A survey on singular perturbations of Markov chains and decision processes

Konstantin E. Avrachenkov[*], Jerzy Filar[†] and Moshe Haviv[‡]

May 28, 2000

Abstract

In this survey we present a unified treatment of both singular and regular perturbations in finite Markov chains and decision processes. The treatment is based on the analysis of series expansions of various important entities such as the perturbed stationary distribution matrix, the deviation matrix, the mean-passage times matrix and others.

1 Background and Motivation

Finite state Markov Chains (MC's) are among the most widely used probabilistic models of discrete event stochastic phenomena. Named after A.A. Markov, a famous Russian mathematician, they capture the essence of the existentialist "here and now" philosophy in the so-called "Markov property" which, roughly speaking, states that probability transitions to a subsequent state depend only on the current state and time. This property is less restrictive than might appear at first because there is a great deal of flexibility in the choice of what constitutes the "current state". Because of their ubiquitous nature Markov Chains are, nowadays, taught in many undergraduate and graduate courses ranging from mathematics, through engineering to business administration and finance.

[*]Department of Mathematics, The University of South Australia, The Levels, South Australia 5095, Australia.

[†]Department of Mathematics, The University of South Australia, The Levels, South Australia 5095, Australia. This research was supported in part by the ARC grant #A49906132.

[‡]Department of Statistics, The Hebrew University, 91905 Jerusalem, Israel and Department of Econometrics, The University of Sydney, Sydney, NSW 2006, Australia.

Whereas a MC often forms a good description of some discrete event stochastic process, it is not automatically equipped with a capability to model such a process in the situation where there may be a "controller" or a "decision-maker" who – by a judicious choice of actions – can influence the trajectory of the process. This innovation was not introduced until the seminal works of Howard [44] and Blackwell [16] that are generally regarded as the starting point of the modern theory of Markov Decision Processes (or MDP's for short). Since then, MDP's have evolved rapidly to the point that there is a now a fairly complete existence theory, and a number of good algorithms for computing optimal policies with respect to criteria such as maximisation of limiting average expected reward, or the discounted expected reward.

The bulk of the, now vast, literature on both Markov Chains and Markov Decision Models deals with the "perfect information" situations where all the model parameters – in particular probability transitions – are assumed to be known precisely. However, in most applications this assumption will be violated. For instance, a typical parameter, ρ, would normally be replaced by an estimate

$$\hat{\rho} = \rho + \varepsilon(n)$$

where the error term, $\varepsilon(n)$, comes from a statistical procedure used to estimate ρ and n is the number of observations used in that estimation. In most of the valid statistical procedures $|\varepsilon(n)| \downarrow 0$ as $n \uparrow \infty$, in an appropriate sense. Thus, from a perturbation analysis point of view, it is reasonable to suppress the argument n and simply concern ourselves with the effects of $\varepsilon \to 0$.

Roughly speaking, the subject of perturbation analysis of MC's and MDP's divides naturally into the study of "regular" and "singular" perturbations. Intuitively, regular perturbations are "good" in the sense that the effect of the perturbation dissipates harmlessly as $\varepsilon \to 0$, whereas singular pertubations are "bad" in the sense that small changes of ε (in a neighbourhood of 0) can induce "large" effects. Mathematically, it can be shown that singular perturbations are associated with a change of the rank of a suitably selected matrix. This can be easily seen from the now classical example due to Schweitzer [76] where the perturbed probability transition matrix

$$P(\varepsilon) = \begin{pmatrix} 1 - \frac{\varepsilon}{2} & \frac{\varepsilon}{2} \\ \frac{\varepsilon}{2} & 1 - \frac{\varepsilon}{2} \end{pmatrix} \xrightarrow{\varepsilon \downarrow 0} P(0)$$

but the stationary distribution matrix

$$P^*(\varepsilon) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \not\to P^*(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Indeed, the rank of $P^*(\varepsilon)$ is 1 for all $\varepsilon > 0$ and near 0, but it increases to 2 at $\varepsilon = 0$; despite the fact that $P(\varepsilon) \to P(0)$. Thus we see that singular perturbations can occur in MC's in a very natural and essential way. The latter point can be underscored by observing the behaviour of $(I - \lambda P)^{-1}$ as $\lambda \to 1$, where P is a Markov (probability transition) matrix. It is well-known (e.g. see [16] or [65]) that this inverse can be expanded as a Laurent series (with a pole of order 1) in the powers $\varepsilon := 1 - \lambda$. Indeed, much of the theory devoted to the connections between the discounted and limiting average MDP's exploits the asymptotic properties of this expansion, as $\varepsilon \downarrow 0$. Of course, the rank of $(I - \lambda P)$ changes when $\lambda = 1 (\varepsilon = 0)$.

It is not surprising, therefore, that the literature devoted to singularly perturbed MC's and MDP's has been growing steadily in recent years. In fact there have been quite a few developments since the 1995 survey by Abbad and Filar [3].

The purpose of this survey paper is to present an up to date outline of a unified treatment of both singular and regular perturbations in MC's and MDP's that is based on series expansions of various important entities such as the perturbed stationary distribution matrix, the deviation matrix, the mean-passage times matrix and the resolvent-like matrix $(I - \lambda P)^{-1}$. ¿From this series expansion perspective, the regular perturbations are simply the cases where Laurent series reduce to power series. Consequently, the capability to characterise and/or compute the coefficients of these expansions and the order of the pole (if any, at $\varepsilon = 0$) becomes of paramount importance.

This survey covers only the results on *discrete time* MC's and MDP's. For a parallel development of the *continuous time* models we refer an interested reader to the comprehensive book of Yin and Zhang [81].

The logical structure of the survey is as follows: in Section 2, perturbations of (uncontrolled) Markov chains are discussed from the series expansions perspective; in Section 3 the consequences of these results are discussed in the context of optimisation problems arising naturally in the (controlled) MDP case; and, finally, in Section 4 applications of perturbed MDP's to the Hamiltonian Cycle Problem are outlined. This last section demonstrates that the theory of perturbed MDP's has applications outside of its own domain.

2 Uncontrolled perturbed Markov chains

2.1 Introduction and preliminaries

Let $P \in R^{n\times n}$ be a transition stochastic matrix representing transition probabilities in a Markov chain. Suppose that the structure of the underlying Markov chain is aperiodic. Let $P^* = \lim_{t\to\infty} P^t$ which is well-known to exist for aperiodic processes. In the case when the process is also ergodic, P^* has identical rows, each of which is the stationary distribution of P, denoted by π. Let Y be the *deviation matrix* of P which is defined by $Y = (I - P + P^*)^{-1} - P^*$. It is well known (e.g. see [55]) that Y exists and it is the unique matrix satisfying $Y(I-P) = I - P^* = (I-P)Y$ and $P^*Y = 0 = Y1$ (where 1 here is a matrix full of 1's) making it the group inverse of $I - P$. Finally, $Y = \lim_{T\to\infty} \Sigma_{t=0}^{T}(P^t - P^*)$. Let M_{ij} be the mean passage time from state-i into state-j. It is also known that when the corresponding random variable is proper, then M_{ij} is finite. Of course, the matrix M is well-defined if and only if the Markov chain is ergodic. In this case, we have $M_{ij} = (\delta_{ij} + Y_{jj} - Y_{ij})/\pi_j$ and, in particular, $M_{ii} = 1/\pi_i$. The above mentioned results can be found in many sources, for instance, see Meyer [64].

We consider (linear) perturbations of the matrix P and their impact on the structure of the process and on various essential matrices such as the stationary distribution matrix, the deviation matrix and the mean passage time matrix. Specifically, for a scalar ε, $0 < \varepsilon < \varepsilon_{\max}$, and for some zero rowsum matrix C, we look at the set of perturbed stochastic matrices $P(\varepsilon) = P + \varepsilon C$ which are assumed to be ergodic for any ε in the above mentioned region. Note that ergodicity is not assumed with regard to $P = P(0)$. Actually, the case where $P(0)$ contains some unrelated chains (with or without transient states) is our main focus. Our goal here is to survey the existing literature on series expansions for $\pi(\varepsilon)$, $P^*(\varepsilon)$, $Y(\varepsilon)$ and $M(\varepsilon)$, which denote the stationary distribution, the limit matrix, the deviation matrix and the mean passage time matrix, repsectively, of $P(\varepsilon)$, for $0 < \varepsilon < \varepsilon_{\max}$ and consider their relationship to the corresponding entities in the unperturbed MC for $P = P(0)$.

The rest of this section is organized as follows. In the next subsection we discuss the regular case, namely the case in which the unperturbed system is ergodic. Then, in Subsection 2.3, we look at the *nearly completely decomposable* (NCD) case, namely the case in which the state space under the unperturbed process is decomposed into a number of ergodic classes. This number is assumed here to be at least two and no transient states are allowed. This assumption is removed in Subsection 2.4, where we allow the unperturbed system to have two or more ergodic classes plus a number of transient states. It will be seen that the presence of transient states induces some interesting phenomena. In Subsections 2.2 through 2.4, we assume that the perturbed process in ergodic. Subsection 2.5, we comment on some issues involving the removal of this assumption.

2.2 The regular case

In this subsection we assume that the unperturbed Markov chain is ergodic. This leads to the case of **regular perturbations**. The following results appear in a seminal paper by Schweitzer Sc1 [74].

Theorem 1 *Assume that the unperturbed Markov chain is ergodic. Then,*

(i) The matrix functions $P^*(\varepsilon)$, $Y(\varepsilon)$ and $M(\varepsilon)$ are analytic in some (undeleted) neighborhood of zero. In particular, they all admit Maclaurin series expansions:

$$P^*(\varepsilon) = \sum_{m=0}^{\infty} \varepsilon^m P^{(*m)} \ , \ Y(\varepsilon) = \sum_{m=0}^{\infty} \varepsilon^m Y^{(m)} \ \text{and} \ M(\varepsilon) = \sum_{m=0}^{\infty} \varepsilon^m M^{(m)}$$

with some coefficient sequences $\{P^{(*m)}\}_{m=0}^{\infty}$, $\{Y^{(m)}\}_{m=0}^{\infty}$ and $\{M^{(m)}\}_{m=0}^{\infty}$.

(ii) The limit matrices $P^*(\varepsilon)$ and the deviation matrix of the perturbed Markov chain admit the following updating formulae

$$P^*(\varepsilon) = P^*(0)[I - \varepsilon U]^{-1}$$

and

$$\begin{aligned} Y(\varepsilon) &= [I - P^*(\varepsilon)]Y(0)[I - \varepsilon U]^{-1} \\ &= Y(0)[I - \varepsilon U]^{-1} - P^*(0)[I - \varepsilon U]^{-1}Y(0)[I - \varepsilon U]^{-1}, \end{aligned}$$

where $U := CY(0)$.

(iii) These updating formulae yield the following expressions for the power series coefficients.

$$P^{(*0)} = P^*(0), \quad Y^{(0)} = Y(0), \quad M^{(0)} = M(0)$$

$$P^{(*m)} = P^{(*0)}U^m, \quad m \geq 0$$

$$Y^{(m)} = Y(0)U^m - P^*(0)\sum_{j=1}^{m} U^j Y(0) U^{m-j}, \quad m \geq 0$$

$$M_{ij}^{(m)} = \frac{1}{\pi_j^{(0)}}(Y_{jj}^{(m)} - Y_{ij}^{(m)}) - \frac{1}{\pi_j^{(0)}}\sum_{l=1}^{m} \pi_j^{(l)} M_{ij}^{(m-l)}, \quad m \geq 0$$

(iv) *The validity of any of the above series expansion holds for any ε, $0 \leq \varepsilon < \min\{\varepsilon_{\max}, \rho^{-1}(U)\}$ where $\rho(U)$ is the spectral radius of U.*

Example 1. For $0 \leq \varepsilon < .25$ let

$$P(\varepsilon) = P(0) + \varepsilon C = \begin{pmatrix} .5 & .5 \\ .5 & .5 \end{pmatrix} + \varepsilon \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix} .$$

Clearly,

$$P^*(0) = P^{(*0)} = \begin{pmatrix} .5 & .5 \\ .5 & .5 \end{pmatrix}$$

Also,

$$Y(0) = Y^{(0)} = \begin{pmatrix} .5 & -.5 \\ -.5 & .5 \end{pmatrix} \quad \textit{and} \quad M(0) = M^{(0)} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} .$$

Hence,

$$U = CY(0) = \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix} .$$

It is easy to see that for $m \geq 1$, $U^m = 3^{m-1}U$ and hence for $m \geq 1$,

$$P^{(*m)} = P^*(0)U^m = 3^{m-1}\begin{pmatrix} .5 & -.5 \\ .5 & -.5 \end{pmatrix} .$$

Also, for $m \geq 1$,

$$Y^{(m)} = 3^{m-1}Y(0)U - 3^{m-2}(m-1)P^*(0)UY(0)U - 3^{m-1}P^*(0)UY(0)$$

$$= 3^{m-1}\begin{pmatrix} 1.5 & -1.5 \\ -1.5 & 1.5 \end{pmatrix} - 3^{m-2}(m-1)\begin{pmatrix} 1.5 & -1.5 \\ 1.5 & -1.5 \end{pmatrix} - 3^{m-1}\begin{pmatrix} .5 & -.5 \\ .5 & -.5 \end{pmatrix} .$$

Finally,

$$M_{12}(\varepsilon) = 2 + 8\varepsilon + ... \quad \text{and} \quad M_{21}(\varepsilon) = 2 + 4\varepsilon + ...$$

2.3 The nearly completely decomposable case

Let $P(0) \in R^{n \times n}$ be a stochastic matrix representing transition probabilities in a completely decomposable Markov chain. By the latter we mean that there exists a partition Ω of the state space into p, $p \geq 2$, subsets $\Omega = \{I_1, \ldots, I_p\}$ each of which being an ergodic class. We assume that the order of the rows and of the columns of P is compatible with Ω, i.e., for p stochastic matrices, $P_{I_1}, \ldots, P_{I_p}$,

$$P(0) = \begin{pmatrix} P_{I_1} & 0 & \cdots & 0 \\ 0 & P_{I_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & P_{I_p} \end{pmatrix}$$

Note that we assume above that none of the states is transient. Let $C \in R^{n \times n}$ be a zero rowsum matrix such that for some $\varepsilon_{\max} > 0$, the matrix $P + \varepsilon C$ is stochastic for $\varepsilon \in (0, \varepsilon_{\max})$ representing transition probabilities in an ergodic Markov chain. For small values of ε, $P(\varepsilon)$ is called *nearly completely decomposable* (NCD) or sometimes *nearly uncoupled*. Clearly, $C_{ij} \geq 0$ for any pair of states i and j belonging to different subsets.

