$

The notation $\big(r(s,a)\big) \to \big(p(1|s,a),\ldots,p(4|s,a)\big)$ gives the reward and the transition probabilities resulting from the choice of an action $a \in A(s)$ in state $s \in S$.
For instance, the choice of an action 2 in state 3 results in an immediate reward of $r(3,2) = 0$ and transition probabilities $p(1|3,2) = 0$, $p(2|3,2) = 0$, $p(3|3,2) = \frac{1}{2}$, $p(4|3,2) = \frac{1}{2}$.
Let the disturbance laws D_1 and D_2 be defined by:

$$[d_1(s'|s,1)]_{s,s'=1}^{4} = \frac{1}{2}\begin{bmatrix}-1&1&0&0\\1&-1&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},$$

$$[d_2(s'|s,1)]_{s,s'=1}^{4} = \frac{1}{2}\begin{bmatrix}-1&0&1&0\\0&-1&0&1\\0&0&0&0\\0&1&0&-1\end{bmatrix},$$

$$[d_1(s'|s,2)]_{s=3,4}^{s'=1-4} = \frac{1}{2}\begin{bmatrix}1&0&-1&0\\0&0&1&-1\end{bmatrix},$$

$$[d_2(s'|s,2)]_{s=3,4}^{s'=1-4} = \frac{1}{2}\begin{bmatrix}0&1&-1&0\\1&0&0&-1\end{bmatrix}.$$

According to the previous notations, we have that:
$S_1 = \{1\}$, $S_2 = \{2\}$, $S_3 = \{3,4\}$, $n = 3$, $I_1 = \{1,2\}$, $I_2 = \{3,4\}$, $m = 2$.

Let π denote the deterministic policy defined by: $\pi(1) = 1$, $\pi(2) = 1$, $\pi(3) = 1$, $\pi(4) = 2$. We have:

$$P(\pi) = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&\frac{1}{2}&\frac{1}{2}\end{bmatrix}, \quad P^*(\pi) = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&\frac{1}{3}&\frac{2}{3}\\0&0&\frac{1}{3}&\frac{2}{3}\end{bmatrix},$$

$$D_1(\pi) = \frac{1}{2}\begin{bmatrix}-1&1&0&0\\1&-1&0&0\\0&0&0&0\\0&0&1&-1\end{bmatrix}, \quad \textit{and}$$

$$D_2(\pi) = \frac{1}{2}\begin{bmatrix}-1&0&1&0\\0&-1&0&1\\0&0&0&0\\1&0&0&-1\end{bmatrix}.$$

From Remark 2.2, we get:

$$M_1(\pi) = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&\frac{1}{3}&\frac{2}{3}\end{bmatrix}, \quad \textit{and} \quad E_1 = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix}.$$

From the expression (3.3), we get:

$$H(\pi) = \begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&\frac{4}{9}&-\frac{4}{9}\\0&0&-\frac{2}{9}&\frac{2}{9}\end{bmatrix}.$$

By using (3.1) and (3.2), we get:

$$\bar{B}_1(\pi) = \frac{1}{2}\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \textit{and} \quad \bar{B}_2(\pi) = \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ \frac{2}{3} & 0 & -\frac{2}{3} \end{bmatrix}.$$

The Cesaro limit matrix $\bar{P}^*(\pi)$ corresponding to the generator $\bar{B}_1(\pi)$ is given by:

$$\bar{P}^*(\pi) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence from $\bar{P}^*(\pi)$, we derive $M_2(\pi)$ and E_2:

$$M_2(\pi) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \textit{and} \quad E_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

By using the expression (3.5), we get:

$$\bar{C}(\pi) = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}.$$

The Cesaro limit matrix $\bar{C}^*(\pi)$ corresponding to the generator $\bar{C}(\pi)$ is given by:

$$\bar{C}^*(\pi) = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{2}{5} & \frac{3}{5} \end{bmatrix}.$$

Finally, by using the expression (3.6), we get:

$$\hat{P}^*(\pi) = \begin{bmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix}.$$

4. Solving The Limit Markov Control Problem

In this Section we shall construct an **"aggregation-disaggregation"** algorithm for solving the limit MCP (L).
Now, we shall present the "classical" policy improvement algorithm for computing an average optimal deterministic strategy in the case of an irreducible MDP Γ (e.g. Derman [10], Denardo [9], Howard [13], Kallenberg [14]).

