Systems & Control Letters 11 (1988) 309–314
North-Holland

Multiple time scale decomposition of discrete time Markov chains *

J.R. ROHLICEK

BBN Laboratories Incorporated, 10 Moulton St., Cambridge, MA 02238, U.S.A.

A.S. WILLSKY

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

Received 18 January 1988

Abstract: The multiple time scale decomposition of discrete time, finite state Markov chains is addressed. In [1, 2], the behavior of a continuous time Markov chain is approximated using a fast time scale, ε-independent, continuous time process, and a reduced order perturbed process. The procedure can then be iterated to obtain a complete multiple time scale decomposition. In the discrete time case presented in this paper, the basic approximation has a 'hybrid' form. In this form, the fast time scale behavior is approximated using an ε-independent, discrete time Markov chain, and the slow behavior is captured by a perturbed, *continuous time* process. Further time scale decomposition then involves the continuous time procedure in [1, 2]. This extension to discrete time chains bridges previous multiple time scale decomposition results, which have dealt exclusively with either continuous time or discrete time processes, and provides a uniform framework for the analysis of both types of systems.

Keywords: Markov process, Discrete time, Aggregation, Perturbation theory, Multiple time scales.

1. Introduction

Consider the state probabilities, $x[t]$, of a discrete time Markov chain which satisfy the difference equation

$$x[t+1]=\Phi^{(0)}(\varepsilon)x[t], \quad t\in \mathrm{N}_0, \tag{1}$$

where $\phi_{ji}^{(0)}(\varepsilon)$ is the one-step transition probability from state i to state j. Assumed that $\Phi^{(0)}(\varepsilon)$ is an analytic function of a small parameter ε. Note that all the entries of $\Phi^{(0)}(\varepsilon)$ are nonnegative and that [1] $\mathbf{1}^{\mathrm{T}}\Phi^{(0)}(\varepsilon)=\mathbf{1}^{\mathrm{T}}$. The solution of this difference equation has the form

$$x[t]=\Phi^{(0)}(\varepsilon)^{t}x[0]. \tag{2}$$

In this paper we construct a multiple time scale decomposition of (2) that is uniformly valid for $t\in[0,\infty)$. The key step is the construction of the first stage of the approximation which has the form

$$\begin{aligned}\Phi^{(0)}(\varepsilon)^{t}=\Phi^{(0)}(0)^{t}&+U^{(0)}\,\mathrm{e}^{\varepsilon A^{(1)}(\varepsilon)t}V^{(0)}\\&-U^{(0)}V^{(0)}+\mathrm{O}(\varepsilon)\end{aligned} \tag{3}$$

where $A^{(1)}(\varepsilon)$ is the generator of a continuous time, aggregated Markov process and $\mathrm{O}(\varepsilon)$ is a function of ε and t which converges to zero uniformly over the interval $t\in[0,\infty)$ as $\varepsilon\downarrow 0$ and $U^{(0)}$ and $V^{(0)}$ are matrices whose probabilistic significance will be made clear. As we will see, this approximation is very similar to the continuous time approximation developed in [1], the sole difference being the form of the first term in (3). The consequence of this observation is that by subsequent recursive application of the procedure in [1] we can obtain a complete time scale decomposition of (2) in which the fast dynamics are captured by a discrete time model generated by $\Phi^{(0)}(0)$ and all subsequent slower dynamics (i.e. with time scale constants of order ε, ε^2, etc.) are captured by continuous time models at the corresponding time scales.

While the use of differential equations to describe the slow behavior of difference equations is not a new idea (see [3] for example), the explicit construction and, more important, the demonstration of the uniform validity of such multiple time scale approximations for Markov chains has not

* This research was conducted under the support of the Air Force Office of Scientific Research under grant AFOSR-0258 and AFOSR-88-0032 and the Army Research Office under grant DAAG-29-84-K005.

[1] Here $\mathbf{1}^{\mathrm{T}}=[1,\ldots,1]$.

been pursued previously. Indeed, although perturbed discrete time chains have been studied by many other authors, (see [4, 5] for example), there has been little connection with approaches to decomposition of continuous time processes. In particular, the notion of considering a scaled time variable has not been stressed.

