Vol. 45 No. 4 SCIENCE IN CHINA (Series F) August 2002

Hybrid singular systems of differential equations

YIN Gang (殷　刚)[1] & ZHANG Jifeng (张纪峰)[2]

1. Department of Mathematics, Wayne State University, Detroit, MI 48202, USA;
2. Institute of Systems Science, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing, 100080, China

Correspondence should be addressed to Zhang Jifeng (email: jif@control.iss.ac.cn)

Received November 26, 2000

Abstract This work develops hybrid models for large-scale singular differential system and analyzes their asymptotic properties. To take into consideration the discrete shifts in regime across which the behavior of the corresponding dynamic systems is markedly different, our goals are to develop hybrid systems in which continuous dynamics are intertwined with discrete events under random-jump disturbances and to reduce complexity of large-scale singular systems via singularly perturbed Markov chains. To reduce the complexity of large-scale hybrid singular systems, two-time scale is used in the formulation. Under general assumptions, limit behavior of the underlying system is examined. Using weak convergence methods, it is shown that the systems can be approximated by limit systems in which the coefficients are averaged out with respect to the quasi-stationary distributions. Since the limit systems have fewer states, the complexity is much reduced.

Keywords: hybrid model, singular system, differential equation, singularly perturbed Markov chain, weak convergence, averaging.

Singular systems of differential equations arise in various applications in physical sciences, engineering, and economic systems. Due to their importance, such systems have been studied extensively and employed in control and optimization tasks. For some of the recent literature, we refer to refs. [1—4] among others. In these references, for a singular matrix A, a system

$$A\dot{x} + Bx = f(t), \quad x(0) = x_0, \tag{0.1}$$

and/or the related control problems are dealt with. Many interesting and important results have been obtained; a wide range of applications have been examined. The main effort up to date has been devoted to deterministic systems, whereas formulation under random disturbances has not received much attention, to the best of our knowledge. In this work, our goals are to develop hybrid systems in which continuous dynamics are intertwined with discrete events under random disturbances and to reduce complexity of large-scale singular systems via singularly perturbed Markov chains.

Often the underlying dynamic systems of various real-world applications are not only time-varying, but also associated with dramatic movements involving discontinuity, which are influenced by uncertain, exogenous discrete events driven by random disturbances. Many of such systems involve noise of pure jump type, especially for those arising in production planning, economics, and stochastic networks. As a result, the parameters of the dynamic systems come from one of the different regimes with transitions among regimes governed by an unobservable jump process. To model such systems, Markovian formulations have been found to be useful. To better reflect reality and to produce mathematically tractable models, we introduce hybrid models of singular differential systems in this paper. Our main motivation stems from taking discrete event interventions into consideration. For example, let us consider a nation's economy, in which the discrete events are coined by economists as discrete shifts. The evolution of the economy often displays dramatic moves (ups or downs), which then is naturally modeled by a Markov chain. Using such a premise, in lieu of considering a fixed matrix A in (0.1), we treat the case that A is time-varying and depends on a continuous-time Markov chain, which is motivated by a wide range of applications in stochastic networks, communication systems, production planning, and manufacturing; see for example, refs. [5—7] and the references therein.

The rapid increase in computational capabilities and an increasingly quantitative approach to problem solving have posed challenging tasks upon us since the underlying systems are frequently large-scale ones. To incorporate various needs into the model, the underlying Markov chain may have a large state space due to the complex nature of the system. To reduce the complexity of such systems, we exploit the hierarchical structure of the underlying models. The essence of the nearly completely decomposable matrix models, originating in Simon and Ando's work[8] is that in a large-scale system, not all components of the system change at the same rate. Thus it is helpful to model the system via time-scale separation. The resulting system becomes one with both fast-time scale and slow-time scale, leading to a singular perturbation formulation. Such ideas have been successfully used in control of large-scale dynamic systems and manufacturing; see, for example, refs. [5, 6, 9, 10] among others. One of the main components in these works is modeling and analysis via the use of singularly perturbed Markov chains. Owing to the prevalence in various applications, such singularly perturbed Markov chains have received resurgent attention lately; see, for example, refs. [6, 7, 10—16], among others.

In this work, by means of hierarchical approach, decomposition, and aggregation, we propose new switching models of singular differential systems under stochastic disturbances via time-scale separation, resulting in models involving singularly perturbed Markov chains. To highlight the two-time scale, we introduce a small parameter $\varepsilon > 0$. To treat control and optimization problems of such hybrid systems, it is foremost to have a thorough understanding of singular differential systems under singularly perturbed Markovian disturbances. By focusing on the asymptotic behavior of the systems as $\varepsilon \to 0$, we use methods of weak convergence to derive limit results. It is shown that the underlying systems can be approximated by limit systems in which the coefficients are averaged out with respect to the quasi-stationary measures of the Markov chain. Note that the small parameter ε may not appear in the original physical problems. It is for the reason of facilitating the analysis and hierarchical decomposition that we introduce it into the systems. How small is an ε considered to be small? In applications, constants such as 0.1 or 0.5 etc. could be considered as small enough. It mainly indicates the relative order of magnitude in the formulation, and provides guidelines. The asymptotic results of the underlying system (as $\varepsilon \to 0$) give insights into the structure of the systems and provide heuristics for various applications. The averaging approach to be presented will be useful in construction of asymptotic or nearly optimal controls of various systems.

The rest of the paper is arranged as follows. Section 1 presents the precise formulation of the problem. Section 2 recalls some preliminary results on singular systems of differential equations and singularly perturbed Markov chains. Section 3 is concerned with the asymptotic behavior of singular differential systems under the influence of singularly perturbed Markov chains. As an application, Section 4 is devoted to an illustrative example; it presents a hybrid/switching Leontief model and the corresponding asymptotic results. Finally, Section 5 concludes the paper with a few further remarks. Throughout the paper, we use K to denote a generic positive constant. Its value may change for different usage. For $z \in \mathcal{R}^{r\times l}$, we use z' to denote its transpose and $|z|$ to denote its norm.

1 Preliminary results

1.1 Singular systems of differential equations

First, we recall a number of definitions. For more discussion on related issues, the reader is referred to ref. [1]. For $A \in \mathcal{R}^{r\times r}$, the index of A, denoted by $\mathrm{Ind}(A)$, is the least nonnegative integer μ such that $\mathcal{N}(A^\mu) = \mathcal{N}(A^{\mu+1})$. Let $A \in \mathcal{R}^{r\times r}$ with $\mathrm{Ind}(A) = \mu$, dim $\mathcal{R}(A^\mu) = \ell_0$, and dim $\mathcal{N}(A^\mu) = \ell_1$ $(\ell_0 + \ell_1 = r)$, where $\mathcal{R}(A)$ and $\mathcal{N}(A)$ denote the range and null space of A, respectively, and $A = G\begin{pmatrix} C & 0 \\ 0 & N \end{pmatrix}G^{-1}$, where C is an $\ell_0 \times \ell_0$ nonsingular matrix and N is an $\ell_1 \times \ell_1$ nilpotent matrix with $\mu = \mathrm{Ind}(N)$. The Drazin inverse of A, denoted by A^D, is defined as $A^D = G\begin{pmatrix} C^{-1} & 0 \\ 0 & 0 \end{pmatrix}G^{-1}$.

For $A, B \in \mathcal{R}^{r\times r}$, and an $\mathcal{R}^r$-valued function $f(\cdot)$, consider

$$A\dot{x} + Bx = f(t), \quad x(t_0) = x_0. \tag{1.1}$$

The vector x_0 is said to be a consistent initial vector if (1.1) has at least one solution. The differential equation in (1.1) is tractable at t_0 if the initial value problem has a unique solution for each consistent initial vector x_0. Moreover, if a system is tractable at a point t_0, then it is tractable at all t (see ref. [1]).

It follows that[1], the corresponding homogeneous differential equation (i.e. a differential equation (1.1) with $f(t) = 0$) is tractable iff there is a $\lambda \in \mathbb{C}$ such that $(\lambda A + B)^{-1}$ exists.

