Stochastics, 1983, Vol. 8, pp. 259–289
0090-2234/83/0804-0259 $18.50/0

Hierarchical Aggregation of Singularly Perturbed Finite State Markov Processes

M. CODERCH, A. S. WILLSKY, S. S. SASTRY AND D. A. CASTANON†

Laboratory for Information and Decisions Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

(*Accepted for publication July 30, 1982*)

In this paper we study the asymptotic behavior of Finite State Markov Processes with rare transitions. We show how to construct a sequence of increasingly simplified models of a singularly perturbed FSMP and how to combine these aggregated models to produce an asymptotic approximation of the original process uniformly valid over $[0, \infty)$.

1. INTRODUCTION AND MOTIVATING EXAMPLE

1.1. Introduction

In this paper we study the asymptotic behavior of continuous time Finite State Markov Processes (FSMP's) with rare transitions. Let $\eta^{\varepsilon}(t)$ be a FSMP with transition probability matrix $\mathscr{P}^{\varepsilon}(t)=\exp\{A_0(\varepsilon)t\}$, where

$$A_0(\varepsilon)=\sum_{p=0}^{\infty}\varepsilon^p A_{0p} \tag{1.1}$$

is the matrix of transition rates, and $\varepsilon\in[0,\varepsilon_0]$ is a small parameter modelling rare transitions in $\eta^{\varepsilon}(t)$. We establish that, if the perturbation in (1.1) is singular (in the sense that the number of ergodic classes of $\eta^{\varepsilon}(t)$ changes at $\varepsilon=0$), then

†Research supported in part by the DOE under grant ET-76-C-01-2295 and in part by AFOSR under grant AFOSR 82-0258. The first author also acknowledges thankfully the support of the Fundación ITP, Madrid, Spain. Dr. D. A. Castanon is presently at Alphatech, Inc., 3 New England Executive Park, Burlington, MA 01803.

i) the limits

$$\lim_{\varepsilon \downarrow 0} \mathscr{P}^{\varepsilon}(t/\varepsilon^k) \triangleq \mathscr{P}_k(t) \qquad k=1,2,\ldots,m$$

are well defined and determine a finite sequence of (in general stochastically discontinuous) FSMP's $\eta_k(t)$, $k=1,2,\ldots,m$, with transition probability matrix $\mathscr{P}_k(t)$;

ii) the limit processes $\eta_k(t)$ can be aggregated to produce a hierarchy of simplified, approximate models of $\eta^{\varepsilon}(t)$ each of which is a FSMP valid at a certain time scale t/ε^k describing changes in $\eta^{\varepsilon}(t)$ at a distinct level of detail; and

iii) the collection of aggregated models $\hat{\eta}_k(t)$, $k=1,2,\ldots,m$, can then be combined to construct an asymptotic approximation to $\eta^{\varepsilon}(t)$ uniformly valid for $t \geqq 0$.

The idea of using aggregated models to describe gross features of the evolution of Markov processes with rare transitions (i.e., with time scale separation) has been explored by several authors (see [1]–[10]). With the exception of [7]–[10], on which we will comment later, the works referred to above deal with the nearly-decomposable case, i.e., the matrix $A(\varepsilon)$ is assumed to be decomposed into $A(\varepsilon)=A_0+\varepsilon B$ with A_0 block-diagonal. For this simple case it was proven in [2] that the rare transitions among weakly interacting groups of states (which take place for times of order t/ε) can be modeled by a Markov process with one state for each block in A_0. In [7] the authors allow the chain A_0 to have transient states (i.e. A_0 not block-diagonal) but because only the time scale t/ε was considered, the presence of such states did not modify the basic result in [2]. In [8] the case where transitions between weakly interacting groups do not take place until times of order t/ε^k for some $k \geqq 0$ was also included and the same ideas were shown to be applicable to more general Markov processes but, as in previous work, only one aggregated model of a given process was considered. The notion of a hierarchy of aggregated models of a process, each associated with a certain time scale t/ε^k, was discussed in [9] and subsequently in [10]. This latter paper, however, differs in substantial ways from the results and approach taken here. In particular, in [10] the focus is essentially entirely in the frequency domain, i.e. on the resolvent of $A(\varepsilon)$. While a set of aggregated models is developed in this way (using a set of recursive calculations somewhat different than ours), reference [10] does not make a rigorous connection between the aggregated models and the construction of an asymptotic approximation to the original process which is valid on the semi-infinite interval $[0, \infty)$.

