NUMERICAL SOLUTION of MARKOV CHAINS, p. 1–15

The first Laurent series coefficients for singularly perturbed stochastic matrices

Konstantin E. Avrachenkov[1], Moshe Haviv[2]

[1] *INRIA Sophia Antipolis, 2004 Route des Lucioles, B.P.93, 06902, France, e-mail: k.avrachenkov@sophia.inria.fr*
[2] *Department of Statistics, The Hebrew University, 91905, Jerusalem, Israel, e-mail: haviv@mscc.huji.ac.il*

KEY WORDS: Markov chains, mean first passage times, deviation matrix, singular perturbations, aggregation/disaggregation, Laurent series

ABSTRACT

There are a few procedures for computing the Laurent series expansions for the mean passage time matrix and for the deviation matrix of a singularly perturbed Markov chain. We suggest here a method for computing the first terms in these expansions in a way which highlights the system dynamics in various time scales.

1. Introduction

There are a few procedures for computing the Laurent series expansions for the mean passage time matrix and for the deviation matrix of a singularly perturbed Markov chain. We suggest here a method for computing the first terms in these expansions in a way that highlights the system dynamics in various time scales.

One usually refers to singular perturbation in Markov chains as the case in which a slight change in the entries of the corresponding transition (stochastic) matrix, results in structural changes in the underlying process. In particular, sets of states which in the original case do not communicate, do so after a perturbation is imposed.* Also, passage times that originally were not well defined random variables, may become so after the perturbation (but with expected values going to infinity when the size of the perturbation goes to zero). For important early works on singularly perturbed Markov chains see [4, 5, 13, 14, 15, 16, 18] and for more complete bibliography on the subject see the recent survey paper [2].

Loosely speaking, dynamics in singularly perturbed Markov chains have a few time scales. One time scale may correspond to the more frequent transitions occurring among states which

*Note that for the reverse case, namely when communicating states stop communicating under the perturbed process, a quantum change is required and hence this possibility is excluded under singular perturbation.

communicate also in the unperturbed case. Another time scale, considered as the higher scale, may correspond to the less frequent transitions occurring among states which do not communicate in the unperturbed case. But things are more subtle in the higher order time scales. In particular, a few of them may exist. The purpose of this paper is to explore further the processes corresponding to the highest time scale. We claim that in this time scale, there is a process among subsets (as opposed to states) which can be analyzed after constructing a suitable transition (stochastic) matrix. This phenomenon was mostly considered in the case when the unperturbed process does not possess transient states (and hence it possesses only recurrent states belonging to various unconnected subsets). This (somewhat intermediate) case of singular perturbation is referred to as *nearly completely decomposable* (NCD) or *nearly uncoupled*. Here we are also interested in the most general case, namely where under the unperturbed case transient states, as well as at least two ergodic classes exist.

A few phenomena exist in this case. First, as observed in [8] and elsewhere, it is possible that some mean passage times are of order larger than ε^{-1} when ε is the order of the perturbation. In this case computing various parameters of the system, in particular, finding Laurent expansions, is usually based on solving system of equations known as the *fundamental equations*. One of the main purposes of this paper is to show that the most singular term of the Laurent expansion for the mean passage time matrix, can be computed by solving successive Markov chains defined on shrinking state spaces. This is on top of the already known fact that the order of the pole (or, equivalently, the degree of singularity at zero) can be determined by a combinatorial (polynomial) algorithm which has as an input the addresses of entries in the unperturbed transition matrix and in the perturbed matrix. Another phenomenon is the fact that the order of the pole of the mean message time matrix and that of the deviation matrix are not identical. In [8] it is shown that the former is greater than or equal to the latter. An example with equality is given there. We give below an example with strict inequality.

The paper is organized as follows. In the next section we review the regular case, namely the case where the unperturbed process is ergodic too. Most of the results stated here will be needed later for our various derivations. They will also be used for comparison purposes. In Section 3 we look at the NCD case. In particular, we derive new expressions for the singular part in the Laurent expansion for the deviation and mean passage time matrices which are based on the aggregated process between subsets. Finally, the general case is considered in Section 4. In particular, the matrix governing transitions at the highest time scale is computed and from it the most singular matrices in the Laurent series expansion for the mean passage time matrix and for the deviation matrix are easily determined.

2. The regular case

This paper deals mainly with singular perturbation of stochastic matrices. However, as a prerequisite to the analysis of this topic, one has first to consider the regular perturbation case. Our following review goes in line with a review made in [2] and which is based mainly on [10] and [15]. Let $P \in R^{n \times n}$ be a stochastic matrix representing transition probabilities in an ergodic Markov chain which for the sake of simplicity of exposition is also assumed to be aperiodic. Let $C \in R^{n \times n}$ be a zero rowsum matrix such that for some $\varepsilon_{\max} > 0$, the matrix $P(\varepsilon) = P + \varepsilon C$ is stochastic for $\varepsilon \in [0, \varepsilon_{\max})$ and the corresponding Markov chain is ergodic. We are interested how various parameters of the Markov chain vary or behave as a function of $\varepsilon \geq 0$ around the origin.