Step 1: Select an arbitrary deterministic strategy π in the MDP Γ.

Step 2: Solve, for the unknowns $\lambda, y_1, y_2, \ldots, y_{N-1}$, the linear system:

$$\lambda + y_s = r\big(s,\pi(s)\big) + \sum_{s'=1}^{N} p\big(s'|s,\pi(s)\big)y_{s'}; \quad s = 1,\ldots,N$$
$$y_N = 0.$$

Step 3: For each $s = 1,\ldots,N$ compute the improved action $\pi'(s)$ that satisfies:

$$r\big(s,\pi'(s)\big) + \sum_{s'=1}^{N} p\big(s'|s,\pi'(s)\big)y_{s'}$$
$$> r\big(s,\pi(s)\big) + \sum_{s'=1}^{N} p\big(s'|s,\pi(s)\big)y_{s'}. \qquad (4.1)$$

If this is not possible for some $s \in \{1,\ldots,N\}$, then set $\pi'(s) = \pi(s)$.

If $\pi' = \pi$, STOP.

Otherwise, $\pi \leftarrow \pi'$ and go to Step 2.

Remark 4.1 *Note that for any state $s \in S$, the inequality (4.1) will be unchanged if we replace the vector $(y_1,\ldots,y_N)$ by any transformed vector obtained by adding a same constant to all components y_s, $s = 1,\ldots,N$. Thus, different transformed vectors can be used for different states.*
The above observation will be of central importance for the construction of the aggregation-disaggregation algorithm.

Our objective in what follows is to construct a new MDP $\bar{\Gamma}$ which satisfies the conditions of the MDP studied in [1].

We define the new MDP $\bar{\Gamma}$ as follows:

The **State Space** of $\bar{\Gamma}$: $\bar{S} := \{1,2,\ldots,n\}$

The **Action Spaces** of $\bar{\Gamma}$: $\bar{A}(\bar{s}) := X_{s\in S^k} A(s)$ *for each* $\bar{s} \in I_k$, $k = 1,\ldots,m$

The **Transition Law** of $\bar{\Gamma}$: *for all* $(\bar{s},\bar{\mathbf{a}},\bar{s}') \in I_k \times \bar{A}(\bar{s}) \times I_k$, $k = 1,\ldots,m$

$$\bar{q}(\bar{s}'|\bar{s},\bar{\mathbf{a}}) := \begin{cases} 1 + \sum_{s'\in S_{s'}} \sum_{s\in S_s} [\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})]_s d_1(s' \mid s,\bar{a}_s) & \bar{s} = \bar{s}' \\ \sum_{s'\in S_{s'}} \sum_{s\in S_s} [\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})]_s d_1(s' \mid s,\bar{a}_s) & \bar{s} \neq \bar{s}' \end{cases} \qquad (4.2)$$

For every $\bar{s} \in I_k$ and $\bar{s}' \in I_{k'}$,

$$\bar{q}(\bar{s}'|\bar{s},\bar{\mathbf{a}}) := 0 \quad \textit{whenever} \quad k \neq k'. \qquad (4.3)$$

The **Rewards** of $\bar{\Gamma}$: For all $\bar{s} \in \bar{S}$ and $\bar{\mathbf{a}} \in \bar{A}(\bar{s})$,

$$\bar{r}(\bar{s},\bar{\mathbf{a}}) := \sum_{s\in S_s} [\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})]_s r(s,\bar{a}_s). \qquad (4.4)$$

Where $\bar{\mathbf{a}} = \{\bar{a}_s \mid s \in S^k\}$ in the case where $\bar{\mathbf{a}} \in \bar{A}(\bar{s})$ and $\bar{s} \in I_k$.