If T is the matrix of transition rates of a FSMP then the rows of P_0 are the different ergodic probability vectors of the process, and

$$-T^{\#}=(T+P_0)^{-1}-P_0=\int_0^{\infty}(\exp\{Tt\}-P_0)\,dt$$

is referred to as the *potential* matrix.

2.2 Matrix perturbation theory

Suppose now that $T(\varepsilon)$, $\varepsilon\in[0,\varepsilon_0]$ is a matrix valued function with an absolutely convergent series of the form:

$$T(\varepsilon)=\sum_{p=0}^{\infty}\varepsilon^p T_p. \tag{2.1}$$

An important problem is the nature of the ε-dependence of the eigenvalues, eigenprojections and eigennilpotents of $T(\varepsilon)$ as $\varepsilon\downarrow 0$. For a detailed account the reader is referred to [12], here we briefly state several results that we use later on.

The number of distinct eigenvalues of $T(\varepsilon)$ is constant except at some isolated values of ε. Without loss of generality let $\varepsilon=0$ be the only exceptional point in $[0,\varepsilon_0]$. The eigenvalues of $T(\varepsilon)$ are continuous functions of ε and at $\varepsilon=0$ several of them may collapse into a single eigenvalue of $T(0)$. Suppose that rank $T(\varepsilon)>\operatorname{rank} T(0)$ and let Γ_0 be a positively oriented contour enclosing zero but no other eigenvalue of $T(0)$. For ε small enough, all eigenvalues of $T(\varepsilon)$ that collapse into the origin as $\varepsilon\to 0$ (referred to as the zero-group of eigenvalues), are inside Γ_0. The matrix

$$P_0(\varepsilon)=-\frac{1}{2\pi i}\int_{\Gamma_0}R(\xi,T(\varepsilon))\,d\xi$$

is therefore equal to the sum of the eigenprojections for eigenvalues of the zero group and it is called the *total projection for the zero group* of $T(\varepsilon)$. In the subsequent sections we will need the following result.

PROPOSITION 2.2 *Let $T(\varepsilon)$ be as in (2.1) and assume that $T_0=T(0)$ has SSNS then,*

$$\frac{T(\varepsilon)P_0(\varepsilon)}{\varepsilon}=\frac{P_0(\varepsilon)T(\varepsilon)}{\varepsilon}=\sum_{n=0}^{\infty}\varepsilon^n\tilde{T}_n \tag{2.2}$$

where

$$\tilde{T}_n = -\sum_{p=1}^{n+1} \sum_{\substack{v_1+\ldots+v_p=n+1 \\ k_1+\ldots+k_{p+1}=p-1 \\ v_i \geq 1,\, k_j \geq 0}} S^{(k_1)} T_{v_1} S^{(k_2)} \ldots S^{(k_p)} T_{v_p} S^{(k_{p-1})} \tag{2.3}$$

where $S^{(0)} = -P_0(0) \triangleq -P_0$ *and* $S^{(k)} = (T_0^{\#})^k$.

Proof See [13].

2.3 Asymptotic approximation of $\exp\{A_0(\varepsilon)t\}$

Let $A_0(\varepsilon)$ be an $n \times n$ matrix having an (absolutely convergent) expansion of the form:

$$A_0(\varepsilon) = \sum_{p=0}^{\infty} \varepsilon^p A_{0p}. \tag{2.4}$$

Construct a sequence of matrices $A_k(\varepsilon)$, $k = 0, 1, \ldots, m$, as follows. Let $P_0(\varepsilon)$ denote the total projection for the zero group of eigenvalues of $A_0(\varepsilon)$ and define:

$$A_1(\varepsilon) \triangleq \frac{P_0(\varepsilon)A_0(\varepsilon)}{\varepsilon} = \frac{A_0(\varepsilon)P_0(\varepsilon)}{\varepsilon} = \frac{P_0(\varepsilon)A_0(\varepsilon)P_0(\varepsilon)}{\varepsilon}.$$