It is interesting to note that there is one restriction that arises in the discrete time case but not in continuous time. Discrete time Markov chains can have periodic components, while this cannot occur in continuous time. However, for a time scale decomposition of $\Phi^{(0)}(\varepsilon)$ to exist, we must assume that $\Phi^{(0)}(0)$ is aperiodic, i.e. that all of its eigenvalues of unit modulus are at the point $\lambda = 1$. If this were not the case then $\lim_{t\to\infty}\Phi^{(0)}(0)^t$, and thus a multiple time scale decomposition, would not exist. With this restriction our result is completely general.

The remainder of this paper is organized as follows. First, the continuous time result of [1] is summarized and the key discrete time result is stated, as is the resulting complete multiple time scale decomposition. The proof of this result is then provided in Section 3, followed by a simple example in Section 4.

2. The basic result

Let $A^{(0)}(\varepsilon)$ be the generator of a continuous time, n-state Markov chain $\eta^{(0)}(\varepsilon, t)$. Let $E_1,\ldots, E_N$ be the ergodic classes of $A^{(0)}(0)$ and let T denote the transient states. Define the $n\times N$ matrix $U^{(0)}$ of ergodic probabilities where the (i, I) element of this matrix is given by

$$u_{iI}^{(0)} = \lim_{t\to\infty} \Pr\left(\eta^{(0)}(t, 0) = i \mid \eta^{(0)}(0, 0)\in E_I\right), \quad (4)$$

i.e. this is the ergodic probability, based on $A^{(0)}(0)$ of state i given that the process begins in E_I (so for example $u_{iI}^{(0)} = 0$ if $i \notin E_I$). Define also the $N\times n$ 'membership' matrix $V^{(0)}(\varepsilon)$ where

$$v_{Ij}^{(0)}(\varepsilon) = \Pr\Big(\eta^{(0)}(\varepsilon, t^*)\in E_I \mid \eta^{(0)}(\varepsilon, 0) = j,$$
$$t^* = \inf_{t\geq 0}\left(t \mid \eta^{(0)}(\varepsilon, t)\notin T\right)\Big). \quad (5)$$

Note that $V^{(0)}(0)$ is the membership matrix of $A^{(0)}(0)$ (e.g. $v_{Ij}^{(0)}(0) = 1$ if $j\in E_I$, $v_{Ij}^{(0)}(0) = 0$ if $j\in E_K$ for $K\neq I$). Note also that

$$\mathbf{1}^{\mathrm{T}} U^{(0)} = \mathbf{1}^{\mathrm{T}}, \quad (6)$$

$$\mathbf{1}^{\mathrm{T}} V^{(0)} = \mathbf{1}^{\mathrm{T}}, \quad (7)$$

Let $\tilde{V}^{(0)}(\varepsilon)$ be any perturbation of $V^{(0)}(\varepsilon)$ that satisfies

$$\mathbf{1}^{\mathrm{T}} \tilde{V}^{(0)}(\varepsilon) = \mathbf{1}^{\mathrm{T}} \quad (8)$$

and where the leading order term of each element of $\tilde{V}^{(0)}(\varepsilon)$ and $V^{(0)}(\varepsilon)$ are equal:

$$\tilde{v}_{Ij}^{(0)}(\varepsilon) = v_{Ij}^{(0)}(\varepsilon)(1 + \mathrm{O}(\varepsilon)). \quad (9)$$

With this notation we can now state the following which is a direct consequence of the analysis in [1]:

Theorem 1. *Given the generator $A^{(0)}(\varepsilon)$ of a continuous time Markov process, define the reduced-dimension $N\times N$ generator*

$$A^{(1)}(\varepsilon) = \frac{1}{\varepsilon}\tilde{V}^{(0)}(\varepsilon)A^{(0)}(\varepsilon)U^{(0)}. \quad (10)$$

Then

$$\mathrm{e}^{A^{(0)}(\varepsilon)t} = \mathrm{e}^{A^{(0)}(0)t} + \left(U^{(0)}\,\mathrm{e}^{A^{(1)}(\varepsilon)\varepsilon t}V^{(0)} - U^{(0)}V^{(0)}\right) + \mathrm{O}(\varepsilon). \quad (11)$$

By recursive application of this result we can generate a finite set of generators $A^{(0)}(0)$, $A^{(1)}(0),\ldots, A^{(k)}(0)$ of decreasing dimension that together provide a complete uniformly accurate multiple time scale approximation of the original process.