Note that for any $\bar{s} \in I_k$, each action $\bar{\mathbf{a}} \in \bar{A}(\bar{s})$ defines a deterministic strategy which maps $s \in S^k$ onto $\bar{a}_s$. Thus in (4.2) and (4.4), $\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})$ is well defined.
The validity of the Transition Law, namely:
$\sum_{\bar{s}'=1}^{n} q(\bar{s}'|\bar{s},\bar{\mathbf{a}}) = 1$, $\bar{s} \in \bar{S}$, $\bar{\mathbf{a}} \in \bar{A}(\bar{s})$ and $q(\bar{s}'|\bar{s},\bar{\mathbf{a}}) \geq 0$; $\bar{s}'$, $\bar{s} \in \bar{S}$; $\bar{\mathbf{a}} \in \bar{A}(\bar{s})$
can be checked by inspection using the assumptions made on the disturbance law D_1.
For every $\bar{s}' \in I_{k'}$, $\bar{s} \in I_k$, and $k \neq k'$; we define:

$$\bar{d}(\bar{s}'|\bar{s},\bar{\mathbf{a}}) := \sum_{s'\in S_{s'}} \sum_{s\in S_s} [\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})]_s d_2(s' \mid s,\bar{a}_s). \qquad (4.5)$$

For every $\bar{s}'$, $\bar{s} \in I_k$; we define:

$$\bar{d}(\bar{s}'|\bar{s},\bar{\mathbf{a}}) := \sum_{s'\in S_{s'}} \sum_{s\in S_s} [\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})]_s d_2(s' \mid s,\bar{a}_s) +$$

$$\sum_{s'\in S_{s'}} \sum_{s\in S_s} \sum_{i\in I_k} \sum_{s_1,s_2\in S_i} [\mathbf{p}^*_{\bar{s}}(\bar{\mathbf{a}})]_s d_1(s_1 \mid s,\bar{a}_s) d_1(s' \mid s_2,\bar{a}_{s_2}) h_{s_1 s_2}(\bar{\mathbf{a}}),$$

where $h_{s_1 s_2}(\bar{\mathbf{a}}) := (H(\bar{\mathbf{a}}))_{s_1 s_2}$.
Now we define the disturbance law $\bar{D}$ for the MDP $\bar{\Gamma}$ by the set:

$$\bar{D} := \big\{\bar{d}(\bar{s}'|\bar{s},\bar{\mathbf{a}}) \mid \bar{s}', \bar{s} \in \bar{S};\ \bar{\mathbf{a}} \in \bar{A}(\bar{s})\big\}.$$

Consequently we can think of $\bar{\Gamma}$ as being the "union" of m MDP's $\bar{\Gamma}_k$, defined by the restriction of the MDP $\bar{\Gamma}$ to the state space I_k; $k = 1,2,\ldots,m$.

Remark 4.2 *It can be seen easily that the constructed MDP $\bar{\Gamma}$ satisfies all the conditions of the MDP studied in [1]. Hence we can apply all the results in [1] to the new MDP $\bar{\Gamma}$.*

Let $\bar{\Pi}$ denote the class of all stationary strategies in the MDP $\bar{\Gamma}$.

For every $\bar{\pi} \in \bar{\Pi}$, we denote by $\bar{Q}^*(\bar{\pi})$ the Cesaro-limit matrix of the transition matrix $\bar{Q}(\bar{\pi})$ which results from the use of the strategy $\bar{\pi}$ in the MDP $\bar{\Gamma}$.

As in Remark 2.2, for all $\bar{\pi} \in \bar{\Pi}$ we have the following representation of $\bar{Q}^*(\bar{\pi})$:

$$\bar{Q}^*(\bar{\pi}) = EM(\bar{\pi})$$

where E is an $n \times m$ matrix with entries:

$$e_{sk} = \begin{cases} 1 & \text{if } \sum_{l=1}^{k-1} m_l < \bar{s} \leq \sum_{l=1}^{k} m_l \\ 0 & \text{otherwise} \end{cases}$$

for $\bar{s} = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, m$, and $M(\bar{\pi})$ is an $m \times n$ matrix with entries:

$$m_{ks}(\bar{\pi}) = \begin{cases} [\bar{\mathbf{q}}_k^*(\bar{\pi}^k)]_s & \text{if } \sum_{l=1}^{k-1} m_l < \bar{s} \leq \sum_{l=1}^{k} m_l \\ 0 & \text{otherwise} \end{cases}$$

for $k = 1, 2, \ldots, m$ and $\bar{s} = 1, 2, \ldots, n$. Of course we set $\sum_{l=1}^{0} m_l := 0$. Note also that from the above definitions we conclude that:

$$M(\bar{\pi})E = I_{m \times m}.$$

The perturbed transition probabilities for the MDP $\bar{\Gamma}$ are defined by:
For every $(\bar{s}, \bar{\mathbf{a}}, \bar{s}') \in \bar{S} \times \bar{A}(\bar{s}) \times \bar{S}$,

$$\bar{q}_\epsilon(\bar{s}' \mid \bar{s}, \bar{\mathbf{a}}) := \bar{q}(\bar{s}' \mid \bar{s}, \bar{\mathbf{a}}) + \epsilon \bar{d}(\bar{s}' \mid \bar{s}, \bar{\mathbf{a}}). \tag{4.6}$$

For each $\bar{\pi} \in \bar{\Pi}$, let us define the $m \times m$ matrix $\bar{B}(\bar{\pi})$ by:

$$\bar{B}(\bar{\pi}) := M(\bar{\pi})\bar{D}(\bar{\pi})E,$$

where $\bar{D}(\bar{\pi}) := \left(\sum_{\bar{\mathbf{a}} \in \bar{A}(\bar{s})} \bar{d}(\bar{s}' | \bar{s}, \bar{\mathbf{a}})\bar{\pi}(\bar{s}, \bar{\mathbf{a}})\right)_{\bar{s}, \bar{s}'=1}^{n}$.

Now let $\bar{\bar{Q}}^*(\bar{\pi})$ denote the Cesaro-limit matrix corresponding to the generator $\bar{B}(\bar{\pi})$, for each $\bar{\pi} \in \bar{\Pi}$.
For each $\bar{\pi} \in \bar{\Pi}$, we define the $n \times n$ matrix:

$$\hat{\bar{Q}}^*(\bar{\pi}) := E\bar{\bar{Q}}^*(\bar{\pi})M(\bar{\pi}). \tag{4.7}$$

Now from Abbad, Bielecki and Filar [1], it follows that for any $\bar{\pi} \in \bar{\Pi}$,

$$\lim_{\epsilon \to 0} \bar{Q}_\epsilon^*(\bar{\pi}) = \hat{\bar{Q}}^*(\bar{\pi}), \tag{4.8}$$

where, $\bar{Q}_\epsilon^*(\bar{\pi})$ is the Cesaro limit matrix of the transition matrix $\bar{Q}_\epsilon(\bar{\pi})$ resulting from the use of the strategy $\bar{\pi}$ in the perturbed MCP.
Hence the limit Markov control problem for the perturbed MDP $\bar{\Gamma}$ is the optimization problem $(\bar{L})$ defined by:

$$\max_{\bar{\pi} \in \bar{\Pi}} [\hat{\bar{Q}}^*(\bar{\pi})\bar{\mathbf{r}}(\bar{\pi})]_{\bar{s}}, \quad \bar{s} \in \bar{S}$$

Let $\bar{\pi}$ be a deterministic strategy in $\bar{\Gamma}$, the corresponding deterministic strategy π in Γ is defined by:

$$\pi(s) := [\bar{\pi}(\bar{s})]_s; \; s \in S_{\bar{s}}; \quad \bar{s} \in I_k; \text{ and } k = 1, \ldots, m.$$

Proposition 4.1 *Let $\bar{\pi}^*$ be an optimal deterministic strategy for the problem $(\bar{L})$, then its corresponding deterministic strategy π^* is optimal for the problem (L).*