If A_{00} has SSNS then $A_1(\varepsilon)$ has a series expansion of the form

$$A_1(\varepsilon) = \sum_{p=0}^{\infty} \varepsilon^p A_{1p}.$$

If A_{10} also has SSNS then

$$A_2(\varepsilon) \triangleq \frac{P_1(\varepsilon)A_1(\varepsilon)}{\varepsilon} = \frac{P_1(\varepsilon)P_0(\varepsilon)A_0(\varepsilon)}{\varepsilon}$$

also has a series expansion

$$A_2(\varepsilon) = \sum_{p=0}^{\infty} \varepsilon^p A_{2p}$$

where $P_1(\varepsilon)$ is the total projection for the zero group of eigenvalues of $A_1(\varepsilon)$. This recursion can continue if A_{20} also has SSNS. The sequence

$$\frac{1}{h}\int_t^{t+h}\mathscr{P}(\tau)\,d\tau-\frac{1}{h}\int_0^h\mathscr{P}(\tau)\,d\tau=\frac{1}{h}(\mathscr{P}(h)-\Pi)\int_0^t\mathscr{P}(\tau)\,d\tau.$$

As $h\downarrow 0$ the left-hand side converges to $\mathscr{P}(t)-\Pi$ and therefore

$$\lim_{h\downarrow 0}\frac{\mathscr{P}(h)-\Pi}{h}\triangleq A \tag{3.13}$$

exists. Taking limits as $h\downarrow 0$ in (3.12) we get

$$\mathscr{P}(t)=\Pi+A\int_0^t\mathscr{P}(\tau)\,d\tau \qquad \forall\, t>0$$

establishing (3.8). Definition (3.13) together with (3.6) and (3.7) give (3.10), and (3.11) follows immediately from (3.8) and the fact that $\Pi\cdot\mathbf{1}=\mathbf{1}$. The positivity of $\mathscr{P}(t)$, i.e.,

$$\frac{1}{h}\mathscr{P}(h)=\frac{1}{h}\Pi\exp\{At\}=\frac{1}{h}\Pi+A+\frac{o(h)}{h}\geqq 0 \quad \text{for} \quad h\geqq 0$$

implies that for h small enough $A+\Pi/h\geqq 0$ establishing (3.12). To prove the converse suppose now that Π and A satisfy (3.9)–(3.12). Then, $\mathscr{P}(t)=\Pi\exp\{At\}$ clearly satisfies (3.3) and (3.4) and the positivity condition follows from (3.12) as indicated below:

$$\begin{aligned}\Pi\exp\{At\}&=\Pi\, e^{-ct}\exp\{(A+cI)t\}\\&=e^{-ct}\Pi\sum_{n=0}^{\infty}\frac{(A\Pi+c\Pi)^n}{n!}t^n\geqq 0.\end{aligned}$$

We shall refer to the projection $\Pi=\lim_{t\downarrow 0}\mathscr{P}(t)$ as the *ergodic projection at zero* and to the matrix

$$A=\lim_{h\downarrow 0}\frac{\mathscr{P}(h)-\Pi}{h} \tag{3.14}$$

as the *infinitesimal generator of* $\mathscr{P}(t)$.

Remarks 1) It follows from (3.9) that Π is the matrix of ergodic probabilities of a Markov chain and as such it determines a partition of E

in terms of ergodic classes, E_i^0, $i=1,\ldots,s$, and transient states, E_T^0,

$$E=\left(\bigcup_{i=1}^{s} E_i^0\right)\cup E_T^0$$

that we will refer to as the *ergodic partition at zero*. As we will see later, this partition corresponds to a classification of states into different types. While the process is in absorbing state (i.e. in an ergodic class E_i^0 with a single element), the process behaves as a stochastically continuous FSMP. Instantaneous transitions occur between states belonging to the same ergodic class at zero, and transient states are visited only during transitions between ergodic classes, with no time spent in them.

2) For stochastically continuous processes $\Pi=I$ and conditions (i)–(iv) only require that the rows of A add up to zero and that all its off-diagonal entries be non-negative. In the general case some off-diagonal entries of A can be negative provided the corresponding entry in Π is non-zero (see Example 3.2 below). The usual interpretation of a_{ij} as the rate of transitions from state i to j is thus no longer valid in the stochastically discontinuous case. To interpret these entries it is first necessary to perform an aggregation as discussed in Section 3.3.