Define

$$\varPhi_A = (\lambda A + B)^{-1}A, \ \ \varPhi_B = (\lambda A + B)^{-1}B, \ \ \text{and} \ \ \varPhi_f = (\lambda A + B)^{-1}f, \tag{1.2}$$

where $\lambda \in \mathbb{C}$ such that $(\lambda A + B)$ is nonsingular. Note that Theorem 3.1.2 in ref. [1] indicates that the solution of singular systems of differential equations is independent of λ. If the homogeneous equation is tractable, then the general solution is given by

$$x(t) = \exp(-\Phi_A^D \Phi_B (t-t_0))\Phi_A \Phi_A^D v,\ \ v \in \mathcal{R}^r. \tag{1.3}$$

A vector $c \in \mathcal{R}^r$ is a consistent initial vector for the homogeneous equation iff

$$c = \Phi_A \Phi_A^D c \text{ for } c \in \mathcal{R}(\Phi_A^\kappa) = \mathcal{R}(\Phi_A \Phi_A^D),$$

where $\kappa = \mathrm{Ind}(\Phi_A)$.

Suppose that $f(t)$ is κ-times continuously differentiable at t_0 with $\kappa = \mathrm{Ind}(\Phi_A)$. Then the nonhomogeneous equation (1.1) has a general solution

$$\begin{aligned} x(t) &= \exp(-\Phi_A^D \Phi_B (t-t_0))\Phi_A \Phi_A^D v + \int_{t_0}^t \exp(-\Phi_A^D \Phi_B (t-s))\Phi_A^D \Phi_f(s)ds + w(t), \\ w(t) &= (I - \Phi_A \Phi_A^D)\sum_{i=0}^{\kappa-1}(-1)^i (\Phi_A \Phi_B^D)^i \Phi_B^D \Phi_f^{(i)}(t), \end{aligned} \tag{1.4}$$

where $v \in \mathcal{R}^r$, and $f^{(i)}$ denotes the ith derivative of $f(\cdot)$. A vector $c \in \mathcal{R}^r$ is a consistent initial vector associated with t_0 for the nonhomogeneous equation iff c is a solution of

$$(I - \Phi_A \Phi_A^D)(c - w(t_0)) = 0.$$

Moreover, the nonhomogeneous equation is tractable at t_0 and the unique solution of the initial value problem with $x(t_0) = x_0$ is given by (1.4) with $v = x_0$.

1.2 Singularly perturbed Markov chains

Our interests are mainly on nonstationary continuous-time Markov chains. Consider a continuous-time Markov chain $\alpha(\cdot)$ with finite state space $\mathcal{M} = \{1, \cdots, m\}$.

For $i, j \in \mathcal{M}$, with $Q(t) = (q_{ij}(t))$, for $t \geqslant 0$, for any real-valued function g on $\mathcal{M}$ and $i \in \mathcal{M}$, write

$$Q(t)g(\cdot)(i) = \sum_{j\in\mathcal{M}} q_{ij}(t)g(j) = \sum_{j\neq i} q_{ij}(t)(g(j) - g(i)).$$

We say that $Q(t)$, $t \geqslant 0$, is a generator of $\alpha(\cdot)$ if $q_{ij}(t)$ is bounded and Borel measurable, $q_{ij}(t) \geqslant 0$ for $j \neq i$, $q_{ii}(t) = -\sum_{j\neq i} q_{ij}(t)$, $t \geqslant 0$, and for any bounded real-valued function g defined on $\mathcal{M}$

$$g(\alpha(t)) - \int_0^t Q(\varsigma)g(\cdot)(\alpha(\varsigma))d\varsigma \tag{1.5}$$

is a martingale. A generator $Q(t)$ is said to be weakly irreducible if, for each fixed $t \geqslant 0$, the system of equations

$$\nu(t)Q(t) = 0, \quad \sum_{i=1}^m \nu_i(t) = 1 \tag{1.6}$$

has a unique solution $\nu(t) = (\nu_1(t), \ldots, \nu_m(t))$ and $\nu_i(t) \geqslant 0$ for $i = 1, \ldots, m$. Such a solution is termed a quasi-stationary distribution (see ref. [6]). The modifier "quasi-" is used since $\nu(t)$ is time dependent and it does not require all of its components be strictly positive. The weak irreducibility, being first introduced in ref. [12], is a generalization of the usual notion of irreducibility. For example, $\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$ is not irreducible but it is weakly irreducible. Such extensions are needed and convenient for various applications. For further discussions on nonstationary Markov chains through piecewise deterministic process formulation, we refer the reader to ref. [17].

2 Formulation

2.1 The model

Suppose that $\varepsilon > 0$ is a small parameter, $\alpha^\varepsilon(t)$ is a continuous-time Markov chain with $\alpha^\varepsilon(0) = \alpha_0$ and finite state space

$$\begin{aligned}\mathcal{M} &= \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_l \cup \mathcal{M}_* \\ &= \{s_{11}, \cdots, s_{1m_1}\} \cup \cdots \cup \{s_{l1}, \cdots, s_{lm_l}\} \cup \{s_{*1}, \cdots, s_{*m_*}\}.\end{aligned} \tag{2.1}$$

Note that for each $i = 1, \cdots, l$, $\mathcal{M}_i$ is the subspace of states in the ith recurrent class, and $\mathcal{M}_*$ is the collection of transient states (To distinguish the recurrent and transient states, we use the index $*$ for the transient states.). Let the generator of $\alpha^\varepsilon(t)$ be

$$Q^\varepsilon(t) = \frac{\tilde{Q}(t)}{\varepsilon} + \hat{Q}(t) \tag{2.2}$$

such that $\tilde{Q}(t)$ and $\hat{Q}(t)$ are themselves generators and

$$\tilde{Q}(t) = \begin{pmatrix} \tilde{Q}^1(t) & & & & \\ & \tilde{Q}^2(t) & & & \\ & & \ddots & & \\ & & & \tilde{Q}^l(t) & \\ \tilde{Q}^1_*(t) & \tilde{Q}^2_*(t) & \cdots & \tilde{Q}^l_*(t) & \tilde{Q}_*(t) \end{pmatrix}. \tag{2.3}$$

In view of (2.2) and (2.3), $\alpha^\varepsilon(\cdot)$ is a Markov chain involving weak and strong interactions. Within each $\mathcal{M}_i$, the transitions take place in a fast pace, whereas the jumps from $\mathcal{M}_i$ to $\mathcal{M}_j$ occur relatively infrequently.

For $0 < T < \infty$, we work with the time horizon $[0, T]$. Consider the singular systems of differential equations of the form

$$A(\alpha^\varepsilon(t))\dot{x}^\varepsilon + B(\alpha^\varepsilon(t))x^\varepsilon = f(t, \alpha^\varepsilon(t)), \quad x^\varepsilon(0) = x_0. \tag{2.4}$$

In (2.4), $x(t) \in \mathcal{R}^r$, $f(\cdot)$ is an $\mathcal{R}^r$-valued function, and for each $\iota \in \mathcal{M}$, $A(\iota)$ and $B(\iota)$ are $r \times r$ matrices. The precise conditions of $A(\cdot)$, $B(\cdot)$, and $f(\cdot)$ will be given in the sequel.

Since $\alpha^\varepsilon(\cdot)$ is a pure jump process, its sample paths are piecewise constant. As a result, in lieu of a fixed set of coefficients, (2.4) has card($\mathcal{M}$) number of regimes (or configurations) for the system coefficients, where card($\mathcal{M}$) denotes the cardinality of $\mathcal{M}$. Between two jumps of $\alpha^\varepsilon(\cdot)$, (2.4) is a deterministic system, which is an example of nowadays popular hybrid systems.

2.2 Rationale

The form (2.2) and (2.3) of the generator is originated from hierarchical decomposition. Suppose that we have a large-scale system. Model the disturbances by a finite-state Markov chain. Since not all states can be transient, there is at least one recurrent state (see ref. [18]). In accordance with the transition rates of the recurrent states, the recurrent states can be grouped into l recurrent classes. The remaining transient states are in $\mathcal{M}_*$. The precise conditions of the generator will be spelt out later.