Probably, the first motivation to study the singular perturbed Markov chains was given by Simon and Ando [77]. They demonstrated that several problems in econometrics lead to the mathematical model based on singularly perturbed Markov chains. The first rigorous theoretical developments of the singularly perturbed Markov chains have been carried out by Pervozvanski and Smirnov [68] and Gaitsgori and Pervozvanskii [32]. In particular, they have shown that the *limiting* probability distribution $\pi^{(0)} = \lim_{\varepsilon \to 0} \pi(\varepsilon)$ can be expressed in terms of the invariant probability distributions of the ergodic classes $I_k, k = 1, ..., p$ and of the stationary distribution of an *aggregated* chain. Similar ideas were also developed in works of Courtois and his co-authors [19, 20, 21] and in the work of Haviv and his co-authors [35, 36, 37, 39, 40, 45]. Schweitzer [75] showed that $\pi(\varepsilon)$ is analytic in some deleted neighbourhood of zero. That is, $\pi(\varepsilon) = \sum_{m=0}^{\infty} \pi^{(m)} \varepsilon^m$, where $\pi^{(0)} = \lim_{\varepsilon \to 0} \pi(\varepsilon)$ and where $\pi^{(m)}$, $m \geq 1$, are zerosum vectors. Note that $\pi_i^{(0)} > 0$ for all i. Moreover, $\{\pi^{(m)}\}_{m=0}^{\infty}$ is a geometric series, that is, for some matrix $U \in R^{n \times n}$, $\pi^{(m)} = \pi^{(0)} U^m$, $0 \leq m < \infty$. See (4) below for an explicit expression for U. Finally, the series expansion holds for $0 < \varepsilon < \max\{\varepsilon_{\max}, \rho^{-1}(U)\}$.

For any subset $I \in \Omega$, let

$$k_I = \Sigma_{i \in I} \pi_i^{(0)} \tag{1}$$

and let γ_I be the subvector of $\pi^{(0)}$ corresponding to subset I rescaled so as its entry-sum is now one. Then, γ_I is the unique stationary distribution of P_I. Note that computing γ_I is relatively easy as only the knowledge of P_I is needed.

Next define the matrix $Q \in R^{p\times p}$ which is

usually referred to as the *aggregate* transition matrix. Each row, and likewise each column in Q corresponds to a subset in Ω. Then, for subsets I and J, $I \neq J$, let

$$Q_{IJ} = \sum_{i\in I}(\gamma_I)_i \sum_{j\in J} C_{ij} \tag{2}$$

QIJ

and let

$$Q_{II} = 1 + \sum_{i\in I}(\gamma_I)_i \sum_{j\in I} C_{ij} = 1 - \sum_{J\neq I} Q_{IJ} \quad . \tag{3}$$

QII

Without loss of generality, assume that Q_{II} is non-negative for all subsets I and hence Q is easily seen to be a stochastic matrix.[1] Moreover, Q is irreducible and the vector $k \in R^p$ (see (1)) is easily checked to be its unique stationary distribution. Often it is convenient to express the aggregated transition matrix Q in matrix terms. Specificaly, let $V \in R^{p\times n}$ be such that its i-th row is full of zeros except for γ_{I_i} at the entries corresponding to subset I_i, and where $W \in R^{n\times p}$ is such that its j-th column is full of zeros except for 1's in the entries corresponding to subset I_j.[2] Then, we can write

$$Q = I + VCW.$$

The aggregate stochastic matrix Q represents transition probabilities between subsets which in this context are sometimes referred to as *macro-states*. However, although the original process among states is Markovian, this is not necessarily the case with the process among macro-states (and indeed typically it is not).[3] Yet, as the following indicates, much can be learned on the original process from the analysis of the aggregate matrix.

[1] Note that the matrix C can be divided by any constant and ε can be multiplied by this constant leading to the same $n \times n$ transtion matrices. Taking this constant small enough guarantees the stochasticity of Q and hence this is assumed without loss of generality. In particular, the stationary distribution of Q is invariant with respect to the choice of this constant. Alternatively, one can define Q_{II} by $-\Sigma_{J\neq I} Q_{IJ}$ and consider Q as the generator of the aggregated process, i.e., the process among subsets (and hence no need to assume anything further with regard to the size of the entries of the matrix C.)

[2] Note that $VW \in R^{p\times p}$ is the identity matrix. Moreover, V and W correspond to orthonormal sets of eigenvectors of $P(0)$ belonging to the eigenvalue 1, V as left eigenvectors and W as right eigenvectors.

[3] The process among macro-states is an example of a partially observable Markov process.

Theorem 2 *Let the perturbed Markov chain be nearly completely decomposable. The stationary distribution $\pi(\varepsilon)$ admits a Maclaurin series expansion in a deleted neighborhood of zero. Specifically, for some vectors $\{\pi^{(m)}\}_{m=0}^{\infty}$ with $\pi^{(0)}$ being a probability vector positive in all its entries and satisfying $\pi^{(0)} = \pi^{(0)}P(0)$, and for some zerosum vectors $\pi^{(m)}$, $m \geq 1$, $\pi(\varepsilon) = \Sigma_{m=0}^{\infty}\pi^{(m)}\varepsilon^m$. Moreover, for $I \in \Omega$, $\pi_I^{(0)} = k_I\gamma_I$.*[4] *Also, the series $\{\pi^{(m)}\}_{m=0}^{\infty}$ is geometric, i.e., for some square matrix U, $\pi^{(m)} = \pi^{(0)}U^m$ for any $m \geq 0$. Actually,*

$$U = CY(0)(I + CWDV), \tag{4}$$

where D is the deviation matrix of the aggregated transition matrix Q. Alternatively,

$$U = CY^{(0)} \tag{5}$$

where $Y^{(0)}$ is the first regular term of the Laurent series expansion for $Y(\varepsilon)$ (see also the next theorem). Finally, the validity of the series expansion holds for any ε, $0 \leq \varepsilon < \min\{\varepsilon_{\max}, \rho^{-1}(U)\}$ where $\rho(U)$ is the spactral radius of U.

We note that a series expression for $\pi(\varepsilon)$ appeared orginally in [75]. The expression for U in (4) taken from [39] sheds more light on the role of the aggregate process played in the original process than its original expression given in [75] (see Equation 3-1 there). Also we note that $U = CY^{(0)}$ is an elegant generalization [76] of the regular case where $U = CY(0)$.

For ε, $0 \leq \varepsilon < \varepsilon_{\max}$, let $Y(\varepsilon)$ be the deviation matrix of $P(\varepsilon)$. This matrix is uniquely defined and the case $\varepsilon = 0$ is no exception. Yet, as we see shortly, there is no continuity of $Y(\varepsilon)$ at $\varepsilon = 0$. In particular, $Y(0)$ has the same shape as P has, namely

$$Y(0) = \begin{pmatrix} Y_{I_1} & 0 & \cdots & 0 \\ 0 & Y_{I_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Y_{I_p} \end{pmatrix} \tag{6}$$

where Y_{I_i} is the deviation matrix of P_{I_i}, $1 \leq i \leq p$.

Theorem 3 *In the case of NCD Markov chains, the matrix $Y(\varepsilon)$ admits a Laurent series expansion in a deleted neighborhood of zero with the order of*

[4]Recall that γ_I is the stationary distribution of P_I and that k in the stationary distribution of the aggreagted matrix Q as defined in (2) and (3).

the pole being exactly one. Specifically, for some matrices $\{Y^{(m)}\}_{m=-1}^{\infty}$ *with* $Y^{(-1)} \neq 0$,

$$Y(\varepsilon) = \frac{1}{\varepsilon}Y^{(-1)} + Y^{(0)} + \varepsilon Y^{(1)} + \varepsilon^2 Y^{(2)} + \cdots \tag{7}$$

for $0 < \varepsilon < \varepsilon_{\max}$. *Moreover,*

$$Y^{(-1)} = WDV,$$

or in a component form,

$$Y_{ij}^{(-1)} = D_{IJ}(\gamma_J)_j \quad , \quad i \in I \ , \ j \in J. \tag{8}$$

A series expression for $Y(\varepsilon)$ was originally developed in [75]. Yet, the expression for $Y^{(-1)}$ given above in (8) is taken from [39]. We find (8) appealing as it explicitly shows the role played by the aggregate matrix (and hence by the process among macro-states) in the original process. This point was already mentioned in [37] (see Equation 3 there).

We now focus our attention on $M(\varepsilon)$. Note that as opposed to $Y(0)$, $M(0)$ is not well-defined as the corresponding mean value (when $\varepsilon = 0$ and states i and j belong to two different subsets) does not exist. Let $E \in R^{p\times p}$ be the mean passage time matrix associated with the aggregated process. I.e., for any pair of subsets I and J ($I = J$ included), E_{IJ} is the mean passage time from the macro-state I into the macro- state J when transition probabilities are governed by the stochastic matrix Q.

The following theorem appears in [10].

Theorem 4 *The matrix* $M(\varepsilon)$ *admits a Laurent series expansion in a deleted neighborhood of zero with the order of the pole being exactly one. Specifically, for some matrices* $\{M^{(m)}\}_{m=-1}^{\infty}$ *with* $M^{(-1)} \neq 0$,

$$M(\varepsilon) = \frac{1}{\varepsilon}M^{(-1)} + M^{(0)} + \varepsilon M^{(1)} + \varepsilon^2 M^{(2)} + \cdots \tag{9}$$

for $0 < \varepsilon < \varepsilon_{\max}$. *Moreover, for* $i \in I$ *and* $j \in J$,

$$M_{ij}^{(-1)} = \left\{ \begin{array}{cc} 0 & if \ J = I \\ E_{IJ} & if J \neq I \end{array} \right. \tag{10}$$

$$M_{ij}^{(m)} = \frac{1}{\pi_j^{(0)}}(Y_{jj}^{(m)} - Y_{ij}^{(m)}) - \frac{1}{\pi_j^{(0)}}\sum_{l=1}^{m+1}\pi_j^{(l)}M_{ij}^{(m-l)} \quad , \quad m \geq -1$$

The following example is taken from [76].

Example 2 Let

$$P(0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & .5 & .5 \\ 0 & .5 & .5 \end{pmatrix} \quad \text{and} \quad C = \frac{2}{7} \begin{pmatrix} -2 & 1 & 1 \\ 3 & -1 & -2 \\ 4 & -3 & -1 \end{pmatrix}.$$

The number of subset equals 2 with $\gamma_{I_1} = 1$ and $\gamma_{I_2} = (0.5, 0.5)$. First, we construct the following matrices.