Example 3.2 The following is a stochastically discontinuous transition probability matrix:

$$\mathscr{P}(t)=\begin{bmatrix} p_1 e^{-\lambda t} & p_2 e^{-\lambda t} & 1-e^{-\lambda t} \\ p_1 e^{-\lambda t} & p_2 e^{-\lambda t} & 1-e^{-\lambda t} \\ 0 & 0 & 1 \end{bmatrix}, \quad p_1+p_2=1, \quad t>0$$

with initial projection and infinitesimal generator given by:

$$\Pi=\begin{bmatrix} p_1 & p_2 & 0 \\ p_1 & p_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \qquad A=\begin{bmatrix} -p_1\lambda & -p_2\lambda & \lambda \\ -p_1\lambda & -p_2\lambda & \lambda \\ 0 & 0 & 0 \end{bmatrix}.$$

For $p_1=p_2=\lambda=1/2$ this is the stochastically discontinuous limit process $\eta_1(t)$ described in Section 1.2.

3.2 Implications of stochastic discontinuity

If we consider a separable version of a stochastically continuous FSMP then its sample functions are easily visualized as piecewise continuous functions taking values in E [19]. The evolution of the process can be thought of as a succession of stays in different states of E, each being of random duration and exponentially distributed. The sequence of states visited follows a Markov chain law with one-step transition probabilities determined by the entries of the generator A. On the contrary, the sample functions of a stochastically discontinuous process are much more irregular. As we will now prove, these processes have instantaneous states, i.e., states in which the process spends no time with probability one. Furthermore, in general, a stochastically discontinuous process spends a non-zero amount of time switching among instantaneous states. The sample functions are therefore nowhere continuous on certain time intervals.

Consider a separable version of a FSMP $\eta(t)$ with initial projection Π and generator A, and let Λ be a separating set. For $t>0$ and $n=0,1,\ldots$, take

$$0=t_{0n}<t_{1n}<\ldots<t_{nn}=t$$

in such a way that the sets

$$\Lambda_n=\{t_{0n},t_{1n},\ldots,t_{nn}\}$$

increase monotonically and $\cup\Lambda_n=\Lambda\cap[0,t]$. Then we have:

$$\begin{aligned}
&\Pr\{\eta(\tau)=i,\forall\tau\in[0,t]\,|\,\eta(0)=i\}\\
&=\Pr\{\eta(\tau)=i,\forall\tau\in[0,t]\cap\Lambda\,|\,\eta(0)=i\}\\
&=\lim_{n\to\infty}\Pr\{\eta(\tau)=i,\forall\tau\in[0,t]\cap\Lambda_n\,|\,\eta(0)=i\}\\
&=\lim_{n\to\infty}\prod_{k=0}^{n-1}p_{ii}(t_{k+1,n}-t_{k,n})\\
&=\exp\left\{\lim_{n\to\infty}\sum_{k=0}^{n-1}\log p_{ii}(t_{k+1,n}-t_{k,n})\right\}
\end{aligned}\tag{3.15}$$

where $p_{ii}(t)$ are the diagonal elements of $\mathscr{P}(t)=\Pi\exp\{At\}$. Equation (3.15) facilitates a classification of the states of $\eta(t)$ according to the diagonal entries, π_{ii}, of Π. If $\pi_{ii}=0$ then $p_{ii}(h)\to0$ as $h\to0$ and therefore (3.15) gives:

$$\Pr\{\eta(\tau)=i,\forall\tau\in[0,t]\,|\,\eta(0)=i\}=0\qquad\forall t>0.$$

If, on the other hand, $0<\pi_{ii}\leq 1$ use (3.14) to write:

$$\frac{p_{ii}(h)}{\pi_{ii}}=1+\frac{a_{ii}}{\pi_{ii}}h+o(h)$$

or

$$\log p_{ii}(h)=\log \pi_{ii}+\frac{a_{ii}}{\pi_{ii}}h+o(h)$$

and it follows from (3.15) that

$$\Pr\{\eta(\tau)=i, \forall\, \tau\in[0,t]\,|\,\eta(0)=i\}=\begin{cases}0 & \text{if} \quad \pi_{ii}<1 \\ \exp\{a_{ii}t\} & \text{if} \quad \pi_{ii}=1\end{cases}. \tag{3.16}$$

DEFINITION 3.3 A state i will be called *instantaneous* if $\pi_{ii}<1$ and regular if $\pi_{ii}=1$. An instantaneous state j will be called *evanescent* if $\pi_{jj}=0$.