In the subsequent analysis, it is required that $\varepsilon \to 0$. In the actual applications, however, $\varepsilon > 0$ is simply a small positive constant. Ref. [6] provides an illustrative example on how to convert a generator into the form (2.2). The procedure uses an elementary matrix row and column operations.

We concern ourselves with the case that the state space $\mathcal{M}$ is large, i.e. $\mathcal{M}$ contains a large number of elements. The amount of computation in handling the underlying system could be intensive. Rather than treating the system directly, we aggregate the states in each recurrent class into one state, resulting in an aggregated process with state space having l elements. Using such an aggregation, we shall derive a limit system that depends only on l possible regimes or configurations. Suppose that $\text{card}(\mathcal{M}) = m$ and $l \ll m$. Then the complexity is substantially reduced.

2.3 Assumptions

We make the following assumptions. The first condition is about the singularly perturbed Markov chain, and the second one is concerned with system (2.4).

(A1) The generators $\tilde{Q}(\cdot)$ and $\hat{Q}(\cdot)$ are uniformly bounded and Borel measurable such that

(a) for each $i = 1, \cdots, l$, $\tilde{Q}^i(t)$ is weakly irreducible;

(b) $\tilde{Q}_*(t)$ is asymptotically stable, i.e. all of its eigenvalues belong to the left half of the complex plane;

(c) for each $i = 1, \cdots, l$, there exist constant matrices $G_*^i \in \mathcal{R}^{m_* \times m_i}$, $G_* \in \mathcal{R}^{m_* \times m_*}$, and matrix-valued function $B(t) \in \mathcal{R}^{m_* \times m_*}$ such that $B(\cdot)$ is Lipschitz continuous, and

$$\tilde{Q}_*^i(t) = B(t) G_*^i \text{ and } \tilde{Q}_*(t) = B(t) G_*.$$

(A2) For each $\iota \in \mathcal{M}$,

(a) there exists $\lambda(\iota)$ such that $\lambda(\iota)A(\iota) + B(\iota)$ is nonsingular;

(b) $f(\cdot, \iota)$ is $(\kappa + 1)$-times continuously differentiable, where

$$\kappa = \max_{\iota \in \mathcal{M}} \text{Ind}\left((\lambda(\iota)A(\iota) + B(\iota))^{-1} A(\iota) \right).$$

Remark 2.1. If the last row of $\tilde{Q}(t)$ disappears (i.e. the corresponding chain has only recurrent states), then the Lipschitz condition in (A1) is no longer needed (see ref. [16] for further details). If a full asymptotic expansion of the solution of the corresponding forward equation (see (3.1) in what follows) is desired, more smoothness of $\tilde{Q}(t)$ and $\hat{Q}(t)$ will be needed (see ref. [6]).

To see the implication of (A1) (c), for $i = 1, \cdots, l$, define

$$a_{m_i}(t) = -\tilde{Q}_*^{-1}(t) \tilde{Q}_*^i(t) \mathbb{1}_{m_i}. \tag{2.5}$$

Using (A1) (c), it is easily seen that $a_{m_k}(t) = a_{m_k}$ is independent of t and is an $(m_* \times 1)$-vector with $a_{m_i} = (a_{m_i,1}, \cdots, a_{m_i,m_*})'$. Moreover (see ref. [15]),

$$a_{m_i,j} \geqslant 0, \quad \text{and} \quad \sum_{i=1}^{l} a_{m_i,j} = 1 \text{ for each } j = 1, \cdots, m_*. \tag{2.6}$$

Thus for each $t \in [0,T]$ and $j = 1, \cdots, m_*$, $(a_{m_1,j}, \cdots, a_{m_l,j})$ is a probability row vector. Its component $a_{m_i,j}$ represents the probability that the chain jumps from the jth transient state s_{*j} to the ith recurrent class $\mathcal{M}_i$.

If $\tilde{Q}(t)$ consists of only one irreducible block, the corresponding system is one subject to fast variations. In this case, the asymptotic study to be presented in what follows still provides a reduction in complexity. It indicates that one can ignore the detailed variations and concentrate only on the average effect.

(A2) is a sufficient condition for tractability. That is, (A2) implies that the system under consideration is tractable.

The formulation is completed. We are now in a position to analyze the underlying system.

3 Limit behavior

This section is divided into four parts. We first recall several results concerning the singularly perturbed Markov chains to be used in our study. Then we obtain the tightness of the process $\{x^\varepsilon(\cdot)\}$. Next, we derive the weak convergence of this process. Finally, we treat a special case in which $\bar{Q}(t)$ consists of a single weakly irreducible block. In what follows, for an integer ℓ, by $D^\ell[0,T]$, we mean the space of $\mathcal{R}^\ell$-valued functions that are right continuous and have left limits, endowed with the Skorohod topology (see refs. [19—22] among others).

3.1 Asymptotic properties of $\alpha^\varepsilon(\cdot)$

Denote $p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 1), \cdots, P(\alpha^\varepsilon(t) = m)) \in \mathcal{R}^{1\times m}$. Then the probability vector $p^\varepsilon(\cdot)$ satisfies the forward equation

$$\begin{aligned} \frac{dp^\varepsilon(t)}{dt} &= p^\varepsilon(t) Q^\varepsilon(t), \\ p^\varepsilon(0) &= p_0, \ p_{0,i} \geqslant 0, \quad \sum_{i=1}^{m} p_{0,i} = 1, \end{aligned} \tag{3.1}$$

$p_{0,i}$ denotes the ith component of p_0.

Define

$$\begin{aligned} \tilde{\mathbb{1}}_* &= \begin{pmatrix} \mathbb{1}_{m_1} & & & \\ & \ddots & & \\ & & \mathbb{1}_{m_l} & \\ a_{m_1} & \cdots & a_{m_l} & 0_{m_*\times m_*} \end{pmatrix}, \\ \tilde{\mathbb{1}} &= \mathrm{diag}(\mathbb{1}_{m_1}, \cdots, \mathbb{1}_{m_l}), \end{aligned} \tag{3.2}$$

where $\mathbb{1}_\iota$ is an $\iota \times 1$ vector with all components being 1, $\mathrm{diag}(\Lambda_1, \cdots, \Lambda_l)$ is a block diagonal matrix with matrix entries $\Lambda_1, \cdots, \Lambda_l$, $0_{m_*\times m_*}$ is an $m_* \times m_*$ zero matrix, and a_{m_i} is given by

(2.6). For each $i = 1, \ldots, l$, denote the quasi-stationary distribution for $\tilde{Q}^i(t)$ by $\nu^i(t)$. Using the partition form

$$\hat{Q}(t) = \begin{pmatrix} \hat{Q}^{11}(t) & \hat{Q}^{12}(t) \\ \hat{Q}^{21}(t) & \hat{Q}^{22}(t) \end{pmatrix},$$

with

$$\hat{Q}^{11}(t) \in \mathcal{R}^{(m-m_*)\times(m-m_*)}, \quad \hat{Q}^{12}(t) \in \mathcal{R}^{(m-m_*)\times m_*},$$

$$\hat{Q}^{21}(t) \in \mathcal{R}^{m_*\times(m-m_*)}, \quad \text{and } \hat{Q}^{22}(t) \in \mathcal{R}^{m_*\times m_*},$$

write

$$\bar{Q}_*(t) = \mathrm{diag}(\nu^1(t), \cdots, \nu^l(t))(\tilde{Q}^{11}(t)\mathbb{1} + \tilde{Q}^{12}(t)(a_{m_1}, \cdots, a_{m_l})), \tag{3.3}$$

and

$$\bar{Q}(t) = \mathrm{diag}(\bar{Q}_*(t), 0_{m_*\times m_*}). \tag{3.4}$$

Define an aggregated process $\bar{\alpha}^\varepsilon(\cdot)$ by

$$\bar{\alpha}^\varepsilon(t) = \begin{cases} i, & \text{if } \alpha^\varepsilon(t) \in \mathcal{M}_i, \\ U_j, & \text{if } \alpha^\varepsilon(t) = s_{*j}, \end{cases}$$

where

$$U_j = \sum_{k=1}^{l} k I_{\{\sum_{i=1}^{k-1} a_{m_i,j} \leqslant U \leqslant \sum_{k=1}^{k} a_{m_i,j}\}},$$

and U is a random variable uniformly distributed in $[0, 1]$. Note that we only aggregate the states in each recurrent class since if the process is currently in one of the transient states, in a short period of time, it will jump into one of the recurrent classes.