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix} \quad W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad Y(0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 0.5 \end{pmatrix}$$

The aggregated transition matrix is given by

$$Q = I + VCW = \begin{pmatrix} 3/7 & 4/7 \\ 1 & 0 \end{pmatrix} \quad .$$

Hence, $k = (7/11, 4/11)$ and $\pi^{(0)} = (7/11, 2/11, 2/11)$. Next, we calculate D, the deviation matrix of Q,

$$D = (I - Q + Q^*)^{-1} - Q^* = \frac{1}{121} \begin{pmatrix} 28 & -28 \\ -49 & 49 \end{pmatrix}$$

and hence, using (8),

$$Y^{(-1)} = WDV = \frac{1}{242} \begin{pmatrix} 56 & -28 & -28 \\ -98 & 49 & 49 \\ -98 & 49 & 49 \end{pmatrix}$$

The matrix E, which is the mean passage time matrix for the aggregated process, equals

$$E = \begin{pmatrix} 11/7 & 7/4 \\ 1 & 11/4 \end{pmatrix}$$

and hence, using (10),

$$M^{(-1)} = \begin{pmatrix} 0 & 7/4 & 7/4 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Finally, from (4) we get that

$$U = CY(0)(I + CWDV) = \frac{1}{77} \begin{pmatrix} 0 & 0 & 0 \\ -2 & 12 & -10 \\ 4 & -24 & 20 \end{pmatrix}$$

2.4 The general case

In this section we impose no assumptions on the ergodic structure of the unperturbed chain. That is, the unperturbed chain may now consist of several ergodic classes plus a set of transient states. Probably, Delebecque and Quadrat [24] were the first who studied the model in this general setting. However, they have analysed only the case of the first order singularity, namely the case where the Laurent series expansion for the perturbed deviation matrix has a simple pole. This analysis is not much different from the analysis of the NCD case. Later Delebecque [25], using the reduction technique of Kato [52], removed the first order singularity assumption and showed how to compute the asymptotic expansion for $\pi(\varepsilon)$ in the general case. However, the reduction technique of [52] does not provide an efficient computational scheme. Other related papers devoted to the case where the unperturbed chain has a transient set are [17, 18, 15, 56, 72, 78].

The following typical example of the general situation was considered extensively in the past (see [39, 34, 40, 12]):

Example 3.

$$P(\varepsilon) = P(0) + \varepsilon C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} .$$

In this example the unperturbed chain contains two ergodic classes (states 2 and 4) and two transient states (states 1 and 3). They all are coupled to a single chain when $\varepsilon > 0$. Moreover, states 2 and 4 (i.e., the ergodic chains in the unperturbed process) communicate under the perturbed case only via states 1 and 3 (i.e., transient states in the unperturbed case). This phonomenon makes this case more involved than the NCD case. For example, the reader is invited to check that the order of magnitude of $M_{24}(\varepsilon)$ is ε^{-2}.

Theorem 5

(i) The stationary distribution $\pi(\varepsilon)$ admits a Maclaurin series expansion in a deleted neighbourhood of zero, that is, for some sequence $\{\pi^{(m)}\}_{m=0}^{\infty}$ with $\pi^{(0)}$ being a probability vector and $\pi^{(m)}$ for $m \geq 1$, being zero sum vectors,

$$\pi(\varepsilon) = \sum_{m=0}^{\infty} \varepsilon^m \pi^{(m)} .$$

Moreover, $\pi^{(0)}P(0) = \pi^{(0)}$ *and for some matrix* U*,*

$$\pi^{(m)} = \pi^{(0)}U^m \quad . \tag{11}$$

pi_geom

(ii) The deviation matrix $Y(\varepsilon)$ *admits a Laurent series expansion in a deleted neighborhood of zero with a non-essential pole, that is, for some nonnegative integer* s *and for some sequence of matrices* $\{Y^{(m)}\}_{m=-s}^{\infty}$ *with* $Y^{(-s)} \neq 0$*,*

$$Y(\varepsilon) = \frac{Y^{(-s)}}{\varepsilon^s} + \frac{Y^{(-s+1)}}{\varepsilon^{s-1}} + \cdots + Y^{(0)} + Y^{(1)}\varepsilon + \cdots \quad .$$

Moreover, the regular part $Y^R(\varepsilon) = Y^{(0)} + \varepsilon Y^{(1)} + \cdots$ *can be expressed by the updating formula*

$$Y^R(\varepsilon) = [I - P^*(\varepsilon)]Y^{(0)}[I - \varepsilon U]^{-1} - P^*(\varepsilon)\sum_{i=1}^{s} U^i Y^{(-i)} \quad . \tag{12}$$

up_Y

The latter provides a computationally efficient recursive formula for the coefficients of the regular part:

$$\begin{aligned} Y^{(m)} &= Y^{(0)}U^m - P_0^*\sum_{j=0}^{m} U^j Y^{(0)} U^{m-j} \\ &+ P^{*(0)}U^m[I - \sum_{i=1}^{s} U^i Y^{(-i)}], \quad m \geq 1. \end{aligned} \tag{13}$$

recur_Y

Furthermore, the quotient matrix U *can be expressed as*

$$U = CY^{(0)} \tag{14}$$

Y0G

(iii) The mean passage time matrix $M(\varepsilon)$ *admits a Laurent series expansion in a deleted neighborhood of zero with a non-essential pole, that is, for some nonnegative integer* t *and for some sequence of matrices* $\{M^{(m)}\}_{m=-t}^{\infty}$ *with* $M^{(-t)} \neq 0$*,*

$$M(\varepsilon) = \frac{M^{(-t)}}{\varepsilon^t} + \frac{M^{(-t+1)}}{\varepsilon^{t-1}} + \cdots + M^{(0)} + M^{(1)}\varepsilon + \cdots \quad .$$

Finally, $t \geq s$*.*

We would like to emphasize that once one has found the value of s and computed the terms $Y^{(-s)}, ..., Y^{(0)}$ and $\pi^{(0)}$, all the other terms in the series expansion for $Y(\varepsilon)$ and $\pi(\varepsilon)$ can be computed by (13). We refer below to a procedure for computing s. The term $\pi^{(0)}$ (which we believe to be the most important coefficient of the expansion) can be computed either by a reduction process [25] or via the determination of the null space of an auxiliary augmented matrix. The latter method will be described in some detail below. Finally, one can find methods for computing $Y^{(-s)}, ..., Y^{(0)}$ (and hence U) (once s and $\pi^{(0)}$ are in hand) by methods outlined in [12] and [39].

Remark 1 *Note that the formulae (11) and (12) are the natural generalization of the regular case. Formula (11) and (14) appeared originally in [76]. Finally, (12) was derived in [12].*

The next theorem shows that the order of poles of $Y(\varepsilon)$ and $M(\varepsilon)$ (denoted above by s and t, respectively), can be computed by a combinatorial algorithm.

Theorem 6 *For $1 \leq i, j \leq n$, let u_{ij} be the order of the pole of $M_{ij}(\varepsilon)$ at zero. Then, there exists an $O(n^4)$ algorithm for computing these orders whose input is the addresses of the nonzero entries in $P(0)$ and in C. Likewise, for $1 \leq i, j \leq n$, let v_{ij} be the order of the pole of $Y_{ij}(\varepsilon)$ at zero. Then,*

$$t \geq \max_{ij}(u_{ij} - u_{jj}) = \max_{ij} v_{ij} = s \quad .$$

The above theorem and a detailed algorithm are given in [34].

Example 3 (cont.). Running the algorithm developed in [34] for computing u_{ij}, $1 \leq i, j \leq 4$, results in

$$(u_{ij}) = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 0 & 2 & 2 \\ 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 0 \end{pmatrix}$$

In particular, $s = t = 2$.

For $m \geq 0$, let $B_m \in R^{n(m+1) \times n(m+1)}$ be the matrix such that each of its $n \times n$ block diagonal matrices equals $I - P(0)$ and each of its $n \times n$ matrices above the diagonal equals $-C$. Namely,

$$B_m = \begin{pmatrix} I-P(0) & -C & 0 & \cdots & 0 & 0 \\ 0 & I-P(0) & -C & \cdots & 0 & 0 \\ 0 & 0 & I-P(0) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I-P(0) & -C \\ 0 & 0 & 0 & \cdots & 0 & I-P(0) \end{pmatrix}.$$

where $I-P(0)$ appears $m+1$ times while $-C$ appears m times. By equating coefficients of the same order in the identity $\pi(\varepsilon) = \pi(\varepsilon)(I - P(0) - \varepsilon C)$, it is easy to conclude that for any $m \geq 0$, $(\pi^{(0)}, \pi^{(1)}, \cdots, \pi^{(m)})$ is a left eigenvector of B_m belonging to the eigenvalue zero. Of course, it is not an unique eigenvector. Also, the dimension of this eigenspace of B_m is non-increasing in m. Let V_m be the subspace of vectors of length n such that the first n entires of vectors in the above mentioned eigenspace of B_m induce. Again, the dimension of V_m is non-increasing in m. Finally, these dimensions are always greater than or equal to 1. The following theorem is taken from [40].

Theorem 7 *(The general case: computing $\pi^{(0)}$).*

$$t + 1 = min\{m;\ dimV_m = 1\} \quad .$$

The above two theorems suuggest a way to compute $\pi^{(0)}$. Specifically, one can first find the value of t by the combinatorial algorithm suggested in [34] and then the left eigenspace of B_t belonging to the eigenvalue zero can be explored. For the latter task some procedures exist. A few of them are based on *reduction* steps. In [39] one reduction step is carried out while [25] and [12] suggest performing a number of reduction steps. The latter number is bounded by t as at the t-th step one gets a non-singular system. Actually, the procedure was defined in the previous subsection and takes care of NCD matrices with one (and terminal) reduction step.

We again note that once $\pi^{(0)}$ and $Y^{(0)}$ are calculated, the other coefficients of the series expansion for $\pi(\varepsilon)$ can be computed by the recurssive formula $\pi^{(m)} = \pi^{(m-1)}U, m \geq 1$, where $U = CY^{(0)}$.

Example 3 (cont.). We have pointed out above that $t = 2$. Hence, solving for the left null space of

$$B_2 = \left(\begin{array}{c|c|c} I-P(0) & -C & 0 \\ 0 & I-P(0) & -C \\ 0 & 0 & I-P(0) \end{array}\right)$$

$$
= \left(\begin{array}{cccc|cccc|cccc}
1 & -1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right)
$$

results in $(0, \alpha, 0, \alpha, *, *, *, *, *, *, *, *)$ for any α.[5] Hence, $\pi^{(0)} = (0, .5, 0, .5)$. In [39] it is shown that the same set of equations which is needed in order to solve for $\pi^{(0)}$ is also all required in order to compute U and $Y^{(-s)}$. For this numerical example, using the methods of either [39] or [12], one obtains

$$
Y^{(0)} = \begin{pmatrix} .5 & -.5 & .5 & -.5 \\ 0 & 0 & 0 & 0 \\ .5 & -.5 & .5 & -.5 \\ 0 & 0 & 0 & 0 \end{pmatrix}
$$

and hence

$$
U = CY^{(0)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 \\ 0.5 & -0.5 & 0.5 & -0.5 \end{pmatrix}
$$

The terms of the singular part of the Laurent expansion for the deviation matrix are:

$$
Y^{(-2)} = \begin{pmatrix} 0 & .25 & 0 & -.25 \\ 0 & .25 & 0 & -.25 \\ 0 & -.25 & 0 & .25 \\ 0 & -.25 & 0 & .25 \end{pmatrix} \quad \text{and} \quad Y^{(-1)} = \begin{pmatrix} .25 & -.5 & -.25 & .5 \\ .25 & 0 & -.25 & 0 \\ -.25 & .5 & .25 & -.5 \\ -.25 & 0 & .25 & 0 \end{pmatrix}
$$

Finally,

$$
M_{21}(\varepsilon) = \frac{Y_{11}(\varepsilon) - Y_{21}(\varepsilon)}{\pi_1(\varepsilon)} = [(\frac{0.25}{\varepsilon} + 0.5 + ...) - (\frac{0.25}{\varepsilon} + 0 + ...)]/[0.5\varepsilon - 0.5\varepsilon^2 + ...] = \frac{1}{\varepsilon} + ...,
$$

[5] The sign '*' corresponds to values not computed (and not relevant for our current purposes.)

$$M_{31}(\varepsilon) = \frac{1}{\varepsilon^2} + \frac{1}{\varepsilon} + ..., \quad M_{41}(\varepsilon) = \frac{1}{\varepsilon^2} + \frac{2}{\varepsilon} + ...,$$

$$M_{12}(\varepsilon) = \frac{1}{\varepsilon} + ..., \quad M_{32}(\varepsilon) = \frac{1}{\varepsilon^2} + \frac{0}{\varepsilon} + ..., \quad M_{42}(\varepsilon) = \frac{1}{\varepsilon^2} + \frac{1}{\varepsilon} + ... \ .$$

We like to point out that the expression for $Y^{(-1)}$ for this example given in HR1 [39] had a numerical error that was corrected in AL [12].

2.5 The case of non ergodic perturbed Markov chains

In this section we briefly discuss the singularity phenomena that can appear if the perturbed Markov chain itself is not ergodic. The reader can find a more detailed exposition in A [9].