Remarks:

1) Notice that this classification is based on the ergodic partition at zero.
2) We have just seen that the sojourn time in instantaneous states is zero w.p.l.
3) Also, the sojourn time in regular states is exponentially distributed. All states of a stochastically continuous process are regular.
4) In Example 3.2, states $\{1,2\}$ are instantaneous, non-evanescent states while 3 is regular.
5) Even though the duration of stays in a given instantaneous state is zero w.p.l., there is, in general, a non-zero probability of finding the process in an instantaneous state at any given time (as in states $\{1,2\}$ of Example 3.2).
6) The probability of finding the process in an evanescent state at any given time is zero. This follows from the fact that $\pi_{ii}=0$ implies $\pi_{ji}=0$, $j=1,\ldots,n$ (i.e. evanescent states are transient states of the chain Π) and because:

$$\mathscr{P}(t)=\Pi\exp\{At\}=\Pi\exp\{At\}\Pi \qquad \forall\, t>0 \tag{3.17}$$

we have

$$p_{ji}(t)=0, \quad \forall\, t>0, \quad j=1,\ldots,n. \tag{3.18}$$

The evanescent states can thus be neglected in the sense that there exists a version of the process $\eta(t)$ with the same finite dimensional distributions which does not take values in the set of evanescent states.

Proof Follows from (3.30) and the fact that (3.33) always exists for stochastically continuous processes [19]. □

In the sequel we shall refer to Π' as the *ergodic projection at* ∞.

For future reference it is important to notice that because

$$\mathscr{P}(t) = \Pi \exp\{At\} - \exp\{At\} - I + \Pi$$

Eq. (3.33) implies that generators of FSMP's are semistable matrices.

4. SINGULARLY PERTURBED FINITE STATE MARKOV PROCESSES

4.1 Regular and singular perturbations

Consider now a stochastically continuous FSMP $\eta^\varepsilon(t)$ that takes values in $E_0 = \{e_1, \ldots, e_{n_0}\}$ with infinitesimal generator of the form:

$$A_0(\varepsilon) = \sum_{p=0}^{\infty} \varepsilon^p A_{0p} \qquad \varepsilon \in [0, \varepsilon_0]. \tag{4.1}$$

The small parameter ε models rare transitions in $\eta^\varepsilon(t)$ and we shall refer to $\eta^\varepsilon(t)$ for $\varepsilon > 0$ as a perturbed version of the process $\eta^0(t)$. Let $\mathscr{P}^\varepsilon(t)$ and $\mathscr{P}^0(t)$ denote the transition probability matrices of $\eta^\varepsilon(t)$ and $\eta^0(t)$ respectively. Our objective is to analyze the behavior of $\eta^\varepsilon(t)$ (or equivalently, that of $\mathscr{P}^\varepsilon(t)$ as $\varepsilon \downarrow 0$ on the time interval $[0, \infty)$.

First, it is straightforward to verify that on any interval of the form $[0, T]$, $\eta^\varepsilon(t)$ can be approximated by $\eta^0(t)$. Precisely,

$$\lim_{\varepsilon \downarrow 0} \sup_{0 \leq t \leq T} \|\mathscr{P}^\varepsilon(t) - \mathscr{P}^0(t)\| = 0 \qquad \forall\, T < \infty \tag{4.2}$$

i.e., the finite dimensional distributions of $\eta^\varepsilon(t)$ converge to those of $\eta^0(t)$ uniformly on $[0, T]$. However, as the example in Section 1.2 illustrates, the behavior of $\eta^\varepsilon(t)$ on the infinite time interval $[0, \infty)$ may differ markedly from that of $\eta^0(t)$. We shall say that $\eta^\varepsilon(t)$ is *regularly perturbed* if

$$\lim_{\varepsilon \downarrow 0} \sup_{t \geq 0} \|\mathscr{P}^\varepsilon(t) - \mathscr{P}^0(t)\| = 0 \tag{4.3}$$

otherwise, we will say that the perturbation is *singular*. In what follows we focus on the singularly perturbed case, since failure of (4.3) is symptomatic of the existence of distinct behavior at different time scales.