For $\varepsilon \in [0, \varepsilon_{\max})$, let $\pi(\varepsilon)$ be the stationary distribution of $P(\varepsilon)$, i.e., $\pi(\varepsilon)$ is the unique row vector satisfying $\pi(\varepsilon) = \pi(\varepsilon)P(\varepsilon)$, $\pi(\varepsilon)\underline{1} = 1$, where $\underline{1}$ is the vector of ones with the appropriate dimension. In other words, the stationary distribution is the probability vector which coupled with the eigenvalue 1 form a left eigensystem of the corresponding stochastic matrix. As the Markov chain is assumed to be ergodic, the stationary distribution is uniquely defined and all its entries are positive. It is well known that $\pi(\varepsilon)$ is analytic around $\varepsilon = 0$. Namely, for some sequence $\{\pi^{(i)}\}_{i=0}^{\infty}$ and any $\varepsilon \in [0, \varepsilon_{\max})$, $\pi(\varepsilon) = \Sigma_{i=0}^{\infty} \varepsilon^i \pi^{(i)}$, where $\pi^{(0)} = \pi(0)$ and where $\pi^{(i)}$, $i \geq 1$, are zerosum vectors. Furthermore, $\{\pi^{(i)}\}_{i=0}^{\infty}$ is a geometric series, i.e., for some matrix $U \in R^{n \times n}$,

$$\pi^{(i)} = \pi^{(0)} U^i, \quad 0 \leq i < \infty. \tag{1}$$

The exact definition of U is given below in (4). †

For $\varepsilon \in [0, \varepsilon_{\max})$, let $Y(\varepsilon) \in R^{n \times n}$ be the deviation matrix associated with the Markov chain represented by $P(\varepsilon)$. The matrix $Y(\varepsilon)$ is defined by

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

where $P^*(\varepsilon) = \lim_{t \to \infty} P^t(\varepsilon)$.‡ The deviation matrix is the unique matrix satisfying

$$Y(\varepsilon)(I - P(\varepsilon)) = (I - P(\varepsilon))Y(\varepsilon) = I - P^*(\varepsilon), \tag{2}$$

with

$$Y(\varepsilon)\underline{1} = \underline{0}, \quad \pi(\varepsilon) = \pi(\varepsilon)Y(\varepsilon). \tag{3}$$

Hence, $Y(\varepsilon) = [I - P(\varepsilon)]^{\#}$, that is, $Y(\varepsilon)$ is the group inverse of $I - P(\varepsilon)$ [14]. Identity (in ε) (2) leads to systems of equations (each system due to a particular coefficient of ε to some order) which can de utilized in order to compute these coefficients. Finally, the deviation matrix received its name from the fact that

$$Y(\varepsilon) = \lim_{T \to \infty} \sum_{t=0}^{T} (P^t(\varepsilon) - P^*(\varepsilon)),$$

i.e., $Y_{ij}(\varepsilon)$ is the difference over an infinitely long horizon, between the expected number of visits in state j when the process initializes in state i and when it initializes in accordance with the stationary distribution.

The deviation matrix Y of the unperturbed matrix P plays a key role in the series expansion for $\pi(\varepsilon)$. Specifically, the quotient matrix U defining the geomeric sequence $\{\pi^{(i)}\}_{i=0}^{\infty}$ is given by

$$U = CY. \tag{4}$$

For $\varepsilon \in [0, \varepsilon_{\max})$, let $M_{ij}(\varepsilon)$ be the mean passage time from state i into state j ($i = j$ included). In the ergodic case the corresponding random variable is well defined and possesses a finite mean for all $\varepsilon \in [0, \varepsilon_{max})$. The matrix function $M(\varepsilon)$ is continuous and analytic around

† In fact, the radius of convergence is $\min\{\varepsilon_{\max}, \rho^{-1}(U)\}$, where $\rho(U)$ is the spectral radius of the matrix U. Yet, for simplicity we denote this radius of convergence by $\varepsilon_{\max}$ too.

‡ It is clear that in the ergodic case $P^*(\varepsilon)$ is the matrix all its rows equals $\pi(\varepsilon)$. Yet, we prefer to define $P^*(\varepsilon)$ as the limit of $P^t(\varepsilon)$ since this definition holds also when the assumption of ergodicity is removed.

zero, i.e., for some series of matrices $\{M^{(i)}\}_{i=0}^{\infty}$ and any $\varepsilon \in [0, \varepsilon_{\max})$, $M(\varepsilon) = \Sigma_{i=0}^{\infty} \varepsilon^i M^{(i)}$ where $M^{(0)} = M(0)$. Note that for any state i, $M_{ii}(\varepsilon) = [\pi_i(\varepsilon)]^{-1}$. The matrices $M(\varepsilon)$ and $Y(\varepsilon)$ are related via

$$M_{ij}(\varepsilon) = \frac{\delta_{ij} + Y_{jj}(\varepsilon) - Y_{ij}(\varepsilon)}{\pi_j(\varepsilon)} \tag{5}$$

where δ is the Kronecker delta. Finally, in the regular case, the matrix functions $Y(\varepsilon)$ and $M(\varepsilon)$ admit Maclaurin series expansions. Namely, for $\varepsilon \geq 0$ small enough

$$Y(\varepsilon) = \sum_{i=0}^{\infty} Y^{(i)} \varepsilon^i \tag{6}$$

for a sequence of matrices $\{Y^{(i)}\}_{i=0}^{\infty}$ with $Y^{(0)} = Y(0)$. A corresponding expansion with $M^{(0)} = M(0)$ exists for $M(\varepsilon)$.

3. The nearly completely decomposable case

Let $P \in R^{n \times n}$ 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 $\Omega = \{\Omega_1, \ldots, \Omega_p\}$ each of which being an ergodic class. We assume that the order of the rows and columns of P is compatible with Ω, that is, for stochastic matrices, P_I, $I = 1, \ldots, p$, we have

$$P = \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & P_p \end{pmatrix}$$

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. Note that ergodicity is assumed here regardless how small ε is, as long as it is slightly positive. In fact, we face here an example of singular perturbation as a small changes in the transition probabilities leads to structural changes. In particular, states that did not communicate under $P = P(0)$, do so under $P(\varepsilon)$ for any ε, $0 < \varepsilon < \varepsilon_{\max}$.

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. Of course, $P(0) = P$. We next review some results which hold in NCD markov chains. Our review repeats the review given in Section 4.2.3. of [2]. Then, in Theorem 2 we state our new results for this model.