For each $i = 1, \cdots, l, *$, $j = 1, \cdots, m_i$, define the occupation measures by

$$o_{ij}^\varepsilon(t) = \begin{cases} \displaystyle\int_0^t I_{\{\alpha^\varepsilon(s)=s_{ij}\}} - \nu_j^i(s) I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}} ds, & \text{if } i = 1, \cdots, l; \\ \displaystyle\int_0^t I_{\{\alpha^\varepsilon(s)=s_{*j}\}} ds, & \text{if } i = *. \end{cases} \tag{3.5}$$

We now present results on the solution of the forward equation, the asymptotics of the occupation measures, and the weak convergence of the aggregated process.

Theorem 3.1. Assume (A1) and (A2). Then

(i) $p^\varepsilon(t) \to p(t)$ as $\varepsilon \to 0$ for all $t \in [0, T]$, where

$$p(t) = (\theta(t)\mathrm{diag}(\nu^1(t), \cdots, \nu^l(t)), 0_{1\times m_*}) \tag{3.6}$$

and where

$$\dot{\theta}(t) = \theta(t)\bar{Q}_*(t), \quad \theta(0) = (p_0\tilde{\mathbb{1}}_*)_r \tag{3.7}$$

with $\theta(t) = (\theta_1(t), \ldots, \theta_l(t)) \in \mathcal{R}^{1\times l}$, and $(p_0 \mathbb{1}_*)_r$ being the partitioned vector corresponding to the recurrent part.

(ii) For each $i = 1, \cdots, l, *,, j = 1, \cdots, m_i$,

$$\sup_{t\in[0,T]} E[o^{\varepsilon}_{ij}(t)]^2 \to 0 \text{ as } \varepsilon \to 0. \tag{3.8}$$

(iii) $\bar{\alpha}^{\varepsilon}(\cdot)$ converges weakly to $\bar{\alpha}(\cdot)$, a Markov chain generated by $\bar{Q}_*(t)$.

Remark 3.1. The proofs of the three parts of the theorem can be found in that of Theorem 3.4, Theorem 4.2, and Theorem 4.3, in ref. [16], respectively. Note that the convergence in assertion (i) above is in the sense of pointwise convergence. The $o^{\varepsilon}_{ij}(t)$ is termed an occupation measure since it measures the amount of time the underlying Markov chain spends in a given state. Assertion (ii) gives a mean squares convergence of the sequence of unscaled occupation measures. Although the aggregated process $\bar{\alpha}^{\varepsilon}(\cdot)$ may not be Markov, its weak limit turns out to be a Markov chain whose generator is an "average" of the generator $\hat{Q}(t)$ with respect to the quasi-stationary distributions.

3.2 Tightness

We aim to prove the weak convergence of $x^{\varepsilon}(\cdot)$. To do so, we need to verify the tightness of the underlying process and then characterize the limit. Rather than working with $x^{\varepsilon}(\cdot)$, we use a device known as N-truncation (see ref. [21] or [22]) and work with a truncated process. We show that the truncated process is tight and converges weakly. Finally, by using the uniqueness of the solution (tractability), we conclude the proof.

To proceed, for any $0 < N$, let $S_N = \{x; |x| \leqslant N\}$. That is, S_N is the sphere with radius N centered at the origin. The process $x^{\varepsilon,N}(\cdot)$ is an N-truncation of $x^{\varepsilon}(\cdot)$, if

$$\lim_{K_0\to\infty} \limsup_{\varepsilon\to 0} P(\sup_{t\in[0,T]} |x^{\varepsilon,N}(t)| \geqslant K_0) = 0,$$

and $x^{\varepsilon,N}(t) = x^{\varepsilon}(t)$ up until the first exit from S_N. Define a smooth real-valued function $q^N(\cdot)$ as $q^N(x) = 1$ if $x \in S_N$; $q^N(x) = 0$ if $x \in \mathcal{R}^r - S_{N+1}$. Consider

$$\begin{aligned} &A(\alpha^{\varepsilon}(t))\dot{x}^{\varepsilon,N}(t) + B(\alpha^{\varepsilon}(t))x^{\varepsilon,N}(t)q^N(x^{\varepsilon,N}(t)) = f(t,\alpha^{\varepsilon}(t))q^N(x^{\varepsilon,N}(t)),\\ &x^{\varepsilon,N}(0) = x_0. \end{aligned} \tag{3.9}$$

Thus, $x^{\varepsilon,N}(\cdot)$ is a process with compact support; it stops at the first time S_{N+1} exits, equal to $x^{\varepsilon}(\cdot)$ up until the first exit from S_N, and "decays" between the spheres S_N and S_{N+1}. The truncation device enables us to work with a bounded process and to obtain the desired result for such a process.

Theorem 3.2. Assume (A1) and (A2). Then $\{x^{\varepsilon,N}(\cdot)\}$ given by (3.9) is tight in $D^r[0,T]$.

Proof. Due to the truncation, $\{x^{\varepsilon,N}(\cdot)\}$ is bounded. Suppose that the nth jump of $\alpha^{\varepsilon}(t)$ happens at the moment τ^{ε}_n with $\tau^{\varepsilon}_0 = 0$. Then $\{\tau^{\varepsilon}_n\}$ is a sequence of $\mathcal{F}^{\varepsilon}_t$-stopping times, where $\mathcal{F}^{\varepsilon}_t$ is the σ-algebra generated by $\{\alpha^{\varepsilon}(s); s \leqslant t\}$ (If the generators are constants, the distribution of $\tau_{n+1} - \tau_n$ is well known and is exponentially distributed. For time-dependent generators, the distribution of $\tau_{n+1} - \tau_n$ can be found in ref. [17]; see also ref. [6] for more details.). For any

$t \in [0,T]$, t will be between two jump times, i.e. there is an n such that $\tau_n \leqslant t < \tau_{n+1} \wedge T$, where $\tau_{n+1} \wedge T = \min(\tau_{n+1}, T)$. Assuming $\alpha^\varepsilon(\tau_n) = s_{ij}$ for some $i = 1, \cdots, l$ or $i = *$, $j = 1, \cdots, m_i$, then $\alpha^\varepsilon(t) = s_{ij}$ for all $t \in [\tau_n, \tau_{n+1} \wedge T)$. By (A2), the system is tractable at τ_n, and hence

$$\begin{aligned} x^{\varepsilon,N}(t) =& \exp(-\Phi_A^D(s_{ij})\Phi_B(s_{ij})(t-\tau_n^\varepsilon))\Phi_A(s_{ij})\Phi_A^D(s_{ij})x^{\varepsilon,N}(\tau_n) \\ &+ \int_{\tau_n}^t \exp(-\Phi_A^D(s_{ij})\Phi_B(s_{ij})(t-s))\Phi_A^D(s_{ij})\Phi_f(s,s_{ij})ds \\ &+ (I - \Phi_A(s_{ij})\Phi_A^D(s_{ij}))\sum_{\iota=0}^{k-1}(-1)^\iota(\Phi_A(s_{ij})\Phi_B^D(s_{ij}))^\iota \Phi_B^D(s_{ij})\Phi_f^{(\iota)}(t,s_{ij}), \end{aligned} \tag{3.10}$$

where $\Phi_A(s_{ij})$, $\Phi_B(s_{ij})$, and $\Phi_f(t,s_{ij})$ are defined in (1.2). Differentiating (3.10) with respect to t yields that

$$\begin{aligned} \dot{x}^{\varepsilon,N}(t) =& -\Phi_A(s_{ij})\Phi_A^D(s_{ij})\exp(-\Phi_A^D(s_{ij})\Phi_B(s_{ij})(t-\tau_n^\varepsilon))\Phi_A(s_{ij})\Phi_A^D(s_{ij})x^{\varepsilon,N}(\tau_n) \\ &- \Phi_A(s_{ij})\Phi_A^D(s_{ij})\int_{\tau_n}^t \exp(-\Phi_A^D(s_{ij})\Phi_B(s_{ij})(t-s))\Phi_A^D(s_{ij})\Phi_f(s,s_{ij})ds \\ &+ \Phi_A^D(s_{ij})\Phi_f(t,s_{ij}) \\ &+ (I - \Phi_A(s_{ij})\Phi_A^D(s_{ij}))\sum_{\iota=1}^{\kappa}(-1)^\iota(\Phi_A(s_{ij})\Phi_B^D(s_{ij}))^\iota \Phi_B^D(s_{ij})\Phi_f^{(\iota)}(t,s_{ij}). \end{aligned} \tag{3.11}$$