DEFINITION 4.1 We will say that $\eta^\varepsilon(t)$ has *well defined behavior at time scale* t/ε^k, $k>0$, if there exists a continuous, time-dependent matrix $Y_k(t)$ such that for any $\delta>0$, $T<\infty$,

$$\lim_{\varepsilon\downarrow 0}\ \sup_{\delta\leq t\leq T} \|\mathscr{P}^\varepsilon(t/\varepsilon^k)-Y_k(t)\|=0. \tag{4.4}$$

Remarks:

1) It is readily verified that the limit matrix $Y_k(t)$ in (4.4) must be the transition probability matrix of some FSMP $\eta_k(t)$ taking values in E_0. Thus (4.4) is equivalent to saying that $\eta^\varepsilon(t/\varepsilon^k)$ converges to some FSMP $\eta_k(t)$ as $\varepsilon\downarrow 0$ in the sense of finite dimensional distributions.
2) As we will see in Section 4.3, $\lim_{\varepsilon\downarrow 0}\mathscr{P}^\varepsilon(t/\alpha(\varepsilon))$ exists for any order function $\alpha(\varepsilon)$ ($\alpha:[0,\varepsilon_0]\to\mathbb{R}^+$, $\alpha(0)=0$ and $\alpha(\cdot)$ continuous and monotone increasing). It turns out, however, that only the limits $Y_k(t)$ for a finite number of positive integers $k=0,1,\ldots,m$ are required to construct an asymptotic approximation to $\mathscr{P}^\varepsilon(t)$ uniformly valid for $t\geq 0$. We shall call t, $t/\varepsilon,\ldots,t/\varepsilon^m$ *the fundamental or natural time scales of the process* $\eta^\varepsilon(t)$.
3) Regularly perturbed processes have trivial time scale behavior. For any order function $\alpha(\varepsilon)$

$$\lim_{\varepsilon\downarrow 0}\sup_{t>0}\|\mathscr{P}^\varepsilon(t/\alpha(\varepsilon))-\Pi_0^\infty\|=0 \tag{4.5}$$

where Π_0^∞ is the ergodic projection at ∞ of the unperturbed process $\eta^0(t)$.

PROPOSITION 4.2 *The process $\eta^\varepsilon(t)$ is singularly perturbed if and only if the number of ergodic classes at ∞ of the perturbed process $\eta^\varepsilon(t)$ is different from that of $\eta^0(t)$ or, equivalently, if* rank $A_0(\varepsilon)\neq$ rank A_{00} *for* $\varepsilon>0$.

Proof See [13] for a proof of the proposition in terms of the rank condition. The statement in terms of the number of ergodic classes at ∞ follows from the fact that this number equals nul $A_0(\varepsilon)$. □

4.2 Multiple time scale behavior and aggregation

In Section 2.3 we indicated that if a matrix $A_0(\varepsilon)$ satisfies the MSST property then $\exp\{A_0(\varepsilon)t\}$ has an asymptotic approximation that clearly displays its multiple time scale behavior [Eqs. (2.9)–(2.11)]. We now prove that generators of FSMP's always satisfy the MSST condition and we the number of states of the successive aggregated models and they can be implemented in a recursive fashion. We illustrate this procedure with an example in the next section.

2) In [10] Delebecque gives a recursive algorithm to compute the aggregated models $\hat{A}_k$ by constructing an array analogous to Table I but using a somewhat simplified version of formula (2.3) (essentially eliminating terms that are cancelled at subsequent stages of the recursion).