The first thing to observe under the NCD case is that the stationary distribution of the unperturbed Markov chain is not uniquely defined. Yet, the limit matrix is well defined (when all classes are aperiodic). However, when $0 < \varepsilon < \varepsilon_{\max}$, the stationary distribution is well-defined. Thus, for $\varepsilon \in (0, \varepsilon_{\max})$, let $\pi(\varepsilon)$ be the unique stationary distribution of $P(\varepsilon)$. Define $\pi(0)$ by $\lim_{\varepsilon \to 0} \pi(\varepsilon)$. Then, $\pi(0)$ is *a* stationary distribution of P (but, of course, not the only one). In particular, $\pi_i(0) > 0$ for all i, $1 \leq i \leq n$. It is well-known that $\pi(\varepsilon)$ is analytic around

$\varepsilon = 0$. In particular, for some sequence $\{\pi^{(i)}\}_{i=0}^{\infty}$ and any $\varepsilon \in (0, \varepsilon_{\max})$, $\pi(\varepsilon) = \Sigma_{i=0}^{\infty}\varepsilon^i\pi^{(i)}$, where $\pi^{(i)}$, $i \geq 1$, are zerosum vectors. Moreover, $\{\pi^{(i)}\}_{i=0}^{\infty}$ is a geometric sequence, i.e., for some matrix $U \in R^{n\times n}$, $\pi^{(i)} = \pi^{(0)}U^i$, $0 \leq i < \infty$. See (8) below for an explicit expression for U.

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

$$k_I = \Sigma_{i\in I}\pi_i^{(0)} \tag{7}$$

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

Next we define the matrix $Q \in R^{p\times p}$ which is usually referred to as the *aggregate* matrix as it represents transition probabilities between subsets. Each row, and likewise each column, in Q corresponds to a subset in Ω. Then for the subsets I and J, with $I \neq J$, let

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

and let

$$Q_{II} = 1 - \sum_{J\neq I} Q_{IJ} \quad .$$

Without loss of generality, assume that Q_{II} is non-negative for all subsets I. Note that the matrix C can be divided by any constant and ε can be multiplied by this constant leading to the same $n \times n$ transition matrices $P(\varepsilon)$. Taking this constant small enough guarantees the stochasticity of Q and hence the above assumption is without loss of generality. In particular, the stationary distribution of Q is invariant with respect to the choice of this constant. Moreover, Q is irreducible and the vector $k \in R^p$ (see (7)) is its unique stationary distribution.

As already mentioned above the aggregate stochastic matrix Q represents transition probabilities between subsets. Yet, in spite of the fact that the original process among states is Markovian, this is not necessarily the case with respect to the process among subsets (and typically it is not). However, as will be discussed shortly, much can be learned on the original process from the analysis of the aggregate matrix.

For $\varepsilon \in [0, \varepsilon_{\max})$, let $Y(\varepsilon)$ be the deviation matrix of $P(\varepsilon)$. This matrix is uniquely defined and note that the case $\varepsilon = 0$ is not excluded. Indeed the requirements (2) and (3) apply here too. However, as we see shortly, in the NCD case there is no continuity of $Y(\varepsilon)$ at $\varepsilon = 0$. In particular, $Y(0)$ has the same structure as P has, namely

$$Y(0) = \begin{pmatrix} Y_1 & 0 & \cdots & 0 \\ 0 & Y_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Y_p \end{pmatrix},$$

where Y_I is the deviation matrix of P_I, $1 \leq I \leq p$. In particular, $Y(0)$ is finite.

Now let $V \in R^{p\times n}$ be such that its I-th row is full of zeros excepts for the entries corresponding to subset I where γ_I appears. Also, let $W \in R^{n\times p}$ be such that its J-th column, $1 \leq J \leq p$, is full of zeros except for the entries corresponding to subset J which is full of ones. Note that $VW \in R^{p\times p}$ is the identity matrix. Hence, V and W correspond to biorthogonal sets of eigenvectors of $P(0)$ belonging to the eigenvalue 1, V as left eigenvectors and W as right eigenvectors. In addition, it is easy to see that $Q = I + VCW$. Finally, denote by $D \in R^{p\times p}$ the deviation matrix of Q.

Now we can give an expression for the matrix U, the quotient matrix of the geometric sequence $\{\pi^{(i)}\}_{i=0}^{\infty}$:

$$U = CY(0)(I + CWDV). \tag{8}$$

An equivalent expression for U appears in [16] (see Equation 3-1 there). Yet, we find the presentation in (8) more appealing as it shows explicitly the role played in the perturbed process by the aggregate matrix (and hence by the process among subsets) via its deviation matrix D. We omit the tedious proof for the equivalence between the two forms for matrix U.

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

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

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

$$Y_{ij}^{(-1)} = D_{IJ}(\gamma_J)_j \ . \tag{10}$$

Note from (10) that for all pairs i and j, $\lim_{\varepsilon \to 0} Y_{ij}(\varepsilon) \to \infty$. In particular, $\lim_{\varepsilon \to 0} Y(\varepsilon) \neq Y(0)$. Theorem 1 appeared first in [16]. Yet, there (see Equation 3-10a) $Y^{(-1)}$ has a different expression. We find that expression (10) explicitly shows the role played in the orginal process by the aggregate matrix (and hence by the process among subsets). Formula (10) appeared first in [7] (see Equation 3 there).

For $0 < \varepsilon < \varepsilon_{\max}$ and for any pair of states i and j, let $M_{ij}(\varepsilon)$ be the mean passage time from state i into state j when transitions are governed by the stochastic matrix $P(\varepsilon)$. Note that as opposed to $Y(0)$, $M(0)$ is not well-defined. This is the case as under $P(0)$ states belonging to different subsets do not communicate.

Let $E \in R^{p \times p}$ be the mean passage time matrix associated with the aggregated process. That is, for any pair of subsets I and J ($I = J$ included), E_{IJ} is the mean passage time from subset I into subset J when transition probabilities are governed by the stochastic matrix Q.