The basic result that is proved in the next section is the following:

Theorem 2. *Let $\Phi^{(0)}(\varepsilon)$ be the one-step transition probability matrix of a discrete time Markov process. Define the continuous time generator*

$$A^{(0)}(\varepsilon) = \Phi^{(0)}(\varepsilon) - I. \quad (12)$$

Then

$$\Phi^{(0)}(\varepsilon)^t = \Phi^{(0)}(0)^t + \left(U^{(0)}\,\mathrm{e}^{A^{(1)}(\varepsilon)\varepsilon t}V^{(0)} - U^{(0)}V^{(0)}\right) + \mathrm{O}(\varepsilon) \quad (13)$$

where $u^{(0)}$, $\tilde{V}^{(0)}(\varepsilon)$ and $A^{(1)}(\varepsilon)$ are derived from $A^{(0)}(\varepsilon)$ as described above.

Again by recursive application of Theorem 1, we obtain the following algorithm.

Algorithm 1.

1. Given $\Phi^{(0)}(\varepsilon)$, compute

$$A^{(0)}(\varepsilon)=\Phi^{(0)}(\varepsilon)-I. \tag{14}$$

Set $k\leftarrow 0$.

2. Given $A^{(k)}(\varepsilon)$ compute $A^{(k)}$ and a suitable $\tilde{V}^{(k)}(\varepsilon)$ as described for $k=0$. From these compute

$$A^{(k+1)}(\varepsilon)=\frac{1}{\varepsilon}\tilde{V}^{(k)}(\varepsilon)A^{(k)}(\varepsilon)U^{(k)}. \tag{15}$$

3. Stop when there are no more time scales (for example if $A^{(0)}(\varepsilon)$ has m ergodic classes for $\varepsilon>0$, stop when $A^{(k)}(0)$ also has m ergodic classes). Otherwise, set $k\leftarrow k+1$. Go to 2.

4. Then

$$\begin{aligned}\Phi^{(0)}(\varepsilon)^t &= \Phi^{(0)^t}+\left(U^{(0)}\,\mathrm{e}^{A^{(1)}\varepsilon t}V^{(0)}-U^{(0)}V^{(0)}\right)\\ &\quad+\left(U^{(0)}U^{(1)}\,\mathrm{e}^{A^{(2)}\varepsilon^2 t}V^{(1)}V^{(0)}\right.\\ &\qquad\left.-U^{(0)}U^{(1)}V^{(1)}V^{(0)}\right)\\ &\quad+\cdots\\ &\quad+\left(U^{(0)}\cdots U^{(k-1)}\,\mathrm{e}^{A^{(k)}\varepsilon^k t}V^{(k-1)}\cdots V^{(0)}\right.\\ &\qquad\left.-U^{(0)}\cdots U^{(k-1)}V^{(k-1)}\cdots V^{(0)}\right)\\ &\quad+\mathrm{O}(\varepsilon)\end{aligned} \tag{16}$$

where $\mathrm{O}(\varepsilon)$ is a function of ε and t which converges uniformly to zero over $t\geq 0$.

Note that in step 1, $A^{(0)}(\varepsilon)$ is a generator of a continuous time Markov process. Though the fast behavior of the continuous time process generated by $A^{(0)}(\varepsilon)$ is very different from that of the discrete time process generated by $\Phi^{(0)}(\varepsilon)$, we will prove that their *slow time scale* behaviors are approximately equal. This fact forms the basis of the argument for using $A^{(1)}(\varepsilon)$ to approximate the slow behavior of the original discrete time process.