In case where the perturbed Markov chain has several ergodic classes plus a non- empty set of transient states, then the perturbed transition matrix can be written in the following canonical form

$$P(\varepsilon) = \begin{bmatrix} P_1(\varepsilon) & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & P_\ell(\varepsilon) & 0 \\ R_1(\varepsilon) & \cdots & R_\ell(\varepsilon) & S(\varepsilon) \end{bmatrix} \begin{array}{l} \} \ \Omega_1 \\ \ \ \vdots \\ \} \ \Omega_\ell \\ \} \ \Omega_T \end{array} \qquad (15)$$

can_pert

where $P_i(\varepsilon) = P_i(0) + \varepsilon C_i$, $R_i(\varepsilon) = R_i(0) + \varepsilon C_{Ri}$, $S(\varepsilon) = S(0) + \varepsilon C_S$. Now note that all invariant measures $m_i(\varepsilon)$ of the perturbed Markov chain can be immediately constructed from the invariant measures of the ergodic classes associated with stochastic matrices $P_i(\varepsilon), i = 1, ..., \ell$. Namely, $m_i(\varepsilon) = [0 \cdots 0 \ \pi_i(\varepsilon) \ 0 \cdots 0]$, where $\pi_i(\varepsilon)$ is uniquely determined by the system

$$\begin{cases} \pi_i(\varepsilon) P_i(\varepsilon) = \pi_i(\varepsilon), \\ \qquad \pi_i(\varepsilon) \underline{1} = 1. \end{cases}$$

This is exactly the perturbation problem under the assumption that there are several ergodic classes plus possibly a set of transient states coupled by the perturbation into a single ergodic class. This issue was covered in the previous subsections.

The challenging task that we face in this case of multi-chain perturbed process is the computation of power series for the *right* eigenvectors of the perturbed Markov chains. The pertubed right eigenvectors contain the probabilities of being absorbed in each of the ergodic classes under the perturbed processes. These quantities exhibit the multi-chain ergodic properties of the perturbed process. For instance, nothing conceptually new would emerge had the set of transient states in the perturbed process been empty.

The right 0-eigenvectors of the perturbed chain can be written in the following form [55]

$$q_i(\varepsilon) = \begin{bmatrix} 0 \\ \underline{1} \\ 0 \\ \varphi_i(\varepsilon) \end{bmatrix} \begin{matrix} \} \ \Omega_i \\ \\ \} \ \Omega_T \end{matrix} \tag{16}$$

reig_pert

where subvector $\varphi_i(\varepsilon)$ is given by

$$\varphi_i(\varepsilon) = (I - S(\varepsilon))^{-1}(\varepsilon) R_i(\varepsilon)\underline{1}, \tag{17}$$

phi_pert

Note that if some ergodic classes become transient after the perturbation, then the matrix valued function $(I - S(\varepsilon))^{-1}$ has a singularity at $\varepsilon = 0$. To further elaborate on this phenomenon, let us consider the structure of the substochastic matrix $S(0)$.

$$S(0) = \begin{bmatrix} \tilde{P}_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \tilde{P}_m & 0 \\ \tilde{R}_1 & \cdots & \tilde{R}_m & \tilde{S} \end{bmatrix}.$$

Blocks $\tilde{P}_1, ..., \tilde{P}_m$ represent the ergodic classes of the original Markov chain which merge with the transient set after the perturbation. Of course, $m = 0$ is possible. Since each of $\tilde{P}_1, ..., \tilde{P}_m$ is a stochastic matrix, we conclude that matrix $I - S(0)$ has zero as an eigenvalue with multiplicity of at least m. Of course, $I - S(0)$ is not invertible. However, the matrix $(I - S(\varepsilon))^{-1}$ exists for small positive (but not zero) values of ε. ¿From the results of [11], it follows that one can expand $(I - S(\varepsilon))^{-1}$ as a Laurent series at $\varepsilon = 0$

$$(I - S(\varepsilon))^{-1} = \frac{1}{\varepsilon^p} U^{(-p)} + ... + \frac{1}{\varepsilon} U^{(-1)} + U^{(0)} + \varepsilon U^{(1)} + ... \tag{18}$$

laur_T

One can use the methods of [11] to calculate the coefficients of the above series. Substituting the expression $R_i(\varepsilon) = R_i(0) + \varepsilon C_{Ri}$ and the Laurent series (18) into formula (17), we obtain the asymptotic expansion

$$\varphi_i(\varepsilon) = \varphi_i^{(0)} + \varepsilon \varphi_i^{(1)} + \varepsilon^2 \varphi_i^{(2)} + ... \ , \tag{19}$$

pow_phi

where

$$\varphi_i^{(m)} = U^{(m)} R_i^{(0)} + U^{(m-1)} C_{Ri}, \quad m \geq 0. \tag{20}$$

phi_coef

The above formulae are valid in the most general setting. Some interesting particular cases are discussed in [9].

3 Singularly Perturbed Markov Decision Processes

In this section we review some results on singular perturbations of Markov decision processes (MDPs) and introduce a unified approach to singular perturbation, Blackwell optimality and branching Markov decision processes. First we briefly intoduce notation and define various optimality criteria. The reader can find a more detailed study on MDPs in the books and surveys Der,HL,How,Kal,Kal2,Put,Vt2 [26, 42, 44, 53, 54, 70, 80] and the references therein.

We consider a discrete-time MDP with a finite state space $\mathbb{X} = \{1, ..., N\}$ and a finite action space $\mathbb{A}(i) = \{1, ..., m_i\}$ for each state $i \in \mathbb{X}$. At any time point t the system is in one of the states $i \in \mathbb{X}$ and the controller or "decision-maker" chooses an action $a \in \mathbb{A}(i)$; as a result the following occur: (a) the controller gains an immediate reward $r(i, a)$, and (b) the process moves to a state $j \in \mathbb{X}$ with probability $p(j|i, a)$, where $p(j|i, a) \geq 0$ and $\sum_{j \in \mathbb{X}} p(j|i, a) = 1$.

A *decision rule* π_t at time t is a function which assigns a probability to the event that any particular action a is taken at time t. In general, π_t may depend on history $h_t = (i_0, a_0, i_1, a_1, ..., a_{t-1}, i_t)$ up to time t. The distribution $\pi_t(\cdot|h_t)$ defines the probability of selecting any action a at time t given the history h_t.

A *policy* (or *strategy*) is a sequence of decision rules $\pi = (\pi_0, \pi_1, ..., \pi_t, ...)$. A policy π is called *Markov* if $\pi_t(\cdot|h_t) = \pi_t(\cdot|i_t)$. If $\pi_t(\cdot|i) = \pi_{t'}(\cdot|i)$ for all $t, t' \in \mathbb{N}$ then the Markov policy π is called *stationary*. Furthermore, a *deterministic* policy π is a stationary policy whose single decision rule is nonrandomized. It can be defined by the function $f(i) = a, a \in \mathbb{A}(i)$.

Let Π, Π^S and Π^D denote the sets of all policies, of all stationary policies and of all deterministic policies, respectively. For the stationary policy $\pi \in \Pi^S$ define the corresponding transition matrix $P(\pi) = \{p_{ij}(\pi)\}_{i,j=1}^N$ and the reward vector $r(\pi) = \{r_i(\pi)\}_{i=1}^N$

$$p_{ij}(\pi) := \sum_{a \in \mathbb{A}(i)} p(j|i, a)\pi(a|i), \quad r_i(\pi) := \sum_{a \in \mathbb{A}(i)} r(i, a)\pi(a|i).$$

Let the *limit matrix* $P^*(\pi)$ and the *deviation matrix* (or *reduced resolvent* as it is sometimes refer to in this setting) $Y(\pi)$ associated with policy π be defined by

$$P^*(\pi) := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} P^{t-1}(\pi)$$

and

$$Y(\pi) := (I - P(\pi) + P^*(\pi))^{-1} - P^*(\pi).$$

The *expected average reward* $g_i(\pi)$ and the *expected discounted reward* $v_i^\lambda(\pi)$ are defined as follows:

$$g_i(\pi) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[P^{t-1}(\pi) r(\pi)\right]_i$$

and

$$v_i^\lambda(\pi) := \sum_{t=1}^{\infty} \lambda^{t-1} \left[P^{t-1}(\pi) r(\pi)\right]_i = \left[(I - \lambda P(\pi))^{-1} r(\pi)\right]_i,$$

where $i \in \mathbb{X}$ is an initial state and $\lambda \in (0,1)$ is a discount factor. One can also use the interest rate $\rho = (1-\lambda)/\lambda \in (0,\infty]$ instead of the discount factor λ. Note that $\lambda \uparrow 1$ is equivalent to $\rho \downarrow 0$. Then, the expected discount reward $v^\rho(\pi) := v^{\lambda(\rho)}(\pi)$ can be expanded as a Laurent series MV,Vt1,LdPut [65, 79, 59] for sufficiently small ρ

$$v^\rho(\pi) = (1+\rho) \left[\rho^{-1} y_{-1}(\pi) + \sum_{m=0}^{\infty} \rho^m y_m(\pi)\right], \tag{21}$$ Laur

where $y_{-1}(\pi) = P^*(\pi) r(\pi)$ and and $y_m = (-1)^m Y(\pi)^{m+1} r(\pi)$ for $m \geq 0$. Note that the above series can be considered as a particular case of resolvent expansion Ka,Rt [52, 73], which in turn can be viewed as a particular case of the inversion of singularly perturbed matrices. Note that in particular $y_{-1}(\pi) = g(\pi)$ and $y_0 = h(\pi)$, where $g(\pi)$ is the expected average reward vector defined above and $h(\pi)$ is the bias vector.

Definition 1 *The stationary policy π_* is called a discount optimal policy for the discount factor $\lambda \in (0,1)$ if $v_i^\lambda(\pi_*) \geq v_i^\lambda(\pi)$ for all $i \in \mathbb{X}$ and all $\pi \in \Pi^S$.*

Definition 2 *The stationary policy π_* is called a gain optimal policy if $g_i(\pi_*) \geq g_i(\pi)$ for each $i \in \mathbb{X}$ and all $\pi \in \Pi^S$.*

It is well known that there exists a deterministic discount policy and a gain optimal policy, each of which can be computed by a number of efficient algorithms Der,Kal,Put [26, 53, 70]. The Laurent expansion (Laur 21) allows us to define more selective optimality criteria MV,Vt1,Vt2 [65, 79, 80].

m-discount **Definition 3** *A policy $\pi^* \in \Pi^S$ is called an m-discount optimal policy for some integer $m \geq 0$ if*

$$[y_m(\pi^*)]_i \geq [y_m(\pi)]_i$$

for all $i \in \mathbb{X}$ and all policies π that are $(m-1)$-discount optimal. By convention, the gain optimal policy is (-1)-discount optimal policy.

In particular, 0-discount optimal policy is called *bias optimal.*

Moreover, it is known that there exists a stationary (and even deterministic) policy that is n-discount optimal for all $n \geq -1$ [65, 79, 80]. It is referred to as the *Blackwell optimal policy* [16] and is defined equivalently in the following way.

Definition 4 *A policy π_* is said to be Blackwell optimal if there exists some $\rho_0 > 0$ such that for all $\rho \in (0, \rho_0]$ $v^\rho(\pi^*) \geq v^\rho(\pi)$ for all $\pi \in \Pi^S$.*

In other words, a Blackwell optimal policy is a policy which is discount optimal policy for any discount factor sufficiently close to one.

The results of singular perturbation theory have several applications to Markov decision processes. The first example is the Laurent series (21) for the expected discount reward. Another important application is singularly perturbed MDP. For the clarity of exposition we restrict ourselves to the case of linear perturbation, namely, we assume that the transition probabilities of the perturbed MDP are given by

$$p^\varepsilon(j|i,a) = p(j|i,a) + \varepsilon d(j|i,a) \geq 0, \tag{22}$$

where $p(j|i,a)$ are transition probabilities of the original unperturbed chain, $\sum_j d(j|i,a) = 0$, and ε is a "small" perturbation parameter. We are especially interested in the case of *singular* perturbations, that is when the perturbation changes the ergodic structure of the underlying Markov chain.

As the following example shows, the policies that are optimal for the unperturbed MDP may not coincide with the optimal policies for the perturbed MDP.

Example 2.1 Let us consider an MDP model with $\mathbb{X} = \{1, 2\}$, $\mathbb{A}(1) = \{a_1, b_1\}$, $\mathbb{A}(2) = \{a_2\}$ and

$$p^\varepsilon(1|1,a_1) = 1, \quad p^\varepsilon(2|1,a_1) = 0;$$
$$p^\varepsilon(1|1,b_1) = 1 - \varepsilon, \quad p^\varepsilon(2|1,b_1) = \varepsilon;$$
$$p^\varepsilon(1|2,a_2) = \varepsilon, \quad P^\varepsilon(2|2,a_2) = 1 - \varepsilon;$$
$$r(1,a_1) = 1, \quad r(1,b_1) = 1.5, \quad r(2,a_2) = 0$$

There are only two deterministic policies, $u = [a_1, a_2]$ and $v = [b_1, a_2]$. For these policies one can easily calculate the average reward vectors. Namely,

$$g^\varepsilon(u) = \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \varepsilon = 0 \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \varepsilon > 0 \end{cases} \quad \text{and} \quad g^\varepsilon(v) = \begin{cases} \begin{bmatrix} 1.5 \\ 0 \end{bmatrix} & \varepsilon = 0 \\ \begin{bmatrix} 0.75 \\ 0 \end{bmatrix} & \varepsilon > 0 \end{cases}$$

Thus, we can see that for $\varepsilon = 0$ the average optimal policy is v, whereas for $\varepsilon > 0$ the average optimal policy is u.