5. AN EXAMPLE

Consider the process $\eta^{\varepsilon}(t)$ in Figure 5. A quick look at the unperturbed version in Figure 6 will convince the reader of the singular nature of the perturbation. The ergodic projection at ∞ of $\eta^0(t)$ determines four ergodic classes $E_1=\{1,2\}$,† $E_2=\{3\}$, $E_3=\{4,5\}$ and $E_4=\{7\}$ and a transient state $E_T=\{6\}$. The aggregated model $\hat{\eta}_1(t)$ valid at time scale t/ε is portrayed in Figure 7 and it has the following infinitesimal generator:

$$\hat{A}_1=U_1BV_1=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/2 & -1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Notice that the aggregation operation in addition to collapsing $\{1,2\}$ and $\{4,5\}$ into two states also prunes the evanescent state $\{6\}$. At the next

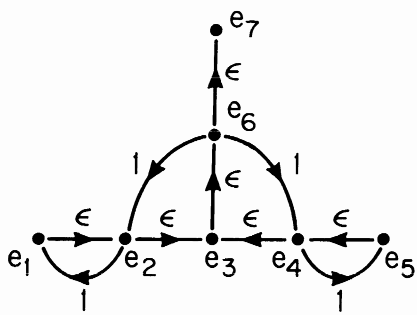

FIGURE 5 The perturbed process $\eta^{\varepsilon}(t)$.

†Notice that if a transient state communicates with only one ergodic class, as states 2 and 4 do in this example, it can be included in that ergodic class for aggregation purposes.

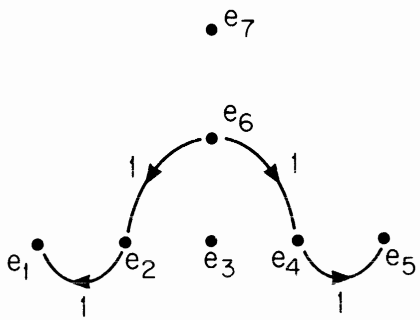

FIGURE 6 The unperturbed process $\eta^0(t)$.

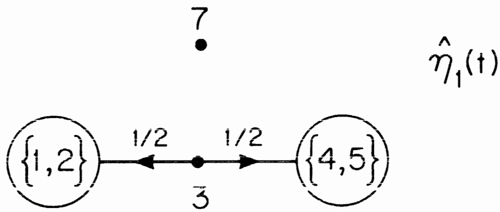

FIGURE 7 Aggregated model valid at time scale t/ε.

stage we get

$$\hat{P}_1 = \lim_{t\to\infty} e^{\hat{A}_1 t} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which leads to the following aggregation partition:

$$E'_1 = \{1, 2\}, \quad E'_2 = \{4, 5\}, \quad E'_3 = \{7\} \quad \text{and} \quad E'_T = \{3, 6\}.$$

The corresponding aggregated model valid at time scale t/ε^2, $\hat{\eta}_2(t)$ has generator:

$$\hat{A}_2 = -U_2 B A_0^{\#} B V_2 = \begin{bmatrix} -1/2 & 1/2 & 0 \\ 1/2 & -1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and it is represented in Figure 8. Finally,

$$\hat{P}_2 = \lim_{t\to\infty} e^{\hat{A}_2 t} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

leads to the next aggregation partition: $E_1'' = \{1, 2, 3, 4, 5, 6\}$, $E_2'' = \{7\}$. The aggregated model valid at t/ε^3 has rates

$$\hat{A}_3 = -U_3 B A_0^{\#} B V_1 \hat{A}_1^{\#} U_1 B A_0^{\#} B V_3 = \begin{bmatrix} -1/2 & 1/2 \\ 0 & 0 \end{bmatrix}$$

and it is portrayed in Figure 9. The hierarchy of models ends here because

$$\operatorname{rank} A_0 + \operatorname{rank} \hat{A}_1 + \operatorname{rank} \hat{A}_2 + \operatorname{rank} \hat{A}_3 = \operatorname{rank} A_0(\varepsilon) = 6 \qquad \varepsilon > 0.$$

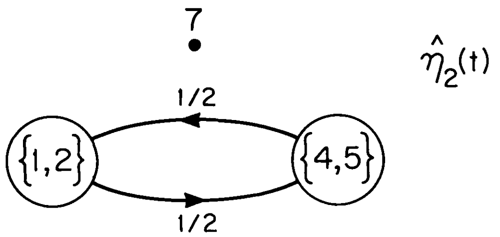

FIGURE 8 Aggregated model valid at time scale t/ε^2.