3. Proof of Theorem 2

The proof is composed of three distinct subsections. First, the 'fast' and 'slow' components of $\Phi^{(0)}(\varepsilon)^t$ are identified. The following two subsections address approximation of these components separately. The superscript $^{(0)}$ is omitted in the derivation to simplify the notation.

3.1. Separation of 'fast' and 'slow' components

The behavior of $\Phi(\varepsilon)^t$ can be separated into 'fast' and 'slow' components. The slow component is associated with eigenvalues which converge to 1 as $\varepsilon\downarrow 0$ while the fast component is associated with those eigenvalue which converge to points within the unit circle. The approach taken here is based on Kato's perturbation results for linear operators [7] and parallels Coderch's approach to separation of time scales in the continuous time, general linear system case [8].

The generator $\Phi(\varepsilon)$ can be expressed as the spectral sum

$$\Phi(\varepsilon)=\sum_i\lambda_i(\varepsilon)P_i(\varepsilon)+D_i(\varepsilon) \tag{17}$$

where $P_i(\varepsilon)$ is the eigenprojection and $D_i(\varepsilon)$ is the eigennilpotent associated with the eigenvalue $\lambda_i(\varepsilon)$. Note that in general, these projections and nilpotents are not analytic functions of ε even if $\Phi(\varepsilon)$ is.

These eigenprojections and nilpotents have the properties that

$$P_i(\varepsilon)P_j(\varepsilon)=\begin{cases}P_j(\varepsilon) & \text{if } i=j,\\ 0 & \text{if } i\neq j,\end{cases} \tag{18}$$

$$P_i(\varepsilon)D_j(\varepsilon)=\begin{cases}D_j(\varepsilon) & \text{if } i=j,\\ 0 & \text{if } i\neq j.\end{cases} \tag{19}$$

The *total projection of the* 1*-group* can be formed as

$$P(\varepsilon)\equiv\sum_{i:\lambda_i(\varepsilon)\to 1}P_i(\varepsilon). \tag{20}$$

Although as stated above the individual projections and nilpotents are not necessarily analytic functions of ε, Kato [7] shows that any *total projection* of an eigengroup of a perturbed matrix is analytic. An *eigengroup* is a set of eigenvalues which converge to a common point. Therefore since $\Phi(\varepsilon)$ is an analytic function of ε, $P(\varepsilon)$ is analytic at $\varepsilon = 0$.

The generator $\Phi(\varepsilon)$ can therefore be decomposed into the sum of two parts

$$\Phi(\varepsilon) = P(\varepsilon)\Phi(\varepsilon) + Q(\varepsilon)\Phi(\varepsilon) \tag{21}$$

where

$$Q(\varepsilon) \equiv I - P(\varepsilon). \tag{22}$$

Recall that by assumption, $\Phi(0)$ is aperiodic. Therefore, all the eigenvalues on the unit circle are in fact concentrated at $\lambda = 1$. The eigenvalues of $Q(\varepsilon)\Phi(\varepsilon)$ therefore converge to points strictly within the unit circle.

Using this decomposition of $\Phi(\varepsilon)$ and the properties of the eigenprojections stated above, the following decomposition is possible

$$\Phi(\varepsilon)^t = (P(\varepsilon)\Phi(\varepsilon) + Q(\varepsilon)\Phi(\varepsilon))^t \tag{23}$$

$$= P(\varepsilon)(P(\varepsilon)\Phi(\varepsilon))^t + Q(\varepsilon)(Q(\varepsilon)\Phi(\varepsilon))^t. \tag{24}$$

In order to prove that the approximation (3) is valid, the two terms in the sum (24) will be treated separately.

3.2. Approximation of the fast behavior

Using the decomposition (24), the fast behavior, which is determined by $Q(\varepsilon)\Phi(\varepsilon)$, can be easily approximated since the ε-dependence is a regular perturbation of $Q(0)\Phi(0)$. The goal is to show that

$$(Q(\varepsilon)\Phi(\varepsilon))^t - (Q(0)\Phi(0))^t = O(\varepsilon) \tag{25}$$

from which follows that

$$Q(\varepsilon)(Q(\varepsilon)\Phi(\varepsilon))^t = Q(0)(Q(0)\Phi(0))^t + O(\varepsilon) \tag{26}$$

$$= Q(0)\Phi(0)^t + O(\varepsilon) \tag{27}$$

$$= \Phi(0)^t - P(0) + O(\varepsilon) \tag{28}$$

$$= \Phi(0)^t + UV + O(\varepsilon). \tag{29}$$

The last equality follows from basic properties of Markov chains, i.e.

$$P(0) = \lim_{t\to\infty} \Phi(0)^t = UV \tag{30}$$

where U is the ergodic probability matrix and V is the corresponding membership matrix.