The definition of $x^{\varepsilon,N}(\cdot)$, the condition on $f(\cdot)$, and the familiar triangle inequality then imply $\sup_{t\in[0,T]}|\dot{x}^{\varepsilon,N}(t)| \leqslant K$.

To proceed, by (A2), for any $t \in [0,T]$, there is a $\lambda(\alpha^\varepsilon(t))$ such that $\lambda(\alpha^\varepsilon(t))A(\alpha^\varepsilon(t)) + B(\alpha^\varepsilon(t))$ is invertible. Thus (3.9) leads to

$$\begin{aligned} \dot{x}^{\varepsilon,N}(t) =& -[\lambda(\alpha^\varepsilon(t))A(\alpha^\varepsilon(t)) + B(\alpha^\varepsilon(t))]^{-1}\lambda(\alpha^\varepsilon(t))B(\alpha^\varepsilon(t))x^{\varepsilon,N}(t)q^N(x^{\varepsilon,N}(t)) \\ &+ [\lambda(\alpha^\varepsilon(t))A(\alpha^\varepsilon(t)) + B(\alpha^\varepsilon(t))]^{-1}\lambda(\alpha^\varepsilon(t))f(t,\alpha^\varepsilon(t))q^N(x^{\varepsilon,N}(t)) \\ &+ [\lambda(\alpha^\varepsilon(t))A(\alpha^\varepsilon(t)) + B(\alpha^\varepsilon(t))]^{-1}B(\alpha^\varepsilon(t))\dot{x}^{\varepsilon,N}(t). \end{aligned} \tag{3.12}$$

It then follows

$$\begin{aligned} x^{\varepsilon,N}(t) =& x_0 - \int_0^t [\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}\lambda(\alpha^\varepsilon(s))B(\alpha^\varepsilon(s))x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))ds \\ &+ \int_0^t [\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}\lambda(\alpha^\varepsilon(s))f(s,\alpha^\varepsilon(s))q^N(x^{\varepsilon,N}(s))ds \\ &+ \int_0^t [\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}B(\alpha^\varepsilon(s))\dot{x}^{\varepsilon,N}(s)ds. \end{aligned} \tag{3.13}$$

For any $\delta > 0$ and $t, u > 0$ with $0 \leqslant u < \delta$, using (3.13), we have

$$\begin{aligned} & E_t^\varepsilon|x^{\varepsilon,N}(t+u) - x^{\varepsilon,N}(t)|^2 \\ \leqslant & 3E_t^\varepsilon\left|\int_t^{t+u}[\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}\lambda(\alpha^\varepsilon(s))B(\alpha^\varepsilon(s))x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))ds\right|^2 \\ &+ 3E_t^\varepsilon\left|\int_t^{t+u}[\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}\lambda(\alpha^\varepsilon(s))f(s,\alpha^\varepsilon(s))q^N(x^{\varepsilon,N}(s))ds\right|^2 \\ &+ 3E_t^\varepsilon\left|\int_t^{t+u}[\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}B(\alpha^\varepsilon(s))\dot{x}^{\varepsilon,N}(s)ds\right|^2 \\ \leqslant & \zeta^\varepsilon(t,u), \end{aligned} \tag{3.14}$$

where $\zeta^\varepsilon(t,u)$ is $\mathcal{F}_t^\varepsilon$-measurable and

$$\limsup_{\varepsilon\to 0} E\zeta^\varepsilon(t,u) = O(u^2) = O(\delta^2). \tag{3.15}$$

In the above, we have used the boundedness of $x^{\varepsilon,N}(\cdot)$ and $\dot{x}^{\varepsilon,N}(\cdot)$, the boundedness of $A(\iota)$ and $B(\iota)$, the continuity of $f(\iota,\cdot)$ for each $\iota \in \mathcal{M}$, and

$$E_t^\varepsilon \left| \int_t^{t+u} h(s)ds \right|^2 \leqslant \zeta^\varepsilon(t,u)$$

such that (3.15) holds, where $h(s)$ is any one of the following functions

$$[\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}\lambda(\alpha^\varepsilon(s))B(\alpha^\varepsilon(s))x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s)),$$
$$[\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}\lambda(\alpha^\varepsilon(s))f(s,\alpha^\varepsilon(s))q^N(x^{\varepsilon,N}(s)),$$
$$[\lambda(\alpha^\varepsilon(s))A(\alpha^\varepsilon(s)) + B(\alpha^\varepsilon(s))]^{-1}B(\alpha^\varepsilon(s))\dot{x}^{\varepsilon,N}(s).$$

It follows

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0} E|x^{\varepsilon,N}(t+u) - x^{\varepsilon,N}(t)|^2 \leqslant \lim_{\delta\to 0} O(\delta^2) = 0. \tag{3.16}$$

The desired tightness of $\{x^{\varepsilon,N}(\cdot)\}$ then follows from the tightness criteria (see refs. [20, 21]).

Since $\{x^{\varepsilon,N}(\cdot)\}$ is tight, we can extract a weakly convergent subsequence by Prohorov's theorem (see refs. [19, 20]). Select such a convergent subsequence. For notational simplicity, still denote the sequence by $x^{\varepsilon,N}(\cdot)$ (i.e. still use ε as its index). In view of (3.15), we also have the following corollary (see Corollary in ref. [20]).

Corollary 3.1. Under the conditions of Theorem 3.2, the limit of $x^{\varepsilon,N}(\cdot)$ has continuous sample paths w.p.1 (with probability one).

3.3 Weak convergence

Theorem 3.3. Under the conditions of Theorem 3.2, $x^{\varepsilon,N}(\cdot)$ converges weakly to $x^N(\cdot)$, a solution of the singular system of differential equations

$$\overline{A}(\overline{\alpha}(t))\dot{x}^N + \overline{B}(\overline{\alpha}(t))x^N q^N(x^N) = \overline{f}(t,\overline{\alpha}(t))q^N(x^N), \quad x(0) = x_0, \tag{3.17}$$

where

$$\begin{aligned}
\overline{A}(\overline{\alpha}(t)) &= \sum_{i=1}^{l}\sum_{j=1}^{m_i}\nu_j^i(t)A(s_{ij}),\\
\overline{B}(\overline{\alpha}(t)) &= \sum_{i=1}^{l}\sum_{j=1}^{m_i}\nu_j^i(t)B(s_{ij}),\\
\overline{f}(t,\overline{\alpha}(t)) &= \sum_{i=1}^{l}\sum_{j=1}^{m_i}\nu_j^i(t)f(t,s_{ij}).
\end{aligned} \tag{3.18}$$

Proof. In view of Corollary 3.1, the limit of the weakly convergent subsequence has continuous paths w.p.1. We proceed to characterize the limit process.