Theorem 2. (*the NCD case*) *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 $M^{(-1)} \neq 0$,*

$$M(\varepsilon) = \frac{1}{\varepsilon}M^{(-1)} + M^{(0)} + \varepsilon M^{(1)} + \varepsilon^2 M^{(2)} + \cdots \tag{11}$$

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

$$M_{ij}^{(-1)} = \begin{cases} 0 & if\ J = I \\ E_{IJ} & if J \neq I \end{cases} \tag{12}$$

and, for the case $J = I$, we have

$$M_{II}^{(0)} = \frac{1}{k_I}M_I(0) + Y_I \sum_{K \neq I} C_{IK} M_{KI}^{(-1)} - \mathcal{J}[Y_I \sum_{K \neq I} C_{IK} M_{KI}^{(-1)}]_{dg}, \tag{13}$$

where $\mathcal{J}$, the matrix full of ones of appropriate dimension and where for any square matrix A, A_{dg} is the diagonal matrix whose diagonal and the diagonal of A, coincide. In particular,

$$[M_{II}(0)]_{ii} = \frac{[M_I(0)]_{ii}}{k_I}. \tag{14}$$

Proof. By (5) and by the fact that the Markov chain is ergodic when $0<\varepsilon<\varepsilon_{\max}$,

$$M_{ij}(\varepsilon)=\frac{\delta_{ij}+Y_{jj}(\varepsilon)-Y_{ij}(\varepsilon)}{\pi_j(\varepsilon)},\quad 0<\varepsilon<\varepsilon_{\max}. \tag{15}$$

Hence, by (9) and (15) the Laurent series (11) exists and

$$M_{ij}^{(-1)}=\frac{Y_{jj}^{(-1)}-Y_{ij}^{(-1)}}{\pi_j^{(0)}}\quad . \tag{16}$$

By (10), $Y_{jj}^{(-1)}=Y_{ij}^{(-1)}$ whenever states i and j are in the same subset, hence $M_{ij}^{(-1)}=0$ in this case. Using (10) again for the case where $J\neq I$, we conclude that the numerator of (16) equals $(D_{JJ}-D_{IJ})(\gamma_J)_j$. By (7) and the definition of γ_J, the denomerator equals $k_J(\gamma_J)_j$. Thus, the right hand side of (16) is equal to $(D_{JJ}-D_{JI})/k_J$. Applying (5) but now to the matrix Q,we conclude that

$$M_{ij}^{(-1)}=\frac{D_{JJ}-D_{JI}}{k_J}=E_{IJ}$$

whenever $i\in I$, $j\in J$ and $J\neq I$.

Next, we derive the expression (13). From [10] we know that $M(\varepsilon)$ satisfies the following identity

$$(I-P(\varepsilon))M(\varepsilon)=-P(\varepsilon)M_{dg}(\varepsilon)+\mathcal{J}.$$

Collecting terms free of ε and taking into account the block structure, we get that

$$(I-P_I)M_{II}^{(0)}=-P_IM_{IIdg}^{(0)}+\mathcal{J}+\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}, \tag{17}$$

for $I=1,\dots,p$. Since $M_{ii}(\varepsilon)=[\pi_i(\varepsilon)]^{-1}$ and $M_{ii}^{(-1)}=0$, we know that $M_{ii}^{(0)}=[k_I(\gamma_I)_i]^{-1}$. Then, the general solution of (17) is given by

$$M_{II}^{(0)}=-Y_IP_IM_{IIdg}^{(0)}+Y_I\mathcal{J}+Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}+\mathcal{J}A,$$

where $A=\operatorname{diag}\{\alpha_1,\dots,\alpha_{n_I}\}$ is a diagonal matrix with some constants $\alpha_1,\dots,\alpha_{n_I}$, to be determined later. Using the properties (2) and (3), we get the following equivalent expression

$$M_{II}^{(0)}=(I-P_I^*-Y_I)M_{IIdg}^{(0)}+Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}+\mathcal{J}A.$$

Looking only at the diagonal elements in the above expression, one determines immediately α_i. Namely,

$$\alpha_i=\frac{1}{k_I}+\frac{Y_{Iii}}{k_I\gamma_{Ii}}-[Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}]_{ii}.$$

And hence, we conclude that

$$\begin{aligned}M_{II}^{(0)}&=(I-Y_I+\mathcal{J}Y_{Idg})M_{IIdg}^{(0)}+Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}-\mathcal{J}[Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}]_{dg}\\&=(I-Y_I+\mathcal{J}Y_{Idg})\frac{1}{k_I}M_{Idg}(0)+Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}-\mathcal{J}[Y_I\sum_{K\neq I}C_{IK}M_{KI}^{(-1)}]_{dg}\end{aligned}$$

$$= \frac{1}{k_I} M_I(0) + Y_I \sum_{K \neq I} C_{IK} M_{KI}^{(-1)} - \mathcal{J}[Y_I \sum_{K \neq I} C_{IK} M_{KI}^{(-1)}]_{dg}$$

as required.

□

Recall that $M_{ij}^{(-1)} = 0$ whenever i and j belong to the same subset. Actually, this is the only case when zero values can be encountered. Put differently, for $i \in I$ and $j \in J$, $M_{ij}^{(-1)} > 0$ if and only if $I \neq J$.

We note that Theorem 2 (see, in particular, (13)) implies that even for the case where states i and j belong to the same subset, which is the only case when the mean passage time is well defined also for $\varepsilon = 0$, $M_{ij}^{(0)}$ does not equal $M_{ij}(0)$. Indeed, even within a subset, the effect of other subsets does not vanish or even diminish when ε goes to zero (at least not as mean passage times are concerned). It is interesting that the effect of the complementary subset of I on the diagonal entries in $M^{(0)}$ is only through k_I (see (14)) while the effect on the off-diagonal entries is more involved (see (13)) and in fact each of the raw data is required in order to compute these values.