The validity of (25) can be argued from the fact that the eigenvalues of $Q(0)\Phi(0)$ are all strictly inside the unit circle. From Kato [7], the Z-transform, given by

$$T(\varepsilon, z) \equiv \left(I - z^{-1}Q(\varepsilon)\Phi(\varepsilon)\right)^{-1} \tag{31}$$

converges uniformly away from the singularities of $T(0, z)$. [2] The difference in (25) can be written as

$$\Delta(\varepsilon, t) \equiv (Q(\varepsilon)\Phi(\varepsilon))^t - (Q(0)\Phi(0))^t \tag{32}$$

$$= \frac{1}{2\pi i}\oint_\Gamma z^{t-1}(T(\varepsilon, z) - T(0, z))\,dz \tag{33}$$

where Γ is a positively oriented contour of length $|\Gamma|$ contained inside the unit circle. Since on the contour $|z^t| \le 1$ for $t \ge 0$,

$$\|\Delta(\varepsilon, t)\| \le \frac{1}{2\pi}|\Gamma| \sup_{z\in\Gamma}\left\|\frac{1}{z}(T(\varepsilon, z) - T(0, z))\right\|$$

$$= O(\varepsilon) \tag{34}$$

and therefore (25)–(29) follow.

3.3. Approximation of the slow behavior

The approximation of the slow behavior determined by $P(\varepsilon)\Phi(\varepsilon)$ is based on its further separation into components that evolve at various time scales. Within each time scale, we employ the matrix equivalent of the scalar approximation

$$(1 + \varepsilon\lambda)^t = e^{\varepsilon\lambda t} + O(\varepsilon) \tag{35}$$

whenever $\mathrm{Re}(\lambda) < 0, \varepsilon \in [0, \varepsilon_0)$, a fact that can easily be verified by series expansion of the terms. Note also that (35) is obviously true for $\lambda = 0$. The result sought in this section is

$$P(\varepsilon)(P(\varepsilon)\Phi(\varepsilon))^t - P(\varepsilon)\,e^{P(\varepsilon)(\Phi(\varepsilon)-I)t} = O(\varepsilon). \tag{36}$$

Before continuing with the general development, we should note that the proof of the validity of the approximation (36) is particularly simple in the special situation when the eigenvalues of $P(\varepsilon)\Phi(\varepsilon)$ are semi-simple over an interval $\varepsilon \in [0, \varepsilon_0)$. Specifically, since the eigenvalues are

[2] Kato states these results in terms of the resolvent $R(\xi, A(\varepsilon)) \equiv (A(\varepsilon) - \xi I)^{-1}$. The Z-transform is more commonly used in the context of Markov chains.

semi-simple, there are no eigennilpotents and therefore

$$P(\varepsilon)\Phi(\varepsilon)=\sum_i \lambda_i(\varepsilon)P_i(\varepsilon), \tag{37}$$

$$(P(\varepsilon)\Phi(\varepsilon))^t=\sum_i \lambda_i(\varepsilon)^t P_i(\varepsilon), \tag{38}$$

$$e^{P(\varepsilon)(\Phi(\varepsilon)-I)t}=\sum_i e^{\lambda_i(\varepsilon)-1)t}P_i(\varepsilon). \tag{39}$$

By matching the terms in these sums and applying the scalar result (35), the approximation (36) follows directly.

When the eigenvalues of $P(\varepsilon)\Phi(\varepsilon)$ are not semi-simple on $\varepsilon \in [0, \varepsilon_0)$, the approximation (36) can still be shown to be valid but the proof is not as straightforward. The basic results which will be employed is the matrix form of the scalar approximation (35) above.