Integrating (2.4) leads to

$$\begin{aligned}
&\int_0^t A(\alpha^\varepsilon(s))\dot{x}^{\varepsilon,N}(s)ds + \int_0^t B(\alpha^\varepsilon(s))x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))ds\\
&= \textstyle\int_0^t f(s,\alpha^\varepsilon(s))q^N(x^{\varepsilon,N}(s))ds.
\end{aligned} \tag{3.19}$$

Owing to the piecewise constant property of $A(\alpha^\varepsilon(t))$, we can write the system as

$$\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t A(s_{ij})\dot{x}^{\varepsilon,N}(s)I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds+\sum_{j=1}^{m_*}\int_0^t A(s_{*j})\dot{x}^{\varepsilon,N}(s)I_{\{\alpha^\varepsilon(s)=s_{*j}\}}ds\\
=&-\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t B(s_{ij})x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds\\
&-\sum_{j=1}^{m_*}\int_0^t B(s_{*j})x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{*j}\}}ds\\
&+\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t f(s,s_{ij})q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds\\
&+\sum_{j=1}^{m_*}\int_0^t f(s,s_{*j})q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{*j}\}}ds. \qquad (3.20)
\end{aligned}$$

To proceed, we treat each of the terms in (3.20) separately. First,

$$\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t A(s_{ij})\dot{x}^{\varepsilon,N}(s)I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds\\
=&\sum_{i=1}^{l}\sum_{j=1}^{m_i}A(s_{ij})\int_0^t \dot{x}^{\varepsilon,N}(s)[I_{\{\alpha^\varepsilon(s)=s_{ij}\}}-\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}]ds\\
&+\sum_{i=1}^{l}\sum_{j=1}^{m_i}A(s_{ij})\int_0^t \dot{x}^{\varepsilon,N}(s)\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}ds. \qquad (3.21)
\end{aligned}$$

Next, an integration by parts leads to

$$\begin{aligned}
&E\left|\int_0^t \dot{x}^{\varepsilon,N}(s)[I_{\{\alpha^\varepsilon(s)=s_{ij}\}}-\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}]ds\right|\\
\leqslant&E\left|\dot{x}^{\varepsilon,N}(t)\int_0^t[I_{\{\alpha^\varepsilon(s)=s_{ij}\}}-\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}]ds\right|\\
&+E\left|\int_0^t\left(\int_0^s[I_{\{\alpha^\varepsilon(u)=s_{ij}\}}-\nu_j^i(u)I_{\{\alpha^\varepsilon(u)\in\mathcal{M}_i\}}]du\right)\ddot{x}^{\varepsilon,N}(s)ds\right|.
\end{aligned}$$

Similar to (3.10) and (3.11), it can be verified that $\sup_{t\in[0,T]}|\ddot{x}^{\varepsilon,N}(t)|\leqslant K$. Then by the boundedness of $\ddot{x}^{\varepsilon,N}(s)$ and applying Theorem 3.1, in particular, (3.8), as $\varepsilon\to 0$, we obtain

$$E\left|\int_0^t \dot{x}^{\varepsilon,N}(s)[I_{\{\alpha^\varepsilon(s)=s_{ij}\}}-\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}]ds\right|\to 0,$$

and the limit is uniformly in $t\in[0,T]$. It follows from (3.21),

$$\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t A(s_{ij})\dot{x}^{\varepsilon,N}(s)I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds\\
=&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t A(s_{ij})\dot{x}^{\varepsilon,N}(s)\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}ds+o(1), \qquad (3.22)
\end{aligned}$$

where as $\varepsilon \to 0$, $o(1) \to 0$ in probability uniformly in $t \in [0,T]$. Likewise,

$$\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t B(s_{ij})x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds \\
&=\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t B(s_{ij})x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}ds+o(1), \\
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t f(s,s_{ij})q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{ij}\}}ds \\
&=\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t f(s,s_{ij})q^N(x^{\varepsilon,N}(s))\nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}ds+o(1),
\end{aligned}\tag{3.23}$$

where as $\varepsilon \to 0$, $o(1) \to 0$ in probability uniformly in $t \in [0,T]$.

By virtue of Theorem 3.1 (iii) and using (3.5), similar to the derivations of (3.21) and (3.22), as $\varepsilon \to 0$,

$$\begin{aligned}
&\sum_{j=1}^{m_*}\int_0^t A(s_{*j})\dot{x}^{\varepsilon,N}(s)I_{\{\alpha^\varepsilon(s)=s_{*j}\}}ds \to 0 \ \text{ in probability,} \\
&\sum_{j=1}^{m_*}\int_0^t B(s_{*j})x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{*j}\}}ds \to 0 \text{ in probability,} \\
&\sum_{j=1}^{m_*}\int_0^t f(s,s_{*j})q^N(x^{\varepsilon,N}(s))I_{\{\alpha^\varepsilon(s)=s_{*j}\}}ds \to 0 \text{ in probability.}
\end{aligned}\tag{3.24}$$

Thus the transient states are asymptotically unimportant.

Working with (3.22), by virtue of the weak convergence of $x^{\varepsilon,N}(\cdot)$ to $x^N(\cdot)$ and the Skorohod representation, we may assume that $x^{\varepsilon,N}(\cdot)$ converges to $x^N(\cdot)$ with probability one, and the convergence is uniform on any bounded time interval. In addition, the weak convergence of $\overline{\alpha}^\varepsilon(\cdot)$ (see Theorem 3.1 (iii)) and the Skorohod representation imply that $I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}} = I_{\{\overline{\alpha}^\varepsilon(s)=i\}}$ converges to $I_{\{\overline{\alpha}(s)=i\}}$. In addition, since $\alpha^\varepsilon(\cdot)$ has piecewise constant sample paths, $(d/dt)A(\alpha^\varepsilon(t)) = 0$ for almost all $t \in [0,T]$. Therefore, as $\varepsilon \to 0$,

$$\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t A(s_{ij})\dot{x}^{\varepsilon,N}(s)\nu_j^i(s)I_{\{\overline{\alpha}(s)=i\}}ds \\
&=\sum_{i=1}^{l}\sum_{j=1}^{m_i}A(s_{ij})x^{\varepsilon,N}(t)\nu_j^i(t)I_{\{\alpha^\varepsilon(t)\in\mathcal{M}_i\}}-\sum_{i=1}^{l}\sum_{j=1}^{m_i}A(s_{ij})x_0\nu_j^i(0)I_{\{\overline{\alpha}^\varepsilon(0)\in\mathcal{M}_i\}} \\
&\to\sum_{i=1}^{l}\sum_{j=1}^{m_i}A(s_{ij})x^N(t)\nu_j^i(t)I_{\{\overline{\alpha}(t)=i\}}-\sum_{i=1}^{l}\sum_{j=1}^{m_i}A(s_{ij})x_0\nu_j^i(0)I_{\{\overline{\alpha}(0)=i\}}.
\end{aligned}\tag{3.25}$$

Similarly, as $\varepsilon \to 0$,

$$\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t B(s_{ij})x^{\varepsilon,N}(s)q^N(x^{\varepsilon,N}(s))\nu_j^i(s)I_{\{\overline{\alpha}^\varepsilon(s)=i\}}ds \\
&\quad\to\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t B(s_{ij})x^N(s)q^N(x^N(s))\nu_j^i(s)I_{\{\overline{\alpha}(s)=i\}}ds,
\end{aligned}\tag{3.26}$$

and

$$\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t f(s,s_{ij})q^N(x^{\varepsilon,N}(s))\nu_j^i(s)I_{\{\overline{\alpha}^{\varepsilon}(s)=i\}}ds$$
$$\to \sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_0^t f(s,s_{ij})q^N(x^N(s))\nu_j^i(s)I_{\{\overline{\alpha}(s)=i\}}ds. \tag{3.27}$$

Combining (3.25)—(3.27), we obtain the equation satisfied by the limit $x^N(\cdot)$, namely,

$$\overline{A}(\overline{\alpha}(s))x^N(s) = A(\overline{\alpha}(0))x_0 - \int_0^t \overline{B}(\overline{\alpha}(s))x^N(s)q^N(x^N(s))ds$$
$$+\int_0^t \overline{f}(s,\overline{\alpha}(s))q^N(x^N(s))ds. \tag{3.28}$$

The desired result thus follows.