The observation that $M_{ij}^{(-1)} = 0$ whenever states i and j belong to the same subset was first made in [11]. This paper also contains alternative representation for the non-diagonal blocks of $M^{(-1)}$ and for the diagonal blocks of $M^{(0)}$. We like to note that formula (12) appeared without a proof in [2] pointing to this paper.

4. The general case: absorbing states communicate via transient states

First we recall that $Y(\varepsilon)$ and $M(\varepsilon)$ always possess Laurent series expansions around zero. This is the case since these functions can be obtained as solutions to linear systems. Furthermore, elements of $Y(\varepsilon)$ and of $M(\varepsilon)$ can be expressed as ratios between finite polynomials.

The next important issue is the order of the poles of $Y(\varepsilon)$ and $M(\varepsilon)$ at $\varepsilon = 0$. Assuming the perturbed process to be irreducible, as we are doing throughout the paper, maybe the first question to address here is if some of Section 3's results still hold in the general case. For example, are the orders of the poles of $Y(\varepsilon)$ and of $M(\varepsilon)$ always smaller than or equal to one? Or, do these orders always coincide?

So far we have assumed that no transient states (under P) exist and this was a sufficient condition for the order of the poles at zero to coincide and to be smaller than or equal to one. Thus, the existence of transient states is a necessary condition for higher order singularity. Yet, as examples show, this is not a sufficient condition and some additional structure (on top of having transient states) is needed in order to encounter higher order of singularities. Indeed, as was shown for instance in [8] and [6], if a recurrent (under P) state j can be reached from another recurrent (under P) state i, where i and j belog to different subsets, and when $P(\varepsilon)$ for $0 < \varepsilon < \varepsilon_{\max}$ governs transition probabilities, only through a path which contains transient (under P) states, higher order poles can be encountered. The following perturbed transition matrix, which was analyzed in [9], [8] and [2], is a good example for such cases.

Example 1

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

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 in particular implies that the expected time it takes to reach state 3 for a process which starts in state 1 is of the order of magnitude of $O(\varepsilon^{-2})$. In other words, the order of the pole of $M_{13}(\varepsilon)$ at zero is two.

□

In the general case a stochastic matrix has the following form

$$P = \begin{pmatrix} Q_0 & Q_1 & Q_2 & \cdots & Q_p \\ 0 & P_1 & 0 & \cdots & 0 \\ 0 & 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & P_p \end{pmatrix},$$

where Ω_0 is the (possibly empty) set of transient states (i.e., $\lim_{t\to\infty} Q_0^t = 0$) and where the rest of the states are as before with $p \geq 1$. In this manuscript we limit ourselves to linear perturbations, that is, $P(\varepsilon) = P(0) + \varepsilon C$ for some matrix C. Yet, for the reduction process defined below, we need to consider analytic perturbations (of lower dimensions) which are not necessarily linear. Thus, although seems redundant at this stage we assume that

$$P(\varepsilon) - I = \sum_{k=0}^{\infty} \varepsilon^k G_k \tag{18}$$

for some matrix sequence $\{G_k\}_{k=0}^{\infty}$. Of course, $G_0 = P(0) - I$, $G_1 = C$ and $G_k = 0$ for $k \geq 2$ in the present setting. Define by $G(\varepsilon)$ the righthand side of (18).

It was shown in [3, 9] that the deviation matrix $Y(\varepsilon)$ of the perturbed Markov chain possesses a Laurent series expansion

$$Y(\varepsilon) = \frac{1}{\varepsilon^s}(Y_0 + \varepsilon Y_1 + \cdots), \tag{19}$$

for some integer $s \geq 0$ and with $Y_0 \neq 0$.[§] The value of s can be determined by the algorithm suggested in [8]. Specifically, it is an $O(n^4)$ algorithm which receives as input the addresses of the positive elements of $P(0)$ and of C and whose output (among other parameters) is s.

[§]Note that due to the fact that now s can vary (as opposed to the regular case where $s = 0$ and the NCD case where $s = 1$), we have adopted different notation here. Specifically, Y_0 is the leading coefficient and subscripts, rather than superscripts, denote the order of the coefficients. The superscript index is preserved for later purposes.

Note that in order to apply the numerical procedure outlined below it is required to compute s first.

As was noticed in Section 2, the deviation matrix $Y(\varepsilon)$ is the unique solution of the system

$$Y(\varepsilon)G(\varepsilon) = P^*(\varepsilon) - I, \tag{20}$$

$$Y(\varepsilon)\underline{1} = 0.$$

It is well known, see for instance [17], that $P^*(\varepsilon)$ can be expanded as a power series

$$P^*(\varepsilon) = P_0^* + \varepsilon P_1^* + \varepsilon^2 P_2^* + \cdots$$

with $P_0^* \neq P^*(0)$ in the singular perturbation case. Upon the substitution of the above series for $P^*(\varepsilon)$, (18) and (19) into (20) and then collecting the terms with the same power of ε, we get the following system of equations for Y_i, $i \geq 0$, known as the *fundamental equations*,

$$\begin{array}{rcll} Y_0G_0 & = & 0 & \quad (F0) \\ Y_1G_0 + Y_0G_1 & = & 0 & \quad (F1) \\ \vdots & \vdots & \vdots & \\ Y_sG_0 + Y_{s-1}G_1 + \cdots + Y_1G_{s-1} + Y_0G_s & = & P_0^* - I & \quad (Fs) \end{array}$$

Note that above we present only $s+1$ fundamental equations although there are in fact infinitely many such equations. The reason behind that is a result found in [9] which says that these $s+1$ sets of equations are all that is needed such that they coupled with the requirement that $Y(\varepsilon)1 = 0$, lead to a unique solution for Y_0 (but not for the other coefficients). See also [3].