Lemma 3. *Consider a matrix A with semi-simple zero eigenvalue and such that all the nonzero eigenvalues have negative real parts. Then*

$$(I+\varepsilon A)^t - e^{\varepsilon A t}=O(\varepsilon). \tag{40}$$

Proof. The case where the eigenvalues are semi-simple has been discussed in the text above. A general proof is available in [2]. □

The key to application of this lemma lies in isolating the various time scales and applying the result to each separately. Define the continuous time Markov generator

$$A^{(0)}(\varepsilon)\equiv\Phi^{(0)}(\varepsilon)-I. \tag{41}$$

$P(\varepsilon)A^{(0)}(\varepsilon)$ can be decomposed into terms

$$P(\varepsilon)A^{(0)}(\varepsilon)=\sum_{i=1}^{K}\varepsilon^i B^{(i)}(\varepsilon) \tag{42}$$

where

$$\varepsilon^i B^{(i)}(\varepsilon)=R^{(i)}(\varepsilon)A^{(0)}(\varepsilon) \tag{43}$$

and

$$R^{(i)}(\varepsilon)\equiv\sum_{i:\lambda_i(\varepsilon)-1=O(\varepsilon^i)}P_i(\varepsilon). \tag{44}$$

All the eigenprojections $R^{(i)}(\varepsilon)$ exist and are analytic at $\varepsilon=0$. This fact follows since $A^{(0)}(\varepsilon)$ is the generator of a continuous time Markov process and therefore satisfies the 'Multiple Semi-Simple Null Structure' condition which in turn guarantees that all these eigenprojections exist at $\varepsilon=0$ [8]. Essentially, $\varepsilon^i B^{(i)}(\varepsilon)$ captures all the eigenvalues of $\Phi^{(0)}(\varepsilon)-I$ which are strictly $O(\varepsilon^i)$.

Since the eigenvalues of $B^{(i)}(\varepsilon)$ are all identically zero or have strictly negative O(1) real parts, the ε-dependence is a real perturbation

$$e^{B^{(i)}(\varepsilon)t}=e^{B^{(i)}(0)t}+O(\varepsilon), \tag{45}$$

$$(I+\varepsilon^i B^{(i)}(\varepsilon))^t = (I+\varepsilon^i B^{(i)}(0))^t+O(\varepsilon). \tag{46}$$

Therefore applying Lemma 3 to the right hand sides of the above equations gives

$$(I+\varepsilon^i B^{(i)}(\varepsilon))^t=e^{\varepsilon^i B^{(i)}(\varepsilon)t}+O(\varepsilon). \tag{47}$$

By decomposing the terms in (36),

$$P(\varepsilon)(P(\varepsilon)\Phi^{(0)}(\varepsilon))^t=\sum_{i=1}^{k}R^{(i)}(\varepsilon)(I+\varepsilon^i B^{(i)}(\varepsilon))^t, \tag{48}$$

$$P(\varepsilon)\, e^{P(\varepsilon)(\Phi^{(0)}(\varepsilon)-I)t}=\sum_{i=1}^{k}R^{(i)}(\varepsilon)\, e^{\varepsilon^i B^{(i)}(\varepsilon)t}, \tag{49}$$

and matching the finite number of terms in these sums proves that the approximation (36) is indeed valid.

Finally, the term

$$P(\varepsilon)e^{P(\varepsilon)(\Phi^{(0)}(\varepsilon)-I)t}=P(\varepsilon)e^{P(\varepsilon)A^{(0)}(\varepsilon)t} \tag{50}$$

is identically the term for the slow behavior of the continuous time process generated by $A^{(0)}(\varepsilon)=\phi^{(0)}(\varepsilon)-I$. Using the results of [1], this can be written as

$$P(\varepsilon)\, e^{P(\varepsilon)A^{(0)}(\varepsilon)t}=U^{(0)}\, e^{\varepsilon A^{(1)}(\varepsilon)t}V^{(0)}+O(\varepsilon) \tag{51}$$

where $A^{(1)}(\varepsilon)$ is a reduced order Markov generator and $U^{(0)}$ and $V^{(0)}$ are the ergodic probability and memberhip matrices determined from $A^{(0)}(0)$.