Next, we let $N\to\infty$ and conclude that $x^N(\cdot)\to x(\cdot)$ and hence the weak convergence of $x^{\varepsilon}(\cdot)$ to $x(\cdot)$. The proof of the following theorem uses the measures induced by $x^N(\cdot)$ and $x(\cdot)$ and the tractability. The details are similar to that of ref. [21] and are omitted.

Theorem 3.4. Suppose that (A1) and (A2) are satisfied. Then the untruncated process $x^{\varepsilon}(\cdot)$ given in (2.4) is also tight in $D^r[0,T]$, and $x^{\varepsilon}(\cdot)$ converges weakly to $x(\cdot)$, a solution of the differential system

$$\overline{A}(\overline{\alpha}(t))\dot{x} + \overline{B}(\overline{\alpha}(t))x = \overline{f}(t,\overline{\alpha}(t)), \quad x(0) = x_0, \tag{3.29}$$

where $\overline{A}$, $\overline{B}$, and $\overline{f}$ are defined in (3.18).

3.4 Limit result under weak irreducibility

In the previous section, we have derived limit results for singular systems under Markovian perturbation with generator $Q^{\varepsilon}(t)$ given by (2.2) and (2.3). We now consider a special case, namely, $\widetilde{Q}(t)$ in (2.3) has only one weakly irreducible block. To fix the notation, suppose that $Q^{\varepsilon}(t) = Q(t)/\varepsilon + Q_0(t)$, where $Q(t)$ is weakly irreducible. Now, all the states are recurrent. Moreover, it is easily seen that for sufficiently small $\varepsilon > 0$, $Q^{\varepsilon}(t)$ is also weakly irreducible. We can now apply the results obtained in Theorem 3.2 and Theorem 3.3. Let the state space of $\alpha^{\varepsilon}(\cdot)$ be $\mathcal{M} = \{1,\ldots,m\}$ and the quasi-stationary distribution corresponding to $Q(t)$ be $\nu(t) = (\nu_1,\ldots,\nu_m)\in\mathbb{R}^{1\times m}$. We obtain the following theorem.

Theorem 3.5. Suppose that the conditions of Theorem 3.3 are satisfied with $\widetilde{Q}(t)$ and $\widehat{Q}(t)$ replaced by $Q(t)$ and $Q_0(t)$ and with the Lipschitz continuity in (A1) being deleted. Then $x^{\varepsilon}(\cdot)$ is tight in $D^r[0,T]$ and $x^{\varepsilon}(\cdot)$ converges weakly to $x(\cdot)$, which is a solution of

$$\overline{A}(s)\dot{x}(s) + \overline{B}(s)x(s) = \overline{f}(s), \quad x(0) = x_0, \tag{3.30}$$

where

$$\overline{A}(t) = \sum_{\iota=1}^{m}\nu_{\iota}(t)A(\iota), \quad \overline{B}(t) = \sum_{\iota=1}^{m}\nu_{\iota}(t)B(\iota), \quad \overline{f}(t) = \sum_{\iota=1}^{m}\nu_{\iota}(t)f(t,\iota). \tag{3.31}$$

Remark 3.2. Note that in this case, the limit system is a deterministic one. Moreover, if $Q(t) = Q$, a constant matrix, explicit form of solution can be given. The solution of (3.30) is given by

$$\begin{aligned} x(t) &= \exp(-\Phi_{\overline{A}}^{D}\Phi_{\overline{B}}t)\Phi_{\overline{A}}\Phi_{\overline{A}}^{D}x_0 + \int_0^t \exp(-\Phi_{\overline{A}}^{D}\Phi_{\overline{B}}(t-s))\Phi_{\overline{A}}^{D}\Phi_{\overline{f}}(s)ds + w(t), \\ w(t) &= (I - \Phi_{\overline{A}}\Phi_{\overline{A}}^{D})\sum_{i=0}^{\kappa-1}(-1)^i(\Phi_{\overline{A}}\Phi_{\overline{B}}^{D})^i\Phi_{\overline{B}}^{D}\Phi_{\overline{f}}^{(i)}(t). \end{aligned} \tag{3.32}$$

4 Hybrid/switching Leontief models

The well-known Leontief model is a dynamic system of a multisector economy (see, for example, ref. [23]). The traditional setup can be stated as follows. Suppose that there are r sectors. Let $x_i(t)$ be the output of sector i at time t and $D_i(t)$ the demand for the product of sector i at time t. Denote $x(t) = (x_1(t), \cdots, x_r(t))' \in \mathbb{R}^r$ and $D(t) = (D_1(t), \cdots, D_r(t))' \in \mathbb{R}^r$. Let a_{ij} be the amount of commodity i that sector j needs to produce and denote $A = (a_{ij})$. If a given sector does not produce a commodity, then A may have a zero row. Thus, A is often a singular matrix. Denote by b_{ij} the proportion of commodity j that is transferred to commodity i. The matrix $B = (b_{ij})$ is termed a Leontief input-output matrix. The Leontief dynamic model is given by

$$A\dot{x} = (I - B)x + D(t), \tag{4.1}$$

with A being a singular matrix.

In the classical Leontief model, the coefficients are fixed. Nevertheless, in reality, more often than not, they are changing with respect to time depending on the trend of the economy. Not only are A, B, and D time varying, but also they are subject to discrete shifts in regime-episodes across which the behavior of the corresponding dynamic systems is markedly different. As a result, a promising alternative than the traditional model is to allow for the possibility of sudden, discrete changes in the values of the parameters resulting in a "hybrid" or "switching model" governed by a Markov chain. In what follows, we propose a hybrid Leontief model with switching regime. The premise of our model is that many of the important movements in economy arise from discrete events. A nation's economy sometimes appear quite calm and at other instances are rather volatile. To describe how this volatility changes over time is by far important. It is easily seen that monetary, fiscal, or income policies, often change in a discontinuous fashion with jump sample paths, which is often referred to as shocks in economics. Economists cannot observe these shifts directly, so these discrete events are governed by hidden random processes. Since the late 1980s and early 1990s, increasing interests on using Markov-based models in economics have been shown. Although most of these efforts are devoted to time series analysis (see refs. [24—27] and the references therein), it is conceivable that the use of Markov-based models will play a more prominent role in the near future. Similar to the consideration of stock market, a continuous-time Markov chain can be used to model the trend of the economy. To illustrate, consider a simple example in which the economy has two possible "states," fast growth phase (denoted by 2) and slow growth phase (denoted by 1). At any given time t, the economy will be in either the fast growth state or the slow growth state governed by the outcome of a Markov chain. Similarly, consider another example with the use of unemployment data, the economy may be said to be in state 1 if the unemployment rate is rising and in state 2 if the unemployment rate is falling. Corresponding to the two states, either for the economic growth or for the unemployment rates, the regimes or configurations of system (4.1) will naturally differ resulting in different coefficients of the linear equations for different regimes. That is, the matrices A, B, and D vary with respect to different regimes governed by a Markov chain $\alpha(\cdot)$. More generally, the economy can have a number of states instead of just two states (e.g. different levels of increases and decreases). This then leads to a hybrid/switching model modulated by a Markov chain with finite state space.

Our next concern is the reduction of complexity. In a multi-sector economy, the state space of $\alpha(\cdot)$ is likely to be very large due to the rapid growth in science and technology. The large number of states of the underlying chain gives a detailed representation of the position of the economy. Nevertheless, the large-scale nature of the system makes the design and control of such systems very difficult tasks. To reduce the complexity of the system, we observe that not all states in the system change at the same speed. Some of them vary rapidly and others change slowly. The inherent fast and slow time scales give us the possibility of grouping the states in the systems in accordance with their transition rates. We introduce a small parameter $\varepsilon > 0$, and let $\alpha(t) = \alpha^\varepsilon(t)$ with the generator given by (2.2) and $\widetilde{Q}(t)$ specified by (2.3). Suppose that we want to control a hybrid system in which the state space of $\mathcal{M}$ consists of m elements, where m is a large number. Using appropriate asymptotic analysis, instead, we can consider a "reduced" system whose state space consists of only l elements. If $l \ll m$, the complexity of the task is dramatically reduced.