Since often we do not know the exact value of the perturbation parameter ε, we are interested in finding the policy which is "close" to the optimal one for ε small but different from zero. We will call it a *suboptimal* policy or a *limit control optimal* policy. A strict definition will be given later in this section.

Most research in this topic was carried out with the assumption that the unperturbed MDP is completely decomposable and that the perturbed process is ergodic. More precisely, one can introduce the following four assumptions:

A1) $\mathbb{X} = \cup_{k=1}^{p} \mathbb{X}_k$, where $\mathbb{X}_k \cap \mathbb{X}_l = \emptyset$ if $k \neq l$, $n > 1$, and $\sum_{k=1}^{p} card(\mathbb{X}_k) = N$

A2) $p(j|i,a) = 0$ whenever $i \in \mathbb{X}_k, j \in \mathbb{X}_l$ and $k \neq l$.

A3) For every $i = 1, ..., p$ the unperturbed MDP associated with the subspace $\mathbb{X}_k$ is ergodic.

A4) The transition matrix $P^{\varepsilon}(\pi)$ is irreducible for any $\pi \in \Pi^S$ and any ε sufficiently small but different from zero, that is the perturbed MDP is ergodic.

The structure defined by assumptions A1)-A4) has a clear interpretation. The perturbed MDP model can be viewed as a complex system consisting of p "weakly-interacting" subsytems associated with $\mathbb{X}_k, k = 1, ..., p$. Note that perturbation $\varepsilon d(j|i,a)$, where i and j are the states of different subsystems $\mathbb{X}_k$ and $\mathbb{X}_l$ respectively, represents the probability of rare transitions between the subsystems, which are independent in the unperturbed chain.

This model of singularly perturbed MDP was first studied by Pervozvanski and Gaitsgory [67]. They proposed an *aggregation - disaggregation* algorithm for the computation of suboptimal policies for the perturbed MDP with average and discount optimality criteria. Furthermore, they have shown that in the case of discount optimality criterion, an optimal policy for the original problem is also suboptimal for the perturbed problem. The latter is not true in general if one uses the gain optimality criterion. In particular, if the perturbation is singular, the suboptimal policy of the perturbed problem can be quite different from the optimal policies of the unperturbed problem (see Example 2.1). The explanation for this phenomenon is that the interaction between weakly-coupled subsytems in the singularly perturbed process makes an effect only after sufficiently long time interval, and, if one uses discounting, the end of the trajectory has no significant contribution to the expected discount reward.

To investigate the above phenomenon, it was proposed in [24] to consider the discount optimality criterion for the case where the interest rate goes to zero (equivalently, the discount factor goes to one) as the perturbation parameter vanishes. In particular, in [24] it is assumed that $\rho = \rho(\varepsilon) = \mu\varepsilon$ for some constant μ. The latter allows one to exhibit the two time scale behaviour of the singularly perturbed MDP. In addition, the model in [24] admits a set of transient states. The dynamic programming equation for the perturbed model is solved in [24] by using an approach based on lexicographical ordering. This concept was first applied to MDPs by Veinott [79] to calculate a Blackwell optimal policy. Moreover, as one can seen there, singular perturbed MDPs have a close relation with Blackwell optimality [16, 65, 79, 70, 80] and Markov branching decision chains (for comprehensive study of Markov branching chains the reader is refered to the paper by Huang and Veinott [45]). Later, based on [25], results which appeared first in [24], were extended in [71] to include the possibility of several time scales. The key point in [25] and [71] is the use of the following dependency of the interest rate on the perturbation parameter: $\rho = \rho(\varepsilon) = \mu\varepsilon^l$, where l represents the order of a time scale. In [4] Altman and Gaitsgory have generalized the results of [67] to constrained Markov decision processes.

Abbad, Bielecki and Filar [1, 3, 14, 2] formulated the *limit control principle*, which provides a formal setting for the concept of suboptimal policies. First, as elaborated in Section 2, we note that for any stationary policy $\pi \in \Pi^S$, there exists a limiting ergodic projection

$$\hat{P}^*(\pi) := \lim_{\varepsilon \to 0} P^{*,\varepsilon}(\pi).$$

The average reward optimization problem for the perturbed MDP can be written in the form

$$g_i^{opt,\varepsilon} = \max_{\pi \in \Pi^S} [P^{*,\varepsilon}(\pi) r(\pi)]_i \quad (L^\varepsilon).$$

The limit control principle says that instead of the above singular program one can consider a well-defined *limit Markov control problem*:

$$\hat{g}_i^{opt} = \max_{\pi \in \Pi^S} [\hat{P}^*(\pi) r(\pi)]_i \quad (L).$$

It is natural to expect that an optimal strategy to (L), if exists, will approximate well the optimal strategy for the perturbed problem (L^ε), in the case where the perturbation parameter is small. In [14] it was shown that this is indeed the case. Specifically, let π_* be any maximizer in (L), then

$$\lim_{\varepsilon \to 0} \max_{i \in \mathbb{X}} |g_i^\varepsilon(\pi_*) - g^{opt,\varepsilon}| = 0.$$

In [1] it is proved that there exists a deterministic policy that solves the limit control problem (L). Recently, Bielecki and Stettner [15] have generalized the limit control principle to MDPs with general Borel state spaces and compact action spaces.

Under the assumptions A1) – A4) the limit Markov control problem (L) can be solved by solving the following linear programming problem (P):

maximise $\sum_{k=1}^{n} \sum_{i \in \mathbb{X}_k} \sum_{a \in A(i)} r(i,a) z_{ia}^k$

subject to:

$$\sum_{i \in \mathbb{X}_k} \sum_{a \in A(i)} (\delta_{ij} - p(j|i,a)) z_{ia}^k = 0, \quad j \in \mathbb{X}_k; k = 1, \dots, n$$

$$\sum_{k=1}^{n} \sum_{j \in \mathbb{X}_\ell} \sum_{i \in \mathbb{X}_k} \sum_{a \in A(i)} d(j|i,a) z_{ia}^k = 0; \quad \ell = 1, \dots, n$$

$$\sum_{k=1}^{n} \sum_{i \in \mathbb{X}_k} \sum_{a \in A(i)} z_{ia}^k = 1$$

$$z_{ia}^k \geq 0 \quad k = 1, \dots, n; \quad i \in \mathbb{X}_k, a \in A(i).$$

It can be shown (see [3]) that an optimal strategy in the limit Markov control problem (L) can be constructed as follows.

Theorem 8 *Let $\{z_{ia}^k | k = 1, \dots, n;\ i \in \mathbb{X}_k;\ a \in A(i)\}$ be an optimal extreme solution to the linear program (P), then the deterministic strategy defined by*

$$f_*(i) = a, \quad i \in \mathbb{X}_k,\ k = 1, \dots, n \Longleftrightarrow z_{ia}^k > 0$$

is optimal in the limit Markov control problem (L).

The linear program (P) is similar to one given by Gaitsgory and Pervozyanski [33]. However, these authors used techniques different from those in [3].

As a result of the solution of problem (L), one gets *limit control optimal* policy $f_* \in \Pi^D$ such that for any other deterministic policy $f \in \Pi^D$ the holds

$$\lim_{\varepsilon \to 0} (g_i^\varepsilon(f_*) - g_i^\varepsilon(f)) \geq 0, \quad i \in \mathbb{X}.$$

However, as was noted in [1], a policy that solves (L), in general, is only suboptimal. Interestingly, there exists a policy that is optimal for all sufficiently small ε but different from zero (see [1]). This policy is called *uniform optimal* [5] and it satisfies the following inequality

$$g_i^\varepsilon(f_*) \geq g_i^\varepsilon(f), \quad i \in \mathbb{X},$$

for any deterministic policy f and all ε sufficiently small but different from zero. We like to emphasize that a uniform optimal policy is limit control optimal, but the converse need not hold, as the following example illustrates.

unif_examp **Example 2.2** Consider $\mathbb{X} = \{1, 2\}$, $\mathbb{A}(1) = \{a_1, b_1\}$, $\mathbb{A}(2) = \{a_2\}$; let

$$\begin{array}{ll} p^\varepsilon(1|1,a_1) = 1, & r(1,a_1) = 10 \\ p^\varepsilon(1|1,b_1) = 1-\varepsilon, & r(1,b_1) = 10 \\ p^\varepsilon(2|1,b_1) = \varepsilon, & \\ p^\varepsilon(1|2,a_2) = 1, & r(2,a_2) = 0 \end{array}$$

Then the stationary policy $u(1) = a_1, u(2) = a_2$ is uniformly optimal with expected average reward $g_i^\varepsilon(u) = 10$. The stationary policy $v(1) = b_1, v(2) = a_2$ is limit control optimal as $\lim_{\varepsilon \to 0} g_i^\varepsilon(v) = 10$, but for every $\varepsilon > 0$,

$$g_i^\varepsilon(v) = \frac{10}{1+\varepsilon} < g_i^\varepsilon(u).$$

A heuristic procedure to determine the uniform optimal policy is proposed in [1]. In contrast, a well-defined computational procedure was proposed in [5], which is based on asymptotic linear programming. The *asymptotic linear programming* was first introduced by Jeroslow [50, 51] and later refined by Hordijk, Dekker and Kallenberg [43]. It is designed to solve the following parametric linear program

$$\max c(\varepsilon)x \tag{23}$$ 1.3

$$A(\varepsilon)x = b(\varepsilon), \quad x \geq 0 \tag{24}$$ 1.4

for ε small but different from zero. The elements of $A(\varepsilon), b(\varepsilon)$ and $c(\varepsilon)$ are assumed to be polynomials of ε. The basic idea of this method is to perform the simplex algorithm over the field of rational functions instead of the field of real numbers. The latter results in a solution that is optimal for all ε sufficiently small but different from zero. We would like to note that instead of asymptotic linear programming one can use its modification, *the asymptotic simplex method* [67, 57, 58, 7], which operates over the field of Laurent series. Moreover, if the asymptotic linear programming or asymptotic simplex method is used, then one can drop Assumptions A1)-A4) and consider the most general situation.

As can be concluded from the above overview of Markov decision processes, there is an intrinsic relationship between the Blackwell optimality and singular perturbations. Actually, the contributions of Delebecque and Quadrat [24, 25, 71] clearly emphasize this connection. It turns out that the Blackwell optimality, singularly perturbed MDPs and branching Markov decision chains can all be considered in a unified framework based on asymptotic simplex method. Towards this end, let us consider the linear programming formulation for the following four MDP models.

Model I: MDP with Blackwell optimality criterion

A Blackwell optimal policy is a discount optimal policy for all discount factor sufficiently close to one, or, equivalently, the policy which is optimal for all interest rate sufficiently close to zero. A Blackwell optimal policy can be determined by using the asymptotic linear programming approach [6, 43, 57, 58]. Namely, one needs to find a solution for the following linear program which is optimal for all interest rate $\rho > 0$ which are sufficiently small:

$$\max \sum_{i,a} (1+\rho) r(i,a) x_{ia}$$

$$\sum_{i,a} [(1+\rho)\delta_{ij} - p(j|i,a)] x_{ia} = 1, \quad x_{ia} \geq 0. \qquad (25)$$

Note that the above linear program can be immediately written in the form (23),(24) with $\varepsilon = \rho$. In [43] a refined version of Jeroslow's asymptotic linear programming method [50, 51] for solving (25) was developed. Later Lamond [58] applied his version of the asymtotic simplex method to the above parametric LP. Of course, our algorithm [7] can be applied to this model as well. As pointed out in [58], in this particular case, the Laurent series for the basis matrix always admits a simple pole. This in turn implies that the complexity of the basis updating is $O(n^2)$, which is comparable with the complexity of the basis updating for the ordinary simplex method.

Model II: Markov branching decision chains.

The Markov branching decision chains were introduced in [45]. These are MDPs with immediate rewards which dependent on the interest rate. Namely, it is assumed that $r(i,a) = r^{\rho}(i,a)$ is a polynomial in the interest rate ρ. To find a policy which is optimal for all sufficiently small ρ (Strong-value policy [45]), we need to solve only slightly modified version of (25), that is

$$\max \sum_{i,a} (1+\rho) r^{\rho}(i,a) x_{ia}$$

$$\sum_{i,a} [(1+\rho)\delta_{ij} - p(j|i,a)] x_{ia} = 1, \quad x_{ia} \geq 0. \tag{26}$$

Again the asymptotic simplex method can be immediately applied. A method of finding the optimal policy proposed by Huang and Veinott [45] also employs the asymptotic programming approach. The method is based on the subsequent solution of the augmented LPs, whose objective functions and right hand sides (but not the constraint coefficient matrix) depend on the parameter.