The system of $s+1$ fundamental equations can be solved by means of a reduction process [1, 3, 5, 9] as outlined next. Extending the definition of the matrices V and W given in Section 3, let $V \in R^{p\times n}$ be such that its I-th row is full of zeros excepts for γ_I at the entries corresponding to subset Ω_I¶, and let $W \in R^{n\times p}$ be such that W_{iJ} is equal to the probability (under the unperturbed transition matrix) that a process which initializes in state i is eventually absorbed into the ergodic subset Ω_J. Of course, if i is recurrent and $i \in \Omega_J$ then $W_{iJ} = 1$. If i is recurrent and $i \notin \Omega_J$ then $W_{iJ} = 0$. Finally, if i is transient, $W_{iJ} = [(I - Q_0)^{-1}Q_J\underline{1}]_i$. Let

$$G_k^{(1)} = V \sum_{p=1}^{k+1} \sum_{\nu_1+\cdots+\nu_p=k+1} G_{\nu_1}Y(0)G_{\nu_2}\cdots Y(0)G_{\nu_p}W.$$

Note that in the case where $G_k = 0$ for $k \geq 2$, we have $G_k^{(1)} = VG_1(Y(0)G_1)^kW$. As was shown in [1, 3], the system (F) is equivalent to the following reduced system ‖ with variables

¶This is exactly the same definition for V as given in the NCD case.

‖The superscript (1) corresponds to the fact that the first reduction step is done and there will be more steps to come.

$Y_i^{(1)}$

$$
\begin{array}{rcll}
Y_0^{(1)}G_0^{(1)} &=& 0 & (RF0)\\
Y_1^{(1)}G_0^{(1)}+Y_0^{(1)}G_1^{(1)} &=& 0 & (RF1)\\
\vdots & \vdots & \vdots & \\
Y_{s-1}^{(1)}G_0^{(1)}+Y_{s-2}^{(1)}G_1^{(1)}+\cdots+Y_1^{(1)}G_{s-2}^{(1)}+Y_0^{(1)}G_{s-1}^{(1)} &=& (P_0^*-I)Q & (RFs-1)
\end{array}
$$

Note that $Y_k^{(1)} \in R^{p\times p}$, $k \geq 0$, and hence the terminology of *reduced system*. The matrix $Y_0^{(1)}$ is uniquely determined by the above equations and the normalization condition $Y_0^{(1)}\underline{1} = 0$. Once $Y_0^{(1)}$ is obtained, Y_0 can be calculated by

$$Y_0 = Y_0^{(1)}V.$$

Note that the system (RF) has s matrix equations in comparison with $s+1$ matrix equations in (F), and hence this is another reason behind the terminology of *reduced system*. In particular, the dimension of aggregated matrices $G_k^{(1)}$ is equal to the number of ergodic sets in the unperturbed Markov chain. As in the NCD case, we refer to $G_0^{(1)} = I+VP(0)W$ as a generator of the aggregated Markov chain.

We can apply the reduction technique again but now to the system (RF). After the second reduction step the number of matrix equations is reduced to $s-1$. Similarly, one can perform s reduction steps. Specifically, define in a recursive manner

$$G_k^{(j)} = V^{(j-1)}\sum_{p=1}^{k+1}\sum_{\nu_1+\cdots+\nu_p=k+1} G_{\nu_1}^{(j-1)}H^{(j-1)}G_{\nu_2}^{(j-1)}\cdots H^{(j-1)}G_{\nu_p}^{(j-1)}W^{(j-1)}, \quad 1\leq j\leq s,$$

where $H^{(j-1)}$ is the deviation matrix corresponding to the generator $G_0^{(j-1)}$. As $G_0^{(j)}$ is a Markov chain generator (see e.g., [5]), let the matrices $V^{(j)}$ and $W^{(j)}$ be defined similarly to the matrices V and W for the original Markov chain. By convention, let $V^{(0)} = V$ and $W^{(0)} = W$. Note that by the nature of the final reduction step, $V^{(s)}$ is a row vector while $W^{(s)}$ is a column vector, the latter being full of ones. Then, the j-th step reduces the fundamental system into the form

$$
\begin{array}{rcll}
Y_0^{(j)}G_0^{(j)} &=& 0 & (RjF0)\\
Y_1^{(j)}G_0^{(j)}+Y_0^{(j)}G_1^{(j)} &=& 0 & (RjF1)\\
\vdots & \vdots & \vdots & \\
Y_{s-1}^{(j)}G_0^{(j)}+Y_{s-2}^{(j)}G_1^{(j)}+\cdots+Y_1^{(j)}G_{s-2}^{(j)}+Y_0^{(j)}G_{s-1}^{(j)} &=& (P_0^*-I)WW^{(1)}\cdots W^{(j-1)} & \\
 & & & (RjFs-1).
\end{array}
$$

The limiting stationary distribution π_0 can be given by the following formula [1, 5]

$$\pi^{(0)} = V^{(s)}V^{(s-1)}\cdots V^{(1)}V. \tag{21}$$

To specify the above formula for each element $\pi_i^{(0)}$, $1\leq i\leq n$, we introduce the integer valued function $I^{(k)}(i)$, $k = 0,\ldots,s-1$. Specifically, let $I^{(k)}(i)$ be the index of the ergodic set in the

$k-th$ reduction step to which state i belongs. Then, formula (21) can be rewritten in the component form

$$\pi_i^{(0)} = V^{(s)}_{I^{(s-1)}(i)} V^{(s-1)}_{I^{(s-1)}(i),I^{(s-2)}(i)} \cdots V^{(1)}_{I^{(1)}(i),I^{(0)}(i)} V_{I^{(0)}(i),i} \tag{22}$$

From (22) one can learn that if state i is transient at some level of the aggregation, the corresponding element $\pi_i^{(0)}$ is equal to zero.

Before stating the next theorem, we next introduce a useful notion of the degree of transience.