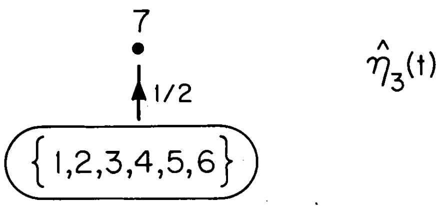

FIGURE 9 Aggregated model valid at time scale t/ε^3.

This example illustrates how a comparatively complex singularly perturbed FSMP can be asymptotically approximated by a collection of very simple FSMP's.

6. CONCLUSIONS

We have presented a methodology for isolating different events in a singularly perturbed FSMP according to their level of rareness. This methodology leads to a hierarchy of reduced-order models for such processes, each describing the evolution of the process with a different degree of detail and being adequate at a certain time scale. The complete (finite) collection of models obtained in this way can then be combined to produce an approximation valid on the infinite time interval $[0, \infty)$. We refer the reader to [21] for the more general case of singularly perturbed linear dynamical systems and for some filtering applications based on the hierarchical description of FSMP's.

Acknowledgements

We acknowledge gratefully many helpful discussions with Profs. G. Verghese and J. C. Willems.

References

[1] H. A. Simon and A. Ando, Aggregation of variables in dynamic systems, *Econometrica* **29** (1961), 111–138.
[2] V. S. Korolyuk, L. I. Polishchuk and A. A. Tomusyak, A limit theorem for semi-Markov processes, *Kibernetika* **5**, No. 4 (1969), 144–145.
[3] V. G. Gaitsgori and A. A. Pervozvanskii, Aggregation of states in a Markov chain with weak interactions, *Kibernetica* **3** (1975), 91–98.
[4] P. V. Kokotovic, Subsystems, time scales and multimodeling, Invited paper, 2nd IFAC Symposium on Large-Scale Systems: Theory and Applications, Toulouse, France (1980).
[5] D. Teneketzis, H. Javid and B. Shridhar, Control of weakly-coupled Markov chains, Proceedings CDC Conference, Albuquerque, New Mexico (1980).
[6] P. J. Curtois, *Decomposability: Queuing and Computer Applications*, Academic Press, New York, 1977.
[7] F. Delebecque and J. P. Quadrat, Optimal control of Markov chains admitting strong and weak interactions, *Automatica* **17** (1981), 281–296.
[8] V. S. Korolyuk and A. F. Turbin, *Mathematical Foundations of Phase Consolidation for Complex Systems*, Naukova Dumka, Kiev, 1978 (in Russian).
[9] D. A. Castanon, M. Coderch, B. C. Levy and A. S. Willsky, Asymptotic analysis, approximation and aggregation methods for stochastic hybrid systems, Proceedings 1980 JACC, San Diego, California.
[10] F. Delebecque, A reduction process for perturbed Markov chains, submitted to *SIAM J. Appl. Math.*
[11] J. Keilson, *Markov Chain Models—Rarity and Exponentiality*, Springer-Verlag, New York, 1978, pp. 132–133.
[12] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1966.
[13] M. Coderch, A. S. Willsky, S. S. Sastry and D. A. Castanon, Hierarchical aggregation of linear systems with multiple time scales, to appear in *IEEE Trans. on Aut. Contr.*
[14] A. A. Dynkin, *Markov Processes*, Springer-Verlag, Berlin, 1965.
[15] W. Doeblin, Sur l'équation matricielle $A(t+s)=A(t)A(s)$ et ses applications aux probabilités en chaine, *Bull. Sci. Math.* **62** (1938), 21–32.

[16] J. L. Doob, Topics in the theory of Markoff chains, *Trans. Am. Math. Soc.* **52** (1942), 37–64.

[[17] A. V. Skorokhod, Constructive methods of specifying stochastic processes, *Russ. Math. Surreys* **20** (1965), 70.

[18] D. Williams, *Diffusions, Markov Processes and Martingales*, Vol. I: *Foundations*, Wiley, New York, 1979.

[19] J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953, Chapter VI.

[20] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups*, Am. Math. Soc., Providence, 1957.

[21] M. Coderch, Multiple time scale approach to hierarchical aggregation of linear systems and finite state Markov processes, Ph.D. Thesis, MIT-LIDS, August 1982.