Model III: Singularly perturbed MDP with average criterion

In Section 3.4 we applied the asymptotic linear programming [50, 51, 43], which is based on the ordering in the field of rational functions, to singularly perturbed MDPs with average criterion. Since one needs to solve the following parametric program

$$\max \sum_{i,a} r(i,a) x_{ia}$$

$$\sum_{i,a} [\delta_{ij} - p(j|i,a) - \varepsilon d(j|i,a)] x_{ia} = 0,$$

$$\sum_{a} x_{ja} + \sum_{i,a} [(\delta_{ij} - p(j|i,a) - \varepsilon d(j|i,a)] y_{ia} = \beta_j,$$

$$x_{ia} \geq 0, \qquad y_{ia} \geq 0, \tag{27}$$

the asymptotic simplex method, based on the Laurent series expansions, can be applied as well. Moreover, since the basis updating in the asymptotic linear programming takes $O(m^4 log(m))$ flops, and the asymptotic simplex method takes $\bar{s}m^2$ flops. The latter method is computationally superior.

Model IV: Singularly perturbed MDP with killing interest rate

In [24, 71] considered singularly perturbed MDPs with "killing interest rate" $\rho(\varepsilon) = \mu\varepsilon^l$, where l is the order of a time scale were considered. This model exhibits the necessity of different control regimes for the different time scales. The papers [24, 71] provided a lexicographical policy improvement algorithm for the solution of the perturbed dynamic programming equation.

Alternatively, the extension of our asymptotic simplex method for the polynomial perturbation can be used in this problem. Here the parametric linear program takes the following form

$$\max \sum_{i,a} (1+\mu\varepsilon^l) r(i,a) x_{ia}$$

$$\sum_{i,a} [(1+\mu\varepsilon^l)\delta_{ij} - p(j|i,a) - \varepsilon d(j|i,a)] x_{ia} = 1, \quad x_{ia} \geq 0. \qquad (28)$$ QD_model

Generalized Model:

Finally, we would like to note that the Models I,II and IV can be viewed as particular cases of the next unified scheme. Let transition probabilities $p^\varepsilon(j|i,a)$, immediate rewards $r^\varepsilon(i,a)$ and interest rate $\rho(\varepsilon)$ of an MDP model be polynomials of the parameter ε. Then a policy which is optimal for all sufficiently small values of parameter ε can be found from the next perturbed linear program.

$$\max \sum_{i,a} (1+\rho(\varepsilon)) r^\varepsilon(i,a) x_{ia}$$

$$\sum_{i,a} [(1+\rho(\varepsilon))\delta_{ij} - p^\varepsilon(j|i,a)] x_{ia} = 1, \quad x_{ia} \geq 0. \qquad (29)$$ gen_model

The above perturbed LP can be efficiently solved by the generalization of the asymptotic simplex method proposed in AlFlAv [7]. Note that we retrieve Model I with $\rho(\varepsilon)=\varepsilon$, $r^\varepsilon(i,a)=r(i,a)$, $p^\varepsilon(j|i,a)=p(j|i,a)$; Model II with $\rho(\varepsilon)=\varepsilon$, $r^\varepsilon(i,a)=\sum_{k=0}^p \varepsilon^k r_k(i,a)$, $p^\varepsilon(j|i,a)=p(j|i,a)$; and Model IV with $\rho(\varepsilon)=\mu\varepsilon^l$, $r^\varepsilon(i,a)=r(i,a)$, $p^\varepsilon(j|i,a)=p(j|i,a)+\varepsilon d(j|i,a)$.

4 Singularly Perturbed MDP's and the Combinatorial Hamiltonian Cycle Problem

In this section we demonstrate that a famous combinatorial optimisation problem such as the Hamiltonian Cycle Problem (HCP) can be regarded as a singularly perturbed Markov Decision Process. We begin with a brief description of only one version of the HCP.

In graph theoretic terms, the problem is to find a simple cycle of N arcs, that is a *Hamiltonian Cycle* or a *tour*, in a directed graph G with N nodes and with arcs (i,j), or to determine that none exist. Recall that a simple cycle is one that passes exactly once through each node comprising the cycle.

It is known that HCP belongs to the NP-complete class of problems and, as such, is considered very difficult from an algorithmic perspective.

In this section we propose the following, unorthodox, perspective of the Hamiltonian cycle problem: Consider a moving object tracing out a directed path on the graph G with its movement "controlled" by a function f mapping the set of nodes $\mathbb{X} = \{1, 2, \ldots, N\}$ into the set of arcs A.

Clearly, we can think of this set of nodes as the state space of a Markov decision process Γ where for each state/node i, the action space

$$A(i) = \{a = j | (i, j) \in A\}$$

is in one to one correspondence with the set of arcs emanating from that node.

Of course, we shall ignore the trivial case $A(i) = \emptyset$, because in such a case, obviously, no Hamiltonian cycle exists. Furthermore, if we restrict the function f above in such a way that $f(i) \in A(i)$, for each $i \in \mathbb{X}$, then we see that f can be thought of as a deterministic strategy f in an MDP Γ. Designating node 1 as the "home node" in G, we shall say that f is a *Hamiltonian cycle in* G if the set of arcs $\{(1, f(1)), (2, f(2)), \ldots, (N, f(N))\}$ is a Hamiltonian cycle in G. If the above set of arcs contains cycles of length less than N, we shall say that f *has subcycles in* G.

Note that if $P(f)$ is the transition probability matrix of a Markov chain induced by f that is a Hamiltonian cycle, then $P(f)$ is irreducible and the long-run frequency of visits to any state $x_i(f) = 1/N$. Of course, if f has subcycles in G, then $P(f)$ contains multiple ergodic classes which complicates the analysis of the Markov decision process Γ, in which we have embedded our graph theoretic problem.

A class of limiting average Markov decision processes that retains most of the desirable properties of the irreducible processes is the so-called "unichained" class. Briefly, a Markov decision process is *unichained* if for every deterministic stationary control $\mathbf{f}$, $P(f)$ contains only a single ergodic class and possibly a nonempty set of transient states. We now perturb the transition probabilities of Γ slightly to create an ε-perturbed process $\Gamma(\varepsilon)$ (for $0 < \varepsilon < 1$) defined by:

$$p^{\varepsilon}(j|i,a) = \begin{cases} 1 & \text{if } i = 1 \text{ and } a = j \\ 0 & \text{if } i = 1 \text{ and } a \neq j \\ 1 & \text{if } i > 1 \text{ and } a = j = 1 \\ \varepsilon & \text{if } i > 1,\ a \neq j, \text{ and } j = 1 \\ 1 - \varepsilon & \text{if } i > 1,\ a = j, \text{ and } j > 1 \\ 0 & \text{if } i > 1,\ a \neq j, \text{ and } j > 1. \end{cases}$$

Note that with the above perturbation, for each pair of nodes i, j (not equal to 1) corresponding to a "deterministic arc" (i,j) our perturbation replaces that arc by a pair of "stochastic arcs" $(i,1)$ and (i,j):

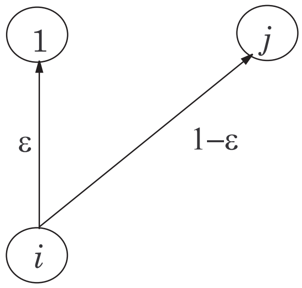

Figure 1:

with weights ε and $(1-\varepsilon)$ respectively ($\varepsilon \in (0,1)$). Note that this perturbation changes Γ to an ε-perturbed Markov decision process $\Gamma(\varepsilon)$.

Example 4.1
Consider the following complete graph G on four nodes (with no self-loops):

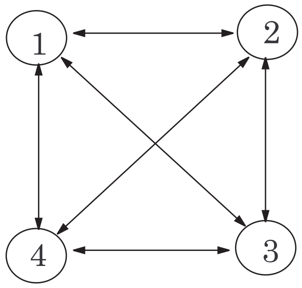

Figure 2:

and think of the nodes as the states of an MDP, denoted by Γ, and of the arcs emanating from a given node as actions available at that state. The Hamiltonian cycle $c_1 : 1 \to 2 \to 3 \to 4 \to 1$ corresponds to the deterministic stationary strategy $f_1 : \{1,2,3,4\} \to \{2,3,4,1\}$ where $f_1(2) = 3$ corresponds to the controller choosing arc (2,3) in state 2 with probability 1. The Markov chain induced by f_1 is given by the transition matrix

$$P(f_1) \quad = \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

which is irreducible. On the other hand, the union of two sub-cycles: $1 \to 2 \to 1$ and $3 \to 4 \to 3$ corresponds to the policy $f_2 : \{1, 2, 3, 4\} \to \{2, 1, 4, 3\}$ which identifies the Markov chain transition matrix

$$P(f_2) \quad = \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

containing two distinct ergodic classes.

As mentioned earlier, the perturbation destroys multiple ergodic classes and induces a unichained, singularly perturbed, Markov decision process $\Gamma(\varepsilon)$. For instance, the policy f_2 now has the Markov chain matrix

$$P(f_2) \quad = \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \varepsilon & 0 & 0 & 1-\varepsilon \\ \varepsilon & 0 & 1-\varepsilon & 0 \end{pmatrix}.$$

Remark 4.1
Our perturbation has made the home node/state 1 rather special. In particular, the home state always belongs to the single ergodic class of $P(f)$ for any $f \in \Pi^S$. Of course, some other states could be transient.

We shall undertake the analysis of the Hamiltonian cycle problem in the "frequency space" of the perturbed process $\Gamma(\varepsilon)$. Recall that with every $f \in \Pi^S$ we can associate the long-run frequency vector $\mathbf{x}(f)$. This is achieved by defining a map $M : \Pi^S \to X(\varepsilon)$ by

$$x_{ia}(f) = \pi_i(\varepsilon, f) f(a|i); \quad f \in \Pi^S$$

for each $i \in \mathbb{X}$ and $a \in A(i)$, where $\pi_i(\varepsilon, f)$ is the i-th element of the stationary distribution vector of the perturbed Markov chain transitions matrix $P(f)$, and $f(a|i)$ is the probability of choosing action a in state i.

Now consider the polyhedral set $\mathbf{X}(\varepsilon)$ defined by the constraints

(i) $\displaystyle\sum_{i=1}^{N} \sum_{a \in A(i)} [\delta(i,j) - p_\varepsilon(j|i,a)]\, x_{ia} = 0;\ j \in \mathbb{X}.$

(ii) $\displaystyle\sum_{i=1}^{N} \sum_{a \in A(i)} x_{ia} = 1.$

(iii) $x_{ia} \geq 0;\ a \in A(i),\ i \in \mathbb{X}.$

Next define a map $\hat{M} : \mathbf{X}(\varepsilon) \to \Pi^S$ by

$$f_{\mathbf{x}}(i,a) \quad = \quad \begin{cases} \frac{x_{ia}}{x_i}; & \text{if } x_i = \sum_{a \in A(i)} x_{ia} > 0 \\ 1; & \text{if } x_i = 0 \text{ and } a = a_1 \\ 0; & \text{if } x_i = 0 \text{ and } a \neq a_1, \end{cases}$$

for every $a \in A(i)$, $i \in \mathbb{X}$ where a_1 denotes the first available action in a given state according to some ordering. The following result can be found in [28] and [31].

Lemma 4.1

(i) *The set* $\mathbf{X}(\varepsilon) = \left\{\mathbf{x}(f) | f \in \Pi^S\right\}$ *and will henceforth be called the (long-run) "frequency space" of* $\Gamma(\varepsilon)$.

(ii) *For every* $\mathbf{x} \in \mathbf{X}(\varepsilon)$,

$$M\left(\hat{M}(\mathbf{x})\right) = \mathbf{x}$$

but the inverse of M *need not exist.*

(iii) *If* $\mathbf{x}$ *is an extreme point of* $\mathbf{X}(\varepsilon)$, *then*

$$f_{\mathbf{x}} = \hat{M}(\mathbf{x}) \in \Pi^D.$$

(iv) *If* $f \in \Pi^D$ *is a Hamiltonian cycle, then* $\mathbf{x}(\mathbf{f})$ *is an extreme point of* $\mathbf{X}(\varepsilon)$.