Definition 1. *For a states i, denote its* degree of transience *by $t(i)$ and define it by*

$$t(i) = \min_{m \geq 0} \{\pi_i^{(m)} > 0\}.$$

Note that $>$ above can be replaced with $\neq$.

Since $\pi_i(\varepsilon) = 1/M_{ii}(\varepsilon)$, it is clear that $t(i)$ is equal to the order of the pole of $M_{ii}(\varepsilon)$ at zero. In particular, its value can be determined by the algorithm presented in [8]. Furthermore, there always exists at least one state i such that $t(i) = 0$, otherwise the elements of $\pi^{(0)}$ would not sum up to one. This issue was also discussed in [19], where the degree of transience was used in order to refine the solution concept of Nash equilibrium suitable for non-cooperative games.

Theorem 3. *The most singular coefficient of the Laurent series for the deviation matrix of the perturbed Markov chain is given by*

$$Y_0 = WW^{(1)} \cdots W^{(s-1)} H^{(s)} V^{(s-1)} \cdots V^{(1)} V, \tag{23}$$

where $H^{(s)} = [-G_0^{(s)}]^{\#}$ is the deviation matrix for the s-th level aggregated Markov chain. Furthermore, let state i belong to some ergodic set of the $(s-1)$-th level aggregated process and let state j have zero degree of transience, i.e., $t(j) = 0$. Then, the most singular coefficient of the Laurent series for $M_{ij}(\varepsilon)$ is given by

$$M_{ij}^{(-s)} = \begin{cases} \dfrac{H^{(s)}_{I^{(s-1)}(j),I^{(s-1)}(j)} - H^{(s)}_{I^{(s-1)}(i),I^{(s-1)}(j)}}{V^{(s)}_{I^{(s-1)}(j)}}, & if \quad I^{(s-1)}(i) \neq I^{(s-1)}(j) \\ 0, & if \quad I^{(s-1)}(i) = I^{(s-1)}(j). \end{cases} \tag{24}$$

Proof. After s reduction steps we obtain the following equation (RsF0)

$$Y_0^{(s)} G_0^{(s)} = (P_0^* - I) W \cdots W^{(s-1)}$$

Since $P_0^* = W \cdots W^{(s-1)} \underline{1} V^{(s)} V^{(s-1)} \cdots V$ [5] and $V^{(k)} W^{(k)} = I$, the righthand side of the above equation can be transformed as follows:

$$(P_0^* - I) W \cdots W^{(s-1)} = (W \cdots W^{(s-1)} \underline{1} V^{(s)} V^{(s-1)} \cdots V - I) W \cdots W^{(s-1)} =$$
$$W \cdots W^{(s-1)} \underline{1} V^{(s)} - W \cdots W^{(s-1)} = W \cdots W^{(s-1)} (\underline{1} V^{(s)} - I)$$

Next we recall that if $G_0^{(s)}$ has a simple zero eigenvalue, the equation

$$Y_0^{(s)} G_0^{(s)} = W \cdots W^{(s-1)} (\underline{1} V^{(s)} - I)$$

coupled with the normalization condition $Y_0^{(s)}\underline{1} = 0$, gives a unique solution for $Y_0^{(s)}$. Hence, applying the group generalized inverse, we get

$$Y_0^{(s)} = W \cdots W^{(s-1)}(\underline{1}V^{(s)} - I)(G_0^{(s)})^{\#} = W \cdots W^{(s-1)}(-G_0^{(s)})^{\#} = W \cdots W^{(s-1)}H^{(s)}$$

Finally, we have

$$Y_0 = Y_0^{(1)}V = Y_0^{(s)}V^{(s-1)} \cdots V = W \cdots W^{(s-1)}H^{(s)}V^{(s-1)} \cdots V,$$

which is expression (23). Using (5) and (22), we obtain (24).

□

Note that an immediate corollary from the above theorem is as follows.

Corollary 1. *Let t_{ij} be the order of the pole of $M_{ij}(\varepsilon)$ at zero. Let j be such that $t_{jj} = 0$ (or, equivalently $t(j) = 0$). Then,*

$$s = \max_{1 \le i \le n} t_{ij}.$$

Example 2. Consider the following perturbed transition matrix

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

As before, we denote by t_{ij} the order of the pole of $M_{ij}(\varepsilon)$ at $\varepsilon = 0$. By the algorithm suggested in [8] we get that (t_{ij}) is given by

$$(t_{ij}) = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 3 & 0 & 1 & 2 \\ 3 & 0 & 1 & 2 \\ 3 & 0 & 0 & 2 \end{pmatrix}.$$

Hence, the order of singularity of $M(\varepsilon)$, given by $\max_{ij} t_{ij}$, is three. The order of singularity of the deviation matrix $Y(\varepsilon)$, denote above by s and given by $\max_{ij}\{t_{ij} - t_{jj}\}$ [8] is then equal to one. In particular, denoting by t the order of singularity of $M(\varepsilon)$ at zero, we have derived an example where $s < t$. Also, $t(j) = t_{jj}$, the degree of transience, can be read from the diagonal of the matrix (t_{ij}). Alternatively, one can apply Corollary 1 to determine that $s = 1$.

Next let us apply Theorem 3 to this example. Here we have

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

and

$$W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Hence,

$$G_0^{(1)} = VCW = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \quad H^{(1)} = [-G_0^{(1)}]^{\#} = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}.$$

As zero is a simple eigenvalue of $G_0^{(1)}$, only one reduction step is required here. This is of course an alternative way to notice that $s = 1$. Next, we calculate $\pi^{(0)}$ and Y_0.

$$\pi^{(0)} = V^{(1)}V = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$$

$$Y_0 = WH^{(1)}V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Inspecting the entries of $\pi^{(0)}$ we see that all transient states i in the unperturbed process are, as always, with $\pi_i^{(0)} = 0$. A phenomenon we observe here is that state 1, although being recurrent in the unperturbed system, is also with $\pi_1^{(0)} = 0$. This is of course a-priori possible (yet, not all recurrent states can have simultaneously this property). In particular, the recurrent state, state 1, as opposed to state 2, possesses some degree of transience in the perturbed Markov chain.