Proof: Let $\bar{\pi}$ and π be any deterministic strategy in the MDP $\bar{\Gamma}$ and its corresponding strategy in the MDP Γ respectively. From the definition (4.2) of the transition law $\bar{q}$, it can be verified by inspection that

$$\bar{Q}(\bar{\pi}) = I_n + M_1(\pi)D_1(\pi)E_1 \tag{4.9}$$

From the definition of the disturbance law $\bar{D}$, it can also be verified by inspection that

$$\bar{D}(\bar{\pi}) = M_1(\pi)\Big\{D_2(\pi) + D_1(\pi)H(\pi)D_1(\pi)\Big\}E_1 \tag{4.10}$$

From (4.9) and (4.10), it follows that:

$$\bar{P}^*(\pi) = \bar{Q}^*(\bar{\pi}), \; M_2(\pi) = M(\bar{\pi}), \quad \textit{and} \quad \bar{C}(\pi) = \bar{B}(\bar{\pi}).$$

From the definition (4.4) of the rewards $\bar{\mathbf{r}}$, it follows that $\bar{\mathbf{r}}(\bar{\pi}) = M_1(\pi)\mathbf{r}(\pi)$.
Now for any $\bar{s} \in \bar{S}$ and $s \in S_{\bar{s}}$, we have that:

$$\begin{aligned} [\hat{\bar{Q}}^*(\bar{\pi})\bar{\mathbf{r}}(\bar{\pi})]_{\bar{s}} &= [E\bar{\bar{Q}}^*(\bar{\pi})M(\bar{\pi})\bar{\mathbf{r}}(\bar{\pi})]_{\bar{s}} \\ &= [Q_2\bar{C}^*(\pi)M_2(\pi)M_1(\pi)\mathbf{r}(\pi)]_{\bar{s}} \\ &= [Q_1Q_2\bar{C}^*(\pi)M_2(\pi)M_1(\pi)\mathbf{r}(\pi)]_s = [\hat{P}^*(\pi)\mathbf{r}(\pi)]_s. \end{aligned}$$

Finally, since from [1] or [2] both problems (L) and $(\bar{L})$ possess optimal deterministic strategies then the Proposition results.

Remark 4.3 *The importance of the previous Proposition stems from the fact that it converts the problem (L) which is the limit Markov control problem for a perturbation of order 2 into the problem $(\bar{L})$ which is the limit Markov control problem for a perturbation of order 1 (ϵ-additive).*
However, the difficulties of the problem $(\bar{L})$ reside in the fact that its action spaces $\bar{A}(\bar{s})$, $\bar{s} \in \bar{S}$ are large, and its transition law $\bar{q}$ and disturbance law $\bar{D}$ involve Cesaro-limit and deviation matrices.
But in what follows we shall show how we can avoid those difficulties.

Now, we shall apply the aggregation-disaggregation algorithm developped in [1] to the problem $(\bar{L})$.

Step 1: Select an arbitrary deterministic strategy π in $\bar{\Gamma}$, and set:

$$[\pi(k)]_{\bar{s}} := \pi(\bar{s}); \quad \bar{s} \in I_k; \quad k = 1, 2, \ldots m.$$

Step 2: For all k, $k' = 1, \ldots, m$; compute

$$q_{kk'}\big(\pi(k)\big) := \begin{cases} 1 + \sum_{\bar{s}' \in I_k} \sum_{\bar{s} \in I_k} [\bar{\mathbf{q}}_k^*\big(\pi(k)\big)]_{\bar{s}} \bar{d}\big(\bar{s}' \mid \bar{s}, \pi(\bar{s})\big) & k = k' \\ \sum_{\bar{s}' \in I_{k'}} \sum_{\bar{s} \in I_k} [\bar{\mathbf{q}}_k^*\big(\pi(k)\big)]_{\bar{s}} \bar{d}\big(\bar{s}' \mid \bar{s}, \pi(\bar{s})\big) & k \neq k' \end{cases}$$

and

$$c\big(k, \pi(k)\big) := \sum_{\bar{s} \in I_k} [\bar{\mathbf{q}}_k^*\big(\pi(k)\big)]_{\bar{s}} \bar{r}\big(\bar{s}, \pi(\bar{s})\big), \tag{4.11}$$

where, $\bar{\mathbf{q}}_k^*\big(\pi(k)\big)$ is the stationary distribution vector corresponding to the k-th block of the transition matrix $\bar{Q}(\bar{\pi})$.