We shall now derive a useful partition of the class Π^D of deterministic strategies that is based on the graphs they "trace out" in G. In particular, note that with each $f \in \Pi^D$ we can associate a subgraph G_f of G defined by

$$\text{arc } (i,j) \in G_f \iff f(i) = j$$

We shall also denote a simple cycle of length m and beginning at 1 by a set of arcs

$$c_m^1 = \{(i_1 = 1, i_2), (i_2, i_3), \ldots, (i_m, i_{m+1} = 1)\}\,; \quad m = 2, 3, \ldots, N.$$

Of course, c_N^1 is a Hamiltonian cycle. If G_f contains a cycle c_m^1 we write $G_f \supset c_m^1$. For $2 \le m \le N$, let $C_m := \left\{f \in \Pi^D | G_f \supset c_m^1\right\}$,namely, the set of deterministic strategies that trace out a simple cycle of length m, beginning at 1, for each $m = 2, 3, \ldots, N$. Of course, C_N is the set of strategies that correspond to Hamiltonian cycles and any single C_m can be empty, depending on the structure of the original graph G. Thus a typical strategy $f \in C_3$, for example, traces out a graph G_f in G that might look like Figure 3 where the dots indicate the "immaterial" remainder of G_f that corresponds to states that are transient in $P(f)$.

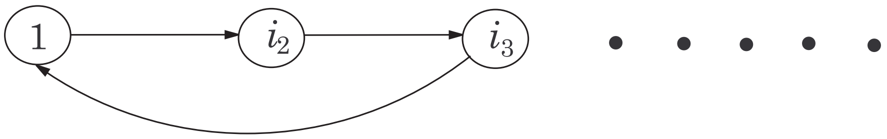

Figure 3:

The partition of the deterministic strategies that seems to be most relevant for our purposes is

$$\Pi^D \quad = \quad \left[\bigcup_{m=2}^{N} C_m\right] \bigcup B, \tag{30}$$

eq2.5.1

where B contains[6] all the deterministic strategies that are not in any of the C_m's. Note that a typical strategy f in B traces out a graph G_f in G that might looks as follows:

[6]It will soon be seen that the strategies in B are in a certain sense "bad" or, more precisely, difficult to analyse, thereby motivating the symbol B.

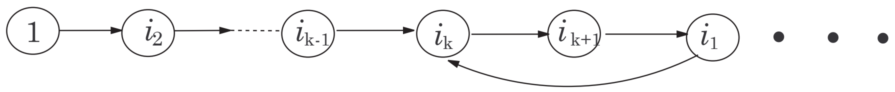

Figure 4:

where the dots again denote the immaterial part of G_f. However, it is important to note that for any $\varepsilon > 0$, the states $1, i_2, \ldots, i_{k-1}$ are not transient in $\Gamma_\alpha(\varepsilon)$.

It is, perhaps, interesting to observe that all strategies in a given set in the partition (30) induce the same long-run frequency $x_1(f)$ of visits to the home node 1. This observation is captured in the following proposition which can be found in [66] and [31]

Proposition 4.2
Let $\varepsilon \in (0,1)$, $f \in \Pi^D$, and $\mathbf{x}(f)$ be its long-run frequency vector (that is, $\mathbf{x}(f) = M(f)$). The long-run frequency of visits to the home state 1 is given by

$$x_1(f) = \sum_{a \in A(1)} x_{1a}(f) = \begin{cases} \dfrac{1}{d_m(\varepsilon)}; & \text{if } f \in C_m,\ m = 2,3,\ldots,N \\ \dfrac{\varepsilon}{1+\varepsilon}; & \text{if } f \in B, \end{cases}$$

where $d_m(\varepsilon) = 1 + \sum_{i=2}^{m} (1-\varepsilon)^{i-2}$ for $m = 2,3,\ldots,N$.

The above proposition leads the following characterizations of the Hamiltonian cycles of a directed graph.

Theorem 9

(i) Let $f \in \Pi^D$ be a Hamiltonian cycle in the graph G. Then $G_f = c_N^1$, $\mathbf{x}(f)$ is an extreme point of $\mathbf{X}(\varepsilon)$ and $x_1(f) = \frac{1}{d_N(\varepsilon)}$.

(ii) Conversely, suppose that $\mathbf{x}$ is an extreme point of $\mathbf{X}(\varepsilon)$ and that $x_1 = \sum_{a \in A(1)} x_{1a} = \frac{1}{d_N(\varepsilon)}$, then $f = \hat{M}(\mathbf{x})$ is a Hamiltonian cycle in G.

(iii) Hamiltonian cycles of the graph G are in $1:1$ correspondence with those points of $\mathbf{X}(\varepsilon)$ which satisfy

(a) $$x_1 = \sum_{a \in A(1)} x_{1a} = \frac{1}{d_N(\varepsilon)}.$$

(b) For every $i \in \mathbb{X}$, $x_i = \sum\limits_{a \in A(1)} x_{ia} > 0$ and $\frac{x_{ia}}{x_i} \in \{0,1\}$ for each $a \in A(i)$, $i \in \mathbb{X}$.

Remark 4.2: It is, perhaps, significant to note that for all $\varepsilon \in (0,1)$, $m = 2, 3, \dots, N-1$

$$\frac{1}{d_m(\varepsilon)} > \frac{1}{d_{m+1}(\varepsilon)} > \frac{\varepsilon}{1+\varepsilon}.$$

Thus Theorem 9 demonstrates that the extreme points $\mathbf{x}$ of $\mathbf{X}(\varepsilon)$ can be "ranked" according to their values of the linear function $l(\mathbf{x}) = \sum_{a \in A(1)} x_{1a}$. Unfortunately, the Hamiltonian cycles (if they exist) may attain only the "second lowest" value of $l(\mathbf{x})$, namely, $\frac{1}{d_N(\varepsilon)}$. The latter problem is, partially, rectified in the following result that can be found in [29].

Theorem 10 *Let $f^* \in \Pi^D$ be a Hamiltonian cycle in the graph G. Then for $\varepsilon \geq 0$ and sufficiently small f^* is a global minimiser of the following perturbed optimisation problem:*

$$\min_{f \in \Pi^D} \{[I - P(f) + P^*(f)]^{-1}_{11}\},$$

where H_{11} denotes the $(1,1)^{th}$ entry of matrix H.

The above theorem shows that if Hamiltonian cycles exist, they are the minimisers of the top left element of the fundamental matrix, over $f \in \Pi^D$, as long as ε is sufficiently near 0. However, from the optimisation point of view, elements of the fundamental matrix are not straightforward to analyse. This leads naturally to the *open problem*: can asymptotic expansions such as the one in Theorem 5, facilitate algorithmic approaches to the optimisation problem in Theorem 10?

To date the best algorithmic results based on this stochastic approach to the HCP are reported in [8]. These results exploit the properties in (i) – (iii) of Theorem 9, above. In particular, it can be checked that the most awkward requirement $x_{ia}/x_i \in \{0,1\}$ for all $i \in \mathbb{X}, a \in \mathbb{A}(i)$ is equivalent $\min\{x_{ia}, x_{ib}\} = 0$ for all $i \in \mathbb{X}$, $a, b \in \mathbb{A}(i)$ and $a \neq b$. This observation immediately leads to the following mixed integer programming formulation of the HCP problem:

$$
\begin{array}{rrcll}
 & \multicolumn{4}{l}{\min \sum_i \sum_a c_{ia} x_{ia}} \\
\text{s.t.} & x & \in & X(\varepsilon) & \\
 & x_1 & = & 1/d_N(\varepsilon) & \\
 & x_{ia} & \leq & M y_{ia} & : \; i \in \mathbb{X}, a \in \mathbb{A}(i) \\
 & y_{ia} + y_{ib} & \leq & 1 & ; \; i \in \mathbb{X}, a, b \in \mathbb{A}(i), a \neq b \\
 & y_{ia} & \in & \{0,1\} & ; \; i \in \mathbb{X}, a \in \mathbb{A}(i).
\end{array}
$$

In the above $M \geq 1/d_N(\varepsilon)$ and c_{ia}'s can be experimented with. In preliminary numerical experiments reported in [8], randomly generated problems with up to 100 nodes and 300 arcs were solved in less than 150 cpu seconds on a Sun Workstation. Recently, Filar and Lasserre [30] proposed a non-standard branch and bound algorithm for the formulation

$$
\begin{array}{ll}
 & \min \sum\limits_i \sum\limits_a c_{ia} x_{ia} \\
\text{s.t.} & \text{(a)} \quad \mathrm{x} \in \mathrm{X}(\varepsilon) \\
 & \text{(b)} \quad \mathrm{x}_1 = 1/\mathrm{d}_\mathrm{N}(\varepsilon) \\
 & \text{(c)} \quad \mathrm{x}_{\mathrm{ia}}/\mathrm{x}_\mathrm{i} \in \{0,1\}; \; \mathrm{i} \in \mathbb{X}, \; \mathrm{a} \in \mathbb{A}(\mathrm{i})
\end{array}
$$

which exploits the structural property that ensures that if the relaxation omitting (c) is solved by the simplex method, then (c) can be violated for at most one i and at most one pair of arcs a, b.

In an interesting, related, development Feinberg [27] has recently considered the embedding of a Hamiltonian cycle problem in a discounted Markov decision process. In that paper the perturbation parameter ε is not necessary but, instead, the discount factor $\lambda \in [0,1)$ plays a crucial role. In particular Feinberg's embedding can be obtained by setting $\varepsilon = 0$ in $p_1^\varepsilon(j|i,a)$'s defined earlier and by setting

$$
r(i,a) = \begin{cases} 1 & \text{if } \mathrm{i} = 1, \; \mathrm{a} \in \mathbb{A}(1) \\ 0 & \text{otherwise} \end{cases}
$$

For any $\pi \in \Pi^S$ the expected discounted reward $v_i^\lambda(\pi)$ is now defined as in Section 3. The analysis in [27] is based on the following interesting observation. Let i_m denote the state/node visited at stage m, then

$$
v_1^\lambda(\pi) = \sum_{m=0}^{\infty} \lambda^m P_1^\pi(i_m = 1),
$$

where $P_1^\pi(\cdot)$ denotes the probability measure induced by π and the initial state $i_0 = 1$, and

$$P_1^\pi(i_m = 1) = \frac{1}{m!}\left[\frac{\partial^m}{\partial\lambda^m}(v_1^\lambda(\pi))\right]_{\lambda=0}.$$

The above lead to novel characterisations of Hamiltonian cycles that are summarised below.

Theorem 11 *With the embedding in Γ_λ described above the following statements are equivalent:*

(i) A policy $\pi = f$ is deterministic and a Hamiltonian cycle in G.

(ii) A policy π is stationary and a Hamiltonian cycle in G.

(iii) A policy f is deterministic and $v_1^\lambda(f) = (1-\lambda^N)^{-1}$ for at least $\lambda \in [0,1)$.

(iv) A policy π is stationary and $v_1^\lambda(\pi) = (1-\lambda^N)^{-1}$ for $2N-1$ distinct discount factors $\lambda_k \in (0,1)$; $k = 1, 2, \ldots, 2N-1$.

The above characterisation naturally leads to a number of mathematical programming formulations of both HCP and TSP that are described in Fein [27]. There is clearly a need to explore the algorithmic potential of these formulations.

References

AbFl [1] M. Abbad and J.A. Filar, "Perturbation and stability theory for Markov control problems", *IEEE Trans. Auto. Contr.*, AC-37, no.9, pp.1415-1420, 1992.

AFs [2] M. Abbad and J.A. Filar, "Algorithms for singularly perturbed Markov control problems: A survey", in *Techniques in discrete-time stochastic control systems* (ed.) C.T. Leondes, Series: Control and Dynamic Systems, v.73, Academic Press, New York, 1995.

AbFlBl [3] M. Abbad, J.A. Filar and T.R. Bielecki, Algorithms for singularly perturbed limiting average Markov control problems. IEEE Trans. on Automatic Control 37:1421-1425, 1992.

Al [4] E. Altman and V.G. Gaitsgory, "Stability and Singular Perturbations in Constrained Markov Decision Problems", *IEEE Trans. Autom. Control*, v.38, pp.971-975, 1993.

AlAvFl [5] E. Altman, K.E. Avrachenkov, and J.A. Filar, "Asymptotic linear programming and policy improvement for singularly perturbed Markov decision processes", *ZOR: Math. Meth. Oper. Res.*, v.49, pp.97-109, 1999.

AHK [6] E. Altman, A. Hordijk, and L.C.M. Kallenberg, "On the value function in constrained control of Markov chains", *ZOR: Math. Meth. Oper. Res.*, v.44, pp.387-399, 1996.

AlFlAv [7] J.A. Filar, E. Altman and K.E. Avrachenkov, "An asymptotic simplex method for singularly perturbed linear programs", submited to *Operations Research Letters*, 1999.

And [8] M. Andramonov, J. Filar, A. Rubinov and P. Pardalos, "Hamiltonian Cycle Problem via Markov Chains and Min-type Approaches"in *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*, Ed. P.M. Pardalos, Kluwer Aacademic Publishers, to appear.

A [9] K.E. Avrachenkov, "Analytic perturbation theory and its applications", PhD Thesis, University of South Australia, 1999.

AH [10] K.E. Avrachenkov and M. Haviv, "The highest singular coefficients in singular perturbation of stochastic marices," (to appear).