4. Example

In this section, a simple two time scale, discrete time Markov chain is decomposed. Consider the

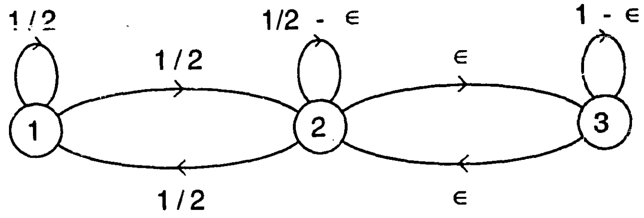

Fig. 1. Discrete time perturbed Markov process.

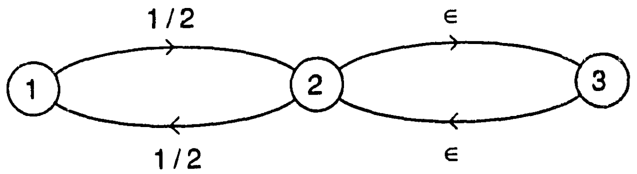

Fig. 2. Associated continuous time process.

process with the transition probability graph illustrated in Figure 1 and with generator

$$\Phi^{(0)}(\varepsilon)=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2}-\varepsilon & \varepsilon \\ 0 & \varepsilon & 1-\varepsilon \end{bmatrix}. \tag{52}$$

The transition rates of the continuous time process generated by $A^{(0)}(\varepsilon)=\Phi^{(0)}(\varepsilon)-I$ are shown in Figure 2. The slow time scale process obtained using the Markov algorithm is shown in Figure 3 and has a generator

$$A^{(1)}(\varepsilon)=\begin{bmatrix} -\frac{1}{2} & 1 \\ \frac{1}{2} & -1 \end{bmatrix}+\mathrm{O}(\varepsilon). \tag{53}$$

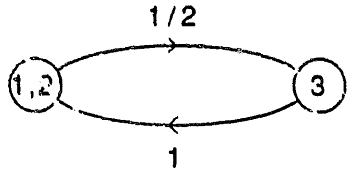

Fig. 3. $\mathrm{O}(1/\varepsilon)$ time scale continuous time process.

The combined approximation is therefore

$$\Phi^{(0)}(\varepsilon)^t$$

$$=\Phi^{(0)}(0)^t+U^{(0)}\,\mathrm{e}^{\varepsilon A^{(1)}(0)t}V^{(0)} - U^{(0)}V^{(0)}+\mathrm{O}(\varepsilon) \tag{54}$$

$$=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}^t \tag{55}$$

$$+\begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}\exp\left(\begin{bmatrix} -\frac{1}{2} & 1 \\ \frac{1}{2} & -1 \end{bmatrix}\varepsilon t\right)\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$-\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}+\mathrm{O}(\varepsilon) \tag{56}$$

References

[1] J.R. Rohlicek and A.S. Willsky, The reduction of perturbed Markov generators: An algorithm exposing the role of transient states, *J. ACM* **35** (3) (1988) 675–696.

[2] J.R. Rohlicek, Aggretation and Time Scale Analysis of Perturbed Markov Systems, Ph.D. Thesis, MIT (1987).

[3] G. Blankenship, Singularly perturbed difference equations in optimal control problems, *IEEE Trans. Automat. Control* **26** (4) (1981) 911–917.

[4] P.J. Courtois, *Decomposability: Queuing and Computer Systems Applications* (Academic Press, New York, 1977).

[5] H. Simon and A. Ando, Aggregation of variables in dynamic systems, *Econometrica* **29** (2) (1961) 111–139.

[6] M. Coderch, A.S. Willsky, S.S. Sastry, and D.A. Castinon, Hierarchical aggregation of singularly perturbed finite state Markov processes, *Stochastics* **8** (1983) 259–289.

[7] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).

[8] M. Coderch, Multiple Timescale Approach to Hierarchical Aggregation of Linear Systems, Ph.D. Thesis, MIT (1982).