Step 3: Solve, for the unknowns $\lambda, y_1, y_2, \ldots, y_{m-1}$, the linear system: $k = 1, \ldots, m$

$$\begin{aligned} \lambda + y_k &= c\big(k, \pi(k)\big) + \sum_{k'=1}^{m} q_{kk'}\big(\pi(k)\big) y_{k'} \quad (4.12) \\ y_m &= 0 \end{aligned}$$

Step 4: For each $k = 1, \ldots, m$ compute the deterministic strategy $\pi'(k)$ obtained after one iteration of the policy improvement algorithm (the starting strategy is $\pi(k)$) for the MDP $\bar{\Gamma}_k$ with reward $\bar{c}_k$ defined by: for any $\bar{s} \in I_k$ and $\bar{a} \in \bar{A}(\bar{s})$,

$$\bar{c}_k(\bar{s}, \bar{a}) := \bar{r}(\bar{s}, \bar{a}) + \sum_{k'=1}^{m} \sum_{\bar{s}' \in I_{k'}} \bar{d}(\bar{s}' \mid \bar{s}, \bar{a}) y_{k'}. \quad (4.13)$$

Step 5: If $\pi'(k) = \pi(k)$ for all $k = 1, \ldots, m$ STOP.
Otherwise $\pi(k) \leftarrow \pi'(k)$; $k = 1, 2, \ldots, m$ and go to Step 2.

Note that Step 4 says the following: let $k = 1, 2, \ldots, m$ be fixed;

Step 4.1: The starting strategy is $\pi(k)$.

Step 4.2: Solve the following linear system for the unknowns $\xi; z_{\bar{s}}, \bar{s} \in I_k$

$$\xi + z_{\bar{s}} = \bar{c}_k\big(\bar{s}, \pi(k)(\bar{s})\big) + \sum_{\bar{s}' \in I_k} \bar{q}\big(\bar{s}' | \bar{s}, \pi(k)(\bar{s})\big) z_{\bar{s}'} \quad (4.14)$$

where $z_{\bar{s}_0} := 0$ *for some* $\bar{s}_0 \in I_k$.

Step 4.3: For each $\bar{s} \in I_k$, find an action $\bar{a} \in \bar{A}(\bar{s})$ that satisfies:

$$\bar{c}_k(\bar{s}, \bar{a}) + \sum_{\bar{s}' \in I_k} \bar{q}(\bar{s}' | \bar{s}, \bar{a}) z_{\bar{s}'} > \bar{c}_k\big(\bar{s}, \pi(k)(\bar{s})\big) + \sum_{\bar{s}' \in I_k} \bar{q}\big(\bar{s}' | \bar{s}, \pi(k)(\bar{s})\big) z_{\bar{s}'} \quad (4.15)$$

Set $\pi'(k)(\bar{s}) := \bar{a}$ if $\bar{a}$ exits.
If $\bar{a}$ does not exit for some $\bar{s} \in I_k$, set $\pi'(k)(\bar{s}) := \pi(k)(\bar{s})$.

Step 4.4: If $\pi'(k) = \pi(k)$, STOP.
Otherwise $\pi(k) \leftarrow \pi'(k)$ and go to Step 4.2.

Note that Step 4.3 seems to be complicated since $\bar{A}(\bar{s})$ is large for each $\bar{s} \in I_k$. However, in what follows we shall prove that for each $\bar{s} \in I_k$ the problem of finding an action $\bar{a} \in \bar{A}(\bar{s})$ in Step 4.3 can be converted to one iteration of the policy improvement algorithm for an MDP defined by Γ_s except for the rewards which will be defined appropriately.
Now, it follows that if $y_k = 0$ then in Step 4.3 the problem of finding an action $\bar{a} \in \bar{A}(\bar{s})$ for each $\bar{s} \in I_k$ can be done by using one iteration of the policy improvement algorithm to the MDP Γ_s in which the rewards are defined by:
For any $s \in S_{\bar{s}}$ and $a \in A(s)$, $r_{\bar{s}}(s, a) := r(s, a) + \sum_{\bar{s}' \in I_k} \sum_{s' \in S_{\bar{s}'}} d_1(s' \mid s, a) z_{\bar{s}'} + \sum_{k'=1}^{m} \sum_{\bar{s}' \in I_{k'}} \sum_{s' \in S_{\bar{s}'}} d_2(s' \mid s, a) y_{k'}$.