AHH [11] K.E. Avrachenkov, M. Haviv and P.G. Howlett, "Inversion of analytic matrix functions that are singular at the origin", submitted to *SIAM J. Matr. Anal. Appl.*

AL [12] K.E. Avrachenkov and J.B. Lasserre, "The fundamental matrix of singularly perturbed Markov chains," *Advances in Applied Probability*, v.31, (to appear).

AL2 [13] K.E. Avrachenkov and J.B. Lasserre, "Perturbation analysis of reduced resolvents and generalized inverses", LAAS Report No. 98520, also submited to *Numerische Mathematik*, 1998.

BlFl [14] T.R. Bielecki and J.A. Filar, "Singularly perturbed Markov control problem: Limiting average cost", *Annals O.R.*, v.28, pp.153-168, 1991.

BlSt [15] T.R. Bielecki and L. Stettner, "Ergodic control of singularly perturbed Markov process in discrete time with general state and compact action spaces", submitted, 1996.

Black [16] D. Blackwell, "Discrete dynamic programming", *Ann. Math. Stat.*, v.33, pp.719-726, 1962.

CW [17] M. Cordech, A.S. Willsky, S.S. Sastry and D.A. Castanon, *Hierarchical aggregation of linear systems with multiple time scales*, IEEE Trans. Autom. Contr., **AC-28**, (1983) 1029-1071.

CdWl [18] M. Cordech, A.S. Willsky, S.S. Sastry, and D.A. Castanon, "Hierarchical aggregation of singularly perturbed finite state Markov processes," *Stochastics.* v.8, pp.259-289, 1983.

Co [19] P.J. Courtois, *Decomposability: queueing and computer system applications*, Academic Press, New York, 1977.

CL [20] P.J. Courtois and G. Louchard, "Approximation of eigencharacteristics in nearly-completely decomposable stochastic systems", *Stoch. Process. Appl.*, v.4, pp.283-296, 1976.

CS [21] P.J. Courtois and P. Semel, "Bounds for the positive eigenvectors of non-negative matrices and their approximation by decomposition", *JACM*, v.31, pp.804-825, 1984.

Dek [22] R. Dekker, *Denumerable Markov decision chains: optimal policies for small interest rate*, Ph.D. Thesis, Univ. of Leiden.

DH [23] R. Dekker and A. Hordijk, "Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards", *Math. Oper. Res.*, v.13, pp.395-420, 1988.

DqQt [24] F. Delebecque and J.P. Quadrat, "Optimal control of Markov chains admitting strong and weak interactions", *Automatica*, v.17, pp.281-296, 1981.

D [25] F. Delebecque, *A reduction process for perturbed Markov chain*, SIAM J. Appl. Math., **43**, (1983) 325-350.

Der [26] C. Derman, *Finite state Markovian decision processes*, Academic Press, New York, 1970.

Fein [27] E.A. Feinberg, "Constrained Discounted Markov Decision Processes and Hamiltonian Cycles", MOR, to appear.

FK [28] J.A. Filar and D. Krass, "Hamiltion cycles and Markov chains," *Mathematics of Operations Research*, Vol. 19, pp. 223–237, 1994.

FiLi [29] J.A. Filar and Ke Liu, "Hamiltonian Cycle Problem and Singularly Perturbed Decision Process", in *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*, IMS Lecture Notes – Monograph Series, USA, 1996.

FL [30] J.A. Filar and J-B Lasserre, "A Non-Standard Branch and Bound Method for the Hamiltonian Cycle Problem", LAAS Report 98247, June 1998.

FiVr [31] J.A. Filar and K. Vrieze, *Competitive Markov Decision Processes,* Springer-Verlag, N.Y., 1996.

GP [32] V.G. Gaitsgori and A.A. Pervozvanskii, "Aggregation of states in a Markov chain with weak interactions", *Cybernetics*, v.11, pp.441-450, 1975. (Translation of Russian original in *Kibernetika*, v.11, pp.91-98, 1975.)

GP2 [33] V.G. Gaitsgori and A.A. Pervozvanskii, *Theory of Suboptimal Decisions*, Kluwer Academic Publishers, 1988.

HH [34] R. Hassin, and M. Haviv, "Mean passage times and nearly uncoupled Markov chains," *SAIM Journal of Discret Mathematics*, Vol. 5, pp. 386–397, 1992.

Ha2 [35] M. Haviv, "An approximation to the stationary distribution of a nearly completely decomposable Markov chain and its error analysis," *SIAM Journal on Algebraic and Discrete Methods*, vol. 7, pp. 589–594, 1986.

Ha3 [36] M. Haviv, "Aggregation/disaggregation methods for computing the stationary distribution of a Markov chain," *SIAM Journal on Numerical Analysis*, vol. 24, pp. 952–966, 1987.

Ha1 [37] M, Haviv, "More on the Rayliegh-Ritz refinement technique for nearly uncoupled matrices," *SIAM Journal of Matrix Analysis and Application*, Vol. 10, pp. 287–293, 1989.

[38] M. Haviv and Y. Ritov, "An approximation to the stationary distribution of a nearly completely decomposable Markov chain and its error bounds," *SIAM Journal on Algebraic and Discrete Methods*, vol. 7, pp. 583–588, 1986.

HR1 [39] M. Haviv and Y. Ritov, "Series expansions for stochastic matrices," unpublished manuscript, 1989.

HR2 [40] M. Haviv and Y. Ritov, "On series expansions for stochastic matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 14, pp. 670–677, 1993.

HV [41] M. Haviv and L. Van der Heyden, "Perturbation bounds for the stationary probabilities of a finite Markov chain," *Advances in Applied Probability*, vol. 16, pp. 804–818, 1984.

HL [42] O. Hernandez-Lerma and J.B. Lasserre, *Discrete- time Markov control processes: basic optimality criteria*, Springer- Verlag, New York, 1996.

HDK [43] A. Hordijk, R. Dekker, and L.C.M. Kallenberg, "Sensitivity analysis in discounted Markovian decision problems", *Spectrum*, v.7, pp.143-151, 1985.

How [44] R.A. Howard, *Dynamic programming and Markov processes*, Cambridge, MA: MIT Press, 1960.

HV [45] Y. Huang and A.F. Veinott, Jr., "Markov branching decision chains with interest-rate-dependent rewards", *Probability in the Engineering and Information Sciences*, v.9, pp.99-121, 1995.

Hu [46] Y. Huang, "A canonical form for pencils of matrices with applications to asymptotic linear programs, *Lin. Alg. Appl.*, v.234, pp.97-123, 1996.

Hu1 [47] J.J. Hunter, "Stationary distributions of perturbed Markov chains", *Lin. Alg. Appl.*, v.82, pp.201-214, 1986.

Hu2 [48] J.J. Hunter, "The computation of stationary distributions of Markov chains through perturbations", *J. Appl. Math. Stoch. Anal.*, v.4, pp.29-46, 1991.

Hu3 [49] J.J. Hunter, "A survey of generalized inverses and their use in applied probability", *Math. Chronicle*, v.20, pp.13-26, 1991.

Jer1 [50] R.G. Jeroslow, "Asymptotic Linear Programming", *Oper. Res.*, v.21, pp.1128-1141, 1973.

Jer2 [51] R.G. Jeroslow, "Linear Programs Dependent on a Single Parameter", *Disc. Math.*, v.6, pp.119-140, 1973.

Ka [52] T. Kato, *Perturbation theory for linear operators*, Springer-Verlag, Berlin, 1966.

Kal [53] L. C. M. Kallenberg, *Linear programming and finite Markovian control problems*, Mathematical Centre Tracts 148, Amsterdam, 1983.

Kal2 [54] L. C. M. Kallenberg, "Survey of linear programming for standard and nonstandard Markovian control problems, Part I: Theory", *ZOR - Methods and Models in Operations Research*, v.40, pp. 1-42, 1994.

KS [55] J.G. Kemeny and J.L. Snell, *Finite Markov Chains,* Von Nostrand, New York, 1960.

KT [56] V.S. Korolyuk and A.F. Turbin, *Mathematical foundations of the state lumping of large systems*, Naukova Dumka, Kiev, 1978, (in Russian), translated by Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

Lam1 [57] B.F. Lamond, "A generalized inverse method for asymptotic linear programming", *Math. Programming*, v.43, pp.71-86, 1989.

Lam2 [58] B.F. Lamond, "An efficient basis update for asymptotic linear programming", *Lin. Alg. Appl.*, v.184, pp.83-102, 1993.

LdPut [59] B. F. Lamond and M. L. Puterman, "Generalized inverses in discrete time Markov decision processes", *SIAM J. Matrix Anal. Appl.*, v.10, pp.118-134, 1989.

Ls [60] J.B. Lasserre, "A formula for singular perturbation of Markov chains", *J. Appl. Prob.*, v.31, pp.829-833, 1994.

LaLu [61] G. Latouche and G. Louchard, "Return times in nearly completely decomposable stochastic processes", *J. Appl. Prob.*, v.15, pp.251-267, 1978.

Lat [62] G. Latouche, "First passage times in nearly decomposable Markov chains", in *Numerical solution of Markov chains*, Workshop 1990, *Pure Prob. Appl.*, v.8, pp.401-411, 1991.

LuLa [63] G. Louchard and G. Latouche, "Geometric bounds on iterative approximations for nearly completely decomposable Markov chains", *J. Appl. Prob.*, v.27, pp.521-529, 1990.

Me [64] C.D. Meyer, "The role of the group generelized inverse in the theory of finite Markov chains," *SIAM Review*, Vol. 17, pp. 443–464, 1975.

MV [65] B. L. Miller and A. F. Veinott, Jr., "Discrete dynamic programming with a small interest rate", *Ann. Math. Stat.* v.40, pp.366-370, 1969.

MF [66] Ming Chen and J.A. Filar (1992),"Hamiltonian Cycles, Quadratic Programming and Ranking of Extreme Points" in *Global Optimization,* C. Floudas and P. Pardalos, eds. Princeton University Press.

PG [67] A.A. Pervozvanski and V.G. Gaitsgori, *Theory of suboptimal decisions*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988, (Translation from the Russian orginal: *Decomposition, aggregation and approximate optimization*, Nauka, Moscow, 1979.)

PS [68] A.A. Pervozvanskii and I.N. Smirnov, "Stationary-state evaluation for a complex system with slowly varying couplings", *Cybernetics*, v.10, pp.603-611, 1974. (Translation of Russian original in *Kibernetika,* v.10, pp.45-51, 1974.)

PhKo [69] R.G. Phillips and P.V. Kokotovic, "A singular perturbation approach to modeling and control of Markov chains", *IEEE Trans. Auto. Contr.*, AC-26, no.5, pp.1087-1094, 1981.

Put [70] M. L. Puterman, *Markov decision processes*, John Wiley & Sons, New York, 1994.

Q [71] J.P. Quadrat, "Optimal control of perturbed Markov chains: the multitime scale case", in *Singular perturbations in systems and control*, ed. M.D. Ardema, CISM Courses and Lectures no.280, Springer-Verlag, Wien - New York, 1983.

RW [72] J.R. Rohlicek and A.S. Willsky, "The reduction of Markov generators: An algorithm exposing the role of transient states", *JACM*, v.35, pp. 675-696, 1988.

Rt [73] U. Rothblum, *Resolvent expansions of matrices and applications*, Linear Algebra Appl., **38**, (1981) 33–49.

Sc1 [74] P.J. Schweitzer, "Perturbation theory and finite Markov chains," *J. Appl. Probability*, Vol. 5, pp. 401–413, 1968.

Sc2 [75] P.J. Schweitzer, "Perturbation series expansion of nearly completely-decomposable Markov chains," Working Paper Series No. 8122, The Graduate School of Management, The University of Rochester, 1981.

Sc3 [76] P.J. Schweitzer, "Perturbation series expansion of nearly completely-decomposable Markov chains," appeared in *Teletraffic Analysis and Computer Performance Evaluation,* by O.J. Boxma, J.W. Cohen and H.C. Tijms (editors), pp. 319–328, Elsevier Science Publishers B.V. (North Holland), 1986.

SA [77] H.A. Simon and A. Ando, "Aggregation of variables in dynamic systems", *Econometrica*, v.29, pp.111-138, 1961.

Vat [78] H. Vantilborgh, "Aggregation with an error of $O(\varepsilon^2)$", *Journal of the Association for Computing Machinery*, v.32, pp.161-190, 1985.

Vt1 [79] A. F. Veinott, Jr., "Discrete dynamic programming with sensitive discount optimality criteria", *Ann. Math. Stat.* v.40, pp.1635-1660, 1969.

Vt2 [80] A. F. Veinott, Jr., "Markov decision chains", in *Studies in Optimization*, eds. G. B. Dantzig and B. C. Eaves, pp. 124-159, 1974.

YZ [81] G.G. Yin and Q. Zhang, "Continuous-time Markov chains and applications: A singular perturbation approach", Series: Applications of Mathematics, v.37, Springer-Verlag, New York, 1998.