In the asymptotic analysis, to obtain a rigorous result, it is necessary to consider the limit as $\varepsilon \to 0$. In the actual applications, ε could be a constant; it need not go to 0. The asymptotic result, however, renders guidance on the control, optimization, and design of the actual system. For further discussions on the interpretation of the time scale separation, see ref. [6].

Using the hybrid/switching Markov chain $\alpha^\varepsilon(t)$, we consider the following hybrid Leontief model

$$\begin{aligned} &A(\alpha^\varepsilon(t))\dot{x}^\varepsilon(t) = (I - B(\alpha^\varepsilon(t)))x^\varepsilon(t) + D(t, \alpha^\varepsilon(t)),\\ &x^\varepsilon(0) = x_0. \end{aligned} \tag{4.2}$$

Assume that conditions (A1) and (A2) hold for the continuous-time model (4.2). Then the weak convergence results discussed in the previous sections hold. We have the following limit result.

Theorem 4.1. Under the conditions of Theorem 3.2, $\{x^\varepsilon(\cdot)\}$ given by (4.2) is tight in $D^r[0,T]$, and $x^\varepsilon(\cdot)$ converges weakly to $x(\cdot)$, which is a solution of the following singular differential equation

$$\begin{aligned} &\overline{A}(\overline{\alpha}(t))\dot{x}(t) = (I - \overline{B}(\overline{\alpha}(t)))x(t) + \overline{D}(t, \overline{\alpha}(t)),\\ &x(0) = x_0, \end{aligned} \tag{4.3}$$

where $\overline{A}(\overline{\alpha}(t))$, $\overline{B}(\overline{\alpha}(t))$, and $\overline{D}(t, \overline{\alpha}(t))$ are as defined in (3.18) with $\overline{f}(t, \overline{\alpha}(t))$ replaced by $\overline{D}(t, \overline{\alpha}(t))$.

Under the conditions of Theorem 3.5, the weak limit is given by

$$\begin{aligned} &\overline{A}(t)\dot{x}(t) = (I - \overline{B}(t))x(t) + \overline{D}(t), \\ &x(0) = x_0, \end{aligned} \tag{4.4}$$

where $\overline{A}(t)$, $\overline{B}(t)$, and $\overline{D}(t)$ are as defined in (3.31) with $\overline{f}(t)$ replaced by $\overline{D}(t)$.

Remark 4.1. Let us give some economic implication of the results obtained. First, the original Leontief model (4.1) is purely deterministic. Thus it may not respond to any stochastic changes and random fluctuations. It appears that a major factor that dominates the states of economy is the trends of the general economy. By using the model (4.2), we consider a refinement of the well-known Leontief model.

The limit systems both under weak irreducibility (see (4.4)) and multi-block $\widetilde{Q}(t)$ (see (4.3)) present a reduction of complexity. The keyword is averaging. The weak irreducible case corresponds to the underlying system that undergoes rapid variations; the limit system is a deterministic one. It indicates we can ignore the detailed fluctuations in the actual system, and examine only its average. In the case of a multi-block $\widetilde{Q}(t)$, we are interested in the situation that card$(\mathcal{M})$ is fairly large, which follows a realistic consideration of complex economic systems. If card$(\overline{\mathcal{M}})$ $\ll$card$(\mathcal{M})$, the complexity of the underlying system is much reduced by means of the averaging approach.

5 Further remarks

This paper has been devoted to singular systems of differential equations. We have focused on two main points. The first one is to develop hybrid models of singular differential systems driven by random disturbances. The second one is the reduction of complexity of large-scale singular systems. Here the machinery we are using is the hierarchical approach via singular perturbation methods. Future work can be directed to the control and optimization problems of large-scale singular systems of differential equations and to nonlinear singular systems.

Acknowledgements Yin Gang was supported by the National Science Foundation of US (Grant No. DMS-9877090), and Zhang Jifeng was supported by the National Natural Science Foundation of China (Grant No. 69725006).

References

1. Campbell, S. L., Singular Systems of Differential Equations, I & II, San Francisco: Pitman, 1980 & 1982.
2. Cheng, Z. L., Hong, L., Zhang, J. F., The optimal regulation of generalized state-space systems with quadratic cost, Automatica, 1988, 24: 707—710.
3. Dai, L., Singular control systems, Lecture Notes in Control and Information Sci., vol. 118, Berlin: Springer-Verlag, 1989.
4. Huang, J., Zhang, J. F., Impulse-free output regulation of singular nonlinear systems, Internt. J. Control, 1998, 71: 789—806.
5. Sethi, S. P., Zhang, Q., Hierarchical Decision Making in Stochastic Manufacturing Systems, Boston: Birkhäuser, 1994.
6. Yin, G., Zhang, Q., Continuous-time Markov Chains and Applications: A Singular Perturbation Approach,New York: Springer-Verlag, 1998.
7. Zhang, Q., Yin, G., On nearly optimal controls of hybrid LQG problems, IEEE Trans. Automat. Control, 1999, 44: 2271—2282.
8. Simon, H. A., Ando, A., Aggregation of variables in dynamic systems, Econometrica, 1961, 29: 111—138.

9. Pervozvanskii, A. A., Gaitsgori, V. G., Theory of Suboptimal Decisions: Decomposition and Aggregation, Dordrecht: Kluwer, 1988.
10. Phillips, R. G., Kokotovic, P. V., A singular perturbation approach to modelling and control of Markov chains, IEEE Trans. Automat. Control, 1981, 26: 1087—1094.
11. Il'in, A., Khasminskii, R., Yin, G., Asymptotic expansions of solutions of integro-differential equations for transition densities of singularly perturbed switching diffusions: Rapid switchings, J. Math. Anal. Appl., 1999, 238: 516—539.
12. Khasminskii, R. Z., Yin, G., Zhang, Q., Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains, SIAM J. Appl. Math., 1996, 56: 277—293.
13. Khasminskii, R. Z., Yin, G., Zhang, Q., Constructing asymptotic series for probability distribution of Markov chains with weak and strong interactions, Quart. Appl. Math., 1997, 55: 177—200.
14. Yin, G., Zhang, Q., Control of dynamic systems under the influence of singularly perturbed Markov chains, J. Math. Anal. Appl., 1997, 216: 343—367.
15. Yin, G., Zhang, Q., Badowski, G., Asymptotic properties of a singularly perturbed Markov chain with inclusion of transient states, Ann. Appl. Probab., 2000, 10: 549—572.
16. Yin, G., Zhang, Q., Badowski, G., Singularly perturbed Markov chains: Convergence and aggregation, J. Mult. Anal., 2000, 72: 208—229.
17. Davis, M. H. A., Markov Models and Optimization, London: Chapman & Hall, 1993.
18. Iosifescu, M., Finite Markov Processes and Their Applications, Chichester: Wiley, 1980.
19. Billingsley, P., Convergence of Probability Measures, 2nd ed., New York: Wiley, 1999.
20. Ethier, S. N., Kurtz, T. G., Markov Processes: Characterization and Convergence, New York: Wiley, 1986.
21. Kushner, H. J., Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory, Cambridge, MA: MIT Press, 1984.
22. Kushner, H. J., Yin, G., Stochastic Approximation Algorithms and Applications, New York: Springer-Verlag,1997.
23. Kendrick, D., On the Leontief dynamic inverse, Quat. J. Economics, 1972, 86: 693—696.
24. Brunner, A. D., Testing for structural breaks in U.S. post-war inflation data, Mimeo, Board of Governors of Federal Reserve System, Washinton D.C., 1991.
25. Cai, J., A Markov model of unconditional variance in ARCH, J. Business Economic Statist, 1994, 12: 309—316.
26. Hamilton, J.D., Susmel, R., Autoregressive conditional heteroskedasticity and changes in regime, J. Econometrics, 1994, 64: 307—333.
27. Hansen, B. E., The likelihood ratio test under nonstandard conditions: Testing Markov trend model of GNP, J. Appl. Economics, 1992, 7: S61—S82.