References

[1] M. Abbad, T. R. Bielecki, and J. A. Filar, *Algorithms for Singularly Perturbed Limiting Average Markov Control Problems*, Proceedings of the 29th CDC, editor IEEE, (1990).

[2] M. Abbad and J. A. Filar, *Perturbation and Stability Theory for Markov Control Problems*, Tech. Rep. 90-13, University of Maryland at Baltimore County, (1990), (accepted by IEEE Transactions on Automatic Control).

[3] R. Aldhaheri and H. Khalil, *Aggregation and optimal control of nearly completely decomposable markov chains*, in Proceedings of the 28th CDC, IEEE, (1989), pp. 1277-1282.

[4] T. R. Bielecki and J. A. Filar, *Singularly Perturbed Markov Control Problem: Limiting Average Cost*, Proceedings of the 28th CDC, editor IEEE, (1989).

[5] D. Blackwell, *Discrete Dynamic Programming* , Annals of Mathematical Statistics, **33**, (1962), pp. 719-726.

[6] M. Cordech, A. Willsky, S. Sastry, and D. Castanon, *Hierarchical aggregation of linear systems with multiple time scales*, IEEE Transactions on Automatic Control, **AC-28** (1983), pp. 1017–1029.

[7] F. Delebecque, *A reduction process for perturbed markov chains*, SIAM Journal of Applied Mathematics, **48** (1983), pp. 325–350.

[8] F. Delebecque and J. Quadrat, *Optimal control of markov chains admitting strong and weak interactions*, Automatica, **17** (1981), pp. 281–296.

[9] E. V. Denardo, *Dynamic Programming*, Prentice-Hall, Eglewood Cliffs, New Jersey, (1982).

[10] C. Derman, *Finite State Markovian Decision Process*, Academic Press, New York, (1970).

[11] N. V. Dijk, *Perturbation Theory for Unbounded Markov Reward Processes with Applications to Queueing*, Adv. Appl. Prob., **20** (1988), pp. 99-111.

[12] N. V. Dijk and M. Puterman, *Perturbation Theory for Markov Reward Processes with Applications to Queueing Systems*, Adv. Appl. Prob., **20** (1988), pp. 79-98.

[13] R. A. Howard, *Dynamic Programming and Markov Processes*, M.I.T.Press, Cambridge, Massachusetts, (1960).

[14] L. C. M. Kallenberg, *Linear Programming and Finite Markovian Control Problems*, Mathematical Center Tracts **148** (1983), Amsterdam.

[15] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, (1980).

[16] P. Kokotovic, *Application of singular perturbation techniques to control problems*, SIAM Review, **26** (1984), pp. 501–550.

[17] A. A. Pervozvanskii and V. G. Gaitsgori, *Theory of Suboptimal Decisions*, Kluwer, Dordrecht, (1988).

[18] R. G. Phillips and P. Kokotovic, *A singular perturbation approach to modelling and control of markov chains*, IEEE Transactions on Automatic Control, **AC-26** (1981), pp. 1087–1094.

[19] J. Rohlicek and A. Willsky, *Multiple time scale decomposition of discrete time markov chains*, Systems and Control Letters, **11** (1988), pp. 309–314

[20] P. J. Schweitzer, *Perturbation theory and finite markov chains*, Journal of Applied Probability, **5** (1968), pp. 401–413.

[21] P. J. Schweitzer, *Perturbation series expansions for nearly completely-decomposable markov chains*, Teletrafic Analysis and Computer Performance Evaluation, (1986), pp. 319–328.