Furthermore, the degrees of transience (see Definition 1) for the states in this example are $t(1) = 2$, $t(2) = 0$, $t(3) = 1$ and $t(4) = 2$. Applying formula (24) of Theorem 3, we get

$$M_{12}(\varepsilon) = \frac{H_{22}^{(1)} - H_{12}^{(1)}}{V_2^{(1)}}\varepsilon^{-1} + \mathrm{o}(\varepsilon^{-1}) = \frac{0-(-1)}{1}\varepsilon^{-1} + \mathrm{o}(\varepsilon^{-1}) = \varepsilon^{-1} + \mathrm{o}(\varepsilon^{-1}).$$

Note that if a fifth state was added such that this state would have related to the fourth as currently the fourth is related to the third, $t(5)$ would be equal to 3 but still the value of s will be preserved at $s = 1$. Also, the values of $t(2)$, $t(3)$ and $t(4)$ would stay unchanged. Finally, in the modified example we would have $t(1) = t(5) = 3$.

□

5. Concluding remarks

Algorithms for computing the order of the poles of the mean passage times and of the deviation matrices, and for computing the coefficients of the corresponding Laurent series are known. The purpose of this paper was to suggest alternative methods so as to shed some more light on the transition probabilities in the highest time scale process, mainly transition between ergodic subsets via transient states. This was done completely for the NCD case and partially for the general case. It particular, for the latter case we dealt only with transition into recurrent classes who receive limit probabilities which are away from the zero under the perturbed process. The question of what are the corresponding coefficients in the case of states with non-zero degree of transience is still open.

Another open issue is finding the order of the individual poles in the Laurent expansion for the deviation matrix while using combinatorial techniques. Specifically, let s_{ij} be the order of the pole $Y_{ij}(\varepsilon)$ at zero. In [8] one can find a combinatorial algorithm (which gets as input the addresses of the positive entries of P and of C) for computing $s = \max_{ij} s_{ij}$. However, we are not aware of a combinatorial algorithm for determining the individual s_{ij} themselves.

REFERENCES

1. Avrachenkov, K.E. (1999), "Analytic Perturbation Theory and its Applications", PhD Thesis, available at `http://www-sop.inria.fr/mistral/personnel/K.Avrachenkov/`.
2. Avrachenkov, K.E., J.A. Filar and M. Haviv (2002), "Singular perturbations of Markov chains and decision processes. A Survey", in *Markov Decision Processes: Models, Methods, Directions and Open Problems*, pp. 113–152, edited by E. Feinberg and A. Schwartz, Kluwer Academic Publisher.
3. Avrachenkov, K.E. and J.B. Lasserre (1999), "The fundamental matrix of singularly perturbed Markov chains," *Advances in Applied Probability*, v.31, pp.679–697.
4. P.J. Courtois, *Decomposability: queueing and computer system applications*, Academic Press, New York, 1977.
5. Delebecque, F., (1981), "A reduction process for perturbed Markov chains", *SIAM J. Appl. Math.*, v.48, pp.325–350.
6. Filar, J.A. and D. Krass (1994), "Hamiltonian cycles and Markov chains," *Mathematics of Operations Research*, Vol. 19, pp. 223–237.
7. Haviv, M. (1989), "More on the Rayliegh-Ritz refinement technique for nearly uncoupled matrices," *SIAM Journal of Matrix Analysis and Application*, Vol. 10, pp. 287–293.
8. Hassin, R. and Haviv, M. (1992), "Mean passage times and nearly uncoupled Markov chains," *SAIM Journal of Discret Mathematics*, Vol. 5, pp. 386–397.
9. Haviv, M. and Ritov, Y. (1993), "On series expansions and stochastic matrices", *SIAM Journal on Matrix Analysis and its Applications.*, V.14, pp.670–676.
10. Kemeny, J.G. and J.L. Snell (1960), *Finite Markov Chains,* Von Nostrand, New York.
11. Latouch, G. and G. Louchard (1978), "Return times in nearl-completely decomposable stochastic processes," *Journal of Applied Probability*, Vol. 15, pp. 251–267.
12. Latouch, G. and G. Louchard (1982), "Random times in nearly-completely decomposable, transient Markov chains," *Cahiers du C.E.R.O.*, Vol. 24, pp. 321–352.
13. McAllister, D.F., Stewart, G.W., and Stewart, W.J., "On a Rayleigh-Ritz refinement technique for nearly uncoupled stochastic matrices", *Lin. Alg. Appl.*, Vol. 60, pp. 1–25, 1984.
14. Meyer, C.D. (1975), "The role of the group generalized inverse in the theory of finite Markov chains", *SIAM Review*, v.17, pp.443–464.
15. Schweitzer, P.J. (1968), "Perturbation theory and finite Markov chains," Vol. 5, pp. 401–413.
16. Schweitzer, P.J. (1981), "Perturbation series expansion of nearly completely-decomposable Markov chains," Working Paper Series No. 8122, The Graduate School of Management, The University of Rochester.
17. Schweitzer, P.J. (1986), "Perturbation series expansion of nearly completely-decomposable Markov chains," in *Teletraffic Analysis and Computer Performance Evaluation,* O.J. Boxma, J.W. Cohen and H.C. Tijms (eds.), Elsevier Science Publisher B.V. (North Holland), (1986), pp.319–328.
18. H.A. Simon and A. Ando, "Aggregation of variables in dynamic systems", *Econometrica*, v.29, pp.111–138, 1961.
19. Young, H.P. (1993), "The evolution of conventions," *Econometrics*, Vol. 61, pp. 57–84.